MDPI

Article

Influential Nodes Identification in Complex Networks via Information Entropy

Chungu Guo [1], Liangwei Yang [1], Xiao Chen [2], Duanbing Chen [1,3,4,5,*] , Hui Gao [1] and Jing Ma [6]

[1]   School of Computer Science and Engineering, University of Electricity Science and Technology of China, Chengdu 611731, China; thespringguo@gmail.com (C.G.); y805797986@gmail.com (L.Y.); ahuigao@gmail.com (H.G.)

[2]   Information Assurance Office of Army Staff, Beijing 100043, China; 18215561894@163.com

[3]   The Center for Digital Culture and Media, University of Electricity Science and Technology of China, Chengdu 611731, China

[4]   Institute of Fundamental and Frontier Sciences, University of Electricity Science and Technology of China, Chengdu 611731, China

[5]   Union Big Data Tech. Inc., Chengdu 610041, China

[6]   Business School, Sichuan University, Chengdu 610064, China; jingma@stu.scu.edu.cn

*   Correspondence:dbchen@uestc.edu.cn

check for updates

Abstract: Identifying a set of influential nodes is an important topic in complex networks which plays a crucial role in many applications, such as market advertising, rumor controlling, and predicting valuable scientific publications. In regard to this, researchers have developed algorithms from simple degree methods to all kinds of sophisticated approaches. However, a more robust and practical algorithm is required for the task. In this paper, we propose the EnRenew algorithm aimed to identify a set of influential nodes via information entropy. Firstly, the information entropy of each node is calculated as initial spreading ability. Then, select the node with the largest information entropy and renovate its *l*-length reachable nodes' spreading ability by an attenuation factor, repeat this process until specific number of influential nodes are selected. Compared with the best state-of-the-art benchmark methods, the performance of proposed algorithm improved by 21.1%, 7.0%, 30.0%, 5.0%, 2.5%, and 9.0% in final affected scale on CEnew, Email, Hamster, Router, Condmat, and Amazon network, respectively, under the Susceptible-Infected-Recovered (SIR) simulation model. The proposed algorithm measures the importance of nodes based on information entropy and selects a group of important nodes through dynamic update strategy. The impressive results on the SIR simulation model shed light on new method of node mining in complex networks for information spreading and epidemic prevention.

Keywords: complex networks; influential nodes; information entropy; SIR model

1. Introduction

Complex networks are common in real life and can be used to represent complex systems in many fields. For example, collaboration networks [1] are used to cover the scientific collaborations between authors, email networks [2] denote the email communications between users, protein-DNA networks [3] help people gain a deep insight on biochemical reaction, railway networks [4] reveal the structure of railway via complex network methods, social networks show interactions between people [5,6], and international trade network [7] reflects the products trade between countries. A deep understanding and controlling of different complex networks is of great significance in information spreading and network connectivity. On one hand, by using the influential nodes, we can make successful advertisements for products [8], discover drug target candidates, assist information

diffusion [9], and even detect essential proteins [10]. On the other hand, by removing some critical nodes, it can greatly reduce the connectivity of the network to restrain the outbreak of epidemics [11] or spreading of rumors [12].

The ongoing COVID-19 epidemics is catching wide attention around the world. Every country is making enormous effort to control the virus spreading. By analyzing social networks, it would be easier for us to control epidemics spreading. In the last decades, propagation dynamics has always been an important research direction. Many mechanisms, such as epidemic spreading [13–16], rumor propagation [17,18], social sudden events spreading [19], and e-commercial advertisements, are all closely related to complex network dynamics. Early in 1760, Daniel Bernoulli studied smallpox vaccine by using ordinary differential equations, and gave the Bernoulli equations [20] , which is one of the earliest propagation dynamics models. Later, Hamer presented the mass-action principle [21,22] when studying the recurring epidemics of measles. A.G. McKendrick and W.O. Kermack formulated a famous modern mathematical epidemic model named the Susceptible-Infected-Recovered (SIR) compartmental model when studying the spreading pattern of the Black Death and the plague in 1906. SIR compartmental model successfully predicted the outbreak of several epidemics [23]. Harding et al. [24] followed the maximum entropy (MaxEnt) principle when simulating on the SIS model to study epidemics spreading on networks. Wang et al. [25] used entropy to incorporate many related factors, and designed a rumor spreading model. Zenil et al. [26] used algorithmic information content to steer the controlling of complex systems. By analyzing airline network data, Brockmann et al. identified the origin place of epidemics [27]. Pastor et al. summarized the methods and models in propagation dynamics [13], providing theoretical support for predicting and controlling propagation.

Identifying a set of influential spreaders in complex networks plays a crucial part in target advertisement, epidemics and rumor control. One naive method is to select top *k* ranked nodes as spreaders according to ranking methods. Lately, there have been many ranking methods to this problem. Some methods consider node's local information, such as degree centrality, and its extending measures [28]. Degree centrality counts the number of node's direct neighbors as its influence. H-index [29] considers 2-nd order neighbors as node's importance. Based on degree centrality, LocalRank [30] takes node's 4-th order neighbors into account. ClusterRank [31] combines Degree centrality and clustering coefficient [32] to measure node's importance in spreading. Coreness centrality [33] decomposes graph by *K*-shell method, and it selects nodes based on their location within the graph. Zeng and Zhang [34] presented algorithm to rank nodes by decomposing networks in a mixed degree decomposition procedure. There are also some path-based methods to quantify nodes' importance, such as Eccentricity centrality [35], Closeness centrality [36], Betweenness centrality [37], and Katz centrality [38]. These methods suffer from relatively high computational cost for calculating shortest path between nodes. Considering the mutual enhancement [39], researchers put forward Eigenvector centrality [40] and cumulative nomination algorithm [41]. Besides, random walk based methods have also been proposed, such as the famous PageRank [42] and LeaderRank [43]. Information entropy has also been used as importance indicator to measure nodes' importance [44,45]. Qiao et al. proposed entropy centrality [46,47] to measure the potential for communication activity between node pair. Ai assumes that the removal of a more important node is likely to cause more structural variation. So he measured node importance by entropy variation [48] which calculates the change of graph entropy when that node is removed. In order to capture a wider range and greater abundance of information, Li et al. put forward entropy and mutual information-based centrality approach (EMI) [49] to take both topological and digital network characteristics into account. Fei et al. used relative entropy [50] to choose appropriate centrality measures when dealing with different networks. Zareie et al. [51] ranked nodes in social networks based on nodes' spreading capability by information entropy. Shetty et al. [52] used entropy to detect influential nodes in low-level, incomplete and noisy social network. By combining the advantages of degree centrality, efficiency centrality, betweenness centrality and correlation centrality through entropy, Fan et al. proposed correlation centrality index [53] to quantify node's importance. Recently, entropy has also been used to rank nodes in

weighted networks [54] and social networks [55]. However, the node set built by simply assembling the nodes and sorting them employed by the aforementioned methods may not be comparable to an elaborately selected set of nodes due to the rich club phenomenon [56], namely, important nodes tend to overlap with each other. Thus, lots of methods aim to directly select a set of nodes are proposed.

Kempe et al. defined the problem of identifying a set of influential spreaders in complex networks as influence maximization problem [57], and they used hill-climbing based greedy algorithm that is within 63% of optimal in several models. Greedy method [58] is usually taken as the approximate solution of influence maximization problem, but it is not efficient for its high computational cost. Chen et al. [58] proposed NewGreedy and MixedGreedy method. Borgatti [59] specified mining influential spreaders in social networks by two classes: KPP-POS and KPP-NEG, based on which he calculated the importance of nodes. Narayanam et al. [60] proposed SPIN algorithm based on Shapley value to deal with information diffusion problem in social networks. Although the above greedy based methods can achieve relatively better result, they would cost lots of time for Monte Carlo simulation. So more heuristic algorithms were proposed. Chen et al. put forward simple and efficient DegreeDiscount algorithm [58] in which if one node is selected, its neighbors' degree would be discounted. Zhang et al. proposed VoteRank [61] which selects the influential node set via a voting strategy. Zhao et al. [62] introduced coloring technology into complex networks to seperate independent node sets, and selected nodes from different node sets, ensuring selected nodes are not closely connected. Hu et al. [63] and Guo et al. [64] further considered the distance between independent sets and achieved a better performance. Bao et al. [65] sought to find dispersive distributed spreaders by a heuristic clustering algorithm. Zhou [66] proposed an algorithm to find a set of influential nodes via message passing theory. Ji el al. [67] considered percolation in the network to obtain a set of distributed and coordinated spreaders. Researchers also seek to maximize the influence by studying communities [68–73]. Zhang [74] seperated graph nodes into communities by using K-medoid method before selecting nodes. Gong et al. [75] divided graph into communities of different sizes, and selected nodes by using degree centrality and other indicators. Chen et al. [76] detected communities by using SHRINK and Kcut algorithm. Later they selected nodes from different communities as candidate nodes, and used CDH method to find final k influential nodes. Recently, some novel methods based on node dynamics have been proposed which rank nodes to select influential spreaders [77,78]. Şirag Erkol et al. made a systematic comparison between methods focused on influence maximization problem [79]. They classify multiple algorithms to three classes, and made a detailed explanation and comparison between methods. More algorithms in this domain are described and classified clearly by Lü et al. in their review paper [80].

