

# Identifying Event-Specific Opinion Leaders by Local Weighted LeaderRank

Wanxia Yang<sup>1,\*</sup>, Sadaqatur Rehman<sup>2</sup> and Wenhui Que<sup>3</sup>

<sup>1</sup>Mechanical and Electrical Engineering college, Gansu Agricultural University, Lanzhou, 730070, China

<sup>2</sup>Department of Computer Science, Namal Institute, Mianwali, 42250, Pakistan

<sup>3</sup>Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing, China

\*Corresponding Author: Wanxia Yang. Email: yanwanxia@163.com

**Abstract:** Identifying event-specific opinion leaders is essential for understanding event developments and influencing public opinion. News articles are informative and formal in expression, and include valuable information on specific events. In this paper, we propose an improved variant of LeaderRank, called local weighted LeaderRank, to measure the event-specific influence of person nodes in a weighted and undirected person cooccurrence network constructed using news articles related to a specific event. Our proposed method measures the influence of person nodes by considering both the cooccurrence strength between persons, and additional local link weight information for each local person node. To evaluate the performance of our method, we use the weighted susceptible infected (WSI) model to simulate the influence-spreading process in real-person cooccurrence networks. The experiment results obtained after measuring the rank correlations between the rank list generated by the simulation results and those generated by the influence measures show that our method identifies event-specific opinion leaders effectively and performs better than other state-of-the-art influence measures, such as weighted K-shell decomposition and the weighted local centrality.

**Keywords:** Opinion leaders; specific event; local weighted LeaderRank; cooccurrence network

## 1 Introduction

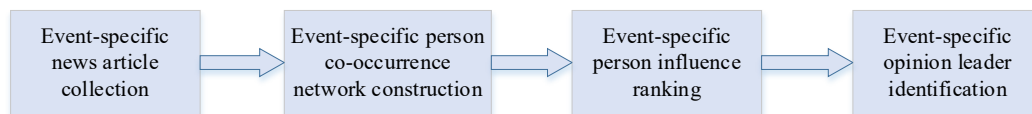
The concept of opinion leaders, which refers to an influential and active minority in the process of information transfer and personal interaction, was first proposed by Lazarsfeld [1]. Since the famous two-step flow propagation pattern was first generated, the concept of opinion leaders has been studied in various fields. Dong et al. [2] investigated its application in innovation propagation, and reference [3] analyzed the impact of opinion leaders in marketing. As the Internet has grown, research on the identification of those influences, based on Internet media information, has evolved. The authors of [4] studied them in the blogosphere, while the authors of [5] identified theirs in the Turkish online discussion space. In Aleahmad et al. [6] discussed how to identify opinion leaders on medical topics using news articles. In recent years, Twitter, the world's most popular microblogging service, has gradually become a research hotspot among scholars. The influences of users on Twitter was studied in [7–9]. One primary approach applied in these studies is social network analysis. In this paper, we explore how social network analysis can be used to identify event-specific opinion leaders based on their cooccurrences in news articles. In this context, cooccurrence refers to common information related to the event specifics, in multiple news articles.

Compared to the general opinion leaders, event-specific opinion leaders are more relevant to a particular event. By identifying the latter, we can understand how a specific event develops and the roles played by those related to it; such knowledge can help influence both public opinion as well as the



development of the event itself. In contrast to the contents on Twitter and other microblogs, news articles feature abundant information, diverse sources, and standardized expressions. By analyzing the content of news articles related to a specific matter, we can obtain valuable information about it, the most important of which is the cooccurrence information about individuals.

In this paper, we propose an approach to identify event-specific opinion leaders using news articles. The framework of the proposed method is shown in Fig. 1. First, we crawl news articles published at a particular time on the Internet. Then, out of those, news articles on a specific event are filtered by relevant keywords and names are extracted with the assistance of named entity recognition tools. Subsequently, we construct an event-specific weighted person cooccurrence network using the improved name recognition results. Furthermore, we propose an enhanced variant of LeaderRank, which we call local weighted LeaderRank (LWLR), which considers the link weight to measure the influences of various nodes in a person's cooccurrence network. Moreover, LWLR makes use of more local information when computing node influence.



**Figure 1:** Framework of the proposed approach

To evaluate the effectiveness of our proposed LWLR in identifying opinion leaders, we use the weighted susceptible infected (WSI) model [10]. The experimental results verify that our proposed method identifies event-specific opinion leaders effectively, and that it performs better than other state-of-the-art methods, such as the weighted k-shell decomposition algorithm and the weighted local centrality model.

The remainder of this paper is organized as follows. The related works are introduced in Section 2. In Section 3, we present in details the construction process of the event-specific person cooccurrence network, which consists of two parts: One for Chinese texts and one for English texts. The LWLR is proposed in Section 4. Section 5 presents the experiments comparing LWLR to other related methods, and reports the results. We provide a brief discussion of the proposed work in Section 6. Finally, we conclude our paper in Section 7.

## 2 Related Work

In recent years, popular metrics such as betweenness centrality and closeness centrality have yielded excellent results; however, they have proven unsuitable for the use on large-scale networks due to their computational complexity. Recently, Wang et al. [11], found the most efficient method for distinguishing the influence ability of nodes with the same k-shell value. Since then, some measures used to identify highly influential nodes have been introduced [12–13], and some weighted decomposition methods for complex networks have been proposed to further improve ranking performance [14–15]. In contrast to the accepted belief that highly connected or core-located nodes are essential spreaders, the social circle breadth metric was proposed to measure the propagating influence of a node by qualitatively combining both its local and global structural properties [16]. However, ranking algorithms that apply to global network information are time-consuming; thus, they are rarely applied. To that effect, ranking algorithms based on local network information have been designed to rank nodes effectively and efficiently. For example, a semi-local centrality measure that takes into account both the nearest and the next-nearest neighbors of a node was designed in [17], and its effectiveness and efficiency for ranking the spreading ability of nodes has been well demonstrated. Semi-local centrality achieves a good tradeoff between low-correlation degree centrality and other time-consuming steps. In directed networks, several ranking methods based on the iterative processes have been proposed, such as PageRank [18], HITS (HITS (Hyperlink-Induced Topic Search)) [19–20] and LeaderRank [21]. An improved variant of LeaderRank called weighted LeaderRank [22–23] was proposed, where it assigns degree-dependent weights to links related to the ground node.

