Ain Shams University
Faculty of Engineering
Computers and Systems Eng. Dept.

Project
Social Media Analysis

CSE 323
Programming with Data Structures
Spring 2019

Background

Fundamentally, “computer science is a science of abstraction.” Computer scientists must create abstractions of real-world problems that can be represented and manipulated in a computer.

Graph Theory is a relatively new area of mathematics, first studied by the super famous mathematician Leonhard Euler in 1735. Since then it has blossomed into a powerful tool used in nearly every branch of science and is currently an active area of mathematics research. A graph is a set of points (we call them vertices or nodes) connected by lines (edges or arcs).

Some of the applications of graph theory are:

- Networks: A network consists of sites that send and receive messages of various types.
- Maps: A map consists of places that are connected by roads.
- Social Media platforms: The relations (friendship) among users can be modeled as a graph of nodes and edges.

Ain Shams University
Faculty of Engineering
Computers and Systems Eng. Dept.

Project
Social Media Analysis

CSE 323
Programming with Data Structures
Spring 2019

Task

The project aims at applying algorithms based on graph theory to compute the **centrality** property for each node in the graph.
Definition of 'central' varies by context/purpose. However, the main aim is to give a score to each node. This score is used to determine the most important nodes in the graph. These techniques are used in social media analysis to determine the set of influencers out of a graph containing millions of users and connections.
Centrality has many definitions. Each definition gives scores to nodes in different ways.
You are asked to implement three centrality metrics on **undirected graphs**.

Centrality in graphs

Degree Centrality

Degree centrality is the simplest centrality measure to compute. Recall that a node's degree is simply a count of how many social connections (i.e., edges) it has. The degree centrality for a node is simply its degree. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1.

Example: TODO

Closeness centrality

Closeness centrality indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network.
For a node x, The centrality C(x) is defined as:

$$C(x) = \frac{1}{\sum_{y} d(y,x)}$$

where $d(y,x)$ is the shortest distance between node x and all other nodes y that are connected to it

Ain Shams University
Faculty of Engineering
Computers and Systems Eng. Dept.

Project
Social Media Analysis

CSE 323
Programming with Data Structures
Spring 2019

Betweenness Centrality

Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.

Example: TODO

Input description [MIGHT BE CHANGED]

The program should read from a file describing the graph.
The first line will contain two integers n,m where n is the number of nodes and m is the number of edges in the graph ($1 < n < 500$, $0 < m < 0.5*n*(n-1)$).
This will be followed by m lines describing the edges of the graph.
Each of the m lines will contain two integers a, b that represents an undirected edge between nodes a and b.
The last line will contain a string describing the type of centrality that will be used (DEGREE - CLOSENESS - BETWEENNESS).

Output description [MIGHT BE CHANGED]

Print n lines showing the centrality measure for all the nodes.
The output will take the form NODE C(NODE).
The lines should be sorted in a way such that the node with the highest centrality score is printed first.

Grades Distribution

TODO

Ain Shams University
Faculty of Engineering
Computers and Systems Eng. Dept.

Project
Social Media Analysis

CSE 323
Programming with Data Structures
Spring 2019

General Rules

- Each team should consist of 4 members.
- You should register your teams using the following form: TODO
- Submission will be during the second week of the summer vacation. Early submissions will be given additional bonus marks.

Sample test cases

TODO