Journal of Advanced Research in Technology and Management Sciences
Volume: 02 Issue: 03 ISSN: 2582-3078 November 2019
Available online at: http://www.jartms.org

SOCIAL NETWORK ANALYSIS OF C2C E-COMMERCE RECOMMENDER SYSTEM BASED ON LINK PREDICTION BY EMPLOYING GRAPH BASED APPROACH

[1]AAREPU LAKSHMAN, [2]Dr. YOGESH KUMAR SHARMA, [3]Dr. B.M.G. PRASAD
[1]Research Scholar, Shri Jagdish prasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
[2]Associate Professor & HOD Dept of CSE & IT, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
[3]Professor, Dept of CSE and Dean PG Courses, Holy Mary Institute of Technology and Science Hyderabad, Telangana, India

ABSTRACT: Social network analysis emerged as an important research topic in sociology decades ago, and it has also attracted scientists from various fields of study like psychology, anthropology, geography and economics. This paper focuses on building a recommendation algorithm for C2C (Consumer-to-Consumer) e-commerce business model by considering special features of C2C e-commerce websites. In this paper, users and their transactions considered as a network based on link prediction using a graph based technique. This is an important task in social network analysis is used to build the recommender system. The proposed tow-level recommendation algorithm, rather than topology of the network, uses nodes' features like category of items, ratings of users, and reputation of sellers. The links predicted can be presented as recommendations to the user. This paper also focuses on how graph algorithm can be used to improve recommendation in ecommerce websites. The method incorporates semantic recommendation using overlap technique based in graph. The results show that the proposed model can be used to predict a portion of future trades between users in a C2C commercial network.

KEY WORD: C2C e-commerce, Social network analysis, link prediction, Graph based algorithm, overlap technique

I. INTRODUCTION

Recommendation in e-commerce means providing the users with products and services they are interested in. Recommendation system in e-commerce has become extremely popular in the recent years. E-commerce websites use different techniques to provide users with better experience in online shopping.

With new technology and improved techniques e-commence is able to provide users product and services based on their interest. Different techniques such as content based, collaborative based and hybrid based are used to give users a better shopping experience [5].

Recommendation based on best-selling items, demographic information of users, and user's past behavior analysis, are samples of techniques used by Recommender systems in ecommerce websites [3]. These techniques can be considered as different types of personalization, since they help the website to adapt itself to the customers. Although most of commercial websites use B2C e-commerce, some of them support C2C ecommerce. Ebay.com, Bizrate.com, etc. are instances of websites in which customers can be buyer and seller at the same time. Current recommendation algorithms (e.g. collaborative filtering, content based algorithms, and hybrid algorithms) can be most useful for B2C e-commerce websites. Since these algorithms only concentrate on items and consumers (but not sellers), there is a lack of recommend systems for C2C e-commerce websites [10].

The solution proposed in this paper for the mentioned problem is based on social network analysis. Social network analysis emerged as an important research topic in sociology decades ago, and their first study was focused on the adoption of medical and agricultural innovations. It is a vast field of research that has attracted researchers from anthropology, economics, psychology, biology and geography, just to mention a few. Link prediction which is an important task in social network analysis, is the problem of predicting the existence of a link between two nodes in a graph where prediction is based on the attributes of the nodes and other observed links [9].

II. RELATED WORK

Recommendation in e-commerce website is a general term for providing users with products and services (queries, novels, policies, movies, images, books, Web pages, etc.) relevant to their taste. User-based recommendation works by recommending items that are liked by the like minded users, and item-based recommendation works by suggesting items based on similar properties.
Recently, several matrix factorization methods have been used for collaborative filtering [4]. These methods are used to combine the user-item rating matrix using low-rank approximations algorithm, and use it to make better predictions.

In content based filtering the properties of the product or services are utilized for recommendation. The properties of a product or services under consideration are analyzed and are matched with other properties of the products present in the database [6]. The similar property products are than displayed as the recommended products. Content based filtering has wide application in e-commerce websites. Sites such as amazon, eBay, snapdel, flipkart utilized content based filtering effectively to recommend product to the users. In order to recommend relevant queries to Web users and make surfing of net easy, a technique called query suggestion has been used by some prominent commercial search engines, such as Yahoo!, Bing, Ask, and Google. Query suggestion predicts the interest of an active user by analyzing information from similar users or items. Query suggestion helps by narrowing down the scope of the search. Query suggestion helps to suggest full queries that have been used by previous users. This helps in preserving query integrity and coherence.

In [8], Guy, studied personalized recommendation of social software items, including bookmarked web-pages, blog entries, and communities. They focused on recommendations that are derived from the user's social network. In their proposed model, Social network information is collected and aggregated across different data sources. In, Arazy et al. [9] proposed a framework for improving social recommender systems using behavioral theories of advice-taking. They identified four factors that impact a user's decision when he receives a recommendation: Homophily, Trust, Reputation, and Tie strength. After computing these relationship factors, the system should calculate a weighted average of them in order to rate each recommendation source.

With so much of data being collected in the recent years and considering of ever changing experience of the users, it is not only possible to recommend products based on content or collaborative recommendation system [10]. That is the reason why most of the users prefer to surf e-commerce sites where integration of both the methods are involved. Most of the ecommerce have adapted to the hybrid recommendation system. These recommendation system can be a combination of collaborative, content based, click through analyses, query suggestions, etc.

III. PROPOSED MRTHOD

In general, link prediction approaches are divided into three categories: methods based on structural measures in the network, methods based on the content or attribute similarity between nodes, and methods based on probabilistic models. In this paper, measures based on structure of network and attributes of nodes are used for link prediction task. Our proposed model for link recommendation (i.e. recommendation of buying items from sellers) in commercial networks is shown in Figure (1). As shown in Figure (1), the recommendation model consists of six stages. It should be noted that stage 2, 3 and 4 can be done in parallel, but other stages must be performed in specified order.

