Article

MINE: Identifying Top-*k* Vital Nodes in Complex Networks via Maximum Influential Neighbors Expansion

Xinyu Huang, **Dongming Chen** *, **Dongqi Wang and Tao Ren**

Software College, Northeastern University, Shenyang 110169, China; neuhxy@163.com (X.H.);
wangdq@swc.neu.edu.cn (D.W.); rent@swc.neu.edu.cn (T.R.)
*   Correspondence: chendm@mail.neu.edu.cn

Abstract: Identifying vital nodes in complex networks is of paramount importance in understanding and controlling the spreading dynamics. Currently, this study is facing great challenges in dealing with big data in many real-life applications. With the deepening of the research, scholars began to realize that the analysis on traditional graph model is insufficient because many nodes in a multilayer network share connections among different layers. To address this problem both efficiently and effectively, a novel algorithm for identifying vital nodes in both monolayer and multilayer networks is proposed in this paper. Firstly, a node influence measure is employed to determine the initial leader of a local community. Subsequently, the community structures are revealed via the Maximum Influential Neighbors Expansion (MINE) strategy. Afterward, the communities are regarded as super-nodes for an iteratively folding process till convergence, in order to identify influencers hierarchically. Numerical experiments on 32 real-world datasets are conducted to verify the performance of the proposed algorithm, which shows superiority to the competitors. Furthermore, we apply the proposed algorithm in the graph of adjacencies derived from the maps of China and USA. The comparison and analysis of the identified provinces (or states) suggest that the proposed algorithm is feasible and reasonable on real-life applications.

Keywords: complex network; multilayer networks; node ranking; influence maximization

1. Introduction

The rapid development of information technology has witnessed a blossom of network science in the last decade. Network theory is an important tool for representing and analyzing the structures and functions of real-world complex systems such as social networks, power grids, transportation networks, economical networks, biological networks and so on [1,2]. For the enormous data scale, most studies merely focus on a small group of important nodes rather than the whole network, which has attracted great focus on the promising topic, i.e., vital nodes identification. Identifying vital nodes in complex networks is crucial in understanding and controlling the spreading dynamics, and it can be further split into two processes: (1) ranking of vital nodes and (2) identification of a small group of nodes that are able to maximize the influence. Similar research topics include node ranking [3], influence maximization [4], node centrality [5] and so on. Inherently, all these topics are focusing on finding vital nodes that exert maximum influence in the network. During the past decades, it has brought fruitful applications such as maximizing the influential scope in viral marketing [6], controlling the outbreak of epidemic diseases [7], preventing cascading failures in power grids [8] and so on [9].

We have entered the era of "big data", in which an obvious feature is that the data are being ubiquitously generated and continuously expanded [10]. Big data has brought us great challenges

in identifying the vital notes within complex networks, summarized as the following: (1) "Volume", the huge data scale has become prohibitive in tracking with the full data set, which may require the power of high-performance computing or distributed processing; (2) "Variety", many real-life complex systems are naturally heterogeneous networks, i.e., a more complicated structure that includes multiple interactions among various channels, which may lead to traditional research on a single channel insufficient to get integrated results; (3) "Velocity", the dynamic changes of networks force an efficient analysis to satisfy the real-time graph mining; (4) "Value", the big data implies great practical application value and requires an outstanding performance for the proposed algorithms.

With the in-depth study, more and more researchers come to realize that the traditional graph model is insufficient to evaluate node importance in many real-world scenarios [11], e.g., the individuals interact with each other across multiple social platforms, which shows the reasonability in modeling with a multiplex social network. The negligence of the multiple relationships between social actors may lead to incorrect or inaccurate results of the most versatile users [12]. Hence, the study on multilayer networks [13,14] has been introduced to find influential spreaders to get more accurate results. Specifically, many real-life complex systems are naturally hierarchical networks, which suggests the reasonability of graph mining with different levels.

In this paper, we design a novel algorithm for identifying top-*k* vital nodes in complex networks hierarchically. Firstly, a node influence measure is employed for ranking influential nodes that may be the initial leaders of local communities. Subsequently, the community structures are revealed via the Maximum Influential Neighbors Expansion (MINE) strategy. Afterward, the community structures are regarded as super-nodes for an iteratively folding process till convergence, in order to identify influencers hierarchically. The purpose of the proposed MINE algorithm is to accurately detect nodes that are able to exert strong influence over both the monolayer and multilayer networks. The main contributions of this paper are listed as follows:

(1) A novel algorithm for identifying vital nodes hierarchically is proposed in this paper, in which community structures derived by maximum influential neighbors expansion strategy are regarded as higher-level nodes.
(2) Based on the simulations on percolation theory and the epidemic model, 32 real-world datasets (16 monolayer networks and 16 multilayer networks) are employed to verify the performance of the proposed algorithm. Experimental results suggest its feasibility and superiority.
(3) The proposed algorithm is applied in the graph of adjacencies derived from the maps of China and USA, respectively, which exhibits the significance in real-life applications.

The rest of this paper is organized as follows. Section 2 summarized the related works, which is further classified into centrality measures, influence maximization in multilayer networks and application areas. Section 3 introduces the definition of the research problem and presents the mathematical models for both monolayer and multilayer networks, and subsequently, the algorithm process for identifying top-*k* vital nodes is introduced in detail. Section 4 exhibits the experiments and analysis, including comparison experiments on 32 real-world datasets, which verify the feasibility and veracity of the proposed algorithm. Moreover, the application in the graph of adjacencies derived from the maps of China and USA is analyzed. Section 5 summarizes the whole paper and provides concluding remarks.

2. Related Works

The identification of top-*k* vital nodes can be resolved into Influence Maximization Problem (IMP), which was first proposed by Domingos and Richardson [15], and later proved by Kempe et al. [16] that it was an NP-hard (i.e., non-deterministic polynomial-time hardness) problem [17]. A plethora of methods for IMP are proposed in the last decades [18,19], which can be mainly classified into centrality measures, link topological ranking measures, entropy measures, and node embedding measures [20,21]. We hereby introduce several typical representative methods in the following subsections.

2.1. Representative Centrality Measures

The simplest way to measure the node influence is by degree centrality [22] (DC), i.e., a focal node with more neighbors implies greater influence. However, several vital nodes, which are holding a lower degree but lying in a "bridge-like" position may exert enormous influence in the information diffusion process [23]. Analogously, closeness centrality [24] (CC) emphasizes the shortest path length of a focal node to all the others in the whole network, which suggests an optimal location is crucial for resource allocation in complex networks. However, both of the betweenness and closeness centralities are holding relatively high computational complexity, which becomes prohibitive in applying for large-scale networks. Some other classical measures such as eigenvector centrality [25] (EC), PageRank [26] (PR), H-index [27] and k-shell decomposition [28] are successful in many real-life scenarios, but are criticized by some certain limitations. Inspired by the gravity formula in physics, Ma et al. propose a gravity model [29] (GR) which considers the gravity between two nodes is proportional to the product of k-core, and inversely proportional to the power of the distance between them. Afterward, the node importance is measured by the gravity accumulation of the neighbors within three steps of a focal node. However, the computing of k-shell decomposition is time-consuming in large-scale networks. To improve the efficiency of GR, Li et al. propose a Local-Gravity [30] (LGR) model by replacing k-core with degree centrality and merely consider the neighbors within R steps. The parameter R can be further determined by half of the network diameter, approximately. Numerical experiments on fourteen datasets suggest its superiority, while the determination of parameter R requires calculating the shortest distance between all pairs of nodes, which also becomes prohibitive in large-scale networks [31]. Fan et al. [32] define two cycle-based node characteristics, i.e., cycle number and cycle ratio, which can be used to measure a node's importance. The research opens a new direction of understanding network structures, while the improvements in computational complexity are expected for further applications. INF [33] is a novel centrality measure that merely considers the local neighboring information of a focal node and can be further applied to multilayer networks. Extensive experiments on real-world datasets suggest its capability in identifying influencers in social networks. Several typical representative measures are summarized and compared, as shown in Table 1.

Table 1. A brief comparison of node centrality measures.

Metric	Equation	Topology	Complexity
DC [22]	$DC(i) = \sum_j^n a_{ij}$	Local	$O(n)$
BC [23]	$BC(i) = \sum_{s \neq i, s \neq t, i \neq t} \frac{g_{st}(i)}{g_{st}}$	Global	$O(nm + n^2 \log n)$
CC [24]	$CC(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}}$	Global	$O(nm + n^2 \log n)$
EC [25]	$EC(i) = k_1^{-1} \sum_j A_{ij} x_j$	Global	$O(n + m)$
PR [26]	$PR_k(i) = \sum_{j=1}^n a_{ji} \frac{PR_{k-1}(j)}{k_j^{out}} \ i = 1, 2, \dots, n.$	Global	$O(n + m)$
H-index [27]	$H(i) = \arg\max_{h \in \mathcal{N}} \{\forall \lvert \Gamma(j) \rvert \geq h, 1 \leq h \leq \lvert \Gamma(i) \rvert, j \in \Gamma(i)\}$	Semi-local	$O(n + m)$
k-core [28]	$k_s(i) = k$, s.t. $i \in S_{core}^k, i \notin S_{core}^{k+1}$	Global	$O(n + m)$
LR [34]	$LR_i = LR_i(t_c) + \frac{LR_g(t_c)}{n}$	Global	$O(n + m)$
GR, GR+ [29]	$G(i) = \sum_{j \in \psi_i} \frac{k_s(i) k_s(j)}{d_{ij}^2}$, $G_+(i) = \sum_{j \in \Gamma_i} G(j)$	Global	$O(n^3)$
LGR [30]	$LG_R(i) = \sum_{d_{ij} \leq R, j \neq i} \frac{k_i k_j}{d_{ij}^2}$	Semi-local	$O(n^2)$
INF [33]	$INF(i) = \sum_{j \in \Gamma(i)} \frac{w_{ij}}{k_j}$	Local	$O(nk)$

Note: n denotes the number of nodes, $\Gamma(i)$ denotes the neighbors of node i, k denotes the averaging degree, a_{ij} denotes the element at the i-th row and j-column of the adjacency matrix A; $\mathcal{N}$ denotes the set of natural numbers; d_{ij} is the shortest path length from node i to node j; g_{st} in BC (i.e., betweenness centrality) is the total number of shortest paths, $g_{st}(i)$ is the shortest path between s and t that passes through node i; in eigenvector centrality (EC), k_1^{-1} depicts the eigenvalue at the stable state; in PageRank (PR), $\sum_{i=1}^N PR_0(i) = 0$, k_j^{out} is the number of edges from node j to i; in k-core, S_{core}^k represents the node set after removing nodes with degree less or equal to k, likewise, S_{core}^{k+1} represents the node set after removing nodes with degree less or equal to $k + 1$; in LR (i.e., LeaderRank), $LR_i(t_c)$ denotes the score of node i at time t, $LR_g(t_c)$ denotes the score of the ground node at steady state; ψ_i in GR denotes the neighbors whose distance to node i is less than or equal to 3; R in Local-Gravity (LGR) denotes the truncation radius, i.e., only nodes within R steps are under consideration; w_{ij} in INF (i.e., the node influence) denotes the weight of edge (i, j) in a weighted network.

2.2. Influence Maximization in Multilayer Networks

Until now, great endeavors have been made to addressing the influence maximization problem in multilayer networks [35–39]. Wang et al. [12] propose a centrality measure (EDCPTD) to identify essential nodes based on tensor decomposition, which incorporates the intralayer edges and interlayer edges simultaneously to measure the node importance in multilayer networks. Tidke et al. [40] propose a consensus-based aggregation method to identify top-k influential nodes in online social networks, where a Twitter dataset covering two topics (Politics and Economy) is employed. The authors exploit heterogeneous surface learning features to rank the nodes with low computational complexity ($O(n)$). Li et al. [41] propose a novel approach to identify the influential nodes in network of networks (NON, a special case of multilayer network) [42], where the influence of each node in different layers is integrated by combination rules of evidence theory. Experimental results on China transport networks show its feasibility in identifying vital cities. Pedroche et al. [43] propose an eigenvector centrality for multiplex networks, which is adapted according to the idea behind the two-layer approach PageRank. Wang et al. [44] design an extended independent cascade (IC) model to simulate the influence diffusion process in multiplex networks, which aims at identifying influential nodes across all the network layers. Notably, modern complex systems tend to have more complicated structures than ever, which brings difficulties to the research on influence maximization problem via a traditional graph model. Hence, identifying vital nodes in multilayer networks is still an open issue.

