International Journal on Artificial Intelligence Tools

The Role of Activity and Similarity in Rating and Social Behavior in Social Recommender Systems

Carine Pierrette Mukamakuza
E-Commerce Research Division, TU Wien, Favoritenstrasse 9-11/194-4
Vienna, Austria
carine@ec.tuwien.ac.at

Dimitris Sacharidis
E-Commerce Research Division, TU Wien, Favoritenstrasse 9-11/194-4
Vienna, Austria
dimitris@ec.tuwien.ac.at

Hannes Werthner
E-Commerce Research Division, TU Wien, Favoritenstrasse 9-11/194-4
Vienna, Austria
werthner@ec.tuwien.ac.at

Social recommender systems exploit two sources of information for making recommendations, the historical rating behavior of users, and the social connections among them. The basic assumption is that if two users are friends, they are likely to share similar preferences. Many recommendation approaches are based on such correlations between the rating and the social behavior of users. However, there is little work in studying whether there actually exist such correlations and how strong they are. In our work, we look at the two views of user behavior, their social connections, and their history of ratings, and investigate two research questions. The first examines if strong activity in one view, e.g., having many friends, implies strong activity in the other view, e.g., having rated many items. The second investigates whether high similarity in one view, e.g., network similarity, implies high similarity in the other view, e.g., rating similarity. We employ various notions of activity and similarity, and identify those that appear to have the stronger impact. Specifically, to some degree, we find that rating behavior determines social behavior, and that the opposite relationship is weaker.

Keywords: Social recommender systems; network analytics; collaborative filtering.

1. Introduction

Traditionally, people turn to their friends, or those whose opinions they trust, when looking for advice and recommendations on a specific domain. *Social Recommender Systems* attempt to mimic this behavior by also drawing information from the social context of the users. The underlying assumption is that the decision process of a user depends not only on her individual preferences, but also on influence she receives from her social connections. This is motivated by the phenomena of homophily

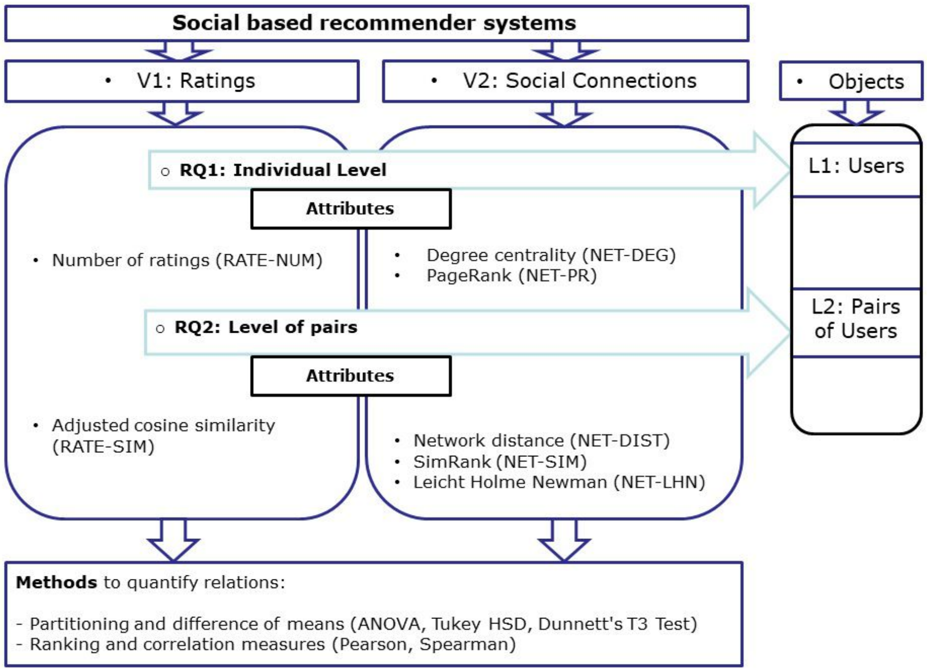


Fig. 1: An illustration of our methodology for studying relationships between rating and social behavior in social recommender systems.

and influence in social networks [1]; the former suggests that users socially connect because they have similar interests, while the latter says that socially connected users tend to develop similar interests.

Specifically, social recommenders exploit two distinct sources of information, the historical rating behavior of users, exactly like collaborative filtering techniques, and the social network, and assume that these sources are correlated, implying that the latter can provide additional information about the former. As an example, consider the role of the neighborhood in user-user collaborative filtering, e.g., in [2]. The users in the *rating neighborhood* of a target user essentially provide additional knowledge about the preferences of the target user, so that the system can make better informed recommendations. In analogy, in social recommenders, e.g., [3], the users connected to the target user, i.e., her *social neighborhood*, also provide complimentary information about the preferences of the target user.

Social recommendations is an active research area in the past few years, e.g., refer to [3,4,5,6,7,8,9,10,11,12]. However, there is little work [13] in studying whether the assumptions put forward by these recommenders are actually true and to what extent. In our work [14,15,16], we seek to formulate concrete research questions that investigate statistical correlations between social and rating behavior of users. The goal is to validate the assumptions typically made, and better understand the connections between social and rating behavior so as to design more effective social recommenders.

Our methodological approach is illustrated in Figure 1. We consider two *views*, the historical rating behavior (V1), and the social connections (V2) of users, corresponding to the two sources of information available to a social recommender. The goal of our study is to examine whether connections between these two views exist. More concretely, we define several *attributes* capturing important aspects of each view, and then observe whether there is a correlation between their value distributions. We discern two types of attributes: those that concern users individually, which we call level 1 (L1) attributes; and those that quantify relations between pairs of users, which we call level 2 (L2) attributes.

Level 1 (L1) attributes characterize users based on their volume of *activity*. For example, users can be highly active raters in the system, providing lots of feedback, or highly active socially, possessing a central position in the network. Level 2 (L2) attributes quantify the *similarity* between pairs of users. For example, two users can have similar preferences in terms of ratings, or be socially similar in terms of their network distance or the number of common friends.

Note that there are two separate ways to classify attributes: based on the view they correspond to (V1 or V2), and based on the level they are defined (L1 or L2). As attributes from different levels cannot be compared, we seek to examine connections at each level separately. Therefore, we pose two research questions. At the first level L1, the first research question (RQ1) asks: Does strong *activity* in one view imply strong activity in the other? Here activity in terms of rating behavior (V1) is captured by the number of ratings of a user (denoted as RATE-NUM). In terms of social connections (V2), activity is quantified by some measure of node centrality; we consider degree centrality (NET-DEG) and PageRank (NET-PR). Essentially, in RQ1 we seek for correlations between “heavy raters” and “popular users” [15].

At the second level, the second research question (RQ2) asks: Does high *similarity* between users in one view implies high similarity in the other? Similarity in terms of rating behavior (V1) quantifies how similar the ratings given by two users is. For this purpose, we use the widely popular adjusted cosine similarity [17] (RATE-SIM), which is related to Pearson’s correlation coefficient. To address the case of implicit feedback, we measure the Jaccard index (RATE-JAC) of the sets of interactions between two users. For the second view, we consider various notions of network similarity, namely network distance (NET-DIST), the SimRank similarity[18] (NET-SIM), and the Leicht Holme Newman index[19,20] (NET-LHN).

