RSGAN-SPH: Incorporating Social Proximity with Reliable Friends Generation via Adversarial Training to Improve Social Recommendation

https://github.com/J-morag/RSGAN-SPH

Based on Generating Reliable Friends via Adversarial Training to Improve Social Recommendation (Yu et al. 2019)

Jonathan Morag, Shachar Wild, Bracha Shapira
Ben-Gurion University of the Negev
{moragj, wildsha}@post.bgu.ac.il, {bracha.shapira}@post.bgu.ac.il

Abstract

In the research area of social recommendation systems, most of the recent studies assume that people share similar preferences with their friends and that the online social relations are helpful in improving traditional recommendation systems. However, the online social networks are quite sparse and a majority of users only have a small number of friends, which makes this assumption often unreliable. Moreover, explicit friends may not share similar interests. Therefore, discovering a number of reliable friends for each user plays an important role in advancing social recommendation. Recently, researchers have proposed a method which focuses on extracting valuable explicit social links, while identifying reliable friends, using both implicit and explicit relations. However, this method only considers the users' common rating data in order to do so, instead of taking into account the social proximity as well, which may limit its performance. In this work, we propose RSGAN-SP and RSGAN-SPH, the integration of the state-of-the-art RSGAN model and social proximity, in the aim to to improve its performances by allowing it use different methods for generating reliable friends. RSGAN-SP considers only the social proximity when identifying reliable friends, and RSGAN-SPH considers both rating data and social proximity to do so, while integrating it using a scheduling algorithm. Moreover, since our study suggests making advantage of another type of information, besides rating data, it further overcomes the sparsity and unreliability problems of explicit social relations, while also attempting to improve the social recommendation performance. We conduct comprehensive experiments on two real-world social media datasets to test our proposed methods, show their merits and verify the positive effects of the generated reliable friends when considering the social proximity between users.

1 Introduction

In today's digital world, the wide use of social media leads to the creation of abundance data across the Internet, such as social feeds and various news. As a result, many users may find it difficult to find relevant information for their specific tasks. To tackle this problem, recommendation systems can identify users' preferences by modeling the historical data of users, such as explicit ratings and implicit feedback. In that manner, users can easily find the content they are actually interested in, instead of dealing with data overload. Nevertheless, most users tend to consume only a small fraction of items among the full collection. Therefore, traditional recommendation systems often suffer from the common problem of data sparsity and thus fail to satisfy the users.

To tackle this thorny problem, a possible solution suggests transferring knowledge from other fields or platforms and incorporate them into traditional recommendation systems. This solution mostly relies on the assumption that people are often influenced by their friends when it comes to decision making (Chevalier and Mayzlin 2006), which makes the social relations very useful for various recommendation tasks.

Following this, several works explored the use of social networks to recommendation systems, by developing various social recommendation methods. TrustWalker (Jamali and Ester 2009) proposed a random walk model for combining trust-based and item-based recommendation, which attempts to overcome the cold start users problem that collaborative filtering (Su and Khoshgoftaar 2009) often suffer from. Another work (Ma et al. 2011) provided general method which makes use of matrix factorization framework with social regularization, that can be extended to incorporate other contextual information, such as social tags.

However, in many occasions, social recommendation is not as successful as one can expect, and sometimes the integration of social connections may harm the recommendation performance (Tang, Hu, and Liu 2013). This issue occurs due to social connections being diverse and also unreliable in some occasions. Many social recommendation systems directly use the explicit social ties when making a recommendation, only based on the principle of homophily (McPherson, Smith-Lovin, and Cook 2001). In other words, they Naïvely rely on the assumption that connected users share similar preferences. This assumption may not be accurate enough, as the online community is quite different from the offline community in terms of the scale and the possibilities in generating connections. As such, in the online social networks, not all of social relations have a positive impact on quality-improving for recommendation, as most of them may hardly reach a consensus with each other in all the aspects of user preferences. Moreover, since the social networks are with open nature, social media users are sometimes mixed with a number of malicious accounts, which may pose a threat to accurate preference inference (Ren et al. 2015). Following this, using explicit social connections directly may lead to unsatisfying results.

Due to the unreliability of using explicit social connections in recommendation systems, several studies attempted to only focus on the informative, relevant relations from explicit social networks, by filtering the noisy ones: finding a wise group of experts in social networks(Yin, Cui, and Huang 2011), learning personalized preference of strong and weak ties for social recommendation (Wang et al. 2017b), and constructing social recommender based on factorization and distance metric learning (Yu et al. 2017).

