11-2013

Information vs interaction: An alternative user ranking model for social networks

Wei XIE
Singapore Management University, wei.xie.2012@smu.edu.sg

Ai Phuong HOANG
Singapore Management University, aphoang.2013@smu.edu.sg

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

DOI: https://doi.org/10.1007/978-3-319-03260-3_20

Citation

XIE, Wei; HOANG, Ai Phuong; ZHU, Feida; and LIM, Ee Peng. Information vs interaction: An alternative user ranking model for social networks. (2013). *Social Informatics: 5th International Conference, SocInfo 2013, Kyoto, Japan, November 25-27, 2013: Proceedings*. 8238, 227-240. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/1975

Information vs Interaction: An Alternative User Ranking Model for Social Networks

Wei Xie, Emmy Hoang Ai Phuong, Feida Zhu and Ee-Peng Lim

Singappre Management University, Singapore
{wei.xie.2012,aphoang.2012,fdzhu,eplim}@smu.edu.sg

Abstract. The recent years have seen an unprecedented boom of social network services, such as Twitter, which boasts over 200 million users. In such big social platforms, the influential users are ideal targets for viral marketing to potentially reach an audience of maximal size. Most proposed algorithms rely on the linkage structure of the respective underlying network to determine the information flow and hence indicate a users influence. From social interaction perspective, we built a model based on the dynamic user interactions constantly taking place on top of these linkage structures. In particular, in the Twitter setting we supposed a principle of balanced retweet reciprocity, and then formulated it to disclose the values of Twitter users. Our experiments on real Twitter data demonstrated that our proposed model presents different yet equally insightful ranking results. Besides, the conducted prediction test showed the correctness of our model.

1 Introduction

The sheer number of indexed webpages online, which is estimated at 3.97 billion [1], has made ranking algorithms indispensable for virtually any practical applications to access individual webpages. Algorithms such as PageRank [11] and HITS [3] have achieved huge success in finding top-ranked authoritative webpages by analyzing the URL linkage structure. Similarly, the recent boom of social network services has posted a need as strong for good algorithms to rank their users for a variety of applications. For example, top-ranked users by social influence are ideal targets for viral marketing to potentially reach an audience of maximal size. Among the social network services, micro-blogging services like Twitter have been the most favorable in terms of marketing due to the fact that information, in the form of tweets, could spread the fastest through the follow links. A number of algorithms have therefore been proposed for the particular setting of Twitter among which TwitterRank [15] has been one of the most noticeable. What TwitterRank and PageRank, including those similar ones they each represent, shared in common is that they both rely on the linkage structure of the respective underlying network, i.e., the URL linkage network for PageRank and the follow link network for TwitterRank.

[1] http://www.worldwidewebsize.com/

A closer examination of these linkage structures shows that they represent primarily how information would flow and tend to be relatively static. For example, the Twitter follow network gives the diffusion of tweets and is relatively static compared to the other user actions such as tweet and retweet. What they fail to capture is the dynamic user interactions constantly taking place on top of these linkage structures, e.g., how users retweet and reply one another. Yet, it is our believe that the dynamic user interactions is also an important part essential to a social network because they reveal more insights into users' social relationship than the underlying linkage structure. For example, it is common that users only interact with a small number of other users with retweet and reply out of the many who follow them and whom they follow, or both. Even among those they indeed interact with, they interact differently, e.g., retweeting with different frequency. Clearly, these user interactions, which are also much more dynamic, shed more interesting insights into their social relationships, e.g., relationship strength, relative social status, etc..

In this paper, we propose an alternative user ranking model based on a user interaction perspective, which could give rather different ranking results compared with the traditional ones, which we would consider them as based on an information flow perspective. Let's look at a simple illustrative example. In Figure 1, nodes represent Twitter users, directed edges in (a) denote follow links and the weighted directed edges in (b) denote the number of times a user has retweeted the other one. For example, It tells from the figure that Dave has retweeted Alice three times while Alice has only retweeted Dave once. Now if we run PageRank algorithm on the underlying follow network, the node of Dave would rank the highest as it is the network hub of the information flow. While this makes perfect sense from the information flow perspective, we argue that, if we examine instead how users interact with each other, then we could have a different ranking of the nodes. For example, suppose we assume the ratio between the number of retweets between two users corresponds to their relative social relation status in the sense that a user with higher relative status would be retweeted more than the other party with relatively lower status. Then, given this assumption, the node of Alice could be the highest ranked one from the user interaction perspective since Alice appears superior to Dave who is a node of importance itself. This example illustrates the difference between the rankings from two different perspective, namely, the information flow one and the user interaction one.

The main contribution of this paper is to re-examine the value of users in social network from the social interaction perspective. In particular, we consider the social interaction in the notion of reciprocity based on the retweet interaction between Twitter users. Reciprocity is a well-established concept in both social science [4] and economics [13]. In our particular Twitter setting, it refers to the mutual adoption of each other's tweets between two users in the form of retweet, the result of which is a boost to both parties' social impact. We formulated the retweet reciprocity, proposed an alternative user ranking model based on retweet reciprocity and developed efficient inference solution. Our experiments

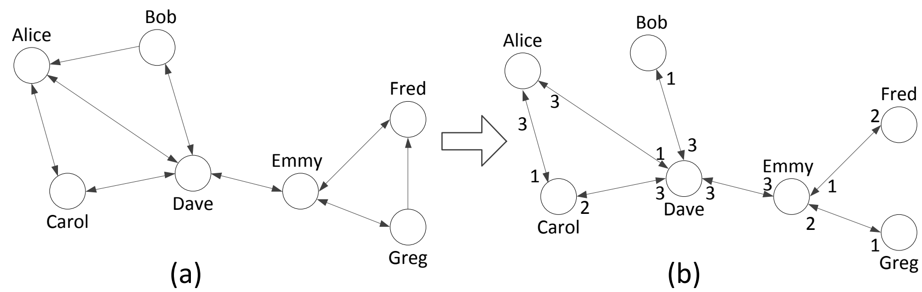

Fig. 1. (a) Twitter follow network. (b) Twitter reciprocal retweet network.

on real Twitter data demonstrated that our proposed model presents different yet equally insightful ranking results.

2 Proposed Solution

2.1 Preliminaries

We consider a set of users $U = \{u_i\}_{i=1}^n$, where n is the size of U, from which we want to find influential users. Denote R as all the behaviour of retweet which are performed by U. Here we only consider the behaviour of retweet without regard to the contents of tweets. In other words, R is represented as a bag of ordered pairs $\{(u_i, u_j)\}$, where a pair (u_i, u_j) means the user u_i retweet a tweet from its original user u_j, which follows the direction of retweet. Based on the retweets R, we construct a retweet network, which actually is a weighted directed graph $G = (U, E, W)$, such that the users U is the set of vertexes of graph G, $E = \{(u_i, u_j)\} \in U \times U$ is the set of retweet behaviour among them, and $W = \{w_{i,j}\}$ where $w_{i,j}$ indicates the number of retweets from u_i to u_j. After getting the retweet network, we construct the reciprocal retweet network by simply removing all the one-way edges, which is illustrated in Figure 1 (b). At last, denote $V = \{v_i\}_{i=1}^n$ as the values of users. In this paper, we aim to disclose such values V from the retweet interactions between users U.

2.2 Model

Principle of Balanced Reciprocity.

We consider the social interaction in the notion of reciprocity based on the retweet interaction between Twitter users. Particularly, here it refers to the mutual adoption of each other's tweets between two users in the form of retweet. However, we can observe inequality of such retweet reciprocity. For example, as shown in Figure 1, Alice and Dave have retweeted each other. But it is in unequal way – Dave has retweeted Alice three times while Alice has only retweeted Dave once. We suppose such inequality reflects the inequality of the users' social status.

For example, Alice may have higher quality tweets than Dave, so Dave retweeted Alice's tweets more than Alice retweeted Dave's. Despite such kind of inequality, there is still a balance between two users's retweet behaviour, which is agreed by both users. In other words, there is a balanced reciprocity between two users, namely, the ratio between the number of retweets between two users corresponds to their relative social relation status. We measure users' social status as their values V. According to the above principle of balanced reciprocity, we conduct the Equation 1 below.

$$\frac{w_{i,j}}{w_{j,i}} = \frac{v_j}{v_i} \tag{1}$$

We suppose the Equation 1 are reflected by the continuous interaction between Twitter users.

Minimizing Error.

For the observed data, we expect that the error of the Equation 1 should be as small as possible for all pairs of users. For easy optimization, we transform Equation 1 as a equivalent linear formulation as shown in the Equation 2 below.

$$w_{i,j} \cdot v_i = w_{j,i} \cdot v_j \tag{2}$$

So the Equation 3 below which is the sum of all the square error of all pairs of users should be minimised.

$$e(V) = \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{i,j} \cdot v_i - w_{j,i} \cdot v_j)^2 \tag{3}$$

So we can infer the values of users V by minimising the error function $e(V)$.

2.3 Inference

In this section, we discuss how to infer the value of each user by minimising the error function $e(V)$. First, it is quite obvious that $e(V) = 0$, if we set all $v_i = 0$, which makes no sense. So here we conduct a penalty function $p(v)$, and append penalty terms at the end of $e(V)$ as the below Equation 4. So we minimise $e^*(V)$ instead of $e(V)$. Here the penalty function $p(v)$ should have such properties: (I) $p(v)$ is very large at $v = 0$, so that v_i is far from 0 by minimising $e^*(V)$; (II) $p(v) > 0$ and $\lim\limits_{v \to +\infty} = 0$, so that there is no penalty when v_i is far from 0; and at last (III) $p(v)$ is a monotonically decreasing function. In this paper, we set $p(v) = M \cdot e^{-kv}$, (where M and k are positive value), but it is not the only formulation for the penalty function $p(v)$.

$$e^*(V) = \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{i,j} \cdot v_i - w_{j,i} \cdot v_j)^2 + \sum_{i=1}^{n} p(v_i) \tag{4}$$

Denote the derivative of $p(v)$ as $p'(v)$. We can get the derivatives of $e^*(V)$ as the Equation 5 below.

$$\frac{\partial e^*}{\partial v_i} = 2\sum_{j=1}^{n} w_{i,j}(w_{i,j} \cdot v_i - w_{j,i} \cdot v_j) + \sum_{i=1}^{n} p'(v_i) \tag{5}$$

Based on the the derivatives of $e^*(V)$, gradient descent method is used to get the optimal value of V.

3 Empirical Evaluation

3.1 Data Set

We use a Twitter data set which contains 3,165,479 users. These users are obtained by a snowball-style crawling starting from a seed set of Singapore local celebrities and active users and tracing their follower/followee links up to two hops. We use the subset of tweets between October 1, 2011 and December 31, 2011, which contains 50,918,021 tweets and 90,205 distinct users and 6,943,189 retweets. Among these 90,205 users, 44,152 users retweet at least one tweet or be retweeted at least once in our data set. We also get the follow links between these 44,152 users, including 653,619 links in total. Using these users and retweets, we constructed the retweet graph described in the Preliminaries Section. The follow graph is also constructed based on the follow links between these users.

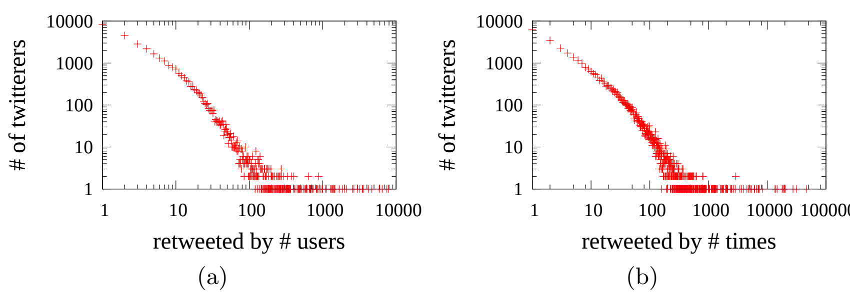

Fig. 2. Power law phenomenon exists in the retweet network.

First we consider Twitter users' popularity in terms of the number of users by whom they were retweeted, and the number of times they were retweeted. In Figure 2(a), we show the distribution of the number of users by whom Twitter users were retweeted. In Figure 2(b), we show the distribution of the number of times Twitter users were retweeted. As illustrated in [10], the pow-law phenomenon was found in the Twitter follow network, similarly, in our data set, from Figure 2(a)(b) we also found that the power-law phenomenon exists in the Twitter retweet network. Intuitively, the value of the Twitter users can be measured by their popularity in both terms above. In Section 3.2 we compared our solution with this intuition.