Recently, hierarchical analysis on large-scale networks has received continuous focus in scientific society [45–47]. A typical k-shell method decomposes a network into hierarchically ordered shells by recursively pruning the nodes with degree being less than the current shell index. Ye et al. [48] apply this measure into time-varying networks, and propose a temporal k-shell decomposition method to identify vital nodes in temporal networks. Extensive experiments on real-world datasets suggest that the superiority of the proposed method. Chen et al. [49] propose a fusion index named as spreading influence related centrality to identify the influence of nodes by extracting and synthesizing topology feature information of traditional centrality indices and spreading influence. Zareie et al. [50] present a hierarchical method (HKS) for the specification of the topological locations of nodes. Experimental results on real-world and artificial datasets suggest that the proposed HKS method is superior to similar approaches and has an acceptable run time. In short, the above-mentioned approaches provide a hierarchical view on vital nodes identification, which suggests a promising field for further studies.

2.3. Application Areas

During the past decades, the identification of vital nodes in complex networks yields fruitful applications in real-life scenarios, such as viral marketing [6], epidemic outbreak [7], cascading failures [8], public safety networks [51] and so on.

Huang et al. [52] propose a community-based influence maximization method for viral marketing, which enables consumers to spread information about a product to other people, much in the same way that a virus spreads from one to another. Experimental results indicate that the community structures benefit in improving the performance of influence maximization. Hu et al. [53] devise a mitigation strategy against cascades in power grids. Different from vital nodes identification, in this work, the vulnerable transmission lines are identified and then strengthened to prevent cascades in power grids. The simulations on the IEEE-39 bus system and the IEEE 118 bus system suggest the feasibility of the proposed method in reducing the power blackout risks. Nikolaou et al. [54] propose an epidemiological model to identify critical airports for controlling global infectious disease outbreak through European airports' systems. Experimental results indicate the effectiveness of European airports and controlling measures should be taken in order to break the chain of infections. Fragkos et al. [55] introduce a novel evacuation-planning mechanism to support the distributed and autonomous evacuation process within the operation of a public safety system. Simulations under several scenarios exhibit the superiority of the proposed framework. In short, vital nodes identification

in complex networks is of paramount importance in a variety of applications and that's our purpose of this writing.

3. Model and Method

The influence maximization problem in complex networks can be roughly explained as finding a group of nodes (regarded as seeds) that can bring the maximum influence on the network dynamics, described as

$$A = \arg \min_{A \subset V} \max\{\sigma(A)\}, \tag{1}$$

where A is a subset of node set V that are identified as seeds, $\sigma(A)$ denotes the final influenced node set. This problem is simplified as top-k vital nodes identification by setting $|A| = k$, which has recently received great attention [40,56–58]. In this paper, both monolayer and multilayer network models are under consideration for addressing the influence maximization problem, as introduced in the following subsection.

3.1. Mathematical Models

The monolayer network model for vital nodes identification in this paper can be represented by

$$G = (V, E), \tag{2}$$

where V is the node set and $E \subseteq V \times V$ is the edge set. The multilayer network model [14] that we utilized in this paper is described as

$$\mathcal{M} = (\mathcal{G}, \mathcal{C}), \tag{3}$$

where $\mathcal{G} = \{G^\alpha; \alpha \in \{1, \ldots, L\}\}$ represents a series of monolayer networks with $G^\alpha = (V^\alpha, E^\alpha)$ depict the monolayer network at layer α, where V^α and E^α represent the nodes and intralayer edges at layer α, respectively. $\mathcal{C}$ is the interlayer edges between the nodes of any two different layers, described as

$$\mathcal{C} = \{E^{\alpha\beta} \subseteq V^\alpha \times V^\beta; \alpha, \beta \in 1, \ldots, L, \alpha \neq \beta\}, \tag{4}$$

where $E^{\alpha\beta}$ represents the interlayers edges between the nodes at layer α and layer β, V^α and V^β represent the node set at layer α and layer β, respectively. Specifically, we utilize the supra-adjacency matrix to model the a multilayer network, described as

$$\mathcal{M} = \begin{bmatrix} A_1 & I_{12} & \cdots & I_{1L} \\ I_{21} & A_2 & \cdots & I_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ I_{L1} & I_{L2} & \cdots & A_L \end{bmatrix} \in \mathcal{R}^{N \times N}, \tag{5}$$

where the diagonal elements $A_1, A_2, \ldots, A_L$ represent the adjacency matrix (i.e., intralayer edges) of layer $1, 2, \ldots, L$, respectively. The non-diagonal elements represent all interlayer edges, which can be represented by $\bigcup_{\alpha,\beta=1,\alpha\neq\beta}^{L} I_{\alpha\beta}$, where $I_{\alpha\beta}$ represents the interlayer edges between layer α and layer β. $\mathcal{R}$ is the set of real numbers, $N = \sum_{1 \leq \alpha \leq L} |V^\alpha|$ is the total number of the nodes.

We employ the Susceptible–Infected–Recovered (SIR) epidemic model [59] for analyzing influence maximization problem. It has three possible states:

- Susceptible (S) state, where a node is vulnerable to infection.
- Infectious (I) state, where a node is infected and tries to infect its susceptible neighbors.
- Recovered (R) state, where a node has recovered (or isolated) and can no longer infect others.

In a network, two nodes are considered to have "contact" if they are connected by an edge in the network. Within these two nodes, if one is infected with a certain epidemic, then the other is in "susceptible" state and maybe later infected [60]. A node is considered to be recovered if it is isolated or immune to the epidemic. In detail, to check the spreading ability of a focal node, we set this node as an infected seed to infect its neighbors. At each time step, an infected node can infect its neighbors with infection probability β, and then the infected nodes are recovered from the epidemic with probability γ. The differential equations are given as

$$\begin{cases} \frac{dS}{dt} = -\beta S I \\ \frac{dI}{dt} = \beta S I - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} . \tag{6}$$

3.2. Algorithm Description

We employ the INF indicator [33] to measure the initial leader of communities and then design a MINE algorithm to identify the vital nodes hierarchically. Suppose k is the number of nodes to be identified, the process of MINE is described as the following steps.

Step 1: Load the network G and calculate the INF indicator for each node;

Step 2: Divide communities by identifying influential nodes and neighbors;

    (a) Select the node i with maximum INF and mark it as the initial core of a community C_i;

    (b) Select a neighbor j that has the maximum INF among the neighbors of node i, and set j's community label as i if $C(i, j) > 0$ is satisfied according to Equation (7);

    (c) Repeat Step 2(b) until all the neighbors of node i are divided;

    (d) Repeat Step 2(a) to (c), assign the undivided nodes with community labels according to Equation (8) until all the nodes are marked with community labels;

Step 3: Identify vital nodes by folding communities hierarchically;

    (a) Regard a community C_i as a node i' in the new folded network G';

    (b) Update the INF of i' with the sum of members' INF in community C_i, as described in Equation (9), add an edge $(i'j')$ in G' if there are connections between community C_i and community C_j in G and set the weight of (i', j') according to Equation (10);

    (c) Repeat Step 3(a) and (b) until the size of the folded network is stable;

    (d) Traverse the nodes in the folded network and replace the INF of node i in the initial network G as the INF of node i' in the final folded network;

Step 4: Sort the nodes by influence in descending order and obtain the top-k nodes as the seeds.

The proposed MINE algorithm first identifies several influential nodes as the initial core of local communities, and then divides the neighbors of these core nodes into a reasonable community. To improve the reasonability of the division process, we define an evaluation function for the contribution of each neighbor j, given as

$$C(i, j) = \sum_{k \in \Gamma(j)} \delta(c_i, c_k), \tag{7}$$

where j is a neighbor of node i, $\Gamma(j)$ denotes the neighbors of node j, c_i and c_k denote the community labels of node i and node k, respectively. $\delta(c_i, c_k) = 1$ if i and k are with the same community label (i.e., $c_i = c_k$), and $\delta(c_i, c_k) = -1$ otherwise.

Inspired by the label propagation algorithm [61], we define a rule for an undivided node j to be with a community label as

$$c_j = \arg\max_{c_k} \sum_{k \in \Gamma(j)} |c_k|, \tag{8}$$

where c_j and c_k denote the community labels of node j and node k, respectively, $\Gamma(j)$ denotes the neighbors of node j, $|c_k|$ represents the amount of c_k.

In the folding process from network G to G', the INF of a node i' in G' is defined as the accumulated INF of members in the community in G, described as

$$INF(i') = \sum_{m \in C_i} INF(m), \tag{9}$$

where C_i denotes a community in network G, m is a member of the community C_i. The weight of an edge (i', j') in G' is defined as

$$w_{i'j'} = \sum_{i \in C_i, j \in C_j, C_i \neq C_j} w_{ij}, \tag{10}$$

where $w_{i'j'}$ represents the weight of an edge (i', j') in the folded network G', C_i and C_j represent two communities in network G, w_{ij} is the weight of an edge (i, j) in network G.

The pseudo-code of the above process is shown as Algorithm 1, with illustration on real-world datasets presented in Figure 1.

3.3. Complexity Analysis

Aiming at addressing the problem of influence maximization, the proposed MINE algorithm firstly ranks the nodes by employing the INF measure and identifies several top-ranking nodes as an initial leader of local communities. Subsequently, the communities are divided by the maximum influential nodes expansion strategy. The community structures are formed by numerous local communities which are further regarded as super-nodes in the folding process till convergence. Suppose m and n are the numbers of edges and nodes, respectively, the average degree of nodes is k. At each folding process, approximate n/k communities are obtained, the network size is shrinking exponentially, described as $n \to n/k$, and the number of edges is reducing as $m \to m - k \times n/k$, i.e., $m \to m - n$. Thus, the folding process with maximum influential neighbors expansion strategy can be converged quickly.

A step-by-step analysis on time complexity is listed as follows: In Step 1, the calculation of INF for all the nodes requires the complexity of $O(nk)$; In Step 2, the sort of influential nodes requires $O(n \log n)$, and ranking the neighbors of the initial core of local communities requires $O(k^2 \log k)$. The operation of contribution function (i.e., Equation (7)) for all the leaders' neighbors requires complexity less than $O((n - n/k) \times k)$, and the assignment of undivided nodes (i.e., Equation (8)) requires $O(k^2)$; In Step 3, the folding process converges quickly, which requires time complexity of $O(k^2 \log(n + k))$; In Step 4, the sorting of vital nodes requires $O(n \log n)$. Thus, the total time complexity of the proposed MINE algorithm is $O(2n(k + \log n) + k^2 \log k(1 + \log n) - n)$. As k is usually far less than n, therefore, the total time complexity can be simplified as $O(n \log n)$. Overall, the complexity of the proposed method is acceptable, which holds a lower computational complexity than BC or CC ($O(nm + n^2 \log n)$).

The storage of a network requires the space complexity of $O(n^2)$, the storage of node influence and community labels requires $O(2n)$. In the folding process, suppose the MINE algorithm converges in l times, this process requires extra $O(l * n^2)$ space complexity. As l is commonly far less than n, thereby the total space complexity for MINE algorithm can be simplified as $O(n^2)$.

(**a**) Two communities in Karate club network.

(**b**) Top-two influencers in the folded Karate club network.

(**c**) Communities in the Enron email network.

(**d**) Top-five influencers in the folded Enron email network.

(**e**) Communities in Football team network.

