

EE 232E Project 2

Social Network Mining

Hengjie Yang, Sheng Chang, Wendi Cui, and Tianyi Liu

May 6, 2018

1 Facebook network

1.1 Structural properties of the facebook network

The facebook network is plotted in Fig. 1.

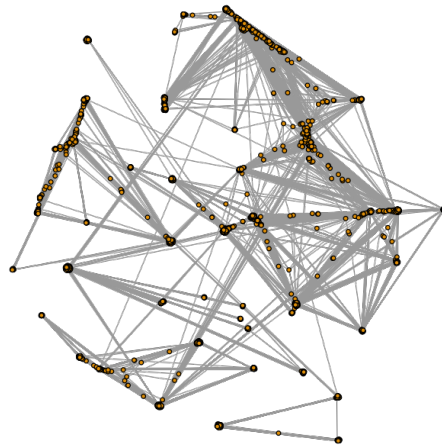


Figure 1: Facebook network

Question 1: Is the facebook network connected? If not, find the giant connected component (GCC) of the network and report the size of the GCC.

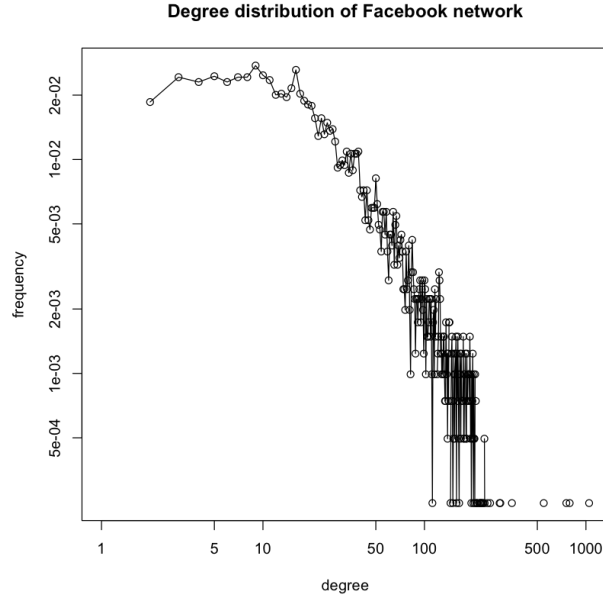


Figure 2: Degree distribution of the Facebook network

The facebook network is connected.

Question 2: Find the diameter of the network. If the network is not connected, then find the diameter of the GCC.

The diameter of the Facebook network is 8.

Question 3: Plot the degree distribution of the facebook network and report the average degree.

The degree distribution of the Facebook network is shown in Fig. 2. The average degree is 43.69101.

Question 4: Plot the degree distribution of question 3 in a log-log scale. Try to fit a line to the plot and estimate the slope of the line.

The degree distribution of Facebook network in log-log scale is shown in Fig. 3. In order to find a line which fits the data, we consider the data starting from 20-th and ends 6 before the end, thus, with linear regression analysis, the line we find is as follows.

$$y = 1.032 - 1.607x \quad (1)$$

where y represents the $\log(\text{frequency})$ and x represents the $\log(\text{degree})$. The estimated slope is -1.607 .

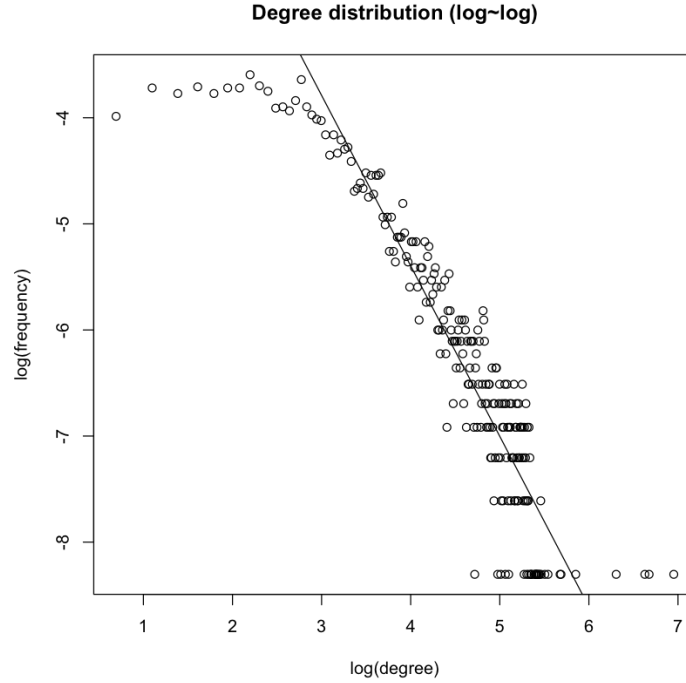


Figure 3: Degree distribution of the Facebook network in log-log scale

1.2 Personalized network

Question 5: Create a personalized network of the user whose ID is 1. How many nodes and edges does this personalized network have?

The personalized network is shown in Fig. 4. The number of nodes is 348, and the number of edges is 2866.

Question 6: What is the diameter of the personalized network? Please state a trivial upper and lower bound for the diameter of the personalized network.

The diameter of the personalized network is 2. A trivial upper bound of the diameter of the personalized network is 2 and the lower bound of the personalized network is 1.

Question 7: In the context of the personalized network, what is the meaning of the diameter of the personalized network to be equal to the upper bound you derived in question 6. What is the meaning of the diameter of the personalized network to be equal to the lower bound you derived in question 6?

The meaning is that: give the core node of the personalized network, when the number of the neighbor nodes is 1, clearly the diameter of this network is 1; If the number of the

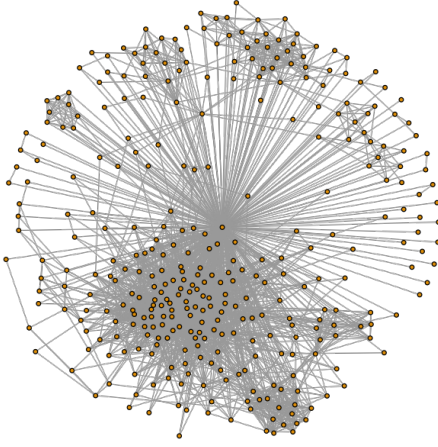


Figure 4: Personalized network of user with ID 1

neighbor nodes is greater than 1, since all neighbor nodes are connected to the core node, thus the diameter of this network is 2.

1.3 Core node's personalized network

1.3.1 Community structure of core node's personalized network

1.3.2 Community structure with the core node removed

1.3.3 Characteristic of nodes in the personalized network

1.4 Friend recommendation in personalized networks

1.4.1 Neighborhood based measure

1.4.2 Friend recommendation using neighborhood based measures

1.4.3 Creating the list of users

1.4.4 Average accuracy of friend recommendation algorithm

2 Google+ Network

2.1 Community structure of personal networks