INFOCURATION SYSTEM CAN SAVE YOUR TIME IN SOCIAL NETWORKS

Research-in-Progress

Jihae Suh
College of Business Administration
Seoul National University, Seoul, Korea
jihaesuh77@gmail.com

Youngseok Choi
College of Business Administration
Seoul National University, Seoul, Korea
aquinas9@snu.ac.kr

Jinsoo Park
Graduate School of Business & College of Business Administration
Seoul National University, Seoul, Korea
jinsoo@snu.ac.kr

Abstract

The wealth of information in modern society is one of its most revolutionary features, and this vastness make not only difficult to acquire high quality data, but access itself since the data is scattered around. In other words, numerous structural holes exist in the network of weak tie. In this paper, we present an algorithm to develop inforcuration system that enables to recommend the user with similar preference to other users in the sub-network and to fill the structural hole between sub-networks by connecting opinion leader based on the topic preference analysis. An Infocuration system is a high-level information system that works closely with network users to assist them in identifying helpful information that solves their problems and reduces the obstacles for establishing or maintaining positive interpersonal relationships in the social network.

Keywords: Infocuration, social network, structural hole, topic preference analysis

Introduction

The wealth of information in modern society is one of its most revolutionary features, and this trend will continue alongside the rapid advancement of information technology. However, large quantities of information cannot ensure that the information is of a high quality. The plethora of information in society is a double-edged sword. High-quality information benefits individuals and organizations, whereas poor-quality information damages individuals and organizations. What is information quality? "Fitness for use" is a widely adopted term, which refers to information quality issues. The phrase emphasizes the significance of the user's perspective of information quality because product or service usage will ultimately determine whether it is fit for use. Additionally, "data quality" describes data that is fit for use by data consumers (Wang and Strong 1996). However, just having access to high-quality information is not sufficient. Users must discover or create effective knowledge from high-quality information, and this knowledge should enable individual users and organizations to resolve their own problems.

Thus far, many approaches have been employed to discover valuable knowledge that meets the needs of users and solves problems. For example, with the increase in the number of cellular phone users, micro-blogging is currently undergoing rapid growth and gaining considerable popularity. Twitter has been able to extend its role quickly as a micro-blogger because it enables people to connect easily, provides real time information, and gives people the option to re-tweet. Over time, the purpose of using Twitter changes, indicating that users employ Twitter to resolve their individual problems and obtain high-quality information rather than simply spreading rumors, reading news, or keeping in touch with friends. However, there is uncertainty regarding whether information-seekers benefit and gain satisfaction from Twitter. As a tool to acquire knowledge, Twitter still has some problems. First, Twitter users share enormous quantities of data; however, it is uncertain whether the information is appropriate. Second, even if the information is fit for the user's purpose, obtaining any useful knowledge from it is difficult. Such information exists in a spare data form, not a connected form. To solve such problems and improve user satisfaction, the role of Twitter should change from building local relationships to developing user-generated collective intelligence. Therefore, a method for finding valuable information and connecting qualified information is required.

Networks rich in structural diversity provide "information benefits" or "vision advantages" that increase performance by presenting diverse perspectives and information (Burt 1992). However, in real situations, high-quality data contains structural holes, which is an absence of ties between two parts of a network, and this kind of data makes it difficult to solve problems and discover knowledge. Therefore, identifying and exploiting a structural hole provides a competitive advantage to users. In other words, sparse networks rich in structural holes, such as the Twitter network, provide opportunities to access a wide variety of content and encourage users to develop both their own individual networks and to join other cluster networks.

