IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 18, 2019, accepted October 15, 2019, date of publication October 25, 2019, date of current version November 6, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2949678*

# Identifying Multiple Influential Users Based on the Overlapping Influence in Multiplex Networks

**JIANJUN CHEN[1], YUE DENG[1], ZHEN SU[2,3], SONGXIN WANG[4], CHAO GAO[1,5], AND XIANGHUA LI[1]**

[1]College of Computer and Information Science, Southwest University, Chongqing 400715, China
[2]Potsdam Institute for Climate Impact Research, 11473 Potsdam, Germany
[3]Institute of Physics, Humboldt University of Berlin, 12489 Berlin, Germany
[4]School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China
[5]Department of Computer Science, Hong Kong Baptist University, Hong Kong

Corresponding author: Xianghua Li (li_xianghua@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976181 and Grant 11931015, in part by the Natural Science Foundation of Chongqing under Grant cstc2018jcyjAX0274 and Grant cstc2019jcyj-zdxmX0025, and in part by the National Training Programs of Innovation and Entrepreneurship for Undergraduates under Grant 201910635036.

**ABSTRACT** Online social networks (OSNs) are interaction platforms that can promote knowledge spreading, rumor propagation, and virus diffusion. Identifying influential users in OSNs is of great significance for accelerating the information propagation especially when information is able to travel across multiple channels. However, most previous studies are limited to a single network or select multiple influential users based on the centrality ranking result of each user, not addressing the overlapping influence (OI) among users. In practice, the collective influence of multiple users is not equal to the total sum of these users' influences. In this paper, we propose a novel OI-based method for identifying multiple influential users in multiplex social networks. We first define the effective spreading shortest path (ESSP) by utilizing the concept of spreading rate in order to denote the relative location of users. Then, the collective influence is quantified by taking the topological factor and the location distribution of users into account. The identified users based on our proposed method are central and relatively scattered with a low overlapping influence. With the Susceptible-Infected-Recovered (SIR) model, we estimate our proposed method with other benchmark algorithms. Experimental results in both synthetic and real-world networks verify that our proposed method has a better performance in terms of the spreading efficiency.

**INDEX TERMS** Multiplex networks, influential users, overlapping influence, shortest path.

## I. INTRODUCTION

The development of online social networks (OSNs) has created a new major interaction medium and formed promising landscape for information dissemination. The engagement of online users generates a huge volume of data for investigating the human behavioral patterns [1]. More importantly, the fact that an opinion or decision of individuals is influenced by their neighbors or friends has a considerable impact on the popularity of new products or brands [2], [3]. Targeting influential users is vital for designing techniques for either accelerating the information diffusion in marketing applications or suppressing the propagation of unwanted contents [4], [5]. The crucial problem is how to select multiple users, called central users, who can influence a massive number of users [6]. The measurement of influential users is beneficial for advertisers to implement effective campaigns. Central users are believed to play a key role in the propagation process. In practice, if a virus attacks a central user with a large degree, betweenness, PageRank or k-shell [7], [8], it would quickly pervade the whole network [9]. If we protect or immunize these users, the propagation scale would be greatly alleviated [10].

Although the propagation dynamics have received more and more attention, most of the studies still remain in a single network [11]. However, in fact, a user often has more than one social account such as Twitter, Facebook and Instagram. More recently, scientists start to consider a particular class of

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

networks in which it is coupled by several subnetworks and vertices have multiple different types of links among layers, called multiplex networks [12], [13]. Due to the multichannel characteristics of information propagation, multiplex social networks can more accurately describe the structure of complex systems and their interactions while the single networks usually neglect these links among the users.

The identification of influential users is generally based on the ranking relative to the centrality measurement. In the last few years, certain centrality measures for multiplex networks have aroused intense interest and discussion to solve the problem of user influence ranking [14]–[19]. However, classical methods for selecting influential users based on the centrality only consider the structural characteristics of individuals, but do not address the overlapping influence among the chosen nodes [20], [21]. The overlapping influence is a common phenomenon in a propagation process, which is invalid for maximizing the influence spreading [22]. For example, as shown in Fig. 1, if two users (i.e., $V_1$ and $V_2$) who share a lot of common friends (i.e., $V_3$, $V_4$ and $V_5$) are selected as the initial spreaders to spread a message in social networks, their common friends would receive the redundant information and the spread of the message is limited. The phenomenon that spreaders affect a group of same users is regarded as the overlapping influence [22].

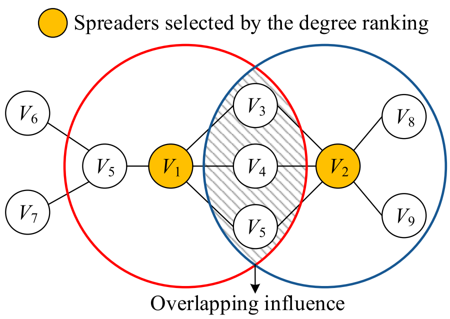



**FIGURE 1. An illustration of the overlapping influence. Two users ($V_1$ and $V_2$) with the highest degree are selected as spreaders. Users in the red and blue circles represent the neighbors of $V_1$ and $V_2$, respectively. However, $V_1$ and $V_2$ share three common friends (i.e., $V_3$, $V_4$ and $V_5$), suppressing the influence maximization. Specifically, when a large amount of redundant information is transmitted to common friends (i.e., $V_3$, $V_4$ and $V_5$), it will cause serious waste of resources.**

More specifically, the collective influence of multiple users is not equal to the sum of influence of each user [23], [24]. Therefore, finding a set of users with the maximum influence is one of the most problems in the field of network science. Distance among users has been confirmed to affect the spreading efficiency especially when the number of users is large [25]. Besides, spreading rates in different layers are of great significance for the propagation scale [26]. Through analyzing these features, we make some improvements to current topology-based strategies by taking the overlapping influence into consideration.

In this paper, we propose a novel approach that considers both the structural properties of individuals and the relative location of the selected users which can effectively reduce the overlapping effect among multiple users in the propagation process. Specifically, the structural property stands for the centrality measure, and the location distribution depends on the effective spreading shortest path (denoted as ESSP). ESSP is defined to quantify the relative distance among users in multiplex netowrks capitalizing on the idea of the spreading rate in the spreading process. Compared with the existed methods, we balance these two metrics and find a certain fraction of influential users rather than measure the influence of a single user. Then the numerical simulation results with Susceptible-Infected-Recovered (SIR) model verify that our proposed method outperforms traditional methods in the spreading efficiency in terms of the maximum spreading influence. The main contribution of this paper can be summarized as:

1) A novel method is proposed by tactfully combining the structural properties and relative locations of multiple users, which aims to minimize the overlapping influence and maximize the spreading influence.
2) The effective spreading shortest path (ESSP) is proposed to measure the collective influence of multiple users and a series of multiplex networks are adopted to verify the effectiveness of the proposed method.

The rest of the paper is organized as follows: Sec. II introduces the related definitions, approaches, and models. Then a novel method for selecting multiple influential users in multiplex networks is proposed in Sec. III. Sec. IV presents the results of simulation experiments based on six networks. Sec. V provides conclusions and future directions of the research.

## II. RELATED WORK

This section provides the theoretical foundation of our work. Specifically, Sec. II-A states the formulation of multiplex networks. Then, Sec. II-B introduces the centrality measures for multiplex networks. Sec. II-C presents the detailed study on the overlapping influence.

### A. FORMULATION OF MULTIPLEX NETWORKS

A multiplex network can be naturally defined as a combination of graphs where the notion of a layer $L_m = \{1, \ldots, L\}$ is introduced as follows: $\{G_\alpha\}_{\alpha=1}^{L} = \{(V_\alpha, E_\alpha)\}_{\alpha=1}^{L}$. Generally, the user sets are the same across the different layers (i.e., $V_\alpha = V_\beta = V, \forall \alpha, \beta \in L_m$). An edge set $E_\alpha \subseteq V_\alpha \times V_\alpha$ represents the interactions of a particular relation among users such as friendship ties, family ties or co-worker ties [12], [13], [27]. For instance, $G_1$ and $G_2$ represent social networks of Twitter and Facebook, respectively. The user sets $V_1$ and $V_2$ share the same users since a user often has more than one social account. However, the edge sets $E_1$ and $E_2$ can be obviously different within the networks due to the various relationships.

The propagation process in a multiplex network can be characterized by traditional propagation models. Here we

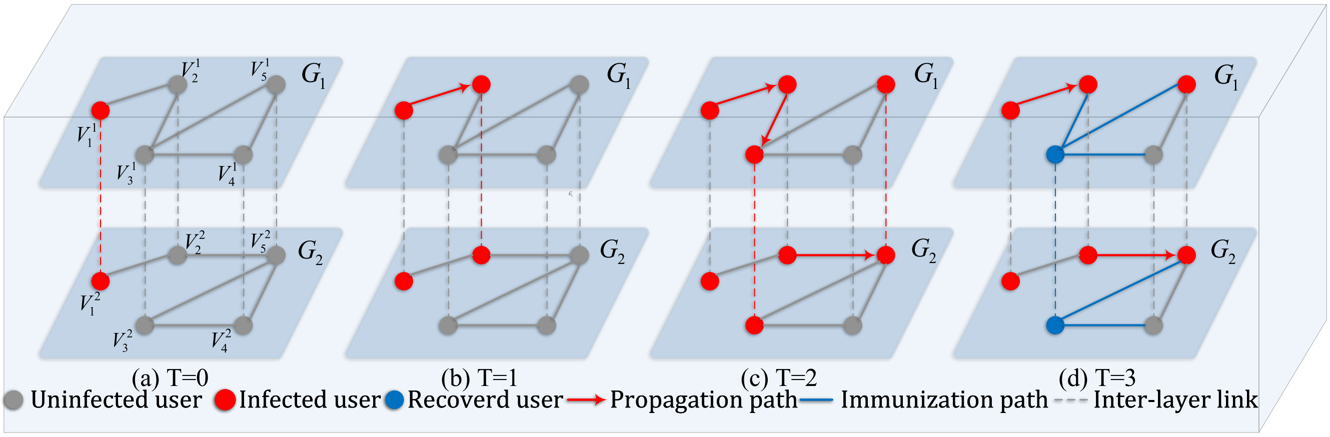



**FIGURE 2.** An illustrative example of information propagation process in a multiplex network. At $T = 1$, the infected user successfully transmit the information to its neighbor (i.e., from $V_1^1$ to $V_2^1$) while the coupled user (i.e., $V_2^2$) in the other layer is infected automatically. At $T = 2$, the information propagates in both layers (i.e., from $V_2^1$ to $V_3^1$ and from $V_2^2$ to $V_5^2$). At $T = 3$, one of the infected users (i.e., $V_3^2$) is recovered and will not be infected anymore.

adopt the Susceptible-Infected-Recovered (SIR) model as the information propagation model [28]. $N$ users in a network can be divided into three states: susceptible, infected, and recovered. The dynamics of the SIR model can be described as Eq. (1):

$$\begin{cases} \dfrac{dS(t)}{dt} = -\mu I(t)S(t) \\ \dfrac{dI(t)}{dt} = \mu I(t)S(t) - \delta I(t) \\ \dfrac{dR(t)}{dt} = \delta I(t) \end{cases} \tag{1}$$

where $\mu$ denotes the probability that an infected user successfully transmits the information to the susceptible neighbors at each step and $\delta$ is the probability of infected individuals recovering and immunizing, which means the recoverd users will not spread the information and be affected by it anymore.

In multiplex networks, the propagation is expected to diffuse among different layers. The probability of infection and recovery may be different in each layer. We experience the infection probability $\mu_\gamma$ and recovery probability $\delta_\gamma$ in a layer $\gamma$. As shown in Fig. 2, if a user (i.e., $V_2^1$) in one layer is infected, at each step, it will spread the information to the susceptible neighbors in all layers (i.e., $V_2^1$ to $V_3^1$ and from $V_2^2$ to $V_5^2$). Besides, the user sets are the same across different layers, which here means the coupled users in different layers share the same state: $S$, $I$ or $R$.

In the applications of maximization influence, such as viral marketing, current methods usually select a group of nodes for protection or infection according to the centrality ranking result of each node [29], [30]. The centrality measures for multiplex network are introduced in the following section.

