Network Science
Spring 2024
Problem class 1

---

1. If you haven't done so already, read Box 2.4 in Barabasi. Now, consider a N-node directed graph without self-loops and a maximum of 1 link between a pair of nodes. Let $\mathbf{A}$ be its adjacency matrix, so $A_{ij} = 1$ if there is a link from node j to node i and $A_{ij} = 0$ otherwise.

   (a) Provide an interpretation of $(\mathbf{A}^T)^2$.
   Solution: Say $\mathbf{B} = (\mathbf{A}^T)^2$. Then B_{ij} contains the number of directed paths from node i to node j of length two.

   (b) Provide an expression using $\mathbf{A}$ and N which gives the total number of node-pairs connected by 1-way paths of length 1 and also an expression for the 2-way paths of length 1. Leave your answer in a form using index notation.
   Solution: Let $\mathbf{C} = \mathbf{A} - \mathbf{A}^T$. Then C_{ij} will be ± 1 if the two nodes are connected by a single directed link. So, $(1/2)\sum_{i=1}^{N}\sum_{j=1}^{N}|C_{ij}|$ will give the total number of 1-way node-pairs. Setting that value to N_1 and defining K as the sum of all elements in $\mathbf{A}$, the total number of node-pairs connected by 2-way links is $(K - N_1)/2$.

2. Sketch/describe 3 small graphs with distinct values for the global and average clustering coefficients
   Solution: E.g. question 5 from problem sheet 1, figure 2.16(b) in book, double-star graph from book, ...

3. The centrality measures introduced in class all avoid assigning centralities with differing signs. Is this important? Why or why not?
   Solution: Having node centralities with differing signs introduces ambiguity. Is the most important node the node with the most positive centrality or the node whose centrality is largest in magnitude? If it is the latter, then a simple centrality measure like the eigenvector centrality won't make sense. If it is the former, there will be peculiar behavior such as a node with negative centrality linking to you and reducing your importance.

   Discrete probability:
We will use a fair amount of discrete probability over the next few weeks. These exercises aim to provide a probability "warm-up".

1. Describe in words how Bernoulli trials, the binomial distribution, and the Poisson distribution are related to each other.
   Solution: The binomial distribution tells us the probability of a certain number

of successes from a given number of Bernoulli trials. The Poisson distribution is a useful approximation for the Binomial distribution when the number of Bernoulli trials is large and the expected number of successes is constant (independent of the number of trials).

2. In how many ways can the letters C,O,L,O,R be arranged so that the two O's are not adjacent?
Solution: Total number of combinations = $5!/2 = 60$ (swapping the O's gives the same 'word'). Forbidden combinations = $4! = 24$. Allowed combinations = 60-24 = 36

3. In a group of 10 students where half of them are mathematics students and the rest computing students, what is the probability of selecting a random subgroup of 5 students so that there are more mathematics students than computing students in the group.
Solution: The total number is $C_5^{10} = 252$ (elsewhere in the course we denote C_5^{10} by $\binom{10}{5}$). To have more mathematics students than computing students, we need to have either 5 mathematics students, 4 mathematics and 1 computing student, 3 mathematics and 2 computing students. The number of possible combinations are 1, $C_4^5 C_1^5 = 25$, $C_3^5 C_2^5 = 100$. The probability is therefore $126/252 = 0.5$. A symmetric argument can be used to argue as we have equal number of mathematics and computing students and it is impossible to have equal number of mathematics and computing students in the subgroup. Each group with more mathematics than computing students can be mapped to another group with more computing than mathematics students.