SCIENCE CHINA
Information Sciences

Relative influence maximization in competitive social networks

Dingda YANG[1,3], Xiangwen LIAO[2,3,*], Huawei SHEN[4], Xueqi CHENG[4] and Guolong CHEN[2,3]

Citation: SCIENCE CHINA Information Sciences **60**, 108101 (2017); doi: 10.1007/s11432-016-9080-3

View online: https://engine.scichina.com/doi/10.1007/s11432-016-9080-3

View Table of Contents: https://engine.scichina.com/publisher/scp/journal/SCIS/60/10

Published by the Science China Press

Articles you may be interested in

Reverse influence maximization algorithm in social networks
Journal of Computer Applications **40**, 1944 (2020);

Influence maximization with limit cost in social network
SCIENCE CHINA Information Sciences **56**, 078102 (2013);

Influence maximization in noisy networks
EPL **123**, 58007 (2018);

Active influence in dynamical models of structural balance in social networks
EPL **103**, 18001 (2013);

Identifying influential nodes in social networks via community structure and influence distribution difference
数字通信和网络 **7**, 131 (2021);

SCIENCE CHINA
Information Sciences

• HIGHLIGHT •

October 2017, Vol. 60 108101:1–108101:3
doi: 10.1007/s11432-016-9080-3

Relative influence maximization in competitive social networks

Dingda YANG[1,3], Xiangwen LIAO[2,3*], Huawei SHEN[4],
Xueqi CHENG[4] & Guolong CHEN[2,3]

[1]*College of Physics and Information Engineering, Fuzhou University, Fuzhou* 350116*, China;*
[2]*College of Mathematics and Computer Science, Fuzhou University, Fuzhou* 350116*, China;*
[3]*Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou* 350116*, China;*
[4]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190*, China*

Received December 16, 2016; accepted April 24, 2017; published online August 9, 2017

Citation Yang D D, Liao X W, Shen H W, et al. Relative influence maximization in competitive social networks. Sci China Inf Sci, 2017, 60(10): 108101, doi: 10.1007/s11432-016-9080-3

In many realistic scenarios, such as political election and viral marketing, two opposite opinions, i.e., positive opinion and negative opinion, spread simultaneously in the same social networks [1, 2]. Consequently, to achieve good word-of-mouth effect, it is desired to maximize the spread of positive opinions while reducing the spread of negative opinions, i.e., maximizing the difference between the spread of positive opinions and the spread of negative opinions.

In this article, we study the relative influence maximization (RIM) problem, which seeks to select initial individuals as a positive seed set under the existence of negative individuals, maximizing the difference between the spread of positive opinions and the spread of negative opinions, i.e., the relative influence. Existing methods approximately solve this problem either by promoting the spread of positive influence [1] or by limiting the spread of negative influence [2]. In this article, we theoretically analyze the intrinsic complexity of this problem and empirically develop efficient method to directly solve the RIM problem in social networks.

To describe the spread of two competitive opinions, we introduce a competitive independent cascade (CIC) model by extending the classical independent cascade (IC) model [3]. In CIC model, each individual is in one of three states, i.e., inactive, P-active and N-active. Individuals in inactive states are not influenced. Individuals in P(N)-active states stand for those who adopt the positive (negative) opinions. The diffusion process of positive and negative opinions unfolds independently as in IC model. When an individual is influenced by both positive and negative opinions simultaneously, negative opinion dominates over positive opinion, following the empirical observations [1].

Given such a competitive diffusion model, a network $G = (V, E)$, a set of initial adopters of negative opinions I_N, and a positive integer k, RIM aims to select an optimal positive seed set I_P with k nodes so that P-active individuals are more than N-active individuals as many as possible in the end of diffusion. Mathematically, the RIM problem could be formalized as

$$I_P = \arg \max_{|I_P|=k, I_P \subseteq V \setminus I_N} \{\sigma_P\left(I_P|I_N\right) - \sigma_N\left(I_P|I_N\right)\}, \tag{1}$$

where $\sigma_P\left(I_P|I_N\right)$ and $\sigma_N\left(I_P|I_N\right)$ are the spread of the positive and negative opinions, respectively.

* Corresponding author (email: liaoxw@fzu.edu.cn)
The authors declare that they have no conflict of interest.

