

SOCIAL BRIDGE: SEARCHING BEYOND FRIEND OF A FRIEND NETWORKS

A THESIS IN  
Computer Science

Presented to the Faculty of the University of  
Missouri Kansas City in partial fulfillment  
of the requirements for the degree

MASTERS OF SCIENCE

By

TEJA SWAROOP MYLAVARAPU

B.Tech, Jawaharlal Nehru Technological University, 2010

Kansas City, Missouri

2012

# SOCIAL BRIDGE: SEARCHING BEYOND FRIEND OF A FRIEND NETWORKS

Teja Swaroop Mylavarapu, Candidate for the Master of Science Degree

University of Missouri-Kansas City, 2012

## ABSTRACT

Social networking has turned into an integral constituent in our lives. There appears to be an imperative demand for finding and linking with others to share one's day-to-day activities. However, currently available search engines for social networking have limited features, such as searches for people mainly by name or finding people within a single domain. With the increasing popularity and complexity of social networks, there is a high demand to enhance current social networks with more advanced features such as, finding people according to their common interests, interaction patterns, or linking someone across domains beyond Friend of a Friend (FOAF) networks.

This thesis aims to develop a social search engine, called the Social Bridge that dynamically generates an integrated social profile that portrays a user's profile of interests and interactions with others and helps him/her in connecting to others who share these common interests and interactions. The Social Bridge expands the FOAF concept of current social networking by defining the social strength that represents the degree of affability among people. Social Bridge is based on the integrated profiles of social networks generated by the level of interactions between friends and their respective interests (e.g., friends, likes, hash tags, etc.) extracted from their Twitter and Facebook profiles. The Social Bridge engine has been implemented using advanced methods and techniques including Information Retrieval Techniques (TF/IDF) and Fuzzy Logic. The Social Bridge framework is compared

with the existing traditional social networking models and the proposed algorithms have proven to be powerful and efficient in finding potential friends for large social networks. The Social Bridge framework has been further evaluated through a survey of social network users for their feedback on its genuineness, correctness, and scalability.

## APPROVAL

The faculty listed below, appointed by the Dean of the School of Computing and engineering, have examined a thesis titled “Social Bridge: Searching Beyond Friend of a Friend Networks,” by Teja Swaroop Mylavarapu, candidate for the Master of Science degree, and certify that in their opinion it is worthy of acceptance.

### Supervisory Committee

Yugyung Lee, Ph.D., (Chair)  
School of Computing and Engineering

Dinakarbandian Deendayal, M.D, Ph.D., M.S  
School of Computing and Engineering

Praveen Rao, Ph.D.  
School of Computing and Engineering

Dennis P. Owens, Ph.D., M.D.  
Department of Psychiatry  
University of Kansas School of Medicine

## TABLE OF CONTENTS

ABSTRACT .....	iii
ILLUSTRATIONS .....	ix
TABLES .....	xi
LIST OF ABBREVIATIONS .....	xii
ACKNOWLEDGEMENTS .....	xiii
CHAPTER	
1. INTRODUCTION	
1.1 Research Motivation .....	1
1.2 Problem Statement .....	2
1.3 Thesis Outline .....	3
2. RELATED WORK .....	4
2.1 Social Computing and Web Mashup on the Internet .....	4
2.2 Social Networking Sites.....	5
2.3 Social Networking Sites as Search Engines.....	5
3. SOCIAL BRIDGE OVERVIEW .....	9
3.1 Conceptual Framework for Social Bridge.....	9
3.2 Network Building in Social Bridge.....	15
3.3 Name Search in Social Bridge .....	26
3.4 Category Search in Social Bridge .....	27
3.5 Advance Search in Social Bridge .....	29
4. SOCIAL BRIDGE FRAMEWORK .....	33
4.1 Introduction .....	33

4.2 Authentication.....	37
4.3 Data Collection .....	39
4.3.1 REST Application Programming Interface.....	40
4.3.2 Graph Application Programming Interface .....	41
4.3.3 Storing Collected Data.....	41
4.4 Data Extraction .....	43
4.5 Data Analysis .....	46
4.6 Network Visualization .....	49
4.7 Dynamic Social Bridge Search.....	51
5. EVALUATION.....	53
5.1 Introduction.....	53
5.2 Experimental Setup.....	53
5.3 Experimental Domains.....	54
5.4 Social Bridge Runtime Performances.....	54
5.5 Social Bridge Search.....	57
5.5.1 Search by Name .....	58
5.5.2 Search by Common Interest.....	59
5.5.3 Search People by Activities .....	62
5.5.3.1 Facebook Interaction.....	62
5.5.3.2 Twitter Interaction .....	63
5.5.4 Search by Integrated Categories .....	64
5.5.5 Search by Bridge Strength .....	65
5.6 Social Bridge Advanced Search.....	69

5.7 Comparison with Other Social Network Search.....	71
5.7.1 Facebook Search .....	71
5.7.2 Twitter Search.....	76
5.7.3 Comparison between Social Bridge Search and Others .....	79
5.8 Social Bridge Survey .....	80
5.8.1 User Profile .....	80
5.8.2 Social Relations and Strength .....	82
5.8.3 Social Bridge Evaluation .....	83
6. CASE STUDY: SOCIAL BRIDGE FOR CLINICAL TRIALS .....	87
6.1 Introduction.....	87
6.2 Communicator for MindFlow .....	88
6.3 Scheduler for Patients .....	93
6.4 Social Bridge in MindFlow.....	95
7. CONCLUSION AND FUTURE WORK .....	99
7.1 Summary .....	99
7.2 Person Identification .....	99
7.3 Hashtag Recommendations.....	100
7.4 Time Complexity and Efficiency.....	101
7.5 Future Work .....	101
REFERENCES .....	102
VITAE.....	105

## ILLUSTRATIONS

Figure	Page
1. Types of Friends .....	13
2. Network Integration .....	14
3. Existing Social Network Connection.....	15
4. Social Bridge Network Connection .....	16
5. Social Bridge Integration .....	17
6. Pseudo Code.....	19
7. Categories Integration.....	21
8. Fuzzy Logic for Twitter .....	22
9. Fuzzy Logic for Facebook .....	24
10. Potential Friends from Integrated Categories .....	29
11. Social Bridge Framework .....	34
12. Authentication Flow.....	38
13. JSON Data Parser .....	45
14. Twitter Visualization .....	50
15. Facebook Visualization.....	50
16. Twitter search Graph.....	51
17. People Search Performance .....	56
18. Term Frequencies of Integrated Categories .....	61
19. Twitter Users' Interactions .....	64
20. Social Bridge Users' Social Strength.....	66
21. Social Bridge Users' Individual Bridge Strength .....	67
22. Twub Cluster Strength .....	69



23. Advanced Search User Interface.....	70
24. Advanced Search Bridge Strength .....	70
25. Social Bridge New Connection.....	71
26. Facebook Search Graph .....	72
27. Facebook Name Search.....	73
28. Facebook Username Search .....	73
29. Facebook Category Search .....	75
30. Twitter Search Graph.....	76
31. Twitter Hashtag Search.....	77
32. Twitter Category Search .....	78
33. Survey Networks.....	80
34. Survey Social Network Hours .....	81
35. Survey Social Network Connection.....	81
36. Survey Social Network Relations and Strengths .....	82
37. Survey Social Network User Friendliness .....	83
38. Survey Social Network Usefulness.....	84
39. Survey Social Network New Friend .....	85
40. Survey Social Bridge Comparison.....	85
41. MindFlow Architecture Diagram.....	89
42. Study Announcement.....	92
43. Session Management .....	93
44. Scheduler for Patients .....	94
45. Appointment Setup .....	95

## TABLES

Table	Page
1. Related Work Search Comparisons.....	6
2. Related Work Social Bridge Mechanisms .....	7
3. Twitter Categories and Weights.....	21
4. Twitter Bridge Strength Variables .....	22
5. Facebook Inferences .....	23
6. Facebook Bridge Strength Variables .....	24
7. Advanced Search Factors.....	30
8. Advanced Search Weights .....	30
9. Hashtag Categories .....	47
10. Facebook and Twitter Data Set.....	54
11. Runtime Performance for Social Bridge Construction and Search .....	54
12. Time-frames for Path Search (Level by Level) .....	55
13. Integrated Friends Search: Search by Name.....	58
14. Integrated Categories Search .....	59
15. Facebook Common Interests.....	62
16. Integrated Strength for Category .....	65
17. Twub Cluster Category Search .....	68
18. Facebook Friends Connection Levels.....	72
19. Summary of the Search Comparison among Social Bridge, Facebook, Twitter .....	79
20. Social Bridge friends Search in MindFlow.....	97
21. Social Bridge Categories Search in MindFlow.....	97

## LIST OF ABBREVIATIONS

FOAF	Friend of a Friend
XML	Extensible Markup Language
JSON	JavaScript Object Notation
URL	Uniform Resource Locator
HTTP	Hyper Text Transfer Protocol
SOAP	Simple Object Access Protocol
RPC	Remote Procedure Call
WSDL	Web Services Description Language
API	Application Programming Interface
REST	Representational State Transfer

## ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the following people who have directly or indirectly helped me in my academic achievements. Firstly, I would like to thank Dr. Yugyung Lee, my mentor and advisor, for her continuous support and guidance throughout my master's program in Computer Science. I sincerely thank Dr. Dinakar Pandian Deendayal, Dr. Praveen Rao, and Dr. Dennis P. Owens for accepting to be a part of my thesis committee and making time for me in their busy schedules. Finally, I would like to thank my parents Girija Mylavarapu and M.S.R. Murty, my little sister Srilekha Mylavarapu, family members, and friends for all their encouragement and support. My achievements would not have been possible without their support.

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the University of Missouri - Kansas City.

## CHAPTER 1

### INTRODUCTION

#### 1.1 Research Motivation

Social networking has become a major part of our lives [1]. We want to connect with people who match our interests and gravitate more towards them. The most common way to search for people is through search engines. However, the existing search engines for social networking have limited searching features, such as searches for people only by their names or within their own domains. The search within a single domain based on a name is not sufficient to find people spread across the network. If someone's name is unknown, it is just impossible to find that person. There is a need to search for people based on their interests and interaction patterns.

Even if people are searched in a single domain based on their names, we cannot completely understand the interaction across different domains (social networking sites) and the associations among different people. We cannot know people's interests, their favorite activities or hobbies, their levels of association with others across various social networking and micro-blogging sites. A person's behavior is not completely understood by studying his profile in a single domain.

The existing social networking sites follow the concept of Friend of a Friend (FOAF) and are able to connect to others who are friends of friends. There are no proper recommendations to search for a person outside the FOAF network. These limitations constrict us from utilizing the features of social computing by not being able to connect to people of our choice.

There is a need to develop a search engine that extends the fundamental concept of searching for a person. There is an absolute need to integrate all the social networking sites and develop a search engine that enables us to search for a person among various networking sites

because people in different organizations rely heavily on social networks [2]. The boundaries of these ubiquitous social networking sites have to be extended to enable us to search for a person on a cross domain network based on name, by common interests, hash tags, favorites, and interaction patterns.

We need to develop a search engine that dynamically generates an integrated social network that portrays user's profiles that comprise their preferences and interactions with others and helps them in connecting with people who share common interests. The Social Bridge expands the FOAF concept of the current social network and introduces the social strength that represents a greater degree of affability among people. Social Bridge has a great potential to help improve the existing mode of connection among people and overcome the limitations. Social Bridge is an endeavor that breaks the frontiers of social networking sites and exemplifies the concept.

## 1.2 Problem Statement

In this thesis, an intelligent integrated search engine called Social Bridge is proposed. It dynamically generates an integrated social profile that portrays a user's profile of interests and interactions with others. This determines the person's interaction towards other people in a cross domain network. The Social Bridge expands the FOAF concept of current social networking and introduces the social strength that represents a greater degree of affability among people.

Given a name, this search engine searches for a person in an integrated social networking site and generates a network that portrays an individual's strengths from all the social networking sites and helps people to get connected. Given a choice of preference, like interests or hash tags, this system semantically searches for people with those preferences and suggests friends. The

choice of search may be further narrowed down like a choice of domain or a choice of categories like interests, hash tags, favorites, retweets, etc., that in turn might find better potential friends.

A visual prototype system has been developed for this thesis for two important domains such as Facebook and Twitter. It has also been compared with an existing traditional social networking model and has proven to be better in terms of efficiency as we have extended the concept of FOAF.

### 1.3 Thesis Outline

In Chapter 2, we present the related work that which compares Social Bridge with Facebook and Twitter. Chapter 3 describes the Social Bridge framework. Chapter 4 explains the Social Bridge implementation. Chapter 5 shows the evaluation and experimental results of measuring the performance of Social Bridge. Chapter 6 discusses the case study of Social Bridge on Clinical Trials. Chapter 7 concludes this thesis and provides the scope for future work on Social Bridge.

## CHAPTER 2

### RELATED WORK

In this chapter, we will review various mechanisms that contribute significantly to the areas of social networks and search engines. In Section 2.1, we discuss the concepts of social computing and Web mashup on the Internet contribute to the process of integrating data from various sources. In Section 2.2, we discuss social networking sites that collect social and professional data from people. In Section 2.3, we discuss social networking sites that also act as search engines that are confined to restrictions and improve the search efficiency in the social network.

#### 2.1 Social Computing and Web Mashup on the Internet

Social computing [3] is mainly explained as graph theory with human beings as nodes and the relationship among them as edges. It is a structure wherein the strength of the relationship between individuals is explained based on the relation (bonding) between them. Social computing has become more widely known because of its relationship to a number of recent trends like the popularity of social software and Web 2.0.

A Web mashup [4] is a web application that gathers content from different sources and presents it in a different way or with a unique outline. The Web is continually growing more open and more social. Because of this, many websites have opened up programming interfaces (API's) that allow developers to get at their core information. The main characteristics of the mashup are combination, visualization, and aggregation. This has paved the way for new concepts to emerge and new challenges to be solved.



## 2.2 Social Networking Sites

Jothi et al. [5] states that social networking sites today can be considered as the best platform for effective communication. Communication is the form of connecting to people or in the form of advertising brand products. Facebook collects all the personal data of users who have registered with them and displays ads as per the choice of the person's profile. Jones et al. [6] claims that Facebook is a threat to privacy. But it is the users' choice of preference in sharing their personal data. There are certain privacy settings in Facebook that might lock down one's personal data and will never be shared with anyone. Twitter is a micro-blogging site that allows 140 characters in the form of tweets to reach different people. Romero et al. [7] claims that tokens in the form of hashtags are spread out to enhance the communication between persons in twitter. Hashtags are one of the key concepts in the micro-blogging world as it is emerging as a winner in helping people of common interests communicate with each other, thereby serving the purpose of its existence.

## 2.3 Social Networking Sites as Search Engines

SHRM (Society for Human Resource Management) [8] surveyed findings that focused on the use of social media in the workplace and conducted surveys on social networking websites and online search engines as a tool for screening potential job candid

ates. Because of the rise in the use of social networking websites and online search engines, more organizations have already developed, or are in the process of developing, policies about the use of these methods for screening job candidates. Integrated social networking search engines might be a potential tool in analyzing a person's professional and social life and emerge as a successful tool in gauging the best candidate for a particular job.

The table below briefly compares the searching factors of Facebook, Twitter, and Social Bridge.

Table 1. Related Work – Search Comparisons

<b>Feature</b>	<b>Facebook</b>	<b>Twitter</b>	<b>Social Bridge</b>
Search a person in a cross domain network	No	No	Yes
Search by interests	No	No	Yes
Search by hashtags	Not Applicable	Yes	Yes
Search by Activities (status message, photo tags)	No	Not Applicable	Yes
Search by association (Retweets and Favorites)	Not Applicable	No	Yes
Searching concept	Friend of a Friend	Friend of a Friend	Extends Friend of a Friend
Determine Individual Strength (Bridge strength)	No	No	Yes
Search in a cross domain network [9]	No	No	Yes
Hashtags classification (Twub Cluster) [10]	Not Applicable	No	Yes
Advance Search	No	No	Yes

The above table clearly explains the various searching mechanisms of Social Bridge, Facebook and Twitter. It clearly explains the inability to search a person in a cross-domain

network in Facebook and Twitter and the ability to search a person in Social Bridge. Table 1 also shows various searching techniques like search by interest, search by hashtags, search by association, advance search which proves the better discovery of people in a cross-domain network using Social Bridge. Social Bridge also determines an Individual Bridge Strength which determines the strength of the person in the social network.

Table 2. Related Work-Social Bridge Mechanisms

<b>Approach</b>	<b>Authors</b>	<b>Field</b>	<b>Mechanism</b>
Social Network Profile Analysis	Gartel et al. [12]	Facebook Likes	Ranking mechanism
Performance activity	Tiffany et al. [1]	Facebook Social Activity	Content posts, Profile updates, Group participation
Social Network Integration	Jung et al. [12]	Twitter and Facebook User Identification	Context Fusion, user scores.
Semantic relevant applications	McMillan et al. [13]	Twitter Hashtags	Hashtags recommendations and classification
Interaction Patterns	Tiffany et al. [1]	Facebook and Twitter (Likes and Hashtags)	User similarity with Likes and Hashtags
Term Frequency-Inverse Document Frequency	Juan Ramos	Twitter and Facebook Likes and Twub Cluster	Determining category frequency in the integrated network

Table 2 explains the various mechanisms applied in building Social Bridge. The table mentions the techniques such as Interaction patterns applied for likes and hashtags, Term Frequency –Inverse Document Frequency (TF-IDF) for determining category frequency in the integrated network, performance activity for various profile groups.

## CHAPTER 3

### SOCIAL BRIDGE FRAMEWORK

Social Bridge is a semantic framework which expands the Friend of a Friend (FOAF) concept of current social networking sites to discover and connect people based on their association, common interest and their interaction.

In Section 3.1, we explained the conceptual framework for Social Bridge. In Section 3.2, we discuss the Network building in Social Bridge. Section 3.3 discusses the search by the name of a person in an integrated profile. Integrated profile is acquired from Facebook and Twitter explaining the person's association and his/her interests. Section 3.4 discusses the search by interests and by Twub Cluster. Section 3.5 discusses the advanced search where the factors (Hashtags, likes, retweets) of search is completely left to the user

#### 3.1 Conceptual Framework for Social Bridge

Social Bridge is a semantic framework which dynamically generates an integrated social network that portrays users' profiles of interests and interactions with others and helps them in connecting to others who share these common interests and interactions. Social Bridge enables us to search for a person on cross domain integrated social networks unlike the ubiquitous search engines. Social Bridge extends the concept of FOAF of current social networking sites and introduces the individual bridge strength that represents the degree of affability among people.

Social Bridge approaches a three dimension methodology to determine a person's strength in a social network. They are i) Association, ii) Common Interest, iii) Interaction.

- i) Association: Association of a person is determined by obtaining the number of friends, followers and people following a person in a social network. This

determines a person's association on how he/she associated with people. Friends play a very key role in determining the strength of association of a person in an integrated social networking site. The number of friends from Facebook and Twitter are extracted and analyzed.

- ii) Common Interests: We also determine the list of common friends from Facebook and Twitter that determines the stronger friends of the lot in a cross-domain network. Individual Social Bridge strength for association of a person in a single and integrated social network is determined. Common interest of the person is obtained by analyzing people's hashtags and likes. Hashtags are obtained from Twitter and likes are obtained from Facebook. These determine the interest of the person and common interests are determined further by identifying common set of interests among all the interests obtained.
- iii) Interaction: Interaction of a person is determined. Interaction of a person is determined by analyzing the person's retweets, favorites, status likes, status comments, photo likes, photo comments, photo tags.

Bridge strength is determined as a sum of fuzzy values of association, common interest and interaction of a person pertaining to a domain (Twitter or Facebook). This determines the person's strength in a single social network.

Based on the association, common interest and interaction, bridge strength for specific domain of a person is determined. Social Bridge connects people beyond FOAF relationships in a single domain. It connects people who are not friends in the existing social network (single domain) by discovering potential friends with the bridge strengths.

i) Potential friend

Potential friend is a person who is currently not connected as a friend in the existing domain, but has similar interests matching to become a friend. Social Bridge enables us to discover new friends in the existing social networking sites.

ii) Friends with common interests

Friends with common interests are people who share common interests among them. Common interests include 'likes' from Facebook or 'hashtags' from Twitter. People having these factors in common are potentially connected as friends.

iii) Friends with common friends (association)

People who have many common friends in common are considered to have stronger association with each other and can be considered for potential connection. These people are connected for having friends common in their profiles.

For connecting people in integrated networks, the following concepts apply.

- i) Integrated Social Network: Integrated Social Network is defined as a social structure formed by nodes (people) and edges (relationships) interconnected to each other. Nodes and edges are structured as an integrated network by gathering data from two different social networks (Facebook and Twitter). People registered across these networks are articulated and made available to a wider integrated network to enhance efficient connection. This social network perspective gives a way of analyzing the structure and connects people from different social networks.

- ii) Node: A node is a terminal point or an intersection point of a graph. It is the fundamental unit of which graphs are formed. In Social Bridge, nodes are people which are connected to one another in a social network.
- iii) Edge: An edge  $e$  is a link between two nodes. A link is the abstraction of a structure supporting movements between nodes. It has a direction that is commonly represented as an arrow. When an arrow is not used, it is assumed the link is bi-directional. In Social Bridge, an edge is considered as a relationship among people (nodes).
- iv) Integrated friend: Integrated friend is a person who is currently not connected as a friend in the cross (different) domain, but has similar interests, common friends matching to become a friend and get connected.
- v) Individual Bridge Strength: Individual Bridge Strength is defined as the mean of integrated individual strength (fuzzy strengths) of the person from all social networking sites in this integrated social networking domain. This is calculated by assessing all the individual strength's (contributing friends, likes, hashtags, retweets, status comments,, status likes, favorites, photo likes, photo comments) of the person from Twitter and Facebook. This bridge strength ascertains the trust of the person in the integrated social networking site.

Relationships in social networks are measured based on the way one person connects to another. People in social networks are commonly associated to each other via a direct friend or through a Friend of a Friend. Friendship determines the basic mode of connection. This friendship, or connection, generally happens if two persons meet in person, or if they are classmates, colleagues, relatives, friends, partners, etc.





interest and their association towards a topic of interest, are analyzed applying fuzzy logic in Social Bridge. Fuzzy logic is applied in determining individual bridge strength which estimates the association, interest and interaction of a person in an integrated social network. Activities in an integrated social network are analyzed considering friends, likes, hashtags, retweets, comments, tags, favorites, etc. Based on these factors, individual bridge strength is determined that helps users to study their profiles in detail.

