

Citation: Wang H, Zhang Y, Zhang Z, Mahadevan S, Deng Y (2015) *Physarum*Spreader: A New Bio-Inspired Methodology for Identifying Influential Spreaders in Complex Networks. PLoS ONE 10(12): e0145028. doi:10.1371/journal.pone.0145028

Editor: Satoru Hayasaka, Wake Forest School of Medicine, UNITED STATES

Received: November 12, 2014

Accepted: November 27, 2015

Published: December 18, 2015

Copyright: © 2015 Wang et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work is partially supported by National Natural Science Foundation of China (Grant Nos. 61174022, 61573290, 61503237), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20131102130002), R&D Program of China (2012BAH07B01), National High Technology Research and Development Program of China (863 Program) (Grant No. 2013AA013801), China State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No. BUAA-VR-14KF-02), the General Research **RESEARCH ARTICLE**

*Physarum*Spreader: A New Bio-Inspired Methodology for Identifying Influential Spreaders in Complex Networks

Hongping Wang^{1®}, Yajuan Zhang^{1®}, Zili Zhang^{1,2}, Sankaran Mahadevan³, Yong Deng¹*

1 School of Computer and Information Science, Southwest University, Chongqing, China, 2 School of Information Technology, Deakin University, Geelong, VIC, Australia, 3 Civil and Environmental Engineering Department, Vanderbilt University, Nashville, TN, United States of America

• These authors contributed equally to this work.

* ydeng@swu.edu.cn; prof.deng@hotmail.com

Abstract

Identifying influential spreaders in networks, which contributes to optimizing the use of available resources and efficient spreading of information, is of great theoretical significance and practical value. A random-walk-based algorithm *LeaderRank* has been shown as an effective and efficient method in recognizing leaders in social network, which even outperforms the well-known PageRank method. As LeaderRank is initially developed for binary directed networks, further extensions should be studied in weighted networks. In this paper, a generalized algorithm *Physarum*Spreader is proposed by combining LeaderRank with a positive feedback mechanism inspired from an amoeboid organism called *Physarum Polycephalum*. By taking edge weights into consideration and adding the positive feedback mechanism, *Physarum*Spreader is applicable in both directed and undirected networks with weights. By taking two real networks for examples, the effectiveness of the proposed method is demonstrated by comparing with other standard centrality measures.

Introduction

Over the years, the study of graphs and networks have drawn increasing attention in a wide variety of scientific disciplines, such as biology, computer science, economics, mathematics and sociology. Meanwhile, network analysis, as a key tool to map and measure those network entities and their connections, has been well studied to provide a visual and mathematical view of networks. However, a major challenge it encounters is how to identify the most efficient spreaders for optimizing the use of available resources and ensuring efficient spread of information [1-5]. With different topology and relations, node centrality can be endowed with various meanings, such as influence [6], importance [7, 8], popularity [9, 10], controllability [11, 12] and spreading efficiency [1]. Studies on the importance of a node for spreading can be of great significance in controlling rumor, disease spreading and information flow in networks [13–16].

Program of the Science Supported by Sichuan Provincial Department of Education (Grant No. 14ZB0322).

PLOS ONE

Competing Interests: The authors have declared that no competing interests exist.

Ever since the idea of centrality was introduced, various centrality measures have been proposed to identify nodes which are more central than others [17-19]. Recently, Landherr et al. [20] conducted a critical review of five common centrality measures in social networks, including degree centrality [17], closeness centrality [17], betweenness centrality [17], eigenvector centrality [21] and Katz's centrality [22]. Degree, the simplest centrality measure, is the number of edges that a node is connected to, which was firstly proposed by Freeman. However, a node with higher degree might not be in a position which can access resources quicker. In order to make up this drawback, a more sophisticated centrality measure *closeness* is developed, which was defined as the inverse total length from a node to all other nodes. Another important centrality measure is *betweenness*, which is calculated as the fraction of the times that a node lies on the shortest paths over the total number of the shortest paths. Besides, by introducing random walks, a revised betweenness centrality is proposed in [23], which counts the frequency of a node traversed by a random walk between two other nodes. And then, a random-walk-based centrality called *LeaderRank* [9] has been proposed, which can identify leaders in social networks better than the well-known PageRank algorithm. After that, Chen et al. [6] developed a semi-local centrality measure as a tradeoff between local degree centrality and other global but time-consuming measures. Ranking influential nodes can be seen as a multi-attribute decision making problem [24]. Due to the efficiency to combine different data [25-29], evidence theory is also widely used for identifying influential spreaders in complex networks [30, 31].

However, those measures described above are only suitable in binary networks. In many real networks, edges are with some form of attributes or weights, rather than simply either present or absent in a pair of nodes. If only binary networks are considered, ignoring the intrinsical weights attached to edges, plenty of valuable information has been lost and the analysis cannot be accurate and comprehensive. Thus, many researchers have turned their attention to centrality for weighted networks [32–36].

In 1991, Freeman et al. [37] introduced a new measure of centrality based on the concept of network flows, flow betweenness, which considered all the independent paths between all pairs of nodes in the network. Based on the degree centrality, Barrat et al. [38] proposed a measure for weighted networks, which is the sum of weights of edges that a node is connected to. Beyond that, Newman [32] and Brandes [39] have generalized the closeness and betweenness centrality for weighted networks by using Dijkstra's algorithm [40] on computing the shortest paths. By taking both edge weights and the number of edges into consideration, a new generalization was proposed by Opsahl et al. [33], using a tuning parameter to balance the relative importance between those two parts. Later, Qi et al. [34] developed a Laplacian centrality method considering "intermediate" environmental information around a node.

In this paper, we proposed a generalized centrality metric, called *Physarum*Spreader, to identify nodes with high spreading performance in networks. *Physarum*Spreader is developed on the basis of a random-walk-based algorithm LeaderRank and a positive feedback mechanism inspired from an amoeboid organism, called *Physarum Polycephalum*. With the integration of the algorithm and mechanism, our *Physarum*Spreader is applicable in both directed and undirected networks with weights. It overcomes the shortcomings of LeaderRank which is only designed for binary network and does not work well for undirected networks. Furthermore, a susceptible-infected-removed (SIR) model is employed to examine the spreading performance of nodes identified by different centrality measures. With simulations on different networks and comparison with other centrality measures, it reveals that *Physarum*Spreader works well in identifying influential nodes with high spreading performance and good tolerance.

The rest of the paper is organized as follows. Section 2 begins with a brief introduction to LeaderRank algorithm and positive feedback mechanism of *Physarum Polycephalum* adopted in our method. Then, procedure of the proposed *Physarum*Spreader for identifying influential

spreaders in networks is depicted in Section 3. And two applications in real networks are presented in section 4. What's more, the spreading effectiveness and robustness are studied. Section 5 concludes the paper.

1 Basic Theory

1.1 LeaderRank for Identifying Leaders [9]

Given a directed network of N nodes and M edges, a *ground node* is then added by establishing bidirectional edges between it and all the other nodes, which assures the modified network as strongly connected. And the modified network consists of N + 1 nodes and M + 2N directed edges. Initially, each node in the network, except for the ground node, is assigned with one unit resource, while the ground node is assigned with no resource. And then each node evenly distributes its resource to neighbors along the outgoing edges. Next is to update resource distribution as summing up the resource each node derives from its incoming edges. This process of distribution and updating of resources continues until steady state is attained. The whole process can be described mathematically as follows.

Assuming $r_i(t)$ denotes the resource of node *i* at time *t*, the initial state (t = 0) of resource distribution can be represented as:

$$\begin{aligned} r_g(0) &= 0, \text{ where } g \text{ is ground node} \\ r_i(0) &= 1, \text{ where } i \in \{1, 2, \cdots, N+1\} \backslash \{g\} \end{aligned} \tag{1}$$

And each node can update its resource according to the following equation:

$$r_i(t+1) = \sum_{j=1}^{N+1} \frac{a_{ji}}{k_j^{out}} r_j(t)$$
(2)

where a_{ij} is the element of the corresponding (N + 1)-dimensional adjacency matrix, which equals 1 if there is a directed link from *j* to *i* and 0 otherwise, and k_i^{out} is the out-degree of node *j*.

When the resource $r_i(t)$ at all nodes converges to a unique steady state at time t_c , the resource at the ground nodes is then evenly distributed to all other nodes, and the final resource distribution on nodes *i* is:

$$R_i = r_i(t_c) + \frac{r_g(t_c)}{N}, \text{ where } i \in \{1, 2, \cdots, N\}$$

$$(3)$$

1.2 Physarum Model for Path Finding

Physarum polycephalum, as a large, single-celled amoeboid organism, can form a dynamic tubular network within the discovered food sources.

Recently, a large amoeboid organism, *Physarum polycephalum*, turned out to be capable of solving many graph theoretical problems [41-45], including finding the shortest path [46-48]. Furthermore, it has been shown experimentally that the network it generates is of high intelligence and performance in road-network [49] and great transport efficiency in vascular network [50], even comparable to or better than the Tokyo rail network [51]. During the process of its path finding and tube selection, *Physarum* will cut off those non-competing long tubes and reinforce shorter tubes. And, with the positive feedback mechanism among the tube length, the flux through tubes and the conductivity of tubes (tube width), shorter tubes result in larger flux; tube with a large flux grow (tube width increases); wider tubes leads to a further increase

of flux as the resistance to the flow decreases in wider tubes. According to *Physarum*'s tubular network, each tube segment is regarded as an edge e_{ij} in graph and its two ends are denoted as nodes *i* and *j*, which the edge connects. For each tube segment, there are two critical attributes: one is the length of the tube L_{ij} and the other is its thickness, which is always represented as conductivity D_{ij} . Based on the theory that thick, short tubes are typically the most effective for transportation, the mathematical model of *Physarum* includes two parts: flux through tubes and adaptation of conductivity according to its flux.

1.2.1 Flux through Tubes [52]. As circulation is based on streaming through network of tubular channels, the flux of sol through the tubes Q_{ij} is approximately modeled as Poiseuille flow:

$$Q_{ij} = \frac{D_{ij}}{L_{ij}} \left(p_i - p_j \right) \tag{4}$$

where p_i is the pressure at node *i*.

By considering the balance of flux through each node, we have

$$\sum_{i} \frac{D_{ij}}{L_{ij}} \left(p_i - p_j \right) = \begin{cases} -I_0, & \text{for } j = s, \\ +I_0, & \text{for } j = t, \\ 0, & \text{otherwise.} \end{cases}$$
(5)

where *s* is the source node that the initial flux I_0 flows out, while *t* is the sink node from which the flux flows in.

1.2.2 Adaptation of Conductivity [52]. In order to model the positive feedback mechanism that tube widens with increasing flux and degenerates with decreasing flux, the conductivity D_{ij} is assumed to change over time according to the flux Q_{ij} :

$$\frac{\mathrm{d}}{\mathrm{d}t}D_{ij} = f\left(|Q_{ij}|\right) - \gamma D_{ij} \tag{6}$$

where γ is a decay rate of the tube. f(Q) is an increasing function with f(0) = 0. More detailed description of f(Q) can be found in [53].

2 PhysarumSpreader

LeaderRank is an efficient method for identifying influential leaders in opinion spreading and outperforms PageRank algorithm, the basis of the Google search engine, in ranking effectiveness and robustness against manipulations and noisy data. However, it is initially designed for binary networks, which is not suitable for weighted networks.

Consider that the resource distribution from each node to its neighbors in LeaderRank, is similar to the flowing of flux through tubes in *Physarum* model of path finding. Nevertheless, *Physarum* model is designed for finding the shortest paths in both binary and weighted networks, which is capable of handling edge weights. Thus, it is natural to consider that adoption of the positive feedback mechanism between conductivity and flux in *Physarum* model may be of great help in overcoming the weakness of LeaderRank in weighted networks.

