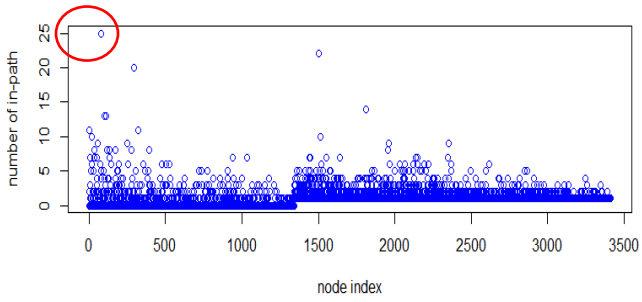
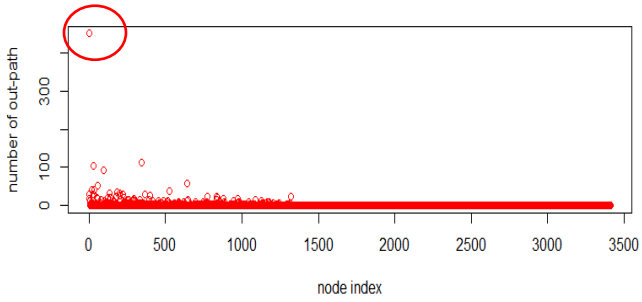
Advanced Lab 2

Ting Lan

UIN: 651457659

Overall structure: This network is a directed network since it points from the user who answer the question to those who asked. After simplifying the network of SAP Community, there are 4120 edges (there are 6090 edges before simplifying the network) and 3415 nodes in the network. There is a giant component in the network that contains 79% of the vertex (2696/3415)





Nodes and edges: Within all the nodes, node '592540' has maximal number of out-path, with other 452 nodes, that is to say user '592540' has answered questions of 452 users in the forum for the most time while there are only 2 in-path for this users since there are only 2 other users who answered the questions asked by user ' 592540'. Also, there is a node with maximal number of in-path, which is 25 answers to the questions post by this user.

Connectivity and Reciprocity: SAP community network is neither weakly connected nor strongly connected since the user does not have to answer all the questions asked by all other users. Therefore, the user is not connected to other users on the questions side if he or she didn't answer the questions. In terms of reciprocity in the directed graph, the reciprocity of the network is only 0.5% of the total number of edges, which means that the reciprocity between users are extremely low.

Diameters between nodes: Under the equal weight of 1 for each edge, the diameter of the nodes is 26 while the inverse logarithm weight gives a diameter of 14.27, which is shorter than that of the equal weighted network. The average path length is 3.98 , in this case, if the clustering is very high, the average path and diameter would not be so large. With the inverse logarithm weight, we have a higher clustering in the network

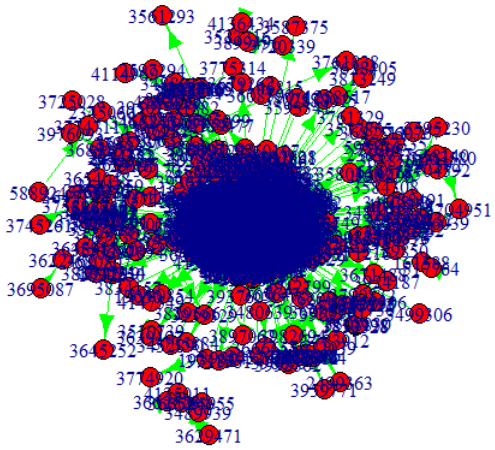
Clique census: In this clique census, this reflect that there are 3320 for cliques of size 2 , 335 for cliques of size 3, 39 cliques of size 4 and only 5 for cliques of size 5. In this case, the largest clique is 5 nodes and the smallest size of clique is 2. This means that for each of the users who asked questions on the SAP forum, at least there is one user who tried to answer their questions. For those cliques of size 5, it means that at most there are 5 users who tried to answer the questions asked by one users.

```
   2    3    4    5
3320  335   39    5
```

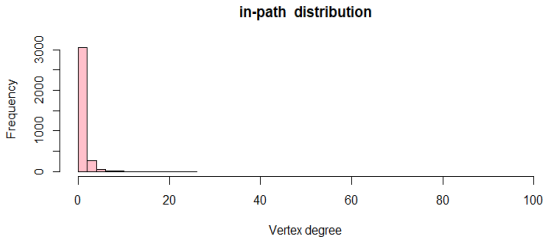
The largest cliques with size 5 are centered at **nodes 592540,22328,351048,2704623 and 3552437**

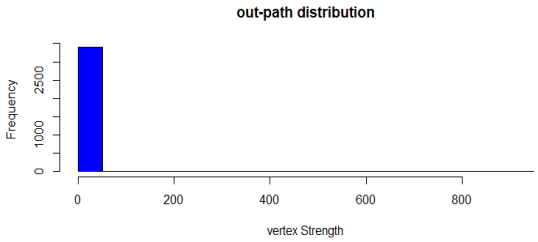
```
[[5]]
+ 5/3415 vertices, named, from 465ec54:
[1] 592540  22328   3510478 2704623 3552437
```

Plotting:

