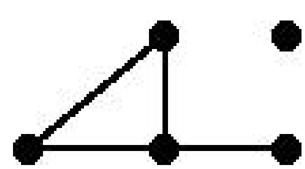
Probability Models, Test 1, due March 7, noon

1. In our model, we choose K out of N nodes at random. Another way of thinking of this is to choose the nodes one by one. In the first step, each node has probability $1/N$ to be chosen; in the second step, each node has probability $1/(N-1)$ to be chosen, and so on, in each step choosing a node according to a uniform distribution. We could generalize this to choosing nodes according to distributions other than the uniform so that some nodes might be more likely to be chosen than others. What would this mean from a biological point of view?

2. Consider the yeast network with $K = 20$ and $p = 0.01$.
(a) Compute the (approximate) speciation probability in our model.
(b) By how many percent does the speciation probability increase if we we double K?
(c) By how many percent must we increase K in order to double the speciation probability?
(d) Compute the speciation probability in Orr's model.
(e) With $p = 0.01$, how many substitutions would we need in our model to get a speciation probability that is as high as Orr's in **(d)**?

3. Consider the 3 networks from Problem 2 on HW2, revisited on HW3. For each network, compute the second-order term $\frac{g''(\mu)}{2}\,\sigma^2$ where $g(x) = 1-(1-p)^x$ and $\sigma^2 = \mathrm{Var}[X]$. You don't have to recalculate the variance but can use the number from the homework solutions.

4. For the network below, let X be the number of interactions (edges) when we choose $K = 3$ nodes and let $p = 0.1$. Find **(a)** the density α, **(b)** $E[X]$ and $Var[X]$, **(c)** the speciation probability with our approximation formula.



5. Suppose we study a particular bacterial population where we know that each interaction has a 50/50 chance of leading to an incompatibility. We also know that there is on average 1 substitution every 20 generations. If 30 populations are studied for 100 generations each and 16 of them experience no speciation events, what is the estimated density of the network?

6. Recall the quantities N_S : the number of edge pairs that share a node and N_D : the number of edge pairs that do not share a node. Find N_S and N_D in the following networks:

(a) The complete network with N nodes. The answer is $N_S = 3\binom{N}{3}$ and $N_D = 3\binom{N}{4}$ and you can try to argue for these formulas directly, or use the method with degrees that we have used in class.

(b) The disjoint network with N nodes.

(c) This network: