Computational Models for Social Network Analysis: A Brief Survey

Jie Tang
Department of Computer Science and Technology, Tsinghua University
Tsinghua National Laboratory for Information Science and Technology (TNList)
jietang@tsinghua.edu.cn

ABSTRACT

With the exponential growth of online social network services such as Facebook and Twitter, social networks and social medias become more and more important, directly influencing politics, economics, and our daily life. Mining big social networks aims to collect and analyze web-scale social data to reveal patterns of individual and group behaviors. It is an inherently interdisciplinary academic field which emerged from sociology, psychology, statistics, and graph theory. In this article, I briefly survey recent progress on social network mining with an emphasis on understanding the interactions among users in the large dynamic social networks. I will start with some basic knowledge for social network analysis, including methodologies and tools for macro-level, meso-level and micro-level social network analysis. Then I will give an overall roadmap of social network mining. After that, I will describe methodologies for modeling user behavior including state-of-the-art methods for learning user profiles, and introduce recent progress on modeling dynamics of user behaviors using deep learning. Then I will present models and algorithms for quantitative analysis on social interactions including homophily and social influence.Finally, I will introduce network structure model including social group formation, and network topology generation. We will introduce recent developed network embedding algorithms for modeling social networks with the embedding techniques. Finally, I will use several concrete examples from *Alibaba*, the largest online shopping website in the world, and *WeChat*, the largest social messaging service in China, to explain how online social networks influence our offline world.

Keywords

Social networks; Big data; Social influence; User behavior

1. INTRODUCTION

The emergence and rapid proliferation of online social applications and media, such as instant messaging (e.g., Snapchat, WeChat, IRC, AIM, Jabber, Skype), sharing sites (e.g., Flickr, Picassa, YouTube, Plaxo), blogs (e.g., Blogger, WordPress, LiveJournal), wikis (e.g., Wikipedia, PBWiki), microblogs (e.g., Twitter, Jaiku, Weibo), social networks (e.g., Facebook, MySpace, Ning),

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3041021.3051101

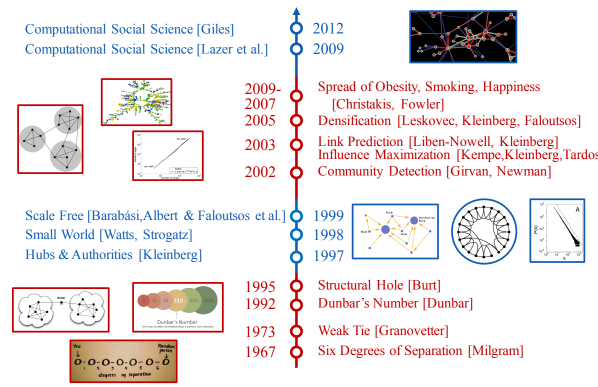

Figure 1: History of social network research.

scientific networks (e.g., DBLP, ArnetMiner), bring many opportunities for studying very large social networks, at the same time also pose a number of new challenges. From the social perspective, the online social networks already become a bridge to connect our physical daily life with the virtual Web space. Facebook has more than 1.65 billion users and Tencent (the largest social networking service in China) has attracted more than 800 million monthly active QQ users and 700 monthly active WeChat users in 2016.

History of social network research. The research of social network analysis can be traced back to fifty years ago. The main efforts were from sociology. For example, Milgram used several years to validate the existence of *small world* phenomenon, also referred to as six-degree of separation by sending mails to thousands of people [50]. Granovetter developed the theory of *Weak tie*, which suggests that weak ties between users are responsible for the majority of the embeddedness and structure of social networks in society as well as the transmission of information through these networks [29]. Even earlier, sociologists Georg Simmel proposed the concept of *structural theories* in sociology, which focus on the dynamic formation of triads [62] in 1910s and Jacob Moreno is the first to develop the sociograms to analyze people's inter-relationships in 1930s. More recently, Burt proposed the theory of structural hole [8] — a user is said to span a *structural hole* in a social network if she is linked to people in parts of the network that are otherwise not well connected to one another. Another interesting social phenomenon of Dunbar's number has been discovered by British anthropologist Robin Dunbar [18], indicating a cognitive limit to the number of people with whom one can maintain stable social relationships.