### B. CENTRALITY MEASURES FOR MULTIPLEX NETWORKS

Recently, centrality ranking measures for multiplex networks have been investigated quite intensively, aiming at extending the classical centrality measures from the single-layer networks. Current approaches for the centrality in multiplex networks can be classified into three main categories: (1) Aggregate multiplex networks and then estimate the centrality metrics on such aggregated network [31], [32]. (2) Treat the layers independently and evaluate the centrality in multiple ways [18]. (3) The general centrality measure for multiple networks is calculated directly considering the interactions among layers [15]–[19], [33]–[35].

For examples, Sola *et al.* [19] introduced a new eigenvector centrality for a multiplex network, assuming a node ranking in a layer is affected by the counterparts in other layers. Arda *et al.* [15] proposed a multiplex PageRank metric based on the consideration of the mutual influence of the same nodes between different layers. De Domenico *et al.* [18] built a mathematical model with tensor and exploited PageRank, betweenness and eigenvector centralities to multiplex networks for ranking the versatile nodes. Sole -Ribalta *et al.* [36] firstly gave a version of multiplex betweenness centrality which estimates the betweenness centrality in each layer independently and aggregated it together. Based on that, Chakraborty and Narayanam *et al.* [34] proposed a general cross-layer betweenness centrality, capitalizing on the idea of the inter-layer shortest path. However, these traditional centrality measures for multiplex networks only focus on the position characteristics of a single node, not addressing the overlapping influence among multiple nodes.

### C. OVERLAPPING INFLUENCE

There are already some studies on the effects of overlap in single networks. Hu *et al.* [25] investigated the relationship of spreading influence and distance between spreaders, the theoretical and experimental results have shown that for regular networks the larger the distance between spreaders is, the more effective the spreading would be. Guo *et al.* [37] introduced an approach which can obviously improve the influence spreading by adding a distance parameter into the coloring method. Zhou *et al.* [23] proposed a collective influence metric of multiple spreaders via evaluating the overlapping influence according to the degree, betweenness, and PageRank. Zhang *et al.* [38] presented an iterative method named 'VoteRank' to identify a set of decentralized spreaders where all nodes vote in a spreader in each turn, and the voting ability of neighbors of selected spreader will decrease in subsequent turn. However, most of the current studies are still limited in a single network. How to extend the study of overlapping influence to multiplex networks is what we concern in this paper.

## III. OI-BASED METHOD

This section introduces the detailed principles of our proposed method. The effective spreading shortest path is first defined in Sec. III-A. Then, Sec. III-B derives the overlapping influence based (OI-based) method to identify multiple influential users.

### A. EFFECTIVE SPREADING SHORTEST PATH

The shortest path between spreaders in multiplex networks shows a great impact on the overlapping influence [21], [25]. The multiplex shortest path indicates a path having the minimum length from the source user in any layer to the destination user in any layer [36]. Specifically, we define a path on a multiplex network consisting of $L$ layers and $N$ users: $p_{x^\alpha\to y^\beta} \in P_{x^\alpha\to y^\beta}$ which starts from a user $x$ in a layer $\alpha$ to a user $y$ in a layer $\beta$. $P_{x^\alpha\to y^\beta}$ denotes a set of all possible paths and we can define a distance $d(p_{x^\alpha\to y^\beta})$ as the length of such path. Therefore, for $\forall\alpha, \beta \in L$ and $\forall x, y \in N$, the multiplex shortest path can be defined in Eq. (2):

$$P^*_{x^\alpha\to y^\beta} = \underset{p_{x^\alpha\to y^\beta}\in P_{x^\alpha\to y^\beta}}{\arg\min} \; d(p_{x^\alpha\to y^\beta}) \tag{2}$$

In this paper, we propose an effective spreading shortest path (denoted as ESSP) in multiplex networks based on the spreading rate. For the purpose of finding a path that starts from a user $x$ and reaches a user $y$ in a shortest time, we usually choose the traditional shortest path in networks. However, in the propagation process in multiplex social networks, a multiplex shortest path may not be the best choice since the different spreading rate in each layer can increase its time consumption.

In the propagation process, the spreading rate in a multiplex network is defined in Eq. (3):

$$\lambda_\gamma = \frac{\mu_\gamma}{\delta_\gamma} \tag{3}$$

where $\gamma \in \{1, 2, ..., L\}$, and $\mu_\gamma$ and $\delta_\gamma$ denote the infection and recovery probability in a layer $\gamma$, respectively.

Suppose that we construct a multiplex network $G$ with different spreading rates of each layer. Assume that there are several paths $p_{x^\alpha\to y^\beta} \in P_{x^\alpha\to y^\beta}$ from a user $x$ in a layer $\alpha$ to a user $y$ in a layer $\beta$. Each link $(i, j)^\gamma$ in a path may occur in different layers. To define the effective spreading path in multiplex networks, the spreading rates can be multiplicative or additive through the path. The maximization of multiplicative spreading rate can be defined in Eq. (4):

$$\max \prod_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \lambda_\gamma \tag{4}$$

which can be transformed into the minimization of the additive spreading rates in Eq. (5):

$$\begin{aligned} \max \prod_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \log\lambda_\gamma &= \max \sum_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \log\lambda_\gamma \\ &= \min(-\sum_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \log\lambda_\gamma) \\ &= \min(\sum_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \log\frac{1}{\lambda_\gamma}) \end{aligned} \tag{5}$$

Therefore, the effective spreading path can be obtained by minimizing the additive spreading rates which changes the spreading rate of each link to $\log\frac{1}{\lambda_\gamma}$. Based on that, the effective spreading path length and the shortest path can be defined in Eq. (6) and Eq. ( 7), respectively.

$$d^s(p_{x^\alpha\to y^\beta}) = \sum_{(i,j)^\gamma\in p_{x^\alpha\to y^\beta}} \log\frac{1}{\lambda_\gamma} \tag{6}$$

