

Lab 7: Networks and Scaling Laws

DUE: Wed, December 1, in class (start of class).

This lab has two parts: building and analyzing a simple network and experimenting with two scaling laws.

Part 1: Book Network

For this part you will use the Amazon.com website to build a network of book titles on a particular controversial subject. Nodes in the network will be book titles and books A and B will be linked if Amazon's page on book A lists book B under "Customers Who Bought This Item Also Bought".

1. Go to Amazon.com and select "Books" in the Search options. Choose a controversial topic and search for books on this topic. (Some examples: "abortion", "creationism", "Obama", "Tea Party". (You can choose one of these or choose your own.)
2. Choose three titles on your topic—one for each side of the topic (e.g., pro- or anti- Obama) and one that is "neutral" or "close to center". You can decide how to classify these books by looking at their descriptions on Amazon. Start your network with three nodes corresponding to these three titles.
3. Make links from each of these nodes to nodes corresponding to three books on the "Customers Who Bought This Item Also Bought" list further down the page. For each new node you create, make links from it to three books on its "Customers Who Bought This Item Also Bought" list. Continue until you have done three iterations (first set of three books, next set of books that link to the first iteration books, next set of books that link to the second iteration books). For your report, draw your network, with the book titles in the nodes.
4. Plot the degree distribution of your network.
5. In a paragraph or so, discuss the structure of your network. Does its degree distribution approximately follow a power law? Does it exhibit clustering? Are there any obvious "hubs"? Are there any "bridges" between clusters?

Part II: Scaling Laws

6. *Test Zipf's law.* Take one paragraph of text from some source and count the frequencies of each word appearing in this paragraph. Make a table similar to that on p. 270 and plot the distribution (frequency vs. rank) similar to the plot on p. 271. On the same plot, plot the distribution you would expect from Zipf's law. Comment on whether or not the text you selected approximately follows Zipf's law. (In your report, give the original text as well as the data and plots you obtain from it.)
7. *Test Benford's law.* Find a set of data on the web, or create your own to test Benford's law, which we discussed in class and which is described in Wikipedia (Go to <http://en.wikipedia.org/> and search for "Benford's law".)

The data set should contain at least 20 items. Some examples of data could be heights of 20 world tallest buildings in feet, or weights in grams of 20 largest mammals, or number of pages in 20 longest books, etc. Plot the frequencies of the leading digits in the first 20 numbers in your data set as well as the frequencies that are expected under Benford's law (given on the Wikipedia web page). Comment on how well your data set follows Benford's law.