The main mechanism behind the presented method is combining the positive feedback mechanism in *Physarum* model and resources distribution mechanism in LeaderRank. Specifically, each node proportionally distributes resources to its outlinks based on their weights and conductivities. Then, a positive feedback mechanism is employed to accelerate convergence of the algorithm. Here, the positive feedback mechanism is an interaction between conductivities and resources along each link. A link with few resources leads to a weak conductivities. The

PLOS ONE

weak conductivities produce a further decrease of resources along the link. Similarly, a link with more resources cases stronger conductivities and further contributes itself to obtain more resources. Finally, the resources of a node are the sum of its inlinks' resources. This process will continue until each nodes' resources are steady.

Therefore, in this paper, based on the primary LeaderRank, an extended algorithm, called *Phy-sarumSpreader*, is proposed for capturing the spreading ability of nodes in weighted networks.

2.1 General Flow of PhysarumSpreader

The general flow of our proposed *Physarum*Spreader is described as follows, along with graphical demonstration of a random directed example network *Net* as shown in Fig 1.

Step 1 Add a ground node into the network by connecting every node through bidirectional links (Fig 2). The weight w_{ig} of the inlink which direction is form node $i(i \in N)$ to ground node is determined by the following equation:

$$w_{ig} = \begin{cases} \frac{\sum_{j=1}^{l_{i}^{out}} w_{ij}}{l_{i}^{out}}, & if \quad l_{i}^{out} > 0\\ 1, & if \quad l_{i}^{out} = 0 \end{cases}$$
(7)







PLOS ONE

where l_i^{Out} denotes the total number of outlinks from node *i* without considering weight and node *j* represents the neighbour of node *i*. If the network is directed, this step guarantees the network to be strongly connected.

Step 2 Initialize all nodes (other than the ground node) with unit of resource and the ground node with a score of 0 (Fig 3).

$$I_g(0) = 0, \text{ where } g \text{ is ground node}$$

$$I_i(0) = 1, \text{ where } i \in \{1, 2, \cdots, N+1\} \setminus \{g\}$$
(8)

Step 3 Distribute each node's flux to its neighbors through the out-going edges according to their edge weights.

$$Q_{ij}(t) = \frac{w_{ij}D_{ij}(t)}{\sum_{j=1}^{N+1} w_{ij}D_{ij}(t)}$$
(9)



Fig 3. Initialization of flux for nodes in Net.

PLOS ONE

where D_{ij} is the conductivity of each edge with initial value as 1. It is notable that the value of w_{ij} varies under different circumstances. If the given network is binary, $w_{ij} = 1$ for all edges in the network. If the network is weighted and the weight refers to the cost of traversing the edge, w_{ij} will be the reciprocal of the edge weight. But if the weight stands for the strength of edge relation, such as the number of social proximities, w_{ij} will be assigned as the weight. The initial flux distribution on each out-going edge of nodes at time t = 0 is calculated, shown in matrix $\mathbf{Q}(\mathbf{0})$:

	[0	0.5	0	0	0.5
	0	0	0.5	0	0.5
$\mathbf{Q}(0) =$	0.5	0	0	0	0.5
	0.33	0.33	0	0	0.33
	0.25	0.25	0.25	0.25	0

Step 4 Adapt the conductivity according the flux through each edge using the following equation:

$$D_{ij}(t+1) = \left[Q_{ij}(t) + D_{ij}(t)\right]/2, \text{ where } i, j \neq g$$
(10)

The edge conductivity at time t = 1 can be calculated by Eq 10, according to the flux distribution of time t = 0. The result is shown in matrix **D**(1):

	[⁰	0.75	0	0	1
	0	0	0.75	0	1
$\mathbf{D}(1) =$	0.75	0	0	0	1
	0.67	0.67	0	0	1
	1	1	1	1	1

Step 5 Update the resources of each node for the next iteration, according to the flux flowing into the node and its current score.

$$I_i(t+1) = \sum_{j=1}^{N+1} Q_{ji}(t) I_j(t)$$
(11)

Step 6 Determine whether the steady state of nodes' score is attained. If it converges to a steady state, the flux of the ground node is evenly distributed to all other nodes. And the final score *S_i* for each node's spreading performance is attainted as:

$$S_i = I_i(t_c) + \frac{I_g(t_c)}{N}, \text{ where } i \in \{1, 2, \cdots, N\}$$
 (12)

If the state is not steady yet, the process continues to Step 3.

The final resources for each node in *Net* is: $S_1(1.0100) > S_2(1.0097) > S_3(1.0002) > S_4(0.9801)$.

3 Comparisons and Tests

In this section, the proposed method is compared with another four approaches (degree, betweenness, k-shell, weighted PageRank) to demonstrate its effectiveness. The four methods will be defined in section 3.1. What's more, all methods are tested on noisy data to evaluate their stability. Two real networks are employed. One of them is a directed and weighted network. It is the network of the 500 busiest commercial airports in the United States. A tie exists between two airports if a flight was scheduled between them in 2002. The weights correspond to the number of seats available on the scheduled flights [54]. You can obtain the data through the hyperlink listed in <u>S1 Text</u>. The other one is an undirected and weighted network reorganized by Newman [55–57]. It is a collaboration network of scientists posting preprints on the high-energy theory archive at <u>www.arxiv.org</u>. These papers appeared in a 5-year window, from 1995 to 1999 inclusive. The data can also be downloaded through the <u>S1 Text</u>. More attributes about the two networks are listed in <u>Table 1</u>.



Table 1. Basic statistics of the two real networks.

Networks	Nodes	Edges	Diameter	Average Clustering Coefficient	Average Degeree
US airports network	1572	28235	9	0.469	17.961
collaborations network	8361	15751	19 7.025	0.636	3.768

doi:10.1371/journal.pone.0145028.t001

3.1 Definitions of the compared centrality measures

In the context of social science, the topology of a social network is represented by an adjacency matrix $A = \{w_{ij}\}_{N \times N}$, where the element $w_{ij} > 0$ if there exists a link from *j* to *i* and $w_{ij} = 0$ otherwise. For an undirected network, A is a symmetric matrix with $w_{ij} = w_{ji}$. If the network is weighted, the element w_{ij} represents the weight of the link from *j* to *i*. Actually, the adjacency matrix A fully describes the topological structure of the social network. Here, the adopted centrality measures to make comparisons are calculated by the following equations:

(1)Degree k_i for a node *i* can be computed as follows:

$$k_{i} = \sum_{j=1}^{N} w_{ij} + \sum_{j=1}^{N} w_{ji}$$
(13)

(2)Betweenness $C_B(i)$ is defined as

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$
(14)

where $\sigma_s t$ is the number of the shortest paths between nodes *s* and *t*, and $\sigma_{st}(i)$ is the number of the shortest paths between *s* and *t* which pass through node *i*.

(3)K-shell [58]: The k-shell index of a node is obtained by a procedure called k-shell decomposition, where we successively prune nodes in the network layer by layer. Concretely, the decomposition starts by removing nodes with degree k = 1. After that, some nodes may have only one link left. So we continue pruning the network iteratively until there are no nodes with k = 1. The removed nodes fall into a k-shell with index $k_S = 1$. With the similar method, we iteratively remove the next k shell $k_S = 2$ and higher k shells until all nodes are pruned. In the decomposition procedure, each node is assigned with a k-shell index. The periphery of the network corresponds to small k_S and the nodes with high k_S define the core of the network.

(4)Weighted PageRank [59] can be calculated from:

$$PR_i(t) = \alpha \sum_{j=1}^{N} \frac{w_{ij} PR_j(t)}{k_j^{out}} + \frac{1-\alpha}{N}$$
(15)

where $k_j^{out} = \sum_{i=1}^{N} w_{ji}$ and α is the jumping probability. $PR_i(t)$ is the probability that node *i* is visited by the random walker at time *t*. As time *t* increases, the probability $PR_i(t)$ will converge to a stationary probability PR_i . This value is defined as the PageRank which are used to determine its ranking relative to other nodes. In the calculation, the conventional choice of α is 0.85. In this paper, α is set as 0.85 for all experiments.

3.2 Effectiveness

A modified susceptible-infected-removed (SIR) model is employed to estimate the spreading influence of the top-ranked nodes in weighted networks. In this model, individuals can be in three discrete states: susceptible, infected or removed. Each individual in the model can be

represented by a node of the network and can only spread infection to its neighbors along the outgoing edges in the network. At each step, each infected node *i* randomly chooses one of its susceptible neighbors, *j*, and infects it with probability λ_{ij} , and then be removed (dead or recovered with immunity) with probability β . The probability λ_{ij} is determined by the following equation [60]:

$$\lambda_{ij} = \left(\frac{w_{ij}}{w_{max}}\right)^{\alpha}, \quad \alpha > 0$$

where α is a positive constant and w_{max} is the largest value of w_{ij} in the network. Since $\frac{w_{ij}}{w_{max}} < 1$, the smaller the α is, the more quickly the infection spreads. The process stops when no infected node is present. Here we use the cumulative number of infected nodes (which includes infected and recover nodes), denoted by *N*, as a function time. Without the loss of generality, α and β are assigned as 0.2 and 1.

We first compare the spreading processed activated by top-ranked [61] nodes from *Physar-um*Spreader and another four centrality measures (degree, betweenness, k-shell, weighted PageRank). Taking weighted PageRank as an example. If among two top-*L* lists by *Physarum*-Spreader and weighted PageRank, there are *n* different nodes, we compare the spreading processed activated by these *n* different nodes, respectively. For example, if L = 5, the top-ranked lists are {97, 1016, 754, 333, 335} and {754, 1222, 97, 841, 1016} for weighted PageRank and *Physarum*Spreader, then n = 3 and the spreading processes are compared with initially infected nodes {333, 335} for weighted PageRank and {1222, 841} for *Physarum*Spreader.

Tables $\underline{2}$ and $\underline{3}$ list the top 50 nodes obtained by different centrality methods in US airports network and collaborations network.

In US airports network, Fig 4a, 4b, 4c and 4d show the spreading results compared with degree centrality, betweenness, k-shell and weighted PageRank corresponding to n = 11, 17, 16 and 10, respectively, when *L* equals to 20. As can be seen, the proposed method slightly outperforms the other four centrality measures. In order to verify the efficacy of the proposed method, we check its efficiency among more nodes. Since it is impossible to check all nodes, we selected the 50 most important nodes to conduct the experiment. Fig 5a, 5b, 5c and 5d display the spreading results corresponding to n = 24, 35, 27 and 18 under L = 50. Here we can see, the proposed algorithm exhibits a good efficiency in L = 50 as well as in L = 20.

We also applied our algorithm in the other network with bigger size: collaborations network. Fig 6a, 6b, 6c and 6d show the spreading results compared with another four centrality measures corresponding to n = 11, 16, 20 and 10, respectively, under L = 20. The figures show that the cumulative number of the infected nodes obtained by our method is a bit bigger than the results calculated by the others except k-shell. However, when L equals to 50 in Fig 7, our method is a little bit worse than other centrality measures but has a quicker spread velocity than k-shell while n = 29, 43, 47 and 27.

With L increasing, more and more important nodes overlap, those overlapped nodes will be removed from the top-ranked lists. It means that the rest different nodes may have less spreading influence than those removed. Thus, the cumulative numbers of infected nodes have little changes in spite of n and L increasing.

3.3 Robustness

To scientifically test the performance of the *Physarum*Spreader algorithm, we measure the change in rankings when links are randomly removed with probability form 0.1 to 0.5. The rankings obtained from the modified network are compared to those from the original



Table 2. Top 50 nodes ranked by different centrality methods for US airports network.

