الجامعـــة الإسلاميـــة للتكنولوجيا
UNIVERSITE ISLAMIQUE DE TECHNOLOGIE
ISLAMIC UNIVERSITY OF TECHNOLOGY
DHAKA, BANGLADESH
ORGANISATION OF ISLAMIC COOPERATION

An Improved Cohesion Based Community Recommendation System

Authors

Kazi Wasif Ahmed **Student Id: 104418**
Md. Mamunur Rashid **Student Id: 104420**

Supervisor

Md. Kamrul Hasan, PhD
Systems & Software Lab (SSL)

Department of Computer Science & Engineering (CSE)
Islamic University of Technology (IUT)

A Thesis submitted to the Department of Computer Science and Engineering (CSE), Islamic University of Technology in Partial Fulfillment of the requirements for the degree of Bachelor of Science in CSE (Computer Science & Engineering)

September, 2014

<u>CERTIFICATE OF RESEARCH</u>

This is to certify that the work presented in this thesis paper is the outcome of the research carried out by the candidates under the supervision of **Md. Kamrul Hasan, PhD, Assistant Professor, Department of Computer Science and Engineering, IUT Gazipur.** It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree or any judgment.

Authors

__________________________ __________________________

Signature of Supervisor

____

Signature of the Head of the Department

____

Abstract

Social Networking Sites (SNS) are the dominating entities in the modern web. The importance of social networking sites in our life is increasing day by day as they are attracting millions of users by their interesting features and activities. It enables the researchers to use the information available in these sites. Online Community is appealing to people as they can enjoy sharing their ideas, view, and know about view of other people. At the same time, they are interested in joining different community. However, with the rapid growth of SNS's resulting in information overload people are in dilemmas to choose right community from huge list of available communities and it is also time consuming. Potential choice of communities is influenced by many factors of user behavior and activeness in Social Networking Sites. The recent surge of research in recommendation algorithms is not surprising. But these algorithms have unsatisfactory results in community recommendation because of lack of intuition in judging rational behavior. Many researches are going on this point to find out recommendation system in various ways. To solve this problem, we introduce cohesion based community recommendation system. In this paper we design a general framework of community recommendation based on cohesion after analyzing the present methods of community recommendation. The main idea of the proposed approach is consisted of following stages- measuring friendship factor, measuring user factor, calculating threshold from present communities of user, community recommendation based on threshold, result analysis. We validated our idea on a small network in Facebook.

Keywords: Social networking, Community, Recommendation system, Cohesion, friendship factor, user factor, threshold

Acknowledgement

At the very beginning we express our heartiest gratitude to Almighty Allah for His divine blessings which allowed us to do this research work to life.

We are grateful and indebted to our supervisor **Md. Kamrul Hasan, PhD,** Assistant Professor**,** Department of Computer Science and Engineering, IUT. Due to his supervision, knowledge and relentless support allowed us to complete this endeavor successfully.

We are thankful to **Prof. Dr. M. A. Mottalib**, Head of the department, Computer Science and Engineering, IUT. Also our appreciation extends to all the respected faculty members of the Department of Computer Science and Engineering, IUT.

Finally we would like to extend our thanks to our friends, students, staffs and everyone else who have contributed to this work in their own way.

Table of Contents

Dedicated To Our Loving Parents

Chapter 1 - Introduction

1.1 – Overview

With the advent of Web 2.0 social networking sites are becoming more popular and interactive. Face-to-face, voice, email, and video communications are traditional medium of interaction between friends, family, and relatives. The traditional medium takes place when two parties had already shared some form of common value: interest, region, family bond, trust, or knowledge of each other. Although, on online social network (SN) two parties initiate communication without the common values between them, they still can freely share their personal information with each other [14]. In the virtual world, joining or creating groups and making friends are a click of a button, which makes online social networking sites, such as Friendster, MySpace, Hi5, and Facebook more and more popular and diverse each day [7]. Therefore, online SN's advantages are user friendliness and flexible in cyberspace where users can communicate with others and create and join groups as their wishes.

Joining Communities is influenced by many different factors like friendship factor, user preference factor with respect to SNS's. And with the invent of all types of virtual communication tools it is becoming more unpredictable who will like which community. This is of more importance to search engine companies and SNS's because the increase of people of similar minds expedites the probability of buying similar products. Hence Product advertisement which is the driving force of modern e-commerce gets easier [1].

1.2 – Problem statement

Community recommendation is a challenging task, due to the dynamics and diversity of community. A community may be formed at any time by an arbitrary number of people with similar interests, and the same person may join multiple types of community in nature. Community recommender system needs to capture not only the preferences of individual group members but also the social interaction and bonding. In our work, we investigate the problem of community recommendation based on different context. It is believed that hundreds of communities are created each day. Therefore, it is difficult for the users to view all the communities to select relevant ones. Again, there is some irreverent community recommendation, which results in poor user experience.

1.3 – Motivation

The motivation behind our work is to reduce the sample space of the list of community based on user interest by applying some filtering methods and considering the friendship factor. In this way will recommend user the effective community they may like to join rather than suggesting irreverent communities, which may make the user, feel annoying. We believe that the chosen parameters will correctly formalize user behavior and mentality which is vital point for recommending community. There is also some business perspective behind our work. If we can successfully recommend community for a user then it is also possible to identify user's test, interest as business point of view to recommend business products for the user.

Chapter 2 - Literature Review

2.1 - Social Networking

A Social Network is a group of people interacting with each other. It consists of activities of a user like messaging, posting, sharing information, joining different communities according to their interest. SNA has its origins in both social science and in the broader fields of network analysis and graph theory. Network analysis concerns itself with the formulation and solution of problems that have a network structure; such structure is usually captured in a graph. Graph theory provides a set of abstract concepts and methods for the analysis of graphs. These, in combination with other analytical tools and with methods developed specifically for the visualization and analysis of social (and other) networks, form the basis of what we call SNA methods. But SNA is not just a methodology; it is a unique perspective on how society functions.

Instead of focusing on individuals and their attributes, or on macroscopic social structures, it centers on relations between individuals, groups, or social institutions.

A social network is a set of people or groups of people with some pattern of contacts or interactions between them. The patterns of friendships between individuals, business relationships between companies, and intermarriages between families are all examples of networks that have been studied in the past.

2.2 - Recommender System

Over the last decade, Recommender Systems became an important research area to find out new approaches of recommendation both in industry and academia. The interest in this area still remains high because of the abundance of practical applications that help users to deal with information overload and provide personalized recommendations, content and services to them. Recommender systems can be traced back to the extensive work in the cognitive science, approximation theory, information retrieval, forecasting theories, and also have links to management science, and also to the consumer choice modeling in marketing.

Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that a user would give to an item (such as music, books or movies) or social element (e.g. people or groups) they had not yet considered, using a model built from the characteristics of an item or the user's social environment. The recommendation problem can be formulated as follows:

Let C be the set of all users and let S be the set of all possible items that can be recommended, such as books, movies, or friends. Let u be a utility function that measures usefulness of item s to user c, i.e.

$$u : C \times S \rightarrow R ,$$

Where R is a totally ordered set (non-negative integers or real numbers within a certain range).

Then for each user $c \in C$, we want to choose such item $s' \in S$ that maximizes the user's utility. More formally:

$$\forall c \in C, \; s'_c = \arg\max u(c, s)$$

In recommender systems the utility of an item is usually represented by a rating, which indicates how a particular user liked a particular item [9].

Generally Recommender systems are divided into two categories.

- Content-based recommendations: the user is recommended items similar to the ones the user preferred in the past.
- Collaborative recommendations: the user is recommended items that people with similar tastes and preferences liked in the past

Content-based filtering methods are based on information about and characteristics of the items that are going to be recommended. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. The system creates a content-based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques.

Simple approaches use the average values of the rated item vector while other sophisticated methods use machine learning techniques such as Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks in order to estimate the probability that the user is going to like the item.

Content-based approach to recommendation has its roots in information retrieval and information filtering research. Because of the significant and early advancements made by the information retrieval and filtering communities and because of the importance of several text-based applications, many current content-based systems focus on recommending items containing textual information, such as documents, Web sites (URLs), and Usenet news messages. Info finder [13] and News weeder are some examples of content-based model. Content-based approach has some limitations like the following:

Limited content analysis: The content must either be in a form that can be parsed automatically by a computer (e.g., text), or the features should be assigned to items manually. Another problem with limited content analysis is that, if two different items are represented by the same set of features, they are indistinguishable. Therefore, since text-based documents are usually represented by their most important keywords, content-based systems cannot distinguish between a well-written article and a badly written one, if they happen to use the same terms.

Over-specialization: The system can only recommend items that score highly against a user‘s profile; the user is limited to being recommended items similar to those already rated.

New User Problem: user has to rate a sufficient number of items before a content-based recommender system can really understand user‘s preferences and present the user with reliable recommendations. Therefore, a new user, having very few ratings, would not be able to get accurate recommendations.

Collaborative Filtering models recommend new items based on previous transactions as well as preference of similar users [cohesion]. This method collect and analyze a large amount of information on users‘ behaviors, activities or preferences and predicting what users will like based on their similarity to other users. User-based collaborative filtering attempts to model the social process of asking a friend for a recommendation. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself.

Collaborative filtering is mainly divided in two categories: memory-based and model-based collaborative algorithms [11]. The entire user-product database is used in memory-based algorithms to make a prediction while the model-based algorithm first

Generates a model of ratings and then predict. Though this approach has demonstrated its usefulness in many applications, it still has limitations that includes,

New user problem: The system must first learn the user‘s preferences from the ratings that the user makes.

New item problem: New items are added regularly to recommender systems.

Collaborative systems rely solely on users ‘preferences to make recommendations.

Therefore, until the new item is rated by a substantial number of users, the recommender system would not be able to recommend it.

Sparsity: The number of ratings already obtained is usually very small compared to the number of ratings that need to be predicted. Effective prediction of ratings from a small number of examples is important. Also, the success of the collaborative recommender system depends on the availability of a critical mass of us.

There are some **hybrid models** also where content-based and collaborative-based models have been unified to compromise their shortcomings [12]. They use components like linear combination of predicted ratings, various voting schemes, incorporating one component as a part of the heuristic for the other. For example, Billsus & Pazzani 2000 uses hybrid recommendation system.

2.3. Social Recommender Systems:

Social recommender systems target the social media and activity domain in social networking sites. This recommendation is often based on personalization techniques. The study of social-based recommender systems is a new area. One key insight is that social-based recommendations should account for a number of dimensions within a user's social network, including social relationship strength, expertise, and user similarity [16].

