User Grouping Behavior in Online Forums[*]

Xiaolin Shi[†] Jun Zhu[‡] Rui Cai[§] Lei Zhang[§]

[†]Dept. of EECS, University of Michigan, Ann Arbor, MI. shixl@umich.edu
[‡]Dept. of Comp. Sci & Tech., Tsinghua University. jun-zhu@mails.thu.edu.cn
[§]Microsoft Research Asia. {ruicai,leizhang}@microsoft.com

ABSTRACT

Online forums represent one type of social media that is particularly rich for studying human behavior in information seeking and diffusing. The way users join communities is a reflection of the changing and expanding of their interests toward information. In this paper, we study the patterns of user participation behavior, and the feature factors that influence such behavior on different forum datasets. We find that, despite the relative randomness and lesser commitment of structural relationships in online forums, users' community joining behaviors display some strong regularities. One particularly interesting observation is that the very weak relationships between users defined by online replies have similar diffusion curves as those of real friendships or co-authorships. We build social selection models, Bipartite Markov Random Field (BiMRF), to quantitatively evaluate the prediction performance of those feature factors and their relationships. Using these models, we show that some features carry supplementary information, and the effectiveness of different features vary in different types of forums. Moreover, the results of BiMRF with two-star configurations suggest that the feature of user similarity defined by frequency of communication or number of common friends is inadequate to predict grouping behavior, but adding node-level features can improve the fit of the model.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behaviorial Sciences; H.2.8 [**Information Systems**]: Database Management—*Database Applications, Data mining*

General Terms

Measurement, Experimentation

Keywords

Social networks, online forums, information diffusion, social selection model

1. INTRODUCTION

Online forums provide a unique type of social environment that enables people to share and access information freely. Users can either start new topics or leave comments in the threads of existing topics. Usually, an online forum has tens or hundreds of distinct *boards* or *communities*. These boards or communities group hundreds to thousands of threads of similar related topics together. Because of the huge numbers of users and the high dynamics of online forums, this type of environment has a rich complexity [6].

As new ideas and controversial discussions are always emerging and propagating among online forums, it is interesting to study the process of *information diffusion* in this social medium. The human behavior of gathering together and forming groups has been an important theme in studying information diffusion, because people taking the same actions as their neighbors is strong evidence that information flow has occurred [1]. Characterizing user grouping behavior in online social environments does not only help researchers to understand many of the sociological problems of human behavior, but also facilitates them to improve various applications in the online environment, such as the recommendation systems [22].

In this paper, we are mainly focusing on three central questions:

i What are the factors in online forums that potentially influence people's behavior in joining communities and how do they impact?

ii What are the relationships between these factors, i.e. which ones are more effective in predicting the user joining behavior, and which ones carry supplementary information?

iii What are the similarities and differences of user grouping behavior in forums of different types (such as news forums versus technology forums)?

By *a user joining a community* in a online forum, we mean the user posting at least once in the community. In this sense, "communities" are explicitly pre-defined, but the joining behavior is temporary and requires little effort. In the previous studies of information diffusion in other social environments, such as LiveJournal and DBLP [1], or a recommendation referral program run by a large retailer [10], the relationships between people are explicit and the actions taken require more commitment. However, the relationships or links in most forum networks are hidden and implicit — there are no well-defined links such as friendship or affiliations [6]. The most obvious relationship among users in online forums is the reply relationship between users. Instead of reflecting strong friendship, the reasons people are linked together by online replies may be because of common interests or different opinions [28, 5, 6].

[*]This work was performed in part while the 1^{st} and the 2^{nd} authors were visiting Microsoft Research Asia.

KDD'09, June 28–July 1, 2009, Paris, France.
Copyright 2009 ACM 978-1-60558-495-9/09/06 ...$5.00.

In order to answer the first question, we analyze several features that can usually be obtained from a forum dataset. Our first discovery is that, despite the relative randomness and arbitrariness, the diffusion curve of influence from users of reply relationships has very similar diffusion patterns as [1], although the reasons that people are linked together are very different. We also investigate the influence of the features associated with communities, which include the size of communities and the authority or the interestingness of the information in the communities. We find that their corresponding information diffusion curves show some strong regularities of user joining behavior as well, and these curves are very different from those of reply relationships. Furthermore, we analyze the effects of *similarity of users* on the communities they join, and find two users who communicate more frequently or have more common friends are more likely to be in the same set of communities.

In order to answer the second question, we construct a bipartite graph, whose two sets of nodes are users and communities, to encompass all the features and their relationships in this problem. Based on the bipartite graph, we build a bipartite Markov Random Field (BiMRF) model to quantitatively evaluate how much each feature affects the grouping behavior in online forums, as well as their relationships with each other. BiMRF is a Markov random graph [4, 25] with edges and two-stars as its configuration, and incorporates the node-level features we have described as in a social selection model [17]. The most significant advantage of using the BiMRF model is that it can explicitly incorporate the dependency between different users' joining behavior, i.e., how a user's joining behavior is affected by her friends' joining behavior. In contrast, the decision trees as used in [1] cannot directly model such dependency. The results of this quantitative analysis shows that different features have different effectiveness in prediction in news forums versus technology forums. Together with results from the quanlitative analysis, we are able to answer the third question. Our work also suggest that BiMRF models can be applied to analyze bipartite networks that are used to represent people and the common membership they belong to in general.

The findings discovered in this paper are useful for improving and designing social network systems. Basically there are two important social functions in a social network system. One is how to recommend similar users, and the other one is how to recommend communities to users. The study of user grouping behavior reveals important features that have great impacts on how users join communities, and therefore provides valuable insights for social system owners to improve user experience. For example, a forum website can provide more social intelligence by recommending top rated posts or large communities to users. It can also remind a user to pay more attention to other users who share similar interests with him. The findings related to the differences between news forums and technology forums also suggest that social systems should be designed with more considerations of diversified user intentions.

The rest of this paper is structured as follows. Section 2 discusses some related work. Section 3 describes the datasets and network analysis results. Section 4 investigates dynamic features related to community-joining behavior. Section 5 presents the BiMRF model with quantitative analysis. Section 6 concludes this paper.

2. RELATED WORK

The relationship between the user behavior and their social environment is the focus of a large body of work recently, such as [3, 6, 20]. The behavior of grouping is particularly interesting in social networks because it is closely related to the topic of *information diffusion* or *epidemics* [23].