The primary function of a social network system such as Twitter is not to search for total strangers on the Internet. Rather, the main purpose is to connect both online and offline to reinforce and maintain the strong ties present in offline friendships and to cultivate those weak ties with acquaintances that provide quality information. Generally, close associations, such as strong ties, often leave people with access only to the same resources as others with whom they are closely tied. Hence, the "strength of weak ties" (Granovetter 1973) characterizes connections to others outside the strong tie network and to the information or resources circulating in other spheres (Burt 1992; Granovetter 1973; McPherson and Smith-Lovin 1987). On Twitter, there are more chances to strengthen weak ties because Twitter users choose to "follow" other users without any consideration for whether the relationship will be accepted. Twitter users with weak ties receive positive outcomes such as solving problems, broadening their knowledge base, providing exposure to ideas and attitudes different from their own, and increasing their ability to recognize and take advantage of new opportunities (Allen 1977; Bruffee 1993; Cohen and Levinthal 1990; Coleman 1998; Haythornthwaite 2002; Koschmann 1996).

In this paper, we present an algorithm that enables to recommend the user with similar preference to other users in the sub-network and to fill the structural hole between sub-networks by connecting opinion leader based on the topic preference analysis. . In addition, a prototype system based on Infocuration algorithm is developed as a proof-of-concept, and finally the feasibility of our approach is tested. The compound word "Infocuration" originates from the words "information" and "curation." The term "curation" depicts tasks of acquiring, classifying, and maintaining valuable objects. An Infocuration system is a high-level information system that works closely with network users to assist them in identifying helpful information that solves their problems and reduces the obstacles for establishing or maintaining positive interpersonal relationships in the social network. Therefore, Infocuration systems perform the following tasks: providing valuable knowledge obtained from scattered information sources, solving a user's problems, and classifying contents using collective intelligence. In addition, it fills structural holes within the Twitter network by strengthening weak ties. Specifically, it recommends active users who possess sufficient subject knowledge to users seeking information through connections to subgroups, which relate to additional subgroups using a method that combines a topic matrix and collaborative filtering. The remainder of this paper adheres to the following structure: after presenting the micro-blogging service of Twitter and some related works, we introduce the methodology of developing an Infocuration system and conclude with a discussion and direction for future work.

Background

Twitter

Among micro-blogging websites, Twitter, a social network service, has been extending its role fast. Twitter was created in 2006 and had over 456 million accounts in 2012 with 175 million tweets per day (Chen et al. 2010). There are several factors to explain the selection of Twitter. First, it has a fast speed. Twitter has a 140-length restriction; therefore, it easily permits users to post and receive messages (called "tweets"), which can be quickly read on almost any mobile device. Second, it has reusable features that permit users to easily "re-tweet" other user's tweets. Third, the Twitter connection is unidirectional (Marcelo et al. 2010). Compared to other users of social networks such as Facebook, LinkedIn, or My Space, users of Twitter simply decide to "follow" other users without giving thought to whether the relationship will be accepted or reciprocated. Fourth, Twitter provides real time information. Users prefer to share information on Twitter because finding real-time information is easy. For example, while shopping, a user tweets about the store location and item, following which real-time information is exchanged and interesting conversation develops immediately among Twitter users. This type of information is more valuable than the data we already possess. However, as time goes by, the objective of using Twitter changes, which indicates that users employ Twitter to solve problems and obtain high-quality information rather than merely spreading rumors, reading news, or keeping in touch with friends.

Related Work on Twitter

Several efforts have been made to understand the Twitter network and its influence. To recommend followers and informative content, Armentano et al. (2011) present a method to generate a set of recommendations targeting users who post tweets that may be of interest to other users. Furthermore, they propose a way to evaluate and compare recommendations of tweet messages (Marcelo et al. 2011). Chen et al. (2010) suggest one approach for recommending URLs of interest that are generated from information streams such as tweets by applying a topic model of target users and a social voting mechanism. A topic model is an algorithm combining topic relevance ranking and score. Geyer et al. (2008) explore a number of topic recommendation systems for improving user engagement within social networks such as Twitter and argue that a social network-based recommendation yields better performance than simple content matching. Phelan et al. (2009) analyze user preferences such as tweets and recent news posted in user specified feeds and compares them to public timelines or Twitter friends for story ranking and recommendation.

