Towards Efficient Content Management in Online Social Communities – A Study of User Interest and Context

Xing Xie, Dongsheng Li, Huanhuan Xia

School of Computer Science
Fudan University, Shanghai 201203 P.R.China

Abstract— **As hundreds of millions of users access online social communities at daily or even real-time basis, large amounts of user data are continuously generated. The fast-growing user-generated content poses great challenges on online social network system design for efficient content management and delivery. For instance, in content-centric online social network, all contents are organized in a temporal order which makes it very time consuming for users to browse all these contents to locate what they really like. By conducting a comprehensive study of online user activities, including content, social, and time characteristics, this paper try to accurately characterize user interest and user context, with the end goal of more efficient content management and real-time content delivery in online social network systems. The detail analysis of user activities is conducted on the real data collected from a popular online social community among Chinese universities with over 63,000 users to demonstrate the advantage and effectiveness of extracting user interest and context characteristics and applying them in designing more efficient content management systems.**

I. Introduction

As online social communities (also called online social networks) have become more and more popular, hundreds of millions of users now participate in online social communities, and a huge number of content items are dynamically generated by individual users and shared with others, such as blog posts, replies, photos, and short videos. One urgent requirement in such systems is efficiently managing the dynamic generated huge number of content items such that the content items of interest to specific users can be easily identified and delivered to end users in real time. For instance, in a popular online social community with thousands of new topics created everyday, users will fell frustrated if they need to browse all those contents to locate what they really like. What they need is such a system that can know their interests and deliver those interested contents to them in a real time way.

To design such content management systems in online social network systems is really a challenging problem, because it requires a deep and thorough understanding of two key characteristics in online social communities: user interest and user context, i.e., what types of content a user is interested in, under what context. However, most of recent research [16], [4], [14], [7], [13], [5] about online social communities only focus on the graph structure of online social networks to reveal the structure characteristics in online social communities, such as power-law node degree distribution, link symmetry, local clusters, and group evolution. Although important, these structure characteristics tell little about the diverse interests of different users.

With the end goal of designing more effective and efficient content management system in online social communities, our work differs from previous studies in that we focus on the content, social, and time characteristics of user activities, in order to accurately identify user interests and context. We also prove such knowledge (i.e., the observed user activity patterns) can be utilized at runtime to quickly identify content items of particular interest to certain users and deliver content items to interested users in real time.

This work makes the following contributions:

- We have collected four months worth of comprehensive user activity data and friendship data from a popular online social community consisting of over 63,000 users, 2 million posts and replies, and 18 million post views;
- We conduct a detailed analysis of user activities, and identify important user interest and context characteristics, which we believe is never done before in large scale online social communities;
- We demonstrate the effectiveness of utilizing user interest and context information for efficiently identifying and delivering user-interested content items in content management systems.

The rest of the paper is organized as follows. Section II presents our methodology for data collection and data analysis. Detailed analysis of user activities in terms of interest and context is presented in Section III, and Section IV evaluates the effectiveness of interest- and context-based content management and delivery. Section V discusses related work, and Section VI concludes this paper.

II. Data Collection and Analysis Methodology

In this section, we describe our methodology for data collection and data analysis.

A. Data Collection

Ri Yue Guang Hua (RYGH)[1] is a very popular and representative online social community among Chinese universities. A

[1] http://bbs.fudan.edu.cn

rich set of social functionalities are supported, such as adding and communicating with friends, viewing and posting articles (and multimedia content) to various social or interest groups. RYGH currently has over 63,000 registered users, over 50% of them visit the website on regular basis. Everyday, there are about 3,000 new topics, 15,000 new posts, and 250,000 view requests for articles.

RYGH has two important properties for user activity analysis:

• This is a relatively large and self-contained social community as it contains mostly current students and alumni of the university and is visited regularly by its users. This allows us to capture a complete picture of the whole community and the activities of individual users.

• After over ten years' evolvement, this online system has established well-structured subcommunities and latent interest groups, covering a complete yet diverse set of information and communication needs of its users. This allows for easy classification of different types of user activities.

