Page Rank Related work:

Google’s PageRank method was developed to evaluate the importance of web-pages via their link structure. The mathematics of PageRank, however, are entirely general and apply to any graph or network in any domain. Thus, PageRank is now regularly used in bibliometrics, social and information network analysis, and for link prediction and recommendation. To do this, Google designed a system of scores called PageRank that used the link structure of the web to determine which pages are important. More generally, we can consider random surfer models on a graph with an arbitrary set of nodes instead of pages, and transition probabilities instead of randomly clicked links. Teleporting is the essential distinguishing feature of the PageRank random walk that had not appeared before in the literature. It ensures that the resulting importance scores always exist and are unique. It also makes the PageRank importance scores easy to compute. Two uses underlie the majority of PageRank applications. In the first, PageRank is used as a network centrality measure. A network centrality score yields the importance of each node in light of the entire graph structure; the goal is to use PageRank to help understand the graph better by focusing on what PageRank reveals as important. It is often compared to or contrasted with a host of other centrality or graph theoretic measures. In the second type of use, PageRank is used to illuminate a region of a large graph around a target set of interest; for this reason, we call this second use a localized measure. It is also commonly called personalized PageRank based on the discussion of personalized teleportation behaviors in the original PageRank manuscript where the random surfer teleports only to pages that are interesting to the user.

Page Rank in social networks:

PageRank serves three purposes in a social network, where the nodes are people and the edges are some type of social relationship. First, it can help solve link prediction problems to find individuals who will become friends soon. Second, it serves a classic role in evaluating the centrality of the people involved to estimate their social status and power. Third, it helps evaluate the potential influence of a node on the opinions of the network. PageRank has been used to rank individuals in the Twitter network by their importance (Java, 2007) and to help characterize properties of the Twitter social network by the PageRank values of their users.

Page rank with Diffusion:

We decided to perform a diffusion analysis and build a new model for the user behavior in a particular social network. In order to perform diffusion, we required page rank and community detection for the prediction model of diffusion.

The network we took into consideration is the Facebook network and obtained the dataset from snap datasets. The URL for the data set is as shown below:

<https://snap.stanford.edu/data/egonets-Facebook.html>

The network data set we took into consideration consist of 4039 nodes and 88234 edges. This is an unsigned network. In order to compute the page rank for the network and also the properties of the network we used R studio.

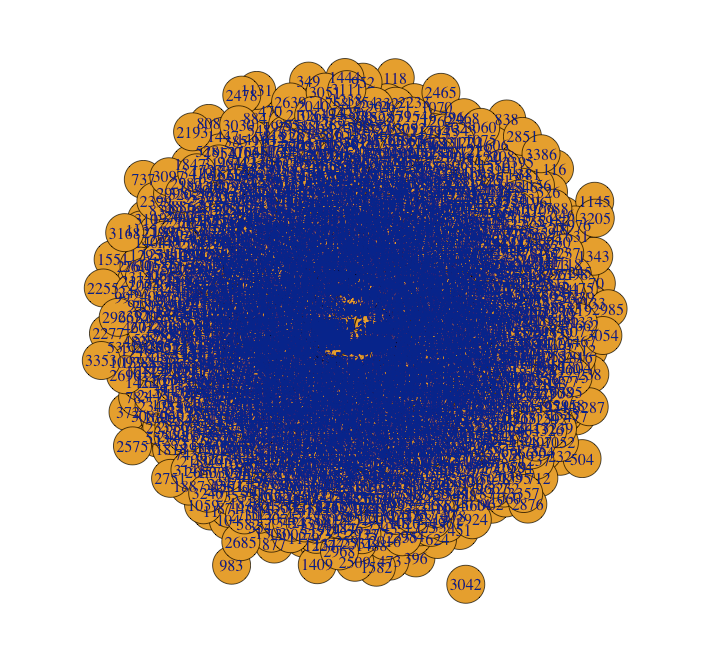


Fig: Graph of the network selected

Considering the above mentioned network, we intend to know the following properties in the network:

1. PageRank
2. Betweenness
3. Centrality
4. Max degree

Page rank:

Max(page\_rank(a1)$vector) 🡪 0.005067841

Which(Max(page\_rank(a1)$vector)) 🡪 3831

Max Degree:

Max(degree(a1)) 🡪 214

max(degree(a1)) 🡪 3831

Betweenness:

max(betweenness(a1)) 🡪 60510.5

which.max(betweenness(a1)) 🡪 3831

From the above results, we can infer that page rank for this network is mainly concentrated on the node 3831 which has the highest value for degree, page rank and betweenness. We decided to plot the graph of page rank value with the index and the graph we obtained as follows:

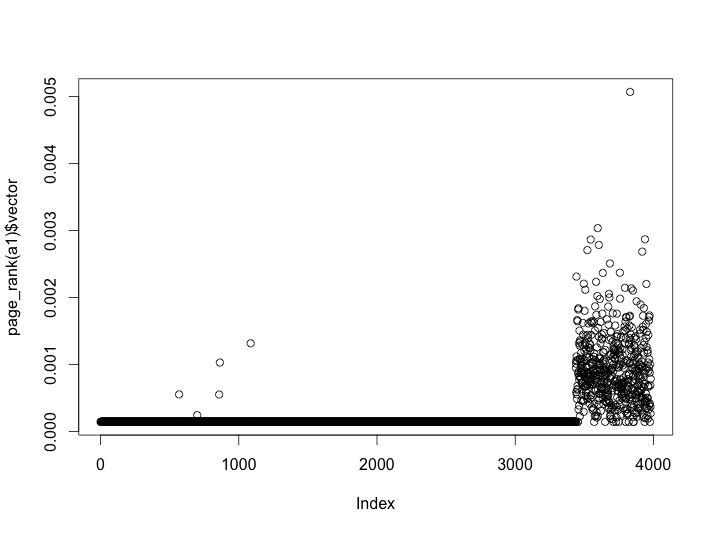


Fig: plot of page\_rank vs index

From the above plot we can see that the nodes which have the maximum page rank are more concentrated towards the right hand side suggesting that the nodes with high index number have high page rank. There is a sudden spike in the page rank value once it reaches the index value of 3500.Most of the nodes have the page rank value which is more dispersed unlike something around 0 for the nodes below 3500. We can see that the node which has the maximum page rank has the index of 3831 with a value of 0.005067.