Exploring Social Influence via Posterior Effect of Word-of-Mouth Recommendations

Junming Huang
Institute of Computing
Technology, Chinese Academy
of Sciences
Beijing, China
mail@huangjunming.com

Xue-Qi Cheng
Institute of Computing
Technology, Chinese Academy
of Sciences
Beijing, China
cxq@ict.ac.cn

Hua-Wei Shen
Institute of Computing
Technology, Chinese Academy
of Sciences
Beijing, China
shenhuawei@ict.ac.cn

Tao Zhou
Web Sciences Center,
University of Electronic
Science and Technology of
China
Chengdu, China
zhutou@ustc.edu

Xiaolong Jin
Institute of Computing
Technology, Chinese Academy
of Sciences
Beijing, China
jinxiaolong@ict.ac.cn

ABSTRACT

Word-of-mouth has proven an effective strategy for promoting products through social relations. Particularly, existing studies have convincingly demonstrated that word-of-mouth recommendations can boost users' prior expectation and hence encourage them to adopt a certain innovation, such as buying a book or watching a movie. However, less attention has been paid to studying the posterior effect of word-of-mouth recommendations, i.e., whether or not word-of-mouth recommendations can influence users' posterior e-valuation on the products or services recommended to them, the answer to which is critical to estimating user satisfaction when proposing a word-of-mouth marketing strategy. In order to fill this gap, in this paper we empirically study the above issue and verify that word-of-mouth recommendations are strongly associated with users' posterior evaluation. Through elaborately designed statistical hypothesis tests we prove the causality that word-of-mouth recommendations directly prompt the posterior evaluation of receivers. Finally, we propose a method for investigating users' social influence, namely, their ability to affect followers' posterior evaluation via word-of-mouth recommendations, by examining the number of their followers and their sensitivity of discovering good items. The experimental results on real datasets show that our method can successfully identify 78% influential friends with strong social influence.

Categories and Subject Descriptors

J.4 [**SOCIAL AND BEHAVIORAL SCIENCES**]: Sociology; H.1.2 [**MODELS AND PRINCIPLES**]: User/Machine Systems—*Human factors*

General Terms

Design, Human Factors, Verification

Keywords

social influence, word-of-mouth recommendation, posterior evaluation

1. INTRODUCTION

Word-of-mouth recommendation, the process that an individual passes personal attitude to another (e.g., encouraging a potential customer to buy a product), has proven in decades to be an effective way to influence the latter's decision making. It was found that, for example, in the shopping site Amazon[1] an increase in book reviews, which is a kind of online word-of-mouth recommendations, results in an increase in relative sales [7]. Consequently, word-of-mouth marketing has been quickly developing via a variety of media including messages, blogs, microblogs and emails. In recent years, the integration of social network services and online sharing communities has gained a great success on a large number of web sites, such as Douban[2] and Goodreads[3], where millions of users cannot only rate books, music and movies, but also share their ratings and reviews with friends or followers. On those social network sites, the phenomenon that users recommend favorites to friends and followers plays an important role in shaping users' behaviors and their collections. Therefore, a deep understanding of word-of-mouth recommendations could be a key to designing a successful marketing strategy in online sharing communities.

---

[1] www.Amazon.com
[2] www.Douban.com
[3] www.Goodreads.com

The pervasive usage of word-of-mouth leads to a highly interesting research topic to understand its intrinsics. In decades various models, such as linear threshold, cascade model and spreading model, have been developed to explain why and how a word-of-mouth message encourages a receiver to adopt a certain innovation (e.g., a new product). It is widely believed that the reason encouraging a receiver to adopt an innovation is that a word-of-mouth message can raise his/her *prior expectation* on the concerned innovation, i.e., how valuable, useful and usable he/she expects the product to be. For example, when facing the question "to buy a BluRay player or an HD-DVD player", a person would more likely choose the former if many of his/her friends have already possessed BluRay players, which often provides a strong evidence in the quality of BluRay players.

However, less attention has been paid to investigating the *posterior effect* of word-of-mouth recommendations on a user's *posterior evaluation*, i.e., whether or not a word-of-mouth recommendation influences the evaluation (e.g., a rating) of the user on a product after he/she actually buys and experiences it. Without such knowledge, it is almost impossible to estimate user satisfaction, which is strongly related to user loyalty, before a word-of-mouth marketing strategy is proposed.

Intuitively, posterior evaluation depends only on how a product matches users' need, and is unlikely related to whether or not a friend has recommended it before. Nonetheless, in this paper we find a counter-intuitive phenomenon that a word-of-mouth recommendation surprisingly causes a change in the posterior evaluation of users. Upon this finding, we propose a framework to quantitatively measure *social influence*, i.e., the ability of an individual to affect a certain follower's posterior evaluation, and further develop a method for identifying individuals with great social influence. By properly estimating individuals' social influence, it is possible to leverage word-of-mouth recommendations in social networks to improve user satisfaction. In general, the main contributions of this paper can be summarized as follows:

1. This paper studies the relationship between word-of-mouth recommendations and users' posterior evaluation and find a counter-intuitive phenomenon that word-of-mouth recommendations are correlated to a raise in users' posterior evaluation.

2. Through elaborately designed hypothesis tests, we prove, for the first time, that word-of-mouth recommendations can significantly prompt users' posterior evaluation.

3. We propose a framework to quantitatively measure individuals' social influence by examining the number of their followers and their sensitivity of discovering good items.

4. We develop a method for identifying influential friends with strong social influence. Upon experiments on real datasets, we demonstrate that our method can identify up to 78% influential friends.

The rest sections are organized as follows: Section 2 summarizes existing studies on word-of-mouth recommendations. Section 3 describes how we find the phenomenon that word-of-mouth recommendations are correlated to a raise in users' posterior evaluation. Section 4 proposes two hypotheses to the reason that causes the phenomenon. The hypotheses are further tested in Section 5. In Section 6 we demonstrate that an influential individual can be identified with his/her sensitivity on good items and the number of followers.