A growing body of researchers use articulated social network information in ways that are similar to our study. Seth and Zhang (2008) use social network information from Orkut, a social networking website (http://www.orkut.com) that is owned and operated by Google, to build personalized recommendations of messages in communities. Their model is based on news media research and considers simplification and diversity features. Groh and Emig (2007) use collaborative filtering techniques to articulate social networks for neighborhood generation. Bonhard et al. (2006) suggest that rating overlap and profile similarity may be a powerful source of information to determine the appropriateness of a recommendation (Geyer et al. 2008). Even though many methods mentioned above have attempted to improve recommendations to social network users, they all have limitations; therefore, our study proposes an Infocuration system to overcome those limitations.

Methodology – Infocuration System on Twitter

Overall Procedure of Infocuration System

The primary purpose of our study is to create a novel way to provide the qualified information or solve the problem of efficient information distribution by filling the structural hole on Twitter network. In this study, we define Infocuration system as a name of qualified information transfer carrier in which Twitter network is organized under a title and subtopics. In other words, the major role of Infocuration is curation that supports the information generated by users to be shared efficiently. To develop this kind of an Infocuration system, we propose the following process. Figure 1 shows the overall procedure of our methodology to build Infocuration.

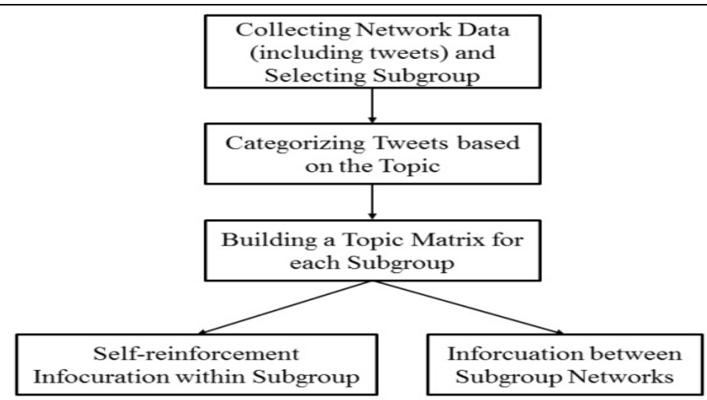


Figure 1. Overall Procedure of the Proposed Infocuration System

Collecting network data and extracting subgroup

Twitter provides an Application Programming Interface (API) that is easy to crawl and collect data. Infocuration crawls and collects profiles and messages of users who address specific topics on Twitter during a specific period. Messages regarding a trending topic are used to discover the group preference. Twitter data includes brief profile about users such as full name, location, a short biography, a web- page and the number of tweets of the user. In addition, it tracts the relationship among followers and followees.

Categorizing messages based on the topic category

Collected messages are categorized to assess which topics they correspond to. Messages use the categories that were provided in the popular twitter news website (Tweetmeme.com) where many researchers have used (Chen et al. 2010; Lerman and Ghosh 2010; Duan et al. 2010). Messages will be categorized by people, and reliability of the categorization will be secured since numerous people carried out the categorization.

Building a topic matrix for each subgroup

Each subgroup's topic matrix shall be defined to identify which tweets exist for the t number of topics when it comes to g number of topics for each subgroup network. Assume that the i^{th} subgroup G_i's total number of people is n_i, while topic matrix is T^i. At this time, T^i can be defined as follows.

$$T^i = N^i\{w_1 * TW^i + w_2 * RT^i + w_3 * MT^i\} \quad – (1)$$

where

- N^i is the $n_i \times n_i$ matrix that manifests the following relationship among the n_i number of people when it comes to the members of the G_i,
- TW^i is the $n_i \times t$ asymmetric matrix that manifests the number of tweets drafted by the members when it comes to each topic,

- RT^i is the $n_i \times t$ asymmetric matrix that manifests the retweeted number when it comes to the tweet drafted by specific members for each topic,
- MT^i is the $n_i \times t$ asymmetric matrix that manifests the frequency that the tweet for each topic is mentioned, and
- Weighting parameter w_1, w_2, w_3 will be decided in a heuristic manner using the weight for all the three elements mentioned above.

Figure 2 shows an example by utilizing the network of i^{th} subgroup, mentioned below in an arbitrary manner. Assumption is that the number of topics selected previously is four.

