A Community Based Social Recommender System for Individuals & Groups

Maryam Fatemi, Laurissa Tokarchuk School of Electronic Engineering and Computer Science Queen Mary University of London London, UK maryam.fatemi@eecs.qmul.ac.uk, laurissa.tokarchuk@eecs.qmul.ac.uk

Abstract- Social networks provide rich new types of data that promise to personalize and improve Recommender Systems. One promising type of data extracted from social networks is relationships. While existing recommenders have capitalized on the use of explicit relationship data such as friendships, the use of implicit relationship data (that gathered from less obvious connections such as co-occurrence) offers great potential for diversifying recommendation for both individuals and groups. Our novel Community Based Social Recommender System (CBSRS) utilizes this new social data to provide personalized recommendations based on communities constructed from the users' social interaction history with the items in the target domain. It allows quick and efficient recommendations to groups as well as individuals. We propose and evaluate such approach using the Internet Movie database (IMDb). We use the underlying social network graph of the movies based on their common reviewers to model the generic network of interests. Communities are then discovered and used as a basis to provide extensive and diverse recommendation for one user, couples or group of users (friends, family, co-workers) offering them movies of their common interests. Finally we demonstrate that the proposed CBSRS increases the accuracy in the results whilst it addresses the cold sparsity of data and start problem for new users and items and provides recommendation for both individuals and groups.

Keywords-Recommender Systems, Online Social Networks

I. INTRODUCTION

Advances in technology have led to the creation of a new generation of services such as wikis, blogs, social networks and social media websites such as Google+, Facebook, YouTube, Wikipedia, Twitter, Amazon, and Flickr. The information available is invaluable but the sheer amount of it (information overload) and its ambiguity due to multiple formats, limits its usage. Recommender systems have been proposed to address this problem by filtering the relevant data and suggesting items that are closer to the interests of the users. They are successfully used in commercial websites such as Amazon, iTunes, Netflix, etc. Typically traditional recommender systems use either the content of items (content based) or the collaboration between users and items such as rating (collaborative based) or a combination of them. However due to the nature of data they use, they all suffer from one or more weaknesses such as cold start, sparsity of data and overspecialized recommendations. Furthermore the general assumption is that users of a domain are independent and identically distributed [14] whereas in reality, people turn to friends they trust for suggestion. Hence, in traditional recommender systems the important social interactions or connections among users are ignored and thus their

recommendations are somewhat unrealistic. The new types of data available in social networks can play an important role in application development, information dissemination and knowledge discovery and could potentially address the shortcomings associated with traditional recommender systems.

This paper proposes a state of the art Community Based Social Recommender System (CBSRS) that utilizes the explicit relationships between users and items to infer the implicit links between the items. Due to being community based, the CBSRS addresses five different challenging recommendation scenarios:

- 1. It provides serendipitous recommendation, overcoming the overspecialized recommendations issue.
- 2. It delivers recommendation to a new user and thus overcoming the cold start problem associated with new users.
- 3. It partially addresses the cold start problem for new items, recommending new items after only having one review.
- 4. It generates recommendation with even a relatively small set of data and thus overcoming the sparsity issue.
- 5. Finally it presents recommendation to groups of users.

CBSRS is applied on the social network of movies in IMDb (Internet Movie Database). The social network is constructed based on the network of reviewers for movies. The resulting network consists of strongly connected sub-structures, called communities. These structures are important as nodes belonging to a community tend to have very similar properties.

The paper is organised as follow: related works and community detection algorithms are provided in section II. Analysing of the online social network of movies and its communities are discussed in section III and IV. CBSRS algorithm is explained in section V and results are evaluated in section VI. We conclude in section VII.

II. RELATED WORK

Traditional collaborative based approaches predict users' interest based on their rating history [7] whilst social recommender systems exploit the social interaction between the users i.e. friendships [6]. Trust based recommenders e.g.[15], assume that users' taste are most likely similar or influenced by their trusted friends and so utilize the inferred implicit or observed explicit trust information to further improve recommendations. Their primary limitation is that they are based on the supposition that users have similar tastes with other users they trust which is not always true since tastes of one user's friends may vary significantly. Therefore, trust-aware recommenders cannot be directly applied to generate results in social recommenders and in reality only a few online systems e.g. Epinions, have implemented the trust mechanism. Circle based recommender is an example of trust based system which uses the different trust circles a user may form in his online social networks [12] and [10] uses the social network information to design two social regularization terms so friends with dissimilar tastes are distinguished in the social regularization terms. As it utilizes all the social connections of each

user, the existence of some social connections may hurt the recommender performance.

