

RESEARCH ARTICLE

Maximizing the Spread of Influence via Generalized Degree Discount

Xiaojie Wang¹, Xue Zhang¹, Chengli Zhao¹, Dongyun Yi^{1,2*}

1 College of Science, National University of Defense Technology, Changsha, Hunan, China, **2** State key laboratory of High Performance Computing, National University of Defense Technology, Changsha, Hunan, China

* dongyun.yi@gmail.com



CrossMark
click for updates

Abstract

It is a crucial and fundamental issue to identify a small subset of influential spreaders that can control the spreading process in networks. In previous studies, a degree-based heuristic called *DegreeDiscount* has been shown to effectively identify multiple influential spreaders and has served as a benchmark method. However, the basic assumption of *DegreeDiscount* is not adequate, because it treats all the nodes equally without any differences. To consider a general situation in real world networks, a novel heuristic method named *GeneralizedDegreeDiscount* is proposed in this paper as an effective extension of original method. In our method, the status of a node is defined as a probability of not being influenced by any of its neighbors, and an index generalized discounted degree of one node is presented to measure the expected number of nodes it can influence. Then the spreaders are selected sequentially upon its generalized discounted degree in current network. Empirical experiments are conducted on four real networks, and the results show that the spreaders identified by our approach are more influential than several benchmark methods. Finally, we analyze the relationship between our method and three common degree-based methods.

OPEN ACCESS

Citation: Wang X, Zhang X, Zhao C, Yi D (2016) Maximizing the Spread of Influence via Generalized Degree Discount. PLoS ONE 11(10): e0164393. doi:10.1371/journal.pone.0164393

Editor: Peter Csermely, Semmelweis University, HUNGARY

Received: June 16, 2016

Accepted: September 23, 2016

Published: October 12, 2016

Copyright: © 2016 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), 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: This work is partially supported by HPCL open fund of NUDT (201403-01), Postgraduate Innovation Fund of Hunan province (CX2015B010), and Postgraduate Innovation Fund of NUDT (B150203).

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

Introduction

In complex networks, models and methods for propagation behavior are always of great theoretical and practical importance. Consider a scenario in advertising: a small IT company develops a cool online application and it wants to let more people know their product. However, the funds for advertising are limited. An economical way to advertise is to deliver the product to a small group of initial users (or spreaders) who are willing to advertise the product by word of mouth. This is often referred to as influence maximization. Theoretically, influence maximization in networks is a specific problem about how to effectively identify a small subset of nodes and maximize their spreading influence. Although much work has been done on measuring the influence of a single node [1–6], methods that can effectively identify multiple influential spreaders are still lacking.

The pioneers of this research area are Domingos and Richardson [7, 8] who studied influence maximization as an algorithmic problem and developed a probabilistic method. Kempe, Kleinberg and Tardos [9] also made a significant contribution to this field. They showed that the problem is an NP-hard discrete optimization problem, and proposed a greedy strategy to select the spreaders that could achieve an approximation guarantee of 63%. Unfortunately, their greedy method encountered a serious drawback in computing efficiency, which limited its wide usage in large-scale networks. Leskovec *et al.* [10] demonstrated that many realistic influence maximization problems exhibit a property of “submodularity”, and they proposed a *Cost-Effective Lazy Forward (CELF)* method to improve the efficiency of the greedy method. Narayanan *et al.* [11] analyzed the Shapley value concept from cooperative game theory and proposed *Shapley value-based Influential Nodes (SPIN)*. Zhao *et al.* [12] attempted to find effective multiple spreaders by generalizing the idea of the coloring problem in graph theory to complex networks. He *et al.* [13] suggested a novel method to identify multiple spreaders from communities in a balanced way. Zhang *et al.* [14] presented an iterative method named VoteRank to identify a set of decentralized spreaders. Chen *et al.* [15] decomposed the local topological structure of nodes and proposed a *DegreeDiscount* heuristic. Numerical experiments showed that *DegreeDiscount* could nearly match the performance of the greedy method, while the computational complexity of the former one was quite low. However, all nodes were treated equally in *DegreeDiscount*, which was a little oversimplified and might reduce the performance of the algorithm.

In this paper, we depict the status of nodes more concisely as a probabilistic score, and propose the *GeneralizedDegreeDiscount* heuristic. We discuss the computational efficiency of our method and demonstrate that the complexity is linearly correlated with the network scale, which makes our method efficient and scalable to large-scale networks. Experiments are performed on several real networks, and the results show that our method can outperform some centrality-based methods.

Materials and Methods

Intuition and theory

Degree is a basic centrality index in the research area of complex networks. It is well known that a node with a higher degree can influence more nodes than a node with a lower degree. Some researches in sociology have shown that selecting nodes with the highest degree as spreaders can result in better spreading influence than many other methods. However, in some recent studies, the authors argued that nodes with the highest degree might not always be the most influential ones. Though the effectiveness of the *Degree* is questionable, the low computational complexity of this strategy results in its widespread use in many business fields. In this section, we try to enhance the performance of this method by using several heuristic strategies.

Let node v be a neighbor of node u . Suppose u has been selected as a spreader. When considering the selection of v as a new spreader, one should formulate a method for calculating the contribution of edge \overline{uv} to the degree of node v . It cannot be counted as 1, as has been done previously. As u has been selected as a spreader: (i) it is no longer necessary for v to influence u . (ii) u may also influence v with some probability, which further weakens the potential influence of v . Based on these considerations, we explore several heuristic strategies, in which the spreaders are selected one by one.

Degree-based heuristics. *DegreeDistance* [16] takes a naive approach to avoid the relative influence between adjacent nodes. For example, if a node has been selected as a spreader, we can ignore its neighbors and consider other nodes. *DegreeDistance* defines a candidate set C and a distance threshold d_{td} . At first, all the nodes are in the candidate set C . In each round, a

node v with maximum degree in C is selected as a spreader, and the nodes within a distance d_{td} to v are removed from C . The procedure ends when all the spreaders have been selected.

In Ref [15], Chen *et al.* proposed two degree-based heuristics, *SingleDiscount* and *DegreeDiscount*. *SingleDiscount* considers a simple adaptive strategy. In each round, a node u with the maximum degree is selected as a spreader. Then, for each $v \in \Gamma(u)$, we do not count \overline{uv} when calculating its degree. In other words, the degree of v will be discounted by 1. This type of degree is named as discount degree, and is denoted by sd_v ,

$$sd_v = d_v - t_v \tag{1}$$

where d_v denotes the original degree of v , and t_v denotes the number of v 's neighbors who have already been selected as spreaders.

DegreeDiscount is specifically designed for the independent cascade model. For a specific spreading probability p , *DegreeDiscount* attempts to conduct a deeper analysis of the local structure of the nodes. Suppose that we want to calculate the potential spreading ability of node v . Let the spreading probability be p . When p is small, the multi-hop neighbors of v can be ignored, and only the nearest neighbors are counted toward the degree. Let $u \in \Gamma(v)$ be a spreader neighbor of v . Obviously, the probability that v is directly influenced by u is p . As a result, u will not only contribute nothing to v , but also weakens the spreading ability of v .

