Social Relations versus Near Neighbours: Reliable Recommenders in Limited Information Social Network Collaborative Filtering for Online Advertising

Dionisis Margaris
Department of Informatics and Telecommunications
University of Athens
Athens, Greece
margaris@di.uoa.gr

Dimitris Spiliotopoulos and Costas Vassilakis
Department of Informatics and Telecommunications
University of the Peloponnese
Tripoli, Greece
dspiliot@uop.gr, costas@uop.gr

***Abstract*—Online advertising benefits by recommender systems since the latter analyse reviews and rating of products, providing useful insight of the buyer perception of products and services. When traditional recommender system information is enriched with social network information, more successful recommendations are produced, since more users' aspects are taken into consideration. However, social network information may be unavailable since some users may not have social network accounts or may not consent to their use for recommendations, while rating data may be unavailable due to the cold start phenomenon. In this paper, we propose an algorithm that combines limited collaborative filtering information, comprised only of users' ratings on items, with limited social network information, comprised only of users' social relations, in order to improve (1) prediction accuracy and (2) prediction coverage in collaborative filtering recommender systems, at the same time. The proposed algorithm considerably improves rating prediction accuracy and coverage, while it can be easily integrated in recommender systems.**

Keywords— Online Advertising, Social Networks, Collaborative Filtering, Limited Information, Near Neighbours, Pearson Correlation Coefficient, Evaluation.

I. INTRODUCTION

Recommender systems (RS) are utilized by online advertising in several aspects, ranging from the type of feedback (product ratings, reviews, user sentiment, and more) to sparsity. Even more, online advertising strategies can be augmented by recommender systems that are handling issues such as popularity bias by inducing greater diversity of recommended items [1]. A widely-used approach for making recommendations, stemming from user behaviour and actions is collaborative filtering. Collaborative filtering (CF) synthesizes the informed opinions of humans (notably these opinions in many cases encompass the aspect of satisfaction), to make personalized and accurate predictions and recommendations [2]. Its biggest advantage is that explicit content description is not required (as in content-based systems): instead, traditional CF relies only on opinions expressed by users on items either explicitly (e.g. a user enters a rating for the item) or implicitly (e.g. a user purchases an item, or clicks an advertisement banner, which indicates a positive assessment). In the context of CF, personalization is achieved by considering ratings of ''similar users'', under the CF fundamental assumption that if users X and Y have similar behaviour (e.g. listening, watching, buying, rating assignment) on some items, they will act similarly on others [2,3].

When traditional CF is employed, typical recommender systems (RSs) assume that users are independent and ignore the social interactions among them, hence they fail to incorporate important aspects that denote interaction, tie strength and influence among users, which can substantially enhance recommendation quality [4,5,6].

Social network (SN) data-based RSs take into account static data from the user profile, such as location, age or gender, complemented with dynamic aspects and contextual information stemming from users' behaviour and/or SN state, such as time, users' mood, items' general acceptance and social influence [7,8] to supplement the traditional CF data (such as users' ratings and items' static characteristics). By taking this information into consideration, in the recommendation process, the SN RSs manage to achieve better user-targeted recommendations.

However, when both the CF and the SN information available is limited, containing only user ratings and user social relations, respectively, (e.g. the information that an external application may have available), a SN CF system faces the much harder task to successfully combine this limited information, in order to produce predictions and recommendations, better than the ones the traditional CF system will produce. Theoretically, the success of such tasks would depend on the types of available SN information and qualitative characteristics such as *denseness/sparseness*. The challenge is to design an approach that performs consistently across the board of datasets that expose existing but limited SN CF information.

In this paper, we (1) propose a simple, yet effective algorithm that combines limited CF information of user ratings and user social relations (friendship, trust, etc.), aiming to consistently improve both the rating prediction quality and prediction

ASONAM '19, August 27–30, 2019, Vancouver, BC, Canada

ACM 978-1-4503-6868-1/19/08…$15.00
https://doi.org/10.1145/3341161.3345620

coverage (i.e. the percentage of the cases for which a personalized prediction can be computed) and (2) evaluate the proposed algorithm against five contemporary SN CF datasets. In our experiments, (a) *sparse* and *dense* CF datasets (a CF dataset's density refers to the number of ratings when compared to the number of users and items in it [9]), as well as (b) *sparse* and *dense* SN datasets (a SN dataset's density refers to the number of relations when compared to the number of users in it [2]), as well as (c) *directed edge* (trusts) and *undirected edge* (friendships) SN datasets were used.

Our results show that the proposed algorithm introduces considerable prediction accuracy gains, reducing both prediction error metrics (MAE and RMSE), as well as coverage increase, when compared to the plain CF algorithm in all tested datasets. The results enable the use of the proposed algorithm in cases where social information is available, yet limited, and may lead to advanced online advertising models that include findings from the social ties both from the recommender and the real time application of online advertising. Certain RS challenges that can also benefit are the simultaneous recommendation from new users and items (cold start problem) and personalization (how vs what is recommended).

It is worth noting that the proposed technique can be combined with other algorithms that have been proposed for improving rating prediction accuracy, recommendation quality or prediction coverage in CF-based systems, focusing either in traditional CF-based systems (clustering techniques, concept drift, etc.) [10,11,12,13,14] or in SN CF-based systems (influence, trust, etc.) [4,5,6,7,15].