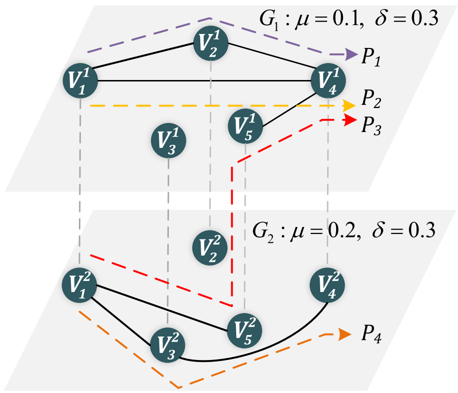
$$p^s_{x^\alpha\to y^\beta} = \underset{p_{x^\alpha\to y^\beta}\in P_{x^\alpha\to y^\beta}}{\arg\min} \; d^s(p_{x^\alpha\to y^\beta}) \tag{7}$$

In order to reflect the rationality of our method, we construct a real situation as shown in Fig. 3. When a user $V_1$ transmits a message to a user $V_4$ through four potential paths, the traditional shortest path in a network does not represent the most effective propagation path due to the fact that the spreading rate in $G_1$ limits its probability of propagation although $V_1^1$ is directly linked with $V_4^1$. Instead, the $P_4$ path in $G_2$ is the path which is most likely to tell $V_4$ in a short period of time based on Eq. (7). The specific calculation process is shown in Alg. 1.

### B. OVERLAPPING INFLUENCE-BASED METHOD

In practice, a user with a large centrality is believed to contribute to the spreading efficiency while the long distance among multiple users is of great significance to suppress the overlapping influences. In order to identify a certain fraction of influential users, we propose a collective influence metric that combines the topology characteristic with the location distribution defined in Eq. (8):

$$\Phi(S) = \sum_{x\in S} c_x \cdot \frac{1}{\frac{1}{2}|S|(|S|-1)} \sum_{x,y\in S;x\neq y} d^*(p_{x^\alpha\to y^\beta}) \tag{8}$$




| Paths | Link count $G_1$ | Link count $G_2$ | Effective spreading path length |
|---|---|---|---|
| $P_1$ | 2 | 0 | 0.95 |
| $P_2$ | 1 | 0 | 0.48 |
| $P_3$ | 1 | 1 | 0.65 |
| $P_4$ | 0 | 2 | 0.35 |

**FIGURE 3. An illustrative example of a multiplex network consisting of links of individuals in $G_1$ and $G_2$. There are four paths from $V_1$ to $V_4$. The table shows the number of links traversed in each layer for the four paths and the effective spreading path length computed by taking infection and recovery probabilities into consideration based on Eq. (6). $P_4$ is shown as the effective spreading shortest path.**

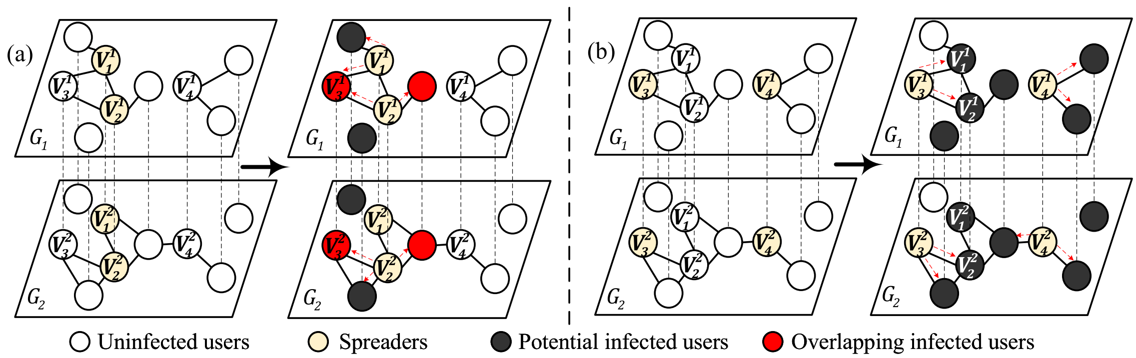



**FIGURE 4. An illustration of our proposed multiple users selection method. We assume that $\lambda_1 = 0.3$ and $\lambda_2 = 0.5$. (a) $V_1$ and $V_2$ are selected as spreaders according to the aggregation degree and they transmit the information to their neighbors. However, since they share multiple common neighbors which means a high overlapping influence, the spreading scale is limited. (b) $V_3$ and $V_4$ are selected as spreaders computed by our OI-based method. In the next time step, it shows a greater probability to expand the spreading scale and reduce the overlapping influence.**

**Algorithm 1** Computation of ESSP Length

**Input**: $G_m$, the spreading rate $\lambda$, the spreader $s$
**Output**: The ESSP length $d^*$

Initialize the min-priority queue $Q$
**while** *$Q$ is not empty* **do**
  $u = Q$.extract_min()
  **for** *each layer $i$ of $L$* **do**
    **for** *each neighbor $v_i$ of $u_i$* **do**
      $length(v_i, u_i)$ is calculated based on Eq. (6)
      $alt = d^*[u_i] + length(v_i, u_i)$
      **if** $alt < d^*[v]$ **then**
        $d^*[v] = alt$
        $Q$.decrease_priority($v$, $alt$)

where $S$ represents a set of users, $c_x$ denotes the centrality of user $x$, and $d^*(p_{x^\alpha \to y^\beta})$ is the value of the effective spreading shortest path length. Based on that, we find the optimal set of users in all possible sets $S'$ which has a maximum collective influence in Eq. (9):

$$S = \underset{S \in S'}{\arg\max}\ \Phi(S) \tag{9}$$

For the purpose of finding an optimal set of users, here we adopt a greedy algorithm to maximize $\Phi$. We initially select top-$K$ users with the highest centralities and add them into the spreader set. At each time step, we temporarily replace a spreader with a new candidate user and compare the collective influence. If the collective influence increases, the spreader set changes, and the iteration continues until $\Phi$ stops increasing.

For identifying a certain fraction of users in multiplex social networks, an example is shown in Fig. 4 to verify the effectiveness of the overlapping influence-based (OI-based) method. The calculation process is specified in Alg. 2 .

