Probability Models, Test 1, due February 23

1(a) In what sense is Orr's model a probability model? In what additional sense is our model a probability model?
(b) In our model, we choose K out of N nodes at random. Another way of thinking of it 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. What would this mean from a biological point of view?

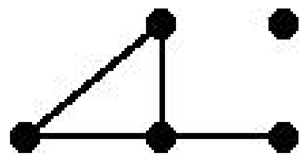
2. Consider the yeast network with $K = 20$ and $p = 0.01$.

(a) Compute the 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)**?
(f) What does the Finnish word *lumipallo* mean?

3. Suppose we study a particular bacterial population where we know that each gene 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?

THERE IS ONE MORE PAGE!

4. For the network below, let X be the number of interactions (edges) when we choose $K = 3$ nodes. Find **(a)** $P(X = j)$ for $j = 0, 1, 2, 3$, **(b)** $E[X]$ and $\mathrm{Var}[X]$, **(c)** the density α, **(d)** the speciation probability, both exactly and with our approximation formula.



5. 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. *Hint:* Argue that $N_S = 3\binom{N}{3}$ and establish a similar type of expression for N_D.

(b) The disjoint network with N nodes.

(c) This network: