Six degrees of Wikipedia
Assignment 1 (Due: February 13, 2013)
Sociology 204: Social Networks

Remember to put your name and precept section on the assignment and to staple it!

So far we have read about the existence of short path lengths between randomly chosen points in a network. The models presented in Watts (1999) suggest why this might be a general property of networked systems. For this assignment you will look for those short path in the links between entries in Wikipedia.

For example, imagine trying to get from the page of Cornell University to the page Princeton University moving only through links on Wikipedia. That one is pretty easy. One path is: Cornell University -> Ivy League -> Princeton University
Another maybe less obvious path of the same length is: Cornell University -> I.M. Pei -> Princeton University
In general, there may be many paths, but always try to find the shortest.

That was a pretty easy example. Sometimes things are harder and you make wrong turns. For example trying to go from Lehman Brothers to Great Depression I went:
Lehman Brothers -> Bankruptcy -> Lehman Brothers -> New York City -> Wall Street -> Great Depression
Note that you can use the back button in your search.

Please describe the paths that you find for the following pairs (and keep track of wrong turns):
1) Princeton University to Kevin Bacon
2) Kevin Bacon to Princeton University
3) Princeton University to Carnegie Hall
4) Carnegie Hall to Princeton University
5) Princeton University to the White House
6) White House to Princeton University
7) Princeton University to Grand Canyon
8) Grand Canyon to Princeton University

9) Now go the website http://beta.degreesofwikipedia.com/ and find the actual shortest path lengths between these pages. Do you notice any systematic patterns in terms of how well you did in comparison to the actual shortest paths? You must use data to support your answer.
10) Travers and Milgram (1969) noted that of the 64 chains that reached the target 16 were sent by Mr. Jacobs, a clothing merchant in the town.
a) Did you observe funneling in your searches to Princeton?
b) Did you observe funneling in your searches from Princeton?

11) Imagine that you have to give your friend some advice about how to solve these search problems. What algorithm/strategy/approach would you advise them to use if they have to get to a random page from Princeton University?

12) The diameter of a graph is the longest possible shortest path in a graph. For example, if you calculated the shortest path between all points on Wikipedia, the diameter would be the longest of these shortest paths. Try to find two pages that have a very long shortest path connecting them (remember you can find the shortest paths here: http://beta.degreesofwikipedia.com/). You can’t use the pages of asteroids or any of the paths that are already listed on the web. Also, you should restrict your attention to the giant component (in other words, do not consider orphaned pages). Finally, you cannot use trial and error. In other words, you cannot repeatedly request random pages and calculate the path length between them. Record the process that you go through. What strategy did you use to find this pair? In your precept we will compare strategies and see which one was most successful.