2.4- Community Recommendation System

A Community consists of group of people having common interest. It may be a hobby, something the community members are passionate about, a common goal, a common project, or merely the preference for a similar lifestyle, geographical location, or profession. Clearly, people join the community because they care about the common interest that glues the community members together. For example, in social networks, communities correspond to groups of friends who attended the same school, or who come from the same hometown [5]. With the advent of Web 2.0, social computing has emerged as one of the vital area of research recently. There is lots of research work on this area like movie recommendation, friend recommendation, and community recommendation etc. Recommendation systems is a method of information filtering system that seek to predict the 'rating' or 'preference' that a user would give to an item (such as music, books, or movies) or social element (e.g. people or groups) they had not yet considered, using a model built from the characteristics of an item or the user's social environment. With the increase of e-commerce recommendation system has gaining more interest. This is due to of possibility of increase sell obtained from successful recommendation. Sites that offer different products such as books, clothes and movies, most often also provides recommendations based on previous brought products. The problem of product, service, friend recommendation and community recommendation or in more global context information, is growing in both commercial and academic research interest [1]. Community Recommendation is recommending community for users they may like to join but not joined yet. There are two main recommendation algorithms content based and collaborative algorithm. Content based algorithm requires textual information as its name suggests and recommend websites newspaper articles and other contents. Collaborative based algorithm recommends products to a user which it believes have been preferred by similar user.

2.5- Related Work

We are quickly stepping towards a digital society. As more facilities are provided more people are getting connected. In the way people getting the opportunity to share their view, preference with each other. So, the community recommendation system is becoming increasingly important. It’s an interesting and challenging area of research. There is some work like forecasting community for user based on interaction in social networking site. Social graph is generated based on social relationship of user in social networking site [2]. Some work focus on bi-directional interaction between user and friends and activeness of friends in the community [3]. There is also some work like recommending friends in social networking sites based on cohesion [1]. Some work recommends groups based on decision trees and feature extraction from user profile [14]. There also exist some recommendation system which use collaborative filtering approach for community recommendation [17].

Limitation of present Approach:

- In the first two research work only the friendship bonding is used for detecting communities for a user to join but the user preference factor is ignored. User preference factor is the calculation of similarity between lists of community user has already joined with the list of community to join.
- Collaborative approach for community detection is suitable to some extent but doesn’t consider the friendship strength. Therefore, the collaborative approach lacks intuition in judging rational behavior of user.
- The parameters used are not appropriate enough to formalize user behavior.

In our work we are considering both friendship bonding and user preference factor. We believe that it will provide better user experience recommending related community for individual. Business perspective behind our work if we can successfully recommend community for a user then it is also possible to identify user’s test, interest as business point of view to recommend business products for the user.

Chapter 3 - Proposed System

In the previous chapter, we have extensively discussed about the existing community recommendation system. We have tried to find out the problems of that system and gain a lot of information about social networking and community recommendation. After analyzing those, we also try to make a new system for suggesting community in social networking sites. In this section, we present our proposed algorithm based on cohesion.

3.1 – Cohesion in Social Community

Cohesion is an abstract term that is easy to grab by intuition but surprisingly difficult to define in a strict manner. Informally cohesion is the sum of all the factors that attract people to join or to be part of a group. Cohesion refers to the degree to which the elements of a module belong together. Modules with high cohesion tend to be preferable because high cohesion is associated with several desirable traits of software including robustness, reliability, reusability, and understandability whereas low cohesion is associated with undesirable traits such as being difficult to maintain, difficult to test, difficult to reuse, and even difficult to understand [1].

In Social Networking, Cohesion is defined in a connected network and it is considered that network with high degree connectedness is more cohesive. Cohesion is an integral part of physical community and it is assumed that cohesion will have the same impact on social networking. So the impact of the cohesion cannot be ignored in online social networking and it is very effective term to recommending community for a person in social networking sites.

There many popular social networking sites like Facebook, MySpace, LinkedIn, twitter, Orkut etc. We have chosen Facebook as our sample social networking site for community recommendation. We have used small network of Facebook users in our research.

3.2 - Frame Work of proposed approach

Frame Work of Proposed Approach:

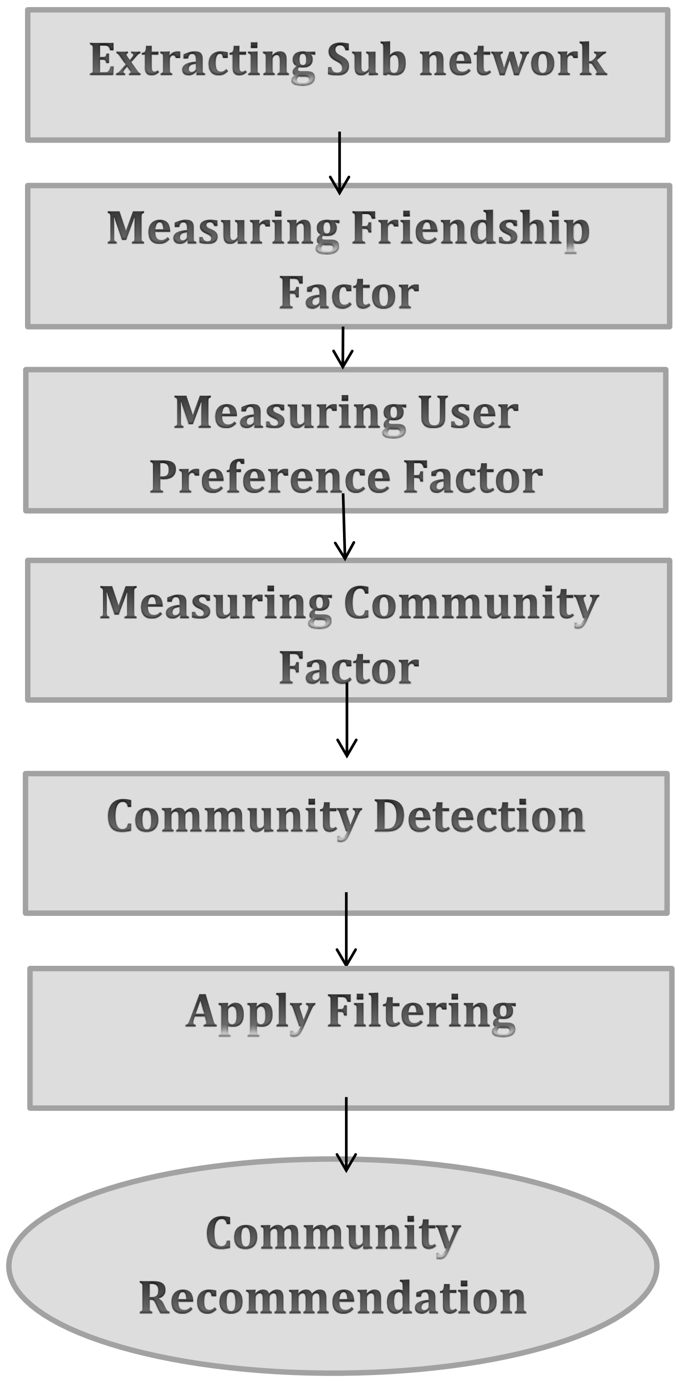


Figure: A Framework of Proposed Community Recommendation System

3.3 - Detail Explanation of Proposed System

In this section, we give a brief idea about our proposed system and how it works. This is explained below:

Utility of Factors: It is believed that the interest, like, dislike mentality of an individual is greatly influenced by the company he keeps. A fundamental property of social networks is that people tend to have attributes similar to those of their friends. There are two underlying reasons for this. First, the process of social influence leads people to adopt behaviors exhibited by those they interact with; this effect is at work in many settings where new ideas diffuse by word-of-mouth or imitation through a network of people. A second, distinct reason is that people tend to form relationships with others who are already similar to them. This phenomenon, which is often termed selection, has a long history of study in sociology [6]. We are using these factors because they play an important role determining friendship strength, which is essential for community detection. For example number of mutual friends, number of common life events, common background, number of photos tagged can determine closeness or bonding between friends. There may be some friends who do not have much activity among them in social networking site but they belong to common background like school, college, or same work place, which can be considered as offline friend's community. Offline friend's community can also be considered as strong bonding friend community. Friends who belong to similar communities tend to have similar interests. We are also keeping track of similar communities an individual is already member. This will give clear idea about an individual's mentality the kind of community he will like to join.

Extracting Sub-network:

Social Networking sites are very large entity with its size. Day by day the size of the network is increasing and as the people are joining there is huge number of information overloading happens on those sites. For experiment of our proposed system, we take the network of a random individual. We get the list of all communities of a user from Facebook using Netvizz and explore and analysis its visual representation with Gephi.

After getting the sub-network, we used the roster method to collect our desired data. (The roster method is representing the elements of set using brackets, {}. For instance, all even numbers under ten would be represented: {2, 4, 6, and 8}. the roster method is often associated with 'roster and rule' which is a way of finding a rule that the elements of a set follow. Sets can generally comprise any list of items (i.e. a grocery list)

Measuring Friendship Factor:

Friendship Factor indicates the quantitative amount friendship between two nodes (friends). In social network, there we get both online friends and also the presence of offline friends. People want the fellowship of their offline friends in their virtual life also. Therefore, factors for both friendships are important for measuring link strength. We calculate both online and offline friendship factor and they are termed as friendship factor.

To measure the strength between two nodes T and T1, the factors are defined below:

$$\text{Friendship}(T,T1) = \frac{OnlineFriendship_{strength}\,(T,T1)+OfflineFriendship_{strength}\,(T,T1)}{2}$$

Measuring online friendship factors:

We measure the strength between the online friends based on the friendship factors we have mentioned earlier and using some strategy. To measure the strength between two nodes T and T1, the factors are defined below:

$$OnlineFriendship_{strength}\,(T,T1) = \frac{\sum_{i=1}^{n} F_n\,(T,T1)}{n}$$

Where, n= number of parameters for calculating friendship strength,

Summary of responses to feature usage:

Answer Options	Response Percent
Comments or Wall (Public Messages)	95.3%
Profile	90.6%
Profile Photo (One)	87.5%
Photos (Additional)	87.5%
Messaging (Private Messages)	85.9%
Friends or Connections	82.0%
Groups	81.3%
Events	75.8%

Source:Success Factors of Online Social Networks-Evan Carroll
INLS 490 Online Social Networks
School of Information and Library ScienceThe University of North Carolina at Chapel Hill(2007)

Factors	Formula
F_1)T, T (1	Number of mutual friends /Total friends of T
F_2)T, T(1	Number of apps used by both/Total number of apps used by T
F_3)T, T(1	Number of photos tagged in/Total number of photos by T
F_4)T, T(1	Number of post on each other's wall/Total number of wall posts except own wall
F_5)T, T(1	Number of common events of T and T1/Total number of events by T
F_6)T, T(1	Number of messages/Total number of messages for T
F_7)T, T (1	Number of common likes between T and T1 /Total number of likes by T
F_8)T, T (1	Number of Common groups between T and T1 /Total number of likes by T

Therefore, $OnlineFriendship_{strength}$ between node (T, T1) can be defined as,

$OnlineFriendship_{strength}$ (T, T1) = $\frac{F_1+F_2+F_3+F_4+F_5+F_6+F_7+F_8}{8}$

Measuring offline friendship factors:

There may be some friends who do not have much activity among them in social networking site but they belong to common background like same school, college, or same work place, which can be considered as offline friends. Offline friendship can also be considered as strong bonding friend community. Friends who belong to similar communities tend to have similar interests. We are also keeping track of similar communities an individual is already member. This will give clear idea about an individual's mentality the kind of community he will like to join. To measure the strength between two nodes T and T1, the factors are defined below:

$$\mathrm{OfflineFriendship_{strength}}\,(\mathrm{T}, \mathrm{T1}) = \frac{\sum_{j=1}^{m} \mathrm{P_m}(\mathrm{T}, \mathrm{T1})}{\mathrm{m}}$$

Where, m= number of parameters for calculating offline friendship;

P_m (T, T1) = =parameter value of m for link from T to T1

Accordingly,

P_1)T, T= (1 Number of common educational institutions for T and T1/ Total number educational institutes of T.

P_2)T, T= (1 Number of common workplaces for T and T1/Total number of workplaces for T.

P_3)T, T= (1 Number of common places lived by T and T1/Number of places lived by T.

Educational background means from where and what school someone came from. Basically it refers all the schools that someone has been. Generally persons with such common backgrounds can be called as old friends. Ralph Waldo Emerson said "It is one of the blessings of old friends that you can afford to be stupid with them". People say school, college and varsity friends are friends for life. Even if they don't know each other in person, people with a common educational background have similar interest and way of thinking due to same educational atmosphere and sometimes same faculties or teachers. There are also a significant number of alumni communities in social network like Facebook. Therefore, it has to have a significant weight in link strength.

The workplace is the physical location where someone works. Such a place can range from a home office to a large office building or factory. The workplace is one of the most important social spaces other than the home. For friendship it's an important factor but some conditions and exceptions remain. There are some possibilities of people working for the same company but as a different department. Due to pressure of work and diversity of sections two friends never get a chance to communicate themselves. On the hand, common places lived by two friends. It is also a controversial issue. Because how much close they are, were they neighbors, were they hang out regularly? Actually we cannot determine any definite

$$OfflineFriendship_{strength}\,(T,T1) = \frac{P_1(T,T1)+P_2(T,T1)+P_3(T,T1)}{3}$$

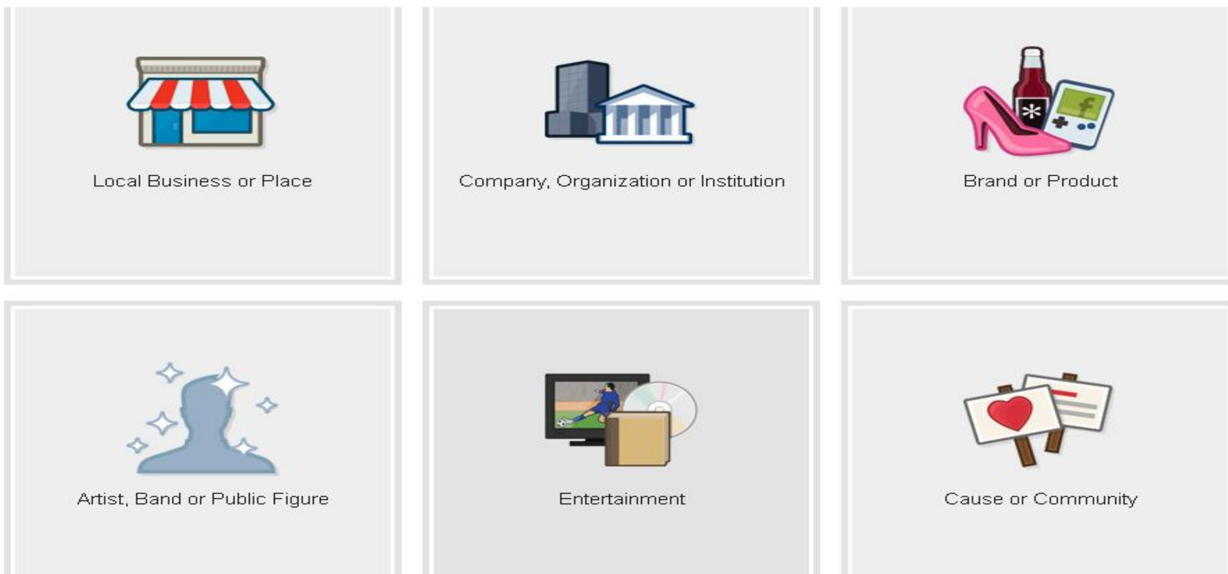
Combining $OnlineFriendship_{strength}\,(T,T1)$ and $OfflineFriendship_{strength}\,(T,T1)$ we can get the Friendship$(T,T1)$, this indicates the connectedness between two friends. Maximum value of Friendship$(T,T1)$ is 1.

$$\mathbf{Friendship}(\boldsymbol{T},\boldsymbol{T1}) = \frac{\boldsymbol{OnlineFriendship_{strength}}\,(\boldsymbol{T},\boldsymbol{T1})+\boldsymbol{OfflineFriendship_{strength}}\,(\boldsymbol{T},\boldsymbol{T1})}{\mathbf{2}}$$

The total link strength for a community is defined as Friendship Factor.

Measuring User Preference factor:

The user factor is calculated based on user preferences to the certain class of community. All the communities are classified into six general categories: local business, organization or institution, Brand or product, public figure, entertainment and cause.