When calculating the potential spreading ability of v , *DegreeDiscount* ignores the differences of v 's neighbors. As all the neighbors are treated equally, the diagram of v and its neighbors $\Gamma(v)$ can be mapped into a star-like subgraph structure. Let $Star(v)$ be the subgraph considered here, and let the edges in the subgraph be the edges incident to v . Let d_v be the degree of node v , t_v be the number of spreader neighbors of v , and p be the spreading probability.

As the candidate node v has t_v spreader neighbors, the probability that v is influenced by these neighbors is $1 - (1 - p)^{t_v}$. In this situation, selecting v as a new spreader may not bring any additional influence. In the opposite situation, selecting v will contribute to the spreading process by v itself and its normal neighbors. The former term can influence 1 node(v itself), and the latter can influence $d_v - t_v$ nodes(normal neighbors) with probability p . Together, the expected number of nodes influenced by v is

$$(1 - p)^{t_v}(1 + (d_v - t_v)p) \tag{2}$$

Under the first order of Taylor expansion, when p is small, the left term can be approximated by $1 - t_v p + o(t_v p)$. After further simplification, the whole equation becomes

$$1 + (d_v - 2t_v - (d_v - t_v)t_v p)p + o(t_v p) \tag{3}$$

Then, the discounted degree of v can be defined as

$$dd_v = d_v - 2t_v - (d_v - t_v)t_v p \tag{4}$$

Note that in the original equation Eq (2), dd_v is always non-negative. However, in the simplified form Eq (4), dd_v may be negative with some special parameters. In this situation, we manually set dd_v to be 0. Fig 1 depicts the local topology considered by *DegreeDiscount*. In this toy model, $d_v = 4$, $t_v = 1$, and

$$dd_v = 4 - 2 \times 1 - (4 - 1) \times 1p = 2 - 3p \tag{5}$$

Generalized Degree Discount. As *DegreeDiscount* models all the neighbors in $\Gamma(v)$ equally, it ignores the differences among them. Take an extreme case as example, let $s, t \in \Gamma(v)$ be two normal neighbors of node v . If all the neighbors of s itself are spreaders and all the

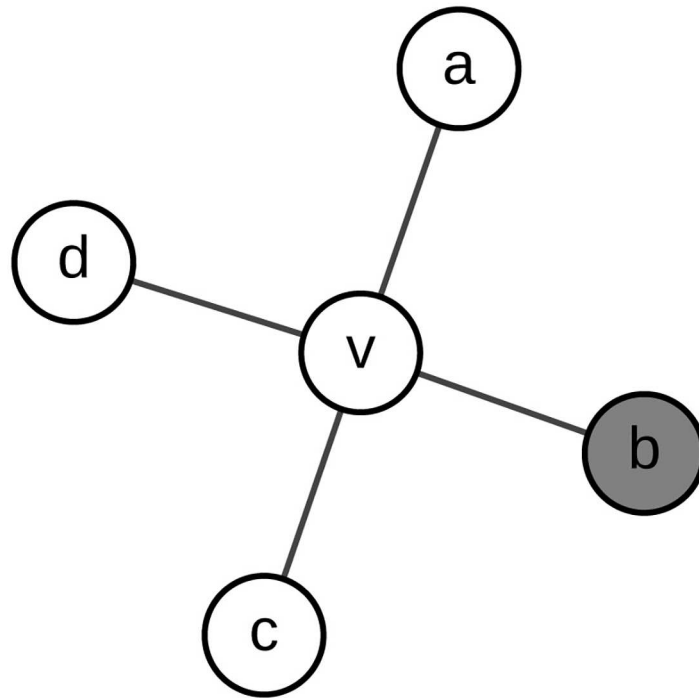


Fig 1. The diagram of *DegreeDiscount*. Nodes filled by gray denote the selected spreaders, and others denote normal nodes.

doi:10.1371/journal.pone.0164393.g001

neighbors of t itself are normal nodes, they should not be treated equally. Obviously, the probability that s is influenced by its own neighbors is far larger than t . When calculating the potential contribution of s , t towards v , the latter one should be given more weight. To make the original model more precisely, we propose the *GeneralizedDegreeDiscount*.

Similar to the analysis in *DegreeDiscount*, the probability that node v is not influenced by its spreader neighbors is $(1 - p)^{t_v}$. If v is not influenced by any of those neighbors, selecting v will enhance the total influence by v itself and its $d_v - t_v$ normal neighbors. For any normal neighbor $w \in \Gamma(v)$, the probability that w is not influenced by its own neighbors is also $(1 - p)^{t_w}$. In other words, w will bring additional influence $(1 - p)^{t_w}$ to v with probability p . Together, the expected number of nodes that will be influenced by v is

$$(1 - p)^{t_v} \left(1 + \sum_{d_v - t_v} (1 - p)^{t_w} p \right) \tag{6}$$

where the summation is over all $d_v - t_v$ normal neighbors of node v .

Departing from the conduction in *DegreeDiscount*, here we consider the second-order Taylor expansion for the left term and the first-order expansion for the right term:

$$\left(1 - t_v p + \frac{1}{2} t_v (t_v - 1) p^2 + o((t_v p)^2) \right) \left(1 + \sum_{d_v - t_v} (1 - t_w p + o(t_w p) p) \right) \tag{7}$$

