Identifying the Degree of Separation with the Tool ‘‘We Are So Close’’

Ms.S.Sridevi.
M.Tech,
Computer Science & Engineering,
Kalasalingam University, Krishnan Kovil, Tamilnadu, India.

***Abstract*— A social networking service is a platform to build social networks or social relations among people who share similar interests, activities, backgrounds or real-life connections. Social networking sites are popular online communication forms among adolescents and emerging adults. Yet little is known about young people's activities on these sites and how their networks of “friends” relate to their other online and offline networks. A social network is a collection of people, each of whom is acquainted with some subset of the others. Such a network can be represented as a set of points denoting people, joined in pairs by lines denoting acquaintance. One could, in principle, construct the social network for a company or firm, for a school or university, or for any other community up to and including the entire world. Since social networking sites, such as MySpace and Facebook, began allowing organizations to create profiles and become active members, organizations have started incorporating these strategies into their public relations programming. A friend of your friend probably knows a friend of others friend; Facebook shrunk the gap between us. This project re-evaluate and extend the six degrees of separation theory by using a real social searching Facebook tool ‘‘we R So Close’’.**

***Index Terms*—Social network, Facebook, Myspace, Degrees of separation.**

1.INTRODUCTION

Social network sites (SNSs) are increasingly attracting the attention of academic and industry researchers intrigued by their affordances and reach. This special theme section of the Journal of Computer-Mediated Communication brings together scholarship on these emergent phenomena. Since their introduction, social network sites (SNSs) such as MySpace, Facebook, Cyworld, and Bebo have attracted millions of users, many of whom have integrated these sites into their daily practices.

As of this writing, there are hundreds of SNSs, with various technological affordances, supporting a wide range of interests and practices. While their key technological features are fairly consistent, the cultures that emerge around SNSs are varied. Most sites support the maintenance of pre-existing social networks, but others help strangers connect based on shared interests, political views, or activities. Some sites cater to diverse audiences, while others attract people based on common language or shared racial, sexual, religious, or nationality-based identities. Sites also vary in the extent to which they incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video-sharing.

Scholars from disparate fields have examined SNSs in order to understand the practices, implications, culture, and meaning of the sites, as well as users’ engagement with them. This special theme section of the Journal of Computer-Mediated Communication brings together a unique collection of articles that analyze a wide spectrum of social network sites using various methodological techniques, theoretical traditions, and analytic approaches. By collecting these articles in this issue, our goal is to showcase some of the interdisciplinary scholarship around these sites.

A social network is a collection of people, each of whom is acquainted with some subset of the others. Such a network can be represented as a set of points (or vertices) denoting people, joined in pairs by lines (or edges) denoting acquaintance. One could, in principle, construct the social network for a company or firm, for a school or university, or for any other community up to and including the entire world.

Social networks have been the subject of both empirical and theoretical study in the social sciences for at least 50 years, partly because of inherent interest in the patterns of human interaction, but also because their structure has important implications for the spread of information and disease. It is clear, for example, that variation in just the average number of acquaintances that individuals have (also called the average degree of the network) might substantially influence the propagation of a rumor, a fashion, a joke, or this year's flu. A famous early empirical study of the structure of social networks, conducted by Stanley Milgram, asked test subjects, chosen at random from a Nebraska telephone directory, to get a letter to a target subject in Boston, a stockbroker friend of Milgram's.

The instructions were that the letters were to be sent to their addressee (the stockbroker) by passing them from person to person, but that they could be passed only to someone whom the passer knew on a first-name basis. Because it was not likely that the initial recipients of the letters were on a first-name basis with a Boston stockbroker, their best strategy was to pass their letter to someone whom they felt was nearer to the stockbroker in some sense, either social or geographical: perhaps someone they knew in the financial industry, or a friend in Massachusetts.

Although there were certainly biases present in Milgram's experiment letters that took a longer path were perhaps more likely to get lost or forgotten, for instance his result is usually taken as evidence of the “small-world hypothesis,” that most pairs of people in a population can be connected by only a short chain of intermediate acquaintances, even when the size of the population is very large. Milgram's work, although cleverly conducted and in many ways revealing, does not, however, tell us much about the detailed

structure of social networks, data that are crucial to the understanding of information or disease propagation. Many other studies have addressed this problem did the same for communities of Utah Mormans, Native Americans, and Micronesian islanders, and there are many other examples to be found in the literature. Surveys or interviews were used to determine friendships.

These networks, however, suffer from a different problem: although they may loosely be said to be social networks in the sense that their structure in some way reflects features of the society that built them, they do not directly measure actual contact between people. Many researchers, of course, are interested in these networks for their own sake, but to the extent that we want to know about human acquaintance patterns, power grids and computer networks are a poor proxy for the real thing. In this project, the system re-evaluate and extend the six degrees of separation theory by using a real social searching Facebook tool ''We R So Close''.

2. Related Work

2.1 A Distributed Storage System for Structured Data

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, both in terms of data size and latency requirements. Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products. In this paper we describe the simple data model provided by Bigtable, which gives clients dynamic control over data layout and format, and we describe the design and implementation of Bigtable.

2.2 Semantic web-based social network access control

The existence of online social networks that include person specific information creates interesting opportunities for various application s ranging from marketing to community organization. On the other hand, security and privacy concerns need to be addressed for creating such applications. Improving social network access control systems appears as the first step toward addressing the existing security and privacy concerns related to online social networks.

To address some of the current limitations, we have created an experimental social network using synthetic data which we then use to test the efficacy of the semantic reasoning based approaches we have previously suggested.

By constructing such an ontology, we model the Social Network Knowledge Base (SNKB). The main advantage for using an ontology for modeling OSN data is that relationships among many different social network concepts can be naturally represented using OWL. Furthermore, by using reasoning, many inferences about such relationships could be done automatically. Our access control enforcement mechanism is then implemented by exploiting this knowledge. In particular, the idea is to define security policies as rules whose antecedents state conditions on SNKB, and consequents specify the authorized actions.

In this paper, we have proposed an extensible fine-grained online social network access control model based on semantic web tools. In addition, we propose authorization, administration and filtering policies that are modeled using OWL and SWRL. The architecture of a framework in support of this model has also been presented. Further, we have implemented a version of this framework and presented experimental results for the length of time access control can be evaluated using this scheme. Further work could be conducted in the area of determining a minimal set of access policies that could be used in evaluating access requests in a further attempt to increase the efficiency of these requests.

2.3 Spreading Information in Social Network

The problem of spreading information in social networks is a topic of considerable recent interest, but the conventional influence maximization problem which selects a set of any arbitrary k nodes in a network as the initially activated nodes might be inadequate in a real-world social network – cyberstalkers try to initially spread a rumour through their neighbours only rather than arbitrary users selected from the entire network. To consider this more practical scenario, Kim and Eiko introduced the optimisation problem to find influential neighbours to maximise information diffusion. We extend this model by introducing several important parameters such as user propagation rate on his (or her) neighbours to provide a more general and practical information diffusion model. We performed intensive simulations on several real-world network topologies (emails, blogs, Twitter and Facebook) to develop more effective information spreading schemes under this model. Unlike the results of previous research, our experimental results shows that information can be efficiently propagated in social networks using the propagation rate alone, even without consideration of the "number of friends" information.

Assume that a spreader $s \in V$ wants to spread a piece of information r through the network $G = (V, E)$ by sharing r with its min (k, d(s)) neighbours only. Node s first tries to assess the influence of information diffusion for each neighbour $v \in N(s)$, respectively, by collecting the information about v. We note that neighbours' influence should be estimated based on s's local information only, rather than the whole network. As online social networks such as Facebook typically provide APIs to get the neighbourhood information about user, s might automatically collect the information about its own neighbours.

After estimating the neighbours' influences, s selects the top min(k, d(s)) nodes with the highest estimated values from N (s) as the most influential neighbours for information diffusion; that is, for the IC model in Section 2, they are chosen as the set of initially activated nodes $S0 \subseteq V$.To accelerate the speed of information diffusion, a possible straightforward approach is to increase the number of initially activated neighbours k. Probably, we can imagine that even the naive Random selection scheme can also be used to efficiently disseminate a piece of information if k increases overtime.

3. Method

The feasibility study is carried out to test whether the proposed system is worth being implemented. The proposed system will be selected if it is best enough in meeting the performance requirements. The feasibility carried out mainly in three sections namely.1. Economic Feasibility, 2. Technical Feasibility, 3. Behavioral Feasibility.