For both research questions, we quantify relationships between the objects of study (users in L1 and user-pairs in L2) between the two views using two methods. In the first, termed *partitioning*, we partition objects based on an attribute of one view (e.g., rating activity RATE-NUM), and investigate how the mean of an attribute of the other view (e.g., mean social activity in terms of NET-DEG) varies across partitions. This shows, for example, if degree centrality (NET-DEG) increases along

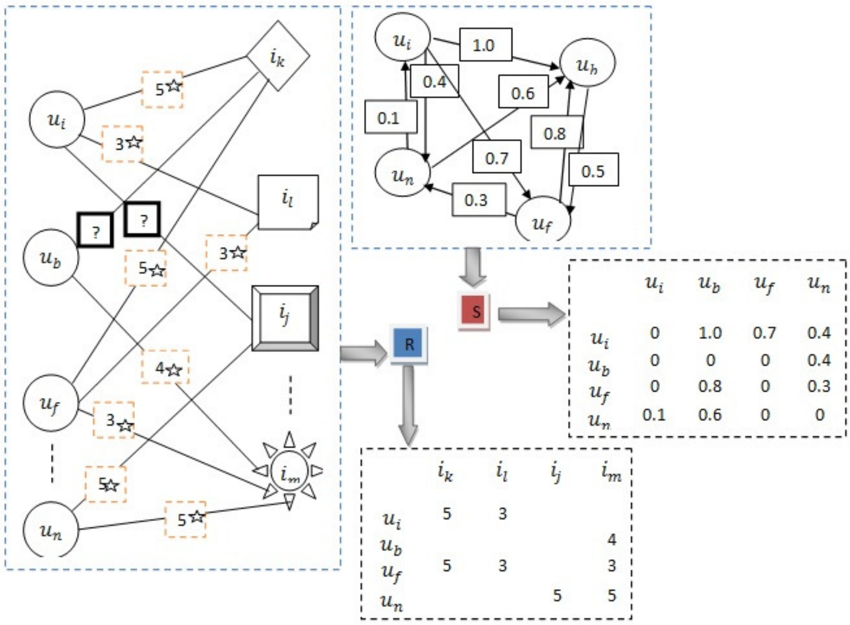


Fig. 2: The ratings matrix and the social adjacency matrix are the two source of information used by a social recommender system.

with the number of ratings (RATE-NUM). In the second method, termed *ranking*, we compile two rankings of objects based on attributes of each view, and study whether statistical correlations between the rankings appear. For example, we may compute how many pair of users are both highly similar (e.g., in the top-100) in terms of rating behavior (RATE-SIM) and in terms of social similarity (NET-LHN).

The results of our study can be summarized as follows. At the level of users, we see that the number of ratings made by a user and her centrality in social network are related, particularly when the latter is measured in terms of the number of social connections. At the second level, we see that rating similarity between pairs of users is related primarily with their network distance and with a particular metric of network similarity, SimRank. At both levels, rating and social behavior seem to be related, and specifically, rating seems to determine social behavior more strongly than the other way around.

The remainder of this paper is structured as follows. Section 2 establishes the necessary background and overviews existing work, and Section 3 describes our methodology. Sections 4 and 5 present experimental results of our first and second research question, respectively, while Section 6 draws conclusions.

2. Related Work

Social recommenders borrow ideas from collaborative filtering techniques, but also use additional information from the social connections of users.

Collaborative Filtering. Social recommender systems borrow ideas from *Collaborative Filtering* (CF), which is the most commonly used method for making recommendations. In CF approaches, users and items with similar rating patterns are taken into account [21] to produce a recommendation for the target user.

The basic entity in CF is the *user-item ratings matrix*, composed of a set of items $I = i_1, ...i_n$ and a set of users $U = u_1, ...u_m$. The ratings matrix $R \in \mathbb{R}^{n \times m}$ contains the ratings given by users to items, where n represents the number of items and m number of users. CF exclusively uses the ratings in R to make recommendations. The left and bottom part of Figure 2 depict the historical rating behavior of users in terms of the ratings matrix R.

Memory-based methods for CF are divided into two categories. *User-user* techniques make the assumption that users had similar tastes in the past they are most likely to have the same tastes in the future, i.e., user preferences then to remain constant and stable over the time. Then to predict ratings of a target user, they utilize the ratings to the target item by a set of the users whose similarity level is closer to the target user, the neighborhood. On the other hand, *item-item* methods uses the target user's profile to compute the target item's similarity to other items rated by the target user.

Model-based methods make predictions by learning parameters describing how ratings are generated. The most famous is the *Matrix Factorization* (MF) technique [22,23,24]. In its simplest incarnation, MF computes a low-rank approximation of the sparse ratings matrix R by multiplication of two matrices: a user-feature matrix U and an item-feature matrix V, both involving $k \ll \min\{m, n\}$ features. A rating R_{ij} is then predicted by the dot product of the user-feature vector U_i and the item-feature vector V_j.

Social Recommender Systems. Social recommenders (SR) operate similar to collaborative filtering systems but differ in that they make recommendations taking also into account the social connections between users. The latter is conveyed by the social adjacency matrix S, where an entry portrays the connection strength between the corresponding users. The top and right part of Figure 2 present an example of connections among users, which are encoded in matrix S. Social recommenders combine information contained in matrices R and S. In the following, we review the most important related work, differentiating between *memory-based* and *model-based* SRs. For an overview of this research area and other associated topics, we refer the reader to [11].

Memory-based social recommenders apply ideas from memory-based collaborative filtering to combine information from the social graph and the past user behavior. In Trust-aware Recommender systems (TaRS) [3], the idea is to treat the social neighborhood of the target user in a manner similar to the rating neighborhood in user-based CF. Following TaRS, several works have recently appeared. An experimental evaluation of several memory-based social recommenders is provided in [12]. The authors also propose to fuse recommendations from friends with recommendations from implicit social relations, and show that such an approach improves accuracy and increases coverage.

Homophily in social networks refers to the notion that similar users tend to be socially connected and vice versa. In the context of social recommenders, the work in [6] studies homophily on two online social media networks, BlogCatalog and Last.fm by extracting communities based on the network ties. Similarly, [8] investigates the presence of homophily in three systems that combine tagging social media with online social networks.

The other important category in SR is *model-based social recommenders*, where model-based collaborative filtering, and predominantly matrix factorization, approaches are used. One of the first works in this direction is SoRec [4] that extends the basic MF model to incorporate the social network. The social adjacency matrix S is factorized into a user-specific matrix U and a factor-specific matrix F, where matrix U is also part of the factorization of the ratings matrix. The latent feature vectors of users are then learnt based on both the rating and social network matrices.

Social trust ensemble [5] builds on the hybrid idea of [3], and defines a linear combination of basic MF predictions with social network predictions. In Social Regularization [25,7], the key idea is to use the basic MF formula for predicting ratings, but use regularization terms to force the learned user feature vectors to be similar between friends. In [9], contextual information and social network information are combined to improve quality of recommendations. In the community-based models of [10], the idea of social regularization is taken one step further. A target user can belong to different communities, and each community should be regularized individually.

There is a common assumption in all of the related work: if two people are socially connected, then they must have similar preferences. This assumption is adopted by all social recommenders but to a different extent. Some of the proposed methods, e.g., [4,5,25], go to the extreme, as they explicitly mandate that two friends should have similar preferences (user features). This approach completely ignores the fact that the degree of influence/homophily may actually vary among friends.

More recent methods based on social regularization, e.g., [7,10,26], acknowledge that not all pairs of friends should be treated equally. Instead, they force two friends to have similar features to the degree that their observed rating behavior is similar. At first, this may seem like a more realistic model, but upon a more detailed inspection, it sort of defeats the purpose of using a social recommender. If two friends exhibit very similar rating behavior, then any basic matrix factorization model, which is agnostic to the two users' social connections, should be able to understand this relationship on its own, and assign similar features to these users anyway. All social regularization does is to make this even more explicit for the underlying model, asking it to ensure that similar friends have similar features. So in this case, social recommenders do not do anything different than plain old collaborative filtering.

Now consider the case of two friends, among whom at least one is a cold-start user so that their observed rating behavior is not similar — at least so far. In this setting, a social regularization-based recommender would mandate that these two friends should not be assigned similar features, again much like basic collaborative filtering would. However, this ignores the possibility of social influence between these friends, which is exactly the premise behind social recommendation: when there is little information in the ratings matrix to work with, augment it with social connections.

The success of such social recommenders brings into question the validity of the social recommendation assumption. When and how should we use social connections to augment collaborative filtering. To answer this, we must examine in detail the relationship between the two sources of information available, the ratings (view V1) and the social connections (view V2). We note that although previous work, e.g., [13], has investigated some of these relationships, they have done so in a non-systematic way, like we propose here. We believe that the findings of our work can be exploited to design more realistic and effective social recommenders.