To capture user post activities, we have developed a web crawler that periodically checks for recently updated webpages for each discussion board (subcommunity). Users' view article activities are identified by checking the URL requests logged by the application server hosting the system. In addition, we have obtained friendship information from each user's profile. Overall, we have collected four months worth of daily user activities across 373 subcommunities. There are approximately 6 GB data in total, and the data details are listed in Table I.

TABLE I
STATS OF CRAWLED DATA

Users	Topics	Posts	Views	Friendship links
63,706	278,442	2,108,086	17,948,242	509,458

B. Analysis Methodology

In online social communities, User-generated-content (UGC) is created by users' activities, which are intuitively influenced by users' interests and social relations. In order to explore the interconnection among activities and interests, we first investigate if users' activities are guided by their interests. This study will give us the knowledge about how users generate their data. Activities alone are also useful for content management and delivery, such as web users will be more active on weekend. So studying the activities' context to find the patterns for activity will be helpful for data managers. In order to know the practicalness of the results we get, we should have a way to do the evaluation. To this end, we porpoise

• Activities V.S. Interests

Exploring the interconnection between activities and interests is not trivial task since users' interests are not apparent in online social communities. We identify user's interests in two ways: 1) using the social groups (subcommunities) information. 2) extracting interests from users' daily activities. With the interests information, we use statistical method, e.g. statistical distribution of interests in users' activities, to show the correlation between activities and interests. Also, we find that users' interests are diverse even in a subcommunity, so that we use a clustering method to find a user's true interest.

• Activity V.S. Context

Furthermore, we study the activities context by using statistical analysis, in order to find different activity patterns in different subcommunities.

• Case Study

In addition, we give a case study for utilizing our study results in a recommender system, an instance of content management and delivery for online social communities. We justify the conclusions that we made in Section III, and then discuss the possible extensions of our conclusions in other domains.

III. USER ACTIVITY ANALYSIS

Through an in-depth analysis of the wide variety of online user activities (i.e., article post, view and reply), the goal of this study is to identify which information extracted from user activity can be used to improve the efficiency of content management and delivery in online social communities.

So in this section, we first analyze the distribution of users' interests extracted from users' activities and conclude that users' concentrated interests can be used to improve the content management by interest-based content storage. We then investigate the context pattern of users activities and observe users' specific context pattern in different subcommunities. In the end, we show that viewing activities are more complete for characterizing most users' interest than posting activities.

The significance of analysis of these features will be further evaluated by the improvement of recommending quality in Section IV.

A. Activity and Interest

In this section, we present that user has diverse activity on his/her interests which means that content can be organized in the unit of interest for flexible storage and efficient access.

At first, we need to introduce a new concept $Interest\ group$. To organize all kinds of content in a online social community, the online social communities' designers often manually categorize the whole community into many subcommunities by their different content domain. In fact, a subcommunity may often have their clear internal structure of interest groups (a interest group consists of items with more similar contents). A user may be interested in all items in a subcommunity or just items in several interest groups. For example, considering a sport community consisting of several subcommunities (e.g., football, basketball, volleyball and so on), Bob may be interested in all items about basketball but Alice may be only interested in items concerned to her own home teams. So we can say user's interests lie in two layers: subcommunity and interest group.

Generally speaking, users tend to have more activities on their interested content, so we can identify users' real interest by the analysis of users' activities distribution in whole community. We first characterize the distribution of user interest across the variety of subcommunities. Figure 1(a) shows the CDF of number of subcommunities that users interested in. As we can see, a majority of users only participate in a few

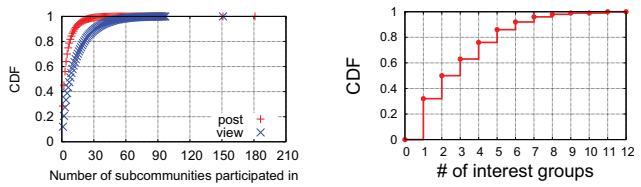

Fig. 1. (a). CDF of number of subcommunity users participated in. (b). CDF of user interest in the Astrology subcommunity.

