# User interactions in social networks


## Srinath Narayanan

### Graduate student, UCSD



## ABSTRACT

The growth of social media has reached unprecedented orders, and it is expected to grow exponentially. With numerous sites allowing users to share, interact, critique and evolve, the social networks are growing ever popular, accounting for nearly 35% of the internet traffic. Users of such interactions tend to establish multiple links based on their geographic, ethnographic and linguistic similarities. Thus it is important to study the activity interaction and activity networks of users, rather than the connectivity. To expand, it is important to analyze how we interact with our network, more than to understand how we are connected. Studies conducted on multiple activities centers like Reddit, Twitter and Facebook reveal that these activity sources vary much more than the connection network; and it grows over time for the same connection network, meaning we interact with different people differently over time. This work is inspired by the paper by Viswanath et.al on evolution of interaction networks at the Max Planck institute and is adapted to my Facebook dataset, to understand how the activity network is sampled from my perspective; keeping myself as the center activity node. This work is submitted for the "Final poster presentation" of ECE-289 class, Fall 2017.


## INTRODUCTION

The inspiration for this topic yields from the enormous growth of social media over the past decade. Facebook, found in 2004, crossed the network threshold of a million by 2008; Twitter was found in 2006 and is now the most popular supplier of quick news-bits and global communication; Tumblr found in 2007 is the go-to place to destress and laugh at microblogs; and Pinterest launched in 2010 is the search space for images, gifs etc. As we see over the past decade these networks have grown enormously and often outgrew all the expert expectations. It is a new wave in how humans interact, and along with the introduction of 4G networking, rapid prototyping of applications, reduced overhead on communications etc., they are changing the landscape of communication.

The growth of these networks fuel the formation of bubbles of small groups of communities, ideally exhibiting the in-out structure [1] where they form strong communities but also display the small-world phenomenon. These networks share hundreds of millions of shared users and grow at a rapid pace and is hard to be completely quantified by any theoretical model. Users are observed to form friendship communities ideally ranging for tens to thousands of friends. They are split as family, close friend, acquaintances etc. and display hierarchy. The average number of Facebook friends one has in 2017 is 120 [2], and the number of twitter followers range from 100-520, since many of them are bots. From the figure below we can see that Facebook and Twitter rule the social nets.

The word cloud of different social media platforms [© Infographic]

Although most social networks allow only a binary state of friendship : Friend or Not-A-Friend, it is not the ideal case. There are many levels (or) strengths to the level of links established. This attribute is tough to be extracted from the geometric outline of the network graph, but is easily extracted from the activity network. The activity outlined in the graph can be described by microscopic and macroscopic observational views. The microscopic elements describe the interaction network, whereas the macroscopic elements describe the evolution of the interactions. This poster is a brief study on these interactions for a public and a private dataset. The intuition behind the analysis aligns with the idea behind social media where we do not necessary interact only with our friends, but are open to more social interactions in groups and pages carrying similar interests which are not captured in general connection networks.

## DATASET

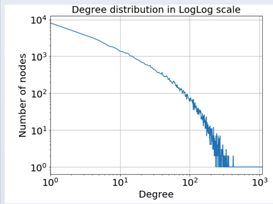
The first dataset considered is the New-Orleans Facebook dataset publically distributed at http://socialnetworks.mpi-sws.org/. It is published as a part of the paper on user interactions by Social Computing Research institute at the Max Planck Institute @ MIP_SWS [3] . It consists of 90269 users with 3646662 friendship links and over 8000000 logged interactions in form of page-likes, page-visits, messages, friend requests etc. over a period of 2 years, from 2008 to 2010. The dataset is well anonymized.

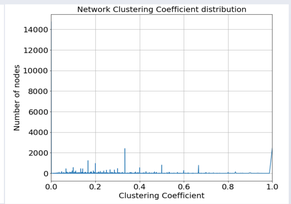
The second dataset is a private dataset of mine, that I extracted using the data archiving feature with Facebook. It consists of an history of my friends lists, my activity lists and events associated with my profile, and is used to construct my activity network.

## EXPERIMENTS & OBSERVATIONS

### 1. NEW ORLEANS Friendship links dataset

The high-level characteristics of the data, such as the degree distribution, centrality measures, clustering coefficient etc. help us understand the structure of the network, and the attributes it displays.




As we can see the network, as expected, displays a power law, which is expected on any real world connection networks. It also displays decent clustering of nodes, with a spread-out distribution.

**The average node degree : 25.64; one has 25 friends on average**

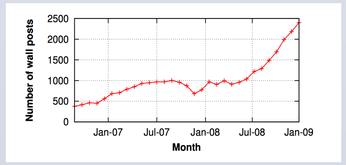
**The average clustering coefficient is : 0.221**

**The diameter is : 7**

**The average shortest distance is : 5.2**

### 2. NEW ORLEANS Activity dataset

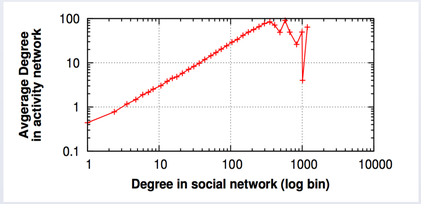
The activity dataset consists of interactions between multiple pairs of users, by means of messaging, tagging, posting on walls etc. This dataset is intuitively significant since 2 users are connected by varying degree of closeness which can be measured by their pair activities. This dataset represented 12.2% of all distinct pairs of users interactions.