3. Research Methodology

Our research methodology, summarized in Figure 1, involves examining the two behavior views, on ratings (V1) and on social connections (V2), at two levels, users (L1) and pairs of users (L2). At each level, we consider several attributes, from both views, describing the object of study (users or pairs of users), and we seek to quantify correlations between attributes of different views. Section 3.1 presents the attributes used, and Section 3.2 discusses how we study correlations. Section 3.3 presents the dataset used in our study.

3.1. *Examined Attributes*

This section presents the attributes used to describe the objects of study, users at L1, pairs of users at L2. At the first level, the attributes capture the level of *activity* of users in terms of their rating behavior and social connections. At the second level, the attributes capture the *similarity* between two users, again in terms of their rating behavior and social connections.

3.1.1. *Attributes Capturing Activity of Users (L1)*

We consider one notion of activity in terms of rating behavior, and two notions in terms of social connections, based on the concept of node centrality [27].

RATE-NUM. For the rating behavior, we consider the number of ratings a user has provided, denoted as RATE-NUM. This essentially, captures how "*heavy*" rater a user is.

NET-DEG. Degree is the most intuitive interpretation of popularity, as it counts the number of (incoming or outgoing) connections a user has. In terms of the adjacency matrix S, the Degree of user u_i is

$$d_i = \sum_{k=1}^{m} (S_{ki} + S_{ik}).$$

NET-PR. PageRank [28] depends on the number of incoming connections of a user as well as their quality, with higher centrality users giving more importance to their outgoing connections; in some sense, the higher its PageRank is the more respected a user is. In terms of the adjacency matrix S, PageRank satisfies the equation

$$x_i = \alpha \sum_{k=1}^{m} \frac{S_{ki}}{\max\{d_k^{out}, 1\}} x_k + \frac{1-\alpha}{m},$$

where $d_i^{out} = \sum_{k=1}^{m} S_{ki}$ is the out-Degree centrality of user i, and α is the damping factor, typically set to 0.85.

3.1.2. *Attributes Capturing Similarity Between Two Users (L2)*

We consider one notion of similarity in terms of rating behavior, and three notions in terms of social connections.

RATE-SIM. The pairwise cosine similarity metric finds the normalized dot product of the rating vectors of two users [29]. This simple defition however has some limitations. It is known that people's tend to rate on different scales. Some people are naturally high raters which means they might rate items highly in general, even if they do not like the item very much. There are some people who tend to rate low, even when they like the items very much. The traditional cosine similarity does not consider the difference in rating scale between different users [30]. The adjusted cosine similarity offsets this drawback by subtracting the corresponding user average from each co-rated pair. Formally, the similarity, denoted as RATE-SIM, we use between users u and $v$is given by:

$$sim(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u).(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u} (r_{ui} - \bar{r}_u)^2}.\sqrt{\sum_{i \in I_v} (r_{vi} - \bar{r}_v)^2}},$$

where I_u and I_v are the sets of items rated by user u and v, r_{ui} is the rating user u gave to item i and $\bar{r}_u$ the average of all ratings given by u.

NET-DIST. Network distance between two users is a minimum number of connections, or links, that separate them in the network. It can also be defined as the length of the shortest path between two users. The algorithm of Floyd-Warshall[31] can be used to determine the network distance of all pairs of users.

NET-SIM. The idea behind *SimRank* is simple: two users are similar if they are referenced by similar users [18,32]. Each user is considered to be completely similar to herself, which gives it a similarity score of 1. The similarity $SR(u, v)$ between users u and v takes values in $[0, 1]$, and satisfies a recursive equation. If $u = v$ then $SR(u, v)$ is defined to be 1. Otherwise,

$$SR(u,v) = \frac{C}{|N(u)||N(v)|} \sum_{u' \in N(u)} \sum_{v' \in N(v)} SR(u', v'),$$

where C is a constant between 0 and 1, and u', v' are in-neighbors of users u and v, belonging to the sets $N(u)$ and $N(v)$, respectively. A detail here is that either u or v may not have any in-neighbors. Since there is no way to assume any similarity between u and v in this case, SimRank is set to $SR(u, v) = 0$, which makes the addition of the main equation to be 0 when $N(u) = \emptyset$ or $N(v) = \emptyset$.

NET-LHN. The *Leicht Holme Newman* index[19,20] counts the expected number of common neighbors between two users. For users u and v the NET-LHN is computed as:

$$LHN(u,v) = \frac{|N(u) \cap N(v)|}{d_u \times d_v},$$

where $N(u)$ is the neighborhood of user u, and d_u is the degree of u. Intuitively, NET-LHN assigns a high similarity score to pairs of users that have many common neighbors[33].

3.2. *Quantifying Relationships*

To establish whether a relationship between objects' attributes between the two views exists, we consider two approaches. In the first, termed *partitioning*, we divide objects into three partition A, B, C according to an attribute of one view, called the *partition attribute*. Partition A contains objects with low activity (users in L1) or similarity (pairs of users in L2), while partition C contains objects with high activity or similarity. In each partition, we compute the mean of an attribute of the other view, called the *test attribute*. Then we examine if the test attribute increases with the partition attribute. For example, for partitions based on RATE-NUM, we compute the average NET-DEG (or NET-PR), and see whether mean NET-DEG increases from partition A though C. To formally test if there is a trend, we first apply ANOVA to investigate whether the mean of the test attribute is significantly different across partitions. ANOVA tells us whether results are significant overall, but it does not reveal exactly where those differences lie. Therefore, if the ANOVA test is positive, we perform a post hoc analysis, Tukey's HSD [34] or Dunett's T3 [35] test, on pairwise differences of the partition means (B-A, C-B, C-A), to whether a trend in the test attribute is significant.

For the second approach, termed *ranking*, we create a rankings of objects based on each attribute, and retain only those that have the highest activity (L1) or similarity (L2); the selected attribute is called the *ranking attribute*. For example, we may construct the ranking of the top-100 most heavy raters (RATE-NUM) in the system. Then, we look for correlations between the two views (ratings and social behavior) in two ways. In the first, we pick two rankings produced by attributes of different views, and count the number of common objects in them. For example, we see how many users are both among the top-100 most heavy raters (RATE-NUM) and the top-100 most well connected users (NET-DEG). In the second ranking method, we pick one ranking, and study the correlation, measured by Pearson's and Spearman's correlation coefficients, between two attributes of different views. For example, we select the 100 most similar pairs according to their ranking (RATE-SIM), and see how their RATE-SIM correlates to NET-SIM.

3.3. *Datasets*

In our study, we use two publicly available datasets, FilmTrust[36] and CiaoDVD[37], collected from traces of user interaction in social recommenders. These data are commonly used in the literature and contain rating activity, i.e., a ratings matrix R, as well as information about the social connections among users, i.e., an adjacency matrix S. In this paper, we report results on the first dataset, FilmTrust, as results on the second show similar trends; additional results on CiaoDVD for the first research question can be found in [15].

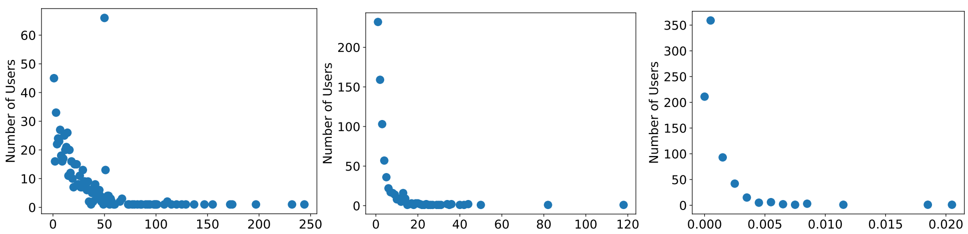
FilmTrust contains data from a social networking site[a] in which users can rate and review movies. FilmTrust essentially contains two sub-datasets, a social network in addition to the user-item ratings matrix. The social connections are bidirectional and capture the trust between users (trustee, trustor). Users can specify a level of trust, but due to sharing policy, the dataset only contains information on whether a connection exists.