Later, in the end of twenty centenary, physicists put a great deal of efforts to study the networks formed by social users. Several state-of-the-art network generation models have been proposed, such as Erdős-Rényi (ER) model (random graph) [21], small-world model [72], Barabási-Albert model (preferential attachment) [5]. Faloutsos et al. [22] also presented a generative model to explain the growing patterns of Internet networks. In computer science, researchers also started to pay attention to the research of large online networks. For instance, PageRank is an algorithm used by Google Search to rank websites in their search engine results [54]; and Hyperlink-Induced Topic Search (HITS) is also a link analysis algorithm to rate web pages according to authority and "hub" degrees. Both algorithms can be used to measure the importance of nodes in a large network.

Starting form the beginning of twenty-first centenary, in particular with the rapid development of online social networks, more and more researchers from multiple different disciplinary united the efforts to study the dynamic patterns underlying the evolutionary social networks. The interdisciplinary research has been developing into a new scientific research direction: network science. A relatively formal definition of the concept **Computational Social Science** was given in [40, 26]. This line of research include user behavior modeling, community detection, link prediction, social influence analysis, and network topology analysis. Community detection in networks is one of the most popular topics of modern network science. Communities are groups of vertices having higher probability of being connected to each other than to members of other groups, though other patterns are possible [52, 27, 53]. Liben-Nowell and Kleinberg [44] systematically investigate the problem of inferring new links among users given a snapshot of a social network. They introduced several unsupervised approaches to deal with this problem based on "proximity" of nodes in a network—or the principle of homophily [39] ("birds of a feather flock together" [48]). Christakis and Fowler [24] have developed the theory of Three-Degree-of-Influence, which posits that "everything we do or say tends to ripple through our network, having an impact even on our three degree of friends (friends' friends' friends)". Domingos, et al. [16], Richardson [56], and Kempe et al. [36] formally defined the problem of *influence maximization*, and proposed two popular social influence models: linear threshold model and independent cascaded model [16, 56, 36]. In both models, the objective is to find a small subset of users (seed users) to adopt a behavior (e.g., adopt a product), and the goal is to trigger a large cascade of further adoptions through the influence diffusion model.

Research roadmap for social network mining. Figure 2 gives the research roadmap for social network mining. In general, existing research centers around individuals, interactions between users and topological structure of the formed network by the users. On the top of social network mining, we can consider many applications such as social prediction [65], social search [15, 68], information diffusion [30], and social advertisement [4]. The underlying theories for social network mining include theories from social science and algorithmic foundations from computer science. In the rest of this article, we review related literature on social theories, user behavior modeling, social tie analysis, and network analysis.

2. RELATED LITERATURE

2.1 Social Theories

A basic principle for mining social networks in this book is to incorporate social theories into data mining (or machine learning) model. For social theories, we mainly consider Social balance [20], Social status [13], Structural hole [8], Two-step information-flow [38], and Strong/Weak tie hypothesis [29, 37].

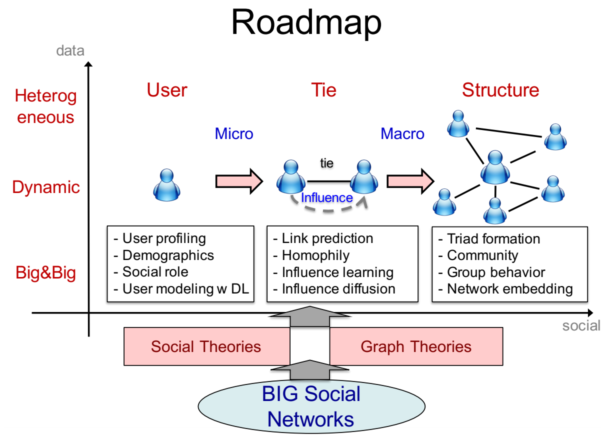

Figure 2: Roadmap of research on social network analysis.

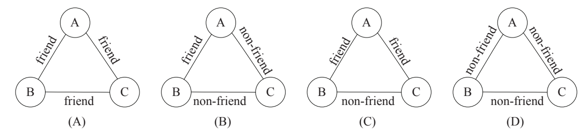

Figure 3: Illustration of structural balance theory. (A) and (B) are balanced, while (C) and (D) are not balanced.

Social balance theory suggests that people in a social network tend to form into a balanced network structure. Figure 3 shows such an example to illustrate the structural balance theory over triads, which is the simplest group structure to which balance theory applies. For a triad, the balance theory implies that either all three of these users are friends—"the friend of my friend is my friend"—or only one pair of them are friends—"the enemy of my enemy is my friend".