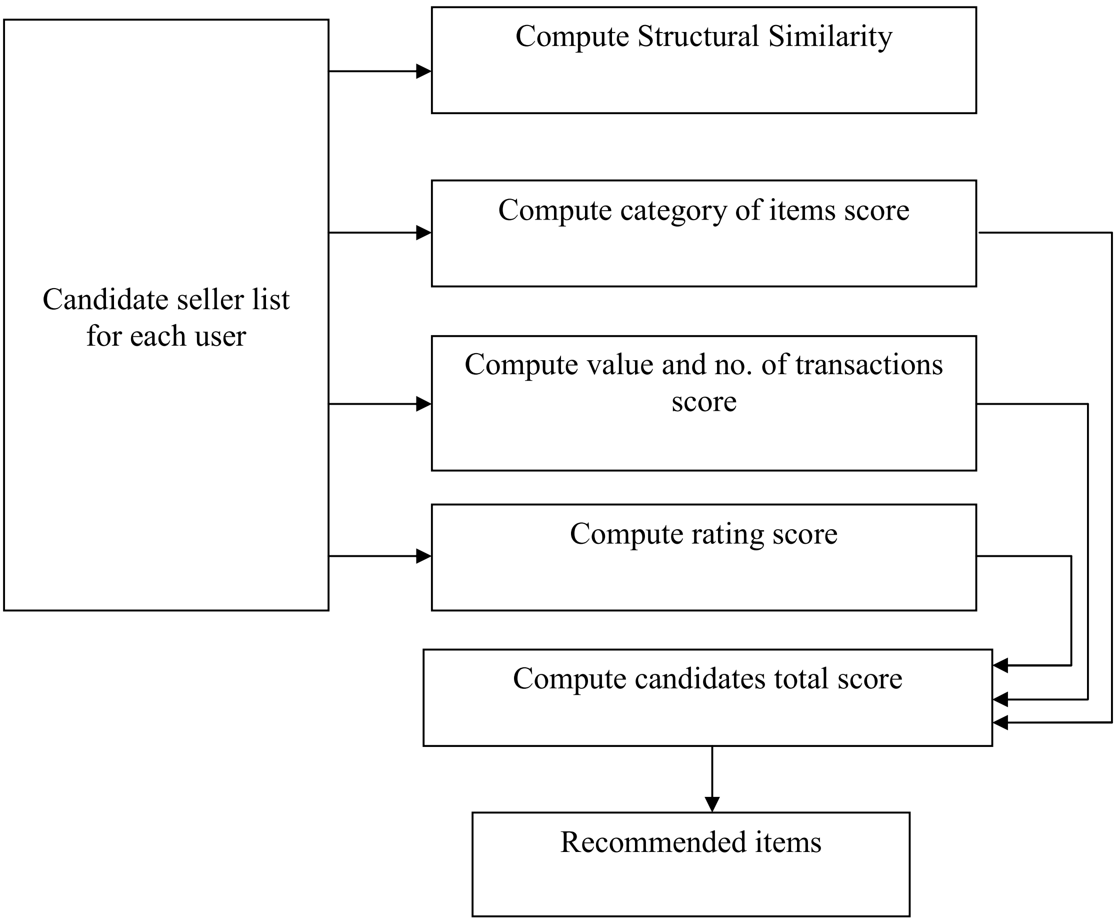


Fig. 1: PROPOSED LINK RECOMMENDATION MODEL FOR C2C COMMERCIAL NETWORKS

A. Computing structural similarity

SimRank is a similarity measures proposed for directed graphs. This measure computes similarities based on an intuitive concept: two objects are similar if they refer to similar objects or they are referred by similar objects. Like PageRank algorithm, SimRank computes similarity of graph nodes using their input and output edges. The first step is to calculate user similarities according to SimRank measure. We use Eq. (1) to calculate similarity of target users to other users in transactions graph.

$$S(u,v) = \frac{C}{|I(u)||I(v)|} \sum_{i=1}^{|I(u)|} \sum_{j=1}^{|I(u)|} S\left(I_i(u), I_j(v)\right) \text{ --- (1)}$$

In Eq. (1), u and v are sample users, $S(u,v)$ is the similarity of u and v, C is the damping factor, $I(u)$ is the list of sellers that user u have purchased items from and $I(v)$is the list of sellers that user v have purchased items from. After computing similarity of each target user like u to other users in graph, there will be a list of n similar users to u We use the list of similar users to specify candidate sellers; for this purpose, we choose sellers that have sold items to users in u's similarity list (but u has bought nothing from them) as candidate sellers. Considering $n = 1$ (size of u's similar users list), candidate sellers for u in this graph would be d and e.

B. Computing Category Scores

There are bunch of algorithms for measuring similarity of such vectors. Here, the significant note is the importance of different categories for each user. For considering the importance of each category for each user, we used two factors:
1. The importance of each category in local network of a user which can be calculated using Eq. (2);
2. Number and value of items bought or sold in each category.

$$L_u(a) = \sum_{v \in N(u)} A(v,a) \text{ --------- (2)}$$

In Eq. (2), $L_u(a)$ is the importance of category a in user's local network, a is a sample category, and $N(u)$ is the list of u's adjacent nodes. $A(v,a)$ is equal to 1 if both u and v have bought or sold an item in category of a, and 0 otherwise; If u and v does not have any category in common, $L_u(a)$ is equal to 1.

Now we can use similarity measures like Jaccard's coefficient to compute the score of candidate sellers for the target user according to category of items. Eq. (4) is the adapted form of Jaccard's coefficient for our problem, Where $S_{category}(u,v)$ is the category score between u and v, c is a category that both u and v have traded in, $A(u)$ is the list of categories that user u have traded in, and $w_u(a)$is the category score of attribute a according to user u:

$$S_{category}(u,v) = \sum_{c \epsilon A(u) \cap A(v)} \frac{w_u(c) + w_v(c)}{\sum_{a \in A(u)} w_u(a) + \sum_{b \in A(v)} w_v(b)} \quad ------(4)$$

C. Computing Reputation Scores

Reputation is a quantity derived from the prior interactions of individuals to others in a community which is globally visible to all members of the network. The type of reputation system in our model is centralized but unlike common centralized reputation systems, every seller may have a different value of reputation form the viewpoint of different users. To compute reputation of each seller, four factors are considered: seller ratings from prior transactions, monetary value of each transaction, total number of seller's transactions, and importance of categories to target user. Eq. (5) shows how we calculate a seller's reputation from perspective of a target user.

$$S_{reputation}(u,v) = \frac{\sum_{i=1}^{s(v)} \left(R_s(v,i) \times D(v,i) \times L\big(u, g(i)\big)\right)}{Q(v)} \quad ------ (5)$$

$$\alpha + \beta = \mathcal{X}$$

$S_{reputation}(u,v)$ is the reputation of user v from user u's view.
$Q(v)$is total sales number of user v, $R_s(v,i)$is the average value of item quality, responsiveness, shipment cost, and shipment time rating that seller v have received in transaction i (a number between [-1,1]). $D(v,i)$ is the monetary value of user v's sales in transaction i. $L\big(u,g(i)\big)$ is the importance of category in which user v have sold an item and is computed similar to $L(a)$in Eq. (2); If u has not traded in category $g(i)$, $L\big(u,g(i)\big)$ value will be equal to 1.