The work in [1] also studies the human behavior of group formation. However, our work differs from it in the following aspects. First, the forum data, which have loose structures and hidden relationships, are different from the two social networks studied in [1]. The relationships between two users and between a user and a community in LiveJournal and DBLP require high commitment. For example, related neighbors have to be real friends in LiveJournal or co-authors in DBLP. In contrast, both user-user and user-community relationships in forums are much weaker because users do not have to exert much effort to have reply relationships with other users or participate in communities online. Second, in addition to the diffusion curves of numbers of related users, our work also studies the diffusion curves of other forum features, the relationships between these features, and how the user behavior differs in news versus technology forums. Finally, instead of using decision trees [1], we use exponential random graph models, which can evaluate more complicated dependency features.

Another work studying user participation behavior is [9]. Instead of considering the relationship between users and communities, their target is to investigate the motivations of user participation on a social media site. The work [2] focuses on users who are heavily engaged in the group, and the behavioral differences between those users and ordinary users. They use a bipartite model to represent the user-group relationship; however, their model is to predict the "long-core" membership.

Users that do not participate publicly in online communities, i.e. people who lurk without posting, may be also interested in those communities. However, they are less positive in both activity and influence[15].

Exponential random graphs [18], which include the simplest Bernoulli random graph or Erdös-Renyi random graph model, Markov random graph [4, 25] and the recent developments [21, 19], have been extensively studied for social network analysis. Traditional use of random graph models is to discover structural statistics of networks, such as triangles and stars. Our work is an application of the homogeneous Markov random graph models [4, 25] with consideration of node-level attributes to give quantitative analysis of the forum data. BiMRF is a social selection model [17], in which individual users may change their joining behavior on the basis of the attributes of others.

3. OVERVIEW OF THE NETWORKS

In this section, we present an overview of the datasets and the bipartite networks of user-community relationships, as well as some structral features of these bipartite networks.

3.1 Datasets Description

The datasets we study are from four online forums or online discussion platforms: Digg[1], Apple Discussions[2], Google Earth Community[3], and Honda-tech[4].

Digg is a news aggregator website, where users can submit news, videos, and pictures. In addition to that, users are able to lead discussions about the content that they are passionate about. All posted items, including news, images,

[1] http://digg.com

[2] http://discussions.apple.com

[3] http://bbs.keyhole.com/ubb/ubbthreads.php/Cat/0

[4] http://www.honda-tech.com

videos and discussion comments can be rated by users by "digging" them. It is a platform on which people can provide content from anywhere on the web, and collectively determine the value of the information. The data we have crawled from Digg is from Oct., 2007 to Jul., 2008. It has 50 communities with topics of a great diversity. More than 200,000 users were active (i.e. posted at least once), and about 48,000 threads were built during that time period.

Unlike Digg, the other three forums focus on topics related to a specific product or technology. Apple Discussion is a platform mainly for Apple users seeking help, answering others' questions or exchanging opinions about Apple products. In our dataset, there are about 350,000 users and about the same number of threads in 331 different communities. The time window of this data ranges from 2001 to 2008. The forum of Google Earth holds discussions about the technology of Google Earth. Our dataset has about 700,000 threads in 54 different communities, and 230,000 users were active from May, 2003 to June, 2008. A fraction of the posts in the Google Earth forum had ratings with them. Finally, Honda-tech is a forum for Honda customers to provide and exchange information and resources. It had 86,000 threads and about 45,000 users from 2001 to 2008. There were 63 communities in this forum. All of the four forums have explicit reply relationships in the datasets we have crawled.

3.2 User-Community Bipartite Network

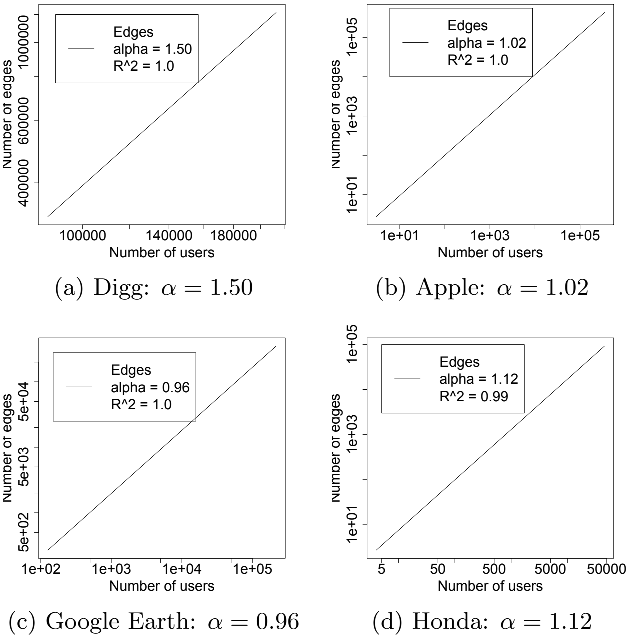
In social networks, bipartite networks or affiliation networks are bipartite graphs that are used to represent the people and the common memberships they belong to, such as the author-scientific article network, the actor-movie network [14]. In our problem, we define the user-community relationship as a bipartite graph: there is an edge between a user u and a community c, if and only if u has ever posted an article or a comment in c. Because little effort or commitment is required to post in online forums, the relationship between users and communities is not as strong as many other bipartite networks of user-membership. However, from the analysis of the bipartite networks, we are able to see some regularities of user joining patterns.

Table 1: Statistics about the bipartite networks.

Forum	Digg	Apple	Google Earth	Honda
User	212,635	349,066	231,976	45,718
Commu.	50	331	54	63
Edge	1,185,167	451,338	345,038	122,946
$\langle k_u \rangle$	5.57	1.29	1.49	2.69
$\langle k_c \rangle$	23703.34	1367.69	6389.59	1951.52
r	-0.2169	-0.0888	-0.2271	-0.0578

Table 1 gives a basic description of the user-community bipartite networks constructed from our forum datasets. $\langle k_u \rangle$ is the average number of communities a user joins, while $\langle k_c \rangle$ is the average number of users a community has. From the values of $\langle k_u \rangle$ and $\langle k_c \rangle$, we see the bipartite graph of Digg is much denser than the other three. This shows that in news forums such as Digg, users are more likely to join multiple communities than in technology forums. r is the value of assortativity, whose concept is defined as the preference of the nodes in a network to have edges with others that are similar under certain measurement [13]. Here we measure similarity with regard to degrees of nodes in the bipartite graph, and get the Pearson correlation coefficient between the degree of the users and the degree of the communities. We see that all four bipartite networks show negative values of r, which implies that in forums, less active users are more likely to join popular communities, while less popular communities are mostly occupied by active users.

We then examine the growth of edges versus the growth of users in the bipartite networks of forums, by looking at whether their α of $e(t) \propto n(t)^{\alpha}$ follow the densification law [11] . In our bipartite networks, we assume that all communities existed since the beginning of our data availability, and that users start to join since the time they had their first post. From Figure 1, we see that the growth of edges is almost linear with respect to the numbers of nodes in the bipartite graphs of the four datasets. Being consistent with their low average degrees of users $\langle k_u \rangle$, this tells us that most users in the technology forums have much more focused interests and mostly stay in single communities. However, this is not the case for the forum of Digg, whose α is 1.5. In fact, we find that there are quite a few users who join almost all of the communities in this forum site.



(a) Digg: $\alpha = 1.50$

(b) Apple: $\alpha = 1.02$

(c) Google Earth: $\alpha = 0.96$

(d) Honda: $\alpha = 1.12$

Figure 1: The growth of edges versus the growth of users in the bipartite networks.

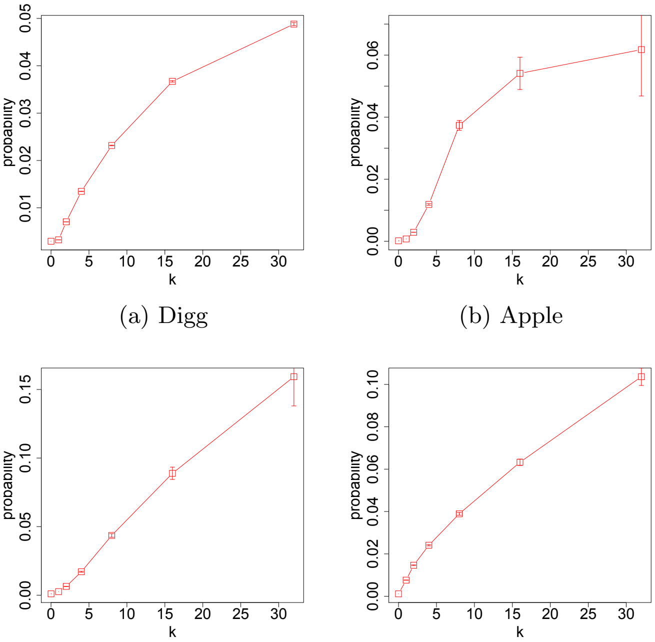
4. COMMUNITY MEMBERSHIP

In the previous section, we analyzed some structural features of the online forum networks. In this section, we study the process of community joining behavior directly. In order to see the dynamics of user behavior, we divide the datasets into 30 time snapshots. The diffusion curves we examine are the relationships between user joining behavior at time t and the related features at the previous time snapshot $t - 1$. These curves show the change of joining probabilities as functions of different features associated with either users or communities. Moreover, we also study the correlations of user similarities and the communities they join.

4.1 Friends of Reply Relationship

We use this feature to describe how users are influenced by the numbers of neighbors with whom they have ever had any reply relationship. Although the reply relationship is not exactly the same as a real friendship, this is usually the most common and explicit user-user relationship that can be extracted from a forum dataset. In addition, as we will show, the reply relationship exhibits similar patterns in its diffusion curves as those of stronger relationships in other social networks, such as friendship or co-authorship in [1].

For every tuple (u, c, t) of user-community relationship at time t, we look at the reply friends of u who were active in c at the previous time snapshot $t-1$. We denote the number of such reply friends as k. By observing all the cases of whether u joins c with k reply friends at the previous time snapshot, we get the joining probability as a function of k. From Figure 2, we see that all four curves exhibit the *law of diminishing returns*. That is, the curves increase fast at the beginning, but more and more slowly towards the end. This is highly consistent with the observations of information diffusion in some other social networks [1, 10]. Moreover, the "S-shaped" behavior at $k = 0, 1, 2$ described in [1] is also observed in the three large datasets, Digg, Apple and Google Earth. The absence of this behavior in Honda may be because of the significantly smaller size of this dataset.



(a) Digg

(b) Apple

(c) Google Earth

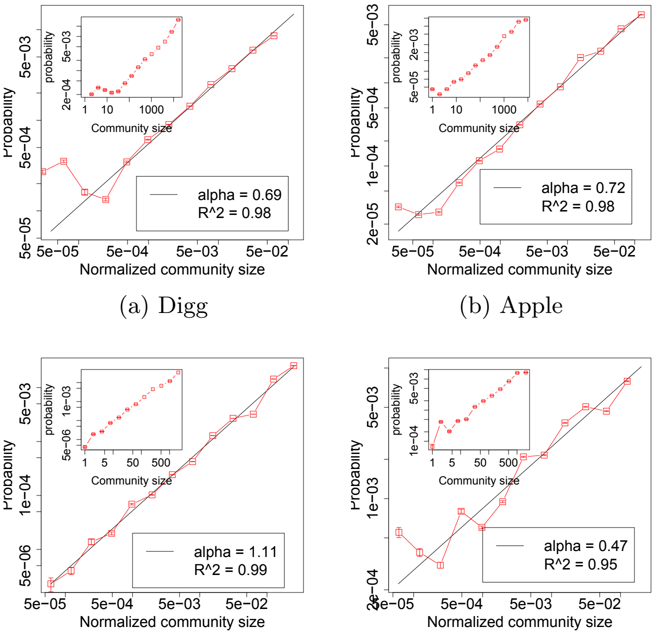
(d) Honda

Figure 2: The probability of a user joining a community in the forum as a function of the number of reply friend k who are active in that community at the previous time snapshot.

An alternative way to connect users is to let users in the same thread form a clique or a complete subgraph. However, this is a looser relationship than the reply relationship because users in the same thread may be interested in different aspects of the thread topic [28]. In fact, we also observe similar diffusion curves when considering the users in the same threads as 'friends', although the probability values are much lower. This is interesting since it suggests that in many social networks, despite the diversity of 'friendships', their diffusion curves may have very similar patterns.

