The Impact of Social Connections in Personalization

Carine Pierrette Mukamakuza
E-Commerce Research Division
TU Wien
Austria
carine@ec.tuwien.ac.at

Dimitris Sacharidis
E-Commerce Research Division
TU Wien
Austria
dimitris@ec.tuwien.ac.at

Hannes Werthner
E-Commerce Research Division
TU Wien
Austria
hannes@ec.tuwien.ac.at

ABSTRACT

Personalization is typically based on preferences extracted from the interactions of users with the system. A recent trend is to also account for the social influence among users, which may play a non-negligible role in shaping one's individual preferences. The underlying assumptions are that friends tend to develop similar taste, i.e., homophily, and that similar users tend to connect to each other, i.e., social selection. In this work, we investigate the conditions under which social influence has a significant impact on the preferences of users. We find that pairs of friends, where one is socially very active whereas the other is not, exhibit stronger correlations in their preferences compared to other pairs of friends, implying thus a stronger mechanism of influence.

CCS CONCEPTS

• **Information systems → Social recommendation**; **Personalization**; **Recommender systems**.

KEYWORDS

Personalization; Recommender Systems; Social influence; Social network analysis

ACM Reference Format:
Carine Pierrette Mukamakuza, Dimitris Sacharidis, and Hannes Werthner. 2019. The Impact of Social Connections in Personalization. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3314183.3323675

1 INTRODUCTION

Personalization based on collaborative filtering typically exploits similarity patterns from historical records of interaction between users and items [9]. A recent trend is to also consider the social aspect, and specifically the tendency of individuals to associate and bond with similar others, a phenomenon called *social selection*, and the tendency of socially connected individuals to exhibit similar preferences, a phenomenon called *homophily* [20]. In such *social-based collaborative filtering* [1, 4, 14–19], a profile for a target user is computed not only based on historical user-item interactions, but also based on the target user's social connections. This is also motivated by the way people often make decisions in real life — choices are often governed by interpersonal influence from social connections, besides individual preferences.

In such systems, there exist two data sources that govern personalization: the historical rating (or feedback) activity of users, and the social connections. In our line of work [21, 22], we seek to quantify the extent to which one source affects the other. The goal is twofold: on the one hand, we seek to validate the assumptions often implicitly made in the literature, and on the other hand, we aim to understand the connections between social and feedback activity so as to design more effective personalization strategies.

In this work, we consider pairs of friends and apply the following methodology. Each pair can be described by *edge attributes* that quantify the similarity between the two connected users. These attributes can be computed based on either the historical rating activity, such as the degree of similarity between the ratings given to items, or on the social connections, such as the number of common friends between the two connected users. Our objective is to associate attributes from one data source (feedback or social activity) to the other, and understand what are the causes for the observed correlations.

To this end, we compute *node attributes* that quantify the level of activity a user exhibits, either in terms of her feedback provided or in terms of her social connections. For example, a user is highly active in terms of feedback, if she has rated many items in the past, while a user is highly active socially, i.e., is popular, if she has a central position in the network [22].

To explain possible correlations in edge attributes and answer questions such as when do two friends influence each other more, we classify friends into three groups, based on the amount of activity (rating or social) the two connected users exhibit, i.e., their node attributes. We consider pairs of friends that are: (LL) both of low activity, (HH) both of high activity, or (LH) one has high and the other low activity. We then investigate whether the rating/social similarities, i.e., the node attributes, differ significantly among the three groups.

The most important finding of our work are that pairs of type LH in terms of social connection exhibit stronger correlations in their rating behavior. This means that there is a stronger force of influence between them. Although the direction of the force cannot be identified using the data available, we conjecture that popular users are the ones that exert influence on unpopular ones.

The remainder of this paper is structured as follows. Section 2 establishes the necessary background and overviews existing work and Section 3 describes our approach. Section 4 presents experimental results of our research question while Section 5 draws the conclusions.

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6711-0/19/06…$15.00
https://doi.org/10.1145/3314183.3323675

2 RELATED WORK

In personalization approaches based on *Collaborative Filtering* (CF), users and items with similar feedback patterns are taken into account to compute a user profile for the target user [9]. The basic entity in CF is the *user-item ratings matrix* $R \in \mathbb{R}^{n \times m}$ that contains the ratings given by m users to n items. The most popular CF method is the *Matrix Factorization* (MF) technique [10, 11, 24], which, in its simplest incarnation, computes a low-rank approximation of the sparse rating matrix R.

Social-aware personalization differ from CF 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 relationship strength between the corresponding users. Social recommenders combine information contained in matrices R and S. In the following, we review the most important related work; for a more complete overview refer to [23].

In trust-aware recommender systems [19], the idea is to treat the social neighborhood of the target user in a manner similar to the rating neighborhood in user-based CF. An experimental evaluation of several memory-based social recommenders is provided in [2]. 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. SoRec [17] 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 rating matrix.

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 [4] studies homophily on two online social media networks, BlogCatalog, and Last.fm by extracting communities based on the network ties. Similarly, [1] investigates the presence of homophily in three systems that combine tagging social media with online social networks. The most recent works [14, 18] apply MF combined with regularization techniques that aim to capture the homophily in the social network.

3 RESEARCH APPROACH

In our work, we wish to investigate when social connections play a role in shaping the preferences of users. We assume we have a dataset consisting of (1) a history of user-item feedbacks (ratings), similar to that typically used in collaborative filtering, and (2) a set of social connections between these users. Our approach is based on viewing such a dataset as a labeled social network, which has the same structure with that implied by the social connections between users, but additionally has attributes for the nodes (users) and the edges (pairs of friends).

The research question we address in this work is the following.

RQ Do user attributes affect the strength of user-user similarities?

In other words, if we know individual aspects about users, e.g., their level of activity in a personalization system, can we infer a pairwise relationship, e.g., the similarity of their observed activities, between friends?

The rest of the section is organized as follows. Section 3.1 presents the augmented social network, then Section 3.2 explains our methodology, while Section 3.3 describes the dataset used.

3.1 Augmented Social Network

We conceptually consider a social network where nodes and edges have additional attributes as defined in the following.

Node Attributes Capturing Activity of Users. We consider one notion of activity in terms of rating behavior, and a notion in terms of social connections, based on the concept of node centrality [5].

RATE-NUM For the rating activity, we consider the number of ratings a user has provided. This essentially, captures how active a user is in the system.

NET-DEG Degree centrality is the most intuitive interpretation of social activity, counting the number of (incoming or outgoing) social connections a user has.

Edge Attributes Capturing Similarity Between Two Friends. We consider two notions of similarity in terms of rating behavior, and two 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 [25]. This simple definition, however, has some limitations. It is known that people 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 [13]. 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.

RATE-PCC Pairwise similarity (RATE-PCC) is the rating similarity when only the common rated items between two users are considered:

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

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. The $\sum_{i \in I_u \cap I_v}$ is the sum of the items that both users have rated in common.

NET-SIM The idea behind *SimRank* is simple: two users are similar if they are referenced by similar users [3, 8]. 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-SIM can be considered as a global pairwise similarity measure.

NET-LHN The *Leicht Holme Newman* index [7, 12] 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 [26]. NET-LHN, in contrast to NET-SIM, can be considered as a local pairwise similarity measure.

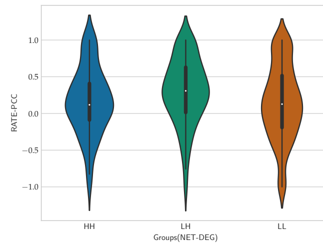
3.2 Methodology

Previous work that exploits social influence between users [2, 22, 23] has demonstrated that there exist correlations between the similarities in terms of the social network and the observed feedback. In terms of our augmented social network, this translates into correlations of the various edge attributes. In this work, we seek to understand when these correlations are stronger. Specifically, we want to see if node attributes can help identify these instances.