D. Computing Ratings Scores

The other metric that can be used to find best sellers, is similarity between candidate sellers and prior sellers that have sold items to users, in terms of ratings they have gained. We assume that buyers rate sellers after each transaction, regarding to the overall satisfaction, quality of goods, delivery time, and seller support. These rating vectors can be used to choose best sellers for target users. Several metrics have been proposed in order to compute similarity of two numeric vectors. Cosine similarity and Pearson correlation are the most common ones.

E. Computing Total Score of Candidates

The List of best sellers for a target user is built using three metrics we discussed before (Category score, Ratings score, and Reputation score). Before using these 3 scores, they should be normalized into a number between [0,1]. Eventually, total score of each candidate is calculated by combining the scores from previous stages:

$$S_{total(u,v)} = \alpha \times S_{cat}(u,v) + \beta \times S_{rat}(u,v) + \gamma \times S_{rep}(u,v) \quad ----- (6)$$

In Eq. (6), $S_{total(u,v)}$ is the total score of seller v for recommending his items to user u, α is category score coefficient, β is ratings score coefficient, Υ is reputation score coefficient, $S_{cat}(u,v)$ is category score of user v according to user u, $S_{rat}(u,v)$ is rating score of user v according to user u, and $S_{rep}(u,v)$ is reputation score of user v according to user u. α, β and Υ must be chosen in range of [0,1] so that $\alpha + \beta + \Upsilon = 1$. By changing value of these coefficients, one can change the influence of each score in building best sellers list.

F. Overlap method

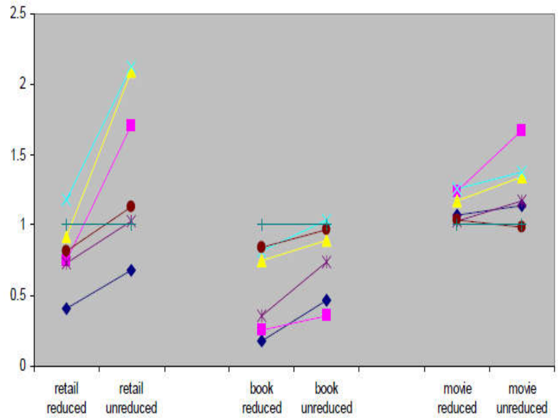
The matching profile click through analyses is provided as recommended products in the form of you may also like recommendation on e-commerce sites. Warehouse stores content of the website this includes products and product catalogue, users, and the usage logs generated by the web server. The matching profiles to an active user which are used to suggested products or services are called recommenders or like minded users. Recommendation engine generates the recommendations of the users by considering parameters such as their profiles, products profiles, click through analyses, recommenders profile, etc. The recommendation module is the algorithm which constantly evolves by learning the users behavior. The users liking can be analyzed from the feedback provided the users knowingly or unknowingly. Clicking of the suggested product, analyzing the similar products, history searched on the site, the products previously brought, etc are all feedback collected by the feedback module. This feedback is taken every time the user is surfing and is used for recommending better relevant products for the users.

Most of the e-commerce websites and search engines lack semantic and sentiment recommendation property. The site under consideration is flickr.com which is a search engine. The proposed system uses overlap semantic method to integrate recommendation and semantics. In the proposed system images are recommended to the users based on the query suggestion. The images data set is stored in the graph format, images from the nodes and the semantic relation between them is depicted by the overlap technique forms the edges. The tag of the images which are provided by the user are stored in an array as nodes. The overlap value which is calculated for the semantically related images, are stored as distance in an array. The graph algorithm which creates a sub graph of the selected images from the data set related to the query used for recommendation. Insert images in dataset along with tags, Store tags as nodes and connect the nodes with one another with the help of overlap formula.

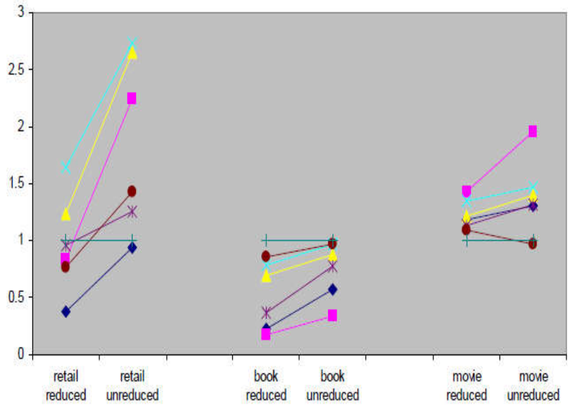
$$overlap = \frac{N(p \cap q)}{min\big(N(p)N(q)\big)}$$