## IV. EXPERIMENTS AND RESULTS

The details of our experiments are stated in this section. Specifically, Secs. IV-A and IV-B describe the performance

**Algorithm 2** The Process of Selecting a Set of Influential Users

**Input**: The centrality of users in $G$, the length of ESSP $d^*$, the number of spreaders $K$
**Output**: The selected spreader set $S$
1 Initialize $S$ to the set of top-$K$ users with the maximum centralities
2 Calculate the initial collective influence $\Phi$ based on Eq. (8)
3 **while** $\Phi$ *increases in the last iteration* **do**
4     **for** $n = 1{:}N$ **do**
5         Temporarily replace spreader with $n$
6         Compute the new collective influence $\Phi$
7         **if** $\Phi$ *increases* **then**
8             Update the user set $S$

metrics and datasets, respectively. The effectiveness of our proposed method is estimated in Sec. IV-C. The parameter analyses are implemented in Sec. IV-D

### A. PERFORMANCE METRICS

In this study, we adopt two metrics to evaluate the performance of the proposed method. For the purpose of comparing the spreading speed for different methods, we define the spreading influence $F(t)$ at time $t$ in Eq. (10):

$$F(t) = \frac{n_I(t) + n_R(t)}{n} \tag{10}$$

where $n$ denotes the total number of nodes in a network, and $n_I(t)$ and $n_R(t)$ are the number of infected and recovered nodes at time $t$, respectively.

The second metric is the maximum spreading influence which is defined in Eq. (11):

$$MI = \frac{n_R}{n} \tag{11}$$

where $n_R$ denotes the total number of the recovered individuals when the spreading process reaches a stable state. A stable state is reached when all the infected nodes are recovered, and thereby there are only susceptible and recovered nodes.

### B. DATASETS

We adopt two synthetic networks and four real networks[1] to verify the validity of our proposed method. Concretely, $G_1$ and $G_2$ are random and scale-free networks generated by a network generator [39], respectively. The probabilities of a newly joined node connecting with an original node are 0.005 and 0.004 in two layers in $G_1$, respectively. In $G_2$, a new node establishes 4 and 3 connected edges with the original nodes in layers, respectively. $G_3$ and $G_4$ are part of the public database that archives and disseminates genetic and protein interaction data from humans and model organisms [40]–[42]. They concern HOMO and Drosophila, respectively. $G_5$ corresponds to the papers with the word "network" in the title or abstract of different arXiv categories up to May 2014. $G_6$ represents the different types of relationships data from twitter focused on People's Climate March in 2014 [43]. Without loss of generality, we only consider the duplex (i.e., double-layer) networks in this study. Specifically, $G_3$ and $G_4$ consist of two layers of the original networks (i.e., Direct_interaction and Physical_association). $G_5$ includes two layers called Physics.data-an and Math-ph. $G_6$ is made up by layers of RT and MT. We extracted part of the original real networks in a way that all nodes exist in both layers and the basic attributes of the six networks are shown in Table 1.

[1] https://comunelab.fbk.eu/data.php

**TABLE 1. Description of multiplex networks in our experiments. *N* and *E* denote the number of nodes and edges in networks, respectively. The number of layers is represented as *L*.**

| Networks | Name | $N$ | $E$ | $L$ | Type |
|---|---|---|---|---|---|
| $G_1$ | ER-ER | 4000 | 75863 | 2 | Random |
| $G_2$ | BA-BA | 5000 | 34971 | 2 | Scale-free |
| $G_3$ | HOMO | 1528 | 12064 | 2 | Genetic |
| $G_4$ | Drosophila | 1386 | 7452 | 2 | Genetic |
| $G_5$ | ArXiv | 3885 | 20047 | 2 | Coauthorship |
| $G_6$ | NYClimateMarch | 4150 | 40328 | 2 | Social |

### C. EXPERIMENTAL RESULTS

As mentioned above, centrality measures for multiplex networks have been widely studied. Here we adopt several methods proposed so far as competitors to evaluate the performance of OI-based method, i.e., the additive multiplex PageRank (addPR) [15], versatile PageRank (verPR) [18], cross-layer betweenness centrality (clyBC) [34], [36], and aggregation degree (aggDeg) [44]. In our experiments, the spreading rates in two layers are $\lambda_1 = 0.1$ and $\lambda_2 = 0.05$, respectively. The experiments are repeated over 100 times and are conducted in the same environment.

Figure 5 shows the proportion of infected and recovered nodes at each time step on six networks with $p = 0.2$ where $p$ represents the ratio of the number of source spreaders. Under the same condition, different methods perform diversely due to the differences of network structures while our proposed OI-based method is more efficacious. It can be seen that from the spreaders with a suppressed overlapping influence, information can spread faster and finally reach a larger scale. In some special cases, a few centrality measures may perform better in the spreading speed than other OI-based centrality method, like addDeg performs better than OI-clyBC. OI-verPR and OI-addPR in $G_6$. However, the OI-based methods combined with the corresponding centralities are still enhanced.

Actually, the spreading speed and maximum spreading influence are not only determined by the influence of