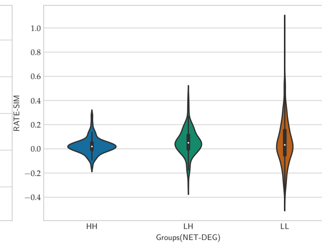
Therefore, we define classes of pairs of friends, based on their node attributes, and then measure whether similarities among edge attributes become stronger or weaker across classes. More concretely, a user is assigned a label L when her activity (node attribute RATE-NUM or NET-DEG) is below some threshold L, label H when her activity is above another threshold H, and no label otherwise; we consider various values for these thresholds. In this way, two friends are classified into four classes:

- **LL** when both have label L,
- **HH** when both have label H,
- **LH** when one has label L and the other label H,
- – when one has no label.

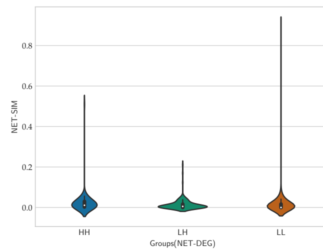
This essentially induces a partition on the edges of the augmented social network. We examine the three classes LL, HH, and LH, to see if for some class we measure stronger/weaker edge-based similarities. As a first step, we plot the distribution of an edge attribute (RATE-SIM, RATE-PCC, NET-SIM, NET-LHN) within the class, and visually explore if any differences across classes appear. Then, we focus on the mean edge attribute for a class, and perform statistical tests (ANOVA followed by pairwise post hoc analysis) to see whether the visual differences across classes are actually significant.



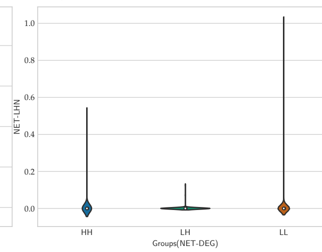
(a) RATE-PCC vs. NET-DEG



(b) RATE-SIM vs. NET-DEG



(c) NET-SIM vs. NET-DEG



(d) NET-LHN vs. NET-DEG

Figure 1: Classes based on NET-DEG

3.3 Data

In our study, we use a publicly available dataset, FilmTrust [6], collected from traces of user interaction in social-based collaborative filtering. The data contain feedback history, i.e., a rating matrix R, as well as information about the social connections among users, i.e., an adjacency matrix S. In total, there are 740 users with 1576 social connections. Across all users, mean NET-DEG is 18, and mean RATE-NUM is 43.5. Across all pairs of friends, mean RATE-PCC is 0.181, mean RATE-SIM is 0.049, mean NET-SIM is 0.0186, while mean NET-LHN is 0.0056.

4 EXPERIMENTAL EVALUATION

Section 4.1 presents the results of our evaluation, while Section 4.2 summarizes the findings.

4.1 Results

Does RATE-PCC depend on NET-DEG? We first consider partitioning pairs of friends based on the NET-DEG. We explore different definition of low (L) and high (H) NET-DEG, based on which we assign pairs of friends into classes LL, LH, and HH. For each class, we compute the mean RATE-PCC. The results are shown in Table 1, where we see that RATE-PCC varies significantly across different classes.

We then fix L and H to their default values of L=10 and H=20, and look deeper into the three classes they induce. Specifically, LL contains pairs of friends where each has less than 10 friends in total; HH contains pairs of friends where each has more than 20 friends in total; LH contains pairs of friends, where one has few ($\leq$ 10) and the other has many ($\geq$ 20) other friends. There are 873 number of pairs examined in total; HH contains 142 pairs, LH has 157 pairs, and LL 574 pairs. The mean RATE-PCC within the classes is 0.162, 0.293 and 0.137 respectively.

Table 1: Mean RATE-PCC of NET-DEG classes

		H	5	10	15	20	30	40	50
		HH	0.18	0.162	0.166	0.162	0.121	0.27	-0.17
L	LL	LH							
5	0.152		0.188	0.212	0.28	0.275	0.27	0.27	0.193
10	0.153		0.201	0.235	0.28	0.29	0.27	0.26	0.201
15	0.132		0.185	0.204	0.258	0.265	0.258	0.27	0.164
20	0.14		0.19	0.201	0.251	0.257	0.25	0.248	0.157
30	0.16		0.192	0.201	0.248	0.255	0.246	0.246	0.154
40	0.17		0.19	0.195	0.23	0.233	0.22	0.233	0.15
50	0.18		0.192	0.2	0.227	0.23	0.214	0.232	0.151

Figure 1a shows the distribution of RATE-PCC between pairs of friends in each of the three classes. While not immediately apparent, the distributions have different means and shape. To quantify this, we perform ANOVA analysis, which shows that the mean RATE-PCC across the classes is significantly different (p-value of 0.00235). Then, post hoc analysis of the results, presented in Table 2, finds that the RATE-PCC similarity of LH pairs of friends is considerably and significantly higher than other pairs of friends. This implies that a pair of friends that is formed by a popular H user and a less popular L user tend to influence each other's rating behavior.

Table 2: RATE-PCC differences across NET-DEG classes

Pair	Diff. of Means	95% CI
LL - LH	-0.1551	$[-0.254, -0.0562]$
LL - HH	-0.0253	$[-0.0129, 0.0786]$
LH - HH	0.1297	[0.0056,0.2539]

Does RATE-SIM depend on NET-DEG? We repeat the previous setup, this time looking at the RATE-SIM between two friends. Table 3 shows the mean RATE-SIM for various definitions of L and H in terms of NET-DEG. Differences exists but are not as dramatic as in the case of RATE-PCC.

Table 3: Mean RATE-SIM of NET-DEG classes

		H	5	10	15	20	30	40	50
		HH	0.05	0.037	0.03	0.025	0.025	0.019	-0.013
L	LL	LH							
5	0.06		0.05	0.042	0.06	0.06	0.06	0.07	0.039
10	0.05		0.05	0.05	0.05	0.05	0.05	0.05	0.027
15	0.05		0.05	0.05	0.06	0.06	0.06	0.06	0.026
20	0.05		0.05	0.05	0.06	0.06	0.05	0.05	0.026
30	0.05		0.05	0.044	0.05	0.05	0.05	0.05	0.023
40	0.05		0.05	0.044	0.05	0.05	0.05	0.05	0.023
50	0.05		0.05	0.044	0.05	0.05	0.04	0.05	0.023

Fixing the definition of L and H to their default values, in Figure 1b, we plot the distribution of RATE-SIM within the three classes. Class HH has a mean RATE-SIM of 0.025, LH of 0.05, and LL of 0.05. That is, mean RATE-SIM is roughly equal for LH and LL categories and higher than HH which has the lowest mean. However, ANOVA results show that the differences are not significant (p-value of 0.148). We conclude that no safe conclusions can be drawn from this experiment.

Does NET-SIM depend on NET-DEG? In this experiment we measure friend similarity in terms of their global network similarity quantified as NET-SIM. Table 4 presents the mean NET-SIM for the various classes previously explored, where we do not observe any meaningful trends.