The existing methods, including betweenness centrality and closeness centrality, are used in social network analysis to identify opinion leaders. Compared with these approaches, our proposed approach implements name recognition and normalization for both English and Chinese news articles. It also uses a more complicated normalization process that causes the resulting cooccurrence network to be closer to reality. Moreover, our proposed method calculates a relevance factor for each news article related to a specific event and measures the node influence in the network using LWLR.

Among the methods for measuring the node influence in a complex network, degree centrality, betweenness centrality, closeness centrality, local centrality, k-shell decomposition, and weighted LeaderRank represent the conventional approaches. In this paper, considering that our experimental data consist of news articles, we add another method called Mentions, which reflects the number of times a particular person is mentioned in news articles related to a specific event over a period of time. Of the above seven measures, the first six are used to analyze node influence in previously unweighted networks. Because the networks addressed in this paper are weighted, we extend the above methods to the weighted variants. Each variant is described in detail below.

Considering a weighted and undirected network  $G = (V, E)$ ,  $N = |V|$  denotes the number of vertices, and  $M = |E|$  denotes the number of edges.  $G$  can be illustrated as an adjacency matrix  $A = \{w_{ij}\} \in R^{n,n}$ , where  $w_{ij} > 0$  if node  $i$  is connected to node  $j$  and  $w_{ij} = 0$  otherwise. We use  $\Gamma_h(i)$  to express the set of neighbors within  $h$ -hops of node  $i$ .

The weighted LeaderRank (WLR) is described in detail below, because it adds weighted variants.

The WLR measure proposed in [21] introduces a weighted mechanism to improve the performance of LeaderRank. In a given network  $G$ , a ground node is added that is connected to each node through a bidirectional link. Thus, the network becomes tightly connected, consisting of  $N+1$  nodes and  $M+2N$  links, and can be expressed by the  $(N+1)$ -dimensional weighted adjacency matrix  $W$ . For any normal node  $i$  and ground node  $g$ ,  $w_{gi} = k_i^a$  and  $w_{ig} = 1$ , where  $a$  is a free parameter. In this paper, we discuss only the case in which  $a = 1$ . Once the weight of every link has been calculated, the score of the path from node  $j$  to node  $i$  is in proportional to the weight  $w_{ji}$ , following a biased random walk

$$s_i(t+1) = \sum_{j=1}^{N+1} \frac{w_{ji}}{\sum_{l=1}^{N+1} w_{jl}} s_j(t) .$$

Similar to LeaderRank, the final scores in the steady state are used to quantify the influences of nodes.

Among the seven methods mentioned above, the method most similar to ours is WLR, which can be applied only to previously unweighted and directed networks. In this paper, we extend WLR to an improved variant, i.e., LWLR. Compared to WLR, our method considers the structure and link weight information of the network. It makes use of more local information when calculating the link weights from the ground node to normal nodes.

### 3 Event-Specific Person Cooccurrence Network Construction

In this paper, we implement a method of event-specific person cooccurrence network construction using both Chinese and English news articles. The network construction process based on Chinese news articles is introduced in detail on a dataset related to the Goujing event (Gou Jing took part in college entrance examination in 1997, and her Gao Kao score reached admission standard of special secondary school in Jining city. However, she did not fill in the application form. Later, her personal identity and college entrance examination scores were falsely used by Qiu Xiaohui), while the process for English-language news articles is introduced in detail on a dataset related to the Policemen neck trampling event in American.

#### 3.1 Cooccurrence Network Construction Using Chinese News Articles

The process of constructing an event-specific person cooccurrence network using Chinese news articles consists of five main steps.

The first step involves crawling Chinese news articles related to the event. We crawled a large number of news articles from Chinese news websites and then filtered these articles using keywords related to a specific event to obtain a set of news articles relative to that event. For instance, in the Gou Jing event, we crawled the news articles published between June 22, 2020, and August 22, 2020, on various well-known Chinese-language news websites and obtained 1,599 news articles related to the event by filtering the collected news articles using the keywords “(Gou Jing)”.

The second step consists of calculating the event-specific relevance factor for each news article. Because the relevance factors of different news articles to the event are not identical, we must calculate the event-specific relevance factor of each news article to determine the strength of the cooccurrence relationships among different news articles. First, we define an event description vector  $V = [c_1, c_2, \dots, c_N]$ , where  $c_i$  represents an event description word and  $N$  is the total number of event description words. Then, we split the texts of all of the event-related news articles into sequences of words using the ICTCLAS tools. After removing the stop words from the word segmentation results, we count the number of articles in which each word appears. The top- $N$  words are regarded as the event description words. In the example considered here, we set  $N = 2000$  for the Gou Jing event. Moreover, the number of articles in which event description word  $c_i$  appears is taken as the weight of that word relative to the event, denoted by  $W(c_i)$ . Thus, the event-specific relevance factor for a certain news article can be calculated as follows:

$$Rel(u) = \sum_{i=1}^{V_N} Fre(c_i) \cdot W(c_i) \quad (1)$$

where  $Fre(c_i)$  is the occurrence frequency of event description word  $c_i$  in news article  $u$  and  $V_N$  is the total number of event description words. The largest value of the event-specific relevance factor for any article is taken to be equal to 1; thus, the relevance factor of article  $u$  is normalized as follows:

$$Rel_N(u) = \frac{Rel(u)}{Rel_{Max}(u)} \quad (2)$$

where  $Rel_{Max}(u)$  denotes the largest value of  $Rel(u)$ .

The third step consists of recognizing the names that appear in news articles and improving the name recognition results. We use the ICTCLAS tools developed by the Chinese Academy of Sciences for Chinese name recognition in Chinese articles; however, its the recognition performance is not perfect because of undesirable behavior such as word segmentation errors with names, problems with name coreference resolution, and name spelling mistakes. To improve the name recognition results for subsequent experiments, we augment the name recognition outcomes generated by the ICTCLAS tools via the following processes.

*Manual correction.* The primary purpose when manually correcting the name recognition results generated by the ICTCLAS tools is to resolve some of the general problems encountered by name recognition algorithms, such as word segmentation mistakes in names, misspelled names, and noun recognition. For example, regarding name recognition mistakes in news articles related to the Gou Jing event, the names “” and “” were segmented as “” and “”. Name spelling mistakes may also occur; for example, the name “” was written as “”. Additionally, nouns such as locations and companies that start with family names may be mischaracterized as names; for example, “”, “”, “”, “”, and “” were recognized as names.

*Shallow name coreference resolution in a single news article.* In this paper, we mainly consider two cases of shallow name coreference resolution in a single news article. In the first case, both the full name and the first name only of a particular person appear in a single news article. We merge these two names into one name based on their containment relationship. In the second case, both the full name of a particular person and an expression consisting of the family name of this person and the word “” or “”

appear in a single news article. In this case, we merge the expression consisting of the family name and the word “'” or “'” into the full name that starts with the same family name. When more than one name appearing in a single news article has the same family name, the expression consisting of family name of that person and word“'” or “'” will be merged into the full name with the same family name, which is closest to the location of the first occurrence of the expression in the article. For instance, “'” and “'”, which appear in a single news article related to the Gou Jing event, are merged into one name represented as “'” after shallow name coreference resolution.

*Shallow name coreference resolution across news articles.* Different names may be recognized in various news articles that actually refer to the same person. In this paper, we consider the cases in which the full name of a particular person appears in one article but only the family name of that person appears in another article. In such cases, we merge the two names into one name based on their containment relationship.

The fourth step consists of calculating the link weights of the cooccurrence network. After the preceding processing steps, we have acquired a relevance factor for each news article and the names of persons that appear in the articles. The cooccurrence network can be described as  $G = (V, E)$ , where  $V$  denotes the set of all the names and  $E$  denotes the set of cooccurrence links between those names. The weight of the cooccurrence link between any two names  $i$  and  $j$  can be calculated as follows:

$$w_{ij} = \sum_{u \in \Gamma_{ij}} Rel(u) \quad (3)$$

where  $\Gamma_{ij}$  denotes the set of news articles in which name  $i$  and name  $j$  both appear, and  $Rel(u)$  denotes the event-specific relevance factor of news article  $u$ . For the Gou Jing event, the constructed cooccurrence network contains 2,180 name nodes. In general, the constructed network will be not fully connected. Therefore, to evaluate the effectiveness of methods of identifying opinion leaders using the WSI model, we need to calculate the maximal connected subgraph of the cooccurrence network.

Accordingly, the fifth step consists of calculating the maximal connected subgraph of the person cooccurrence network. The obtained maximally connected subgraph is the person cooccurrence network that must be constructed to finally identify event-specific opinion leaders. For the Gou Jing event, there are 2,180 name nodes and 2,765 links between name nodes, and the average degree in the obtained maximally connected subgraph is 19.75. Based on the maximally connected subgraph, we can rank the influences of the name nodes in a network using the node influence ranking measures considered in this paper.

### 3.2 Cooccurrence Network Construction Using English News Articles

The process of constructing event-specific person cooccurrence network using English news articles primarily consists of a set of five steps similar to those applied to the Chinese news articles. These steps are described in detail below.

The first step involves crawling datasets to identify English news articles related to the event. We crawled a large number of news articles published between May 28, 2020, and June 28, 2020, on certain well-known English-language news websites and filtered those news articles with the keywords “Policeman” and “neck trampling” to obtain 1,083 news articles related to the Policeman neck trampling event.

The second step is to calculate the event-specific relevance factor for each news article. First, we convert the uppercase letters in each news article related to the event into lowercase letters and then split the texts into sequences of words, with characters that are not alphanumeric, such as spaces, commas and periods, serving as separators. After removing the stop words in each news article, we stem all the words in each news article using the Porter stemmer. Then, we calculate the event-specific relevance factor for a particular English news article using a calculation similar to that for the Chinese news articles, as shown in Eq. (1) and Eq. (2).

The third step consists of recognizing names in news articles and improving the name recognition results. We apply the Stanford Named Entity Recognize, which uses a 3-class model, for English name recognition in English articles; however, its recognition performance is not perfect because of various

undesirable behaviors in the recognition results, such as name segmentation mistakes, problems with name coreference resolution, and name spelling mistakes. To improve the name recognition outcomes for the subsequent experiments, we apply the following processing steps.

*Manual correction.* The primary purpose of manually correcting the name recognition results is to solve problems such as mistakes in the name segmentation, spelling, and the noun recognition, such as the mischaracterization of locations and magazines as names. For instance, in the case of name recognition in news articles related to the Policeman neck trampling event, the phrase “mayor of Minneapolis Jacob Frey” was not accurately split to produce the name “Jacob Frey”. Name spelling mistakes also occur; for example, the name “George Floyd” was misspelled as “George Freud”. Additionally, nouns such as locations and magazines can be recognized as names; for example, “Colorado” and “Denver” were recognized as names.

*Shallow name coreference resolution in a single news article.* Different ways of referencing the same person, such as the person’s given name, family name, or full name, may appear within a single English news article. We merge these different forms of address for a particular person into a single form based on their containment relationship. For instance, “George Floyd” and “Floyd”, which appeared in a single news article related to the Policeman neck trampling event, were merged into one name represented as “George Floyd” after shallow name coreference resolution.

*Shallow name coreference resolution across news articles.* Different names recognized in different news articles may refer to the same person. In this study, we considered cases in which the full name of a particular person appears in one article while only the given name or the family name of that person appears in another article by merging these two names into one name based on their containment relationship. For instance, “Tim Walz” and “Walz”, which appeared in news articles related to the Policeman neck trampling event, were merged into one name, “Tim Walz”, after shallow name coreference resolution.

The fourth step consists of calculating the link weights of the cooccurrence network, while the fifth step is to extract the maximal connected subgraph of the person cooccurrence network. These two steps are identical to those used for the Chinese news articles; hence their descriptions are not repeated here. For the Policeman neck trampling event, there are 1,876 name nodes and 8,743 links, and the average degree of the obtained maximally connected subgraph is 14.72. Based on the maximally connected subgraph, we can rank the influence of name nodes in the network using the node influence ranking measures considered in this paper.

#### 4 Local Weighted LeaderRank

According to the definition of WLR in Section 2, the weight of a link from a normal node to the ground node can be calculated as  $w_{ig} = \frac{1}{D_i} \cdot \sum_{j \in \Gamma_i} w_{ij}$ , where  $w_{ij}$  denotes the weight of the cooccurrence link between normal node  $i$  and normal node  $j$ ,  $\Gamma_i$  denotes all normal neighboring nodes of node  $i$ , and  $D_i = |\Gamma_i|$  denotes the total number of normal neighboring nodes of node  $i$ . The weights of the links from the ground node to the normal nodes can be calculated as  $w_{gi} = \alpha \cdot WD_i + (1 - \alpha) \sum_{j \in \Gamma_i} WD_j$ , where  $WD_i = \sum_{j \in \Gamma_i} w_{ij}$ ,  $\Gamma_i$  denotes all normal neighboring nodes of node  $i$ , and  $\alpha$  is a tuning parameter with a range of [0,1]. Once the weight of every link has been determined, the score from node  $j$  to node  $i$  is calculated as follows:

$$s_i(t+1) = \sum_{j=1}^{N+1} \frac{w_{ji}}{\sum_{l=1}^{N+1} w_{jl}} s_j(t) \quad (4)$$

where  $s_i(t)$  denotes the score of node  $i$  in time step  $t$ . The process starts with an initialization procedure in which the scores of all the nodes other than the ground node are set to 1 and the score of the ground

node score is set to 0, the process soon converges to a unique steady state expressed as  $s_i(t_c)$ , where  $t_c$  is the convergence time. In the steady state, we achieve an equal distribution of the ground node score relative to all other node scores to retain scores for the nodes of interest. Thus, we determine that the final score of a node in the person cooccurrence network is the influence score  $s$ , calculated as follows:

$$s_i = s_i(t_c) + \frac{w_{gi}}{\sum_{i=1}^N w_{gi}} s_g(t_c) \quad (5)$$

where  $s_g(t_c)$  is the score of the ground node in the steady state.

Based on the above properties, compared with WLR and the other methods introduced in Section 2, LWLR presents several advantages in ranking performance: (1) LWLR considers the link weight information throughout the entire cooccurrence network. (2) When calculating the link weights from the ground node to the normal nodes, LWLR considers both the link weight information between the focal node and its neighboring nodes and the link weights from the nodes adjacent to the focal nodes to their own adjacent nodes.

## 5 Experiment and Result

### 5.1 Data

To evaluate the performance of LWLR for identifying event-specific opinion leaders from news articles, we carried out experiments on two datasets. (1) To compile a Gou Jing event dataset, we crawled a large number of news articles related to that event published between June 22, 2020, and August 22, 2020, on various well-known Chinese-language news websites and filtered those news articles using the keywords “” and “”. Finally, we obtained 1,599 related news articles and used them to construct a person cooccurrence network for the Gou Jing event, abbreviated as the PCNGJE for convenience, via the aforementioned method. (2) To compile a Policeman neck trampling dataset, we crawled a large number of news articles published on an English-language news website between May 28, 2020, and June 28, 2020, and filtered the collected news articles with the keywords “Policeman” and “neck trampling” to obtain 1,083 news articles related to the Policeman neck trampling event. We used these articles to construct a person cooccurrence network for the Policeman neck trampling event (PCNPTE) via the aforementioned method. The basic statistics of these two real-world networks are presented in Tab. 1.

**Table 1:** Basic statistics of two real-world networks, including the number of nodes  $N$ , the number of links  $M$ , the average degree  $\langle k \rangle$ , the minimum value among the link weights  $E_{MIN}$ , the maximum value among the link weights of links  $E_{MAX}$ , and the average value of the link weights  $\langle E \rangle$

Networks	N	M	$\langle k \rangle$	$E_{MIN}$	$E_{MAX}$	$\langle E \rangle$
PCNGJE	2180	2765	19.75	0.052	32.75	0.877
PCNPTE	1876	8743	14.72	0.170	6.983	0.614

### 5.2 Evaluation with the WSI Model

Because the constructed person cooccurrence network is a weighted network, we use a weighted variant of the susceptible-infected (SI) model, i.e., the WSI model, to evaluate the influences of the nodes in the network to obtain the node rankings and identify opinion leaders. In the WSI model, each node can be in only one of two states: susceptible or infected. Initially, all nodes are in the susceptible (S) state, except for one node that is in the infected (I) state. In each time step, the infected nodes influence their susceptible neighbors with the probability of  $\beta = \beta_B \cdot \left(\frac{w_{ij}}{w_{Max}}\right)^\phi$ , where  $\beta_B$  is the base infection rate,  $w_{ij}$  is the weight of the link from node  $i$  to node  $j$ ,  $w_{Max}$  is the maximum link weight in the cooccurrence network, and  $\phi$  is a free parameter that is set to 0.25 in this paper. For minimal  $\beta_B$  values, the influence spread will reach only a very small fraction of the nodes in specific time steps. The obtained node

influence rankings will thus fail to effectively distinguish the influence of the nodes. In contrast, under large  $\beta_B$  values, the influence spread will reach a significant fraction of the nodes in specific time steps. In this case, the role of any individual node is no longer essential, and the spread will cover almost the entire network, independent of its origin. Hence, the obtained ranking list of the nodes in terms of their influence will also be ineffective. Therefore, we use relatively small values of  $\beta_B$  in the WSI model in this paper, that is,  $\beta_B \in [0.01, 0.1]$ . The number of infected nodes in the WSI model is denoted by  $F\langle k \rangle$ , which converges to a steady value as  $t$ , the number of elapsed time steps, increases. That is, all the nodes will eventually be infected. We define  $F\langle t_c \rangle$  as the node influence score, where  $t_c$  denotes a certain time step in which the value of  $F\langle t_c \rangle$  and the slope of  $F\langle t_c \rangle$  are both substantial. Larger values of  $F\langle t_c \rangle$  for a specific node obtained by averaging over 100 independent realizations indicate that this node is more influential in the network and has an increased the likelihood of being an event-specific opinion leader.

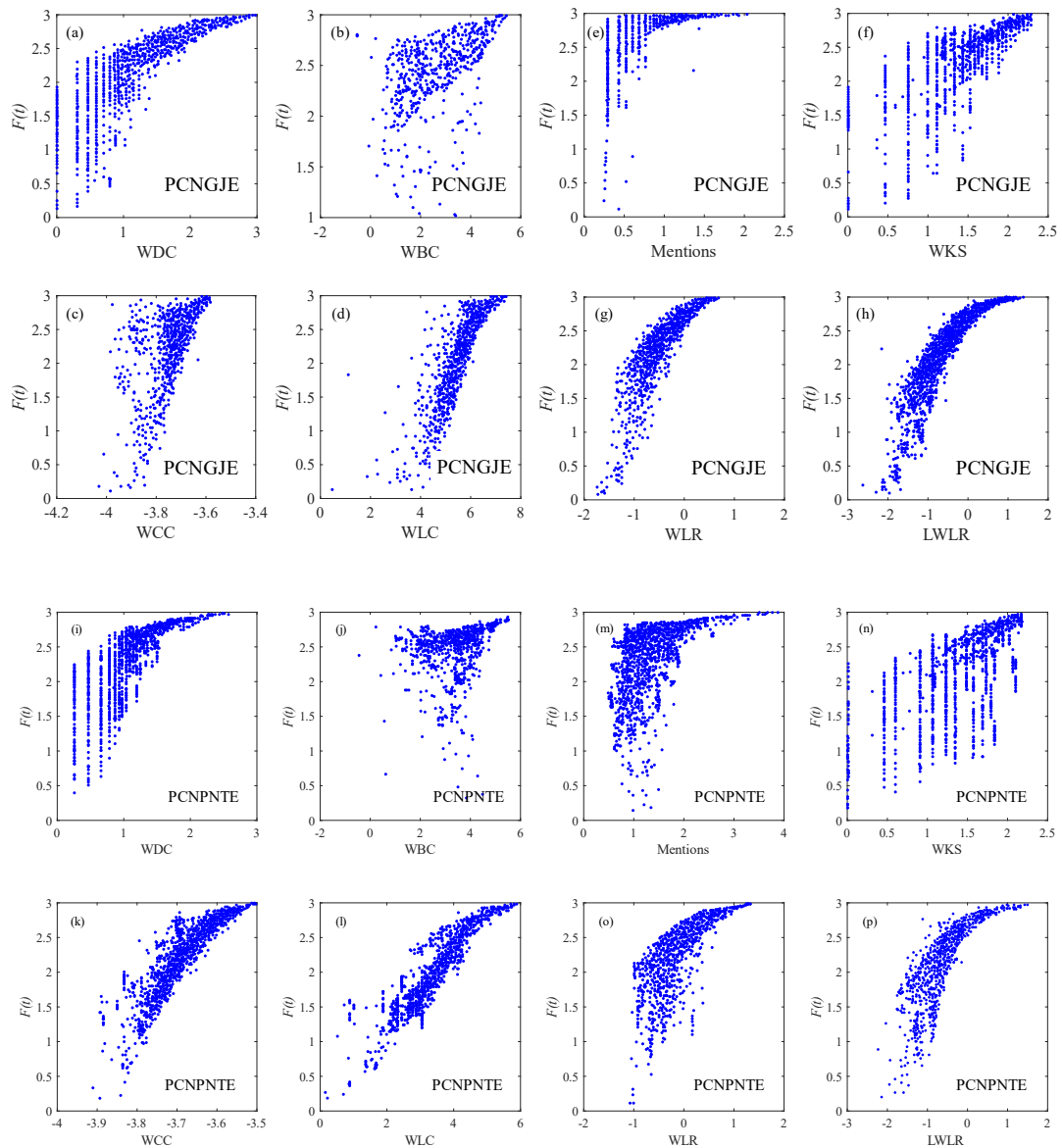
### 5.3 Results

#### 5.3.1 Effectiveness

We used the WSI model to evaluate the node influence score in the networks based on the two aforementioned datasets. In this experiment, we set the free parameter to  $a = 0.25$  and the base infection rate to  $\beta_B = 0.1$ . During the simulated spreading process, a single node was initially set as infected in each realization; the spreading process then propagated from that single node. For the Gou Jing event, we set  $t_c = 9$ , while for the Policeman neck trampling event, we set  $t_c = 10$ . The influence score for each node was obtained by averaging over 100 independent realizations. We calculated the node influence scores in each cooccurrence network using LWLR and the other methods considered in this paper. as introduced previously. The correlations between the influence scores of the nodes as calculated using a particular influence measure and the influence scores of the nodes as calculated using the WSI model are shown in Fig. 2.

Clearly, for the PCNGJE, the correlations between the influence scores of the nodes as calculated using the WDC, WBC, WKS, and Mentions models and the influence scores of the nodes as calculated using the WSI model are not desirable. For instance, the original scores of some nodes with high ranks are lower than those of some nodes with low ranks. The correlations between the influence scores of the nodes as calculated using the WCC, WLC, WLR, and LWLR models and the real influence scores of the nodes as calculated using the WSI model are better. However, WLR and LWLR perform better than the WCC and WLC. LWLR outperforms WLR, especially in term of the fraction of nodes with low scores, because LWLR considers more local information of the focal node, which improves its the ability to distinguish among the influences of nodes with low-to-medium influence scores. The WDC, WBC, WKS and Mentions models perform even worse on the PCNPTE than on the PCNGJE. The WCC and WLR perform slightly better than these four techniques but slightly worse than the WLC and LWLR. Thus, the WLC and LWLR are the two best methods among those considered in this paper. We next briefly analyze the performances of these two methods in ranking nodes with high ranks; for further details, see Section 5.3.2. Clearly, LWLR is better than the WLC at distinguishing the influences of high-ranking nodes. These results indicate that LWLR generally performs better than do the other methods considered in this paper on the PCNPTE. Overall, the above experiments based on the two persons cooccurrence networks can verify that our proposed method is able to effectively identify opinion leaders in networks.





**Figure 2:** The correlation between the influence scores of nodes calculated by all eight influence measure methods considered in this paper and the influence scores of nodes calculated by the WSI model based on two real-word networks. The results shown are the averages of 100 independent realizations

To quantitatively assess the correlations between the ranked lists produced by our method, each other method considered for comparison in this paper, and the WSI model, we calculated Kendall's tau coefficient,  $\tau$ . Kendall's tau coefficient considers a set of joint observations from two random variables, i.e.,  $X$  and  $Y$  (in our case,  $X$  denotes the node influence scores calculated for all nodes using a particular influence measure, and  $Y$  denotes the node influence scores calculated using the WSI model for all nodes). Any pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$  are considered to be concordant if the ranks of both elements are in agreement, i.e., if both  $x_i > x_j$  and  $y_i > y_j$  or if both  $x_i < x_j$  and  $y_i < y_j$ . They are discordant if  $x_i > x_j$  and  $y_i < y_j$  or if  $x_i < x_j$  and  $y_i > y_j$ . If  $x_i = x_j$  or  $y_i = y_j$ , then this pair of observations is neither concordant nor discordant. Kendall's tau coefficient  $\tau$  is calculated as follows:

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \quad (6)$$

where  $n_c$  and  $n_d$  represent the numbers of concordant and discordant pairs, respectively. A higher value of  $\tau$  indicates a more accurate result. The ideal situation is  $\tau = 1$ , indicating that the ranked list generated by the tested measurement method is exactly identical to that generated by the WSI model.

For the PCNGJE and PCNPNTE, the Kendall's tau values for all eight influence measures considered are shown in Tab. 2; Our proposed method achieves the maximum Kendall's tau value for the PCNGJE. The ranked list generated by our proposed method is closer to that generated by the WSI model than generated by any other method considered for comparison in this paper. For the PCNPNTE, the WLC yields the highest tau value. However, the value of tau for our proposed method is very close to this maximum tau value and larger than the tau values of any other method except for the WLC. These results show that the proposed method offers a competitive performance on the PCNPNTE. Generally, the proposed method performs better than any other method except the WLC in regard to the accuracy of the ranked list of all nodes. In addition, compared to the WLC, our proposed method shows more stable performance on different networks.