Another social psychological theory is the *theory of status* [13, 31, 42]. This theory is based on the directed relationship network. Suppose each directed relationship is labeled by a positive sign "+" or a negative sign "-" (where sign "+"/"-" denotes the target node has a higher/lower status than the source node). Then status theory posits that if, in a triangle on three nodes (also called triad), we take each negative edge, reverse its direction, and flip its sign to positive, then the resulting triangle (with all positive edge signs) should be acyclic. Figure 4 illustrates four examples. The first two triangles satisfy the status ordering and the latter two do not satisfy it.

Roughly speaking, a user is said to span a *structural hole* in a social network if she is linked to people in parts of the network that are otherwise not well connected to one another [8]. Such user is also referred to as *structural hole spanner* [46]. Arguments based on structural holes suggest that there is an informational advantage to have friends in a network who do not know each other. A sales manager with a diverse range of connections can be considered as a structural hole spanner, with a number of potentially *weak ties* [29] to individuals in different communities. More generally, we can think about Web sites such as eBay as spanning structural holes, in that they facilitate economic interactions between people who would otherwise not be able to find each other.

The *two-step information-flow theory* is first introduced in [38] and further elaborated in literature [33, 34]. The theory suggests that ideas (innovations) usually flow first to *opinion leaders*, and

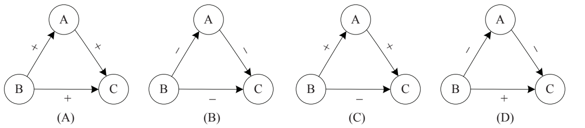

Figure 4: Illustration of status theory. **(A) and (B) satisfy the status theory, while (C) and (D) do not satisfy the status theory. Here positive "+" denotes the target node has a higher status than the source node; and negative "-" denotes the target node has a lower status than the source node. In total there are 16 different cases.**

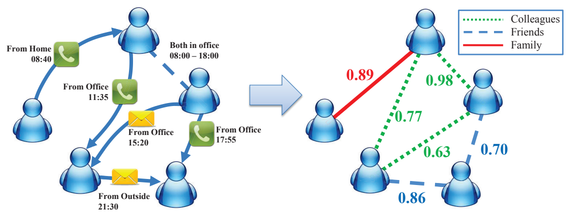

Figure 5: An example of inferring social ties in a mobile communication network. The left figure is the input of the task, and the right figure is the output of the task of inferring social ties.

then from them to a wider population. In the enterprise email network, for example, managers may act as opinion leaders to help spread information to subordinates.

Interpersonal ties generally come in three varieties: strong, weak, or absent. *Strong tie hypothesis* implies that one's close friends tend to move in the same circles that she/he does, while *Weak tie hypothesis* argues that weak ties are responsible for the majority of the embeddedness and structure of social networks in society as well as the transmission of information through these networks [29].

2.2 Social Tie Analysis

Mining social ties is an important problem in social network analysis. Based on the strong/weak tie hypothesis, there is a bunch of research conducted in recent years. The goal of *social tie analysis* is to automatically recognize the semantics associated with each social relationship. Awareness of the semantics of social relationships can benefit many applications. For example, if we could have extracted friendships between users from the mobile communication network, then we can leverage the friendships for a "word-of-mouth" promotion of a new product. Figure 5 gives an example of relationship mining in mobile calling network. The left figure is the input of the problem: a mobile social network consisting of users, calls and messages between users, and users' location logs, etc. The objective is to infer the type of the relationships in the network. In the right figure, the users who are family members are connected with red-colored lines, friends are connected with blue-colored dash lines, and colleagues are connected with green-colored dotted lines. The probability associated with each relationship represents our confidence on the detected relationship types.

There are several works on mining the relationship semantics. Diehl et al. [14] try to identify the manager-subordinate relationships by learning a ranking function. They define a ranking objective function and cast the relationship identification as a relationship ranking problem. Menon et al. [49] proposed a log-linear matrix model for dyadic prediction. They use matrix factorization to derive latent features and incorporate the latent features for predicting the label of user relationships. Wang et al. [71] proposed a probabilistic model for mining the advisor-advisee relationships from the publication network. The proposed model is referred to as time-constrained probabilistic factor graph model (TFGM), which supports both supervised and unsupervised learning. Eagle et al. [19] presented several patterns discovered in mobile phone data, and try to use these pattern to infer the friendship network. Tang et al. [69] developed a classification framework of social media based on differentiating different types of social connections. However, these algorithms mainly focus on a specific domain, while our model is general and can be applied to different domains. Moreover, these methods also do not explicitly consider the correlation information between different relationships.