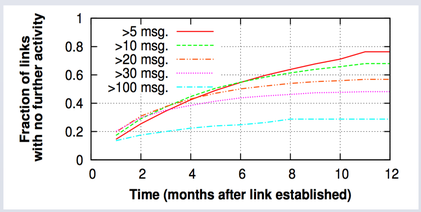

This network displays surprising observations such as :

1. Users on an average have 100 activity links with others users a month.
2. The dominant and central nodes are not shared in friendship and activity networks. The most popular friend is not the most active.
3. The fraction of total activity is clustered among different pairs of users, indicating social equilibrium.
4. The activity distribution for a distinct pair of users drops exponentially with time, excluding a given set of backbone nodes for the user. This likely explains the fact that users don't interact with their acquaintances a lot, but interact with their close friends frequently.
5. The fraction of user pairs with no pair activity like tagging etc., increases exponentially with number of months of friendship. This suggests that friends in Facebook, need not be close friends in real life.

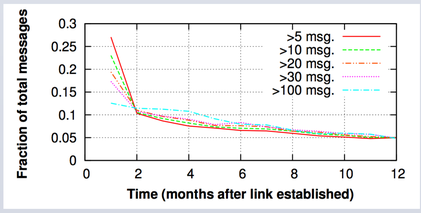

The graph on the left is the Friendship graph and the one on the right is the Activity network, color coded with same cluster indices. We can see that the most connected nodes (5170, 2332 etc.) are not the most active ones. The most active nodes 133, had a node degree of 13, which is below average.




The graph above shows how the activity of each node corresponds to the degree of each node in the network. This indicates that most people are moderately active, and their active does increase with their connections.




Plotting the links with no activity after the initial connection, shows that we tend to neglect and ignore our Facebook connections, and we don't attend to them, after accepting their requests.et




In the above graph we can see that the links / events provides boosts of activity and messaging between the entities, and it starts to fade as time goes on. It also indicates a formation of a social equilibrium, where we communicate with each level of friendships in a different manner.

### 3. My private Facebook dataset

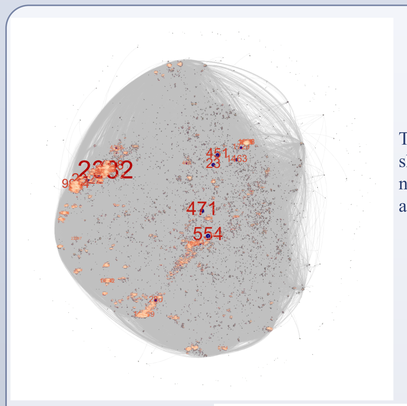
Inspired by the results I tried some experiments on my own social network, sampled with me as the center. The data was partially extracted from the Facebook's archive option, and using Netvizz.

**My node degree : 690**

**My clustering coefficient : 0.42**

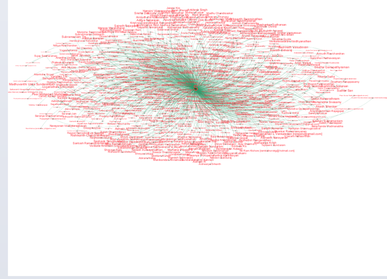
**My diameter : 6**

**My average shortest distance is : 4.6**




The image on the left shows my activity network with my node as the center.

The image on the right shows my personal friendship network labelled with my friend names ranked on the time-stamp when I connected with them.



In the image above I saw that my school friends had large labels since I connected with them early, but were farther away since I had connected with them early enough.

## CONCLUSIONS

This work tried to experiment and study if the friendship (or) connection networks actually exhibit real-life interaction properties. We studied the interaction between users in the New-Orleans Facebook dataset and my personal Facebook connection dataset. We identify that the activities are not uniformly generated, and there is a right-skew; where a minority of user-pairs generate a majority of content. Also we see that users tend to fade-off in a mutual friendship with time, where the number of interactions exponentially decreases, and the number of messages saturates. Finally we visualized my personal Facebook data and identified that my activity is heavily clustering and full of traffic. Also we saw how my high-school and grad school friends were spread out in my network.

## REFERENCES


1. Viswanath et.al, "On the Evolution of User Interaction in Facebook"
2. A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In Proc. of IMC, 2007.
3. Dataset - http://socialnetworks.mpi-sws.org/
4. H. Chun, H. Kwak, Y-H. Eom, Y-Y. Ahn, S. Moon and H. Jeong. Online Social Networks: Sheer Volume vs Social Interaction. In Proc. of IMC, 2008.
5. R. Kumar, J. Novak and A. Tomkins. Structure and Evolution of Online Social Networks. In Proc. of SIGKDD, 2006.


## ACKNOWLEDGEMENTS


I would like to thank Professor Massimo and his TAs for an excellent ECE 289 class. I am grateful for their guidance and answers. I would also like to thank UCSD for the juypterhub infrastructure and poster printing services.