

CS-E4650 - Assignment 4

Arttu Häkkinen - 596077

November 26, 2021

Task 1

a)

i) Nodes with the highest node degree are: 1477 with node degree of 43 and 1443 with node degree of 43. High node degree means high number of neighbors.

ii) Nodes with the highest weighted degree are: 1437 with weighted degree of 221.0 and 1563 with weighted degree of 216.8. High weighted degree means high sum of all edge weights.

iii) Nodes with the highest closeness centrality are: 1477 with closeness centrality of about 0.957 and 1443 with closeness centrality of about 0.957. High closeness centrality means that the average distance to other nodes is small.

iv) Nodes with the highest betweenness centrality are: 1443 with betweenness centrality of about 10.27 and 1477 with betweenness centrality of about 9.289. High betweenness centrality means that the node appears on many of the shortest paths between other nodes.

v) The top-two nodes with highest weighted degree (1437 and 1563) are not among the top degree nodes because the edge between these two nodes has such high edge weight of about 101.8. This causes the weighted degree to have really high value for these two nodes but the normal node degree is still not in the top.

vi) The most critical nodes for the information flow are the nodes 1477 and 1443.

b)

Running the modularity function finds in total 4 communities, assigning each node into one of these communities labeled from 0 to 3.

Nodes that belong to the community

- 0 are {1459, 1516, 1452, 1498, 1468, 1522, 1477, 1483, 1457, 1479, 1495, 1458, 1461, 1453, 1751}.
- 1 are {1520, 1502, 1451, 1493, 1501, 1637}.
- 2 are {1486, 1428, 1471, 1592, 1482, 1435, 1585, 1578, 1426, 1441, 1511, 1430, 1750}.
- 3 are {1439, 1563, 1489, 1465, 1431, 1480, 1434, 1437, 1443, 1427, 1429, 1475}.

Let's try to interpret these communities based on the metadata of the nodes. Use the attached python script to combine the metadata from 'schoolclass5meta.txt' with the resulting data table after running the modularity function. This way we can easily see whether the found communities by the function make any sense based on the background variables.

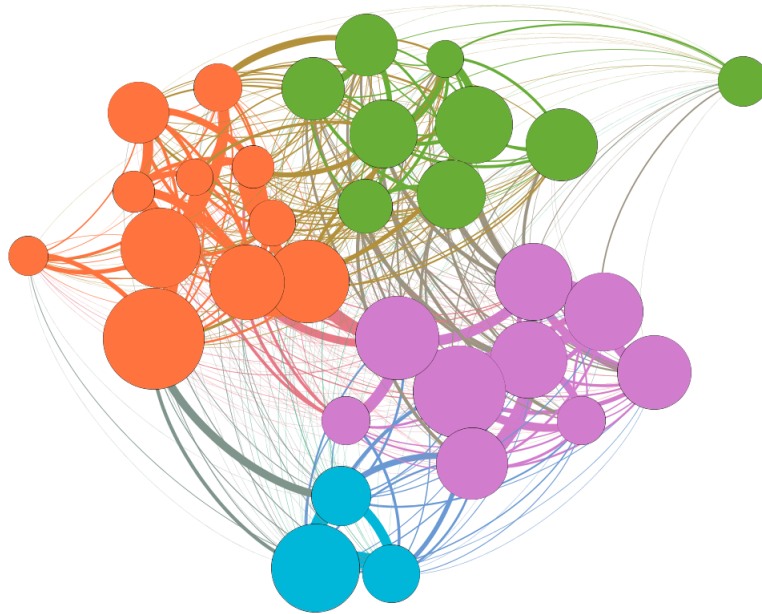
Let's first focus on the largest community 0. It seems to consist of 15 members that are all from class 5A. The members in this community are males and females, so being male or female doesn't have an effect to this community.

The community 1 consists of 6 students. They all are also from class 5A, and 5 of them are females. The last member of the group has undefined gender.

The community 2 is clearly all boys except one from the class 5B. All members are from class 5B and they are male.

The community 3 is formed by the one boy from 5B that didn't belong to community 2 and all girls from class 5B.

c)



As my experiment with gephi, I tried some of the visualization techniques presented in the gephi's quick start guide [1]. First I changed the layout of

the graph using Force Atlas from the layout module and changed the repulsion strength to 10000 to expand the graph. Then I adjusted the node sizes of the graph based on the values of the nodes' betweenness centrality ranging from 10 to 100. This had to been implemented in action by runnning the layout module again and checking the boc "Adjust by Sizes". Now the nodes didn't overlap anymore and I think this is really crucial for clear visualization. Next, I added community detection, so that the calculated modularity classes / communities would be more clear also visually. This was done by selecting from node partition module the modularity class and pressing the apply button. This way it is nice that the different community nodes and edges are colored. It makes the graph lot more visually clear. Finally, I added a filter so that nodes with less than 34 edges adjacent to them would not be shown in the graph. This makes it easier to get the big picture without too much detailed information visually from the graph. The resulting image can be found above:

References

- [1] <https://gephi.org/users/quick-start/>