CSC 641 Network Science
Fall 2016
Dr. Natarajan Meghanathan

Assignment 2
Network Analysis and Visualization using Gephi

Due: October 25, 2016: 6 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. I suggest downloading the 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 in the course website.

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 - Li-Jing Chang
(2) Dolphin Network - Alnazer Elbedairy
(3) Football Network - Carlos Martinez
(4) Les Mis Network - Osho Oyeyemi
(5) Political Book Network - Di Wu
(6) Senator Press Meets Network - Fei Yang
(7) UK Faculty Network - Chen Fang

FYI: Gephi works only with Java JDK 1.7 and not with Java 1.8. So, in the next page, I have described how to change the configuration file of Gephi to run it with Java JDK 1.7.
Also, a brief description of each of the above networks is given at the end of the assignment description.

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) <u>Report and discussion:</u> Compile a report for the network metrics evaluated for the classical network assigned to you. Include screenshots for all the figures and layouts.
(2) <u>Video(s):</u> 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.

Running Gephi along with Java JDK 1.7

You need Java Development Kit (JDK) 1.7 to run Gephi. If you have version 1.8, it would not work.

So, install JDK 1.7 and you would have Gephi running perfectly on windows 7, 8 and 10 (x64) with Java 1.7 (x64).

The good news is that you do not need to uninstall Java 1.8, you can install Java 1.7 at the same time.

Instructions:

1. Head over to http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html and install Java 1.7

2. Go to your Gephi installation folder (probably **C:\Program Files (x86)\Gephi-0.8.2**) and locate the folder **etc**

3. Inside **etc** folder you will find a file named **gephi.config**, open this file with notepad.

4. Search for the keyword **#jdkhome="/path/to/jdk"**

5. Once you find this part of code, remove the **#** from the beginning since the machine will not execute this code and it will consider it as comment.

6. Replace the text inside double quotations **/path/to/jdk** with the directory address of your Java folder. In order to find your Java folder, check your windows drive (probably **C:**), and go to **Program Files**, and find the folder **Java**, inside that folder you will find a folder that starts with**"jdk1.7"**. You need to copy the address of this folder and paste it instead of **/path/to/jdk** in the **gephi.config** file.

For me it was **C:\Program Files\Java\jdk1.7.0_67.**

7. Save the config file, if you are not allowed to do so save it on some other place and once done, replace it on the original file.

So to sum up, remove the # from gephi.config, replace the JDK path with your Java 1.7 folder, and you're done!

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.

> M. E. J. Newman, "Finding Community Structure in Networks using the Eigenvectors of Matrices," *Physical Review E*, vol. 74, no. 3, 036104, September 2006. DOI: http://dx.doi.org/10.1103/PhysRevE.74.036104.

2) 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.

> D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 3, pp. 396-405, September 2003.

3) 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.

> M. Girvan and M. E. J. Newman, "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821-7826, June 2002. DOI: 10.1073/pnas.122653799.

4) 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.

> D. E. Knuth, *The Stanford GraphBase*: *A Platform for Combinatorial Computing*, 1st Edition, Addison-Wesley, Reading, MA, December 1993.

5) 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).

> V. Krebs, "Proxy Networks: Analyzing One Network to Reveal Another," *Bulletin de Méthodologie Sociologique*, vol. 79, pp. 61-40, July 2003.

6) 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.

> J. Grimmer, "A Bayesian Hierarchical Topic Mode for Political Texts: Measuring Expressed Agendas in Senate Press Releases," *Political Analysis*, vol. 18, no. 1, pp. 1-35, January 2010. DOI: 10.1093/pan/mpp034.

7) 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.

> T. Nepusz, A. Petroczi, L. Negyessy and F. Bazso, "Fuzzy Communities and the Concept of Bridgeness in Complex Networks," *Physical Review E*, vol. 77, no. 1, 016107, January 2008.
> DOI: http://dx.doi.org/10.1103/PhysRevE.77.016107