International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12989-12996

A Community Based Content Recommender Systems

Satya Keerthi Gorripati
Information Technology Department, GMR Institute of Technology, Rajam, Andhra Pradesh, India.

Orcid Id: 0000-0002-1127-1055

Valli Kumari Vatsavayi
Department of Computer Science and Systems Engineering,
Andhra University, Visakhapatnam, Andhra Pradesh, India.

Orcid Id: 0000-0002-7252-8301

Abstract

With the expansion of content sources on web 2.0, is an important issue of content-based filtering where the framework abilities to take user ratings from user's activities in regards to one content source and utilize them over the other content types. When the service is restricted to recommend the content of an indistinguishable case, for the user is now utilizing, the value from proposed framework is altogether less than for this content whatever the user using is now, he will be recommended to use other content types. For instance, recommend news articles based on history of browsing, playing similar characteristic music to the song supplied by the user as an initial seed and providing document recommendations. Many content-based recommender systems are active on websites like Pandora Radio, Internet Movie Database and Rotten Tomatoes etc. In spite of many successful recommenders there is even a need for an accurate one. Recent studies focus on combining social annotation through community detection with collaborative recommender systems. In this manuscript, we propose a framework, which merges both community detection and content recommendation in order to amend the existing community based recommendation. The efficiency of the proposed approach is compared against the traditional approaches. Experimental results show progress in resolving the issues faced by the collaborative approaches.

Keywords: Community Detection, Content based, Recommendation, Rating, Visualization perusing

INTRODUCTION

Advances in technology has added to the establishment of a new era of services such as web logs, wikis, social media and social networking sites such as Facebook, Twitter, Wikipedia, YouTube, Amazon, Flickr, and Google+. The information accessible is extremely valuable, however the sheer amount of it (called information overload) and its uncertainty due to various formats, and constrains. Recommender systems have been suggested to address this data over load issue by filtering suitable information and recommending items that are nearer to the pastimes of the users [1]. These recommender systems are effectively implemented in popular websites such as Amazon, Flip kart and Netflix etc. Typically, conventional recommender systems use either the collaboration between items and users (collaborative based) or an integration of them (hybrid based) or the content of items (content based). Yet they experience the ill effects of sparsity, cold start and over specialized recommendations [2].

Moreover the general premise is that users of a domain are identical and independent, whereas in genuineness, people turn to friends, they trust for the hint [3, 4]. These social interactions among users are completely ignored in traditional recommender and thus they provide unrealistic recommendations [5, 15]. The information produced from social networks can take on a major role in knowledge discovery, information diffusion and in application development and could address the shortcomings related with conventional recommenders.

This paper proposes a Community Based Content Recommender System (CBCRS) that uses an user interacted item inside a community, and giving a recommendation that is similar in content to that item and belongs to the same community. Due to being content community based, the CBCRS tends to the accompanying downsides in recommendation scenario:

1. The user profile is constructed by exploiting the ratings made by him, but not by seeing the closest neighbors of the active user (like collaborative approach).

2. It is capable of recommend new user and thus defeating the cold-user problem related with new users.

3. It generates recommendations by considering moderately small set of data, but not the whole set of neighbors and in this way conquering the sparsity issue.

The Framework is applied in the social network of movies in MovieLens data set wherein the network is a collection of interactions which can be established based on edge list or adjacency matrix when a community algorithm is applied on it, it results a tightly connected clique called community. These social systems are critical as nodes within a community tend to possess identical attributes.

RELATED WORK

User preference is made based on preferences and interests in view of preferences given by the user to certain items along with the content related to them in content-based recommender systems. Preferences/Tastes can be gathered either explicitly from users, by requesting that they rate an item, or implicitly analyzing his doings. By comparing the user profile in contradiction to topographies of all items recommendations are then generated for him. A substance can be represented using keyword-based models, in which the recommender makes a Vector Space Model (VSM) portrayal of specific features where an item is mapped by a vector in a multi-dimensional distance. These dimensions represent the features used to key out the details. This implies, the system determines a significance of the user's interest towards the items. For example, in the movie recommendation, the features refers to an item can be actors, director or genre. With this feature the content-based recommender system may actually handle cold-start problem [8].

