

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.

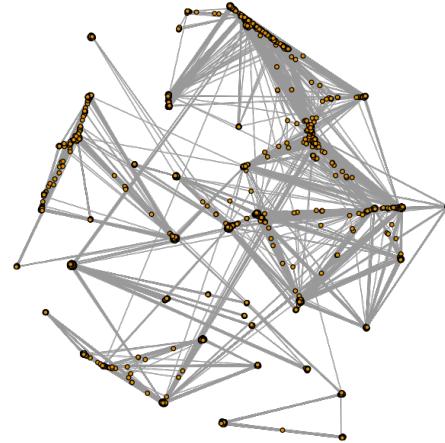


Figure 1: Facebook network

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.

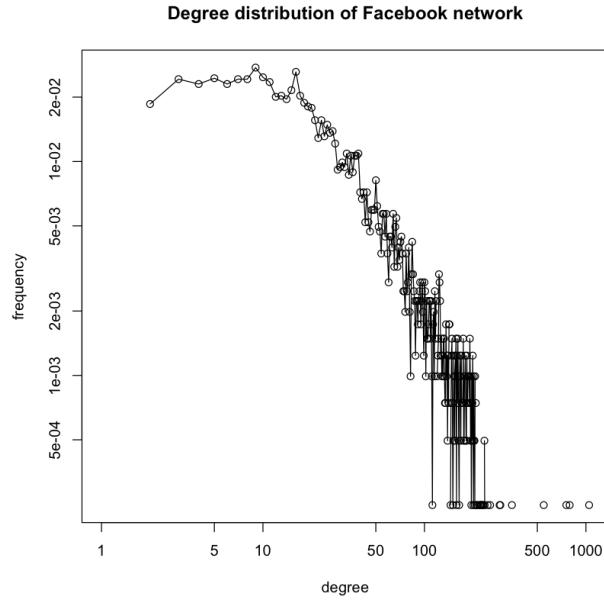


Figure 2: Degree distribution of the Facebook network

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 .

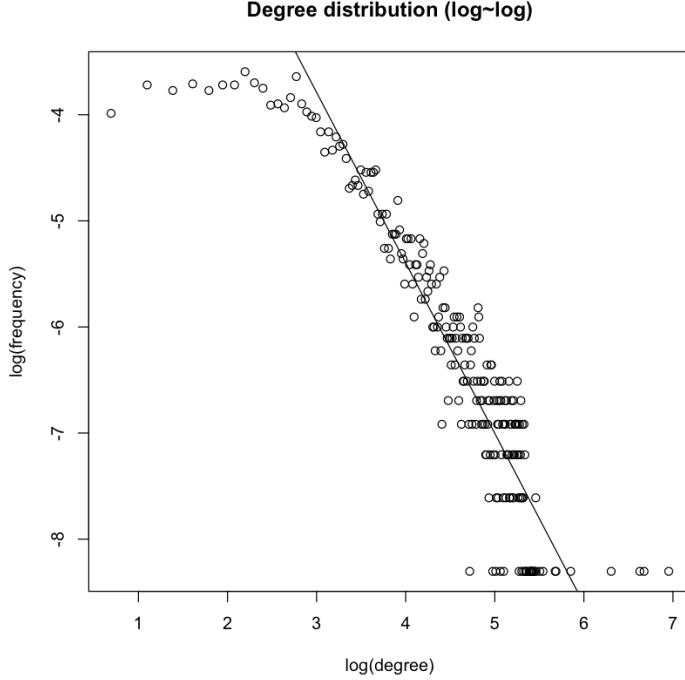


Figure 3: Degree distribution of the Facebook network in log-log scale

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 netowrk 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

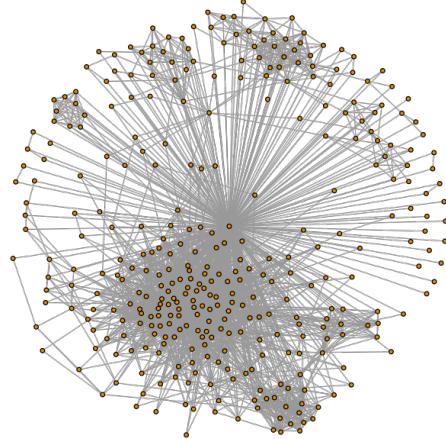


Figure 4: Personalized network of user with ID 1

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

There are 40 core nodes in the Facebook network, which is the nodes that have more than 200 neighbors i.e. the degree of the nodes is greater than 200. And the average degree of the core nodes is 279.

