

CIS600 Principles of Social Media and Data Mining
Spring 2023

Week 4 Homework - Network Analysis

In this exercise, you will analyze an email communication network between employees in a college. Each node will represent an employee and each directed edge between two nodes will represent an individual email. In the directed graph, the sender will be represented in the left node and the email recipient will be represented in the right node.

- Q1. Load up the data of the directed multigraph from email.
- Using networkx, load up the directed multigraph from email_net.txt. Make sure the node names are strings. The function should return a directed multigraph networkx graph.
- Q2. Find out the number employees and number of emails are represented in the graph?
- Q3.
- a) In the university, email communication network has been created as an employee sends an email to another employee, a communication channel has been created, allowing the sender to provide information to the receiver, but not vice versa. Based on the emails sent in the data, find out whether it is possible for information to go from every employee to every other employee.
- b) Now assume that a communication channel established by an email allows information to be exchanged both ways.
- Based on the emails sent in the data, find out whether it is possible for information to go from every employee to every other employee?
- Q4. Find out the number of nodes in the largest (in terms of nodes) weakly connected component?
- Q5. Find out the number of nodes in the largest (in terms of nodes) strongly connected component?
- Q6. Using the NetworkX function `strongly_connected_component_subgraphs`, find the subgraph of nodes in a largest strongly connected component. Call this graph `G_sg`.
- Q7. Find out the average distance between nodes in `G_sg`?
- Q8. What is the largest possible distance between two employees in `G_sg`?