EE 232E Graphs and Network Flows

HW3

Weikun Han 804774358
Duzhi Chen 004773782
2017/5/2

Introduction

In this assignment, we are going to study the properties of a real network. The data of this real network is in the form of directed edge list. This format contains node1, node2 and the weight of the edge from node1 to node2.

Questions

1. Is this network connected? If not, find out the giant connected component (strongly connected). And in the following, we will deal with this giant connected component.

   This network is not connected. We find that the GCC contains 10487 vertices out of the total 10501 vertices of this network.
   Figure 1 is the GCC of this network we find.

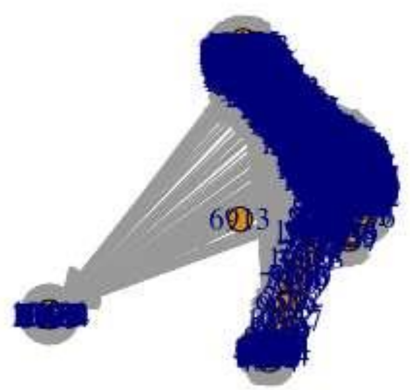

Figure 1, the GCC of the real network

2. Measure the degree distribution of in-degree and out-degree of the nodes. (plot and briefly analyze)

This graph is directed, so we calculate the in and out degree distributions separately. As the figures shown below, we can see that the in and out degree distributions of this GCC are almost the same.

Figure 2 and 3 are the in and out degree distributions of the GCC.