© Science China Press and Springer-Verlag Berlin Heidelberg 2017
https://engine.scichina.com/doi/10.1007/s11432-016-9080-3

To ensure that the objective function (1) is non-negative, we transform it into an equivalent formulation

$$I_P = \arg \max_{|I_P|=k, I_P \subseteq V \setminus I_N} \{\sigma_P (I_P|I_N) + \sigma_{R_N} (I_P|I_N)\}, \quad (2)$$

where $\sigma_{R_N} (I_P|I_N) = \sigma_N (\emptyset|I_N) - \sigma_N (I_P|I_N)$. For convenience, we define $f(I_P|I_N) = \sigma_P(I_P|I_N) + \sigma_{R_N}(I_P|I_N)$.

We first analyze the properties of the RIM problem. When we remove the early adopters of negative opinions, CIC model is reduced to IC model and RIM problem is reduced to influence maximization problem, which has been proved to be NP-hard in [3]. Hence, RIM is also NP-hard. Following the practice in [4], we construct a subgraph G_i in advance and get the final results with a deterministic process. In this way, we proved that the function $f(I_P|I_N)$ is monotone and submodular in subgraph G_i. Limited by space, the details of the proof are omitted. A combination of monotone and submodular functions also keeps the original properties. Hence, $f(I_P|I_N)$ is monotone and submodular in G.

Method. Given that $f(I_P|I_N)$ is non-negative, monotone and submodular, we propose a greedy algorithm called GreedyRIM that achieves $(1 - 1/e)$ approximation ratio. Referring to [4], we determine diffusion subgraph G_i in advance. For the selection of ith-node, we scan the subgraph with CIC to obtain the influence of positive and negative nodes eventually by adding each node into the set I_P. To get an accurate approximation of the influence, we generate subgraphs for sufficient large R times and use the average over these R subgraphs as the final result. Finally, we add one node v into the selected set I_P such that v together with I_P maximizes $f(I_P|I_N)$.

The time complexity of GreedyRIM is $O(Rm + kRnm_1)$, where n is the number of nodes, m is the number of edges and m_1 is the maximum number of edges in each subgraphs. So this time-consuming algorithm is not suitable for large-scale social networks. An efficient solution is to utilize heuristics [5]. Two widely-used heuristics are random and degree based centrality. Random heuristic is the simplest, but is not solid. High-degree heuristic performs better than other heuristics, such as "Distance centrality". However, high-degree heuristic together with its improvement, e.g., SingleDiscount, is from the perspective of network topology, neglecting the practical situation that opposite opinions spread.

In this article, we propose a Distance-Sensitive heuristic centrality, referred to DS. The main idea of DS is that it considers the spread of the positive and negative opinions at the same time. The DS centrality of v is computed in the following way:

$$\mathrm{DS}(v) = (1 + p^x) \cdot \sum_{d=1}^{D} \left(\log N_d(v) \cdot p^d\right), \quad (3)$$

where p is the diffusion probability, x is the distance from node v to set I_N, and $N_d(v)$ is the number of reachable nodes from node v within distance d. The conceptual justification of DS has two aspects. In the first aspect, the centrality of v is inversely proportional to the distance from set I_N, i.e., it prefers to select the nodes near the set I_N so that it is helpful for limiting the spread of negative opinions. In the second aspect, the centrality of v is proportional to the number of neighbors and the nodes that are close to the v are more likely be activated. DS is estimated only on the regions of nodes, guaranteeing its efficiency.

We conduct experiments of algorithms on two collaboration networks (NetHEPT[1)] and Geom[2)]) and synthetic networks. We assign a base diffusion probability p, such that the diffusion probability from u to v is $p_{uv} = 1 - (1 - p)^{c_{uv}}$, where c_{uv} is the number of papers that the two authors collaborated. To void randomness of selection of negative nodes, 50 nodes with the highest degree are set as individuals with negative opinions initially. For different algorithm, we compare the relative influence of different k ranging from 1 to 100. All experiments are run on a Linux server with 2.8 GHz AMD Opteron(tm) Processor 6320 CPU and 32 G memory.

We first evaluate the effectiveness of GreedyRIM in collaboration networks. We elaborate two typical baselines, i.e., the greedy positive influence maximization algorithm (GreedyPIM) and the greedy negative influence minimization algorithm (GreedyNIM). They are general forms of existing methods on maximizing the positive influence and minimizing the negative influence [1,2]. Figure 1(a) and (b) show that GreedyRIM outperforms GreedyPIM and GreedyNIM in two networks. The results in the inset show that GreedyRIM algorithm requires fewer number of initial adopters of positive opinions to defeat negative influence (relative influence is greater than zero) than other algorithms. GreedyRIM is superior because it takes into account the spread of positive opinions and the existing of negative opinions. Existing methods, i.e., promoting the spread

1) http://www.arXiv.org.