1.3.1 Community structure of core node's personalized network

We aim to find the community structure and compute the modularity scores using Fast-Greedy, Edge-Betweenness, and Infomap community detection algorithms for each of some core nodes' personalized network.

For Node ID 1, the community figures based on different algorithms are shown in Fig 5, Fig 6 and Fig 7.

For Node ID 108, the community figures based on different algorithms are shown in Fig 8, Fig 9 and Fig 10.

For Node ID 349, the community figures based on different algorithms are shown in Fig 11, Fig 12 and Fig 13.

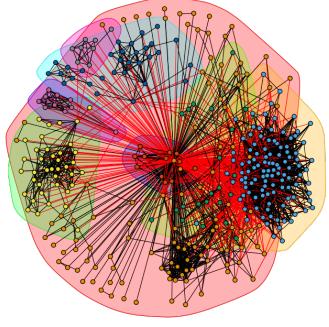


Figure 5: community structure using
Fast-Greedy algorithms

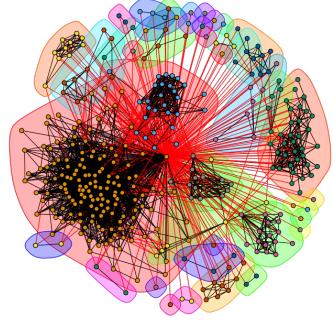


Figure 6: community structure using
Edge-Betweenness algorithms

For Node ID 484, the community figures based on different algorithms are shown in Fig 14, Fig 15 and Fig 16.

For Node ID 1087, the community figures based on different algorithms are shown in Fig 17, Fig 18 and Fig 19.

What's more, all the modularity scores from above core nodes' personalized network computed by different algorithms is shown in Table 1.

Table 1: The modularity scores for core nodes' personalized network

Node ID	Fast-Greedy	Edge-Betweenness	Infomap
1	0.41310	0.35330	0.38912
108	0.43592	-	0.50842
349	0.25171	0.13352	0.20375
484	0.50700	0.48909	0.51527
1087	0.14553	0.02762	0.02690

1.3.2 Community structure with the core node removed

In this part, we aim to explore the effect on the community structure of a core node's personalized network when the core node itself is removed from the personalized network.

For Node ID 1, the community figures based on different algorithms are shown in Fig 20, Fig 21 and Fig 22.

For Node ID 108, the community figures based on different algorithms are shown in Fig 23, Fig 24 and Fig 25.

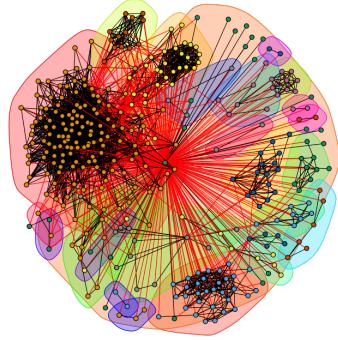


Figure 7: community structure using Infomap algorithms

For Node ID 349, the community figures based on different algorithms are shown in Fig 26, Fig 27 and Fig 28.

For Node ID 484, the community figures based on different algorithms are shown in Fig 29, Fig 30 and Fig 31.

For Node ID 1087, the community figures based on different algorithms are shown in Fig 32, Fig 33 and Fig 34.

What's more, all the modularity scores from above core nodes' personalized network computed by different algorithms is shown in Table 2.