Most of the non-greedy strategy methods suffer from a possibility that some spreaders are so close that their influence may overlap. DegreeDiscount and VoteRank use iterative selection strategy. After a node is selected, they weaken its neighbors' influence to cope with the rich club phenomenon. However, these two algorithms roughly induce nodes' local information. Besides, they do not further make use of the difference between nodes when weakening nodes' influence. In this paper, we propose a new heuristic algorithm named EnRenew based on node's entropy to select a set of influential nodes. EnRenew also uses iterative selection strategy. It initially calculates the influence of each node by its information entropy (further explained in Section 2.2), and then repeatedly select the node with the largest information entropy and renovate its l-length reachable nodes' information entropy by an attenuation factor until specific number of nodes are selected. Experiments show that the proposed method yields the largest final affected scale on 6 real networks in the Susceptible-Infected-Recovered (SIR) simulation model compared with state-of-the-art benchmark methods. The results reveal that EnRenew could be a promising tool for related work. Besides, to make the algorithm practically more useful, we provide EnRenew's source code and all the experiments details on https://github.com/YangLiangwei/Influential-nodes-identification-in-complex-networks-via-information-entropy, and researchers can download it freely for their convenience.

The rest of paper is organized as follows: The identifying method is presented in Section 2. Experiment results are analyzed and discussed in Section 3. Conclusions and future interest research topics are given in Section 4.

2. Methods

2.1. Spreading Model

The best way to measure the influence of a set of nodes in complex networks is through propagation dynamic process on real life network data. A susceptible infected removed model (SIR model) is initially used to simulate the dynamic of disease spreading [23]. It is later widely used to analyze similar spreading process, such as rumor [81] and population [82]. In this paper, the SIR model is adopted to objectively evaluate the spreading ability of nodes selected by algorithms. Each node in the SIR model can be classified into one of three states, namely, Susceptible nodes (S), Infected nodes (I), and Recovered nodes (R). At first, set initial selected nodes to infected status and all others in network to susceptible status. In each propagation iteration, each infected node randomly choose one of its direct neighbors and infect it with probability μ. In the meantime, each infected node will be recovered with probability β and won't be infected again. In this study, $\lambda = \frac{\mu}{\beta}$ is defined as infected rate, which is crucial to the spreading speed in the SIR model. Apparently, the network can reach a steady stage with no infection after enough propagation iterations. To enable information spreads widely in networks, we set $\mu = 1.5\mu_c$, where $\mu_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$ [83] is the spreading threshold of SIR, $\langle k \rangle$ is the average degree of network. When μ is smaller than μ_c, spreading in SIR could only affect a small range or even cannot spread at all. When it is much larger than μ_c, nearly all methods could affect the whole network, which would be meaningless for comparison. Thus, we select μ around μ_c on the experiments. During the SIR propagation mentioned above, enough information can be obtained to evaluate the impact of initial selected nodes in the network and the metrics derived from the procedure is explained in Section 2.4.
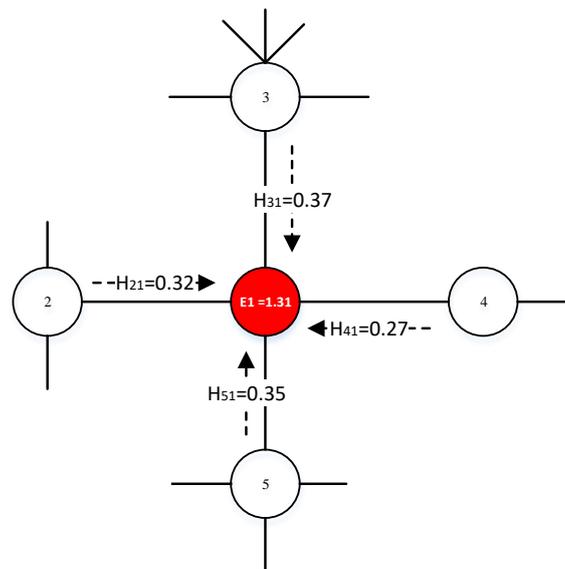
2.2. EnRenew Algorithm

The influential nodes selecting algorithm proposed in this paper is named EnRenew, deduced from the concept of the algorithm. EnRenew introduces entropy and renews the nodes' entropy through an iterative selection process. EnRenew is inspired by VoteRank algorithm proposed by Zhang et al. [61], where the influential nodes are selected in an iterative voting procedure. VoteRank assigns each node with voting ability and scores. Initially, each node's voting ability to its neighbors is 1. After a node is selected, the direct neighbors' voting ability will be decreased by $\frac{1}{\langle k \rangle}$, where $\langle k \rangle = \frac{2*m}{n}$ is the average degree of the network. VoteRank roughly assigns all nodes in graph with the same voting ability and attenuation factor, which ignores node's local information. To overcome this shortcoming, we propose a heuristic algorithm named EnRenew and described as follows.

In information theory, information quantity measures the information brought about by a specific event and information entropy is the expectation of the information quantity. These two concepts are introduced into complex network in Reference [44–46] to calculate the importance of node. Information entropy of any node v can be calculated by:

$$E_v = \sum_{u \in \Gamma_v} H_{uv} = \sum_{u \in \Gamma_v} -p_{uv} \log p_{uv}, \tag{1}$$

where $p_{uv} = \frac{d_u}{\sum_{l \in \Gamma_v} d_l}$, $\sum_{l \in \Gamma_v} p_{lv} = 1$, Γ_v indicates node v's direct neighbors, and d_u is the degree of node u. H_{uv} is the spreading ability provided from u to v. E_v is node v's information entropy indicating its initial importance which would be renewed as described in Algorithm 1. A detailed calculating of node entropy is shown in Figure 1.

Figure 1. It shows how the red node's (node 1) entropy is calculated in detail. Node 1 has four neighbors from node 2 to node 5. Node 1's information entropy is then calculated by $E_1 = H_{21} + H_{31} + H_{41} + H_{51} = 0.32 + 0.37 + 0.27 + 0.35 = 1.31$.

Algorithm 1: EnRenew

Input: number of initial spreaders *num*, Graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ and
$\quad\quad E = \{e_1, \ldots, e_m\}$

Output: S contains *num* spreaders

//Initialization;

$S = \varnothing$

for $(u, v) \in E$ **do**
$\quad\quad p_{uv} = \frac{d_u}{\sum_{l \in \Gamma_v} d_l}$
$\quad\quad H_{uv} = -p_{uv} \log p_{uv}$
end

//Selecting spreaders;

while $|S| < num$ **do**
$\quad\quad$**for** $v \in V$ **do**
$\quad\quad\quad\quad E_v = \sum_{u \in \Gamma_v} H_{uv}$
$\quad\quad$**end**
$\quad\quad$add v to S, where $v = \arg\max_{v \in V} E_v$
$\quad\quad V \leftarrow V \backslash \{v\}$
$\quad\quad$**for** v's l-length reachable node u^l $(1 \leq l \leq diameter(G), u^0 = v)$ **do**
$\quad\quad\quad\quad H_{u^{l-1}u^l} -= \frac{1}{2^{l-1}} \frac{H_{u^{l-1}u^l}}{E_{\langle k \rangle}}$
$\quad\quad$**end**
end

Output S;

Simply selecting the nodes with a measure of degree as initial spreaders might not achieve good results. Because most real networks have obvious clumping phenomenon, that is, high-impact nodes in the network are often connected closely in a same community. Information cannot be copiously

disseminated to the whole network. To manage this situation, after each high impact node is selected, we renovate the information entropy of all nodes in its local scope and then select the node with the highest information entropy, the process of which is shown in Algorithm 1.

$E_{\langle k \rangle} = -\langle k \rangle \cdot \frac{1}{\langle k \rangle} \cdot \log \frac{1}{\langle k \rangle}$ and $\langle k \rangle$ is the average degree of the network. $\frac{1}{2^{l-1}}$ is the attenuation factor, the farther the node is from node v, the smaller impact on the node will be. $E_{\langle k \rangle}$ can be seen as the information entropy of any node in $\langle k \rangle$-regular graph if $\langle k \rangle$ is an integer.

From Algorithm 1, we can see that after a new node is selected, the renew of its l-length reachable nodes' information entropy is related with H and $E_{\langle k \rangle}$, which reflects local structure information and global network information, respectively. Compared with VoteRank, EnRenew replaces voting ability by H value between connected nodes. It induces more local information than directly set voting ability as 1 in VoteRank. At the same time, EnRenew uses $\frac{H}{E_{\langle k \rangle}}$ as the attenuate factor instead of $\frac{1}{\langle k \rangle}$ in VoteRank, retaining global information.

2.3. Computational Complexity Analysis

Computational complexity (usually time complexity) is used to describe the relationship between the input of different scales and the running time of the algorithm. Generally, brute force can solve most problems accurately, but it cannot be applied in most scenarios because of its intolerable time complexity. Time complexity is an extremely important indicator of an algorithm's effectiveness. Through analysis, the algorithm is proved to be able to identify influential nodes in large-scale network in limited time. The computational complexity of EnRenew can be analyzed in three parts, initialization, selection and renewing. n, m and r represent the number of nodes, edges and initial infected nodes, respectively. At start, EnRenew takes $O(n \cdot \langle k \rangle) = O(m)$ for calculating information entropy. Node selection selects the node with the largest information entropy and requires $O(n)$, which can further be decreased to $O(\log n)$ if stored in an efficient data structure such as red-black tree. Renewing the l-length reachable nodes' information entropy needs $O(\langle k \rangle^l) = O(\frac{m^l}{n^l})$. As suggested in Section 3.3, $l = 2$ yields impressive results with $O(\frac{m^2}{n^2})$. Since selection and renewing parts need to be performed r times to get enough spreaders, the final computational complexity is $O(m + n) + O(r \log n) + O(r \langle k \rangle^2) = O(m + n + r \log n + \frac{rm^2}{n^2})$. Especially, when the network is sparse and $r \ll n$, the complexity will be decreased to $O(n)$.

2.4. Performance Metrics

The algorithm's performance is measured by the selected nodes' properties including its spreading ability and location property. Spreading ability can be measured by infected scale at time t $F(t)$ and final infected scale $F(t_c)$, which are obtained from SIR simulation and widely used to measure the spreading ability of nodes [61,84–88]. L_S is obtained from selected nodes' location property by measuring their dispersion [61].

Infected scale $F(t)$ demonstrates the influence scale at time t and is defined by

$$F(t) = \frac{n_{I(t)} + n_{R(t)}}{n}, \tag{2}$$

where $n_{I(t)}$ and $n_{R(t)}$ are the number of infected and recovered nodes at time t, respectively. At the same time step t, larger $F(t)$ indicates more nodes are infected by initial influential nodes, while a shorter time t indicates the initial influential nodes spread faster in the network.

$F(t_c)$ is the final affected scale when the spreading reaches stable state. This reflects the final spreading ability of initial spreaders. The larger the value is, the stronger the spreading capacity of initial nodes. $F(t_c)$ is defined by:

$$F(t_c) = \frac{n_{R(t_c)}}{n}, \tag{3}$$

where t_c is the time when SIR propagation procedure reaches its stable state.

L_S is the average shortest path length of initial infection set S. Usually, with larger L_S, the initial spreaders are more dispersed and can influence a larger range. This can be defined by:

$$L_S = \frac{1}{|S|(|S|-1)} \sum_{u,v \in S, u \neq v} l_{u,v},$$

(4)

where $l_{u,v}$ denotes the length of the shortest path from node u to v. If u and v is disconnected, the shortest path is replaced by $D_{GC} + 1$, where D_{GC} is the largest diameter of connected components.

3. Results and Discussions

3.1. An Example Network

An example network shown in Figure 2 is used to show the rationality of nodes the proposed algorithm chooses. The first three nodes selected by EnRenew is distributed in three communities, while those selected by the other algorithms are not. We further run the SIR simulation on the example network with EnRenew and other five benchmark methods. The detailed result is shown in Table 1 for an in-depth discussion. This result is obtained by averaging 1000 experiments.
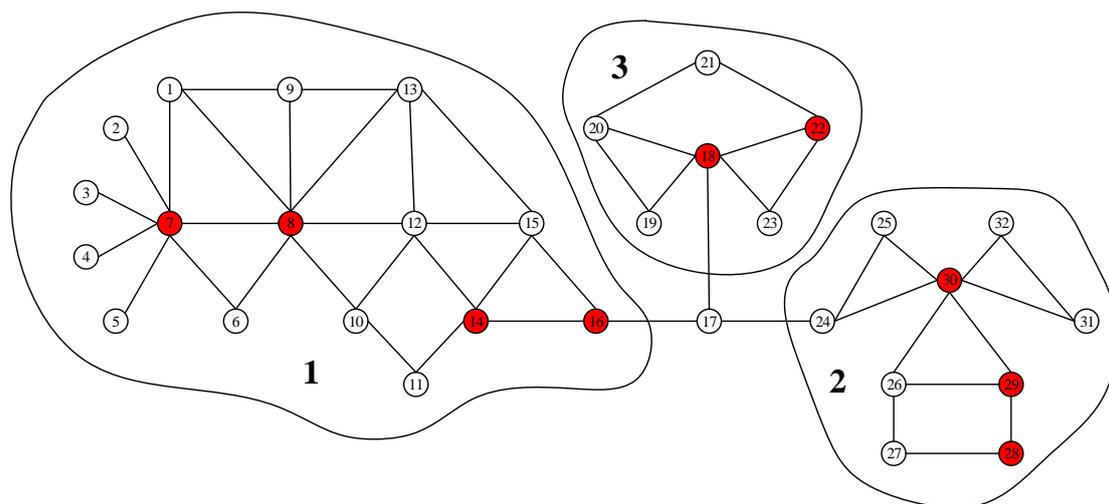
Figure 2. This network consists of three communities at different scales. The first nine nodes selected by EnRenew are marked red. The network typically shows the rich club phenomenon, that is, nodes with large degree tend to be connected together.

Table 1. Experiment results on example network shown in Figure 2. The first three nodes selected by EnRenew are distributed in three communities. The best results of Community Distribution, Average Distance and $F(t_c)$ are marked in bold.

Method	Initial Spreading Nodes									Community Distribution			Average Distance	$F(t_c)$
	1	2	3	4	5	6	7	8	9	1	2	3		
Adaptive Degree [58]	8	30	7	18	12	16	13	26	24	5	3	1	3.64	18.22
PageRank [42]	7	8	30	18	12	13	26	29	22	**4**	**3**	**2**	4.33	17.56
h-index [29]	18	12	13	17	9	8	30	16	15	6	1	1	2.75	15.64
k-shell [33]	7	8	30	18	12	13	10	26	17	5	2	1	3.72	17.97
VoteRank [61]	7	30	8	18	12	14	26	22	12	5	2	2	3.94	18.36
Greedy [58]	17	6	27	21	16	32	1	11	23	4	2	2	**4.92**	**21.33**
EnRenew (*l* = 2)	**8**	**30**	**18**	7	16	22	14	28	29	**4**	**3**	**2**	4.27	18.81

Table 2 shows the experiment results when choosing 9 nodes as the initial spreading set. Greedy method is usually used as the upper bound, but it is not efficient in large networks due to its high time complexity. EnRenew and PageRank distribute 4 nodes in community 1, 3 nodes in community 2, and 1 node in community 3. The distribution matches the size of community. However, the nodes selected by the other algorithms tend to cluster in community 1 except for greedy method. This will induce spreading within high density area, which is not efficient to spread in the entire network. EnRenew and PageRank can adaptively allocate reasonable number of nodes based on the size of the community just as Greedy method. Nodes selected by EnRenew have the second largest average distance except Greedy, which indicates EnRenew tends to distribute nodes sparsely in the graph. It aptly alleviates the adverse effect of spreading caused by the rich club phenomenon. Although EnRenew's average distance is smaller than PageRank, it has a higher final infected scale $F(t_c)$. Test result on PageRank also indicates that just select nodes widely spread across the network may not induce to a larger influence range. EnRenew performs the closest to Greedy with a low computational cost. It shows the proposed algorithm's effectiveness to maximize influence with limited nodes.

Table 2. Topological features of networks.

Networks	n	m	$\langle k \rangle$	k_{max}	$\langle c \rangle$
CEnew	453	2025	8.94	237	0.646
Email	1133	5451	9.62	71	0.22
Hamster	2426	16631	13.711	273	0.538
Router	5022	6258	2.492	106	0.012
Condmat	23133	93497	8.083	281	0.633
Amazon	334863	925872	5.530	549	0.397

Note: n and m are the total number of nodes and edges, respectively, and $\langle k \rangle = \frac{2*m}{n}$ stands for average node degree and $k_{max} = \max_{v \in V} d_v$ is the max degree in the network and Average clustering coefficient $\langle c \rangle$ measures the degree of aggregation in the network. $\langle c \rangle = \frac{1}{n} \sum_{i=1}^{n} \frac{2*I_i}{|\Gamma_i|*(|\Gamma_i|-1)}$, where I_i denotes the number of edges between direct neighbors of node i.

3.2. Data Description

Table 2 describes six different networks varying from small to large-scale, which are used to evaluate the performance of the methods. CEnew [89] is a list of edges of the metabolic network of C.elegans. Email [90] is an Email user communication network. Hamster [91] is a network reflecting friendship and family links between users of the website http://www.hamsterster.com, where node and edge demonstrate the web user and relationship between two nodes, respectively. Router network [92] reflects the Internet topology at the router level. Condmat (Condense Matter Physics) [93] is a collaboration network of authors of scientific papers from the arXiv. It shows the author collaboration in papers submitted to Condense Matter Physics. A node in the network represents an author, and an edge between two nodes shows the two authors have collaboratively published papers. In the Amazon network [94], each node represents a product, and an edge between two nodes represents two products were frequently purchased together.

3.3. Analysis of Influence Range When Renewing

We firstly conduct experiments on the parameter l, which is the influence range when renewing the information entropy. If $l = 1$, only the direct neighbors' importance of selected node will be renewed, and if $l = 2$, the importance of 2-length reachable nodes will be renewed and so forth. The results with varying parameter l from 1 to 4 on four networks are shown in Figure 3.

It can be seen from Figure 3 that, when $l = 2$, the method gets the best performance in four of the six networks. In network Email, although the results when $l = 3$ and $l = 4$ are slightly better comparing with the case of $l = 2$, the running time increases sharply. Besides, the three degrees of influence (TDI) theory [95] also states that a individual's social influence is only within a relatively small range. Based on our experiments, we set the influence range parameter l at 2 in the preceding experiments.
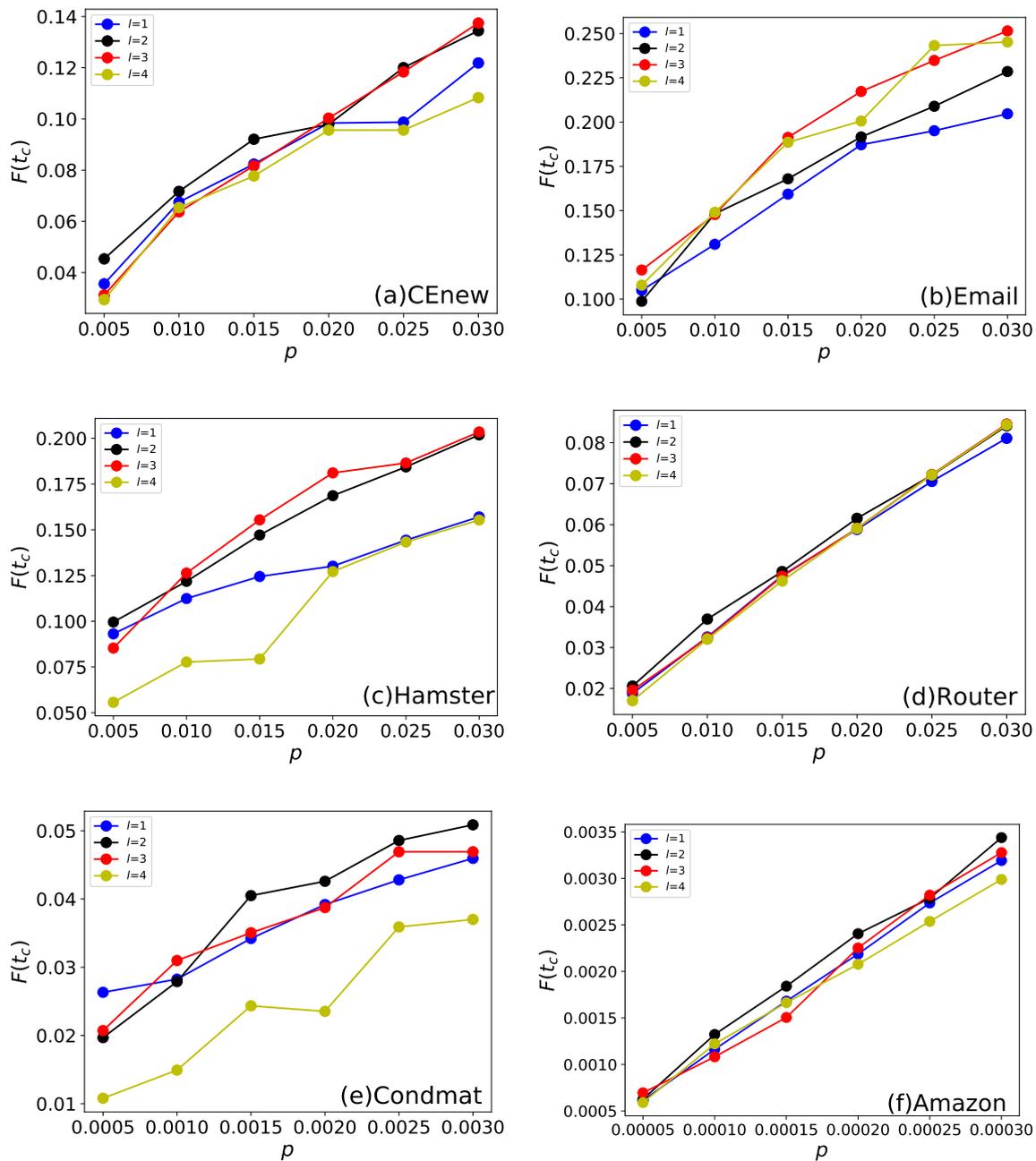
Figure 3. The figure shows how EnRenew's parameter l influences final affected scale $F(t_c)$ on the six networks. Each subfigure shows experiment results on one network. p is the ratio of initial infected nodes. The results are obtained by averaging on 100 independent runs with spread rate $\lambda = 1.5$ in SIR. With specific ratio of initial infected nodes p, larger final affected scale $F(t_c)$ means more reasonable of the parameter l. The best parameter l differs from different networks. In real life application, l can be used as an tuning parameter.

3.4. Comparation with Benchmark Algorithms

Many factors affect the final propagation scale in networks. A good influential nodes mining algorithm should prove its robustness in networks varying in structure, nodes size, initial infection set size, infection probability, and recovery probability. To evaluate the performance of EnRenew, VoteRank , Adaptive Degree, *k*-shell, PageRank, and h-index algorithms are selected as benchmark methods for comparing. Furthermore, greedy method is usually taken as upper bound on influence

maximization problem, but it is not practical on large networks due to its high time computational complexity. Thus, we added Greedy method as upper bound on the two small networks (CEnew and Email).
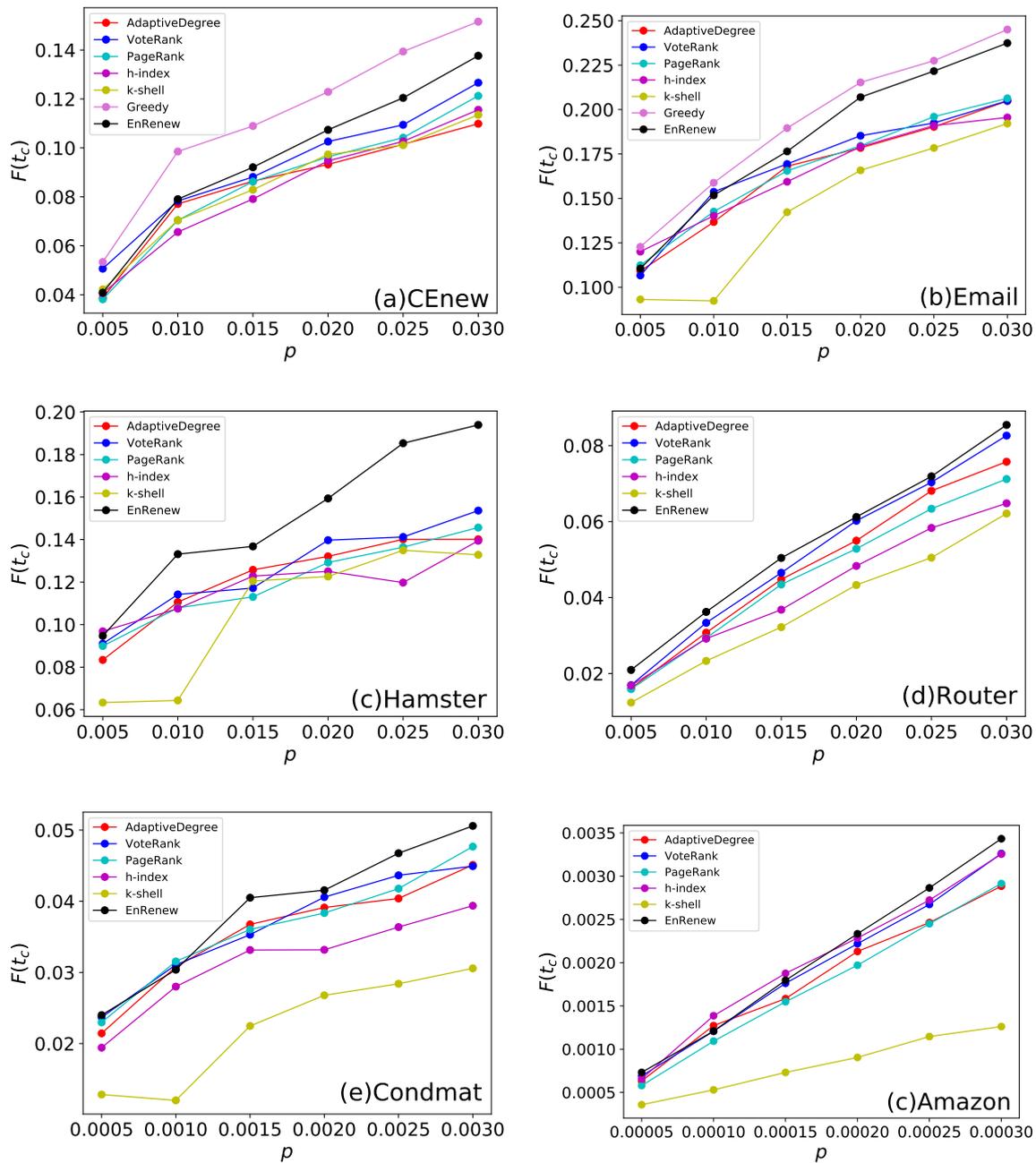
The final affected scale $F(t_c)$ of each method on different initial infected sizes are shown in Figure 4. It can be seen that EnRenew achieves an impressing result on the six networks. In the small network, such as CEnew and Email, EnRenew has an apparent better result on the other benchmark methods. Besides, it nearly reaches the upper bound on Email network. In Hamster network, it achieves a $F(t_c)$ of 0.22 only by ratio of 0.03 initial infected nodes, which is a huge improvement than all the other methods. In Condmat network, the number of affected nodes are nearly 20 times more than the initial ones. In a large Amazon network, 11 nodes will be affected on average for one selected initial infected node. But the algorithm performs unsatisfactory on network Router. All the methods did not yield good results due to the high sparsity structure of the network. In this sparse network, the information can hardly spread out with small number of initial spreaders. By comparing the 6 methods from the Figure 4, EnRenew surpasses all the other methods on five networks with nearly all kinds of p varying from small to large. This result reveals that when the size of initial infected nodes varies, EnRenew also shows its superiority to all the other methods. What is worth noticing is that EnRenew performs about the same as other methods when p is small, but it has a greater improvement with the rise of initial infected ratio p. This phenomenon shows the rationality of the importance renewing process. The renewing process of EnRenew would influence more nodes when p is larger. The better improvement of EnRenew than other methods shows the renewing process reasonability redistributes nodes' importance.