subcommunities – 90% (63%) users have posted (viewed) messages in less than 12 subcommunities, which means that most users have concentrated interests on subcommunities. The similar interest concentration lies in interest group layer. We use the Astrology subcommunity in RYGH as an example which consists of 12 implicit interest groups (i.e., 12 zodiac groups), and a user is interested in one interest group if he or she has activities in it. Figure 1(b) shows the content cumulative distribution of the 12 implicit interest groups in the Astrology subcommunity. Most users (78%) are interested in no more than 4 interest groups which indicates more concentrated user interests within the subcommunity.

Next, we study the distribution of users' activities among the interests that they are interested in. Let S_i be the total number of activities user i participate in all k interests, and $S_{i,j}$ be the number of activities of user i in interest j, we use the *disparity measure* $Y(k,i)$ [9] to characterize the diversity of activity in one's interests, as follows.

$$Y(k,i) = \sum_{j=1}^{k} (\frac{S_{i,j}}{S_i})^2$$

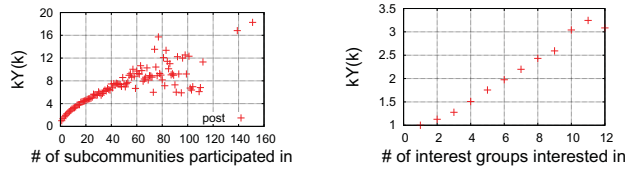

Fig. 2. (a). Disparity of users' participation in every subcommunities. (b). Disparity of users' participation in every interest groups.

If user activities are evenly distributed among the interests, $kY(k,i) = 1$; and if the distribution is highly nonuniform, $kY(k,i) \to k$. Figure 2(a) shows the result on subcommunity layer and Figure 2(b) shows the result on interest group layer. As we can see, both of the distribution of user activity among the interests exhibit great disparity. This study demonstrates that, online user activities are mainly driven by their specific interests. So more flexible storage and efficient access can be obtained by organizing content in the unit of interest. By using the analysis of interest groups, the improvement of recommending quality will be shown in Section IV.

B. Activity and Context

In this section, our study focuses on investigating the context patterns of user activities and we explain why users' context information must be used to characterize one's real interest.

There are two general causes leading to users' specific context pattern: the schedule of user's specific identity and

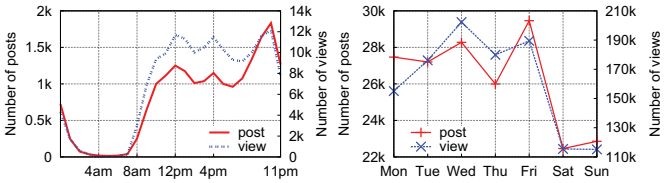

Fig. 3. (a). User daily context activity. (b). User weekly context activity.

social events in different subcommunities. Figure 3(a) shows the statistics of the daily activities of the online users. Consistent hour-by-hour patterns are observed – user activities peak at around 12 pm, 4 pm, and 10–11 pm. Such patterns are correlated with the daily activities of college students. The influence of user's specific identity to users' context pattern can also be observed in user weekly activities. As shown in Figure 3(b), users are more active during weekdays. The total user activities decrease significantly during weekends, as many of the students are out of town or occupied by other events.

Social events in different subcommunities also influence users' context pattern a lot. For instance, in *Stock*, users are active during the trading time, 9:30–11:30 am and 1–3 pm of the stock market in China. Similar social- and/or content-dependent activity patterns can also be observed in other subcommunities of RYGH. Because there are different time of access peak, we can improve the efficiency of content management by allocating more efficient storage resource and access strategy to a subcommunity when it comes to its access peak.

Now, we explain why users' context information must be used to characterize one's real interest. In online social community, most users only browse a small number of posts even though some of ignored items may be of interest to them. Marking those "ignored" content as uninterested leads to incorrect user interest estimation, which can be called the false negative problem. Leveraging users' time context information can help identify users' online session. And only the content items belonging to the user's online sessions are considered to identify the interest of an online user.

