

EE 232E
Graphs and Network Flows
Project 1
Winter 2016

Liqiang Yu, Rongjing Bai, Yunwen Zhu
904592975, 204587519, 104593417

05-14-2016

Contents

1	Problem 4	3
2	Problem 6	4
2.1	Clustering Coefficient	4
2.1.1	Global Clustering Coefficient	4
2.1.2	Local Clustering Coefficient	4
2.1.3	Network average clustering coefficient	6
2.2	Density	6
2.3	Resutls	6

1 Problem 4

After deleting the core node 1, the personal network of core node 1 is shown in figure 1 After applying Fast Greedy, Edge Betweenness and Infomap algorithms

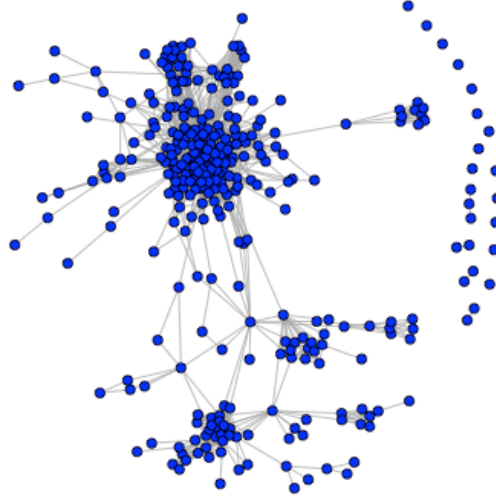


Figure 1: in degree distribution

to the personal network of core node 1 after the deletion, we can explore the community structures and compare them with the network before the deletion. The results are shown in figure 2. The left column are the community structures after applying three kinds of algorithms and the right column are the comparison results of the community distribution before and after the deletion.

From figure 2 we can see that edge betweenness and infomap method almost produced the same results, but fast greedy produced the distribution that moved right a little bit.

Moreover, the modularity results before and after the deletion are shown in table 1. From the results we can see that infomap produced exactly the same results, however fast greedy and edge betweenness both produced larger results after the deletion, which makes sense. Because the modularity is, quoted from wikipedia, "designed to measure the strength of division of a network into mod-

ules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.” After deleting the core node, which is pivot node of the its personal network, the network should be easier to divide into modules and therefore the modularity should increase.

Table 1: The modularity comparison results

	fast-greedy	edge-betweenness	infomap
Before deletion	0.4131014	0.3533022	0.4180077
After deletion	0.4418533	0.4161461	0.4180077

2 Problem 6

In this problem, we choose to use the cluster coefficient and density to define two types of communities in the 41 personal networks. We try to find the community indices that have maximal and minimal results and compare them to draw the conclusion.

2.1 Clustering Coefficient

2.1.1 Global Clustering Coefficient

The global clustering coefficient is based on triplets of nodes. A triplet consists of three connected nodes. A triangle therefore includes three closed triplets. Therefore the global clustering coefficient is the number of closed triplets (or 3 x triplets) over the total number of triplets.

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets}}$$

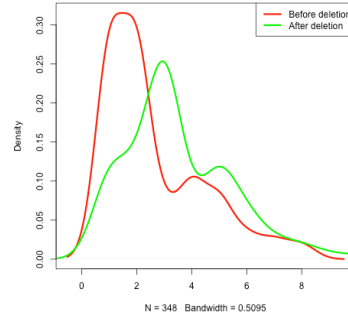
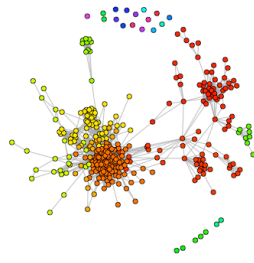
2.1.2 Local Clustering Coefficient

The local clustering coefficient of a vertex in a graph quantifies how close its neighbors are to being a clique. A graph $G = (V, E)$ formally consists of a set of vertices V and a set of edges E between them. A edge e_{ij} connects vertex v_i with v_j . The neighborhood N_i for a vertex v_i is defined as its immediately connected neighbors as follows

$$N_i = \{v_j : e_{ij} \in E \cup e_{ji} \in E\}$$

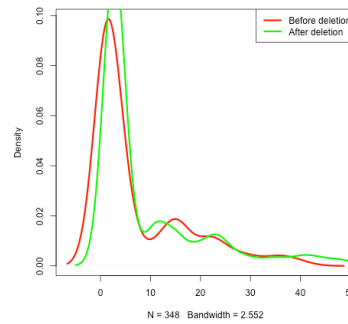
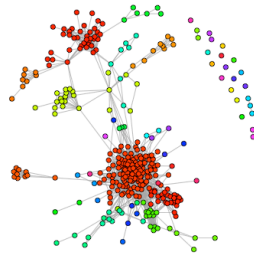
we define k_i as the number of vertices in N_i . So the local clustering coefficient for undirected graph is

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$



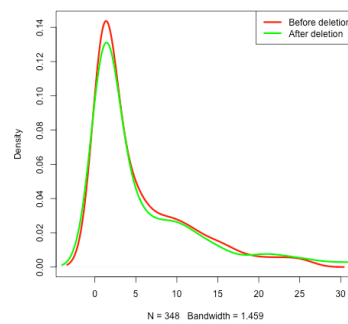
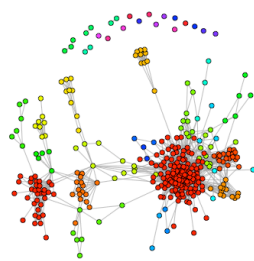
The community structure for fast-greedy method

The comparison results for fast-greedy method



The community structure for edge-betweenness method

The comparison results for edge-betweenness method



The community structure for infomap method

The comparison results for infomap method

Figure 2
5

2.1.3 Network average clustering coefficient

As an alternative to the global clustering coefficient, the overall level of clustering in a network is measured as the average of the local clustering coefficients of all the vertices n

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

2.2 Density

Network density describes the portion of the potential connections in a network that are actual connections. A potential network is a connection that could potentially exist between two nodes - regardless of whether or not it actually does. That is, for a cluster c , the density $d(c)$ is defined as

$$d(c) = \frac{2 \times E(c)}{|c|(|c| - 1)}$$

where $E(c)$ is the number of edges in cluster c and $|c|$ is the number of vertices in cluster c . High density means there are dense connections inside the cluster therefore a network with high density has large possibility to be a clique.

2.3 Results

The results are shown in table 2. We calculate both the cluster coefficient and density of each community that is larger than 10 in each personal network and store the community indices of the largest and smallest values.

Table 2: The two type communities' indices

type1 cluster	6 3 2 4 1 1 1 2 1 2 1 1 2 2 2 2 1 3 2 1 1 1 2 3 1 2 1 2 3 2 3 1 1 2 1 1 2 1 2 1 2
type1 density	6 3 2 4 1 1 1 2 1 2 1 1 2 2 2 2 1 3 2 1 1 1 2 3 1 2 1 2 3 2 3 1 1 2 1 1 2 1 2 1 2
type2 cluster	7 4 3 5 4 3 2 1 3 1 2 3 1 1 1 2 1 3 2 8 2 1 1 2 1 2 1 1 1 2 2 1 2 2 1 2 1 2 1
type2 density	7 4 3 5 4 3 2 1 3 1 2 3 1 1 1 2 1 3 2 8 2 1 1 2 1 2 1 1 1 2 2 1 2 2 1 2 1 2 1

Type 1 are indices with largest values and type 2 are indices with smallest values. From the table 2 we can see that clustering coefficient and density gave exactly the same results. So we can draw the conclusion that both clustering coefficient and density can distinguish two types of communities, one is densely connected inside, maybe a network of high school classmates and another sparselyl connected inside, maybe the fans of a sport team.