J Supercomput (2018) 74:3654–3665
https://doi.org/10.1007/s11227-016-1858-9

Cognitive social network analysis for supporting the reliable decision-making process

Mucheol Kim[1] · **Sangyong Han**[2]

Published online: 9 September 2016

Abstract With the advancement of the Internet and web technologies, social networks have gained attention as a new paradigm for user-centered information systems. As the amount of accessible information increases, the need for personalized information increases as well. Under such circumstances, social networks that are based on trust between users are increasingly utilized to provide efficient and reliable information management. This paper proposes the cognitive social network analysis that analyzes relationships between users with typical properties. The proposed approach analyzes users' habitual activities and creates a local social network. The framework then integrates the local networks via the friend of a friend, thereby creating a global social network. User relationships in the global network are reinforced to maximize information sharing. To evaluate the performance of the information shared in the proposed autonomic cognitive social network framework, the accuracy of the information associated with social network was measured using the ROC Curve. In future, we should analyze the social influence factors from relationship between community and users.

Keywords Sentiment analysis · Social network analysis · Decision support system · Web technology

✉ Sangyong Han
hansy@cau.ac.kr

Mucheol Kim
mucheol.kim@gmail.com

1 Division of Media Software, Sungkyul University, Anyang, Korea

2 School of Computer Science and Engineering, Chung-Ang University, Seoul, Korea

1 Introduction

The rapid development of web technology in the 20th century, a lot of people had a chance to communicate with each other through the web [1–3]. As a result, web data has dramatically increased, and people's interest in an open platform such as an online social network has risen [4–6]. In general, an online social network refers to a human network which is being formed through information sharing, exchange and trading through the Internet. While people form limited social relationships such as friends, family and colleagues in a real world, a wide variety of open social relationships can be established in the online environment [7]. Users would display their propensity or express their opinions on various experiences through social networking services such as blog, twitter and facebook. Therefore, these social media and social activities can be understood as a brand-new means of collaboration and communication [8,9].

In traditional inter-user relations, psychological and social trustworthiness was perceived qualitatively based on past experiences, and then future behavior was predictable. In the online social network environment, on the contrary, because of diverse relationships which are formed from various sources, it's been hard to recognize relationships based on qualitative background knowledge. To develop a reliable relationship model in the online social network environment, therefore, it is important to abstractize social and psychological information into electronic and qualitative information [10]. For the qualification of social relationship in the online social network environment, recognizable data from diverse sources should be used. Specifically, they can be classified into the followings: behavioral interaction information, personal preferences and information and environmental activity information. First, the behavioral interaction information means a direct or indirect interaction with other users in the online social network environment. This information includes an act of adding a counterpart or exchanging opinion on twitter, coauthoring a paper through joint research and evaluation of the same contents in the film evaluation system. In contrast, personal information refers to attributes in which users stipulate their own scope. In personal information, relationships are formed in assumption that those with the same attributes may have a similar propensity [11]. With the personal information, users can form a group with the same attributes and group with static characteristics. In addition, these groups differ in terms of affinity levels depending on the members' characteristics [12–14]. Hence, environmental information represents that users' environmental factors influence their social activities, which means that users' diverse social activity records reflect users' environmental background. Unlike personal information, this environmental information can form dynamic groups [15].

In this paper, a model which can analyze users' preferences and support decision-making on the results is proposed. A trust and reputation model is utilized for the quantification and abstractization of relationship between users. In the process, data from diverse sources are analyzed, and inter-user relationships are recognized. Depending on information characteristics, meanwhile, the information is classified into personal determinant, behavior determinant and environmental determinant. Through an interaction among the three determinants, inter-user relationship is expressed with trust values. The trust values obtained through the user-centered network can extensible using the FOAF relationship. As a result, user knowledge is extended, and the knowledge acquired from the extended social network can be used in decision-making. Therefore, users can be freed from most irrelevant data, reducing a risk in information acquisition.

2 Related work

In this chapter, the bases of the proposed paper, which are social network, trust-based system and decision-making support system are reviewed.

For the effective management of an information system, many researchers in the social network environment have been interested in the trust-based quantification and abstractization of inter-user relationship. In particular, it is difficult to apply inter-user trust factors which were naturally formed in the offline environment to the online social network as they are. Therefore, it is critical to convert social and psychological information into digital and quantitative information to measure interpersonal trust factors. Resnick and Zeckhauser [16] proposed a method to improve the performances of the recommendation system using the feedback in the calculation of system reliability in the online trading system. They proposed a model with which reputation can be obtained through both positive and negative responses of buyers and sellers, which can be obtained through trade and revealed an aspect of specializing someone's unclear common evaluations on a particular target. They also made the proposed algorithm applicable to eBay's reputation system. O'Donovan and Smyth [17] derived inter-user trustworthiness with the similarity level of attributes in the user profile. In addition, [11] proposed a reliability inference model in the social network environment, which is applicable to the recommendation system. To derive trust values with the indirectly connected user, in this paper, connected users' assessment values were combined. In addition, an algorithm which calculates reputation, an objective user assessment factor, was suggested by combining the trust values of searched multiple paths. To manage the relationship model of an online social network with dynamic preference, data mining technique-based repair and maintenance methods were proposed [18,19]. Furthermore, it was attempted to improve a relationship model's reliability through studies which predict a new inter-user link which is slated to be formed in the future based on the conventional relationship model [20].

This kind of trust management system has been applied to diverse applications. Yoo et al. [21] used the social trust management system in identifying spam mails to filter unwanted information. While conventional spam filtering systems were designed based on contents only, this system in this paper increased accuracy in the identification of spam mails by analyzing an inter-user relationship model. In [22–25], moreover, a social trust management method was applied to the recommendation system to acquire information which meets user preferences.

3 Cognitive social networks for sentiment analysis

In this chapter, autonomous cognitive social networks are discussed Fig. 1. The cognitive social networks derived from sociology are divided into behavioral determinant, personal determinant and environmental determinant. To improve the reliability of

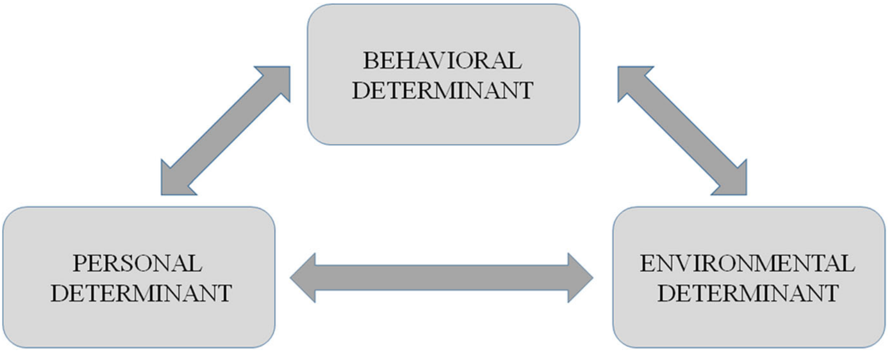

Fig. 1 Cognitive social networks with three determinant elements

inter-user relationship, this paper quantified and abstractized cognitive social networks.

3.1 Behavioral determinant for cognitive social networks

In this section, various types of user behavioral interaction patterns are analyzed and abstractized and quantified in trust values. In terms of the behavioral interaction patterns, the definition of trust relationship based on the results of the rating evaluation is handled.

The behavioral interaction patterns which become the backbone of social networks are extracted from these diverse data, and the results are used in network construction. In a real world, users' opinions contain an entity's sentiment. Emotional expressions change depending on target, circumstance and environment. To quantify the emotions, therefore, they should be expressed in the form of vector which reveals a direction, not a scalar format as shown in Fig. 2. In the proposed method as well, the behavioral