2. RELATED WORKS

There is a long history of research focusing on the phenomenon, mechanics and usage of word-of-mouth recommendations. Earlier research observed the phenomenon that a word-of-mouth can effectively affect the receiver's decision making [16]. Consequently, word-of-mouth marketing, the strategy for promoting products with word-of-mouth recommendations, has been rapidly developing to encourage purchase by increasing a potential customer's prior expectation on the concerned product. The media of word-of-mouth recommendation is not limited to traditional oral conversation, but extended to online book reviews [7], emails [20], blogs [1], tweets [4], etc.

The wide usage of word-of-mouth has attracted great research attention to explain its intrinsic mechanics, which becomes practically possible with the availability of tremendous data from modern online communities [11]. Richardson and Domingos modeled the network value of individuals in social network when a word-of-mouth marketing is proposed [12, 24]. Lescovec empirically analyzed email promotions in a person-to-person network and figured out how purchase probability increases with incoming recommendations from friends [20]. Various models have been developed to explain the mechanism how a word-of-mouth message encourages a receiver to adopt a certain innovation, such as linear threshold [15], cascade model [13] and spreading model [23, 31, 8].

So far, lots of existing studies have been focusing on quantitatively estimating social influence of single nodes in social networks with topological and behavioral methods [26]. In many papers, it is believed that influence is related with the topological roles of the nodes in social networks, and can thus be calculated by link analysis, such as PageRank, HITS and LeaderRank [22]. More other studies attempted to estimate social influence by mining user behavior history to examine the effect of an individual on another's prior expectation and further decision making in a variety of scenarios, such as blogs [1], tweets [4], tags [14], books [17] and product reviews [3]. In addition, researchers have found that social influence could vary across different topics and items. Romero et al. found that tags of different types and topics spread in different mechanics on Twitter [25], Song and Tang found social influence can be of different strengths and directions given different topics [28, 29], and Cui et al. calculated social influence on item-level granularity with a modified matrix factorization algorithm [10].

Other research results focus on influence maximization, the task to find a small set of seed nodes to trigger a viral marketing that can reach a maximal range of nodes on an underlying social network, given the way information flows (e.g., defining the successful rate to propagate an innovation through every edge). Such a set of influential individuals are preferred when a merchant proposes a viral marketing. Many solutions have been developed to iteratively select seeds with a greedy algorithm, which guarantees a near-

optimal selection, such as Kempe et al. [19] and Leskovec et al. [21].

Researchers also paid attention to discover influence hidden behind the correlation between a user's actions and his friends' behaviors. However, it is difficult to tell whether the correlation comes from influence or homophily, since the discovery of causality relation between variables is always a hard job and requires careful experimental design [18]. Crandall pointed out that influence and similarity can interplay [9]. Anagnostopoulos excluded influence with a high probability through an empirical analysis on the photo tagging website Flickr [2].

Sociologist analyzed the motive that why an individual sends a word-of-mouth recommendation in social networks. Many hints revealed that whether or not a word-of-mouth recommendation is presented mostly depends on the degree a recommender himself perceives the valuable, usefulness and usability of the referred product [6]. Wasko found reasons to share opinions also include reputation building or enjoyable feeling [30]. Those findings help us simplify our model in Section 5.1.

Although extensive studies have focused on addressing the prior expectation of word-of-mouth recommendations, less attention has been paid to investigating how word-of-mouth recommendations affect users' posterior evaluation. Bone found that a word-of-mouth comment from authority could bias users' post-usage performance judgments [5], i.e., users' posterior evaluation. However, a more common question, i.e., whether or not an ordinary user's word-of-mouth recommendation can affect his/her friends' posterior evaluation, is still unclear. We intend to answer it in this study.

3. PRELIMINARY STUDY

In this section, we demonstrate the counter-intuitive phenomenon that word-of-mouth recommendations can cause significant posterior effect on users. More specifically, word-of-mouth recommendations can obviously prompt users' posterior evaluation.

3.1 Data collection

We collect data from the aforementioned two online communities, Douban and Goodreads, where users can rate books and movies, and follow or make friends with other users. Friends are treated as mutual followers. The collected data contain user activities before Aug 2010 when we crawled the two web sites. We started crawling from the member lists of several hottest discussion groups on those web sites, and iteratively expanded to users who follow, or are followed by, the crawled ones. Among the $1,014,890$ Douban users and $912,973$ Goodreads users crawled, we remove inactive users who have rated less than 5 items and followed less than 5 other users, since their inactivity may lead to unreliable statistics. The detailed information about the datasets are presented in Table 1. In both datasets, the distribution of item popularity (how many times an item is collected) is observed in a power law as shown in Fig. 1.

On Douban, a user clicks the "recommend" button on an item's page to recommend it. Afterwards, all followers will get a notification of the recommendation. A total of $165,905(15.2\%)$ Douban items are recommended by at least one user, as shown in Fig. 2. On Goodreads, when a user rates a book, all followers will get a recommendation message like "Adam loves *Harry Potter 7*". If a user rates an

Table 1: The two datasets crawled from Douban and Goodreads.

Property	Douban	Goodreads
Number of users	552,827	308,789
Number of directed edges	22,541,688	4,999,871
Number of items	1,089,914	1,648,546
Number of collections	111,001,250	26,900,929

item that has been *recommended* by anyone he/she follows, i.e. a Douban friend clicks the "recommend" button or a Goodreads friend rates it, we consider it as a *rating with word-of-mouth recommendation*. Otherwise, it is a *rating without word-of-mouth recommendation*.

In order to make ratings from different users comparable, we transform all ratings into Z-scores as follows, which will be used in the remained sections if no further specification is given.

$$r_{u,i} = \frac{R_{u,i} - \bar{R}_u}{\sigma_u}, \qquad (1)$$

where $r_{u,i}$ is the Z-score, corresponding to $R_{u,i}$ that is the observed rating user u posted on item i; $\bar{R}_u$ and σ_u are the mean and deviation among all ratings posted by user u, respectively.