Table 2: The modularity scores for core nodes' personalized network

Node ID	Fast-Greedy	Edge-Betweenness	Infomap
1	0.41310	0.35330	0.38912
108	0.45812	0.52132	0.52068
349	0.24569	0.15056	0.24657
484	0.53421	0.51544	0.54344
1087	0.14819	0.03249	0.02737

1.3.3 Characteristic of nodes in the personalized network

We aim to explore characteristics of nodes in the personalized network using two measures. These are two measures. One is embeddedness of a node that is defined as the number of mutual friends a node shares with the core node. Another that is dispersion of a node is defined as the sum of distances between every pair of the mutual friends the node shares

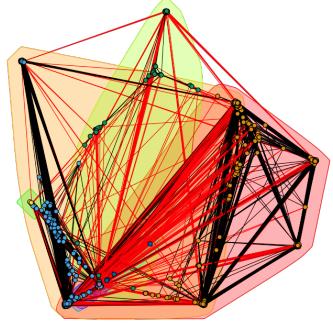


Figure 8: community structure using
Fast-Greedy algorithms

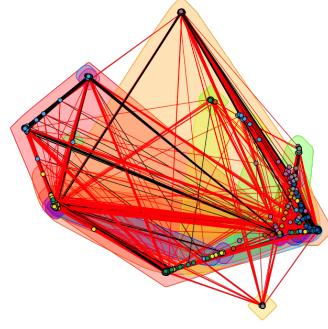


Figure 9: community structure using
Edge-Betweenness algorithms

with the core node. The distances should be calculated in a modified graph where the node (whose dispersion is being computed) and the core node are removed.

The expression relating the Embeddedness of a node to its degree is given as follows.

$$Embedd(i) \leq Degree(i) - 1 \quad (2)$$

For code code 1's personalized network, the distribution of embeddedness and dispersion is shown in Fig 35 and Fig 36.

For code code 108's personalized network, the distribution of embeddedness and dispersion is shown in Fig 37 and Fig 38.

For code code 349's personalized network, the distribution of embeddedness and dispersion is shown in Fig 39 and Fig 40.

For code code 484's personalized network, the distribution of embeddedness and dispersion is shown in Fig 41 and Fig 42.

For code code 1087's personalized network, the distribution of embeddedness and dispersion is shown in Fig 43 and Fig 44.

For each of the core node's personalized network, apply Fast-Greedy algorithm to detect the community structure of the personalized network and use colors and highlight the node with maximum dispersion and the edges incident to this node to plot this community.

For code code 1's personalized network, the community structure is shown in Fig 45. For code code 108's personalized network, the community structure is shown in Fig 46. For code code 349's personalized network, the community structure is shown in Fig 47. For code code 484's personalized network, the community structure is shown in Fig 48. For code code 1087's personalized network, the community structure is shown in Fig 49.

Similarly, this time we highlight the node with maximum embeddedness and the node with maximum ratio of dispersion to embeddedness.

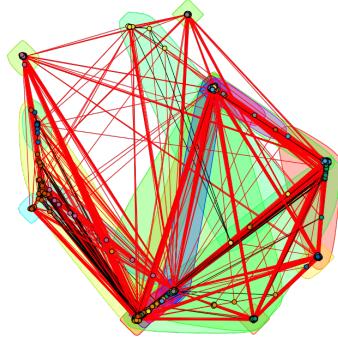


Figure 10: community structure using Infomap algorithms

For code code 1's personalized network, the community structure is shown in Fig 50. For code code 108's personalized network, the community structure is shown in Fig 51. For code code 349's personalized network, the community structure is shown in Fig 52. For code code 484's personalized network, the community structure is shown in Fig 53. For code code 1087's personalized network, the community structure is shown in Fig 54.

