الجامعـــة الإسلاميـــة للتكنولوجيا

UNIVERSITE ISLAMIQUE DE TECHNOLOGIE

ISLAMIC UNIVERSITY OF TECHNOLOGY

DHAKA, BANGLADESH

ORGANISATION OF ISLAMIC COOPERATION

A Cohesion Based Friend Recommendation System

Authors

Md. Abu Naser	Student Id: 084405
Md. Nafiz Hamid	Student Id: 084421

Supervisor

Md. Kamrul Hasan, PhD

Systems & Software Lab (SSL)

Assistant professor

Department of Computer Science & Engineering (CSE)

Islamic University of Technology (IUT)

Co-Supervisor

Hasan Mahmud

Systems & Software Lab (SSL)

Assistant Professor

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, 2012

CERTIFICATE OF RESEARCH

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** and co-supervision of **Hasan Mahmud, 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

Md. Abu Naser **Md. Nafiz Hamid**

Signature of Co-Supervisor

Hasan Mahmud
Assistant professor
Department of CSE, IUT

Signature of Supervisor

Md. Kamrul Hasan, PhD
Assistant professor
Department of CSE, IUT

Signature of the Head of the Department

Prof. Dr. M. A. Mottalib
Head, Department of CSE, IUT

Abstract

Social network sites have attracted millions of users with the social revolution in Web 2.0. A social network is composed by communities of individuals or organizations that are connected by a common interest. Online social networking sites like Twitter, Facebook and Orkut are among the most visited sites in the Internet. In the social network sites, a user can register other users as friends and enjoy communication. However, the large amount of online users and their diverse and dynamic interests possess great challenges to support such a novel feature in online social networks. In this paper, we design a general friend recommendation framework based on cohesion after analyzing the current method of friend recommendation. The main idea of the proposed method is consisted of the following stages- measuring the link strength in a network and find out possible link on this network that is yet to be established; detecting communities among the network using modularity and recommending friends.

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 and co-supervisor **Hasan Mahmud,** Assistant Professor, Department of Computer Science and Engineering, IUT. Their supervision, knowledge and relentless support allowed us to complete this endeavor successfully. Their patience and encouragement allow us to stand where we stand today.

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

Social Networking Sites (SNS) are the dominating entities in the modern web. Online friendship is now similarly appealing to people as offline friendship if not more. People tend to enjoy the fellowship of their real life friends in a virtual world. At the same time they are interested in making online friends. But With the rapid growth of SNS's resulting in information overload people are in dilemmas to choose the right friend and rightly so.

Potential choice of friends is influenced by many intrinsic as well as exogenous factors with respect to SNS's. And with the invent of all types of virtual communication tools it is becoming more unpredictable who will like whom. 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. The recent surge of research in recommendation algorithms is therefore not surprising. Popular movie recommendation site Netflix (www.netflix.com) had even conducted competition with $ 1M grand prize for a 10% improvement on their movie recommendation algorithm.

However, recommendation of products and recommendation of human beings are as disparate as can be. Historically there has been two main recommendation algorithms- content based and collaborative algorithm. Content based algorithm requires textual information as its name suggests and recommends websites newspaper articles and other contents. Collaborative based algorithm recommends products to a user which it believes have been light by similar users. Both of these algorithms have yielded unsatisfactory results in friend recommendation because of their lack of intuition in judging rational beings.

In the last few years, social networks have been increasing in both size and services. Social networking services (SNSs) such as Facebook, MySpace, Twitter, Flickr, YouTube and Orkut are growing in popularity and importance and to some extent they are also contributing to a change in human social behavior. Online social friends are more influencing and people are now eager to make friendship with their same mentality. As the size is growing, it is very difficult to find out good friends and sometimes the user gets negative result. So the user of this social networking site expects that the system administrator will help them by friend recommendation system. Many researches are going on this point to find out recommendation system on various ways. We also try to improve the recommendation system on the basis of cohesiveness of friendship.

1.2 - Problem Statement

In our overall thesis, we investigate the problem of friend recommendation in modern context. It is believed most of the SNSs deploy trivial FOF (Friends of Friend) algorithm for friendship recommendation [9, 18, 33]. Many others use search engines that provide an abundance of friends as recommendations. Hence, we do a thorough research on the existing distinct friend recommendation algorithms. We show that there has been considerable research in this field from different perspectives. For example, Topology based recommending systems have already been articulated by Researchers while context-based approaches have also been discussed.

Therefore, our aim in this paper is two-fold. First, we do a comprehensive literature review in friend recommendation in case of SNSs, what were the techniques, what were their strengths and what were their weaknesses. We also try to classify them in some classes. Second, we propose a new algorithm where we suggest friends in terms of cohesion. First, we measure link strength of an extracted network by some of the parameters that constitute cohesion. Second, we augment the network by the calculated link strength and 'litient Conjecture' rule. Finally, we detect the community using louvaine method and recommend friends within the same community

1.3 - Motivation

In this research, we want to clarify what we mean by “cohesion” and why it is a suitable method for friend recommendation. Cohesiveness is an abstract term which is easy to grab by intuition but surprisingly difficult to define in a strict manner. Essentially, Cohesiveness is the sum of all the factors that attract people to a certain group and in terms of SNSs, we can classify them into certain clusters. Kurt Lewin [30] defined cohesiveness as how a member perceives her relationship with a certain group. Therefore, cohesiveness is an integral part of a community and we believe it has the same impact on an online community. Interestingly, there has been no research on how cohesiveness plays a part in people liking each other to date in our knowledge. Hence, we have put forward an approach that tries to identify cohesive subgroups and then recommend friends within that subgroup to each other who are yet to be connected.

1.4 - Research Challenges

Recommending people on social networking sites is worth studying because it is different from traditional recommendations of books, movies, restaurants, etc. due to the social implications of “friending”. For example, before adding a friend, one often has to consider how the other person would perceive this action and whether he or she would acknowledge the friendship.

Furthermore, the most important challenge in designing a recommender system for a social network is the privacy of users. With the ever increasing web crimes and identity theft, people are becoming more and more skeptical and careful in sharing their personal information. Hence, unless a user can trust the system with their data, there will be missing attributes which creates a stymie in generating recommendation.

Moreover, a major hindrance in research of social networking sites is collection of data because of the privacy issue. Exploitation of social network data is the fragmentation of the population of social network users into numerous proprietary and closed social networks. This issue is compounded by the fact that each new game or media application tends to build its own social network around it rather than building upon the rich data available about existing social relationships. So it is difficult to find out real data of user in case of research. Spam detection and advertisement detection are research challenges that need extra attention from the research community. Since users and data production increase, spam (irrelevant in-formation) and advertisements will continue growing.

As social networks will continue to evolve, discovering communities and constructing specific social graphs from large scale social networks will continue to be a dynamic research challenge. Also, online social communities face critical social and ethical issues that need special care and delicate handling.

Chapter 2 – Literature Review

2.1 - Social Networking

With the advent of Web 2.0, social computing has emerged as one of the hot research topics recently. It involves the collecting, extracting, accessing, processing, computing and visualizing of social signals and information. SNSs are an online phenomenon which provides social network based services to support easy message posting, information sharing and inter-friend communication.

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.

Social Network Sites are defined as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system [24]. Social Networking sites (SNS's) provide users with opportunity to connect with their offline friends as well as making new friends with latent ties who otherwise would never have met them. They also supplement their relationships with close relations and help to maintain the social capital [25]. People tend to trust the opinions of friends they know rather than the opinions of strangers.

Key Terminologies of Social Network

As social networking sites like facebook, twitter, google plus etc are getting popular day by day, social networking analysis are becoming an important subject for the researcher. There are some basic terminologies in social networking analysis. Some of them are discussed below.

Vertex (pl. vertices): The fundamental unit of a network, also called a site(physics), a node (computer science), or an actor (sociology).

Edge: The line connecting two vertices. Also called a bond (physics), a link (computer science), or a tie (sociology).

Directed/undirected Edge: An edge is directed if it runs in only one direction (such as a one-way road between two points), and undirected if it runs in both directions.Directed edges, which are sometimes called arcs, can be thought of as sporting arrows indicating their orientation. A graph is directed if all of its edges are directed. An undirected graph can be represented by a directed one having two edges between each pair of connected vertices, one in each direction.

Degree: The number of edges connected to a vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. In a few recent articles, the degree is referred to as the \connectivity" of a vertex, but we avoid this usage because the word connectivity already has another meaning in graph theory. A directed graph has both an in-degree and an out-degree for each vertex, which are the numbers of in-coming and out-going edges respectively.

Diameter: The diameter of a network is the length (in number of edges) of the longest geodesic path between any two vertices. A few authors have also used this term to mean the average geodesic distance in a graph, although strictly the two quantities are quite distinct.

Degree centrality: A node's (in-) or (out-) degree is the number of links that lead into or out of the node. In an undirected graph they are of course identical. Often used as measure of a node's degree of connectedness and hence also influence and/or popularity.

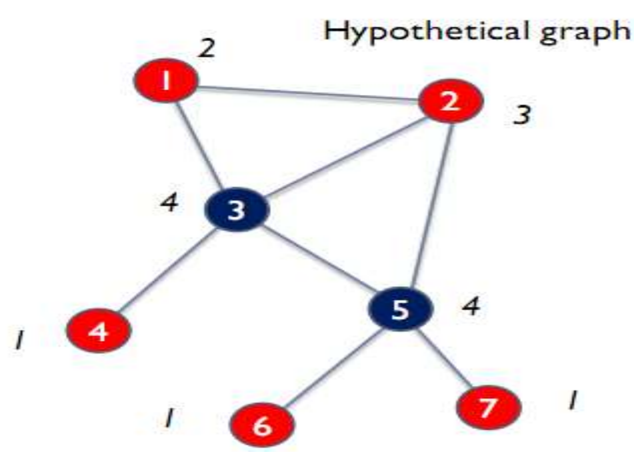
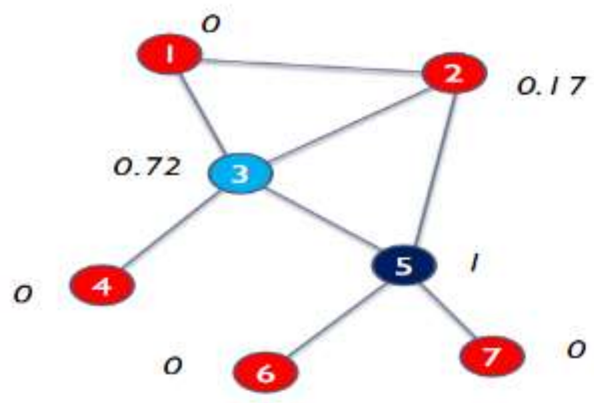


Figure: Degree Centrality

It is useful in assessing which nodes are central with respect to spreading information and influencing others in their immediate 'neighborhood'.