3.2 Higher ratings on items with recommendations

Firstly, we take a glance at the ratings of two sample movies *The Pursuit of Happyness* and *Lost in Translation* on Douban. The former movie has been rated by over $200,000$ users with a mean rating of 4.2 out of 5 stars, ranked top 30%, whilst the latter gets a relatively low mean rating of 3.9, ranked top 50%. Fig. 3 shows the rating distributions of these two movies and Table 2 reports ratings from 7 users who have rated both movies, where ratings with word-of-mouth recommendation are underlined. An obvious trend can be revealed that with word-of-mouth recommendations Douban users are more likely to present a higher rating to the recommended movies, indicating that they enjoy recommended movies more than unrecommended ones. This trend is even clearer on the lower-rated movie *Lost in Translation*. Specifically, the probability to rate it with 5 stars increases to 4 times if a recommendation is present.

To examine whether or not the above observed trend reflects a general tendency, we report the probability distributions of all (normalized) ratings with and without word-of-mouth recommendations on the two collected datasets. As
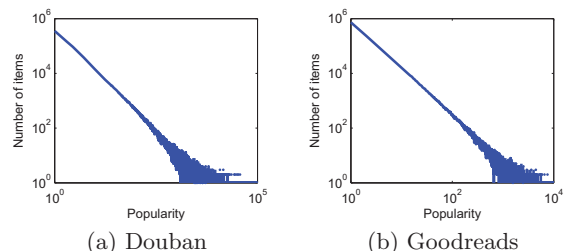

(a) Douban   (b) Goodreads

Figure 1: The distributions of popularity (the times an item is collected).
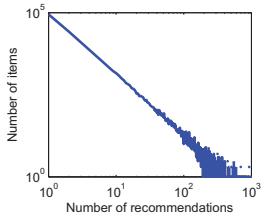
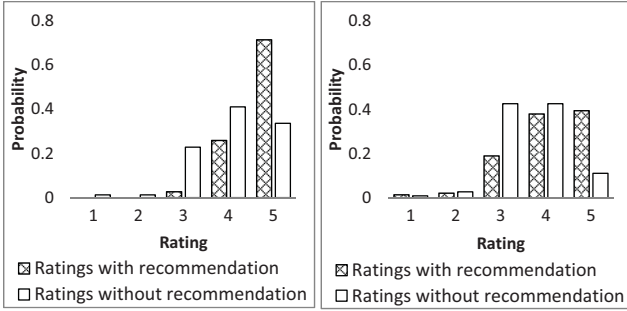**Figure 2: The distribution of recommendation count of Douban items.**

Figure 3: The probability distributions of Douban user ratings of a high-rated movie *The Pursuit of Happyness* (left) and a low-rated movie *Lost in Translation* (right).

Table 2: The ratings of two sample movies. The ratings with recommendation are underlined.

	Mean rating	u1	u2	u3	u4	u5	u6	u7
The Pursuit of Happyness	4.2	<u>5</u>	<u>5</u>	3	<u>5</u>	4	<u>5</u>	<u>5</u>
Lost in Translation	3.9	<u>4</u>	5	<u>4</u>	<u>4</u>	<u>4</u>	<u>4</u>	2

(a) Douban

(b) Goodreads

Figure 4: The probability distributions of ratings, posted with (solid curves) or without (dash curves) a word-of-mouth recommendation.
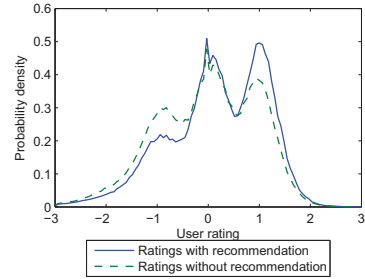
shown in Fig. 4, an individual is more likely to present a high rating (less likely a low rating) to an item with a word-of-mouth recommendation, compared with an item without. Besides, the mean ratings without recommendations are almost zero in both Fig. 4(a) and 4(b), which is straightforward given the way we normalize ratings. In contrast, the mean ratings with recommendations are raised by 0.176 and 0.081 in two datasets respectively. Upon such observations, we can assert that as compared with an item without a word-of-mouth recommendation, an item recommended by a friend is tended to be given a higher rating, implying that *a word-of-mouth recommendation is correlated to a raise in user posterior evaluation*.

An obvious explanation is that a recommendation implies the high quality of an item, which straightforwardly leads to a higher rating. Fig. 5(a) reveals that an item recommended at least once has a much larger collection count (solid curve) than than an item without any recommendation (dash curve), and the collection count of an item is strongly related to its recommendation count. Furthermore, the mean rating of items with at least one recommendation (solid curve and dash curve) is higher than that of items without recommendations (dash dot curve), as is shown in Fig. 5(b), although the ratings seem not correlated to the recommendation counts.
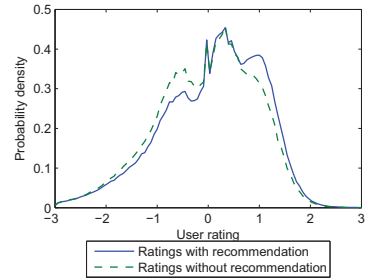
However, the above explanation only considers whether or not an item is recommended and ignores a possibly key factor that who recommends it. Therefore, in the following we examine the difference between the rating of an item recommended by the rater's friends, and that of an item recommended by a stranger that the rater never knows. Fig. 5(b) shows that a rating of an item with a word-of-mouth recommendation (solid curve) is even higher than that of an

item not directly recommended to the rater (dash curve). In a word, the tendency should not be simply explained with "a recommended item is preferred", since word-of-mouth recommendations play an indispensable role in influencing user ratings.

4. POSTERIOR EVALUATION MODELS

In this section we propose two possible models in order to formally explain the phenomenon observed in the previous section, whose validity and effectiveness will be further tested in the next section.

In the first model, a word-of-mouth recommendation and a higher rating of an item both attribute to certain common but unknown factors. In the other model, a word-of-mouth recommendation directly influences the rater to present a higher rating. The key difference between these two models lie in whether or not there is a *causality* relation between a word-of-mouth recommendation and an increase of the average user rating of an item.

The first model, which can be easily thought of, indicates that an individual's higher rating and his/her friend's rec-
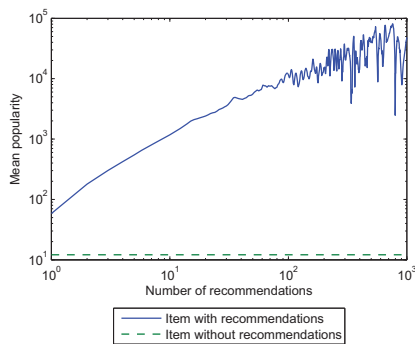
(a) Collection count w.r.t. recommendation count.
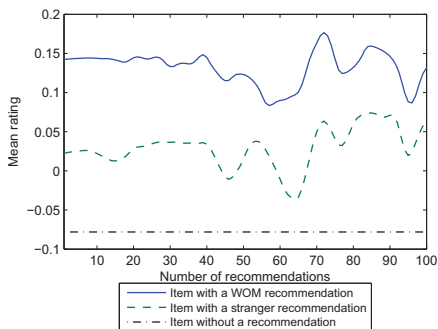
(b) Mean rating w.r.t. recommendation count.

Figure 5: The popularity and ratings w.r.t. recommendation count of items on Douban.

ommendation are both results of a group of unknown factors, including item quality, shared interests and any other reason. Take item quality for example. A recommendation can be viewed as a strong hint regarding the quality of the recommended item. If an item is recommended, it is more probable that the item is good enough to worth a higher rating. Another example is shared interests. It is observed in social networks that friends often share common interests more or less. If a friend of a user loves a movie and thus recommends it, it is quite probable that the movie also matches the user's taste, and thus a higher rating can be expected. Without loss of generality, we abstract such a model in Fig. 6(a), where a single variable c denotes a combination of all common factors. In this model, an individual's higher rating, denoted as r, and whether his/her friend recommends, denoted as m', are *independent* results from common unknown factors. Later on, we refer to this model as the ***independent model***.

The second explanation which sounds more consequential is that a friend's word-of-mouth recommendation is a direct cause of an individual's higher rating: m' directly influences r, as shown in Fig. 6(b).[4] For instance, the existence of a word-of-mouth recommendation may bring extra benefit to the receiver (e.g., reading a book that a friend recommends

---

[4]We ignore the inverse case that a user rating r leads to m' since we only consider a word-of-mouth recommendation when it is presented before the concerning user rates.
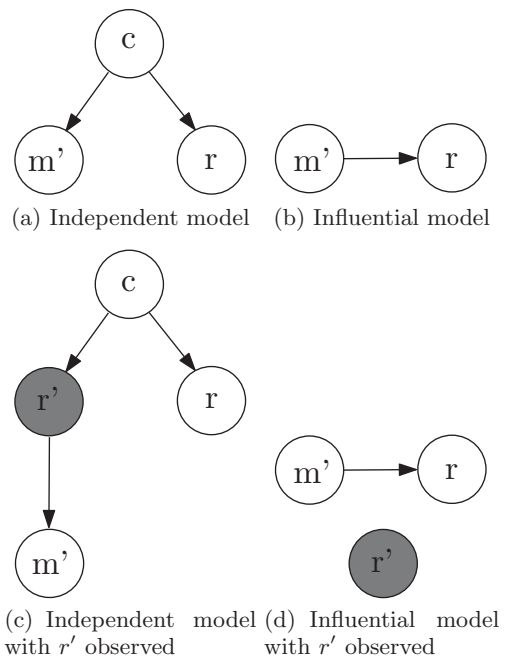

(a) Independent model   (b) Influential model

(c) Independent model (d) Influential model
with r' observed   with r' observed

Figure 6: Two possible models and corresponding conditional dependence relation between r and m', given r' observed.

might strengthen their friendship). Throughout the paper, this model is referred to as the ***influential model***. Note that we do not assume that a word-of-mouth recommendation is the only factor of user posterior evaluation.

5. MODEL VERIFICATION

In this section, we examine which of the two aforementioned models is supported by the real data using statistical hypothesis tests.

5.1 Designing statistical hypothesis tests

According to Fig. 6(a) and Fig. 6(b), in order to judge which model is more appropriate to explain the real data, we need to examine the conditional independence between m' and r given c. Unfortunately, such a conditional independence cannot be directly examined since the hidden variable c is unobservable. To circumvent this problem, we introduce the recommender's rating, denoted as r', to modify the models. In some cases an individual u rates an item previously recommended by his friend (i.e., $m' = 1$), r' denotes the rating that the friend posted to the item. In other cases an individual u rates an item with no friend recommendation (i.e., $m' = 0$), r' denotes the rating posted to the item by a randomly chosen friend of u who might have recommended the item to u. The modified independent model is shown in Fig. 6(c). Such a modification is reasonable from two aspects: 1) r' depends on c symmetrically as r in the original independent model; 2) m' fully depends on r' due to a common practice that whether or not a user recommends an item is determined by his/her rating on this item [6, 30]. In the modified independent model, whether m' and r are independent conditioned on r' is equivalent with whether they are independent conditioned on c, according to Bayesian net-

work property. In addition, for the influential model where m' is a direct factor node of r, the dependence of r on m' always hold no matter any other variables are observed, as shown in Fig. 6(d). Therefore, we can judge which of the independent model and the influential model is more appropriate by examining the conditional independence between m' and r given r'.