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

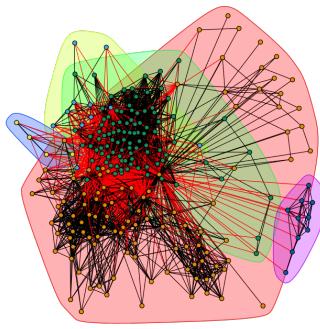


Figure 11: community structure using
Fast-Greedy algorithms

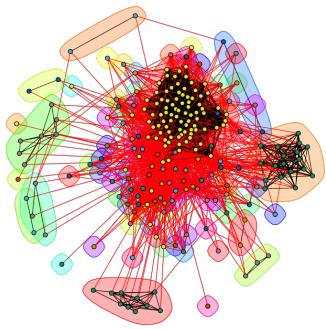


Figure 12: community structure using
Edge-Betweenness algorithms

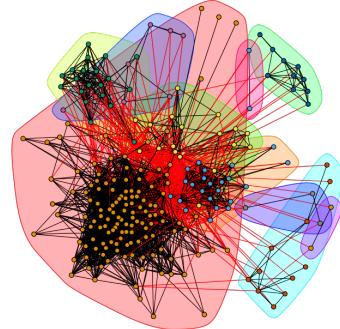


Figure 13: community structure using Infomap algorithms

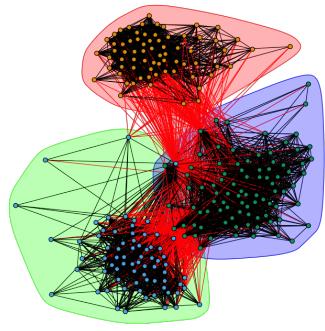


Figure 14: community structure using
Fast-Greedy algorithms

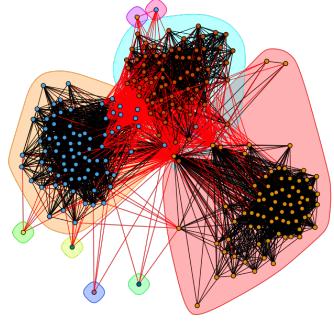


Figure 15: community structure using
Edge-Betweenness algorithms

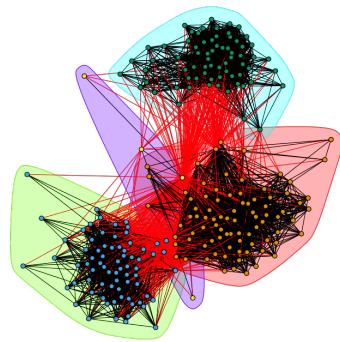


Figure 16: community structure using Infomap algorithms

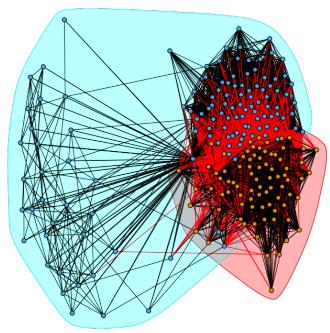


Figure 17: community structure using
Fast-Greedy algorithms

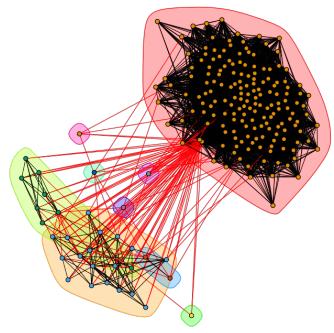


Figure 18: community structure using