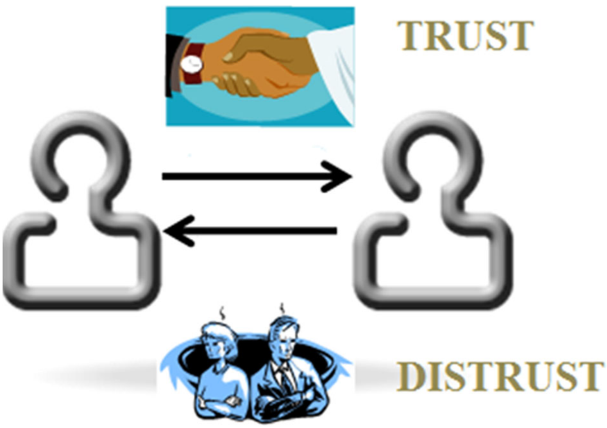

Fig. 2 Social relationship between individuals with positive and negative trust

Table 1 User sentiment analysis with user rating result

Rating	Meaning	User sentiment
1	Horrible	Negative sentiment
2	Bad	
3	Not bad	Neutral sentiment
4	Interesting	Positive sentiment
5	Fantastic	

information of the rating evaluation system is analyzed and expressed in the form of vector as well. (Behavioral determinants belong to the interaction with unilateral implicit attributes.)

In general rating evaluation systems, users express their emotions on the contents through rating evaluation. In this study, the evaluation results on the contents are recognized in five stages as shown in Table 1. The user sentiment on them is classified into three categories: positive, negative and neutral.

In positive rating, contents are matched with user preferences. Therefore, positive emotions are expressed. In negative rating, contents are against user preferences so that negative emotions are expressed. In neutral rating, user preferences and contents differ so that emotional values are vague. Therefore, the inter-user relationship based on similarity analysis on emotional preferences is expressed as the results of the rating on the same contents and sentiment preferences. In other words, whether or not the sentiment and opinion expressions on the same contents are homogenous becomes a factor which expresses an inter-user relationship. In this study, the inter-user relationship (i.e., the level of sentiment and opinion matching) is quantified and abstractized using trust values. The general trust values are defined as follows: [10].

Definition 1 Trust is defined as the firm belief in the competence of an entity to act dependably, reliably and securely within a specific context.

In this study, behavioral trust ($T_{\text{Sentiment}}$) is recognized as the similarity level of the sentiment which is being expressed by users through diverse contents. Then, rating on the contents is simplified into whether or not user opinions and emotional preferences are matched. In this study, inter-user trustworthiness is derived through the quantity of the matched opinions on the same contents ($s_{A,B}$) while emotional preferences are obtained through the matched quantity ($n_{A,B}$) (Eq. 1). In the social network designed based on these calculated trust values, an emotional relationship model is diagramed, and information can be shared with those with similar preferences.

$$T_{BD}(A, B) = \frac{2s_{A,B} + n_{A,B}}{2R_A} \times 10 \tag{1}$$

$T_{A,B}$ can be obtained by combining the quantity of the contents in which the opinions on reliability between User 'A' and User 'B' are matched and quantity of the contents in which similar opinions are expressed. The components have the following meaning:

R_A: Quantity of the contents evaluated by User 'A'.

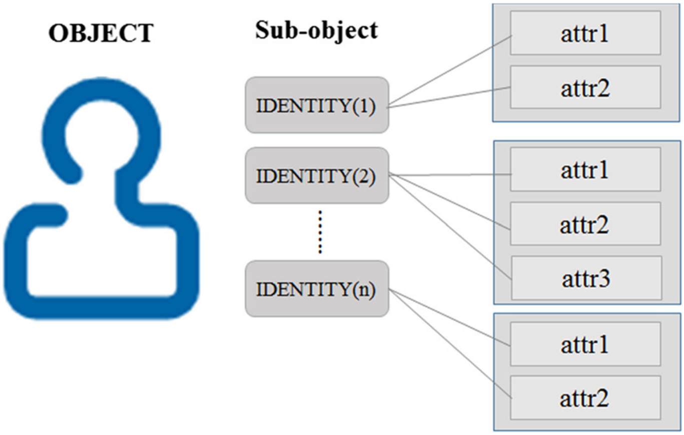

Fig. 3 Hierarchical structure of personal information in cognitive social networks

$s_{A,B}$: Quantity of the contents with the same rating by User 'A' and User 'B'.

$n_{A,B}$: Number of cases in which positive or negative emotions are expressed even through the same contents of User 'A' and User 'B' are differently rated.