Timestep experiment is made to assess the propagation speed when given a fixed number of initial infected nodes. The exact results of $F(t)$ varying with time step t are shown in Figure 5. From the experiment, it can be seen that with same number of initial infected nodes, EnRenew always reaches a higher peak than the benchmark methods, which indicates a larger final infection rate. In the steady stage, EnRenew surpasses the best benchmark method by 21.1%, 7.0%, 30.0%, 5.0%, 2.5% and 9.0% in final affected scale on CEnew, Email, Hamster, Router, Condmat and Amazon networks, respectively. In view of propagation speed, EnRenew reaches the peak at about 300th time step in CEnew, 200th time step in Email, 400th time step in Hamster, 50th time step in Router, 400th time step in Condmat and 150th time step in Amazon. EnRenew always takes less time to influence the same number of nodes compared with other benchmark methods. From Figure 5, it can also be seen that k-shell also performs worst from the early stage in all the networks. Nodes with high core value tend to cluster together, which makes information hard to dissipate. Especially in the Amazon network, after 100 timesteps, all other methods reach a $F(t)$ of 0.0028, which is more than twice as large as k-shell. In contrast to k-shell, EnRenew spreads the fastest from early stage to the steady stage. It shows that the proposed method not only achieve a larger final infection scale, but also have a faster infection rate of propagation.

In real life situations, the infected rate λ varies greatly and has huge influence on the propagation procedure. Different λ represents virus or information with different spreading ability. The results on different λ and methods are shown in Figure 6. From the experiments, it can be observed that in most of cases, EnRenew surpasses all other algorithms with λ varying from 0.5 to 2.0 on all networks. Besides, experiment results on CEnew and Email show that EnRenew nearly reaches the upper bound. It shows EnRenew has a stronger generalization ability comparing with other methods. Especially, the EnRenew shows its impressing superiority in strong spreading experiments when λ is large.

Generally speaking, if the selected nodes are widely spread in the network, they tend to have an extensive impact influence on information spreading in entire network. L_S is used to measure dispersity of initial infected nodes for algorithms. Figure 7 shows the results of L_S of nodes selected by different algorithms on 6 different networks. It can be seen that, except for the Amazon network, EnRenew always has the largest L_S, indicting the widespread of selected nodes. Especially in CEnew, EnRenew performs far beyond all the other methods as its L_S is nearly as large as the upper bound.

In regard to the large-scale Amazon network, the network contains lots of small cliques and *k*-shell selects the dispersed cliques, which makes *k*-shell has the largest L_S. But other experimental results of *k*-shell show a poor performance. This further confirms that EnRenew does not naively distribute selected nodes widely across the network, but rather based on the potential propagation ability of each node.
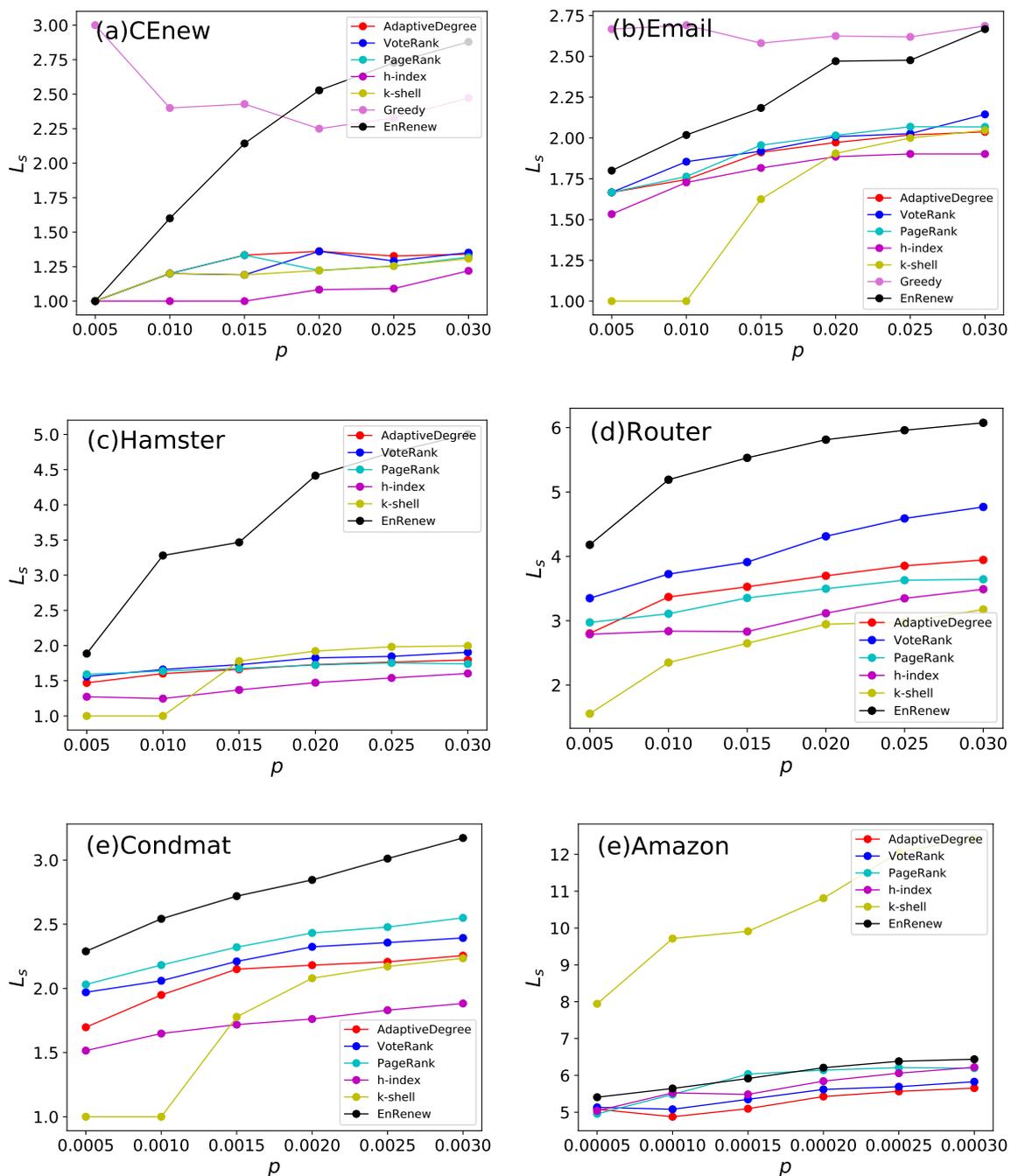

