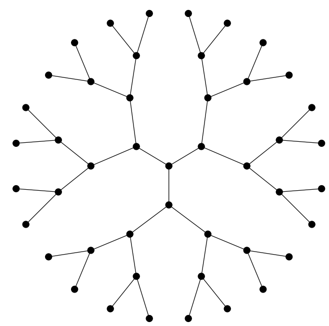
ECS 253 / MAE 253, Network Theory and Applications
Spring 2016
Common Problem Set # 2, Due May 3

Problem 1: The Cayley tree

A Cayley tree is a symmetric regular tree emanating from a central node of degree k. Every node in the network has degree k, until we reach the nodes at the maximum depth d that have degree one and are called the leaves of the network. The figure below shows a Cayley tree with $k = 3$ with depth $d = 4$.



For a Cayley degree of degree k and depth d calculate:

a) How many vertices are exactly distance one from the center?
b) How many vertices are exactly distance two from the center?
c) How many vertices are exactly distance l from the center?
d) What is $n(l)$ the total number of vertices contained within distance l from the central vertex? (Include the central vertex in this count.)
e) Present an argument that the Cayley tree has a small world diameter by showing that $d \sim \log(n)/\log(k)$. (It does not have to be a rigorous, but show your reasoning.)

Problem 2: Finite size scaling.

Consider a network where node degree k follows a power law degree distribution $p_k = (\gamma - 1)p^{-\gamma}$, with $\gamma > 1$. We will approximate k as continuous and then let P_K denote the cumulative distribution function (CDF) which is the probability a node will have degree less than or equal to K,

$$P_K = \int_1^K p_k dk.$$

Here we will work out an estimate for the maximum node degree, $K_{\max}$, that one would expect to see in a network of size N with a power law degree distribution. Operationally, we define the expected value of $K_{\max}$ for a network of size N to be the value of degree when we expect only one node bigger than this value:

$$N(1 - P_{K_{\max}}) \approx 1.$$

Using this show that $K_{\max} \approx N^{1/(\gamma-1)}$ and evaluate the expression explicitly for $\gamma = 2, 3, 4$.

Problem 3: Analysis of a real-world network

For this problem you must find a data set of a real-world network. It could be a recommendation network of books constructed via amazon.com, a flight network for an airline, a collaboration network of scientists or movie actors, a protein-interaction/gene-interaction network, a piece of the Amtrak rail network, a Facebook network, etc. The network should have somewhere between 200 to 1000 nodes.

a) Describe your data set and where/how you obtained it. Is this a directed or undirected graph? Are there several components, or is it all one connected component?

b) How many nodes and edges are present? What is the average degree? (If it is a directed graph give values for both average in- and out-degree.)

c) Plot the degree distribution (again, if directed, plot both in- and out-degree distributions). Identify the distribution that best fits your data, choosing from Gaussian, exponential, power law. (If you want to get more sophisticated, consider also power law with a cutoff and log normal distributions.)

d) Visualize the network. Try to use color or size to display interesting attributes of your data (degree, age, high-clustering, etc). You may want to label the nodes with their identities.

e) Run a community detection algorithm on your network (igraph has a number of algorithms implemented, for networkx an additional package has to be installed, e.g. http://perso.crans.org/aynaud/communities/). How many communities did you find? What is the size distribution of the communities? Use the visualization of point d) and color code the communities. Can you interpret what you found?