

CptS 591: Elements of Network Science

Spring 2020

Assignment 2: Centrality and Ranking.

Due by: Mon, Mar. 09, 2020, 6pm

There are three problems in this assignment.

For Problem 1, you will work with four real-world networks. Three of these are networks selected from Mark Newman's collection of Network Data (<http://www-personal.umich.edu/~mejn/netdata/>). The networks in the collection are given certain descriptive names. The three networks chosen for this assignment are called: (1) Political blogs, (2) Neural network, and (3) Internet. Read the brief descriptions there so you get an idea for what the networks are modeling. The data sets are in GML format, which is a format the package `igraph` understands.

You are free to choose the fourth real-world network for Problem 1, but it is recommended to be a social or information network. Examples include a social network in a certain community/context (school, workplace, prison, etc.), a Facebook network, a twitter network, a who-calls-whom network, who-trusts-whom network, a citation network, etc.

For Problems 1 and 2, you are expected to use the software tool `igraph`. For some of the calculations and plots you may need to use `Numpy` (which is a package for scientific computing with Python), `MatLab` or a similar environment. Problem 3 does not require any software tool. It is an essay-type question based on reading a paper.

Your submission: will be one PDF file consisting of your solutions and an appendix describing your procedures (e.g. listing of Python/R/Matlab codes).

Problem 1 (45%). First, briefly describe the fourth real-world network you chose to work with for this problem and provide a URL to where it is located. For each of the four real-world networks you study, identify the node(s) in the network with the *highest* scores in terms of the following centrality measures: (i) Degree, (ii) Eccentricity, (iii) Closeness, (iv) Betweenness, (v) Katz index, (vi) PageRank, (vii) Kleinberg's Authority score, and (viii) Kleinberg's Hub score.

Try to find out in what way the nodes identified by these measures as the most important might be important in the network. Are there cases in which the different centrality measures identify the same node(s) as the most important? Discuss your results.

Problem 2 (45%). Generate two types of random graphs using `igraph`'s methods for the *Erdos-Renyi* and *Barabasi-Albert* (preferential attachment) models. (The Barabasi-Albert model is a simple stochastic algorithm that generates a scale-free graph, a graph with a Power-Law degree distribution.) Let the number of nodes in each case be 20. Choose the parameters in the two models such that the two graphs you generate have roughly the same number of edges. In the same fashion, generate another set of two graphs, this time with the number of nodes in each case being 40, instead of 20.

1. For each of the four graphs you generated, compute all eigenvalues and corresponding eigenvectors of the Laplacian of the graph (e.g. using `MatLab`'s (or `Numpy`'s) *eig*

function). Do not include the results of this computation in your submission, but you will need the result for the subsequent subproblems.

2. Complete the following table for each graph. (In case the graph you generated consists of more than one component, in populating the table, consider the largest connected component.)

Graph	n	m	d_{min}	d_{max}	l	D	cc_g	λ_2	λ_n
-------	-----	-----	-----------	-----------	-----	-----	--------	-------------	-------------

where n is the number of nodes; m is the number of edges; d_{min} is the minimum degree in the graph; d_{max} is the maximum degree in the graph; l is the average path length; D is the diameter; cc_g is the global clustering coefficient; λ_2 is the second smallest eigenvalue (the algebraic connectivity); and λ_n is the largest eigenvalue. Discuss the observations you make out of these data.

3. For this subproblem you will focus on the 20-nodes graphs in each of the two models. You will generate two figures, one for each model. In each figure, plot two quantities together: the eigenvector corresponding to the second-smallest eigenvalue (λ_2) and the eigenvector corresponding to the largest eigenvalue (λ_n). In the plots, the x-axis would show vertex ids and the y-axis shows value of the eigenvector. Compare and contrast the plots of the two graphs, and discuss any observations you make.

Problem 3 (10%). Read the first four sections of the SIAM Review article “PageRank Beyond the Web” by David Gleich (note a copy of the article is posted under the “Papers” page on the course’s website). From the very wide range of applications of PageRank discussed in section 4 of the article, pick three that you personally found most fascinating (or interesting) and tell me why. For each application you picked, a couple of sentences stating your reasons is adequate, but you are welcome to write a few more sentences, if you wish.