Who will Trade with Whom? Predicting Buyer-Seller Interactions in Online Trading Platforms through Social Networks

Christoph Trattner[1], Denis Parra[2], Lukas Eberhard[1] and Xidao Wen[3]
[1]Graz University of Technology, Austria
[2]Pontificia Universidad Catolica de Chile, Chile
[3]University of Pittsburgh, USA
ctrattner@know-center.at, dparra@ing.puc.cl, lukas.eberhard@tugraz.at, xiw55@pitt.edu

ABSTRACT

In this paper we present the latest results of a recently started project that aims at studying the extent to which links between buyers and sellers, *i.e.* trading interactions in online trading platforms, can be predicted from external knowledge sources such as online social networks. To that end, we conducted a large-scale experiment on data obtained from the virtual world SecondLife. As our results reveal, online social network data bears a significant potential (28% over the baseline) to predict links between buyers and sellers in online trading platforms.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

Keywords

Seller buyer link prediction; online social networks; Second Life

1. INTRODUCTION

Although a considerable amount of research has been conducted recently on how to predict links between users in networks, research utilizing data from more than one type of knowledge source is rare [2]. Especially in the context of online trading platforms, little work has been conducted so far to unveil the potential of external knowledge sources, such as online social networks, to predict buyer-seller interactions between users (e.g., [1, 3]). We contribute to this sparse field of research by presenting results of a recently started project that aims to understand the extent to which trading interactions are formed and can be predicted by employing data from external knowledge sources such as online social networks. To tackle this challenge we trained a binary classifier that learned the relations between pairs of users based on a number of features induced from online social network and trading network. Since it is nearly impossible to obtain rich large-scale real-world online social and trading network datasets by Web crawling, our investigations focused on the virtual world Second Life (SL), where we can easily find and mine both data sources. To the best of our knowledge, this is the first large-scale study that shows the potentials and limitations of a rich set of content- and network-oriented features to predict links between buyers and sellers from online social network data.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577364.

2. DATA COLLECTION PROCESS

We conducted our experiments on two different types of datasets - the SL online social network MySecondLife[1] and the SL Marketplace[2]. Similar to eBay, every seller in the SL marketplace owns her own sub-page – called the seller's store – where all items offered are presented to the general public. As with other trading platforms such as Amazon, sellers in the SL Marketplace have the possibility to apply meta-data information such as price, title, or description to their products. Customers in turn are able to provide reviews or ratings to products. In order to crawl all stores and corresponding meta-data information and interactions from the SL marketplace, we exploited the fact that every store has a unique URI built from the URL pattern http://marketplace.secondlife.com/stores/STORE_ID, where STORE_ID is an incremental integer starting at 1. With this exploit at hand, we were able to download 131,087 complete store (=seller) profiles with corresponding 268,852 trading interactions.

The online social network MySecondLife was introduced by Linden Labs, in July 2011. It is aimed to compared to Facebook regarding postings and check-ins but aims only at residents of the virtual world. Hence, users can interact with each other by sharing text messages and commenting or loving these messages. A user profile can be accessed through a unique URL, https://my.secondlife.com/en/USER_ID, where USER_ID depicts the user's name. The necessary names were extracted from the SL Marketplace dataset. In order to gather a large fraction of the whole network we also extracted all interaction partners recursively until no new user could be found. All over, 169,035 complete user profiles with overall 3,175,304 interactions were downloaded.

3. EVALUATION METHODOLOGY

As highlighted, it is our interest to predict buyer-seller interactions between users in an online trading platform based on features

[1]https://my.secondlife.com/
[2]https://marketplace.secondlife.com/

		Feature (Description)	InfoGain
Online Social Network	Network	Adamic Adar	0.0027
		Number of common neighbors (in)	0.0024
		Number of common neighbors (out)	0.0033
		Jaccard's Coefficient neighbors (in)	0.0024
		Jaccard's Coefficient neighbors (out)	0.0034
		Preferential Attachment score (in)	**0.0259**
		Preferential Attachment score (out)	0.0046
		Reciprocity of user communication	0.0050
		Number of total neighbors (in)	**0.0248**
		Number of total neighbors (out)	0
	Content	Number of common regions	0.0009
		Jaccard's Coefficient checked-in regions	0.0015
		Number of total checked-in regions	0
		Number of common favored regions	0.0070
		Jaccard's Coefficient favored regions	0.0070
		Number of total favored regions	0.0070
		Number of common groups	**0.0182**
		Jaccard's Coefficient groups	**0.0190**
		Number of total groups	0.0035
		Number of common interests	0
		Jaccard's Coefficient interests	0
		Number of total interests	0.0019
		Total number of Interactions	0.0074
Trading Network	Network	Adamic Adar	0.0070
		Number of common neighbors (in)	**0.0152**
		Number of common neighbors (out)	0.0080
		Jaccard's Coefficient neighbors (in)	**0.0152**
		Jaccard's Coefficient neighbors (out)	0.0081
		Preferential Attachment score (in)	**0.4564**
		Preferential Attachment score (out)	0.0085
		Number of total neighbors (in)	**0.3101**
		Number of total neighbors (out)	**0.1556**
	Cont.	Cosine Similarity product categories	**0.1575**
		Cosine Similarity product prices	**0.1105**
		Cosine Similarity product ratings	**0.1263**