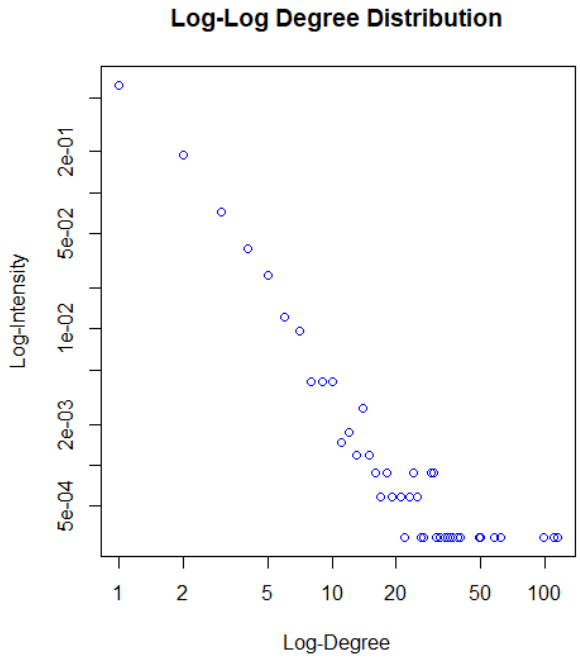


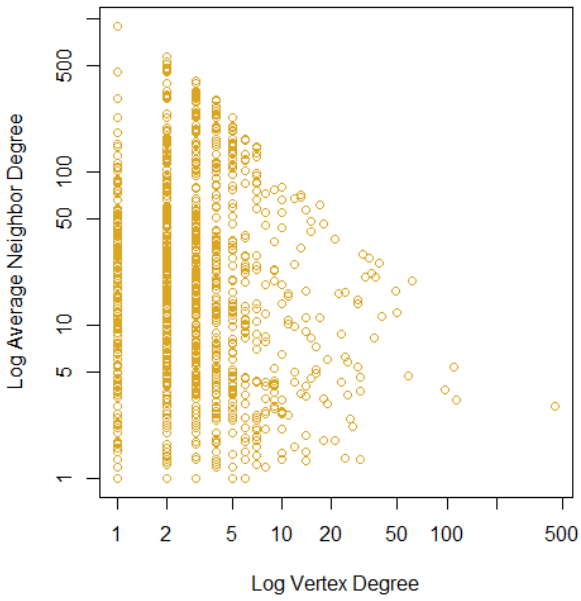
The distributions of network measures among nodes and edges



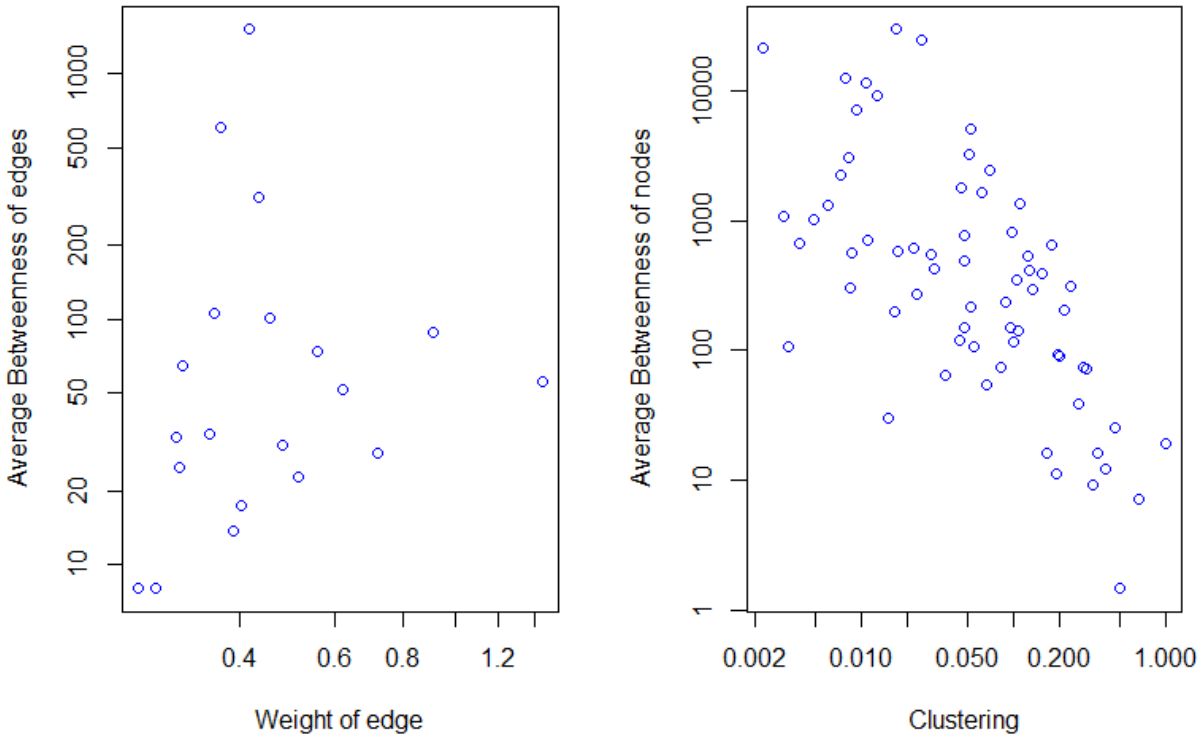


As we can see from the histogram of the degree distribution, most of the difference within the node degree distribution comes from the number of in-path. This means that the number of answers each user has got from the forum are so different while the number of answers they give to someone else are almost the same. Therefore, most of the users are asking questions in the forum while there are only a few users who answer questions post by others.





Transitivity: The global transitivity of the network is only 0.9%, which means that less than 1% of the connected triples close to form triangle. This means that it barely happens that if two users answer the same questions asked by another user, then this two users might tend to answer the question asked by the other.



From the graph, there is a roughly negative correlation between clustering and average betweenness of nodes. On the right side, the higher average betweenness f nodes means that the nodes are popular in the network because the questions asked by them is answered by many other users. However, since all the users are only focusing on the questions, there are low connection between all users who are answering the same question.