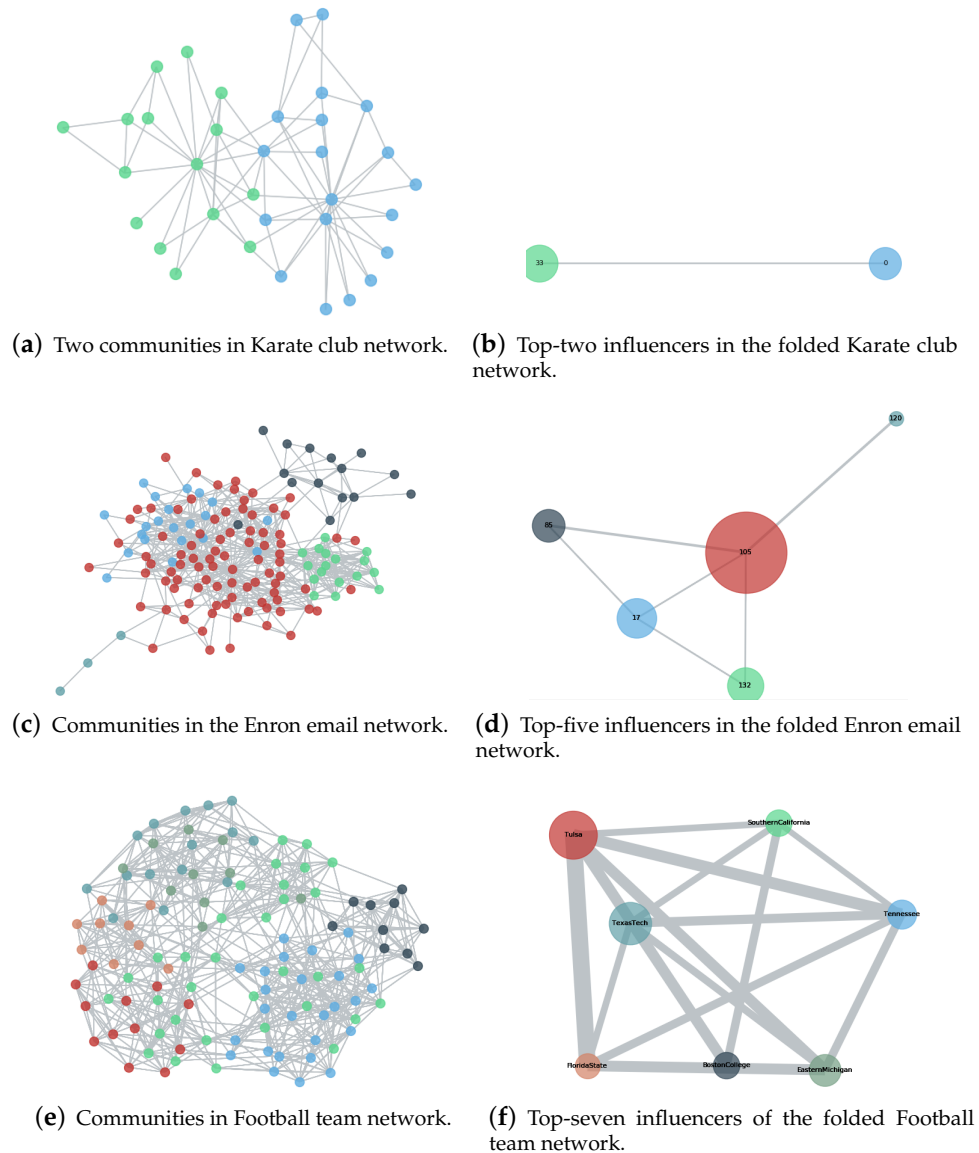(**f**) Top-seven influencers of the folded Football team network.

Figure 1. Illustration of the proposed Maximum Influential Neighbors Expansion (MINE) algorithm. Communities in three real-world networks (i.e., Karate Club network [62], Enron email network [63] and College Football Team network [64]) are presented in the left three panels the and top-ranking nodes are highlighted in the right three panels, which suggests that the MINE algorithm aims to identifying vital nodes hierarchically.

Algorithm 1 MINE

   Input: A network G, the number of seeds k;
   Output: A dict S that contains k items, where the key of each item is a node and the value is the node's
            influence;

1   ... ; // Load nodes V, edges E from G; # **Step 1**
2   **for** i **in** V **do**
3   // Calculate the influence for each node
4      $i.group$ = **null**; // Initilize each node with no community label
5   **for** j **in** $i.neighbors$ **do**
6          $i.inf$ += $\gamma w_{ij}/k_j$; // $\gamma = 1$ for monolayer networks, while for multilayer networks, γ is
                determined by the edge type [33]

7   I = sort(V, key=inf)$\downarrow$; // Sorted V by node influence in descending order; # **Step 2**
8   **for** i **in** I **do**
9   **if** $i.group$ == **null then**
10          $i.group$ = i; // Mark the top influential nodes as the initial core of a community
11          $i.neighbors$ = sort($i.neighbors$, key=inf)$\downarrow$;
12   **for** j **in** $i.neighbors$ **do**
13   // Divide j to i's community if j is not divided and $C(i,j) > 0$ is satisfied according to
                Equation (7)
14   **if** $j.group$ == **null and** $C(i,j) > 0$ **then**
15                  $j.group$ = $i.group$;

16   **for** i **in** I **do**
17   **if** $i.group$ == **null then**
18          $labels$ = {}; // A temp variable to store the labels of neighbors
19   **for** j **in** $i.neighbors$ **do**
20   **if** j **in** $labels$ **then**
21                  $labels[j]$ += 1;
22   **else**
23                  $labels[j]$ = 1;
24          $i.group$ = $\arg\max_{c \in labels} labels[c]$;

25   C = getCurrentCommunity(G); // Get current community by $node.group$
26   $V^l = V$; $E^l = E$; // # **Step 3**
27   **for** c **in** C **do**
28   // Fold the network by community structures
29      $V^{l+1}.add\_node(c, inf = \sum_{m \in c} m.inf)$;
30      $c.core.inf = c.inf$; // Update the influence of nodes in V
31   **for** c' **in** $c.neighbors$ **do**
32          $E^{l+1}.add\_edge((c,c'), w_{cc'} = \sum_{i \in c, j \in c'} w_{ij})$; // Update the wieght of edges between c and c'

33   $G^{l+1} = (V^{l+1}, E^{l+1})$; // Generate a folded network
34   **while** $\textbf{len}(V^{l+1}) < \textbf{len}(V^l)$ **do**
35   ... ; // Divide communities according to the processes in Step 2 and folding process in Step 3
36   l + +;
37      $G^{l+1} = (V^{l+1}, E^{l+1})$; // Re-generate a folded network based on the community structures
38   **if** $V^{l+1} == V^l$ **then**
39   break;

40   S = sort(V, key=inf)$\downarrow[:k]$; // Obtain top-k influential nodes; # **Step 4**
41   **return** S;

---

4. Experiments and Analysis

The proposed algorithm is implemented in Python 3.8.3, with relevant libraries NetworkX 2.4 and MultinetX. The operating system is 64 bit Windows 10, with configuration of Intel (R) Core (TM) i7-7700 CPU @ 3.60 GHz (8 CPUs), 3.6 GHz and 16 GB ram installed. The goal of the experiments is to compare the performance of the proposed MINE algorithm with competitors.

4.1. Experimental Datasets

In this paper, 32 real-world datasets are employed to verify the performance of the proposed algorithm, which are classified into two groups: monolayer networks and multilayer networks. The statistics of the two groups of datasets are shown in Tables 2 and 3, respectively. All the datasets are available at https://www.neusncp.com/user/file?id=311&code=mine.

Table 2. Statistics of 16 monolayer network datasets that are utilized in this paper.

Dataset Name	$|V|$	$|E|$	$<k>$	$<d>$	$<C>$	r	$|H|$	β_c
Club [62]	34	78	4.5882	2.4082	0.5706	–0.4756	1.6933	0.1477
Dolphins [65]	62	159	5.1290	3.3570	0.2590	–0.0436	1.3268	0.1723
911 [66]	69	159	4.6087	2.4672	0.4698	–0.0380	1.7304	0.1434
Lesmis [67]	77	254	6.5974	2.6411	0.5731	–0.1652	1.8273	0.0905
Voles [63]	89	143	3.2135	0.5610	0.6083	0.2877	1.3340	0.3043
Escherichia [68]	97	212	4.3711	5.5369	0.3675	0.4116	1.2367	0.2270
Polbooks [69]	105	441	8.4000	3.0788	0.4875	–0.1279	1.4207	0.0915
Football [64]	115	613	10.6609	2.5082	0.4032	0.1624	1.0069	0.1027
Enron [63]	143	623	8.7133	2.9670	0.4339	–0.0195	1.4829	0.0839
NEUSNCP [70]	167	465	5.5689	2.8810	0.3122	–0.3619	3.3624	0.0564
Jazz [71]	198	2742	27.6970	2.2350	0.6175	0.0202	1.3951	0.0266
USAir [72]	332	2126	12.8072	2.7381	0.6252	–0.2079	3.4639	0.0231
NS [73]	379	914	4.8232	6.0419	0.7412	–0.0817	1.6630	0.1424
DMLC [74]	659	1570	4.7648	2.6370	0.3279	–0.1914	14.8897	0.0143
Game of Thrones [75]	796	2823	7.0930	3.4162	0.4859	–0.1154	4.1941	0.0348
Power [69]	4941	6594	2.6691	18.9892	0.0801	0.0035	1.4504	0.3483

Note: $|V|$ and $|E|$ depict the number of nodes and edges, respectively. $<k>$ denotes the average degree; $<d>$ is the average shortest path; $<C>$ is the average clustering index; r is the assortativity coefficient; $|H|$ is the degree heterogeneity and β_c represents the epidemic threshold of the Susceptible-Infected-Recovered (SIR) model; Club represents a graph constructing from W.W. Zachary's investigation on the karate club of a university in USA in 1970-72, which contains 34 nodes and 78 edges based on the observation on the relationship of club members; Dolphins contains an undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand, as compiled by Lusseau et al. in 2003; 911 represents the social relationships of terrorists at Hamburg Cell (which started the planning for 9/11 and ultimately participated in the attack); Lesmis is the coappearance network of characters in the novel "Les Miserables"; Voles is an animal interaction network where the nodes represent voles and there will be an edge between two voles if they have ever been caught in at least one common trap; Escherichia is transcriptional regulation networks in cells orchestrate gene expression, where nodes represent operons, and each edge is directed from an operon that encodes a transcription factor to an operon that it directly regulates; Polbooks is a network of books about US politics published around the time of the 2004 presidential election and sold by Amazon, where nodes represent books and there will be an edge between two books if they are frequent co-purchased by the same buyers; Football describes the network of American football games between Division IA colleges during regular season Fall 2000; Enron is an email network collected from Enron company, where nodes represents; NEUSNCP collects the relationships of members in a social platform; Jazz represents the network derived from the collaboration patterns of jazz musicians; USAir is an undirected, weighted network representing US airports, where the nodes represent airports and the edges represent air travel connections among them; NS represents the largest component of coauthorships between scientists working on network theory and experiment; Game of Thrones represents the relationships of characters in the first five volumes of the novel "A Song of Ice and Fire"; DMLC represents the inferred links by small/medium-scale protein-protein interactions; Power is an undirected, unweighted network representing the topology of the Western States Power Grid of the USA.

Table 3. Statistics of 16 multilayer datasets that are utilized in this paper.