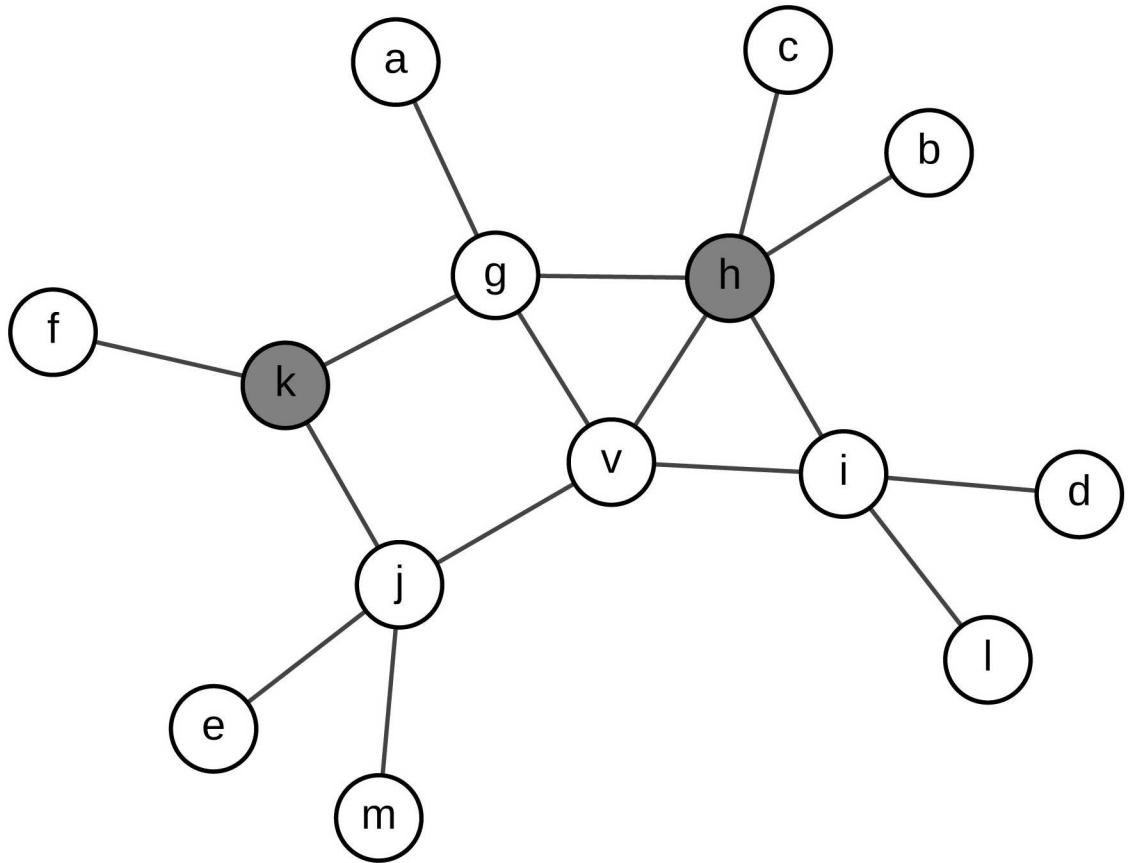


Fig 2. The diagram of *GeneralizedDegreeDiscount*. Nodes filled by gray denote the selected spreaders, and others denote normal nodes.

doi:10.1371/journal.pone.0164393.g002

After further simplification, the equation becomes

$$1 + \left(d_v - 2t_v - (d_v - t_v)t_v p + \frac{1}{2}t_v(t_v - 1)p - \sum_{d_v-t_v} t_w p \right) p + o(t_v^2 p^2) + \sum_{d_v-t_v} o(t_w p^2) \quad (8)$$

Then, the generalized discounted degree of v can be defined as

$$gdd_v = d_v - 2t_v - (d_v - t_v)t_v p + \frac{1}{2}t_v(t_v - 1)p - \sum_{d_v-t_v} t_w p \quad (9)$$

Similar to the situation in *DegreeDiscount*, the simplified equation of gdd_v may also be less than zero. In our real implementations, we set $gdd_v = 0$ in this situation. Fig 2 depicts the local topology considered by *GeneralizedDegreeDiscount*. In this toy model, $d_v = 4$ and $t_v = 1$. Note that the summation is over all normal neighbors, i.e., $\{g, i, j\}$ in this figure. As $t_g = 2$, $t_i = 1$ and $t_j = 1$, the generalized discounted degree of v is

$$gdd_v = 4 - 2 \times 1 - (4 - 1) \times 1p + \frac{1}{2} \times 1 \times (1 - 1)p - (2 + 1 + 1)p = 2 - 7p \quad (10)$$

Compared to the formulation of *DegreeDiscount*(Eq (4)), the formulation of *GeneralizedDegreeDiscount*(Eq (9)) adds the last two terms. As the consideration of the latter one is deeper than the former one, *GeneralizedDegreeDiscount* should be more effective than *DegreeDiscount*. However, the difference between them is not so significant in real situations. In reality, a small fraction of spreaders must be selected and their influences broadcast with low probability. Because the number of spreaders is not large, usually $t_v \ll d_v$ for all nodes. Thus, in Eq (9), the fourth and fifth terms are smaller than the third term, which makes *GeneralizedDegreeDiscount* just similar to *DegreeDiscount*. In the Results section, we will compare the two methods numerically.

Computational efficiency

The *GeneralizedDegreeDiscount* is implemented in Algorithm 1. If we want to select l spreaders, the algorithm must run for l rounds. Let N be the number of nodes and $\langle k \rangle$ be the average degree. In each round, the selection scheme costs $O(N)$, the neighbor finding scheme costs $O(\langle k \rangle^2)$, and for each of those neighbors, the updating process costs $O(\langle k \rangle)$. Then, the total time cost of the algorithm is $O(l(N + \langle k \rangle^2 + \langle k \rangle^2 \cdot \langle k \rangle)) \approx O(l(N + \langle k \rangle^3))$. In many networks, the average degree is far less than the number of nodes: $\langle k \rangle \ll N$. Thus the time cost of *GeneralizedDegreeDiscount* will be nearly $O(lN)$, which is just linearly correlated with the scale of the network.

Algorithm 1 GeneralizedDegreeDiscount(G, l, p)

```

Initialize  $S = \emptyset$ 
for all  $v \in V$  do
     $gdd_v = d_v$ 
     $t_v = 0$ 
end for
// iteratively select  $l$  spreaders
for  $i = 1$  to  $l$  do
    // select the node with the maximum generalized discount degree  $gdd_v$ 
     $u = \arg \max_v \{gdd_v \mid v \in V \setminus S\}$ 
     $S = S \cup \{u\}$ 
     $NB = \emptyset$ 
    // find the nearest and next nearest neighbors of  $u$  and update  $t_v$  for  $v \in \Gamma(u)$ 
    for all  $v \in \Gamma(u)$  do
         $NB = NB \cup \{v\}$ 
         $t_v = t_v + 1$ 
        for all  $w \in \Gamma(v)$  do
             $NB = NB \cup \{w\}$ 
        end for
    end for
    // update  $gdd_v$  for all  $v \in NB$ 
    for all  $v \in NB$  do
         $sum_{t_w} = 0$ 
        for all  $w \in \Gamma(v)$  do
            if  $w \notin S$  then
                 $sum_{t_w} = sum_{t_w} + t_w$ 
            end if
        end for
         $gdd_v = d_v - 2t_v - (d_v - t_v)t_v p + \frac{1}{2}t_v(t_v - 1)p - sum_{t_w} p$ 
        if  $gdd_v < 0$  then
             $gdd_v = 0$ 
        end if
    end for
end for
return  $S$ 

```

Results and Discussion

To evaluate the performance, we simulate the experiments using the Susceptible-Infected-Recovered(SIR) model. The SIR model was originally proposed as a model of the dynamics of the spread of disease. Due to the similarities between epidemic transmission and the spread of information, we use SIR to measure the spreading influence of individual nodes. In the SIR model, a node may assume one of three states(susceptible, infected and recovered). Specifically, susceptible individuals S in the model is analogous to individuals who are not aware of the information. Infected individuals I can be analogous to information carriers who are willing to spread information to their neighbors. Recovered individuals R are those who had previously received the information but later lost interest. To better simulate the real-world spreading process, we use the SIR model with limited contact [17]. At each time step, each infected node will randomly select a neighbor to contact, and will transmit the disease to its neighbor with probability p if the neighbor is susceptible. After the transmission process, the infected node will become a recovered node with probability q . The effective spreading rate λ is defined as p/q . When there are no infected nodes, the process stops, and we use the fraction of recovered nodes to measure the spreading influence.

Data Description

To evaluate the influences of different groups of spreaders selected by various methods, we conduct the experiments on the following four networks from different fields.

1. Enron [18]: An email communication network which covers all the email communication within Enron Corporation. Nodes in the network are email addresses and edges represent the email communications among them.
2. Cond-mat [19]: A collaboration network of scientists posting preprints to the condensed matter archive at arxiv.org between January 1, 1995 and March 31, 2005. Nodes in the network represent the scientists and edges represent the collaborations among them.
3. Gnutella [20]: A snapshot of the Gnutella peer-to-peer file sharing network at August 31 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.
4. Epinions [21]: A who-trust-whom online social network of the general consumer review site Epinions.com. All the trust relationships interact and form the Web of Trust, which is then combined with review ratings to determine which reviews are shown to the users.