Rank	Degree	Betweenness	K-shell	Weighted PageRank	PhysarumSpreader
1	97	76	27	97	754
2	1016	584	65	1016	1222
3	604	1331	97	754	97
4	654	1433	106	333	841
5	754	207	126	335	1016
6	422	591	148	654	774
7	333	604	176	1219	1219
8	907	428	181	76	76
9	606	892	184	606	333
10	335	1520	206	1222	654
11	866	126	207	866	1544
12	368	1542	212	752	790
13	1062	1129	261	1064	33
14	752	654	267	907	325
15	184	343	271	267	582
16	841	1557	306	422	599
17	445	1006	325	841	985
18	1222	427	328	368	207
19	267	686	333	1062	1457
20	1219	136	335	184	866
21	207	431	368	604	132
22	261	60	405	1250	241
23	212	754	422	445	427
24	176	1238	445	573	335
25	850	166	497	774	907
26	1557	206	582	212	606
27	1064	866	604	427	573
28	1072	722	606	1190	1250
29	1250	1051	624	1050	136
30	998	93	649	328	264
31	584	1287	654	132	1006
32	870	1032	752	846	267
33	1148	1566	754	1369	1064
34	838	1263	774	790	60
35	328	573	787	1287	422
36	774	107	838	779	368
37	624	394	841	850	1029
38	1287	365	846	1557	1426
39	271	1125	850	1457	78
40	106	466	863	1244	492
41	910	1294	866	985	184
42	846	422	870	968	661
43	1212	907	907	584	39
44	614	498	910	838	604
45	1192	184	985	870	968
46	148	670	998	261	445
47	76	752	1005	1258	1062

(Continued)



Table 2. (Continued)

Rank	Degree	Betweenness	K-shell	Weighted PageRank	PhysarumSpreader
48	306	710	1008	1239	752
49	1369	503	1016	1263	212
50	126	36	1017	1148	342

doi:10.1371/journal.pone.0145028.t002

Table 3. Top 50 nodes ranked by different centrality methods for collaborations network.

Rank	Degree	Betweenness	K-shell	Weighted PageRank	PhysarumSpreader
1	87	1571	6790	24	24
2	480	1832	6791	87	1794
3	168	24	6792	763	547
4	24	975	6793	997	546
5	997	3473	6794	546	473
6	481	500	6795	530	415
7	656	1603	6796	547	763
8	926	1368	6797	480	474
9	1571	926	6798	168	106
10	546	87	6799	469	926
11	975	1617	6800	433	123
12	473	276	6801	1424	2855
13	530	1337	6802	1201	87
14	547	481	6803	1794	480
15	884	193	6804	675	1340
16	930	930	6805	816	469
17	1603	2544	6806	975	39
18	39	480	6807	926	124
19	123	6487	6808	2117	1452
20	956	387	6809	106	479
21	106	168	6810	415	3940
22	675	752	6811	164	1087
23	1171	165	6812	2254	197
24	1201	307	6813	2855	38
25	1516	4201	310	177	420
26	2117	1190	5687	1941	3402
27	38	5248	5688	1698	1548
28	310	502	5689	884	431
29	763	1738	5690	1754	1754
30	1393	2286	5691	896	195
31	2070	5908	5692	1528	1641
32	1286	4490	5693	1479	37
33	185	1275	5694	473	1511
34	193	1515	5695	474	1870
35	431	6505	5696	1444	1512
36	433	3140	5697	336	178
37	469	997	5698	38	517
38	4201	185	5699	41	1571

(Continued)



Table 3. (Continued)

Rank	Degree	Betweenness	K-shell	Weighted PageRank	PhysarumSpreader
39	221	890	5700	1736	1642
40	449	166	5701	1452	518
41	1126	1393	5702	39	530
42	1255	1479	5703	23	322
43	1479	3770	5704	1854	2116
44	1570	404	53	85	2217
45	1698	501	530	956	3717
46	1754	948	546	616	321
47	35	1058	547	631	1640
48	41	956	1831	592	433
49	371	103	4836	1241	956
50	474	1247	5248	420	168

doi:10.1371/journal.pone.0145028.t003

network, by measuring the impact I_R on ranking, as given by [9]





Fig 4. Comparison of epidemic spreading efficiency among different measures when L = 20, in US airports network. Each result *N* is obtained by averaging over 100 implementations with $\alpha = 0.2$ and $\beta = 1$. (a):*Physarum*Spreader vs Degree, n = 11. (b):*Physarum*Spreader vs Betweenness, n = 17. (c): *Physarum*Spreader vs K-shell, n = 16. (d):*Physarum*Spreader vs Weighted PageRank, n = 10.



Fig 5. Comparison of epidemic spreading efficiency among different measures when *L* = 50, in US airports network. Each result *N* is obtained by averaging over 100 implementations with α = 0.2 and β = 1. (a):*Physarum*Spreader vs Degree, *n* = 24. (b):*Physarum*Spreader vs Betweenness, *n* = 35. (c): *Physarum*Spreader vs K-shell, *n* = 27. (d):*Physarum*Spreader vs Weighted PageRank, *n* = 18.

3 and R'_i correspond to the rankings obtained respectively from the original and modified graph. We measure I_R for *Physarum*Spreader, weighted PageRank, degree, betweenness and k-shell subject to the same modifications.

As shown in Figs <u>8</u> and <u>9</u>, I_R increase with the number of links removed. In Fig <u>8</u>, we can observe that our method is obviously more tolerant than weighted PageRank, betweenness and k-shell but worse than degree.

Here we can see, robustness of the presented approach is also better than weighted PageRank, in Fig 9. However, when p > 0.2, the tolerance of our method underperform degree and is a little worse than k-shell while p > 0.3. Notably, it seems that betweenness becomes the most tolerant method, but it's not true. Due to the topology of the original network, there are 4568 nodes whose value of betweenness equal to 0. It means that the rankings of these nodes remain unchanged when the network is modified. Hence, the tolerance of betweenness is evaluated invalidly, when it is tested in collaborations network by this measure.

In practical application, a valuable ranking algorithm should not be effective only, but also be robust. To further demonstrate the proposed method's advantages in terms of robustness, we demonstrate *Physarum*Spreader algorithm in a representative example called *sybil attack* [62]. Consider a situation that spammers deliberately gain disproportionately high rank by creating huge fake entities. To simulate this attack, each time node *i* creates v(v = 10, 50, 100) fake entities which only direct to node *i* with weight equaling to 1. The rank of node *i* is denoted by



Fig 6. Comparison of epidemic spreading efficiency among different measures when L = 20, in collaborations network. Each result *N* is obtained by averaging over 100 implementations with $\alpha = 0.2$ and $\beta = 1$. (a):*Physarum*Spreader vs Degree, n = 11. (b):*Physarum*Spreader vs Betweenness, n = 16. (c): *Physarum*Spreader vs K-shell, n = 20. (d):*Physarum*Spreader vs Weighted PageRank, n = 10.

PLOS ONE

 r_i first and then it is represented by r'_i after manipulations. Obviously, r'_i is less than or equal to r_i and if there are smaller differences between them, the approach will be more robust. Only the top-100 users (i = 1, 2, 3, ..., 100) are studied under this attack.

As shown in Figs $\underline{10}$ and $\underline{11}$, the vertical axis displays the manipulated rank of a user after creation of v fake fans and the horizontal axis shows its original rank. As we can see, in both networks of US airports network and collaborations network, *Physarum*Spreader is the most robust algorithm among others against manipulations as its smaller change of rank.

4 Conclusions

In this paper, focus is placed on identifying influential spreaders in weighted networks and an extended algorithm called *Physarum*Spreader has been proposed based on LeaderRank and a positive feedback mechanism inspired from *Physarum Polycephalum*. In order to investigate the performance of the proposed method, two weighted real networks that one is directed and the other is undirected, have been used as test network data sets. Furthermore, comparison with four well-known centrality measures are also studied with the help of an epidemic spreading model. Experimental results indicate that *Physarum*Spreader is effective in identifying influential spreaders. In addition, the proposed method has a good robustness compared with other measures.



Fig 7. Comparison of epidemic spreading efficiency among different measures when L = 50, in collaborations network network. Each result N is obtained by averaging over 100 implementations with $\alpha = 0.2$ and $\beta = 1$. (a):*Physarum*Spreader vs Degree, n = 29. (b):*Physarum*Spreader vs Betweenness, n = 43. (c):*Physarum*Spreader vs K-shell, n = 47. (d):*Physarum*Spreader vs Weighted PageRank, n = 27.







Fig 9. The impact ranking as a function of number of links randomly removed for collaborations network. All data points are average values over 100 independent runs with error bars showing the standard deviations.

PLOS ONE



Fig 10. The impact ranking as a function of number of links randomly removed for US airports network. All data points are average values over 100 independent runs with error bars showing the standard deviations.



Fig 11. The impact ranking as a function of number of links randomly removed for collaborations network. All data points are average values over 100 independent runs with error bars showing the standard deviations.

PLOS ONE

Supporting Information

S1 Text. Dataset web pages. (TXT)

Acknowledgments

The authors greatly appreciate the reviews' suggestions and the editor's encouragement. The work is partially supported by National High Technology Research and Development Program of China (863 Program) (Grant No. 2013AA013801), National Natural Science Foundation of China (Grant Nos. 61174022, 61573290, 61503237), China State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No. BUAA-VR-14KF-02).

Author Contributions

Conceived and designed the experiments: HW YZ YD. Performed the experiments: YZ ZZ. Analyzed the data: HW YD SM. Wrote the paper: YZ HW.

References

1. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. Nature Physics. 2010; 6:888–893. doi: <u>10.1038/nphys1746</u>