4.2 Community Sizes

It is intuitive to expect that more popular information diffuses among the network at a faster pace. We examine this hypothesis in this part. We use *community size* as the measurement to quantify the 'popularity' of information.



(a) Digg

(b) Apple

(c) Google Earth

(d) Honda

Figure 3: The probability of a user joining a community in the forum as a function of the normalized community size at the previous time snapshot. The insets show the probability before normalization.

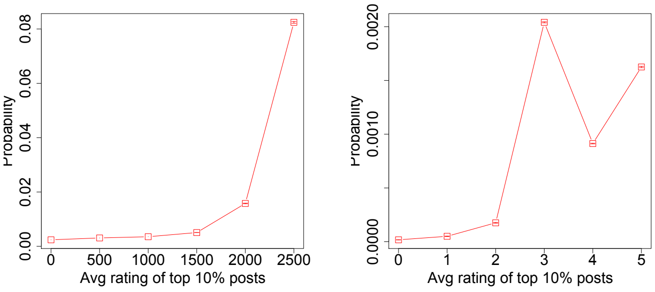
By the *community size* at a time snapshot, we mean the number of users who have posted at least one article or comment in that community during that time snapshot. We call these users *active users*. The total sizes of all the communities at different time snapshots vary a lot, which may be because of both the limitation of the datasets and the effect of exponential growth of social communities. So we further normalize the community size over the sum of the sizes of all communities at that time snapshot.

Similar to the diffusion curves of the reply relationship, for every user-community tuple at each time snapshot t, we look at the normalized size of the community at time $t-1$, and get the user joining probability as a function of it. The curves are shown in Figure 3. The insets are the diffusion curves of absolute community size. From the figure, we can see that all the curves can be fitted by straight lines in the log-log scale. That is, if we use p and s to denote the probability of joining and normalized community size respectively, we have $p \propto s^{\alpha}$. We find that α is less than 1 in three of the figures, and larger but close to 1 in Google Earth. This tells us that the growth of the joining probability is sub-linear or linear with respect to the normalized community size.

4.3 Average Ratings of Top Posts

Aside from the popularity of information, we are also interested in how the *authority* or *interestingness* of information impacts user behavior. Usually, in a social environment such as forums, the evaluation of the authority or interestingness of information is the result of the *wisdom of crowds*, since the ratings are the cumulative results of the users. In our datasets, Digg and Google Earth have rating systems, but their rating systems have some differences. First of all, the range of the ratings in Google Earth is from 0 to 5, while there is no upper bound of the ratings in Digg since those ratings are just the number of times a post has been "digged" by the users who like it. So the influence from the ratings in Digg may be confounded with the influence of community size, while Google Earth does not. Moreover, Digg allows ratings on starting posts as well as replies; while Google Earth only allows ratings on starting posts.

Users usually only see the ratings of starting posts before reading more of a thread. So we only consider the ratings of starting posts as evidence for the authority of information. What is more, the data shows that in a community, the distribution of the ratings of starting posts is highly skewed, with most starting posts having very low scores and only a small fraction of them having high scores. Based on this fact, we choose the posts with top 10% ratings in each community at every time snapshot, and get the average of the ratings. Similar to the analysis of the previous two features, we plot the probability for users joining a community at a time snapshot as a function of the average rating of the top 10% posts in the community at the previous time snapshot. Figure 4 shows the resulting curves.



(a) Digg (b) Google Earth

Figure 4: The probability of a user joining a community as a function of the average rating of the top 10% high rating posts in the community at the previous time snapshot.

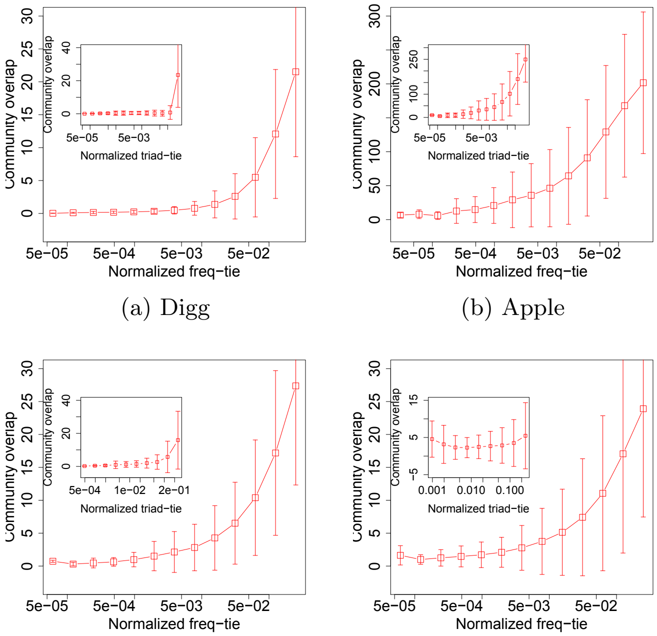
It is interesting to see that there is a smoothly increasing curve for Digg, and the curve grows much faster after the average rating reaching a point around 2000. This curve shows a pattern that is called *critical mass*. On the other hand, the curve for Google Earth does not consistently increase as the one of Digg does. But still, the probability is much higher when the average rating is at 3, 4 or 5 than that of when the average rating is at 0, 1, or 2. This difference between Digg and Google Earth might be due to people's different purposes in the two types of forums. In Google Earth, people are mainly seeking answers to their particular questions that may be only related to the topics in limited communities, so although the scores of the posts in the community matter, they do not have much difference after a threshold. However, the purposes people have for joining communities in Digg are more diverse. In addition, the front page of Digg enables users to read interesting topics without being aware of the communities they are in [9]. So increasing interestingness of the posts may be able to attract more users.

4.4 Similarities of Users

In the previous part, we have studied how certain features affect the probability of users joining communities. Those features are associated with either a single user or a single community. In this part, we analyze the features with dependency: if two users are 'similar' in a certain way, what is the correlation of the sets of communities they join?

To define the 'similarities' of users, two criteria are used. The first one is the number of times two users reply to each other's posts, normalized over the total number of articles or comments the two users have posted. The second one is the number of common friends that the two users have in the reply network, normalized over a half of the sum of the numbers of friends the two users have in total. For easy reference, we will name these two types of user similarity *frequency-user-similarity* and *triad-user-similarity* respectively. Each similarity measures takes values between 0 and 1. In order to get rid of noise introduced by trivial behavior, all users who only post once are ignored.

In order to know whether more similar users are more likely to join the same communities, we compare their similarities versus the overlaps of communities they have joined. For two users u_1 and u_2, let the sets of the communities they have joined be S_1 and S_2, and the absolute overlap of their communities be $S_1 \cap S_2$. However, we need to account for the fact that some users may have little 'similarity' but large community overlap because they participate in almost all of the communities in a forum. So we normalize the absolute overlap by the expected overlap. The expected overlap can be obtained as $O_e = (|S_1| \times |S_2|)/(|S|)$, where S is the set of all communities in the forum. Then the normalized overlap O_n can be got by the equation: $O_n = \frac{(|S_1 \cap S_2| - O_e)}{O_e}$.



(a) Digg (b) Apple

(c) Google Earth (d) Honda

Figure 5: The user similarities versus the community overlaps. The main plots use the communication frequency between users as the user similarity, and the insets use the number of common friends.

Figure 5 shows the relationships between user similarities and the normalized community overlaps. The correlation is positive for all forums and similarity measures, which means that more similar users are more likely to be in the same communities. We have to note that Figure 5 only shows the static correlations of user similarities and their community overlaps. This is different from the dynamic diffusion curves that we see in Figure 2 - 4. In fact, by computing the correlations between the user similarity at time $t-1$ and their community overlap at time t, we find they are neutrally correlated. This means two users either communicating more frequently or having more common reply friends at certain time are not more likely to join the same new communities in the following time snapshot. We will use a statistical model to further investigate this problem in Section 5.

4.5 Summary

In this section, we have shown how the community joining behavior is influenced by features associated with users and communities. The empirical diffusion curves show that these features are affecting human behavior in various ways. It is particularly interesting to see that the feature of reply friend has similar diffusion curves as those of real friend relationships in other types of social networks.

Moreover, we have analyzed the features of dependency. User-user similarities defined by their frequencies of communications and numbers of common friends are both positively correlated with the overlaps of the communities that the users have joined. However, there is no correlation between the user similarity and the sets of communities the users are going to join.

So far, we have examined the features separately. We will now consider them together to answer such questions as which feature best predicts user behavior and what correlations can be made with multiple features. We use a bipartite Markov random field model to study these problems.

5. STATISTICAL USER GROUPING MODEL

In this section, we present a bipartite MRF (BiMRF) model, also known as a social selection random graph [17], to examine the quantitative effects of different features on the user grouping behavior in online forums. In addition to predicting user behavior based on the features observed, these models help reveal relationships between the features. Based on these relationships, we observe that the features have different effects in information diffusion in news and technology forums. As we shall see in Section 5.2, the advantage of using BiMRF models in our problem is that they can explicitly incorporate the dependency between related users' joining behavior, i.e. how a user's joining behavior affects her friends' joining behavior. The decision tree as used in [1] cannot explicitly model such dependency.

5.1 Bipartite Markov Random Fields

In social network analysis, exponential random graph ($p^\star$) models have been extensively studied, including the simplest Bernolli random graph or the Erdös-Renyi model and the Markov random graph [4, 25] and its new specifications [21, 19]. In machine learning society, a Markov random graph is a Markov random field (MRF) with edges represented as random variables. In the sequel, we will obey this convention and point its connection to random graph models.

Based on the bipartite networks we have described in Section 3.2, we define the bipartite MRF (BiMRF) as follows. BiMRF is a bipartite graph and the vertices at one side are associated with the variables $U = \{U_i\}_{i=1}^{M}$ which represent users, and the vertices at the other side are associated with variables $C = \{C_j\}_{j=1}^{N}$ which represent communities. In the same spirit as the previous analysis, given the observed features, we treat the joining behavior at different time snapshots independently in the BiMRF model. Figure 6 shows the model's graph at time t. We will use the tuple (u, c, t) to denote the user-community relationship at time t. In our model, each user is a d-dimensional feature vector $u_i = [u_{i1}, \ldots, u_{id}]^\top$, of which the feature values can change over the time t as we have discussed in previous sections. Different users can be connected together, for example, if their similarity (by some measurement) is above some threshold. Each community c_j can have its features (e.g. community sizes) and can also connect to other communities if we have similarity defined between them, and their similarity is large enough. We will use O to denote all the observations, including users and their features, communities and their features, and connection structure of users and of communities. We introduce a set of random variables $E = \{E_{ij}^t : 1 \le i \le N\ ,\ 1 \le j \le M\ and\ 1 \le t \le T\}$, and each r.v. is an indicator: $e_{ij}^t = 1$ if the user u_i joins the community c_j at time t; otherwise, it is 0. Let $\{e\}$ denote an instance of the random variables E. By the basic theory of random fields [8], given the observations O, BiMRF defines a conditional distribution as follows:

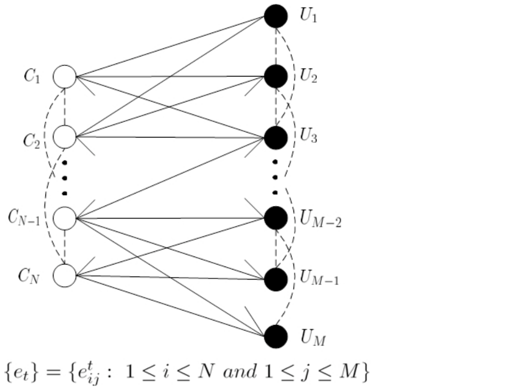


Figure 6: A bipartite MRF model with N communities and M users at time t. $\{e_t\}$ is an instance of the connections between users and communities at time t. The dashed edges are observed evidence.

$$p(\{e\}|O) = \frac{1}{Z(\mathbf{w})} \exp\Big(\sum_{k=1}^{K} w_k f_k(\{e\}, O)\Big)$$

where f_k are feature functions, which can be real or binary (here we assume they are binary, i.e. true or false), and w_k are their weights, which will be learned from a given training dataset. As we have mentioned, BiMRF treats the joining behavior at different time snapshots independently given the observed features. Thus, $p(\{e\}|O) = \prod_{t=1}^{T} p(\{e_t\}|O)$.