Table 4: Mean NET-SIM of NET-DEG classes

		H	5	10	15	20	30	40	50
		HH	0.02	0.02	0.02	0.02	0.016	0.013	0.002
L	LL	LH							
5	0.03		0.013	0.014	0.014	0.014	0.01	0.01	0.01
10	0.02		0.015	0.016	0.014	0.014	0.012	0.01	0.01
15	0.02		0.016	0.018	0.017	0.017	0.016	0.014	0.01
20	0.02		0.016	0.019	0.018	0.017	0.017	0.014	0.01
30	0.02		0.018	0.02	0.02	0.02	0.018	0.014	0.01
40	0.02		0.018	0.02	0.02	0.02	0.017	0.015	0.01
50	0.02		0.018	0.02	0.02	0.02	0.017	0.014	0.01

We next fix L and H to their default values, and plot the distribution of NET-SIM within the three induced classes in Figure 1c. Classes LL and HH have a mean of 0.02, while LH has a mean of 0.014, i.e., they are roughly equal. ANOVA finds they do not significantly differ (p-value of 0.466). Any differences in terms of NET-SIM across NET-DEG classes are not significant.

Does NET-LHN depend on NET-DEG? In the last experiment with classes defined on NET-DEG, we measure pairwise similarities in terms of the local network similarity NET-LHN. Table 5 presents the mean NET-LHN for the studied classes.

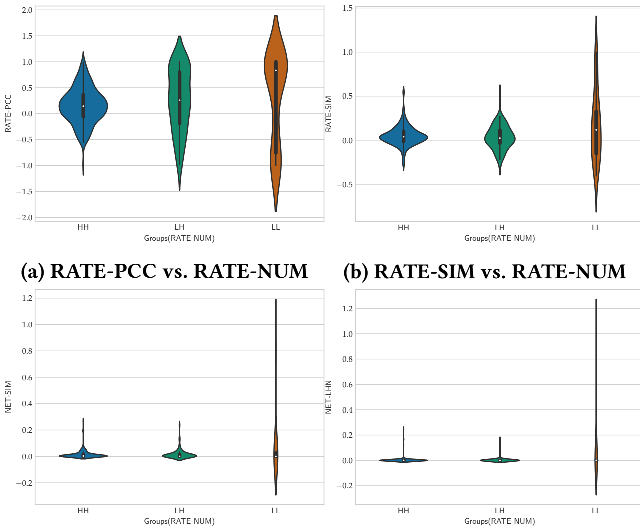
Table 5: Mean NET-LHN of NET-DEG classes

		H	5	10	15	20	30	40	50
		HH	0.005	0.007	0.008	0.007	0	0	0
L	LL	LH							
5	0.011		0.001	0.001	0.0007	0.0008	0	0	0
10	0.006		0.004	0.005	0.001	0.001	0.001	0	0
15	0.007		0.004	0.005	0.003	0.003	0.003	0.002	0
20	0.007		0.004	0.006	0.004	0.003	0.004	0.002	0
30	0.007		0.005	0.007	0.006	0.006	0.004	0.001	0
40	0.007		0.005	0.006	0.006	0.006	0.003	0.001	0
50	0.007		0.005	0.006	0.006	0.005	0.003	0.001	0

We fix L and H to their default values for NET-DEG, and drawn the distribution of NET-LHN across the three classes in Figure 1d. ANOVA reports no significant differences for the mean NET-DEG values.

Does RATE-PCC depend on RATE-NUM? In the following set of experiments, we classify pair of friends based on their number of provided ratings, RATE-NUM. First, we consider pairwise similarity in terms of RATE-PCC. Table 6 includes the mean RATE-NUM for different definitions of L and H in terms of RATE-NUM. Except when L=5, we note that the mean RATE-PCC is roughly the same across classes.

We fix L and H to their default values L=10 and H=30, and examine the three classes they define. We have 576 pairs in total, with class HH containing 444 pairs, class LH has 94 pairs, and class LL has 38 pairs. The mean value of RATE-PCC for each class is 0.156,