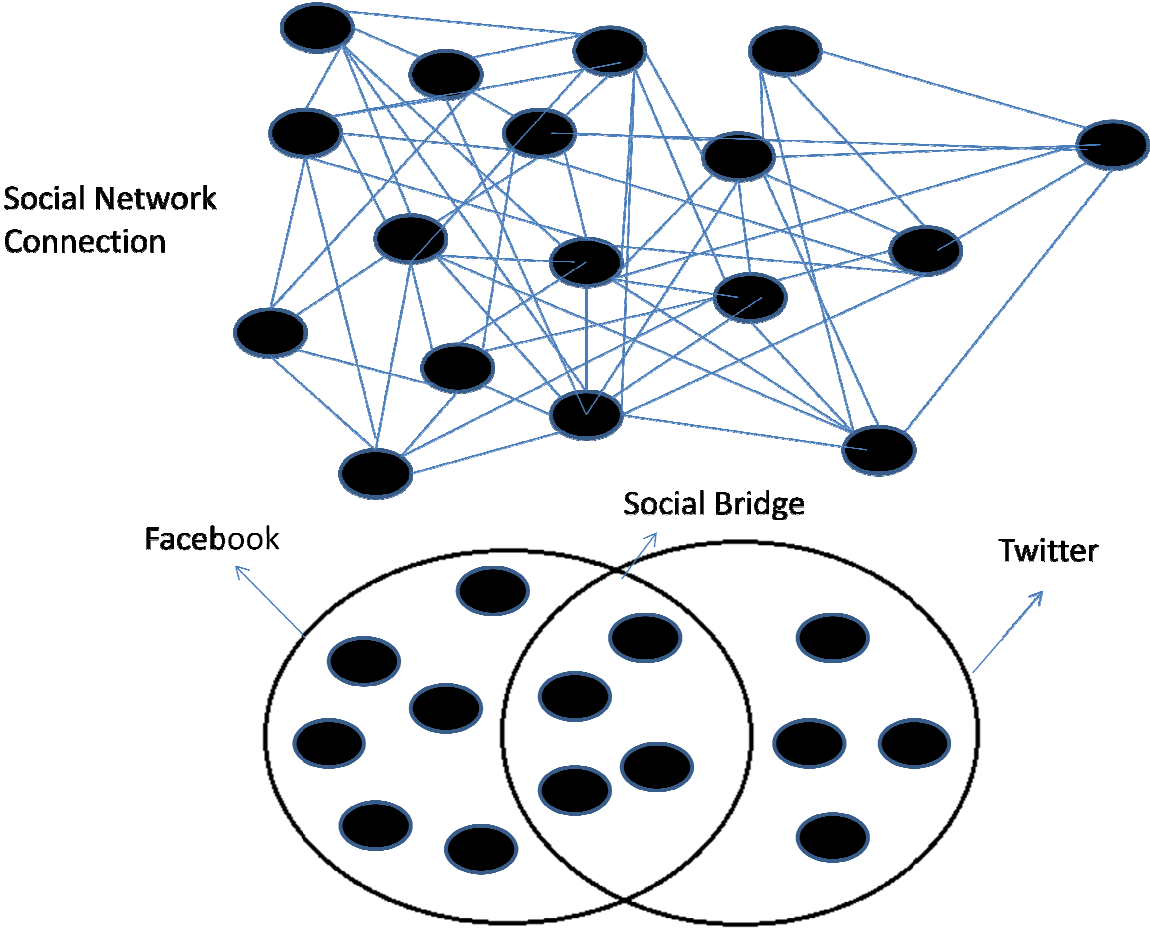


Figure 2. Network Integration

### 3.2 Network Building in Social Bridge

Survey results show that people in the existing social network are connected with one another through Friend of a Friend or with people whom they meet in person.

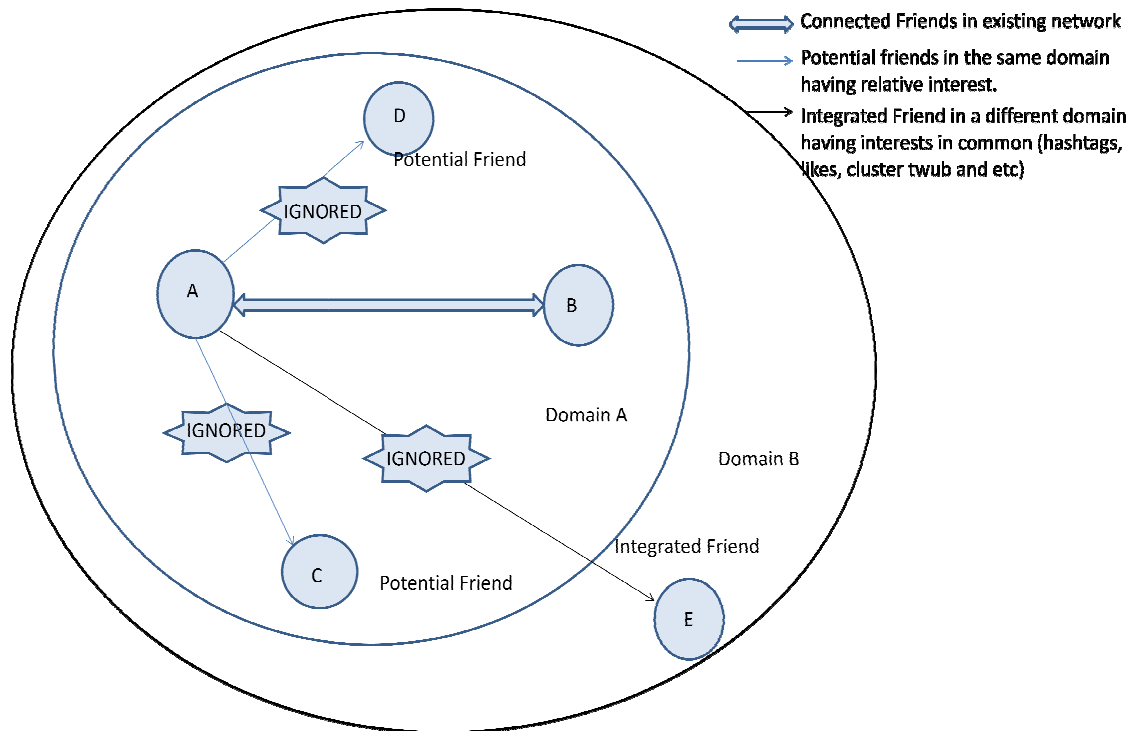


Figure 3. Existing Social Network Connection

The figure above shows the connection among people in the existing network where friends 'A' and 'B' get connected with Friend of a Friend. However, there are people who match with person 'A' within the same domain and from a different domain. People within the same domain are termed as potential friends within the domain and people from different domains are termed as integrated friends. The existing social networking sites are missing out on the method to connect to people in this regard. There is a potential possibility to connect people who are not in a person's friends list and with whom we share common interests (likes, hashtags, Twub Cluster).

Social Bridge reaches out to different people in connecting friends by finding new potential friends within the same domain and by discovering people from various other domains in the form of integrated friends.

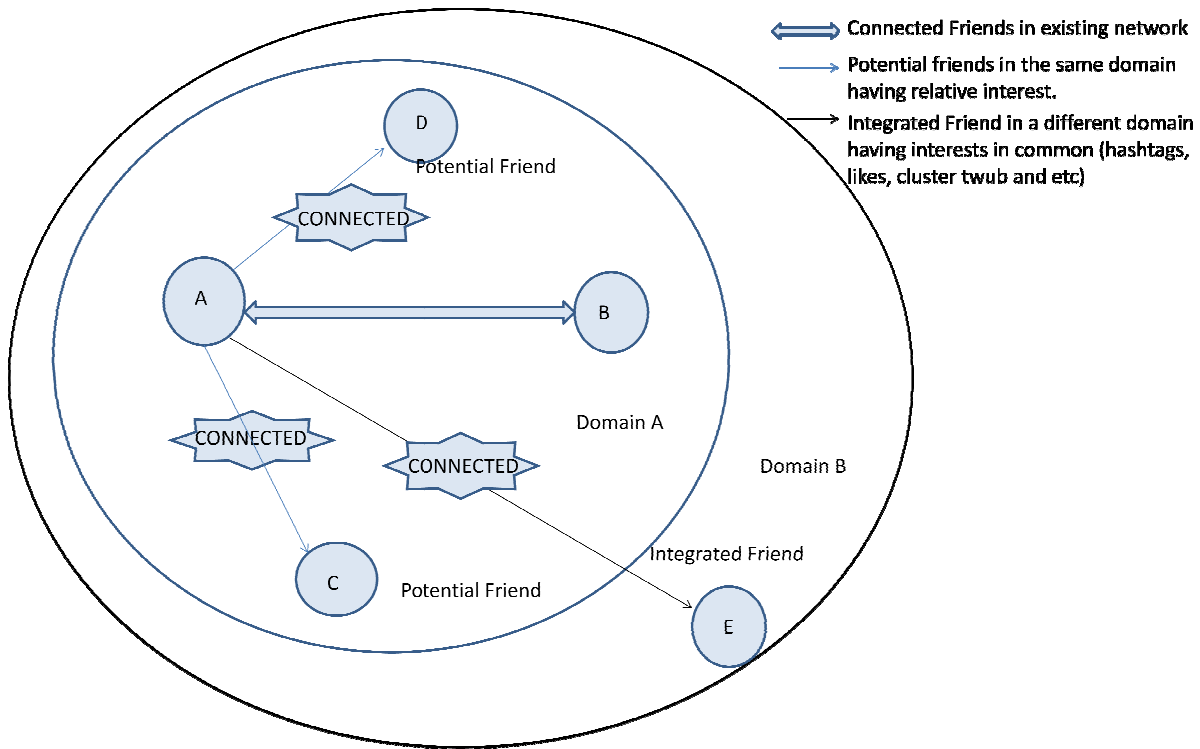


Figure 4. Social Bridge Network Connection

The above diagram displays that people from the same domain and people from various other integrated domains are discovered and potentially connected. We determine the level of association of people towards their other friends by determining their common interests these people share among themselves, the number of times they have commented on their messages, photos, videos, etc. This only determines a person's interaction on how he/she interacts with a person and how active is the person in the social network. We can determine the person's

responsiveness and his activity in the social network. Social Bridge initially builds a network by gathering data from different people from same domain and then gathers data from cross domain.

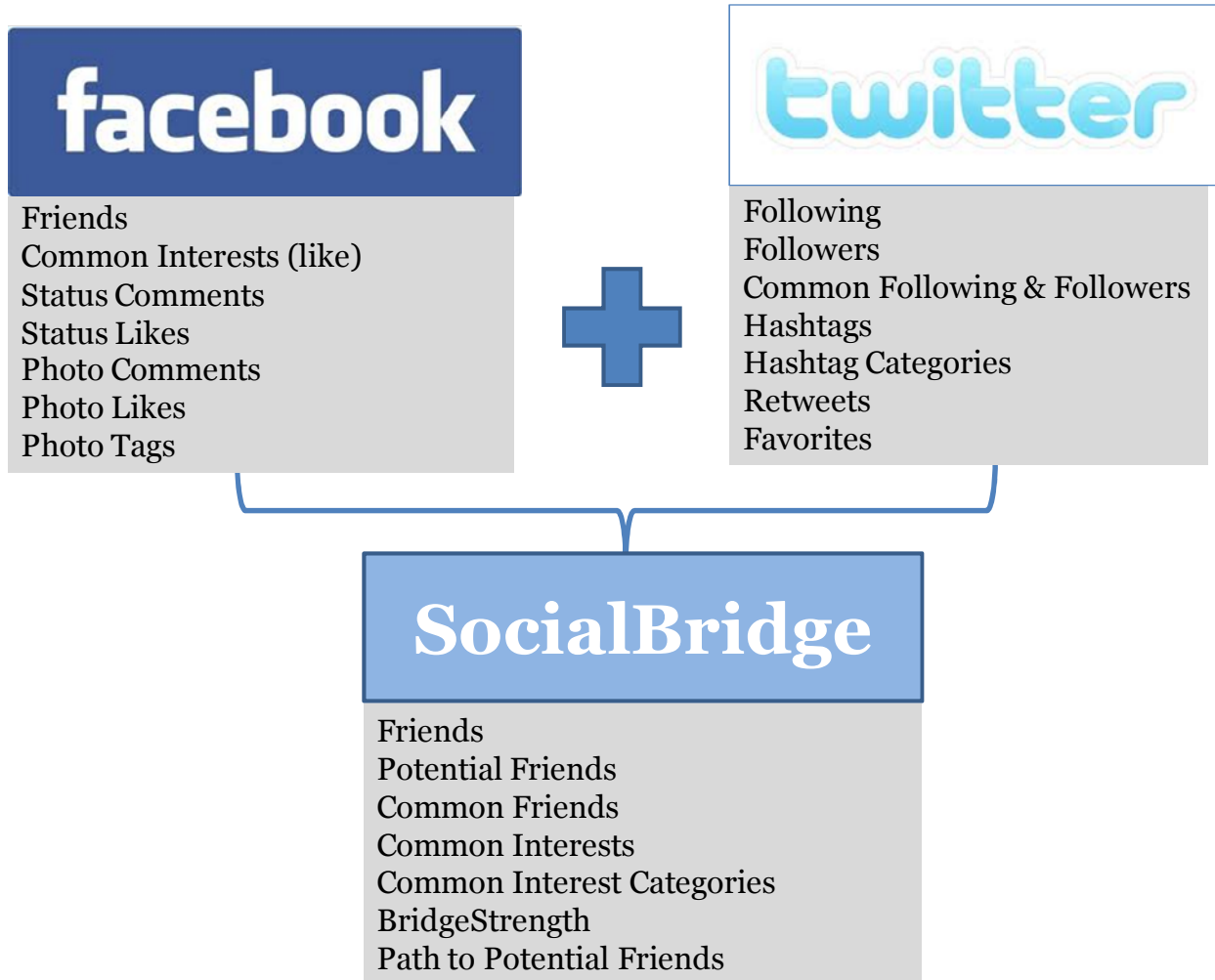


Figure 5. Social Bridge Integration

We mainly determine individual strength by analyzing the likes, hashtags. Term Frequency-Inverse document frequency is analyzed to determine the importance of likes or hashtags of a person on an overall basis.

Term Frequency is defined as the number of times a term  $t$  has occurred in a particular document  $d$ . The inverse document frequency is a measure of whether the term is repetitive or rare across the set.

### **Term Frequency:**

Term Frequency for likes is calculated by counting each and every category occurrence to the total number of categories enlisted.

$$TF_u(l) = \text{categoryCount}(u, l_i);$$

$u$  = username,  $l_i$  = like category

### **Inverse Document Frequency:**

The inverse document frequency is a measure of whether the term is repetitive or rare across the documents.

$$IDF_u = \text{Log}\left(\frac{\text{userCount}(A)}{\text{userCount}(l_i)}\right);$$

$A$  = All categories,  $l_i$  = like category

### **Term Frequency - Inverse Document Frequency:**

$$tf * idf(u) = (TF_u(l_i) \times IDF_u(l_i))$$

$$tf * idf(u) = \text{categoryCount}(u, l_i) \times \text{Log}\left(\frac{\text{userCount}(A)}{\text{userCount}(l_i)}\right)$$

`categoryCount()` is a method that counts the number of times the category has occurred in the document to determine the frequency. `userCount()` determines the number of users pertaining to the specific category.

Pseudo code for the above Term Frequency – Inverse Document Frequency is below:

```
for (int i = 0; i < FB_likes_distinct.Count; i++)
{
    FB_ppl_categcount = fb.getCategorypplcount(FB_likes_distinct[i].ToString());
    TF = Convert.ToDouble(count_likes_distinct[i]) / Convert.ToDouble(pivot_indistinct_likes.Count);
    IDF = Math.Log(((FB_totalppl_count) / (FB_ppl_categcount)));
    TF_IDF = TF * IDF;
}
```

Figure 6. Pseudo Code

Term Frequency for ‘likes’ is calculated by counting every category occurrence of the person to the total number of categories enlisted. A high term frequency indicates the user likes a category that is of high frequency. A high Inverse Document Frequency indicates the ‘likes’ category is rare among people.

An example which calculates Term Frequency-Inverse Document Frequency is explained. We determine people who have ‘Athlete’ as their interested category. Using Term Frequency, the ratio of the count of the category ‘Athlete’ to the total count of all the interested categories of the particular user is evaluated. This is later evaluated on a percentage basis. People having a high ratio are given the first priority over the other. This sorting enables the person searching to quickly connect to the highest person on the list as he/she is the strongest in the particular category.

These factors help us in assessing the person’s inclination towards a particular topic among all the topics available. This determines how frequently a person associates with a particular category among all of them. If the frequency towards a particular category is high, then this person has more interest towards that particular category than towards the other

categories. So, if a user is searching for a person who is interested in that particular category, then this person can be the best connection choice.

There is no proper classification of hashtags in Twitter. Hashtags classification is done in the form of Twub Cluster in Social Bridge. All the hashtags of the user are collected and categorized within a range of 175 categories. All the hashtags of the user are collected and categorized within a range of 175 categories that are deduced from Facebook. All the categories of Facebook have been considered in order to build Twub Cluster. In Social Bridge, we have classified hashtags in our own Twub Cluster. Twubs are Twitter groups built around content aggregated from hashtags. This search enables us to search for a person with interested categories (Music, Sports, and Health, etc).

Twub Cluster is defined as a repository of hashtags classification that includes the clustering of hashtags of various users into the different categories. Classification is done based on the meaning of the hashtag and mapping it with the associated category. Apart from the hash tags classification, we have a section that suggests more hash tags in the form of ‘recommended hashtags’.

These sets of words are very useful in getting the level of association of this person with any other person in our database. With these sets of hash tags, we can search for people who have tweeted with these hash tags and classify their interests. By doing this, we can get more people who are interested in these categories and analyze their association towards each other. After obtaining the ‘likes’ and ‘hash tags categories’ of these people, we determine the strength of association of this person towards those categories.



When the names of people are searched, their hashtags are further analyzed and the set of categories to which their hashtags are mapped are retrieved. We also determine the term frequency and inverse document frequency of the hashtag from the various sets of categories and determine the frequency of a particular category from the various categories to which the persons searched are associated. This also enables the user to know the association of any person towards a particular category from the various categories available.

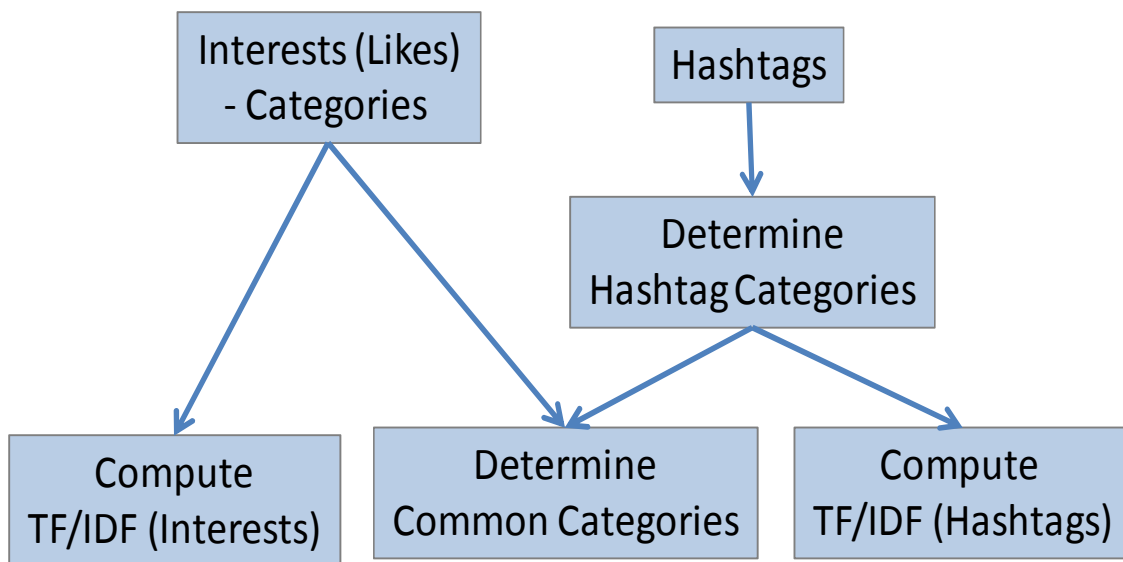


Figure 7. Categories Integration

After obtaining the association, common interests and interaction of a person, we further calculate the individual bridge strength of the person. These values are next carried forward to calculate Fuzzy Logic, which determines the level of interaction of this person with all his/her friends. This individual bridge strength is calculated for all the domains on a cross domain platform that shows all the person’s details in an integrated social network.

**Fuzzy logic inferences for Twitter:**

Table 3. Twitter Categories and Weights

Weights	Categories					
	Interaction ( $W_i$ )		Interests ( $W_n$ )	Networking ( $W_n$ )		
	Retweets ( $W_{t1}$ )	Favorites ( $W_{t2}$ )	Hashtags ( $W_{n1}$ )	Friends ( $W_{n1}$ )	Followers ( $W_{n2}$ )	Common ( $W_{n3}$ )
	0.01	0.01	0.01	0.01	0.01	0.01

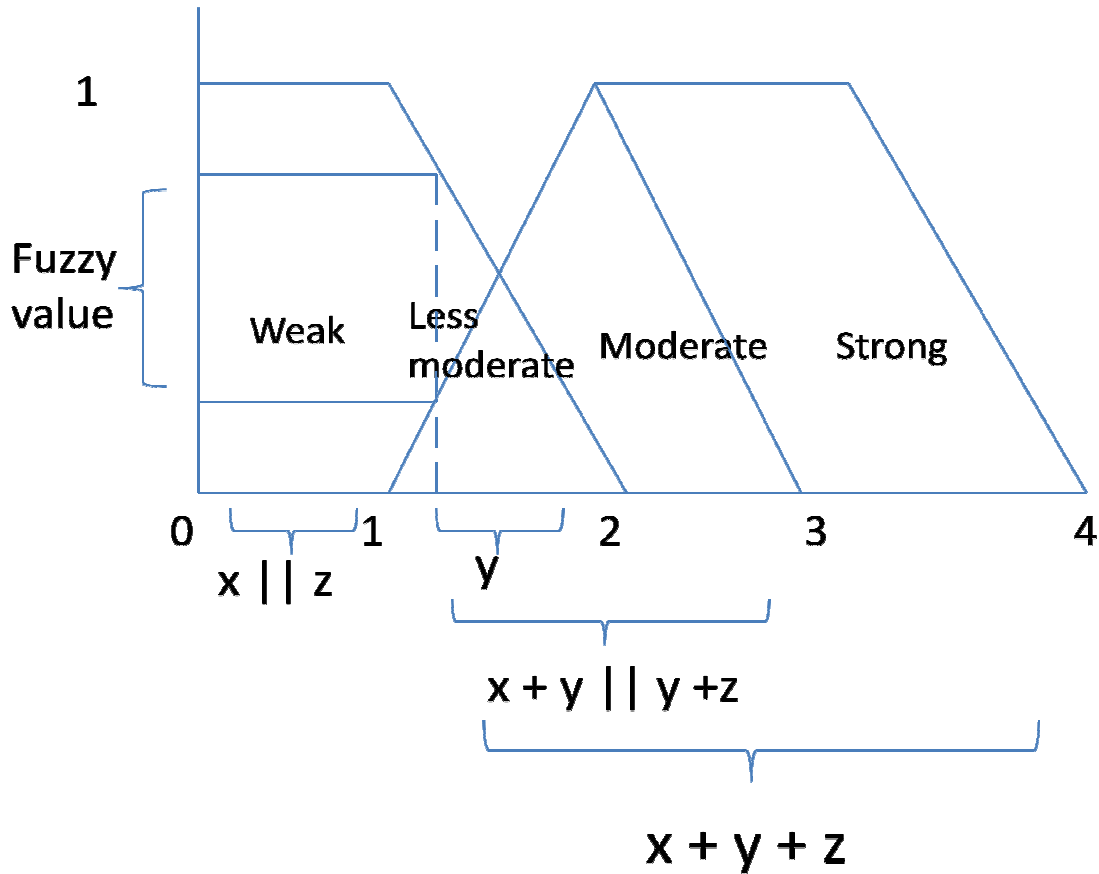


Figure 8. Fuzzy Logic for Twitter

The following table shows the variables used for the Twitter bridge strength computation.

Table 4. Twitter Bridge Strength Variables

X	This represents the count of interaction. $0 \leq x \leq 1$
Y	This represents the count of common interests $1 \leq y \leq 2$
Z	This represents the count of association. $0 \leq z \leq 1$ .

$x + y$	This determines the sum of the counts of interaction and common interests excluding association The range of the values is $1 < x + y < = 3$ .
$y + z$	This determines the sum of the counts of association and common interests excluding interaction The range of the values is $1 < y + z < = 3$ .
$x + y + z$	This determines the sum of counts of association, common interests and interaction. The range of the values is $1 < x + y + z < = 4$

The TwitterBridgeStrength is defined as follows:

$$\text{TwitterBridgeStrength}(x, y, z) = \sum_{k=1}^n W_k(oc_k)$$

Where,  $W_k$  is the weight of each variable and  $oc_k$  is the number of occurrences of interactions, interests, and friends.

Measure the range of the value and thereby determine the Fuzzy Strength that also depends on the occurrences.

If TwitterBridgeStrength lies between 0 and 1 = Weak

If TwitterBridgeStrength lies between 1 and 2 = Less Moderate

If TwitterBridgeStrength lies between 2 and 3 = Moderate

If TwitterBridgeStrength lies between 3 and 4 = Strong

Table 5. Facebook Inferences

Weights	Inferences						
	Interaction					Interests	Networking
	Status likes	Status comments	Photo comments	Photo likes	Photo tags	Likes	Friends
	0.01	0.01	0.01	0.01	0.01	0.01	0.01

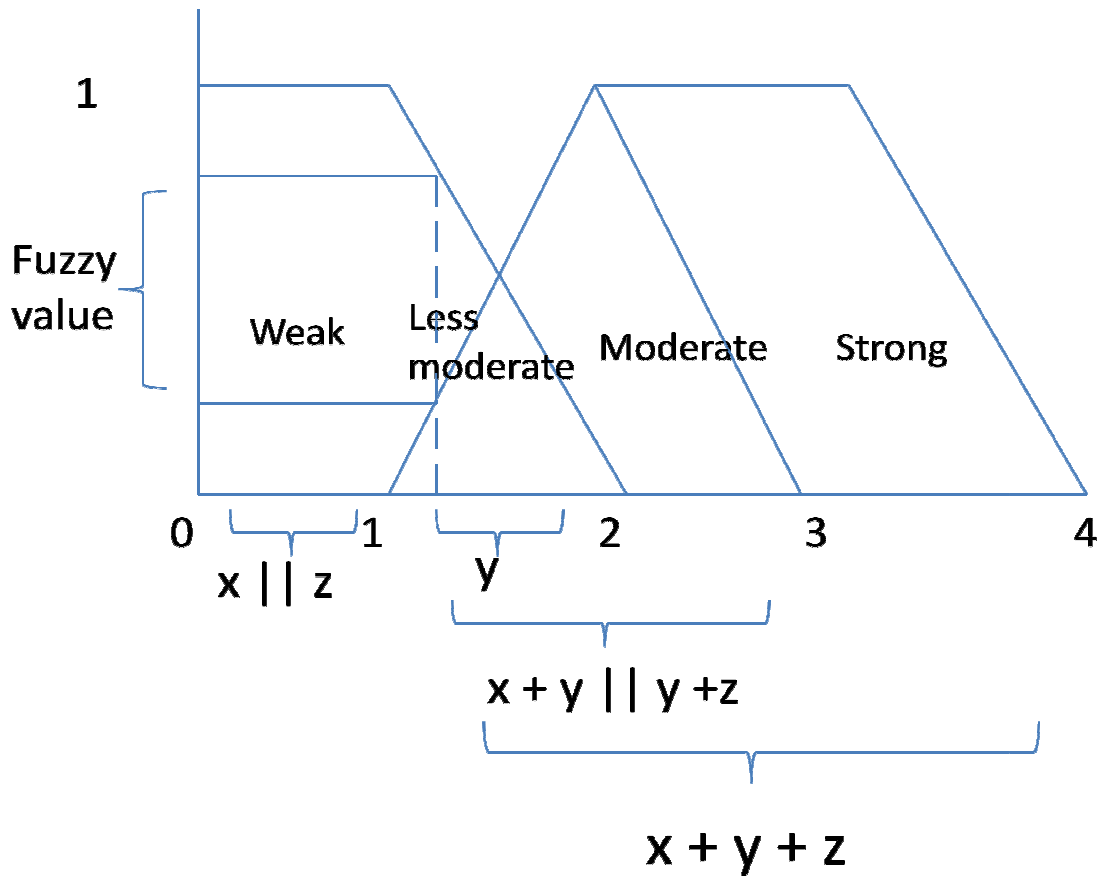


Figure 9. Fuzzy Logic for Facebook

The following table shows the variables used for the Facebook bridge strength computation.

Table 6. Facebook Bridge Strength Variables

X	This represents the count of interaction. The range of the values is $0 \leq x \leq 1$
Y	This represents the count of common interests. The range of the values is $1 \leq y \leq 2$
Z	This represents the count of association. The range of the values is $0 \leq z \leq 1$
$x + y$	This determines the sum of the counts of interaction and common interests excluding association. The range of the values is $1 < x + y \leq 3$ .
$y + z$	This determines the sum of the counts of association and common interests excluding interaction. The range of the values is $1 < y + z \leq 3$ .

$x + y + z$	This determines the sum of counts of association, common interests and interaction. The range of the values is $1 < x + y + z \leq 4$
-------------	---

The FacebookBridgeStrength is defined as follows:

$$\text{FacebookBridgeStrength}(x, y, z) = \sum_{k=1}^n W_k(oc_k)$$

Where,  $W_k$  is the weight of each variable and  $oc_k$  is the number of occurrences of interaction, interests and friends

Measure the range of the value and thus determine the Fuzzy Strength that also depends on the occurrences.

If FacebookBridgeStrength lies between 0 and 1 = Weak

If FacebookBridgeStrength lies between 1 and 2 = Less Moderate

If FacebookBridgeStrength lies between 2 and 3 = Moderate

If FacebookBridgeStrength lies between 3 and 4 = Strong

This bridge strength ascertains the trust of the person in the integrated social networking site. This means that the higher the bridge strength, the higher the strength and the lower the bridge strength, the lower the strength. Bridge strength enables the people searching for a particular person to understand a person's level of interaction in the integrated social networking site.

$$\text{IndividualBridgeStrength} = \left( \frac{\text{FacebookBridgeStrength} + \text{TwitterBridgeStrength}}{2} \right)$$

The individual bridge strength determines the person's individual strength in the integrated social networking site that is within a range of 0 – 4.

- If IndividualBridgeStrength tends towards 0, this means the association of the person in the integrated social networking site is weak.

- If IndividualBridgeStrength tends towards 4, this means the association of the person in the integrated social networking site is strong and this person can be contacted for future connections.

This individual strength enables us to determine a defined value within a range of 0-4 that determines the strength of association. A higher IndividualBridgeStrength means a higher association, and a lower IndividualBridgeStrength means a lower association.

### 3.3 Name Search in Social Bridge

After an integrated network has been built, the network is ready to connect and discover new people. One of the ways to connect people in this integrated social network is by discovering people by name. This search lists a person's complete information (integrated profile) that lets the user understand the person's integrated profile comprising details from Facebook and Twitter. Associated strengths from likes, hashtags, Twub Cluster, and overall individual bridge strength is displayed, which determines the person's activity in the integrated site.

This search initially starts by checking the person's details from two domains, namely, Twitter and Facebook. This method gathers all the data from both domains and integrates them together to display the integrated profile of the person searched. This method displays the list of the person's association, common interests and interaction which determines the person's individual bridge strength in the integrated social networking site. Strengths that are evaluated are assigned to the person searched and determine the person's level of association in the integrated social networking site.

A person of higher strength implies higher interaction and association in the social networking sites and a person of lower strength implies lower interaction. This is calculated by applying the concept of fuzzy logic to determine the individualistic bridge strength.

As we can observe from the above diagram, person 'E' is searched in the integrated social network and is discovered during the search by name. The whole integrated profile is observed and analyzed by the user who searched for the person and decides on the friend connection. This search enables us to connect to new persons from different domains to enhance the mode of connection.

### 3.4 Category Search in Social Bridge

What if we do not know the name of the person to connect? What if we do not know whom to contact to get connected? What if we want a set of persons who are interested only in a set of a particular category?

Social Bridge introduces a unique way of searching for people even without knowing their names or location. Social Bridge connects to people whom the user might not know at all. It enables the connection of people from all over the world by extending the concept of Friend of a Friend. Even though the person is not in the user's friend list, Social Bridge enables us to connect them together.

Search is carried forward in three ways:

- a) Search by 'likes'
- b) Search by 'hashtags'
- c) Search by 'likes' and 'Twub Cluster'

a) Search by 'likes':

This search follows an approach to search for people exclusively in Facebook and determines those who share relatively strong interests within the category search. An intensive searching technique is followed that retrieves people who have their Term Frequency strongly associated towards the specific category searched over all the other categories pertaining to that person. People of higher value are given priority over others. This lets the user choose people who have a strong association towards the specific category.

b) Search by 'hashtags':

This search follows an approach to search for people exclusively in Twitter and determines those who share relatively strong interests with the hashtag searched. This searching technique retrieves people whose term frequencies are strongly associated towards the specific hashtag that was found by searching through all the other hashtags pertaining to those people. All the people are enlisted and those of higher value are given priority over the others. This lets the user to choose people who have a strong association towards the specific hashtag, unlike in Twitter that retrieves tweets containing the hashtag.

c) Search by 'likes' and 'Twub Cluster'

This search follows a slightly different technique for discovering people. When a person searches for people within a specific category, this search looks for the category in Likes of Facebook and Twub Cluster of Twitter. An intensive search is carried out and people from different domains are retrieved for the end user. Their Term Frequencies are determined and people having similar interests from both domains and of higher frequencies, are given higher priority over the others. This is considered unique as the search is carried forward



from two different domains and people having relative strong interests are retrieved for the end user.

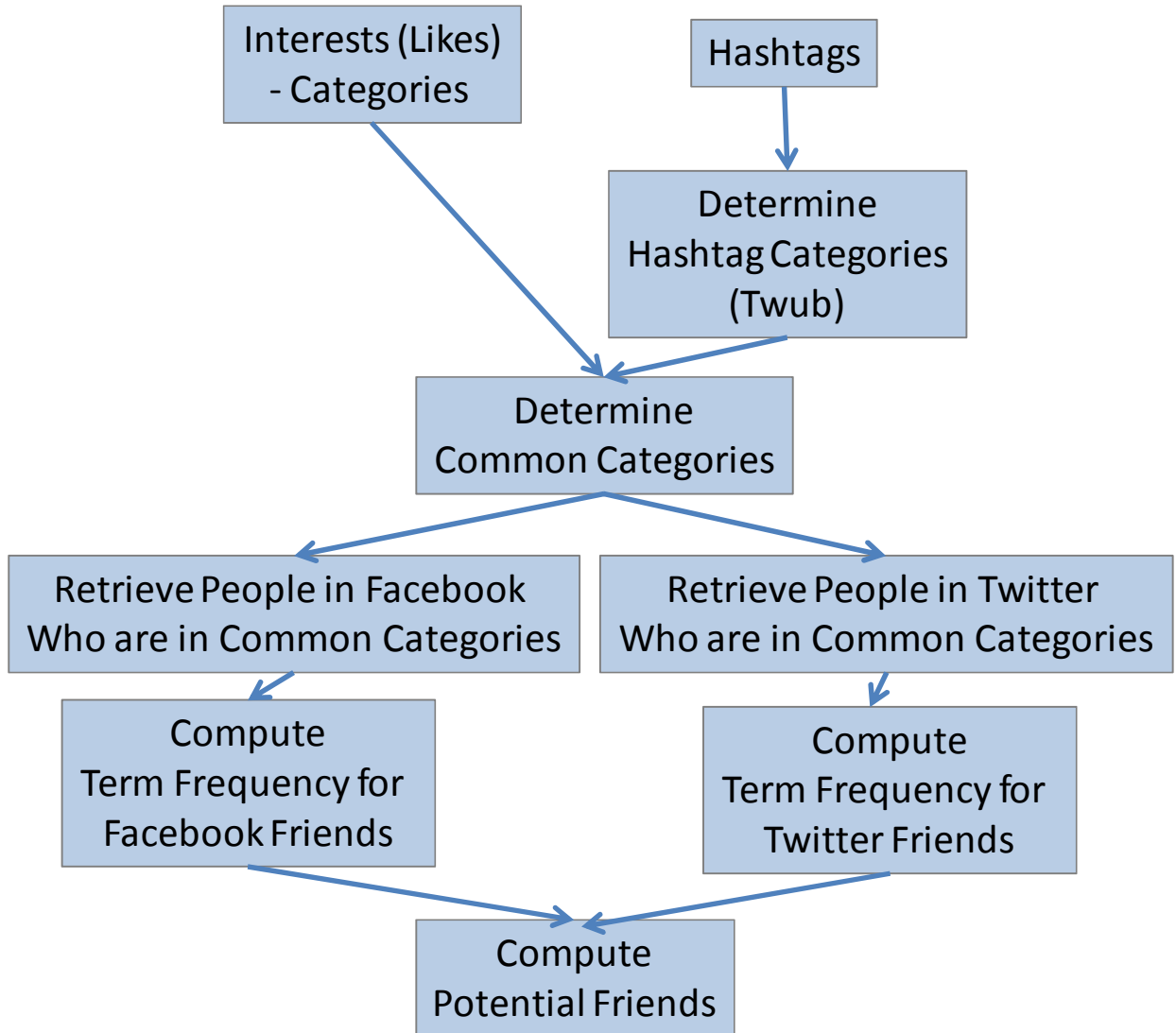


Figure 10. Potential Friends from Integrated Categories

### 3.5 Advance Search in Social Bridge

A Social Bridge advanced search enables a person to search for someone else with the choices and options of the user's choice. The choice is completely up to the user in determining the person's individual bridge strength.

The Social Bridge advanced search is advanced in many ways. It enables users to expand their searches to many categories by choosing the desired categories. Search factors are divided into two categories: the two social networking sites known as Twitter and Facebook. Each factor is given a range of values to choose and thereby decides the level of contribution of the certain factor in that particular search. Users are given the flexibility to discover people with certain factors.

Table 7. Advanced Search Factors

Contribution factors for Social Bridge advanced search are

Twitter	Facebook
Hashtags	Likes
Following and followers	Friends
Retweets and favorites	Comments and tags (photos and status messages)

For the above inferences, weights of the inferences vary and depend upon the user's choice of preference. Certain ranges of weights from 0-100% are given for every factor with a specific constraint so that the sum of all the factors adds up to 100% to determine the set of persons who are interested in the particular fields chosen by the user.

Table 8. Advanced Search Weights

Twitter		Facebook	
Factors (inferences)	Weights	Factors (inferences)	Weights
Hashtags	0 -100	Likes	0 – 100
Friends and followers	0 – 100	Friends	0 – 100
Retweets and favorites	0 – 100	Comments and tags	0 – 100
Sum (equal to)	100	Sum (equal to)	100

From Table 8, we can infer that the weights affecting the range vary from 0-100 for each factor in the social networking sites. A user may choose from the set of all the choices and determine the range of strengths. For instance, in Twitter, a user might be interested in knowing people who have more friends and less hashtags and in Facebook; a user might be interested in people associated with many friends and lesser likes. This means the user is searching for a people who have the characteristics of a follower, but not those of a leader. Depending upon the choices of the user, the strengths vary and result in different people from the standard search. Their sum, in total, should be equal to 100 in order to determine the individual bridge strength for the integrated social network.

An advanced search is carried out by obtaining the weights and the factors and the search in the integrated social network is calculated and people who match the search with the weights matching their profile are displayed to the end user. The advanced search is slightly different to the original search as the results obtained in the former solely depend on the user's choice of preference.

**Twitter:**

Variables:

$T_1$ : Retweets, Favorites,  $T_2$ : Hashtags,  $T_3$ : Friends, Followers, Common friends and followers

$$\text{User\_weight} = w(x) = \sum_{n=1}^3 T_n$$

If  $(w(x) == 100)$

$$FV(x)_{(\text{Twitter})} = w(x) * \sum_{k=1}^n (oc_k)$$

Where,

$oc_k$  = number of occurrences from user profile

**Facebook:**

Variables:

$F_1$  = Status likes, status comments, photo comments, photo likes, photo tags,  $F_2$  = Likes,  $F_3$  = Friends

Calculations:

User\_weight =  $h(y) = \sum_{n=1}^3 F_i$

If ( $h(y) == 100$ )

$F(y)_{(\text{Facebook})} = h(y) * \sum_{k=1}^n (oc_k)$

Where,

$oc_k$  = number of occurrences

Individual bridge strength determines a person's individual integrated strength in the integrated social network. Based on the inferences and weights chosen by the user, individual bridge strength is calculated and people corresponding to the results of the search are displayed based on their calculated strengths.

$$IndividualBridgeStrength = \left( \frac{F(y)(\text{Facebook}) + F(x)(\text{Twitter})}{2} \right)$$

People in the integrated network are displayed based on the order of their individual bridge strengths (descending order). People having higher strengths are given priority when compared to those with weaker strengths. With this discovery of new people, potential and integrated friends are discovered and by extending the concept of FOAF, new relations are established.

## CHAPTER 4

### SOCIAL BRIDGE IMPLEMENTATION

#### 4.1 Introduction

Social Bridge is implemented by generating an integrated social profile that portrays people profile of interests and interactions with others and helps them in connecting to people who share common interests and interactions. Social Bridge search's and connects people from various different social and micro-blogging sites to form an integrated social network. Social Bridge visualizes the current strength of a person based on his/her association, common interests, interaction and semantically suggests new friends who can be potentially connected.

This is a mashup of different networking sites and involves steps like (1) authenticating the user with a secure protocols, (2) acquiring data from the authenticated users using different Representational State Transfer (REST) Application Programming Interfaces (API), (3) extracting the desired data from the complete data set for analysis (detailed study) of the person, (4) analyzing the data set of different social and micro-blogging sites to determine an integrated fuzzy strength, and (5) integrating profiles from different social networking sites to determine integrated bridge strength and discover new people, (6) visualizing their individual and integrated networks, (7) determining various searching techniques to discover and study people and their profiles.

An integrated social network framework contains seven major components, namely, authentication, data collection, data extraction, data analysis, social network integration, and dynamic network visualization. Figure 11 illustrates all the major components of the integrated social network. In the following section, we give a short description about the prominence and functionality of each component.

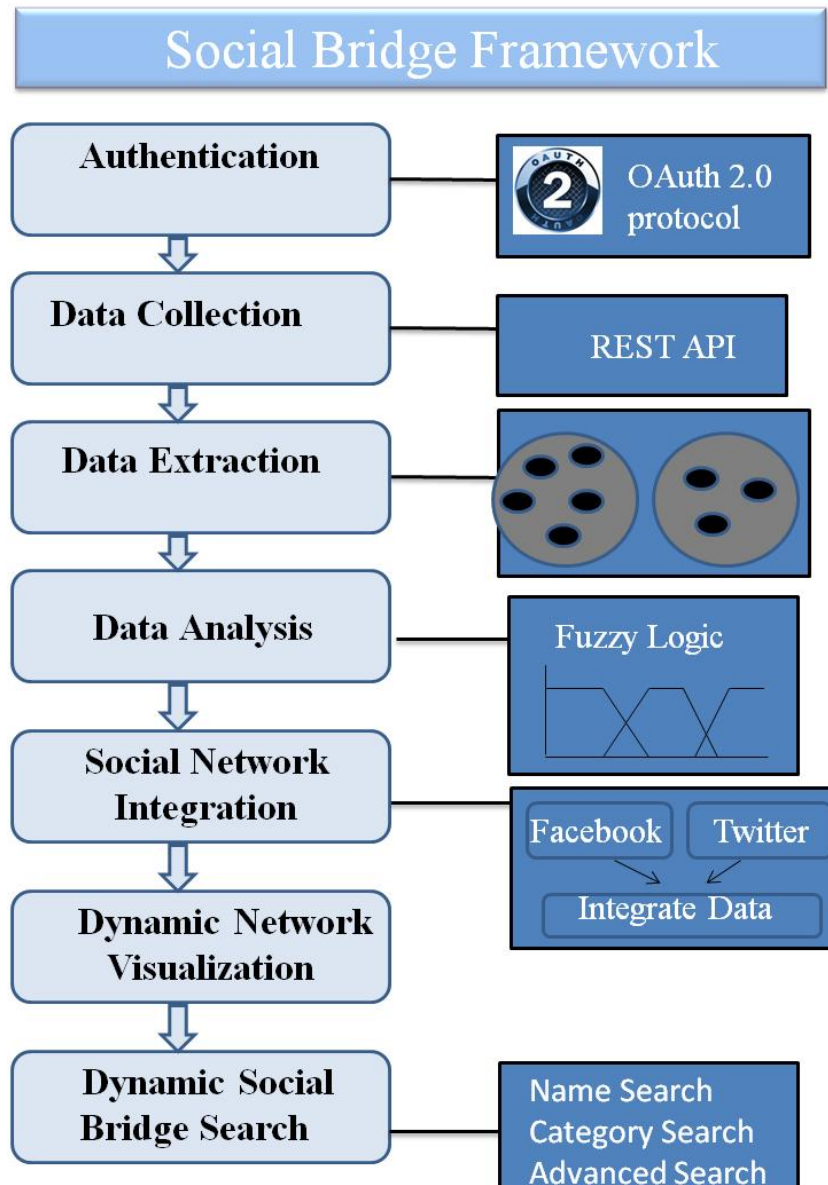


Figure 11. Social Bridge Framework

The authentication component plays the key role before the data collection starts. In order to participate in this system, the user is first authenticated with OAuth 2.0 protocol [14]. OAuth 2.0 is a secure protocol that is open for authorization. This authenticates users and allows them to share their resources without disclosing their credentials. This authentication lets the system

access the data of the authenticated user. This is considered as the most prominent step of the whole system.

The data collection component collects the data of the authenticated user using the REST API. REST is an architectural style for designing client-server network applications. Depending upon the requirement and the permissions of the user, the data is transmitted over HTTP protocol based on REST architecture. Various amounts of the user's data such as profile details, interests/categories, friends, followers, hashtags, photos, videos, profile feeds, news feeds, etc. are collected. Various constraints in accessing the data are solved here.

The data collected thus far from different REST APIs have all the details of each person stored in different formats. Various formats in which the data are returned from the API include JSON (JavaScript Object Notation), XML (Extensible Markup Language), Atom [15], and RSS (RDF Site Summary) [16].

In this thesis, we have been using JSON as the format and have used various libraries to parse the data and extract the desired content that is useful for further analysis. The data that are unnecessary for analysis are cleaned in this stage and the exact required data are sent further.

