A survey of social network and information dissemination analysis

Chin. Sci. Bull. (2014) 59(32):4163–4172
DOI 10.1007/s11434-014-0368-5

csb.scichina.com
www.springer.com/scp

Review

Computer Science & Technology

A survey of social network and information dissemination analysis

Binxing Fang · Yan Jia · Yi Han · Shasha Li · Bin Zhou

Received: 15 October 2013 / Accepted: 10 March 2014 / Published online: 7 May 2014

Abstract In recent years, with the rapid growth of social network services (SNS), social networks pervade nearly every aspect of our daily lives. Social networks are influencing today's societal and cultural issues, and changing the way of people seeing themselves. To fully understand the running mechanisms of social networks, in this paper, we aim at series of high knitted and important elements of online social networks. We mainly focus on 3 important but also open research problems, they are (1) structural properties and evolving laws, (2) social crowds and their interaction behaviors and (3) information and its diffusion. In this paper, we review the related work on the 3 problems. Then, we briefly introduce some interesting research directions and our progress on these research problems.

Keywords Social network · Topological structure · Crowds · Information dissemination

SPECIAL TOPIC: Network and Information Security

B. Fang
School of Computer, Beijing University of Posts and Telecommunication, Beijing 100876, China

Y. Jia · Y. Han (✉) · S. Li · B. Zhou
College of Computer, National University of Defense Technology, Changsha 410073, China
e-mail: yihan@nudt.edu.cn

Y. Han
Department of Computer Science, Peking University, Beijing 100871, China

Y. Han
National Engineering Laboratory of Information Content Security Technologies, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

1 Introduction

In recent years, with the rapid growth of social network services, such as Weibo, Twitter and Facebook, social networks pervade nearly every aspect of our daily lives. The role of social networks has been rapidly becoming beyond platforms of communication and making connections. It plays an important role as indispensable tools for professional networking, social recommendations, or advertisement. Internet-based social networks consist with the most important virtual society of maintaining social relationships. Meanwhile, social networks also have far-reaching effects for national security and social development.

With billions of different connections, individuals constitute a "relational structure" on social networks, which includes a large number of complex relationships, such as social communities, social ties, or linkage farms, etc. Based on the relationship of social network structure, the connected individuals gather with a large number of ongoing events. They influence each other by interactions. Thus, the individuals form a variety of networking crowds with common behavioral characteristics. Based on relational structure and social networking crowds, various kinds of information have been quickly published and disseminated, which leads to the formation of the social media. Virtual world gives feedback to the reality societies. Thus, the virtual and reality keep interacting and making influence to each other.

Social networks influence today's societal and cultural issues, and change the way of people seeing themselves. To fully understand the running mechanisms of social networks, in this paper, we aim at series of high knitted and important elements of online social networks. They are topological structure, crowds and information.

Firstly, we are doing research on looking for properties, modeling approaches, and evolving laws in online social networks. Secondly, we focus on mining the crowds' formation mechanism, group behaviors, and interaction within crowds. Furthermore, we are going to find the information dissemination patterns within various social networks as well as the mutual influence interaction with traditional media.

The above research work is tightly related with many fields, including computer science, sociology, management, psychology and other disciplines. The study on social network analysis can reveal the complex relationship and interaction laws among topological structure, crowds and information. Furthermore, our research can also provide an important theoretical support to social network and information dissemination analysis.

2 Survey on social network analysis

A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the ties between these actors. Existing researches on social networks involve three perspectives: topological structure and its evolution; crowds' interaction and mutual influence; information and its dissemination.

2.1 Structure and structure evolution theory of social network

In terms of structure and its evolution of social network, existing researches involve social network modeling, social network structure characteristic analysis, community detection and evolution theory.

2.1.1 Social network modeling

Graph theory is widely applied in the field of social network modeling. Researchers tried to use Graph theory to quantitatively analyze social networks and have achieved promising results.

Based on Ref. [1], Graph theory was firstly introduced into human social network analysis in 1930s, by Moreno et al. A few years later, Harary et al. explored a directed graph model which characterized asymmetric relations among social actors as directed edges and first defined the concept of "centrality" in 1960s. In 2008, Tong et al. [2] modeled relations between social actors which belong to different categories with a bipartite graph and proposed methods to measure proximity between two actors in the same category. Sequentially, a mathematical model of tripartite structures was proposed by Ghoshal et al. [3] in 2009. The model represented tripartite structures in social networks as random hypergraphs. In 2012, Kim and Leskovec [4] proposed a multiplicative attribute graphs model to build relationships between node attributes and network structure. The experiments showed that proper MAG models could generate networks following power-law distribution. Moreover, in 2005, Pei et al. [5] studied the problem of multi-social network modeling. They developed a cross-graph model for several interesting applications such as cross-market customer segmentation and investigated a heuristic algorithm to mine cross-graph quasi-cliques.

