

CS 429/529 Assignment 3

Assignment Release Date: Monday, October 24, 2022

Assignment Submissions Due: Thursday, November 3, 2022 (Due 23:59 Istanbul Timezone)

Assignment Submission: Submit on Moodle

Questions: gozde.yazici@bilkent.edu.tr, miraykas@cs.bilkent.edu.tr

Maximum Points: 100

Important Notes Before You Start:

- Your homework assignment file(s) name should include your full name (first and last name) and the assignment number. Your submission should be submitted as a single pdf file on Moodle using the following name template:
FirstName_LastName_A#no.pdf
- Homework assignments are to be done individually, partnering on this homework assignment is not allowed.
- **Penalty points:** Submissions not following the requested file naming/format will receive a 10 point cut.

Zachary's Karate Club Dataset

The data for this assignment is about the network of friendships between the 34 members of a karate club at a US university, as described by Wayne Zachary in 1977. This is a famous benchmark dataset with more than 5000 citations. The original work in which this dataset is collected can be found [here](#) if you are interested in further reading. The paper describes information flow for conflict and fission in small groups.

The same network data is given in two different formats: karate.xml.zip is compatible with ORA, karate.gml.zip is compatible with Gephi. Pick the one that works for you. If you want to use another tool, you can export the data from one of these tools in csv format and continue from there as well. **Treat the data as an undirected, binary network with no self-loops. This is important as it would change how the metrics are computed.**

Exercise 1: Compute Network and Node Level Metrics

(60 points) Using Gephi or ORA or another software you may like, calculate the following metrics:

- a) Degree distribution (include the distribution graph in your submission)
- b) Network Density
- c) Degree centrality (Note: Use the unscaled version of the centrality metric value in your submission if you are using ORA.)
- d) Betweenness centrality (Note: Use the unscaled version of the centrality metric value in your submission if you are using ORA.)
- e) Closeness centrality
- f) Page Rank centrality

10 points for each metric above. If the metric is a node level metric, include a table for the centrality values of top 10 nodes.

Exercise 2: Network Visualization

(30 points) Provide a visualization of the network where the nodes' sizes are adjusted based on betweenness centrality (e.g. The bigger the betweenness of a node, the bigger is its size in the visualization) The visualization should have readable node labels.

Exercise 3: Interpretation and Analysis

(10 points) When you look at the three nodes that are ranked highest for each centrality metric, what did you observe? Did different centralities yield the same or different results? Explain why and comment on what the differences/similarities mean.