The user factor for a certain community is calculated in terms of category of the community and number of user community present in that category which determines the user preference to that category of community. For example:

Select a group and detect if the group belongs to any cluster. User factor of a group for a target user can be defined as

$$\text{User Preference Factor}(C_n) = \frac{\text{Number of members in the cluster of selected group Cn}}{\text{Total number of groups of target user}}$$

Suppose, the user belongs to 46 groups which are divided into 9 classes or clusters

Cluster no.	1	2	3	4	5	6	7	8	9
No. of groups	11	5	9	4	7	2	3	4	1

If the group to be recommended is C_1 and it is belonged to cluster 3 then

$$\textit{User Factor}(C_1) = \frac{No.\, of\ memeberes\ in\ cluster\ no.\, 3}{Total\ no.\, of\ groups\ of\ user} = \frac{9}{46} = .20$$

Measuring Community factor:

Combining the friendship factor and user factor community factor is calculated for each community. The equation for community factor is given below:

Community Factor (c1) = *User Preference Factor (C_1)* Friendship Factor (C_1)*

Community Detection:

Initially we get the list of all communities of user and his/her friends in a sub-graph. Then we make an adjacency matrix (an **adjacency matrix** is a means of representing which vertices (or nodes) of a graph are adjacent to which other vertices).

We have an advantage of using a bipartite graph where

U_x are the users or his/her friends and total number of users is x

And C_y is the communities and total number of communities is y.

For adjacency matrix

$$A_{x,y} = \begin{cases} 1, & \text{if Ux is a member of Cy} \\ 0, & \text{otherwise} \end{cases}$$

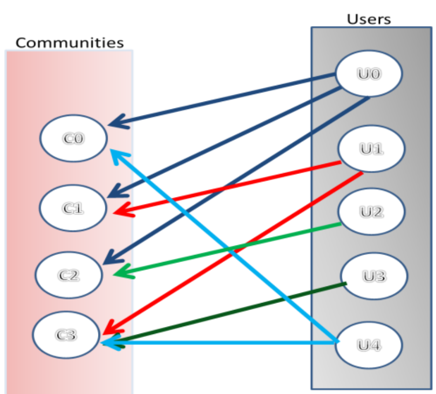


Figure: Users → Community relationships

$$A = \begin{array}{c} \\ U0 \\ U1 \\ U2 \\ U3 \\ U4 \end{array} \begin{array}{c} \begin{array}{cccc} C0 & C1 & C2 & C3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

For a particular individual determine all the communities. Suppose for U0 the communities are C0, C1, and C2. Now find the communities which could be recommended for U0. It is determined by rejecting the communities of U0 from total communities held by U0 and his/her friends.

Possible community recommended =

Total communities held by U0 and his/her friends - Communities of U0.

= { C0, C1, C2, C3 } - { C0, C1, C2 }

= {C3}

$$A = \begin{array}{cc} & \begin{array}{cccc} C0 & C1 & C2 & C3 \end{array} \\ \begin{array}{c} U0 \\ U1 \\ U2 \\ U3 \\ U4 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad A = \begin{array}{cc} & \begin{array}{cccc} C0 & C1 & C2 & C3 \end{array} \\ \begin{array}{c} U0 \\ U1 \\ U2 \\ U3 \\ U4 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Now the question is: which communities would be recommended?

Next step is:

Calculate $Link_{strength}\,(U0, Un)$ for every friends of U0 for n=number of friends.

Place the values $Link_{strength}\,(U0, Un) = Zn$ where Ax,y =1

Nullify the row for U0 except the possible recommended communities.

$$A = \begin{array}{cc} & \begin{array}{cccc} C0 & C1 & C2 & C3 \end{array} \\ \begin{array}{c} U0 \\ U1 \\ U2 \\ U3 \\ U4 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \Longrightarrow A = \begin{array}{cc} & \begin{array}{cccc} C0 & C1 & C2 & C3 \end{array} \\ \begin{array}{c} U1 \\ U2 \\ U3 \\ U4 \end{array} & \begin{bmatrix} 0 & z1 & 0 & z1 \\ 0 & 0 & z2 & 0 \\ 0 & 0 & 0 & z3 \\ z4 & 0 & 0 & z4 \end{bmatrix} \end{array}$$

Determine Community factor for every possible recommended communities by adding the values of Zn through the column. i.e.

$$CommunityFactor_{Cy} = \sum_{x=0}^{x} Z_x \ if \ A_{x,y} > 0$$

Where $CommunityFactor_{Cy}$ is community factor for Cy=0…n

For community C3

$$CommunityFactor_{C3} \ = z1 + z2 + z3$$

If $CommunityFactor_{C3}$ is greater than the threshold then this community is to be processed for further filtering.

Threshold can be found by summing up the community factor of present communities of user and then divide it by total number of present communities.

$$Threshold = \frac{\sum_{i=0}^{n} CommunityFactor_{Cn}}{n}$$

Here, $CommunityFactor_{Cn}$ are the Community factors of present communities of user and n= total number of total communities by user. Here, we define threshold value based on user profile and number of community an individual belongs to.

Calculating value this for all community in the list, we get a new list of community, which have value larger than the threshold.

Apply Filtering:

Then we rank the communities based on their value. The communities, which have value below threshold they are removed from the list. Then again, we use some filtering strategy to select efficient community for user.

- Using user preference rank the communities according to the type or genre based on mostly liked type of communities by user.
- A filtration is to be done according to location and gender for appropriate recommendation

Community Recommendation:

After performing the all steps of our proposed method, a list of community for user is generated. Then finally, we recommend the communities to the user.