2.1.2 Structural characteristics analysis of social network

With the development of online social network services, a lot of researches are done in the fields of structural characteristic analysis on large-scale social network. These researches have verified that social network is a kind of complex network in terms of structural features, such as "six degrees of separation", "scale-free network", "power-law distribution", etc. For example, Dodds et al. [6] verified six degrees of separation on organizational network with 60,000 nodes in 2003. In 2005, Liben et al. [7] studied path length on Yahoo Online Community. They found that the average path length was 8 and the qualified path length was 10, larger than 6 when the community was in the largest scale. In 2007, Golder et al. [8] reported that the median of degrees was 144 and average degree was 179.53 on Facebook; in 2010, Charu and Wang [9] considered in-degree in directed graph as a factor reflecting the popularity of a node and found that popularity also followed power-law distribution in social network. In 2012, Liu et al. [10] analyzed event-based social network on Meetup data. The result indicated that the node degrees followed long-tail distribution and social interaction showed strong locality.

2.1.3 Community detection on social network

According to the definition of community structure, community detection can be categorized into non-overlapping community detection and overlapping community detection.

In terms of non-overlapping community detection, a widely concerned algorithm was explored by Clauset et al. [11] in 2004, which tried to find community in social network by searching for a network division with maximum modularity. However, optimization algorithm would suffer resolution limitation problem as figured out by Fortunato and Barthélemy [12] in 2007. That means that modularity optimization may fail to identify modules smaller than a scale which depends on the total number L of links of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined.

In terms of overlapping community detection, in 2005, Palla et al. [13] proposed to uncover the overlapping community structure of complex networks with a K-clique based approach and the experiment demonstrated that community structures have non-trivial correlations and specific scaling properties. In 2009, Lancichinetti et al. [14] pointed out that networks often showed a hierarchical organization, with communities embedded within other communities. Furthermore, they presented the first algorithm that finds both overlapping communities and the hierarchical structure based on the local optimization of a fitness function. In the same year, Shen et al. [15] tried to address the problem by an algorithm (EAGLE) which dealt with the set of maximal cliques and adopts an agglomerative framework.

However, all above algorithms were explored under the assumption that the networks were complete which is not always true in the real world applications. Recently, researchers have paid more attention to community detection in some typical networks. For example, in 2012, Chen et al. [16] studied the community detection problem on sparse undirected unweighted networks, in which the in-cluster and across-cluster edge densities were very small. And, they presented a convex optimization formulation, essentially a weighted version of low-rank matrix decomposition, to address the problem. Lin et al. [17] studied the problem of detecting communities in incomplete information networks with missing edges. A hierarchical clustering approach was explored, which estimated the distance between nodes with a distance metric learned from the observed edges in the local information regions.

2.1.4 Structure evolution analysis on social network

In the fields of structure evolution analysis on social network, researchers studied the statistical laws of social network evolution and proposed evolution models at macro, meso and micro levels.

At macro level, Leskovec et al. [18] developed a graph-based method in 2005 to characterize dynamic association and its time evolution between the number of nodes and network diameter. In 2006, Chakrabarti et al. [19] proposed to use data mining technology to model both time and structure evolution on social network. In 2007, Chi et al. [20] extended similarity measurement and firstly proposed evolutionary spectral clustering which used graph cut to measure community structure and evolution. In 2012, Allamanis et al. [21] analyzed the impact of location on the forming of social relations. They found that the process of gravitational attachment which considered both node degrees and distances could simulate the location based evolution of social network.

At meso level, in 2007, Tantipathananandh et al. [22] listed various structural changes at the meso level on social network according to human experience. Then, they built a cost optimization model to simulation structure evolution on social network, which integrated individual cost, group cost and color changing cost on corresponding structural changes. Bródka et al. [23] developed an event-based method in 2013 for group evolution discovery, which analyzed group continuing, shrinking, growing, splitting, merging, discovering and forming, and proposed a measurement on group changing. In the same year, Chen et al. [24] presented a principled approach to detect overlapping temporal community structure in time-evolving networks, which found the overlapping temporal community structure that maximized a quality function associated with each snapshot of the network subject to a temporal smoothness constraint.

At micro level, Leskovec et al. [25] studied microscopic evolution on social network in 2008 and pointed out that the generation of edges was inversely proportional to the distance between nodes.

2.2 Group interaction and mutual influence on social network

Existing researches on group interaction and mutual influence in social network involve two aspects. One is group behavior modeling and behavior feature analysis and the other is group sentiment modeling and interaction.

2.2.1 Group behavior modeling and behavior feature analysis

Most researches on group behavior modeling focus on analysis and models of users' social network selection and migration. For example, Ellison et al. [26] studied the impact factors for users' social network selection in 2007. They divided resources accumulated from various group behavior relations into bridging, bonding and maintained social capital and found that bridging social capital was the most important for individuals to select an online social network. While, studies of Steinfield et al. [27] in 2008 indicated that besides social capital outcomes, users' self-esteem and users' satisfaction on real-life are also psychological variables to influence users' social network selection. In 2009, Kang et al. [28] studied the role of self-image congruity and regret in social network selection. Chen et al. [29] figured out that the main reasons of users' migration include satisfaction of present social network, attraction of other networks and the cost of migration, etc. In 2012, Abbasi et al. [30] studied heterogeneity and migration among 7 popular online social network, including Facebook, Twitter, YouTube, etc. They modeled the migration of individual/group to present how users migrate from one online social network to another.