FilmTrust contains 1,508 users, 2,071 items, 35,497 ratings, and 1,853 social connections. As there exist 635 users with no social connections, and 133 with no ratings history, we exclude them from our analysis. That is, we only consider the 740 users that have rated at least one item and trust, or are trusted by, at least another person. This results in 273,430 pairs of users to considered in L2. The mean number of ratings per user is 23.5 with the minimum and the maximum being 1 and 244. The ratings scale is from 0.5 to 4 with a step 0.5, and the mean rating score over all ratings is 3.0.

Figure 3 shows the distributions of RATE-NUM, NET-DEG, and NET-PR on the dataset. We note that the mean NET-DEG is 4.7, with min and max values of 1 and 118, and that the mean NET-PR is 0.0012, with the min and max values of 0 and 0.21. These right-skewed distributions show that the majority of users give few ratings and have low centralities, and that at the same time there exist several users that are very heavy raters and very central.

[a] `http://trust.mindswap.org/FilmTrust`

Figure 4 shows scatter plots of RATE-SIM with each attribute (NET-DIST, NET-SIM, NET-LHN) of V2. We note that a point in these plots represents a pair of users. The color of a pair corresponds to the group (A, B, or C) it is partitioned into.



(a) RATE-NUM (b) NET-DEG (c) NET-PR

Fig. 3: Probability distribution of a user having specific values of NUM-RATE, NET-DEG, and NET-PR

4. The Role of Activity in Rating and Social Behavior (L1)

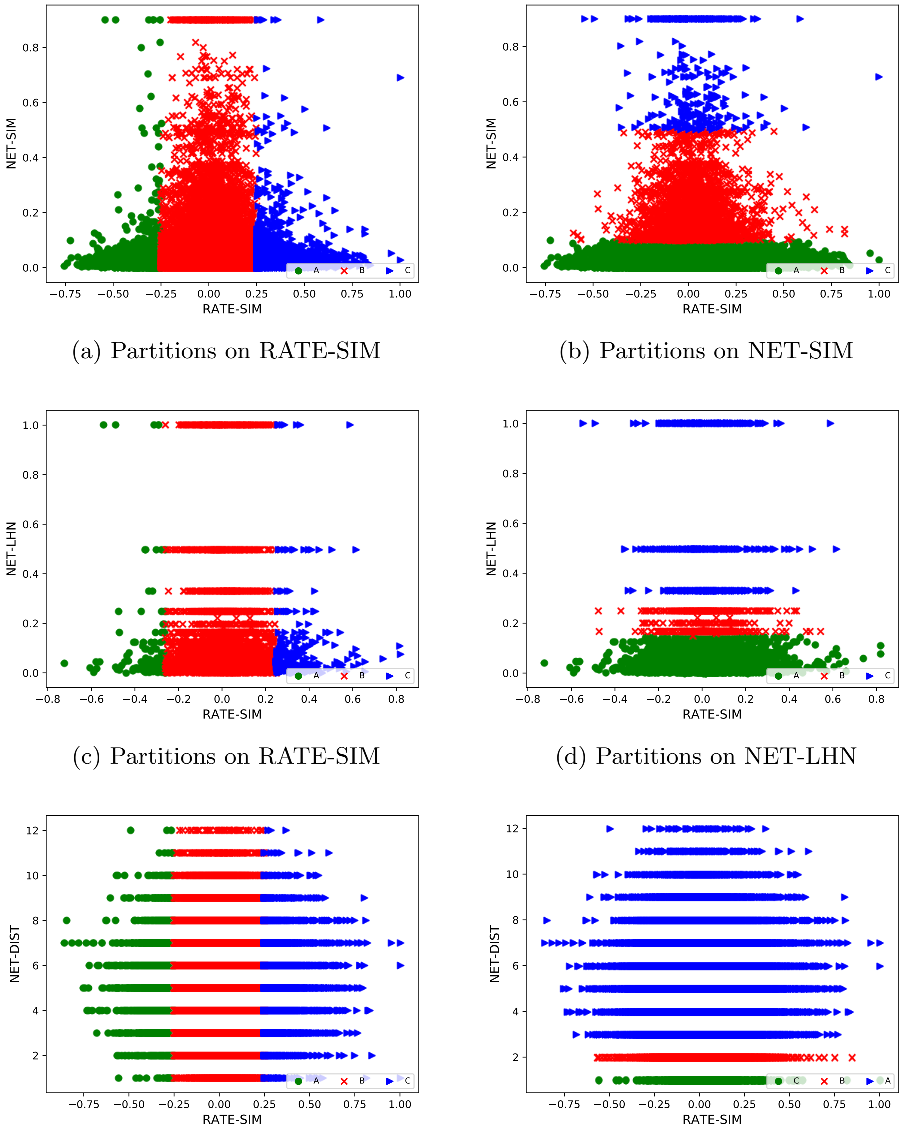
We present the results from applying our methodology on the level of users.

4.1. *Results from Partitioning Method*

To assess the relationship between activity in the rating and in the social behavior, we consider three distinct divisions, one per each attribute of L1, RATE-NUM, NET-DEG, NET-PR. A division splits users into three partitions, A, B, C, in increasing value of the partitioning attribute. We first determine the lower and upper terciles (3-quantiles) of the partitioning attribute and divide accordingly. Partitions are thus balanced, with each containing roughly 1/3 of all users. Descriptions of the partitions are shown in Table 1.

Does the mean NET-DEG differ across RATE-NUM partitions? In the first experiment, we partition users according to their RATE-NUM, and compute the mean NET-DEG in each partition. Then, we apply ANOVA to investigate whether the mean NET-DEG is significantly different across partitions. The results show an F value of 24.4 that provides significant evidence against the hypothesis that the means are equal (p-value in the order of 10^{-11}).

Following this result, we investigate whether the mean NET-DEG increases from partitions A through C. We apply the Tukey's HSD test to check every pair of partitions and see if the difference of their mean NET-DEG is significant. The difference of means and its corresponding 95% confidence interval for each pair are shown in Table 2. As suspected partitions A and B, containing non-heavy raters, have mostly similar mean NET-DEG and no significant difference is observed. However, there is



(a) Partitions on RATE-SIM (b) Partitions on NET-SIM

(c) Partitions on RATE-SIM (d) Partitions on NET-LHN

(e) Partitions on RATE-SIM (f) Partitions on DIST-NET

Fig. 4: correlation of RATE-SIM values and Network similarity(NS) values among partitions

a significant difference when we compare either A or B with partition C of heavy raters.

Does the mean RATE-NUM differ across NET-DEG partitions? We also

Table 1: Description of Partitions for Activity of Users

	lower div	upper div
RATE-NUM	11	30
NET-DEG	1	4
NET-PR	5.4×10^{-4}	1.1×10^{-3}

Table 2: Tukey's HSD Test on Mean NET-DEG among RATE-NUM Partitions

Pair	Diff. of Means	95% CI	p-value
B - A	0.861	$[0.716, 2.439]$	0.40
C - B	3.565	$[1.987, 5.143]$	$\mathbf{4 \times 10^{-6}}$
C - A	4.427	$[2.849, 6.004]$	$\approx \mathbf{0}$

study the reciprocal association. The ANOVA analysis based on the mean RATE-NUM among partitions based on NET-DEG gives an F value of 11.49 providing significant evidence against the hypothesis that the means are equal (p-value in the order of 10^{-5}). The Tukey's HSD test, depicted in Table 3, shows that partitions A and B of non-popular users have mostly similar mean RATE-NUM and no significant difference is observed. However, there is a significant difference when we compare B with C, and of course A with C, implying that popular (in terms of NET-DEG) users tend to be heavier raters.

Table 3: Tukey's HSD Test on Mean RATE-NUM among NET-DEG Partitions

Pair	Diff. of Means	95% CI	p-value
B - A	-0.313	$[-5.729, 5.103]$	0.99
C - B	9.729	$[4.317, 15.145]$	$\mathbf{8.1 \times 10^{-5}}$
C - A	9.416	$[3.91, 14.832]$	$\mathbf{1.4 \times 10^{-4}}$

Does the mean NET-PR differ across RATE-NUM partitions? We repeat the previous setup, this time measuring popularity by means of NET-PR. Table 4 present the results. The findings are similar, except with slightly lower significance: rating heaviness implies social popularity.