The rest of the paper is structured as follows: section 2 overviews related work, while section 3 presents the proposed algorithm. Section 4 evaluates the proposed algorithm using 5 contemporary datasets and finally, section 5 concludes the paper and outlines future work.

II. Related Work

SN data have been identified as an important means for increasing recommendation accuracy and broadening recommendation variety in contemporary RSs.

The need for extending recommendation algorithms to take into account the SNs structure is established in [15] and [16]; in these works, measures for user tie strength and social influence are also proposed. In more detail, [15] presents a predictive model that maps social media data to tie strength, distinguishing between strong and weak ties with generally accepted accuracy and illustrates how modelling tie strength can improve social media design elements, including privacy controls, message routing, friend introductions and information prioritization, while [16] measure social influence via social cues on an economically relevant form of user behaviour and average rates of response, demonstrate the substantial consequences of including minimal social cues in advertising and measure the positive relationship between a consumer's response and the strength of their connection with an affiliated peer. [17] examines the role of SNs in the recommendation process within a large-scale field experiment that randomizes exposure to signals about friends' information and the relative role of strong and weak ties. [18] develops a way to increase recommendation effectiveness by combining SN information and CF methods, by collecting data about users' preference ratings and their SN relationships from a SN Web site and by developing approaches for selecting neighbours and amplifying friends' data. [19] investigates the role of SNs relationships in developing a track recommendation system based on a common CF item recommendation method, by taking into account both the social annotation and friendships inherent in the social graph established among users, items and tags. [20] enhances a content-based RS by including the trust between individuals, users' interaction and aspects of each user's personality in the recommendation algorithm. [21] outlines a conceptual RS design within which the structure and dynamics of a SN contribute to the dimensions of trust propagation, source's reputation and tie strength between users, which are then taken into account by the system's prediction component to generate recommendations. [22] proposes a trust-aware system for personalized user recommendations in SNs. In this work, trust (and distrust) between people is a central concept in the algorithm that assists members of a community to make decisions about other members of the same community. [23] examines how social data, ratings and reviews can be combined to create personalized rankings of reviews, tailored to each individual user, thus providing each user with efficient access to the reviews that are deemed most suitable for him. [24] develops a way to increase recommendation effectiveness by incorporating SN information into CF. It collects data about users' preference ratings and their SN relationships from a social networking website and then it evaluates CF performance with diverse neighbour groups combining groups of friends and nearest neighbours (NNs). [25] proposes a method to incorporate social trust information (i.e., trusted neighbours explicitly specified by users) in providing recommendations. More specifically, ratings of a user's trusted neighbours are merged to complement and represent the preferences of the user and to find other users with similar preferences (i.e., similar users). In addition, the quality of merged ratings is measured by the confidence considering the number of ratings and the ratio of conflicts between positive and negative opinions. Further, the rating confidence is incorporated into the computation of user similarity. The prediction for a given item is generated by aggregating the ratings of similar users. [3] proposes an algorithm for query personalization, based on SN information. The algorithm exploits both the users' choices of items (browsing and rating) and the influence information from SNs to suitably adapt queries for each user. Query adaptation is performed via (re)writing the query sorting specification to order the qualifying data according to their projected interest for the user. [7] extends a SN-based RS, which utilizes the information in SN to improve the performance of traditional RSs. Two methods were proposed to improve performance, classifying the correlations between pairs of user ratings to improve the accuracy of the system and making the system robust to sparse data.

However, all these works require additional information, concerning either the users (age, location, gender, etc.) or the items ((sub)category, price, reliability, etc.) or the SN (tie strength, social influence, etc.).

The present work fills this gap by presenting an algorithm that takes into account only the basic information a CF system can store, the *user-item-rating* tuple with the basic information a SN CF can store, the *user-user-relation* tuple. To the best of our knowledge, no other algorithm combines limited CF information with limited SN information, in order to increase prediction accuracy and coverage.

III. The Proposed Algorithm

In CF, predictions for a user U are computed based on U's *NNs*, i.e. a set of users that have rated items similarly to U. The similarity metric between two users U and V is typically based on the Pearson Correlation Coefficient (PCC) [2], which is expressed as:

$$sim(U,V) = \frac{\sum_k (r_{U,k} - \overline{r_U}) * (r_{V,k} - \overline{r_V})}{\sqrt{\sum_k (r_{U,k} - \overline{r_u})^2 * \sum_k (r_{V,k} - \overline{r_V})^2}} \quad (1)$$

where k ranges over items that have been rated by both U and V, while $\overline{r_U}$ and $\overline{r_V}$ are the mean values or ratings entered by users U and V, respectively. Then, for user U, his NN users NN_U are selected out of the users with whom a positive similarity has been computed.

Afterwards, for computing a rating prediction $p_{U,i}$ for the rating of user U on item i, the computation is expressed as:

$$p_{U,i} = \overline{r_u} + \frac{\sum_{V \in NN_u} sim(U,V) * (r_{V,i} - \overline{r_V})}{\sum_{V \in NN_u} sim(U,V)} \quad (2)$$

The proposed algorithm introduces the concept of *SN NNs, of a user U*, i.e. users who share a social relation (i.e. trust or friendship) with user U, who will be denoted as *SN_NNs*; for clarity, we will denote the initial CF NNs of a user U, as *CF_NNs*.