3.2 Personal determinant for cognitive social networks

In social networks, the scope and characteristics of each user or user groups are decided by their attributes [26]. In other words, they can be perceived as a concept of granular computing just like the hierarchical structure of social networks (Fig. 3). In other words, the objects of social networks have sub-objects which reveal their preferences, and the sub-objects also show each object's preferences. Therefore, the inter-user relationship can be expressed as a group with a small unit such as class and cluster depending on their sub-object and attributes. The static organizational identity which can be extracted from user profile, such as occupation, age, nationality and blood type becomes the standard which reveals personal preferences. To recognize the preference of the group of users in similar conditions, it is needed to analyze relationship among them, in which personal information is matched. In this section, therefore, users' personal information is collected in smartphone environment, and its relationship is analyzed to calculate cohesion among the static organizations.

Based on the calculated cohesion, the affinity of the user group is selected. In other words, the influence depending on whether or not personal information attributes are shared is calculated and applicable to the inter-user relationship. Then, the influence is called, 'affinity on the attribute'. Ankolekar et al. [27] defined it as follows:

Definition 2 Affinity represents the degree of homogeneity in the network, with respect to a particular attribute.

High affinity means that the groups are coherent or highly correlated. Therefore, affinity can be calculated through analysis on how much user attributes are linked or shared.

$$\text{User} = \{a_1, a_2, \ldots, a_p\} \tag{2}$$

Users have information on attributes which express different domains, such as occupation, gender, age and address, which is a set of identifiable attributes that classify users or groups. If assumed that there are 'p' attributes, each user can be expressed as a set of attributes as shown Eq. 11. To evaluate intimacy depending on how much the attributes are matched, then the inter-user links are calculated depending on the matched attributes. In addition, the clustering coefficient is calculated and reflected based on the number of users who share the attributes. The affinity by the attribute group can be calculated as follows:

$$S_a = \frac{|\left\{\left(\text{User}_i, \text{User}_j\right) \in E : \text{ s.t. User}_i.a_p = \text{User}_j a_p|}{|E|} \tag{3}$$

First of all, to calculate inter-user affinity, the relevance of the users who share these attributes (i.e., link relationship) is analyzed. In Eq. 3, the number of links owned by the users with the same attribute is calculated, which is called 'fraction of links'. The calculations refer to users' index of empathy, which forms a group of the same attributes. Here, E is a set of all links, and a_p refers to the pth attribute value.

$$N(e) = \frac{e(e-1)}{2} \tag{4}$$

If assumed that nodes are randomly distributed, the number of all creatable links ($N(e)$) on the number of nodes (e) with the same attribute can be calculated as shown in Eq. 4. To calculate the influence of the users with each attribute, we calculated the clustering coefficient of a group of the users who share the attribute.

$$E_a = \frac{\sum_{i=0}^{k} N(T_i)}{N(|U|)} \tag{5}$$

As shown in Eq. 5, the clustering coefficient on each attribute is calculated as the number of links, which can be created by the user group who shares each attribute to be prepared against the creatable links for all nodes. Here, U refers to the sum of all users who share the attribute ($U = \sum_{i=0}^{k} T_i$).

$$\text{Affinity}_a = \frac{S_a}{E_a} \tag{6}$$

The affinity on each attribute is calculated using the fraction of links (Eq. 6) on the number of the users with the same attributes. Then, high affinity means that the links in the group with similar attributes are closely related with user attributes.

$$Q_a(A, B) = \begin{cases} 1 & \text{if } A.a_i \text{ and } B.a_i \text{ are same} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Even though the number of user groups is the same, the importance differs depending on their social influence. To find out the static group-based trustworthiness between User 'A' and User 'B', a group which is created by the shared attributes should be recognized using Eq. 7.

$$T_{\text{Affinity}}(A, B) = \sum_{i=1}^{n} \text{Affinity}_{\text{Attr}_i} \cdot Q_{\text{Attr}_i}(A, B) \tag{8}$$

In conclusion, inter-user trust values are calculated through the number of common groups between users and affinity of the attribute (Eq. 8). In other words, after adding diversity and the weighted values of the group to user preferences, the social influence of the static group can be applied to a trust model as sentiment information.

$$T_{\text{PD}}(A, B) = T_{\text{Sentiment}}(A, B) \times (1 + T_{\text{Affinity}}(A, B)) \tag{9}$$

The trust value between User 'A' and User 'B' can be calculated by combining behavioral pattern-based trust value and affinity-based trust value (Eq. 9). Then inter-user trust relationship is formed based on behavioral information by reflecting the influence of the user group in weighted values.