Social network information has been widely studied in many other research tasks. In [8] a generic framework for incorporating social context information (authors' identities and their social networks) has been proposed by adding regularization constraints to the text-based predictor to predict review quality. Other social recommenders include [5, 6, 14, and 16]. Although the method in [16] is called social recommendation, it essentially does not utilize any social network information but explores similar users to generate recommendations. [14] is actually a trust aware method as it utilizes trust information in its analysis whereas [5] only utilizes very simple heuristics in making recommendations and [6] uses neighbourhood information to generate results. CBSRS presented here exploits the new dimension of data which is the social interactions that exists between the entities of a domain. This novel data (relationships between users and items) allows the algorithm to build a social network and interpret users' interest based on their link with other similar users, the community they belong and some network /item features. This view on social recommendation is novel and none of the previous work discussed is capable of addressing it.

Communities, sub-structures are one of the important features derived from social networks and are widely defined as group of nodes which are more densely tied to one another than any other part of the network. Partitioning of nodes into subgroups is important in understanding the likely behaviour of the network as whole [1]. Graph clustering approaches, hierarchical clustering, and modularity-based methods [3] are a few examples of method to identify communities. One of the popular methods widely used is by Palla et al [2] which identifies the overlapping communities based on clique percolation. Lancichinetti et al. [3] agglomerative algorithm also finds overlapping and hierarchical communities based on the local optimisation of a fitness function. Blondel et al method Louvain [4] is also widely applied and is based on the modularity optimisation and can analyse significantly larger networks very accurately with considerably less computational time. First it finds small communities by optimising modularity using local information and secondly it aggregates small communities. These steps are repeated until maximum modularity optimisation is achieved.

Similar to our work, [9] uses a simple community detection method to address the cold start problem associated to new users. The modularity-based community detection method identifies the communities of users in different social network domains and incorporates them to predict the rating of items in another domain where the user has few rating history. The social network is based on friendship relationships explicitly chosen by each user. Traditional collaborative based recommender is used to calculate the predicted rating of the items within each community. By using data from different domains the authors argue that some previous history for the users are imported and therefore this could resolve the cold start problem. However this method is partial as the objects in different domain could be different, communities formed may have different structure and a user may have irrelevant or no other history in other domains. Similarly, YouTube recommender system proposed by [11] is based on community detection and the information extracted from the network of co-commentator. The YouTube Recommender Network (YRN) is built using only 25 comments per videos. This method uses the co-reviewing relationships but its very limited as the YRN does not reflect the whole structure of the YouTube. Furthermore the recommendation is based on the degree and neighboring nodes and thus it has the potential to provide repetitious and overspecialized recommendations of nodes with high degree and ignores other nodes with a lower degree such as new additions.



Figure 1. Degree Distribution in IMDb Network of Movies.

III. BUILDING THE SOCIAL NETWORK OF MOVIES

CBSRS for the movies is applied on the Internet Movie Database (IMDb). Information from IMDb was in collected in Nov 2011 and updated in 2013. In total 21,894 movies and 276,795 reviews from 133,318 users were used. More than 40% of users had more than one movie reviewed, and average reviews per movies were 12.64. In social network of IMDb, movies are presented as nodes and are linked if at least one reviewer has commented on both of them. The edges are weighted by the number of common reviewers between a movie pair; e.g. if 6 unique reviewers had commended on them then the weight would be 6. For a reviewer R who has reviewed a set of movies $M, R(M) = \{M_1, M_2, \dots, M_n\}$, there exists n(n-1)/2 possible links amongst the n movies. Unlike previous studies, the novelty in this works is that the nodes are not the users but are items and hence communities are groups of items (movies) and the social interaction of the users are the linking factor. For 21,894 movies, 3,860,809 edges exist. The visualisation and some of the measurement was performed using Gephi (http://gephi.org). Some movies only had one reviewer and hence were not connected to the core. These outliers were eliminated due to their low popularity and isolated position. The remaining set contained 7,474 movies and 2,380,762 edges.

IV. NETWORK MEASUREMENTS AND ITS COMMUNITIES

Many different measurements can be carried out on a network; here we limit the measurement to a few relevant metrics to identify the influential characteristics of the network. Networks which follow the power law distribution share a common structure of a relatively small amount of nodes connected by large number of edges that are not random. Other important characteristics are whether the resulting network is considered to be small-world and scale-free. A network is small world if most of the nodes who are not neighbours can be reached from every other node by a small number of steps. Usually small-world networks have a small diameter and high clustering. Scale-free networks are a group of power-law networks where the high-degree nodes tend to be connected to other high degree nodes. A high clustering coefficient suggests that the network has tightly knit groups. The resulting network follows the power-law distribution with $\lambda = 2.089$, Figure 1. The Average Path Length is 2.181 and Network Diameter, 6; considerably small for the size of the network. The Clustering Coefficient is 0.828, which is much higher than an equivalent random network. These observations confirm that the network is small-world and scale-free. This is expected, as users who reviewed movies have similar interest and thus the movies have many similar features. In general the observed result suggests that the structure of the IMDb network of movies is a social graph with a number of small tightly clustered communities with high degree nodes. The resulting network of movies is considered as a generic network of the interests of the users in IMDb and is used to identify interest based communities.