Since the dashed edges in Figure 6 are fixed and the probability $p(\{e\}|O)$ is defined on the connections between nodes on different sides, we call the model as a Bipartite MRF. A dashed edge is added if the similarity of the two users or the two communities at either side is above some threshold.

5.2 Feature Function Definition

We now define the feature functions for modeling user-community behavior. The features we use in BiMRF include three singleton features and two types of user similarity. The *singleton features* are those either associated with users or communities. They are *reply-friend*, *normalized-community-size* and *top-post-rating* in our models, and will be denoted as rf, cs, tp respectively. The two *dependency features* are the two types of user similarity, i.e. *frequency-user-similarity* and *triad-user-similarity*, and we use us_f and us_t to denote them. We use the same bins as those used in the analysis of Section 4, i.e. we take linear-bin to define the feature function of *top-post-rating*, and the log-bin to define the other feature functions. Suppose the basis of the logarithm is b (e.g., 2 in our model). Two representative feature functions are defined in Table 2. Since f_k^{cs} are defined on an individual (u, c, t) tuple (i.e., a single joining event), we call these feature functions *singleton* feature functions; while f_k^{us} are defined on more than one joining events, thus we call these feature functions *dependency* feature functions. These dependency feature functions explicitly model the dependency between different users' joining behavior. In decision trees [1], however, such dependency cannot be directly modeled.

To avoid functions which appear sparsely in the datasets,

Table 2: The two representative feature functions in BiMRF. cs **denotes the features of** *normalized community-size* **and** us **denotes the two types of user similarity who are the same in defining feature functions.**

Categories	Features	Feature functions
Singleton	*community-size*	$f_k^{cs}(e_{ij}^t = 1, c_i, u_j, t) = \begin{cases} 1 & \text{if } b^{-k} \leq \text{CommuSize}(c_i, t) \leq b^{-k+1} \\ 0 & \text{otherwise} \end{cases}$
Dependency	*user similarity* *(frequency or triad)*	$f_k^{us}(e_{ij}^t = 1, e_{il} = 1, c_i, u_j, u_l, t) = \begin{cases} 1 & \text{if } b^{-k} \leq \text{UserSim}(u_j, u_l, t) \leq b^{-k+1} \\ 0 & \text{otherwise} \end{cases}$

we set an upper bound (e.g., 512) for the reply-friend feature and a lower bound (e.g., 2^{-19}) for the other four features. The features that are beyond this bound are defined in one feature function and will be treated the same in BiMRF.

These feature functions have a close connection to the configurations in Markov random graph [4, 25]. The singleton feature functions correspond to the dyad configuration and the dependency functions correspond to two-star configurations. In each type of configuration, we consider node-level attributes as in a social selection model [17].

5.3 Model Fitting and Testing

Model fitting is to learn the parameters from a given dataset. In this case, the data set is a pairing of observations and edges, i.e., $\mathcal{D} = \{\langle\{e\}, O\rangle\}$. The best model to fit the data is the one with the maximum conditional likelihood: $\mathcal{L} = \log p(\{e\}|O)$. The optimization problem can be done with gradient ascent methods, such as the L-BFGS [12]. Since the probability is an exponential family distribution, the gradient is: $\frac{\partial \mathcal{L}}{\partial w_k} = E_{\hat{p}}[f_k] - E_p[f_k]$, where $E_p[.]$ is the expectation with respect to the model distribution $p(\{e\}|O)$, and $E_{\hat{p}}[.]$ is the expectation with respect to the empirical distribution on the given data corpus.

Without *dependency* feature functions, the Bipartite MRF models reduce to logistic regression models, also known as Bernoulli graphs in social network society. In this case, the model distribution, or the marginal probabilities as required in the objective function and its gradients, can be easily computed for each (u, c, t) independently.

With *dependency* feature functions, the BiMRF model is a homogeneous Markov random graph [4, 25] with two-star configurations. In each configuration, we consider node-level attributes as in a social selection model [17]. In a Markov random graph, the marginal probabilities on different edges, i.e., different (u, c, t) tuples, are coupled together. In other words, the event that a user joins a community at a particular time depends on the joining events of the related users or the communities at that time. Thus, we cannot compute the marginal probabilities of different edges independently.

For Markov random graphs, various estimation methods have been studied in social networks, such as the pseudo-likelihood method [24] and the Monte Carlo maximum likelihood estimation [26]. In this paper, we use variational methods [7], which are among the most popular inference methods in the graphical model literature. The mean field approximation bears the form of pseudo-likelihood function [24]. But unlike the pseudo-likelihood method, mean field marginal probabilities are computed iteratively using the coupled mean filed equations given initial values.

5.3.1 Mean Field Inference

To illustrate how mean field inference works in BiMRF, we use the user similarity feature as an example. The following derivations can be easily extended to other BiMRF models. The BiMRF model defines the following joint distribution:

$$p(\{e\}|O) \propto \exp\Big\{ \sum_{k=1}^{K^{us}} w_k^{us} \sum_{ijlt} f_k^{us}(e_{ij}^t, e_{il}^t, c_i, u_j, u_l, t) \Big\}$$

We define the factorized variational distribution $q(\{e\}|O) = \prod_{ijt} q(e_{ij}^t|O)$ as an approximation to the joint distribution. To find the best approximation $q^\star$, we minimize the KL-divergence: $KL(q(\{e\}|O)||p(\{e\}|O))$. The optimization problem can be solved by an alternating minimization method. Specifically, at each step we solve the problem with respect to only one marginal distribution $q(e_{ij}^t)$ and keep all others fixed. Then, we can get the following coupled mean field equations by using q_{ij}^t to denote $q(e_{ij}^t = 1|O)$:

$$q_{ij}^t \propto \exp\Big(\sum_{k=1}^{K^{us}} w_k^{us} \sum_l q_{il}^t f_k^{us}(e_{ij}^t = 1, e_{il}^t = 1, c_i, u_j, u_l, t)\Big)$$

These coupled mean field equations reflect our intuition that the event that user j joins community i at time t is dependent on whether other connected users l join the community at that particular time. We iteratively solve the coupled equations to get a fixed point solution, which gives the (approximate) marginal probabilities.

Table 3: Distributions of the number of related users on different datasets for frequency-user-similarity.

#Related Users	Digg	Google Earth	Apple	Honda
≤ 20	63.93	97.03	96.81	69.80
≤ 40	75.89	98.80	98.51	81.05
≤ 60	82.30	99.29	98.98	86.30
≤ 100	88.40	99.59	99.34	91.48

5.3.2 Prediction

Given a learned model, we can do prediction on unseen (u_j, c_i, t) tuples and get the marginal probability that an edge exists $p(e_{ij}^t = 1|O)$. This is the probability that the user u_j joins the community c_i at time t. Since joining events are rare, the probabilities $p(e_{ij}^t = 1|O)$ are much smaller than 0.5. We cannot use a threshold (e.g., 0.5) to decide whether a user joins a community. Instead, we use the ordering metrics ROC Area (ROCA) and Average Precision (AP) to evaluate the goodness of the models. We evaluate the results of the features individually as well as with different combinations. In each experiment, we randomly sample 70 percent of the (u, c, t) tuples as training data and predict on the rest in each dataset.

An issue with regard to the models with user-similarity features is that we need to take care of the large number of related users as defined by user-similarity. For example, for the frequency-user-similarity, the maximum number of related users on Digg is 43,269, 15,205 on Apple, 3740 on Google Earth, and 2236 on Honda. These large numbers will destabilize the computation when performing mean field inference. Fortunately, as shown in Table 3, for frequency-user-similarity, most of the users have small numbers of related users. The case of the triad-user-similarity is similar. Thus, we can use a pruning method to remove those rare users who have a large number of related users. In this experiment, we apply a simple strategy. We remove a user's user-similarity features if the number of her related users is larger than K, which is a pre-specified parameter. We set K

Table 4: Evaluation results of different BiMRF models on the four datasets.

BiMRF Models	Digg		Google Earth		Apple		Honda	
	ROCA	AP	ROCA	AP	ROCA	AP	ROCA	AP
$\{cs\}$	0.700	0.00536	0.860	0.00697	0.912	0.00296	0.833	0.00542
$\{rf\}$	0.718	0.00922	0.520	0.00128	0.522	0.00025	0.640	0.00743
$\{cs, rf\}$	0.800	0.01295	0.862	0.00738	0.913	0.00310	0.853	0.01257
$\{us_f\}$	0.442	0.00271	0.477	0.00188	0.473	0.00014	0.467	0.00147
$\{us_t\}$	0.474	0.00911	0.467	0.00235	0.467	0.00018	0.483	0.00179
$\{us_f, cs\}$	0.699	0.00540	0.861	0.00734	0.912	0.00296	0.831	0.00542
$\{us_t, cs\}$	0.705	0.00551	0.860	0.00698	0.912	0.00296	0.832	0.00536
$\{us_f, rf\}$	0.570	0.00362	0.545	0.00122	0.532	0.00015	0.561	0.00162
$\{us_t, rf\}$	0.703	0.00708	0.526	0.00117	0.531	0.00015	0.588	0.00179
$\{us_f, cs, rf\}$	0.796	0.01276	0.861	0.00744	0.899	0.00295	0.851	0.01248
$\{us_t, cs, rf\}$	0.800	0.01301	0.862	0.00724	0.906	0.00307	0.853	0.01177

at 20 in our experiments. We tried other parameters (e.g., 40 or 60), and the results do not change much.

5.4 Observations

Singleton features. In Section 4 we have seen the diffusion curves related to reply friends and community sizes, however, we cannot compare their effectiveness in predicting the user joining behavior from those curves. The BiMRF models help us do this. From the first two rows of Table 4, we see that for Google Earth, Apple, and Honda, the *community-size* feature predicts user joining behavior much better than *reply-friend* does. In particular, *reply-friend* has very little effect in Google Earth and Apple (their ROCA values are around 0.5). In contrast, *reply-friend* performs slightly better than *community-size* in Digg. Furthermore, by comparing the first three rows of Table 5, we see that although *top-post-rating* performs worse than *community-size* in Digg and Google Earth, it is better than *reply-friend* in Google Earth while worse than *reply-friend* in Digg.

These observations suggest that in the three technology forums, users' joining behavior correlates more closely with the features associated with communities, such as community sizes and average ratings of the top posts in the communities, rather than the number of reply friends of users. On the other hand, in a news forum such as Digg, the user behavior has a stronger correlation with the number of their reply friends. The possible reasons for this difference are as follows. First, in both Google Earth and Apple, about 53% - 54% of users have only one post, while there are about 27% such users in Honda and 33% such users in Digg. This may explain the relatively poor performance of *reply-friend* in Google Earth and Apple, since there are large fractions of users who do not have any reply friends before joining any community, and they do not have any further activity after getting some reply friends. Second, from the low average degrees of users and the almost linear growth of edges versus the users in Section 3, we know that most users in the three technology forums like to stay in one community from the time they joined, i.e. they do not tend to switch their focus or expand their interests to different communities. In this way, the properties of the communities are more essential for users to decide which community to join at the very beginning, because no matter how many reply friends they gain, it is not likely for them to follow their reply friend to other communities. However, users do not have such focused interests in Digg, so their interests are more likely to change to other communities as their reply friends do.

Table 4 and 5 list the main results of the models using different features. Table 5 shows the results related to *top-post-rating*, which appears only in Digg and Google Earth. From the quantitative measures in these two tables, we make several observations regarding the features.

Table 5: Evaluation results of the top-post-rating, and user-similarity on Digg and Google Earth.

BiMRF Models	Digg		Google Earth	
	ROCA	AP	ROCA	AP
$\{tp\}$	0.639	0.00404	0.760	0.00229
$\{cs\}$	0.700	0.00536	0.860	0.00697
$\{rf\}$	0.718	0.00922	0.520	0.00128
$\{tp, cs\}$	0.708	0.00568	0.882	0.01040
$\{tp, rf\}$	0.774	0.01155	0.765	0.00250
$\{tp, cs, rf\}$	0.804	0.01371	0.884	0.01080
$\{tp, us_f\}$	0.642	0.00418	0.765	0.00236
$\{tp, us_t\}$	0.647	0.00454	0.761	0.00230
$\{tp, cs, rf, us_f\}$	0.802	0.01378	0.885	0.01044
$\{tp, cs, rf, us_t\}$	0.804	0.01375	0.883	0.01075

Dependency features. The results (the fourth and fifth rows of Table 4) of BiMRF models using two user similarities tell us that these dependency features perform poorly in predicting, e.g., their ROCA scores are all below 0.5. Note that although Figure 5 shows that there are positive correlations between the similarities of users and the overlaps of the communities they belong to, those correlations are static and do not reflect the dynamic relationship of the users' similarities and their future joining behavior. And our analysis in Section 4 gets the neutral correlations between the user similarities in a time snapshot and the overlaps of communities they are going to join in the next time snapshot.