(a) RATE-PCC vs. RATE-NUM (b) RATE-SIM vs. RATE-NUM

(c) NET-SIM vs. RATE-NUM (d) NET-LHN vs. RATE-NUM

Figure 2: Classes based on RATE-NUM

Table 6: Mean RATE-PCC of RATE-NUM classes

		H	5	10	20	30	50	70	100
		HH	0.157	0.162	0.152	0.156	0.191	0.134	0.124
L	LL	LH							
5	0.93		0.68	0.66	0.743	0.705	0.613	0.573	0.49
10	0.322		0.233	0.236	0.273	0.257	0.302	0.355	0.254
20	0.19		0.201	0.208	0.217	0.204	0.206	0.24	0.192
30	0.2		0.19	0.19	0.191	0.192	0.21	0.216	0.182
50	0.19		0.172	0.176	0.17	0.174	0.18	0.191	0.164
70	0.176		0.167	0.171	0.17	0.17	0.182	0.198	0.164
100	0.185		0.17	0.172	0.172	0.174	0.185	0.193	0.164

0.257, 0.322, respectively, and Figure 2a draws the distribution of RATE-PCC within the classes. ANOVA shows that the three classes do not differ significantly in terms of their mean RATE-PCC (p-value of 0.068). The conclusion is that classes based on RATE-NUM do not differ substantially in terms of their RATE-PCC.

Does RATE-SIM depend on RATE-NUM? We next consider whether there are differences across RATE-NUM classes in terms of the RATE-SIM, instead of RATE-PCCs. Table 7 presents the mean RATE-SIM for different definition of classes.

Table 7: Mean RATE-SIM of RATE-NUM classes

		H	5	10	20	30	50	70	100
		HH	0.045	0.045	0.047	0.048	0.042	0.033	0.0077
L	LL	LH							
5	0.91		0.075	0.065	0.093	0.097	0.077	0.057	0.032
10	0.167		0.049	0.057	0.037	0.033	0.04	0.041	0.01
20	0.08		0.05	0.048	0.042	0.039	0.031	0.034	0.028
30	0.07		0.048	0.046	0.042	0.041	0.035	0.035	0.029
50	0.06		0.048	0.047	0.045	0.045	0.035	0.036	0.03
70	0.054		0.048	0.048	0.047	0.047	0.038	0.038	0.031
100	0.052		0.047	0.047	0.047	0.047	0.039	0.036	0.032

We fix L and H to their default values, and plot the distribution of RATE-SIM for each class in Figure 2b. In addition, we perform ANOVA and find that the means of classes differ significantly (p-value of $< 10^{-5}$). However, post hoc analysis, shown in Table 8, finds that the magnitude of the differences is not significant. Hence, we cannot draw any safe conclusions in this experiment.

Table 8: RATE-SIM differences across RATE-NUM classes

Pair	Diff. of Means	95% CI
LL - LH	0.1333	-0.0560,0.3226]
LL - HH	0.1195	[-0.0595,0.2984]
LH - HH	-0.0138	[-0.0515,0.0239]

Does NET-SIM depend on RATE-NUM? Next, we consider global network pairwise similarity between friends. Table 9 shows mean NET-SIM for the different definitions of RATE-NUM-based classes.

Table 9: Mean NET-SIM of RATE-NUM classes

		H	5	10	20	30	50	70	100
		HH	0.017	0.016	0.018	0.015	0.01	0.011	0.003
L	LL	LH							
5	0.35		0.02	0.02	0.021	0.022	0.023	0.02	0.0009
10	0.12		0.03	0.016	0.016	0.016	0.015	0.011	0.001
20	0.03		0.02	0.015	0.016	0.016	0.016	0.017	0.018
30	0.02		0.02	0.016	0.018	0.016	0.016	0.015	0.014
50	0.02		0.02	0.017	0.018	0.016	0.016	0.017	0.012
70	0.02		0.02	0.017	0.018	0.016	0.016	0.016	0.011
100	0.02		0.02	0.017	0.018	0.017	0.016	0.016	0.011

Again, we fix L and H to their defaults, and plot NET-SIM distributions for the three induced classes in Figure 1c. As before, while ANOVA shows that the means are not equal with high significance (p-value of $< 10^{-13}$), post-hoc analysis, presented in Table 10, shows non-significant differences.