Nevertheless, social relations still suffer from being almost as sparse as the user feedback and filtering explicit relations further decrease the available data. Therefore, the studies mentioned above do not make a significant contribution in improving recommendation quality. To tackle this problem, another alternative suggested identifying implicit friends across the social network, with whom a user can have similar tastes, while not having explicit social links with them. The first study (Ma 2013) proposed to use rating profiles to search for implicit friends when social networks are not available, while other work (Yu et al. 2018) adopted network embedding techniques to uncover implicit friends for each user.Although these studies managed to overcome the sparsity problem of explicit social relations, neither of them has an adaptive evaluation mechanism to assess the quality of identified implicit friends. The recommendation model receives these extracted social links based on the simple assumption that they are much more dependable in reflecting users' preferences, rather than the original explicit friends, even if the searching policy may be independent from the recommendation process. Moreover, Many of the social recommendation methods treat the training phase as a static process and do not consider the change of similarity or proximity between two users during training. After several optimization iterations, some relations will not be informative enough for the model. It may even become noises due to the complexity of the models and by utilizing multiple tuning parameters.

To overcome this issue, another approach suggested to not only uncover reliable friends but to also dynamically assess these relationships. RSGAN (Reliable Social recommendation framework) (Yu et al. 2019) proposed a social recommendation framework for Top-N recommendation which focuses on reliable friends identification with a dynamic evaluation mechanism. As the observed social networks are generally very sparse, they identified a few highly reliable friends as seeded friends from both the observed and unobserved social networks, who are quite likely to boost the recommendation performance. Nevertheless, the fact that it does not utilize enough information, besides ranking data, may limit its social recommendation performance.

Many social recommendation systems suffer from the same disadvantage: Data sparsity due to users experiencing and assessing only small number of items, which may limit the recommendation performance. In our work, we propose the integration of social proximity concept (assessment of common number of friends) into the state-of-the-art RSGAN model in order to improve its performances by allowing it to investigate utilizing more information when identifying reliable friends, which may further assist in overcoming the sparsity problem of social data. Social proximity represents social closeness in social network, which can be inferred from their local network structure. We explore our proposed improvement in a similar way to RSGAN. We combine GAN, social proximity and our approach as social recommendation framework for the RSGAN architecture to improve its results further. This, to the best of our knowledge, has not been tested yet.

The remainder of this paper is organized as follows: In Section 2, we provide a brief overview of related work (definitions and background); in Section 3, we describe the proposed method that was used in this work; in Section 4, we describe the datasets and evaluation methods used in our experiments; in Section 5, we present the results that we obtained from our experiments; in Section 6, we provide discussion that arises from our results and finally, in Section 7, we give our conclusions from this study, and we offer future research directions.

2 Related Work

In this section, we briefly review related work on three aspects: social recommendation, adversarial learning in recommender systems and social proximity. Next, we will provide an overview of a framework which integrates the first two fields, called RSGAN, which our work is based on. Our proposed methods incorporate this framework, along with all three fields.

2.1 Social Recommendation

Previous studies that dealt with social recommendation mostly discussed how the explicit social relations can be used to improve recommendation performance. These studies simply assume that explicitly connected users are supposed to share similar preferences due to the principle of homophily (McPherson, Smith-Lovin, and Cook 2001). Examples of these studies are TrustMF (Yang et al. 2016) and SoRec (Ma et al. 2008), which cofactorize the rating matrix and the relation matrix by sharing a common latent space in which purchase and social information meet with. STE (Ma, King, and Lyu 2009) proposed an ensemble that considers user's essential preference as the linear combination of its own explicit preference and those of its friends. Another study which used fusing strategy, social regularization (Ma et al. 2011) managed to minimize the gap between the personal taste of a user and the average taste of its friends. It did so by utilizing weighted social regularized terms.

Other studies proposed a few models that explain the social influence on feedback of users from various perspectives (Chen et al. 2018), (Xiao et al. 2017). They noticed that user exposure to certain items has a influences recommendation, and then social connections are utilized to help identify users' exposure to items, rather than considering preferences, which may be less restrictive. Furthermore, another study has shown that social information may also used to model the order of items to be recommended (Zhao, McAuley, and King 2014). The writers developed a social Bayesian personalized ranking method, which gives higher ranks to items that their friends prefer.

On the other hand, many following studies concentrated on discovering valuable information from social relations while also identifying reliable social connections. One of these studies (Wang et al. 2017b) utilized the EM algorithm in the aim to differentiate strong ties and weak ties from overall social ties. Another study (Liu et al. 2018) proposed a new concept, essential preference space. This was done to describe the multiple preferences of particular users in social recommender systems. Moreover, based on the intuition that social tie inherently has various facets indicating multiple trust relationships among users, another study (Tang, Gao, and Liu 2012) suggested to distinguish between multi-faceted trust in search of experts of different types.