The data analysis component analyses the data that are passed after extraction. This holds the key to the entire network. The data collected from social and micro-blogging sites are first stored in relative databases corresponding to the user. The integrated social network knowledge base contains a huge collection of data and a relational database is chosen to sustain the scalability. The data and the person's details, obtained from the earlier steps, are used to design ontologies that determine the classes and properties. Ontologies represent the information as a set of classes in a domain and their corresponding relationships.

The user's individual data that are first analyzed include basic profile details, friends, followers, hashtags, common friends and followers, interests, retweets, favorites, photos, videos, status messages, likes, comments, tags, etc. All these details of both the social and micro-blogging sites are analyzed, and then bridge strength of the individual is determined.

After analyzing an individual's data, the same user is now analyzed with respect to other users who have participated earlier in our system. A list of common friends, common followers, common interests, common hashtags, common messages, and comments are taken into consideration. With this information, the Fuzzy Strength of this person, with respect to others, is determined.

After the analysis of data for different social and micro-blogging sites is completed, Social Network Integration [12] is accomplished. This component integrates all the data of an individual gathered from various networking sites and integrates them into an integrated social network. This step combines all the data and determines an integrated Fuzzy Logic [17] for the person.

In addition, the system semantically suggests friends who might potentially be interested in connecting with this person. This is semantically accomplished, based on the analysis done earlier. This component visualizes the individual and integrated social networks. The designed ontologies are utilized to determine class-property-class relationship and to visualize the person's individual and integrated network.

The individuals are considered as classes (nodes) and their information (such as interests, hashtags, etc.) as properties (edges). A directed graph is determined between classes based on the details obtained and its respective strength. Strength and its ratio are determined as a label. This



form of representation makes the system more understandable and thus, easier for people to follow it and its purpose.

## 4.2 Authentication

This is the first step in the integrated social network. OAuth 2.0 is an open standard for authorization. OAuth 2.0 is the latest evolution of the OAuth protocol. This authorization allows us to share and access users' private details (profile details, profile feeds, news feeds, friends, followers, likes, comments, retweets, mentions, favorites, etc.) without disclosing their secret usernames/passwords. This authorization protocol has been widely embraced by social networking giants like Facebook, Twitter, LinkedIn, Foursquare, MySpace, etc. Even colossal software firms like Google, Microsoft, and Yahoo! follow this protocol.

We have developed a web application called the 'Integrated Network' that acts as an interface for the users to give access to our integrated social network. The application is integrated with OAuth 2.0 and is completely safe for users to participate in our network as we follow all the protocols of OAuth 2.0.

Users utilize the application which acts as the interface to our network and submit their credentials. This authentication lets us gain access to their private resources. After authentication, a code is generated that passed as a request parameter to their personal social networking site (like Facebook), asking the user to give access to the application.

Only after this access, is an access token generated. This token is a random number, generated by the respective social networking site, to give approval to any third party application that requires ingress to gain private data with the authorization of the user. This token holds the

key for the entire process. With this token, the application can access private and shared data that has been made accessible by the user to other applications.

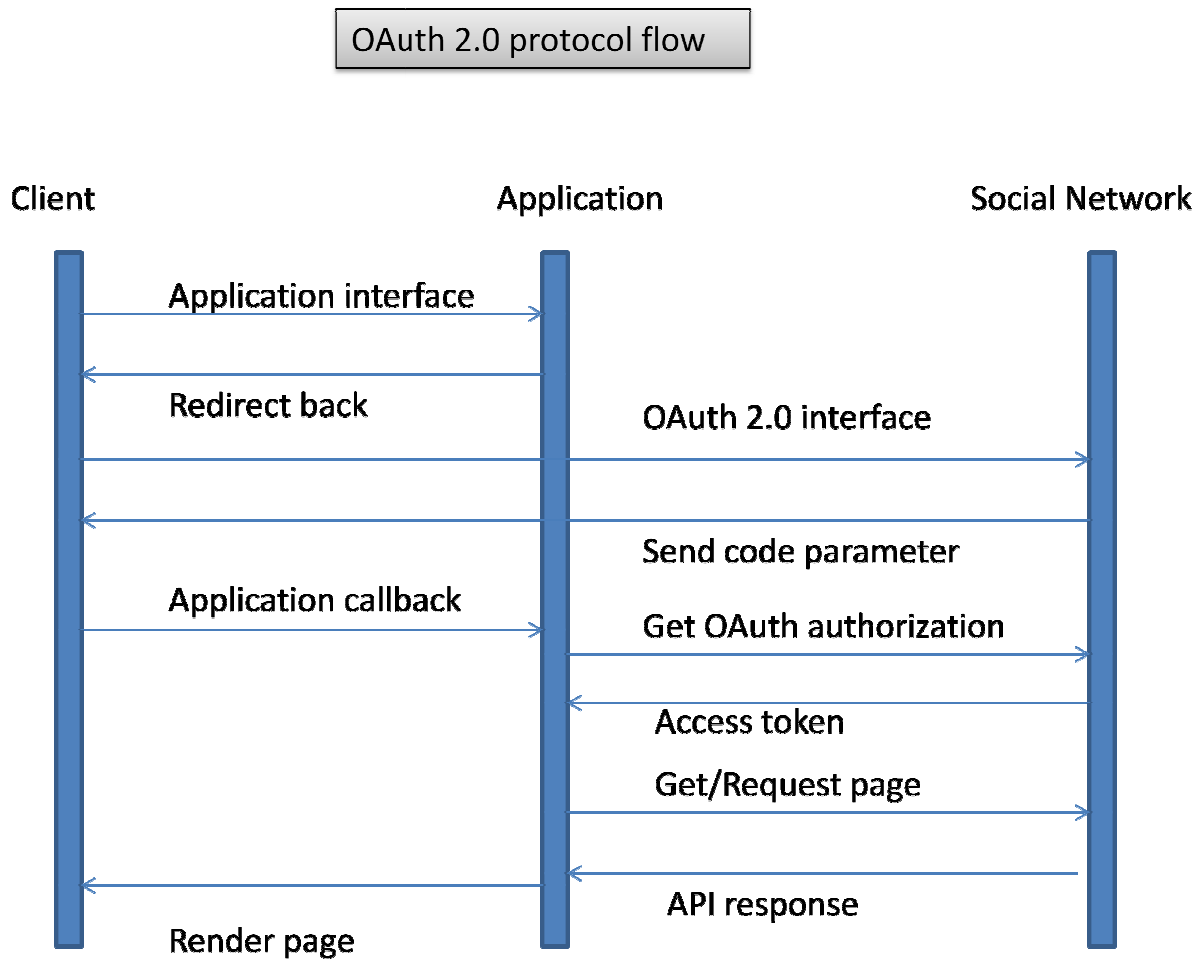


Figure 12. Authentication Flow

**Client-Application flow:** Client interacts with the application and is redirected back upon request.

**Client-Network flow:** Client accesses the OAuth protocol and obtains a code after entering credentials.

**Client-Application-Network flow:** Client calls back the application, the application gains access to the network and the network renders an access token.

**Application-Network flow:** Application requests details/page, network responds through REST API.

**Time frame:** The access token generated is as determined by the protocol is valid for 2 hours. However, we can make this token permanent by obtaining special permission from the user. The user, while accessing this application requests permanent access to his/ her profile. The permission obtained is granted as 'offline\_access' by the user.

**Security:** This access token can be disabled by the user at any time by changing his/her password or any of the credentials. It can also be disabled by removing the application from his/her respective social networking site.

### 4.3 Data Collection

This is the second step in the integrated social network. This component is applicable only to those users who have gone through the process of authentication. Once the user is authenticated, the data and details of the user are collected from the API of the social networks.

An API is a set of standards or instructions for accessing a web application. API's are released by software companies for other software developers to use their services and data, and develop their own projects. It is a software to software interface where applications communicate with each other.

#### 4.3.1 REST Application Programming Interface

In our project, we have used REST Application Programming Interface of Twitter [18]. After the authentication, based on our requirements, we gather data from the micro-blogging site. This site returns data in the form of REST API.

REST stands for Representational State Transfer. It is a stateless, client-server protocol to communicate between machines. It uses the stateless HTTP protocol for communication between client and server or between machines. REST is the best alternative to RPC or SOAP and WSDL based web services. The World Wide Web based on HTTP is the best example for REST-based architecture.

The Twitter REST API methods allow developers to access core Twitter data. This includes updating timelines, status data, and user information. It also enables us to access the user's friends, followers, tweets, retweets, favorites, mentions, hashtags, and many more. Example for Twitter Representational State Transfer API to access timeline of the user:

[https://api.twitter.com/1/statuses/user\\_timeline.json?screen\\_name=<screen\\_name>](https://api.twitter.com/1/statuses/user_timeline.json?screen_name=<screen_name>)

The above REST API allows us to access the timeline of the user with a particular screen name. There are many other different API's with which we can access the personal details of the user with their permission. The data here is returned in JSON (JavaScript Object Notation) format that is later parsed in the following components.

#### 4.3.2 Graph Application Programming Interface

In our project, we have used Graph Application Programming Interface of Facebook [19]. After the authentication, based on our requirements, we gathered data from the social networking site. This site returns data in the form of Graph API.

The Graph API forms the core of the Facebook platform. It approves the software developers requesting API's to read data from Facebook. It is a simple and complete view of the Facebook graph of the user with objects embedded inside. This API describes all the details of the user and his association with people surrounding him inside the social network. Every object in the social networking site has a unique object enhancing the ease of access thereby avoiding confusion.

Example for Facebook Graph Application Programming Interface to access the profile feeds of the user:

[https://graph.facebook.com/me/feed?access\\_token=<access\\_token>](https://graph.facebook.com/me/feed?access_token=<access_token>)

This above Graph API returns the profile feed of the user. The access token appended is the permission token that is generated after the OAuth 2.0 authentication. With this access token, the details of the user are accessed through API. The data here is returned in the form of JSON.

#### 4.3.3 Storing Collected Data

The above mentioned API returns data in the form of JSON files. The files are stored in SQL Server on the cloud and are locked with security passwords, thereby not allowing everyone to get hold of and access the files. The data from files is extracted using C# programming language. We have chosen C# as the programming language in building Social Bridge. C# is a web based programming language and is very efficient in implementing web based concepts.

We have utilized the most advance technology of cloud computing [20] in our project. To keep the data safe, we have selected a secure way of storing the data. Only the administrator of the cloud can access the data and the files are completely private. The files are stored in the root

folder of the cloud that is accessed by only the administrator. Permissions are also changed in the IIS server and the directory browsing for the files is disabled. The files can be first downloaded on to the root folder and can later be moved over to the cloud. Since we deal with huge amounts of data as the number of users increases with time, we secure all the files in a single folder and secure it in order to meet the scalability of the project.

Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth.

In order to make the files public, we have to edit their security settings. A new user has to be added under the security tab and has to be named IIS\_IUSRS in order that he/she can be given all the controls to read, write, and modify. Later, the directory browsing the file in the IIS server should also be enabled. This will make the file completely public and can be accessible by anyone using the cloud. Web configuration files need to be checked if there are any further problems with the security of files in a folder.

In this way, the files collected from social networking sites and micro-blogging sites are stored in a secure manner over the cloud. We are following two security standards in order to protect the user's data from becoming public. One is the OAuth 2.0 protocol that is followed by the user during the time that the access token generated and the other is cloud security protocols followed by us to maintain the confidentiality of the user's data. The scalability of the project is met by maintaining all the data in a folder and changing the security settings as a whole.

We have used an Amazon Elastic Cloud Compute (EC2) instance and an IBM Academic Skills Cloud for our project. Amazon's cloud and IBM's cloud seem to be very efficient in terms of security, performance, and scalability. The ease and access of data is very convenient in both clouds.

#### 4.4 Data Extraction

All the collected data stored on the cloud are stored in JSON file format. There are four different file formats in which REST API returns for the data to be extracted: JSON (JavaScript Object Notation), XML (Extensible Markup Language), Atom, RSS (RDF Site Summary).

For our project, we have chosen JSON format over other file formats. The reasons for choosing JSON over other file formats are, JSON is quick, fast and easy to learn, parsing JSON files is faster than XML parsing, JSON is good at representing complex data types and it is also a valid subset of JavaScript.

The JSON format is generally used for serializing and transmitting structured data over a network. The main purpose is to exchange data between client applications and the web server. So, when it came to API, JSON was unanimously chosen over XML.

- i) Graph API returns data only in the form of JSON.

The data obtained can be extracted on the client side and server side. Client side extraction can be done through JavaScript and server side extraction can be done through various libraries. 'NewtonSoft' is one library that allows us to parse the required data on the server side efficiently. The below pseudo code shows a method to deserialize the JSON file and extract the data and store it in a string format.

```
String json = File.ReadAllText("<File Path>");  
<Class name> fv = JsonConvert.DeserializeObject<<Class name>>(json2);
```

Generic classes are used in parsing the data and extracting the required content. REST and Graph API return serialized objects and we deserialize those objects and extract the required content. A small example is shown below that shows the data collected before extraction and after extraction.



```

{ "data": [
  {
    "name": "Integrated Network",
    "category": "Application",
    "id": "157658170991930",
    "created_time": "2012-02-23T02:57:16+0000"
  },
  {
    "name": "24",
    "category": "Tv show",
    "id": "14176232250",
    "created_time": "2011-11-18T08:51:44+0000"
  },
  {
    "name": "Cric",
    "category": "Sports league",
    "id": "146067842159104",
    "created_time": "2011-10-13T19:30:56+0000"
  },
  {
    "name": "Charlie Wilson's War",
    "category": "Movie",
    "id": "124254674251376",
    "created_time": "2011-08-20T04:21:22+0000"
  },
]
}

```



Name	Category
Integrated Network	Application
24	TV Show
Cric	Sports League
Charlie Wilson's war	Movie

Figure 13. JSON Data Parser

Extracted data is stored in the form of tables in the database. We identify only the required data and parse only the required data for analysis, and discard the rest of them. The data identification is done, based on the data returned.

Using REST API, we get the details of profiles, friends, followers, retweets, mentions, favorites, and hashtags of the authenticated user in the micro-blogging site. Depending on the API and on the data returned, we use appropriate classes and methods and parse the data dynamically. Using Graph API, we get the details of profile feeds, friends, likes (interests),

comments, tags, status messages, news feeds, photos, videos, and check-ins of the authenticated user in the social networking site.

#### 4.5 Data Analysis

This is the fourth component of the entire system that holds the key to the entire system. After parsing the required data and storing it in the SQL Server of our database, we divide the project into two sub categories: Individual Strength and Connecting Unknown Potential Candidates.

##### i) Individual Strength

The individual strength of the user is calculated based on the details provided by the user to our system. All the details and data of Facebook and Twitter are collected, extracted, and stored. Individual Strength of Facebook: The details extracted for the person with respect to Facebook are profile feeds, friends, likes (interests), comments, tags, status message, news feeds, photos, videos, and check-ins. For the analysis, we first get the basic profile feeds of the person and his/her basic details that include first name and last name, gender, location, and profile ID.

The person's friends are enlisted. All the details of the user's friends (like their name, location, and id), are captured. This helps us in analyzing the number of friends the user has in a particular social networking site and helps us in evaluating his/her strength

We next analyze the user's likes [11] page. All the interests of the user are analyzed here. Category and name are the factors evaluated. We get the count of the categories and the number of times the user has liked that category. The count factor contributes significantly in the

evaluation. Term Frequency and Inverse Document Frequency are analyzed. TF\*IDF is a numerical statistical weight often used in information retrieval and text mining.

Next, the user's interaction with friends is determined. The factors considered are status comments, status likes, photo tags, photo likes, and photo comments. Frequency is determined based on the number of times the user interacts with his/her friends for the above factors. Frequency weight is determined by multiplying the weight 0.01 to the frequency of the interactions. Based on the above calculations, a fuzzy strength is determined for the individual with Fuzzy Logic. Individual strength of Twitter: The details extracted for a person with respect to Twitter are profile feeds, friends, followers, hashtags, retweets, and favorites.

For the analysis [21], we first get the basic profile feeds of the person and his/her basic details that include their first name and last name, location, and profile id. The person's friends are enlisted. All the details of the user's friends such as their names, locations, and ids are captured. Friends are persons with whom the user is interested in connecting.

The person's followers are captured. This enables us to know the number of followers in Twitter. A person's following is assumed to be people who are interested in connecting to the current person. People who are common to both friends and followers are assumed as people who are strongly connected to the current person.

Hash tags are analyzed by determining all the hash tags the person has tweeted. These hash tags help in determining the interest of the user pertaining to different topics. We have classified these hash tags into categories. Our very own library has been built that classifies the hash tags of the person into different categories. A library containing different hash tags are considered and are classified into a set of categories.

Table 9. Hashtag Categories

<b>Hashtag</b>	<b>Category (Twub Cluster)</b>
#exam	Education
#basketball	Sports
#church	Community
#technology	Computers/Technology

The above diagram depicts the classification of the hash tags into a set of categories. Term Frequency and Inverse Document Frequency are also analyzed. Next, analysis is done on retweets and favorites. The person with whom the current user has re-tweeted and the number of persons who are all the favorites of the current user is determined.

With all the values in place, fuzzy logic is calculated and the individual strength is determined. By combining the individual integrated strength of Facebook and Twitter, we determine the integrated individual strength of the person who portrays the person's activity in all the social networks combined.

ii) Connecting unknown potential candidates

This step involves connecting one person with another who does not know each other. We have extended the concept of Friend of a Friend. Here, persons are connected to one another based on their choice of interests.

This is a two-step process.

- 1) Determining the field of interest (likes) with whomever the user wants to connect. A list of categories is displayed for the user to choose his/her field of interest.
- 2) Determining the hash tags with whomever the user wants to connect. Either an exact hash tag or a part of it is required for the information retrieval.

Likes and hash tags determine the fields of interest of the person. This shows that the person tends towards these particular topics and is interested in connecting to people who also have interest in these topics. We next search through our database and retrieve people whose interests match with the interests selected based on category criteria. Term Frequency of the like (interest) is determined by the strength of the person's overall likes. The same is calculated for hashtags that depict the strength of the person's hash tag category over all the hash tag categories.

Sorting in descending order, all the names of the people who are interested in those particular categories are selected and displayed. This list includes people who are not friends of the user. They can be directly connected in this system based on the interests chosen. This connection enables us to extend the concept of Friend of a Friend and expands the fundamental boundaries set. Connections between people are done between different social networking sites that do not exist in today's ubiquitous system.

#### 4.6 Network Visualization

An individual network is generated based on the integrated data determined. A Gephi tool is utilized for displaying data in the form of graphs. All the factors such as strength of association with friends, TF-IDF for likes and hash tags, association with a person based on re-tweets, favorites, status messages, photo likes, comments and tags are considered. The graph is generated and the associated values are stored in the form of CSV, PDF, Gephi, and JPEG files.

Google Chart Application Programming Interface is also utilized in visualizing a person's association, common interests and interaction dynamically. These charts are generated dynamically which represent the person's details in both the social networking sites. The below

figure shows a person activity in Twitter that includes Association (Friends and Following), Common Interests (hashtags), and Interaction (Retweets and Favorites).

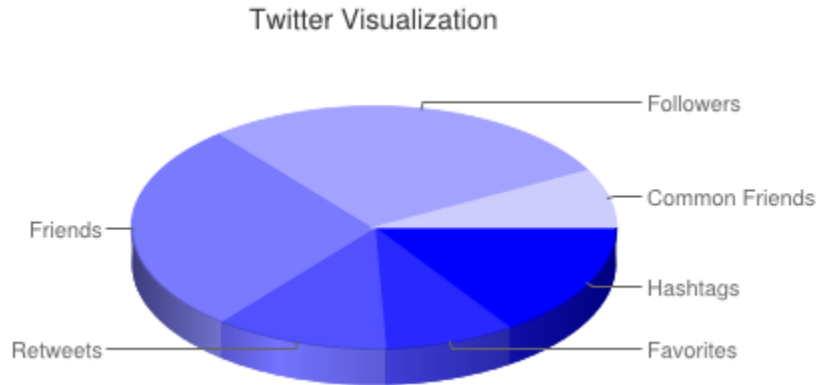


Figure 14. Twitter Visualization

The below figure shows a person activity in Facebook which includes Association (Friends), Common Interests (Likes), Interaction (Status comments, Status likes, Photo comments, Photo likes, Photo tags).

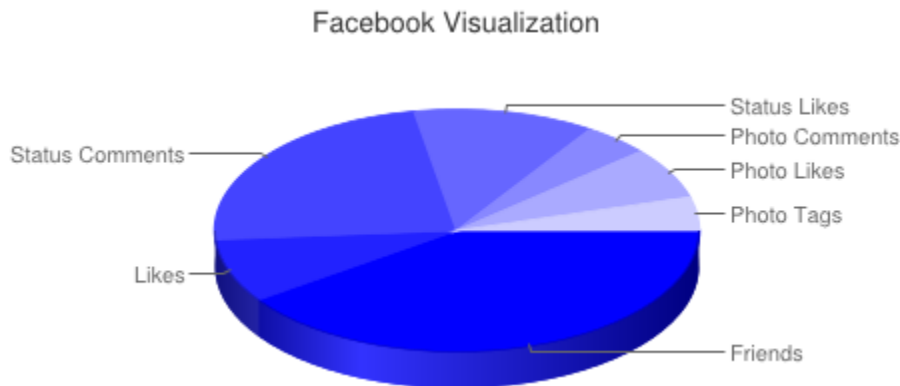


Figure 15. Facebook Visualization

Cytoscape [22] software has also been utilized in visualizing the network for Social Bridge. Cytoscape is an open source bioinformatics software platform for visualizing interaction

networks. Cytoscape was originally created at the Institute of Systems Biology in Seattle in 2002. Cytoscape is a very good platform which helps in visualizing networks dynamically and generates SVG files which can be embedded in a web page. The diagram below shows the association of a person towards his friends in Twitter. It displays the set of friends a person has in a network and their connection.

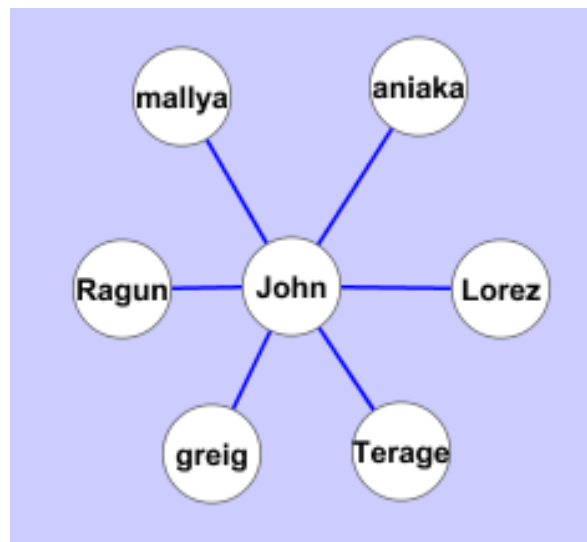


Figure 16. Twitter Search Graph

#### 4.7 Dynamic Social Bridge Search

A dynamic social bridge search has been included that enables the user to search for people having registered. Searches include

- a) Search by Name
  - a. It searches for a person based on the user's name.
  - b. It determines a person's entire profile from two social networking sites (which current search engines do not do).
- b) Search by Common Interest

- a. It searches for a person based on ‘hashtags’ or ‘Likes’.
  - b. It searches for a person based on integrated interested categories (Likes and Twub Cluster)
  - c. It semantically suggests more hashtags as ‘Hashtag Recommendations’ for the person searched who enhances the searching choices for the end user.
- c) Search by Activities
- a. It determines people’s associations towards their friends or followers. It determines the way people have interacted with their friends by comments, likes, retweets, etc. in the integrated social network.
- d) Search by Bridge Strength
- a. It determines the individual bridge strength of people and determines their activities in an integrated site.
- e) Search by User’s Selected Categories (Advanced Search)
- a. It gives choices to the user to decide upon the search categories selected.

These various searching techniques are implemented in Social Bridge to discover new people across existing and cross domain social networks. Social Bridge is also implemented as search engine of which the three factors association, common interests, interaction play a key role in determining the individual bridge strength of a person and discover new people. Implementing these searches enables to discover new people and connect people who might be potential or integrated friends.



## CHAPTER 5

### EVALUATION

#### 5.1 Introduction

In this chapter, the Social Bridge engine will be compared with the existing traditional social networking models and the proposed algorithms have proven to be powerful and efficient in finding potential friends for large social networks. The Social Bridge engine will be further evaluated through a survey of social network users for their feedback on its genuineness, accuracy, and scalability.

Section 5.2 explains the Experimental Setup that is used for evaluating the performance of the system. Section 5.3 explains the experimental domains on which the Social Bridge was evaluated. Section 5.4 explains the integrated Social Bridge run time performances. Section 5.5 explains the Social Bridge search. Section 5.6 displays results based on an Advanced Search in the Social Bridge. Section 5.7 compares the searching mechanisms of Social Bridge with other social networking sites Section 5.8 displays the anonymous survey results of Social Bridge.

## 5.2 Experimental Setup

The Social Bridge system has been hosted on the Amazon cloud (EC2) instance that had IIS 7 web server installed on a 32 bit Windows Server 2008 R2 operating system running on a machine with a processing speed of 2.75 GHz and 2.66 GHz and 1.66 GB of RAM.

The Social Bridge system has been also hosted on IBM cloud (Academic Skills) that also had an IIS 7 web server installed on a 64 bit Windows Server 2008 R2 operating system running on a machine with a processing speed of 2.27 GHz (2 processors) and 4.00 GB of RAM. For the web interfaces of the Social Bridge system, we used the Visual Studio 2010 Professional edition IDE with C# as a code behind the language. To store the data, we made use of the Microsoft SQL server 2008 express edition that was freely available.

## 5.3 Experimental Domains

The data extraction has been conducted with a total of 42 Facebook and 23 Twitter users. The data collections were incrementally conducted with all the people through the secure OAuth 2.0 protocol that we discussed earlier.

Table 10. Facebook and Twitter Data Sets

Facebook	Registered Users	42	Twitter	Registered users	23
	Total Friends	12424		Friends	647
	Total Likes	1451		Followers	448
	All Status Count	774		Common Friends and Followers	41
	Photo Tags	287		Hashtags	17
	Profile Feeds	261		Hashtag Categories	913
				Retweets	76
		Favorites	80		
		Photo Tags	287		
		Profile Feeds	261		

#### 5.4 Social Bridge Runtime Performances

After a number of iterations, we observed the time frame taken to search for a person at various levels. These time frames are recorded to note the efficiency of Social Bridge that depicts the speed of the search for a person at various levels and retrieves their information.

Table 11. Runtime Performance for Social Bridge Construction and Search

Type	Minimum Time (Min : Sec : Msec)	Maximum Time (Min : Sec : Msec)
Data Collection	00:00:06	00:00:22
Data Extraction	00:00:11	00:00:25
Network Building time	00:21:11	00:57:43
Search by Name	00:23:32	1:13 :78
Search by Interest (Likes or Hashtags)	00:00:02	00:01:00
Search by common interest (Twub Cluster and Likes category)	00:00:16	00:00:23
Search by user selected category (Advanced)	04:56:22	05:24:47

Search)		
Search for Path for Potential Friends	00:57:46	03:42:39

Table 12. Time-frames for Path Search (Level by Level)

Level	Individual Iteration (sec)	#People	Cumulative Iteration (sec)
1	0.1074843	1	1.9879498
2	0.0841045	479	2.0418176
3	29.0761262	5966	34.5328086
4	207.5684948	12,747	227.1228610

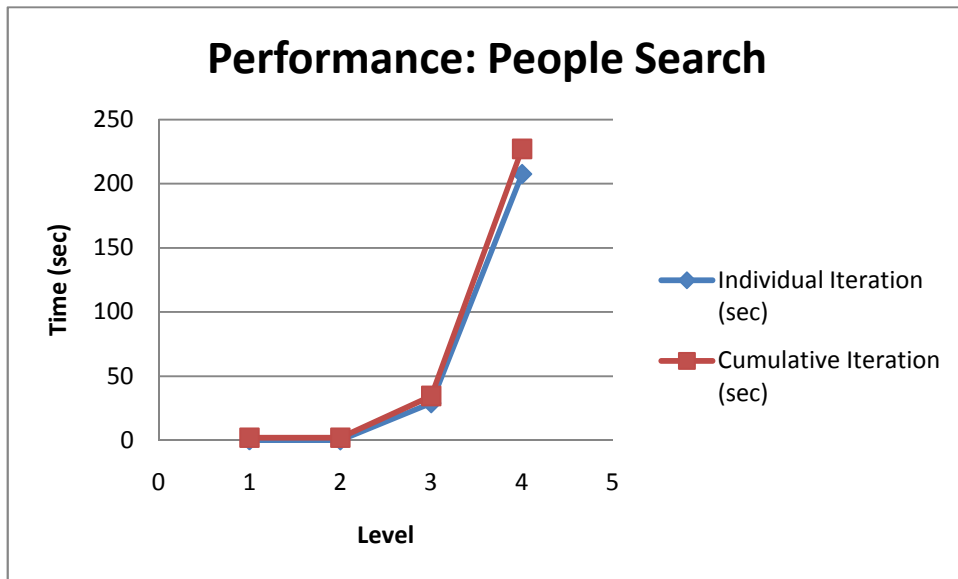


Figure 17. People Search Performance

Individual iteration represents the time frame to search for a person directly at the specified level if the level is known in advance. Cumulative iteration represents the complete search for a person in a given environment where the level of the person is completely unknown. Cumulative iteration depicts the search of all sets of lists of all the people and their friends until the person is found at a desired level. This determines the efficiency of Social Bridge for information retrieval.

This can be further enhanced by reducing the time complexity of the network. The efficiency remains the same but the complexity can be reduced further using various time and space efficient algorithms that utilize less memory space and iterate all the names of the people in an efficient manner. The algorithm implemented in Social Bridge is tested to be efficient for 12,747 members iterating through the loop. This can be further increased and more time efficient algorithms may be applied as the existing social networking sites do not provide such a search.

## 5.5 Social Bridge Search

The Social Bridge search enables us to search for all the details of a person in an integrated social networking site. This search combines all the details of the person collected from Facebook, and from Twitter. The person is checked for similarity in both social networking sites by checking their first name, last name, and their location. These three factors have been taken into consideration to identify the person's uniqueness in both social networking sites.

The features that the Social Bridge search engine provides are as follows:

- a) Search by Name

- a. Searches for a person based on user name.
  - b. Determines a person's entire profile from two social networking sites (which current search engines do not do).
- b) Search by Common Interest
- a. Searches for a person based on 'hashtags'.
  - b. Searches for a person based on 'likes'.
  - c. Searches for a person based on integrated interested categories (Likes and Twub Cluster)
  - d. Semantically suggests more hashtags as 'Hashtag Recommendations' for the person searched that enhance the searching choices for the end user.
- c) Search by Activities

This determines a person's activities towards friends or followers. It determines the way a person has interacted with his/her friends by comments, likes, retweets, etc. in the integrated social network.

d) Search by Bridge Strength

Determines a person's individual bridge strength and determines his/her activities in an integrated site. This search is carried forward by searching people based on Bridge Strength.

e) Search by User's Selected Categories (Advanced Search)

Choice is given to the user to decide upon the search factors. The choice is left to the user in determining weights and in choosing the factors.

### 5.5.1 Search by Name

Typing a person’s name will activate the search for the person and searches for the names of friends from both Facebook and Twitter. It first enlists all the names of the people who have registered with Social Bridge and displays them. It displays friends from Facebook, followers from Facebook, friends from Twitter, common friends, followers from Twitter, and common integrated friends from both of these social networks.

Table 13. Integrated Friends Search

Person name	Friends Category	Count	Friends’ Names
Peter	Facebook Friends	25	Jois, Roger, Gudiba, Vrnkmr, Varanasi, Rnjit, Nutalap, John, Moi, Avjana, Krshn, Sanguine, Algun, Racham, Garlap, Nori, Shekhar, Thantar, Bhamidi, Rupan, Myla, Maity, Arch, Pochanap, Babin
	Twitter Followers	24	Pochanap, Arch, Krshn, Moi, Tarun, Allen, Babin, Shaik, geeta_ch, achar, Tej, Racham, dock, ardith, Rams, tesh, Mantr, uday, Rampras, Kulkarni, daram, chakrav, chauthan, voteforsearchengine
	Twitter Common Friends and Followers	7	Arch, Racham, Babin, Krshn, Moi, Allen, geeta_ch
	Common Integrated Friends	6	Varanasi, Thanthar, Babin, Moi, Krshn, Arch

On the grounds of confidentiality, real names have been masked. Table 13 displays the entire list of friends of the person associated with his/her networking account across different domains. By obtaining the information of common integrated friends, we can know the person’s strong association towards his/her friends. We can understand how strongly he/she interacts with his/her friends.

Common friends and followers on Twitter give the names of people with whom this is very strongly associated on Twitter. This means that those persons are strongly associated with the person searched. The common integrated friends’ list gives the name of friends who are common to both Facebook and Twitter. With this list, we can assume that this person is very

strongly associated to the persons on this list. These are the lists of people with a count of six who are friends in both Twitter and Facebook. This means that these friends are connected to searched for person in both domains and they are in very strong association with one another.

### 5.5.2 Search for Common Interest

After searching for friends, Social Bridge also enables us to search for a person's likes, hashtags, Twub Clusters and further hashtag recommendations.

Table 14. Integrated Categories Search

Category	Likes and Hashtags
Facebook Likes	Website, University, TV show, TV channel, Travel/leisure, Telecommunication, Sport, Song, Software, Society/culture, School, Retail and consumer merchandise, Restaurant/cafe, Public figure, Product/service, Personal blog, Organization, Non-profit organization, News/media, Musician/band, Musical instrument, Music chart, Music, Movie general, Movie, Local business, Government organization, Games/toys, Field of study, Education, Consulting/business services, Computers/technology, Company, Community, Comedian, Club, Clothing, City, Cause, Cars, Camera/photo, Book, Baby goods/kids goods, Author, Athlete, Arts/humanities, Artist, App, Album, Actor/director
Category	Likes and Hashtags
Twitter Hashtags	erlestanleygardener, erlestanleygardener, sydneysheldon, sydneysheldon, information, civilization, information, school, cars, cars, school, schools, coding, microsoft, google, flute, guitar, athlete, sql2012, bingbar, cryptography, energy, athelete, coding, football, cricket, sports, google, bing, Microsoft
Hashtag Recommendations	substance, subject matter, noesis, message, knowledge, information measure, content, collection, cognition, assemblage, aggregation, accusation, accusal, accumulation, entropy, selective information, data, info, society, social process, excellence, refinement, culture, civilisation, substance, subject matter, noesis, message, knowledge, information measure, content, collection, cognition, assemblage, aggregation, accusation, accusal, accumulation, entropy, selective information, data, info, time period, period of time, period, educational institution, education, edifice, building, body, animal group, shoal, school day, schooltime, schooling, schoolhouse, wheeled vehicle, motor vehicle, compartment, automotive vehicle, cable car, elevator car, gondola, railroad car, railway car, railcar, motorcar, machine, automobile, auto, wheeled vehicle, motor



	vehicle, compartment, automotive vehicle, cable car, elevator car, gondola, railroad car, railway car, railcar, motorcar, machine, automobile, auto, time period, period of time, period, educational institution, education, edifice, building, body, animal group, shoal, school day, schooltime, schooling, schoolhouse, time period, period of time, period, educational institution, education, edifice, building, body, animal group, shoal, school day, schooltime, schooling, schoolhouse, writing, committal to writing, steganography, secret writing, cryptography, search engine, woodwind instrument, woodwind, wood, wineglass, groove, channel, fluting, champagne flute, flute glass, transverse flute, stringed instrument, contestant, jock, writing, scientific discipline, science, committal to writing, steganography, secret writing, coding, cryptology, cryptanalytics, cryptanalysis, strength, sprightliness, spirit, physical phenomenon, liveliness, life, healthiness, good health, forcefulness, force, executive department, drive, doe, energy department, department of energy, vitality, vim, muscularity, get-up-and-go, push, zip, vigour, vigor, writing, committal to writing, steganography, secret writing, cryptography, field game
Twub Cluster	author, author, author, author, education, education, education, Other, Other, Other, Other, Other, computers/technology, internet/software, internet/software, movies/music, movies/music, Other, internet/software, internet/software, computers/technology, transportation, Other, computers/technology, sports, sports, Other, internet/software, internet/software, internet/software, Other

The diagram above depicts all the likes of the person in Facebook and the category of names to which the ‘like’ belongs. This tells us the area of interest this person is interested in. It also shows the list of hash tags this person has tweeted. All the hash tags within the tweets in Twitter pertaining to this person are enlisted above. These hash tags are further classified within different categories. A library named ‘Twub Cluster’ has been created to classify the myriad set of hash tags into a defined set of categories.

The table below determines a person’s Term Frequencies of common categories of Twitter and Facebook. This shows how interested and how strongly this person is towards these set of categories.

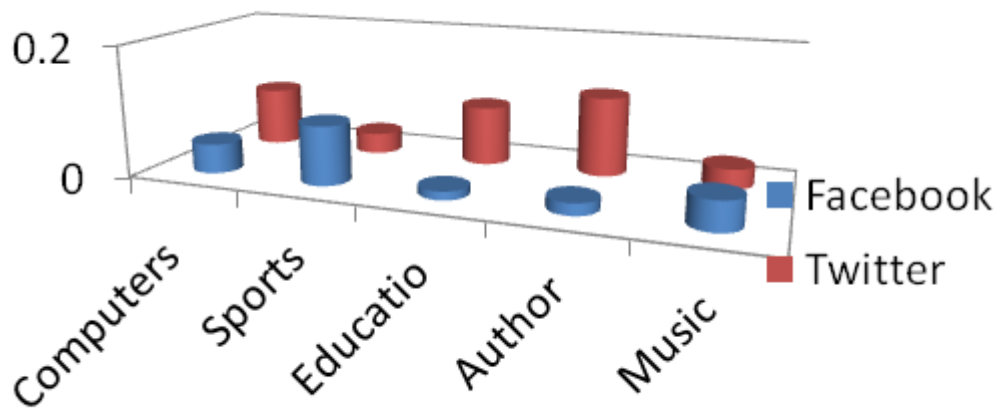


Figure 18. Term Frequencies of Integrated Categories

Figure 18 depicts the Term Frequency of person ‘Peter’ in an integrated network that shows the strength of a person towards a particular category. A high value in Term Frequency (TF) implies that the person is interested more in a particular category when compared to all the other categories. A low TF value implies that the person’s interest towards the particular category is very low.

### 5.5.3 Search for People by Activities

Social Bridge enables us to search for the level of interaction between the person searched for and with all his/her friends enlisted in Facebook and Twitter. It displays the intensity and the level of interaction between friends within the system. This enables us to know the friend’s interactions to help us determine the stronger friends among all of them.

#### 5.5.3.1. Facebook Interaction

Table 15 shows the interaction of a person with a set of friends within the domain. This search enables us to determine the common categories (likes), between friends, the number of

times a friend has liked a status, commented on a status, commented on a photo, liked a photo, tagged in a video, and tagged in a photo. This search helps us in the actual level of interaction between friends and determines their mutual strengths. This search, coupled with the search of friends, will help us determine the strength of the friendship between the person searched for and all the friends associated with him/her.

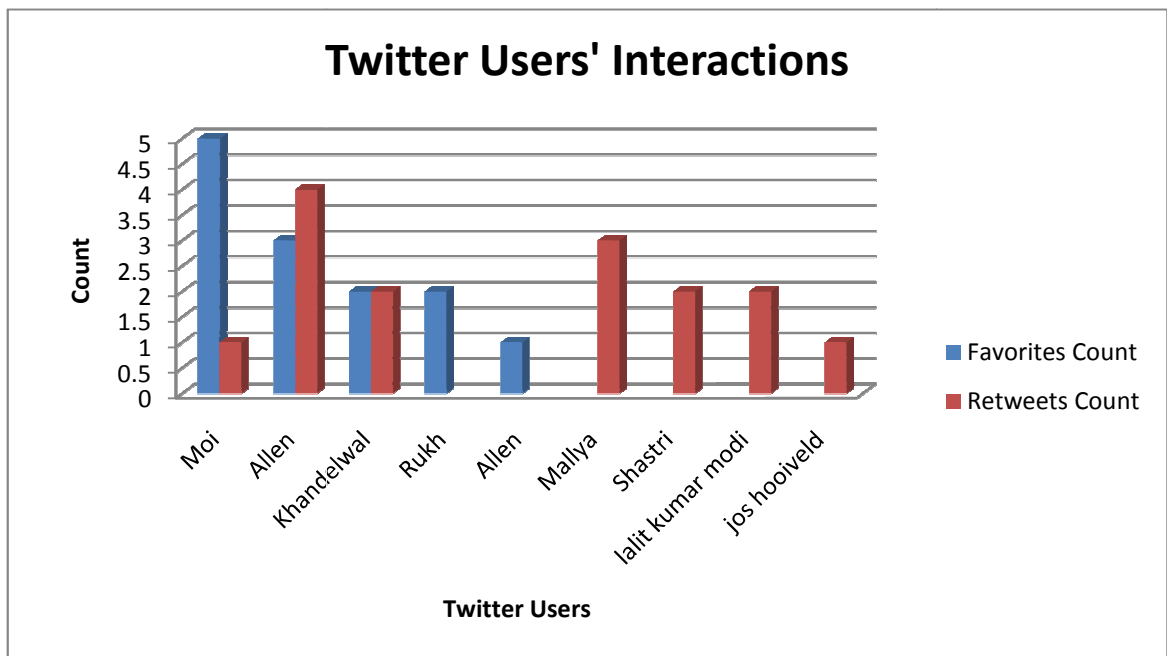
Table 15. Facebook Common Interests

Person Name	Common Categories	Status Comment	Status Liked	Photo comment	Photo Liked	Photo Tagged
Rupan	Community Movie Musician/band TV show	12	6	4	3	2
Nori	Athlete Author Community Computers/technology Musician/band Professional sports team TV show University	6	3	1	4	1
Racham	App Athlete Author Community Computers/technology Education Interest Movie Musician/band Product/service Professional sports team Sport TV show University	4	2	5	0	1
Krshn	Athlete Community Computers/technology Education Movie Product/service Professional sports team University	16	2	3	1	3
Pochanap	App Author Computers/technology Education Health/medical/pharmaceut icals Movie Musician/band	2	5	0	4	7
Arch	App Athlete Author Book Club Community Computers/technology Education Field of study	14	6	4	1	8

	Movie Musician/band Product/service Society/culture TV show University					
Satty	App Community Education Movie Product/service	4	10	3	0	5
Geeta_ ch	Author Book Community Movie Musician/band Society/culture TV show University	10	2	0	2	0

### 5.5.3.2 Twitter Interaction

Figure 19 shows the interaction of a person with a set of friends within Twitter. This search enables us to determine a count of the favorites with a person and a retweets count with a person. This search helps us in the actual level of interactions as favorites and retweets between friends determine their mutual strength. This search, coupled with the search of friends in the friends search, will help us determine the strength of the friendship between the person searched and all the friends associated with him/her.



### Figure 19. Twitter User's Interactions

These are arranged in descending order and display the strength of interaction between the people searched for and their friends or followers. It can be determined that the person topping the chart is the person who has the maximum number of interactions. The person standing in last place is the person with whom the interaction is the lowest.

#### 5.5.4 Search by Integrated Categories

A search for a person from an integrated data is also enabled in Social Bridge. This search, when given a category, will result in people who share common interests in Facebook as well as in Twitter. It also determines people who feel strongly towards a particular category in their own domain. For example, Table 16 shows the search results for people towards the category 'Author.'

Table 16. Integrated Strength for Category

Person name	Strength
Pochanap	7.29
Ranj	5.55
Arch	4.12
Moi	2.11
Thantar	1.01

The above table shows the results of people who share common interests from both the social networking sites and their respective strengths towards the category. People ranking highly are strongly associated towards that category and then people with lower strengths follow.

This helps us in determining the set of people who have strong interests in both the social networking sites.

### 5.5.5 Search by Common Interests

The Social Bridge semantically searches for people with similar interests in an integrated cross-networked domain and suggests friends who share common interests. This search is made exclusively for people who are not in the friends list of the person searched. This enables this social network search to reach out to different people and find new friends who are not already known. In this way, new people with similar interests are discovered and connected. The figure below shows a sample list of people who share common interests towards the category ‘Athlete’.

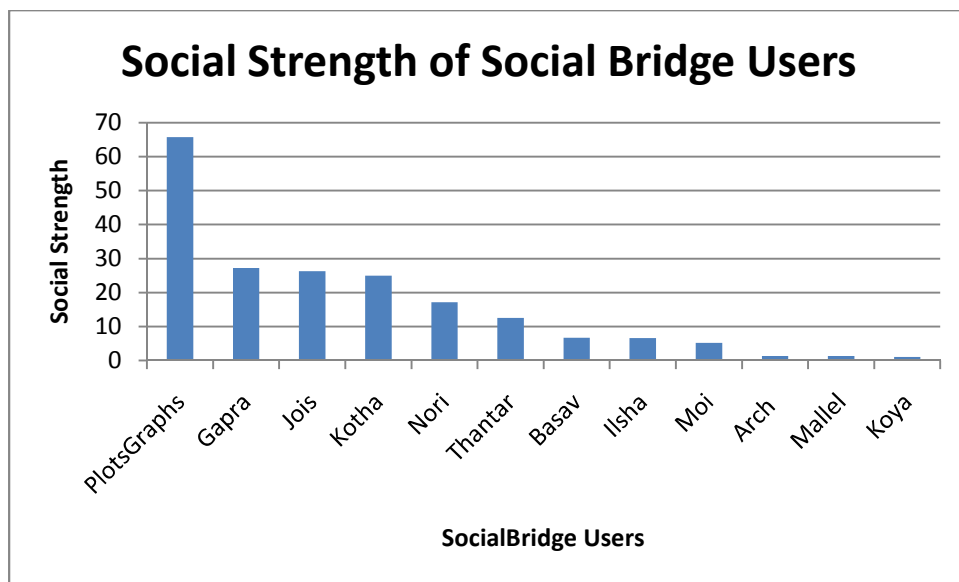


Figure 20. Social Bridge Users’ Social Strengths

The social strength data are sorted and the people who have a greater interest towards a particular category are given a higher strength and placed above the others. An evaluation is done by listing out all the persons who have ‘Athlete’ as their interested category. Next, using

Term Frequency, the ratio of the count of the category ‘Athlete’ to the total count of all the interested categories of the particular user is evaluated. This is later evaluated on a percentage basis. People having a high ratio are given the first priority over the other. This sorting enables the person searching to quickly connect to the highest person on the list as he/she is the strongest in the particular category.

Facebook does not determine the individual strength of the persons who are actively participating in the social networking sites. It only determines the friendship activity, but does not explain an individual’s active participation on the social networking site. It does not determine any specific strength or association of an individual towards others.

In Social Bridge, we determine the individual strength of the person and determine his/her strength of association with another person. The strength is determined applying Fuzzy Logic. A set of ranges is determined for people participating and their individual bridge strength is displayed. The figure below shows an example of individual strengths that are calculated on a percentage basis.

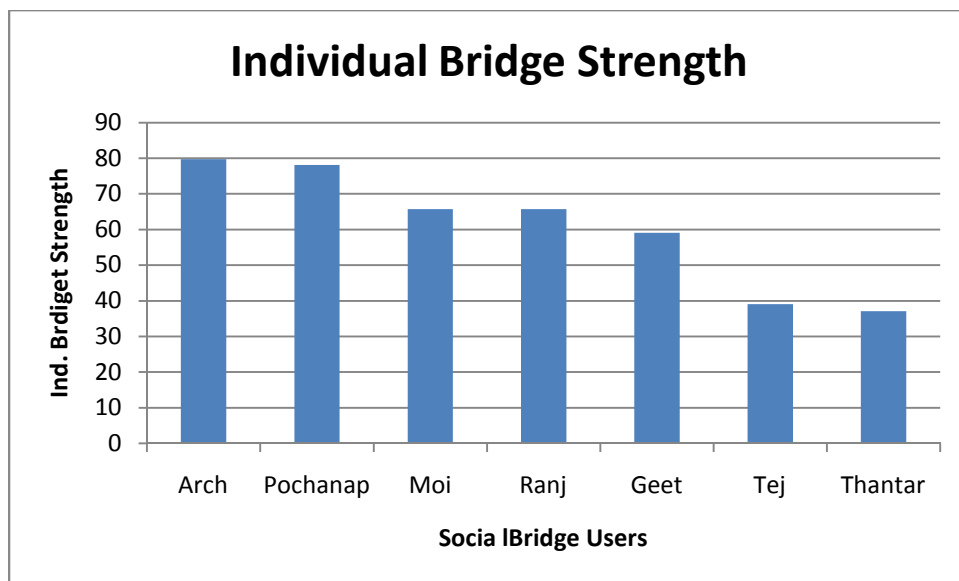


Figure 21. Social Bridge Users' Individual Bridge Strengths

The figure above displays the individual bridge strengths of people with strengths aligned in descending order. The person with the highest strength is displayed on top, which makes him/her the strongest among all. Strongest is defined here as the most active participation among all the others. The corresponding person with the highest strength is the most active person in the integrated social network and can be further considered to get connected.

Search is enabled with respect to hashtag or with respect to hashtag category from the Twub Cluster. This search is one of the most efficient searches in retrieving a person from the huge database without knowing his/her name. This will enable us to search with either a hashtag or with the category of hashtags.

Consider an instance when a user wants to search for a person, but does not know the exact hashtag but knows the similar category into which the hashtag might be categorized. So, in order to facilitate such a search, we have introduced another search that enables the user to search for a person with respect to categories from our Twub Cluster. From our 175 categories, the user is free to choose his own set of categories and search for a person who falls under those categories or for the person who tweeted with a hashtag.

For example, searching for people under the category 'Internet/Software' will enlist people who tweeted with the hashtag that falls into the category and their Twub Cluster strength towards it.

Table 17. Twub Cluster Category Search

---

Twub Cluster Category = 'Internet/Software'
---

---



Person's Name	Hashtag Tweeted
Arch	Google, Microsoft, Bing, Bing Bar, sql2012
Hagup	Cyber, Cyber world
Pendra	Microsoft, Sql2012
Khendel	Bing
Jith	Microsoft
Greig	Web

If the same search is carried out in Social Bridge to search for a person who has tweeted with the hashtag '#Cricket', Social Bridge brings out the results shown in the following figure.

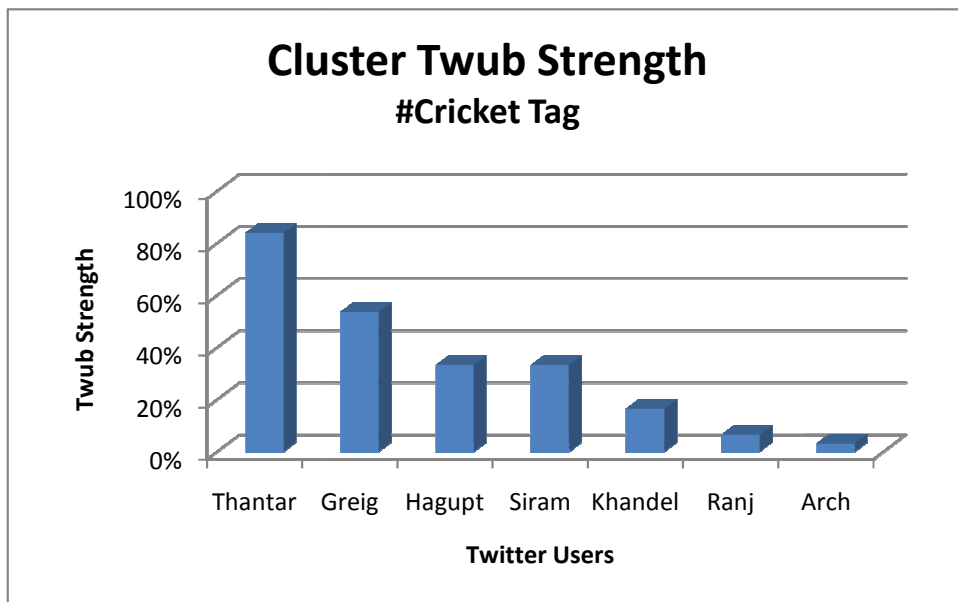


Figure 22. Twub Cluster Strength

The above search looks for people, who have tweeted with the hashtag 'Cricket' and their Twub strength is calculated as we discussed earlier. This search is more efficient than the search in Twitter as it decides a determined approach in the form of Term-Frequency that establishes

the number of times the hashtag has been tweeted by the particular person and estimates the person's strength towards the particular hashtag. This is absolutely more efficient than Twitter.

### 5.6 Social Bridge Advanced Search

A Social bridge advanced search enables a person to search for someone with the choices and options per the user's choice. The choice completely belongs to the user in determining the person's individual bridge strength. For a search where the user chooses the factors below, the results obtained are shown in the corresponding graph.

Choose your weights:

Twitter:		Facebook:	
Hashtags:	20% ▼	Likes:	60% ▼
Friends and followers:	40% ▼	Friends:	20% ▼
Retweets and favoritess:	40% ▼	Comments and tags:	20% ▼

Figure 23. Advanced Search User Interface

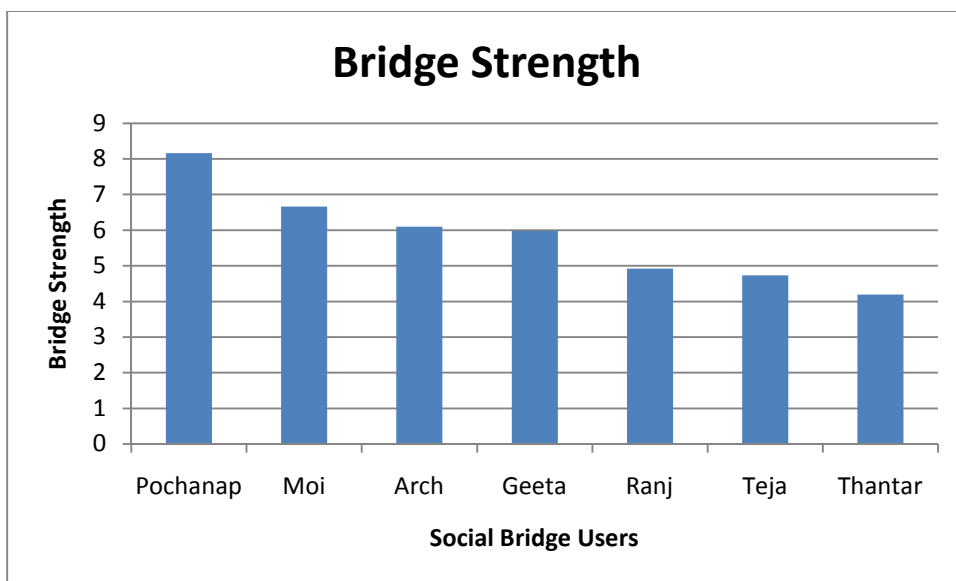


Figure 24. Advanced Search Bridge Strength

The figure below shows the way a person is connected to another in Social Bridge. Levels indicate the people placement in existing social networks and the connecting lines show how those levels can be bypassed and connected directly to the person in Social Bridge.

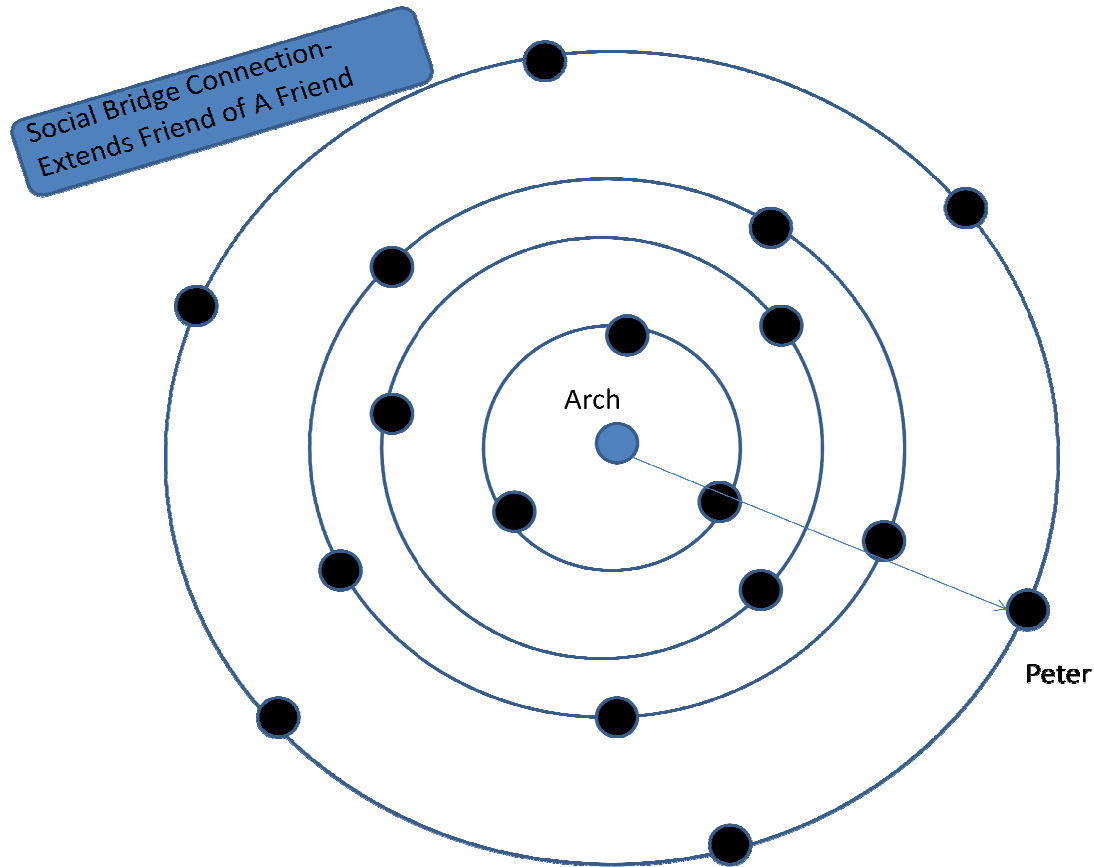


Figure 25. Social Bridge New Connection

## 5.7 Comparison with Other Social Network Searches

### 5.7.1 Facebook Search

**Search by Name:** Facebook enables us to search for a person with a name in its own domain. It has certain restrictions and constraints pertaining to Friend of a Friend (FOAF), e.g., search only

by name, etc. It searches for a person within its own network or searches for a person who is a Friend of a Friend. It can search up to two levels and retrieves the person's name.

The below diagram shows a sample of a 'friend search' in Facebook. This sample is scaled down to five friends for a person and their friends.

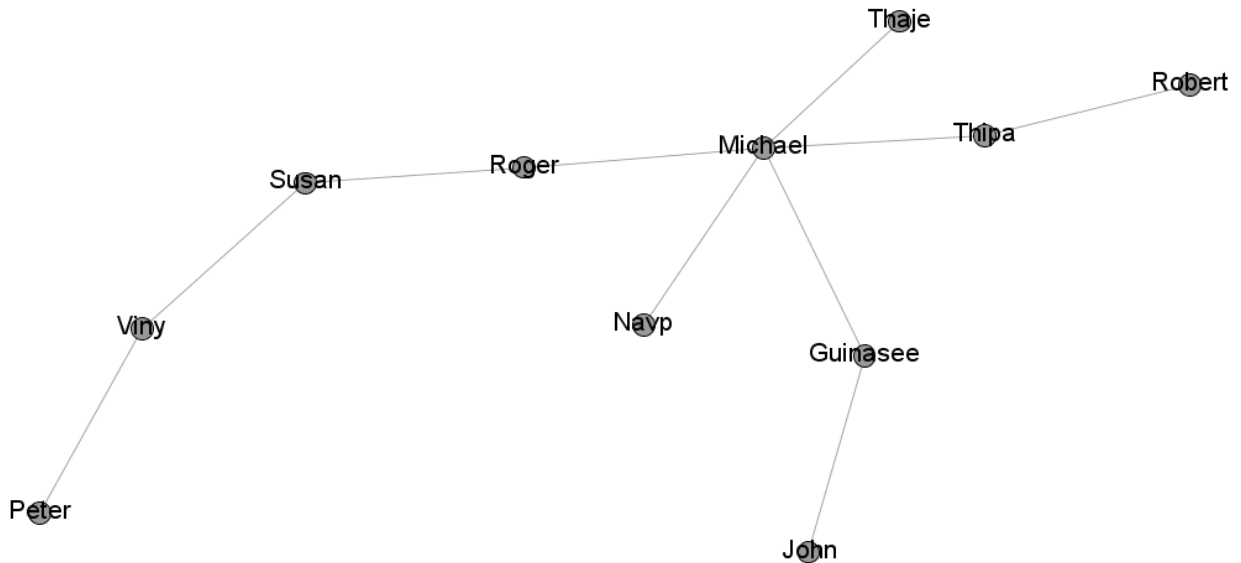


Figure 26. Facebook Search Graph

In this case, a person called 'Michael' is searching for a person called 'Peter' in Facebook. The above diagram shows the person search in Facebook. It shows 'Michael's' friends connected to his direct friends at level 1 ('Roger', 'Thilpa', 'Guinasee', 'Thaje', and 'Navp'). He is indirectly connected to 'Susan', 'John' and 'Robert' at level 2. Then the person is again indirectly connected at level 3 to 'Viny' and once again indirectly connected to 'Peter' at level 4.

Table 18. Facebook Friends Connection Levels

Person's	Level	of	Friends' Name
----------	-------	----	---------------

Name	Friendship	
Michael	1	Roger', 'Thilpa', 'Guinasee', 'Thaje', and 'Navp'
	2	'Susan', 'John' and 'Robert'
	3	'Viny'
	4	'Peter'

Facebook retrieves the information of a person when the exact full name of the person is typed.

Figure below shows the outputs of the search.



Figure 27. Facebook Name Search

If only a person's user name is known and the person is not in the friend's list of the searched for person, then Facebook returns 0 results and shows the image below. Since the user name search is not enabled in Facebook, it is really difficult to search for a person if the full name of the person is unknown.



Figure 28. Facebook Username Search

In Social Bridge, a person is found even though he/she is not in the friend's list of the person searched. A search is enabled with a username or with the person's full name from the entire list of integrated social networking sites. The accuracy of finding a person is extended to a level of four (as of now), where people are searched and are retrieved for efficient connections. The level can be further extended if we so choose.

**Search by Common Interest:** There is absolutely no scope in searching for a person on Facebook to find a person with similar interests or similar likes. Considering the case of searching for a person who has a deep interest towards 'Athlete', the search result in Facebook is as follows:

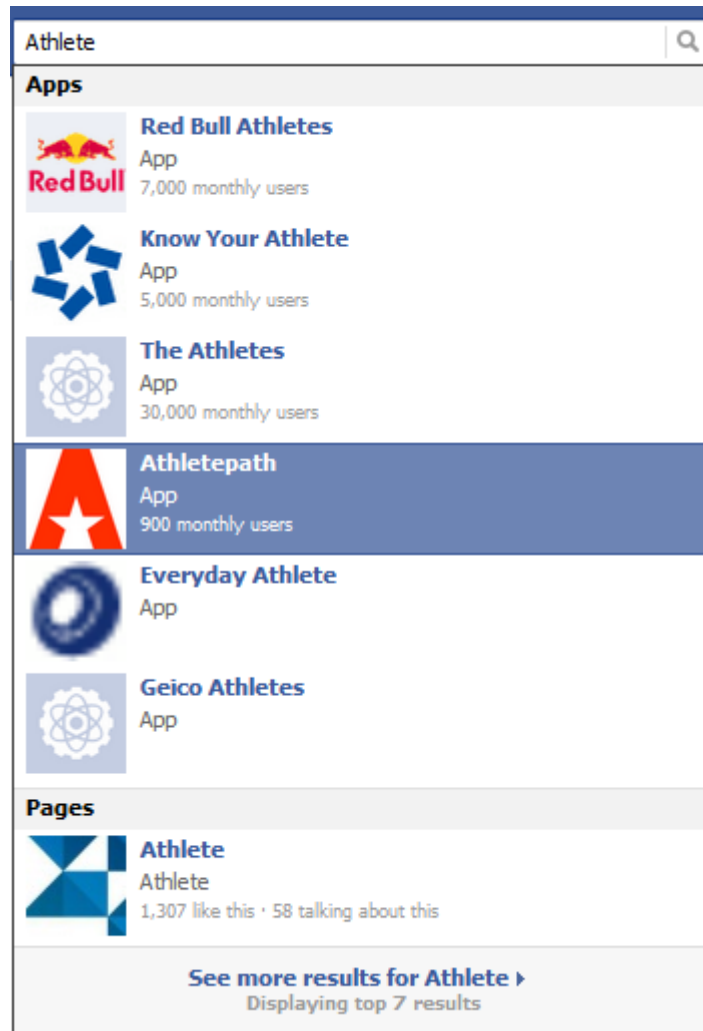


Figure 29. Facebook Category Search

The above diagram shows different pages and links to pages where 'Athlete' is created as a Page or as various applications. However, we do not get accurate and efficient information about people liking 'Athlete'. There is an absolute need to connect different people irrespective of their locations who have similar interests and common topics among them that they can share with each other through social networking sites.

### 5.7.2 Twitter Search

**Search by Name:** Twitter also enables us to search for a person with a name in its own domain. It helps us to search for a person via the person's user name. It extends the concept of Friend of a Friend and but searches for a person within its own domain. Given a user name, a person can be searched for on Twitter and one can directly connect to the person searched for. Figure 30 shows a search result for 'search by name' via Twitter. A user name is given and the search is carried out to retrieve the person's details in Twitter. The diagram below shows a person's friends and followers (up to three).

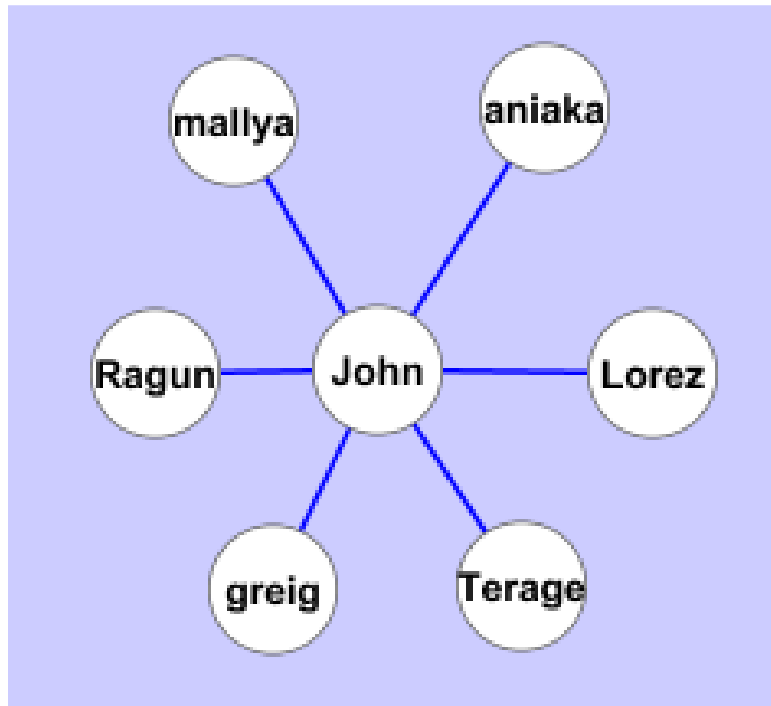


Figure 30. Twitter Search Graph

**Search by Hashtags:** When we search for a person based on Hashtags, the results in Tweets containing the particular hashtag. It just displays all the latest tweets that are tweeted with the particular hashtag that are active. This does not help us find persons who are relatively active in



the particular hashtag. It does not differentiate people who have very actively tweeted the particular hashtag and are more strongly associated towards the hashtag. It also does not enlist the set of persons who have actively tweeted with the particular hashtag. For example, if we want to search for people who have actively tweeted with the hashtag '#Cricket', Twitter's results are



Figure 31. Twitter Hashtag Search

**Search by Hashtag Classification:** There is no scope for absolute classification of hashtags in Twitter. All the hashtags created by various users are not clustered and are stored in the repository without categorization in Twitter. Searching for a hashtag based on its category is not enabled. A random search with a particular category results in tweets containing the word

searched. For example, searching for hashtags or people with respect to hashtag '#Internet/Software' results in tweets containing hashtag '#Internet/Software' in Twitter.



Figure 32. Twitter Category Search

This result just gives you a lot of unnecessary data that is redundant and is hardly of any benefit to the user. We cannot identify people who are interested in the particular hashtag or category. There is a need for hashtag classification as tweets classified by these hashtags are not categorized.

### 5.7.3 Comparison between Social Bridge Searches

Table 19. Summary of the Search Comparison among Social Bridge, Facebook, and Twitter

Social Bridge	Facebook	Twitter
<ul style="list-style-type: none"> <li>• Search by Name</li> <li>• Search by Common Interest</li> <li>• Search by Bridge Strength</li> <li>• Search by User Selected Features</li> <li>• Search by hashtag or Twub Cluster</li> </ul>	<ul style="list-style-type: none"> <li>• Search by Name</li> </ul>	<ul style="list-style-type: none"> <li>• Search by Name</li> <li>• Search by Hashtags</li> </ul>
	<ol style="list-style-type: none"> <li>1. Cannot search for a person with similar interests.</li> <li>2. Cannot discover a person if the name is unknown.</li> <li>3. Does not determine the strength of connectivity between persons</li> <li>4. Cannot find a person in a cross domain network like Twitter or LinkedIn.</li> <li>5. Has no scope for advanced search.</li> </ol>	<ol style="list-style-type: none"> <li>1. Cannot search for a person with similar interests.</li> <li>2. No proper classification of hashtags.</li> <li>3. Does not give information of the person's activity and association with the person searched.</li> <li>4. It does not semantically suggest more hash tags when we search for a hash tag.</li> <li>5. Search for hashtags results in tweets containing hashtags; do not determine people who are interested in the particular hashtag.</li> <li>6. No proper search of a person with respect to friends and followers of a person</li> <li>7. Cannot find a person in a cross domain network like Twitter or LinkedIn.</li> </ol>

## 5.8 Social Bridge Survey

We have conducted an anonymous survey with a total of 15 social network users (10 male and 5 female). All of them have at least one social network site account. Five questions were asked about their social network behaviors and six questions were about the Social Bridge Framework.