In the fields of individual behavior feature analysis, earlier researches such as Schramm [31] analyzed individual behavior from the view of communication studies and proposed methods to construct individual behavior feature pattern abstractly. Recently, with the development of Internet and large collection of web log, more researches have been done to quantitatively analyze and model online individual behavior. For example, in 2003, Eirinaki and Vazirgiannis [32] compared individual behavior feature mining functions in various web services. In 2008, Jiang et al. [33] proposed a practical method to mine individual preferences and regular pattern of their behaviors in a multidimensional space where users' preferences on some categorical attributes were unknown, from some superior and inferior examples provided by a user. Wong et al. [34] proposed an online skyline analysis based method in 2009 to measure the uniqueness of attributes to mine individual behavior features in a group.

2.2.2 *Group sentiment modeling and interaction analysis*

Existing researches on group sentiment modeling and interaction involve analysis the relationship between individual interaction and group sentiment forming, sentiment modeling and herd behavior analysis.

As early as 1960, Schramm [31] has built relatively complete theoretical system on group and interaction from the view of communication studies. With the rising of online social network, the problem attracted researchers' attentions again. Studies on online social network emerged in large numbers. For example, in 2007, Ellison et al. [26] figured out that online interaction could statistically support association among users and promote the forming of group sentiment instead of isolating offline users. In 2009, Camerer [35] investigated a game theory based method to model and analyze influences among cascading behavior, individual effect and group effect on a social network. In 2011, Bollen et al. [36] performed analysis on Twitter and pointed out that users could message their sentiment polarity to connected users so that they would gradually hold the same or similar subjective feelings. While, Xu et al. [37] detected sentiment community by using the sentiment polarity held by users as features.

In the field of sentiment modeling, Gryc and Moilanen [38] developed a text polarity mining algorithm in 2010 which used community division on blog space as a classification feature. The experiments analyzed polarity to Obama on a dataset consisting of 16741 bloggers and 2,800,000 blogs. In 2011, Aoki and Uchida [39] modeled diverse emotions with emotional vectors, but they only utilized emoticons to construct vectors without text content information. In 2012, Nguyen et al. [40] presented a strategy of building statistical models from the social media dynamics to predict collective sentiment dynamics. They modeled the collective sentiment change without delving into micro analysis of individual tweets or users and their corresponding low level network structures. Experiments on large-scale Twitter data show that the model can achieve above 85 % accuracy on directional sentiment prediction.

Herd behavior described how individuals in a group can act together without planned together. Group sentiment and interaction played important roles in herd behaviors. Recent researches on herd behavior on social network mainly focus on its generation environment, impact factors and mechanism. In 2008, Chen [41] studied herd behavior in online shopping from three different views of online ratings, sales volume and recommendation sources. The results indicated that online ratings, sales volume and decision of other users could influence consumers' decision. Furthermore, consumers preferred recommendations from other users and outside the website instead of expert and the website. In 2009, Yang et al. [42] used cellular automata to simulate herd behavior and the experiments showed that the results of herd behaviors are sensitive to the original state. In 2010, Sharawneh and Williams [43] proposed a collaborative filtering algorithm to model herd behavior in online shopping which integrated credit of opinion leader and social network information. The study indicated that the recommendation based on opinion leader could predict users' selection behavior more accurately. In 2012, Pei and Wu [44] proposed a dynamic model to reflect the average state of emergency crowd behavior combined with characteristics of crowd behavior. The result showed that infection capability decided the infection among individuals, while infection rate and cure rate were the keys to the emergency crowd. In the same year, Lewis et al. [45] analyzed social selection and peer influence on online social network using stochastic actor-based models. Using data on the Facebook activity of a cohort of college students over 4 years, they found that students who shared certain tastes in music and in movies, but not in books, were significantly likely to befriend one another. Meanwhile, they found little evidence for the diffusion of tastes among Facebook friends—except for tastes in classical/jazz music.

2.3 Information and information dissemination on social network

Existing researches on social network information and its dissemination involve three aspects: information and its energy, information dissemination model and the influence of information diffusion.

2.3.1 *Information and information energy on social network*

Existing researches on information and its energy mainly focus on two fields: symbolic representation of information, and information energy and its evolution.

Researches on symbolic representation of information can go back to 1920s. In 1923, Cassirer [46] analyzed the relationship between information symbols and their meanings. In 1940s, information theory was developed by Shannon [47] and Jaynes [48] to find fundamental limits on signal processing operations such as compressing data and on reliably storing and communicating data. A key measure of information is entropy, which is usually expressed by the average number of bits needed to store or communicate one symbol in a message. Entropy can be used not only to measure the amount of information but also to quantify the uncertainty involved in predicting the value of a random variable. Since the inception of the theory, it has broadened to find applications in many other areas, including statistical inference, natural language processing, model selection in ecology, and other forms of data analysis. Though information theory is well-developed, it's still limited to describe equilibrium. Actually, information and information entropy are always changing with time and space.

In the past decades, researches on information energy and information evolution also attracted researchers' attention. In 2004, Danchin et al. [49] proposed to use state entropy to characterize status and potential of information dissemination on social network. They pointed out that information energy was enhanced with the complement and evolution of heterologous heterogeneous information in the process of information dissemination. In the same year, Xing [50] introduced the non-equilibrium statistical information theory. The theory used information entropy evolution equation to present and qualify information evolution mechanism. In 2011, Lin et al. [51] proposed an integrated method to analyze topic evolution and information propagation path on social network. The method used Gaussian conditional random field models to analyze text content, influence and topic evolution simultaneously. In Meanwhile, Jo et al. [52] tried to model and analyze the process of information evolution through a topic topological graph. Kumar et al. [53] also proposed a novel analysis method which presented topic with multi-frequency close set of words and used improved PageRank algorithm to calculate the evolution of various sub-topic. And, Hong [54] developed a machine learning based algorithm to predict how long a message would be forwarded, which combined message content, publish time, meta information of the publisher and message and etc. Romero [55] studied topic evolution based on hashtag in Twitter.

2.3.2 *Social network information dissemination model*

Existing researches on information dissemination models mainly focus on infection model, topological graph based model, and statistical inference based models.

Infection model was first introduced in 2000 by Hethcote [56] and then many flavors were investigated such as SIR, SIS, SIRS, etc. [57]. In infection models, an uninfected individual would be infected if the probability of it to be infected by infected ones is larger than a threshold, in which, S, I, and R indicate 3 types of individuals. They are susceptible, infected, and removed, respectively. In 2004, Gruhl et al. [58] proposed a SIRS-based model to simulate topic diffusion in blog space and investigated algorithm to estimate reading ratio and infection ratio of a blog. However, in 2010, Zhou et al. [59] indicated that the threshold in infection models on scale-free networks was close to zero, which means that even small source of infection would spread throughout the network.

In terms of topological graph based models, in 2006, Kumar et al. [60] studied the evolution of the network topology on online social network, including the evolution of degree distribution, clustering coefficient and etc.

In terms of statistical inference based models, in 2000, Doucet et al. [61] proposed BRPT algorithm, which utilized decomposition of dynamic probabilistic network to reduce the dimension of the sampling distribution, and local sampling to estimate local distribution in order to improve the computational efficiency. In 2001, Kask et al. [62] investigated Buchet-Elimination algorithm; and in the same year, Murphy and Weiss [63] developed Junction tree based static information propagation algorithm, which could simulate information propagation in middle-scale network efficiently. In 2003, Paskin [64] proposed CBK algorithm which introduced conditional independence of local structures to improve the calculation accuracy of the dynamic information propagation.

Furthermore, in the fields of communication studies, in 1948, Lasswell [65] introduced 5W propagation mode theory. In 1960, Schramm [31] further established mass propagation mode based on control theory, which introduced a feedback effect and paid more attention on the role of information codecs and feedback in information dissemination.

2.3.3 *Influence analysis in information dissemination on social network*

Existing methods of influence analysis in information dissemination can be categories into three genres: Probabilistic methods, econometric methods, and methods of communication.

Probabilistic methods model the procedure of information dissemination on social network with a probabilistic graph. The sensitivity and influence of information dissemination are analyzed through qualifying parameter sensitivity and the structure of the graph. In 2006, McCandless et al. [66] proposed a Monte Carlo simulation based information propagation algorithm. However, the algorithm couldn't reveal nonlinear relationships when analyzing influence in information diffusion. In the same time, researchers tried to find ways to analyze influence in information diffusion in probabilistic structure models. For example, Wang [67] studied Bayesian networks and found it sensitive to parameters. The experiment results showed that sensitivity analysis was efficient to parameter analysis in Bayesian networks. While, Wang's sensitivity analysis of Bayesian network approaches involved only a single parameter. In 2012, Renooij [68] extends sensitivity analysis of Bayesian network to more than one parameter using a dynamical probabilistic structure. In 2011, Bakshy et al. [69] investigated the attributes and relative influence of 1.6 M Twitter users by tracking 74 million diffusion events that took place on the Twitter follower graph over a two month interval in 2009. In 2012, Aral and Walker [70] presented a method that used in vivo randomized experimentation to identify influence and susceptibility in networks. The experiment results revealed that influential people with influential friends may be instrumental in the spread of this product in the network.

Utility theory in economics is another effective theory to measure the influence of information dissemination. In 1992, Shenoy [71] developed a utility assessment based resolving method to estimate influence of information propagation. In 1999, Madsen [72] investigated an entropy-based function for influence estimation. In 2006, Maes and Philippe [73] proposed an information analysis method based on Multi-Agent causal probabilistic graphical model, which only use part of the information.

In the fields of communication studies, since the 1970s, there has been a research climax in dissemination effect resulting in a series of theories. In 1970, knowledge gap theory was proposed by Tichenor et al. [74]. They believed that the increase of information in society is not evenly acquired by every member of society: people with higher socioeconomic status tend to have better ability to acquire information. In 1974, Neumann [75] introduced the theory of "spiral of silence". The theory is one that explores hypotheses to determine why some groups remain silent while others are more vocal in forums of public disclosure. In 1972, McCombs and Shaw [76] developed agenda-setting theory which describes the ability of the news media to influence the salience of topics on the public agenda. That is, if a news item is covered frequently and prominently the audience will regard the issue as more important.

3 Research problems of online social network analysis

We consider an online social network consists of 3 elements, they are (1) social structures, denoted by *SocStruct*; (2) crowds, denoted by *CommInter*; and (3) information, denoted by *InforPro*. The social network is represented by a quintuple:

$$G = \langle P, N, I, F, EQ \rangle. \tag{1}$$

In Eq. (1), P is the set of individuals. $N = \langle V, E\rangle$ is the network structure. V is the set of vertices. E is the set of edges. $I := \langle \text{Content,Source} \rangle$ denotes the information on social networks. F denotes the impact factor controlling the network evolving. EQ denotes kinetic equation that determines the evolution of social networking. Please note that, the evolution of social networking can be regarded as a complex system. EQ is an abstract representation, which represents a kind of similar problems.

