

CSC 641 Network Science

Spring 2017

Dr. Natarajan Meghanathan

Assignment 1

Network Analysis and Visualization using Gephi

Due: March 21, 2017: 4 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 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 - Ashley Abraham
- (2) Anna Karenina Network - Jonathan Adams
- (3) Copper Field Network - Uzochukwu Akzbueze
- (4) Dolphin Network - Mahzabin Akhter
- (5) Facebook Network - Omar Aljawfi
- (6) Football Network - Benjamin Garlington
- (7) Les Mis Network - Md Hasan
- (8) Political Book Network - Anthony Lam
- (9) Senator Press Meets Network - Peter Lowe
- (10) Teenage Women Friends Network - Md Rahman
- (11) UK Faculty Network - Maruf Sumon
- (12) Jazz Band Network - Ying Zhang

FYI: 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 Betweenness 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 Google Drive sent to my email address: natarajan.meghanathan@jsums.edu.

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) Anna Karenina Network: This a network of 138 characters (vertices) in the novel *Anna Karenina*; there exists an edge between two vertices if the corresponding characters have appeared together in at least one scene in the novel.
D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*, 1st Edition, Addison-Wesley, Reading, MA, December 1993.
- 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.
D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*, 1st Edition, Addison-Wesley, Reading, MA, December 1993.
- 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.
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.
- 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.
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](https://doi.org/10.1073/pnas.122653799).
- 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.
D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*, 1st Edition, Addison-Wesley, Reading, MA, December 1993.
- 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).
V. Krebs, "Proxy Networks: Analyzing One Network to Reveal Another," *Bulletin de Méthodologie Sociologique*, vol. 79, pp. 61-40, July 2003.
- 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.
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](https://doi.org/10.1093/pan/mpp034).
- 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.
M. Pearson and L. Michell, "Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-taking," *Drugs: Education, Prevention and Policy*, vol. 7, no. 1, pp. 21-37, 2000. DOI: [10.1080/dep.7.1.21.37](https://doi.org/10.1080/dep.7.1.21.37).
- 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.
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>
- 12) Jazz Band Network: This is a network of 198 Jazz bands (vertices) that recorded between the years 1912 and 1940; there exists an edge between two bands if they shared at least one musician in any of their recordings during this period.
P. Geiser and L. Danon, "Community Structure in Jazz," *Advances in Complex Systems*, vo. 6, no. 4, pp. 563-573, July 2003.