4. Methodology

4.1 Economic Feasibility

Economic analysis is the most frequently used method for evaluating effectiveness of the proposed system. More commonly known as cost benefit analysis. This procedure determines the benefits and saving that are expected from the system of the proposed system. The hardware in system department if sufficient for system development.

4.2Technical Feasibility

This study center around the system's department hardware, software and to what extend it can support the proposed system department is having the required hardware and software there is no question of increasing the cost of implementing the proposed system. The criteria, the proposed system is technically feasible and the proposed system can be developed with the existing facility.

4.3Behavioral Feasibility

People are inherently resistant to change and need sufficient amount of training, which would result in lot of expenditure for the organization. The proposed system can generate reports with day-to-day information immediately at the user's request, instead of getting a report, which doesn't contain much detail.

5. Design Approach

5.1 *Dataset Preprocessing*

Before executing the main steps of our experiments, the system had a pre-process procedure to check and filter out invalid data. The filtering function was responsible for analyzing the participant's information in order to delete all fake accounts. This preprocess step was very important for minimizing the influence from invalid data and increasing the accuracy of our experiment. Therefore, the current graph database analyzed in this project is simply graph of real Facebook users who used our ''We R So Close tool''at the time when the experiments were performed a real account is the account that represents actual individuals.

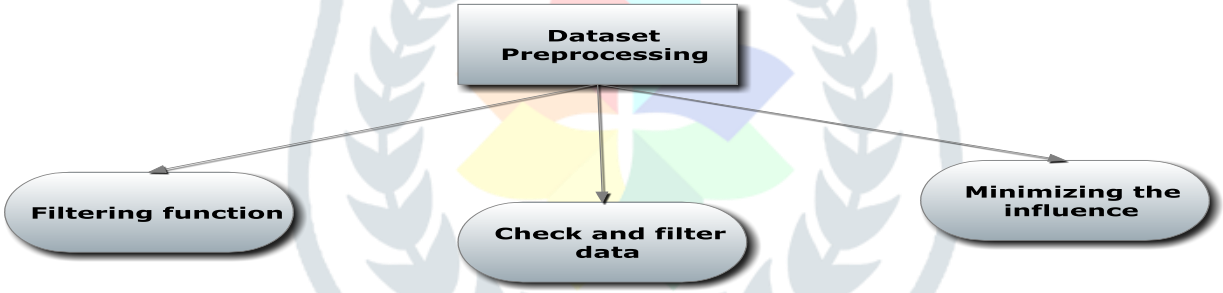

Figure 1.Dataset preprocessing

5.2 Hierarchical Function

A hierarchical database model is a data model in which the data is organized into a tree-like structure. The data is stored as records which are connected to one another through links. A record is a collection of fields, with each field containing only one value. The Hierarchical Data Model is a way of organizing a database with multiple one to many relationships. The structure is based on the rule that one parent can have many children but children are allowed only one parent.

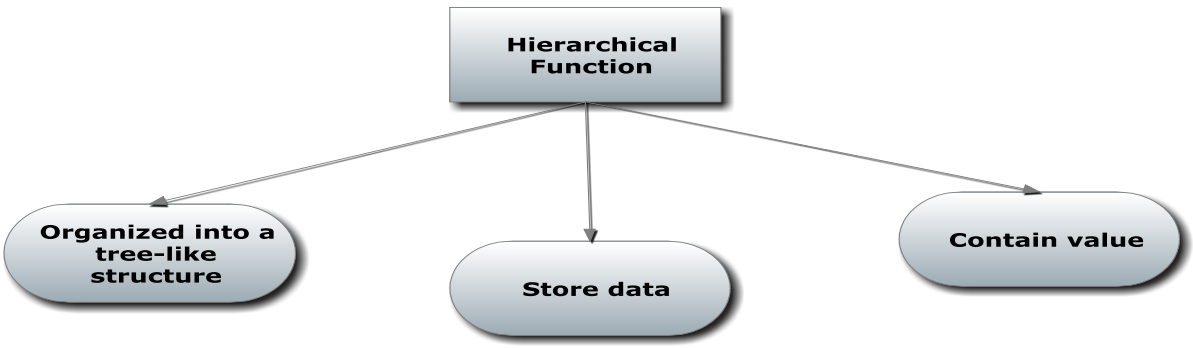

Figure2.Hierarchical Function

5.3 Find Shortest Neighbors

The degree of separation between two nodes is the length of the shortest path between them. The degree of clustering of some node is defined as the real number of edges between these its neighbors divides the maximal number of edges. Once a participant provided consent, s/he was directed to the tool main page that had two main searching functions; the basic searching function found out the path between any two active Facebook users by using the shortest path algorithm and returned all intermediaries between them. The advanced search function required to select a profession from a list in order to show the path between the participants connected to someone who worked in the selected profession. Hence, we got a classification of professions from the bureau of employment and vocational training in order to identify clearly the list of professions provided by this function; each profession has a title and an id number. This function found out the path between two persons who have a rare-special feature that might be a rare profession; it returned all the shortest paths between them.

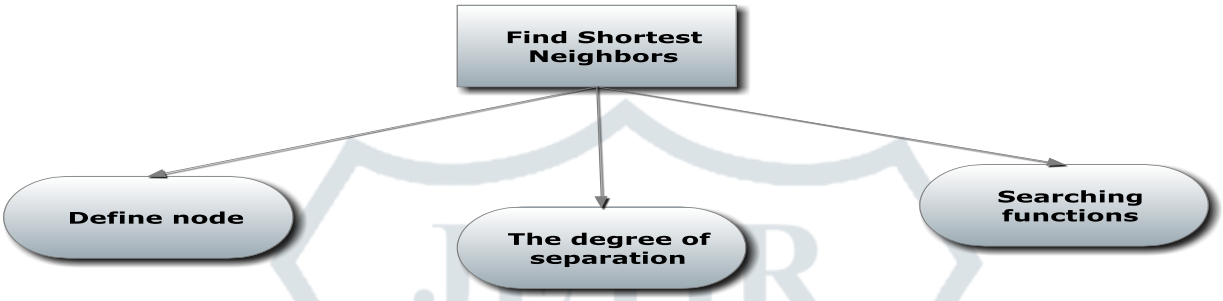

Figure3.Find Shortest Neighbours

5.4 Find fake Account

The function checked the profile pictures folder for the participant. If there is only one photo in the whole profile it makes it quite clear that the account may be ‘‘fake account’’. It checked the info. If the function found that there is no ideal links given regarding school or educational institutions or workplace and that the user is looking for dating and interested in both men and women, it shows signs of ‘‘fake account’’. It checked out the friend list. If found that maximum of the friends are of the opposite gender, it can be assumed that the profile is used either for fun or for random dating ‘‘fake account’’. It looked out for recent wall posts, if it sees loads of people asking ‘‘thanks for adding me, do I know you’’

And yet the posts remains unanswered, it is bound to be a ‘‘fake one’’.It checked the user profile status updates, wall posts and comments. If the user has not updated a status for quite a long time and has not been involved in any wall posting or commenting of other statuses, it means that the profile is likely to be ‘‘non-real account’’.It checked the log file, if the account only used for games, it is clear to be ‘‘non-real account’’.It looked at the recent activities. If it is that the user has just been adding random users and making new friends, and that there are no pages liked or groups joined, it suggests that the user is determined by just adding people and hence the profile is likely to be ‘‘non-real account’’.It checked the account information, if we find that identical information stored in our database then this profile is considered as ‘‘duplicate account’’.

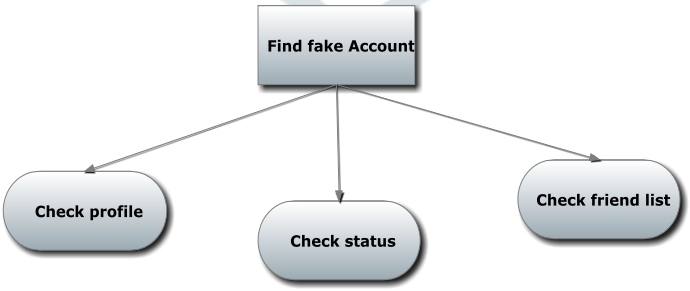

Figure 4.Find Fake Account

6.Implementation of Facebook tool

In proposed system, the system re-evaluates and extend the six degrees of separation theory by using a real social searching Facebook tool ‘‘We R So Close’’. It works in facebook platform. The graph database is used to store the collected data. For this process the system considering the two points: considering the fake accounts; finding someone with special rare features. And also it extends the theory from different viewpoints.