Table 10: NET-SIM differences across RATE-NUM classes

Pair	Diff. of Means	95% CI
LL - LH	0.1034	-0.0302,0.2369]
LL - HH	0.1047	[-0.0234,0.2328]
LH - HH	0.0013	[-0.0088,0.0114]

Does NET-LHN depend on RATE-NUM? The last experiment studies local network pairwise similarity between friends. Table 11 shows mean NET-LHN for the different definitions of RATE-NUM-based classes.

For fixed L and H, Figure 2d plots the distribution of NET-LHN in the three classes. ANOVA finds that they all have roughly equal means, and thus we conclude that no dependence on RATE-NUM is exhibited.

4.2 Discussion

In our evaluation, we have divided pairs of friends in multiple ways into LL, HH, and LH classes, and examine if pairwise similarities across classes show significant differences. The main conclusions

Table 11: Mean NET-LHN of RATE-NUM classes

		H	5	10	20	30	50	70	100
		HH	0.005	0.005	0.006	0.003	0.004	0.0008	0
L	LL	LH							
5	0.125		0.004	0.004	0.005	0.005	0.008	0.005	0
10	0.09		0.015	0.003	0.003	0.004	0.003	0.003	0
20	0.015		0.007	0.002	0.003	0.004	0.006	0.007	0.013
30	0.009		0.007	0.004	0.005	0.003	0.004	0.005	0.008
50	0.008		0.006	0.005	0.006	0.004	0.004	0.005	0.005
70	0.006		0.007	0.004	0.005	0.004	0.004	0.005	0.004
100	0.006		0.006	0.004	0.005	0.004	0.004	0.005	0.004

drawn are the following. In all definitions of classes, we observe some differences in how pairwise similarities are distributed. However, not all of them are significant. When classes are defined according to network activity NET-DEG, only similarities measured by feedback similarity RATE-PCC are found to be significant and have a large impact. Specifically, pairs of friends that belong to class LH tend to have higher RATE-PCC than pairs in the other classes. When classes are defined according to feedback activity RATE-NUM, pairwise similarities in terms of RATE-SIM and NET-SIM are found to be significant; however their impact does not appear to be considerable.

In conclusion, we see that if a user with low social activity is connected with a user with high social activity, we expect their feedback similarity, in terms of RATE-PCC, to be almost two times as high as other pair of friends. Although we cannot be certain of the direction of influence, we conjecture that it flows from the more socially active user to the less active one.

The results obtained here could be exploited to provide more effective personalization. Specifically, we have found that to some extent network-based similarity can substitute feedback-based similarity, and thus be used as a proxy for determining the similarity between friends in terms of their preferences. Moreover, the similarity strength increases when one friend is much less active than the other. These findings could be applied in a collaborative filtering approach, where tastes of similar minded users are aggregated. One idea would be to consider in this aggregation the strength of influence between two friends, computed based on their network similarity and their level of feedback activity.

5 CONCLUSION

This paper provides some in-depth insights into the impact of social connections in the preferences expressed by users. In order to measure the influence among pairs of friends, we label users with L (low) and H (high) based on their feedback activity and their social activity. We then divide pairs of users into classes HH (high-high), LL (low-low), LH (low-high), and investigate whether various pairwise similarity measures (based on either their feedback or their social activity) tend to become stronger. The main outcome of our work, is that a pair of friends that belong to class LH in terms of social activity, tend to be more similar in their feedback activity, compared to other pairs.