These studies demonstrate that utilizing only explicit social relations for social recommendation is not efficient enough. Thefore, considering implicit relations between users is needed in order to enhance the quality of the recommendations. In our work, we do learn both implicit and explicit interactions between users, and by using RSGAN-SP and RSGAN-SPH, we create meaningful social ties, which also further help to overcome the sparsity problem in recommendation systems.

2.2 Adversarial Training in Recommender Systems

Another field which our work deals with is called adversarial learning (Goodfellow et al. 2014), which achieved positive impact in several areas, such as: computer vision and natural language processing. This type of learning consists of two components: the generator and the discriminator. The generator aims to imitate the real data distribution, while the discriminator attempts to differentiate fake examples from the real data. This field is generic and thus can be applied to almost any programming problem. As such, several past works have explored it in the recommender systems area.

One of the first works in this field is IRGAN (Wang et al. 2017a), which is an influential IR model. In addition to that, it is GAN-based. As such, it consisted of two models: the generative model and the discriminator model. The generator selected the informative negative samples, while the discriminator was used to classify which samples are negative and which are positive. RSGAN, the algorithm which our proposed methods are based on, follows similiar actions, only that it deals with social recommendations, rather than IR.

Following IRGAN, another model, called GraphGAN (Wang et al. 2018) was introduced. It suggested using an alternative form of softmax, which is called graph softmax. It was used to overcome the limitations of traditional softmax function, which can be proven satisfying desirable properties of normalization, graph structure awareness, and computational efficiency. This softmax's form was used to to accelerate the training phase, which as result can improve the overall computing efficiency. While this work managed to reduce the training time phase, it is not suitable for our problem (generating reliable friends), since its not graph-based. However, it lead to the use of Gumbel-Softmax (Jang, Gu, and Poole 2017) in RSGAN, which our work is based on. Gumbel-Softmax can approximate categorical samples by adding noises sampled from Gumbel(0 , 1) to the probability vector produced by the generator, which is known as the reparameterization trick. Same as RSGAN, our suggested methods use it to sample the item consumed by the friend. It allows penalizing an item only has a limited negative impact on the chances of the friend's other items to be chosen.

While the above proposed models dealt with item sampling, other work, CFGAN (Chae et al. 2018) attempted to directly learn user profiles with GAN, rather than sampling items to advance the recommendation model. It is a generic collaborative filtering framework based on Generative Adversarial Networks, which was used to improve recommendations' accuracy. This method is similar to our in the sense that it attempts to learn the user profiles. However, our suggested methods utlizing social proximity and ratings' latent factors, rather than collaborative filtering.

All reviewed works demonstrate that utilizing adversarial training in recommender systems may further assist in making efficient recommendations for users and as a result improve overall performance. In our work, although we only deal with social recommendations and do not perform the exact same procedures, we do learn the advantages of GAN and utilize it in our proposed methods, same as RSGAN, which they are based on. By using the generator, we are able to generate recommended generated item for a particular user, while the discriminator receives it and outputs a ranked item list as recommendations with user feedback and the generated item as input.

Reliable Social Generative Adversarial Network

Next, we will describe the Reliable Social Generative Adversarial Network (RSGAN) (Yu et al. 2019) model in further detail since it is used as the base for our work. RSGAN architecture, which can be seen in 1, follows the foundations of the Generative Adversarial Network and is based on social recommendation field.

RSGAN has two major components: the generator and the discriminator. The discriminator takes charge of ranking the candidate items and producing a recommendation list for the user while the generator is responsible for generating a reliable friend and then sampling an item consumed by this friend which is also likely to be consumed by the current user to enhance the discriminator. The discriminator punishes the generated friends if items consumed by them are not helpful for advancing the discriminator, and returns gradients to the generator in order to reduce the probabilities of generating such friends. In other words, the generator is used to produce friends that can possibly enhance the social recommendation model, and a discriminator is responsible for assessing these generated friends and ranking the items according to both the current user and his friends' preferences.