Dataset Name	$|L|$	$|V|$	$|E|$	$|E_A|$	$|E_C|$	$<k>$	$<d>$	$<C>$	β_c
Padgett [76]	2	26	46	35	11	3.5385	2.6923	1.1550	0.3239
HepatitusCVirus [77]	3	129	152	125	27	2.3566	2.3582	12.4008	0.0354
Oryctolagus [77]	3	151	145	138	7	1.9205	0.5642	11.7317	0.0464
LondonTransport [78]	3	399	472	441	31	2.3659	14.2989	1.1597	0.5735
CKM [79]	3	674	2005	1370	635	5.9496	1.1677	1.1879	0.1648
C.Elegans [80]	3	791	3857	3108	749	9.7522	3.5346	1.6895	0.0646
Plasmodium [81]	3	1206	2489	2486	3	4.1277	3.9403	2.9832	0.0884
HumanHerpes4 [77]	4	261	299	241	58	2.2912	2.9561	18.5515	0.0241
BOS [81]	4	369	360	311	49	1.9512	0.4116	2.0429	0.3349
DanioRerio [77]	5	180	206	173	33	2.2889	0.3325	1.6161	0.3705
Xenopus [77]	5	582	710	568	142	2.4399	2.8018	2.0464	0.2504
Gallus [81]	6	367	411	348	63	2.2398	1.277	12.5034	0.0370
Rattus [81]	6	3263	4670	3956	714	2.8624	3.9011	30.0001	0.0118
Candida [81]	7	418	446	386	60	2.1340	2.1034	28.0935	0.0170
PierreAuger [82]	16	965	8022	7153	869	16.6259	4.8105	2.8617	0.0215
EUAirTransportation [83]	37	2034	15199	3588	11611	14.9449	3.5087	1.6187	0.0431

Note: $|L|$ represents the number of layers; $|V|$ and $|E|$ are the number of nodes and edges, respectively; $|E_A|$ and $|E_C|$ are the number of intralayer edges and interlayer edges, respectively; $<k>$ depicts the average degree; $<d>$ depicts the average shortest path; $<C>$ represents the average clustering coefficient and β_c represents the epidemic threshold of the SIR model; Padgett is a two-layered network consists of 2 layers (marriage alliances and business relationships) describing florentine families in the Renaissance; HepatitusCVirus represents the multiplex genetic and protein interactions network of the *Hepatitus C virus*; Oryctolagus represents the multiplex genetic and protein interactions network of the *Oryctolagus Cuniculus*, also known as European rabbit; LondonTransport is collected from the official website of Transport for London (https://www.tfl.gov.uk/), where the nodes represent train stations in London and edges represent existing routes between stations; CKM is collected by Coleman, Katz and Menzel (CKM, for short) on medical innovation, considering physicians in four towns in Illinois, Peoria, Bloomington, Quincy and Galesburg; C.Elegans represents the edges of the metabolic network of *Caenorhabditis elegans*; Plasmodium represents the protein interactions network of the *Plasmodium Falciparum*; HumanHerpes4 represents the multiplex genetic and protein interactions network of the *Epstein-Barr virus*, i.e., human herpes-virus 4 (HHV-4); BOS represents the multiplex genetic and protein interactions network of the *Bos Linnaeus*; DanioRerio represents the multiplex genetic and protein interactions network of the *Danio Rerio (zebrafish)*, a tropical freshwater fish, composed of the following five layers: association, suppressive genetic interaction defined by inequality, direct interaction, additive genetic interaction defined by inequality and physical association; Xenopus represents the multiplex genetic and protein interactions network of the *Xenopus Laevis*, a species of African aquatic frog of the Pipidae family; Gallus represents the interactions among *Gullus* with six layers: direct interaction, physical association, synthetic genetic interaction defined by inequality colocalization, colocalization, additive genetic interaction defined by inequality, and association; Rattus and Candida represent the multiplex genetic and protein interactions network of the *Rattus Norvegicus* and *Candida Albicans*, respectively; PierreAuger represents the multiplex coauthorship network in the internal report repository of the Pierre Auger Collaboration (http://www.auger.org/), the largest team of scientists working about ultra-high energy cosmic rays; EUAirTransportation represents the air transportation network in Europe composed of thirty-seven different layers, where each one is corresponding to a different airline operating.

4.2. Experimental Contents

To verify the performance of the proposed MINE algorithm, comparison experiments are conducted on the above-mentioned monolayer and multilayer networks, respectively. Firstly, a sequence of ranked nodes is calculated by each measure. Subsequently, the vital nodes are removed in turn by the ranking in each sequence, which may result in the network being broken into several subgraphs. We record the number of subgraphs at each removal and make a comparison on all the competitors. This process repeats until all the nodes are removed. Hence, the variation tendency of the subgraphs' number exhibits the influence of a focal centrality measure. The experimental results are plotted in Figures 2 and 3.
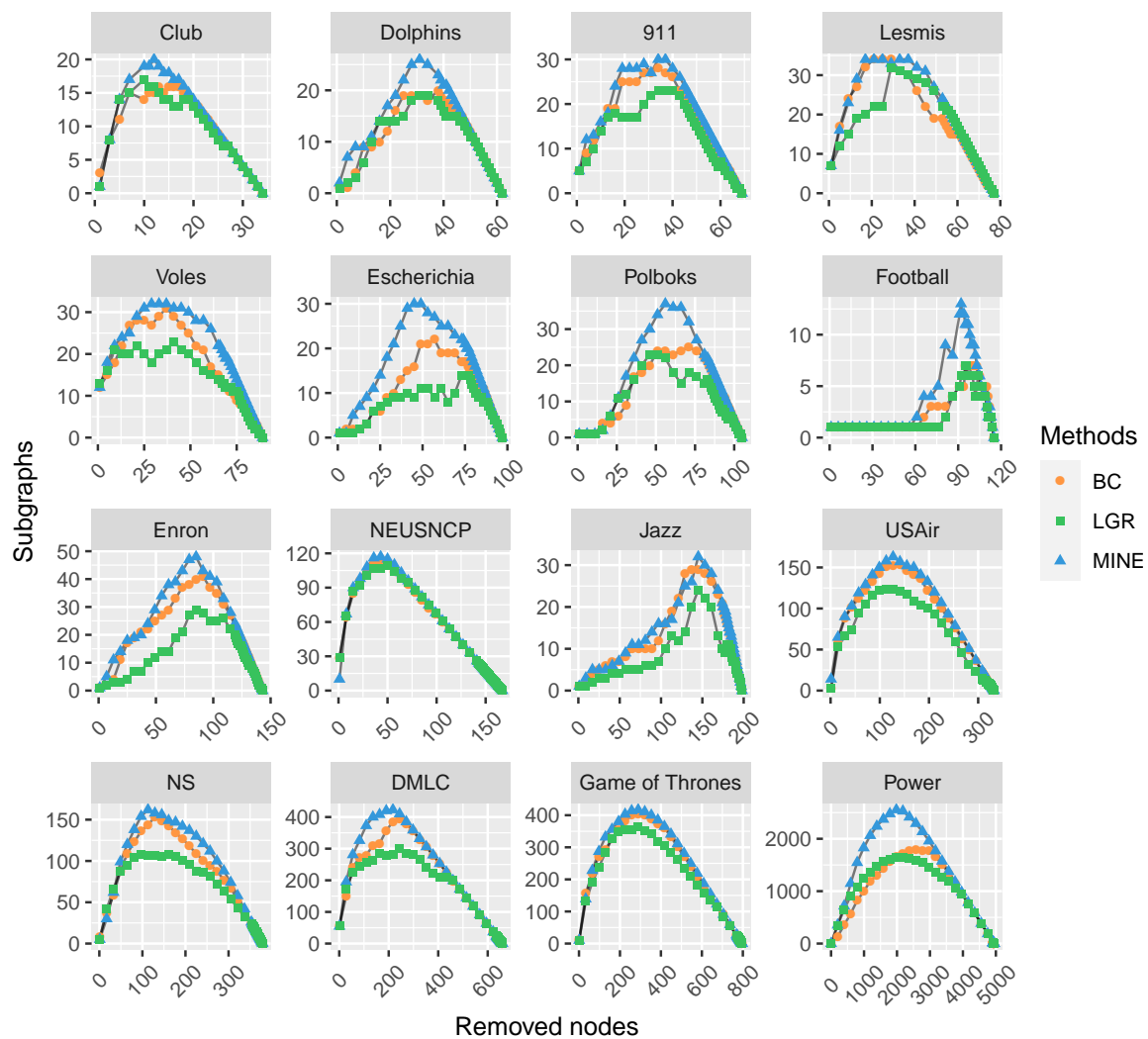
Figure 2. The comparison of varying subgraphs with the removal of the most influential nodes repeatedly with each centrality method on 16 monolayer networks. The proposed MINE algorithm (marked with blue triangles) shows superiority to the BC and LGR methods by obtaining more subgraphs in the removal process.

As shown in Figures 2 and 3, with the removal of vital nodes (in descending order) computed by each measure, the number of derived subgraphs is increasing and reaches a maximum when the network is totally broken up (i.e., there are no edges). Afterward, the number of subgraphs is reducing, and finally reaches zero when all the nodes are removed. Hence, we compare the maximum numbers of subgraphs obtained by each measure, and find that the proposed MINE algorithm (as marked with blue triangles) is able to get greater values than the competitors on all the datasets except HepatitusCVirus, which suggests the superiority of MINE to the competitors. Furthermore, the maximum number of subgraphs by the proposed MINE algorithm on the HepatitusCVirus dataset is very close to the optimal result (i.e., obtained by LGR, as marked with green squares) and following the same trend, which suggests the feasibility of the proposed MINE algorithm.
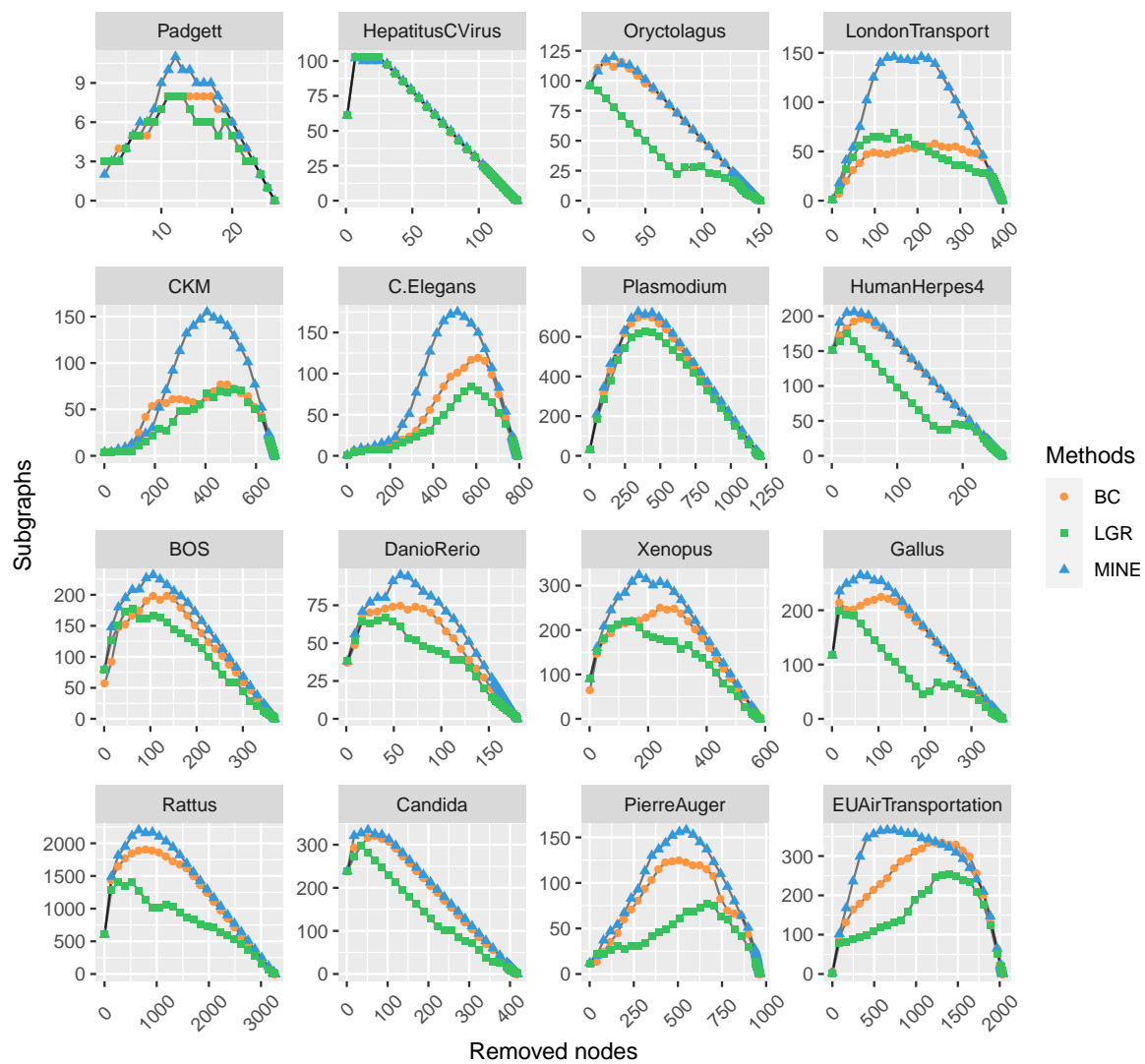
Figure 3. The comparison of varying subgraphs with the removal of the most influential nodes repeatedly with each centrality method on 16 multilayer networks. The proposed MINE algorithm (marked with blue triangles) shows superiority to the BC and LGR methods on the majority of datasets except for HepatitusCVirus, while it is close to the performance of the LGR method (marked with green squares).