The ratings entered by the *SN_NNs* are then utilized in the rating prediction formulation process for user U. More specifically, the rating prediction process proceeds as follows:

Step 1: Initially, we compute the *mean-centered rating* [26] of item i in the *CF_NN*; we will denote this quantity as $p^{CF}_{U,i}$. The mean-centered rating metric of users with positive PCC is computed as:

$$p^{CF}_{U,i} = \frac{\sum_{V \in CF\_NN_U} sim(U,V)_{CF} * (r_{V,i} - \overline{r_V})}{\sum_{V \in CF\_NN_U} sim(U,V)_{CF}} \quad (3)$$

and corresponds to the second term of equation (2). Note that the mean-centered rating computed by equation (3) cannot be used by itself as a rating prediction: this metric quantifies how item i is perceived by users included in the *CF_NNs set,* with regards to the average of ratings that each user within the *CF_NNs set* has entered. Therefore, a positive value for $p^{CF}_{U,i}$ indicates that item i is ranked by users within the *CF_NNs set* better than their averages and vice versa.

Step 2: Subsequently, we follow the same procedure for computing the mean-centered rating of item i in the *SN_NNs* set; this will be denoted as $p^{SN}_{U,i}$. Formally, this is computed as shown in equation (4):

$$p^{SN}_{U,i} = \frac{\sum_{V \in SN\_NN_U} sim(U,V)_{SN} * (r_{V,i} - \overline{r_V})}{\sum_{V \in SN\_NN_U} sim(U,V)_{SN}} \quad (4)$$

As far as the $sim(U,V)_{SN}$ value is concerned, this will be set to 1.0 (all SN_NNs will be treated the same), unless an explicit value is available in the SN dataset. Considering additional information derived by the SN, such as tie strength, influence, trust value, common/mutual relations, etc. [3,4,5,6], in order to compute-tune the optimal $sim(U,V)_{SN}$ value, will be part of our future work.

Step 3: Finally, the weighted average of $p^{CF}_{U,i}$ and $p^{SN}_{U,i}$ is computed, and the result is adjusted by the mean value of ratings entered by U ($\overline{r_U}$), in order to formulate the rating prediction $p_{U,i}$, as shown in equation (5):

$$p_{U,i} = \begin{cases} \overline{r_U} + p^{CF}_{U,i}, & if\ SN\_NNS = \emptyset \\ \overline{r_U} + p^{SN}_{U,i}, & if\ CF\_NNS = \emptyset \\ \overline{r_U} + w_{CF} * p^{CF}_{U,i} + w_{SN} * p^{SN}_{U,i}, & if \\ \quad SN\_NNS \neq \emptyset \wedge CF\_NNS \neq \emptyset & \end{cases} \quad (5)$$

In equation (5) the w_{CF} parameter denotes the weight assigned to the prediction computed by considering only the *CF_NNs* of U, while the w_{SN} parameter, which is complementary to the w_{CF} parameter ($w_{SN} + w_{CF} = 1.0$), is assigned to the prediction computed by considering only the *SN_NNs*, respectively. If no *CF_NNs* of U's that have rated the item i exist, then the prediction is based exclusively on the ratings of his *SN_NNs*, and vice versa.

In the next section, we investigate candidate w_{SN} values, in order to identify the optimal setting for this parameter and assess the performance of the proposed algorithm.

IV. Performance Evaluation

In this section, we report on the experiments that were designed to measure the prediction improvement, introduced by the presented algorithm, due to the inclusion of the limited SN information.

In this comparison we consider the following aspects:

1. The coverage of the algorithm, i.e. the percentage of the cases for which a personalized prediction can be computed [27].
2. Prediction accuracy; for this comparison, we use two well-established error metrics, namely the mean absolute error (MAE), and the Root Mean Squared Error (RMSE) [2]. The fundamental difference between these two metrics is that the RMSE metric 'punishes' big mistakes more severely, while the MAE metric does not discriminate between big and small errors (the RMSE metric was used, among others, in the Netflix competition [28]).

The prediction coverage and accuracy of the proposed algorithm are compared against those of the plain CF algorithm [2], without the use of the SN relations, which is used as a baseline.

TABLE I. DATASETS SUMMARY

Dataset name	#Users	#Items	#Ratings	Avg. #Ratings / User	Density	#Social Relations	Avg. #Social Relations / User	Type of Items	Type of Relations
Ciao [32]	30K	73K	1.6M	53.4	0.07%	40K	1.3	General	Trust
FilmTrust [33]	1.5K	2.1K	35K	23.5	1.13%	1.8K	1.2	Movies	Trust
Epinions [34,35]	665K	134K	665K	13.5	0.009%	487K	9.9	General	Trust
LibraryThings [35,36]	83K	506K	1.7M	20.5	0.004%	130K	1.6	Books	Trust
Dianping Social Rec 2015 [7,37]	148K	11K	2.1M	14.5	0.13%	2.5M	17	Restaurants	Friendship

To compute the MAE and the RMSE we employ the standard "hide one" technique [29,30,31]: each time, we hide one rating in the database and then predict its value based on the ratings of other non-hidden items. In our first experiment, random ratings of each user are hidden (5 tries per user) and then their values are predicted. To further validate our results, we conduct an additional experiment in every dataset, including the ratings' timestamps, where the last rating of each user in the database is hidden and then its value is predicted. The results of these two experiments are in close agreement (less than 2% difference in results) and therefore we report only on the results of the first experiment, for conciseness purposes.