For simplicity, we treat all the networks as undirected and unweighted, and discard self-loops and multiple links. Only the largest connected component of each network is considered. A Brief overview of the networks is shown in Table 1.

Table 1. The basic topological features of four real networks. N and M are the numbers of nodes and edges. $\langle k \rangle$ is the average degree. d_{max} denotes the network diameter and $\langle d \rangle$ denotes the average shortest path length. r and cc are the assortative coefficient and clustering coefficient, respectively.

network	N	M	$\langle k \rangle$	d_{max}	$\langle d \rangle$	r	cc
Enron	33696	180811	10.732	13	4.025	-0.116	0.085
Cond-mat	36458	171736	9.421	18	5.499	0.177	0.243
Gnutella	62561	147878	4.727	11	5.936	-0.093	0.004
Epinions	75877	405739	10.695	15	4.308	-0.041	0.066

doi:10.1371/journal.pone.0164393.t001

Benchmark methods

In complex networks, many centrality indexes have been defined to measure the importance of nodes and links. It is believed that nodes with higher centrality are more influential than common nodes. Accordingly, one naive solution for the influence maximization problem is to select the *top-l* nodes with the highest centrality indexes. In this paper, we use centrality-based methods as the benchmark methods.

Degree is a basic local centrality index for nodes. The higher degree a node has, the more important it is. In a social network, a person with more followers or friends is likely to have a larger influence.

Betweenness [22, 23] measures the extent to which a node is located on the shortest paths between pairs of nodes in networks.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{11}$$

where σ_{st} denotes the number of shortest paths between a pair (s, t) of nodes, and $\sigma_{st}(v)$ denotes the number of shortest paths between any pair of nodes that pass through v .

Closeness [24] is an evaluation of the geometric location of nodes.

$$C_C(v) = \frac{1}{\sum_{u \neq v} d(v, u)} \tag{12}$$

where $d(v, u)$ denotes the distance between nodes v and u . Some researchers have proposed other definitions of closeness to measure the locations of nodes [25, 26].

PageRank [27] evaluates the status of nodes in the random walking process in networks, which is also a core algorithm in the many search engines.

$$PR(v) = d \sum_{u \in \Gamma(v)} \frac{1}{k_u^{out}} PR(u) + (1 - d) \frac{1}{N} \tag{13}$$

where N denotes the total number of nodes, $\Gamma(v)$ denotes the set of neighbors of v , k_u^{out} denotes the out-degree of node u , and d is a dumping factor. In real implementations, we set $d = 0.85$.

Coreness [28] is a well-established centrality index that focuses on the structure of networks. Kitsak *et al.* found that the most efficient spreaders are those located within the core of the network as identified by the *k*-shell decomposition analysis. The decomposition runs in an iterative way. Nodes are assigned to *k* shells according to their remaining degrees, which are obtained by the successive pruning of nodes with degrees smaller than k_s . However, the performance of *k*-shell decomposition is not stable, and many studies have sought to enhance its effectiveness [29–31]. Recently, Liu *et al.* [32] analyzed the structure of core-like groups in networks, and improved the accuracy of the *k*-shell decomposition by filtering out the redundant links. In Ref [33], Lü *et al.* discovered an important relation among degree, H-index and coreness. By constructing a suitable operator, they proved that degree, H-index and coreness were the initial, intermediate and steady states of a special sequences, respectively.

Effectiveness

We use the SIR model to compare the effectiveness of *GeneralizedDegreeDiscount* with *DegreeDistance*, *SingleDiscount*, *DegreeDiscount* and several centrality-based methods discussed before. In each implementation, a fraction of the nodes is selected as spreaders, and the information spreads according to the SIR process described above. The spreading influence is used

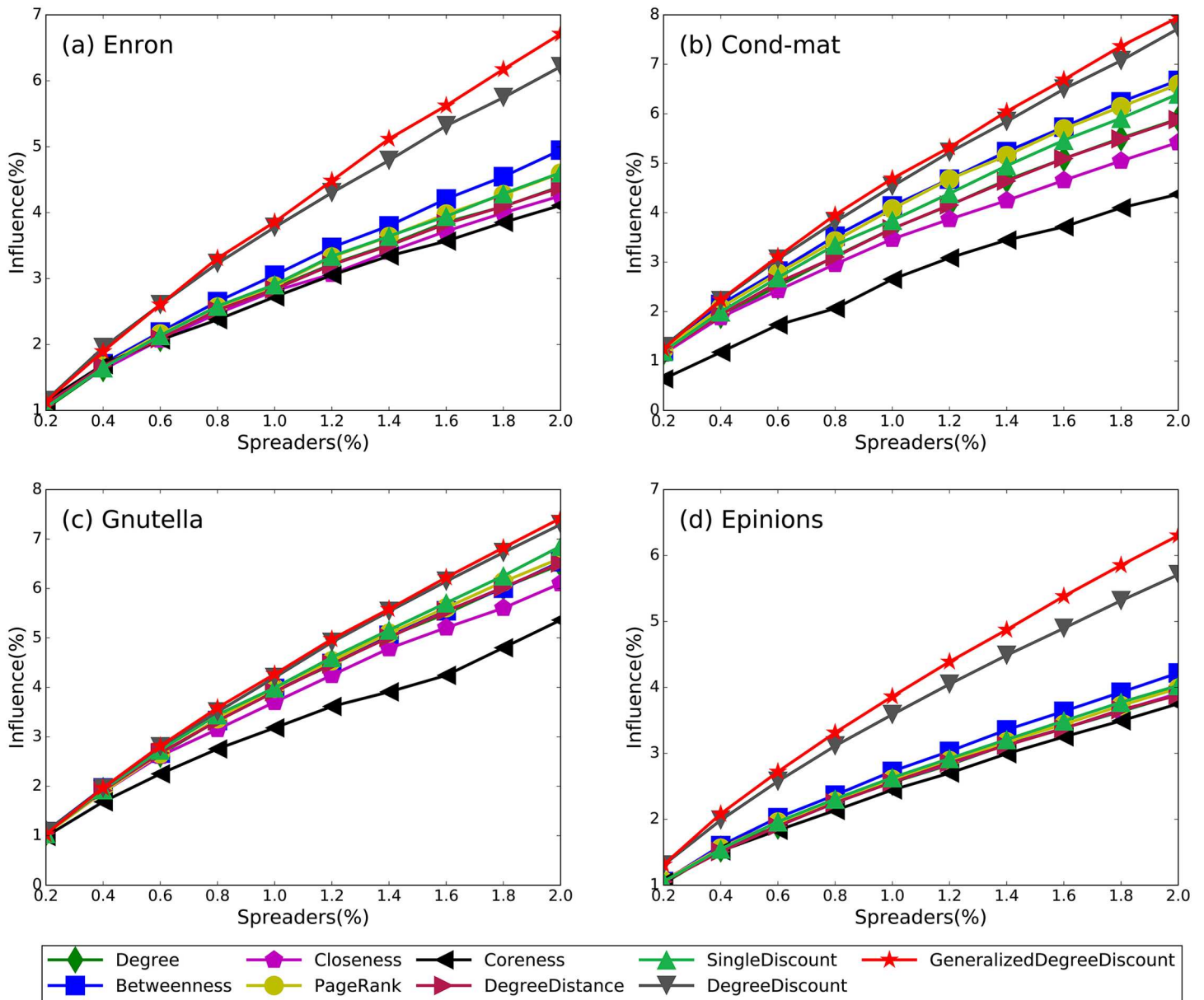


Fig 3. The spreading influence of nine methods on four networks with different fractions of spreaders. The parameters are $\lambda = 1.1$, $q = 1/\langle k \rangle$ for all networks, and all results are obtained by averaging over 200 implementations of the SIR model.

doi:10.1371/journal.pone.0164393.g003

to measure the effectiveness of the methods. For each method, the SIR process is repeated many times to ensure the stability of the results.

