



Department of Computer Science

CS 429/529 – Dynamic and Social Network Analysis

Assignment 4

Alperen CAN

21601740

Instructor: Miray KAŞ

Teaching Assistant: Gözde YAZICI

Bilkent University

22.11.2022 (dd/mm/yy)

EXERCISE 1

In this assignment, I run the Les Miserables network by using different grouping algorithms, which are the betweenness based Girvan-Newman algorithm and the modularity based Louvain algorithm.

Initially, I generated reports on ORA by choosing *Locate Groups* for the Les Miserables dataset as follows:

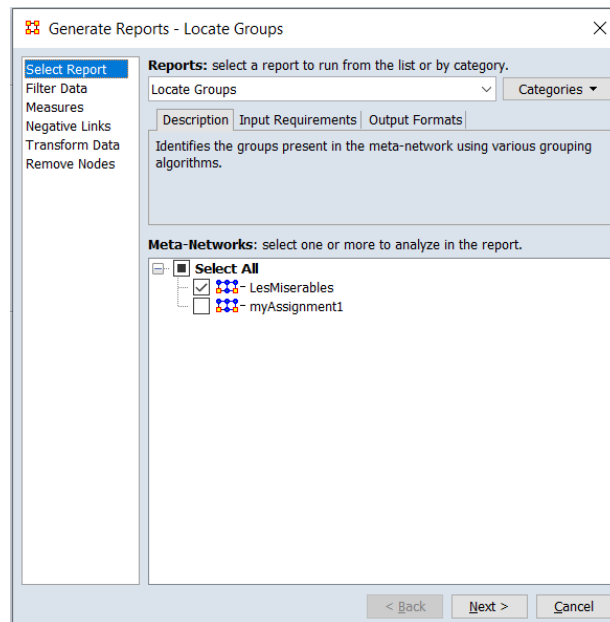


Figure 1: Generate Reports – Locate Groups

Since Louvain algorithm led to 6 different groups, I also chose 6 groups on the *Find This Many Groups* option of the Girvan-Newman algorithm as below, as Ms. Gözde Yazıcı stated at her e-mail.

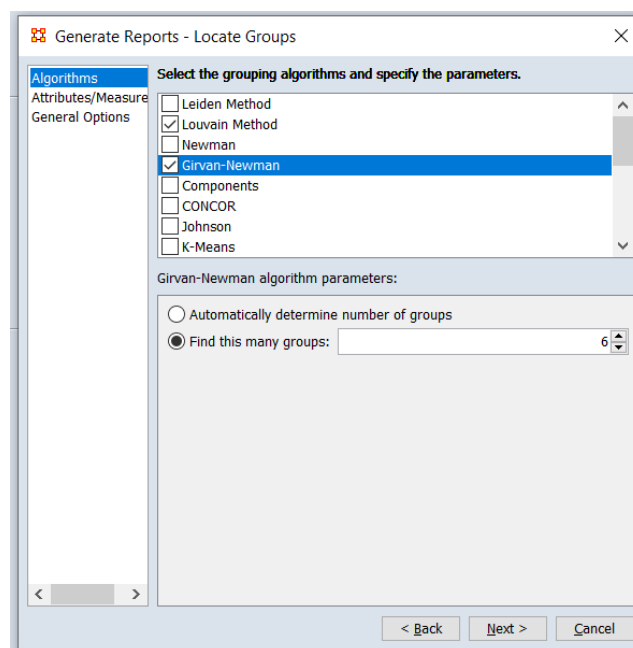


Figure 2: Specifying the Cluster Number for the Girvan-Newman Algorithm

Following, from the *General Options* tab at the left, I enabled Dendrogram since it asked for this assignment.