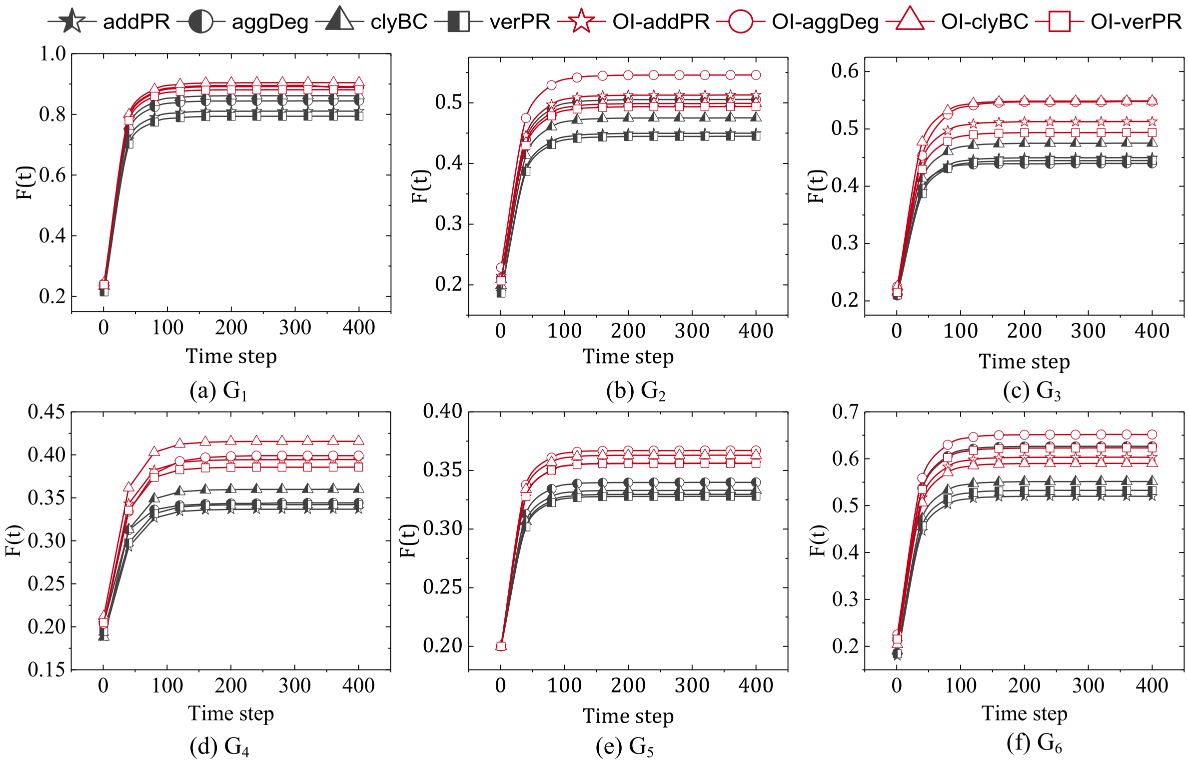



**FIGURE 5.** **The simulation results for spreading influence $F(t)$ at each time step on six networks under different methods. 20% users are selected as spreaders and the spreading rates in different layers are 0.1 and 0.05, respectively. The results are averaged over 100 independent runs. From these results, we can conclude that our OI-based method can spread information faster and reach a larger spreading scale than the corresponding algorithm. Besides, the OI-based methods basically outperforms than other benchmark algorithms.**

spreaders, but also by their relative location. For this reason, multiple users selected by our proposed method have high centralities and are relatively scattered where the effective spreading shorted path (ESSP) is introduced due to different infection rates in layers. It leads to a small overlapping influence and users have a strong ability to spread the information than others. Consequently, our proposed method outperforms a series of competitors.

### D. PARAMETER ANALYSIS

There are two important parameters in our proposed method which are (1) the size of spreaders ($p$), and (2) spreading rate ($\lambda$). This section mainly discusses the effect of these parameters on the OI-based method. The size of spreaders are set from 0.1 to 0.5 of the total users, with an interval of 0.05. In order to study the effect of the spreading rate on our OI-based method, we define the ratio of spreading rates in two layers as $r = \frac{\lambda_1}{\lambda_2}$. $r$ is set from 1.0 to 3.5 interval of 0.5 where $\lambda_1$ and $\lambda_2$ are initially set to 0.025 and 0.05, respectively. When a parameter is chosen for the parameter analyses, the remaining parameters are the same as Sec. IV-C and remain unchanged.

Figure 6 shows the maximum spreading influence when the spreading process get stable against $p$ ranging from 0.1 to 0.5. It is obvious that the maximum spreading influence of our proposed methods is generally larger than the multiplex centralities. More specifically, if $p$ is too small, information cannot be effectively spread no matter how to choose spreaders and the centrality measurements may even performs better than OI-based methods. According to the hypothesis of overlapping influences, when $p$ is small, the overlapping influence is weak and different methods have similar performance. However, when the size of spreaders is large, the overlapping influence is obviously significant and the influential users selected by our proposed method have a low overlapping influence. Consequently, OI-based methods will get wider *MI* than the original methods.

Without loss of generality, we adopt four real-world netoworks to study the effect of the spreading rate on OI-based method. Comparisons of the maximum spreading influence with a various $r$ among different methods are shown in Fig. 7. Results show that the OI-based methods can achieve a wider spreading influence *MI* especially when $r$ and spreading rate are large. Notably, different methods show a similar performance when the spreading rate is small and the maximum spreading influence of OI-based methods increase noticeably after $r = 1$. It is because that when $r = 1$, the spreading rate is the same so that the influence of the effective spreading shortest path (EESP) cannot be fully reflected.