Now we design a statistical hypothesis test to examine the conditional independence between m' and r given r'. In our test, each data point is a triple $\langle m', r, r' \rangle$. We first divide the observed data points into distinct partitions such that the data points in each partition share a unique or similar r' value and therefore follow, and can well approximate, the conditional probability distribution $p(m', r|r')$ if the partition size is large enough for convergence. For this purpose, we discretize the real-valued r' into distinct values in two ways, namely, equal-interval partitioning or equal-frequency partitioning. For the equal-interval partitioning, the values of r' in each partition are confined within a fixed small interval. We discard partitions containing less than 100 data points due to their low statistical reliability. For the equal-frequencies partitioning, each partition has a fixed number of data points. We ignore partitions where the interval of r' is too large (greater than 0.05) to guarantee the values of r' in the same partition are close enough.[5]

For each partition, we test the independence between $p(m'|r')$ and $p(r|r')$. Since m' can take only two values, namely, 0 and 1, $p(m'|r')$ is independent from $p(r|r')$ if and only if $p(r|m' = 0, r')$ is identical to $p(r|m' = 1, r')$ for any given r'. Whether or not $p(r|m' = 0, r')$ is identical to $p(r|m' = 1, r')$ can be examined by testing the empirical measure of two samples, $\{r|m' = 0, r'\}$ and $\{r|m' = 1, r'\}$, which serve as good estimators when the sample size is large enough. Specifically, we employ two different statistical hypothesis test methods to examine: the location t-test to examine whether there exists a significant difference between the means of the two samples, and the Kolmogorov-Smirnov test to examine whether there exists a significant difference between the empirical distribution functions of the two samples.

The null hypothesis in our test is that there exists no statistically significant difference between $p(r|m' = 1, r')$ and $p(r|m' = 0, r')$ for any given r'. If the null hypothesis holds, it implies that given a certain value of r', r identically distributes, regardless of the values of m' and therefore the independent model is supported. On contrary, the alternative hypothesis that the two samples are significantly different disproves the conditional independence and supports the influential model. We test the null hypothesis at the 0.05 significance level against the alternative hypothesis in all partitions for a consistent conclusion. We will declare the influential model holds if tests in almost all partitions reject the null hypothesis, and vice versa. In practice we report the reject ratio as a final result, i.e. the percentage of partitions where statistical hypothesis tests reject the null hypothesis.

In summary, the statistical hypothesis test is conducted as follows:

---

[5]The discretization of recommender rating r' is valid. Our normalization preprocess from integer ratings to real number Z-scores guarantees that the Z-score space of the whole dataset is still countable. Therefore, discretizing Z-scores does not greatly violate value dependence.
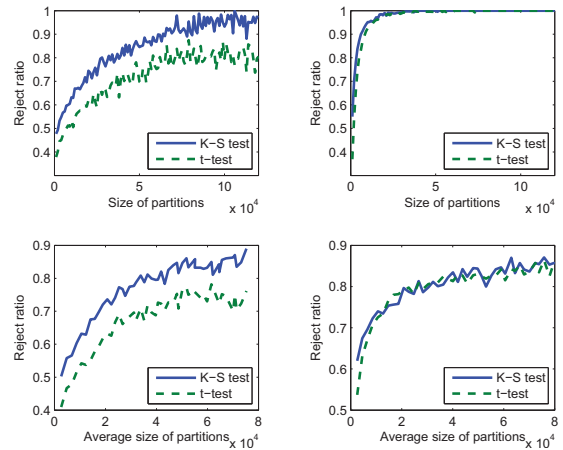

Figure 7: The reject ratio, the percentage of partitions where a two-sample test rejects the null hypothesis. Tests are conducted on two datasets (Left: Douban; Right: Goodreads) and with two partitioning ways (Top: equal-frequency; Bottom: equal-interval) respectively.

1. Slice data into partitions. All data points in each partition share a unique r' value.

2. Divide each partition into two samples, one with $m' = 1$ and the other with $m' = 0$.

3. In each partition, employ the two-sample t-test and a two-sample Kolmogorov-Smirnov test to examine the null hypothesis, i.e. the two samples are identically distributed.

4. Report the percentage of partitions where statistical hypothesis tests reject the null hypothesis.

5.2 Results of statistical hypothesis tests

In our test, we only consider the ratings falling into the range $[-3, 3]$ that accounts for most of the ratings from the Douban and Goodreads datasets. As a consequence, we obtain $186,054$ data points with $m' = 1$ and $2,278,023$ with $m' = 0$ for the Douban dataset. For Goodreads, the numbers of the data points are $6,759,850$ and $2,899,793$, respectively.

Fig. 7 illustrates the test result. When we divide Goodreads dataset into partitions with a fixed size of $20,000$ data points, the Kolmogorov-Smirnov test rejects the null hypothesis in $461(98.5\%)$ partitions among 468 ones, and the t-test rejects in 97.2% of all partitions. Similarly, when the Goodreads dataset is divided into partitions with an equal interval of 0.03, which results in an average size of around $48,000$ data points in each partition as shown on the axis label, the Kolmogorov-Smirnov test rejects the null hypothesis in 155 partitions among 185 ones, covering 83.8% partitions and 97.5% data points, while the t-test rejects in $152(82.2\%)$ partitions and 97.5% data points.

Test results are obvious positive, reading that on both datasets and with both partitioning ways, the two adopted statistical hypothesis tests consistently reject the null hypothesis in favor of the alternative hypothesis in a majority

of partitions. The percentage of partitions where the null hypothesis is rejected increases with the number of data points contained in each partition, because larger samples can better establish a reliable result. When the partition size is large enough, the tests on almost all partitions consistently reject the null hypothesis, strongly supporting the influential model.

The fact that tests on Douban dataset fail to reject the null hypothesis in all partitions might attribute to the imbalance and size of Douban dataset. The Douban dataset is much more imbalanced than the Goodreads dataset, given less than 14% data points are observed with $m' = 1$. Besides, the number of Douban data points tested is less than one fourth of that of Goodreads dataset, therefore it is reluctant to afford a large amount of large partitions to establish reliable statistics results. With the obvious trend that a higher reject ratio comes with a larger size of partition, tests on Douban is expected to reach a almost-hundred-percent reject ratio with a larger dataset.

