

Recommending Sellers to Buyers in Virtual Marketplaces Leveraging Social Information

Lukas Eberhard
Graz University of Technology, Austria
lukas.eberhard@tugraz.at

Christoph Trattner
Know-Center, Austria
ctrattner@know-center.at

ABSTRACT

Social information such as stated interests or geographic check-ins in social networks has shown to be useful in many recommender tasks recently. Although many successful examples exist, not much attention has been put on exploring the extent to which social impact is useful for the task of recommending sellers to buyers in virtual marketplaces. To contribute to this sparse field of research we collected data of a marketplace and a social network in the virtual world of Second Life and introduced several social features and similarity metrics that we used as input for a user-based k -nearest neighbor collaborative filtering method. As our results reveal, most of the types of social information and features which we used are useful to tackle the problem we defined. Social information such as joined groups or stated interests are more useful, while others such as places users have been checking in, do not help much for recommending sellers to buyers. Furthermore, we find that some of the features significantly vary in their predictive power over time, while others show more stable behaviors. This research is relevant for researchers interested in recommender systems and online marketplace research as well as for engineers interested in feature engineering.

Keywords

Online Marketplaces; Recommender Systems; Online Social Networks

1. INTRODUCTION

Utilizing social data for the task of recommending certain types of entities to people has gained great popularity recently [9, 13, 16, 24]. Although a growing body of research exists, exploring new methods and algorithms to recommend items to people more efficiently, not much attention has yet been paid to the usefulness of certain social information available in social networking sites for the task of recommending items or people to people. Especially in the context of e-commerce websites and online marketplaces, the value of social information available in external social networking platforms is to a great extent yet unexplored. Most of the current

research still leverages information that is available within the e-commerce platform, ignoring useful social information [14]. To contribute to this area of research, we present in this paper a work in progress of a research effort that aims at understanding the usefulness of social signals for recommendations in e-commerce websites. We focus upon social signals that are typically available in online social network sites such as Facebook. In particular, we are interested in understanding the usefulness of social information such as likes, comments, group joins, interest statements, geographic check-ins and corresponding similarity features for the task of recommending sellers to buyers in online marketplaces. For the recommendation task we have chosen a user-based k -nearest neighbor (KNN) collaborative filtering (CF) approach.

Problem Statement. In this paper we deal with the following problem: for any user visiting an online marketplace (for whom we also have social networking information available) at a certain time we generate a list of new sellers (sellers the potential user has not observed previously) that the user will most probably buy from in the future. To do so, we try to introduce several social features and similarity metrics from the social networking activities of the user. We use them to train a set of user-based KNN CF models based on these features to generate an optimal list of top- N relevant new sellers to the buyer at a given time to investigate what types of features are the most useful ones at that point of time.

Research Questions. The following research questions were posed:

- **RQ1: Recommending Sellers to Buyers.** Knowing that social networking information and corresponding features can help in recommending products to people in online marketplaces [14], to what extent are certain social information and corresponding user-similarity features useful in a user-based KNN CF setting for the task of recommending sellers to buyers?
- **RQ2: Feature Performance Over Time.** To what extent are these features and the corresponding CF approaches useful over time? This question is typically neglected in recommender systems research, but one which we argue is important to ask, since online marketplaces are typically very dynamic where new sellers and buyers appear nearly every day.

Results. Based on a number of experiments in the virtual world of Second Life (SL) we find that not all social information and corresponding similarity metrics are useful in a user-based CF setting to recommend new sellers to buyers in the marketplace of SL. In fact we find that social information such as joined groups or stated interests induced from the online social network are almost as useful as historical information such as product categories di-

rectly induced from the marketplace. Interestingly, compared to a MostPopular (MP) baseline, location-based social information is not very suitable to tackle the defined problem. This is in line with previous observations that people in virtual worlds are not bound to certain places due to the possibility to teleport to places [2].

Contributions. The main contributions of this work are many-fold, but can be broken down to the following points:

- First, we believe this study is unique in a way that it tackles the problem of recommending new sellers to buyers in marketplaces.
- Second, the study contributes to a better understanding of the seller-buyer recommendation problem by investigating the extent to which social information is useful in a user-based CF setting through a number of offline experiments—a feature that has not been investigated yet.
- Finally, the study shows to what extent the induced social features are useful over time—an important property that to the best of our knowledge has not been reported yet.

The paper is structured as follows: In Section 2 we provide an overview of relevant related work in this area. The datasets used in this work are described in Section 3. Section 4 provides a detailed description of the experimental setup. The results of our experiments are presented in Section 5. Finally, Section 6 reports some conclusions that can be drawn from this work and highlights some future directions which are worth to be further explored.

2. RELATED WORK

Using social information to provide or improve recommenders is a relatively new strand of research. Most notable work in this direction has been performed recently in the context of, for instance, recommending points-of-interest to people (e.g., [8]), recommending tags to people (e.g., [10]), or predicting social interactions (e.g., [4, 19]) or relations (e.g., [22]).

In the context of e-commerce not much work has been performed yet and only a few studies exist typically focusing on algorithmic advances to predict the rating or ranking of items people might prefer [9, 13, 14, 16, 24]. Studies investigating the extent to which social information is useful for the task of recommending sellers to buyers are rare and to the best of our knowledge only one other research effort (apart from our own preliminary research investigations using direct seller-buyer features and machine learning approaches [21]) exists so far.

The study of Guo et al. [11] was performed to investigate the predictive power of social features such as direct and indirect interactions between sellers and buyers on the Chinese website Taobao (one of the world’s largest electronic marketplaces) to recommend sellers to buyers. Among other things, Guo et al. find that direct seller-buyer interactions and product meta-data information are the best features to tackle the task. Although, their work is similar to our own one, many significant differences can be found.

First, contrary to our study, the work of Guo et al. relies on social network data that has been directly induced from interactions between users in Taobao. Compared to this, our study is based on features and interactions that were induced from an external social networking platform that is independent from the marketplace itself. Second, we study a much richer set of features induced from social information such as user check-ins, user interests, group joins, etc. Information that has to the best of our knowledge not been leveraged yet for this kind of task. Third, we use our features in the context of a user-based CF method, a well-established and robust recommendation approach often used in e-commerce websites, while

Table 1: Basic statistics of the marketplace dataset.

Marketplace Dataset	
#Users	87,300
#Trading Interactions	268,852
#Trading Relations	219,889
#Sellers	17,914
#Buyers	77,645
#Sellers+Buyers	8,259
#Product Categories	22
#Products	120,762
Average #Products per Seller	≈ 7
Average #Purchases per Seller	≈ 15
Average #Purchases per Buyer	≈ 3

the study of Guo et al. use a machine learning approach called RankSVM to generate a list of preferred sellers. Finally, we show the extent to which the induced features and similarity metrics are useful over time, a concept that to the best of our knowledge has been neglected yet in all of the related works.

3. DATASETS

In our study we rely on two datasets obtained from the virtual world SL. The main reasons for choosing SL over real world sources are manifold, but mainly due to the fact that currently there are no other datasets available that comprise marketplace and social data of users at the same time.

3.1 Marketplace Dataset

SL provides an online trading platform called Second Life Marketplace where SL users are able to trade with virtual goods. Similarly to online shopping platforms such as eBay a user can be a seller, a buyer, or both. To collect this kind of information we gathered all store sites of the SL Marketplace with a web crawler. This crawler detected 131,087 stores/sellers, whereof 36,330 had at least one product in supply and 17,914 sold at least one product (for our study we only relied on those). Overall 1,725,449 products in 22 different categories (e.g., “Avatar Accessories” or “Vehicles”), with different prices and user ratings were found, from which 120,762 were purchased at least once. The total number of noticed purchases was 268,852 with 77,645 different buyers. Due to the fact that a seller can also be a buyer and a buyer can also be a seller, 8,259 users acted as both seller and buyer. The total number of involved users was 87,300. This obtained data stretches from July 2005 to February 2013. A basic overview of the marketplace dataset is provided in Table 1. Linking all sellers with their buyers based on the product reviews was our basic idea for the marketplace network for the experiments in this paper. Figure 1 shows the purchase distribution for the marketplace users, the transacted purchases over time and the distribution of overall users with the fraction of new ones in a period of more than seven years. It exhibits that the SL Marketplace became more popular over time since the absolute number of purchases ascends correspondingly.

