Network Science
PHYS 5116, Fall 2019

Prof. Albert-László Barabási, Dr. Emma K. Towlson,
Dr. Sebastian Ruf, Dr. Michael Danziger, Dr. Louis Shekhtman

Assignment 1, due by October 4th, no later than 6 pm

Please write your name at the top of your assignment before handing it in. Staple all pages together. If you hand in a digital/scanned copy, you must hand in a **single**, combined document. *If you attempt to hand in more than one file, we will only grade one, at random.*

1. **Matrix formalism**
   Let $\mathbf{A}$ be the $N \times N$ adjacency matrix of an unweighted network, either directed or undirected, with N nodes. Let be $\mathbf{1}$ a column vector of N elements all equal to 1, that is $\mathbf{1} = (1, 1, \ldots, 1)^T$, where the superscript T indicates the operation *transpose*. In terms of these quantities and by using matrix operations (multiplicative constants, multiplication row by column, simple matrix operations like transpose and trace, etc. (no sum symbol $\sum$ allowed, anywhere!), write expressions for each of the following:

   a) (directed) the vectors $\mathbf{k_{in}}$ and $\mathbf{k_{out}}$ whose elements are the in and out degrees of each node;
   b) (directed) the number of bidirectional links in the network;
   c) (either directed or undirected) the number of selfloops in the network;
   d) (either directed or undirected) the $N \times N$ matrix $\mathbf{P}(n)$ for integers $n \geq 0$, where the element $P_{ij}(n)$ is the number of paths from node i to node j of length exactly n.
   e) (undirected) the vector $\mathbf{x}$ of length N whose elements are the number of triangles x_i node i is involved in (*Hint:* What is a triangle in terms of paths?);
   f) (undirected) the vector $\mathbf{c}$ of length N whose elements are the local clustering coefficients for each node i (*Hint:* use your previous answers);

2. **Graph Representation**
   The adjacency matrix is a useful graph representation for many analytical calculations. However, when we need to store a network in a computer, we can save computer memory by offering the list of links in a $L \times 2$ matrix, whose rows contain the starting and end point i and j of each link. Construct for the networks (a) and (b) in Fig. 1 the following:

   a) The corresponding adjacency matrices
   b) The corresponding link lists
   c) Determine the average clustering coefficient of the network shown in Fig. 1(a).
   d) If you switch the labels of nodes 5 and 6 in Fig. 1(a), how does that move change the adjacency matrix? And the link list?

e) What kind of information can you not infer from the link list representation that you can infer from the adjacency matrix.

f) In the left network, how many paths (with possible repetition of nodes and links) of length 3 exist staring from node 1 and ending at node 3? And in the right network?

g) With the help of a computer, count the number of cycles of length 4 in both networks.

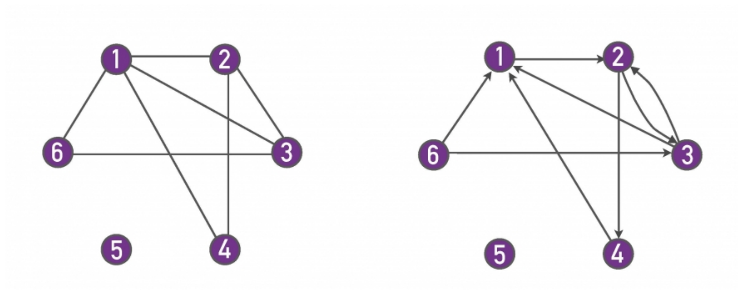

FIG. 1. Two graphs (a) with 6 nodes and 7 edges (b) with 6 nodes and 8 directed edges

3. **Circle network**
Consider a network with N nodes placed on a circle, so that each node connects to m neighbors on either side (consequently each node has degree $2m$). Fig. 2 shows an example of such a network with $m = 2$ and $N = 20$. Calculate the average clustering coefficient $\langle C \rangle$ of this network and the average shortest path $\langle d \rangle$. For simplicity, assume that N and m are chosen such that $(n-1)/2m$ is an integer. What happens to $\langle C \rangle$ if $N \gg 1$? And what happens to $\langle d \rangle$?

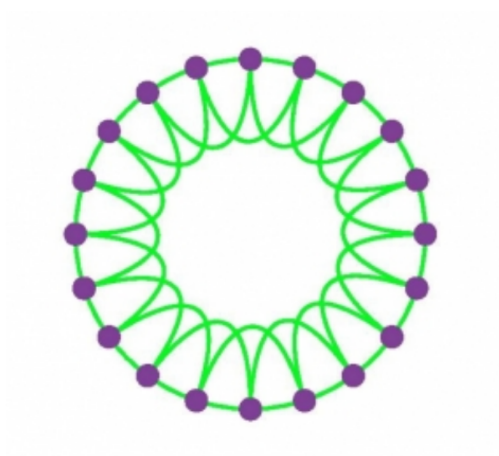

FIG. 2. A circle network with $m = 2$ and $N = 20$.

4. **Friendship paradox**
The degree distribution p_k expresses the probability that a randomly selected node has

k neighbors. However, if we randomly select a link, the probability that a node at one of its ends has degree k is $q_k = Akp_k$, where A is a normalization factor.

a) Find the normalization factor A, assuming that the network has a power-law degree distribution with $2 < \gamma < 3$, with minimum degree k_{min} and maximum degree k_{max}.

b) In the configuration model q_k is also the probability that a randomly chosen node has a neighbor with degree k. What is the average degree of the neighbors of a randomly chosen node?

c) Calculate the average degree of the neighbors of a randomly chosen node in a network with $N = 10^4$, $\gamma = 2.3$, $k_{min} = 1$, and $k_{max} = 1000$. Compare the result with the average degree of the network, $\langle k \rangle$.

d) How can you explain the "paradox" of (c) - that is a node's friends have more friends than the node itself?