We apply the SIR model to compare the performance of competitive measures for further study on the influence maximization problem. Initially, k nodes are set as infected seeds to infect neighboring nodes with probability β. Afterward, the infected nodes are recovered and never be infected again with probability γ. This spreading process repeats until there are no more infected nodes in the network and the number of final recovered nodes are utilized to evaluate the effect of a focal measure. To analyze the varying parameter k with the recovered nodes, we conducted experiments on the above-mentioned 32 datasets and set the ratio as $\beta = \beta_c$ and $\gamma = 1$. Specifically, the range of k is set in $[0, 0.2N]$ according to Pareto Distributions [84], where N is the total number of nodes for each dataset. We implement 1000 independent runs to eliminate accident errors and the comparison results are plotted in Figures 4 and 5.
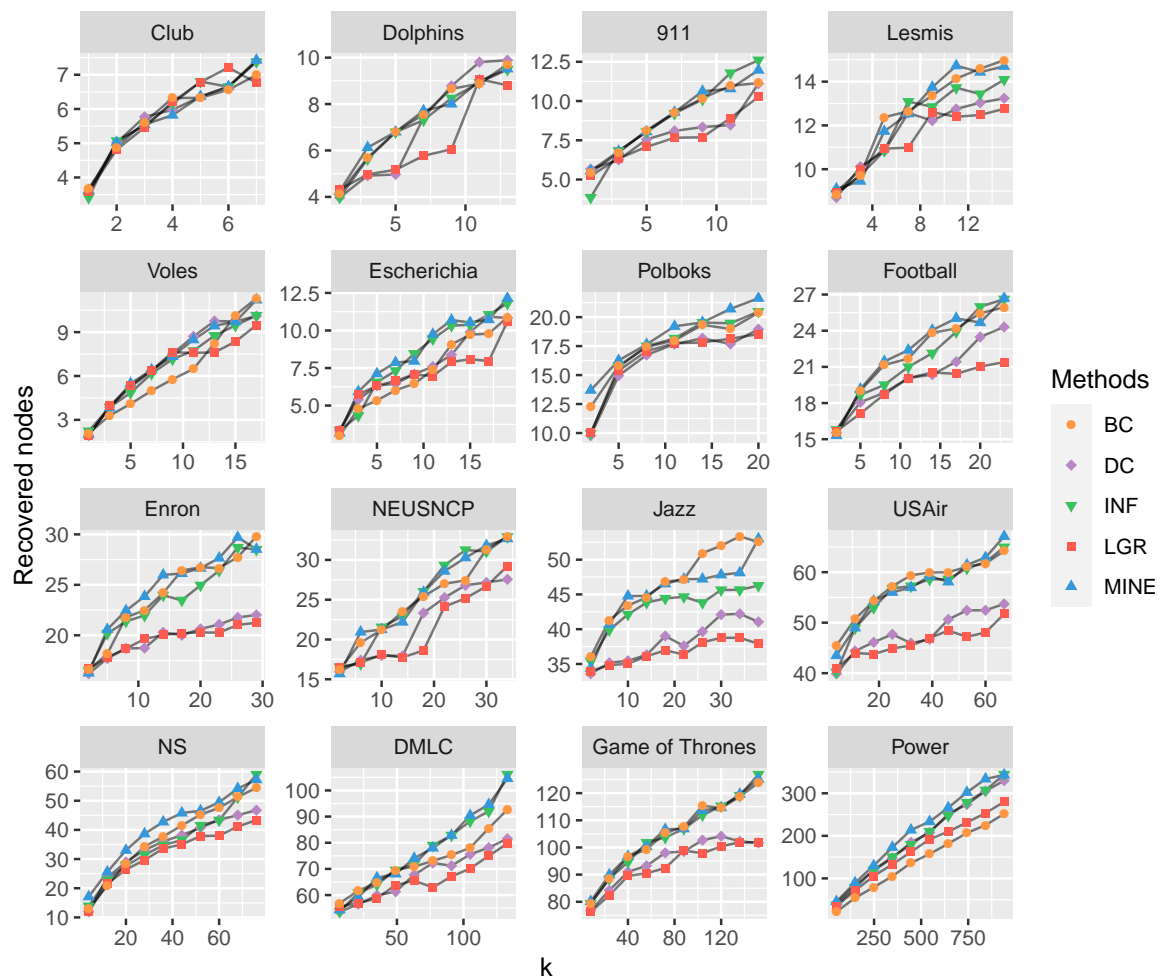
Figure 4. The comparison of recovered nodes with varying *k* (from 0 to 20 percent of total nodes) on 16 monolayer networks. The proposed MINE algorithm (marked with blue triangles) shows superiority to the competitors on Polbooks, Enron, NS, DMLC, Game of Thrones and Power datasets. While on the Jazz dataset, the MINE algorithm is second to BC (marked with orange circles) when *k* is in the range of [26, 34], while MINE achieves more recovered nodes (averaged 52.9888) than that of BC (averaged 52.5331) when *k* is set as 38 (i.e., twenty percent of the number of nodes in the Jazz dataset).

The performance of each measure is evaluated by the height of each curve. As shown in Figures 4 and 5, with the increasing of *k*, the recovered nodes are increasing likewise. In most cases, the proposed MINE algorithm shows superiority to the competitors, by obtaining relatively more recovered nodes with the increasing of "infected" nodes.

To compare the efficiency of the proposed MINE algorithm with competitors, we record the execution time (in seconds) for each measure on the 32 datasets, respectively, and make a brief comparison as presented in Table 4.

As shown in Table 4, DC holds the minimum runtime on almost all the datasets and BC holds maximum runtime in the majority of the experimental datasets. To compare the total runtime comprehensively, we accumulated the runtime of each measure on all the datasets, and make an intuitional comparison as shown in Figure 6.
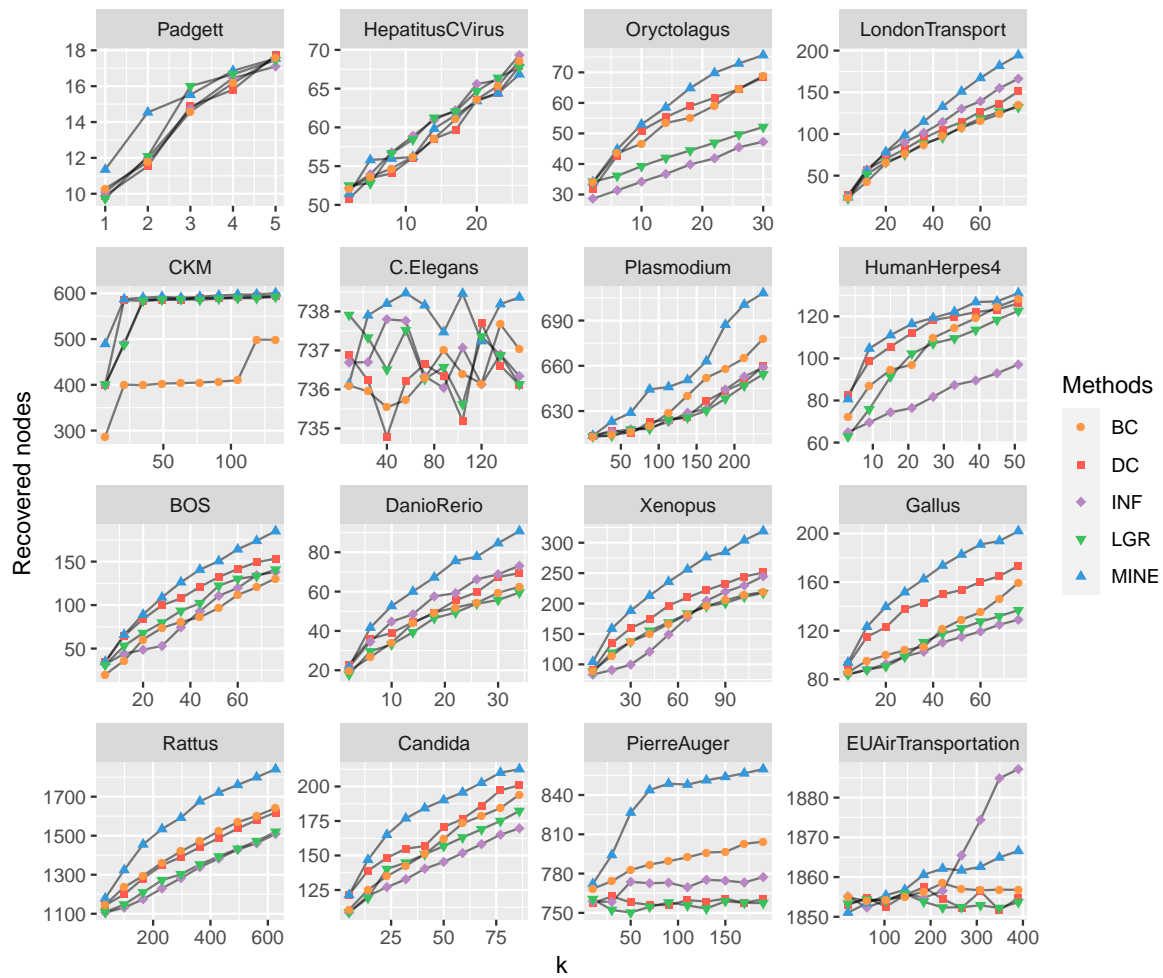
Figure 5. The comparison of recovered nodes with varying *k* (from 0 to 20 percent of total nodes) on 16 multilayer networks. The proposed MINE algorithm (marked with blue triangles) shows superiority to the competitors on BOS, Candida, C.Elegans, DanioRerio, Gallus, HumanHerpes4, LondonTransport, Oryctolagus, Plasmodium, PierreAuger, Rattus and Xenopus datasets, while on the EUAirTransportation dataset, the MINE algorithm is second to INF (marked with purple diamonds).
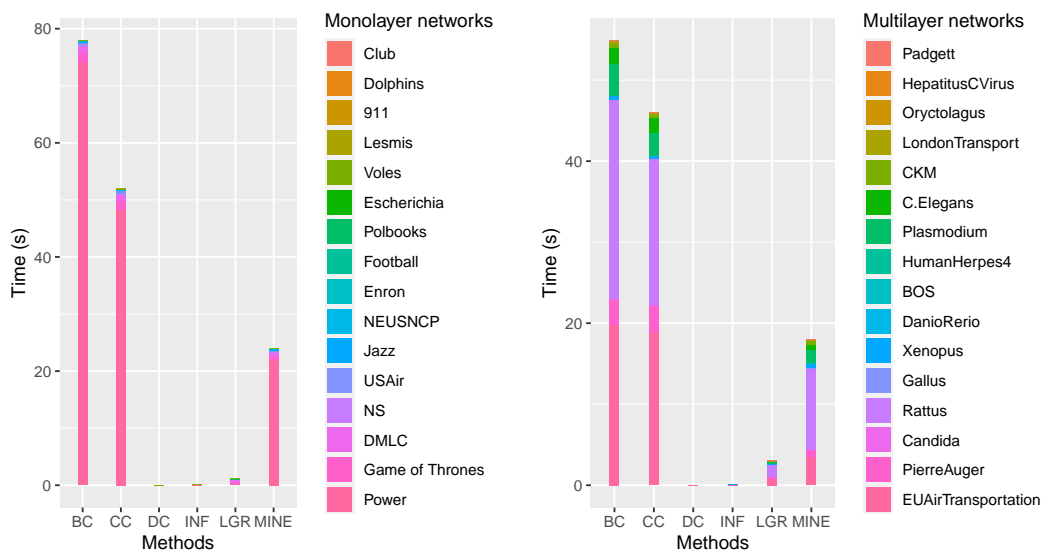

Figure 6. Comparisons of accumulated execution time by performing the competitive measures on the grouped datasets.