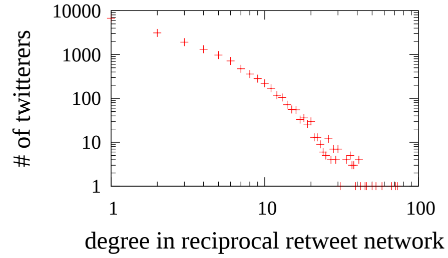

Fig. 3. Power law phenomenon exists in the reciprocal retweet network.

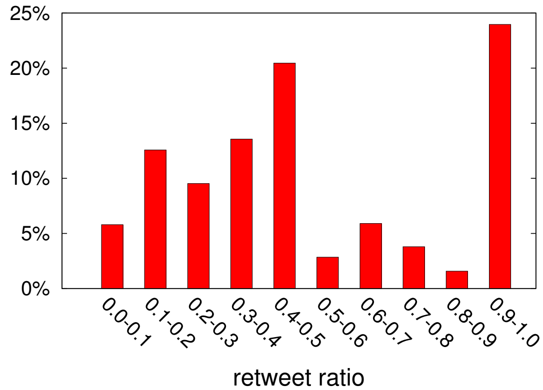

Fig. 4. Retweet ratio.

As we tried to disclose the users' values from their reciprocal retweet behaviour, then we constructed the retweet reciprocal network based on the retweet network, by simply removing all the one-way edges in retweet network. As shown in Figure 3, the power-law phenomenon also exists in this reciprocal retweet network. For each pair of users who retweeted each other, we studied the ratio of the numbers of tweets retweeted by them (the smaller one divided by the larger one). Figure 4 showed the distribution of these retweet ratios. From Figure 4, we can see that the bars at regions 0.9-1.0 and 0.4-0.5 are higher than others, because a lot of users only retweet others one or two times, which leads to the popularity of "1:1" and "1:2". However, in general, the retweet ratios vary from 0.0 to 1.0, which implies the difference of inherent values of different users.

As the scale is too large to get the optimal values for all the users in our dataset in practical time, here we considered a sub reciprocal retweet network of top 3,000 users who have most numbers of users by whom they were retweeted. Based on this sub network, we calculated the values for all these 3,000 users.

	1	2	3	4	5	6
Ours	**SoSingaporean**	**NaomiNeo_**	imwhywhy	speishi	heedyjoee	ShilinKEY
PR	Cursedwithsex	mrbrown	**stcom**	**NaomiNeo_**	TommyWee	xavlur
HS	**SoSingaporean**	**fakeMOE**	**NaomiNeo_**	xavlur	**stcom**	Cursedwithsex
UB	**SoSingaporean**	**stcom**	mrbrown	**fakeMOE**	xavlur	**NaomiNeo_**
RB	**SoSingaporean**	**stcom**	mrbrown	**NaomiNeo_**	Cursedwithsex	BvsSG
FN-PR	**SoSingaporean**	Xiaxue	mrbrown	**stcom**	**fakeMOE**	JoannePeh
FN-HS	**SoSingaporean**	Xiaxue	mrbrown	**fakeMOE**	**stcom**	BvsSG
TR	**stcom**	mrbrown	Cursedwithsex	Xiaxue	bongqiuqiu	humsyourlife

Table 1. The top 6 users ranked by different methods.

Correlation	**Ours**	**PR**	**HS**	**UB**	**RB**	**FN-PR**	**FN-HS**	**TR**
Ours	1.00000	0.02944	-0.05467	0.03128	-0.03412	0.10967	0.11280	-0.06659
PR	0.02944	1.00000	0.52186	0.46060	0.36742	0.24603	0.23552	0.26955
HS	-0.05467	0.52186	1.00000	0.65668	0.50010	0.18476	0.19378	0.34078
UB	0.03128	0.46060	0.65668	1.00000	0.57979	0.25932	0.27937	0.31034
RB	-0.03412	0.36742	0.50010	0.57979	1.00000	0.15418	0.17571	0.29854
FN-PR	0.10967	0.24603	0.18476	0.25932	0.15418	1.00000	0.84048	0.49512
FN-HS	0.11280	0.23552	0.19378	0.27937	0.17571	0.84048	1.00000	0.47802
TR	-0.06659	0.26955	0.34078	0.31034	0.29854	0.49512	0.47802	1.00000

Table 2. Kendall rank correlation of results between rank lists by different methods.

