CSC 641 Network Science, Fall 2019

Dr. Natarajan Meghanathan

Assignment 3: Network Analysis and Visualization using Gephi

Due: October 29, 2019: 7.30 PM (hard deadline; no postponement) **Max. Points:** 100

In this assignment, you will use the Gephi tool to analyze and visualize real-world networks. Gephi is a complete stand-alone application (available at: http://gephi.github.io/) that you can download and install on your computer. Though you can download the latest version and do this assignment, I suggest downloading an earlier version (actually a beta version) directly from the following link: https://github.com/gephi/gephi/releases/download/v0.8.2/gephi-0.8.2-beta.setup.exe.

A demo video of loading the .csv file of a real-world network and analyzing it in Gephi is posted at https://www.youtube.com/watch?v=IoStjWFMtCo.

Each student is assigned a real-world network (the .csv file is in the zip folder for this assignment - labeled based on the name of the network).

Use "Fruchterman Reingold" as the layout algorithm

(1) Word Adjacency Network - Jeremy Champion
(2) Anna Karnenina Network - Marcus Deanes
(3) Copper Field Network - Rashad Evans
(4) Dolphin Network - Allee Gammons
(5) Facebook Network - Amanuel Gebre
(6) Football Network - Quincy Jones
(7) Les Mis Network - Devario Lewis
(8) Political Book Network - Chidiebere Maduka
(9) Senator Press Meets Network - Augustine Ukpebor
(10) Teenage Women Friends Network - Vallimanalan Valliappan
(11) UK Faculty Network - Ran Xu
(12) US Airports Network - Xuecen Zhang

Network Metrics: You will determine the following and show the appropriate visualization:
(1) Degree distribution (node degree vs. probability of node degree) and a plot of the same in Excel
(2) A suitable network layout depicting the tradeoff and/or correlation between node degree and local cluster coefficient
(3) A suitable network layout depicting the tradeoff and/or correlation between Closeness Centrality and Betweeness Centrality values
(4) A suitable network layout depicting the different communities of nodes in your network and the Eigenvector Centrality values of the nodes.
(5) What is your average path length, network diameter and modularity score?

Submission:

(1) Report and discussion: Compile a report for the network metrics evaluated for the classical network assigned to you. Include screenshots for all the figures and layouts.
(2) Video(s): Record video(s) demonstrating your analysis of the classical network. If the demonstration runs for a longer time, you could record separate videos (one for each of the two networks) and upload them to Googledrive sent to my email address: natarajan.meghanathan@jsums.edu.

Desktop Recording Software

If you are not able to record in Canvas, you could use any of the desktop recording software and upload your recorded video.

You could try using one of the **desktop recording software** (or anything of your choice):

CamStudio: http://sourceforge.net/projects/camstudio/files/legacy/

Debut: http://www.nchsoftware.com/capture/index.html

Description of the Real-World Networks

1) Word Adjacency Network: This is a network of 112 words (adjectives and nouns, represented as vertices) in the novel David Copperfield by Charles Dickens; there exists an edge between two vertices if the corresponding words appeared adjacent to each other at least once in the novel.

2) Anna Karnenina Network: This a network of 138 characters (vertices) in the novel Anna Karnenina; there exists an edge between two vertices if the corresponding characters have appeared together in at least one scene in the novel.

3) Copperfield Network: This is a network of 87 characters in the novel David Copperfield by Charles Dickens; there exists an edge between two vertices if the corresponding characters appeared together in at least one scene in the novel.

4) Dolphin Network: This is a network of 62 dolphins (vertices) that lived in the Doubtful Sound fiord of New Zealand; there is an edge between two vertices if the corresponding dolphins were seen moving with each other during the observation period.

5) Facebook Network: This is a network of the 187 friends (vertices) of a user in Facebook. There exists an edge between two nodes if the corresponding people are also friends of each other.

6) Football Network: This is a network of 115 football teams (nodes) of US universities that played in the Fall 2000 season; there is an edge between two nodes if the corresponding teams have played against each other in the league games.

7) Les Miserables Network: This is a network of 77 characters (nodes) in the novel Les Miserables; there exists an edge between two nodes if the corresponding characters appeared together in at least one of the chapters in the novel.

8) Politics Books Network: This is a network of 105 books (vertices) about US politics sold by Amazon.com around the time of the 2004 US presidential election. There exists an edge between two vertices if the corresponding two books were co-purchased by the same buyer (at least one buyer).

9) Senator Press Release Network: This is a network of 92 US senators (vertices) during the period from 2007 to 2010. There exists an edge between two senators if they issued at least one joint press release.

10) Teenage Female Friendship Network: This is a network of 50 female teenage students (vertices) who studied as a cohort in a school in the West of Scotland from 1995 to 1997. There exists an edge between two vertices if the corresponding students reported (in a survey) that they were best friends of each other.

11) UK Faculty Friendship Network: This is a network of 81 faculty (vertices) at a UK university. There exists an edge between two vertices if the corresponding faculty are friends of each other.

12) US Airport Network (1997): This is a network of the 332 airports (vertices) in the US in 1997. There exists an edge between two vertices if there is a direct flight connection between the two corresponding airports.