Table 4: Tukey's HSD Test on Mean NET-PR among RATE-NUM Partitions

Pair	Diff. of Means	95% CI	p-value
B - A	0.000226	$[-8.6 \times 10^{-5}, 0.00053]$	0.20
C - B	0.000517	$[2.04 \times 10^{-4}, 0.00082]$	$\mathbf{3.2 \times 10^{-4}}$
C - A	0.000742	$[4.3 \times 10^{-4}, 0.00105]$	$\mathbf{1.0 \times 10^{-5}}$

Does the mean RATE-NUM differ across NET-PR partitions? Finally, we consider NET-PR partitions and study whether they contain users with significantly different RATE-NUM. Results are presented in Table 5. As in the case of NET-DEG partitions, popularity implies heaviness.

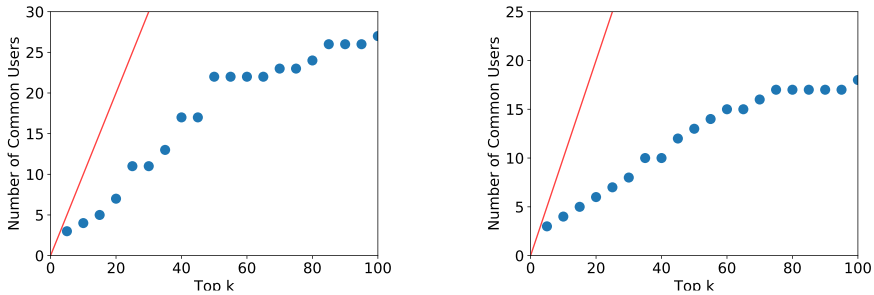
Table 5: Tukey's HSD Test on Mean RATE-NUM among NET-PR Partitions

Pair	Diff. of Means	95% CI	p-value
B - A	-0.239	$[-5.521, 5.043]$	0.99
C - B	9.948	$[4.666, 15.230]$	$\mathbf{3.3 \times 10^{-5}}$
C - A	9.709	$[4.427, 14.991]$	$\mathbf{5.3 \times 10^{-5}}$

4.2. *Results from Ranking Method*

We now seek correlations among the top-100 users according to RATE-NUM, NET-DEG, and NET-PR; these rankings contain about 13% of the users.

How many common users exist among the top-100 heavy and the top-100 popular? First, we consider the number of common users across these rankings, with the results shown in Figures 5. We see that the number of common users increases with a much lower rate that the maximum possible (drawn as the red line). Hence, the ratio of common users is higher when we look at the top of the rankings.



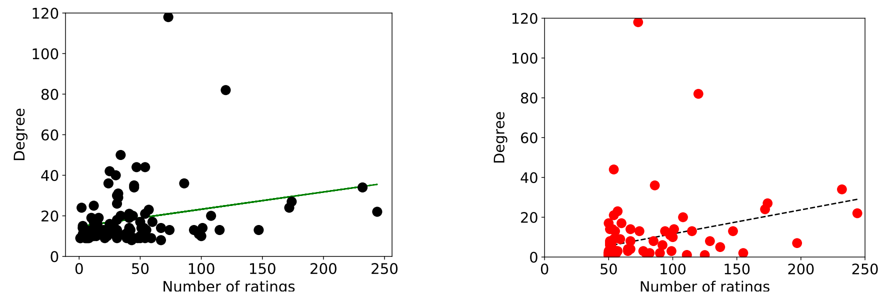
(a) RATE-NUM and NET-DEG (b) RATE-NUM and NET-PR

Fig. 5: Number of common users among the Top-K heaviest and most popular (NET-DEG, NET-PR) users

Are RATE-NUM and NET-DEG correlated? We investigate whether high rating (RATE-NUM) and social (NET-DEG) activity are correlated. Figure 6a shows the values of NET-DEG and RATE-NUM for each user among the top-100 most popular users (according to NET-DEG), while Figure 6b shows the corresponding scatter plot for the top-100 most active users by RATE-NUM. In both figures, we draw the linear regression line, and also measure Pearson's and Spearman's correlation coefficients. The very popular users have weak Pearson's and Spearman's correlation values of 0.25 and 0.27 with low significance (p-values of 0.01 and 0.07). In contrast, the very heavy users have weak Pearson's but strong Spearman's correlation values of 0.3 and 0.67 with high significance (p-values of 0.002 and ≈ 0).

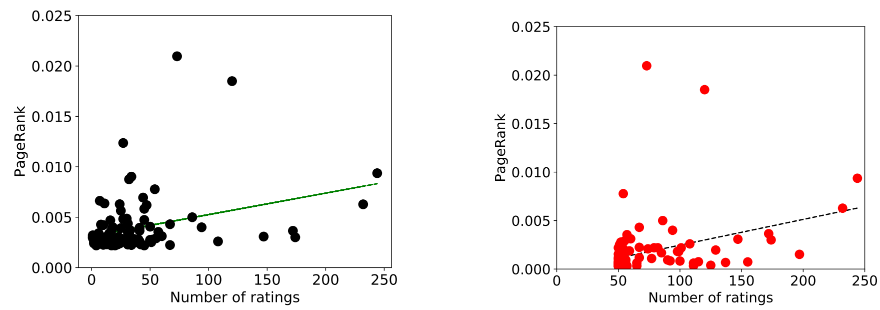
Are RATE-NUM and NET-PR correlated? We repeat the previous setup but this time define social activity by NET-PR. Figure 7 shows the results, where the top-100 users by NET-PR have an insignificant weak correlation between RATE-NUM and NET-PR. On the other hand, the very heavy raters exhibit moderate to strong correlations (Pearson's and Spearman's correlations of 0.35 and 0.60) with high significance (p-values 0.004 and ≈ 0).

As a conclusion, we note that we have observed moderate to strong correlations in the most active raters (top-100 by RATE-NUM) between their rating (RATE-NUM) and their social (NET-DEG and NET-PR) activity. This correlation is not so much linear, as is rank-based (higher Spearman's than Pearson's correlation values).



(a) Top-100 users by NET-DEG (b) Top-100 users by RATE-NUM

Fig. 6: Scatter Plots (RATE-NUM, NET-DEG)



(a) Top-100 users by NET-PR (b) Top-100 users by RATE-NUM

Fig. 7: Scatter Plots (RATE-NUM, NET-PR)

Table 6: Description of Partitions for Similarity between Pairs of Users

	upper div	lower div
NET-DIST	2	3
NET-SIM	0.25	0.15
NET-LHN	0.5	0.1
RATE-SIM	0.25	-0.25

5. The Role of Similarity in Rating and Social Behavior (L2)

We present the results from applying our methodology on the level of pairs of users.

5.1. *Results from Partitioning Method*

We partition pairs of users based on four attributes, three on network similarity, NET-DIST, NET-SIM, NET-LHN, and one on rating similarity, RATE-SIM. Each division splits pairs into three groups, A, B, C, with C having the most similar pairs of users (in the case of NET-DIST, this means pairs with the lowest NET-DIST). Partitions are unbalanced, and shown in Table 6.

Does the mean RATE-SIM differ across NET-DIST partitions? In this experiment, we partition pairs of users according to their distance. In total there are 222,103 number of pairs considered; note that we exclude pairs with RATE-SIM zero. Partition C contains 1,866 pairs of friends; partition B contains 10,842 pairs of users with NET-DIST of 2 (friends of friends); partition A contains 209,395 pairs of users with NET-DIST greater than 2.

Partition C, which contains pairs of friends, has the highest mean RATE-SIM of 0.041. In partition B, the mean RATE-SIM drops to 0.022, while among all other pairs, in partition A, the mean RATE-SIM is 0.020. Therefore, we observe an increase in the mean RATE-SIM as the network distance decreases, a phenomenon which we investigate.

ANOVA shows that the mean RATE-SIM across NET-DIST partitions changes significantly (p-value in the order of 10^{-16}). Following this finding, we perform Dunnett's T3 test to check the significance of the pairwise differences between means; Table 7 presents the pairwise mean differences and their 95% confidence intervals. We observe that all differences are significant (the intervals do not contain the null hypothesis value of zero), with partition C having the largest mean RATE-SIM over the other partitions.