Betweeness centrality: The number of shortest paths that pass through a node divided by all shortest paths in the network. It shows which nodes are more likely to be in communication paths between other nodes. It is also useful in determining points where the network would break apart.



Closeness centrality: The mean length of all shortest paths from a node to all other nodes in the network It is a measure of reach, i.e. how long it will take to reach other nodes from a given starting node.

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 [32].

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 [26] and News weeder [27] 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 [1]. 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 [28]. 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 [29]. 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 - Friend Recommendation System

With the rapid growth of social networks, users of SNSs may easily get overwhelmed by the excessive volume of information. The friendship can significantly affect the quality of recommendations. Therefore, the recommendation of better friend is the essential factor of social network sites to find truly valuable information. There are many friend recommendation systems for social networking sites.

Topological characteristics of Social networks have been well researched [13, 14] while the theoretical research of complex systems has also grown [15, 16]. Liben -Nowell et al [2] defined the link prediction problem as given a snapshot of a social network at time t, predicting the edges that will be added to the network during the time interval from time t to a time t'. A topology based method was applied to approach this problem. As many as nine different predictors (graph distance, common neighbors, Jaccard's co-efficient [17], Adamic /Adar [18], preferential attachment [19], Katz measure [20], hitting time, page [21] and Jeh [22]) were used and their results were caparisoned. Some other supervised [6, 7] and unsupervised [8] models have also been proposed. Lu et al [23] proposed a supervised learning model that can predict links using multiple sources where auxiliary networks also exist.

Silva et al [9] introduced a new clustering index and user calibration procedure with Genetic Algorithm to suggest friends. Knowledge of the structure and topology of these complex networks combined with quantitative properties such as size, density, average path length or cluster coefficient is used in this approach. Albeit a topology based approach, the innovative approach towards this problem certainly opened a new possibility. Chin et al [10] proposed the SCAN method to find potential cohesive subgroups which can be used further to friend recommendation. This method was invented in a social hypertext context.

Moreover, Graph based features have also been used in proposing efficient friend recommendation system. Lin et al [12] proposed WMR- a graph based friend recommendation algorithm. They show weighted minimum-message ratio (WMR) which generates a limited, ordered and personalized friend lists by the real message interaction number among web members.

In Collaborative and Structural recommendation of friends [11] uses the link structure of a social network and content-based recommendation using mutual declared interests. They investigate the problem of link recommendation in such weblog-based social networks and describe an annotated graph-based representation for such networks. It uses graph feature analysis to recommend links (u, v) given structural features of individual vertices and joint features of the start and end points of a candidate link, such as distance between them.

Spiritual and Social context was used by Kwon et al [4] to propose a method that measures friendship strength and then suggests friends from a list. The main idea of the proposed method is consisted of the following three stages; (1) computing the friendship score using physical context; (2) computing the friendship score using social context; (3) combining all of the friendship scores and recommending friends by the scoring values.

Moreover, Yu Zheng et al [34] proposed a system of friend recommendation based on location.-GeoLife2.0. It is a GPS-data-driven social networking service where people can share life experiences and connect to each other with their location histories. By mining people's location history that can measure the similarity between users and perform personalized friend recommendation for an individual.

However, the previous approaches did not consider cohesion for friend recommendation in social networking sites. The friend recommender system is needed to tailor towards cohesion as it has impact on different interaction medium of social networking.

Chapter 3 – Proposed System

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

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 friends for a person in social networking sites.

From the definition we can see that cohesion can be thought of two parameters. One is connectedness and another is density. Whenever this parameter comes, the term modularity is suitable to explain cohesion. Because modularity is one measure of the structure of networks or graphs. It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity is used for detecting community structure in networks with strong cohesion. We have also used this method to finding out cohesive communities in networks for recommending friends.

