CS529 - Assignment 4

Kutay Taşcı

22101359

Overview

In this assignment we are going to experiment with clustering on networks. For this assignment we are going to use Les Miserables dataset first, then we are going to use the dataset we used in term project. For clustering, we used Girvan-Newman (Betweenness-based) and Louvain (Modularity-based) algorithms. With using Gephi we are going to perform the experiments given in assignment paper.

Assignment consists of two experiments. In the first experiment we are going to use Les Miserables dataset for testing grouping algorithms. For each algorithm we are going to perform our experiments and given visualizations. In the second part we will discuss about structural equivalence and cohesive groups. Then, we will further discuss this topic for our term projects data.

Exercise-1

In this exercise we are going to perform grouping on Les Miserables dataset.

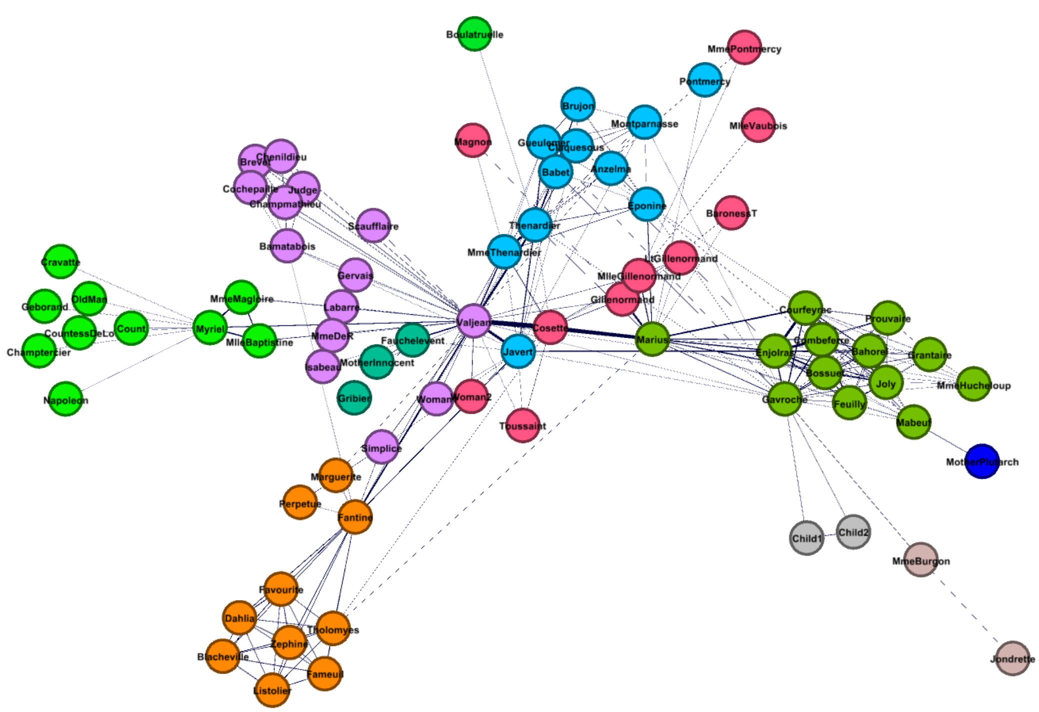
A) In this part you can find the list of nodes matched with the group ID's. You can find the nodes and groups id's for corresponding algorithms bellow.

Id	Label	Girvan-Newman	Louvain
0	Myriel	0	0
1	Napoleon	0	0
2	MlleBaptistine	0	0
3	MmeMagloire	0	0
4	CountessDeLo	0	0
5	Geborand	0	0
6	Champtercier	0	0
7	Cravatte	0	0
8	Count	0	0
9	OldMan	0	0
10	Labarre	4	3
11	Valjean	4	3
12	Marguerite	1	1
13	MmeDeR	4	3
14	Isabeau	4	3
15	Gervais	4	3
16	Tholomyes	1	1
17	Listolier	1	1