Communities are very important because nodes belonging to the same community tend to have similar properties. The community



Figure 2. Genre Distribution in the Top 9 Largest Communities.

membership of movies forms the basis for recommendation in CBSRS proposed here. It allows for diverse recommendations whilst still constraining them within the interest map of each community. The results are more effective as compared to using tags or item similarities which tend to be static, not standardized and often too specialized. Louvain method [4] is used to identify communities as previous work in [13] indicates that it can provide the most accurate results for the large dataset and communities have better structures and diversity. A total of 3718 overlapping communities were identified with the maximum modularity of Q = 0.249. The low Q is due to the large number of the users and their diverse interests in the movies. The largest community discovered, had 513 movies and the smallest only 1. The top three movies with the highest degree in the largest community were: "Casino Royale(2006)", "The Dark Knight(2008)" and "Batman Begins(2005)" and top three genres were Thriller followed by Adventure and Action. Interestingly "'G' Men(1935)" had the highest betweenness centrality followed by "Avatar(2009)" and "The Searchers(1956)". All communities contained diverse set of genres and therefore recommendations using these communities have the potential for serendipitous recommendations. While user based recommendations are limited to only other users' rated items, the CBSRS can produce diverse results covering more items. On average the communities had 10.5 unique genres. A genre analysis graph for the top 9 largest communities of the IMDb social network of movies is shown in Figure 2.

V. THE COMMUNITY BASED SOCIAL RECOMMENDER System for Movies

Movies are recommended based on their community membership, the degree utility value, adjacent nodes, and star ranking. Degree utility value determines how important or popular a node is in a network whereas adjacent nodes shows how close match the two movies are in term of interest of the reviewers and the strength of their linked is determined by the weight of their link (number of the common reviewers). The degree of a node also shows the number of reviews it received and thus its popularity and importance. Star ranking also indicates how good a movie is according to reviewers' point of view. Therefore when CBSRS is generating the recommendations, it considers the following factors: the popularity of a movie within the community (utility value), the closest match of interests of reviewer to that movie (adjacent nodes) and the overall quality of a movie (star ranking).

<u>Community Membership</u>: is defined as a set of all communities a node is a member of:

 $L_c(m) = \{C_1, ..., C_{n_m}\}$

A node can be member of more than one community. The main community $C_{M}(m)$ is defined as the community in which the movie *m* has the highest aggregated sum of edge weights and degrees:

$$C_{M}(m) = \forall (c_i \in L_c(m) \& x \in c_i) :$$

$$\max_{1 \le i \le n_m} (edge(x_{ci}) + degree(x_{ci}))$$

<u>Degree Utility Set:</u> is the sorted list of movies based on their degree utility values. Degree utility value is defined as degree of the nodes in the community to the total degrees of all nodes in the network:

$$u(x) = \frac{degree(x)}{\sum_{j=1}^{N} degree(x_j)}$$
total number of nodes in the network

N·

The higher the utility value the more important the node is. And degree utility list for community C_i , $L_U(C_i)$ is list of the movies x_j in the community C_i , sorted according to their degree utilities $u(x_j)$ as:

$$L_U(C_i) = \{ x_{1_u}, ..., x_{n_u} \}, Where : x_j \in C_i$$

<u>Ranked Adjacency Set:</u> is defined as a set of nodes which are linked by one edge within the same community. The ranking is determined by the weight of the edge connecting them.

$$L_{A}(m) = \{x_{1_{a}, \dots, x_{n_{a}}}\},$$

Where : $(m \& x_j \in C_i) \& (Edge(m, x_j) = 1)$

The nodes adjacency of a movie is important as it indicates that the two connected movies are of the same interest to the reviewers. Therefore the larger the weight, the higher the strength of connection between the two nodes.

<u>Star Ranking Set</u>: is the sorted list of movies within the same community C_i based on the average user ranking $r(x_i)$ for each movie. It is used to decide which movie has a higher quality and popularity according to the reviewers' point of view.

$$L_{R}(C_{i}) = \{x_{1_{i}}, ..., x_{n_{i}}\}, Where : x_{j} \in C_{i}$$

It is logical to recommend the adjacent nodes plus other prominent nodes by the merit of their high degree utility value in the community of the specific movie m. The final recommendations, Mis a set of 10 movies from the combined lists of above. It consists of movies in set M_1 and M_2 , where M_1 is the top 10 common movies from all the lists and M_2 is empty unless number of movies in M_1 is less than 10, when M_2 is then populated. For the simplicity it is decided to choose an even number of movies from each list, to ensure that a diverse set of movies is chosen for recommendation.

$$M = M_1 \cup M_2$$

$$M_1(x_j) = \{ Top_{10} x_j (L_U(C_M(m)) \cap L_A(m) \cap L_R(C_M(m))) \}$$

 $\begin{aligned} M_2 = \{ \ Top_z x_j(L_U(\overline{C_M(m)})) \ U \ Top_z x_j(L_A(m)) \ \cap \ Top_z x_j(L_R(\widehat{C_M(m)})) \ \}, \\ where: \ z = (10 - Len(M_l)/3) \ \& \ (x_j \not\in M_l) \end{aligned}$

As each set represents one dimension, (popularity, closer interest and quality) choosing different number of movies from each set would dominate recommendations towards the features associated with that list. However whether the users would prefer more diverse recommendations including less popular movies to recommendations closest to their interest with no diversity in the items or would simply prefer quality to the popularity, is a study that is very much relative to users taste. Future works will investigate the effect of choosing different number of movies from each list and its effect on the recommendations according to users' perspectives.

CBSRS addresses three different challenging scenarios: recommends items (movies) to a new, recommends to a user with previous reviewing history or recommends items to group of users. Each of these scenarios will be discussed in the sections below.

A. Recommendation to a New User

Suppose user R with no reviewing or ranking history is browsing movie *m* on IMDb. Traditional collaborative based recommender would not be able to provide recommendation. Similarly content based recommender could only provide limited recommendation only if *m* had enough information to match it to similar movies. CBSRS though provides a set of *M* diverse movies independent of *m* or *R* and only based on the generic network of interests previously mapped, hence addressing the cold start problem for new users. It first identifies the main community of movie *m*, $C_M(m)$, the community in which *m* has the highest aggregated sum of edge and degree. Then for all the movies in $C_M(m)$ the degree utility list



Figure 3. Set_1 Communities and Genre Distributions.

 $L_U(C_M(m))$, ranked adjacency list $L_A(m)$ and $L_R(C_M(m))$ are populated. The final recommendation set *M* is then prepared.

B. Recommendation to a User with History

CBSRS produces results for any user regardless of the number of movies he previously reviewed, as even with as little as one review it is enough for the algorithm to start recommending. Suppose that $P_R(x)$ is defined the set of x_1 to x_n movies previously reviewed by user *R*, then for all the movies in $P_R(x)$, communities are listed as:

 $L_C(x) = \{C_1, C_2, \dots, C_{n_R}\}, where x \in P_R(x)$

The main community for *R* is defined as the community which contains the highest number of movies previously reviewed by *R* and the highest aggregated sum of edge weights and degrees. Finally for all the movies in $C_M(x)$ the degree utility list $L_U(C_M(x))$ and star ranking list $L_R(C_M(x))$ are populated. The ranked adjacency list $L_A(x)$ is then populated as all movies adjacent to movies previously reviewed by *R* in the main community, sorted according to their weights:

 $L_A(x) = \{x_1, ..., x_z\}$, where $(x_j \in P_R(x) \& x_j \in C_M(x))$

CBSRS will then prepare list M containing 10 movies from the three sets calculated above.

C. Recommendation to Groups of Users

CBSRS also provides recommendation to group of users up to ten. For groups larger than ten the number of recommended movies in *M* would be increased proportional to the number of users so as to capture the interest of all the users in the group. This is a novel approach in recommending to groups of users. For users R_1 to R_k , first using the same approach as explained in section V.B, the main communities for each users are identified as $C_{MRI}(x)$ to $C_{MRk}(x)$. Then k separate lists of degree utility lists of $L_{UI}(C_{MRI}(x))$ to $L_{Uk}(C_{MRk}(x))$ and ranked adjacency lists of movies within the main community are prepared. The final recommendation M is the top 10 common movies from all the lists. If M is less than 10 movies then the remaining movies is chosen equally from each lists. This will ensure that a diverse set of movies are chosen to accommodate all the tastes of the users within the group and will guarantee that the recommendation is not limited or biased towards the dominant or active users but also takes into account the interests of the less active users with perhaps fewer reviewed items.



Figure 4. Set_2 Communities and Genre Distributions.

VI. COMPARING RESULTS OF CBSRS WITH TRADITIONAL RECOMMENDER SYSTEMS

In this section the recommendation provided by the proposed method is compared with the traditional recommender systems, item and user based collaborative filtering. Two sets of 30 random users and their reviewed movies are chosen. Set 1 had 544 movies, with 29,837 edges and Set 2 had 415 movies with 13,677 of edges. Louvain community detection method was applied on both sets and it discovered 10 communities for each sets. The largest community in Set 1 had 162 members and the smallest only 2 movies and Set_2 largest community had 117 members and smallest had 2 isolated movies. Their clustering coefficients were 0.943 and 0.949, network diameters 4 and average shortest paths 2.007, 2.291, respectively. This exercise was performed to demonstrate that even though the two sets of 30 users were chosen randomly and number of users in each set is very small, they are still highly connected and their movies forms communities. It is further shown that CBSRS outperforms other recommender systems by providing recommendations with even the small set of data and random users with mixed reviewing history thus overcoming the sparsity issue. The results are compared with the two well-known recommender systems for both sets of data Set 1, Set 2 and the whole network.

Figure 3 and figure 4 show the network of movies based on the common reviewers for both Set 1 & 2. Genre distribution analysis of the communities for both sets shows that the larger communities share a good diversity in the movies they have and are not limited to one genre even for a small set of users with mixed reviewing history from 2 to 177 in Set 1 and 2 to 127 reviews in Set 2. This further confirms that CBSRS provides diversified recommendations not limited to one category. On average communities in Set 1 had 11 genres and communities in Set 2 had 11.2 genres. Two typical collaborative filtering recommenders from the Apache Mahout library (http://mahout.apache.org/), user based and item based, were used to provide the top 10 recommendations for users in Set 1, Set 2 and for all users and the outcome was compared with CBSRS. User based recommender provides recommendations based on how similar items are to items based on the items a user have already. Then more similar items are recommended, whereas item based recommender looks at how similar users to users are based on the items they showed interest in then items from similar users are recommended. The user based recommender with Tanimoto

Table 1. Evaluation Metrics for Set 1, Set 2 and All Users.

| Method | Precision | Recall | nDCG |
|------------------------|---------------|---------------|---------------|
| | Set_1= 0.655 | $Set_l = 0.8$ | |
| CBSRS | Set_2=0.540 | $Set_2=0.7$ | - |
| | All=0.855 | All=0.912 | |
| | Set_1= 0.0142 | Set_1= 0.0142 | Set_1= 0.0271 |
| Item Based Recommender | Set 2=0.04 | Set 2=0.04 | Set 2=0.028 |
| | All=0.109 | All=0.109 | All=0.117 |
| User Based Recommender | No results | No results | No results |

coefficient similarity (Jaccard coefficient) was applied on the datasets. This method did not provide any recommendation for the users in Set_1, Set_2, even though the number of movies involved was 544, 415 in Set_1 & 2 respectively. One reason that the algorithm provided no result was because users were chosen purely at random with different number of reviews from 2 to 177 in Set_1 and 2 to 127 in Set_2, and the set of data was too small to find similar users. It also could not any provide recommendations on the whole dataset. Item based recommender with Log Likelihood similarity was then applied to the Set_1 & 2. This method could provide recommendations from θ to 10 items for the users in Set_1, Set_2 and all users. Here we compare the result of item based recommender to the result of CBSRS.

Using the small 60 users set (two sets of 30 users), the total number of unique movies recommended for both sets was 238. Ideally it was expected that given the number of recommendations, at least half of the results would be unique. Also the algorithm provided recommendations only for 83% of the users and none for the remaining 17%. Evaluating the recommendations on the whole dataset, it could not provide recommendations for 28% of users and for the 72% not the 10 items as some users had only 1. The top most reviewed movie by 8 users out of 30 in Set 1 were: "Inception" followed by "Machete", "Avatar", "The Dark Knight", "Transformers: Revenge of the Fallen" and "Prince of Persia: the Sands of Time". Similarly the top most reviewed movie by 7 users in Set 2 were: "Star Wars: Episode III-Revenge of the Sith" followed by "Batman Begins", "Avatar", "Kill Bill: Vol. 1"," The Lord of the Rings: The Return of the King" and "Troy". For evaluation, in each set, 80% of data was used as the training set and the remaining 20% as the test set. Precision, recall and nDCG on data of Set 1, Set 2 and all users' dataset are listed in Table 1. Four random users from Set 1 & 2 with their reviewed movies are chosen for a more detailed analysis here. Due to space limitation of this paper four users were chosen with rather small set of movies. U2 and U11 are from Set 1 and U52 and U56 are from Set 2:

- P₁₂={ Snakes on a Plane (2006), The Fast and the Furious: Tokyo Drift (2006), Torque (2004)}, Genres(Action, Thriller, Crime).
- *P*_{UII}={ Red River (1948), Rob Roy (1995), Shanghai Noon (2000), The Treasure of the Sierra Madre (1948), Willow (1988)}, Genres(Action, Adventure, Western).
- P_{U52}={ Apocalypto (2006), Blood In, Blood Out (1993), Fist of Golden Monkey (1983), Chou (2001), The Wild Bunch (1969)}, Genres(Action, Drama, Western).
- *P*_{U56}={ An Innocent Man, Spartacus, The Towering Inferno (1974)}, Genres(Action, Drama, Thriller).