- Wang P, Yu X, Lu J. Identification and Evolution of Structurally Dominant Nodes in Protein-Protein Interaction Networks. Biomedical Circuits and Systems, IEEE Transactions on. 2014; 8(1):87–97. doi: 10.1109/TBCAS.2014.2303160
- Wang P, Lü J, Yu X. Identification of important nodes in directed biological networks: A network motif approach. PloS one. 2014; 9(8):e106132. doi: <u>10.1371/journal.pone.0106132</u> PMID: <u>25170616</u>
- Zhou T, Medo M, Cimini G, Zhang ZK, Zhang YC. Emergence of scale-free leadership structure in social recommender systems. PLoS One. 2011; 6(7):e20648. doi: <u>10.1371/journal.pone.0020648</u> PMID: <u>21857891</u>
- Deng Y, Liu Y, Zhou D. An Improved Genetic Algorithm with Initial Population Strategy for Symmetric TSP. Mathematical Problems in Engineering. 2015; 2015:212794. doi: <u>10.1155/2015/212794</u>
- Chen D, Lü L, Shang MS, Zhang YC, Zhou T. Identifying influential nodes in complex networks. Physica A: Statistical Mechanics and its Applications. 2012; 391:1777–1787. doi: <u>10.1016/j.physa.2011.09.017</u>
- Kolaczyk ED, Chua DB, Barthélemy M. Group betweenness and co-betweenness: Inter-related notions of coalition centrality. Social Networks. 2009; 31(3):190–203. doi: <u>10.1016/j.socnet.2009.02.003</u>
- Wang F, Antipova A, Porta S. Street centrality and land use intensity in Baton Rouge, Louisiana. Journal of Transport Geography. 2011; 19:285–293. doi: <u>10.1016/j.jtrangeo.2010.01.004</u>
- 9. Lü L, Zhang YC, Yeung CH, Zhou T. Leaders in Social Networks, the Delicious Case. PLos One. 2011; 6:e21202. doi: <u>10.1371/journal.pone.0021202</u> PMID: <u>21738620</u>
- Zhang H, Fiszman M, Shin D, Miller CM, Rosemblat G, Rindflesch TC. Degree centrality for semantic abstraction summarization of therapeutic studies. Journal of Biomedical Informatics. 2011; 44:830– 838. doi: 10.1016/j.jbi.2011.05.001 PMID: 21575741
- 11. Zio E, Sansavini G. Component Criticality in Failure Cascade Processes of Network Systems. Risk Analysis. 2011; 31(8):1196–1210. doi: 10.1111/j.1539-6924.2011.01584.x PMID: 21371060
- Pathak PH, Dutta R. Centrality-based power control for hot-spot mitigation in multi-hop wireless networks. Computer Communications. 2012; 35(9):1074–1085. doi: <u>10.1016/j.comcom.2012.01.023</u>
- Pei S, Makse HA. Spreading dynamics in complex networks. Journal of Statistical Mechanics: Theory and Experiment. 2013; 2013(12):P12002. doi: <u>10.1088/1742-5468/2013/12/P12002</u>
- Sporns O, Honey CJ, Kötter R. Identification and classification of hubs in brain networks. PloS one. 2007; 2(10):e1049. doi: <u>10.1371/journal.pone.0001049</u> PMID: <u>17940613</u>
- 15. Pei S, Muchnik L, Andrade JS Jr, Zheng Z, Makse HA. Searching for superspreaders of information in real-world social media. Scientific reports. 2014; 4:5547. doi: 10.1038/srep05547 PMID: 24989148
- Deng Y. A Threat Assessment Model under Uncertain Environment. Mathematical Problems in Engineering. 2015; 2015:878024, <u>http://dx.doi.org/10.1155/2015/878024</u>. doi: 10.1155/2015/878024
- Freeman LC. Centrality in social networks: Conceptual clarification. Social Networks. 1978; 1(3):215– 239. doi: 10.1016/0378-8733(78)90021-7
- Wang G, Shen Y, Luan E. A measure of centrality based on modularity matrix. Progress in Natural Science. 2008; 18:1043–1047. doi: <u>10.1016/j.pnsc.2008.03.015</u>
- Estrada E, Hatano N. A vibrational approach to node centrality and vulnerability in complex networks. Physica A: Statistical Mechanics and its Applications. 2010; 389(17):3648–3660. doi: <u>10.1016/j.physa.</u> <u>2010.03.030</u>
- Landherr A, Friedl B, Heidemann J. A critical review of centrality measures in social networks. Business & Information Systems Engineering. 2010; 2(6):371–385. doi: <u>10.1007/s12599-010-0127-3</u>
- Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks. 2001; 23(3):191–201. doi: <u>10.1016/S0378-8733(01)00038-7</u>
- 22. Katz L. A new status index derived from sociometric analysis. Psychometrika. 1953; 18(1):39–43. doi: 10.1007/BF02289026
- Newman MEJ. A measure of between centrality based on random walks. Social Networks. 2005; 27 (1):39–54. doi: 10.1016/j.socnet.2004.11.009
- Du Y, Gao C, Hu Y, Mahadevan S, Deng Y. A new method of identifying influential nodes in complex networks based on TOPSIS. Physica A Statistical Mechanics & Its Applications. 2014; 399(4):57–69. doi: <u>10.1016/j.physa.2013.12.031</u>
- Su X, Mahadevan S, Xu P, Deng Y. Dependence assessment in Human Reliability Analysis using evidence theory and AHP. Risk Analysis. 2015; 35:1296–1316. doi: <u>10.1111/risa.12347</u> PMID: <u>25847228</u>
- Jiang W, Yang Y, Luo Y, Qin X. Determining Basic Probability Assignment Based on the Improved Similarity Measures of Generalized Fuzzy Numbers. International Journal of Computers Communications & Control. 2015; 10(3):333–347. doi: 10.15837/ijccc.2015.3.1656