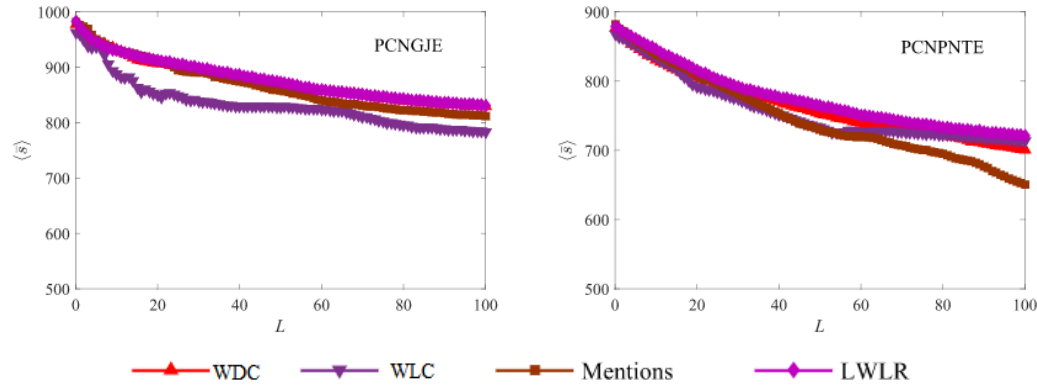
**Table 2:** The Kendall's tau coefficient values for all eight influence measures for both real-word networks

Networks	WDC	WBC	WCC	WLC	WKS	Mentions	WLR	LWLR
PCNGJE	0.7140	0.3692	0.6238	0.7651	0.7081	0.4211	0.7521	0.8432
PCNPNTE	0.5508	0.2713	0.7547	0.8414	0.5149	0.2703	0.6217	0.8095

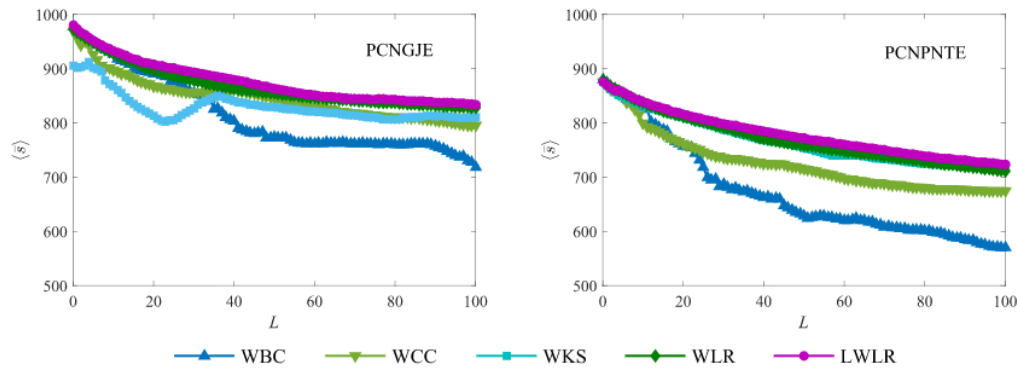
### 5.3.2 Comparison of the Influence of the Top-L Nodes

Because Kendall's tau coefficient evaluates only the overall accuracy of the ranked list of all nodes in the network as generated by a particular method, it cannot efficiently assess the ranking accuracy of the top-ranked nodes. In this paper, we specifically need to identify the top-ranked opinion leaders accurately because they play a significant role in event development. Thus, to test the real influence of the top-ranked nodes, we introduce  $\langle \bar{s} \rangle$ , which is the mean value of the average influence scores of the top-L nodes as ranked by each influence measurement method.

Here, we contrast the  $\langle \bar{s} \rangle$  values of our method with those of the other influence measurement methods on the PCNGJE and PCNPNTE. To illustrate the results more clearly, we separately present the comparisons with the local metrics, i.e., the WDC, the WLC and Mentions, as shown in Fig. 3, and with the global metrics i.e., the WBC, the WCC, the WKS and WLR, as shown in Fig. 4. The curve for a good influence measurement method should slope downward; that is, the mean value of the average influence scores of the top-L nodes should decrease as L increases. Clearly, Fig. 3 and Fig. 4 show that the proposed method outperforms the other methods throughout almost the entire range of L values on both the PCNGJE and the PCNPNTE. On the PCNGJE, the methods closest to our proposed method in terms of their performance are WLR and WDC, while on the PCNPNTE, WLR, the WDC and the WKS are the methods closest to ours. The WLC performs better in regard to the overall ranking for all nodes in the network; however, its performance is not satisfactory for distinguishing the influences of the top-ranked nodes. In brief, by comparing the  $\tau$  and  $\langle \bar{s} \rangle$  results of the various considered methods, we can conclude that our proposed LWLR method yields a ranked list of nodes in terms of their influence that is closer to reality and better identifies the top-ranked opinion leaders in the network than the lists generated by the other considered methods.



**Figure 3:** The mean of the average influence  $\langle \bar{s} \rangle$  of the top- $L$  nodes as ranked based on our local weighted LeaderRank (LWLR) measure and other local metrics, namely, the weighted degree centrality (WDC), the weighted local centrality (WLC), and Mentions. These results were obtained by averaging over 100 independent realizations



**Figure 4:** The mean of the average influence  $\langle \bar{s} \rangle$  of the top- $L$  nodes as ranked based on our local weighted LeaderRank (LWLR) measure and various global metrics, namely, the weighted k-shell centrality (WKS), the weighted betweenness centrality (WBC), the weighted closeness centrality (WCC) and weighted LeaderRank (WLR). These results were obtained by averaging over 100 independent realizations

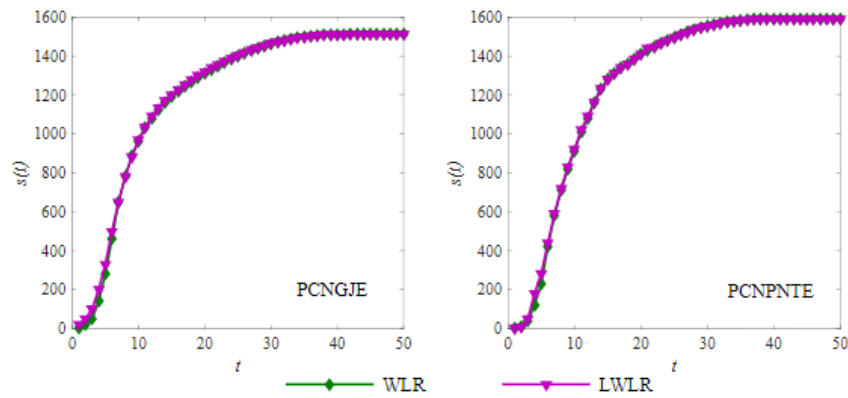
### 5.3.3 Comparison of the Spreading Influence of the Top-10 Nodes

From the analysis above, we know that LWLR ranks the influences of the top-ranked nodes better than the other considered methods do. However, this conclusion considers only the node influence scores in a specific time step of the influence-spreading process of the WSI model, that is, the node influence scores in time step  $t_c$ . To analyze the overall performances of the nodes that appear in the top-ranked list generated by our proposed method throughout the influence-spreading process, we conducted a comparative analysis of the influences of the top 10 nodes obtained by our method and those obtained by WLR. We chose WLR for this comparison because among all the considered methods, its performance is the most similar to that of LWLR in identifying the top 10 nodes. In these experiments, we initially set the nodes appearing in the top 10 list generated by either WLR or LWLR (but not appearing in both lists) as infected, recorded the influence scores throughout the spreading process, and obtained the mean of these spreading influence scores by averaging over 100 independent realizations. Note that if the effect of the common nodes that appear in both ranking lists is not considered, the differences between these two methods can be more effectively distinguished. The simulation results obtained for the two real-world networks are shown in Fig. 5. To reveal the differences between the WLR and LWLR methods more clearly, the number of spreading step  $t$  was limited to 50. In Fig. 5, the curves corresponding to LWLR all lie above those corresponding to WLR for both networks, and the standard deviations for LWLR are

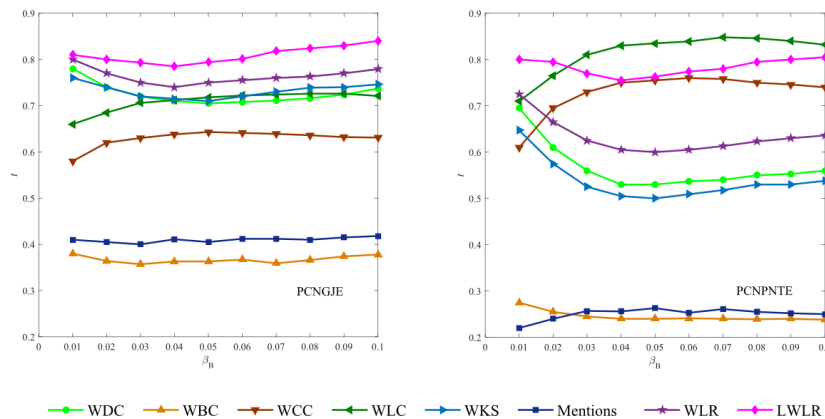
almost all smaller than those for WLR throughout the spreading process. Thus, the nodes that appear in the top-ranked list based on the LWLR measure have a stronger influence on the spreading ability, and they propagate their influence more quickly and broadly, further verifying the effectiveness of our method.

5.3.4 Kendall's Tau with Different Base Infection Rates  $\beta_B$

All of the preceding experiments were performed with the base infection rate set to  $\beta_B = 0.1$ . To consider the impact of different values of  $\beta_B$  on Kendall's tau with regard to the ranked list generated by a specific influence measurement method and the real ranked list generated by the WSI model, we investigated the performance of the WSI model for  $\beta_B$  values in the range of  $[0.01, 0.1]$ . Then, we obtained the values of Kendall's tau under different values of  $\beta_B$ , as shown in Fig. 6. Our method yields the largest Kendall's tau value throughout the entire range of  $\beta_B$  on the PCNGJE, while on the PCNPNTNTE, our method achieves better Kendall's tau values than those obtained with the WLC and performs slightly better than the WLC for relatively small values of  $\beta_B$ , while both methods outperform the other considered methods over the entire range of  $\beta_B$ . In summary, our approach performs better in ranking node influences and in identifying event-specific opinion leaders in a network over the entire range of  $\beta_B$ .



**Figure 5:** The number of infected nodes  $s(t)$  as a function of time  $t$  under the SI model. The initially infected nodes appear either in the top-10 list generated by weighted LeaderRank (WLR, in the brown right-facing triangle) or in the list generated by local weighted LeaderRank (LWLR, in the magenta rhombus) but not appearing in both lists. These results were obtained by averaging over 100 independent realizations



**Figure 6:** Kendall's tau  $\tau$  values obtained by comparing the ranked lists generated by all eight influence measurement methods considered in this paper and the ranked list generated by the WSI model based on two real-world networks, when the spreading probability  $\beta_B$  is in the range of 0.01 to 0.1. These results were obtained by averaging over 100 independent realizations

## 6 Discussion

In this paper, we propose the LWLR measure for identifying opinion leaders in event-specific person cooccurrence networks and compare its performance with that of other influence measures such as the WDC, WBC, and WLC. Among the compared methods, the WLC achieves performance comparable to that of LWLR in the overall node ranking but does not perform as well as LWLR in the ranking of the top-L nodes. This result is likely attributable to the implementation of the WLC. The WLC acquires a tremendous amount of local information about the focal node, which improves its performance in distinguishing the influences of all the nodes in the network with regard to the overall ranking but may negatively affect the influences of some closely linked nodes. For instance, the WLC does not discriminate well among the influences of the top-L nodes in the two real-world networks considered in this paper because these top-L nodes are linked quite closely together and the information regarding the local structures and link weights of these nodes is lacking. Therefore, the WLC scores for these nodes are highly similar and less capable of serving as a basis for discrimination. Due to this problem, the WLC scores worse than the proposed LWLR measure in the overall evaluation. After all, the top-L nodes in the ranking list, which play an essential roles in the development of an event, are highly significant and important for identifying event-specific opinion leaders.