3.2 Comparison with related algorithms

In this section, we conducted the comparisons against related algorithms. All the algorithms studied include:

1. **Proposed Method** Base on the reciprocal retweet network, we calculate the score for each user using our proposed method. In the experiment, the parameters of the penalty function $M = 1e6$, $k = 15$, and the number of iterations is 1000.

2. **PageRank** Base on the weighted retweet network, (A retweets B means a pointer from A to B), we calculate the PageRank score for each user. In the experiment, the residual probability is set to 0.85, and the epsilon is set to 1e-9.

3. **HITS** Base on the unweighted retweet network, (A retweets B means a pointer from A to B), we calculate the authority score for each user using HITS algorithm. In the experiment, the epsilon is set to 1e-9.

4. **Users-based** In this method, we rank the users based on the number of users by whom they were retweeted.

5. **Retweets-based** In this method, we rank the users based on the number of times they are retweeted by others.

6. **PageRank based on follow network** In this method, we construct the follow link network for the users. The PageRank score is calculated based on this graph. In the experiment, the residual probability is set to 0.85, and the epsilon is set to 1e-9.

7. **HITS based on follow network** In this method, we construct the follow link network for the users. The authority score of HITS is calculated based on this graph. In the experiment, the epsilon is set to 1e-9.

8. **TwitterRank** In this method, we construct the follow link network for the users, and use LDA [2] to learn topics from all the tweets of these users. Then set the weight of links as mentioned in TwitterRank, which is based on users's topic profile. Combining the ranking lists in different topics, an aggregation of TwitterRank is calculated. The number of topics $T = 50$, Dirichlet hyper-parameters $\alpha = 50/T$, $\beta = 0.1$, and the residual probability is set to 0.85.

For ease of presentation, our proposed method is denoted as **Ours**, and the related algorithms are abbreviated to **PR**, **HS**, **UB**, **RU**, **FN-PR**, **FN-HS** and **TR** respectively. Table 1 lists the top users ranked by all the methods above. Due to the limitation of the space, only top 6 users are listed.

Case Studies We first evaluate the top "valuable" users in our solution (the row starts with "Ours"). The top one is "*SoSingaporean*", who has more than 121,000 followers, and actively shares everything funny, unique and localised in Singapore. "*SoSingaporean*" is so popular that 7,476 users out of 44,152 users in our dataset retweeted him, however, he only retweeted 36 users back. The second one is "*NaomiNeo_*", who is very active online celebrity and tweeted over 34,000 tweets. She also has a lot of followers (more than 64,000) and was retweeted by 3,545 users in our dataset. It is reasonable that such kinds of users are at the top positions in our results.

Then we compare the results of different methods. Although these 8 methods make use of different information (**Ours** based on the reciprocal retweet network of 3,000 users, **PR** and **HS** based on the retweet network of 3,000 users, **UB** and **RU** based on the whole retweet network, **FN-PR** and **FN-HS** based on the whole follow network, **TR** based on the whole follow network, as well as all the tweets), except **PR** and **TR**, all the methods rank "*SoSingaporean*" as the top one, which shows the inherent value of "*SoSingaporean*". Besides, "*NaomiNeo_*" and "*stcom*" are ranked in the top 6 users by most of the methods.

The other case in our experiment is "*fakeMOE*", which is ranked low in our result, but is ranked top by other methods such as PageRank (top 7 in PageRank, not shown in Table 1). "*fakeMOE*" spoofs the official Twitter account of Ministry of Education (MOE). Being followed by more than 22,000 Twitter users and retweeted by 4,671 users in our dataset, "*fakeMOE*" is definitely a hub user, so that it succeed to disguise itself as influential user in the eyes of

PageRank and other methods. However, it is actually just a fluffing account, and few real influential Twitter users retweet this account. By examining the retweet interactions between "*fakeMOE*" and other users, our method rank it low.