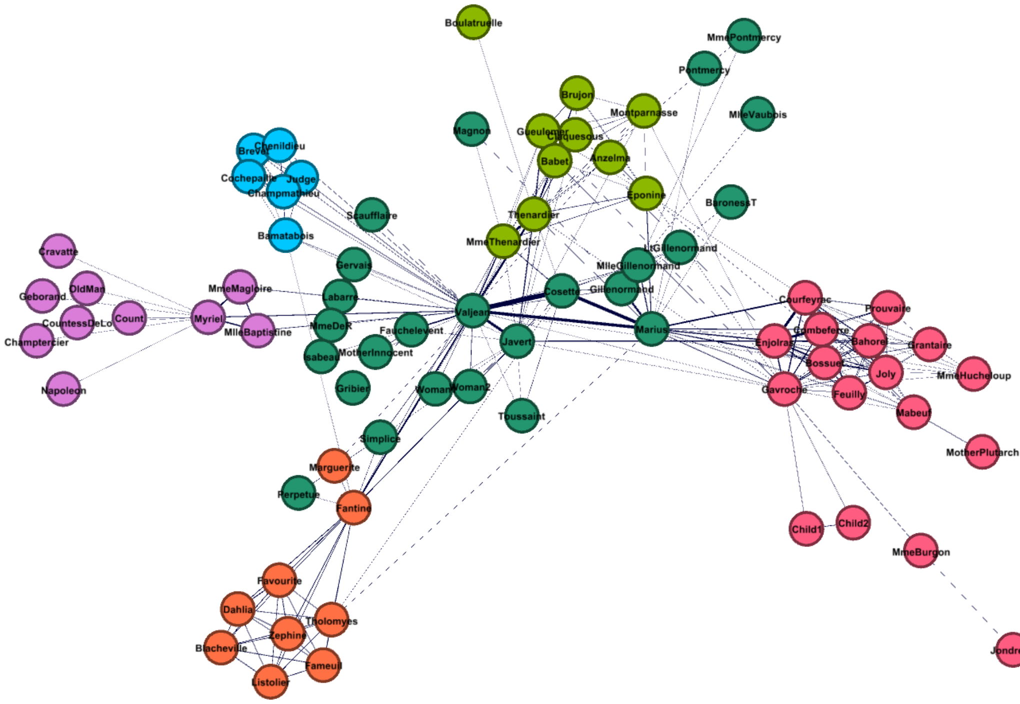
18	Fameuil	1	1
19	Blacheville	1	1
20	Favourite	1	1
21	Dahlia	1	1
22	Zephine	1	1
23	Fantine	1	1
24	MmeThenardier	10	4
25	Thenardier	10	4
26	Cosette	5	3
27	Javert	10	3
28	Fauchelevent	6	3
29	Bamatabois	4	2
30	Perpetue	1	3
31	Simplice	4	3
32	Scaufflaire	4	3
33	Woman1	4	3
34	Judge	4	2
35	Champmathieu	4	2
36	Brevet	4	2
37	Chenildieu	4	2
38	Cochepaille	4	2
39	Pontmercy	10	3
40	Boulatruelle	7	4
41	Eponine	10	4
42	Anzelma	10	4
43	Woman2	5	3
44	MotherInnocent	6	3
45	Gribier	6	3
46	Jondrette	2	5
47	MmeBurgon	2	5
48	Gavroche	3	5
49	Gillenormand	5	3
50	Magnon	5	3
51	MlleGillenormand	5	3
52	MmePontmercy	5	3
53	MlleVaubois	5	3
54	LtGillenormand	5	3
55	Marius	3	3
56	BaronessT	5	3
57	Mabeuf	3	5
58	Enjolras	3	5

59	Combeferre	3	5
60	Prouvaire	3	5
61	Feuilly	3	5
62	Courfeyrac	3	5
63	Bahorel	3	5
64	Bossuet	3	5
65	Joly	3	5
66	Grantaire	3	5
67	MotherPlutarch	8	5
68	Gueulemer	10	4
69	Babet	10	4
70	Claquesous	10	4
71	Montparnasse	10	4
72	Toussaint	5	3
73	Child1	9	5
74	Child2	9	5
75	Brujon	10	4
76	MmeHucheloup	3	5

