

# Exploiting weak ties in trust-based recommender systems using regular equivalence

Tomislav Duricic  
Graz University of Technology  
Graz, Austria  
tduricic@tugraz.at

Dominik Kowald  
Know-Center GmbH  
Graz, Austria  
dkowald@know-center.at

Emanuel Lacic  
Know-Center GmbH  
Graz, Austria  
elacic@know-center.at

Elisabeth Lex  
Graz University of Technology  
Graz, Austria  
elisabeth.lex@tugraz.at

## ABSTRACT

User-based Collaborative Filtering (CF) is one of the most popular approaches to create recommender systems. CF, however, suffers from data sparsity and the cold-start problem since users often rate only a small fraction of available items. One solution is to incorporate additional information into the recommendation process such as explicit trust scores that are assigned by users to others or implicit trust relationships that result from social connections between users. Such relationships typically form a very sparse trust network, which can be utilized to generate recommendations for users based on people they trust. In our work, we explore the use of regular equivalence applied to a trust network to generate a similarity matrix that is used for selecting  $k$ -nearest neighbors used for item recommendation. Two vertices in a network are regularly equivalent if their neighbors are themselves equivalent and by using the iterative approach of calculating regular equivalence, we can study the impact of strong and weak ties on item recommendation. We evaluate our approach on cold-start users on a dataset crawled from Epinions and find that by using weak ties in addition to strong ties, we can improve the performance of a trust-based recommender in terms of recommendation accuracy.

**Problem & objective.** Ever since their introduction, user-based Collaborative Filtering (CF) approaches have been one of the most widely adopted and studied algorithms in the recommender systems literature [9]. CF is based on the intuition that those users, who have shown similar item rating behavior in the past, will likely give similar ratings to items in the future. The basis of CF is to retrieve the  $k$ -nearest neighbors of a target user for whom the recommendations are generated and to recommend items from these  $k$  neighbors, which were rated highly by them but have not yet been rated by the target user. An issue of CF is the cold-start user problem, i.e., novel users, who have rated zero or only a small number of items [10] and whose ratings cannot thus be exploited to find similar users.

As a remedy, trust-based CF methods exploit trust statements expressed by users on platforms such as Epinions [7]. Such trust statements can be explicit, i.e., users assign trust scores to others or implicit, i.e., users engage in social connections with others

they trust. Based on explicit and implicit trust statements, we can generate trust networks and recommend items for users based on people they trust [6]. Since trust networks are often sparse, a particular property of trust, namely transitivity, can be exploited to propagate trust in the network by forming weak ties between users. In this way, new connections are established between users, who do not share a direct link, but are weakly connected via intermediary users [2, 7].

In our work, we focus on the first step of CF, i.e., finding the  $k$ -nearest neighbors. We explore the power of weak ties to find similar neighbors by utilizing a similarity measure from network science referred to as "Katz similarity" (KS) [8]. Although Katz himself never discussed it, KS captures regular equivalence of nodes in a network and can be applied in many different settings [4, 5].

**Approach & method.** Firstly, we utilize the trust connections to create an adjacency matrix where each entry represents a directed trust link between two users. Secondly, we apply the KS measure on the created trust adjacency matrix. More specifically, we calculate the pairwise similarities between users by using the iterative approach for calculating KS:

$$\sigma^{(l_{max}+1)} = \sum_{l=0}^{l_{max}} (\alpha A)^l \quad (1)$$

The iterative approach provides the possibility to set the maximum used path length ( $l_{max}$ ). This approach effectively gives us the ability to define the maximum path length used for forming weak ties between users who are not directly connected. We use the resulting similarity matrix and apply various row normalization ( $L_1$ ,  $L_2$ ,  $max$ ) and degree normalization techniques (in-degree, out-degree, and combined degree normalization) to get a better distribution of similarity values and better evaluation results concerning recommendation accuracy in return. Lastly, we apply an additional method to increase the similarity values derived from weak ties [1]:

$$\sigma_{boost} = A + \hat{\sigma}_{norm} \quad (2)$$

where  $\hat{\sigma}_{norm}$  is calculated by setting the values of strong tie similarities in  $\sigma$  to 0, normalizing the resulting matrix and then setting them to 1. With this approach, we achieve that each entry in  $\sigma_{boost}$

has a similarity value of 1 between pairs of nodes for which there exists an explicit trust connection in  $A$  while also increasing the importance of similarity values derived from weak ties. We evaluate these approaches on the Epinions dataset presented in [7] and compare results for  $l_{max} = 1$  (using only strong ties) and  $l_{max} = 2$  (using strong ties in combination with weak ties derived from paths of length 2).

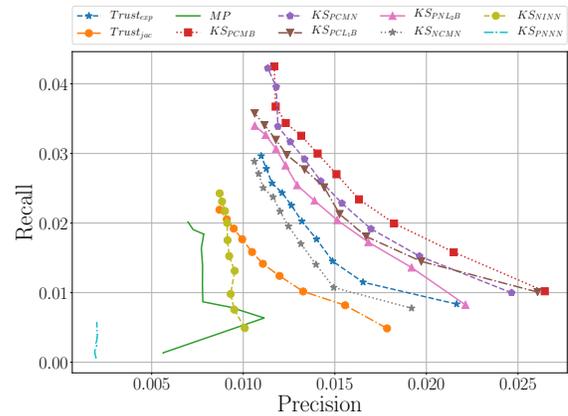
**Results & discussion.** In our study, we evaluate 33 approaches for various combinations of  $l_{max}$  values and normalization techniques. We compare these approaches with three different baselines:  $MP$  (recommending most popular items),  $Trust_{exp}$  (CF using trust connections for finding top  $k$  similar neighbors) and  $Trust_{jac}$  (CF using Jaccard coefficient on explicit trust values for finding top  $k$  similar neighbors). However, in Table 1, we only report the results for a subset of these approaches that provide the most insightful findings. All of the evaluation results are reported for  $n = 10$ , i.e., for 10 recommended items.

Approach	$l_{max}$	Degree norm.	Row norm.	Boost	nDCG	R	P
$Trust_{exp}$					.0224	.0296	.0110
$Trust_{jac}$					.0176	.0219	.0087
$MP$					.0134	.0202	.0070
$KS_{PCMB}$	2	Combined	Max	Yes	<b>.0303</b>	<b>.0425</b>	<b>.0117</b>
$KS_{PCMN}$	2	Combined	Max	No	.0295	.0422	.0113
$KS_{PCL1,B}$	2	Combined	L1	Yes	.0273	.0358	.0106
$KS_{PNL2,B}$	2	No degree	L2	Yes	.0257	.0340	.0106
$KS_{NCMN}$	1	Combined	Max	No	.0213	.0289	.0106
$KS_{NINN}$	1	In degree	N/A	No	.0161	.0243	.0087
$KS_{PNNN}$	2	No degree	N/A	No	.0036	.0057	.0020

**Table 1: Evaluation results for a subset of the 33 evaluated KS-based CF approaches**

The  $Trust_{exp}$  baseline uses only strong ties (explicit trust connections) for making recommendations and one of our main finding from the results of the conducted experiments is that by incorporating weak ties using paths of maximum length 2 from the target node (i.e. similar to adding friends of friends into the neighborhood), we can improve the quality of the recommendations in terms of recommendation accuracy with the best approach being the  $KS_{PCMB}$ . In the best performing approach, we set  $l_{max}$  to 2, apply combined degree (sum of in and out degrees) normalization, remove the strong ties, then perform *max* row normalization and then add the strong ties with the similarity value of 1. We also find that if we don't employ weak ties, i.e.,  $l_{max}$  is set to 1, we achieve better results when we do not apply degree and row normalization (i.e., basically the  $Trust_{exp}$  baseline). However, if  $l_{max}$  is set to 2, we can observe improvements in almost all of the cases except when no row normalization is applied, e.g., in the case of  $KS_{PNNN}$ .

Finally, in Figure 1, we show the performance of all approaches listed in Table 1 in form of Recall-Precision plots for different number of recommended items (i.e.,  $n = 1 - 10$ ). The results clearly show that the best performing algorithm (i.e.,  $KS_{PCMB}$ ) again outperforms all of the other approaches also for a smaller number of recommended items (i.e., for  $n < 10$ ).



**Figure 1: Recall-Precision plots of the described approaches for  $n = 1 - 10$  recommended items.**

**Conclusion & future work.** In this paper, we explored the use of Katz similarity (KS), a similarity measure of regular equivalence in networks, for selecting  $k$ -nearest neighbors in a Collaborative Filtering (CF) algorithm for cold-start users. We used an iterative approach to compute KS since it provides the ability to restrict the length of paths in the network used for similarity calculation. Consequently, we can investigate weak ties of arbitrary length. We found that KS can be a useful measure for neighbor selection if used with degree normalization and row normalization. In summary, with our work, we aimed to shed light on how to exploit weak ties in social networks to increase the performance of trust-based recommender systems. For future work, we plan to run additional experiments using different values for  $l_{max}$  and to explore the use of recently popularized node embeddings (e.g., Node2Vec [3]) to identify weak ties between users.

**Keywords.** Weak ties; Katz similarity; Trust-based recommenders

## REFERENCES

- [1] Tomislav Duricic, Emanuel Lacic, Dominik Kowald, and Elisabeth Lex. 2018. Trust-based collaborative filtering: tackling the cold start problem using regular equivalence. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 446–450.
- [2] Jennifer Ann Golbeck. 2005. *Computing and applying trust in web-based social networks*. Ph.D. Dissertation.
- [3] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 855–864.
- [4] Ilire Hasani-Mavriqi, Dominik Kowald, Denis Helic, and Elisabeth Lex. 2018. Consensus dynamics in online collaboration systems. *Computational social networks* 5, 1 (2018), 2.
- [5] Denis Helic. 2014. Regular equivalence in informed network search. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. IEEE, 1088–1093.
- [6] Neal Lathia, Stephen Hailes, and Licia Capra. 2008. Trust-based collaborative filtering. In *IFIP international conference on trust management*. Springer, 119–134.
- [7] Paolo Massa and Paolo Avesani. 2007. Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, 17–24.
- [8] Mark Newman. 2010. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA. 217–220 pages.
- [9] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative filtering recommender systems. In *The adaptive web*. Springer, 291–324.
- [10] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 253–260.