In the literature, many content-based recommendation algorithms have been proposed. One such traditional algorithm is established on a KNN approach in which the user preference of an unknown item is calculated based on all items known by the user in the catalogue [9, 10]. According to its similarity with the known items every item's contribution is used to predict the preference score [11]. This similarity can be evaluated using several measures like Cosine, Pearson, etc. which completely depend on users' rating history and miss the social interactions between users. There is a recent trend in combining recommendation from social networks and community detection to improve personalized recommendations related to users belonging to the similar community.

Many static community detection algorithms have been suggested which can expose the users' preferences based on clustering model and hybrid approach [12, 13, 14]. Sahebi [6] proposed an approach using community detection to provide a solution to the cold start problem. Qiang [16] also defined a new approach based on multi-label propagation algorithm for static community detection. The notion behind using this algorithm is to utilize the overlapping community structures to recommend items. Louvian community detection applied on social network from the Internet Movie Database to provide personalized recommendations based on the evolved communities. A memory based community proposed by Zhao [17] based on matrix factorization using Latent Dirichlet allocation method on social network.

The above methods only deal with static networks and on collaborative approaches. A very little literature is focuses on content based approaches which provide actionable information filtering and are the basis for text mining. The proposed framework focus on this issue in order to recommend items for the users based on content.

Basic Concepts

Content-Based systems focus on properties of items on what the user has favoured in the past. Similarity of items is determined by evaluating the similarity in their properties. An item profile is to be constructed for each point which is a record or collection of records representing important characteristics of the item that are easily seen [18].

The chief component of the CF-based recommender systems lies in its capability to decide which users accept the most common with a given use. Hybrid CF is the combination of both CF and content based approaches. Many advances are there working out the similarity between users. The following are the most common approaches:

Correlation-based: This correlation is calculated by comparing the active users' rating with co-rated items. The following equation illustrates the correlation between user u and its adjacent user v, where m is the list of co-rated items between u and v.

$$PCC(u,v) = \frac{\sum_{p \in m} (r_{u,p} - \overline{r_u})(r_{v,p} - \overline{r_v})}{\sqrt{\sum_{p \in m} (r_{u,p} - \overline{r_u})^2} \sqrt{\sum_{p \in m} (r_{v,p} - \overline{r_v})^2}}$$

Cosine-based: Users are considered as vectors in the N dimensional item-space. The similarity is calculated by finding cosine between the vectors. The similarity between user u and v in the rating matrix is denoted by

$$COS(u,v) = \frac{\vec{r_u} . \vec{r_v}}{\left\|\vec{r_u}\right\| \left\|\vec{r_v}\right\|}$$

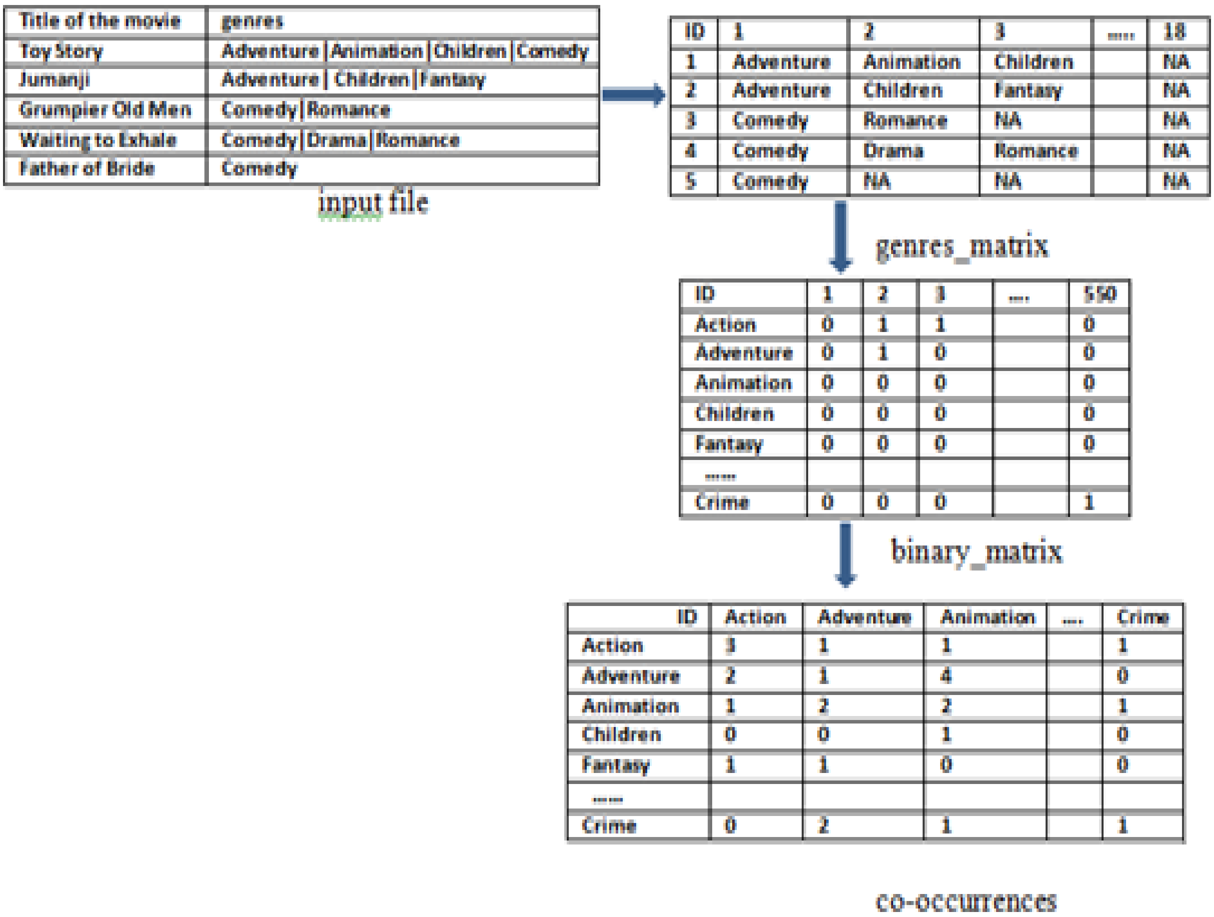
The traditional approaches for the collaborative filtering recommendation, consider top-k nearest neighbourhood in the entire dataset. To reduce this subset authors has proposed community based recommender systems. In this framework we are proposing content based community recommender system to exploit the benefits of both content and collaborative approaches.

PROPOSED WORK

A. The pre-processing Step

In this framework we build a content-based recommender engine based on movie genres only. It may be possible to include multiple attributes that have been determined to be important. A user profile is necessary to define what a user is inclined to prefer, and can be constructed based on the viewing behaviour or user's preferences. We assume if a user rate an item if he prefers to promote that movie. These genre preferences are used to build a network. In this work, the vertices of the network are assumed to be items and hence the communities are tight subgroups of similar genres.

The data available in the movieLens contains all genres related to that movie. A genre_matrix is created for each movie to indicate which genre was present in that movie. Afterward the genre_matrix and rating matrix is converted into a binary matrix. Finally the user profile builds which is the dot product of the genres and the rating information as shown in Figure.1.



Title of the movie	genres
Toy Story	Adventure|Animation|Children|Comedy
Jumanji	Adventure|Children|Fantasy
Grumpier Old Men	Comedy|Romance
Waiting to Exhale	Comedy|Drama|Romance
Father of Bride	Comedy

ID	1	2	3	.....	18
1	Adventure	Animation	Children		NA
2	Adventure	Children	Fantasy		NA
3	Comedy	Romance	NA		NA
4	Comedy	Drama	Romance		NA
5	Comedy	NA	NA		NA

ID	1	2	3	....	550
Action	0	1	1		0
Adventure	0	1	0		0
Animation	0	0	0		0
Children	0	0	0		0
Fantasy	0	0	0		0
......					
Crime	0	0	0		1

ID	Action	Adventure	Animation	....	Crime
Action	3	1	1		1
Adventure	2	1	4		0
Animation	1	2	2		1
Children	0	0	1		0
Fantasy	1	1	0		0
......					
Crime	0	2	1		1

(a)

```
> result[3,]
 [1] 0 1 1 0 0 0 0 1 1 1 0 1 0 0 0 1 1 0 0 0 0 0 1 1
```

(b)

Figure.1: (a) A Sample of generating feature matrix and co-occurrence information and (b) the sample user profile for 25 movies.

The network is a collection of interactions and these are constructed based on the co-occurrences of genres when applied a community detection algorithm on it, it results a tightly connected clique, called communities [19, 20, 21] as shown in Figure 2.

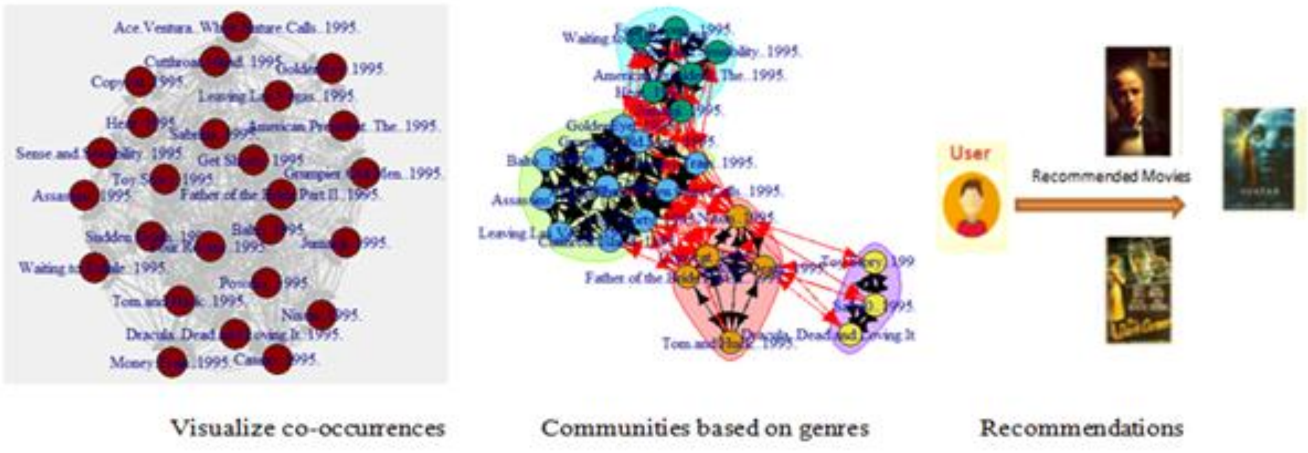


Figure 2: A Sample recommendations.

B. Community Detection Step

The relation between two vertices is modelled using the co-ratings between genres. Indeed an edge between vertices occurs when at least one movie has the same genre for both of them. Based on the network relations, a community can be defined as a set of items that are strongly connected. The advantage is that a community is restricted to only genres (e.g. Horror, Romance, Sci-Fi, etc.) in our framework. In doing this we use a Label propagation algorithm [22, 23] which is a popular and efficient algorithm to detect disjoint as well as overlapped communities[24,25] in static networks. The drawback lies in disjoint communities is, based on frequent interactions the vertices are forced to belong to a single community even though it has participation with multiple communities. Despite of its disadvantages, we aim to find the disjoint communities in static networks, founded along the combination of genres every item (or vertex) belongs to an undivided community.

C. Recommendation Step

In this step, the learned patterns and user profile will be exploited to find the inclination of each user towards each genre. These characteristics will help the recommender system to predict the future preferences. A highest rated item (called target item) for each user is identified. The items which are not rated by the current user and belong to the target items' community are treated as candidate items. The list of neighbor items in order to recommend items to candidate items is restricted to community, but not the whole set of neighbors.