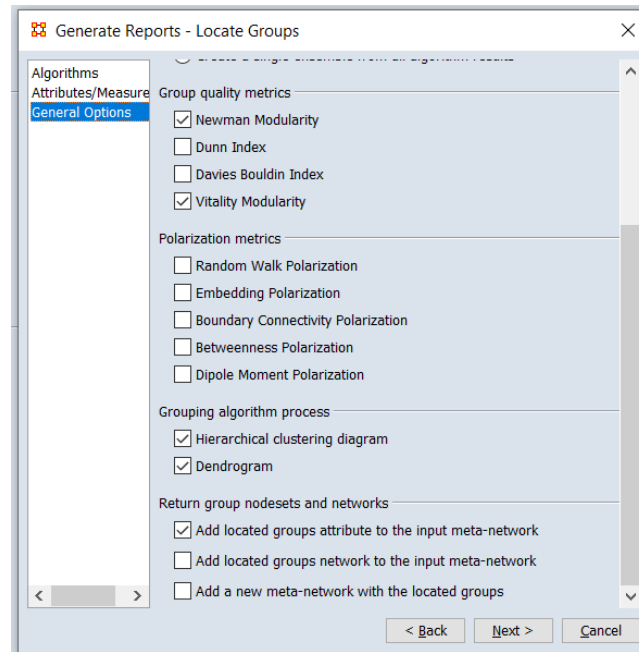


Figure 3: Enabling Dendrogram Option

a) The list of nodes and the groups they belong to are as follows:

a.1) Girvan-Newman Clustering Algorithm

Group	Size	Members
1	23	55, 27, 25, 24, 26, 41, 68, 69, 70, 71, 49, 51, 75, 54, 39, 42, 43, 72, 50, 52, 56, 40, 53
2	17	48, 58, 62, 64, 63, 65, 57, 59, 61, 66, 60, 76, 47, 73, 74, 46, 67
3	16	11, 29, 34, 35, 36, 37, 38, 31, 12, 30, 33, 10, 13, 14, 15, 32
4	10	0, 2, 3, 1, 4, 5, 6, 7, 8, 9
5	8	23, 16, 17, 18, 19, 20, 21, 22
6	3	28, 44, 45

Figure 4: Nodes and their Groups for the Girvan-Newman Algorithm

a.2) Louvain Method Clustering Algorithm

Group	Size	Members
1	22	11, 55, 27, 26, 49, 51, 28, 54, 39, 43, 72, 33, 44, 52, 56, 10, 13, 14, 15, 32, 45, 53
2	17	48, 58, 62, 64, 63, 65, 57, 59, 61, 66, 60, 76, 47, 73, 74, 46, 67
3	11	25, 24, 41, 68, 69, 70, 71, 75, 42, 50, 40
4	11	23, 16, 17, 18, 19, 20, 21, 22, 31, 12, 30
5	10	0, 2, 3, 1, 4, 5, 6, 7, 8, 9
6	6	29, 34, 35, 36, 37, 38

Figure 5: Nodes and their Groups for the Louvain Algorithm

Next, from the *Node Appearance* menu, I chose *Node Color*, then by *Attribute or Measure*.

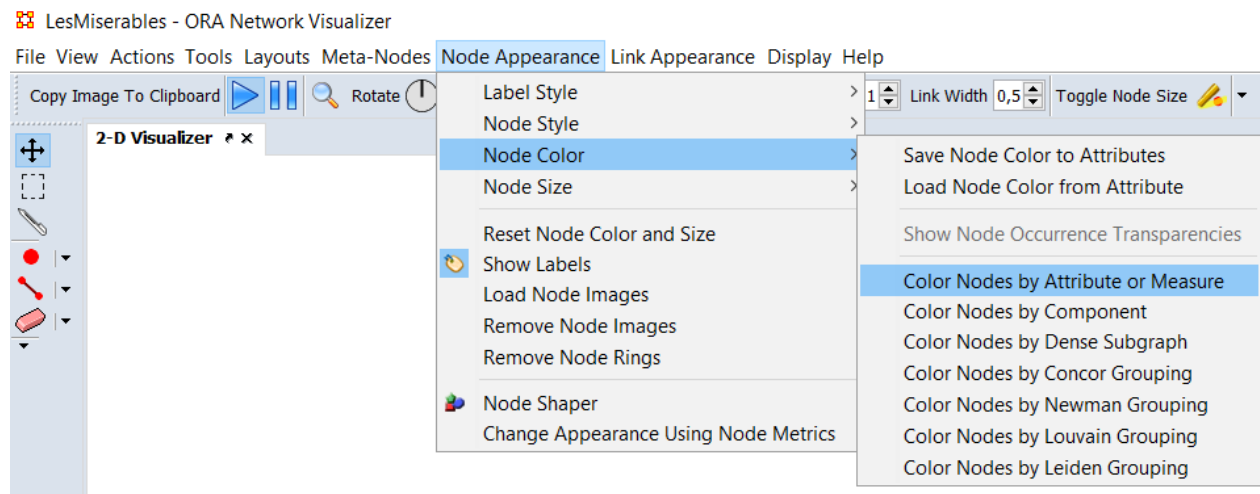


Figure 6: Node Appearance Menu

Then I chose Girvan-Newman group as below:

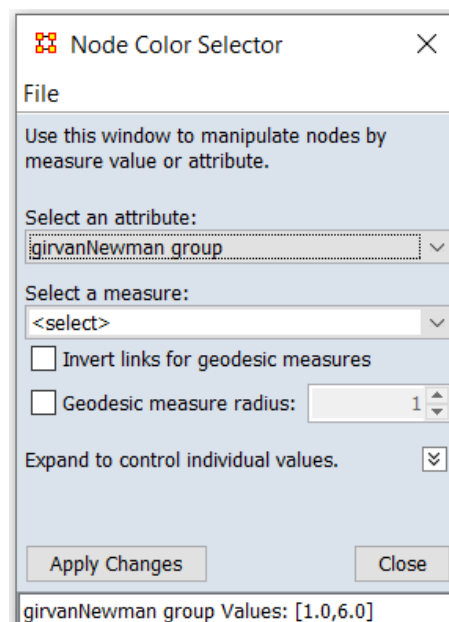
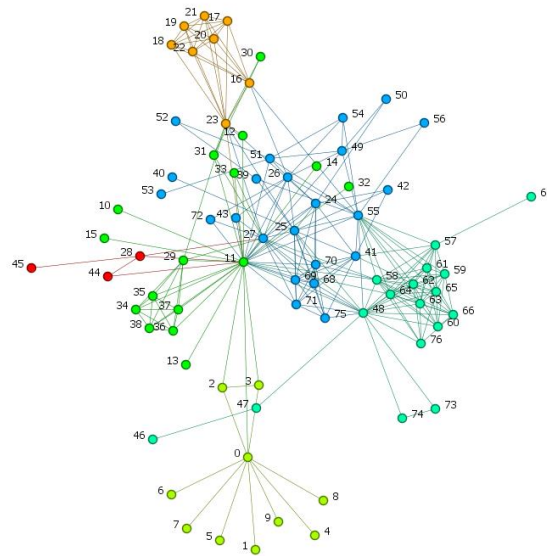


Figure 7: Coloring Nodes by Girvan-Newman Algorithm

After applying the changes, I obtained the network visualization with different node colors based on the Girvan-Newman algorithm.

b) Visualization of the network with Girvan-Newman algorithm is below:

LesMiserables



powered by ORA

Figure 8: Les Miserables Network with Node Colors Based on Girvan-Newman Algorithm

Subsequently, I repeated the same steps to group the nodes based on the Louvain Algorithm as follows:

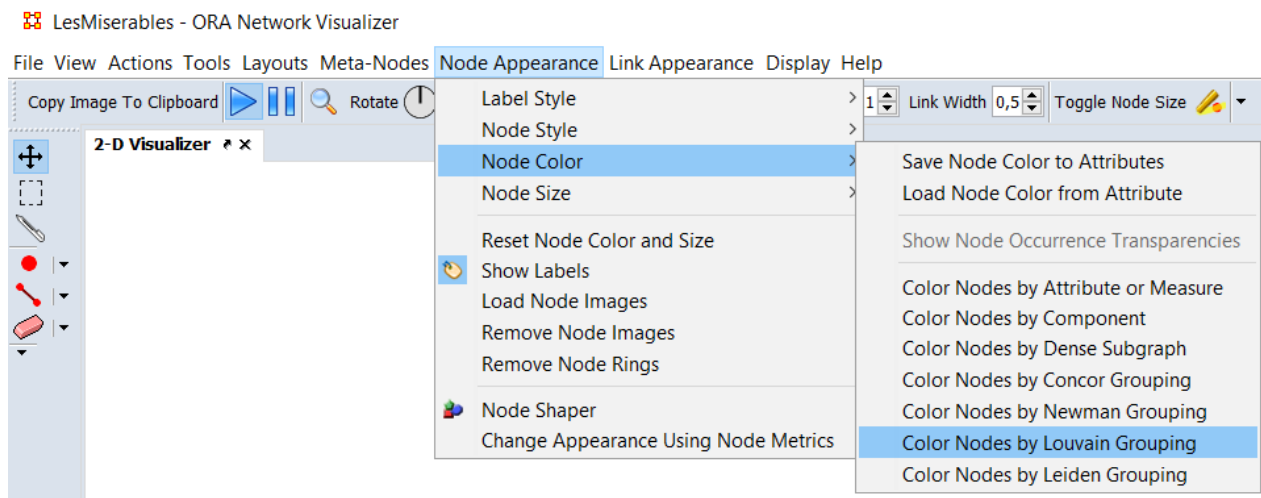
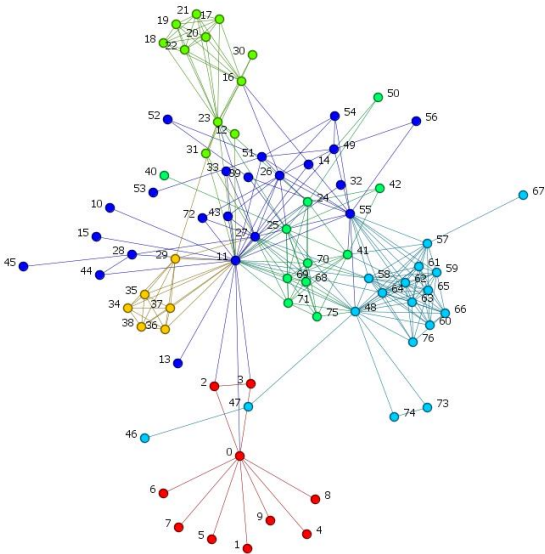


Figure 9: Coloring Nodes by Louvain Algorithm

c) Visualization of the network with Louvain algorithm is below:

LesMiserables



powered by ORA

Figure 10: Les Miserables Network with Node Colors Based on Louvain Algorithm

Later, from the .html file which ORA generated, I was able to see the dendrogram.

d) The dendrogram created by ORA (after manually specifying the cluster number as 6) is below:

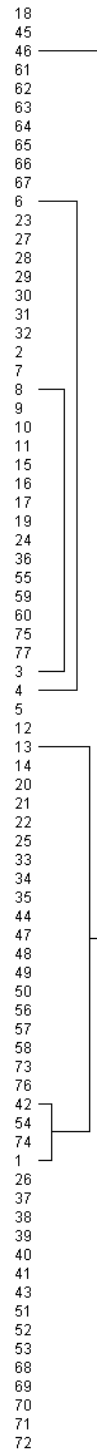


Figure 11: The Dendrogram of the Network based on Girvan-Newman Algorithm with 6 Clusters

- e) Having analyzed the groups founded by different algorithms, I can easily assert that there are strong relations among these groups. In order to see it, we can create a node by group matrix and extract one mode networks from the two mode network, as we did in the previous assignment. Similarly, we can also compare the node groups of Girvan-Newman and Louvain algorithms to see the relation.

Group	Size	Members
1	23	55, 27, 25, 24, 26, 41, 68, 69, 70, 71, 49, 51, 75, 54, 39, 42, 43, 72, 50, 52, 56, 40, 53
2	17	48, 58, 62, 64, 63, 65, 57, 59, 61, 66, 60, 76, 47, 73, 74, 46, 67
3	16	11, 29, 34, 35, 36, 37, 38, 31, 12, 30, 33, 10, 13, 14, 15, 32
4	10	0, 2, 3, 1, 4, 5, 6, 7, 8, 9
5	8	23, 16, 17, 18, 19, 20, 21, 22
6	3	28, 44, 45

Figure 12: Nodes and their Groups for the Girvan-Newman Algorithm

Group	Size	Members
1	22	11, 55, 27, 26, 49, 51, 28, 54, 39, 43, 72, 33, 44, 52, 56, 10, 13, 14, 15, 32, 45, 53
2	17	48, 58, 62, 64, 63, 65, 57, 59, 61, 66, 60, 76, 47, 73, 74, 46, 67
3	11	25, 24, 41, 68, 69, 70, 71, 75, 42, 50, 40
4	11	23, 16, 17, 18, 19, 20, 21, 22, 31, 12, 30
5	10	0, 2, 3, 1, 4, 5, 6, 7, 8, 9
6	6	29, 34, 35, 36, 37, 38

Figure 13: Nodes and their Groups for the Louvain Algorithm

For example, from the tables it can be easily seen that the 2nd group is exactly the same. Similarly, the nodes in the 4th group of Girvan-Newman algorithm (nodes 0,2,3,1,4,5,6,7,8,9) is also together a group in the Louvain Algorithm (Group 5), even without any addition of different nodes. Likewise, if we look at to the 5th group of Girvan-Newman Algorithm, the nodes 23,16,17,18,19,20,21,22 also appears in the 4th group in the Louvain algorithm, but this time with the addition of only 3 different nodes, which are 31,12,30.

We can also find other similarities, but in my opinion these examples are sufficient to show that there is definitely a relation between the groups generated by different algorithms. These relations and similarities can be explained by the fact that there are always some nodes in a network which are strongly connected, that is, there are always some strong ties within the community. While some weak ties among the nodes which are bridging across the communities are tend to change more when different grouping methods are applied, the nodes which are strongly connected in a network are highly likely to appear in similar groups, even though the clustering is done based on other methods and attributes. That is the reason why actually we are seeing that strongly connected node groups also appear together as different grouping algorithms applied, just like in the node group example that I provided below. Following images are cut from the visualizations (from Figure 9 and 10). Note that node 13 (bottom right) is changed the group when a different clustering method is applied, but the nodes with strong ties are more likely to maintain the group. Because, betweenness and modularity are not very independent from each other, just like many other metrics in the Social Network Analysis field.

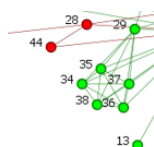


Figure 14: Same group of nodes in Girvan-Newman

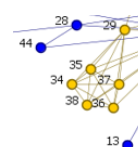


Figure 15: Same group of nodes in Louvain

EXERCISE 2

a)

Having analyzed the structural equivalence and cohesive groups concepts, I can assert that two or more nodes are structurally equivalent if they have connections with only the same nodes in the network. For example, from the figure that I included, I can state that we can talk about structural equivalence for the nodes B and D, because we see that they are similar in a way that they connected to the same nodes, which are A, C and E.

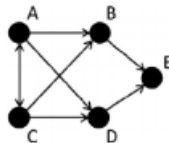


Figure 16: An Example Network

In other words, I think I can describe them as almost *identical*, based on their environment. It is a fact that generally our aim is to predict the outcomes in the Social Network Analysis field. In order to do it, we must take into account of a node's position and relations to other nodes. Here, I think this concept is probably important due to the idea that if two nodes have similar environments, then maybe they will have similar behaviors as well. So, in order to predict one, we can observe its identical and take lessons from it, then predict its structurally equivalents' behaviors.

On the other hand, cohesion means the togetherness of the nodes. Therefore, rather than being identical, a cohesive group refers to the nodes with strong relationships in a network, which is related with the community concept. The member of these groups have mutual and frequent ties compared to the other nodes in the network. In other words, I think we can simply define them as the nodes which have a more *closeness*.

Now we can come to the question that in what situations structurally equivalent groups can be more meaningful than cohesive groups. As we know, Social Network Analysis have use cases in government and also in the health sector. For example, in terms of governmental use cases, assume the government focused on a malicious group, such as drug dealers. In order to dissolve this group to reduce the frequent communications and trades, analyzing the center of this network which has the most number of connections and strongest ties (cohesive group) is important. However, eliminating a bridge node between this community and other ones may not go well as expected, because there might be some other bridging nodes who exactly have the same incoming and outgoing links to other nodes, that is, there might be some structural equivalents of the bridges who are eliminated. In this case, information and trade flow will not be damaged as planned. Therefore, rather than focusing the cohesive group, focusing on the structurally equivalent groups may be a good start point to not waste effort. So that we would know that eliminating some bridges which do not have the structural equivalents (weaker bridges) may result in better outcomes.

Same applies for the health sectors, such as protein gene interactions. If there are some substances which are beneficial to human genome, analyzing their cohesive groups may give hints. However, since there are key and lock relationships between enzymes and substrates, all nodes in the cohesive groups may not give the outcomes that we expect, but the ones which have in and out connections with the same nodes (structural equivalents) may give better hints to detect the most beneficial ones.

- b)** As we explained in our project short proposal, our aim is to do a network analysis about the misinformation regarding covid-19 vaccines. Therefore, we are aiming to gather tweets which spread misinformation. Eventually, we will detect the most influential users on misinforming the others, such as the ones having many retweets – that is – the ones who plays an intermediary role. In our network, there are also some sub-groups. For example, within the network of anti-vaxxers, there are also different cohesive groups such as the ones who are mostly anti-mRNA vaccines such as Biontech, or the ones who thinks the vaccines which affects with traditional methods such as Sinovac are not effective at all. In order to fully analyze the spreading of misinformation, we will focus on these cohesive groups separately by measuring the necessary metrics. However, focusing on the important bridge nodes may not result in better consequences since there may be multiple users who are also connected to the nodes with same incoming and outgoing links, retweeting and liking constantly, behaving like both energizers and also the bridges. Therefore, as I mentioned in the previous part, emphasizing the accounts with same incoming and outgoing links, that is, who follows and be followed by the same users will enable us to obtain better results, since these nodes can be considered as structurally equivalent nodes. Of course, since Twitter is a real life network with numerous connections, it may not be possible that two nodes are exactly the same structural equivalent, but still, I believe that analyzing them will definitely benefit our work.

We are collecting our sample data by the Python code that we wrote. We are going to improve it. Currently, it detects some parts of the tweets including “covid vaccine”. We are currently waiting a reply from Twitter to be able to fully obtain our data, as Ms. Kaş stated in the feedback that she gave us. This is the example code segment:

```
import snsrape.modules.twitter as sntwitter
import pandas as pd

query = "covid vaccine lang:en until:2021-01-01 since:2020-01-01"
tweets = []
limit = 10;

for tweet in sntwitter.TwitterSearchScraper(query).get_items():

    if len(tweets) == limit:
        break

    else:
        tweets.append([tweet.date, tweet.user.username, tweet.content])

data = pd.DataFrame(tweets, columns=['Date', 'User', 'Tweet'])

print(data)
```

Some sample data are as follows:

Brains behind new **5G** data communications networks described below!
New Bill Gates sponsored **corona virus** vaccine, w/nano tech, will run everything and control everyone who are still necessary, like bots to serve the elite? Get your vaccine now?

 [Get the facts about COVID-19](#)




The Rise of AI

There's an AI revolution sweeping across the world. Yet few people know the real story about where thi...

[youtube.com](#)



Rep Andy Biggs 

@RepAndyBiggsAZ

In 2021, the nine most terrifying words in the English language:

"I'm from the government, have you been vaccinated yet?"

1:55 PM · Jul 7, 2021 · Twitter Web App



tokidoki340
@shiroi115

...

Replying to @shiroi115

Experts all agree that the Covid-19 vaccine is not a vaccine. It was designed to make you sick with multiple diseases, including Cancer, Alzheimer's. This shot is a synthetic pathogen that is injecting the disease into your cells and it cannot be removed.

12:13 AM · Jul 14, 2021 · Twitter for iPhone

2 Retweets 2 Likes

References:

[1]

<http://homepage.ntu.edu.tw/~khsu/network/reading/wasserman7.pdf>

[2]

https://www.researchgate.net/publication/304193547_Structural_Equivalence_Meaning_and_Measures