### 5.8.1 User Profile

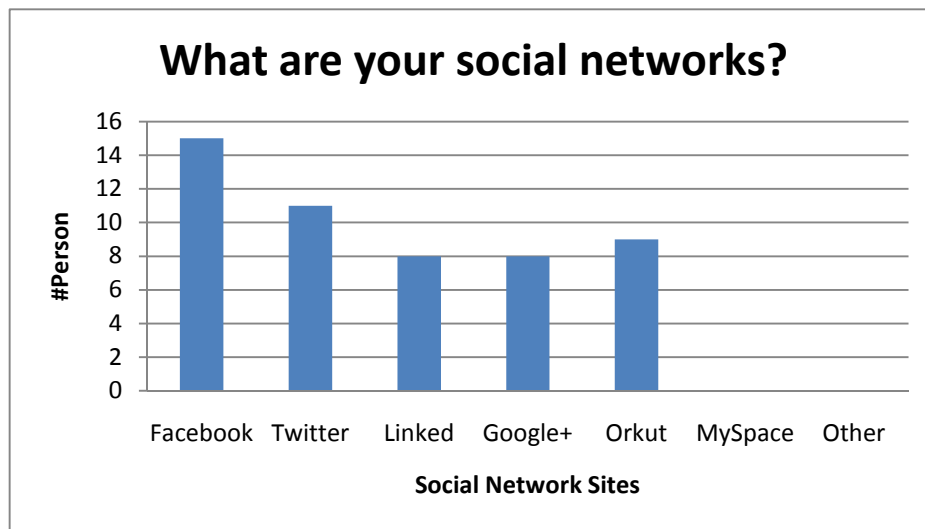


Figure 33. Survey Networks

The result shows that the majority of people have Facebook, Twitter, and Orkut accounts. These are followed by LinkedIn and Google Plus.

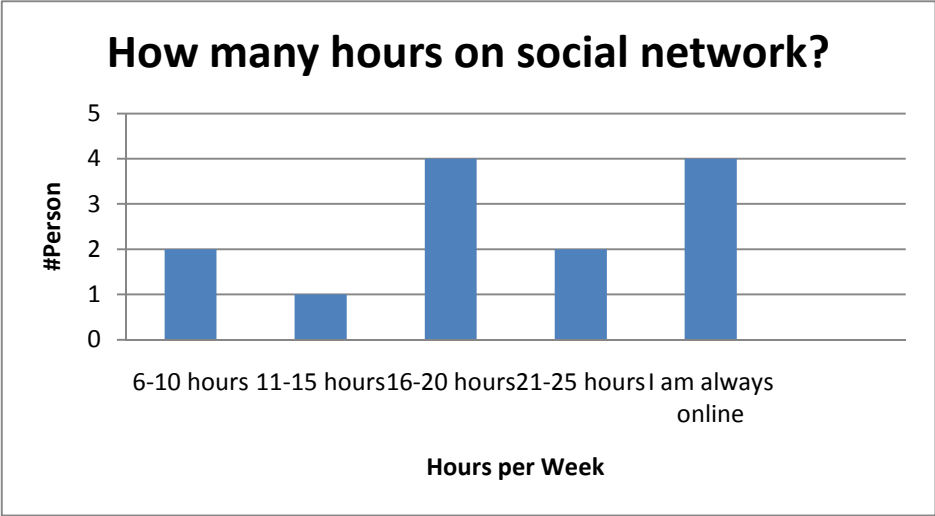


Figure 34. Survey Social Network Hours

The results show that the maximum number of people are either always online or spend at least 16-20 hours online per week.

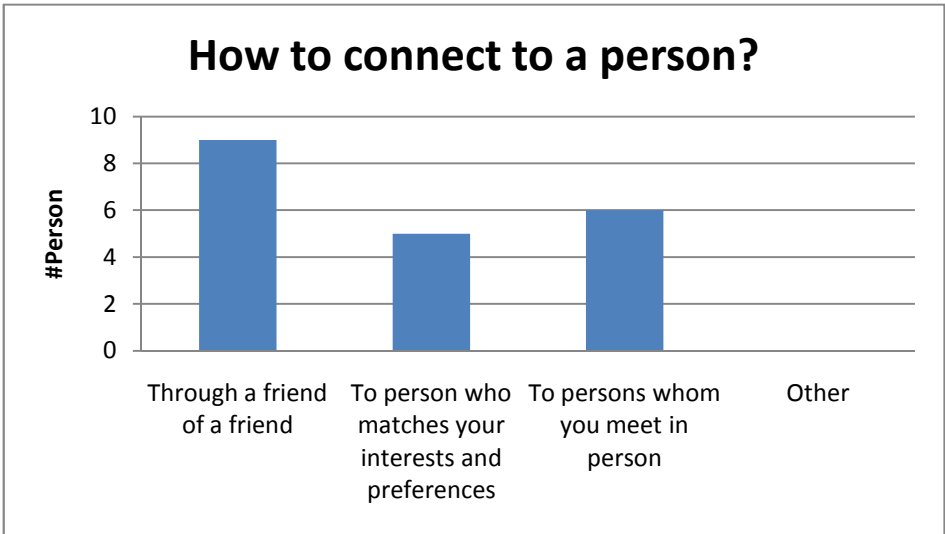


Figure 35. Survey Social Network Connection

The following options are given for the question: i) to keep in touch with friends and family, ii) to meet new people, iii) to make professional and business contacts, iv) to share photos, videos, and music, v) to play games, vi) to discover new music, books, films, and other entertainment, vii) to find information and share feedback about brands and products, viii) To promote a business or cause. The survey result shows that majority of people use social network to keep in touch with friends and to meet and discover new people.

### 5.8.2 Social Relations and Strength

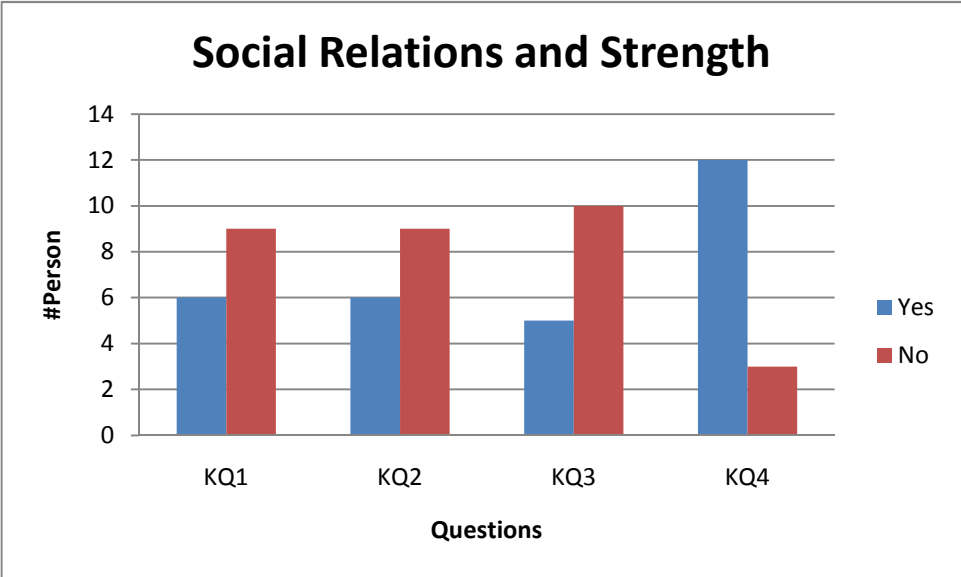


Figure 36. Survey Social Network Relations and Strengths

We asked the users the following four questions on their awareness of their social relationships and their interest in social strength with friends.

KQ1: Do you know how strongly you are connected on your existing social networking sites (like Facebook or Twitter)?

KQ2: Do you know how strongly you are connected to all your friends?

KQ3: Do you know how to integrate friends from all other networking sites into one and make new friends?

KQ4: Are you interested in knowing your existing networking strength and connecting with new people with whom you share similar interests?

The results show that 64.3% of people do not know how strongly they are connected on the social network sites they utilize. The same number of people also does not know how they are connected to their friends. 71.4% do not know how to integrate people from different social networking sites. The survey result shows that 78.6% are interested in knowing how to connect to new persons and know their strength.

### 5.8.3 Social Bridge Evaluation

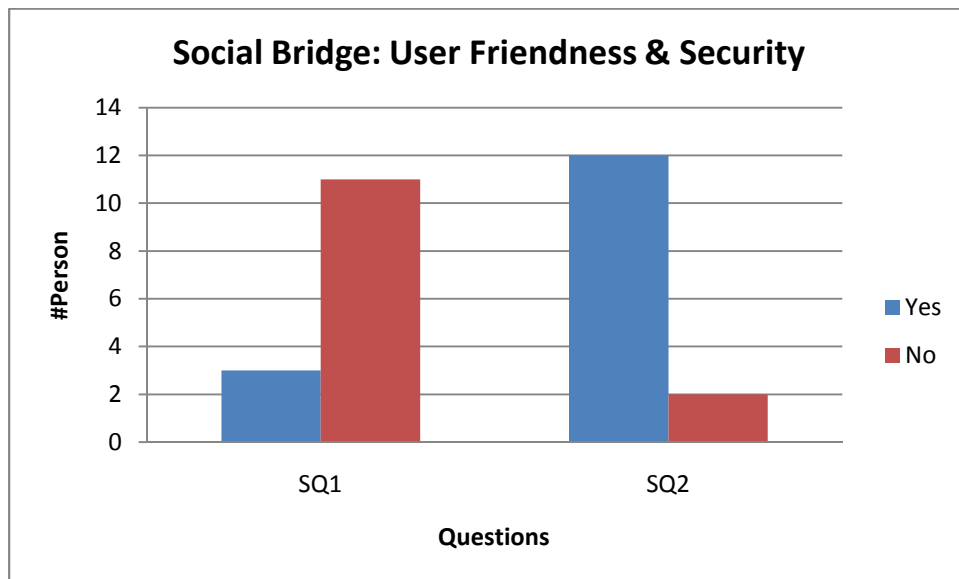


Figure 37. Survey Social Network User Friendliness

78.6% did not face difficulty in understanding the flow of the system. The rest had difficulty in choosing hashtag categories from Twub Cluster as there were no options. 85.7% felt that the application was secure as it had 'https' as the protocol. The rest had questions as they reported that they did not know how to evaluate security in web browsing.

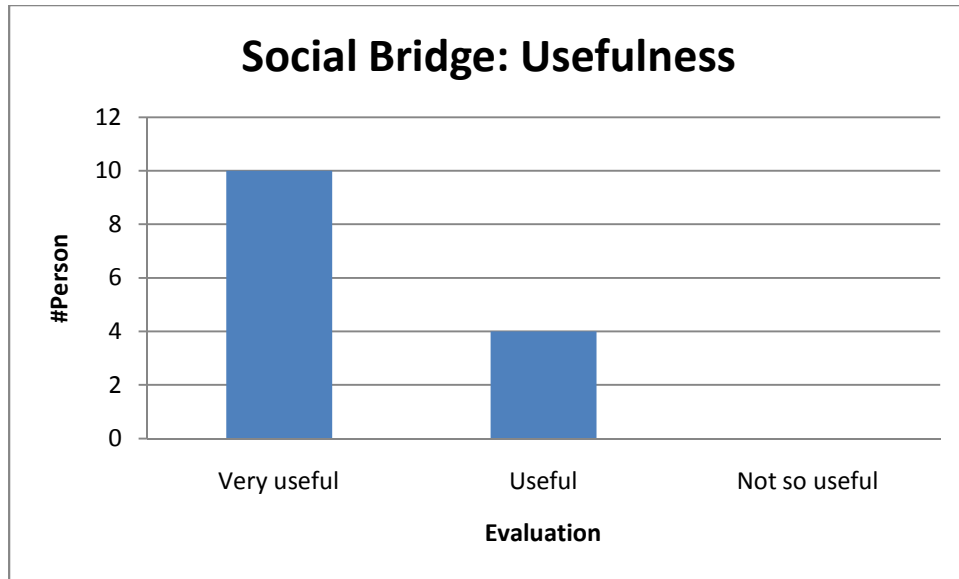


Figure 38. Survey Social Network Usefulness

71.4% felt Social Bridge was very useful and 28.6% felt it was useful as it enabled them to know about a person. Completely from different sites and discover new people. Many of them found Advanced Search very useful as it had the choice left to user.



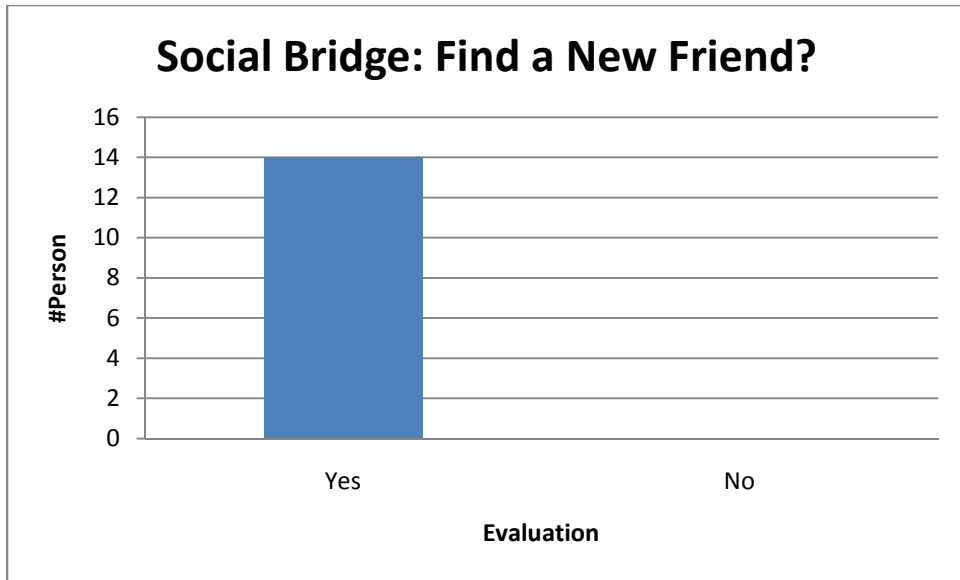


Figure 39. Survey Social Network New Friend

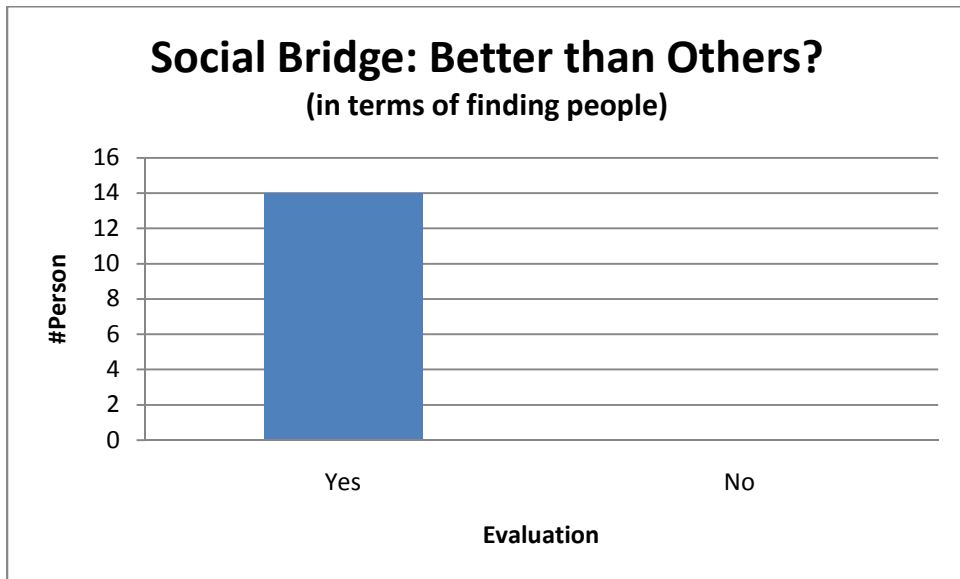


Figure 40. Survey Social Bridge Comparison

100% answered yes as they were able to discover new people outside their own friends' network. 100% felt our system was useful and better than the other social networks as it searches for people from both of them. Some people also felt that the user interface of the system could be better.

## CHAPTER 6

### CASE STUDY: SOCIAL NETWORKS FOR CLINICAL TRIALS

#### 6.1 Introduction

There is a need for human subjects for clinical trials and recruiting them is one of the most important parts in the discovery of new diseases. The approach of going through an elaborate questionnaire process in clinical trial recruitment is not efficient due to the lack of a systematic approach. MindTrial [23] (an intelligent online system for clinical trials), is an effort to move a part of the recruitment flow to an online intelligent model, primarily for clinical trials with an emphasis on mental health disorders.

Recruiting different human subjects in the MindTrial system is done intelligently by asking interactive questions in the form of Quiz, Wizard, or Education components. Various interactive questions are asked that help the recruiter in analyzing the mental and physical stability of the person.

Although much health related websites are available on the internet, there is no such website that flows through a procedure of scheduling a patient through various focus groups and connects them all together through an integrated social networking site. A component called MindFlow [24] is introduced in this online intelligent system for recruiting human subjects and making them undergo a vigorous procedure of workflow that determines the identification of suitable subjects for clinical trials.

MindFlow acts as a bridge to connect people who are interested in particular topics of interest. Social Bridge acts as an efficient mechanism to discover and identify people before giving their details and scheduling an appointment using the scheduler of MindFlow.

Social Bridge is initially used to discover people who might be interested in particular topics related to clinical trials. We therefore use Social Bridge, which is an integrated search engine, to discover people who show active interest towards clinical trials and may participate in focus meetings. People are thus discovered and are invited into Mindflow from where they can schedule an appointment.

## 6.2 Communicator for MindFlow

A ‘communicator’ has been developed that lets the administrators of clinical trials set up a definite schedule for various focus group meetings. A Telerik [25] RadScheduler has been specifically designed for this purpose. The steps involve administrator maintaining the system by creating different sorts of events and announcing them for the patients. Events include different sessions like screening, focus group system evaluations, treatment events, lab test events, study announcements, etc.

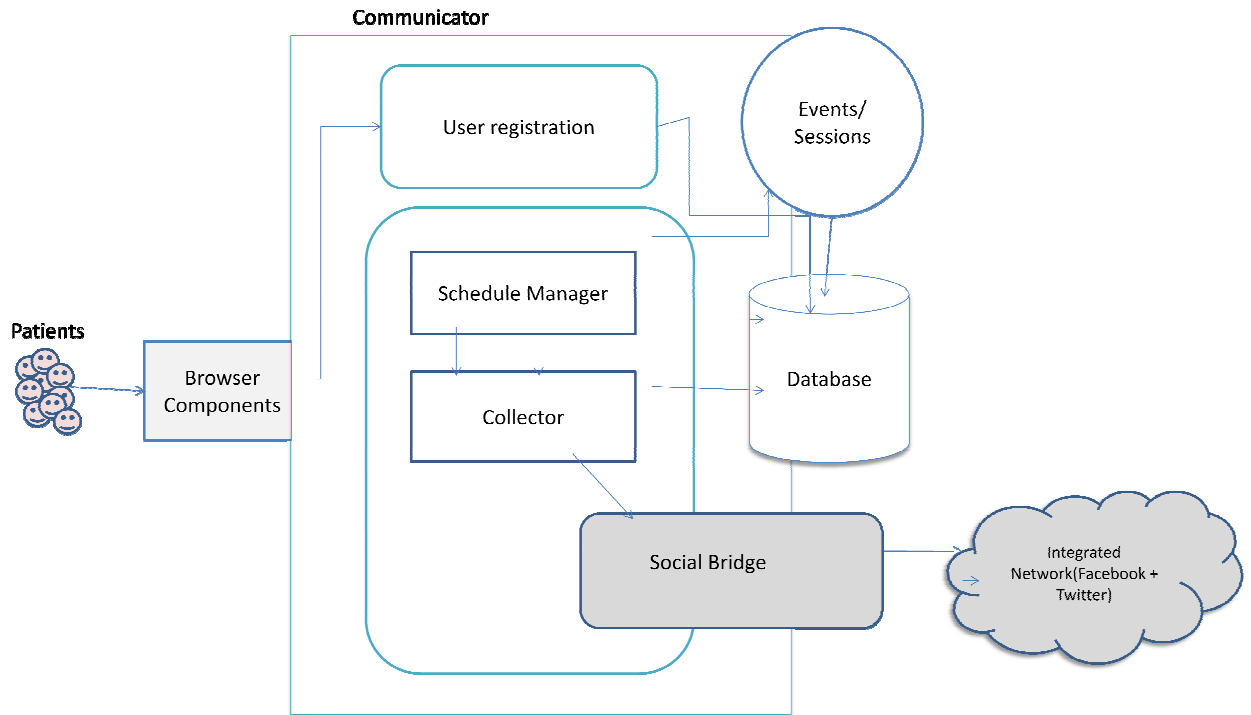


Figure 41. MindFlow Architecture Diagram

The Communicator contains three main components based on its functionalities. These are: Schedule Manager, Collector, and Social Bridge.

**Schedule Manager:** Schedule manager is one of the most important components of the Communicator. Initially, it takes the user input for the schedule of some future phases (dates of all events and the participants for those events). In case if some event gets cancelled or not completed, the user has to just mention that an event was incomplete and the schedule manager will automatically modify the schedule accordingly (that it will take more time for completion of that phase and the necessary timeline) and present the modified schedule instantly. The user though has the flexibility to make further changes in the schedule which eventually gets stored in the database and email notification and text messages are sent to the corresponding receivers mentioning the changes in the schedule. Another important feature of the schedule manager is

scheduling of meetings and other events. A researcher or anybody from the management, having access to create an event, may do so using the schedule manager for some particular group of participants and can notify them through the Communicator.

**Collector:** The collector gathers all relevant information from the participant's data stored in the system database based on the intended activity of the user. This data may include participant's name, address, phone number, gender, age, occupation, ethnicity, email ID, social network ID, service currently availing or willing to avail etc. The collector collects such information and presents them to the user through a graphical user interface. The user may go through each participant's data and enter his current health information as per the medical reports, prescribed medication (as suggested by the researching physicians). He may also update a patient's current trial status and put any restriction to a specific activity. This will activate a pop-up to warn that user in case he tried to attempt the specific activity. The user has his choice of preference for selecting participants from the existing ones, i.e. the collector provides the user a better query management. For example, the user can choose participants according to the gender, ethnicity, location, age group, occupation and/or service consumption etc for the convenience of the management to schedule for some activity.