In order for the generator to predict friends with good quality, the writers identified a few of seeded friends as the ground truth, who are highly probable to be helpful in improving the social recommendation performance. They

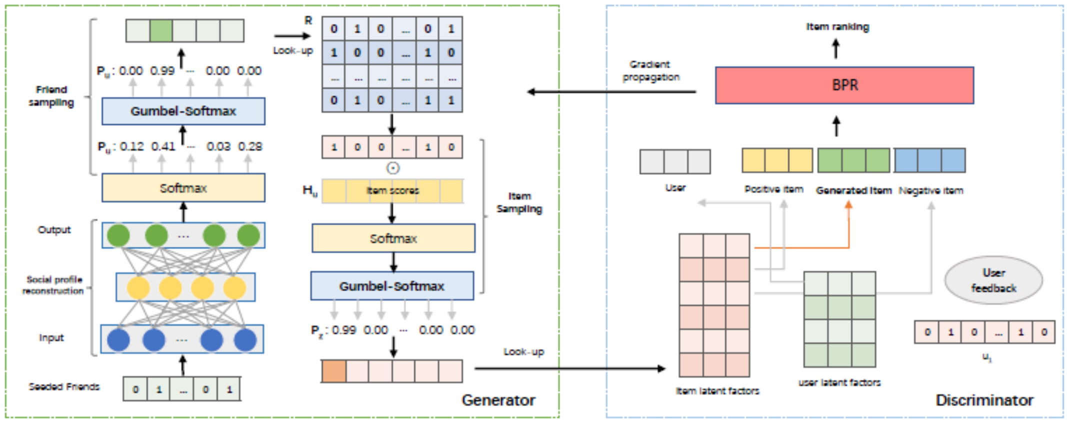

Figure 1: An overview of RSGAN, as shown in (Yu et al. 2019). Our suggested algorithms enhance the method for generating the seeded friends.

identified them by calculating Cosine Similarity (Dehak et al. 2010) for each given users pair, using latent factor of their items ratings. Afterwards, they encoded them into vector which had a length equal to the number of users in the dataset, with binary values. Value of 1 indicates that a friend seeded, where 0 indicates this is not a close friend. As mentioned earlier, this vector was used as the initial input for RSGAN.

RSGAN's main advantages are integration of Social Recommendation with Adversarial Training, whereas previous work focused on one or the other. It also managed to identify reliable friends in both the observed and unobserved social networks and overcame the sparsity and unreliability problems of explicit social relations. However, it is not without disadvantages. Their current solution for identifying seeded friends relies on Cosine Similarity between user ratings latent factors. Not only that this similarity only reflects the angle between vectors, it only takes into account the ratings data of a user, without considering its social proximity with other users in the network. By doing so, they may dismiss crucial type of information that may assist them in identifying better seeded friends. Moreover, using another type of information, besides ratings, may further assist in overcoming the sparsity and unreliability problems of explicit social relations and also improve the overall social recommendation performance. We will now present several studies which dealt with the social proximity field, to further discuss its advantages.

2.3 Social Proximity

Social Proximity is defined as social closeness in social network. The strength of a social tie, which specifies the intensity and the depth of interaction between two people, can be estimated from their local network structure (Granovetter 1973). A review of the most recent research has found that a direct connection to others cannot sufficiently explain the influence or diffusion of behavior in a network (Montgomery et al. 2020). Past studies have shown that people who are "close" in some sense in a social network are more likely to perform similar actions than more distant people. In other words, the degree of a social proximity of two users is related to the similarity of their activities.

lerman et al. (Lerman et al. 2012) has shown that given friends' activity, knowing their proximity to the user can help better predict which URLs the user will forward. They represented a network by a directed, unweighted graph G = (V, E) with V nodes and E edges, where a node represents a user and edge connecting node u and v indicates these two users being friends online. The greater the number of paths connecting u and v, the more likely they are to share information, and the closer they are considered to be in the network. They used the following proximity metrics: number of common neighbors (CN), fraction of common neighbors, or Jaccard (JC) coefficient, and the Adamic-Adar (AA) score, which weighs each common neighbor by the inverse of the logarithm of its degree. The writers found that taking into account friends' proximity to the user can improve prediction, and that most gain is achieved by the attention-limited metrics. They concluded that proximity also has predictive power. People who are close to each other in a social network are more likely to act in a similar way. This paper lead us to utilize Jaccard Similarity as a proximity metric between two given users in our research, as it was shown to efficiently assist in the prediction of social activity. Highest value of this metric indicates a stronger proximity.

Another work that utilized social proximity dealt with its exposure in a social network as a mechanism driving peer influence of adolescent smoking (Khalil, Jones, and Fujimoto 2021). It conceptualized individuals' proximity exposure as an individual's level of exposure to all smokers in a network. This exposure was determined based on the shortest distance between the individual and the smokers in the network. For that purpose, they constructed path models (Granovetter 1973) for predicting smoking behavior. This study has shown that proximity exposure can predict smoking even among nonsmokers without direct ties to friends who smoke. Since social proximity can indeed be used for predicting similar smoking behaviour, it may also generalizes for other type of behaviours, such as: consumption behaviour, and thus assist us in our research. We will now propose our methods to diffuse RSGAN's issues, while taking advantage of social proximity.