3.3 Environmental determinant for cognitive social networks

In this Chapter, we confirmed the identity of the user's identity to recognize the tendency of the user. We extracted a user's activity in communities or groups and generated the clusters. It means that each cluster have similar users who have similar identities or tendencies. In other words, it means the separator can be distinguished from others [28]. It is suitable for clustering, it is unsupervised learning technique, because the dynamic nature of the environmental determinants and unspecified attribute [29,30].

In this paper, we extract the set of identities based on the activities of the users. Further, they might form the identity groups using K-means clustering methodology [31]. We assume that each identity group is consisted on the set of terms they represent distinguish characteristics [32]. Then, we analyzed the relationships between users with cosine similarity [33].

4 Case study and performance analysis

Figure 4 displays the overview of the entire process for selecting a movie with the directional social network analysis based recommendations. The proposed processes are separated to 3-phases. First process is for constructing a social network with trust and distrust with users' feedbacks. After watching a movie, users might evaluate the movie with five phased rate. After that we analyzed the relationship with trust and distrust with their evaluations and preferences. We constructed social networks with the extracted relationship between users.

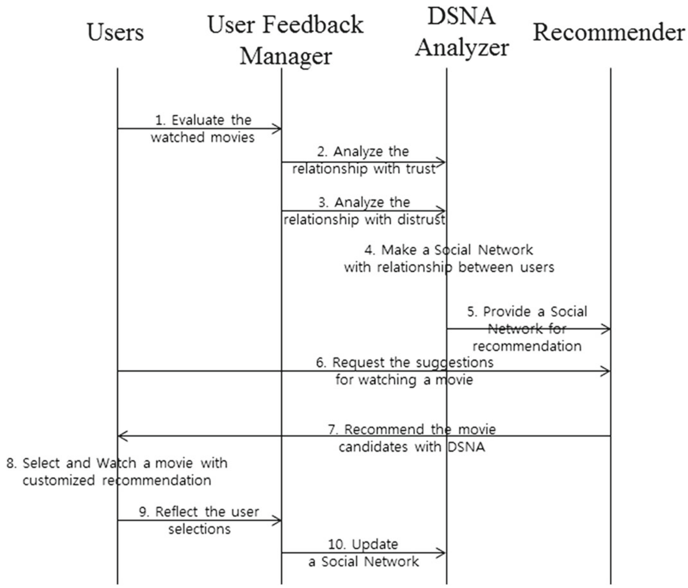

Fig. 4 Case study for customized recommendation with directional social network analysis

Second process recommends the movie candidates with directional social network analysis (DSNA). Last one is the process for collecting the users' selections against the recommendations and updating social networks. It could realize the ecosystem for our analysis and recommendation process.

To evaluate the performance of the proposed method, we examined if information suitable to user preferences is provided using the receiver operating characteristic (ROC) curve. The ROC curve a graph which represents true positive rate (TPR) against false positive rate (FPR). In general, the ROC Curve shows the sensitivity and specificity of test results so that it is a useful tool in evaluating performances in pattern recognition or machine learning [34]. The TPR decides if the results obtained from the proposed system meet user preference or intention (i.e., the sensitivity of the result). In other words, 100 % probability means that all results meet user intention.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \quad (\mathrm{TP{:}\ True\ Positive,\ FN{:}\ False\ negative}) \tag{10}$$

On the contrary, the FPR refers to the specificity of the information obtained from the system (Eq. 10). In other words, it is an evaluation factor on the cases in which the opposite results from the user intention occur. In this paper, the specificity and sensitivity of the results obtained from the proposed system were evaluated using the ROC Curve:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \quad (\mathrm{FP{:}\ False\ positive,\ TN{:}\ True\ negative}) \tag{11}$$

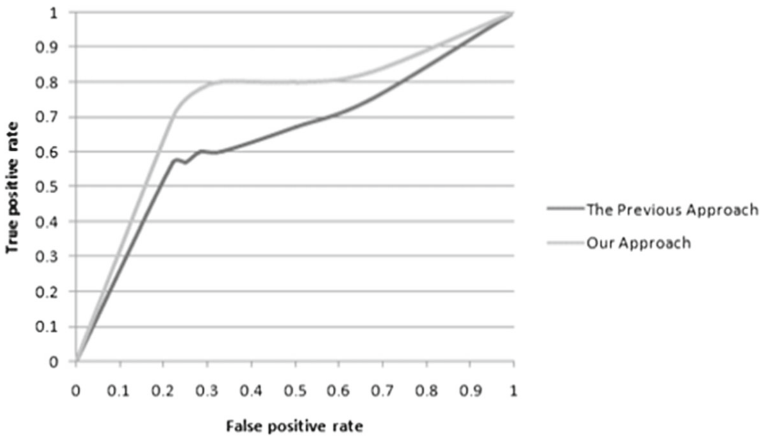

Fig. 5 ROC curve-based performance evaluation in Movielens data set

Using the Movielens data set [35], a social network was built and analyzed. In this paper, the data which were used by 1000 users in evaluating a total of 3952 films were inputted in the construction of an analysis social network. A total of 1000 users upload the results of the rating evaluation on 155 films in average. In addition, they include user profile such as gender, age and occupation. We were extracting two different types of data:

1. Trust value of the inter-user relationship, which is the behavioral determinant that can be derived with the results of the rating evaluation.
2. Trust value of the affinity-based inter-user relationship which is the personal determinant that can be inferred from the user profile.

The first information represents the level of emotional similarity on the same contents among the users. The second information shows affinity with users in the similar group.

In this paper, the sensitivity and specificity of the results obtained using the Movielens data set-based technique and conventional social network-based filtering technique were tested. As done in previous tests, the ROC Curve calculated FPR and TPR by changing the Top-K of the results derived through the proposed social network from 1 to 20 (Fig. 5). As concluded in this evaluation, the proposed method is higher than the conventional method in terms of both sensitivity and specificity, which was also confirmed with AUC values. In terms of AUC value, the proposed method (0.8469) increased by 11 % from the conventional filtering method (0.7619).

5 Conclusions and future work

With web development, a social network has been in the spotlight as a new information system paradigm the so called, ‘open platform’ and ‘user participative system’. In particular, open social relationship has been formed among people with diverse preferences through social networking services such as twitter and facebook. As users face a deluge of information from typical resources, reliable and valuable information recommendation grows in importance. Then we proposed a cognitive social network analysis approach which can explicitly reveal user relationship and related decision-making support methodology after investigating social information from diverse sources and analyzing user preferences and sentiment. In the proposed method, the trust relationship is divided into behavioral interaction information-based determinant, personal information-based determinant and environmental determinant. In future, we should analyze the social influence factors from relationship between community and users. For these analysis we need to recognize the interactions between users in the community and its tendency. As a result, we could realize the enhanced cognitive social networks and suggest the knowledge for the reliable decision-making.

Acknowledgements This work was supported by Institute for Information & communications Technology Promotion (IITP) Grant funded by the Korea Government(MSIP) (No. R0126-15-1007, Curation commerce based global open market system development for personal happiness enhancement).