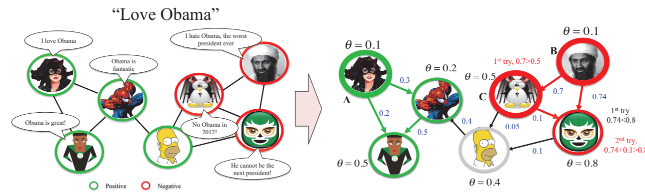

Figure 6: An example of social influence for political mobilization. The left figure is the input of the task, and the right figure is the output: influence probability between users, individual conformity of each user, and key influencers (A, B, C).

Another research branch is to predict and recommend unknown links in social networks. Liben-Nowell et al. [44] study the problem of inferring new interactions among users given a snapshot of a social network. They develop several unsupervised approaches to deal with this problem based on measures for analyzing the "proximity" of nodes in a network. The principle is mainly based on similarity of either content or structure between users. Backstrom et al. [3] propose a supervised random walk algorithm to estimate the strength of social links. Leskovec et al. [41] employ a logistic regression model to predict positive and negative links in online social networks, where the positive links indicate the relationships such as friendship, while negative indicates opposition. However, these works consider only the black-white social networks, and do not consider the types of the relationships.

Recently, Hopcroft et al. [32] explore the problem of reciprocal relationship prediction and Lou et al. [47] extend to study how social relationships develop into triadic closure. They propose a learning framework to formulate the problem of reciprocal relationship prediction into a graphical model and evaluate the proposed method on a Twitter data set. The framework is demonstrated to be very effective, i.e., it is possible to accurately infer 90% of reciprocal relationships in a dynamic network. Tang et al. [66] further propose a general framework for classifying the type of social relationships by learning across heterogeneous networks. The idea is to use social theories (e.g., social balance theory, social status theory, structural hole theory, two-step flow theory, and strong/weak tie) as bridge to connect different social networks. Social theory-based features are defined and incorporated into a triad-based factor graph model to infer the type of social relationships in different networks.

2.3 Social Influence Analysis

Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally [35]. Recently, social influence analysis has attracted a lot interests from both research and industry communities. In general, existing research on social influence analysis can be classified into three categories: influence test, influence measure, and influence diffusion models. Figure 6 shows an example of social influence for political mobilization. The left figure is the input of the task: opinion of each user for "Obama" in the social network, and the right figure is the output: influence probability between users on this topic "Obama", individual conformity of each user, and key influencers (A, B, C).

Influene Test. Several efforts have been made for identifying the existence of the social influence in the online social networks. For example, Anagnostopoulos et al. [1] gives a theoretical justification to identify influence as a source of social correlation when the time series of user actions is available. They propose a shuffle test to prove the existence of the social influence. Singla and Richardson [63] study the correlation between personal behaviors and their interests. They found that in the online system people who chat with each other (using instant messaging) are more likely to share interests (their Web searches are the same or topically similar), and the more time they spend talking, the stronger this relationship is. Bond et al. [7] reported results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users. They found that when one is aware that their friends have made the political votes, their likelihood to vote will significantly increase. Crandall et al. [12] further investigate the correlation between social similarity and influence. More recently, some efforts have been made for analyzing the dynamics in the social network. For example, Scripps et al. [59] investigate how different pre-processing decisions and different network forces such as selection and influence affect the modeling of dynamic networks. Other similar work can be referred to [17].

Influence Measure. The goal of influence measure is to quantify the strength of influence between users. Tang et al. [67] introduced the problem of topic-based social influence analysis. They proposed a Topical Affinity Propagation (TAP) approach to describe the problem via using a graphical probabilistic model. However, these works neither consider heterogeneous information nor learn topics and influence strength jointly. Goyal et al. [28] and Saito et al. [57] measure the pairwise influence between two individuals based on the idea of independent cascade model [36]. Liu et al. [45] also study the problem of measuring the influence on different topics. They propose a generative graphical model which leverages both heterogeneous link information and textual content associated with each user in the network to mine topic-level influence strength. Based on the learned direct influence, we further study the influence propagation and aggregation mechanisms: conservative and non-conservative propagations to derive the indirect influence. Xin et al. [61] study the indirect influence using the theory of quantum cognition. Myers et al. [51] propose a probabilistic model to quantify the external influence out-of-network sources. Belak et al. [6] investigate and measure the influence on the cross-community level so as to to provide a coarse-grained picture of a potentially very large network. They present a framework for cross-community influence analysis and evaluate the proposed method on an ten-year data set from the largest Irish online discussion system Boards.ie. Zhang et al. [74] propose the notion of social influence locality and apply it for modeling users' retweeting behaviors in the social networks. They develop two instantiation functions based on pairwise influence and structural diversity.

Influence Diffusion Models. Social influence has been applied in the application of influence maximization in viral marketing. Domingos and Richardson [16, 56] are the first to study influence maximization as an algorithmic problem. Kempe et al. [36] take the first step to formulate influence maximization as a discrete optimization problem. Leskovec et al. [43] and Chen et al. [10, 11] make efforts to improve the efficiency of influence maximization. Gruhl et al. [30] propose a time-decayed diffusion model for blogging writing, and use an EM-like algorithm to estimate the influence probabilities. Yang et al. [73] study the interplay between users' social roles and their influence on information diffusion. They propose a Role-Aware INformation diffusion model (RAIN) that integrates social role recognition and diffusion modeling into a unified framework.

2.4 User Modeling and Actions

User modeling describes the process of building up a user model to characterize user's skills, declarative knowledge, and specific needs to a system [23].

A number of models have been proposed to model users' behaviors in dynamic social networks. Sarkar et al. [58] develop a generalized model associating each entity in Euclidean latent space and use kernel functions for similarity in latent space to model friendship drifting over time. Tan et al. [65] study how users' behaviors (actions) in a social network are influenced by various factors such as personal interests, social influence, and global trends. They propose a Noise Tolerant Time-varying Factor Graph Model (NTT-FGM) for modeling and predicting social actions, which simultaneously models social network structure, user attributes and user action history for better prediction of the users' future actions. Tan et al. [64] have investigated how users' sentiment can be inferred in the social network by incorporating the social network information. Scripps et al. [59] present a model to investigate how different pre-processing decisions and different network forces such as selection and influence affect the modeling of dynamic networks. They also demonstrate the effects of attribute drifting and the importance of individual attributes in forming links over time.

Group analysis is based on the view that deep lasting change can occur within a carefully formed group whose combined membership reflects the wider norms of society. There is an interest, in group analysis, on the relationship between the individual group member and the rest of the group resulting in a strengthening of both, and a better integration of the individual with his or her community, family and social network. Shi et al. [60] study the pattern of user participation behavior, and the feature factors that influence such behavior on different forum data sets. Backstrom et al. [2] propose a partitioning on the data that selects for active communities of engaged individuals.

3. SUMMARY

To summarize, mining big social networking data represents an interesting and important research direction. There are still many challenges and also potential future directions on this topic, including how to achieve a trade off between big and maybe "small" data? what is the fundamental relationship between micro individual behavior and the macro network phenomenon? how to reveal real causality from the correlated data?

4. REFERENCES

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

[2] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins. Preferential behavior in online groups. In *WSDM'08*, pages 117–128, 2008.

[3] L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In *WSDM'11*, pages 635–644, 2011.

[4] 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.

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

[6] V. Belak, S. Lam, and C. Hayes. Cross-community influence in discussion fora. In *ICWSM'12*, pages 34–41, 2012.

[7] 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.

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

[9] M. E. Califf and R. J. Mooney. Relational learning of pattern-match rules for information extraction. In *Proceedings of Association for the Advancement of Artificial Intelligence (AAAI'99)*, pages 328–334, 1999.

[10] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD'10*, pages 1029–1038, 2010.

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

[12] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *KDD'08*, pages 160–168, 2008.

[13] J. A. Davis and S. Leinhardt. The structure of positive interpersonal relations in small groups. In J. Berger, editor, *Sociological Theories in Progress*, volume 2, pages 218–251. Houghton Mifflin, 1972.

[14] C. P. Diehl, G. Namata, and L. Getoor. Relationship identification for social network discovery. In *AAAI*, pages 546–552, 2007.

[15] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *science*, 301(5634):827–829, 2003.

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

[17] Y. Dourisboure, F. Geraci, and M. Pellegrini. Extraction and classification of dense communities in the web. In *WWW'2007*, pages 461–470, 2007.

[18] R. I. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of human evolution*, 22(6):469–493, 1992.

[19] N. Eagle, A. S. Pentland, and D. Lazer. Inferring social network structure using mobile phone data. *PNAS*, 106(36), 2009.