3.2 Online Social Network Dataset

Similarly to the real world, users in the virtual world of SL are able to establish social links through an online social networking platform called My Second Life. It was introduced by Linden Lab in 2007 and can be compared with other online social networks such as Facebook or Google+. This platform gives SL users the opportunity to present personal information on their user profiles or to interact with other users on the so-called Feed, which can be

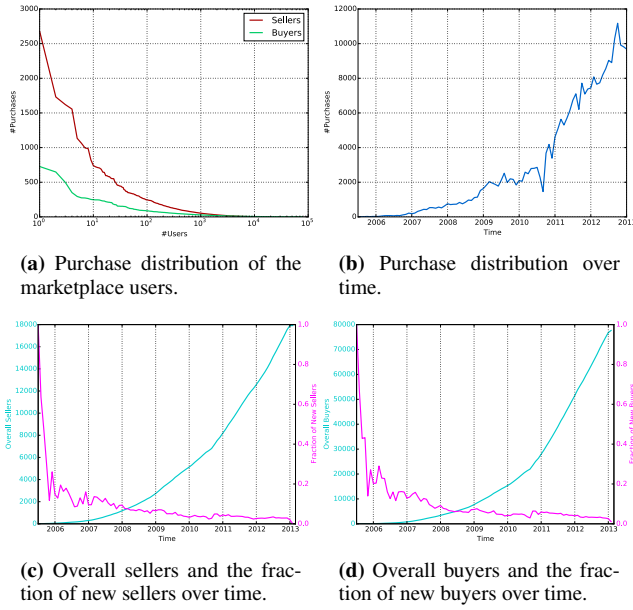


Figure 1: Distributions of purchases, sellers and buyers of the marketplace dataset.

compared with the Timeline in Facebook. A considerable difference to Facebook exists concerning friendship relations. Such a relation type does not exist in My Second Life [18].

At the end of March 2013 we crawled the SL profiles of users who had not changed their profiles to private, based on the crawling methodology as described in our previous work [22]. For each user we obtained the stated interests, the joined groups, the feed interactions with others (text messages, pictures, comments, likes) and the preferred in-world locations—so-called favored regions. Also in-world check-ins can be shared, which is a similar concept to Foursquare check-ins in Facebook. The basic statistics of this dataset are available in Table 2.

4. EXPERIMENTAL SETUP

In this section we provide a detailed description of our experimental setup. First, we describe the recommender approaches used to tackle the task of recommending sellers to buyers. After that, we introduce the similarity features we have chosen in the two provided datasets, that form the basis of our recommendation approach in order to tackle the task of recommending sellers to buyers. Finally, we describe the evaluation methodology and the metrics used in our study.

4.1 Recommendation Approaches

Baseline. As baseline we chose a simple MP recommender approach that recommends the most popular sellers to a potential buyer. Popularity was computed in terms of the number of purchase transactions the user performed.

User-Based CF. The main approach we adopted in order to tackle the task of recommending sellers to buyers is a user-based CF approach [15]. The basic idea of this approach is that buyers who are similar to each other will behave in a similar manner in the marketplace [17]. Out of the different CF approaches, we used the non-probabilistic user-based KNN algorithm, where for each user we first find the k -nearest similar users and create a ranked list of their sellers. Afterwards, we recommend only the top- N sellers of

Table 2: Basic statistics of the online social network dataset.