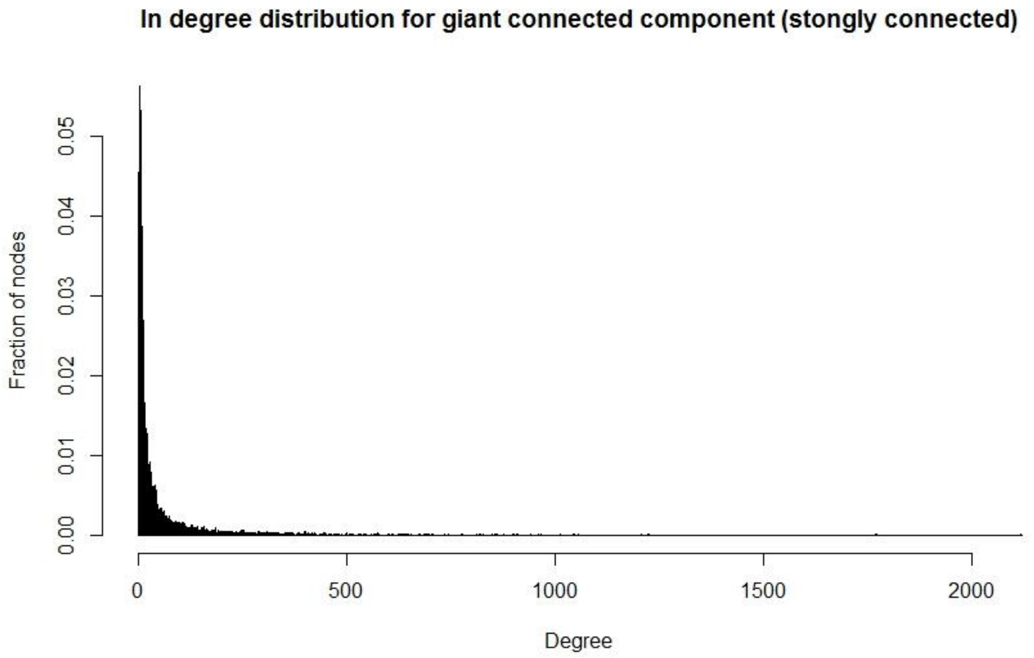

Figure 2, the out-degree distribution of the GCC

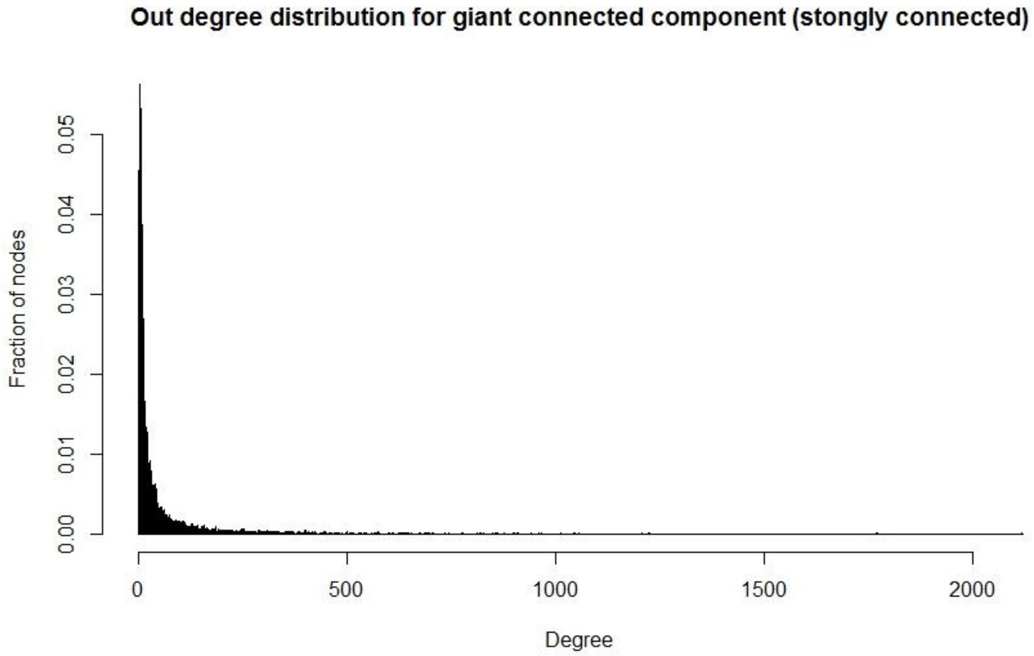

Figure 3, the out-degree distribution of the GCC

3. We would like to measure the community structure of the network.
   We convert the graph into an undirected graph.

For option 1, we use the label propagation method to calculate the community structure.

The length of the community structure is: 5

The modularity of the community structure is: 0.0001765917

Community sizes:

Index	1	2	3	4	5
Size	10471	7	3	3	3

Table 1, the community sizes for the label propagation algorithm of option 1

Figure 4 shows the community structure calculated by option 1.

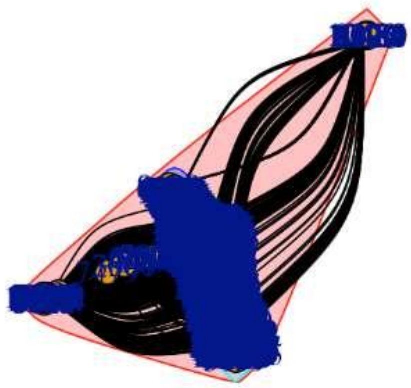

Figure 4, the community structure calculated using the label propagation

For option 2, we calculated the weight of two edges between two nodes as the square root of their product, which converted the directed graph into an undirected one. Then apply both the fast-greedy algorithm and the label propagation algorithm.

Figure 5 and 6 show the community structures computed by the fast-greedy algorithm and the label propagation algorithm separately.

For fast-greedy algorithm

The length of the community structure is: 8

The modularity of the community structure is: 0.3287988

Community sizes:

Index	1	2	3	4	5	6	7	8
Size	1856	1666	1022	2266	731	1236	633	1077

Table 2, the community sizes for fast-greedy algorithm of option 2

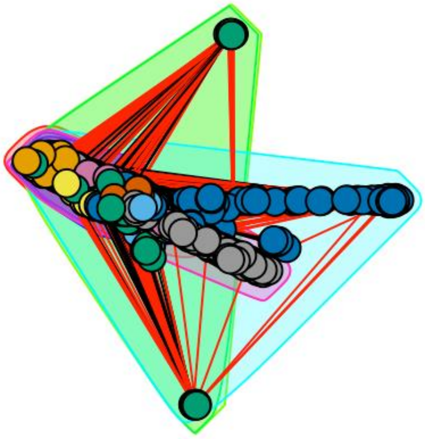

Figure 5, the community structures computed by fast-greedy of option 2

For label propagation algorithm
The length of the community structure is: 6
The modularity of the community structure is: 0.0001290506
Community sizes:

Index	1	2	3	4	5	6
Size	10467	5	6	3	3	3

Table 3, the community sizes for label propagation algorithm of option 2

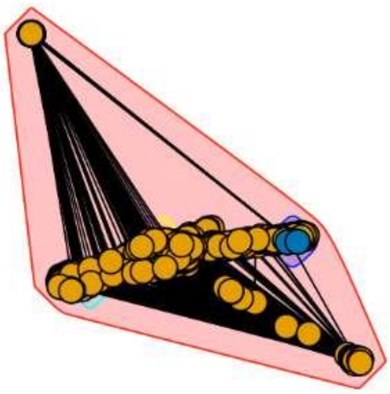

Figure 6, the community structures computed by label propagation of option 2

4. Find the largest community computed from fast-greedy-community with option 2. Isolate the community from other parts of the network to form a new network, and then find the community structure of this new network. This is the sub-community structure of the largest community.

We delete other nodes that do not belong to this community to form a new network.
Community sizes:

Index	1	2	3	4	5	6	7	8
Size	306	457	313	365	426	347	47	5

Table 4, the community sizes for fast-greedy algorithm with option 2
Figure 7 shows the community structures computed by fast-greedy algorithm with option 2

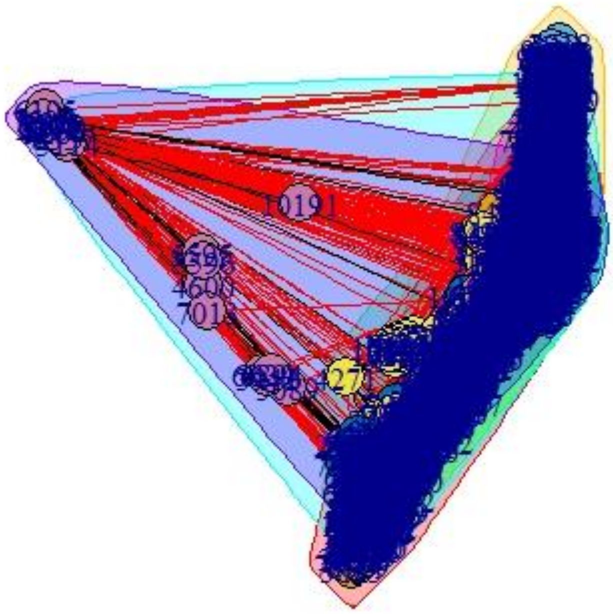

Figure 7, the community structure of the largest sub-community

5. Find all the sub-community structures of the communities (found by fast-greedy-community with option 2) whose sizes are larger than 100.

Figure 8 to 15 show all the sub-community structures whose sizes are larger than 100

Community 1 sizes:

Index	1	2	3	4	5	6	7
Size	244	452	413	490	84	136	37

Table 5, the community sizes of the sub-community structure 1

The length of the community structure is: 7

The modularity of the community structure is: 0.2249632

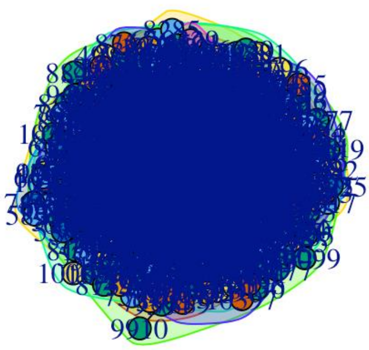

Figure 8, the community structure of sub-community 1

Community 2 sizes:

Index	1	2	3	4	5	6	7	8	9
Size	357	491	347	294	128	28	13	5	3

Table 6, the community sizes of the sub-community structure 2

The length of the community structure is: 9

The modularity of the community structure is: 0.3711271

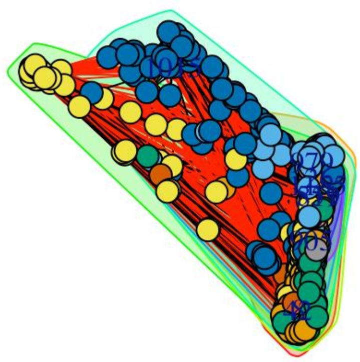

Figure 9, the community structure of sub-community 2

Community 3 sizes:

Index	1	2	3	4	5	6	7	8	9
Size	318	209	30	193	118	44	41	15	10
Index	10	11	12	13	14	15	16	17	18
Size	6	6	4	3	3	3	11	4	4

Table 7, the community sizes of the sub-community structure 3

The length of the community structure is: 18

The modularity of the community structure is: 0.5128581

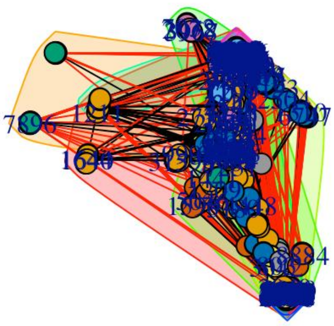

Figure 10, the community structure of sub-community 3

Community 4 sizes:

Index	1	2	3	4	5	6	7	8
Size	306	457	313	365	426	347	47	5

Table 8, the community sizes of the sub-community structure 4

The length of the community structure is: 8

The modularity of the community structure is: 0.3595153

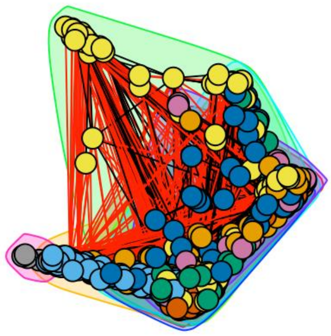

Figure 11, the community structure of sub-community 4

Community 5 sizes:

Index	1	2	3	4	5	6	7	8
Size	60	247	138	56	55	50	74	16
Index	9	10	11	12	13	14	15	
Size	11	4	4	2	7	3	4	