**Social Bridge:** With the increased popularity of today's social networking websites, it has become important to reach the participants through the social networks. Keeping a trial process overview in the documented form in these networks will not only help the participants to have a good understanding of the whole process, but will also help to get involvement of new participants. The management, through the Communicator, may publish participant-specific information or some schedule on his home page of the social network (with his consent) or send it as a private message through the social network itself. The management can also post some

useful statistics and success stories of clinical trials on their website which will further motivate the participants. Video representation of all such information can be made available through the Communicator. It is often seen that encouraging feedback from the cured and previous satisfied participants, helps to keep the moral of current participants intact. The Communicator allows the management to gather information from all the participants who have similar interests and can suggest friendships accordingly

Another advantage of associating social networks with the clinical trial is that when a participant likes some fact about the trial, he can express his liking in the social network. This will definitely grab attention of other users of the network and may gain popularity if others also 'like' the occurrence of such an event or fact. With this, the participant will feel good about his involvement in the clinical trial and may encourage others to take part in it too.

The social informer component of the Communicator will play a significant role in user profiling while having access to the participant's social network profile. The job of the social informer will not be limited to informing the participants about the news and events of a clinical trial, but it will also be able to help the management to take communication decisions by extracting information from the participant's profile. It can be observed from the social networks how the participants communicate with their friends (chat, message, use of cell phones etc) and how their social behavior like the number of friends over time, frequency of posts and updates etc is. Also analysis of several attributes like age, gender, location and ethnicity are helpful. The following are examples of the effect of user profiling. If the collected data shows that people between age group 16 - 35 prefer to text their friend more than talking over phone, then the management may also contact them through text messages more often. Again, based on the geographical location of the participants, the management can guess their social behaviors,

cultural habits, technological advancements leading to their common ways of communication etc. Also different studies reveal distinct characteristics regarding communication among men and women. In such regards interpretation of social network data may improve communication strategy for the management.

This Communicator acts as a base for the entire system and enables the administrator to keep a check on the progress of the proceedings. The administrator is held responsible for the maintenance of the scheduler as he/she sets up the event at his/her convenience. An announcement is made in regard to the event generated. This announcement will be broadcasted to all the patients who are related to this study.

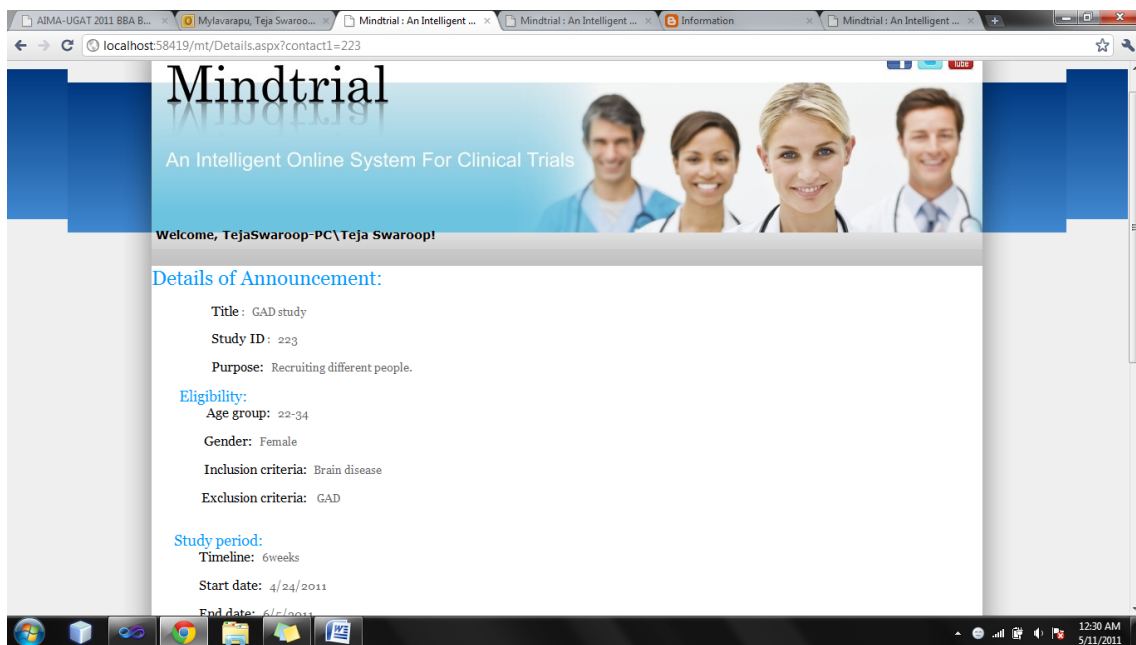


Figure 42. Study Announcement



Based on the announcement, different sessions are created and managed for the setup of different schedules. Various parameters, like name of the session conductor, event id, event type, location, and room number are considered during the setup of a single session in a schedule. Figure 43 displays management for the scheduler that is setup corresponding to an announcement.



Figure 43. Scheduler Session Management

### 6.3 Scheduler for Patients

After logging into the system, patients may want to setup an appointment for a particular event. A list of available appointments pops up for the patients so that they can choose one. These appointments are those that have been previously created by the administrator. Depending upon the patient's availability and convenience, only one appointment can be chosen.

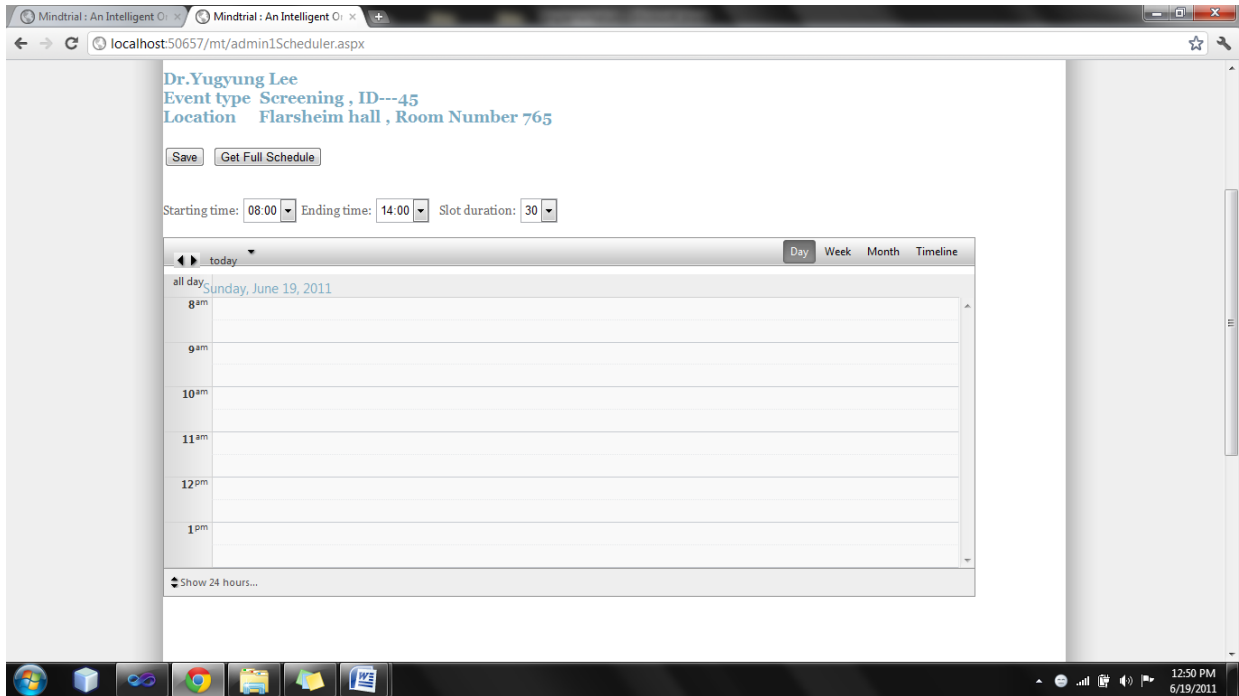


Figure 44. Scheduler for Patients

The patient may later revisit the scheduler and re-schedule the appointment if the former appointment does not suit his calendar. After selecting a convenient date, the scheduler tool will pop up for the patient to select a time on the chosen date. There will be many different slots, of which some are filled while some are free. The patient is free to setup an appointment on any available free date and may even modify it later. It is not allowed to modify other patient's appointment and the scheduler is designed in such a way that if any of the patients try to modify another person's appointment, an error message will pop up.

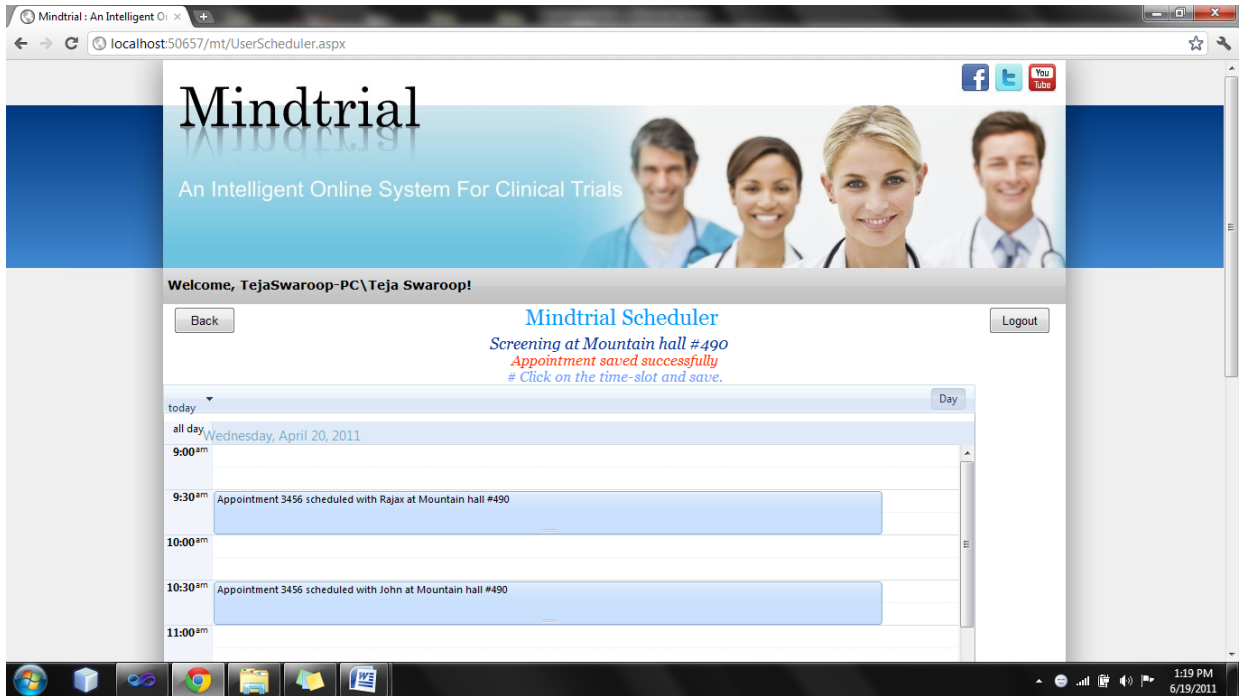


Figure 45. Appointment Setup

The above screenshot confirms the appointment of the patient with the scheduler system. A confirmation email will be sent to the patient about the appointment with the system. A reminder email is also sent to the patient one day before the appointment and thereby, confirming his/her participation. In this way, a patient schedules an appointment with the system and the corresponding even coordinator is immediately notified. The patient may meet the coordinator as per the appointment date and time and further discuss health issues or things of appropriate concern.

#### 6.4 Social Bridge in MindFlow

As humans today interact with computers, social computing has laid a strong base in its development and completely justifies its existence. Social computing is mainly explained as a graph theory as actors (or individuals) as nodes and their relationships as edges. A Social

network is defined as a structure whereby the strength of the relationship between individuals is explained based on the bonding between them.

Social computing has become more widely known because of its relationship to a number of recent trends like the popularity of social software and Web 2.0. Developing a system based on the mashup of different social networking sites and retrieving information based on the trend and popularity of the subjects, demands the tracking of recurring changes in a group.

A Web mashup is a web application that gathers content from different sources and presents it in a different way or with a unique outline. The Web is continually growing more open and more social. Because of this, many websites have opened up programming interfaces (APIs) that allow developers to get at their core information. The main characteristics of a mashup are combination, visualization, and aggregation. This has paved the way for new concepts to emerge and new challenges to be solved. These concepts are applied in MindFlow to determine the strength of the association of one patient with another.

Authentication of patients is done through OAuth 2.0 protocol that lets the users share their personal public data without disclosing their passwords. After proper authentication of the patient, an access token is generated and stored in a file on an Amazon Elastic Compute Cloud. This cloud has an IIS 7 Web server installed on a 32 bit Windows Server 2008 R2 operating system that runs on a machine with a processing speed of 2.75 GHz and 2.66 GHz and 1.66 GB of RAM.

Depending on the access token, various friends' lists are generated that are made public by the user. Social Bridge enables us to search for people via their friends, likes, interests, hash

tags, ways of interaction, strengths of association, etc. Social Bridge first generates a list of all Facebook and Twitter friends and their common friends. Social Bridge, as a search engine, enables us to search with respect to one person. A person’s name is texted in, and all the details that have been personally made public by that person are displayed, thus protecting the privacy of the user.

Table 20. Social Bridge Friends Search in MindFlow

Person Name	Friends Names
Facebook Friends	Gudib, Arch, Thilpa, Moi, Kanch, Thanthar, Maity
Twitter Friends	Pragya, Khandel, Varan, Allen, Tach, Moi, Thilpa
Twitter Followers	Varan, Khandel, Moi
Twitter common following and followers	Varan, Khandel, Moi
Facebook and Twitter combined friends	Moi, Thilpa

Table 20 gives a list of all the friends who are connected via their social networking sites. With this list, we can search for a person who has also scheduled an appointment or who is at least interested in clinical trials or may want to schedule an appointment later. We can access these people basic details and contact them by tweeting or sending an e-mail to them.

Next, we determine the person’s ‘likes’ and hash tags. This list will actually judge a person by his/her interests and the categories he/she is interested in knowing more about. A person’s ‘likes’ and hashtags that determine the person’s personal interests are taken into consideration.

Table 21. Social Bridge Categories Search in MindFlow

Category	Likes and Hashtags
Facebook Likes	Hospital, Clinic, Health, Medical Pharmacy, Home Improvement, Website, Video, University, Sports venue, Public figure, Professional

	sports team, Product/service, Personal blog, Organization, Non-profit organization, Movie, Local business, Games/toys, Education
Twitter Hashtags	Hospital, nurse, doctor, house surgeon, chetan, book, student, teacher, teachers, coding, dell, Microsoft
Twub Cluster	Hospital, Hospital, Hospital, Medicine, author, Other, Other, education, education, computers/technology, computers/technology, internet/software, Other

In this way, patients who are identified in MindFlow and who are willing to connect to people by accessing their details can connect to various numbers of people by discovering new people within their own field by using various searching techniques of Social Bridge. It is an endeavor to connect people interested in similar categories into one group and let them communicate with each other.

## CHAPTER 7

### CONCLUSION AND FUTURE WORK

#### 7.1 Summary

In this thesis, the Social Bridge Framework is proposed that dynamically generates an integrated social network to portray a user's profile of preferences and interactions with others. Social Bridge is a semantic framework which integrates data from two social networking sites, namely, Facebook and Twitter to discover and connect people. The experimental and evaluation results prove that Social Bridge gathers and contains adequate and sufficient information to develop an integrated social network that searches for a person based on the choices of that person. This search engine successfully extends the concept of Friend of a Friend (FOAF) between people and eliminates the boundaries between different social networking sites. This search engine also scrutinizes a person to the maximum with his/her shared data and has also portrays the strengths and methods of association of this person towards others. Moreover, the comparison of searches in Facebook, Twitter, and Social Bridge from Table 1 in Section 2.3 clearly depicts the efficiency of our system and the novel approach to connect with a person. However, there are certain issues like Person identification, Hashtags recommendations which needs to be discussed to build Social Bridge.

#### 7.2 Person Identification

Identifying a person in a single domain or in a single social networking site is considered an easy job. This identification requires a person name or the person's unique id in the network or e-mail id or any factor which uniquely identifies the person. A person in a single domain is unique and all his details are retrieved very easily.

But, the scenario changes a bit during the integration of various social networking sites into one integrated site and develop the same into a search engine. This is because integration is not very easy. We have to carefully consider the factors affecting the change in search of a person. Various factors like unique id in a network, e-mail id cannot be considered to search a person in an integrated network because these factors might vary for various social networking sites.

Search by first name and last name, search by hashtags, search by Twub Cluster, search by Likes is considered unique for all the social networking sites. Supporting this search, search by gender, and search by location is also considered unique during the search of a person in an integrated social networking site.

### 7.3 Hashtags Recommendations

During the search of a person based on hashtags, the hashtags are classified and are searched based on the tweets of person in the repository. This search also enables hashtag recommendation which enlists a set of recommended hashtags for the hashtags searched.

The hashtags recommendation is enabled using WordPress API which results in the synonyms of the hashtag searched. This search is highly recommended as this lets the user to search for more number of people based on more number of hashtags. But if the hashtag has no meaning, then there are no words returned only the direct names of the persons who have tweeted with the hashtag are returned.



## 7.4 Time Complexity and Efficiency

The search for a person in Facebook is confined to the concept of Friend of a Friend and it is not possible to search for a person who is four layers away. This search has been very successfully implemented in Social Bridge in a time of 3 minutes 47 seconds.

This can be further enhanced by reducing the time complexity of the network. The efficiency remains the same but the complexity can be reduced further using various time and space efficient algorithms which utilize less memory space and iterate through all the names of people in an efficient manner.

The algorithm implemented in Social Bridge is tested efficient for 12,747 members iterating through the loop. This can be further increased and better time efficient algorithms may be applied as the existing social networking sites do not provide such search.

## 7.5 Future Work

Social Bridge undoubtedly leaves a lot of scope for the future. Some of the features that might be extended are as follows:

1. Extending the integration of social networking sites to LinkedIn and Google+.
2. Modeling an interactive visualization of Social Bridge to facilitate better understanding.
3. Enhancing the choice of preference during the search of a person by including factors like search by connections, mentions, genders, groups, jobs, companies, etc.
4. Evaluating the performance of the developed tool for Social Bridge.

## REFERENCES

- [1] Tiffany A. Pempek, Yevdokiya A. Yermolayeva, Sandra L. Calvert, College students' social networking experiences on Facebook, *Journal of Applied Developmental Psychology* Vol.30 (2009) pp.227–238
- [2] Meredith M. Skeels Biomedical & Health Informatics, University of Washington Seattle, Jonathan Grudin, Microsoft Research, When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn[Case Study]
- [3] Manoj Parameswaran, Andrew B. Whinston SOCIAL COMPUTING: AN OVERVIEW, *Communications of the Association for Information Systems* (Volume 19, 2007) 762-780
- [4] Tim O'Reilly, What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies*, No. 1, p. 17, first Quarter 2007
- [5] P. Sri Jothi, M. Neelamalar and R. Shakthi Prasad, Analysis of social networking sites: A study on effective communication strategy in developing brand communication, *Journal of Media and Communication Studies* Vol. 3(7), pp. 234-242, July 2011, URL: <http://www.academicjournals.org/jmcs/PDF/pdf2011/July/Jothi%20et%20al.pdf>
- [6] Harvey Jones, Jos\_e Hiram Soltren, Facebook: Threats to Privacy, December 14, 2005,[Survey Paper] URL: <http://groups.csail.mit.edu/mac/classes/6.805/student-papers/fall05-papers/facebook.pdf>

- [7] Daniel M. Romero, Brendan Meeder, Jon Kleinberg, Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, URL:<http://www.cs.cornell.edu/home/kleinber/www11-hashtags.pdf>
- [8] SHRM: Society for Human Resource Management, <http://www.shrm.org/research/surveyfindings/articles/pages/theuseofsocialnetworkingwebsitesandonlineSearchenginesinscreeningjobcandidates.aspx>
- [9] Kaveri Subrahmanyama,b,!, Stephanie M. Reich c, Natalia Waechter b,d,e, Guadalupe Espinoza, Online and of!ine social networks: Use of social networking sites by emerging adults!
- [10] Simon Carter, Manos Tsagkias, Wouter Weerkamp, Twitter hashtags: Joint Translation and Clustering
- [11] Mike Gartrell, Richard Han, Qin Lv, Shivakant Mishra. SocialNews: Enhancing Online News Recommendations By Leveraging Social Network Information. Technical Report, CU-CS-1084-11
- [12] Jason J. Jung (Knowledge Engineering Laboratory Department of Computer Engineering Yeungnam University), Integrating Social Networks for Context Fusion in Mobile Service Platforms, Journal of Universal Computer Science, vol. 16, no. 15 (2010), 2099-2110
- [13] Collin McMillan, Mark Grechanik, Denys Poshyvanyk, Chen Fu, Qing Xie, Exemplar: A Source Code Search Engine For Finding Highly Relevant Applications, IEEE Transactions on Software Engineering, Vol. X, No. X, Month 201X
- [14] Open Standard for Authorization, <http://oauth.net/2/>
- [15] Atom in Wikipedia, [http://en.wikipedia.org/wiki/Atom\\_\(standard\)](http://en.wikipedia.org/wiki/Atom_(standard))

- [16] RSS in Wikipedia, <http://en.wikipedia.org/wiki/RSS>
- [17] Sunhong Kim and Sangyong Han, The method of inferring trust in web-based social network using Fuzzy Logic
- [18] Twitter REST API. <https://dev.twitter.com/docs/api>
- [19] Graph api - facebook developers. <https://developers.facebook.com/docs/reference/api/>.
- [20] Cloud Computing, [http://www.wikinvest.com/concept/Cloud\\_Computing](http://www.wikinvest.com/concept/Cloud_Computing)
- [21] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: Experiments on recommending content from information streams. In Proceedings of the 28th international conference on Human factors in computing systems, pages 1185–1194. ACM, 2010.
- [22] Cytoscape Open Source Software, <http://www.cytoscape.org/>
- [23] Lee, Y., Dinakarbandian, D., Katakam, N., Owens, D. MindTrial: An Intelligent System for Clinical Trials, American Medical Informatics Association (AMIA). 2010 Annual Symposium
- [24] Yugyung Lee, Sourav Jana, Teja Mylavarapu, Deendayal Dinakarbandian, Dennis Owens MindFlow: Intelligent Workflow for Clinical Trials in Mental Healthcare, Hawaii International, Conference on System Sciences – 45, 2012
- [25] Telerik, .NET reporting software, <http://www.telerik.com/>

## VITAE

Teja Swaroop Mylavarapu was born on January 12<sup>th</sup>, 1989, in Andhra Pradesh, India. He spent most of his childhood and teenage years in Hyderabad, India. He then completed his Bachelors in Computer Science from Bandari Srinivas Institute of Technology, affiliated to Jawaharlal Nehru Technological University at Hyderabad, India. After his under graduation he moved to United States for his higher studies. He is currently pursuing his Masters in Computer Science at the University of Missouri – Kansas City.

He worked as a software developer intern for Cerner Corporation during summer 2011 and then continued as an intern and researcher for Midwest Psychiatric Research Group, INC from spring 2011. Through this intern he worked with Dr. Yugyung Lee on the research project titled MindTrial funded by National Institute of Mental Health (NIMH). He was awarded the outstanding student award for his academic and overall performance at UMKC appreciating his performance.