The other case is one local influential news media "*stcom*" (The Straits Times), which has more than 233,000 followers, and is retweeted by 6,455 users in our dataset, doesn't appear in the top 20 list, and actually it is ranked as the 59th one by our method. By exploring the reciprocal retweet network, we found that "*SoSingaporean*" and "*NaomiNeo_*" behave like a hub in the network, i.e. they interact with some "agents" and these "agents" interact with others. Figure 5 presents the 2-layer eco-network of "*SoSingaporean*", from which we can see the reciprocal retweet network is much sparser than the follow network. However, contrary to "*SoSingaporean*" and "*NaomiNeo_*", "*stcom*" only connected with two nodes (one is an art journalist who and the other is a geek). We can only infer the "value" of "*stcom*" from these two nodes, so that we can not infer the accurate "value" of "*stcom*". The reason maybe the serious news media such as "*stcom*" retweet others very carefully and very rarely, because of the consideration of public influence. For these kind of users, the lack of such interaction behaviour makes us hard to infer the accurate "values" of them.

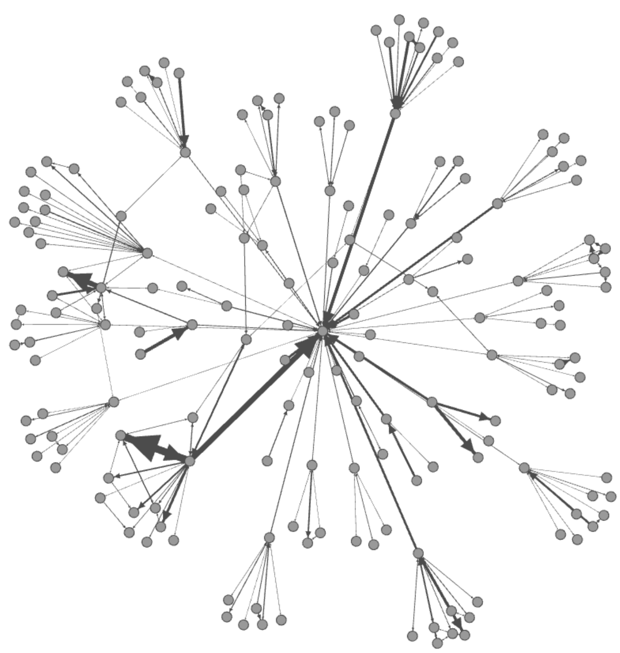

Fig. 5. The eco-network of "*SoSingaporean*"

Correlation To further study the relationship between these methods, we study the correlation between the rank lists generated by them. The correlation is measured by the Kendall rank correlation coefficient [9], which takes value in the range of $[-1, 1]$. If the two rank lists are exactly the same, the correlation coefficient is 1; if the two rank lists are independent, the correlation coefficient is 0; and it is -1 if two rank lists are opposite to each other. The larger the correlation coefficient is, the stronger the similarity of the two rank lists are.

Table 2 lists the Kendall correlation coefficients between all the rank lists generated by all the methods studied. It is observed that the rank lists generated by **PR**, **HS**, **UB** and **RU**, which all are based on the retweet network, are similar; and that the rank lists generated by **FN-PR**, **FN-HS** and **TR**, which are based on the follow network, are similar. As expected, the rank list generated by our solution does not overlap with results of all other methods, because our model aims to capture totally different values of the users in the network. However, compared with **PR**, **HS**, **UB**, **RU** and **TR**, our result is more similar to **FN-PR** and **FN-HS**. The differences between the networks these method based on and the differences between the modes these methods used lead to the differences between the results of them. And they represent the values of Twitter users in different dimensions.

3.3 Retweet Behaviour Prediction

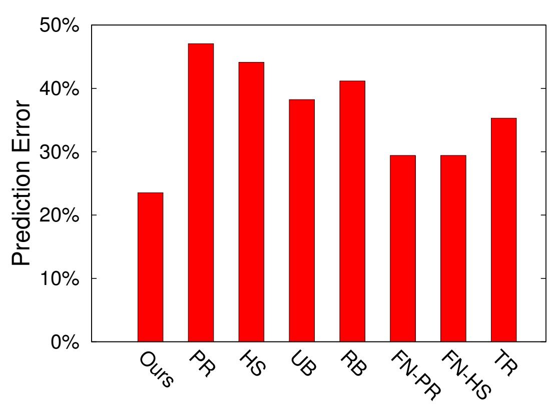

Fig. 6. Prediction error of different methods.