Edge-Betweenness algorithms

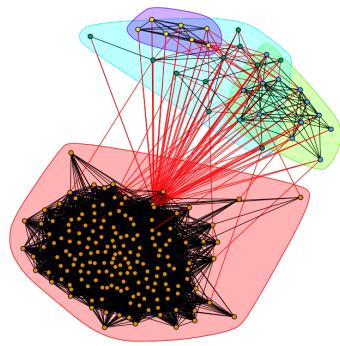


Figure 19: community structure using Infomap algorithms

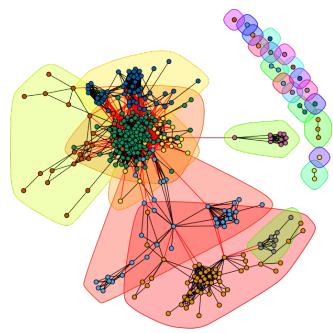


Figure 20: community structure using
Fast-Greedy algorithms

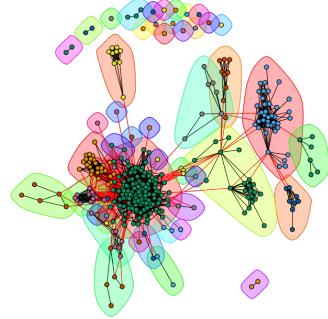


Figure 21: community structure using
Edge-Betweenness algorithms

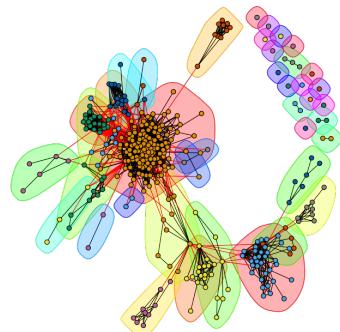


Figure 22: community structure using Infomap algorithms

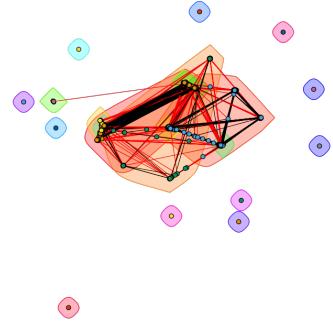


Figure 23: community structure using
Fast-Greedy algorithms

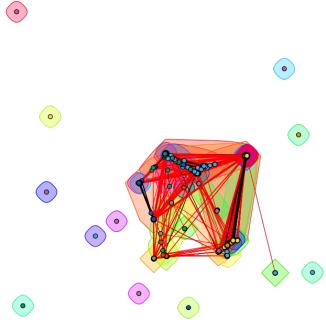


Figure 24: community structure using
Edge-Betweenness algorithms

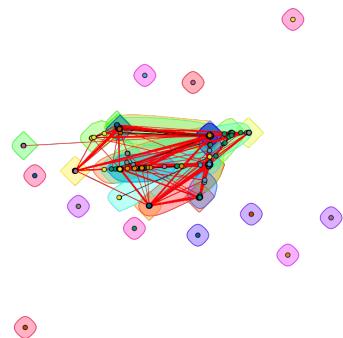


Figure 25: community structure using Infomap algorithms

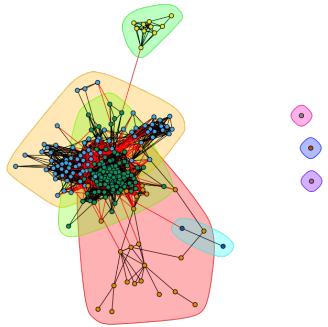


Figure 26: community structure using
Fast-Greedy algorithms

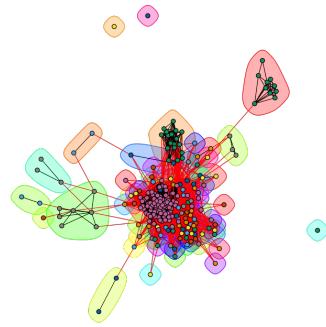