The avgDegreeJob computed the degree of separation between two nodes such as at least one of these nodes had a job property that was listed on the joblist. The joblist is a text file which contains all the target jobs. The avgDegreeJob returned the average degree of separation, standard deviation, and the spending time. TheavgClusterDegree computed the degree of clustering of the target database and returned the average degree of clustering and the spending time. The function examined the account by some weighted criteria; each criterion had a pre-defined point, the function computed the probability of removing that account and made a decision based on the account.

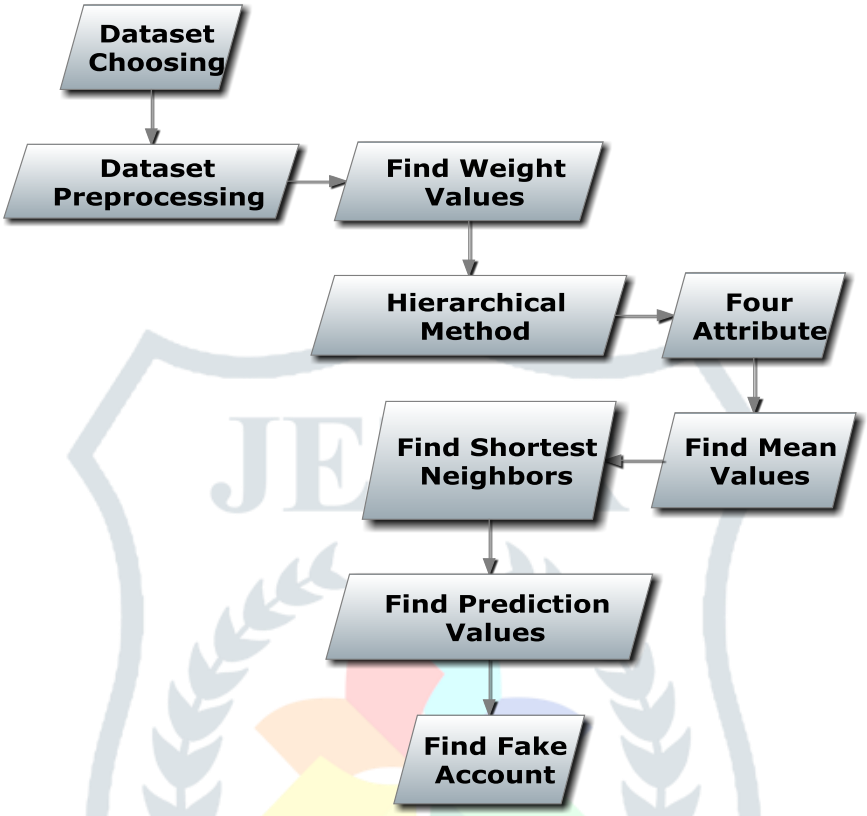

Figure 5.Block Diagram of Identify Degrees of Separation

To achieve the goals of this study, the following basic information was required to be filled out one time by participants who visited our tool for the first time; the participant elementary school; junior high school; the participant current and previous profession. Simultaneously, the system retrieved the participants' friends list and stored it in the Neo4j graph database. The decision to restrict our study to real users allows us to eliminate accounts that have been abandoned in early stages of creation; fake accounts; duplicate accounts, and focus on accounts that plausibly represent actual individuals. The graph does not include'' pages'' that people may ''like'' or ''groups'' that people may join. Even a person who works in rarest profession can be found within 4 degrees of separation; not only are there few degrees of separation between any two people even with rare-special features, but that individuals can successfully navigate these short paths, even though they have no way of seeing the entire network. These results bolstered our previous conclusion, less than four degrees separating you and others in the world.

The sequence diagram of the facebook tool "we are so close" is shown in Figure6.once a data is preprocessed through the filtering function and then proceed to hierarchical function to process the data in a organized like tree structure. Next to define a node to find shortest neighbors and then check the profile to find the fake account. The profile process all the data to store in the degrees of separation and then filter particular data for identifying the status of the profile. The advanced searching functions check the friends list to identify the fake and original account through this the dataset will preprocess all the values and store data in a dataset.