Table 4. Runtime comparison of DC, BC, CC, LGR, INF and MINE on 32 real-world datasets.

Network	DC	BC	CC	LGR	INF	MINE
Club	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Dolphins	<0.0001	<0.0001	**0.0156**	**0.0156**	<0.0001	**0.0156**
911	<0.0001	**0.0156**	<0.0001	<0.0001	<0.0001	**0.0156**
Lesmis	<0.0001	**0.0156**	**0.0156**	<0.0001	<0.0001	**0.0156**
Voles	<0.0001	<0.0001	<0.0001	<0.0001	**0.0156**	**0.0156**
Escherichia	<0.0001	**0.0312**	0.0156	<0.0001	<0.0001	0.0156
Polbooks	<0.0001	**0.0469**	0.0312	<0.0001	0.0156	0.0156
Football	<0.0001	**0.0469**	**0.0469**	<0.0001	0.0312	0.0312
Enron	<0.0001	**0.0625**	0.0469	<0.0001	0.0156	0.0312
NEUSNCP	<0.0001	**0.0938**	0.0625	<0.0001	0.0312	0.0156
Jazz	<0.0001	0.1719	0.2188	<0.0001	0.0781	**0.3750**
USAir	<0.0001	0.3594	**0.4531**	0.0156	0.0938	0.2188
NS	<0.0001	**0.3438**	**0.3438**	0.0156	0.0156	0.0625
DMLC	<0.0001	**1.0312**	0.9219	<0.0001	0.4219	0.7656
Game of Thrones	<0.0001	**1.8750**	1.5312	0.0156	0.2344	0.4219
Power	<0.0001	**73.8594**	48.2344	0.0156	0.1562	21.9531
Padgett	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
HepatitusCVirus	<0.0001	**0.0312**	**0.0312**	<0.0001	0.0156	0.0156
Oryctolagus	<0.0001	**0.0156**	<0.0001	<0.0001	**0.0156**	**0.0156**
LondonTransport	<0.0001	**0.4219**	0.2969	<0.0001	<0.0001	0.2031
CKM	<0.0001	**0.4844**	0.3281	<0.0001	0.0469	0.4219
C.Elegans	<0.0001	**2.0000**	1.9219	0.0156	0.1406	0.7500
Plasmodium	<0.0001	**3.8281**	2.7812	<0.0001	0.1562	1.5625
HumanHerpes4	<0.0001	**0.0938**	0.0781	<0.0001	0.0469	0.0312
BOS	<0.0001	0.0469	0.0156	<0.0001	0.0156	**0.1094**
DanioRerio	<0.0001	0.0156	0.0156	<0.0001	<0.0001	**0.0312**
Xenopus	<0.0001	0.3750	0.2969	<0.0001	0.0156	**0.4375**
Gallus	<0.0001	**0.1094**	0.0938	<0.0001	0.0469	**0.1094**
Rattus	0.0156	**24.4375**	17.9062	<0.0001	1.5000	9.8750
Candida	<0.0001	**0.2500**	0.1719	<0.0001	0.0938	0.1250
PierreAuger	<0.0001	**3.0312**	3.3438	0.0156	0.1719	0.9375
EUAirTransportation	<0.0001	**19.7656**	18.7344	0.0312	0.7500	3.4062

Note: The maximum values of each row are in bold. <0.0001 in the cell represents the runtime (recorded in seconds) is less than 0.0001 s. With the increasing of data size, the runtime of BC and CC are increasing dramatically, while the proposed MINE algorithm is acceptable.

As shown in Figure 6, the left panel exhibits the accumulated execution time by performing DC, CC, DC, INF, LGR and MINE measure on 16 monolayer networks, and analogously, the execution time on 16 multilayer networks is presented in the right panel. The accumulated execution time for the competitive measures can be roughly sorted as DC, INF, LGR, MINE, CC and BC (in ascending order). Although inferior to DC, INF and LGR measures on the accumulated execution time, the proposed MINE algorithm is acceptable for common applications with its performance in influence maximization and superiority to BC and CC on efficiency.

4.3. Discussion

Influence Maximization is a research hotspot of complex networks with a variety of applications in very different fields, such as information diffusion, social network analysis, viral marketing, epidemic control and so on. Degree centrality [22], as the simplest measure, mainly considers a focal node is crucial if it has large amounts of neighbors. However, it is probably to ignore some "bridge-like" nodes (i.e., nodes that connect different components) if they have few neighbors. To overcome this shortcoming, betweenness [23] and closeness [24] centralities are proposed which are able to identify vital nodes from a global perspective. These two measures seem to be more reasonable, however, they are also criticized by high computational complexity, which become prohibitive in large-scale networks. Local gravity [30] is a balanced measure, however, the determination of parameter *R*

requires computing network diameter, which is also a time-consuming process. INF performs well in ranking nodes but lacks nodes selecting strategy, which is insufficient to exert maximum influence. Thus, the MINE algorithm is proposed in this paper, which firstly computes the local neighboring information of a focal node, and then considers the community structures hierarchically by a folding process. The proposed MINE algorithm is capable of identifying vital nodes dispersedly and exerting strong influence. Experimental results on 32 real-world datasets indicate the feasibility and efficiency of the proposed algorithm.

Firstly, the experiments of comparing subgraphs with the removal of vital nodes exhibit the superiority of the proposed algorithm, i.e., the vital nodes identified by the MINE algorithm are easier to exert strong influence than the competitors. By comparing the maximum number of subgraphs when removing the nodes according to the MINE algorithm, the networks are more easily broken up, as shown in Figures 2 and 3. Secondly, we apply the SIR model to evaluate influence maximization effect for each measure (as shown in Figures 4 and 5), with the increasing of infected seeds, the MINE algorithm can obtain relatively more recovered nodes, which suggests that the proposed algorithm is superior to the competitors in both monolayer and multilayer networks. Although inferior to BC on the Jazz dataset and INF on the EUAirTransportation dataset, it is also quite competitive. By analyzing the structures of these two networks, we find that the nodes in Jazz network are densely connected (i.e., the average degree of 27.6970), and most of the nodes are holding the same number of neighbors (approximately 28 neighbors), which brings difficulties to identify vital nodes to exert strong influence. The EUAirTransportation network has 37 layers, and the clustering efficiency is relatively lower (1.6187) than that of other multilayer networks. The proposed MINE algorithm is capable of identifying influencers hierarchically, while the EUAirTransportation network is composed of airlines that fail to form a "tree-like" structure in the folding process. However, the INF indicator considers the local neighboring information of a focal node, which is more capable of identifying vital nodes in the EUAirTransportation network. Nonetheless, the MINE algorithm outperforms other competitors (i.e., BC, CC, DC, and LGR) except INF. Finally, we compare the running time of each indicator on the 32 real-world datasets. Experimental results show the efficiency of the proposed measure.

4.4. Application

Inspired by the graph of adjacencies (i.e., the network of shared borders between countries, states, or provinces) [85], we collect the adjacencies of provinces in China, where each province is represented as a node and there is an edge between any two nodes if the corresponding provinces are geographically adjacent (i.e., share a border). By performing with the proposed MINE algorithm, the vital provinces are plotted in dark blue, as shown in Figure 7, and the full comparisons of provinces by different centrality measures are given in Table 5.

Meghanathan [86] proposes a normalization-based approach to obtain a comprehensive centrality ranking of the states in the graph of adjacencies from the map of USA. As concluded in his research, (i) the state of Missouri is the top-ranked node concerning all the commonly studied centrality measures, (ii) the state of Idaho has the lowest non-zero local clustering coefficient, indicating that the state is the most critical state with respect to facilitating communication between its neighboring states, (iii) the vital states measured by a single centrality is likely to be insufficient to exert strong influence on the whole network. Likewise, the vital nodes (marked in dark blue) identified by the proposed MINE algorithm in the graph of adjacencies related to the lower 48 United States are presented in Figure 8, and the full comparisons are shown in Table 6.
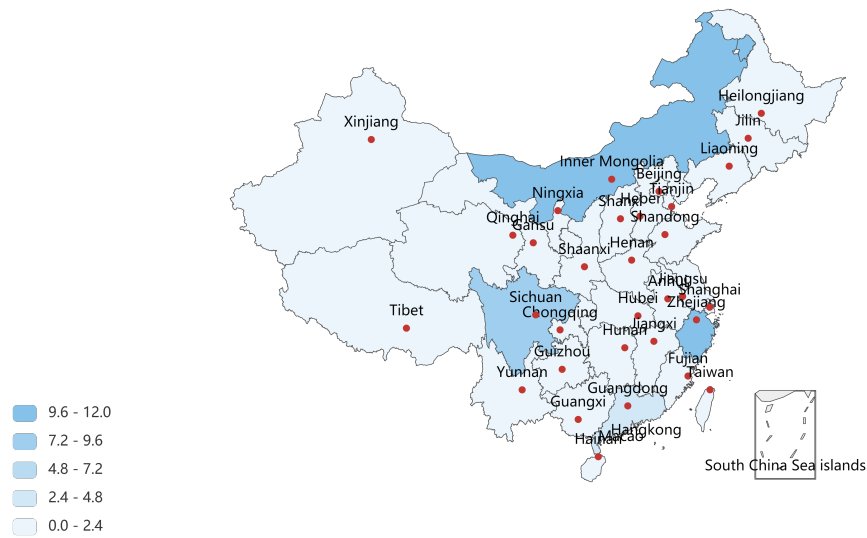
Figure 7. The vital provinces of China measured by the proposed MINE algorithm, where Inner Mongolia, Sichuan, Zhejiang and Guangdong hold more importance in the graph of the adjacencies (formed by 34 vertices and 73 edges). By performing with the proposed MINE algorithm, we can find that Inner Mongolia holds the maximum importance in the derived network model, which may result from its huge span in longitude (97°12′ E–126°04′ E) and adjacencies to eight other provinces, i.e., Gansu, Ningxia, Shaanxi, Shanxi, Hebei, Liaoning, Jilin and Heilongjiang. Zhejiang is the second influential province measured by the MINE algorithm, which suggests its vital position in the southeast of China.
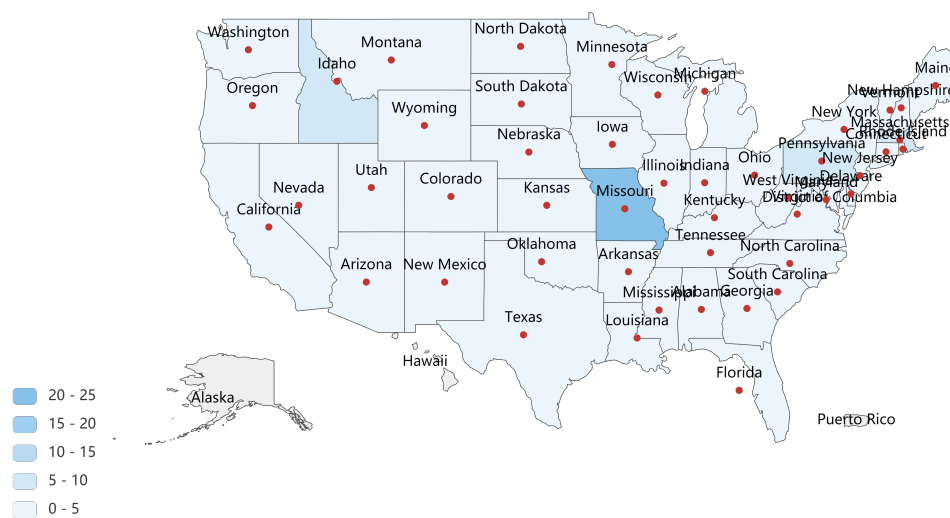

Figure 8. The influence of the lower 48 United States is measured by the proposed MINE algorithm based on the graph of adjacencies (formed by 48 vertices and 104 edges). By performing with MINE algorithm, Missouri (with maximum importance which is same to the conclusion of Meghanathan's research), following by Massachusetts and Idaho (also pointed out by Meghanathan as a vital state) are identified as the top-3 vital states, which are respectively located in the middle, east and west of the lower 48 United States. By activating the top-3 nodes and their neighbors in the graph of adjacencies, we are able to get a maximum collection of activated states that covers the majority of the United States, which are probably to exert strong influence.