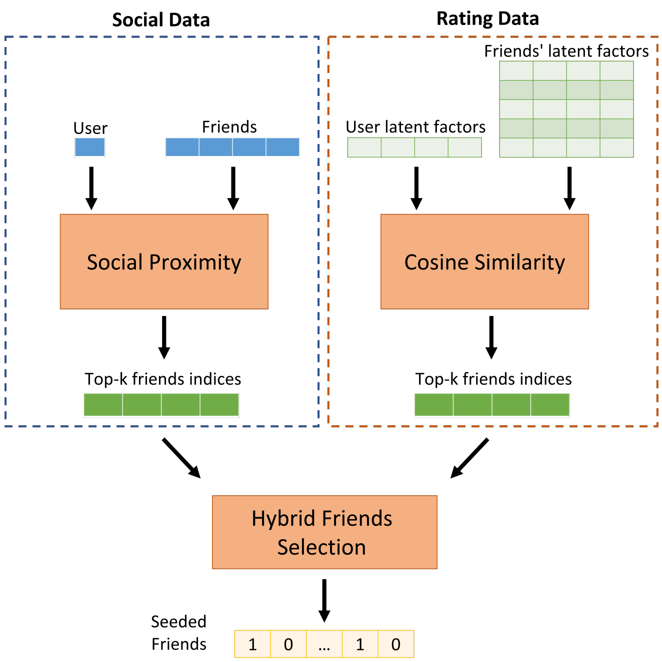

Figure 2: RSGAN-SPH's Overall Pipeline for Identifying Seeded Friends.

3 Method

We propose the integration of proximity in social networks into the RSGAN architecture, in the form of two models, *RSGAN with Social Proximity (RSGAN-SP)* and *RSGAN with Social Proximity Hybrid (RSGAN-SPH)*. Our work tries to overcome a common disadvantage in social recommend systems and in RSGAN specifically, which is the sparsity problem, as well as not utilizing enough information for assessing closeness of users.

We use the *Jaccard Smilarity* measure for estimating the proximity between two users u_1 and u_2 in a social network, given *friends*(u_i) as the set of known friends for user u_i:

$$\mathit{JaccardSim}(u_1, u_2) = \frac{\mathit{friends}(u_1) \cap \mathit{friends}(u_2)}{\mathit{friends}(u_1) \cup \mathit{friends}(u_2)}$$

This measure is inexpensive to compute given two users, as only their immediate neighbors in the social network have to be considered. Therefore, it can be computed on-the-fly when necessary, without the need for any expensive preprocessing or bookkeeping.

We integrate *JaccardSim* into the process for selecting seeded friends as input for RSGAN's generator component. RSGAN chooses seeded friends according to the cosine similarity between latent factor vectors generated from the ratings data of the users. For RSGAN-SP, we replace this measure with *JaccardSim*, choosing the k-best (highest similarity) friends.

RSGAN-SP adds a social factor that is not present in RSGAN, but it also removes a ratings factor that is present in RSGAN. To mitigate this disadvantage, we also introduce RSGAN-SPH. RSGAN-SPH incorporates both rating data and social data into the seeded friends selection process. First, two sorted lists of k-best friends are selected using *JaccardSim* for one, and ratings data for the other. We then select a final list of k seeded friends by choosing friends one by one, alternating between the two lists, each time taking the best candidate from each list that has not yet been chosen for the final list. Duplicates are handled by choosing the next most promising friend from the list. This process continues iteratively until k seeded friends are chosen. This process is described in detail in algorithm 1. The overall pipeline of RSGAN-SP for identifying a user's seeded friends can be seen in figure 2.

Algorithm 1: Hybrid Friends Selection

```
l1 ← top k by ratings;
l2 ← top k by social;
res ← [];
curr list ← l1;
while len(res) < k do
    if len(curr list) > 0 then
        candidate ← pop best from curr list;
        while candidate in res and len(curr list) > 0
          do
            candidate ← pop best from curr list;
        if candidate not in res then
            append candidate to res;
    curr list ← l2 if curr list is l1 else l1;
return res;
```

4 Evaluation

In this section, we describe the experimental plan we used to evaluate our models.

4.1 Experiment Setup

Method Implementation

We implemented RSGAN-SP and RSGAN-SPH, along with its components, using python 2.7 and TensorFlow 1.15. Our implementation follows the models' structure of RSGAN. Our implementation and code for experiments are publicly available.