Fig 3 shows the numerical results of nine methods on four networks. The proposed *GeneralizedDegreeDiscount* outperforms all other methods on all four networks for almost all fraction of spreaders. Especially, as the fraction of spreaders increases, our method shows better and better performance. The only exception is in Enron network, when the fraction of spreaders is

small, the performance of *GeneralizedDegreeDiscount* is slightly worse than *DegreeDiscount*. Compared with *DegreeDiscount*, the performance of *GeneralizedDegreeDiscount* is consistently better. One promising phenomenon observed in our method is that as the fraction of spreaders becomes larger, the performance differences becomes more significant. Numerical results confirm that *GeneralizedDegreeDiscount* is indeed an effective extension of *DegreeDiscount*. In all networks, *Coreness* and *Closeness* perform the worst among all methods. In Ref [31], Liu *et al.* found that nodes in high shells may not be influential because of the existence of core-like groups: groups of nodes that link very locally within themselves. For nodes in the core-like groups, the *Coreness* cannot reflect their location importance in the network, which reduces the accuracy of the k-shell decomposition process. Moreover, if nodes in the highest shell tend to link with one another, their influence areas may overlap significantly. Obviously, selecting those nodes as spreaders may cause a large fraction of the network to be overlooked. The situation for *Closeness* is similar: nodes with high closeness values often distribute closely with one another.

In addition, we test the validity of our method with different effective spreading rates. We fix the fraction of spreaders to be 1% of the scale of the networks and vary the effective spreading rate λ . The results are shown in Fig 4. As in the previous experiments, the proposed *GeneralizedDegreeDiscount* shows a clear advantage in maximizing the spreading influence over all networks under various effective spreading rates. Even if the effective spreading rates are below 1.0 in four networks, *GeneralizedDegreeDiscount* performs better than other methods.

Obviously, *GeneralizedDegreeDiscount* is an adaptive method which recalculates the gdd_v during each step of the spreaders selection processes, while the centrality-based benchmark methods are not. In this part, more comparisons are done among our methods and adaptive versions of *Degree*, *Betweenness* and *Closeness*. To make them adaptive, a simple node-removing process is conducted: in each iteration, the node with the maximum centrality is selected as a spreader, and then we remove it from the network and recalculate the new centrality. The whole process ends until all the spreaders are selected. In fact, the adaptive version of *Degree* is the same as *SingleDiscount*. Fig 5 shows the numerical results. Unlike the previous results, when considering the top spreaders with low effective rate, our *GeneralizedDegreeDiscount* does not perform well. Especially in Gnutella network, the performance of *GeneralizedDegreeDiscount* is worse than *Betweenness-adaptive* and *Closeness-adaptive*. As the clustering coefficient of Gnutella is so small, the spreaders selection process in the early iteration of *GeneralizedDegreeDiscount* is just similar to *Degree*, which may limit the performance of our method. In Figs 3 and 4, it can also be seen that the performance differences between *GeneralizedDegreeDiscount* and other methods are not so remarkable under a small number of spreaders and low effective spreading rate. How to identify multiple influential spreaders in networks with low clustering coefficients is a challenging problem, and we leave it for the future.

Relations with other methods

In this subsection, we perform numerical comparisons among four degree-based methods: *Degree*, *SingleDiscount*, *DegreeDiscount* and *GeneralizedDegreeDiscount*. Though *DegreeDistance* is also a degree-based method, we do not consider it because there is no clear formulation to describe this method. The mathematical formulations of the four are listed below.

- *Degree*

$$d_v = d_v$$

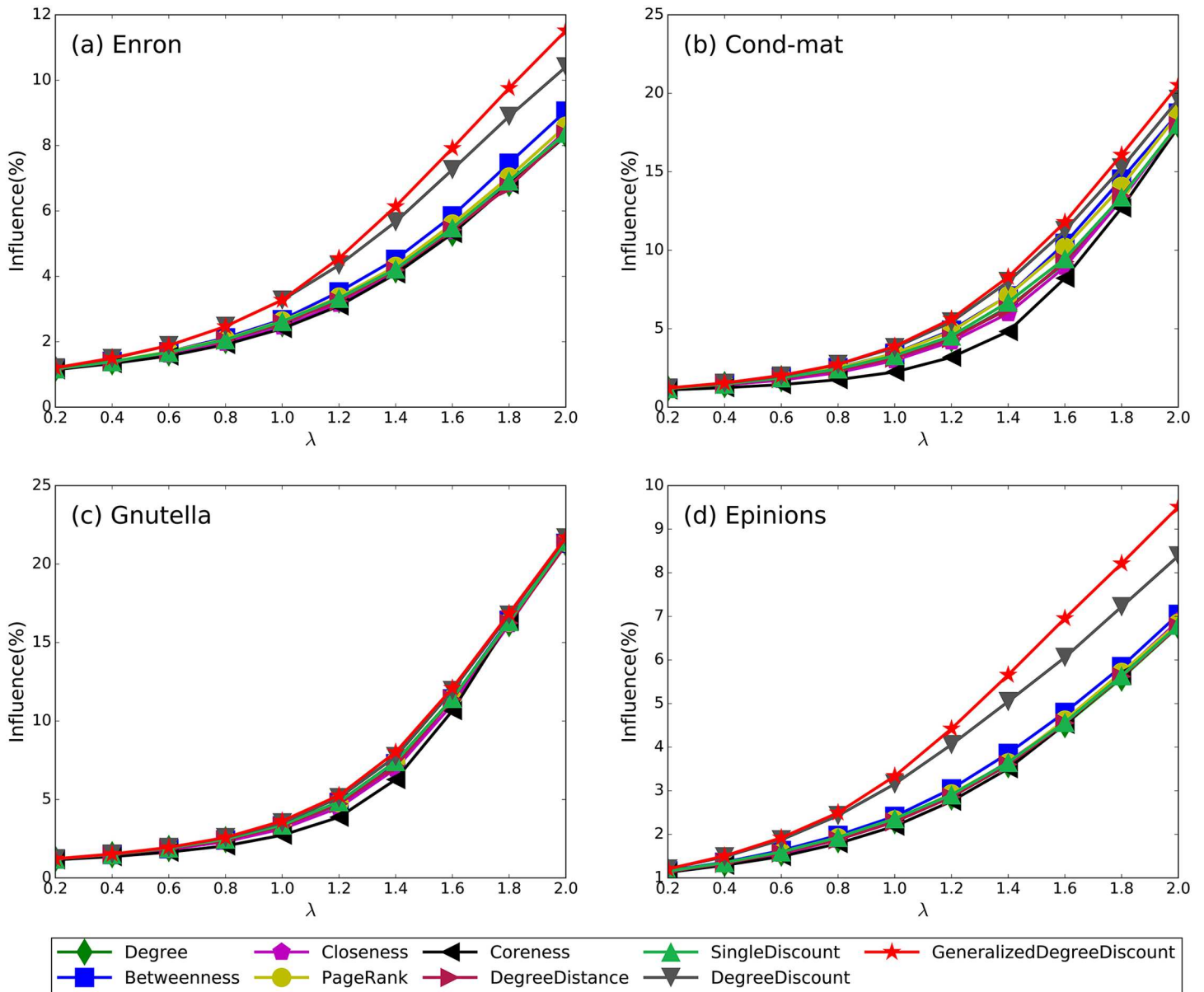


Fig 4. The spreading influence of nine methods on four networks under different effective spreading rates. The parameters are $\lambda = 1.1$, $q = 1/(k)$ for all networks, and all results are obtained by averaging over 200 implementations of the SIR model.