Online Social Network Dataset	
#Users	152,509
#Postings (Text Messages/Pictures)	226,668
#Comments	348,106
#Likes	1,494,044
#Group Joins	1,869,281
#Stated Interests	227,596
#Check-in Postings	466,930
#Unique Check-in Regions	13,251
#Users with Check-ins	36,430
#Stated Favored Regions	337,732
#Unique Favored Regions	22,742
#Users who stated Favored Regions	76,093

the list that are new to the target user (i.e., the user is not a customer of these sellers).

In particular, we calculated the similarity values between the user pairs $sim(u, v)$ based on the user similarity features proposed in Section 4.2 (e.g., constructing the neighborhood). We defined the k -nearest neighbors of a buyer b as $neighbors(b)$ and the coefficient $S_{s,b}$ is 1, if b is a customer of seller s , and 0 otherwise. Based on these values, we ranked each seller s of the k most similar buyers to b using the following formula [17]:

$$pred(b, s) = \sum_{n \in neighbors(b)} sim(b, n) \cdot S_{s,n} \quad (1)$$

In our experiments we applied various numbers for the parameters k and N . In this paper we only present the results with the best performance of our CF approach that was obtained when setting $k = 100$ and $N = 5$ (see Section 5).

4.2 Similarity Features

In this section we describe in detail the features we induce from our two datasets which form the basis for our user-based CF approach as introduced in the previous section. Utilizing different features from different data sources in our CF method not only allows us to compare the predictive power of each feature but also helps us to understand what type of data source (in our setting online social network vs. marketplace data) is the most valuable one. As shown in our previous work [19], similarities between users can be derived in two different ways. Either we calculate similarities between users on the content (e.g., user interests, products purchased or groups they joined, denoted further as *homophilic features*), or on the network structure of the respective network, denoted as *network features*. In the following, we describe the types of similarity features we induced from the SL Marketplace dataset and from the SL online social network.

Network Features. As features for the structure of a network we used the following measures, where $N(u)$ are the neighbors of a user u in the network. We denoted incoming neighbors as $N^-(u)$ and outgoing neighbors as $N^+(u)$:

- *Adamic Adar* [1, 6]:

$$sim(u, v) = \sum_{z \in N^-(u) \cap N^-(v)} \frac{1}{\log(|N^-(z)|)} \quad (2)$$

- *Jaccard's Coefficient* [7, 18, 20]:

$$sim(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (3)$$

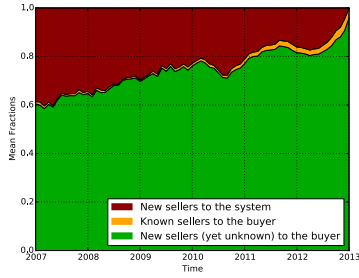


Figure 2: Mean fractions of sellers who are new to the buyer or system over time. As shown, over 60% of the sellers are new to the buyers and only a few sellers are known to them. This trend is increasing in time, showing the potential of a recommender system that recommends sellers to buyers in the SL Marketplace.

- *Preferential Attachment Score* [3, 6]:

$$\text{sim}(u, v) = |N^-(u)| \cdot |N^+(v)| \quad (\text{or vice versa}) \quad (4)$$

- *Interactions* [18]:

$$\text{sim}(u, v) = |\text{interactions}(u, v)| \quad (5)$$

- *Reciprocity* [6]:

$$\text{sim}(u, v) = \begin{cases} 1 & \text{if link in both directions} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The *Jaccard's Coefficient* feature of the *network features* was split into incoming and outgoing features. This means that, either only the incoming neighbors or only the outgoing neighbors of the users in the network were considered.

Homophilic Features. We constructed the following content-based similarity features, where $V(u)$ is a vector of items of a user u .

- *Jaccard's Coefficient*:

$$\text{sim}(u, v) = \frac{|V(u) \cap V(v)|}{|V(u) \cup V(v)|} \quad (7)$$

This feature was applied in the case of the SL online social network to the user's interests, joined groups, check-ins and favored regions.