Table 5. Comparison of provinces of China, measured by DC, BC, CC, LGR, INF and MINE.

Province	DC	BC	CC	LGR	INF	MINE
Inner Mongolia	**0.2424**	0.1823	0.3793	<u>366.0000</u>	<u>2.1845</u>	**11.2417**
Shaanxi	**0.2424**	**0.2438**	<u>0.4400</u>	**474.0000**	1.5512	1.5512
Hebei	<u>0.2121</u>	0.1733	0.3626	269.5000	2.1250	2.1250
Sichuan	<u>0.2121</u>	0.1170	0.3708	322.0000	1.4417	7.8988
Guangdong	<u>0.2121</u>	<u>0.2065</u>	0.3474	231.0000	**2.8333**	4.2619
Gansu	0.1818	0.0914	0.3667	271.5000	1.3095	1.3095
Henan	0.1818	0.1291	0.4177	295.5000	1.1012	1.1012
Hubei	0.1818	0.1856	**0.4459**	324.0000	0.9917	0.9917
Anhui	0.1818	0.1112	0.3882	250.5000	1.2000	1.2000
Jiangxi	0.1818	0.1658	0.3976	270.0000	1.0929	1.0929
Hunan	0.1818	0.1204	0.3976	265.5000	1.1262	1.1262
Chongqing	0.1515	0.0611	0.4074	237.5000	0.8012	0.8012
Guizhou	0.1515	0.0326	0.3511	181.2500	1.0095	1.0095
Zhejiang	0.1515	0.0589	0.3267	147.5000	1.3333	<u>10.5976</u>
Shanxi	0.1212	0.0119	0.3626	165.0000	0.5595	0.5595
Shandong	0.1212	0.0665	0.3626	138.0000	0.7262	0.7262
Qinghai	0.1212	0.0059	0.3028	113.0000	0.8929	0.8929
Jiangsu	0.1212	0.0342	0.3173	97.0000	1.1167	1.1167
Yunnan	0.1212	0.0431	0.3300	119.0000	0.8429	0.8429
Tibet	0.1212	0.0195	0.3000	100.0000	0.9762	0.9762
Fujian	0.1212	0.0737	0.3204	109.0000	1.5095	1.5095
Guangxi	0.1212	0.0588	0.3333	125.0000	0.7595	0.7595
Jilin	0.0909	0.0009	0.2821	60.0000	0.9583	0.9583
Liaoning	0.0909	0.0073	0.3028	81.7500	0.6012	0.6012
Ningxia	0.0909	0.0000	0.3402	103.5000	0.4167	0.4167
Xinjiang	0.0909	0.0039	0.2920	64.5000	0.6667	0.6667
Heilongjiang	0.0606	0.0000	0.2797	37.5000	0.4583	0.4583
Beijing	0.0606	0.0000	0.2705	30.5000	0.6429	0.6429
Tianjin	0.0606	0.0000	0.2705	30.5000	0.6429	0.6429
Shanghai	0.0606	0.0000	0.2661	28.0000	0.4500	0.4500
Hangkong	0.0606	0.0000	0.2619	28.5000	0.6429	0.6429
Macao	0.0606	0.0000	0.2619	28.5000	0.6429	0.6429
Taiwan	0.0303	0.0000	0.2444	8.5000	0.2500	0.2500
Hainan	0.0303	0.0000	0.2598	13.0000	0.1429	0.1429

Note: The maximum values of each column are in bold, and the second maximum values of each column are underlined. By comparing the vital provinces measured by these centrality measures, we can infer that the proposed MINE algorithm is able to identify influential nodes dispersedly, which may benefit the influence maximization effect.

By respectively comparing the vital provinces (or states) in China and the USA, we are able to deduce that the proposed MINE algorithm is capable of identifying vital nodes dispersedly, which may benefit the information diffusion process and exert a positive effect in influence maximization.

Table 6. Comparison of the lower 48 United States, measured by DC, BC, CC, LGR, INF and MINE.

State	DC	BC	CC	LGR	INF	MINE
Missouri	**0.1702**	**0.3703**	**0.3561**	**520.0000**	1.4024	**24.9286**
Kentucky	<u>0.1489</u>	<u>0.3437</u>	<u>0.3431</u>	<u>374.5000</u>	1.3679	1.3679
Tennessee	<u>0.1489</u>	0.1862	0.3219	364.0000	1.3845	1.3845
Idaho	0.1277	0.0886	0.2271	198.0000	1.5667	7.0167
Wyoming	0.1277	0.1268	0.2717	273.0000	1.1167	1.1167
Colorado	0.1277	0.0658	0.2717	259.5000	1.2000	1.2000
Nebraska	0.1277	0.1606	0.3133	307.5000	1.0417	1.0417
South Dakota	0.1277	0.0590	0.2655	231.0000	1.3333	1.3333
Oklahoma	0.1277	0.1147	0.3032	271.5000	1.2083	1.2083
Iowa	0.1277	0.0820	0.3013	282.0000	1.1583	1.1583
Arkansas	0.1277	0.0733	0.2956	268.5000	1.2679	1.2679
Pennsylvania	0.1277	0.3018	0.2655	193.5000	1.5167	5.7690
Nevada	0.1064	0.0119	0.1992	137.5000	1.2000	1.2000
Utah	0.1064	0.0385	0.2327	183.7500	0.9500	0.9500
Illinois	0.1064	0.0451	0.3092	215.0000	0.9345	0.9345
Ohio	0.1064	0.1772	0.3032	178.7500	1.0929	1.0929
West Virginia	0.1064	0.1446	0.3013	182.5000	1.0095	1.0095
Georgia	0.1064	0.0380	0.2568	131.2500	<u>1.6429</u>	3.1857
New York	0.1064	0.2280	0.2238	127.5000	1.3667	1.3667
Massachusetts	0.1064	0.0634	0.1895	92.5000	**1.7000**	<u>7.1000</u>
Oregon	0.0851	0.0056	0.1918	83.0000	1.2000	1.2000
Montana	0.0851	0.0184	0.2293	122.0000	0.8333	0.8333
Arizona	0.0851	0.0481	0.2227	100.0000	0.9833	0.9833
Kansas	0.0851	0.0160	0.2956	160.0000	0.6250	0.6250
New Mexico	0.0851	0.0752	0.2582	126.0000	0.8333	0.8333
Minnesota	0.0851	0.0296	0.2655	108.0000	0.9167	0.9167
Texas	0.0851	0.0205	0.2527	109.0000	0.9167	0.9167
Wisconsin	0.0851	0.0475	0.2765	111.0000	0.9500	0.9500
Mississippi	0.0851	0.0148	0.2626	116.0000	0.8929	0.8929
Indiana	0.0851	0.0142	0.2883	120.0000	0.8762	0.8762
Virginia	0.0851	0.0581	0.2883	125.0000	0.8429	0.8429
Alabama	0.0851	0.0202	0.2527	102.0000	1.0929	1.0929
North Carolina	0.0851	0.0451	0.2733	112.0000	1.0929	1.0929
Maryland	0.0851	0.0334	0.2527	96.0000	0.9500	0.9500
California	0.0638	0.0031	0.1888	51.7500	0.7000	0.7000
North Dakota	0.0638	0.0046	0.2315	63.0000	0.6667	0.6667
Louisiana	0.0638	0.0032	0.2398	63.7500	0.6667	0.6667
Michigan	0.0638	0.0397	0.2626	63.7500	0.7000	0.7000
Delaware	0.0638	0.0017	0.2176	53.2500	0.7500	0.7500
New Jersey	0.0638	0.0032	0.2186	60.7500	0.7000	0.7000
Vermont	0.0638	0.0389	0.1880	50.2500	0.7333	0.7333
Connecticut	0.0638	0.0194	0.1873	47.2500	0.9000	0.9000
New Hampshire	0.0638	0.0426	0.1615	34.5000	1.5333	1.5333
Washington	0.0426	0.0000	0.1873	31.5000	0.4167	0.4167
Florida	0.0426	0.0000	0.2080	26.5000	0.4500	0.4500
South Carolina	0.0426	0.0000	0.2186	26.5000	0.4500	0.4500
Rhode Island	0.0426	0.0000	0.1604	21.5000	0.5333	0.5333
Maine	0.0213	0.0000	0.1395	5.0000	0.3333	0.3333

Note: The maximum values of each column are in bold, and second maximum values of each column are underlined. As identified by DC, BC, CC, LGR, Missouri seems to be the most vital states and Kentucky is next most important one. However, Kentucky is close to Missouri in geography, which limits the scope of influence expansion. By performing with INF measure, Massachusetts and Georgia are identified as the top-2 vital states, which are located in the east of the lower 48 United States and limited to the local areas. The proposed MINE algorithm identifies Missouri (44.9286), Massachusetts (7.1000) and Idaho (7.0167) as top-3 vital states, which are located dispersedly and are easy to exert strong influence.

5. Conclusions

Aiming at solving the influence maximization problem in complex networks, this paper proposes a novel MINE algorithm based on the folding of community structures hierarchically. Specifically, the complexity of MINE is $O(n \log n)$, which suggests its feasibility in large-scale networks. Extensive experiments on 32 real-world datasets are conducted, including 16 monolayer networks and 16 multilayer networks, respectively. The experimental results suggest that the proposed algorithm outperforms competitors in most cases. Afterward, the runtime comparison of the proposed method with competitors verifies the efficiency of the proposed MINE algorithm. The applications in the graph of adjacencies of provinces (states) in China and USA also exhibit its value in real-life scenarios. Overall, the proposed MINE algorithm is capable of identifying vital nodes in complex networks and is of paramount significance in the influence maximization problem. The contribution of this work is likely to benefit many real-world applications, such as promoting network evolutions, preventing the spreading of rumors, etc.

As part of future works, the improvement of identifying community structures can be further studied (e.g., the complexity of $O(n)$ or $O(\log n)$ may provide significant runtime savings). The simplification of the folding process in MINE algorithm is also expected. In a word, we hope the findings in this work will help to improve the research in this promising field.

References

1.　Lü, L.; Chen, D.; Ren, X.L.; Zhang, Q.M.; Zhang, Y.C.; Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **2016**, *650*, 1–63. [CrossRef]

2.　Basaras, P.; Iosifidis, G.; Katsaros, D.; Tassiulas, L. Identifying influential spreaders in complex multilayer networks: A centrality perspective. *IEEE Trans. Netw. Sci. Eng.* **2017**, *6*, 31–45. [CrossRef]

3.　Ren, X. Review of ranking nodes in complex networks. *Chin. Sci. Bull.* **2014**, *59*, 1175–1197. [CrossRef]

4.　Morone, F.; Makse, H.A. Influence maximization in complex networks through optimal percolation. *Nature* **2015**, *524*, 65–68. [CrossRef]

5.　Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **2010**, *32*, 245–251. [CrossRef]

6.　Richardson, M.; Domingos, P. Mining knowledge-sharing sites for viral marketing. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 61–70.