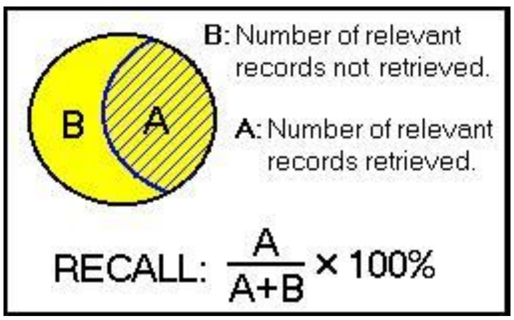
3.4 - Challenges

There are some challenges we have to face during our research like:

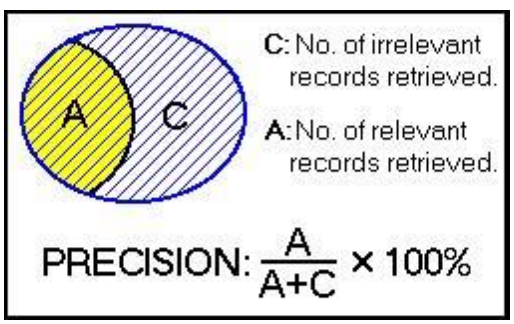
- Attempting to second-guess a mysterious, perverse and profoundly human form of behavior: the personal response to a work of art is a challenge task for a recommendation system.
- There are some limitations for determining the factor of friendship strength as there are so many things to consider and recommending community for a new user will be the most challenging task.
- The major problem in collecting the Facebook data was privacy concerns. At the same time, the format of the Facebook data was the most congenial to our research method. So, we have collected our data using roster method within a small network of user in Facebook.
- Our proposed model doesn't solve the fresh start problem. So, user must be member of at least one community and user should have some friends because we recommending based on cohesion and user preference.
- If the target user has no friend or friends do not belong to any community then our recommended system fails to calculate the threshold value.

Chapter 4- Experiment and Result Analysis

We validated our idea on small network in Facebook. A take the list of present community of user. Then eliminate a number community from the list which the user present and check if the community is recommended using our proposed model. Then we validated the result in terms of precision and recall and final score. As we are not using any global threshold value, the value of threshold will change based users present community list according to user factor and friend factor.



RECALL is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.



PRECISION is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

F_1 score (also **F-score** or **F-measure**) is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score: *p* is the number of correct results divided by the number of all returned results and *r* is the number of correct results divided by the number of results that should have been returned.

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

Initially we took one target user and eliminate one random community from his/her existing communities and calculate the precision and recall. Then we estimate the average of that and then we eliminate any two communities from his/her community list randomly and calculate the mean of precision and recall. We do the similar execution until we delete half of the existing communities from the user.

No. of communities eliminated	Precision	Recall
1	0.6876	0.8750
2	0.7442	0.8036
3	0.7473	0.7440
4	0.8154	0.6289

Avg. Precision= 0.7486 Avg. Recall=0.7629 F1=0.7557

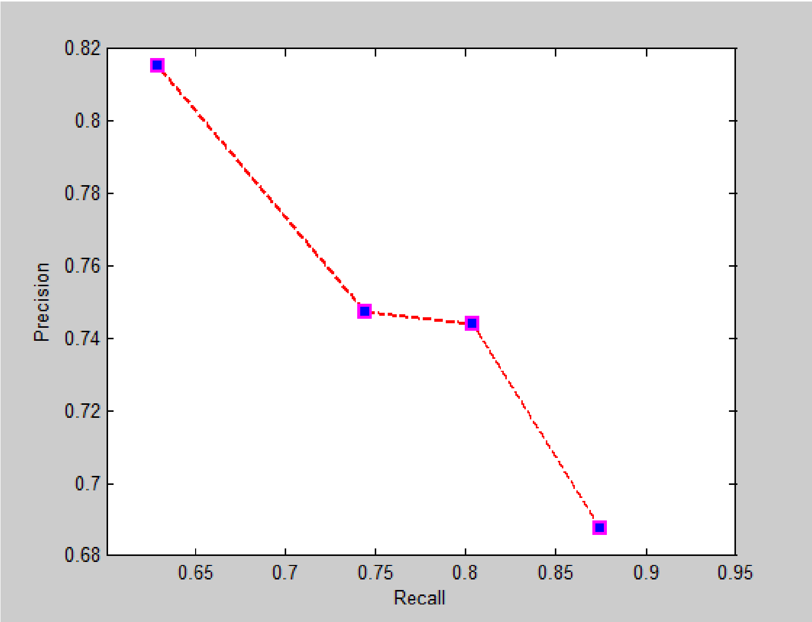
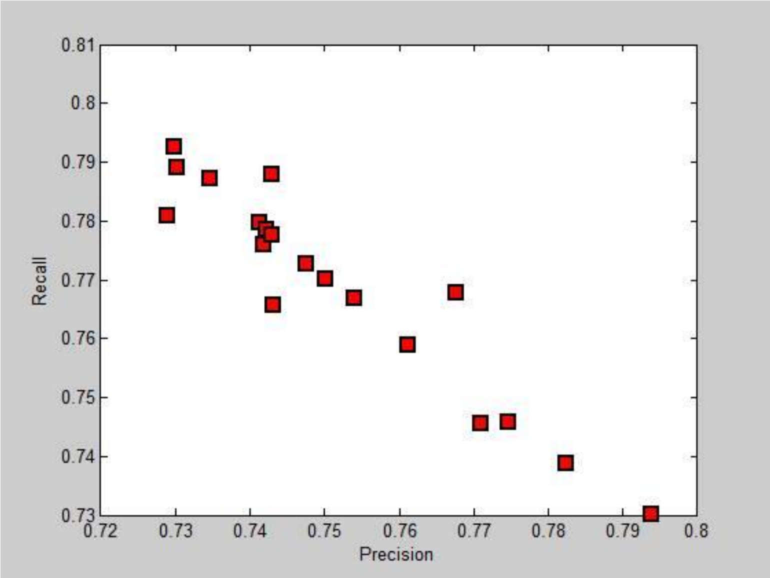
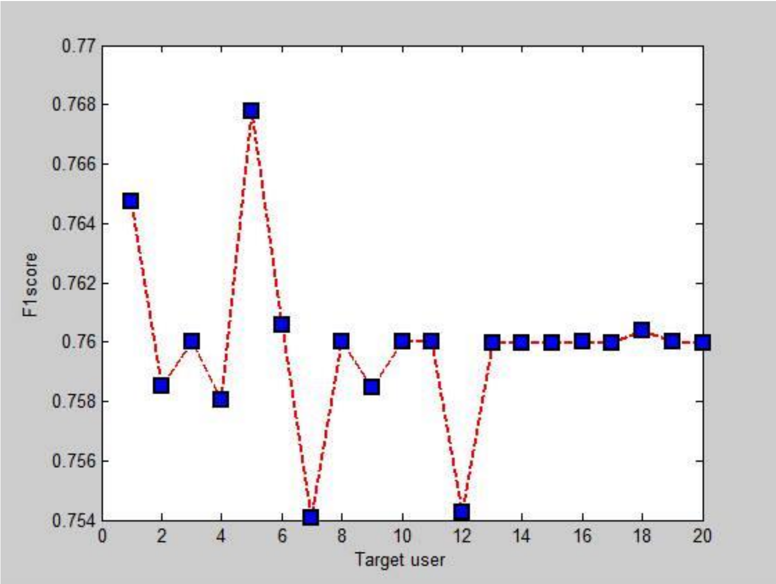