- *Cosine Similarity* [18]:

$$\text{sim}(u, v) = \frac{V(u) \cdot V(v)}{\|V(u)\| \|V(v)\|} \quad (8)$$

This measure was used in the case of the SL Marketplace to the user's product categories, product prices and product ratings.

4.3 Evaluation

The evaluation protocol we followed in this paper is one usually used in order to evaluate a recommender system offline in a time-based manner [5]. First, we considered in our evaluation only users who were present in both datasets (marketplace and online social network) in order to have a fair comparison of the two data sources. Second, for the sellers to buyers recommendation task we considered only those sellers as relevant who have not been observed by the buyer before (i.e., we only recommend sellers to buyers with no trading transactions in the past). Figure 2 presents the mean

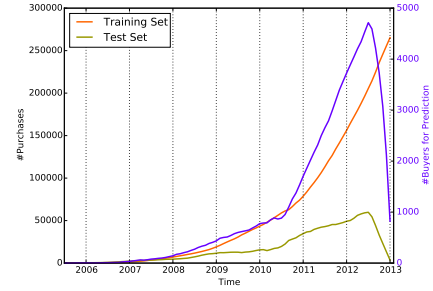


Figure 3: The sizes of the training and test sets and the number of buyers for whom a recommendation can be computed over time. As shown, until 2009 training and test sets are relatively small.

fractions of sellers who are new to the buyer or system over time. As shown, a huge fraction of sellers (over 60%) are always new to the buyer showing the potential of a seller to buyer recommender systems.

We split the SL Marketplace dataset in training and test samples according to the timeline. Consequently, we did the same with the SL online social network. The methodology we follow here is to train our recommender on all historical data available at some point in time t and to use the next forthcoming n months in time for testing. In particular, we generated recommendations every month over the time line (using all historical purchase events for training) beginning in 2007 until 2013 and used the purchase events of the next 6 forthcoming months for testing.

Figure 3 shows the sizes of the training and test sets with respect to the number of purchases and the number of buyers for whom a recommendation can be computed over time. Since the available data is very sparse at the beginning of our timeline, we consider only the results between 2009 and 2013.

In order to determine the predictive power of our recommendation approach two evaluation metrics typically used in recommender systems were employed. In particular, we used the F_1 score ($F1@5$) and the User Coverage to show the extent to which the corresponding similarity features and datasets perform [12].

5. RESULTS

In this section we present the results of our experiments. First, we show how the datasets and the corresponding induced similarity features from the SL online social network and the SL Marketplace perform in the context of our user-based CF approach for the task of recommending sellers to buyers—here we are interested in the social data source and the corresponding social information (RQ1). After that, we show how well these features perform over time (RQ2).

5.1 RQ1: Recommending Sellers to Buyers

Figure 4 shows the mean values of the F_1 score (left y-axis) for each used feature (x-axis) with the respective User Coverage (right y-axis) from 2009 to 2013. As shown, *homophilic features* such as *Groups Jaccard* or *Interests Jaccard* as found in the social network are very valuable similarity features in a user-based CF setting to recommend sellers to buyers efficiently compared to the MP baseline. They are even to the same extent useful as historical features, such as *CosSim Product Categories* induced directly from the SL Marketplace. Interestingly, when comparing location-based social features to a MP baseline, *Favored Regions Jaccard* just show little improvement, while *Check-ins Jaccard* could not improve the results. This is in line with previous observations that people in