To fully understand the running mechanisms of social networks, in this paper, we mainly focus on 3 important but also open research problems, they are (1) structural properties and evolving laws; (2) crowds and their interaction behavior and (3) information and its dissemination. In this section, we briefly introduce above research problems. Figure 1 shows the research problems and interesting directions.

3.1 Structural properties and evolving laws

In existing social network modeling and evolution analysis work, online social network cannot be accurately modeled with existing methods due to its large scale, structural complexity, and multi-dimensional evolution. Therefore, new theories and new models should be developed to efficiently model and analyze online social network.

In order to fully understand the structural properties and evolving laws, some essential questions is open as follows. What kinds of modeling methods have the ability of reflecting both the nature of social networks, but also support the calculation and analysis? What calculation

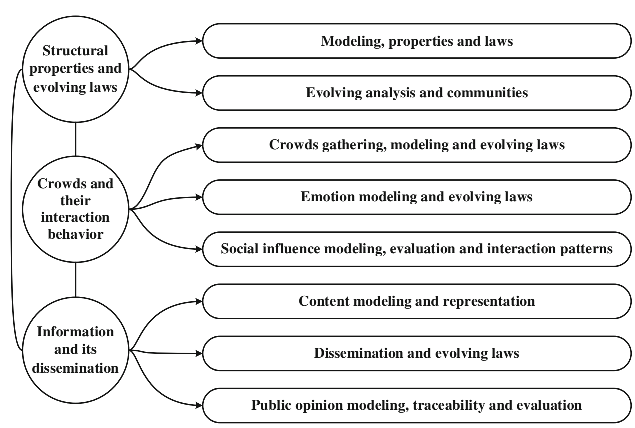

Fig. 1 Research problems

method can accurately characterize the evolution of social network? The problem can be formulated by a quintuple:

$$SocStruct := \langle P, N, I, \lambda, EQ_s \rangle. \quad (2)$$

In Eq. (2), $\lambda := \lambda(p, n, i)$ denotes the structural factor controlling the network evolving, in which $p \in P$, $n \subset N$ and $i \in I$. $EQ_S := \{eq \in N(t) | N(t+1) = (N(t), P, I, \lambda)\}$ is the kinetic energy equation that determines the evolution of social networking structure. The research on structure analysis can be carried out from the following 2 aspects.

(1) Modeling, properties and laws. Based on the techniques of heterogeneous network analysis, topological folding and decomposition, combined with, the social network hybrid graph model can be designed by considering both crowds evolution and information dissemination. Moreover, based on network hybrid graph model, some common characteristics of social networks can be extracted by adopting both multi-attribute association analysis and hierarchical constraint extraction method.
(2) Evolving analysis and communities. We consider the real community structure should be overlapping. Thus, virtual community discovery methods can be developed by density estimation on microstructure and hybrid measurement analysis on mesostructure of social networks. In temporal-spatial space, a hybrid probabilistic graph model with multiple constraints can be designed. Taking kinetic energy equation of group evolution and information dissemination into account, we are trying to design a structural kinetic energy equation for common social networks.

3.2 Crowds and their interaction behavior

In the fields of group behavior modeling and behavior feature analysis, existing researches focus on users' preference modeling and their behaviors of social network selection and migration, but ignored how to model users' behavior when a group is growing up. In the fields of group sentiment modeling and interaction analysis, some phenomenon, such as interactive evolution and public sentiment drift, can't be characterized accurately by existing methods so that it is hard to analyze public opinion precisely. So, new theories and new methods are needed to support information-driven public behavior analysis.

In order to fully understand the crowds and their interaction behaviors, some essential questions is open as follows. In social networks, how is the presence or formation of crowds represented? How are the interactions within/among groups measured? Such interactions affect the process of group evolving. How is it calculated? The problem can be formulated as follows:

$$ComInter := \langle P, N, I, \eta, EQ_c \rangle. \quad (3)$$

In Eq. (3), $\eta := \eta(p, n, i)$ denotes the structural factor controlling the network evolving, in which $p \in P$, $n \subset N$, $i \in I$. $EQ_c := \{eq \in P(t) | P(t+1) = (P(t), N, I, \eta)\}$ is the kinetic energy equation that determines the evolution of crowds. The research on crowd analysis can be carried out from the following 3 aspects.

(1) Crowds gathering, modeling and evolving laws. Based on the research on situational awareness and group cognitive approach, taking the group cohesion calculation method into consideration, the social networking groups gathering mechanism and modeling methods can be designed. Moreover, based on the research on mechanisms of group gathering, taking the kinetic energy equation for structure evolution and information dissemination into account, we are trying to give a kinetic energy equation for crowd evolution.
(2) Emotion modeling and evolving laws. Based on the common cognitive space and behavioral rules, combined with machine learning methods, we are going to conduct research on multidimensional emotion modeling methods for social groups. Secondly, based on formation mechanism of group emotion and cost-benefit theory on various social roles, taking the environmental awareness and information dissemination evolution method into account, we are going to study the emotional evolutionary computation methods.
(3) Social influence modeling, evaluation and interaction patterns. Based on analysis on features of network structure and media, taking the background environment and public cognition of audience on specified information content into account, we are going to conduct research on community influence and its effects assessment. Based on above models on community influence and information dissemination, taking the dynamic game theory into account, we are going to find some interesting social networking group interaction patterns.