To verify the effectiveness of our model, in the section we conduct a user retweet behaviour prediction test. In this test, an assumption is made that for a pair of Twitter users A and B who retweet each other, if A's value is larger than B's value, then B will retweet A more than A retweet B. Based on this assumption, according to the values of Twitter users, we can predict their retweet behaviour, i.e. predict whether A retweet B more than B retweet A.

First, for all the pairs of Twitter users who retweet each other, we randomly choose 1% pairs. For each pair in them, if for both two Twitter users in this pair, their degrees in the reciprocal retweet network are no less than 3, then we remove all the retweets between them. We do this so that even all retweets between them are removed, they also have at least 2 neighbours in the rest network, which can help us to infer their values. Then based on this new network, the values of Twitter users are recalculated for our method and all other methods, except **FN-PR** , **FN-HS** and **TR**, which are based on the follow network rather than the retweet network. At last, for each pair of users between whom all the retweets are removed, according to the values provided by each method, prediction is made. (We ignore the cases that two users retweet each other equally.) Based on the ground truth (the retweets between the Twitter users in each removed pair), we verify the prediction of each method in the term of prediction error rate.

Figure 6 shows the prediction error rates of different methods. We can observe three points as below. 1) Our method outperforms other methods to a large extent, which shows that our model has better understanding of the Twitter users' retweet behaviours. 2) Though slightly worse than our method, **FN-PR** and **FN-HS** outperform all other four methods. This result is consistent with the correlation coefficients between different methods shown in Section 3.2, recalling that the rank lists generated by **FN-PR** and **FN-HS** are more similar to our method's. It also implies the inherent relationship between users' state in follow graph and their retweet behaviours. 3) The error rates of all methods presented are less than 50%, which is the expectation of randomly guessing. Under this consideration, our assumption makes sense, that in general higher-value Twitter users retweet lower-value Twitter users less.

4 Related Work

As the online social network such as Facebook and Twitter grows fast these years, there are several works to study the influential or valuable users in the online social network for purposes such as maximising the spread of influence [8] and viral marketing [14]. PageRank [11] and HITS [3], which are originally used to rank the web pages in the network which is made up of web pages, are naturally used in this new scenario to rank the users in the network which is made up of users. TwitterRank [15] extends PageRank by introducing a new dimension of the topics of tweets. However, both PageRank and HITS are derived based on their own assumptions. PageRank assumes there is a surfer randomly visiting the web pages. HITS considers a academic scenario in which there are two roles: authority and hub. Due to the limitations of the assumptions, PageRank and HITS doesn't take into account the interactions of users in social network, which may be the key point to disclose the values of users in social network. Under this consideration, we proposed a quite different model, in which Twitter users' retweet behaviours are treated as reciprocal social behaviours. In this scenario, the inherent values of users are determined by the continuous interaction between them.

Besides, other works which consider the social features includes [12] and [7]. [12] is an application of HITS in the Twitter setting. It identifies influential users who are able to diffuse information quickly and influent others effectively. It introduces the “passive users” who are reluctant to be influenced. In this model, higher value in the ranking implies that that particular user cans even influent most passive users. [7] models vitality and susceptibility in Twitter. In this model viral information diffusion is due to viral users, viral items and susceptible users. These models provide the other directions to measure the users in the social network.

5 Discussion

In this section, we try to explain the model built in Section 2.2 from an economic perspective. In particular, we try to measure the economic value of the retweet, which leverages the power of word of mouth to help information dramatically spread over the whole Twitter network, and makes some tweets to reach a large number of audiences and to gain huge influential impacts.

In fact, each tweet has its own influential economic value. For example, when a satisfied iPhone user posted a positive tweet about iPhone; that tweet potentially reaches a large number of followers, triggering viral marketing effect for Apple. Eventually, this can help increasing the sales of Apple. In this case, the resulting difference in Apple's sales reflects the economic value of that tweet. Not just only the tweet, retweet also plays an important role in this picture as the power of original tweet is strengthened exponentially by the number of retweets. In term of economic value, it is fair to expect that retweet even has higher value than the tweet itself. The original tweet, most of the time, only expresses anticipatory or evaluative opinion of individual [6]. On the other hand, if someone retweets that original tweet, that action implies that the original opinion is verified, adopted and forwarded to other users.