In summary, the two-sample tests in most partitions reject the null hypothesis in favor of the alternative hypothesis that r behaves differently with different m' values even when r' is observed, which suggests that the influential model correctly captures the causality relation hidden in the real data. That is to say, the posterior evaluation of a user directly depends on whether or not a friend has previously recommended the concerning item to him/her.

6. IDENTIFYING INFLUENTIAL FRIENDS

In the previous section, we have shown that word-of-mouth recommendations can influence users' posterior evaluation on the items they have experienced, which are specifically reflected by the increase of ratings presented by the receivers of the recommendations. Now it is of high interest to know what kind of persons are influential via their word-of-mouth recommendations and to identify the influential friends for each user.

6.1 Social influence measurement

Before proceeding, we need to quantify the raise in posterior evaluation brought by each word-of-mouth recommendation. We use $U_{u,i}^-$ to denote the rating on the item i from the user u without recommendations from his friends and $U_{u,i}^+$ to denote the rating with recommendations. Then the raise caused by word-of-mouth recommendation can be calculated by

$$x_{u,i} = U_{u,i}^+ - U_{u,i}^-. \tag{2}$$

Unfortunately, we cannot observe $U_{u,i}^+$ and $U_{u,i}^-$ simultaneously. To address this problem, we consider only the cases where $U_{u,i}^+$ can be observed, i.e., the user u rated the item i after the item has been recommended by his/her friends at least once[6]. For the unobserved $U_{u,i}^-$, we estimate its value with $\hat{r}_{u,i}$, the expected rating a user would have posted without recommendations from his friends. In practice, $\hat{r}_{u,i}$ can be calculated with well-studied collaborative filtering recommendation algorithms. Actually, any recommendation algorithm can be used with the only constraint

that the used algorithm should not leverage the knowledge about the friendships. In this paper, we employ the probabilistic matrix factorization algorithm [27], which has proven an effective and accurate algorithm especially for the sparse cases. For the sake of convenience, we also replace the $U_{u,i}^+$ with $r_{u,i}$ and the raise in posterior evaluation caused by word-of-mouth recommendation can be written as

$$\hat{x}_{u,i} = r_{u,i} - \hat{r}_{u,i}. \tag{3}$$

For the probabilistic matrix factorization algorithm, we train it with the training set consisting of only the ratings posted without recommendations. In this way, the trained model estimate the unobserved ratings $\hat{r}_{u,i}$ without the knowledge about friendships. Under such consideration, the two training datasets obtained for Douban and Goodreads contain $23, 417, 433$ and $16, 698, 014$ ratings, respectively. With the trained model, we estimate $\hat{r}_{u,i}$ for each observed $r_{u,i}$ and calculate the raise in posterior evaluation caused by word-of-mouth recommendations using Eq. (3).

Then, for any edge (u, v) which represents the friendship between two users u and v in the friendship network, we can measure the influence of u on v as the average raise in posterior evaluation over all items I_{uv} recommended to v by u and rated by v after the recommendation. Specifically, the *social influence* of u on v is calculated as

$$f_{u,v} = \frac{\sum_{i \in I_{uv}} \hat{x}_{v,i}}{|I_{uv}|}, \tag{4}$$

where $|I_{uv}|$ denotes the number of items in I_{uv}.

Given social influence calculated on all edges, salient asymmetry is observed that there usually exists an individual, between a pair of mutually connected friends, having a much greater social influence upon the other than what the other has on him. We quantify this phenomenon with the ratio of social influence on each pair of bidirectional edges, i.e., $\frac{max(f_{u,v}, f_{v,u})}{min(f_{u,v}, f_{v,u})}$. In 83% Douban pairs and 77% Goodreads pairs[7], there exists an individual whose social influence on the other is at least as twice as that in the reverse direction. As indicated by this phenomenon, the influence among two friends often acts in an one-way manner, although two friends can recommend favorite products to each other and thus have the potential to influence each other.

Furthermore, we label (u, v) an *influence edge*, if $f_{u,v}$ is larger than 0.1 and $|I_{uv}|$ is no less than 5. In this case, u serves as the *influential side* and v *influenced side*. Please note that those roles are assigned in a particular edge rather than a global setting. For the two used datasets, 0.095% and 2.19% edges are labeled as influence edges which involve 0.078% and 7.23% users, respectively.

6.2 Identification of influential friends

Now we check what correlates with the direction of influence among friends. Intuitively, there are possibly two kinds of factors, namely, the social positions of users in the friendship network and their personal characteristics independent on other individuals. Specifically, for the first kind of factors, we consider the out-degree[8], PageRank and Lead-

---

[6]For ease of identifying the influential source, in practice we ignore ratings posted with two or more friend recommendations, which are only a small portion of our observations.

[7]We ignore edges with $|I_{uv}| < 5$ whose f could be unreliable.
[8]We also analyze in-degree. Since Douban and Goodreads user-user networks are both observed with high reciprocity, the feature in-degree has almost the same result as out-degree.

erRank. For the second kind of factors, we consider the collection size, collection frequency (number of items collected per day) and sensitivity. Note that sensitivity, proposed in this paper, represents an individual's ability to quickly discover new valuable contents. Specifically, the sensitivity S_u of a user u is defined as

$$S_u = \frac{\sum_{i \in I(u)} |\{v : i \in I(v), T(v,i) > T(u,i)\}|}{\sum_{i \in I(u)} |\{v : i \in I(v)\}|}, \quad (5)$$

where $I(u)$ denotes the set of items collected by user u and $T(u,i)$ is the date when u collects item i. The sensitivity of a user can be explained as the percentage of other users who rates an item later than him/her. For example, if an individual u is the tenth one among all the 100 users who collect *The King's Speech* and the 20th among the 50 users collecting *Black Swan*, the sensitivity of u can then be calculated as $(90 + 30)/(100 + 50) = 0.8$.

With respect to each feature t, we examine the influential-side-to-influenced-side ratio $R_{u,v}^t$ on each influence edge pointing to v from u. Take the out-degree as example, $R_{u,v}^{out-degree} = k_u^{out}/k_v^{out}$. Fig. 8 shows the probability distribution of logarithm of $R_{u,v}^{out-degree}$ and $R_{u,v}^{sensitivity}$. It can be observed that the influential side of an influence edge is more probable to have a higher out-degree and sensitivity than the influenced side, reflected by the majority population of positive values of logarithm ratio (deep grey area in figure). The ratios with respect to other features also provide evidence on the effectiveness of these features when identifying the influential friends.

A more direct method for discriminating the effectiveness of these features is the accuracy in predicting which individual is the influential side in such a pair of friends, that are (a) mutually connected so that either might serve as the influential side, and (b) one edge between them is labeled as an influence edge. The results are shown in Table 3. We can see that almost all the considered features can correctly predict most of the pairs. This implies that a more active individual, from either the social perspective or the personal perspective, is more probable to be the influential side among two friends. We can also see that out-degree and sensitivity play the most important role in detecting influential sides. This tells us that an influential friend is more eager to explore new items and followed by many other users.

Table 3: The prediction accuracy of influence edges with user features.

Feature	Douban	Goodreads
Out-degree	**69.17%**	**71.16%**
Collection size	51.59%	59.18%
Collection frequency	46.18%	47.43%
Sensitivity	**69.41%**	**77.67%**
PageRank	68.98%	70.35%
LeaderRank	68.76%	70.28%
Random guess	50 %	50 %

Focusing on out-degree and sensitivity, we detail the experimental results. For each pair of mutually connected individuals u and v, with one and only one labeled influence edge (u,v), we define the gap between u and v as $G_{u,v}^f = f_u - f_v$ with respect to each feature f, where f_u is the value of feature f of user u. Given a threshold δ for the gap, we predict
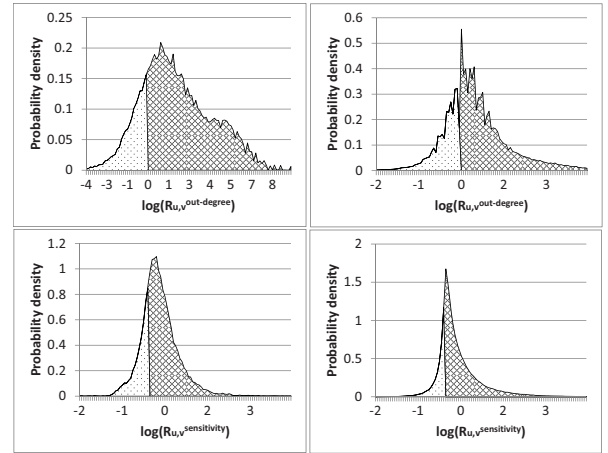

Figure 8: The out-degree and sensitivity of the influential and influenced sides in bidirectionally connected user pairs. Douban (left) and Goodreads (right).

edge (u,v) as an influence edge if $G_{u,v}^f > \delta$, or edge (v,u) as an influence edge if $G_{v,u}^f = -G_{u,v}^f > \delta$. Otherwise, we leave the pair unpredicted. Different gap thresholds lead to different prediction *accuracy*, i.e., the percentage of correctly predicted pairs among all predicted pairs, and *cover rate*, namely, the percentage of predicted pairs among all pairs. Generally speaking, a larger gap leads to a lower cover rate and a higher accuracy of prediction. This reflects that a tighter judge condition results in more confidence. That is to say, the more imbalanced that two friends are in their out-degrees or sensitivity, the more probable that the influence comes from the more social and sensitive side to the other. Fig. 9(a) depicts the accuracy and the cover rate for the two tested datasets. When the cover rate is 100%, the out-degree succeeds at predicting 69.17% or 71.16% of all predicted pairs for the two datasets. The accuracy of the feature sensitivity is slightly higher than the out-degree, being 69.41% or 77.67%. Note that the accuracy of both features are significantly higher than a random predictor with an expected accuracy of 50%.

Note that the out-degree characterizes the social position of an individual, while the sensitivity describes his/her personal characteristics. An interesting question is that whether these two kinds of features can complement each other when identifying the influential friends. In order to answer this question, we combine these two features to predict the influential side of each influence edge. Intuitively, an influential friend may have both higher out-degree and higher sensitivity. Therefore, we integrate these two features into one by combining their influential-to-influenced ratios as follows

$$R_{u,v}^{combined} = (R_{u,v}^{sensitive})^\lambda (R_{u,v}^{out-degree})^{(1-\lambda)}, \quad (6)$$

where the coefficient λ takes a real value in the range $[0,1]$. Then, we predict the influential side as the one with a larger $R_{u,v}^{combined}$. Fig. 9(b) presents the prediction results with varied λ and the gap threshold $\delta = 0$. We can see that the prediction accuracy has a single maximum with respect to the value of λ for the two tested datasets. This indicates that a proper combination of out-degree and sensitivity can

better distinguish an influential side than using only one of these two features. Therefore, it is necessary to combine the social position of individuals and their personal characteristics when identifying the influential friends for social recommendations.
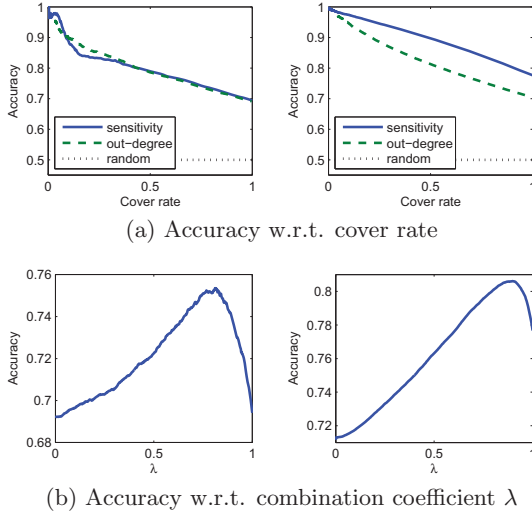

(a) Accuracy w.r.t. cover rate

(b) Accuracy w.r.t. combination coefficient λ

Figure 9: The prediction accuracy of influential sides with larger out-degree and sensitivity on Douban (left) and Goodreads (right), respectively.