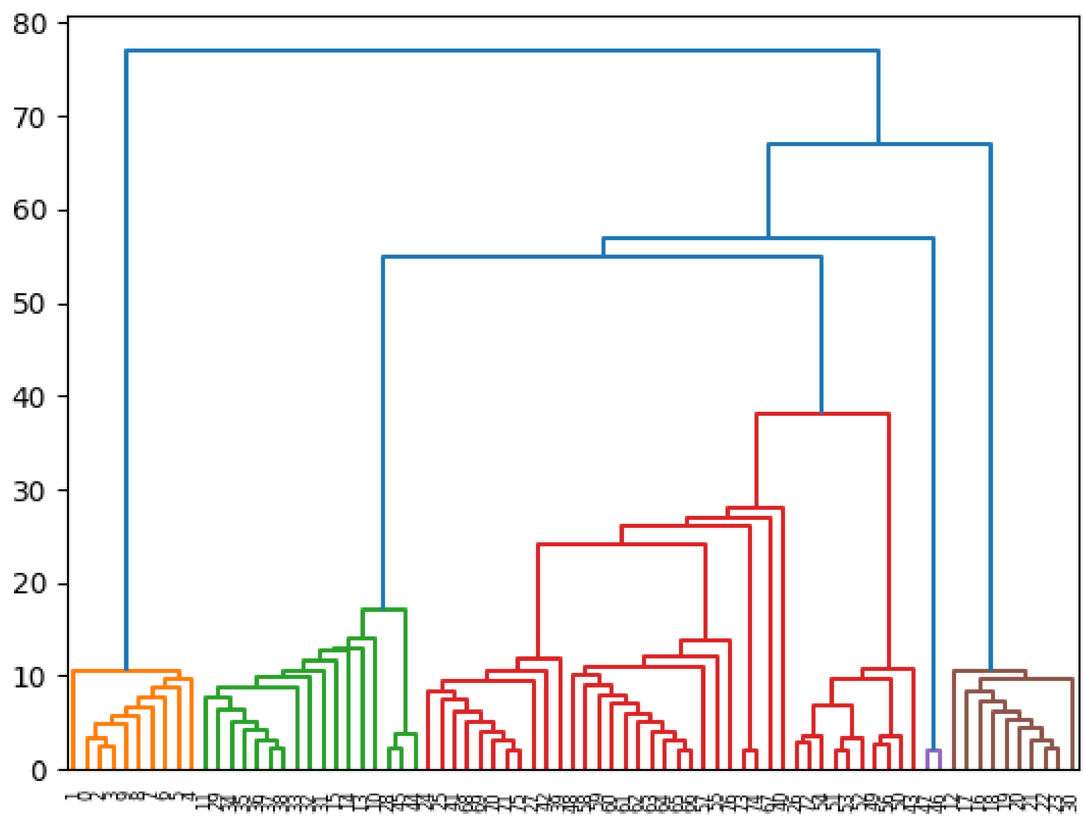
B) Visualization of the network. Colored by Girvan-Newman grouping.



