ECS 253 / MAE 253, Network Theory and Applications

Spring 2023

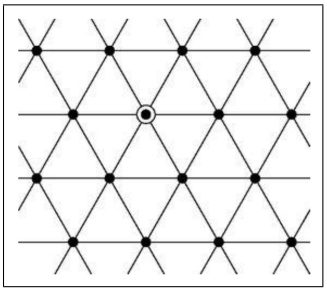
Common Problem Set # 3, Due May 15

Problem 1: Clustering coefficient

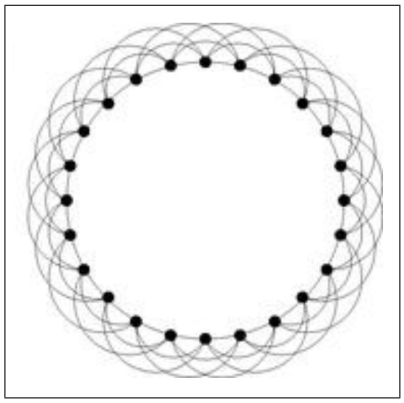
Recall the clustering coefficient C_i for node i:

C_i = (number of links between neighbors of i, excluding i) /
(total number of links that could exist between neighbors).

(a) Calculate the clustering coefficient for a vertex in this triangular lattice:



(b) Consider the regular ring that is the starting point for the Watts-Strogatz small-world network, where each node is connected to its c nearest neighbors. The example shown here is for c=6. Consider the case c=4, where a node is connected to its nearest neighbors and second nearest neighbors. What is C_i for a node in the ring with c=4?



Problem 2: 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 and networkx have options). 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?