C. Viewing activity and Posting activity

In above analysis, we can see viewing activity and posting activity have similar trend but not exactly the same. So it is important to compare them to see which is more complete for characterizing most users' interest.

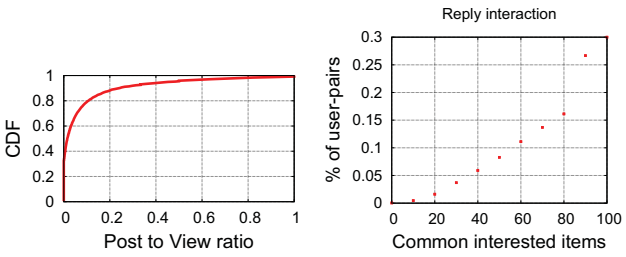

Fig. 4. (a). CDF of post to view ratio. (b). Percentage of user-pairs with interaction.

To see the difference between the viewing activity and posting activity, we first present users' posting activity to

viewing activity ratio. From Figure 4(a), we can see that there are 32% of users who don't have posting activity and 88% of users whose posting activities are less than 20% of their viewing activities. This observation means that most users view many interested items but do not always comment on them. So it's impossible for extracting their complete interest only from posting activity. The incomplete problem of posting activity can also be seen in Figure 4(b). We can see that in all of user-pairs who have common interested (viewed) items there are only 0.037% of user-pairs have reply interaction (posting activity). Even though for those more similar user-pairs who have more than 100 common interested items there are 30% of user-pairs which is still a low percentage. This conclusion explains why we use viewing activity to define users' similarity in Section IV. Another interesting conclusion gained from the trend of increasing monotonically in Figure 4(b) is that more similar user-pairs have larger possibility to interact with each other.

IV. A Case Study In a Recommender System

In this section, we present the experiment results by leveraging the conclusions made in Section III to further justify that our user interest analysis and context analysis can help to improve the quality and efficiency of content management and delivery. We evaluate our analysis in a recommender system which is a typical instance of content management and delivery in online social communities, then we discuss the possibility of extending our work to other content management and delivery domains. The results are mainly presented in the following aspects:

- *Leveraging user interest analysis:* We show that both the recommendation quality and efficiency can be improved by clustering items and users into interest groups, which means our method can more accurately identify user content interest.
- *Leveraging user context analysis:* We show that the context analysis can help to gather real-time information, and can also help to improve recommendation quality, which means our context extraction method can identify user context session more accurately.

A. Recommender Systems

Recommender system is designed to help to filter items (articles, news, books, movies, etc.) that are interesting to individual users. It can detect user interests based on historical activities and delivery new items based on user interests. Amazon [15], Google News [10] develop their commercial recommender systems to provide personalized content delivery.

As the recommender systems are designed to identify user interest, better user interest identification method and context extraction method can help to improve recommendation quality. So, the recommender system is a good instance for us to evaluate our user interest analysis and user context analysis.

B. User Interest Identification

Here we compare the recommendation performance by leveraging the user interest identification method proposed in the previous section with one of state-of-the-art Collaborative Filtering (CF) algorithm, i.e., MinHash based recommendation algorithm [10].

As a probabilistic clustering method to cluster users, the basic idea of MinHash is to randomly permute the whole set of items (S) and for each user u compute its hash value h(u) as the index of the first item under the permutation that belongs to the users' interested item set. Each obtaining hash bucket corresponds to a cluster, that puts two users together in the same cluster with probability equal to their item-set overlap similarity. We can also repeat this step in multiple times and hash each user u into multiple clusters. After obtaining the users clusters, predicting the rating of an item t for user u is computed as Equation 1:

$$P_{u,t} = \sum_{Ci:u \in Ci} \omega(u, Ci) \sum_{v,u \in Ci} R_{v,t} \quad (1)$$

where $\omega(u, Ci)$ is proportional to the fractional membership of the user u to cluster Ci, and $R_{v,t}$ is 1 if the user v likes the item t and 0 otherwise.

The original MinHash algorithm considers the whole items set and users set in a subcommunity, while we consider the items set and users set in a interest group when using MinHash algorithm. We call the modified recommendation algorithm as $MinHash - IG$. The detailed interest group identification algorithm is ignored here since it is not the emphasis of this paper.

In Collaborative Filtering (CF), $Precision$,$Recall$ and *F-measure* are three of the main evaluation metrics for performance in the binary recommendation case. They are defined as follows:

$$\begin{aligned} Precision &= \frac{|S_u \cap S_r|}{|S_r|} \\ Recall &= \frac{|S_u \cap S_r|}{|S_u|} \\ \textit{F-measure} &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned}$$

where S_u stands for the items that user u are interested in, and S_r stands for the items that we recommend.

As $Precision$ and $Recall$ can not reflect the recommendation quality alone, we choose the *F-measure* which combines the $Precision$ and $Recall$ as the evaluation metric to evaluate our recommendation quality.

In this experiment, we consider six subcommunities in RYGH. These six subcommunities with diverse popularity, activity patterns and internal structures (i.e., interest groups) can provide a comprehensive setup for performance evaluation of the proposed user interest identification method.

TABLE II
Recommendation performance of MinHash-IG and MinHash

Subcommunity	MinHash	MinHash-IG	Improvement
Astrology	0.376	0.432	14.9%
Music	0.493	0.515	4.4%
OMTV	0.408	0.425	4.1%
Auto	0.317	0.323	1.7%
Joke	0.426	0.437	2.6%
Movie	0.453	0.470	3.6%

General speaking, the more items system recommends, the higher recall we can get while the F-measure value changes a little. In order to eliminate the influence of different recall, we calculate the average F-measure for 10 different recall (i.e., from 10% to 100%) and the performance results of six subcommunities are shown in Table II. As we can see, the CF algorithm using the user interest identification method consistently outperforms MinHash algorithm among all six subcommunities, with an average of 5.2% improvement in *F-measure*. And the improvement is dramatical for those subcommunities with more clear latent interest groups (e.g., the subcommunity of Astrology).

Using the user interest identification method, user interests are focused on part of the items, and their similar users are focused on a subset of users which have common interests on this part of items, so that the recommendation efficiency can also be improved. Figure 5 shows the comparison of computation time between our method and MinHash algorithm. As we can see, our method outperforms MinHash algorithm consistently in all the six subcommunities. On average, less than $1/5$ time is used in MinHash-IG compared with MinHash algorithm.

C. User context Extraction

The user interest identification method can detect user interests in the content dimension, but in online social communities, most users only browse part of the items, which indicates that we should detect user interests in the time dimension. The user context extraction method can help to detect user interest in the time dimension, which means to identify user interests even more accurately. The experiment using the user context extraction method justifies the conclusion.

TABLE III
ONLINE RECOMMENDATION PERFORMANCE OF MINHASH-IG AND MINHASH

Subcommunity	MinHash	MinHash-IG	Improvement
Astrology	0.376	0.405	7.6%
Music	0.493	0.565	14.5%
OMTV	0.408	0.487	19.4%
Auto	0.317	0.363	14.4%
Joke	0.426	0.454	6.4%
Movie	0.453	0.543	19.8%

After the real-time context information extraction, we can filter the items that are generated when a user is offline, so that the maybe wrong decisions about these items can be filtered. In this case, the recommendation quality can be further improved as shown in figures III. Our method outperforms MinHash algorithm much more in all the six subcommunities. On average, an improvement of 13.7% can be achieved.

D. Possible Extensions

As mentioned in this section, our analysis is not limited to the collaborative filtering domain. Many applications which need to manage and deliver large amounts of data to individual users or clients can benefit from our analysis, such as content delivery network (CDN), search engine, P2P file sharing, web service and so on. We discuss some of them in the following.