Table 9, the community sizes of the sub-community structure 5

The length of the community structure is: 15

The modularity of the community structure is: 0.3980826

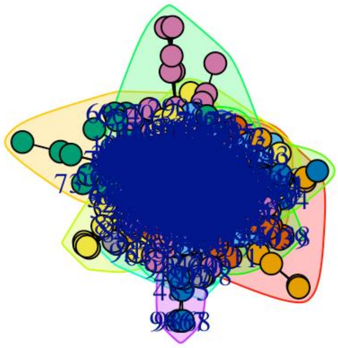

Figure 12, the community structure of sub-community 5

Community 6 sizes:

Index	1	2	3	4	5	6	7	8	9
Size	275	294	184	47	167	66	100	93	10

Table 10, the community sizes of the sub-community structure 6

The length of the community structure is: 9

The modularity of the community structure is: 0.3986121

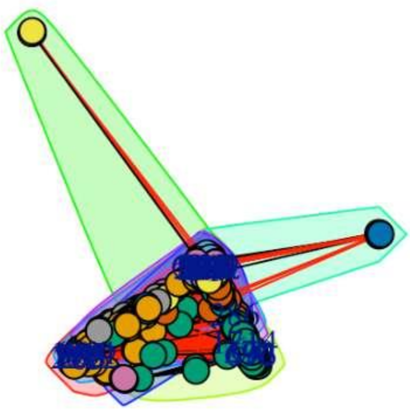

Figure 13, the community structure of sub-community 6

Community 7 sizes:

Index	1	2	3	4	5	6	7	8
Size	162	64	153	65	60	50	38	12
Index	9	10	11	12	13	14	15	16
Size	7	3	3	4	3	3	3	3

Table 11, the community sizes of the sub-community structure 7

The length of the community structure is: 16

The modularity of the community structure is: 0.4796647

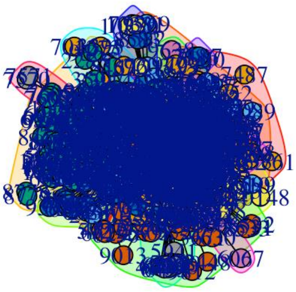

Figure 14, the community structure of sub-community 7

Community 8 sizes:

Index	1	2	3	4	5	6	7
Size	190	253	145	153	75	49	88
Index	8	9	10	11	12	13	
Size	80	16	6	11	4	7	

Table 12, the community sizes of the sub-community structure 8

The length of the community structure is: 13

The modularity of the community structure is: 0.5036454

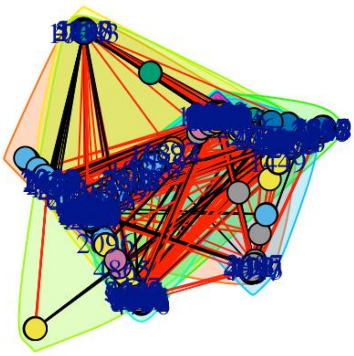

Figure 15, the community structure of sub-community 8

6. Both fast-greedy-community and label-propagation-community assume that each node belongs to only one community. In practice, a node can belong to two or more communities at the same time. There is no command in igraph that can detect overlapped communities. Here we are going to use personalized PageRank to study the overlapped communities' structures.

We use the netrw tool to solve this problem with the community data we get from problem 3. First we create the personalized connected component and find the visiting probability of each nodes of the GCC we generated at first; then we select the top 30 nodes with the highest probability and compute the value M with them. At last, we set the threshold to find nodes which are belonged to multiple communities.

For the label propagation algorithm

Table 13 below shows that the label propagation algorithm determined 24 qualified nodes with a threshold of 0.2.

4687	7979	8915	10176	10177	10178
10179	10348	10349	10350	10351	10352
10401	10402	10403	10464	10465	10466
10468	10475	10476	10477	10486	10496

Table 13, nodes belonged to multiple communities determined by the label propagation algorithm

For the fast-greedy algorithm

Table 14 shows that the fast-greedy algorithm determined 128 qualified nodes with a threshold of 0.1.

68	149	151	726	868	1406	1507	1518	2106	2266
2356	2997	3768	3769	3879	3907	3969	4040	4161	4250
4297	4312	4356	4365	4407	4583	4586	4642	4665	4993
5095	5471	5648	5850	5902	5963	5997	6005	6315	6797
6803	6818	6825	6897	6914	6919	7007	7051	7082	7158

Table 14, nodes belonged to multiple communities determined by the fast-greedy algorithm

From these two tables, we can see that the fast-greedy algorithm has a larger threshold which can determine more nodes belonged to multiple communities. This is consistent with the results from the previous problems that the fast-greedy algorithm tends to divide more sub-communities than the label propagation algorithm does. Also, we can see that there are some continuous indices of nodes in these tables such as 10176-10179, 10348-10352, 10401-10403, 10464-10466 and 10475-10477 in the Table 13. This indicates that there is a spatial continuity of vertices.