3.2 - Frame Work of proposed approach

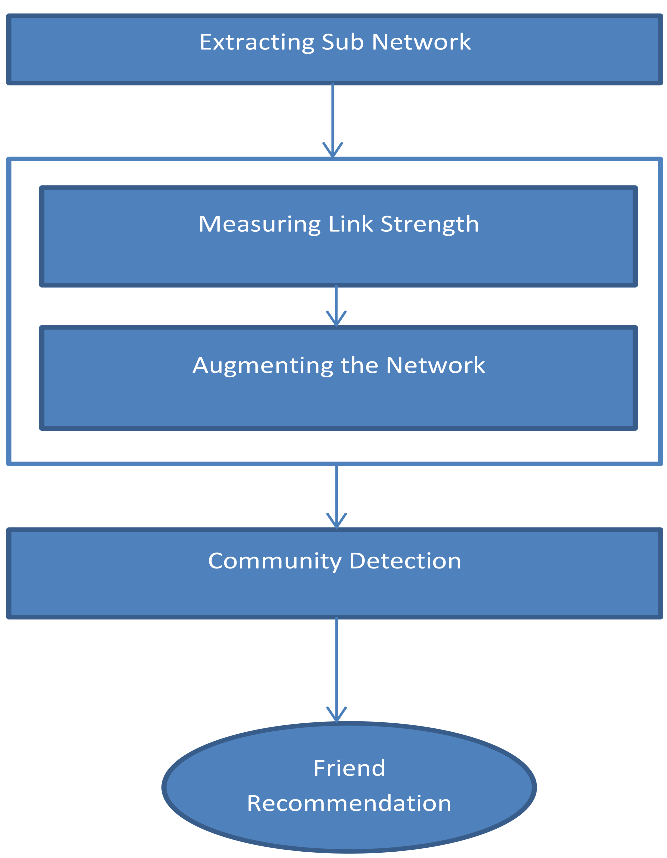


Figure: A Framework of proposed friend recommendation system.

3.3 - Detail Explanation of Proposed System

3.3.1 - 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 overload happens on those sites.

For experiment of our proposed system, we take the whole network of a random individual. After getting the whole network of a client for who are going to suggest friends, we extract the sub-network of 'x' people from the visualized graph.

3.3.2 - Measuring Link Strength

This is an important step of our proposed friend recommendation system. In this step we first collect data from the users of the extracted graph and requested them to provide our desired data of mutual friend, group information and application use.

For measuring the link strength between to users, we will consider these three parameters.

A. **Mutual Friend**: Mutual friend means common friend. In the social networking sites, as there are several common friends between the users. The traditional social networking sites mainly use this 'mutual friend' information to suggest friend.

   Formally, we can define mutual friend as like, w is a mutual friend of u and v, if and only if w is friends to both u and v.

Mathematically, $\boldsymbol{Mutual}\,(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\ \ \boldsymbol{iff}\ \ \boldsymbol{f}(\boldsymbol{u},\boldsymbol{w}) \wedge \boldsymbol{f}(\boldsymbol{v},\boldsymbol{w})$

B. **Group**: In social networking sites we can see that there are many types of groups for different purposes. People join these groups whenever they feel important on that. Sometimes for professional, ideological and academic reasons people join these groups that should count in calculating the strength between the links.

C. **Application use**: People join these type of social network not only for interaction but also they want to get interesting things that can make their time enjoyable. So this social networking sites always try to introduce different application that user can get interest. It is observable that people are using this type of application very frequently.

Formula of measuring link strength with example

We measure the link strength of a client node T and any node friend T1 by the following rules-

$$\boldsymbol{Link_{strength}\,(T,\,T_1) = Weight_1 * f_1\,(T,\,T_1) + Weight_2 * f_2 + Weight_3 * f_3\,(T,\,T_1)}$$

Where,

$f_1\,(T,\,T_1)$= Number of mutual friends / Total friends of T

$f_2\,(T,\,T_1)$= Number of groups both have joined / Total groups joined by T

$f_3\,(T,\,T_1)$= Number of same apps both use / Total apps used by T

$Weight_1\ (mutual\ friend)$= 0.5; $Weight_2(group)$= 0.3; $Weight_3\ (App\ use)$= 0.2

The weights for different parameters have set empirically.

Example: Subject A has 7 mutual friends with Subject B. There are 3 groups where they both have joined and 4 apps they both use. And the total number of friends of A is 12. The total group joined by A is 5 while A used a total number of 6 applications.

So, the link strength between A to B can be easily calculated using the rules specified earlier.

$$Link_{strength}\left(A,\ B\right) = 0.5 * \frac{7}{15} + 0.3 * \frac{3}{5} + 0.2 * \frac{4}{6} = 0.546$$

3.3.3 - Augmenting the network

After measuring the link strength, we augment the network with links between people that we think can happen but not present in the network. To consider possible links for a client T, we will only take into account his 'friends of friends', thereby creating triads, increasing clustering co-efficient. To determine whether a link can happen, we have used the '**Lenient Conjecture'** rule.

Lenient Conjecture rule:

For a client T, a link between him and his friend of friend T2 can happen if

$$\frac{Link_{strength}(T, T_2) + \ Link_{strength}(T_2,\ T)}{2} > Threshold$$

Where,

$$Threshold = \ \frac{\sum_{i=0}^{n} Existing\ Link\ _{strength(i)}\ for\ T}{n}$$

Explanation of 'Lenient Conjecture' rule:

Let, Subject A is friend with Subject B & C. Hence, Subject A's Threshold to be friend with another node. So the threshold of A can be calculated by this rule

$$Threshold = \frac{Link_{\ strength(A,B)} + Link_{\ strength(A,C)}}{2}$$

= (0.561 + 0.588)/2 = 0.5745

Now, let, Subject D is a friend of friend of Subject A. For D have to be A's friend the following condition must be true-

$$\frac{Link_{strength}(A, D\ ) + \ Link_{strength}(D,\ A)}{2} > Threshold$$

But here, (0.478+ 0.4125)/2 > 0.5745. 0.442 is not greater than 0.5745. So the link between A and D is not possible and we will not augment this connection.

3.3.4 - Community Detection & Friend recommendation:

In the study of networks, such as social networks a number of different characteristics have been found to occur commonly, including the small-world property, heavy-tailed degree distributions, and clustering, among others. Another common characteristic is community structure. In the context of networks, community structure refers to the occurrence of groups of nodes in a network that are more densely connected internally than with the rest of the network.

