Assignment 1: Community Detection

Rishav Saha
Serial No: 22573
Data Analytics (E0-259)

August 25, 2024

1 Report

Q1. Run the Girvan-Newman Algorithm on these data sets.

Ans. Code present in Assignment1.py file
Number of Communities on wiki dataset after one iteration: 24
Number of Communities on fastfm-asia dataset till self defined stopping criteria : 26

Q2. Come up with an automated algorithm to determine the right set of communities using the Girvan-Newman method. What would be your stopping criterion?

Ans. The first graph, the wiki-Vote data, includes 24 connected components at the beginning, but the fastfm dataset only has one connected component. I ran the process until we had at least two more components, 26 in the case of the first data and 3 in the case of the second, but if the communities did not increase after 15 iterations, we terminated the loop and the nxm matrix.

Q3. Visualize the resulting dendrogram to come up with an appropriate stopping criterion.

Ans. The wiki-Vote dataset yields 25 communities after the second iteration and 26 communities after the fourteenth iteration. The dendrogram is visualized in Plot 1. For the fastfm dataset, the number of communities didn't increase even after the 15th iteration, so we terminated the loop. Therefore, the dendrogram is empty for it

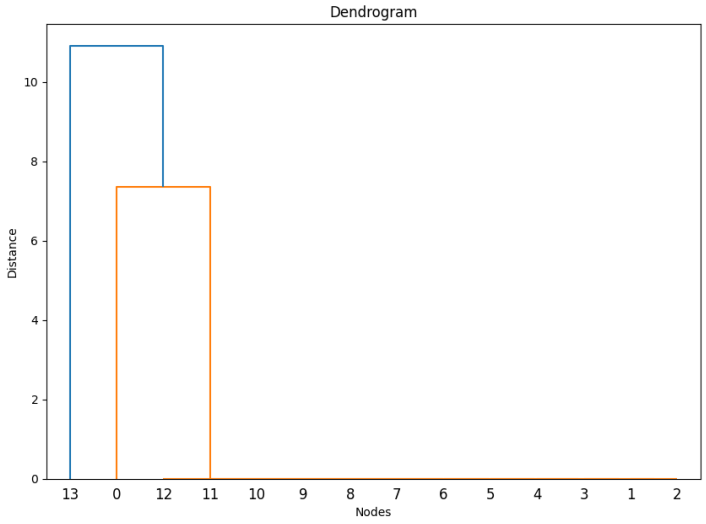

Figure 1: Plot1

Q4. Do the same for the Louvain algorithm. Show the communities you get after one iteration.

Ans. Code Present in Assignment1.py file
Number of Communities of wiki dataset after one iteration of louvain: 165
Number of Communities of lastfm-asia dataset after one iteration of louvain : 148

Q5. How would you pick the best decomposition of nodes into communities?

Ans. The time complexity of the optimized Girvan-Newman technique is $O((n^2)m)$, which climbs to $O(n^4)$ for a complete graph. However, the time complexity of the optimized Louvain technique is O(n log n), which is substantially lower. So, if there are too many nodes, we should utilize Louvain's technique; if the graph is small, we can use either of them. Also, Girvan-Newman's algorithmic implementation uses BFS, therefore if the graph already has a tree-like structure, we may use Girvan-Newman.

Q6. What was the running time of the Girvan-Newman algorithm versus the Louvain algorithm on the datasets you were given?

Ans. A single iteration of Girvan-Newman takes around 40 minutes. Based on my stopping conditions, the algorithm operates for around 6 hours. Running Louvain's Algorithm for one iteration takes approximately 5 hours.

Q7. In your opinion, which algorithm gave rise to better communities, and why?

Ans. From an efficiency perspective, Louvain's algorithm gives rise to better communities because of its time complexity.