

E209: Assignment on Community Detection

Submit your assignment in Python 3.9.12.

Data sets:

- (i) <https://snap.stanford.edu/data/wiki-Vote.html>
- (ii) <https://snap.stanford.edu/data/feather-lastfm-social.html>

Resources you can use for graph visualisation: <https://networkx.org/>, graphviz etc. There are prebuilt implementations of some of the algorithms we require in networkx etc., please do not use those. Use **networkx** and other libraries only for visualisation.

Consider the data sets given above.

1. Run the Girvan-Newman Algorithm on these data sets.
2. Come up with an automated algorithm to determine the right set of communities using the Girvan-Newman method. What would be your **stopping criterion**?
3. Visualise the resulting dendrogram to come up with an appropriate stopping criterion.
4. Do the same for the Louvain algorithm. Show the communities you get after one iteration.
5. How would you pick the best decomposition of nodes into communities?
6. What was the running time of the Girvan-Newman algorithm versus the Louvain algorithm on the data sets you were given?
7. In your opinion which algorithm gave rise to better communities, and why?

You will submit your code and your report as a single zipped file named: YourName_LastFiveDigitsOfYourSRNumber.zip. Follow the submission guidelines. Submission can be done only via Moodle. (In particular, NO EMAIL SUBMISSIONS.) The .zip file should contain the following.

- The python file should be named Assignment1.py.
- The Report file should be named Report1.pdf.
- If there are any plots, they should be PlotY.png, where Y is the plot number.
- Remember to sign the submission statement. Remember to upload the file and click on the submit button.