In this step of our proposed system, we detect community in the social networks using the state-of-the-art 'Louvaine Method'. This method uses modularity and link strength to detect the community among the networks.

The Louvain method is a simple, efficient and easy-to-implement method for identifying communities in large networks. The method unveils hierarchies of communities and allows tozoom within communities to discover sub-communities, sub-sub-communities, etc. It is today one of the most widely used method for detecting communities in large networks.

The method is a greedy optimization method that attempts to optimize the "modularity" of a partition of the network. The optimization is performed in two steps. First, the method looks for "small" communities by optimizing modularity locally. Second, it aggregates nodes belonging to the same community and builds a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained and a hierarchy of communities is produced.

After detecting community, we will recommend a client with people from same communities who are yet to be friends. Also, Also for people already friends with all of their neighbors in the same community, we have recommended them with people with the highest $\text{Link}_{\text{strength}}$ from other communities.

Chapter 4 – Experiment and Result Analysis

4.1 - Experiment Details

We experiment our whole proposed system based on the popular social networking site Facebook. For extraction of the sub-network, we took the whole facebook network of a random individual and visualize it with Industrial-Strength research tool Gephi 0.8.1. Then, we cut off a random sub-network of 10 people.

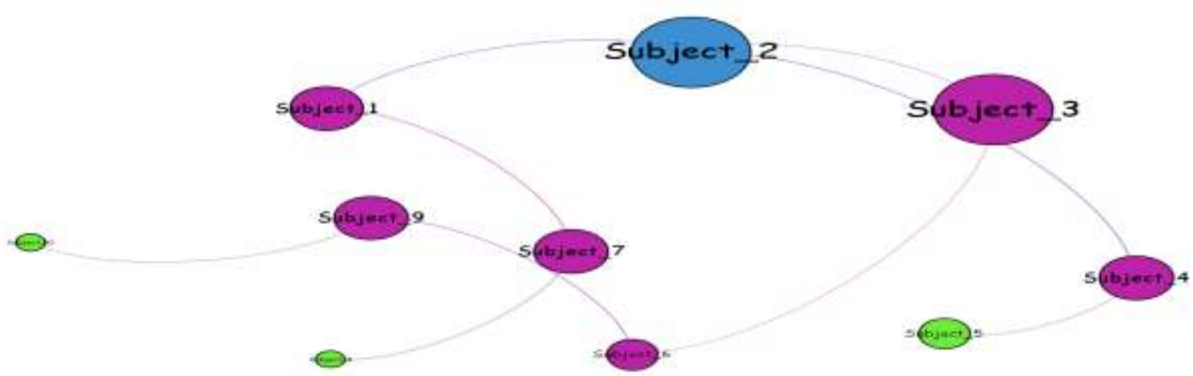


Figure: Extracted sub-network of 10 people

After extraction of sub-network of 10 people we use the roster method to collect our desired data that is needed to calculate link strength. We requested each member of the network to provide the information of mutual friend, groups and application use. Some of the data of Subject_2 is shown on the following table.

Connection	Mutual friend	Total Friend	Same group	Total group	Same App. use	Total App. use
Subject_2-Subject_3	8	20	3	8	2	10
Subject_2-Subject_4	4	20	2	8	4	10
Subject_2-Subject_1	9	20	3	8	2	10
Subject_2-Subject_9	8	20	5	8	1	10

Table: Data for calculating link strength

After collecting data from each users of the network, using perl scripts we calculate the link strength of the connection between existing links. Then we got the result between different links. Some of the results are shown below:

Link	Strength
Subject_1- Subject_2	0.561904761904762
Subject_2- Subject_3	0.471008403361344
Subject_3 – Subject_6	0.335
Subject _6- Subject_9	0.575238095238095
Subject_9- Subject_10	0.536134453781513
Subject _1- Subject_7	0.588095238095238