Figure 27: community structure using
Edge-Betweenness algorithms

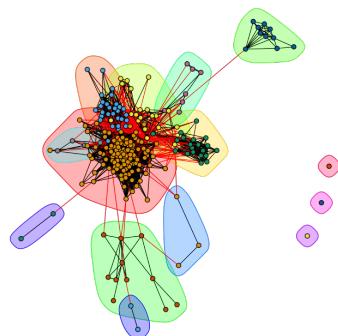


Figure 28: community structure using Infomap algorithms

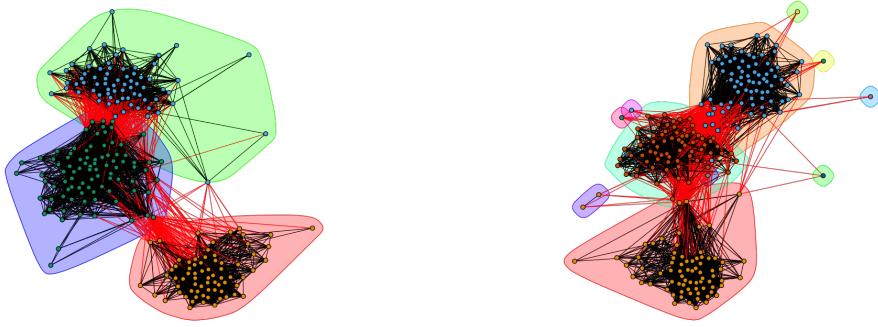


Figure 29: community structure using
Fast-Greedy algorithms

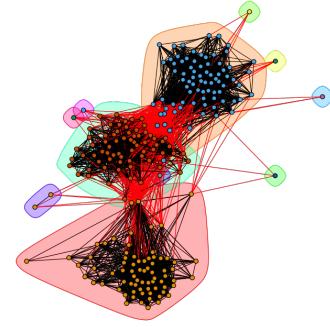


Figure 30: community structure using
Edge-Betweenness algorithms

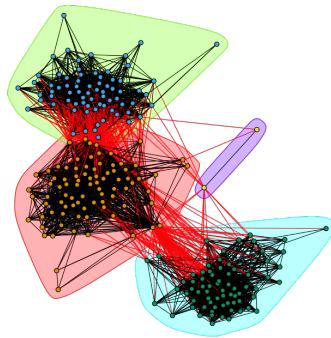


Figure 31: community structure using Infomap algorithms

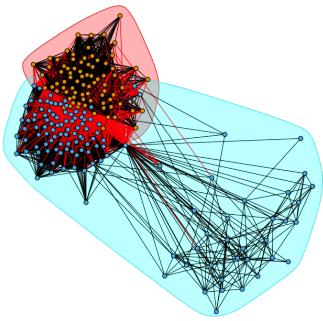


Figure 32: community structure using
Fast-Greedy algorithms

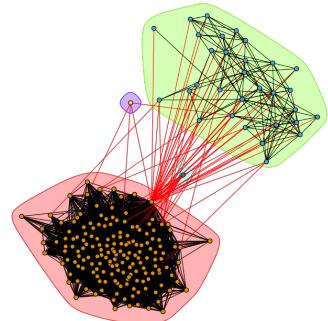


Figure 33: community structure using
Edge-Betweenness algorithms

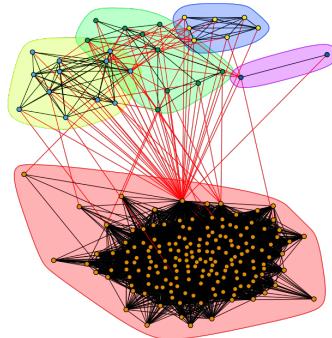


Figure 34: community structure using Infomap algorithms

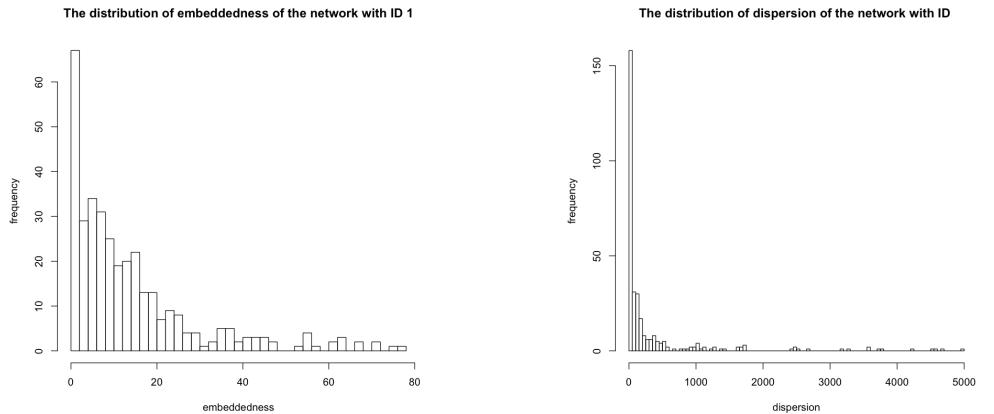


Figure 35: The distribution of embeddedness

Figure 36: The distribution of dispersion

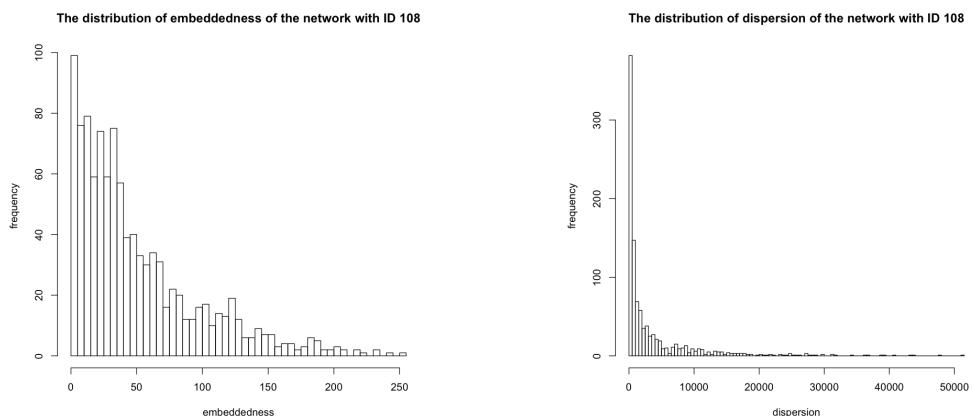


Figure 37: The distribution of embeddedness

Figure 38: The distribution of dispersion

The distribution of embeddedness of the network with ID 349

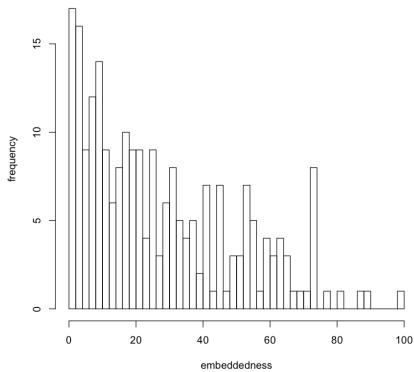


Figure 39: The distribution of embeddedness

The distribution of dispersion of the network with ID 349

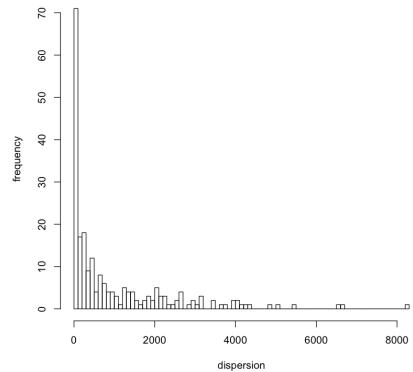


Figure 40: The distribution of dispersion

The distribution of embeddedness of the network with ID 484

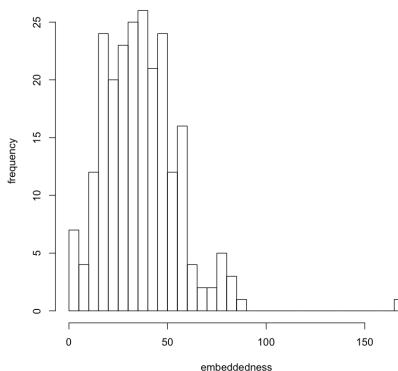


Figure 41: The distribution of embeddedness

The distribution of dispersion of the network with ID 484

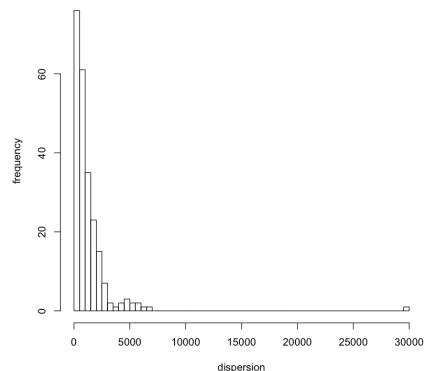


Figure 42: The distribution of dispersion

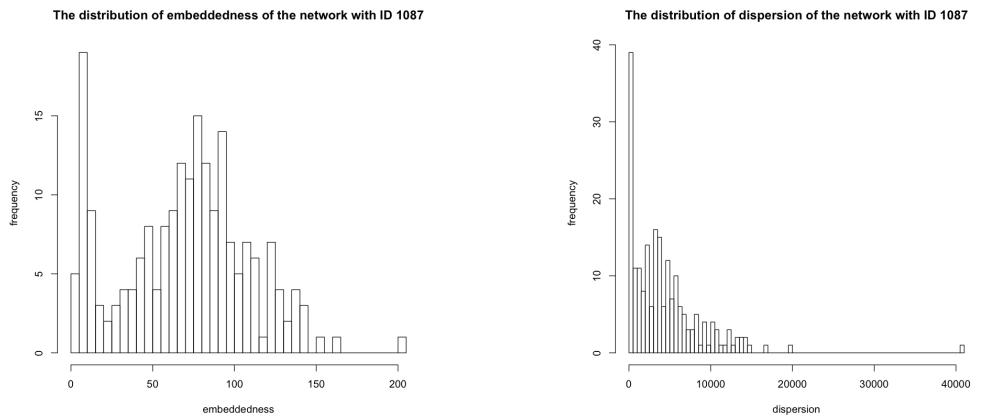


Figure 43: The distribution of embeddedness