D. A Running Example

A simple example is shown to illustrate the procedure using the proposed approach. Let us consider the user ratings of the movie ratings shown in Figure 3, discovered three communities namely: C_1, C_2 and C_3 as shown in Figure 4. The rating scale from 1 to 5 directs the level of user's concern in items, where 1 represents the unconcerned and 5 represents the most concerned item. With the intention to recommend items to the first user, we need the concerned item and candidate items. The item with the highest rating is the target/concerned item and the items belongs to the community of the target item and are not rated by the active user are the candidate item. In our example the target item for the first user is *Leaving Lasvegas* and candidate items are *Powder, Copycat, Sense and sensibility, Casino, Waiting to exhale, etc.*

In the process of computing predictions for the first user the first row is considered shown in Figure 3 and accordingly the concerned item and candidate items are taken from community C_3 shown in Figure 4 and the similarities are computed using Correlation-based shown in Table 1. For simplicity we have used the starting letter of the item in the following equation to compute the prediction for the candidate item Powder. After predicting all candidate items the highest preferred items are recommended to the active user.

	toys	Juma	Grum	Wait	Fat	Heat	Sabr	Sudd	Golc	Ame	Drac	Balto	Nixo	Cutt	Casin	Sens	Four	Ace Ve	Mone	Get S	copyca	Assas	Powc	Leavi	titan
1	-	3.5	3.5	4	-	-	-	-	4	3	4	-	4.5	-	-	-	4	3.5	-	-	2	-	-	5	3
2	3	3	3	3.5	-	-	-	-	-	-	-	-	4	3.5	4	3	3	3	3.5	-	-	-	-	-	-
3	-	4	5	5	4	-	3	-	3	-	-	-	-	-	-	-	4	5	5	4	-	3	-	3	-
4	4	-	4.5	4	-	3	-	-	3	4.5	-	-	-	-	-	4	-	4.5	4	-	3	-	-	3	4.5
5	-	-	-	-	4	3.5	4	4	-	-	-	-	-	-	-	-	-	-	-	3.5	3.5	4	4	-	-
6	-	-	-	-	4	3	4	3	-	-	-	-	-	-	-	-	-	-	-	4	3	4	3	-	-

Figure 3: User recommendations for 25 movies

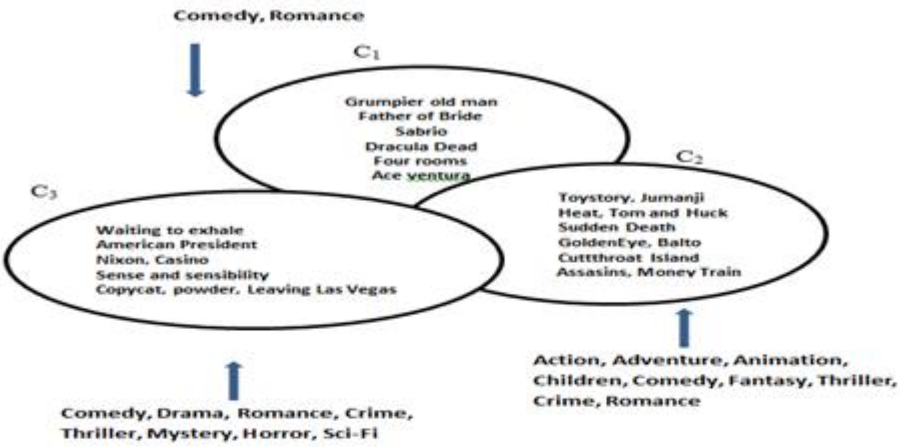


Figure 4: Identified communities based on genres

Table 1: A sample of similarity measures for the target item

Candidate items	Waiting to exhale	American President	Nixon	Leaving Las Vegas	Four rooms	…..
Powder	0.17	0.42	0.22	0.24	0.19	
Casino	0.24	0.19	0.58	0.6	0.13	
Sense and Sensibility	0.2	0.4	0.62	0.4	0.17	

$$P_{u_1,P} = \frac{\sum_{j\in\{W,A,N,L\}} S(P,j)*r_{u_1,j}}{\sum_{j\in\{W,A,N,L\}} |S(P,j)|}$$

$$= \frac{\sum_{j\in\{W\}} S(P,j)*r_{u_1,W} + \sum_{j\in\{A\}} S(P,j)*r_{u_1,A} + \sum_{j\in\{N\}} S(P,j)*r_{u_1,N} + \sum_{j\in\{L\}} S(P,j)*r_{u_1,L}}{\sum_{j\in\{W\}} |S(P,j)| + \sum_{j\in\{A\}} |S(P,j)| + \sum_{j\in\{N\}} |S(P,j)| + \sum_{j\in\{L\}} |S(P,j)|}$$