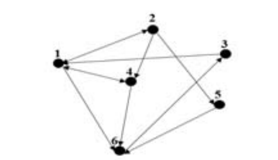


Figure 2. Network Example

N_i for this subgroup network can be expressed as follows.

$$N_i = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

TW^i, RT^i, MT^i's matrix for this network will be 6×4 matrix. Each component will be expressed in the tw_{ij}, rt_{ij}, mt_{ij} format:

$$TW^i = \begin{bmatrix} tw^i_{11} & tw^i_{12} & tw^i_{13} & tw^i_{14} \\ tw^i_{21} & tw^i_{22} & tw^i_{23} & tw^i_{24} \\ tw^i_{31} & tw^i_{32} & tw^i_{33} & tw^i_{34} \\ tw^i_{41} & tw^i_{42} & tw^i_{43} & tw^i_{44} \\ tw^i_{51} & tw^i_{52} & tw^i_{53} & tw^i_{54} \\ tw^i_{61} & tw^i_{62} & tw^i_{63} & tw^i_{64} \end{bmatrix}, RT^i = \begin{bmatrix} rt^i_{11} & rt^i_{12} & rt^i_{13} & rt^i_{14} \\ rt^i_{21} & rt^i_{22} & rt^i_{23} & rt^i_{24} \\ rt^i_{31} & rt^i_{32} & rt^i_{33} & rt^i_{34} \\ rt^i_{41} & rt^i_{42} & rt^i_{43} & rt^i_{44} \\ rt^i_{51} & rt^i_{52} & rt^i_{53} & rt^i_{54} \\ rt^i_{61} & rt^i_{62} & rt^i_{63} & rt^i_{64} \end{bmatrix},$$

$$MT^i = \begin{bmatrix} mt^i_{11} & mt^i_{12} & mt^i_{13} & mt^i_{14} \\ mt^i_{21} & mt^i_{22} & mt^i_{23} & mt^i_{24} \\ mt^i_{31} & mt^i_{32} & mt^i_{33} & mt^i_{34} \\ mt^i_{41} & mt^i_{42} & mt^i_{43} & mt^i_{44} \\ mt^i_{51} & mt^i_{52} & mt^i_{53} & mt^i_{54} \\ mt^i_{61} & mt^i_{62} & mt^i_{63} & mt^i_{64} \end{bmatrix}$$

According to the Formula (1), topic matrix T^i's (3, 2) component t^i_{32} can be obtained by inner-product between the N_i's 3rd row vector and the weighted sum of the TW^i, RT^i, MT^i's 2nd column vectors.

$$t^i_{32} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \cdot \left\{ w_1 \begin{bmatrix} tw^i_{12} \\ tw^i_{22} \\ tw^i_{32} \\ tw^i_{42} \\ tw^i_{52} \\ tw^i_{62} \end{bmatrix} + w_2 \begin{bmatrix} rt^i_{12} \\ rt^i_{22} \\ rt^i_{32} \\ rt^i_{42} \\ rt^i_{52} \\ rt^i_{62} \end{bmatrix} + w_3 \begin{bmatrix} mt^i_{12} \\ mt^i_{22} \\ mt^i_{32} \\ mt^i_{42} \\ mt^i_{52} \\ mt^i_{62} \end{bmatrix} \right\}$$

In other words, as this formula shows, the effect between-ness of the network level is calculated in the formula by factoring in the members' out degree for the specific topic by multiplying the N_i that manifests the relationship of Following among the members when it comes to the Formula (1). In other words, the 3rd member does not own only his or her writing, when it comes to the 2nd topic, but also even the writing of the 1st member who is following. As such, we can obtain the each subgroup's topic matrix. Let's assume that these topic matrix are T^1, T^2, …T^9. We want to carry out two types of Infocuration with this Matrix.