For our experiments we used a laptop computer equipped with one dual core Intel Celeron N2840@2.16 GHz CPU, with 4 GB of RAM and one 240 GB solid state hard disk with a transfer rate of 375 MBps, which hosted the datasets and ran the rating prediction algorithms.

In the remainder of this section, we present and discuss the results obtained from applying the algorithm presented above on five datasets.

The five datasets used in our experiments have the following properties:

- They contain both user-item ratings and social relations between SN users.
- They are up to date (published the last 10 years) and are widely used for benchmarking in SN CF research.
- They vary with respect to the type of dataset item domain (movies, music, books, restaurants, etc.), size, CF density and SN density.

Table I summarizes the basic properties of these datasets.

In the following paragraphs, we report on our findings regarding the performance of the algorithm proposed in this work, versus the plain CF algorithm using only the CF information of the aforementioned datasets.

A. The "Ciao" Dataset

In this dataset, the coverage of the plain CF algorithm is 28.8%; in the rest of the cases, the users have no CF_NNs [2], hence no prediction could be computed for them. When the SN_NNs are taken into account, the coverage raises to 30%. In terms of relative increase, this equals to 4.1%, which is consider totally adequate, since the dataset's average SN relations per user equals to 1.3, which is considered as very low.

The coverage in all variants of the proposed algorithm, having different settings for w_{SN}, is the same; this is expected, since the w_{SN} parameter does not affect the number of SN neighbours (which is the only parameter affecting the coverage), but rather the degree to which the SN_NNs contribute to the final prediction.

As far as the $sim(U,V)_{SN}$ value, used in formula (4), this is set to the exact values contained in the "trusts" file included in the dataset.

Fig. 1 illustrates the effect of the value of the w_{SN} parameter on the quality of recommendations produced by the proposed algorithm, as this is reflected by the MAE and the RMSE metrics, respectively, using the plain CF algorithm's performance as a yardstick.

Considering the quality of the rating predictions, the MAE and RMSE metrics improve when $w_{SN} \leq 80\%$, while for $w_{neg} > 80\%$ they deteriorate slightly. The best performance is obtained when w_{SN}=30%, where the MAE drops by 1.84% and the RMSE is reduced by 1.34%.

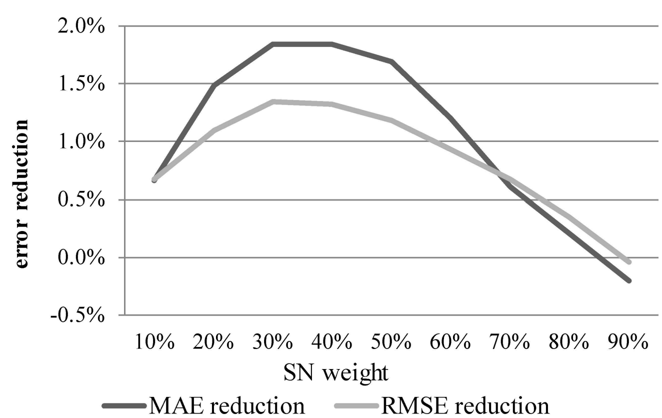

Fig. 1. MAE and RMSE reduction achieved by the proposed algorithm, under different w_{SN} values for the "Ciao" dataset.

Overall, in this dataset the proposed technique offers considerable coverage increase and improvement in the rating prediction quality, despite the low *average SN relations per user* number.

B. The "FilmTrust" Dataset

In this dataset, the coverage of the plain CF algorithm is 79.3%; while the proposed algorithm increases coverage to 83.1%, achieving a 4.7% increase in comparison to the coverage of the plain CF algorithm. Again, due to the low number of SN relations per user (equals to 1.2), in this dataset, this increase is totally adequate.

As far as the $sim(U,V)_{\text{SN}}$ value, used in formula (4), again, this is set to the exact values contained in the "trusts" file included in the dataset.

Fig. 2 illustrates the effect of the value of the w_{SN} parameter on the quality of recommendations produced by the proposed algorithm, as this is reflected by the MAE and the RMSE metrics, again using the plain CF algorithm's performance as a yardstick. Considering the quality of the rating predictions, the MAE and RMSE metrics improve when $20\% \leq w_{SN} \leq 80\%$. The best performance is obtained when w_{SN}=40%, where the MAE drops by 1.35% and the RMSE is reduced by 1.48%. Overall, as previously, in this dataset the proposed technique offers considerable coverage increase and, additionally, an improvement in the rating prediction quality can be obtained, besides its low number of *SN relations per user*.

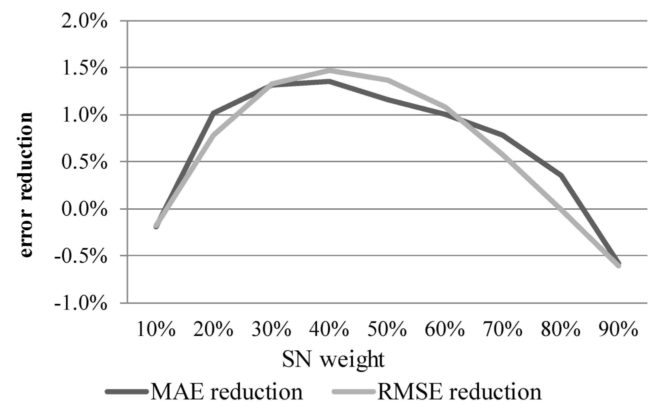

Fig. 2. MAE and RMSE reduction achieved by the proposed algorithm, under different w_{SN} values for the "FilmTrust" dataset.