Does the mean RATE-SIM differ across NET-SIM partitions? We partition pairs of users according to their NET-SIM. We have 204,608 number of pairs in total; note that we exclude pairs where RATE-SIM or NET-SIM is zero. Group C contains 752 pairs of users with the high similarity values; group B has 5,836 pairs of users with the medium NET-SIM; and group A contains 198,020 pairs of users with the lowest NET-SIM. In each partition, we compute the mean RATE-SIM,

Table 7: Dunnett's T3 Test on Mean RATE-SIM among NET-DIST Partitions

Pair	Diff. of Means	95% CI
C - B	0.0184	[0.0174, 0.0193]
C - A	0.0203	[0.0193, 0.0214]
B - A	0.0020	[0.0010, 0.0029]

and observe that the means are different across the groups. ANOVA shows that the mean RATE-SIM across NET-SIM partitions changes significantly (p-value of 2×10^{-16}). Group C has the most similar pairs of users with a mean of 0.024; in group B the mean is 0.022, which is greater than the mean 0.018 of the group A. We thus observe moderate differences between the groups. Dunnett's T3 test shows that these differences are also significant, as shown in Table 8.

Table 8: Dunnett's T3 Test on Mean RATE-SIM among NET-SIM Partitions

Pair	Diff. of Means	95% CI
B - A	0.0035	[0.0024, 0.0046]
C - A	0.0061	[0.0051, 0.0072]
C - B	0.0026	[0.0015, 0.0037]

Does the mean RATE-SIM differ across NET-LHN partitions? We partition pairs of users according to their NET-LHN. We have 10,932 number of pairs in total; note that we exclude pairs where RATE-SIM or NET-LHN is zero. Group C contains 1,124 pairs of users; Group B has 1,030 pairs of users with the medium NET-LHN and the last Group A contains 8,778 pairs of users with the lowest NET-LHN values. The mean RATE-SIM is computed in each group, and we see that the means are roughly equal; A and C have mean RATE-SIM of 0.022, while B has mean RATE-SIM of 0.023. The ANOVA verifies that differences are not significant, and thus we perform no post hoc test.

For the next three experiments, we partition pairs of users according to their rating similarity (RATE-SIM).

Does the mean NET-DIST differ across RATE-SIM partitions? RATE-SIM partitions encompasses 222,103 number of pairs in total. Group C contains 12,768 pairs of users; group B has 203,051 pairs of users and group A contains 6,284 pairs of users with the lowest RATE-SIM values. The mean NET-DIST is computed in each group. Group A which contains most dissimilar pairs of users has the highest mean of 5.25, group B with neither similar nor dissimilar pairs has the mean of 4.93, and group C which contains the most similar users has the mean of 5.16. The results of ANOVA shows a significant difference of means across partitions with a p-value of 2×10^{-16}.

Dunnett's T3 test, depicted on Table 9, shows that all pairwise differences are significant. We observe some moderate variation in the magnitude of the differences. Specifically, groups A and C (of strong similarity or dissimilarity) have the highest mean NET-DIST of 5.25 and 5.16, compared to 4.94 of group B. This implies that pairs of users that are moderately similar (group B) tend to be somewhat closer in terms of network distance. The most important result however is negative. Looking at highly similar raters, we find no relationship about their network position: they can either be directly connected or very far from each other. This is in contrast to the opposite direction of correlation between RATE-SIM and NET-DIST (Table 7), where direct connection of users implies higher similarity in rankings.

Table 9: Dunnett's T3 Test on Mean NET-DIST among RATE-SIM Partitions

Pair	Diff. of Means	95% CI
B - A	-0.3098	[-0.3224,-0.2971]
C - A	-0.0876	[-0.1003,-0.07501]
C - B	0.2221	[0.2094,0.2348]

Does the mean NET-SIM differ across RATE-SIM partitions? We conduct the same experiment on RATE-SIM partitions by exploring the mean NET-SIM. There are 204,608 number of pairs examined in total, with group C containing 11,676 pairs, group B with 186,724 pairs, and group A with 6,208 pairs. Mean NET-SIM is roughly equal across groups: B has the highest mean of 0.032, followed by A with 0.031, and C with 0.030. Results of ANOVA show statistical significance (p-value of 2.64×10^{-13}), and post hoc analysis results are shown in Table 10. The differences are not always significant, and their strength is very small. This leads us to the conclusion that the three groups do not differ.

Table 10: Dunnett's T3 Test on Mean NET-SIM among RATE-SIM Partitions

Pair	Diff. of Means	95% CI
B - A	0.0004	$[-8.03 \times 10^{-6}, 0.0008]$
C - A	-0.0010	$[-1.47 \times 10^{-3}, -0.0006]$
C - B	-0.0014	$[-1.84 \times 10^{-3}, -0.0010]$

Does the mean NET-LHN differ across RATE-SIM partitions? The last experiment is based on 10,932 pairs of users, with group A having 604 pairs, group B having 9,376 pairs, and group C having 952 pairs. The mean NET-LHN in these partitions are roughly equal, with values 0.10, 0.12, 0.10, respectively. Although ANOVA sees significant differences (p-value of 1.73×10^{-15}), as well as Dunnett's T3 test (Table 11), the differences of means are generally small and not indicative of correlation.

Table 11: Dunnett's T3 Test on Mean NET-LHN among RATE-SIM Partitions

Pair	Diff. of Means	95% CI
B - A	0.0189	[0.0123,0.0254]
C - A	0.0064	[0.0001,0.0126]
C - B	-0.0125	[-0.0189,-0.0061]

5.2. *Results from Ranking Method*

We next present results by looking at the rankings created by attributes of similarity among pairs of users.

For the first set of experiments, we look at ratings (top-100 and top-1000) based on RATE-SIM, and look for correlations between RATE-SIM and one of the social similarity attributes.

Are RATE-SIM and NET-SIM correlated for top RATE-SIM pairs? We observe a moderate correlation for the top-100 RATE-SIM pairs (Figure 8a). The correlation, however weakens as we look at the top-1000 pairs (Figure 8b).

Are RATE-SIM and NET-LHN correlated for top RATE-SIM pairs? A weak correlation between RATE-SIM and NET-LHN is observed for top-100 most similar pairs by rating (RATE-SIM) on Figure 8c. Again the correlation significantly weakens as we increase the number of examined pairs to 1000 (Figure 8d).

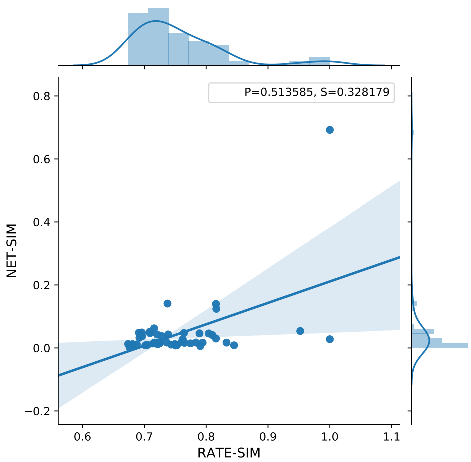
Are RATE-SIM and NET-DIST correlated for top RATE-SIM pairs? For the top-100 similar pairs, we observe a very weak positive correlation between network similarity in terms of NET-DIST (recall than NET-DIST is a measure of dis-similarity) and RATE-SIM. As we increase the number of pairs to 1000, the correlations weaken.

In the last set of experiments, we look at three rankings induced by an attribute measuring social similarity, namely by NET-SIM, NET-LHN, and NET-DIST.

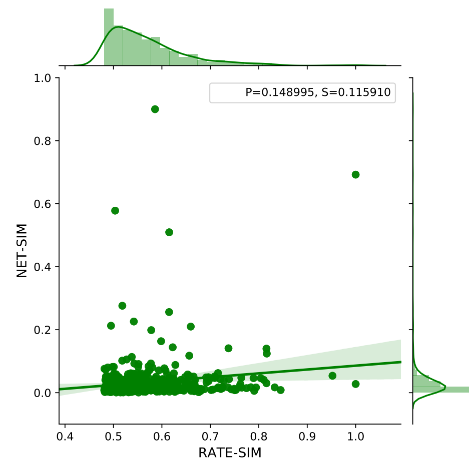
Are NET-SIM and RATE-SIM correlated for top NET-SIM pairs? All top-100 pairs have NET-SIM of 0.9, which means we cannot compute correlations between the tested attributes (Figure 9a). When we increase the number of examined pairs to 1000, we observe very weak correlations (Figure 9b).

