Homework 4: due 4/19

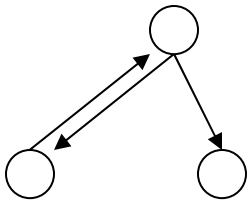
Molecular Evolution:

1. Why the empirical percentage of sequence difference between species needs to be rescaled when used as evolutionary distance?
2. For the Jukes-Cantor and Kimura models, the uniform distribution is stationary. Is this proposition true to other models such as theHKY or the 12-parameter model?
3. Show that, for Jukes-Cantor, $k = 3\alpha t$ is the expected number of substitutions in $[0,t]$, so is a distance.

Network motifs

In class we discussed network motifs: Graph substructures that are present in significant numbers in real biological networks when compared to random graphs. In this problems you will design an algorithm for determining the significance of these network motifs for a new biological interaction network.

2. a. How many possible network motifs are there for three **connected** nodes? Note that we assume directed edges and that we allow edges do go in both directions but we do not allow self influence nodes. See example of one motif below. Draw the set of resulting motifs.



2.b. Present an algorithm that given an undirected graph enumerates all the connected subgraphs with *k* nodes. Discuss its correctness.

2.c. How can the algorithm you presented in 2.b. be used to determine the significance of a specific motif? Explain and discuss what exactly you will be comparing to.