Figure 4. This experiment compares different methods by final affected scale $F(t_c)$ on the six networks. Each subfigure shows experiment results on one network. p is the ratio of initial infected nodes. The results are obtained by averaging on 100 independent runs with spread rate $\lambda = 1.5$ in Susceptible-Infected-Recovered (SIR). With specific ratio of initial spreading nodes p, larger final affected scale $F(t_c)$ indicates that the selected nodes are more advantageous to spreading. It can be seen that EnRenew surpasses all the other benchmark methods on the six networks. On the two small network, EnRenew nearly reaches the upper bound.
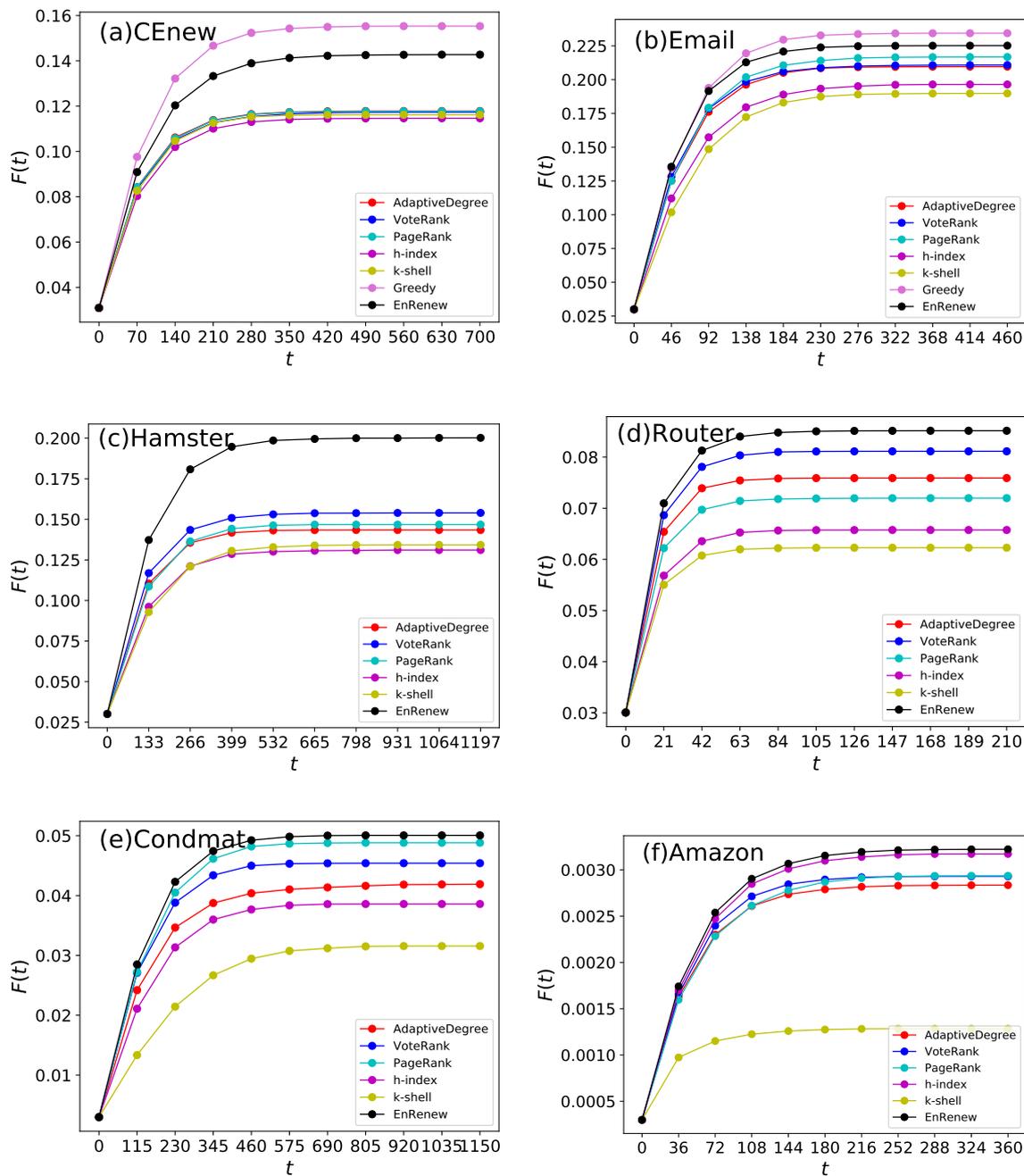
Figure 5. This experiment compares different methods regard to spreading speed. Each subfigure shows experiment results on one network. The ratio of initial infected nodes is 3% for CEnew, Email, Hamster and Router, 0.3% for Condmat and 0.03% for Amazon. The results are obtained by averaging on 100 independent runs with spread rate $\lambda = 1.5$ in SIR. With the same spreading time t, larger $F(t)$ indicates larger influence scale in network, which reveals a faster spreading speed. It can be seen from the figures that EnRenew spreads apparently faster than other benchmark methods on all networks. On the small network CEnew and Email, EnRenew's spreading speed is close to the upper bound.
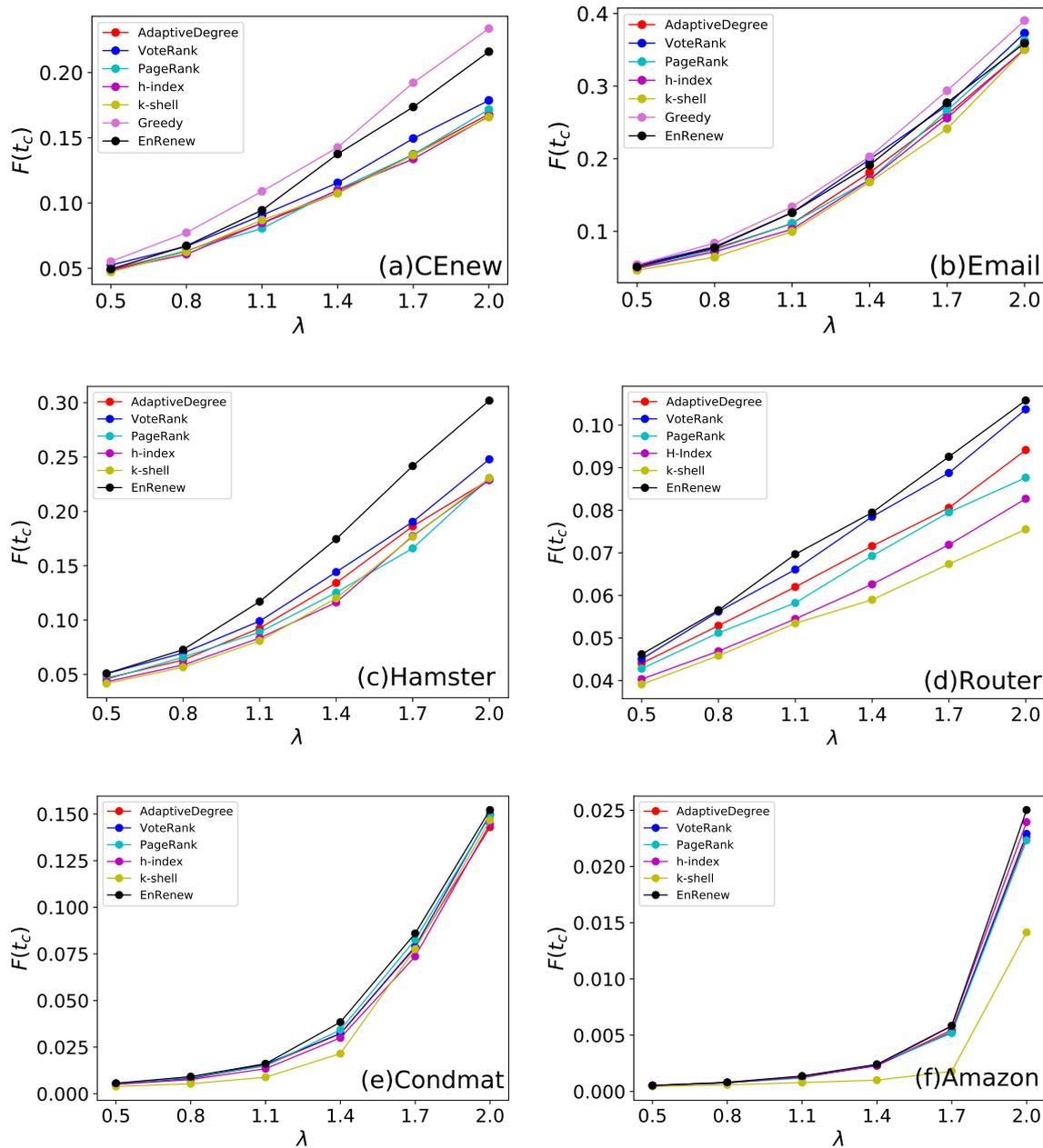
Figure 6. This experiment tests algorithms' effectiveness on different spreading conditions. Each subfigure shows experiment results on one network. The ratio of initial infected nodes is 3% for CEnew, Email, Hamster and Router, 0.3% for Condmat, and 0.03% for Amazon. The results are obtained by averaging on 100 independent runs. Different infected rate λ of SIR can imitate different spreading conditions. EnRenew gets a larger final affected scale $F(t_c)$ on different λ than all the other benchmark methods, which indicates the proposed algorithm has more generalization ability to different spreading conditions.

Figure 7. This experiment analysis average shortest path length L_S of nodes selected by different algorithms. Each subfigure shows experiment results on one network. p is the ratio of initial infected nodes. Generally speaking, larger L_S indicates the selected nodes are more sparsely distributed in network. It can be seen that nodes selected by EnRenew have the apparent largest L_S on five networks. It shows EnRenew tends to select nodes sparsely distributed.

4. Conclusions

The influential nodes identification problem has been widely studied by scientists from computer science through to all disciplines [96–100]. Various algorithms that have been proposed aim to solve peculiar problems in this field. In this study, we proposed a new method named EnRenew by introducing entropy into a complex network, and the SIR model was adopted to evaluate the algorithms. Experimental results on 6 real networks, varying from small to large in size, show that EnRenew is

superior over state-of-the-art benchmark methods in most of cases. Besides, with its low computational complexity, the presented algorithm can be applied to large scale networks. The EnRenew proposed in this paper can also be well applied in rumor controlling, advertise targeting, and many other related areas. But, for influential nodes identification, there still remain many challenges from different perspectives. From the perspective of network size, how to mine influential spreaders in large-scale networks efficiently is a challenging problem. In the area of time-varying networks, most of these networks are constantly changing, which poses the challenge of identifying influential spreaders since they could shift with the changing topology. In the way of multilayer networks, it contains information from different dimensions with interaction between layers and has attracted lots of research interest [101–103]. To identify influential nodes in multilayer networks, we need to further consider the method to better combine information from different layers and relations between them.