The result of item based recommendations is listed in Table 2. For users U11 and U56 there was no recommendation, with the small sets. For U2, the preference was on movies with genres of Action and Thriller and the recommendation was also more on Action, Adventure and Western and the recommendation was more on Action, Adventure, Comedy, and Fantasy and less on Thriller. Even though the number of reviewed items for U11 was greater than U2 and in more diverse categories, item based recommendarions for the 4 users are also listed in Table 2. Looking at the genres of the

Table 2. Item Based Recommendation.

| User | Using Set_1 & Set_1 | 2 | All Users Set | |
|------|--|-------------------------------|---|----------------------------------|
| User | Movies | Genres | | Genres |
| U2 | Assault on Precinct 13 (2005) Daybreakers (2009) Fearless G.I. Joe: The Rise of Cobra Highlander (1986) Robin Hood (2010) Street Kings The Warriors Unstoppable Wanted (2008) | Action Thriller Drama | Die Hard 4.0, Superman Returns, Stealth, Eon Flux, Poseidon, V for Vendetta (2005), Beowulf (2007), The Sentinel, 16 Blocks, Fantastic 4: Rise of the Silver Surfer | Action Thriller Adventure |
| U11 | No Recommendation | - | The Fugitive Gladiator Goldfinger The Poseidon Adventure Star Wars: Episode IV - A New Hope Indiana Jones & the Temple of Doom The Terminator Aliens Predator Indiana Jones and the Last Crusade | Action Adventure Thriller |
| U52 | Astro Boy (2009), Balls of Fury (2007) Beowulf (2007) Dragonheart (1996) Knight and Day (2010) Machete (2010) Pineapple Express (2008) Rocky Balboa (2006) Snakes on a Plane (2006) Timeline (2003) | Action Adventure Comedy | Out for Justice 3:10 to Yuma The Searchers Kelly's Heroes The Crazies The Osterman Weekend The Pope of Greenwich Village Nine Deaths of the Ninja Big Trouble in Little China | Action Drama Adventure |
| U56 | No Recommendation | - | War of the Worlds A Night to Remember Magnum Force The Spy Who Loved Me Moonraker For Your Eyes Only First Blood Indiana Jones & the Temple of Doom Rambo: First Blood Part II Cobra | Action, Thriller Adventure |

recommended movies they are all of the similar genres even though the users reviewed movies with different genres. Unlike the traditional item based recommender, CBSRS recommends 10 movies to 100% of users in all 3 sets. Number of unique items recommended to the 60 random users in Set 1 & 2 was 453, indicating the results were a closer match to the interest of the users rather than a general recommendation for users of seemingly similar taste. The main community for each user is identified as discussed in V.B and three lists of degree utility set, ranked adjacency set and star rating set are populated. The final recommendations are listed in Table 3. It is noted that result contains movies from diverse set of genres while including movies from the same interest of the users. The recommendations for U11 and U56 using CBSRS on Set 1 & 2 and item based recommender using all users' dataset have two and one common movie recommended (no results when small set was used). Recommendations for the other two users were not similar.

CBSRS is also capable of providing recommendation to a group for users based on their common interest as explained in Section V.C. CBSRS was applied to 60 groups of two and three users randomly chosen from Set 1 & 2. In all groups CBSRS was successful in providing 10 recommendations. An example of such groups is listed in Table 4, for the group of users consisting of U2 and U56. The traditional collaborative filtering methods such as the item based can only provide recommendation for one user and cannot recommend items to a group of users. Our novel CBSRS will always provide recommendation even if the users may not share any common reviewed items. This is due to its community based nature and the fact that it uses the generic network of interests. CBSRS also always provide recommendation to anonymous users who may be browsing a specific movie. As an example suppose that a user is browsing the information of a popular movie "The Dark Knight", CBSRS would follow the steps discussed in section V.A. and produces results. First the main community for the movie is identified and then the three lists of degree utility set, adjacency set and star rating are populated. The final recommendation is provided