[a]Advantages

- It can re-verify, extend, and generalize the small world theory.
- It can compute a huge amount of social data.
- It is not limited to the tracing of linear chains.
- It is possible in reality to re-evaluate the six degrees of separation theory.

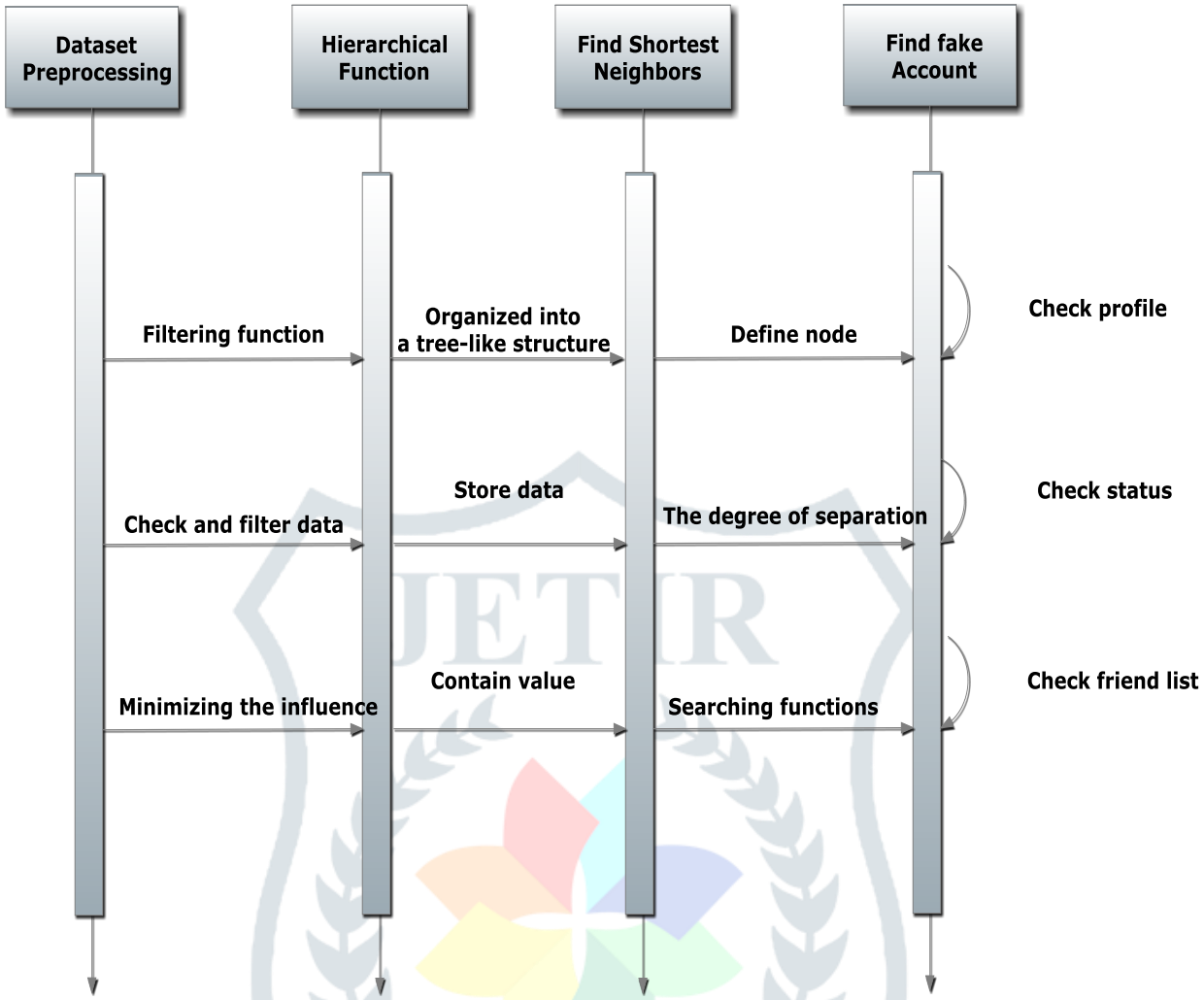

Figure 6.The Procedure sequence diagram

7. Experimental Results

The graph does not include‘‘pages’’ that people may ‘‘like’’ or ‘‘groups’’ that people may join. Even a person who works in rarest profession can be found within 4 degrees of separation; not only are there few degrees of separation between any two people even with rare-special features, but that individuals can successfully navigate these short paths, even though they have no way of seeing the entire network. These results bolstered our previous conclusion, less than four degrees separating you and others in the world.

Figure7.Compare Shortest Neighbors

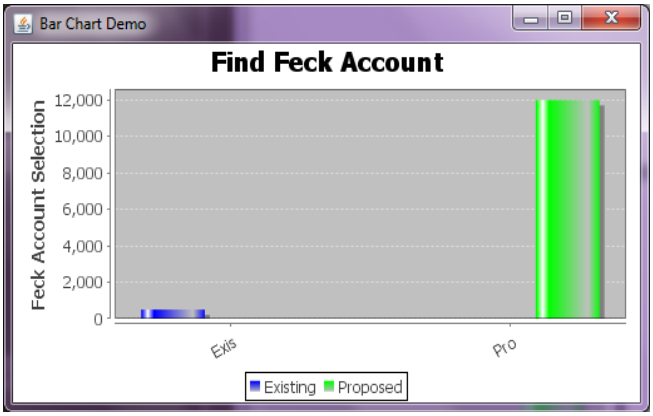

Figure 8.Find Fake Account

The following are the advantages to be used to identify the fake account in Facebook tool.1.It can re-verify, extend, and generalize the small world theory,2.It can compute a huge amount of social data,3.It is not limited to the tracing of linear chains,4.It is possible in reality to re-evaluate the six degrees of separation theory.

8. Conclusion

In this project, the system has extended the small theory from different viewpoints. A social network is a collection of people, each of whom is acquainted with some subset of the others. Such a network can be represented as a set of points (or vertices) denoting people, joined in pairs by lines (or edges) denoting acquaintance. One could, in principle, construct the social network for a company or firm, for a school or university, or for any other community up to and including the entire world. Since social networking sites, such as MySpace and Facebook, began allowing organizations to create profiles and become active members, organizations have started incorporating these strategies into their public relations programming. A friend of your friend probably knows a friend of others friend; Facebook shrunk the gap between us. It re-evaluated the degree of separation theory using data collected by an online Facebook tool. Identifying the degree of separation from a different viewpoint by considering not only the degree of separation between two normal-persons or famous-persons, but also between two persons with very rare-special features, and observing the influence of the high-degrees node.

9. Future Work

In future, to develop a communication process for two original facebook account in a secured way and encrypt all the data in a duplication of code to generate in a voice message through private key separation process. The key should provide a unique identity to two accounts and to process a communication to two people through video chatting and calling.

10. References

[1] Hyoungshick Kim **"**Don't Count The Number of Friends When You Are Spreading Information in Social Networks**"**, 2014.

[2]Erdös, P. &Rényi, A. (1959). On random graphs, I. PublicationesMathematicae (Debrecen), 6, 290–297.Facebook. (2013).Company Timeline. <https://www.facebook.com/data> Retrieved03.09.13.

[3]Karinthy, F. (1929). Chains. Everything is different. In: AtheneumPress.Leavitt, N. (2010). Will NoSQL databases live up to their promise? Computer, 43(2), 12–14.

[4]Leskovec, J., & Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. In Proceedings of the 17th international conference on World Wide Web (pp.915–924).Beijing, China: ACM.Mathworld. (2011). Regular Graph. <http://mathworld.wolfram.com/RegularGraph.html> Retrieved 03.09.13.

[5] Stonebraker, M., & Hong, J. (2009). Saying good-bye to DBMSs, designing effective interfaces. Communications of the ACM, 52(9), 12–13.Sysomos. (2010). Six Degrees of Separation,TwitterStyle.<http://www.sysomos.com/insidetwitter/sixdegrees/> Retrieved 03.09.13.

[6] Gurevitch, M. (1961). The social structure of acquaintanceship networks. Massachusetts: Massachusetts Institute of Technology.Theguardian. (2008). Proof! Just six degrees of separation between us. <http://www.guardian.co.uk/technology/2008/aug/03/internet.email> Retrieved 03.09.13.