Author Contributions: Conceptualization, C.G. and D.C.; methodology, C.G. and L.Y.; software, C.G. and L.Y.; validation, C.G., L.Y. and J.M.; formal analysis, C.G. and L.Y.; investigation, X.C. and H.G.; resources, X.C. and D.C.; writing–original draft preparation, L.Y.; writing–review and editing, C.G., L.Y. and J.M.; visualization, C.G. and X.C.; supervision, D.C. and H.G.; funding acquisition, X.C. and D.C. All authors have read and agreed to the published version of the manuscript.

References

1. Silva, L.N.; Malacarne, A.; Silva, J.W.S.; Kirst, F.V.; Debortoli, R. The Scientific Collaboration Networks in University Management in Brazil. *Creat. Educ.* **2018**, *9*, 1469–1483. [CrossRef]
2. Guimera, R.; Danon, L.; Diazguilera, A.; Giralt, F.; Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **2003**, *68*, 065103. [CrossRef]
3. Sathyapriya, R.; Vijayabaskar, M.; Vishveshwara, S. Insights into protein–DNA interactions through structure network analysis. *PLoS Comput. Biol.* **2008**, *4*, 1–15. [CrossRef] [PubMed]
4. Ghosh, S.; Banerjee, A.; Sharma, N.; Agarwal, S.; Ganguly, N.; Bhattacharya, S.; Mukherjee, A. Statistical analysis of the Indian railway network: A complex network approach. *Acta Phys. Pol. B Proc. Suppl.* **2011**, *4*, 123–138. [CrossRef]
5. Scott, J. Social network analysis. *Sociology* **1988**, *22*, 109–127. [CrossRef]
6. Borgatti, S.P.; Mehra, A.; Brass, D.J.; Labianca, G. Network analysis in the social sciences. *Science* **2009**, *323*, 892–895. [CrossRef] [PubMed]
7. Vidmer, A.; Zeng, A.; Medo, M.; Zhang, Y.C. Prediction in complex systems: The case of the international trade network. *Phys. A Stat. Mech. Its Appl.* **2015**, *436*, 188–199. [CrossRef]
8. Leskovec, J.; Adamic, L.A.; Huberman, B.A. The dynamics of viral marketing. *ACM Trans. Web (TWEB)* **2007**, *1*, 5. [CrossRef]
9. Kimura, M.; Saito, K.; Nakano, R.; Motoda, H. Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Discov.* **2010**, *20*, 70. [CrossRef]
10. Csermely, P.; Korcsmaros, T.; Kiss, H.; London, G.; Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacol. Ther.* **2013**, *138*, 333–408. [CrossRef]
11. Cohen, R.; Havlin, S.; Benavraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **2003**, *91*, 247901. [CrossRef] [PubMed]
12. Kumar, P.; Verma, P.; Singh, A. A Study of Epidemic Spreading and Rumor Spreading over Complex Networks. In *Towards Extensible and Adaptable Methods in Computing*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 131–143.
13. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925. [CrossRef]
14. Wang, W.; Tang, M.; Stanley, H.E.; Braunstein, L.A. Unification of theoretical approaches for epidemic spreading on complex networks. *Rep. Prog. Phys.* **2017**, *80*, 036603. [CrossRef] [PubMed]

15. Ren, G.; Wang, X. Epidemic spreading in time-varying community networks. *Chaos Interdiscip. J. Nonlinear Sci.* **2014**, *24*, 023116. [CrossRef]

16. Zhang, H.F.; Xie, J.R.; Tang, M.; Lai, Y.C. Suppression of epidemic spreading in complex networks by local information based behavioral responses. *Chaos Interdiscip. J. Nonlinear Sci.* **2014**, *24*, 043106. [CrossRef]

17. Cui, A.X.; Wang, W.; Tang, M.; Fu, Y.; Liang, X.; Do, Y. Efficient allocation of heterogeneous response times in information spreading process. *Chaos Interdiscip. J. Nonlinear Sci.* **2014**, *24*, 033113. [CrossRef]

18. Borge-Holthoefer, J.; Moreno, Y. Absence of influential spreaders in rumor dynamics. *Phys. Rev. E* **2012**, *85*, 026116. [CrossRef]

19. Wu, J.; Zheng, M.; Zhang, Z.K.; Wang, W.; Gu, C.; Liu, Z. A model of spreading of sudden events on social networks. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 033113. [CrossRef]

20. Dietz, K.; Heesterbeek, J. Daniel Bernoulli?s epidemiological model revisited. *Math. Biosci.* **2002**, *180*, 1–21. [CrossRef]

21. Fine, P.E. Herd immunity: History, theory, practice. *Epidemiol. Rev.* **1993**, *15*, 265–302. [CrossRef]

22. Hamer, W.H. *Epidemic Disease in England: The Evidence of Variability and of Persistency of Type*; Bedford Press: London, UK, 1906.

23. Anderson, R.M.; Anderson, B.; May, R.M. *Infectious Diseases of Humans: Dynamics and Control*; Oxford University Press: Oxford, UK, 1992.

24. Harding, N.; Nigmatullin, R.; Prokopenko, M. Thermodynamic efficiency of contagions: A statistical mechanical analysis of the SIS epidemic model. *Interface Focus* **2018**, *8*, 20180036. [CrossRef] [PubMed]

25. Wang, C.; Tan, Z.X.; Ye, Y.; Wang, L.; Cheong, K.H.; Xie, N.g. A rumor spreading model based on information entropy. *Sci. Rep.* **2017**, *7*, 1–14. [CrossRef] [PubMed]

26. Zenil, H.; Kiani, N.A.; Marabita, F.; Deng, Y.; Elias, S.; Schmidt, A.; Ball, G.; Tegnér, J. An algorithmic information calculus for causal discovery and reprogramming systems. *iScience* **2019**, *19*, 1160–1172. [CrossRef] [PubMed]

27. Brockmann, D.; Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **2013**, *342*, 1337–1342. [CrossRef]

28. Everett, M.G.; Borgatti, S.P. Extending centrality. *Model. Methods Soc. Netw. Anal.* **2005**, *35*, 57–76.

29. Lü, L.; Zhou, T.; Zhang, Q.M.; Stanley, H.E. The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **2016**, *7*, 10168. [CrossRef]

30. Chen, D.; Lü, L.; Shang, M.; Zhang, Y.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 1777–1787. [CrossRef]

31. Chen, D.B.; Gao, H.; Lü, L.; Zhou, T. Identifying influential nodes in large-scale directed networks: The role of clustering. *PLoS ONE* **2013**, *8*, e77455. [CrossRef]

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

33. 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*, 888–893. [CrossRef]

34. Zeng, A.; Zhang, C. Ranking spreaders by decomposing complex networks. *Phys. Lett. A* **2013**, *377*, 1031–1035. [CrossRef]

35. Hage, P.; Harary, F. Eccentricity and centrality in networks. *Soc. Networks* **1995**, *17*, 57–63. [CrossRef]

36. Sabidussi, G. The centrality index of a graph. *Psychometrika* **1966**, *31*, 581–603. [CrossRef] [PubMed]

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

38. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [CrossRef]

39. Wittenbaum, G.M.; Hubbell, A.P.; Zuckerman, C. Mutual enhancement: Toward an understanding of the collective preference for shared information. *J. Personal. Soc. Psychol.* **1999**, *77*, 967. [CrossRef]

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

41. Poulin, R.; Boily, M.C.; Mâsse, B.R. Dynamical systems to define centrality in social networks. *Soc. Networks* **2000**, *22*, 187–220. [CrossRef]

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

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

44. Nie, T.; Zheng, G.; Zhao, K.; Lu, Z.M. Using mapping entropy to identify node centrality in complex networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *453*, 290–297. [CrossRef]

45. Chen, D.; Xiao, R.; Zeng, A.; Zhang, Y. Path diversity improves the identification of influential spreaders. *EPL* **2013**, *104*, 68006. [CrossRef]

46. Qiao, T.; Shan, W.; Zhou, C. How to identify the most powerful node in complex networks? A novel entropy centrality approach. *Entropy* **2017**, *19*, 614. [CrossRef]

47. Qiao, T.; Shan, W.; Yu, G.; Liu, C. A novel entropy-based centrality approach for identifying vital nodes in weighted networks. *Entropy* **2018**, *20*, 261. [CrossRef]

48. Ai, X. Node importance ranking of complex networks with entropy variation. *Entropy* **2017**, *19*, 303. [CrossRef]

49. Li, Y.; Cai, W.; Li, Y.; Du, X. Key Node Ranking in Complex Networks: A Novel Entropy and Mutual Information-Based Approach. *Entropy* **2020**, *22*, 52. [CrossRef]

50. Fei, L.; Deng, Y. A new method to identify influential nodes based on relative entropy. *Chaos Solitons Fractals* **2017**, *104*, 257–267. [CrossRef]

51. Zareie, A.; Sheikhahmadi, A.; Fatemi, A. Influential nodes ranking in complex networks: An entropy-based approach. *Chaos Solitons Fractals* **2017**, *104*, 485–494. [CrossRef]

52. Shetty, J.; Adibi, J. Discovering important nodes through graph entropy the case of enron email database. In Proceedings of the 3rd International Workshop on Link Discovery, Chicago, IL, USA, 21–25 August 2005; pp. 74–81.