3.3 Information and its dissemination

In the fields of information diffusion model, computational efficiency and accuracy of existing models can be further improved. In addition, dynamical characteristics, such as temporal, spatial and bidirectional characteristics can be further mined to promote deeper analysis of information diffusion mechanism. Furthermore, the influence among concurrent multi-source information is also a new research direction.

In order to fully understand the information and its dissemination, some essential questions is open as follows.

How to calculate the information content of the form of expression? Information on social networks propagation process is calculated with the situation what? How to use computational methods to characterize the information content and the interaction between information dissemination? The problem can be formulated as follows:

$$InforPro := \langle P, N, I, \varphi, EQ_I \rangle. \quad (4)$$

In Eq. (4), $\varphi := \varphi(p, n, i)$ denotes the structural factor controlling the network evolving, in which $p \in P$, $n \subset N$, $i \in I$. $EQ_I := \{eq \in I(t) | I(t+1) = (I(t), N, P, \varphi)\}$ is the kinetic energy equation that determines the evolution of information propagation. The research on information propagation analysis can be carried out from the following 3 aspects.

(1) Content modeling and representation. Based on the cognitive space and its gradient model, taking the multi-information feature matching methods into consideration, we are exploring the computable representation model of the information content. Moreover, based on the information content representation model, combined with non-equilibrium entropy method, the energy calculation model can be defined with the information quality, activity and entropy.

(2) Dissemination and evolving laws. Based on the various topological structures and timing multi-iterative algorithm, taking dynamics equations on structural and crowds' evolution, we are conducting research on finding the kinetic equations of information dissemination. Then, based on the kinetic equations of information dissemination, combined with cognitive space and its gradient model, we are doing research on evolution models of information dissemination.

(3) Public opinion modeling, traceability and evaluation. Based on the kinetic equations for information dissemination and dynamic game theory, combined with feedback control mechanism, we are conducting public opinion persuasion, information dissemination and its traceability study. Then, we analyze the evolutionary models of information dissemination, take various statistical methods on diffusion range, groups, emotional changes and other aspects into account, and evaluate of the effectiveness of information dissemination model.

4 Conclusion

Recently, online social networking services and micro-blogging service receive a lot of attentions. This has led to a rising prominence of social networks analysis in academia, politics, homeland security and business. Social network analysis has been studied extensively from variable aspects such as degree distribution analysis, social entity ranking, community extraction, pattern discovery and so on.

Social network and information dissemination analysis is an interesting and useful research field. Understanding the information dissemination in social networks requires a set of ideas for reasoning about network structure, strategic behavior, and the feedback effects they produce across large populations.

Our work opens various venues for future work. We introduced 3 important aspects and some open research problems, including the social structure, crowds and information traveling through them. Recently, crowd-sourcing prediction on social networks attracted a lot of attention. Analyzing the information dissemination among multiple social networks is also an interesting research field.

Acknowledgement The work was supported by National Basic Research Program of China (2013CB329601 and 2013CB329606) and the National Natural Science Foundation of China (91124002, 61372191, and 61303190).

References

1. Barnes J (1983) Graph theory in network analysis. Soc Netw 5:235–244
2. Tong H, Papadimitriou S, Yu PS et al (2008) Fast monitoring proximity and centrality on time-evolving bipartite graphs. Stat Anal Data Min 1:142–156
3. Ghoshal G, Zlatic V, Caldarelli G et al (2009) Random hypergraphs and their applications. Phys Rev E 79:066118
4. Kim M, Leskovec J (2012) Multiplicative attribute graph model of real-world networks. Internet Math 8:113–160
5. Pei J, Jiang D, Zhang A (2005) Mining cross-graph quasi-cliques in gene expression and protein interaction data. In: Proceedings of the 21st international conference on data engineering, 2005 Apr 05–08, National Center of Sciences, Tokyo, Japan, pp 353–356
6. Dodds P, Watts D, Sabel C (2003) Information exchange and robustness in organizational networks. Proc Natl Acad Sci USA 100:12516–12521
7. Liben D, Novak J, Kumar R et al (2005) Geographic routing in social networks. Proc Natl Acad Sci USA 102:11623–11628
8. Golder S, Wilkinson D, Huberman B (2007) Rhythms of social interaction: messaging within a massive online network. In: Proceedings of the 3rd communication technology conference (CT2007), 2007 Jun 28–30. Springer, East Lansing, pp 41–66
9. Charu A, Wang H (2010) Managing and mining graph data. Springer-Verlag, New York
10. Liu X, He Q, Tian Y et al (2012) Event-based social networks: linking the online and offline social worlds. In: Proceedings of the 18th international conference on knowledge discovery and data mining, 2012 Aug 12–16. ACM, Beijing, pp 1032–1040
11. Clauset A, Newman M, Moore C (2004) Finding community structure in very large networks. Phys Rev Lett 70:066–111

12. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. Proc Natl Acad Sci USA 104:36–41
13. Palla G, Dernyi I, Farkas I (2005) Uncovering the overlapping community structure of complex network in nature and society. Nature 435:814–818
14. Lancichinetti A, Fortunato S, Kertesz J (2009) Detecting the overlapping and hierarchical community structure in complex networks. New J Phys 11:033015
15. Shen HW, Cheng XQ, Cai K et al (2009) Detect overlapping and hierarchical community structure in networks. Phys A 388:1706–1712
16. Chen TD, Sanghavi S, Xu H (2012) Clustering sparse graphs. NIPS 25:2213–2221
17. Lin WQ, Kong XN, Yu PS et al (2012) Community detection in incomplete information networks. In: Proceedings of the 21st international conference on world wide web, 2012 Apr 16–20. ACM, Lyon, pp 341–350
18. Leskovec J, Kleinberg J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining, 2005 Aug 21–24, Chicago, IL, USA, pp 177–187
19. Chakrabarti D, Kumar R, Tomkins A (2006) Evolutionary clustering. In: Proceedings of the 12th international conference on knowledge discovery and data mining, 2006 Aug 20–23, Philadelphia, PA, USA, pp 554–560
20. Chi Y, Song X, Zhou D et al (2007) Evolutionary spectral clustering by incorporating temporal smoothness. In: Proceedings of the 13th international conference on knowledge discovery and data mining, 2007 Aug 24–27, San Jose, CA, USA, pp 153–162
21. Allamanis M, Scellato S, Mascolo C (2012) Evolution of a location-based online social network: analysis and models. In: Proceedings of the 2012 conference on internet measurement, 2012 Nov 14–16, Boston, MA, USA, pp 145–158
22. Tantipathananandh C, Berger-Wolf T, Kempe D (2007) A framework for community identification in dynamic social networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, San Jose, CA, USA, pp 717–726
23. Bródka P, Saganowski S, Kazienko P (2013) GED: the method for group evolution discovery in social networks. Soc Netw Anal Min 3:1–14
24. Chen YD, Kawadia V, Urgaonkar R (2013) Detecting overlapping temporal community structure in time-evolving networks. CoRR 3:1–7
25. Leskovec J, Backstrom L, Kumar R et al (2008) Microscopic evolution of social networks. In: Proceedings of the 14th international conference on knowledge discovery and data mining, 2008 Aug 24–27, Las Vegas, NV, USA, pp 462–470
26. Ellison N, Steinfield C, Lampe C (2007) The benefits of facebook "friends": social capital and college students' use of online social network sites. J Comput Med Commun 12:1143–1168
27. Steinfield C, Ellison N, Lampe C (2008) Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. J Appl Dev Psychol 29:434–445
28. Kang YS, Hong S, Lee H (2009) Exploring continued online service usage behavior: the roles of self-image congruity and regret. Comput Hum Behav 25:111–122
29. Chen ZY, Yang YP, John L (2009) Cyber migration: an empirical investigation on factors that affect users' switch intentions in social networking sites. In: Proceedings of the 42nd Hawaii international conference on system sciences, 2009 Jan 5–8, Big Island, HI, USA, pp 1–11
30. Abbasi M, Chai S, Liu H (2012) Real-world behavior analysis through a social media lens. In: Yang SJ, Greenberg AM, Endsley M (eds) Social computing, behavioral—cultural modeling and prediction. Lecture notes in computer science. Springer, Berlin, pp 18–26
31. Schramm W (1960) Mass communications, 2nd edn. University of Illinois Press, Urbana
32. Eirinaki M, Vazirgiannis M (2003) Web mining for web personalization. ACM Trans Internet Technol 3:1–27
33. Jiang B, Pei J, Lin X et al (2008) Mining preferences from superior and inferior examples. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, 2008 Aug 24–27, Las Vegas, NV, USA, pp 390–398
34. Wong RC, Pei J, Fu A et al (2009) Online skyline analysis with dynamic preferences on nominal attributes. IEEE Trans Knowl Data Eng 21:35–49
35. Camerer C (2009) Behavioral game theory: experiments in strategic interaction. Princeton University Press, Princeton
36. Bollen J, Gonalves B, Ruan G et al (2011) Happiness is assortative in online social networks. Artif Life 17:237–251
37. Xu K, Li J, Liao S (2011) Sentiment community detection in social networks. In: Proceedings of the 2011 conference, 2011 Mar 4–7, Berlin, Germany, pp 804–805
38. Gryc W, Moilanen K. Leveraging textual sentiment analysis with social network modeling: sentiment analysis of political blogs in the 2008 U.S. presidential election. In: Proceedings of the From Text to Political Positions: Text analysis across disciplines, 2010, vol 11, pp 47–70
39. Aoki S, Uchida O (2011) A method for automatically generating the emotional vectors of emoticons using blog articles. Int J Comput 5:346–353
40. Nguyen LT, Wu P, Chan W et al (2012) Predicting collective sentiment dynamics from time-series social media. In: Proceedings of the first international workshop on issues of sentiment discovery and opinion mining, 2012 Aug 12–16, Beijing, China, pp 6–15
41. Chen YF (2008) Herd behavior in purchasing books online. Comput Hum Behav 24:1977–1992
42. Yang KS, Zhu KY, Fu C et al (2009) Simulation of group decision conformity based on cellular automata model. Syst Eng 9:115–124
43. Al-Sharawneh J, Williams M (2010) Credibility-based social network recommendation: follow the leader. In: Proceedings of Australia conference on information system, 2010 Dec 1–3, Brisbane, Australia, paper 24
44. Pei Y, Wu G (2012) Research on emergency crowd behavior based on epidemic dynamics. Intell Decis Technol 2:99–106
45. Lewis K, Gonzalez M, Kaufman J (2012) Social selection and peer influence in an online social network. Proc Natl Acad Sci USA 109:68–72
46. Cassirer E (1923) Philosophie der symbolischen Formen. Wissenschaftliche Buchgesellschaft, German (in German)
47. Shannon CE (1948) A mathematical theory of communication. Bell Syst Technol J 27:379–423
48. Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 4:620–630
49. Danchin E, Giraldeau L, Valone TJ et al (2004) Public information: from nosy neighbors to cultural evolution. Science 5683:487–491
50. Xing XS (2004) Non-equilibrium statistical information theory. Acta Phys Sin 9:2852–2863 (in Chinese)
51. Lin CX, Mei Q, Jiang Y et al (2010) Inferring the diffusion and evolution of topics in social communities. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, 2010 Jul 25–28, Washington, DC, USA, pp 1019–1028
52. Jo Y, Hopcroft JE, Lagoze C (2011) The web of topics: discovering the topology of topic evolution in a corpus. Proceedings of the 20th international conference on world wide web, 2011 Mar 28–Apr 1, Hyderabad, India, pp 257–266