Figure 44: The distribution of dispersion

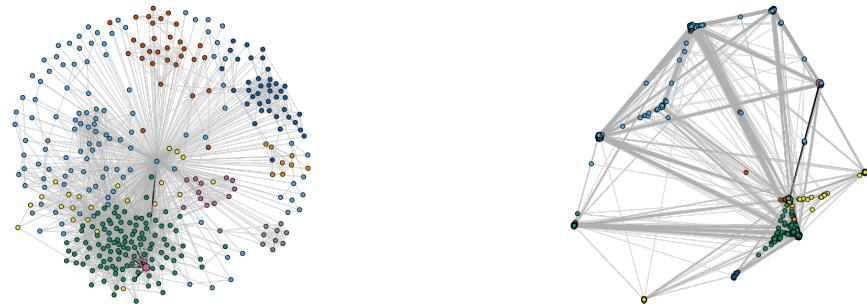


Figure 45: The community structure of No. 1

Figure 46: The community structure of No. 108

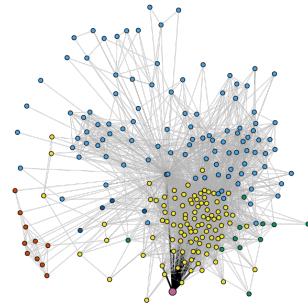


Figure 47: The community structure of
No. 349

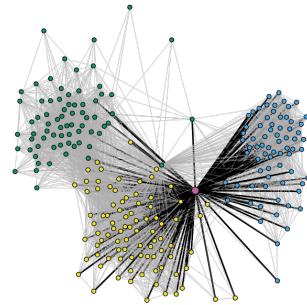


Figure 48: The community structure of
No. 484

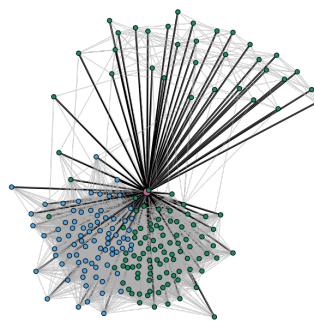


Figure 49: The community structure of No. 1087

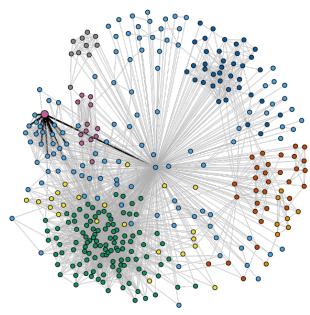


Figure 50: The community structure of
No. 1

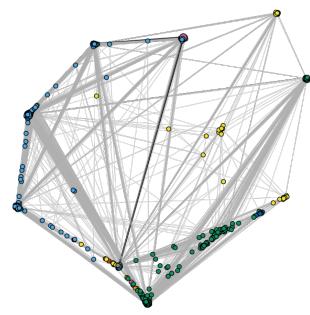


Figure 51: The community structure of
No. 108

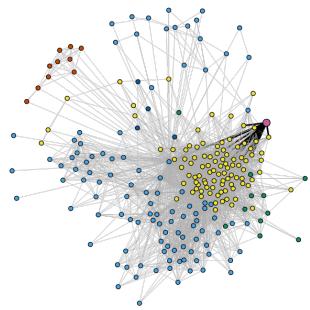


Figure 52: The community structure of
No. 349

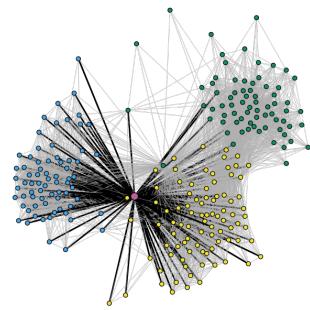


Figure 53: The community structure of
No. 484

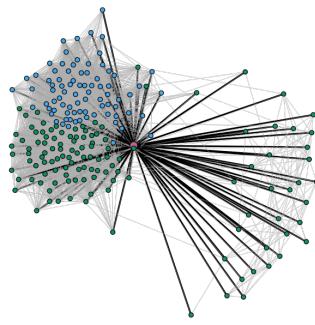


Figure 54: The community structure of No. 1087