A Project Proposal on

Facebook Data Exploratory Analysis

By: Team Member Name: **Sneha Godbole**

Email id: **snegodbo@umail.iu.edu**

Table of Contents

[1. Abstract 3](#_Toc462435257)

[2. Introduction 3](#_Toc462435258)

[**Motivation** 3](#_Toc462435259)

[**Data Introduction: Facebook New Orleans Network** 4](#_Toc462435260)

[**Data statistics** 4](#_Toc462435261)

[**Data Limitations** 4](#_Toc462435262)

[3. Background & Related Work 5](#_Toc462435263)

[**Related Work** 5](#_Toc462435264)

[**Dataset** 5](#_Toc462435265)

[4. Research Questions & working hypothesis 5](#_Toc462435266)

[REFERENCES 6](#_Toc462435267)

# **Abstract**

Online social networks have become extremely popular; many well known as well as emerging web sites provides integration with social networks like Facebook, LinkedIn, Twitter and so on, in order for users to share the content. Users of these networks often establish hundreds to even thousands of social links with other users. Many researchers are working on areas to explore use of actual interaction between users, rather than mere friendship – to distinguish between strong and weak links. There are many studies on how an activity network is structurally different from the social network itself. I found a very good research paper in regards with this, studying the evolution of activity between users in the Facebook social network to capture this notion. The paper finds that links in the activity network tend to come and go rapidly over time, and the strength of ties exhibits a general decreasing trend of activity as the social network link ages. For example, only 30% of Facebook user pairs interact consistently from one month to the next. Interestingly, the paper also finds that even though the links of the activity network change rapidly over time, many graph-theoretic properties of the activity network remain unchanged. The overall research and analysis being done on this data looks very exciting to me, and I am going to perform exploratory analysis of the dataset being used for this paper, to validate their results as well as to find few more interesting patterns / insights from the dataset based on concepts learned on Network Science course so far.

# **Introduction**

## **Motivation**

Online social networks have become a popular way for users to connect, express themselves, and share content. Popular social networks like Facebook, LinkedIn have hundreds of millions of registered users and are growing at a rapid pace. As these networks grow and mature, users have been observed to form hundreds to even thousands of friendship links.

For example, I have around 800+ active friends on Facebook. However, I do not actively interact with all of them every month! Some of them are my best friends, some average friends and some just acquaintances. Our ties too vary, from very strong to very weak. The interactions / activities we do (sharing, liking post, writing on wall, etc.) with them is dependent on the type of tie often times. So, apart from a connectivity graph – friends’ connection graph – the activity linkage / graph is very important in order to understand the actual “actual” / “active” friends among my circle. Some research work mentions this network as “activity network” which has interesting properties / patterns as compared to “connectivity networks” (example: being a friend on Facebook or LinkedIn).

I found an interesting research paper, from [Research Papers Link](https://icon.colorado.edu/#!/networks) , under social network domain, studying the evolution of activity between users in the Facebook social network to capture this notion. The overall research and analysis being done on this data looks very exciting to me, and I am going to perform exploratory analysis of the dataset being used for this paper, to validate their results as well as to find few more interesting patterns / insights from the dataset based on concepts learned on Network Science course so far.

## **Data Introduction: Facebook New Orleans Network**

The research paper focuses on the New Orleans regional network in Facebook for the study. The data is collected over two separate periods. First, during December 29th, 2008 and January 3rd, 2009. In this period, the information about friendship links (i.e., who is friends with whom) was collected from a single user in the New Orleans network and visited all friends of the user and their friends in a breadth-first-search (BFS) fashion, consistent with previous social network crawls. Second, the wall feature in Facebook as a form of user interaction in the social network were crawled. In Facebook, a user’s friends can post comments to the user’s wall; these comments appear on the user’s wall and can be seen by others who visit the user’s profile. Therefore, wall posts represent a broadcast-style messaging service within the site. This second crawl between January 20th, 2009 and January 22nd, 2009, and had downloaded the entire wall history for all users whom were previously discovered in earlier crawl. Each wall post entry in the data set contains information about the wall owner, the user who made the post, the time at which the post was made, and the post content.

## **Data statistics**

The total data gathered information about 90,269 users and 3,646,662 friendship links between those users. This accounts for 52% of the users in the New Orleans network based on the statistics provided by Facebook. However, not all users make their profiles public; this data could view the wall postings of 60,290 (66.7%) users. These 60,290 users were connected together by 1,545,686 links in the social network with an average node degree of 25.3. In the paper I studied and the implementation I am targeting to do, I consider only this set of users. The wall post data spans from September 26th, 2006 to January 22nd, 2009. In total, 838,092 wall posts were observed, for an average of 13.9 wall posts per user. This covers communication between 188,892 distinct pairs of users, representing 12.2% of the links in the social network. The remaining 87.8% of the link in the social network did not exhibit any wall activity.

## **Data Limitations**

First, in Facebook, users can interact in many ways (e.g., messaging, applications, photo uploads, and chat). While wall posting is one of the most popular methods of user inter- actions, I do not know if it is representative of other forms of interaction. Second, our Facebook crawl was limited to the subset of people who made their profiles visible to the people in the same regional network, which is the default privacy setting in Facebook. Third, the data collected is only for users in the giant connected component of the New Orleans networks.

# **Background & Related Work**

## **Related Work**

I found a very good research paper and literature review which gave me an aspiration, from [Research Papers Link](https://icon.colorado.edu/#!/networks), under Social network domain category, named Facebook Wall Posts 2009. This research paper is named  **"On the evolution of user interaction in Facebook”,** was published in WOSN’09, August 17, 2009, Barcelona, Spain [[1]](#_REFERENCES). The direct link to paper PDF is [Paper Link](http://www.mpi-sws.org/~gummadi/papers/wosn23-viswanath.pdf) The paper finds that links in the activity network tend to come and go rapidly over time, and the strength of ties exhibits a general decreasing trend of activity as the social network link ages. For example, only 30% of Facebook user pairs interact consistently from one month to the next. Interestingly, the paper also finds that even though the links of the activity network change rapidly over time, many graph-theoretic properties of the activity network remain unchanged.

## **Dataset**

The dataset which was used for this research work is available at [Dataset Link](http://konect.uni-koblenz.de/networks/facebook-wosn-wall) , [Dataset1](http://socialnetworks.mpi-sws.org/data-wosn2009.html),and I will use same dataset for my validations as well as experiments. Also I have checked the dataset it seems good to me.

# **Research Questions & working hypothesis**

The working hypothesis for my work is I will be able to validate the research and literature survey published in [Paper Link](http://www.mpi-sws.org/~gummadi/papers/wosn23-viswanath.pdf). The implementation will lead to interesting exploratory analysis and I will get some meaningful insights from the friendship as well as activity network.

I am planning to get and visualize following meaningful insights from the data after implementation –

1. Number of wall post per month by all the users.
2. Calculating and validate following network properties for Friendship Network
   1. Clustering Coefficient
   2. Degree Distribution
3. Rate of interaction over time based on fraction of wall posts sent since link formation between / among users.
4. For Activity network, analyze and visualize the network diagram to get meaningful insights on who is a more popular user in this friendship network, based on user activities (wall posts). I will build a network graph for activity network and find out a user with most incoming walls for this insight.

# **REFERENCES**

[1]. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. "On the evolution of user interaction in Facebook." Proc. 2nd ACM SIGCOMM Workshop on Social Networks, 37-42 (2009) - [Link](http://www.mpi-sws.org/~gummadi/papers/wosn23-viswanath.pdf)