2) Jones B. Computational Geometry Database, February 2002. http://jeffe.cs.illinois.edu/compgeom/biblios.html.

https://engine.scichina.com/doi/10.1007/s11432-016-9080-3

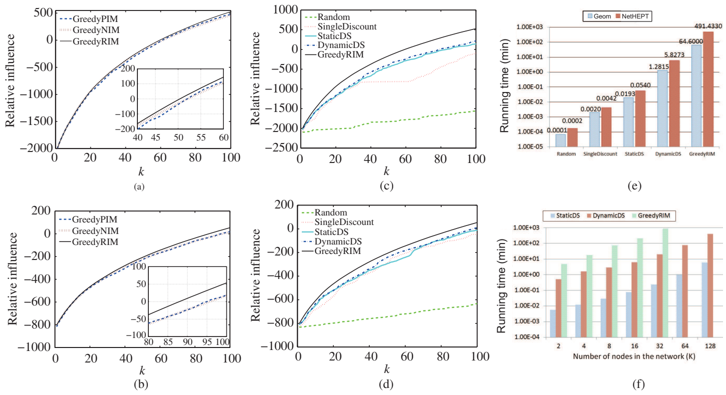

Figure 1 (Color online) Experimental results. (a) Relative influence of greedy algorithms with different k in NetHEPT; (b) relative influence of greedy algorithms with different k in Geom; (c) relative influence of heuristics with different k in NetHEPT; (d) relative influence of heuristics with different k in Geom; (e) running times with k =100 in NetHEPT and Geom; (f) running times with k =100 in synthetic networks.

of positive influence and limiting the spread of the negative influence, cannot achieve the relative influence maximization.

Next, we compare DS with SingleDiscount, which is a state-of-the-art heuristic. The results of random selection, referred to Random, are also reported. We subdivide DS into a static one (StaticDS), and dynamic one (DynamicDS), to distinguish whether the DS centrality is calculate dynamically. As shown in Figure 1(c) and (d), Random performs poorly in both two networks. SingleDiscount heuristic does not perform well in NetHEPT neither. DynamicDS has the best effectiveness, and StaticDS performs better than SingleDiscount for the overwhelming majority of cases. In fact, though SingleDiscount performs quite well in finding influential nodes [6], it still ignores that both positive opinions and negative opinions can diffuse in the networks.

GreedyRIM algorithm is a time consuming process, indicating that it is not suitable for large-scale social networks. All heuristics are orders of magnitude faster than GreedyRIM algorithm as shown in Figure 1(e). DS heuristic not only has an acceptable performance in effectiveness, but also has a drastic advantage in running time. We further test the scalability of DS in synthetic scale-free networks. The results in Figure 1(f) show that GreedyRIM reaches its feasibility limit in 32K-node network. The running time of DynamicDS could be tolerated in small scale networks, but not in very large networks. However, StaticDS can scale up well, so we choose StaticDS as a feasible and effective solution in large-scale networks.

Conclusion. In this article, we study RIM problem under CIC model. We propose a greedy algorithm, i.e., GreedyRIM, by utilizing the monotone and submodular properties of the objective function. We further propose a new Distance-Sensitive based (DS) heuristic. Experiments show that GreedyRIM achieves good effectiveness and DS heuristic gets high efficiency and better effectiveness than other heuristics.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61472400, 61300105), Research Fund for Doctoral Program of Higher Education of China (Grant No. 2012351410010), Key Project of Science and Technology of Fujian (Grant No. 2013H6012), Key Laboratory of Network Data Science & Technology, and Chinese Science and Technology Foundation (Grant No. CASNDST20140X).

References

1 Chen W, Collins A, Cummings R, et al. Influence maximization in social networks when negative opinions may emerge and propagate. In: Proceedings of 2012 SIAM International Conference on Data Mining, Mesa, 2011. 379–390

2 Budak C, Agrawal D, El Abbadi A. Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web, Hyderabad, 2011. 665–674

3 Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, 2003. 137–146

4 Cheng S Q, Shen H W, Huang J M, et al. Staticgreedy: solving the scalability-accuracy dilemma in influence maximization. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, 2013. 509–518

5 Cheng S Q, Shen H W, Huang J M, et al. IMRank: influence maximization via finding self-consistent ranking. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, Gold Coast, 2014. 475–484

6 Chen W, Wang Y J, Yang S Y. Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, 2009. 199–208

https://engine.scichina.com/doi/10.1007/s11432-016-9080-3