C. The "Epinions" Dataset

In this dataset, the coverage of the plain CF algorithm is 14.7%; while the proposed algorithm increases coverage to 19.45%, achieving a 32% increase in comparison to the coverage of the plain CF algorithm. The main reason for this major increase (the largest increase in our five datasets), is clearly the large number of the dataset's *average SN relations per user*, which equals to 9.9.

As far as the $sim(U,V)_{\text{SN}}$ value, used in formula (4), this is set to the exact values contained in the "trusts" file included in the dataset.

Fig. 3 illustrates the effect of the value of the w_{SN} parameter on the quality of recommendations produced by the proposed algorithm, as this is reflected by the MAE and the RMSE metrics, respectively, again using the plain CF algorithm's performance as a yardstick.

Considering the quality of the rating predictions, the MAE and RMSE metrics improve in all cases, while the best performance is obtained when w_{SN}=60%, where the MAE drops by 3.2% and the RMSE is reduced by 2.41%. We can clearly observe that the larger the number of the *SN relations per user* in the SN dataset is, the higher the proposed algorithm's prediction accuracy is achieved.

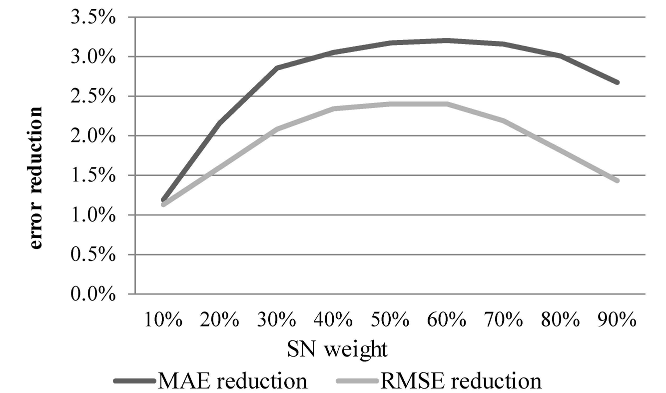

Fig. 3. MAE and RMSE reduction achieved by the proposed algorithm, under different w_{SN} values for the "Epinions" dataset.

Overall, as in the previous cases, in this dataset the proposed technique offers considerable coverage increase and, additionally, an improvement in the rating prediction quality can be obtained.

D. The "LibraryThing" Dataset

In this dataset, the coverage of the plain CF algorithm is 58.5%; while the proposed algorithm increases coverage to 60.7%, achieving a 3.8% increase in comparison to the coverage of the plain CF algorithm. Again, due to the low number of SN relations per user (equals to 1.6), in this dataset, the increase is considered adequate.

As far as the $sim(U,V)_{\text{SN}}$ value, used in formula (4), this is set to 1.0, since there are no exact values in existence in the "edges" file included in the dataset.

Fig. 4 illustrates the effect of the value of the w_{SN} parameter on the quality of recommendations produced by the proposed algorithm, as this is reflected by the MAE and the RMSE metrics, again using the plain CF algorithm's performance as a yardstick.

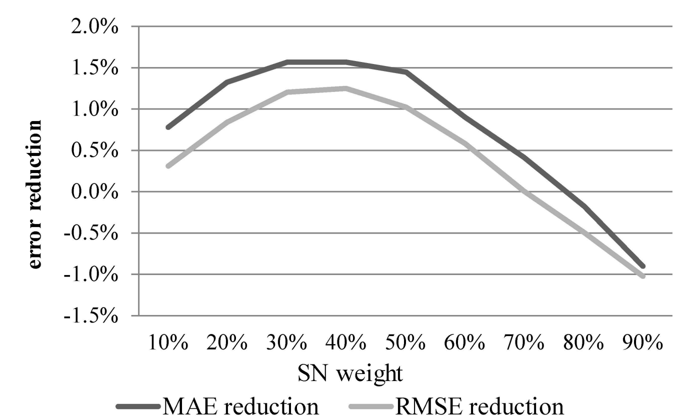

Fig. 4. MAE and RMSE reduction achieved by the proposed algorithm, under different w_{SN} values for the "LibraryThing" dataset.

Considering the quality of the rating predictions, the MAE and RMSE metrics improve when w_{SN}≤70%. The best performance is obtained when w_{SN}=40%, where the MAE drops by 1.56% and the RMSE is reduced by 1.25%. Overall, in this dataset, the proposed technique offers both coverage increase and rating prediction quality improvement, consistent with the previous datasets.

E. The "Dianping_SocialRec_2015" Dataset

In the last dataset, the coverage of the plain CF algorithm is 71.2%; while the proposed algorithm increases coverage to 75.2%, achieving a 5.6% increase in comparison to the coverage of the plain CF algorithm.

The main reason for this moderate increase, even though the large number of the dataset's *average SN relations per user*, is the high initial prediction coverage (71.2%); which leaves no room for large improvement.

As far as the $sim(U, V)_{\text{SN}}$ value, used in formula (4), this is set to 1.0, since there are no exact values in existence in the "user" file included in the dataset.

Fig. 5 illustrates the effect of the value of the w_{SN} parameter on the quality of recommendations produced by the proposed algorithm, as this is reflected by the MAE and the RMSE metrics, respectively, again using the plain CF algorithm's performance as a yardstick.

Considering the quality of the rating predictions, the MAE and RMSE metrics improve in all cases, while the best performance is obtained when w_{SN}=50%, where the MAE drops by 3.25% and the RMSE is reduced by 2.73%. Again, we can clearly observe that the larger the number of the *SN relations per user* in the SN dataset is, the higher the proposed algorithm's prediction accuracy is.

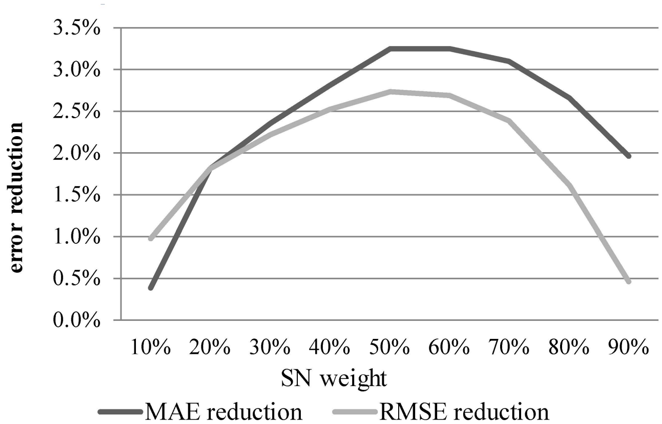

Fig. 5. MAE and RMSE reduction achieved by the proposed algorithm, under different w_{SN} values for the "Dianping_SocialRec_2015" dataset.

Overall, again, in this dataset the proposed technique offers considerable coverage increase and, additionally, an improvement in the rating prediction quality can be obtained.

F. Results Overview

In this section, we overview the results presented in the previous paragraphs.

The inclusion of the SN_NNs achieves an average coverage increase of 10.14%, ranging from 4.1%, for the "Ciao" dataset, to 32.5%, for the "Epinions" Dataset.

As far as the optimal w_{SN} parameter, achieving the largest prediction error reduction, it is observed to be in the range of 30-40%, when the number of the *Social Relations per User* parameter is relatively low, when compared to the *Ratings per User* parameter. On the other hand, when the number of the *Social Relations per User* parameter is relatively high, when compared to the *Ratings per User* parameter, the optimal w_{SN} value must be greater than the one of the w_{CF}, computed to achieve the optimal results when set in the range 50 and 60%.

The evaluation results have clearly shown that when the presented algorithm is applied to *dense* SNs datasets, where the number of *Social Relations per User* is relatively high, when compared to the CF dataset's *Ratings per User* number, the users' *social relations* are proven to be more reliable *recommenders* than the CF *Nearest Neighbours*. On the contrary, when a *sparse* SN dataset is used, where the number of *Social Relations per User* is relatively low, the users' CF *Nearest Neighbours* are proven to be more reliable *recommenders*. More specifically, in the first case (*dense* SN datasets), the algorithm achieves its best results when the weight-importance of the social relations, w_{SN}, is set to the range of 50-60%, while in the second case (*sparse* SN datasets) the algorithm achieves its best results when the weight-importance of the social relations, w_{SN}, is set to the range of 30-40%. When using these settings, the proposed algorithm provides a substantial increase in coverage, ranging from 4.1% to 32.5%, with an average of 10.14%, while at the same time offers improvements regarding rating prediction quality; the MAE decreases by 2.24% and the RMSE by 1.84% on average. Furthermore, the algorithm was found to achieve optimal results when the density of the SN is relatively high, proving that SN information is a valuable tool in RSs.

This clearly proves that, in limited information SN CF systems, when the SN dataset is "weaker" than the CF dataset, the *Nearest Neighbours*, derived by the CF dataset, are more reliable *recommenders* than the *social relations*. On the contrary, when the SN dataset is "stronger" than the CF dataset, the *social relations* are proven to be more reliable *recommenders* than the CF *Nearest Neighbours.*

With this information in hand, the proposed algorithm achieves an average MAE reduction of 2.24%, across the five datasets tested, ranging from 1.35% to 3.25%. Similarly, the average RMSE reductions range from 1.25% to 2.73% with an average of 1.84%. Furthermore, the experiments clearly indicate that the algorithm achieves its best results (in prediction coverage and accuracy) when the density of the SN is relatively high.

V. Conclusion And Future Work

In this paper, we have proposed a simple, yet effective algorithm that successfully combines limited CF information, concerning users' ratings on items, with limited SN information, concerning users' social relations, in order to improve (1) prediction accuracy and (2) prediction coverage in CF RSs, at the same time. The proposed algorithm requires no additional information, such as users' demographic information (age, gender, nationality, location, etc.), items' characteristics (price, category, reliability, etc.) or SN's contextual information (tie strength, influence, etc.). These characteristics are both useful and undemanding of elusive or atypical social information, thus the algorithm is usable for online advertising and other types of real-time applications.

The proposed algorithm has been validated through a set of experiments, aiming to quantify the obtained gains in prediction accuracy and coverage, gain insight on the effect that this combination has in the rating prediction quality and also identify the optimal weights that the SN information should be assigned.