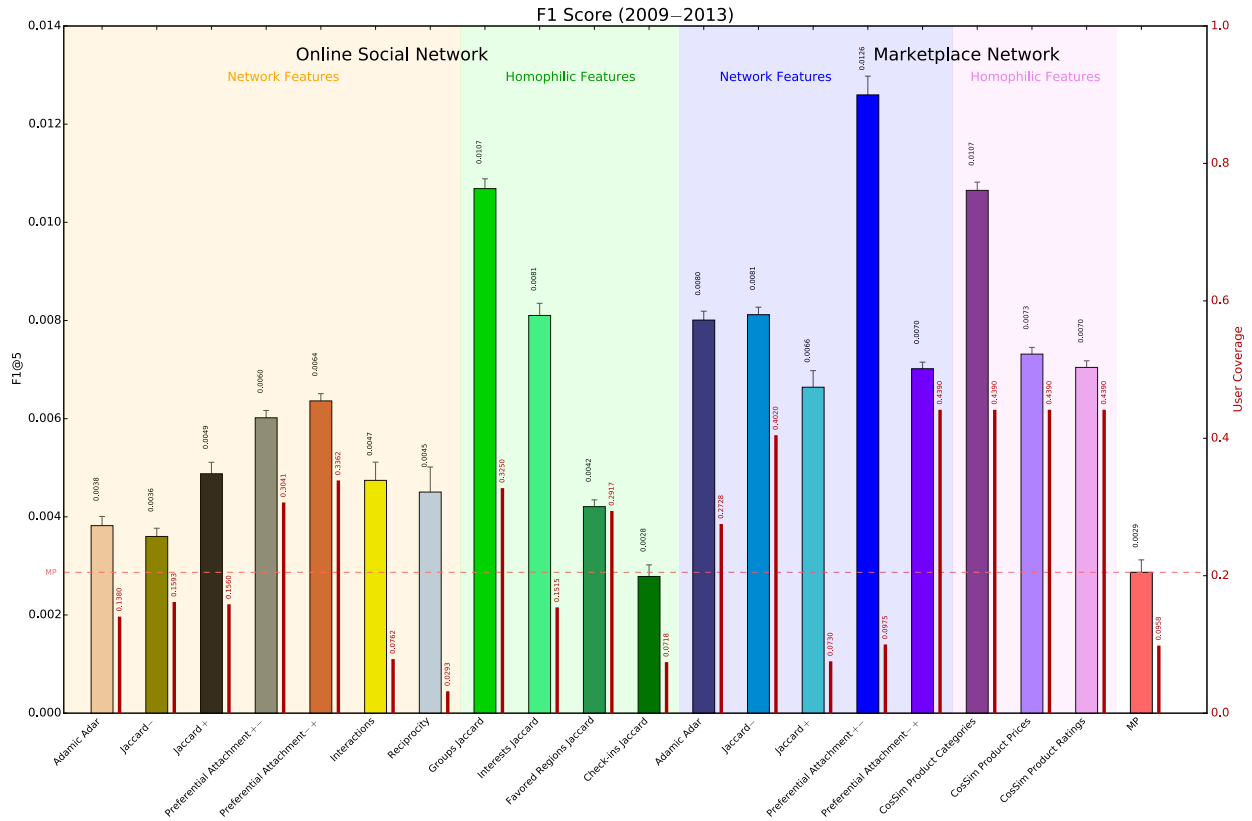


Figure 4: Mean values of the F_1 score of the similarity features and their respective User Coverage over four years. As shown, homophilic features such as *Groups Jaccard* or *Interests Jaccard* as found in the social network are very valuable similarity features in a user-based CF approach to recommend sellers to buyers efficiently compared to the MP baseline (dashed line for comparison). They are even to the same extent useful as historical features such as the *CosSim Product Categories* induced directly from the SL Marketplace.

virtual worlds are not bound to certain places due to the possibility to teleport to places [2].

Note that the User Coverage for the MP is under 100%. This can happen, since in our recommender task we only consider sellers which are not yet known to the buyer (see Figure 2).

5.2 RQ2: Feature Performance Over Time

Figure 5a shows the *network features* of the online social network over time. Although the performance of the features varies over time, it indicates that most of our *network features* of the social dataset are above the baseline at each point in time. As Figure 5b shows, the joined groups and stated interests are powerful information regarding sellers to buyers recommendations. One potential explanation for the improving performance of the *Groups Jaccard* feature compared to the MP over time could be the increasing amount of data available (see Figure 3).