doi:10.1371/journal.pone.0164393.g004

- *SingleDiscount*

$$sd_v = d_v - t_v$$

- *DegreeDiscount*

$$dd_v = d_v - 2t_v - (d_v - t_v)t_v p$$

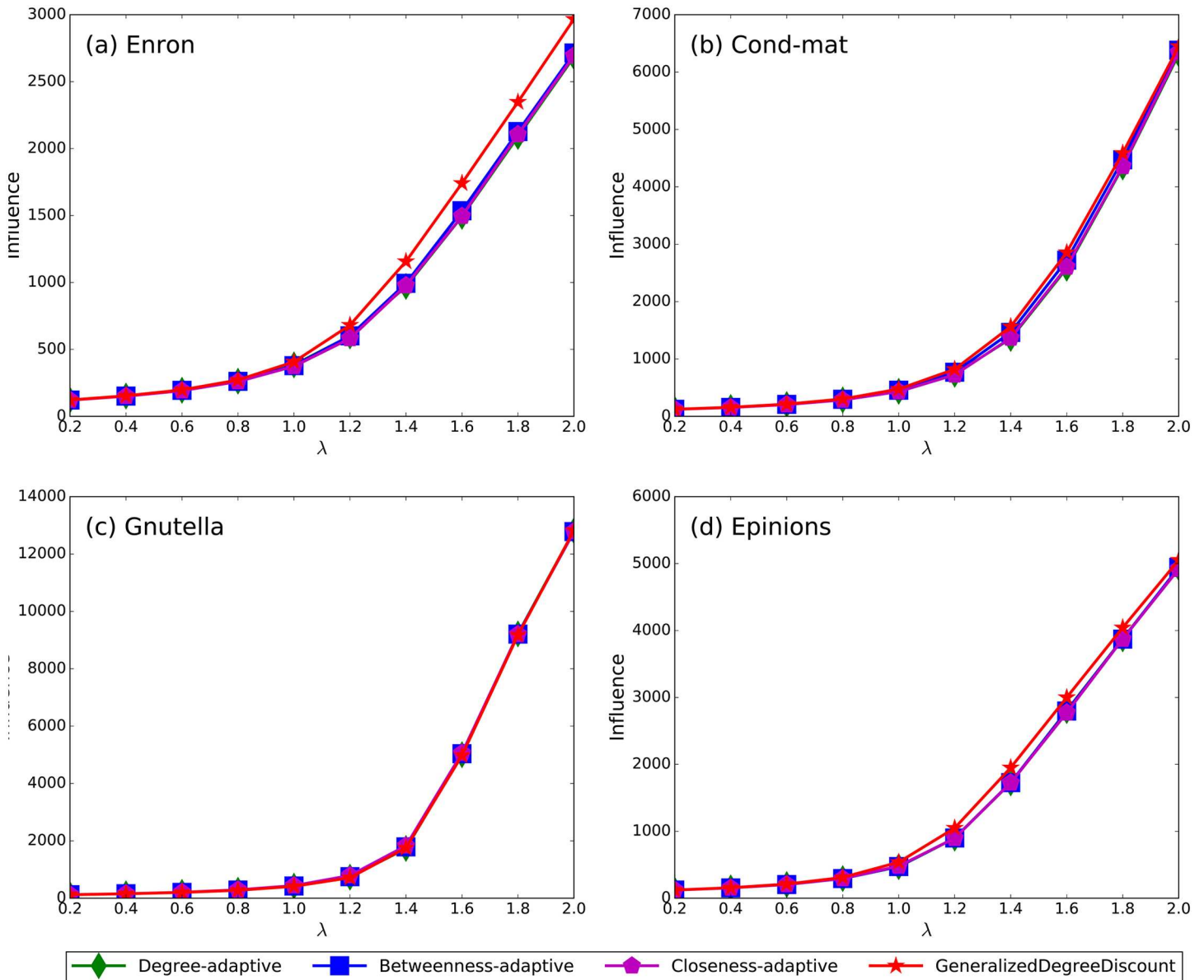


Fig 5. The spreading influence of GeneralizedDegreeDiscount and three adaptive centrality-based methods under different effective spreading rates. The numbers of spreaders are 100 in all networks, and the results are obtained by averaging over 200 implementations of the SIR model.

doi:10.1371/journal.pone.0164393.g005

- *GeneralizedDegreeDiscount*

$$gdd_v = d_v - 2t_v - (d_v - t_v)t_v p + \frac{1}{2}t_v(t_v - 1)p - \sum_{d_v - t_v} t_w p$$

Obviously, the complexity of these methods increases one by one. These formulations indicate that *DegreeDiscount* has more terms in common with *GeneralizedDegreeDiscount* than the other two methods. In an extreme case, when $p = 0$, *GeneralizedDegreeDiscount* is exact the same as *DegreeDiscount*. To better clarify the difference between the methods, we set the