Table 3. CBSRS Recommendation Results

| | Using Set_1 & Set_2 | | All Users Set | |
|------|--------------------------------|------------------|------------------------------|-------------------------------|
| User | Movies | Genres % | Movies | Genres % |
| | Avatar (2009), | | Transformers, | |
| | RoboCop (1987), | | V for Vendetta, | Action=33.33 |
| | Machete (2010), | Action=32.14 | The Bourne Ultimatum, | Thriller=20.0 |
| | The Expendables, | Thriller=21.43 | The Incredible Hulk, | Sci-Fi=13.33 |
| 112 | Terminator Salvation (2009), | Sci-Fi=14.29 | Wanted, | Adventure=10.0 |
| 02 | V for Vendetta (2005), | Adventure=10.7 | Tropic Thunder, | Fantasy=10.0 |
| | Hellboy II: The Golden Army, | Fantasy=10.71 | Pirates of the Caribbean: At | Crime=6.67 |
| | The Warriors (1979), | Drama=7.12 | World's End, | Comedy=3.33 |
| | Highlander (1986), | Crime=3.57 | Rambo, | War=3.33 |
| | Assault on Precinct 13 | | Apocalypto | |
| | Death Race (2000), | | Saving Private Ryan, | Action -2.22 |
| | Star Wars: Episode IV, | | Wild Wild West, | Thrillor=20.0 |
| | Mad Max (1979) , | Action -21.25 | Casino Royale, | A dwantura=10.0 |
| | Escape from New York, | Action -31.23 | The Adventures of Robin | Crima=10.0 |
| | The Terminator (1984), | Adventure=15.62 | Hood, | Drama=10.0 |
| U11 | Terminator 2: Judgment Day, | Thrillor=15.62 | The Towering Inferno, | Eantaci - 2 22 |
| | The League of Extraordinary | Crime=6.25 | Patriot Games, | History=3.33 |
| | Gentlemen (2003), | Eantacy =6.25 | The Crow, | Pomance=3.33 |
| | Casino Royale (2006), | Sport=3.12 | Volcano, | Sci-Fi=3 33 |
| | Die Another Day, | 5pon 5.12 | Payback, | War=3 33 |
| | Conquest of the Planet of Apes | | Star Trek V | Wai 5.55 |
| | The Island, | Action =33.33 | Saving Private Ryan | Thriller=16.13 |
| | The Lord of the Rings: The | Thriller=20.0 | Star Wars: Enisode I - The | Sci-Fi=12.90 |
| | Return of the King, | Crime=13.33 | Phantom Menace. | Adventure=9.68 |
| | Sin City (2005), | Adventure=6.67 | Wing Commander. | Drama=6.45 |
| | 300 (2006), | Fantasy=6.67 | Godzilla. | Horror=6.45 |
| U52 | Casino Royale (2006), | Sci-Fi=6.67 | Armageddon. | Comedy=3.23 |
| | V for Vendetta (2005), | Comedv=3.33 | Vampires. | Fantasy=3.23 |
| | Die Hard 4.0, | Drama=3.33 | Starship Troopers, | History=3.23 |
| | The Bourne Ultimatum, | History=3.33 | Deep Impact, | Romance=3.23 |
| | Shoot 'Em Up, | War=3.33 | Tropic Thunder | War=3.23 |
| | war of the worlds (2005) | 1 1 21 25 | - C : P 1 | western=3.23 |
| | Star Wars: Episode III - | Action $= 31.25$ | Casino Royale, | Action= 27.03 |
| | Revenge of the Sith, | 1 nriller=18.75 | Quantum of Solace, | Adventure=16.22 |
| | 1 in Farma (1007) | Adventure=9.57 | | 1 nrmer = 10.22 |
| | Air Force One (1997), | Advantura=6.25 | Tomorrow Never Disc | Drama=15.51 Diography=5.41 |
| 1156 | A View to a Kill | Auventure=0.25 | Gladiatar | Crimo=5.41 |
| 030 | Top Gup (1986) | Sci $Fi=6.25$ | Goldfinger | History=5.41 |
| | Lathal Waanan | Eastern 2 12 | Armagaddan | Eastacy=2.41 |
| | The Living Davlights | Comedu=3.33 | Octopusev | $P_{omance=2.70}$ |
| | They Live (1988) | Horror=3.33 | The Adventures of Pohin | Sci $Fi=2.70$ |
| 1 | Loose Cannons | Romance=3.33 | Hood | $W_{estern=2.70}$ |

by aggregating the lists. The result is then compared to that of the Amazon (data retrieved in Feb 2013) when viewing the information of "*The Dark Knight*" movie and are listed in table 5. Amazons and other movies database usually recommend movies which are most recent and popular whereas CBSRS looks at the generic interest network taking into account the users' taste. Therefore it can identify both the current trend and the established trends in interests and provide a mixed recommendation.