- Jiang W, Luo Y, Qin X, Zhan J. An improved method to rank generalized fuzzy numbers with different left heights and right heights. Journal of Intelligent & Fuzzy Systems. 2015; 28:2343–2355. doi: <u>10.</u> <u>3233/IFS-151639</u>
- Deng Y. Generalized evidence theory. Applied Intelligence. 2015; 43(3):530–543. doi: <u>10.1007/</u> <u>s10489-015-0661-2</u>
- Deng X, Hu Y, Deng Y, Mahadevan S. Supplier selection using AHP methodology extended by D numbers. Expert Systems with Applications. 2014; 41(1):156–167. doi: <u>10.1016/j.eswa.2013.07.018</u>
- Wei D, Deng X, Zhang X, Deng Y, Mahadevan S. Identifying influential nodes in weighted networks based on evidence theory. Physica A Statistical Mechanics & Its Applications. 2013; 392(10):2564– 2575. doi: <u>10.1016/j.physa.2013.01.054</u>
- Gao C, Wei D, Hu Y, Mahadevan S, Deng Y. A modified evidential methodology of identifying influential nodes in weighted networks. Physica A Statistical Mechanics & Its Applications. 2013; 392(21):5490– 5500. doi: <u>10.1016/j.physa.2013.06.059</u>
- 32. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E—Statistical, Nonlinear and Soft Matter Physics. 2001; 64(1):016132. doi: <u>10.1103/</u><u>PhysRevE.64.016132</u>
- Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010; 32:245–251. doi: <u>10.1016/j.socnet.2010.03.006</u>
- Qi X, Fuller E, Wu Q, Wu Y, Zhang CQ. Laplacian centrality: A new centrality measure for weighted networks. Information Sciences. 2012; 194:240–253. doi: <u>10.1016/j.ins.2011.12.027</u>
- Zhang Y, Zhang Z, Wei D, Deng Y. Centrality measure in weighted networks based on an amoeboid algorithm. Journal of Information & Computational Science. 2012; 9(2):369–376.
- Li Q, Zhou T, Lü L, Chen D. Identifying influential spreaders by weighted leaderrank. Physica A: Statistical Mechanics and its Applications. 2014; 404:47–55. doi: <u>10.1016/j.physa.2014.02.041</u>
- Freeman LC, Borgatti SP, White DR. Centrality in valued graphs: A measure of betweenness based on network flow. Social Networks. 1991; 13:141–154. doi: <u>10.1016/0378-8733(91)90017-N</u>
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proceedings of the National Academy of Science. 2004; 101:3747–3752. doi: <u>10.1073/pnas.</u> <u>0400087101</u>
- Brandes U. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology. 2001; 25 (1):163–177. doi: 10.1080/0022250X.2001.9990249
- Dijkstra EW. A note on two problems in connection with graphs. Numerische Mathematik. 1959; 1 (1):269–271. doi: <u>10.1007/BF01386390</u>
- Baumgarten W, Ueda T, Hauser MJ. Plasmodial vein networks of the slime mold Physarum polycephalum form regular graphs. Physical Review E. 2010; 82(4):046113. doi: <u>10.1103/PhysRevE.82.046113</u>
- Watanabe S, Tero A, Takamatsu A, Nakagaki T. Traffic optimization in railroad networks using an algorithm mimicking an amoeba-like organism, Physarum plasmodium. BioSystems. 2011; 105(3):225–232. doi: <u>10.1016/j.biosystems.2011.05.001</u> PMID: <u>21620930</u>
- Nakagaki T, Yamada H, Tth a. Maze-solving by an amoeboid organism. Nature. 2000; 407(6803):470– 470. doi: <u>10.1038/35035159</u> PMID: <u>11028990</u>
- Zhang Y, Zhang Z, Deng Y. An Improved Maze Solving Algorithm Based on An Amoeboid Organism. In: Proceedings of the 23nd Chinese Control and Decision Conference. Mianyang, China; 2011. p. 1440–1443.
- 45. Wang H, Lu X, Zhang X, Wang Q, Deng Y. A Bio-Inspired Method for the Constrained Shortest Path Problem. The Scientific World Journal. 2014; 2014(Article ID 271280):11 pages. doi: <u>10.1155/2014/</u> <u>271280</u>
- Nakagaki T, Yamada H, Tth A. Path finding by tube morphogenesis in an amoeboid organism. Biophysical chemistry. 2001; 92(1–2):47–52. doi: <u>10.1016/S0301-4622(01)00179-X</u> PMID: <u>11527578</u>
- Nakagaki T, lima M, Ueda T, Nishiura Y, Saigusa T, Tero A, et al. Minimum-risk path finding by an adaptive amoebal network. Physical review letters. 2007; 99(6):068104. doi: <u>10.1103/PhysRevLett.99.068104</u> PMID: <u>17930872</u>
- Miyaji T, Ohnishi I. Physarum can solve the shortest path problem on Riemannian surface mathematically rigorously. International Journal of Pure and Applied Mathematics. 2008; 47(3):353–369.
- 49. Tero A, Kobayashi R, Nakagaki T. Physarum solver: A biologically inspired method of road-network navigation. Physica A: Statistical Mechanics and its Applications. 2006; 363(1):115–119. doi: <u>10.1016/j.physa.2006.01.053</u>
- Baumgarten W, Hauser MJ. Functional organization of the vascular network of Physarum polycephalum. Physical biology. 2013; 10(2):026003. doi: <u>10.1088/1478-3975/10/2/026003</u> PMID: <u>23406784</u>

- Tero A, Takagi S, Saigusa T, Ito K, Bebber DP, Fricker MD, et al. Rules for Biologically Inspired Adaptive Network Design. Science. 2010; 327(5964):439–442. doi: <u>10.1126/science.1177894</u> PMID: <u>20093467</u>
- Tero A, Kobayashi R, Nakagaki T. A mathematical model for adaptive transport network in path finding by true slime mold. Journal of theoretical biology. 2007; 244(4):553–564. doi: <u>10.1016/j.jtbi.2006.07</u>. 015 PMID: 17069858
- Tero A, Kobayashi R, Nakagaki T. A mathematical model for adaptive transport network in path finding by true slime mold. Journal of Theoretical Biology. 2007; 244(4):553–564. doi: <u>10.1016/j.jtbi.2006.07</u>. <u>015</u> PMID: <u>17069858</u>
- 54. Colizza V, Pastor-Satorras R, Vespignani A. Reaction—diffusion processes and metapopulation models in heterogeneous networks. Nature Physics. 2007; 3(4):276–282. doi: <u>10.1038/nphys560</u>
- Newman ME. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences. 2001; 98(2):404–409. doi: <u>10.1073/pnas.98.2.404</u>
- Newman ME. Scientific collaboration networks. I. Network construction and fundamental results. Physical review E. 2001; 64(1):016131. doi: <u>10.1103/PhysRevE.64.016131</u>
- 57. Newman ME. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical review E. 2001; 64(1):016132. doi: 10.1103/PhysRevE.64.016132
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. Nature Physics. 2010; 6(11):888–893. doi: 10.1038/nphys1746
- Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer networks and ISDN systems. 1998; 30(1):107–117. doi: <u>10.1016/S0169-7552(98)00110-X</u>
- Yan G, Zhou T, Wang J, Fu ZQ, Wang BH. Epidemic Spread in Weighted Scale-Free Networks. Chinese Physics Letter. 2005; 22:510–513. doi: 10.1088/0256-307X/22/2/068
- Yang R, Wang BH, Ren J, Bai WJ, Shi ZW, Wang WX, et al. Epidemic spreading on heterogeneous networks with identical infectivity. Physics Letters A. 2007; 364(3):189–193. doi: <u>10.1016/j.physleta.</u> <u>2006.12.021</u>
- Douceur J. The Sybil Attack. In: Peer-to-Peer Systems. vol. 2429. Springer Berlin Heidelberg; 2002. p. 251–260.