53. Wenli, F.; Zhigang, L.; Ping, H. Identifying node importance based on information entropy in complex networks. *Phys. Scr.* **2013**, *88*, 065201. [CrossRef]

54. Hu, P.; Mei, T. Ranking influential nodes in complex networks with structural holes. *Phys. A Stat. Mech. Its Appl.* **2018**, *490*, 624–631. [CrossRef]

55. Wang, Z.; Du, C.; Fan, J.; Xing, Y. Ranking influential nodes in social networks based on node position and neighborhood. *Neurocomputing* **2017**, *260*, 466–477. [CrossRef]

56. Colizza, V.; Flammini, A.; Serrano, M.; Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2006**, *2*, 110–115. [CrossRef]

57. 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.

58. Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 199–208.

59. Borgatti, S.P. Identifying sets of key players in a social network. *Comput. Math. Organ. Theory* **2006**, *12*, 21–34. [CrossRef]

60. Narayanam, R.; Narahari, Y. A shapley value-based approach to discover influential nodes in social networks. *IEEE Trans. Autom. Sci. Eng.* **2010**, *8*, 130–147. [CrossRef]

61. Zhang, J.X.; Chen, D.; Dong, Q.; Zhao, Z.D. Identifying a set of influential spreaders in complex networks. *Sci. Rep.* **2016**, *6*, 27823. [CrossRef]

62. Zhao, X.; Huang, B.; Tang, M.; Zhang, H.; Chen, D. Identifying effective multiple spreaders by coloring complex networks. *EPL* **2014**, *108*, 68005. [CrossRef]

63. Hu, Z.; Liu, J.; Yang, G.; Ren, Z. Effects of the distance among multiple spreaders on the spreading. *EPL* **2014**, *106*, 18002. [CrossRef]

64. Guo, L.; Lin, J.; Guo, Q.; Liu, J. Identifying multiple influential spreaders in term of the distance-based coloring. *Phys. Lett. A* **2016**, *380*, 837–842. [CrossRef]

65. Bao, Z.; Liu, J.; Zhang, H. Identifying multiple influential spreaders by a heuristic clustering algorithm. *Phys. Lett. A* **2017**, *381*, 976–983. [CrossRef]

66. Zhou, H.J. Spin glass approach to the feedback vertex set problem. *Eur. Phys. J. B* **2013**, *86*, 1–9. [CrossRef]

67. Ji, S.; Lü, L.; Chi, H.Y.; Hu, Y. Effective spreading from multiple leaders identified by percolation in the susceptible-infected-recovered (SIR) model. *New J. Phys.* **2017**, *19*, 073020. [CrossRef]

68. Li, R.; Qin, L.; Yu, J.X.; Mao, R. Finding influential communities in massive networks. *Int. J. Very Large Data Bases* **2017**, *26*, 751–776. [CrossRef]

69. Bozorgi, A.; Samet, S.; Kwisthout, J.; Wareham, T. Community-based influence maximization in social networks under a competitive linear threshold model. *Knowl. Based Syst.* **2017**, *134*, 149–158. [CrossRef]

70. Lv, J.; Yang, B.; Yang, Z.; Zhang, W. A community-based algorithm for influence blocking maximization in social networks. *Cluster Comput.* **2017**, *22*, 5587–5602. [CrossRef]

71. Pan, Y.; Li, D.; Liu, J.; Liang, J. Detecting community structure in complex networks via node similarity. *Phys. A Stat. Mech. Its Appl.* **2010**, *389*, 2849–2857. [CrossRef]

72. Tang, L.; Li, S.; Lin, J.; Guo, Q.; Liu, J. Community structure detection based on the neighbor node degree information. *Int. J. Mod. Phys. C* **2016**, *27*, 1–11. [CrossRef]

73. Wang, Y.; Cong, G.; Song, G.; Xie, K. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–28 July 2010; pp. 1039–1048.

74. Zhang, X.; Zhu, J.; Wang, Q.; Zhao, H. Identifying influential nodes in complex networks with community structure. *Knowl.-Based Syst.* **2013**, *42*, 74–84. [CrossRef]

75. Gong, M.; Song, C.; Duan, C.; Ma, L.; Shen, B. An efficient memetic algorithm for influence maximization in social networks. *IEEE Comput. Intell. Mag.* **2016**, *11*, 22–33. [CrossRef]

76. Chen, Y.C.; Peng, W.C.; Lee, S.Y. Efficient algorithms for influence maximization in social networks. *Knowl. Inf. Syst.* **2012**, *33*, 577–601. [CrossRef]

77. Hu, Y.; Ji, S.; Jin, Y.; Feng, L.; Stanley, H.E.; Havlin, S. Local structure can identify and quantify influential global spreaders in large scale social networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7468–7472. [CrossRef]

78. Chen, D.; Sun, H.; Tang, Q.; Tian, S.; Xie, M. Identifying influential spreaders in complex networks by propagation probability dynamics. *Chaos* **2019**, *29*, 033120. [CrossRef] [PubMed]

79. Erkol, I.; Castellano, C.; Radicchi, F. Systematic comparison between methods for the detection of influential spreaders in complex networks. *Sci. Rep.* **2019**, *9*, 1–11. [CrossRef] [PubMed]

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

81. Zhao, L.; Cui, H.; Qiu, X.; Wang, X.; Wang, J. SIR rumor spreading model in the new media age. *Phys. A Stat. Mech. Its Appl.* **2013**, *392*, 995–1003. [CrossRef]

82. Ouboter, T.; Meester, R.; Trapman, P. Stochastic SIR epidemics in a population with households and schools. *J. Math. Biol.* **2016**, *72*, 1177–1193. [CrossRef]

83. Castellano, C.; Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **2010**, *105*, 218701. [CrossRef]

84. He, J.L.; Fu, Y.; Chen, D.B. A novel top-k strategy for influence maximization in complex networks with community structure. *PLoS ONE* **2015**, *10*, e0145283. [CrossRef]

85. Namtirtha, A.; Dutta, A.; Dutta, B. Identifying influential spreaders in complex networks based on kshell hybrid method. *Phys. A Stat. Mech. Its Appl.* **2018**, *499*, 310–324. [CrossRef]

86. Yu, H.; Cao, X.; Liu, Z.; Li, Y. Identifying key nodes based on improved structural holes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2017**, *486*, 318–327. [CrossRef]

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

88. Jiang, L.; Zhao, X.; Ge, B.; Xiao, W.; Ruan, Y. An efficient algorithm for mining a set of influential spreaders in complex networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *516*, 58–65. [CrossRef]

89. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabási, A.L. The large-scale organization of metabolic networks. *Nature* **2000**, *407*, 651. [CrossRef]

90. Kunegis, J. Konect: The koblenz network collection. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 1343–1350.

91. 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; pp. 4292–4293.

92. Spring, N.; Mahajan, R.; Wetherall, D. Measuring ISP topologies with Rocketfuel. In Proceedings of the ACM SIGCOMM Computer Communication Review, Pittsburgh, PA, USA, 19–23 August 2002; Volume 32, pp. 133–145.

93. Leskovec, J.; Kleinberg, J.; Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data (TKDD)* **2007**, *1*, 2. [CrossRef]

94. Yang, J.; Leskovec, J. Defining and evaluating network communities based on ground-truth. *Knowl. Inf. Syst.* **2012**, *42*, 181–213. [CrossRef]

95. Christakis, N.A.; Fowler, J.H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **2007**, *357*, 370–379. [CrossRef]

96. Zhong, L.; Shang, M.; Chen, X.; Cai, S. Identifying the influential nodes via eigen-centrality from the differences and similarities of structure. *Phys. A Stat. Mech. Its Appl.* **2018**, *510*, 77–82. [CrossRef]

97. Yang, Y.; Wang, Z.; Pei, J.; Chen, E. Tracking Influential Individuals in Dynamic Networks. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2615–2628. [CrossRef]

98. Zhao, X.; Liu, F.; Wang, J.; Li, T. Evaluating influential nodes in social networks by local centrality with a coefficient. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 35. [CrossRef]

99. Raj, E.D.; Babu, L.D.D. A survey on topological properties, network models and analytical measures in detecting influential nodes in online social networks. *Int. J. Web Based Communities* **2017**, *13*, 137–156. [CrossRef]

100. Erkol, Ş.; Faqeeh, A.; Radicchi, F. Identifying influential spreaders in noisy networks. *arXiv* **2018**, arXiv:1803.02253.

101. Salehi, M.; Sharma, R.; Marzolla, M.; Magnani, M.; Siyari, P.; Montesi, D. Spreading processes in multilayer networks. *IEEE Trans. Netw. Sci. Eng.* **2015**, *2*, 65–83. [CrossRef]

102. 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]

103. 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]