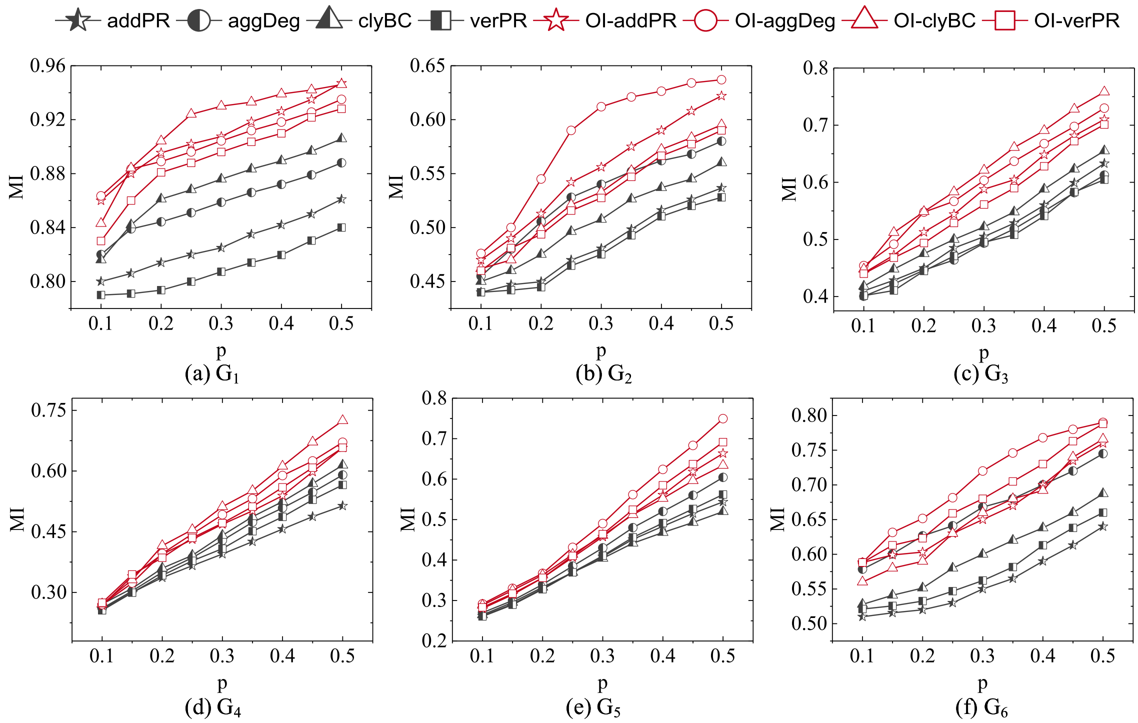



**FIGURE 6.** **The simulation results for the maximum spreading influence *MI* on six networks with different number of spreaders. The spreading rates in different layers are 0.1 and 0.05, respectively. The results are averaged over independent 100 runs. Results show that regardless of the size of spreaders, OI-based method performs better than other algorithms in terms of the maximum spreading influence. The advantages of the OI-based method are more prominent, especially when the proportion of initial spreaders is larger,.**

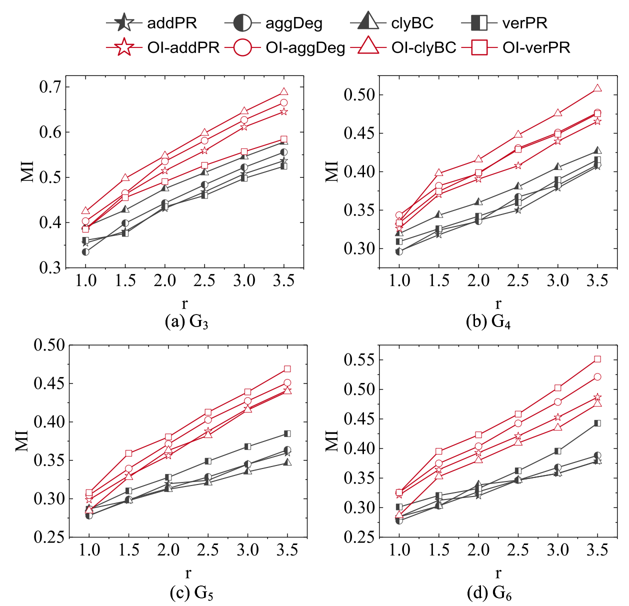



**FIGURE 7.** **The simulation results for the maximum spreading influence *MI* on four real-world networks under different methods. 20% users are selected as spreaders and the ratio *r* of spreading rates in two layers is set from 1.0 to 3.5, respectively. The results are averaged over 100 independent runs. Results show that in different spreading rate of each layer, OI-based method outperforms other algorithms in terms of the maximum spreading influence.**

Based on the above analyses, we can conclude that the parameters can make a tremendous difference in the performances of the OI-based methods. As the fraction of spreaders and the spreading rates get larger, OI-based methods perform better than other methods more significantly.

## V. CONCLUSION

Identifying multiple influential users rather than only one user can be widely applied to the influence maximization and the strategy of advertising. Therefore, how to minimize the overlapping influence among these selected users is a tough task. In this paper, a novel method combines the individual topology characteristics with the relative location is proposed to select multiple influential users in multiplex social networks. According to the proposed method, multiple users are of great collective influence when individuals have great centralities and the selected users are dispersed which leads to a suppressed overlapping influence and the least waste of resources. Therefore, we utilize the spreading rate to define the effective spreading shortest path in multiplex networks. The experimental evaluation of the proposed method is carried out against several competitors proposed so far for multiplex networks. Simulation results of six networks have shown that OI-based method outperforms other benchmark algorithms in the spreading influence. Future work will focus on the study of overlapping links and nodes in multiplex networks. The overlapping influence of the interactions in different layers among the same nodes and a series of applications based on the proposed effective spreading shortest path (ESSP) are what we concern in the next step.

## REFERENCES

[1] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Goncalves, A. Flammini, and F. M. Menczer, "Detecting and tracking political abuse in social media," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 297–304.

[2] C. Gao and J. Liu, "Network-based modeling for characterizing human collective behaviors during extreme events," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 1, pp. 171–183, Jan. 2017.

[3] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Inf. Fusion*, vol. 43, pp. 57–65, Sep. 2018.

[4] M. R. Subramani and B. Rajagopalan, "Knowledge-sharing and influence in online social networks via viral marketing," *Commun. ACM*, vol. 46, no. 12, pp. 300–307, 2003.

[5] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 1103–1108.

[6] C. Gao, L. Zhong, X. Li, Z. Zhang, and N. Shi, "Combination methods for identifying influential nodes in networks," *Int. J. Modern Phys. C*, vol. 26, no. 6, 2015, Art. no. 1550067.

[7] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, pp. 888–893, Aug. 2010.

[8] M. A. Al-Garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, G. Mujtaba, M. U. S. Khan, and S. U. Khan, "Analysis of online social network connections for identification of influential users: Survey and open research issues," *ACM Comput. Surv.*, vol. 51, no. 1, 2018, Art. no. 16.

[9] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. D'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, "Statistical physics of vaccination," *Phys. Rep.*, vol. 664, pp. 1–113, Dec. 2016.

[10] C. Gao, J. Liu, and N. Zhong, "Network immunization with distributed autonomy-oriented entities," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 7, pp. 1222–1229, Jul. 2011.

[11] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, 2006.

[12] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Phys. Rep.*, vol. 544, no. 1, pp. 1–122, 2014.

[13] Z. Wang, L. Wang, A. Szolnoki, and M. Perc, "Evolutionary games on multilayer networks: A colloquium," *Eur. J. Phys. B*, vol. 88, p. 124, May 2015.

[14] J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floría, "Evolution of cooperation in multiplex networks," *Sci. Rep.*, vol. 2, Aug. 2012, Art. no. 620.

[15] A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi, "Multiplex PageRank," *PLoS ONE*, vol. 8, no. 10, 2013, Art. no. e78293.

[16] M. K.-P. Ng, X. Li, and Y. Ye, "MultiRank: Co-ranking for objects and relations in multi-relational data," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1217–1225.

[17] J. Iacovacci, C. Rahmede, A. Arenas, and G. Bianconi, "Functional multiplex PageRank," *Europhys. Lett.*, vol. 116, no. 2, 2016, Art. no. 28004.