In these experiments, five datasets containing both CF information (user-item-rating), and SN information (user-user-relation) and using two types of social relations, trust (directed) and friendship (undirected), were used to examine the behaviour of the proposed algorithm in this class of datasets. The algorithm has proven to successfully adapt to the characteristics of the dataset, yielding promising results both in the case that the SN dataset is "weaker" than the CF dataset and in the case that the SN dataset is "stronger" than the CF dataset,

By design, the proposed algorithm can be directly incorporated in a CF-based RS, when the SN information, concerning the users' relations, becomes available, since (1) it does not require any additional information about the users or the items, (2) it requires minimal additional dataset pre-processing time, computing only the partial prediction created by the SNs' *Nearest Neighbours*, (3) it takes minimal additional storage space (only the user-user-relation tuples are additionally stored), (4) it is easy to implement, through the modification of existed CF-based systems and (5) it can be easily combined with other algorithms that have been proposed for improving rating prediction accuracy and/or coverage.

Our future work will focus on creating an unsupervised version of the algorithm; we envision this algorithm to be able to automatically determine the optimal values assigned to the weight-importance of the CF's *Nearest Neighbours* (w_{CF}) and the SN's *Nearest Neighbours* (w_{SN}), based on the values of their characteristics (density, ratings per user, relations per user, etc.). Furthermore, we will investigate the computation-tuning of the $sim(U,V)_{\mathrm{SN}}$ parameter value, considering additional information derived by the *SNs* domain (such as users' tie strength, influence, common-mutual relations, contextual information, demographic data, etc.). Finally, we are planning to evaluate the aforementioned future work under additional user similarity metrics, such as Hamming Distance, Spearman Coefficient, Cosine Similarity and Euclidean Distance [2, 38, 39, 40], where those are proposed by the literature as more suitable for the additional information. The above can also be utilized in broader applications of prediction methods using social media data [41, 42, 43, 44].

REFERENCES

[1] B. Paudel, T. Haas and A. Bernstein. "Fewer flops at the top: Accuracy, diversity, and regularization in two-class collaborative filtering". Proceedings of the 11th ACM Conference on Recommender Systems, pp. 1–6. ACM, 2017.

[2] J.L. Herlocker, J.A. Konstan, L.G. Terveen and J.T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Transactions on Information Systems,vol. 22(1), pp. 5-53, 2004.

[3] D. Margaris, C. Vassilakis and P. Georgiadis, "Query personalization using social network information and collaborative filtering techniques," Future Generation of Computer Systems, vol. 78, pp. 440- 450, 2018.

[4] J. He and W.W. Chu, A social network-based recommender system (SNRS), Annals of Information Systems, vol. 12, pp. 47-74, 2010.

[5] E. Bakshy, I. Rosenn, C. Marlow and L. Adamic, "The role of social networks in information diffusion," Proceedings of the 21st international conference on World Wide Web, pp. 519-528, 2012.

[6] D. Margaris, C. Vassilakis and P. Georgiadis, "Knowledge-Based Leisure Time Recommendations in Social Networks," Current Trends on Knowledge-Based Systems: Theory and Applications, pp. 23-48, 2017.

[7] H. Li, D. Wu and N. Mamoulis, "A revisit to social network-based recommender systems," Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14), pp. 1239-1242, 2014.

[8] D. Margaris, C. Vassilakis and P. Georgiadis, "Recommendation Information Diffusion in Social Networks Considering User Influence and Semantics," Social Network Analysis and Mining, vol. 6(1), pp. 1-22, 2016.

[9] J. Scott, "Social network analysis," Sociology, Vol 22(1), pp. 109-127, 1988.

[10] S. Gong, "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering," Journal of Software, vol 5(7), pp. 745-752, 2010.

[11] A. Das, M. Datar, A. Garg and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering," Proceedings of the 16th international conference on World Wide Web, pp. 271-280, 2007.

[12] D. Margaris, P. Georgiadis and C. Vassilakis, "A Collaborative Filtering Algorithm with Clustering for Personalized Web Service Selection in Business Processes," Proceedings of the 9th IEEE International Conference on Research Challenges in Information Science (IEEE RCIS 2015), pp. 169-180, 2015.

[13] L. Li, L. Zheng, F. Yang and T. Li, "Modeling and broadening temporal user interest in personalized news recommendation," Expert Systems with Applications, vol. 41 (7), pp. 3168-3177, 2014.

[14] Y. Song, A. Elkahky and X. He, "Multi-ratedeep learning for temporal recommendation," Proceedings of the 39th International ACM Conference on Research and Development inInformation Retrieval (ACM SIGIR 2016), pp. 909-912, 2016.

[15] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," Proceedings of the 2009 ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 211–220, 2009.

[16] E. Bakshy, D. Eckles, R. Yan and I. Rosenn, "Social influence in social advertising: Evidence from field experiments," Proceedings of the 13th ACM Conference on Electronic Commerce, pp. 146–161, 2012.

[17] E.Bakshy, I. Rosenn, C. Marlow and L. Adamic, "The role of social networks in information diffusion," Proceedings of the 21st international conference on World Wide Web, pp. 519-528, 2012.

[18] F. Liu and H. Joo Lee, "Use of social network information to enhance collaborative filtering performance," Expert Systems with Applications, vol. 37(7), pp. 4772-4778, 2010.

[19] I. Konstas, V. Stathopoulos and J.M. Jose, "On social networks and collaborative recommendation," Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–202, 2009.

[20] L. Quijano-Sanchez, J.A. Recio-Garcia and B. Diaz-Agudo, "Group recommendation methods for social network environments," Proceedings of the 3rd Workshop on Recommender Systems and the Social Web, pp. 1-24, 2011.

[21] O. Arazy, N. Kumar and B. Shapira, "Improving social recommender systems," IT Professional, vol. 11(4), 2009.