REFERENCES

[1] Luca Maria Aiello, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. 2012. Friendship prediction and homophily in social media. *TWEB* 6, 2 (2012), 9:1–9:33. https://doi.org/10.1145/2180861.2180866

[2] Shaikhah Alotaibi and Julita Vassileva. 2016. Personalized Recommendation of Research Papers by Fusing Recommendations from Explicit and Implicit Social Network. In *Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016).*

[3] Ioannis Antonellis, Hector Garcia-Molina, and Chi chao Chang. 2007. *Simrank++: Query rewriting through link analysis of the click graph.* Technical Report 2007-32. Stanford InfoLab. http://ilpubs.stanford.edu:8090/868/

[4] Halil Bisgin, Nitin Agarwal, and Xiaowei Xu. 2010. 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.* 533–536. https://doi.org/10.1109/WI-IAT.2010.61

[5] Paolo Boldi and Sebastiano Vigna. 2014. Axioms for Centrality. *Internet Mathematics* 10, 3-4 (2014), 222–262. https://doi.org/10.1080/15427951.2013.865686

[6] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. 2013. A Novel Bayesian Similarity Measure for Recommender Systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence IJCAI.* 2619–2625.

[7] Wei-Feng Guo and Shao-Wu Zhang. 2016. A general method of community detection by identifying community centers with affinity propagation. *Physica A: Statistical Mechanics and its Applications* 447 (2016), 508 – 519. https://doi.org/10.1016/j.physa.2015.12.037

[8] Glen Jeh and Jennifer Widom. 2002. SimRank: A Measure of Structural-context Similarity. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02).* ACM, New York, NY, USA, 538–543. https://doi.org/10.1145/775047.775126

[9] Rong Jin, Joyce Y. Chai, and Luo Si. 2004. An automatic weighting scheme for collaborative filtering. In *SIGIR.* 337–344. https://doi.org/10.1145/1008992.1009051

[10] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD.* ACM, 426–434.

[11] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* 42, 8 (2009), 30–37.

[12] E. A. Leicht, Petter Holme, and M. E. J. Newman. 2006. Vertex similarity in networks. *Phys. Rev. E* 73 (Feb 2006), 026120. Issue 2. https://doi.org/10.1103/PhysRevE.73.026120

[13] Philip Lenhart and Daniel Herzog. 2016. Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations. In *Proceedings of the 3rd Workshop on New Trends in Content-Based Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2016), Boston, MA, USA, September 16, 2016.* 3–10. http://ceur-ws.org/Vol-1673/paper1.pdf

[14] Hui Li, Dingming Wu, Wenbin Tang, and Nikos Mamoulis. 2015. Overlapping Community Regularization for Rating Prediction in Social Recommender Systems. In *RecSys.* 27–34. https://doi.org/10.1145/2792838.2800171

[15] Xin Liu and Karl Aberer. 2013. 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.* 781–802. http://dl.acm.org/citation.cfm?id=2488457

[16] Hao Ma, Irwin King, and Michael R. Lyu. 2009. Learning to recommend with social trust ensemble. In *SIGIR.* 203–210. https://doi.org/10.1145/1571941.1571978

[17] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. 2008. SoRec: social recommendation using probabilistic matrix factorization. In *CIKM.* 931–940. https://doi.org/10.1145/1458082.1458205

[18] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *WSDM.* 287–296. https://doi.org/10.1145/1935826.1935877

[19] Paolo Massa and Paolo Avesani. 2007. Trust-aware recommender systems. In *RecSys.* 17–24. https://doi.org/10.1145/1297231.1297235

[20] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.

[21] Carine Pierrette Mukamakuza. 2017. 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 '17).* ACM, New York, NY, USA, 322–326. https://doi.org/10.1145/3079628.3079706

[22] Carine Pierrette Mukamakuza, Dimitris Sacharidis, and Hannes Werthner. 2018. Mining User Behavior in Social Recommender Systems. In *WIMS.* ACM, 37:1–37:6.

[23] Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). 2015. *Recommender Systems Handbook.* Springer. https://doi.org/10.1007/978-1-4899-7637-6

[24] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *NIPS.* 1257–1264. http://papers.nips.cc/paper/3208-probabilistic-matrix-factorization

[25] Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *WWW.* 285–295. https://doi.org/10.1145/371920.372071

[26] Daniel Schall. 2015. Social Network-Based Recommender Systems. In *Springer International Publishing.*