## 7 Conclusions

This paper presents an approach for identifying event-specific opinion leaders in news articles. We first detail how to construct a person cooccurrence weighted network using news articles on a specific event. Then, we propose an improved variant of WLR, called LWLR, to measure the event-specific influences of persons in the cooccurrence network. Compared to the WLR, our proposed method uses both the link weight information of the cooccurrence network and additional local information of the focal node. To evaluate the effectiveness of the proposed method, we have applied it to two real-world event-specific person cooccurrence networks and simulated the spreading process with the WSI model. The Kendall's tau( $\tau$ ) coefficient is used to measure the rank correlations between the ranked lists created from the simulation results and the ranked lists created using different influence measurement methods. We show that our LWLR measure ranks the influences of nodes more precisely than do the WDC, WKS, WBC, and WCC measures do, while it achieves an accuracy comparable to that of the WLC. Furthermore, by comparing the average influence of the top-L nodes in the ranked lists generated by the proposed method and those generated by the other methods, we demonstrate that our method outperforms the other tested influence measures. We also compare the spreading processes of the top 10 nodes in the ranked lists generated by our method and in those generated by another method and show that the top 10 nodes in the ranked list generated by our method have a stronger spreading ability and can propagate their influence more quickly and broadly, further verifying the effectiveness of the proposed approach. Finally, we investigate the values of Kendall's tau coefficient ( $\tau$ ) as calculated between the ranked lists produced from the simulation results and the ranked lists produced based on the various influence measures when the base  $\beta_B$  value is in the range of [0.01, 0.1]. The results show that the proposed method achieves an accuracy comparable to that of the WLC and outperforms the other methods (except the WLC) over the entire range of  $\beta_B$ . In conclusion, the experimental results prove the effectiveness of our method for identifying opinion leaders in an event-specific person cooccurrence networks constructed using news articles related to a specific event.

**Funding Statement:** This research was supported by the National Natural Science Foundation of China (No. 61862002).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] S. M. Aghdam and N. J. Navimipour, "Opinion leaders selection in the social networks based on trust relationships propagation," *Karbala International Journal of Modern Science*, vol. 2, no. 2, pp. 88–97, 2016.
- [2] L. J. Dong and Z. Y. Peng, "Research on the application of opinion leaders in new media marketing of library," *Library Work and Study*, vol. 6, pp. 105–108, 2017.
- [3] Sh. W. Chen, D. H. Glass and M. McCartney, "Characteristics of successful opinion leaders in a bounded confidence model," *Physica A: Statistical Mechanics and Its Applications*, vol. 449, pp. 426–436, 2016.
- [4] H. U. Khan and A. Daud, "Finding the top influential bloggers based on productivity and popularity features," *New Review of Hypermedia and Multimedia*, vol. 23, no. 3, pp. 189–206, 2017.
- [5] O. Z. Gökçe, E. Hatipoğlu, G. Göktürk, B. Luetgert and Y. Saygin, "Twitter and politics: Identifying turkish opinion leaders in new social media," *Turkish Studies*, vol. 15, no. 4, pp. 671–688, 2014.
- [6] A. Aleahmad, P. Karisani, M. Rahgozar and F. Oroumchian, "Olfinder: Finding opinion leaders in online social networks," *Journal of Information Science*, vol. 42, no. 5, pp. 659–674, 2016.
- [7] S. A. Semenkovich and O. A. Tsukanova, "On the algorithms of identifying opinion leaders in social networks," *Procedia Computer Science*, vol. 162, pp. 778–785, 2019.
- [8] J. Zhang, "Personalised product recommendation model based on user interest," *Computer Systems Science and Engineering*, vol. 34, no. 4, pp. 231–236, 2019.
- [9] E. Qi and M. Deng, "R&D investment enhance the financial performance of company driven by big data computing and analysis," *Computer Systems Science and Engineering*, vol. 34, no. 4, pp. 237–248, 2019.
- [10] Q. C. Wu and F. Zhang, "Dynamical behavior of susceptible–infected–recovered–susceptible epidemic model on weighted networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 491, pp. 382–390, 2018.
- [11] Z. X. Wang, Y. Zhao, J. K. Xi and C. J. Du, "Fast ranking influential nodes in complex networks using a k-shell iteration factor," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 171–181, 2016.
- [12] Q. Ma and J. Ma, "Identifying and ranking influential spreaders in complex networks with consideration of spreading probability," *Physica A: Statistical Mechanics and Its Applications*, vol. 465, pp. 312–330, 2017.
- [13] H. Zhang, G. Chen and X. Li, "Resource management in cloud computing with optimal pricing policies," *Computer Systems Science and Engineering*, vol. 34, no. 4, pp. 249–254, 2019.
- [14] B. Wei, J. Liu, D. J. Wei, C. Gao and Y. Deng, "Weighted k-shell decomposition for complex networks based on potential edge weights," *Physica A: Statistical Mechanics and its Applications*, vol. 420, pp. 277–283, 2015.
- [15] K. Angelou, M. Maragakis and P. Argyrakis, "A structural analysis of the patent citation network by the k-shell decomposition method," *Physica A: Statistical Mechanics and its Applications*, vol. 521, pp. 476–483, 2019.
- [16] S. Wang, F. Wang, Y. Chen, C. Liu, Z. Li *et al.*, "Exploiting social circle broadness for influential spreaders identification in social networks," *World Wide Web*, vol. 18, no. 3, pp. 681–705, 2015.
- [17] J. f. Sheng, J. Y. Dai, B. Wang and S. Z. Jun, "Identifying influential nodes in complex networks based on global and local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 541, 2020. <https://doi.org/10.1016/j.physa.2019.123262>.
- [18] R. Elbarougy, G. Behery and A. E. Khatib, "Extractive arabic text summarization using modified pagerank algorithm," *Egyptian Informatics Journal*, vol. 21, pp. 73–81, 2020.
- [19] H. Schmid-Petri, S. Adam, U. Reber, T. Häussler, D. Maier *et al.*, "Homophily and prestige: an assessment of their relative strength to explain link formation in the online climate change debate," *Social Networks*, vol. 55, pp. 47–54, 2018.
- [20] S. S. Li, Cui, T. J. and J. Liu, "Research on the clustering analysis and similarity in factor space," *Computer Systems Science and Engineering*, vol. 33, no. 5, pp. 397–404, 2018.
- [21] S. Xu and P. Wang, "Identifying important nodes by adaptive leaderrank," *Physica A: Statistical Mechanics and its Applications*, vol. 469, pp. 654–664, 2017.
- [22] S. Kant, T. Mahara, V. Kumar Jain, D. Kumar Jain and A. Kumar Sangaiah, "Leaderrank based k-means clustering initialization method for collaborative filtering," *Computers & Electrical Engineering*, vol. 69, pp. 598–609, 2018.
- [23] K. H. Zhang and K. Q. Li, "The optimization reachability query of large scale multi-attribute constraints directed graph," *Computer Systems Science and Engineering*, vol. 33, no. 2, pp. 71–85, 2018.