DataSets

We based our experiments on three publicly available social recommendation datasets: LastFM, FilmTrust, and Epinions. Because Epinions contains many unique users and items, and RSGAN produces larger models given more unique users and items, we could not experiment on the full Epinions dataset using the hardware at our disposal. Therefore, we used a random sample of one-tenth of the users and one-tenth of the items in the Epinions dataset. We also created a version of the LastFM that is biased towards social relation data, by using a random sample of 25% of the ratings data while preserving all social relation data. We used this additional dataset to test our models in a setting where we hypothesized they would be most useful, that is, one where social relation data is relatively dense and ratings data is relatively sparse. Table 1 shows for each dataset we used, the number of unique users, the number of unique items, the number of ratings data, the number of relations data, and the ratio between the number of relations data and the number of ratings data.

Dataset	#Users	#Items	#Ratings	#Relations	%Relations
LastFM	1,892	17,632	92,834	25,434	0.27
LastFM-biased	1,892	17,632	23,209	25,434	1.10
Epinions(1/10)	1,816	3,732	78,877	168,254	2.13
FilmTrust	1,642	2,071	35,497	1,853	0.05

Table 1: Comparison of datasets. Relations % is the ratio between the number of relation data and the number of ratings data.

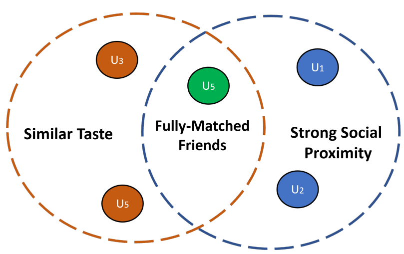

Figure 3: Fully-matched (seeded) friends, as identified by our proposed method, RSGAN-SP.

Baseline Methods

We compare our two methods, RSGAN-SP and RSGAN-SPH, to RSGAN framework. We use the setting as described in the the RSGAN paper (Yu et al. 2019). Since we want to test our different components and compare them to RSGAN components, the same hyper-parameters from the original paper are used, in the aim to examine the advantage of utilizing social proximity in identifying fully-matched friends.

Evaluation Methods and Metrics

We conducted all our our experiments using 3-fold cross validation (Friedman et al. 2001). Out of the data selected for training for each fold, 10% was reserved for validation. All results are shown as the average of the test results of each fold. We compare the performance of the models on the following metrics: Precision@10, Recall@10, and *Normalized Discounted Cumulative Gain* (nDCG) for first 10 results (Wang et al. 2013).

$$Precision = \frac{T_p}{T_p + f_p}$$

$$Recall = \frac{T_p}{T_p + f_n}$$

$$DCG = \sum_{i=1}^{p} \frac{rel_i}{\log_2(i+1)}$$

Where rel_i represents the relevance value of a result, and p represents the rank position of the result.

$$IDCG = \sum_{i=1}^{|REL_p|} \frac{rel_i}{\log_2(i+1)}$$

Where $|REL_p|$ represents the list of all relevant results.

$$nDCG = \frac{DCG_p}{IDCG_p}$$

All models were trained and tested using the same parameters. Regularization coefficient: 0.001, batch size: 512, latent factors: 50, generator sigmoid units: 200, Gumbel-Softmax (Jang, Gu, and Poole 2016) temperature: 0.2, initial learning rate: 0.001 . All models were trained for 100 epochs.

5 Results

In this section, we will present our experimental results.

5.1 standard Datasets

We first experimented using the three standard datasets presented in section 3, LastFM, Epinions, and FilmTrust. This experiment is meant to compare our algorithms with RSGAN on common datasets.

The results of this experiment are shown in the first 9 rows (3 first datasets) of table 2. We can see that RSGAN performed better than RSGAN-SP and RSGAN-SPH according to all metrics we used, on both LastFM and Epinions, while RSGAN-SPH outperformed the other algorithms on FilmTrust. RSGAN outperforms RSGAN-SP on LastFM and Epinions, while on FilmTrust there is no clear winner between the two. On all datasets, RSGAN-SPH performed better than RSGAN-SP. We assume the relative strength of the algorithms is affected by how dense or informative the social data is in each dataset. Thus, it is possible that the social relations data in FilmTrust is more informative than LastFM and Epinions. It is possible that this is due to social relations in FilmTrust being explicit trust statements given by users, and thus they are often informative for inferring the user's tastes.