Table 1: Features and sets used in the experiment (in = incoming links, out = outcoming links). Features marked in bold have an information gain > 0.01.

induced from an online social network as well as a trading network. To that end, we induced two different types of feature sets from our two data sources: network- and content-oriented feature sets. As highlighted in Table 1 both network-oriented feature sets consist of popular link prediction approaches such as the Adamic Adar measure, the Preferential Attachment score, or the Common Neighbors measure [2]. Our content-oriented feature sets mainly consist of simple similarity measures e.g., based on the users favorite locations, stated interests or groups.

For the prediction task we randomly selected 10,000 user-pairs (which were present in both networks) having a trading interaction and 10,000 having no trading interaction. This procedure results in balanced datasets for the test and training data, and therefore in a baseline of 50% for the prediction task when guessing at random. To validate our results, we employed three supervised algorithms: J48, Logistic Regression and SVM (with linear kernel) using the WEKA machine learning suite with 10-fold cross validation.

4. RESULTS

The results of our experiment can be found in Table 2. As shown, the Logistic Regression (Logistic) approach reached the highest estimates in all feature sets and networks. Furthermore, we observe that utilizing online social network features alone and using a logistic regression model allows us to predict trading interactions with 0.641 AUC, which is an improvement of +28% over

Feature Sets		J48	Logistic	SVM
Online Social Network	Network	0.608	**0.610**	0.595
	Content	0.592	**0.602**	0.553
	Combined	0.625	**0.641**	0.578
Trading Network	Network	0.863	**0.899**	0.792
	Content	0.702	**0.713**	0.711
	Combined	0.854	**0.893**	0.797
Combined		0.868	**0.901**	0.716

Table 2: AUC results for predicting trading interactions between buyer and seller with different feature sets and learning algorithms. Best results are highlighted in bold.

the baseline, when guessing at random. Moreover, we can see that network-oriented features outperform content-oriented features in both datasets. We can also clearly see that features induced directly from the trading network outperform the online social network features and set (see also InfoGain values in Table 1). Finally, we can observe that the combination of trading- and network-oriented feature sets only slightly improves the model trained on the trading feature set alone.

5. CONCLUSIONS AND FUTURE WORK

In this work we presented results of research aimed at predicting online trading interactions using features from the trading network as well online social network data. Considering data from the SecondLife virtual world, we studied 35 features categorized into network- and content-oriented ones. Features from the trading network yielded better performance than those from the online social network, with logistic regression outperforming both SVM and J48. Although the combination of features from both social and trading networks did not show a significant improvement over trading network data alone, our results indicate that the online social network data improve the predictive accuracy of trading interactions over random guessing by 28%. This result highlights the potential of online social network data in a recommendation scenario: For instance, a user that navigates for the first time to an online trading platform, i.e., under a cold-start situation, can be recommended potential sellers by exploiting her online social network. In our future work, we plan to study network motifs and other features such as message content to broaden our understanding of how people link with each other in online trading platforms. Furthermore, we are interested in using our findings for building a recommender system for social online marketplaces.

Acknowledgments: The authors would like to thank Michael Steurer for assistance in crawling the SecondLife dataset. This work is supported by the Know-Center.

6. REFERENCES

[1] S. Guo, M. Wang, and J. Leskovec. The role of social networks in online shopping: information passing, price of trust, and consumer choice. In *ACM Conference on Electronic Commerce*, pages 157–166, 2011.

[2] M. Steurer and C. Trattner. Who will interact with whom? a case-study in second life using online social network and location-based social network features to predict interactions between users. In *Ubiquitous Social Media Analysis*, volume 8329 of *Lecture Notes in Computer Science*, pages 108–127. Springer Berlin Heidelberg, 2013.

[3] Y. Zhang and M. Pennacchiotti. Predicting purchase behaviors from social media. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 1521–1532, 2013.