[22] M. Eirinaki, M.D. Louta, I. Varlamis, "A trust-aware system for personalized user recommendations in social networks," IEEE

Transactions on Systems, Man, and Cybernetics: Systems, vol. 44(4), pp. 409-421, 2014.

[23] V. Suresh, S. Roohi, M. Eirinaki and I. Varlamis, “Using social data for personalizing review rankings,” Proceedings of the 6th Workshop on Recommender Systems and the Social Web, pp. 1-4, 2014.

[24] F. Liu and H.J. Lee, “Use of social network information to enhance collaborative filtering performance,” Expert Systems with Applications, vol. 37(7), pp. 4772-4778, 2010.

[25] G. Guo, J. Zhang and D. Thalmann, “Merging trust in collaborative filtering to alleviate data sparsity and cold start,” Knowledge-Based Systems, vol. 57, pp. 57-68, 2014.

[26] C.C. Aggarwal, “Neighborhood-Based Collaborative Filtering,” Recommender Systems: The Textbook, pp. 29-70, 2016.

[27] M. Ge, C. Delgado-Battenfeld and D. Jannach, “Beyond accuracy: Evaluating recommender systems by coverage and serendipity,” Proceedings of the 4th ACM conference on Recommender systems, pp. 257–260, 2010.

[28] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, “Large-Scale Parallel Collaborative Filtering for the Netflix Prize,” Proceedings of the 4th international conference on Algorithmic Aspects in Information and Management (AAIM '08), pp. 337 – 348, 2008.

[29] K. Yu, A. Schwaighofer, V. Tresp, X. Xu and H.P. Kriegel, “Probabilistic Memory-Based Collaborative Filtering,” IEEE Transactions on Knowledge and Data Engineering, vol. 16(1), pp. 56-69, 2004.

[30] D. Margaris and C. Vassilakis, “Enhancing User Rating Database Consistency through Pruning,” Transactions on Large-Scale Data- and Knowledge-Centered Systems, Volume XXXIV, pp. 33-64, 2017.

[31] D. Margaris and C. Vassilakis, “Improving Collaborative Filtering's Rating Prediction Accuracy by Considering Users' Rating Variability,” Proceedings of the 4th IEEE International Conference on Big Data Intelligence and Computing (IEEE DataCom 2018), pp. 1022-1027, 2018.

[32] G. Guo, J. Zhang, D. Thalmann and N. Yorke-Smith, “ETAF: An Extended Trust Antecedents Framework for Trust Prediction,” Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '14), pp. 540-547, 2014.

[33] G. Guo, J. Zhang and N. Yorke-Smith, “A Novel Bayesian Similarity Measure for Recommender Systems,” Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI '13), pp. 2619-2625, 2013.

[34] M. Simon, E. Guillot, L. Médini and F. Laforest. “RED : a Rich Epinions Dataset for Recommender Systems”, Archives-ouvertes, https://hal.archives-ouvertes.fr/hal-01010246 (accessed March 17, 2019).

[35] C. Cai, R. He, J. McAuley, “SPMC: Socially-aware personalized Markov chains for sparse sequential recommendation,” International Joint Conference on Artificial Intelligence (IJCAI '17), pp. 1476-1482, 2017.

[36] T. Zhao, J. McAuley and I. King, “Improving latent factor models via personalized feature projection for one-class recommendation,” Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 821-830, 2015.

[37] H. Li, D. Wu, W. Tang and N. Mamoulis, “Overlapping community regularization for rating prediction in social recommender systems,” Proceedings of the 9th ACM Conference on Recommender Systems, pp. 27-34, 2015.

[38] D. Margaris and C. Vassilakis, “Exploiting Internet of Things Information to Enhance Venues' Recommendation Accuracy,” Service Oriented Computing & Applications, vol. 11(4), pp. 393-409, 2017.

[39] D. Margaris and C. Vassilakis, “Improving Collaborative Filtering's Rating Prediction Quality by Considering Shifts in Rating Practices,” Proceedings of the 19th IEEE International Conference on Business Informatics, pp. 158-166, 2017.

[40] D. Margaris and C. Vassilakis, “Exploiting Rating Abstention Intervals for Addressing Concept Drift in Social Network Recommender Systems,” Informatics, vol. 5(2), Article no. 21, 2018.

[41] D. Antonakaki, D. Spiliotopoulos, C.V. Samaras, S. Ioannidis and P. Fragopoulou, "Investigating the Complete Corpus of Referendum and Elections Tweets," Proceedings of the IEEE/ACM Conference on Advances in Social Networks Analysis and Mining, pp. 100-105, 2016.

[42] G. Schefbeck, D. Spiliotopoulos and T. Risse, "The Recent Challenge in Web Archiving: Archiving the Social Web," Proceedings of the International Council on Archives Congress, pp. 20-24, 2012 .

[43] D. Spiliotopoulos, E. Tzoannos, P. Stavropoulou, G. Kouroupetroglou and A. Pino, “Designing User Interfaces for Social Media driven Digital Preservation and Information Retrieval,” Proceedings of the 13th International Conference on Computers Helping People with Special Needs, pp. 581-584, 2012.

[44] D. Antonakaki, D. Spiliotopoulos, C.V Samaras, P. Pratikakis, S. Ioannidis and P. Fragopoulou, "Social Media Analysis during Political Turbulence," PLoS ONE, vol. 12(10), pp. 1-23, 2017.