Now that retweet has big economic value, and there is no such thing as a free lunch, we suggested that there would be an underlying “*virtual retweet market*”, on which Twitter users's retweeting behaviours are based. In this virtual market, for common benefits, Twitter users “*trade*” with each other by exchanging their retweeting behaviours, i.e. retweeting each other. For example, A retweeted B 3 times, and in reciprocation, B retweeted A 2 times. In this case, A makes a deal with B using its 3 retweets for B's 2 retweets. After conducting such kind of trade, their influence is extended by increasing the numbers of their audiences from the followers of others. We further assume that this virtual market is a free price system without external effects [5], in which the prices of good and services are eventually determined by the exchanging behaviour of users [1]. In this system, the interchange of retweeting behaviours determines the prices, which reflect the economic “*values*” of Twitter users. We can mathematically formalise the “*trade*” behaviours between Twitter users as Equation 2 in Section 2.2. It means that each pair of Twitter users make a fair “*trade*” according to their economic “*values*”.

As electronic commerce develops quickly, it is very possible that this virtual retweet market comes true as a real market in the future. In this scenario, our model can be the basis of this market by calculating the prices of Twitter users' retweeting behaviours.

6 Conclusions and future work

Finding the valuable users in social network is a quite motivated problem due to the potential commercial interest. Rather than from a perspective of information flow, this paper re-examine the value of users in social network from the social interaction perspective. In particular, we consider the social interaction in the notion of reciprocity based on the retweet interaction between Twitter users. We formulated the retweet reciprocity, proposed an alternative user ranking model based on retweet reciprocity and developed efficient inference solution. Our experiments on real Twitter data demonstrated that our proposed model presents different yet equally insightful ranking results. The conducted prediction test also showed the correctness of our model. Besides, we also discuss the meaning of our proposed model from an economic perspective, and explain Twitter users' retweeting behaviour as economic behaviour.

Our paper is just a preliminary study, which still needs a lot of improvements. First, as the experimental results show, there are still some real influential users such as "*stcom*" are not ranked top in our ranking list, which is due to the lack of enough interactions of these users. We plan to incorporate the different kinds of interactions in a social platform, and find influential users by combining all such kinds of interactions. Second, we use gradient descent method to infer the values of users, which is not efficient enough to handle large scale social data. We also plan to improve this by developing approximate efficient algorithm. Third, in near future, social networks will evolve dramatically. Future work of this research will consider the interaction of users in community as well as focus on the interaction between communities. At last, one feasible direction is to add the topic dimension as in TwitterRank [15], and study the interactions between users in different topics.

References

1. Altvater, E., Camiller, P.: The future of the market: An essay on the regulation of money and nature after the collapse of'actually existing sociali. (1993)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3** (2003) 993–1022
3. Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., Gibson, D., Kleinberg, J.: Automatic resource compilation by analyzing hyperlink structure and associated text. Computer Networks and ISDN Systems **30**(1) (1998) 65–74
4. Easley, D., Kleinberg, J.: Networks, crowds, and markets: Reasoning about a highly connected world (2012)
5. Gregory, P.R., Stuart, R.C.: Comparing economic systems in the Twenty-first century. Houghton Mifflin (2004)

6. Hennig-Thurau, T., Wiertz, C., Feldhaus, F.: Exploring the twitter effect: an investigation of the impact of microblogging word of mouth on consumers early adoption of new products. Available at SSRN 2016548 (2012)
7. Hoang, T.A., Lim, E.P.: Virality and susceptibility in information diffusions. In: Sixth International AAAI Conference on Weblogs and Social Media. (2012)
8. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2003) 137–146
9. Kendall, M.G.: A new measure of rank correlation. Biometrika **30**(1-2) (1938) 81–93
10. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, ACM (2010) 591–600
11. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. (1999)
12. Romero, D., Galuba, W., Asur, S., Huberman, B.: Influence and passivity in social media. Machine Learning and Knowledge Discovery in Databases (2011) 18–33
13. Sahlins, M.D.: Stone age economics. Aldine de Gruyter (1972)
14. Shakarian, P., Paulo, D.: Large social networks can be targeted for viral marketing with small seed sets. In: Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, IEEE (2012) 1–8
15. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on Web search and data mining, ACM (2010) 261–270