As expected, the values of all features of the marketplace network are above the baseline most of the time as Figures 5c and 5d reveal. The *Preferential Attachment Score+ -* feature of the *network features* and the *CosSim Product Categories* feature of the *homophilic features* are the most suitable features for sellers predictions. Except for the—for us unaccountable—peak at year 2010, both features also become more suitable for our recommendation task from time to time.

As shown in Figure 5, the User Coverage of the MP approach slightly decreases over time. The reason for this behavior is the

strong increase of buyers in the system in 2011 (see Figures 1d and 3).

6. CONCLUSIONS

In this paper we extended our understanding of the signals available in social networking sites for the task of recommending sellers to buyers in dynamic online marketplaces. We approached this by conducting several offline experiments over time by employing a user-based KNN CF method using several user similarity metrics that have been derived from social information such as likes, comments, joined groups, checked-in places or stated interests. As our experiments reveal, most types of the social information we used are useful for the task of recommending new sellers to buyers in online marketplaces. Furthermore, we find that the methods vary significantly over time raising the question, if better time-dependent alternatives can be found that better adapt to the statistical properties of our dataset.

Limitations & Future Work: One of the limitations of our study is that we conducted our experiments only on one dataset. Applying our methods to other types of datasets would be an interesting extension of our work. Another limitation are the features for the predictions task, for which we believe better time-dependent alternatives could be found [23]. Finally, it would be interesting to apply machine learning to this kind of recommendation task (e.g., in the form of a learning to rank method that combines features) [8] and to study the extent to which direct (as proposed in our previous work [21]) vs. indirect features compare with each other.

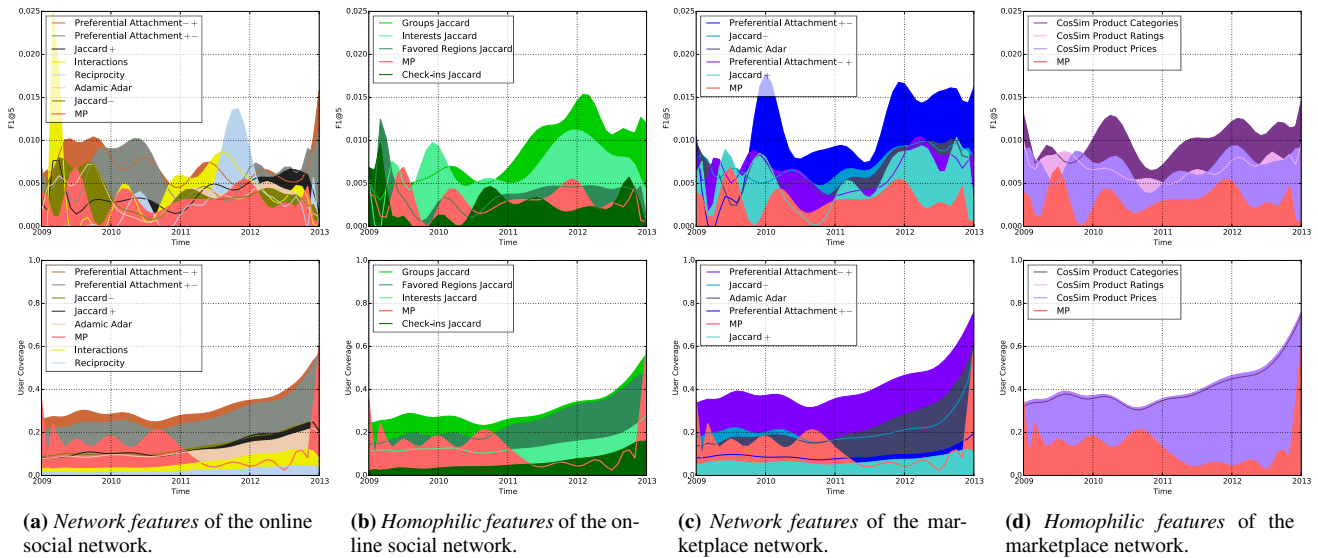


Figure 5: F₁ score and User Coverage for the induced similarity features in the different data sources (online social network & marketplace network) over time. As shown, the network features of the online social network and the trading network oscillate over time while homophilic features behave more stable. Furthermore, some trends over time can be observed. Features such as Groups Jaccard, Preferential Attachment+/- of the online social network and CosSim Product Categories of the marketplace perform better over the years.