Fig: Average precision and recall of a target user by elimination of existing groups

If we take the average precision recall graph for 20 target users eliminating one up to maximum half of the present communities with all combinations and taking average we get a graph like this:



The Average F1 score graph for 20 target user is generated like this:



Chapter 5 - Conclusion and Future Work

This is an interesting area of research. This research will help the users to be a member of a community of their own interest. There are some limitations for determining the factor of friendship strength as there are so many things to consider and recommending community for a new user will be the most challenging task. We are working to solve the problems and build an effective community recommendation system for the betterment of the users.

References:

[1] Md. Nafiz Hamid, Md. Abu Naser, Md. Kamrul Hasan and Hasan Mahmud, A Cohesion Based Friend Recommendation System, Springer Journal, Social networking and Mining, February 2014.

[2] Sanam Kadge and Gresha Bhatia, Graph Based Forecasting for Social Networking Site, 2012 International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20, Mumbai, India.

[3] Marjaneh Safaei, Merve Sahan and Mustafa Ilkan, Social Graph Generation & Forecasting using Social Network Mining, 2009 33rd Annual IEEE International Computer Software and Applications Conference.

[4] Nitai B. Silva, Ing-Ren Tsang, George D.C. Cavalcanti, and Ing-Jyh Tsang, A Graph-Based Friend Recommendation System Using Genetic Algorithm, WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Barcelona, Spain.

[5] Jaewon Yang, Julian McAuley and Jure Leskovec, Community Detection in Networks with Node Attributes, Stanford University, Data Mining (ICDM), 2013 IEEE 13th International Conference.

[6] David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, Siddharth Suri, Feedback Effects between Similarity and Social Influence in Online Communities, KDD '08 Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM New York, NY, USA ©2008.

[7] Wellman, B., Boase, J., Chen, W.: The networked nature of community. IT&Society,vol. 1(1), pp. 151- 165 (2002).

[8] Gediminas Adomavicius and Alexander Tuzhilin, Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, Journal IEEE Transactions on Knowledge and Data Engineering archive, Volume 17 Issue 6, June 2005

[9] Krulwich, B., and Burkey, C., Learning user information interests through hextraction of semantically significant phrases,In Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access, Stanford, Calif., March 1996.

[10] Lang, K ―Newsweeder: Learning to filter netnews, In Proceedings of the 12th International Conference on Machine Learning, Tahoe City, Calif., USA, 1995.

[11] Breese, J. S., Heckerman, D., and Kadie, C, Empirical Analysis of Predictive Algorithms for Collaborative Filtering‖ in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43-52.

[12] Melville, Prem Mooney, Raymond J Nagarajan, Ramadass,Content-Boosted Collaborative Filtering for Improved Recommendations,Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002), pp. 187-192, Edmonton, Canada, July 2002.

[13] Chin, A. and Chignell, M.,Automatic detection of cohesive subgroups within social hypertext: A heuristic approach., New Review of Hypermedia and Multimedia, Vol. 14, No. 1, July 2008, 121-143.

[14] Jessica Marie Vitak, BA, Facebook "Friends": How online identities impact offline relationships, A Thesis submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University in partial fulfillment of the requirements for the degree of Master of Arts in Communication, Culture and Technology, April 22, 2008.

[15] Group Recommendation system for Facebook, Enkh-Amgalan Baatarjav, Jedsada Chartree, and Theraput Meesumrarn, Department of Computer Science and Engineering University of North Texas, Denton, Texas, 76203, USA.

[16] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, Karim Seada, "Enhancing Group Recommendation by Incorporating Social Relationship Interactions", Proceedings of the 16th ACM International Conference on Supporting Group Work (GROUP) 2010.

[17] Wen-Yen Chen, Dong Zhang and Edward Y. Chang, Combinational Collaborative Filtering for Personalized community Recommendation, KDD '08 Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA ©2008.