Assignment 4

Gengcong Yan - 1009903
CS-E4650 Methods of Data mining

November 28, 2021

1 Task 1

1.1 Task1.a

The graphs shown in Fig 1 and 2 represent different measures in questions, the bigger the nodes are, the higher the corresponding values are. Answers follow:

1. Node degree. Nodes [1477, 1443, 1502, 1457] have most degrees.

2. Weighted degree. Nodes [1437, 1563] have most weighted degrees.

3. Closeness centrality. Nodes [1477, 1443, 1502, 1457] have the largest closeness centrality.

4. Betweenness centrality. Nodes [1443, 1477] have the largest betweenness centrality.

5. Although the node with the most degrees has more nodes connected to them, but they are not ranked at the top because the weights between nodes are all small. when the weights are considered for calculation in weighted degree measure, Nodes [1437, 1563] are the most important nodes because some of the edges have a significant weight of 100 or more.

6. In my opinion, the nodes with most degrees are the most critical for the information flow. They are [1477, 1443, 1502, 1457]. The connection of nodes to nodes does not mean that they are close and become good friends. But because they are able to have connections with the most nodes, information spreads through them the fastest.

1.2 Task1.b

From the cluster figure in Fig 3, we can see there are 4 communities rendered in different colors, which represents different groups of students with different features. All students belong to 5A or 5B. The groups in Fig 3 shows students with same gender in the same class are more connected to others student. Moreover, you can even know which class they belongs to from the figure. **The green group and blue group represent males and females in class 5B, respectively. The purple group and orange group represent males and females in class 5A, respectively.**

1.3 Task1.c

1. I visualized all of graphs in the questions above via Gephi, the images are in the corresponding sections.

2. The gender of person 1637 in the class 5A are unkown. Through the observation in the communities graph above, I think 1637 are a girl in class 5A.

3. We filter the graph based on degree and keep those with most. From the updated graph in Fig 4, I guess the remaining person in the graph may have taken on certain administrative duties in the class. They interacted more with the students in their
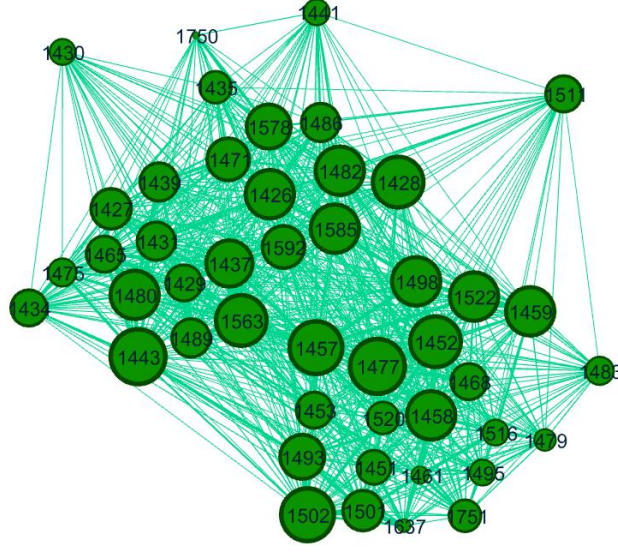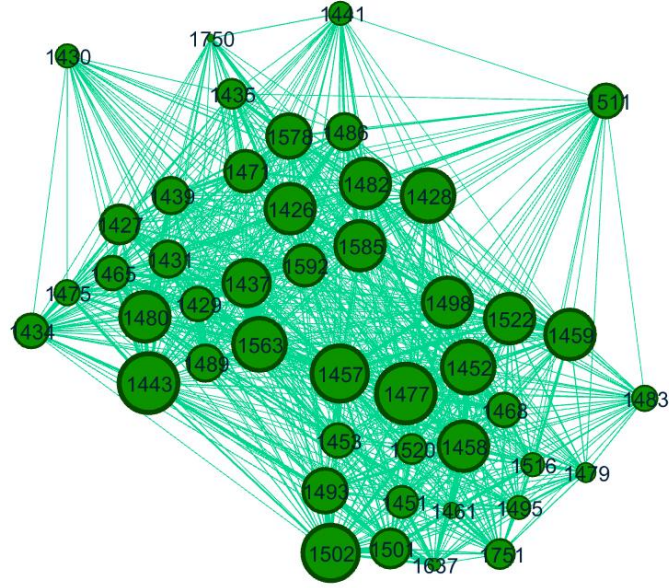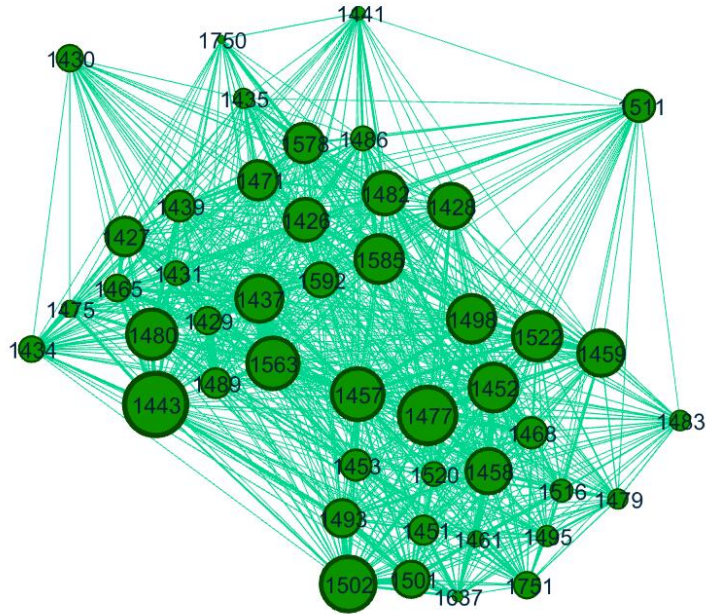
(a) Node degree
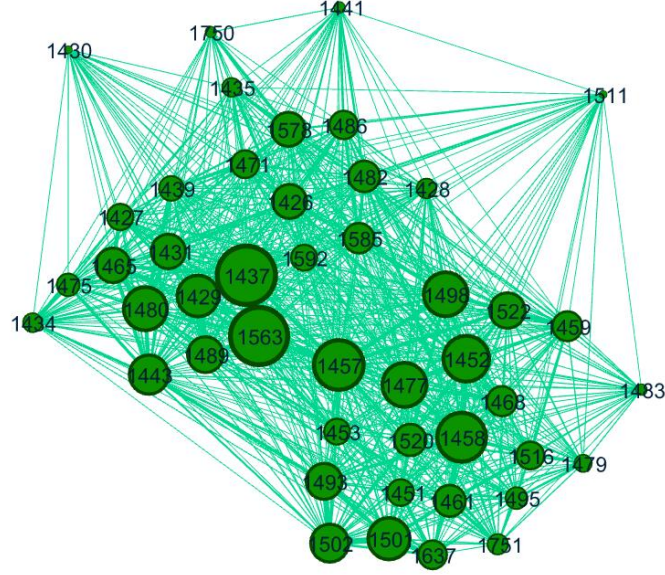
(b) Weighted Degree

Figure 1: Graph of classmates relationship

own class leading them to have more degrees. And they may also often communicate with each other about class work, interact more with other class administrators.

(a) Closeness centrality

(b) Betweenness centrality
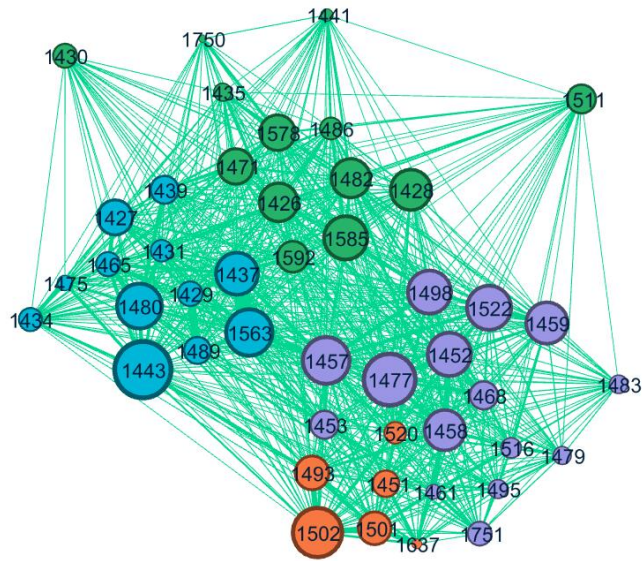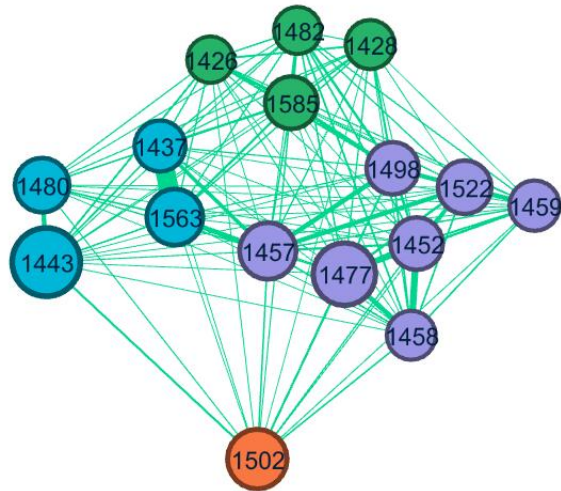
Figure 2: Graph of classmates relationship

Figure 3: Communities in Graph

Figure 4: Filtered graph
(Degree between (39,43))