By a close examination of the joining probabilities predicted by the BiMRF models, we see that many (u, c, t) tuples have a probability larger than 0.1, which is much larger than the average joining probability in the datasets. This means that the BiMRF models with only the dependency features, which correspond to two-star configurations in Markov random graphs, are inadequate for the online forum data. But these models can be improved by incorporating node-level attributes, as shown by the results of BiMRF models with both dependency and singleton features in Table 4 and 5. This suggests that user-similarity has a weak effect on joining behavior in online forums, and thus adding the dependency features of user similarity does not help improve the performance. Finally, we must point out that the naïve mean field we are using in BiMRF makes a very strong independence assumption about the variational distribution q. This may give a poor approximation to the true distribution. Extending to the generalized mean field [27], which incorporates more structural dependency in q, could be helpful to get a better approximation.

Feature combinations. By combining all the singleton features (as in the third row of Table 4 and the sixth row of Table 5), we see that the results are significantly better. This is especially true for Digg. Thus we can conclude that all these three singleton features carry supplementary information with each other, although in the three technology forums, *community-size* outperforms other two significantly.

6. CONCLUSIONS AND DISCUSSIONS

In this paper, we investigated the user participation behavior in diverse online forums. Our study of the structural features of their user-community bipartite networks suggest that, compared with news forums, users' interests in technology forums are more focused in single communities instead of crossing communities. Moreover, the diffusion curves show how the features of reply friends and some attributes associated with the communities have influence on community joining. Although a reply friendship is a much looser relationship [6], it has similar diffusion curves of *diminishing returns* as real friendship and co-authorship in [1]. Furthermore, the statistical BiMRF models present some interesting relationships among these features. In particular, reply friend and community attributes have about the same effectiveness in prediction in the news forum, while in the other three technology forums, features associated with communities are more effective in prediction. These features also provide supplementary information in our model. Finally, our analysis of two-star dependency social selection models suggests that the weak user-similarity features cannot fit the forum data well by themselves and adding node-level features can improve the fit.

As our analysis shows that user preference of information is tied with their related users in the past, and different types of information attract users in different ways, our work provides suggestions on building social systems, such as personalized recommendation systems [16]. Moreover, using the methodology presented in this paper, more detailed studies can be conducted to evaluate other features that may affect users' social behaviors. For example, the user interactions in our study is based on their explicit reply information. In fact, similar analysis can be done based on more hidden behaivors such as browsing, which is known to website owners. They can then use the insights gained from such data to inform their recommendations to users who lurk without posting. Utilizing textual analysis in forum data, and investigating user behavior related to diffusion of discussion topics is also a future direction.

7. ACKNOWLEDGEMENTS

We would like to thank Lada Adamic for many insightful suggestions and discussions. We also would like to thank Joseph Z. Xu for lots of valuable comments and help. Partial Support for this work was provided by MURI award FA9550-08-1-0265 from the Air Force Office of Scientific Research.

8. REFERENCES

[1] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *SIGKDD*, 2006.

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

[3] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *SIGKDD*, 2008.

[4] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81:832–842, 1986.

[5] K.-I. Goh, Y.-H. Eom, H. Jeong, B. Kahng, and D. Kim. Structure and evolution of online social relationships: Heterogeneity in unrestricted discussions. *Physical Review E*, 73(6):066123, 2006.

[6] V. Gómez, A. Kaltenbrunner, and V. López. Statistical analysis of the social network and discussion threads in slashdot. In *WWW*, 2008.

[7] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul. *An introduction to variational methods for graphical models*. M. I. Jordan (Ed.), Learning in Graphical Models, Cambridge: MIT Press, Cambridge, MA, 1999.

[8] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001.

[9] K. Lerman. User participation in social media: Digg study. In *SMA07*, 2007.

[10] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC*, 2006.

[11] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *SIGKDD*, 2005.

[12] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, (45):503–528, 1989.

[13] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003.

[14] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *the National Academy of Sciences*, 99:2566–2572, 2002.

[15] B. Nonnecke, D. Andrews, and J. Preece. Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1):7–20, 2006.

[16] J. M. O'Brien. The race to create a 'smart' google. *FORTUNE Magazine*, 2006.

[17] G. Robins, P. Elliott, and P. Pattison. Network models for social selection processes. *Social Networks*, 23:1–30, 2001.

[18] G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph ($p^\star$) model for social networks. *Social Networks*, 29:173–191, 2007.

[19] G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison. Recent development in exponential random graph ($p^\star$) model for social networks. *Social Networks*, 29:192–215, 2007.

[20] P. Singla and M. Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In *WWW*, 2008.

[21] T. A. B. Snijders, G. L. Robins, and M. S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, pages 99–153, 2006.

[22] X. Song, B. L. Tseng, C.-Y. Lin, and M.-T. Sun. Personalized recommendation driven by information flow. In *SIGIR*, 2006.

[23] D. Strang and S. A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24:265–290, 1998.

[24] D. Strauss and M. Ikeda. Pseudo-likelihood estimation for social networks. *Journal of the American Statistical Association*, 85:204–212, 1990.

[25] S. Wasserman and P. Pattison. Logit models and logistic regressions for social networks, I. an introduction to exponential random graphs and ($p^\star$). *Psychometrika*, 62:401–425, 1996.

[26] S. Wasserman and G. Robins. An introduction to random graphs, dependence graphs, and $p^\star$. *Models and Methods in Social Network Analysis*, pages 148–161, 2005.

[27] E. P. Xing, M. I. Jordan, and S. Russell. A generalized mean field algorithm for variational inference in exponential families. In *UAI*, 2003.

[28] K. Zhongbao and Z. Changshui. Reply networks on a bulletin board system. *Phys. Rev. E*, 67(3):036117, 2003.