Are NET-LHN and RATE-SIM correlated for top NET-LHN pairs? As before, looking at the top-100 pairs by NET-LHN, we cannot draw any conclusions (Figure 9c), as all pairs have the highest NET-LHN value of 1. Increasing the number of pairs to 1000, we begin to see weak negative correlations between NET-LHN and RATE-SIM (Figure 9d), implying that higher RATE-SIM is related to lower NET-LHN.

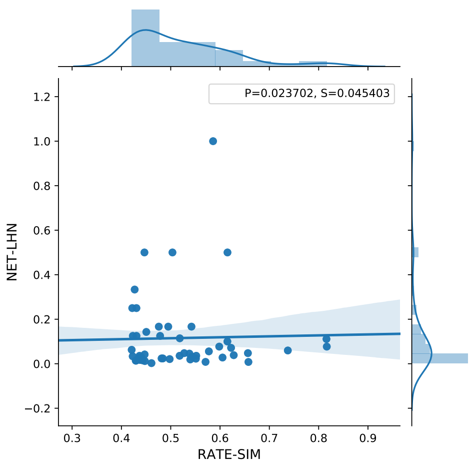
Are NET-DIST and RATE-SIM correlated for top NET-DIST pairs? We look at the top-100 and top-1000 pairs that have the lowest NET-DIST, respectively in Figures 9e and 9f. In both cases, this means pairs of friends with distance of 1.



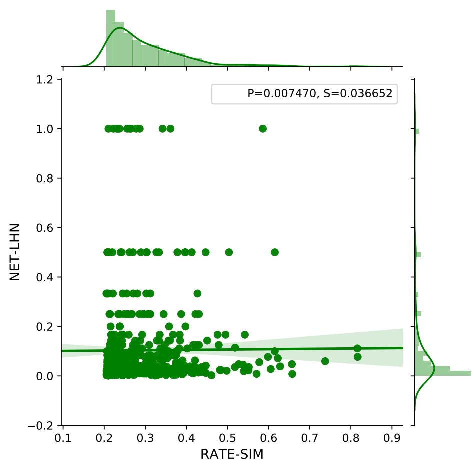
(a) Ranking on RATE-SIM top100



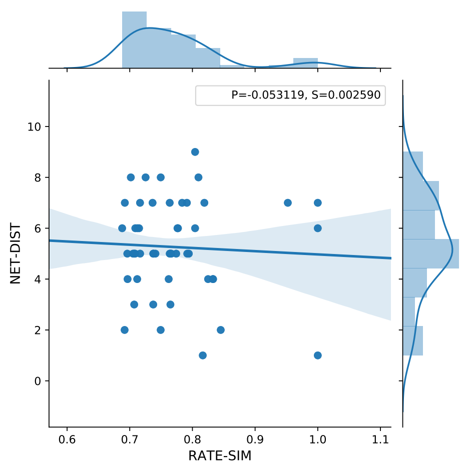
(b) Ranking on RATE-SIM top1000



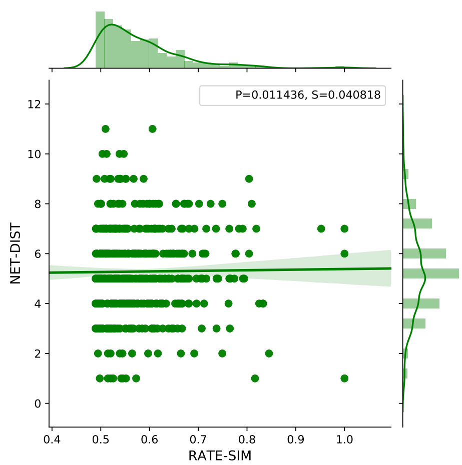
(c) Ranking on RATE-SIM top100



(d) Ranking on RATE-SIM top1000



(e) Ranking on RATE-SIM top100



(f) Ranking on RATE-SIM top1000

Fig. 8: V1: Ranking correlation results

As a result, we cannot draw conclusions on correlation between NET-DIST and RATE-SIM using the ranking method.

6. Conclusion

This work marks an important step towards studying the effects of social connections in rating behavior in social recommender systems. We look at two levels of behavior, that of users (level L1) and that of pairs of users (level L2), and investigate whether the studied objects observe any correlation according to different attributes measuring activity (in L1) and similarity (in L2).

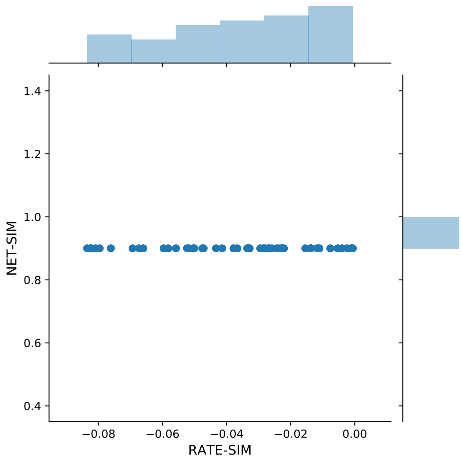
The most significant results are the following. At the first level of users, we see that rating activity (number of ratings made) and social activity (centrality in social network) are related, particularly when the latter is measured in terms of the number of social connections (degree centrality). The relationship is stronger in one direction: high rating activity implies that a user enjoys a more central role in the social network (has more connections).

At the second level, we see that rating similarity between pairs of users is related primarily with their network distance and with network SimRank, but not with the Leicht Holme Newman index. Specifically, we see that as rating similarity increases, so does the social similarity, and that when we focus on pairs of similar raters, we see moderate positive correlations of rating and SimRank similarity. On the other hand, we do not see trends and correlations on the opposite directions; i.e., social similarity does not seem to imply rating similarity.

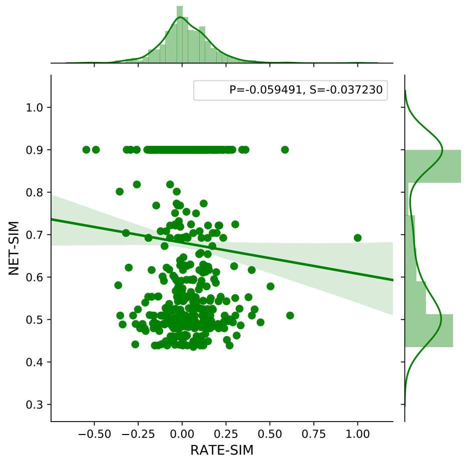
At both levels, we see that rating is related to social behavior, and that the direction of influence is stronger in one way: to some degree, rating behavior determines social behavior. In the future, we plan to focus on pairs of friends, and see if the levels of activity in a pair's members can determine whether the pair exhibits high or low similarity. For instance, we plan to investigate if two friends, among which, one is highly central (with many social connections) and the other is not, exhibits a higher than average similarity of their rating behavior, implying thus a possible social influence.

References

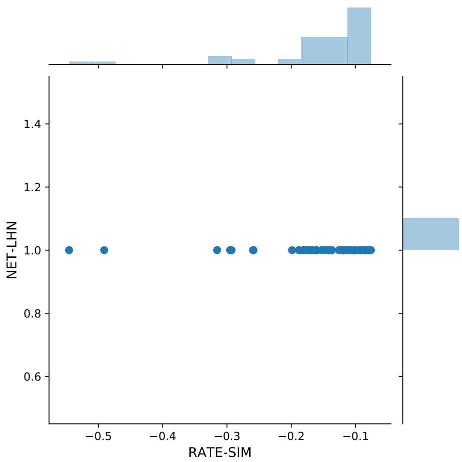
1. D. A. Easley and J. M. Kleinberg, *Networks, Crowds, and Markets - Reasoning About a Highly Connected World* (Cambridge University Press, 2010).
2. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, Grouplens: An open architecture for collaborative filtering of netnews, in *CSCW '94, Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, NC, USA, October 22-26, 1994*, eds. J. B. Smith, F. D. Smith and T. W. Malone (ACM, 1994), pp. 175–186.
3. P. Massa and P. Avesani, Trust-aware recommender systems, in *RecSys*2007, pp. 17–24.
4. H. Ma, H. Yang, M. R. Lyu and I. King, Sorec: social recommendation using probabilistic matrix factorization, in *CIKM*2008, pp. 931–940.