Content delivery network, also named as content distribution network, is used to speed up clients' web access by caching static content from the web server in a local or nearer server. One of the challenges in CDN is the low hit rate problem[1]. Our interest-based clustering can be used to improve the hit rate in CDN. Users in different areas may have different preferences which may be considered as 'interest'. For instance, the CDN server in Los Angeles should cache more news about the LA Lakers for NBA (National Basketball Association) fans, but the server in Boston should cache more news about the Boston Celtics. In this case, our analysis about user interests can be beneficial to CDN.

Search engine is a kind of application that is designed to help users to find information on World Wide Web. Personalized search is becoming more and more popular in recent years, Pitkow et al. argued that two challenges in personalized search are the contextualization and individualization in [19]. The user interest analysis can help to find user true interest, thus can achieve better individualization. Our context information analysis may also help in contextualization.

V. RELATED WORK

Our work draws upon research in such areas concerning social networks, including user activity analysis, and social community applications.

Online social communities have been growing at an enormous speed and drawing significant attention in the recent past. Most studies have focused on analyzing the graph structure of online user friendship networks and the dynamics of online social networks [16], [4], [14], [7], [13], [5]. Several important characteristics of online social networks have been identified and confirmed, including power-law degree distribution, small-world, scale-free, clustering, and weak ties. Although important and useful in many domains, these analysis cannot capture the inherent characteristics of users' content interests and the contexts of their content access requests.

A few recent studies began to investigate user activities in online social networks [6], [17]. Hyunwoo Chun et al. compared the explicit friendship network and the implicit activity network of Cyworld [9]. Adamic et al. analyzed the content characteristics and user interaction patterns at Yahoo Answers [2]. Different from existing work considering single-type user activity networks, our work studies the dynamic structures of a variety types of content, context, and social online activities, from individual users, interest groups, subcommunites, to the whole social network.

An in-depth understanding of the dynamics of user activities is essential to the design of next-generation Internet, and will potentially benefit many emerging applications, such as recommender systems [10], [18], information filtering [3], [11], and Internet marketing [8]. Let us consider recommender systems as an example. The goal of a recommender system is to automatically identify a set of items (e.g., content, products, or users) which are of interest to a certain user. Memory-based methods [12] compute user similarities and predict user's preference based on the opinions of others. Model-based methods [10], on the other hand, cluster users with similar interests into the same group. Item-based method [15]

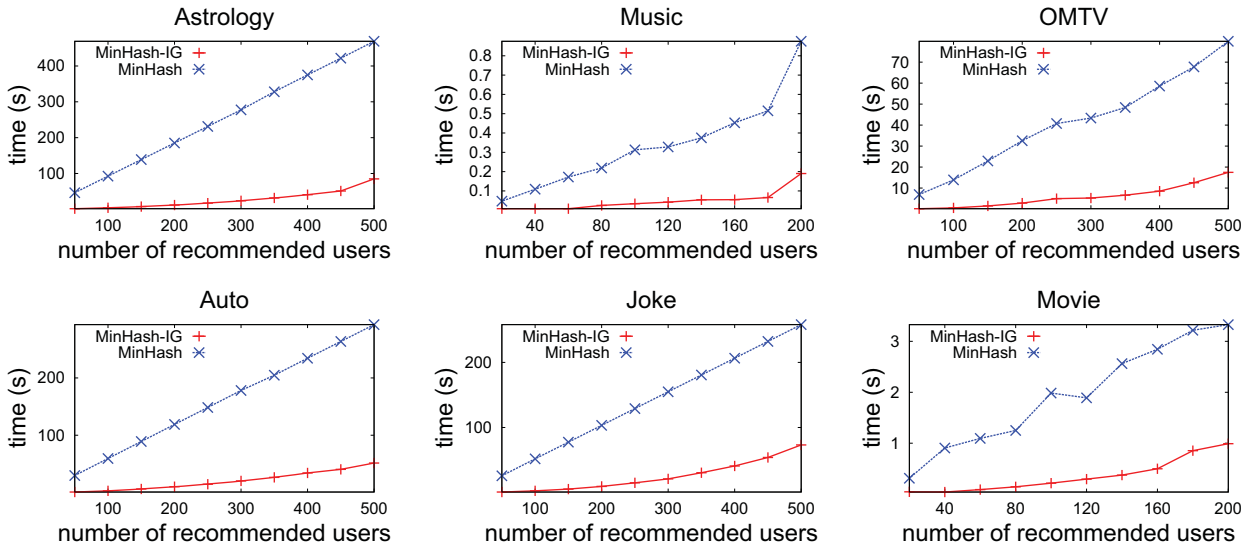

Fig. 5. Computation performance of MinHash-IG and MinHash in six subcommunities.