53. Kumar S, Zafarani R, Liu H (2011) Understanding user migration patterns in social media. In: Proceedings of the 25th AAAI conference on artificial intelligence, 2011 Aug 7–11, San Francisco, CA, USA, pp 1–6
54. Hong LJ (2011) Predicting popular messages in Twitter. In: Proceedings of the 20th international conference on world wide web, 2011 Mar 28–Apr 1, Hyderabad, India, pp 57–58
55. Romero DM (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th international conference on world wide web, 2011 Mar 28–Apr 1, Hyderabad, India, pp 695–704
56. Hethcote H (2000) The mathematics of infectious diseases. SIAM Rev 42:599–653
57. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. In: Proceedings of the 9th international conference on knowledge discovery and data mining, 2003 Aug 24–27, Washington, DC, USA, pp 137–146
58. Gruhl D, Guha R, Liben-Nowell D et al (2004) Information diffusion through blogspace. In: Proceedings of the 13th international conference on world wide web. ACM, pp 491–501
59. Zhou T, Fu ZQ, Niu YW et al (2010) Survey on transmission dynamics in complex networks. J Syst Eng 6:742–754 (in Chinese)
60. Kumar R, Novak J, Tomkins A (2006) Structure and evolution of online social networks. In: Proceedings of the 12th international conference on knowledge discovery and data mining, 2006 Aug 20–23, Philadelphia, PA, USA, pp 611–617
61. Doucet A, Freitas N, Murphy K (2000) Rao-blackwellished particle filtering for dynamic bayesian networks. In: Proceedings of the 16th conference on uncertainty in artificial intelligence, 2000 Jun 30–Jul 3, Stanford, CA, USA, pp 176–183
62. Kask K, Dechter R, Larrosal J et al (2001) Buchet-tree elimination for automated reasoning. Artif Intell 125:91–131
63. Murphy K, Weiss Y (2001) The factored frontier algorithm for approximate inference in DBNs. Technical report, Computer Science Department, University of California, Berkeley, pp 176–183
64. Paskin M (2003) Thin junction tree filters for simultaneous localization and mapping. Technical report, Electronics and Computer Science, University of California, Berkeley
65. Lasswell HD (1948) The structure and function of communication in society. Commun Ideas 37:215–228
66. McCandless LC, Gustafson P, Levy A (2006) Bayesian sensitivity analysis for unmeasured confounding in observational studies. Stat Med 26:2331–2347
67. Wang HQ (2004) Building Bayesian networks: elicitation, evaluation, and learning. Doctor dissertation, University of Pittsburgh, Pittsburgh
68. Renooij S (2012) Efficient sensitivity analysis in hidden markov models. Int J Approx Reason 53:1397–1414
69. Bakshy E, Hofman J, Mason W et al (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining, 2011 Feb 9–12, Hong Kong, China, pp 65–74
70. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. Science 6092:337–341
71. Shenoy P (1992) Valuation-based systems for bayesian decision analysis. Oper Res 40:63–84
72. Madsen A (1999) Lazy propagation. A junction tree inference algorithm based on lazy evaluation. Artif Intell 113:203–245
73. Maes S, Philippe L (2006) Multi-agent causal models for dependability analysis. In: Proceedings of the 1st international conference on availability, reliability and security, 2006 Apr 20–22, Vienna, Austria, pp 794–798
74. Tichenor J, George A, Clarice O (1970) Mass media flow and differential growth in knowledge. Public Opin Q 34:159–170
75. Neumann E (1974) The spiral of silence a theory of public opinion. J Commun 24:43–51
76. McCombs M, Shaw D (1972) The agenda-setting function of mass media. Public Opin Q 36:176–187