$$= \frac{0.17*4+0.42*3+0.22*4.5+0.24*5}{|0.17|+|0.42|+|0.22|+|0.24|}$$

$$= 3.9333$$

EXPERIMENTAL STUDY

In the experiments, we looked into the performance of our approach with state-of-art algorithms and traditional content based recommender algorithm. We propose to use the ml-small dataset in MovieLens available through the website (https://grouplens.org/datasets/movielens/). When the framework is applied to this dataset it has discovered six communities and the clustering coefficient is 0.8742. Using the traditional approaches users were recommended the movies which cover almost all genres in the dataset as shown in Table 2.Whereas the users are mostly recommend the genres belongs to the community of the active user in the proposed approach as shown in Figure 5 and comparison of community approach is shown in Table3.

Table 2: Top 5 Recommendations for the first user using traditional algorithms and the proposed approach

User	IBCF		UBCF		Proposed Approach	
	Movies	Genres	Movies	Genres	Movies	Genres
1	Father of Bride Get Shorty Heat copycat Cutthroat	Action Adventure Romance Comedy Crime Thriller Horror Mystery Thriller Drama	Money Train Get Shorty Father of Bride Powder Sudden Death	Action Drama | Sci-Fi Comedy Crime Thriller Drama	Father of Bride Heat Casino Sense and sensibility Get Shorty	Comedy Drama Romance Crime Thriller Mystery Horror | Sci-Fi

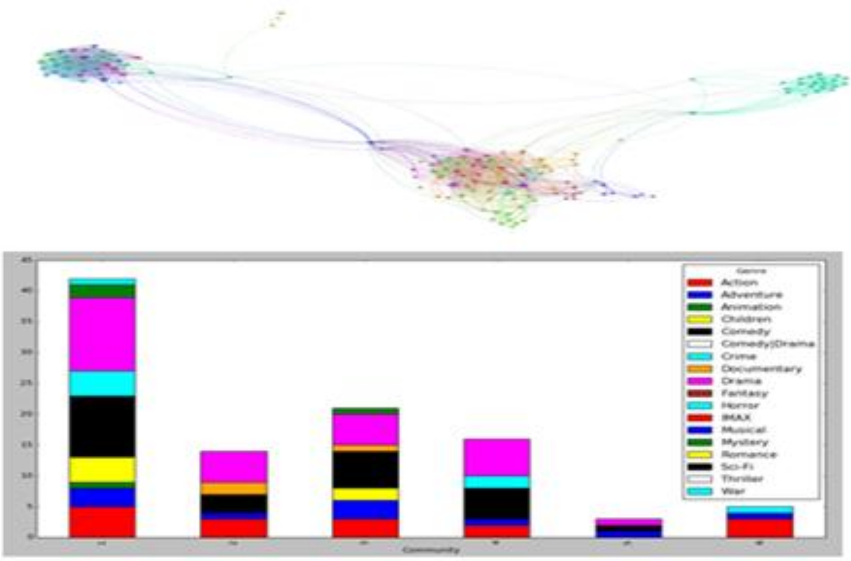


Figure 5: Genre distributions for all communities

Table 3: Comparison

Method	Precision	Recall
Community based	0.672	0.812
IBCF	0.587	0.723
UBCF	0.432	0.532

Conclusion And Future Work

In this paper we propose community based content recommender system which provides fast and accurate recommendation services to users. The novelty of this work is to incorporate the static community detection approach in the context of recommender systems. To achieve that, we have firstly created a feature matrix with columns representing genre name and a binary value indicate that genre was present in each movie. User profile gets created by applying the dot product operation on feature matrix and rating matrix. Secondly, the network interactions collected from the co-occurrences of the genres and apply community detection algorithm to reveal the tight sub-structures having similar features. Finally, the predictions for the candidate items are calculated from the active users' high-level semantic genre features community. The proposed approach has shown a significant progress in accuracy when compared with state-of-the-art recommendation approaches. Our work can be extended to explore low-level visual features than typical expert annotation method.

REFERENCES

[1] Fatemi, Maryam, and Laurissa Tokarchuk. "A Community Based Social Recommender System for Individuals & Groups." *Social Computing (SocialCom), 2013 International Conference on*. IEEE, 2013.

[2] Tang, Jiliang, Xia Hu, and Huan Liu. "Social recommendation: a review." *Social Network Analysis and Mining* 3.4 (2013): 1113-1133.

[3] Moradi, Parham, et al. "A trust-aware recommender algorithm based on users overlapping community structure." *Advances in ICT for Emerging Regions (ICTer), 2016 Sixteenth International Conference on*. IEEE, 2016.

[4] Wang, Yingjie, et al. "A trust-based probabilistic recommendation model for social networks." *Journal of Network and Computer Applications* 55 (2015): 59-67.

[5] Ma, Hao, et al. "Recommender systems with social regularization." *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011.

[6] Sahebi, Shaghayegh, and William W. Cohen. "Community-based recommendations: a solution to the cold start problem." *Workshop on recommender systems and the social web, RSWEB*. 2011.

[7] Abdrabbah, Sabrine Ben, Raouia Ayachi, and Nahla Ben Amor. "Collaborative Filtering based on Dynamic Community Detection." *Dynamic Networks and Knowledge Discovery* (2014): 85.

[8] Yuan, Xiumei, and Pujun Wu. "Content-Based Recommendation Model in Micro-blogs Community." *Management of e-Commerce and e-Government (ICMeCG), 2012 International Conference on*. IEEE, 2012.

[9] Hahsler, Michael. "Developing and testing top-n recommendation algorithms for 0-1 data using recommenderlab." *NSF Industry University Cooperative Research Center for Net-Centric Software and System* (2011).

[10] Suchal, Ján, and Pavol Návrat. "Full text search engine as scalable k-nearest neighbor recommendation system." *IFIP International Conference on Artificial Intelligence in Theory and Practice*. Springer Berlin Heidelberg, 2010.

[11] Ahn, Hyung Jun. "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem." *Information Sciences* 178.1 (2008): 37-51.

[12] Harenberg, Steve, et al. "Community detection in large-scale networks: a survey and empirical evaluation." *Wiley Interdisciplinary Reviews: Computational Statistics* 6.6 (2014): 426-439.

[13] Deshmukh, Dipika, and D. R. Ingle. "A Community Detection and Recommendation System."

[14] Lalwani, Deepika, Durvasula VLN Somayajulu, and P. Radha Krishna. "A community driven social recommendation system." *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE, 2015.

[15] Bok, Kyoungsoo, et al. "Social group recommendation based on dynamic profiles and collaborative filtering." *Neurocomputing* 209 (2016): 3-13.

[16] Qiang, H., Yan, G.: A method of personalized recommendation based on multi-label propagation for overlapping community detection. In: the 3rd International Conference on System Science Engineering Design and Manufacturing Informatization, 1, pp. 360–364. October (2012)

[17] G. Zhao, G., Lee, M. L., Hsu, W., Chen, W., Hu, H.: Community-based user recommendation in uni-directional social networks. IN: the 22th ACM international conference on Conference on information knowledge management, pp. 189–198. October 2013.

[18] Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." *Recommender systems handbook*. Springer US, 2011. 73-105.

[19] Alvari H, et al. (2014), Community detection in dynamic social networks: A game-theoretic approach, In Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM International Conference on, pp. 101-107.

[20] Aston, N. and Hu, W. (2014), Community Detection in Dynamic Social Networks, Communications and Network, Vol.6, No. 02, pp. 124-136.

[21] Aston, N., et al. (2014), Overlapping community detection in dynamic networks. Journal of Software Engineering and Applications, Vol. 7, No. 10, pp.872-882.

[22] Xie, J., and Szymanski, B. K. (2013), Labelrank: A stabilized label propagation algorithm for community detection in networks. In IEEE 2nd workshop on Network Science, pp. 138-143.

[23] Raghavan, U. N., et al. (2007), Near linear time algorithm to detect community structures in large-scale networks, Physical Review E, Vol. 76, No. 3, p.036106.

[24] Angadi, A., & Varma, P. S. (2015). Overlapping community detection in temporal networks. *Indian Journal of Science and Technology*, *8*(31).

[25] Angadi, A., & Varma, P. S. (2016). Finding Hubs and Outliers in Temporal Networks. *Indian Journal of Science and Technology*, *9*(20).