5.2 Social-biased Dataset

We wanted to check our hypothesis regarding the results of the previous experiment, by artificially increasing the relative density of social data in one of the datasets, so that it may be compared directly with the original version. For this reason we also experimented with LastFM-biased (see section 4), where social data is more dense relative to ratings data when compared with the original LastFM dataset. The ratio between the number of relations data and the number of ratings data is 0.27 in LastFM and 1.10 in LastFM-biased.

We see that with LastFM-biased, RSGAN-SP outperforms RSGAN on both precision@10 and nDCG, only losing on recall@10, whereas it lost on all three metrics with LastFM. RSGAN-SPH outperforms both other algorithms on LastFM-biased on all metrics. This leads us to conclude

Dataset	Algorithm	Precision@10	Recall@10	nDCG
LastFM	RSGAN	**0.1063**	**0.0644**	**0.1140**
	RSGAN-SP	0.1049	0.0634	0.1129
	RSGAN-SPH	0.1049	0.0636	0.1138
Epinions(1/10)	RSGAN	**0.0500**	**0.0788**	**0.0779**
	RSGAN-SP	0.0492	0.0775	0.0755
	RSGAN-SPH	0.0498	0.0781	0.0766
FilmTrust	RSGAN	0.4026	0.5112	0.5432
	RSGAN-SP	0.4026	0.5106	0.5437
	RSGAN-SPH	**0.4047**	**0.5161**	**0.5442**
LastFM-biased	RSGAN	0.0134	0.0329	0.0236
	RSGAN-SP	0.0141	0.0326	0.0239
	RSGAN-SPH	**0.0145**	**0.0339**	**0.0247**

Table 2: Results of our experiments on datasets LastFM, Epinions, and FilmTrust, and LastFM-biased, using RSGAN, RSGAN-SP, and RSGAN-SPH.

that the more dense social data in a dataset compared to ratings data, the better our algorithms will perform relative to RSGAN. Even though the ratings data is sparse in LastFM-biased, RSGAN-SPH is able to use that data (in addition to relations data) to outperform both RSGAN and RSGAN-SP, demonstrating the value of considering both social and ratings data when selecting seeded friends.

6 Discussion

Our experiments showed that RSGAN-SPH always outperformed at least one other algorithm, likely because it uses the most informative friends based on both ratings data and social data, and is therefore reasonably well adapted to both datasets where social data is more informative and datasets where ratings data is more informative.

Our results show that RSGAN-SPH can outperform RSGAN, however, its performance depends on properties of each dataset. We think datasets that have more dense or informative social relation data will likely favor RSGAN-SPH.

RSGAN-SPH may be useful for real world application where social data is plentiful but ratings data is sparse. For instance, if a social networking website wishes to introduce a merchandise store, it may have a large amount of social relations data, but very few interactions with the store upon which recommendations may be based.

7 Conclusion and Future Work

In this work, we presented RSGAN-SP and RSGAN-SPH, the integration of social proximity, to the state-of-the-art RSGAN model. We explored the use of both rating and socail data for identifying reliable friends. Both RSGAN-SP and RSGAN-SPH allow the RSGAN model, which only considers rating data, to utilize social proximity to identify reliable friends in social networks. While our first method, RSGAN-SP, suggests using only social data, RSGAN-SPH makes use of both rating and social data, by combining it using scheduling algorithm. To the best of our knowledge, this was not done before. Making use of another type of information allows RSGAN to further overcome the data sparsity problem, which traditional social recommendation systems often suffer from, and thus improve the social recommendation performance. Although RSGAN-SP did not show improvement in results, RSGAN-SPH did manage to improve the social recommendation for some of our examined social network datasets, where both rating and social data are informative enough. This implies that social proximity can be useful for identifying reliable friends, when incorporated with users' rating.

Future work could explore more methods for assessing social proximity among users, as well as different algorithms for integrating it with users' rating data. Moreover, it may be possible further explore the use of social proximity in identifying reliable friends, by examining more ranking metrics.

References

Chae, D.-K.; Kang, J.-S.; Kim, S.-W.; and Lee, J.-T. 2018. Cfgan: A generic collaborative filtering framework based on generative adversarial networks. In *Proceedings of the 27th ACM international conference on information and knowledge management*, 137–146.

Chen, J.; Feng, Y.; Ester, M.; Zhou, S.; Chen, C.; and Wang, C. 2018. Modeling users' exposure with social knowledge influence and consumption influence for recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 953–962.

Chevalier, J. A., and Mayzlin, D. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research* 43(3):345–354.

Dehak, N.; Dehak, R.; Glass, J. R.; Reynolds, D. A.; Kenny, P.; et al. 2010. Cosine similarity scoring without score normalization techniques. In *Odyssey*, 15.