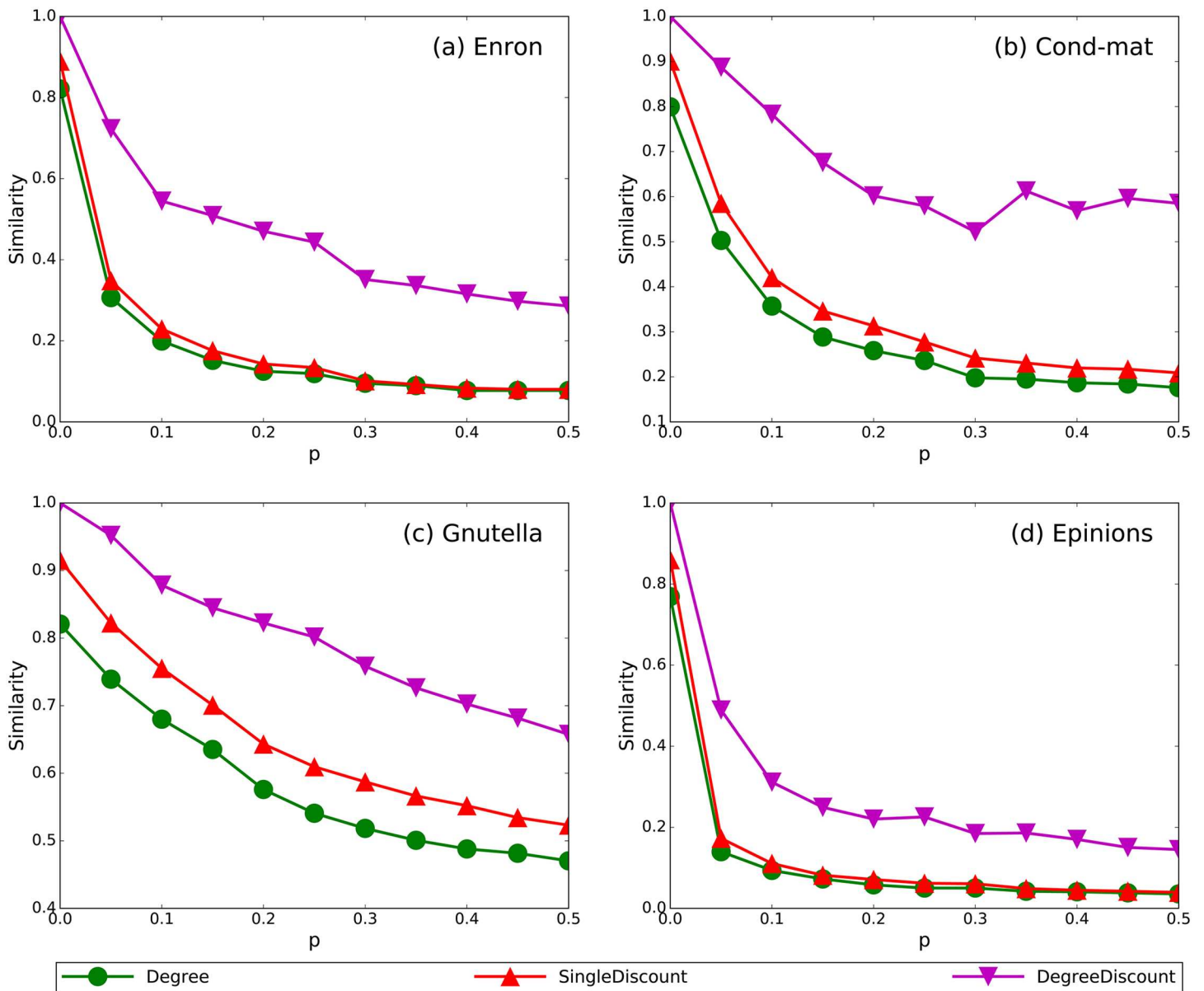


Fig 6. The similarities between GeneralizedDegreeDiscount and Degree, SingleDiscount, DegreeDiscount. The fraction of spreaders is 1%.

doi:10.1371/journal.pone.0164393.g006

fraction of spreaders to be 1% and calculate the similarity (the fraction of commonly selected spreaders) between GeneralizedDegreeDiscount and other methods. Fig 6 shows the results for the four networks. In all the networks, GeneralizedDegreeDiscount shows the best similarity with DegreeDiscount, normal similarity with SingleDiscount, and the worst similarity with Degree.

Conclusion

In this paper, we propose a novel degree-based heuristic, GeneralizedDegreeDiscount, which selects multiple spreaders and maximizes their spreading influence. In our method, when

evaluating the potential influence of a candidate node v , the way in which its neighbors are treated depends on whether it has been selected as a spreader or not. Taking both of the situations into consideration, *GeneralizedDegreeDiscount* uses a heuristic scheme to evaluate the potential influence of all individuals in the network.

We analyze the computational complexity of our method and show that it is just linearly correlated with the network scale. Then, the performance of our method is evaluated in four real networks from different fields. Results show that our method outperforms several centrality-based methods and other heuristic methods in all cases, no matter how many spreaders we choose to select or what the effective spreading rate is.

The theoretical analysis about influence maximization problem is still lacking. Although it has long been proven that there is a strong connection between the spreading process and the percolation process [34, 35], few researches have discussed the relationship between influence maximization and percolation. Recently, Morone and Makse pointed out that the influence maximization problem could be mapped onto optimal percolation problem in random networks [36], which might shed light on a new trend of future researches [37, 38]. Besides, we have witnessed the rapid development of theories and methods for temporal networks [39, 40]. Further researches on the influence maximization problem in temporal networks may also be a promising direction [41, 42].

Supporting Information

S1 File. Enron. Data of Enron email communication network.
(RAR)

S2 File. Cond-mat. Data of Cond-mat collaboration network.
(RAR)

S3 File. Gnutella. Data of Gnutella peer-to-peer network.
(RAR)

S4 File. Epinions. Data of Epinions social network.
(RAR)

Acknowledgments

The authors would like to thank Yanyuan Su and Yirun Ruan for useful comments. This work is partially supported by the major pre-research project of National University of Defense Technology, HPCL open fund of National University of Defense Technology(201403-01), Postgraduate Innovation Fund of Hunan province(CX2015B010), and Postgraduate Innovation Fund of NUDT(B150203).

Author Contributions

Conceptualization: XJW.

Formal analysis: XJW XZ.

Funding acquisition: DYY.

Methodology: XJW.

Project administration: DYY.

Software: XJW.

Supervision: CLZ.

Validation: XJW.

Writing – original draft: XJW.

Writing – review & editing: XZ.

References

1. Yang J, Yao C, Ma W, Chen G. A study of the spreading scheme for viral marketing based on a complex network model. *Physica A: Statistical Mechanics and its Applications*. 2010; 389(4):859–870. doi: [10.1016/j.physa.2009.10.034](https://doi.org/10.1016/j.physa.2009.10.034)
2. 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(4):1777–1787. doi: [10.1016/j.physa.2011.09.017](https://doi.org/10.1016/j.physa.2011.09.017)
3. Li Q, Zhou T, Lü L, Chen D. Identifying influential spreaders by weighted LeaderRank. *Physica A: Statistical Mechanics and its Applications*. 2014; 404(24):47–55. doi: [10.1016/j.physa.2014.02.041](https://doi.org/10.1016/j.physa.2014.02.041)
4. Lü L, Zhang YC, Yeung CH, Zhou T. Leaders in Social Networks, the Delicious Case. *PLoS ONE*. 2011; 6:e21202. doi: [10.1371/journal.pone.0021202](https://doi.org/10.1371/journal.pone.0021202) PMID: [21738620](https://pubmed.ncbi.nlm.nih.gov/21738620/)
5. Chen DB, Gao H, Lü L, Zhou T. Identifying Influential Nodes in Large-Scale Directed Networks: The Role of Clustering. *PLoS ONE*. 2013; 8:e77455. doi: [10.1371/journal.pone.0077455](https://doi.org/10.1371/journal.pone.0077455) PMID: [24204833](https://pubmed.ncbi.nlm.nih.gov/24204833/)
6. Chen DB, Xiao R, Zeng A, Zhang YC. Path diversity improves the identification of influential spreaders. *Epl*. 2013; 104(6):5580–5596. doi: [10.1209/0295-5075/104/68006](https://doi.org/10.1209/0295-5075/104/68006)
7. Domingos P, Richardson M. Mining the network value of customers. In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM; 2001. p. 57–66.
8. Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM; 2002. p. 61–70.
9. 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*. ACM; 2003. p. 137–146.
10. Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N. Cost-effective outbreak detection in networks. In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM; 2007. p. 420–429.
11. Narayanam R, Narahari Y. A shapley value-based approach to discover influential nodes in social networks. *Automation Science and Engineering, IEEE Transactions on*. 2011; 8(1):130–147. doi: [10.1109/TASE.2010.2052042](https://doi.org/10.1109/TASE.2010.2052042)
12. Zhao XY, Huang B, Tang M, Zhang HF, Chen DB. Identifying effective multiple spreaders by coloring complex networks. *EPL (Europhysics Letters)*. 2015; 108(6):68005. doi: [10.1209/0295-5075/108/68005](https://doi.org/10.1209/0295-5075/108/68005)
13. He JL, Fu Y, Chen DB. A Novel Top-k Strategy for Influence Maximization in Complex Networks with Community Structure. *PloS one*. 2015; 10(12):e0145283. doi: [10.1371/journal.pone.0145283](https://doi.org/10.1371/journal.pone.0145283) PMID: [26682706](https://pubmed.ncbi.nlm.nih.gov/26682706/)
14. Zhang JX, Chen DB, Dong Q, Zhao ZD. Identifying a set of influential spreaders in complex networks. *Scientific Reports*. 2016; 6:27823. doi: [10.1038/srep27823](https://doi.org/10.1038/srep27823) PMID: [27296252](https://pubmed.ncbi.nlm.nih.gov/27296252/)
15. 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*. ACM; 2009. p. 199–208.
16. Sheikahmadi A, Nematbakhsh MA, Shokrollahi A. Improving detection of influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*. 2015; 436:833–845. doi: [10.1016/j.physa.2015.04.035](https://doi.org/10.1016/j.physa.2015.04.035)
17. 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: [10.1016/j.physleta.2006.12.021](https://doi.org/10.1016/j.physleta.2006.12.021)
18. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*. 2009; 6(1):29–123. doi: [10.1080/15427951.2009.10129177](https://doi.org/10.1080/15427951.2009.10129177)