References

1. Yager RR (2008) Granular computing for intelligent social network modeling and cooperative decisions. In: Presented at the International IEEE Conference Intelligent Systems
2. Adar E, Ré C (2007) Managing uncertainty in social networks. IEEE Data Eng Bull 30(2):15–22
3. Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323(5916):892–895
4. De Rijke M, Weerkamp W (2008) ECIR 2008 tutorials: search and discovery in user-generated text content. In: Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval, pp 714–715
5. Langville AN, Meyer CD (2011) Google's PageRank and beyond: the science of search engine rankings. Princeton University Press, Princeton
6. Staab S, Domingos P, Mike P, Golbeck J, Ding L, Finin T, Josh A, Nowak A, Vallacher RR (2005) Social networks applied. IEEE Intell Syst 20(1):80–93
7. Breslin J, Decker S (2007) The future of social networks on the internet: the need for semantics. IEEE Internet Comput 11(6):86–90
8. Cheung CM, Lee MK (2009) Understanding the sustainability of a virtual community: model development and empirical test. J Inf Sci 35(3):279–298
9. Xu Z, Tresp V, Rettinger A, Kersting K (2010) Social network mining with nonparametric relational models. In: Advances in social network mining and analysis, pp 77–96
10. Grandison T, Sloman M (2000) A survey of trust in internet applications. IEEE Commun Surv Tutor 3(4):2–16
11. Golbeck J (2009) Trust and nuanced profile similarity in online social networks. ACM Trans Web (TWEB) 3(4):12
12. Borcea-Pfitzmann K, Hansen M, Liesebach K, Pfitzmann A, Steinbrecher S (2006) What user-controlled identity management should learn from communities. Inf Secur Tech Rep 11(3):119–128
13. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 44–54
14. Lampinen A, Tamminen S, Oulasvirta A (2009) All my people right here, right now: management of group co-presence on a social networking site. In: Proceedings of the ACM 2009 International Conference on Supporting Group Work, pp 281–290
15. Toahchoodee M, Abdunabi R, Ray I, Ray I (2009) A trust-based access control model for pervasive computing applications. In: Data and applications security XXIII, pp 307–314
16. Resnick P, Zeckhauser R (2002) Trust among strangers in Internet transactions: empirical analysis of eBay's reputation system. Adv Appl Microecon 11:127–157

17. O'Donovan J, Smyth B (2005) Trust in recommender systems. In: Proceedings of the 10th International Conference on Intelligent User Interfaces, pp 167–174
18. Bae J, Kim S (2009) A global social graph as a hybrid hypergraph. In: NCM'09. Fifth International Joint Conference, pp 1025–1031
19. Bourqui R, Gilbert F, Simonetto P, Zaidi F, Sharan U, Jourdan F (2009) Detecting structural changes and command hierarchies in dynamic social networks. In: International Conference on Advances in Social Network Analysis and Mining, pp 83–88
20. Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031
21. Yoo S, Yang Y, Lin F, Moon IC (2009) Mining social networks for personalized email prioritization. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 967–976
22. Debnath S, Ganguly N, Mitra P (2008) Feature weighting in content based recommendation system using social network analysis. In: Proceedings of the 17th International Conference on World Wide Web, pp 1041–1042
23. Kim M, Seo J, Noh S, Han S (2012) Identity management-based social trust model for mediating information sharing and privacy enhancement. Secur Commun Netw 5(8):887–897
24. Tanis M, Postmes T (2005) A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. Eur J Soc Psychol 35(3):413–424
25. Arthur D, Motwani R, Sharma A, Xu Y (2009) Pricing strategies for viral marketing on social networks. In: Internet and network economics, pp 101–112
26. Kim M, Park SO (2013) Group affinity based social trust model for an intelligent movie recommender system. Multimed Tools Appl 64(2):505–516
27. Ankolekar A, Szabo G, Luon Y, Huberman BA, Wilkinson D, Wu F (2009) Friendlee: a mobile application for your social life. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services
28. Mislove A, Viswanath B, Gummadi KP, Druschel P (2010) You are who you know: inferring user profiles in online social networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp 251–260
29. Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput J 41(8):578–588
30. Li Y, Chung SM, Holt JD (2008) Text document clustering based on frequent word meaning sequences. Data Knowl Eng 64(1):381–404
31. Nigam K, McCallum AK, Thrun S, Mitchell T (2000) Text classification from labeled and unlabeled documents using EM. Mach Learn 39(2–3):103–134
32. Newman ME (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025102
33. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv (CSUR) 31(3):264–323
34. Seo Y (2010) Advanced access control mechanism based on trust and risk for collaboration networks and online social networks. Ph. D thesis, Chung-Ang University
35. MovieLens Data Set. http://www.cs.umn.edu/Research/GroupLens