Friedman, J.; Hastie, T.; Tibshirani, R.; et al. 2001. *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27.

Granovetter, M. S. 1973. The strength of weak ties. *American journal of sociology* 78(6):1360–1380.

Jamali, M., and Ester, M. 2009. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 397–406.

Jang, E.; Gu, S.; and Poole, B. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*.

Jang, E.; Gu, S.; and Poole, B. 2017. Categorical reparametrization with gumble-softmax. In *International Conference on Learning Representations (ICLR 2017)*. OpenReview. net.

Khalil, G. E.; Jones, E. C.; and Fujimoto, K. 2021. Examining proximity exposure in a social network as a mechanism driving peer influence of adolescent smoking. *Addictive Behaviors* 117:106853.

Lerman, K.; Intagorn, S.; Kang, J.-H.; and Ghosh, R. 2012. Using proximity to predict activity in social networks. In *Proceedings of the 21st international conference on World Wide Web*, 555–556.

Liu, C.-Y.; Zhou, C.; Wu, J.; Hu, Y.; and Guo, L. 2018. Social recommendation with an essential preference space. In *Thirty-second AAAI conference on artificial intelligence*.

Ma, H.; Yang, H.; Lyu, M. R.; and King, I. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, 931–940.

Ma, H.; Zhou, D.; Liu, C.; Lyu, M. R.; and King, I. 2011. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 287–296.

Ma, H.; King, I.; and Lyu, M. R. 2009. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 203–210.

Ma, H. 2013. An experimental study on implicit social recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 73–82.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27(1):415–444.

Montgomery, S. C.; Donnelly, M.; Bhatnagar, P.; Carlin, A.; Kee, F.; and Hunter, R. F. 2020. Peer social network processes and adolescent health behaviors: A systematic review. *Preventive medicine* 130:105900.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28:91–99.

Su, X., and Khoshgoftaar, T. M. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009.

Tang, J.; Gao, H.; and Liu, H. 2012. mtrust: Discerning multi-faceted trust in a connected world. In *Proceedings of the fifth ACM international conference on Web search and data mining*, 93–102.

Tang, J.; Hu, X.; and Liu, H. 2013. Social recommendation: a review. *Social Network Analysis and Mining* 3(4):1113–1133.

Wang, Y.; Wang, L.; Li, Y.; He, D.; Chen, W.; and Liu, T.-Y. 2013. A theoretical analysis of ndcg ranking measures. In *Proceedings of the 26th annual conference on learning theory (COLT 2013)*, volume 8, 6. Citeseer.

Wang, J.; Yu, L.; Zhang, W.; Gong, Y.; Xu, Y.; Wang, B.; Zhang, P.; and Zhang, D. 2017a. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, 515–524.

Wang, X.; Hoi, S. C.; Ester, M.; Bu, J.; and Chen, C. 2017b. Learning personalized preference of strong and weak ties for social recommendation. In *Proceedings of the 26th International Conference on World Wide Web*, 1601–1610.

Wang, H.; Wang, J.; Wang, J.; Zhao, M.; Zhang, W.; Zhang, F.; Xie, X.; and Guo, M. 2018. Graphgan: Graph representation learning with generative adversarial nets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Xiao, L.; Min, Z.; Yongfeng, Z.; Yiqun, L.; and Shaoping, M. 2017. Learning and transferring social and item visibilities for personalized recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 337–346.

Yang, B.; Lei, Y.; Liu, J.; and Li, W. 2016. Social collaborative filtering by trust. *IEEE transactions on pattern analysis and machine intelligence* 39(8):1633–1647.

Yin, H.; Cui, B.; and Huang, Y. 2011. Finding a wise group of experts in social networks. In *International Conference on Advanced Data Mining and Applications*, 381–394. Springer.

Yu, J.; Gao, M.; Rong, W.; Song, Y.; and Xiong, Q. 2017. A social recommender based on factorization and distance metric learning. *IEEE Access* 5:21557–21566.

Yu, J.; Gao, M.; Li, J.; Yin, H.; and Liu, H. 2018. Adaptive implicit friends identification over heterogeneous network for social recommendation. In *Proceedings of the 27th ACM international conference on information and knowledge management*, 357–366.

Yu, J.; Gao, M.; Yin, H.; Li, J.; Gao, C.; and Wang, Q. 2019. Generating reliable friends via adversarial training to improve social recommendation. In *2019 IEEE International Conference on Data Mining (ICDM)*, 768–777. IEEE.

Zhao, T.; McAuley, J.; and King, I. 2014. Leveraging social connections to improve personalized ranking for collaborative filtering. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, 261–270.