Computational Models for Social Influence Analysis

[Extended Abstract]

Jie Tang[‡] and Jimeng Sun[†]
[‡]Department of Computer Science and Technology, Tsinghua University
[†]School of Computational Science and Engineering, Georgia Institute of Technology
jietang@tsinghua.edu.cn, jsun@cc.gatech.edu

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Social influence, Social network, Information diffusion, Probabilistic models

1. ABSTRACT

Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally. In this article, we survey recent research progress on social influence analysis. In particular, we first give a brief overview of related background knowledge, and then discuss what is social influence. We try to answer this question in terms of homophily and the process of influence and selection. After that, we focus on describing computational models for social influence including models for influence probability learning and influence diffusion. Finally, we discuss potential applications of social influence.

Preliminaries First we introduce some basic knowledge for social network analysis, including related theories in sociology, fundamental models underlying social networks, and algorithms or measures for quantifying social influence. For social theories, we introduce social balance theory [10], social status theory [21], structural hole theory [6, 22], and two-step information flow theory [19]. For social network models, we introduce the Erdös-Rényi (ER) model [11], Small-World model [28], and the Barabási-Albert (BA) model [4]. For algorithms, we review several fundamental problems in graph theory and the corresponding algorithms to solve them including the Ford-Fulkerson algorithm for maximum flow in a flow network [12], the push-relabel maximum flow algorithm for finding k-densest subgraph [13], and the greedy algorithm for the set covering problem [1]. We will also go through standard measures and concepts of social networks and their connection to social influence such as centrality, clustering coefficient, closeness and betweenness. These measures are fundamental concepts about social network analysis, and are also deeply related to the importance or influence of nodes or edges in the networks.

Definition and Existential Test As there is no a formal definition for social influence, we discuss its definition in terms of several related concepts such as homophily [20], conformity [8, 27], and selection. We further describe methodologies for verifying the existence of influence in various social networks. The methods include shuffle test [2] and randomization test [24]. We will give real world examples to demonstrate how the social influence behaves in different social networks. For example, Bond et al. [5] conducted a randomized controlled trial by delivering political mobilization messages to 61 million Facebook users. Their results verified the existence of social influence on political voting behavior – when one is aware that their friends have made the political votes, their likelihood to vote will significantly increase. Bakshy et al. [3] also conducted randomized controlled trials to verify the existence of social influence on customer responses to advertising in Facebook.

Computational Models We now focus on describing the computational models for social influence analysis, with an emphasis on *influence quantification* and *influence diffusion*. In particular, for influence quantification, we introduce several popular methods for learning the influence probability between users. For example, Tang et al. [26] presented a Topical Affinity Propagation (TAP) approach to quantify the topic-level social influence in large networks. Goyal et al. [14] presented a method to learn the influence probabilities by counting the number of correlated social actions. Tan et a. [25] proposed a model to learn and distinguish the effects of influence, correlation, and uses' action dependency. Kutzkov et al. [18] extended influence probability learning to the stream data. They showed that the influence probabilities could be learned with one pass over the streaming data using only $\mathcal{O}(n \log n)$ space, where n is the number of nodes in a network.

For influence diffusion, we start with several state-of-the-art epidemic models such as Susceptible-Infectious-Recovered (SIR) [17], Susceptible-Infectious-Susceptible (SIS), and Susceptible-Infectious-Recovered-Susceptible (SIRS). Then, we focus on the two popular influence maximization model including independent cascaded model and linear threshold model. The problem of influence maximization has been formally defined as an algorithmic problem by Domingos and Richardson [9, 23]. Kempe et al. [16] further presented the independent cascaded model and the linear threshold model, and theoretically proved the NP-hardness of the two models. They defined the problem using submodular functions, with which a natural greedy strategy could obtain a solution that is provably within $(1 - 1/e)$ of optimal solution. Chen et al. [7] further developed efficient algorithms to

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577273.

approximately solve the influence maximization problem. Finally we will also discuss several extensions of the basic cascaded and linear threshold models, e.g., [15, 30, 29].

Applications Finally we use several real applications as examples to further demonstrate the usefulness of social influence analysis. We empirically study social influence in more than 10 datasets including Twitter, Weibo, Flickr, Gowalla, Coauthor, Mobile, Slashdot, Enron, Epinions, etc. We will share our experience and interesting findings when studying social influence on these data.

2. FIVE CHALLENGES

Understanding the fundamental mechanisms of social influence is a fundamental issue in social network analysis and represents a new and interesting research direction. As for the future work, there are still many challenges and also potential opportunities. Here we list of five major challenges.

- **Big network.** As social networks increasingly becoming larger, it is important to study how the existing computational models for social influence can scale up to handle the "big networks".
- **Globality vs. Locality.** Most existing works focus on studying peer influence, but ignore group effect. Influence could exist in one's personal circles or the whole society, as a resultant of various factors. It would be interesting to study the problem from different perspectives.
- **Dynamic evolution.** Social networks are rather dynamic. It is important to design computational models to capture the dynamic pattern underlying the social influence phenomenon.
- **Social theories.** It would be intriguing to connect the social influence phenomenon with some other social theories such as social status and structural holes so as to understand dynamic change of the network structure.
- **Applications.** There are many potential applications based on the results of social influence. It is important to demonstrate the effectiveness of applying influence in different applications.

Acknowledgements Jie Tang is supported by the National High-tech R&D Program (No. 2014AA015103), Natural Science Foundation of China (No. 61222212), and a research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology, and a research fund supported by Huawei Inc.

3. REFERENCES

[1] N. Alon, D. Moshkovitz, and S. Safra. Algorithmic construction of sets for k-restrictions. *ACM Trans. Algorithms*, 2(2):153–177, 2006.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *KDD'08*, pages 7–15, 2008.

[3] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: evidence from field experiments. In *EC'12*, pages 146–161, 2012.

[4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[5] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295–298, 2012.

[6] R. S. Burt. *Structural Holes: The Social Structure of Competition*. Harvard University Press, 1992.

[7] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD'09*, pages 199–207, 2009.

[8] R. B. Cialdini and N. J. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004.

[9] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD'01*, pages 57–66, 2001.

[10] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[11] P. Erdos and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[12] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[13] A. V. Goldberg. Finding a maximum density subgraph. Technical report, Massachusetts Institute of Technology, 1984.

[14] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *WSDM'10*, pages 241–250, 2010.

[15] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. A data-based approach to social influence maximization. *Proc. VLDB Endow.*, 5(1):73–84, 2012.

[16] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD'03*, pages 137–146, 2003.

[17] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.

[18] K. Kutzkov, A. Bifet, F. Bonchi, and A. Gionis. Strip: stream learning of influence probabilities. In *KDD'13*, pages 275–283, 2013.

[19] P. F. Lazarsfeld, B. Berelson, and H. Gaudet. *The people's choice: How the voter makes up his mind in a presidential campaign*. Columbia University Press, New York, USA, 1944.

[20] P. F. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. *M. Berger, T. Abel, and C. H. Page, editors, Freedom and control in modern society, New York: Van Nostrand*, pages 8–66, 1954.

[21] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *CHI'10*, pages 1361–1370, 2010.

[22] T. Lou and J. Tang. Mining structural hole spanners through information diffusion in social networks. In *WWW'13*, pages 837–848, 2013.

[23] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD'02*, pages 61–70, 2002.

[24] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[25] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In *KDD'10*, pages 1049–1058, 2010.

[26] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD'09*, pages 807–816, 2009.

[27] J. Tang, S. Wu, and J. Sun. Confluence: Conformity influence in large social networks. In *KDD'13*, pages 347–355, 2013.

[28] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, pages 440–442, Jun 1998.

[29] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li. Social influence locality for modeling retweeting behaviors. In *IJCAI'13*, pages 2761–2767, 2013.

[30] H. Zhuang, Y. Sun, J. Tang, J. Zhang, and X. Sun. Influence maximization in dynamic social networks. In *ICDM'13*, 2013.