7. CONCLUSIONS AND DISCUSSIONS

In this paper, we empirically analyzed the effects of word-of-mouth recommendations on real-world online social networks and content sharing communities. Specifically, our analysis indicates that word-of-mouth recommendations cannot only improve the prior expectation of users, who receive the recommendations, on the recommended items, but also raise their posterior evaluation. We further verified that the raise of the posterior evaluation is directly caused by word-of-mouth recommendations, through elaborately designed statistical hypothesis tests. Furthermore, we quantified the raise of posterior evaluation by comparing the real ratings presented by receivers of the word-of-mouth recommendations and the estimated ratings obtained by a classical collaborative filtering recommendation method. The latter are actually employed to approximate the ratings in the case without word-of-mouth recommendations. Based on the quantification, we proposed a method for investigating a user's social influence, i.e., the ability of the user to affect a follower's posterior evaluation, by examining his/her sensitivity of discovering good items and the number of his/her followers. Through experiments on the Douban and Goodreads datasets, we demonstrated that our method can successfully identify up to 78% influential friends.

This paper provides a new research perspective, i.e., the posterior evaluation, instead of the prior expectation which has been widely researched into. This new perspective is important for devising online word-of-mouth recommendation methods, aiming to enhance users' satisfaction and loyalty. However, it is still an open problem to identify the potential individuals whose opinions can influence their social con-

tacts. In addition, it is of high interest to investigate how social influence propagates in social networks and how multiple word-of-mouth recommendations accumulate, which are critical to trigger a viral marketing cascade. Whether or not social influence changes across different topics is another interesting problem. Our future work will focus on these important research issues.

In addition, our findings are helpful to developing new applications. For example, shopping sites can promote a product to potential customers by asking their influential friends to make online word-of-mouth recommendations, instead of using an automatic recommendation system. More importantly, such a marketing strategy might have the potential to improve product sales and overcome the sparsity problem in personalized recommendation, as long as it is carefully designed to avoid fraudulent recommendations.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *Proceedings of the international conference on Web search and web data mining*, WSDM '08, pages 207–218, New York, NY, USA, 2008. ACM.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 7–15, New York, NY, USA, 2008. ACM.

[3] C.-m. Au Yeung and T. Iwata. Strength of social influence in trust networks in product review sites. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 495–504, New York, NY, USA, 2011. ACM.

[4] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 65–74, New York, NY, USA, 2011. ACM.

[5] P. F. Bone. Word-of-mouth effects on short-term and long-term product judgments. *Journal of Business Research*, 32(3):213 – 223, 1995.

[6] T. Brown, T. Barry, P. Dacin, and R. Gunst. Spreading the word: Investigating antecedents of consumers' positive word-of-mouth intentions and behaviors in a retailing context. *Journal of the Academy of Marketing Science*, 33(2):123–138, 2005.

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

[8] G. Cimini, M. Medo, T. Zhou, D. Wei, and Y.-C. Zhang. Heterogeneity, quality, and reputation in an adaptive recommendation model. *The European*

Physical Journal B - Condensed Matter and Complex Systems, 80:201–208, 2011.

[9] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 160–168, New York, NY, USA, 2008. ACM.

[10] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, SIGIR '11, pages 185–194, New York, NY, USA, 2011. ACM.

[11] C. Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424, 2003.

[12] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, pages 57–66, New York, NY, USA, 2001. ACM.

[13] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.

[14] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 241–250, New York, NY, USA, 2010. ACM.

[15] M. Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420–1443, 1978.

[16] P. M. Herr, F. R. Kardes, and J. Kim. Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research: An Interdisciplinary Quarterly*, 17(4):454–62, 1991.

[17] J. Huang, X.-Q. Cheng, J. Guo, H.-W. Shen, and K. Yang. Social recommendation with interpersonal influence. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI 2010)*, pages 601–606, Amsterdam, The Netherlands, 2010. IOS Press.

[18] D. D. Jensen, A. S. Fast, B. J. Taylor, and M. E. Maier. Automatic identification of quasi-experimental designs for discovering causal knowledge. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 372–380, New York, NY, USA, 2008. ACM.

[19] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[20] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1, May 2007.

[21] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 420–429, New York, NY, USA, 2007. ACM.

[22] L. Lu, Y.-C. Zhang, C. H. Yeung, and T. Zhou. Leaders in social networks, the delicious case. *PLoS ONE*, 6(6):e21202, 06 2011.

[23] M. Medo, Y.-C. Zhang, and T. Zhou. Adaptive model for recommendation of news. *EPL (Europhysics Letters)*, 88(3):38005, 2009.

[24] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, pages 61–70, New York, NY, USA, 2002. ACM.

[25] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, WWW '11, pages 695–704, New York, NY, USA, 2011. ACM.

[26] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24:33–60, 2005. 10.1007/s10462-004-0041-5.

[27] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, volume 20, 2008.

[28] X. Song, B. L. Tseng, C.-Y. Lin, and M.-T. Sun. Personalized recommendation driven by information flow. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 509–516, New York, NY, USA, 2006. ACM.

[29] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 807–816, New York, NY, USA, 2009. ACM.

[30] M. M. Wasko and S. Faraj. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1):35–57, 2005.

[31] D. Wei, T. Zhou, G. Cimini, P. Wu, W. Liu, and Y.-C. Zhang. Effective mechanism for social recommendation of news. *Physica A: Statistical Mechanics and its Applications*, 390(11):2117 – 2126, 2011.