Infocuration

Self-reinforcement infocuration within subgroup networks

This Infocuration process induces the connection among the members whose preference for topics is similar within the subgroup, and this enables strengthening of the degree of knowledge sharing within the subgroup. Let us assume that the i^{th} subgroup's topic matrix T^i is as follows.

$$T^i = \begin{bmatrix} t^i_{11} & t^i_{12} & t^i_{13} & t^i_{14} \\ t^i_{21} & t^i_{22} & t^i_{23} & t^i_{24} \\ t^i_{31} & t^i_{32} & t^i_{33} & t^i_{34} \\ t^i_{41} & t^i_{42} & t^i_{43} & t^i_{44} \\ t^i_{51} & t^i_{52} & t^i_{53} & t^i_{54} \\ t^i_{n1} & t^i_{n2} & t^i_{n3} & t^i_{n4} \end{bmatrix}$$

In this matrix, collaborate filtering techniques are used to recommend each member to follow another member with the most similar preference by calculating the similarity among the T^i row vectors (no recommendation is made if already in the Following relationship). Through the process, weak-ties can be strengthened between users who have similar preferences in the subnetwork.

Infocuration between subgroup network

- Filling the structural hole between similar groups

During the process, the method recommends a user to follow based on the similarity of the topic preference among subgroups. The user with the highest centrality in each group will be selected as a followee. To calculate the value that represents the topic preference by individual groups, the topic matrix for each group is used. Each topic matrix factors the topic preference by all the individuals who comprise a subgroup. Likewise, the topic matrix's transpose and the topic matrix are multiplied ($T^{i^T} T^i$) to reduce individuals' dimension, saving only the dimension of the topic. $T^{i^T} T^i$'s eigenvectors. Eigenvector that corresponds to the largest values among them is referred to the subgroup *Gi*'s topic vector TV_i. *g* number of groups assume topic vector, and the calculation of the similarity level among these vectors helps to identify the groups with the most similar topic interests with each group. Later, Infocuration is completed by mutually recommending the opinion leaders (that is, the people with the highest centrality in the group) of the two groups that are closest to each other.

- Topic Association Rule mining using Topic Matrix and the Recommendation based on Completed Topic Matrix's

Row vector for all groups can be perceived as the vector that manifests the topic preference for each of individuals. Accordingly, it is possible to induce association among the topics when it comes to the topic preference through the analysis of row vector for all the topic matrices. First, top row vector of all the subgroup's members is made into table to make the correlation matrix among each topic (Figure 3). Because each component of the Topic row vector signifies topic, association among the topics is pulled out by making the correlation matrix among table's rows. When the correlation is larger than association threshold λ, rule is decided by assuming that there is an assoiaiton among them.

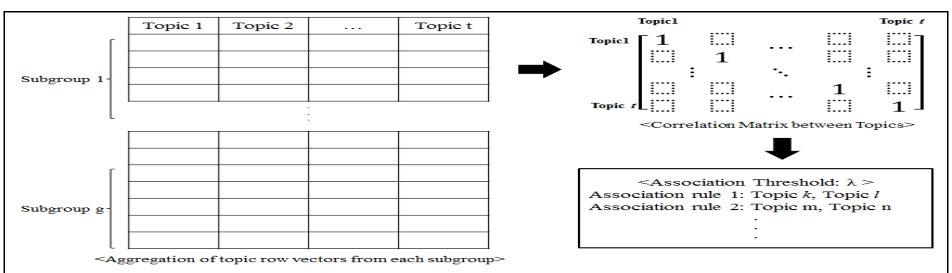


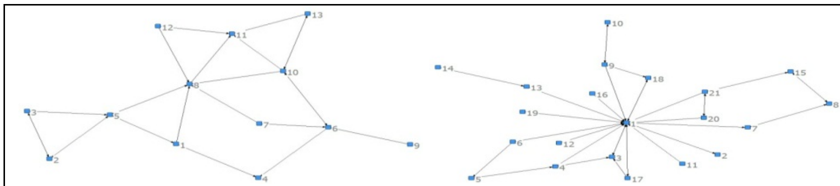
Figure 3. Process of Association Extraction

New link among the subgroups will be recommended by utilizing association rule mining. Each of the columns of the Top Matrix will be added to calculate which group is most interested in which group. Then, cross recommendation will be carried out when the association rule is performed. For example, assume that the subgroup

1 is interested considerably in the k^{th} topic while subgroup 2 is most interested in the 1^{st} topic. However, if the result of the association rule shows that there is an association between topic *l* and *k*, the link among the two groups is recommended. That is, recommendation is made to link the person who contributed the most to the topic *l* among the members of the Subgroup 2 and the person who contributed the most to the topic *k* among the members of the 1.