7. REFERENCES

- [1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.
- [2] L. Balby Marinho, C. Trattner, and D. Parra. Are real-world place recommender algorithms useful in virtual world environments? In *Proc. RecSys’15*, pages 245–248. ACM, 2015.
- [3] A.-L. Barabasi and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.
- [4] K. Bischoff. We love rock ‘n’ roll: analyzing and predicting friendship links in Last.fm. In *Proc. WebSci’12*, pages 47–56. ACM, 2012.
- [5] P. Campos, F. Diaz, and I. Cantador. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2):67–119, 2014.
- [6] J. Cheng, D. M. Romero, B. Meeder, and J. Kleinberg. Predicting Reciprocity in Social Networks. In *Proc. SocialCom’11*, pages 49–56, 2011.
- [7] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *Proc. UbiCom’10*, pages 119–128. ACM, 2010.
- [8] A. Q. de Macedo, L. B. Marinho, and R. L. T. Santos. Context-aware event recommendation in event-based social networks. In *Proc. RecSys’15*, pages 123–130. ACM, 2015.
- [9] J. Delporte, A. Karatzoglou, T. Matuszczyk, and S. Canu. Socially enabled preference learning from implicit feedback data. In *Machine Learning and Knowledge Discovery in Databases*, pages 145–160. Springer, 2013.
- [10] W. Feng and J. Wang. Incorporating heterogeneous information for personalized tag recommendation in social tagging systems. In *Proc. SIGKDD’12*, pages 1276–1284. ACM, 2012.
- [11] S. Guo, M. Wang, and J. Leskovec. The role of social networks in online shopping: information passing, price of trust, and consumer choice. In *Proc. EC’11*, pages 157–166. ACM, 2011.
- [12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, Jan. 2004.
- [13] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. RecSys’10*, pages 135–142. ACM, 2010.
- [14] E. Lalic, D. Kowald, L. Eberhard, C. Trattner, D. Parra, and L. B. Marinho. Utilizing online social network and location-based data to recommend products and categories in online marketplaces. In *Mining, Modeling, and Recommending Things in Social Media*, pages 96–115. Springer, 2015.
- [15] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.
- [16] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proc. WSDM’11*, pages 287–296. ACM, 2011.
- [17] J. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web*, volume 4321 of *Lecture Notes in Computer Science*, pages 291–324. Springer Heidelberg, 2007.
- [18] M. Steurer and C. Trattner. Acquaintance or Partner? Predicting Partnership in Online and Location-Based Social Networks. In *Proc. ASONAM’13*, pages 1–8. ACM, 2011.
- [19] M. Steurer and C. Trattner. Predicting Interactions in Online Social Networks: An Experiment in Second Life. In *Proc. MSM’13*, pages 5:1–5:8. ACM, 2013.
- [20] M. Steurer, C. Trattner, and D. Helic. Predicting Social Interactions from Different Sources of Location-based Knowledge. In *Proc. SOTICS’13*, pages 8–13, 2013.
- [21] C. Trattner, D. Parra, L. Eberhard, and X. Wen. Who will trade with whom?: Predicting buyer-seller interactions in online trading platforms through social networks. In *Proc. WWW’14*, pages 387–388, 2014.
- [22] C. Trattner and M. Steurer. Detecting partnership in location-based and online social networks. *Social Network Analysis and Mining*, 5(1), 2015.
- [23] T. Tylenda, R. Angelova, and S. Bedathur. Towards time-aware link prediction in evolving social networks. In *Proc. SNMA’09*, page 9. ACM, 2009.
- [24] Y. Zhang and M. Pennacchiotti. Predicting Purchase Behaviors from Social Media. In *Proc. WWW’13*, pages 1521–1532, 2013.