7.　Pastor-Satorras, R.; Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **2002**, *65*, 036104. [CrossRef] [PubMed]

8.　Albert, R.; Albert, I.; Nakarado, G.L. Structural vulnerability of the North American power grid. *Phys. Rev. E* **2004**, *69*, 025103. [CrossRef]

9.　Salavati, C.; Abdollahpouri, A.; Manbari, Z. BridgeRank: A novel fast centrality measure based on local structure of the network. *Phys. A Stat. Mech. Its Appl.* **2018**, *496*, 635–653. [CrossRef]

10.　Thai, M.T.; Wu, W.; Xiong, H. *Big Data in Complex and Social Networks*; CRC Press: Boca Raton, FL, USA, 2016.

11.　Liu, W.; Suzumura, T.; Ji, H.; Hu, G. Finding overlapping communities in multilayer networks. *PLoS ONE* **2018**, *13*, e0188747. [CrossRef]

12. Wang, D.; Wang, H.; Zou, X. Identifying key nodes in multilayer networks based on tensor decomposition. *Chaos Interdisc. J. Nonlinear Sci.* **2017**, *27*, 063108. [CrossRef]

13. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [CrossRef]

14. Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [CrossRef] [PubMed]

15. Domingos, P.; Richardson, M. Mining the network value of customers. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 26–29 August 2001; pp. 57–66.

16. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 137–146.

17. Li, X.; Zhang, X.; Zhao, C.; Yi, D.; Li, G. Identifying highly influential nodes in multilayer networks based on global propagation. *Chaos Interdisc. J. Nonlinear Sci.* **2020**, *30*, 061107. [CrossRef] [PubMed]

18. Ford, L.R.; Fulkerson, D.R. Maximal flow through a network. In *Classic Papers in Combinatorics*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 243–248.

19. Corley, H., Jr.; Chang, H. Finding the n most vital nodes in a flow network. *Manag. Sci.* **1974**, *21*, 362–364. [CrossRef]

20. Peng, S.; Zhou, Y.; Cao, L.; Yu, S.; Niu, J.; Jia, W. Influence analysis in social networks: A survey. *J. Netw. Comput. Appl.* **2018**, *106*, 17–32. [CrossRef]

21. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.

22. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **1972**, *2*, 113–120. [CrossRef]

23. Freeman, L.C. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35–41. [CrossRef]

24. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [CrossRef]

25. Bonacich, P. Power and centrality: A family of measures. *Am. J. Sociol.* **1987**, *92*, 1170–1182. [CrossRef]

26. Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]

27. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. [CrossRef] [PubMed]

28. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888. [CrossRef]

29. Ma, L.l.; Ma, C.; Zhang, H.F.; Wang, B.H. Identifying influential spreaders in complex networks based on gravity formula. *Phys. A Stat. Mech. Its Appl.* **2016**, *451*, 205–212. [CrossRef]

30. Li, Z.; Ren, T.; Ma, X.; Liu, S.; Zhang, Y.; Zhou, T. Identifying influential spreaders by gravity model. *Sci. Rep.* **2019**, *9*, 8387. [CrossRef] [PubMed]

31. Maji, G.; Mandal, S.; Sen, S. A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques. *Expert Syst. Appl.* **2020**, *161*, 113681. [CrossRef]

32. Fan, T.; Lü, L.; Shi, D. Towards the cycle structures in complex network: A new perspective. *arXiv* **2019**, arXiv:1903.01397.

33. Huang, X.; Chen, D.; Wang, D.; Ren, T. Identifying Influencers in Social Networks. *Entropy* **2020**, *22*, 450. [CrossRef]

34. Lü, L.; Zhang, Y.C.; Yeung, C.H.; Zhou, T. Leaders in social networks, the delicious case. *PLoS ONE* **2011**, *6*, e21202. [CrossRef]

35. Al-Garadi, M.A.; Varathan, K.D.; Ravana, S.D.; Ahmed, E.; Chang, V. Identifying the influential spreaders in multilayer interactions of online social networks. *J. Intell. Fuzzy Syst.* **2016**, *31*, 2721–2735. [CrossRef]

36. Singh, S.S.; Singh, K.; Kumar, A.; Biswas, B. MIM2: Multiple influence maximization across multiple social networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *526*, 120902. [CrossRef]

37. Chen, X.; Zhou, J.; Liao, Z.; Liu, S.; Zhang, Y. A Novel Method to Rank Influential Nodes in Complex Networks Based on Tsallis Entropy. *Entropy* **2020**, *22*, 848. [CrossRef]

38. Molaei, S.; Farahbakhsh, R.; Salehi, M.; Crespi, N. Identifying Influential Nodes in Heterogeneous Networks. *Expert Syst. Appl.* **2020**, *160*, 113580. [CrossRef]

39. Hosni, A.I.E.; Li, K.; Ahmad, S. Minimizing rumor influence in multiplex online social networks based on human individual and social behaviors. *Inf. Sci.* **2020**, *512*, 1458–1480. [CrossRef]

40. Tidke, B.; Mehta, R.; Dhanani, J. Consensus-based aggregation for identification and ranking of top-k influential nodes. *Neural Comput. Appl.* **2019**, *32*, 1–27. [CrossRef]

41. Li, M.; Zhang, Q.; Deng, Y. Evidential identification of influential nodes in network of networks. *Chaos Solitons Fractals* **2018**, *117*, 283–296. [CrossRef]

42. Battiston, F.; Nicosia, V.; Latora, V. Structural measures for multiplex networks. *Phys. Rev. E* **2014**, *89*, 032804. [CrossRef]

43. Pedroche, F.; Tortosa, L.; Vicent, J.F. An eigenvector centrality for multiplex networks with data. *Symmetry* **2019**, *11*, 763. [CrossRef]

44. Wang, S.; Liu, J.; Jin, Y. Finding influential nodes in multiplex networks using a memetic algorithm. *IEEE Trans. Cybern.* **2019**. [CrossRef]

45. Xie, W.B.; Lee, Y.L.; Wang, C.; Chen, D.B.; Zhou, T. Hierarchical clustering supported by reciprocal nearest neighbors. *Inf. Sci.* **2020**, *527*, 279–292. [CrossRef]

46. Fu, G.; Hou, C.; Yao, X. Learning topological representation for networks via hierarchical sampling. In Proceedings of the IEEE 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.

47. Chen, H.; Perozzi, B.; Hu, Y.; Skiena, S. Harp: Hierarchical representation learning for networks. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

48. Ye, Z.; Zhan, X.; Zhou, Y.; Liu, C.; Zhang, Z.K. Identifying vital nodes on temporal networks: An edge-based K-Shell decomposition. In Proceedings of the IEEE 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 1402–1407.

49. Chen, X.; Tan, M.; Zhao, J.; Yang, T.; Wu, D.; Zhao, R. Identifying influential nodes in complex networks based on a spreading influence related centrality. *Phys. A Stat. Mech. Its Appl.* **2019**, *536*, 122481. [CrossRef]

50. Zareie, A.; Sheikhahmadi, A. A hierarchical approach for influential node ranking in complex social networks. *Expert Syst. Appl.* **2018**, *93*, 200–211. [CrossRef]

51. Tsiropoulou, E.; Koukas, K.; Papavassiliou, S. A socio-physical and mobility-aware coalition formation mechanism in public safety networks. *EAI Endorsed Trans. Future Internet* **2018**, *4*, 154176. [CrossRef]

52. Huang, H.; Shen, H.; Meng, Z.; Chang, H.; He, H. Community-based influence maximization for viral marketing. *Appl. Intell.* **2019**, *49*, 2137–2150. [CrossRef]

53. Hu, P.; Fan, W.L. Mitigation strategy against cascading failures considering vulnerable transmission line in power grid. *Phys. A Stat. Mech. Its Appl.* **2020**, *540*, 123230. [CrossRef]

54. Nikolaou, P.; Dimitriou, L. Identification of critical airports for controlling global infectious disease outbreaks: Stress-tests focusing in Europe. *J. Air Transp. Manag.* **2020**, *85*, 101819. [CrossRef] [PubMed]

55. Fragkos, G.; Apostolopoulos, P.A.; Tsiropoulou, E.E. ESCAPE: Evacuation strategy through clustering and autonomous operation in public safety systems. *Future Internet* **2019**, *11*, 20. [CrossRef]

56. Peng, B.; Chen, W. Adaptive influence maximization with myopic feedback. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 13 December 2019; pp. 5574–5583.

57. Guo, C.; Yang, L.; Chen, X.; Chen, D.; Gao, H.; Ma, J. Influential Nodes Identification in Complex Networks via Information Entropy. *Entropy* **2020**, *22*, 242. [CrossRef]

58. Tidke, B.; Mehta, R.; Dhanani, J. Multimodal ensemble approach to identify and rank top-k influential nodes of scholarly literature using Twitter network. *J. Inf. Sci.* **2020**, *46*, 437–458. [CrossRef]

59. Hethcote, H.W. The mathematics of infectious diseases. *SIAM Rev.* **2000**, *42*, 599–653. [CrossRef]

60. Guan-Rong, C.; Xiao-Fan, W.; Xiang, L. *Introduction to Complex Networks: Models, Structures and Dynamics*; Higher Education Press: Beijing, China, 2012.

61. Raghavan, U.N.; Albert, R.; Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **2007**, *76*, 036106. [CrossRef]

62. Zachary, W.W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [CrossRef]

63. Rossi, R.; Ahmed, N. The network data repository with interactive graph analytics and visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.

64. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef] [PubMed]

65. Lusseau, D.; Schneider, K.; Boisseau, O.J.; Haase, P.; Slooten, E.; Dawson, S.M. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **2003**, *54*, 396–405. [CrossRef]

66. Tsvetovat, M.; Kouznetsov, A. *Social Network Analysis for Startups: Finding Connections on the Social Web*; OŘeilly Media, Inc.: Newton, MA, USA, 2011.

67. Knuth, D.E. The Stanford GraphBase: a platform for combinatorial algorithms. In Proceedings of the SODA, Austin, TX, USA, 25–27 January 1993; Volume 93, pp. 41–43.

68. Shen-Orr, S.S.; Milo, R.; Mangan, S.; Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **2002**, *31*, 64–68. [CrossRef]

69. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440. [CrossRef] [PubMed]

70. Huang, X.; Chen, D.; Ren, T. Social network coalescence based on multilayer network model. *J. Nonlinear Convex Anal.* **2019**, *20*, 1465–1474.

71. Gleiser, P.M.; Danon, L. Community structure in jazz. *Adv. Complex Syst.* **2003**, *6*, 565–573. [CrossRef]

72. Colizza, V.; Pastor-Satorras, R.; Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **2007**, *3*, 276–282. [CrossRef]

73. Newman, M.E. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [CrossRef]

74. Cho, A.; Shin, J.; Hwang, S.; Kim, C.; Shim, H.; Kim, H.; Kim, H.; Lee, I. WormNet v3: A network-assisted hypothesis-generating server for Caenorhabditis elegans. *Nucleic Acids Res.* **2014**, *42*, W76–W82. [CrossRef]

75. Beveridge, A.; Shan, J. Network of thrones. *Math Horizons* **2016**, *23*, 18–22. [CrossRef]

76. Action, R. the Rise of the Medici. *Am. J. Sociol.* **1993**, *98*, 1259–1319.

77. Stark, C.; Breitkreutz, B.J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [CrossRef] [PubMed]

78. De Domenico, M.; Solé-Ribalta, A.; Gómez, S.; Arenas, A. Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8351–8356. [CrossRef] [PubMed]

79. Coleman, J.; Katz, E.; Menzel, H. The diffusion of an innovation among physicians. *Sociometry* **1957**, *20*, 253–270. [CrossRef]

80. Chen, B.L.; Hall, D.H.; Chklovskii, D.B. Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4723–4728. [CrossRef]

81. De Domenico, M.; Nicosia, V.; Arenas, A.; Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **2015**, *6*, 1–9. [CrossRef]

82. De Domenico, M.; Lancichinetti, A.; Arenas, A.; Rosvall, M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **2015**, *5*, 011027. [CrossRef]

83. Cardillo, A.; Gómez-Gardenes, J.; Zanin, M.; Romance, M.; Papo, D.; Del Pozo, F.; Boccaletti, S. Emergence of network features from multiplexity. *Sci. Rep.* **2013**, *3*, 1344. [CrossRef]

84. Newman, M.E. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [CrossRef]

85. Newman, M. *Networks*; Oxford University Press: Oxford, UK, 2018.

86. Meghanathan, N. Complex Network Analysis of the Contiguous United States Graph. *Comput. Inf. Sci.* **2017**, *10*, 54–76. [CrossRef]