IV. RESULTS

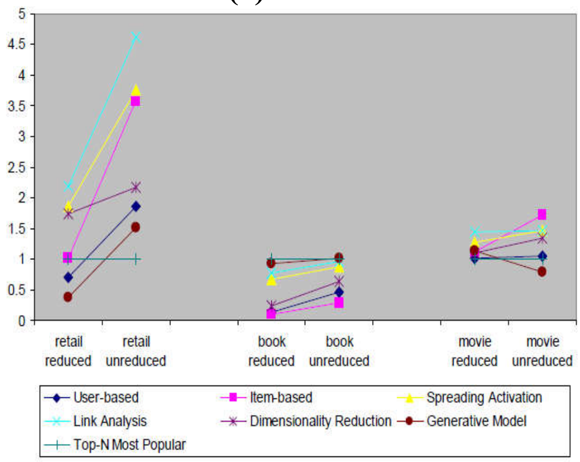
To provide a summary of each algorithm's overall performance across different datasets, we reported the average rank of each algorithm with respect to the four measures.



(a) Precision



(b) Recall



(c)Rank Score

Fig. 2: RELATIVE PERFORMANCE MEASURES

For example, for the precision measure the link analysis algorithm's average rank is 1.50, which corresponds to the average of its ranks for individual datasets (1, 1, 3, 1, 2, and 2). Boldfaced average ranks are the top 2 average ranks. As collaborative filtering algorithms can only recommend products to the consumers that appeared in the training transactions, for consumers with no future purchases in the testing set appeared in the training set, no successful recommendation is possible. To make the performance measures more meaningful, we only evaluate recommendations for target consumers for whom successful recommendations are possible. Therefore, for the same dataset the reduced and unreduced training sets resulted in different numbers of target consumers.

V. CONCLUSION

Modeling users' behavior and recommending items based on their interests is one of the most efficient strategies in ecommerce. Recommender systems made progress over the last few years in e-commerce websites. The goal of the system is recommending items from most appropriate sellers to target customers. The proposed system is able to successfully integrate recommendation and semantics. The proposed model proves that recommendation can be improved if semantic factor while recommending product or services is integrated in the system. The portion of predicted links by our model may seem to be low, but one should consider that even this low percent can end up to a big number of new transactions in large scale commercial networks. As the effect of using weight of links in a graph for link prediction is not deterministic yet, other possible future work can be using the weight of links between users for computing similarities at the stage of creating candidates list. The Results of our experiments show that the proposed model can predict some of the missing links in commercial networks.

VI. REFERENCES

[1] Antonii Rzheuskyi ; Nataliia Kunanets ; Mykola Stakhiv, " Recommendation System "Virtual Reference", 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)**,** Volume: 2**,** Year: 2018**.**
[2] S. Parameswaran; E. Luo; T. Nguyen, "Patch Matching for Image Denoising Using Neighborhood-based Collaborative Filtering," in IEEE Transactions on Circuits and Systems for Video Technology,vol.PP,no.99,pp.11doi:10.1109/TCSVT.2016. 2610038
[3] A. Boroujeni, G. Assadat, and S. A. Hashemi Golpayegani, "Design Context Aware Recommender System Model in M-commerce Platform Using Collaborative Filtering Approach in Social Networks," Caspian Journal of Applied Sciences Research, vol. 2, 2013.
[4] H. Ma, I. King and M. R. Lyu, "Mining Web Graphs for Recommendations," in IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 6, pp. 1051-1064, June 2012. doi: 10.1109/TKDE.2011.18
[5] G. Shani and A. Gunawardana, "Evaluating recommendation systems," Recommender systems handbook, Springer, 2011, pp. 257–297
[6] P. W. Yau and A. Tomlinson, "Towards Privacy in a Context- Aware Social Network Based Recommendation System," Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, Boston,MA,2011,pp.862-865.
[7] O. Arazy, N. Kumar, and B. Shapira, "Improving social recommender systems," IT professional, vol. 11, pp. 38- 44, 2009.
[8] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, et al., "Personalized recommendation of social software items based on social relations," in Proceedings of the third ACM conference on Recommender systems,2009, pp. 53-60.
[9] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," Journal of the American society for information science and technology, vol. 58, pp. 1019-1031, 2007.
[10] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," in Applications of Data Mining to Electronic Commerce, ed: Springer, 2001, pp. 115-153.