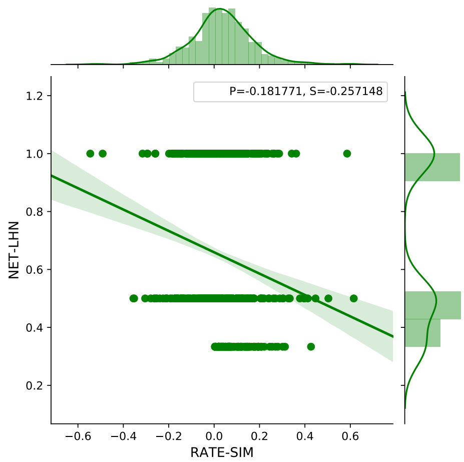
(a) Ranking on NET-SIM top100



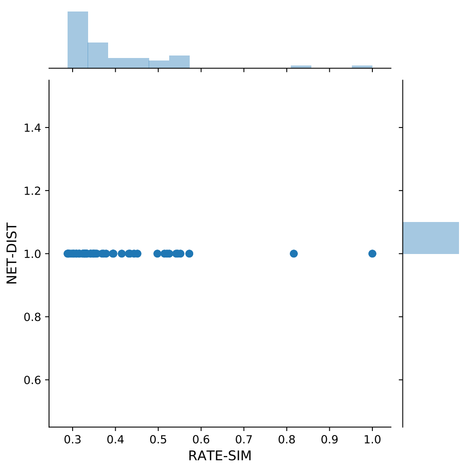
(b) Ranking on NET-SIM top1000



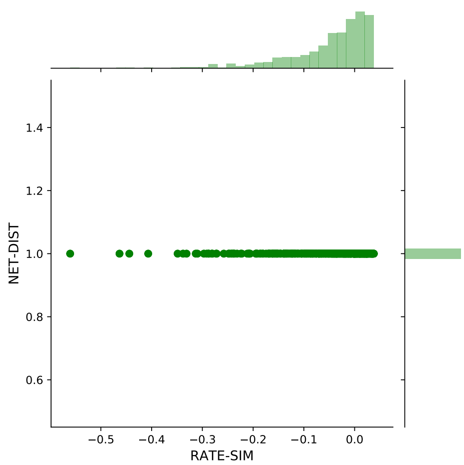
(c) Ranking on NET-LHN top100



(d) Ranking on NET-LHN top1000



(e) Ranking on NET-DIST top100



(f) Ranking on NET-DIST top1000

Fig. 9: V2 : Ranking correlation results

5. H. Ma, I. King and M. R. Lyu, Learning to recommend with social trust ensemble, in *SIGIR*2009, pp. 203–210.
6. H. Bisgin, N. Agarwal and X. Xu, Investigating homophily in online social networks, in *2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010, Toronto, Canada, August 31 - September 3, 2010, Main Conference Proceedings*2010, pp. 533–536.
7. H. Ma, D. Zhou, C. Liu, M. R. Lyu and I. King, Recommender systems with social regularization, in *WSDM*2011, pp. 287–296.
8. L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines and F. Menczer, Friendship prediction and homophily in social media, *TWEB* **6**(2) (2012) 9:1–9:33.
9. X. Liu and K. Aberer, Soco: a social network aided context-aware recommender system, in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*2013, pp. 781–802.
10. H. Li, D. Wu, W. Tang and N. Mamoulis, Overlapping community regularization for rating prediction in social recommender systems, in *RecSys*2015, pp. 27–34.
11. I. Guy, Social recommender systems, in *Recommender Systems Handbook*, eds. F. Ricci, L. Rokach and B. Shapira (Springer, 2015) pp. 511–543.
12. S. Alotaibi and J. Vassileva, Personalized recommendation of research papers by fusing recommendations from explicit and implicit social network, in *UMAP*2016.
13. A. Said, E. W. D. Luca and S. Albayrak, How social relationships affect user similarities, in *SRS Workshop*2010.
14. C. P. Mukamakuza, Analyzing the impact of social connections on rating behavior in social recommender systems, in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization UMAP*, (ACM, New York, NY, USA, 2017), pp. 322–326.
15. C. P. Mukamakuza, D. Sacharidis and H. Werthner, Mining user behavior in social recommender systems, in *WIMS* (ACM, 2018), pp. 37:1–37:6.
16. C. P. Mukamakuza, D. Sacharidis and H. Werthner, The impact of social connections in personalization, in *UMAP Adjunct* (ACM, 2019).
17. M. D. Ekstrand, M. Ludwig, J. A. Konstan and J. Riedl, Rethinking the recommender research ecosystem: reproducibility, openness, and lenskit, in *RecSys* (ACM, 2011), pp. 133–140.
18. G. Jeh and J. Widom, Simrank: A measure of structural-context similarity, in *SIGKDD* (ACM, New York, NY, USA, 2002), pp. 538–543.
19. W.-F. Guo and S.-W. Zhang, A general method of community detection by identifying community centers with affinity propagation, *Physica A: Statistical Mechanics and its Applications* **447** (2016) 508 – 519.
20. E. A. Leicht, P. Holme and M. E. J. Newman, Vertex similarity in networks, *Phys. Rev. E* **73** (Feb 2006) p. 026120.
21. R. Jin, J. Y. Chai and L. Si, An automatic weighting scheme for collaborative filtering, in *SIGIR*2004, pp. 337–344.
22. R. Salakhutdinov and A. Mnih, Probabilistic matrix factorization, in *NIPS*2007, pp. 1257–1264.
23. Y. Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model, in *SIGKDD*2008, pp. 426–434.
24. Y. Koren, R. M. Bell and C. Volinsky, Matrix factorization techniques for recommender systems, *IEEE Computer* **42**(8) (2009) 30–37.
25. M. Jamali and M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in *RecSys*2010, pp. 135–142.
26. H. Zhao, Q. Yao, J. T. Kwok and D. L. Lee, Collaborative filtering with social local

models
27. P. Boldi and S. Vigna, Axioms for centrality, *Internet Mathematics* **10**(3-4) (2014) 222–262.
28. L. Page, S. Brin, R. Motwani and T. Winograd, The pagerank citation ranking: Bringing order to the web (1999).
29. B. M. Sarwar, G. Karypis, J. A. Konstan and J. Riedl, Item-based collaborative filtering recommendation algorithms, in *WWW* 2001, pp. 285–295.
30. P. Lenhart and D. Herzog, Combining content-based and collaborative filtering for personalized sports news recommendations, in *RecSys CBR*2016, pp. 3–10.
31. A. Pradhan and G. K. Mahinthakumar, Finding all-pairs shortest path for a large-scale transportation network using parallel floyd-warshall and parallel dijkstra algorithms, *Journal of Computing in Civil Engineering* **27**(3) (2013) 263–273.
32. I. Antonellis, H. Garcia-Molina and C. chao Chang, Simrank++: Query rewriting through link analysis of the click graph, Technical Report 2007-32, Stanford InfoLab (2007).
33. D. Schall, Social network-based recommender systems, in *Springer International Publishing*2015.
34. J. W. Tukey, Comparing individual means in the analysis of variance, *Biometrics* (1949) 99–114.
35. C. W. Dunnett, Pairwise multiple comparisons in the unequal variance case, *Journal of the American Statistical Association* **75**(372) (1980) 796–800.
36. G. Guo, J. Zhang and N. Yorke-Smith, A novel bayesian similarity measure for recommender systems, in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence IJCAI*2013, pp. 2619–2625.
37. G. Guo, J. Zhang, D. Thalmann and N. Yorke-Smith, ETAF: an extended trust antecedents framework for trust prediction, in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014* 2014, pp. 540–547.