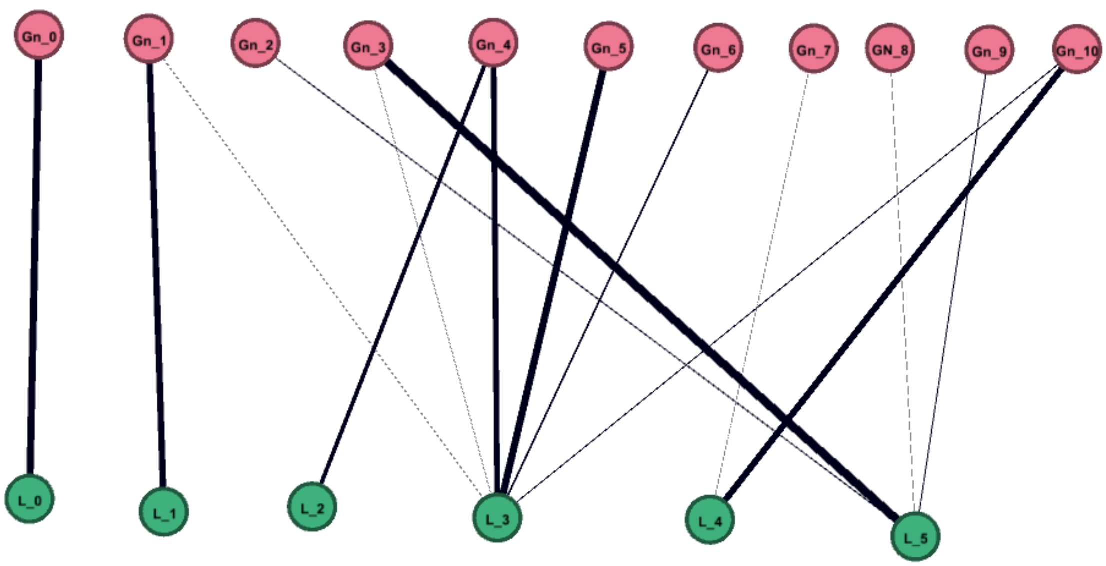
C) Visualization of the network where nodes are colored by Louvain grouping.



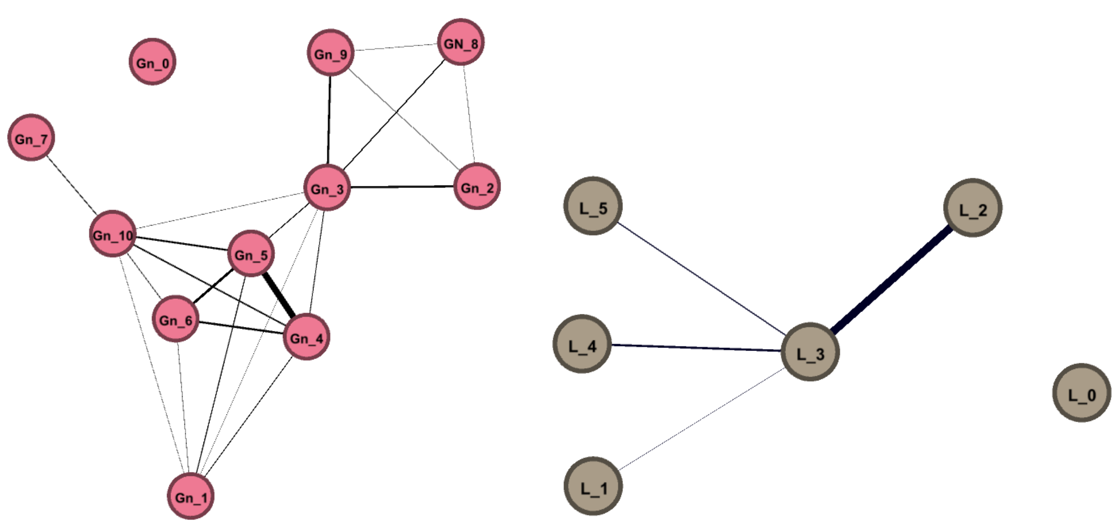
D) For this part, we have generated a dendrogram for Girvan-Newman grouping. Since generating a dendrogram are not implemented in Gephi, I have used python to visualize Girvan-Newman Partitioning.



E) In this part we are going to analyze the relationship between the groups found by the algorithms. For this task we constructed a 2-mode network. Nodes of this network corresponds to groups created by different algorithms. Edges on the other hand represents the common people in these groups, weights of the edges are the amount of common people. GN_# corresponds to groups created by Girvan-Newman algorithm and L_# corresponds to groups created by Louvain algorithm.