[18] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," *Nature Commun.*, vol. 6, Apr. 2015, Art. no. 6868.

[19] L. Solá, M. Romance, R. Criado, J. Flores, A. G. del Amo, and S. Boccaletti, "Eigenvector centrality of nodes in multiplex networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 23, no. 3, 2013, Art. no. 033131.

[20] C. Gao, Z. Su, J. Liu, and J. Kurths, "Even central users do not always drive information diffusion," *Commun. ACM*, vol. 62, no. 2, pp. 61–67, 2019.

[21] G. J. Baxter, G. Bianconi, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, "Correlated edge overlaps in multiplex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 94, no. 1, 2016, Art. no. 012303.

[22] D. T. Nguyen, H. Y. Zhang, S. Das, M. T. Thai, and T. N. Dinh, "Least cost influence in multiplex social networks: Model representation and analysis," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 567–576.

[23] M.-Y. Zhou, W.-M. Xiong, X.-Y. Wu, Y.-X. Zhang, and H. Liao, "Overlapping influence inspires the selection of multiple spreaders in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 508, pp. 76–83, Oct. 2018.

[24] K. Jung, W. Heo, and W. Chen, "IRIE: Scalable and robust influence maximization in social networks," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 918–923.

[25] Z.-L. Hu, J.-G. Liu, G.-Y. Yang, and Z.-M. Ren, "Effects of the distance among multiple spreaders on the spreading," *Europhys. Lett.*, vol. 106, no. 1, 2014, Art. no. 18002.

[26] D. Zhao, L. Li, S. Li, Y. Huo, and Y. Yang, "Identifying influential spreaders in interconnected networks," *Phys. Scripta*, vol. 89, no. 1, 2013, Art. no. 015203.

[27] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, no. 4, 2013, Art. no. 041022.

[28] L. Stone, R. Olinky, and A. Huppert, "Seasonal dynamics of recurrent epidemics," *Nature*, vol. 446, no. 7135, pp. 533–536, 2007.

[29] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Phys. Rep.*, vol. 650, pp. 1–63, Sep. 2016.

[30] C. Gao, J. Liu, and N. Zhong, "Network immunization and virus propagation in email networks: Experimental evaluation and analysis," *Knowl. Inf. Syst.*, vol. 27, no. 2, pp. 253–279, 2010.

[31] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 89, no. 3, 2014, Art. no. 032804.

[32] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, "Navigability of interconnected networks under random failures," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 23, pp. 8351–8356, 2014.

[33] A. Reiffers-Masson and V. Labatut, "Opinion-based centrality in multiplex networks: A convex optimization approach," *Netw. Sci.*, vol. 5, no. 2, pp. 213–234, 2017.

[34] T. Chakraborty and R. Narayanam, "Cross-layer betweenness centrality in multiplex networks with applications," in *Proc. IEEE 32nd Int. Conf. Data Eng.*, May 2016, pp. 397–408.

[35] M. Wu, S. He, Y. Zhang, J. Chen, Y. Sun, Y.-Y. Liu, J. Zhang, and H. V. Poor, "A tensor-based framework for studying eigenvector multicentrality in multilayer networks," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 31, pp. 15407–15413, 2019.

[36] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, "Centrality rankings in multiplex networks," in *Proc. ACM Conf. Web Sci.*, 2014, pp. 149–155.

[37] L. Guo, J.-H. Lin, Q. Guo, and J.-G. Liu, "Identifying multiple influential spreaders in term of the distance-based coloring," *Phys. Lett. A*, vol. 380, nos. 7–8, pp. 837–842, 2016.

[38] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 27823.

[39] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Sci.*, vol. 286, no. 5439, pp. 509–515, 2011.

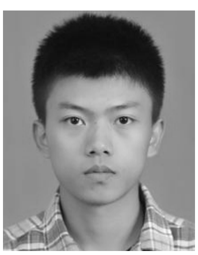
[40] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Commun.*, vol. 6, Apr. 2015, Art. no. 6864.

[41] M. De Domenico, M. A. Porter, and A. Arenas, "MuxViz: A tool for multilayer analysis and visualization of networks," *J. Complex Netw.*, vol. 3, no. 2, pp. 159–176, 2014.

[42] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: A general repository for interaction datasets," *Nucleic Acids Res.*, vol. 34, pp. D535–D539, Jan. 2006.

[43] E. Omodei, M. De Domenico, and A. Arenas, "Characterizing interactions in online social networks during exceptional events," *Frontiers Phys.*, vol. 3, p. 59, Aug. 2015.

[44] P. Bródka, K. Skibicki, P. Kazienko, and K. Musiał, "A degree centrality in multi-layered social network," in *Proc. Int. Conf. Comput. Aspects Social Netw.*, Oct. 2011, pp. 237–242.



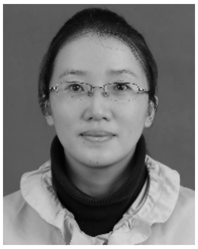
**JIANJUN CHEN** is currently pursuing the bachelor's degree with the College of Computer and Information Science, Southwest University, Chongqing, China. His research interests include complex networks and computational problem solving.

**YUE DENG** is currently pursuing the bachelor's degree with the College of Computer and Information Science, Southwest University, Chongqing, China. Her research interests include data-driven modeling and statistical analysis.

**ZHEN SU** received the master's degree from Southwest University, China, in 2009. He is currently pursuing the Ph.D. degree with the Potsdam Institute for Climate Impact Research, Humboldt University of Berlin, Germany. His research interests include complex networks and statistical analysis.

**SONGXIN WANG** received the Ph.D. degree from Jilin University, in 2002. He is currently an Associate Professor with the School of Information Management and Engineering, Shanghai University of Finance and Economics. His research interests include nature-inspired computing, complex network analysis, and knowledge representation.

**CHAO GAO** is currently a Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. He has been a Research Fellow with the Humboldt University of Berlin, Germany, and a Postdoctoral Research Fellow with the Computer Science Department, Hong Kong Baptist University. His current research interests include complex social networks analysis, nature-inspired computing, and data-driven complex systems modeling.



**XIANGHUA LI** is currently an Associate Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. Her research interests include nature-inspired computing and statistical analysis.