Table: Link Measurement result

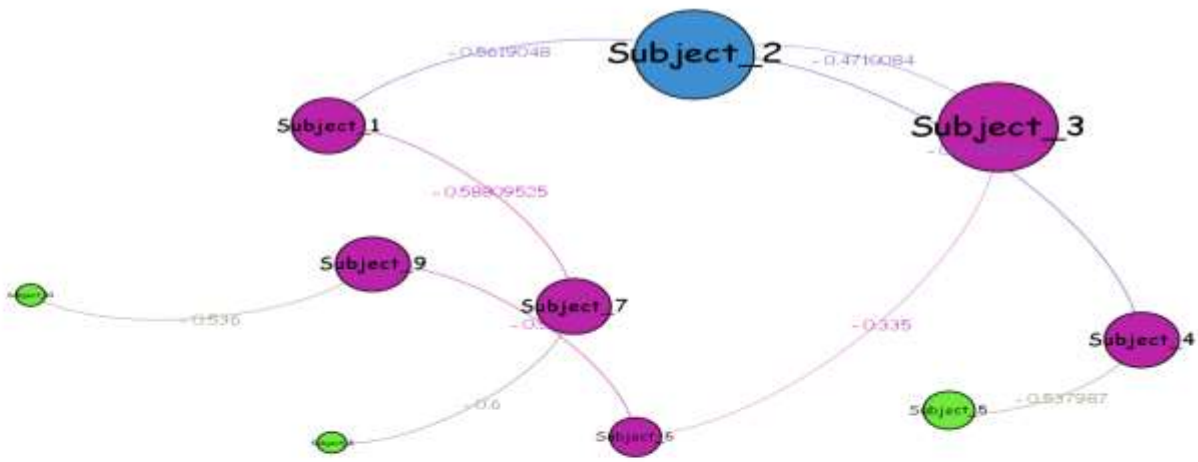


Figure: Link strength connection

After that we augment the network that means we create some connections between different nodes (people) that are not present in the network but yet they can be friend of each other. In case of augmenting the network we do not randomly create a connection. We use the lenient conjecture rule that check the possibility of link creation between two connections. For the calculation we took the link strength in both direction and took the average link strength between them. If the new link strength is greater than the threshold value then we will create connection and augment the network. If the link strength is less than the threshold value, we do not recommend those links. Some of the results are shown in the following table.

Link	Strength1	Strenght2	Avg. between link	Threshold	Link Possibility
Subejct_1-Subject_3	0.561	0.588	0.578	0.4875	Possible
Subject_1-Subject_4	0.1041	0.114	0.1091	0.165	Not Possible
Subject_1-Subject_8	0.471	0.527	0.678	0.703	Not Possible
Subject_2-Subject_7	0.471	0.527	0.678	0.334	Possible

After that getting the augmented network, we applied the state-of-the-art louvaine method that is a simple, efficient and easy-to-implement method for identifying communities in large network. This louvaine method is implemented in different social networking tools. Gephi is one of the tools that we used to implement the louvaine method to detect community.

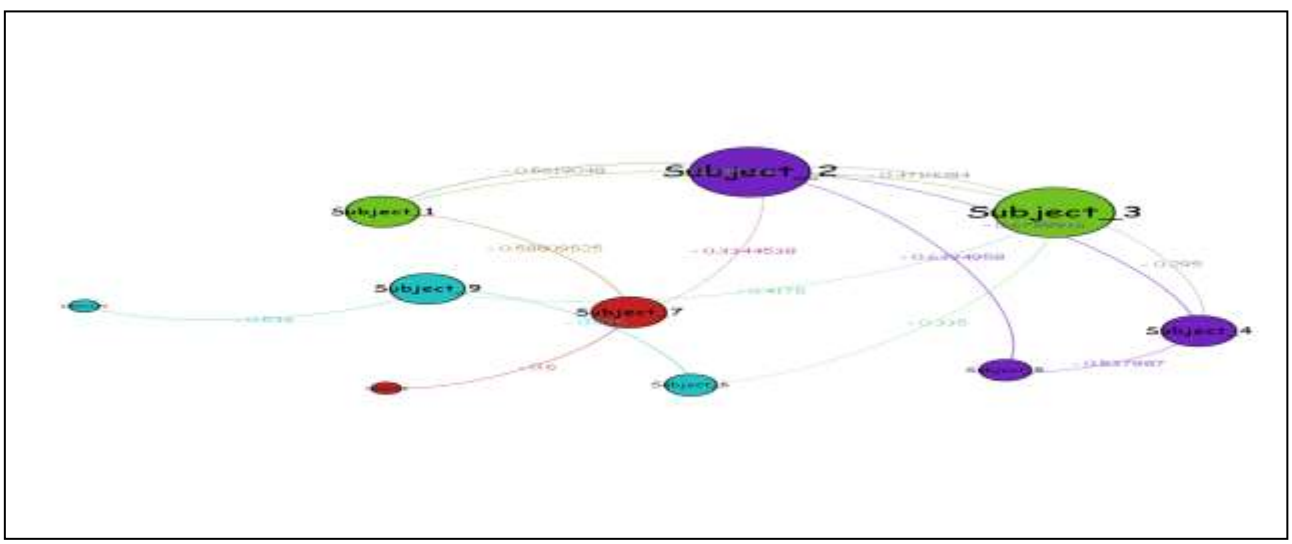


Figure: Community Detection

Using the Gephi tool we detect the community in the network in the figure the same color consist of same community. Then finally we recommend friends within the same community those are yet to be friend. In the time of recommending friends, we check link strength where the link strength is so high they are more cohesive and very good possibility to become friends.

4.2 - Result Analysis:

We have analyzed our result with the user acceptance value and find out the percentage of success. We have also analyses our system using any of the one parameter (mutual friend, group, application use) and got the following result.

	Mutual Friend	**Group**	**App use**	**All three parameters together**
Community Detected	3	5	3	4
Recommended	8	4	7	14
Accepted	5	3	4	10
Acceptance Rate	62.5%	75%	57.14%	71.42%

From the result we can see that if we use only the mutual friend that means the trivial FOF (friend of friend) system then the acceptance rate is less than if we use also group and application use parameter. Though we have experimented our system with a small sub-network because of the unavailability of social networking data, but according to the definition of cohesiveness it is expected that with the combination of this three parameters, the calculation of link strength and recommending friends will be better.