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

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

[22] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM'99*, pages 251–262, 1999.

[23] G. Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1-2):65–86, 2001.

[24] J. H. Fowler and N. A. Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. In *British Medical Journal*, 2008.

[25] L. Getoor and B. Taskar. *Introduction to statistical relational learning*. The MIT Press, 2007.

[26] J. Giles et al. Making the links. *Nature*, 488(7412):448–450, 2012.

[27] M. Girvan and M. E. Newman. Community structure in social and biological networks. *PNAS*, 99(12):7821–7826, 2002.

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

[29] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[30] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW'04*, pages 491–501, 2004.

[31] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW'04*, pages 403–412, 2004.

[32] J. Hopcroft, T. Lou, and J. Tang. Who will follow you back? reciprocal relationship prediction. In *CIKM'11*, pages 1137–1146, 2011.

[33] E. Katz. The two-step flow of communication: an up-to-date report of an hypothesis. *In Enis and Cox(eds.), Marketing Classics*, pages 175–193, 1973.

[34] E. Katz and P. F. Lazarsfeld. *Personal Influence*. The Free Press, New York, USA, 1955.

[35] H. C. Kelman. Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 2(1):51–60, 1958.

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

[37] D. Krackhardt. *The Strength of Strong ties: the importance of philos in networks and organization in Book of Nitin Nohria and Robert G. Eccles (Ed.), Networks and Organizations*. Cambridge, Harvard Business School Press, Hershey, USA, 1992.

[38] 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.

[39] 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.

[40] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Life in the network: the coming age of computational social science. *Science*, 323(5915):721, 2009.

[41] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW'10*, pages 641–650, 2010.

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

[43] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD'07*, pages 420–429, 2007.

[44] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.

[45] L. Liu, J. Tang, J. Han, and S. Yang. Learning influence from heterogeneous social networks. *Data Mining and Knowledge Discovery*, 25(3):511–544, 2012.

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

[47] T. Lou, J. Tang, J. Hopcroft, Z. Fang, and X. Ding. Learning to predict reciprocity and triadic closure in social networks. *TKDD*, 2013, (accepted).

[48] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[49] A. K. Menon and C. Elkan. A log-linear model with latent features for dyadic prediction. In *ICDM*, pages 364–373, 2010.

[50] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[51] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *KDD'12*, pages 33–41, 2012.

[52] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64(2):025102, 2001.

[53] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69(066133), 2004.

[54] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford University, 1999.

[55] A. Popescul and L. H. Ungar. Statistical relational learning for link prediction. In *IJCAI03 Workshop on Learning Statistical Models from Relational Data*, volume 149, page 172, 2003.

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

[57] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *KES '08*, pages 67–75, 2008.

[58] P. Sarkar and A. W. Moore. Dynamic social network analysis using latent space models. *SIGKDD Explor. Newsl.*, 7(2):31–40, 2005.

[59] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In *KDD'2009*, pages 747–756, 2009.

[60] X. Shi, J. Zhu, R. Cai, and L. Zhang. User grouping behavior in online forums. In *KDD'09*, pages 777–786, 2009.

[61] X. Shuai, Y. Ding, J. Busemeyer, S. Chen, Y. Sun, and J. Tang. Modeling indirect influence on twitter. *IJSWIS*, 8(4):20–36, 2012.

[62] G. Simmel. *Sociological Theory*. 7th ed. New York: McGraw?Hill, 2008.

[63] P. Singla and M. Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In *WWW'08*, pages 655–664, 2008.

[64] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li. User-level sentiment analysis incorporating social networks. In *KDD'11*, pages 1397–1405, 2011.

[65] 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.

[66] J. Tang, T. Lou, and J. Kleinberg. Inferring social ties across heterogeneous networks. In *WSDM'12*, pages 743–752, 2012.

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

[68] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: Extraction and mining of academic social networks. In *KDD'08*, pages 990–998, 2008.

[69] L. Tang and H. Liu. Relational learning via latent social dimensions. In *KDD'09*, pages 817–826, 2009.

[70] B. Taskar, M. F. Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In *NIPS*, 2003.

[71] C. Wang, J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. Mining advisor-advisee relationships from research publication networks. In *KDD'10*, pages 203–212, 2010.

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

[73] Y. Yang, J. Tang, C. W.-k. Leung, Y. Sun, Q. Chen, J. Li, and Q. Yang. Rain: Social role-aware information diffusion. In *AAAI'15*, 2015.

[74] 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.