From this 2-mode network we are going to derive two one-mode networks from this bipartite network. Nodes of this one node networks are groups clustered by algorithms. Edges will correspond to intersections of these groping. For example GN_1 cluster has edges both on L_1 and L_3, this will generate an edge between L_1 and L_3 in generated one-mode network.



Above you can see the interaction between groups clustered by different algorithms. As you can see L_0 and GN_0 is completely same with each other. Other clusters interact with each other, because of the intersections in the boundary nodes. Louvain network is less complex because number of clusters are less. We can see the relation between L_0 and GN_0 clearly. Also L_3 is we can see from this graph that L_3 is located in the center of the network. Because it interacts with most of the different clusters.

Exercise-2-A

In this exercise we are going to explain the difference between structural equivalence and structural cohesion (cohesive groups.) We are first going to explain these terms. Then we are going to compare them and finally analyze the situations where equivalent groups be more meaningful than cohesive groups.

The most common definition of structural equivalence is based on the same-neighborhood or social niche idea. Given an undirected graph G(V,E), a pair of vertices u and v are structurally equivalent if N(u) = N(v), where N(x) is the set of vertices adjacent to x. In other words, u and v are connected to the same vertices. For directed graphs, we require that Ni(u) = Ni(v) and No(u) = No(v), meaning that u and v have incoming arcs from the same vertices and outgoing arcs to the same vertices. (Sailer, 1978)

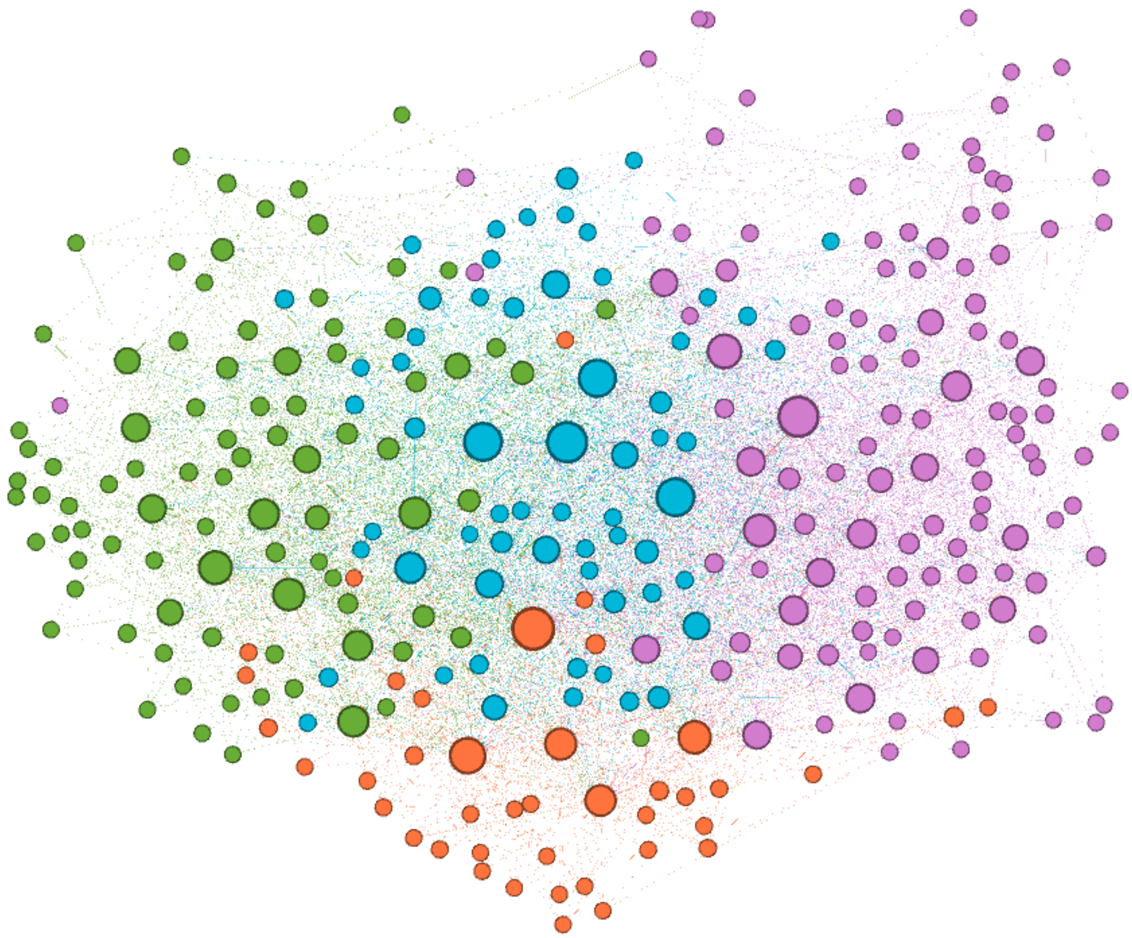
Structural cohesion is the measure of cohesion in social groups. It is defined as the minimal number of actors in a social network that need to be removed to disconnect the group (James Moody, 2015). By this definition we can say that k-cliques are more cohesive than k-node rings. Because we can disconnect components of the ring structure by removing fewer edges than a k-clique. A ring can be disconnected with removing 2 edges, on the other hand for disconnecting k-clique we need to remove k-1 edges.

We can explain the difference between two groups in two perspectives. Firstly, nodes in structurally equivalent groups performs same behaviors since their interactions are overlapping. This leads to indistinguishableness, where we can't tell the difference between nodes by looking at their links. This is not the case for cohesive groups, where nodes in a group can have different interactions with overall network. Secondly, cohesive groups have dense connections with other groups members, these dense connection leads to high connectivity. But structurally equivalent groups can have sparse connections within group members.

By the light of our previous discussions, we can say that structural cohesion is more effective on human-human interactions in social groups. Since, they are better at representing tight connections in social groups and intergroup connections. On the other hand, structural equivalence is more fit to problems where connections between nodes has similar interactions. Like hierarchical networks and information networks. In these networks we can group actors based on their in/out links with other actors in different levels.

Exercise-2-A

In this part we are going to use a sample of the data we will use for our term project, for discussing cohesive groups and structural equality. The dataset we are using in our term project is a social media follower/following dataset for a student club, gathered from Instagram. Bellow you can find a visualization for our network and some metrics to understand our data.



Average Degree: 26.5

Network Diameter: 6

Graph Density: 0.086

Modularity: 0.227

Avg. Clustering Coefficient: 0.615

Avg. Path Length: 2.13

By looking at modularity we can say that groups in this network are fuzzy and overlapping. Also, with using clustering coefficient and average path length we can observe that our network has small world properties.

In this form structural equivalence in this network is not very meaningful. The reason behind this is we have cleaned our dataset to eliminate unrelated people from the network. Before that there existed, nodes which follows one individual or nodes that are followed by other actors. These parts of the network had structural equivalence and can be grouped by their neighbor similarity. But, since these nodes are not related with our project we removed them from our network. Most of the nodes that were resembling structural equivalence was Instagram pages, famous people or individuals outside of the group that are only interacting with one node (social group of that individual outside of student club).

In case of structural cohesion, our sample network has high cohesion as itself. Since modularity is low, groups are fuzzy and overlapping. Thus, cost of disconnecting the groups in network is high. For subgroups overlapping structure prevents the cohesive subgroup structure. We can address this sample data as a one big cohesive group, but subgroups are fuzzy and overlapping. Probably reason behind this is a network of one student club, where in group social interactions are dense. In theory, if we project this network over all existing student clubs in that university. Most probably each student club will resemble a cohesive group separately.

Bibliography

James Moody, J. C. (2015). *Clustering and Cohesion in Networks: Concepts and Measures.* Durham, NC, USA: Elsevie.

Sailer, L. D. (1978). *Structural equivalence: Meaning and definition, computation and application.* Elsevier.