and hybrid methods have also been proposed [10], [18]. In this work, we use the MinHash based recommendation algorithm [10] to investigate the structures of user activities (see Section III). As shown in Section IV, our analysis on user activity can be used to improve the design of recommender systems.

VI. Conclusions and Future Work

Content management is an important and challenging task in online social network systems, as a huge number of content items are dynamically generated by individual users and shared among many users. In this work, we present a comprehensive study on a popular online social community. Focusing on user activities, we extract important user interest and context characteristics, and demonstrate that such characteristics can be utilized at run time to improve the quality and efficiency of content management and content delivery in online social communities, thus addressing an increasingly important issue in information explosion and online social community evolution.

References

[1] M. Abrams, C. R. Standridge, G. Abdulla, S. Williams, and E. A. Fox. Caching proxies: Limitations and potentials. In *Proceedings of the Fourth Interntional Conference on World Wide Web*, Boston, MA, December 1995.

[2] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and Yahoo answers: Everyone knows something. In *WWW '08: Proc. of the 17th intl. conf. on World Wide Web*, pages 665–674. ACM, 2008.

[3] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR '06: Proc. of the 29th annual intl. ACM SIGIR conf. on Research and development in information retrieval*, pages 19–26. ACM, 2006.

[4] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *WWW '07: Proc. of the 16th intl. conf. on World Wide Web*, pages 835–844. ACM, 2007.

[5] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *KDD '06: Proc. of the 12th ACM SIGKDD intl. conf. on Knowledge discovery and data mining*, pages 44–54. ACM, 2006.

[6] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *IMC '07: Proc. of the 7th ACM SIGCOMM conf. on Internet measurement*, pages 1–14, 2007.

[7] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proc. of WWW '09*, pages 721–730, Madrid, Spain, 2009. ACM.

[8] J. A. Chevalier and D. Mayzlin. The effect of word of mouth on sales: Online book reviews. *Marketing Research*, 43(3):345–354, August 2006.

[9] H. Chun, H. Kwak, Y.-H. Eom, Y.-Y. Ahn, S. Moon, and H. Jeong. Comparison of online social relations in volume vs interaction: A case study of cyworld. In *IMC '08: Proc. of the 8th ACM SIGCOMM conf. on Internet measurement*, pages 57–70. ACM, 2008.

[10] A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: Scalable online collaborative filtering. In *WWW '07: Proc. of the 16th intl. conf. on World Wide Web*, pages 271–280, 2007.

[11] S. Han, Y. yeol Ahn, S. Moon, and H. Jeong. Collaborative blog spam filtering using adaptive percolation search. In *WWW2006 Workshop on the Weblogging Ecosystem*, 2006.

[12] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR '99: Proc. of the 22nd Annual intl. ACM SIGIR Conf. on Research and Development in information Retrieval*, pages 230–237. ACM, 1999.

[13] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *Proc. of KDD '06*, pages 611–617, Philadelphia, PA, USA, 2006. ACM.

[14] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *Proc. of KDD '08*, pages 462–470, Las Vegas, Nevada, USA, 2008.

[15] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.

[16] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *IMC '07: Proc. of the 7th ACM SIGCOMM conf. on Internet measurement*, pages 29–42, 2007.

[17] J. B. Pena-Shaff and C. Nicholls. Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, 43(3):243–265, 2004.

[18] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles. Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach. In *In Proc. of the Sixteenth Conf. on Uncertainty in Artificial Intelligence*, pages 473–480. Morgan Kaufmann, 2000.

[19] J. Pitkow, H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel. Personalized search. *Commun. ACM*, 45(9):50–55, 2002.