VII. CONCLUSION

In this paper we proposed a novel Community Based Social Recommender System (CBSRS) which utilizes the new types of data available in social networks. It uses the implicit relationships between the items derived from the direct interactions of the users with them, and depicts the social network graphs of items. More specifically it produces the interest map of users in movies in the form of the generic network of interests and uses it as a basis to detect communities and provide a more accurate and personalized recommendations. It is also demonstrated that due to its nature of being community based, the proposed CBSRS addresses the five different issues other recommender system had failed. Namely, it recommendation, provides serendipitous overcomes the overspecialized and repetitious recommendation results. recommends to a new user and thus can address the cold start Table 4. CBSRS Recommendation Results for Group Users

| Tuble in edoted internation results for Group esers | | | |
|---|---|-----------------|--|
| User | Movies | Genres % | |
| | Avatar (2009), | Action =28.57 | |
| | RoboCop (1987), | Adventure=28.57 | |
| | A View to a Kill, | Thriller=28.57 | |
| | Star Wars: Episode III - Revenge of the Sith, | Drama=28.57 | |
| U2 | V for Vendetta (2005), | Sci-Fi=28.57 | |
| & U56 | The Warriors (1979), | Crime=28.57 | |
| | King Kong (2005), | Fantasy=28.57 | |
| | Blood Work, | Mystery=28.57 | |
| | The Expendables, | Romance=28.57 | |
| | Hellboy II: The Golden Army | War=28.57 | |

Table 5. Top 5 Recommendations from Amazon and the CBSRS

| Movie | CBSRS | Amazon Recommender |
|--------------|---------------------------|--|
| | Avatar (2009), | Batman Begins (2005), |
| The Dark | Iron Man, | The Dark Night Rises, |
| Knight | Quantum of Solace (2008), | Batman Forever, |
| (2008) | Star Trek, | Batman & Robin, |
| | Watchmen | Batman |
| | Casino Royale (2006), | Charlie's Angels, |
| Charlie's | You Only Live Twice, | Miss Congeniality 2, |
| Angels: Full | Goldfinger (1964), | Miss Congeniality, |
| Throttle | Mad Max (1979), | My Super Ex-Girlfriend, |
| | xXx | Lara Croft Tomb Raider: The Cradle of Life |

problem for new users. It also partially addresses the cold start problem for the new items as the new movies can be connected to the generic network of interest after having as little as one review. The sparsity issue associated with recommenders is also addressed here as CBSRS can almost always provide recommendation even with few data available. Finally it provides recommendation to groups as well as individuals. The recommended items have diverse categories whilst they are closest match to the interests of the users, also taking into account the overall popularity and quality perceived of the items amongst all the reviewers of the network. In our experiments on the available data on IMDb, we showed significant improvements over existing approaches in recommender systems.

ACKNOWLEDGMENTS

We would like to thanks the BT Research Centre for their scientific collaboration.

REFERENCES

- P. Pons and M. Latapy. Computing Communities in Large Networks Using Random Walks. In ISCIS '05, Turkey, 2005.
- [2] G. Palla, I. Derenyi, I. Farkas and T. Vicsek. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. Nature, 435:814, 2005.
- [3] A. Lancichinetti, S. Fortunato and J. Kertész. Detecting the Overlapping and Hierarchical Community Structure in Complex Networks. New Journal of Physics, vol. 11 033015, 2009.
- [4] V. D. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre. Fast Unfolding of Communities in Large Networks- Louvain Method. Journal of Statistical Mechanics: Theory and Experiment, P1000, 2008.
- [5] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman. Personalized Recommendation of Social Software Items Based on Social Relations. In RecSys '09, New York, USA, 2009.
- [6] F. Liu and H. J. Lee. Use of Social Network Information to Enhance Collaborative Filtering Performance. Expert System Applications, 37(7):4772–4778, 2010.
- [7] Y. Zhang, B. Cao, and D. Y. Yeung. Multi-Domain Collaborative Filtering. In UAI '10, USA, 2010.
- [8] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting Social Context for Review Quality Prediction. In WWW '10, USA, 2010.
- [9] S. Sahebi and W. Cohen. Community Based Recommendations: a Solution to the Cold Start Problem. In RSWEB '11, USA, 2011.
- [10] H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King. Recommender Systems with Social Regularization. In WSDM '11, Hong Kong, China, 2011.
- [11] S. Qin, R. Menezes and M. Silaghi. A Recommender System for YouTube Based on its Network of Reviewers. In SocialCom '10, 2010.
- [12] X. Yang, H. Steck, and Y. Liu. Circle-based Recommendation in Online Social Networks. In KDD '12, China, 2012.
- [13] M. Fatemi and L. Tokarchuk. An Empirical Study on IMDb and its Communities Based on the Network of Co-Reviewers. In MPM '12, Switzerland, 2012.
- [14] H. Ma, H. Yang, M.R. Lyu, and I. King. SoRec: Social Recommendation Using Probabilistic Matrix factorization. In CIKM '08, New York, USA, 2008.
- [15] M. Jamali, and M. Ester. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks. In RecSys '10, Spain, 2010.
- [16] X. Xin, I. King, H. Deng, and M. R. Lyu. A Social Recommendation Framework Based on Multi-Scale Continuous Conditional Random Fields. In CIKM '09, Hong Kong, China, 2009.