Sample Test: Infocuration Scenario

We conducted a simple pilot test using real data to confirm the feasibility of the proposed method. The self-reinforcing nature of Infocuration among groups can be explained in detail by extracting two groups comprising 15 to 20 persons each. According to the method mentioned above, we crawl the tweets that are produced during a specific period (two months) into two different groups. Subsequently, categories are assigned through a round of crosschecking by three graduate students who manually tagged every category (a total of nine categories) in relation to each tweet. Consequently, the association rule among topics could be extracted, and the topic vector of individuals was calculated (the related raw data are not presented here owing to the page limit). Moreover, the following relationship network of each group was extracted. Thus, this study intends to show the following Infocuration Scenario. First, the following network of each group is presented in Figure 4 below (the identity of each participant is shown numerically to protect privacy).



Group1 Group 2

Figure 4. Following Relationship Network of Two Groups

(1) Self-reinforcement Infocuration within Subgroup Network

In self-reinforcement Infocuration, two members are defined as having similar tendencies if the cosine similarity exhibits a topic preference among members that is greater than 0.5. For Group 1, only members 8, 10, and members 5, 13 exhibited the similarity of over 0.5. Although members 8 and 10 were found to have the highest topic preference similarity, a following relationship was already established; therefore, Infocuration was irrelevant in their case. Despite members 5 and 13 exhibiting the second highest topic preference similarity, they did not have a direct following relationship; therefore, Infocuration was possible. Compared to Group 2, the similarity was above 0.5 among most members because Group 2 was a homogenous group and the topics of interest (lifestyle) for this group matched those of member 1. As for Group 2, additional Infocuration is unnecessary because most members are connected within two hops based on their connection with member 1.

(2) Infocuration between Subgroup Networks

- Filling the structural hole between similar groups

Based on the results of the extraction of eigenvectors in each group's topic matrix and the calculation of cosine similarity, the value was found to be as low as 0.32. This suggests the topic pattern—of interest to both groups—is not similar, and therefore no mutual recommendation is made.

-Topic Association Rule mining using Topic Matrix and the Recommendation based on Completed Topic Matrix's

The association among the topics derived through the analysis of individual topic vector, (Lifestyle, World & Business) and (Channels, World & Business) showed far higher values (Association Threshold=0.6). The Infocuration between Groups 1 and 2 can be made using this association rule. Group 2 is interested mainly in lifestyle and has the highest centrality for this topic (out-degree: 6, in-degree: 16, and between-ness centrality: 163.5). In Group 1, member 10 generated the largest number of tweets for topics related to world and business and showed the highest centrality among users who wrote topic relevant tweets (out-degree: 5, in-degree: 3, and between-ness centrality: 22). Thus, Infocuration occurred through mutual recommendation by member 10 in Group 1 and member 1 in Group 2.

Conclusion and Future Work

Currently, the task of realizing an Infocuration system is underway. This study is not limited to the realization of the IT artifact. Instead, our study accompanies the developing conceptualization of Infocuration. As the social network increasingly contains more information, the process of searching for quality information will become ever more complicated. In this context, Infocuration will provide users with practical knowledge and experience that can enable them to resolve problems more effectively. Moreover, Infocuration will play a very important role in strengthening weak ties that are important for the advancement of social networks.

However, the following steps needs to be taken to establish a better Infocuration system. First, it is necessary to present the topic structure considering depth and valence of the topic and check Infocuration Scenario related to various topics by extracting more groups and refining the algorithm. For a larger scale pilot test, the methodology that automatically distinguishes the topic of each tweet should be applied. Second, the process of measuring the satisfaction of real users with Infocuration should be added and the result should be used to improve the Infocuration method. If these points are revised and added, Infocuration systems will play a key role in dramatically enhancing the efficiency of searching for information through social media.

Reference

Allen, T. J. 1977. *Managing the flow of technology*, Cambridge: MIT Press.

Bonhard, P., Harries, C., McCarthy, J., and Sasse, M.A. 2006. “Accounting for taste: Using Profile Similarity to Improve Recommender Systems,” in *Proceedings of ACM CHI’06*, Montreal, Canada, pp. 1057-1066.

Bruffee, K. 1993. *Collaborative learning*, Johns Hopkins Press, Baltimore.

Burt, R. S., *Stuructural Holes*, Cambridge, MA; Harvard University Press.

Cha, M., Haddadi, H., Benevenuto, F. and Gummad, K. P. 2010. “Measuring user influence on twitter: The million follower fallacy,” in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, pp. 10-17.

Chen, J., Nairn, R., Nelson, L., Bernstein, M, and Chi, E. H. 2010. “Short and Tweet: Experiments on Recommending Content from Information Streams,” in *Proceedings of CHI'2010*, Atlanta, GA, pp. 1185-1194.

Coleman, J. S. 1988. “Social Capital in the Creation of Human Capital," *American Journal of Sociology* (94:1), *Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure,* pp.S95-S120.

Cohen, W., and Levinthal, D. 1990. “Absorptive capacity: A new perspective on learning and innovation,” *Administrative Science Quarterly* (35:1), pp. 128-152.

Duan, Y., Jiang, L., Qin, T., Zhou, M., and Shum, H. Y. 2010. “An empirical study on learning to rank of tweets,” in *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China. pp. 295-303,

Geyer, W., Dugan, C., Millen, D. R., Muller, M., and Freyne, J. 2008. “Recommending Topics for self-descriptions in online user profiles,” in *Proceedings of the 2008 ACM conference on Recommender systems*, Lausanne, Switzerland, pp.23-25.

Granovetter, M. S., 1973. "The Strength of Weak Ties," *American Journal of Sociology* (78:6), pp. 1360-1380.

Groh, G., & Ehmig, C. 2007. “Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering,” in *Proceedings of ACM Group*, pp. 127-136.

Haythornthwaite, C., and Wellman, B. 2002. “The Internet in everyday life: An introduction”. In The Internet in everyday life, B.Wellman & C.Haythornthwaite (eds.),. Oxford : Blackwell., pp. 3-44.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., and Riedl, J. 1997. “GroupLens: applying collaborative filtering to Usenet news,” *Communications of ACM* (40: 3), pp. 77-87.

Kwak, H., Lee, C., Park, H. and Moon, S. 2010. “What is Twitter, a social network or a news media?,” in *Proceedings of the 19th international conference on World wide web*, Raleigh, North Carolina, USA., pp. 26-30.

Lerman, K., Ghosh, R. 2010. “Information contagion: An empirical study of the spread of news on Digg and Twitter social networks,” in *Proceedings of International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA, pp. 90-97.

Marcelo, G. Armentano, D., Godoy, and Amandi, A. 2011. “Towards a Followee Recommender System for Information Seeking Users in Twitter,” in *Proceedings of the Workshop on Semantic Adaptive Social Web*, Girona, Spain.

McPherson J. M., and Smith-Lovin L. 1987. “Homophily in voluntary organizations: status distance and the

composition of face-to-face groups," *American Sociology Review* (52:3), pp. 370-379.
Phelan, O., McCarthy, K., and Smyth, B. 2009. "Using Twitter to recommend real-time topical news," in *Proceedings of the 3rd ACM Conference on Recommender Systems*, New York, pp. 385–388.
Seth, A., and Zhang, J. 2008. "A Social Network Based Approach to Personalized Recommendation of Participatory Media Content," in *Proceedings of the* ICWSM'08, Seattle, WA.
Wang, R. Y., and Strong, D. M. 1996. "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems* (12:4), pp. 5-34.