Assignment1 Report

Aditya Manjunatha

August 2024

Resources used for the assignment

I used the following resources for the assignment:

- Copilot / chatgpt

- Theory for Girvan Newman

- Sample implementation for Girvan Newman

- Original paper of Louvain algorithm

- Class Slides

Question 1

I was only able to get the complete results on applyting girvan newman algorithm on the lstfm-edges dataset.When i ran it on the wiki-Vote dataset, it was taking 14+ hours to run until my laptop crashed ;(. So i will report only partial results for this dataset.

Question 2

Come up with an automated algorithm to determine the right set of communities using the Girvan-Newman method.

I used the change in modularity to determine the stopping condition. So every time i removed the edge of highest centraility, I checked whether the modularity increased or not. If it increased, I continued removing the edges, else I stopped and returned the community matrix.

Question 4

Do the same for the Louvain algorithm. Show the communities you get after one iteration.

The communities I got for both data-sets are stored in the Q4-communities.txt file which i have attached in the zip file.
It is a list , where each element is a list of nodes in the community.
Totally 57 communties have been formed in the Wiki-Vote dataset and 677 communities have been formed in the lastfm-edges dataset.

Question 5

How would you pick the best decomposition of nodes into communities?

Picking the best decomposition of nodes into communities involves evaluating how well the detected communities reflect the underlying structure of the network.
One way to do it is by maximizing the modularity of the network, more the modularity means stronger is the division of nodes into communities.
So given a decomposition of nodes into communities, we can calculate the modularity of the network and compare it with the modularity of a random network.

Question 6

What was the running time of the Girvan-Newman algorithm versus the Louvain algorithm on the data sets you were given?

The running time of the Girvan-Newman algorithm was 14+ hours for the wiki-Vote dataset until my laptop crashed and i couldnt run it on the LSTFM dataset.

The running time of the Louvain algorithm was 50 minutes for the wiki-Vote dataset and about 7-8 minutes for the LSTFM dataset.

Question 7

In your opinion which algorithm gave rise to better communities, and why?

I personally feel that the louvain algorithm is better , becuause it uses a bottomup approach as oppeosed to the top-down approach of the Girvan-Newman algorithm.
It is computationally expensive for large networks, as calculating edge betweenness is costly
I also noticed that the Girwan newman algorithm was creating mistakes in in few places like the bridge edge between 2 connected components.
So it tends to create more single node communities and is sensitive to the removal of certain edges, which may lead to less stable communities.
Again, this might be an error in my programming.
Louvain is Extremely fast and efficient, even for large networks, it directly optimizes modularity, often leading to high-quality community structures.
So Girwan newman is better if you are working on smaller graphs and Louvain is better for larger graphs.
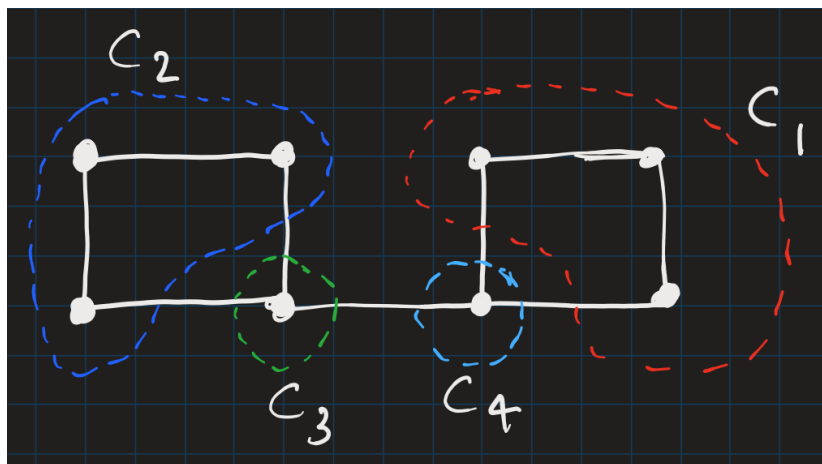

Figure 1: Clearly there should have been only 2 communities, but in few edge cases, like this Girvan fails and creates 4