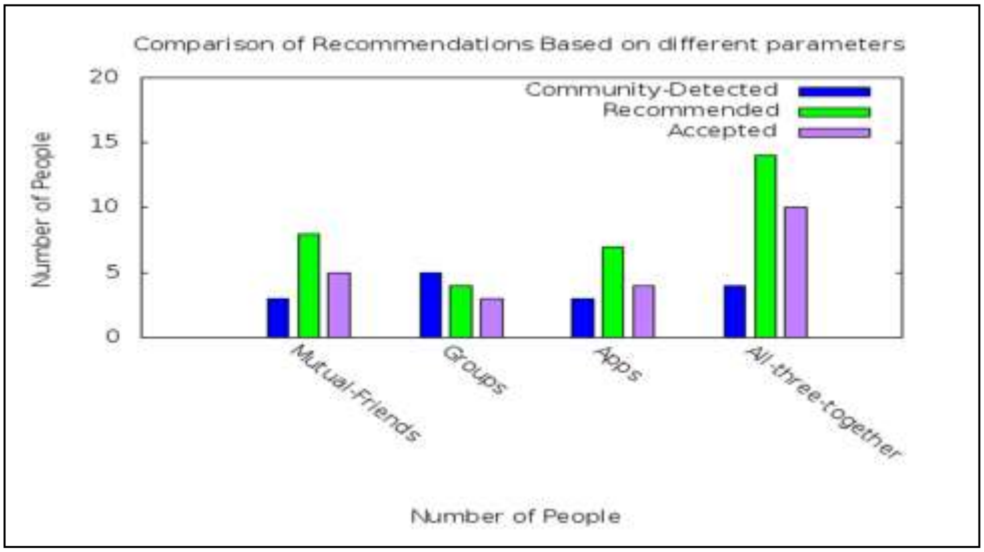


Figure: Result analysis

Chapter 5 – Conclusion and Future Work

With the revolution of web 2.0, Social Networking Sites are getting popularity increasingly. People are joining in those sites to share their views to online friends like their real life friends. For this reason, the size and information of these sites are increasing day by day. Users of these sites hope that the system administrator will provide them recommendation system to make friends. But the recommendation system of human being is not same as the product recommendation as the psychology of human being is different for every person. After realizing this issue, we review the existing system of friend recommendation system in SNSs. And we find that cohesion can be a good measurement in case of social networking recommendation system. So we are proposing this cohesion based friend recommendation system. We hope that this frame work of recommendation system will improve the quality of friend suggestion and will help user to social networking sites.

Though Social networking is now so famous and not that much research has conducted so there are vast scopes to find new technique and upgrade the system. For future work, it is important to test the proposed mechanism more intensively in a larger network using several test groups. Also there may research on improving the existing algorithm of friend recommender system. There is also scope on work NOT to recommend a possible good recommendation than to recommend a bad one. If the system recommends some people who are not at all related to the user then the user might lose their faith from the sites.

The work we have done can be extended in future. There is a huge scope to improvise this technique as friend recommendation is a new research area and create an application on the proposed system. We have used modularity which has the resolution limit so sometimes it cannot detect small communities in a network. Also using unique parameters, we can improve the robustness of the link strength of the network. We have used a sub-network as our experiment; it is extensible to larger network to find out the perfect result of the recommendation system and using a working application.

References

[1] **Herlocker, J. L., Konstan, J. A., and Riedl, J**, "Explaining Collaborative Filtering Recommendations," In Proceeding of ACM 2000 Conference on Computer Supported Cooperative Work, 2000.

[2] **Liben-Nowell, D., and Kleinberg**, J. “The link prediction problem for social networks.” Proceedings of the twelfth international conference on Information and knowledge management, pp 556-559. 2003.

[3] **Guy, I., Ronen I., and Wilcox E**. “Do you know? Recommending people to invite into your social network” Proc. IUI pp. 77-86. 2009.

[4] **Kwon, J. and Kim, S**. “Friend Recommendation Method using Physical and Social Context”. IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.11, November 2010.

[5] **Wolfgang Woerndl and Georg Groh**, “Utilizing Physical and Social Context to Improve Recommender Systems”. In WI- IATW '07: Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, pp. 123-128, 2007.

[6] **M. Hasan, V. Chaoji, S. Salem, and M. Zaki**. “Link prediction using supervised learning.” In Workshop on Link Analysis, Counterterrorism and Security (SDM), 2006.

[7] **H. Kashima and N. Abe**. “A parameterized probabilistic model of network evolution for supervised link prediction.” In ICDM '06, 2006.

[8] **L. Katz**. “A new status index derived from sociometric analysis.” Psychometrika, 18:39-43, 1953.

[9] **Silva, N.,Tsang, I.,Cavalcanti, G., and Tsang, I** .” A Graph-Based Friend Recommendation System Using Genetic Algorithm”. WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Barcelona, Spain CEC.

[10] **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.

[11] **W. H. Hsu, A. L. King, M. S. R. Paradesi, T. Pydimarri, and T.Weninger**. “Collaborative and structural recommendation of friends using weblogbased social

network analysis.” In AAAI Spring Symposia 2006 on Computational Approaches to Analysing Weblogs, 2006.

[12 **Chingching Lin, Shuchuan Lo** “WMR—A Graph-based Algorithm for Friend Recommendation”. Procedings of the 2006 IEEE conference on web intelligence.

[13] **Ahn,Y.Y., Han,S., Kwak, H., Moon, S., and Jeong, H.** “Analysis of Topological Characteristics of Huge Online Social Networking Services”. Proceeding of the 16th International World Wide Web Conference, (Banff, Alberta, Canada, May 8-12, 2007).WWW '07. ACM Press, New York, NY, 835-844, 2007.

[14] **Wilson, M., and Nicholas, C.** “Topological Analysis of an Online Social Network for Older Adults”. Proceeding of the 2008 ACM workshop on Search in Social Media. Napa Valley, California, USA, October 30, 2008). SSM'08. ACM Press, New York, NY, 51-58, 2008.

[15] **Boccaletti, S Latora, V Moreno, Y Chavez, M Hwang, D.** ”Complex networks: Structure and dynamics”.

[16] **Mitchell, Melanie**. “Complex systems: Network thinking”.

[17] **Salton, G., & McGill, M.J. (1983).** ”Introduction to modern information retrieval”. New York: McGraw-Hill.

[18] **Adamic, L.A., & Adar, E. (2003).** “Friends and neighbors on the Web”. Social Networks, 25(3), 211_230.

[19] **Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T.(2002).** “Evolution of the social network of scientific collaboration”. Physica A, 311(3-4), 590-614.

[20] **Katz, L. (1953).** “A new status index derived from sociometric analysis”. Psychometrika, 18(1), 39-43.

[21] **Brin, S., & Page, L. (1998).** “The anatomy of a large-scale hypertextual Web search engine”. Computer Networks and ISDN Systems, 30(1-7), 107-117.

[22] **Jeh, G., & Widom, J. (2003).** “Scaling personalized Web search”. In Proceedings of the 12th International World Wide Web Conference (WWW12) (pp. 271-279). New York: ACM Press.

[23] **Lu, Zhengdong**, “Supervised Link Prediction Using Multiple Sources”.

[24] **Boyd, Danah; Ellison, Nicole**, "Social Network Sites: Definition, History, and Scholarship", Journal of Computer- Mediated Communication, Vol.13, No.1, 2007

[25] **Catanese, Salvatore Meo, Pasquale De Ferrara, Emilio Fiumara, Giacomo.** "Analyzing the Facebook Friendship Graph."

[26] **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.

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

[28] **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.

[29] **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.

[30] muskingum.edu/~psych/psycweb/history/lewin.htm

[31] **Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl**. "Analysis of recommendation algorithms for e-commerce".EC'00, October 17-20, 2000, Minneapolis, Minnesota. Copyright 2000 ACM1-58113-272-7/00/0010

[32]**Gediminas Adomavicius** and **Alexander Tuzhilin.** "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions".

[33] "Social Networks Overview: Current Trends and Research Challenges**"** November 2010 Coordinated by the "nextMEDIA" CSA, Supported by the Future Media Networks cluster. NEXT-Media is supported by FP7, DG Information Society, Unit D2 Networked Media.

[34] **Yu Zheng, Yukun Chen, Xing Xie, Wei-Ying Ma** "GeoLife2.0: A Location-Based Social Networking Service".

[35] http://en.wikipedia.org/wiki

Appendix A

Gephi

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs. Like Photoshop but for data, the user interacts with the representation; manipulate the structures, shapes and colors to reveal hidden properties. The goal is to help data analysts to make hypothesis, intuitively discover patterns, and isolate structure singularities or faults during data sourcing. It is a complementary tool to traditional statistics, as visual thinking with interactive interfaces is now recognized to facilitate reasoning. This is software for Exploratory Data Analysis, a paradigm appeared in the Visual Analytics field of research.

This visualization tool has application on the following areas like,

Exploratory Data Analysis: intuition-oriented analysis by networks manipulations in real time.

Link Analysis: revealing the underlying structures of associations between objects, in particular in scale-free networks.

Social Network Analysis: easy creation of social data connectors to map community organizations and small-world networks.

Biological Network analysis: representing patterns of biological data.

Gephi runs on Windows, Linux and Mac OS X and it is open-source and free. The agronomic interface is based on Netbeans UI and the language of this tool is Java.

For more information and software download, Please follow the link,
https://gephi.org/

Appendix B

Pajek

Pajek is a program for Windows for analysis and visualization of large networks having some thousands or even millions of vertices. In Slovenian language the word pajek means spider. The latest version of Pajek is freely available, for noncommercial use at its home page: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Pajek should provide tools for analysis and visualization of such networks: collaboration networks, organic molecule in chemistry, protein-receptor interaction networks, genealogies, Internet networks, citation networks, diffusion (AIDS, news, innovations) networks, data-mining (2-modenetworks) etc.

The main goals of Pajek are:

- To support abstraction by(recursive) decomposition of a large network into several smaller networks that can be treated further using more sophisticated methods;
- To provide the user with some powerful visualization tools;
- To implement as election of efficient (sub quadratic) algorithms for analysis of large networks.

With Pajek we can find clusters (components, neighborhood of important vertices, cores) in a network, extract vertices that belong to the same clusters and show them separately, possibly with the parts of the context(detailed local view), shrink vertices in clusters and show relations among clusters(global view).

Besides ordinary (directed, undirected, mixed) networks Pajek supports also multi relational networks, 2-modenetworks (bipartite graphs–networks between two disjoint sets of vertices), and temporal networks (dynamic graphs networks changing over time).

For more information and download the manual and command list of Pajek, Please go through the following link:
vlado.fmf.uni-lj.si/pub/networks/**pajek**/doc/**pajek**man.pdf