19. Newman ME. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*. 2001; 98(2):404–409. doi: [10.1073/pnas.98.2.404](https://doi.org/10.1073/pnas.98.2.404)
20. Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*. 2007; 1(1):2. doi: [10.1145/1217299.1217301](https://doi.org/10.1145/1217299.1217301)
21. Richardson M, Agrawal R, Domingos P. Trust management for the semantic web. In: *The Semantic Web-ISWC 2003*. Springer; 2003. p. 351–368. doi: [10.1007/978-3-540-39718-2_23](https://doi.org/10.1007/978-3-540-39718-2_23)
22. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*. 1977; p. 35–41.
23. Freeman LC. Centrality in social networks conceptual clarification. *Social networks*. 1979; 1(3):215–239. doi: [10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
24. Bavelas A. Communication patterns in task-oriented groups. *Journal of the acoustical society of America*. 1950; 22:725. doi: [10.1121/1.1906679](https://doi.org/10.1121/1.1906679)
25. Dangalchev C. Residual closeness in networks. *Physica A: Statistical Mechanics and its Applications*. 2006; 365(2):556–564. doi: [10.1016/j.physa.2005.12.020](https://doi.org/10.1016/j.physa.2005.12.020)
26. Holme P, Kim BJ, Yoon CN, Han SK. Attack vulnerability of complex networks. *Physical Review E*. 2002; 65(5):056109. doi: [10.1103/PhysRevE.65.056109](https://doi.org/10.1103/PhysRevE.65.056109)
27. Brin S, Page L. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*. 2012; 56(18):3825–3833. doi: [10.1016/j.comnet.2012.10.007](https://doi.org/10.1016/j.comnet.2012.10.007)
28. 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](https://doi.org/10.1038/nphys1746)
29. Zeng A, Zhang CJ. Ranking spreaders by decomposing complex networks. *Physics Letters A*. 2013; 377(14):1031–1035. doi: [10.1016/j.physleta.2013.02.039](https://doi.org/10.1016/j.physleta.2013.02.039)
30. Liu JG, Ren ZM, Guo Q. Ranking the spreading influence in complex networks. *Physica A: Statistical Mechanics and its Applications*. 2013; 392(18):4154–4159. doi: [10.1016/j.physa.2013.04.037](https://doi.org/10.1016/j.physa.2013.04.037)
31. Liu Y, Tang M, Zhou T, Do Y. Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Scientific Reports*. 2015; 5:9602. doi: [10.1038/srep09602](https://doi.org/10.1038/srep09602) PMID: [25946319](https://pubmed.ncbi.nlm.nih.gov/25946319/)
32. Liu Y, Tang M, Zhou T, Do Y. Improving the accuracy of the k-shell method by removing redundant links: From a perspective of spreading dynamics. *Scientific Reports*. 2015; 5:13172–13172. doi: [10.1038/srep13172](https://doi.org/10.1038/srep13172) PMID: [26277903](https://pubmed.ncbi.nlm.nih.gov/26277903/)
33. Lü L, Zhou T, Zhang QM, Stanley HE. The H-index of a network node and its relation to degree and coreness. *Nature communications*. 2016; 7:10168. doi: [10.1038/ncomms10168](https://doi.org/10.1038/ncomms10168) PMID: [26754161](https://pubmed.ncbi.nlm.nih.gov/26754161/)
34. Pastorsatorras R, Vespignani A. Immunization of complex networks. *Physical Review E*. 2002; 65:36104. doi: [10.1103/PhysRevE.65.036104](https://doi.org/10.1103/PhysRevE.65.036104)
35. Newman MEJ. Spread of epidemic disease on networks. *Physical Review E*. 2002; 66(1). doi: [10.1103/PhysRevE.66.016128](https://doi.org/10.1103/PhysRevE.66.016128)
36. Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. *Nature*. 2015; 524(7563):65. doi: [10.1038/nature14604](https://doi.org/10.1038/nature14604) PMID: [26131931](https://pubmed.ncbi.nlm.nih.gov/26131931/)
37. Ji S, Lü L, Yeung CH, Hu Y. Effective spreading from multiple leaders identified by percolation in social networks. *arXiv preprint arXiv:150804294*. 2015;.
38. Hu Y, Ji S, Feng L, Yuliang J. Quantify and Maximise Global Viral Influence Through Local Network Information. *arXiv preprint arXiv:150903484*. 2015;.
39. Holme P, Saramäki J. Temporal networks. *Physics reports*. 2012; 519(3):97–125. doi: [10.1016/j.physrep.2012.03.001](https://doi.org/10.1016/j.physrep.2012.03.001)
40. Holme P. Modern temporal network theory: a colloquium. *The European Physical Journal B*. 2015; 88(9):1–30. doi: [10.1140/epjb/e2015-60657-4](https://doi.org/10.1140/epjb/e2015-60657-4)
41. Ren G, Wang X. Epidemic spreading in time-varying community networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2014; 24(2):023116. doi: [10.1063/1.4876436](https://doi.org/10.1063/1.4876436)
42. Starnini M, Machens A, Cattuto C, Barrat A, Pastor-Satorras R. Immunization strategies for epidemic processes in time-varying contact networks. *Journal of theoretical biology*. 2013; 337:89–100. doi: [10.1016/j.jtbi.2013.07.004](https://doi.org/10.1016/j.jtbi.2013.07.004) PMID: [23871715](https://pubmed.ncbi.nlm.nih.gov/23871715/)