**Problem 1:**

Consider a reversible reaction $A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$ in a closed system, where the total number of molecules A and B is equal to N.

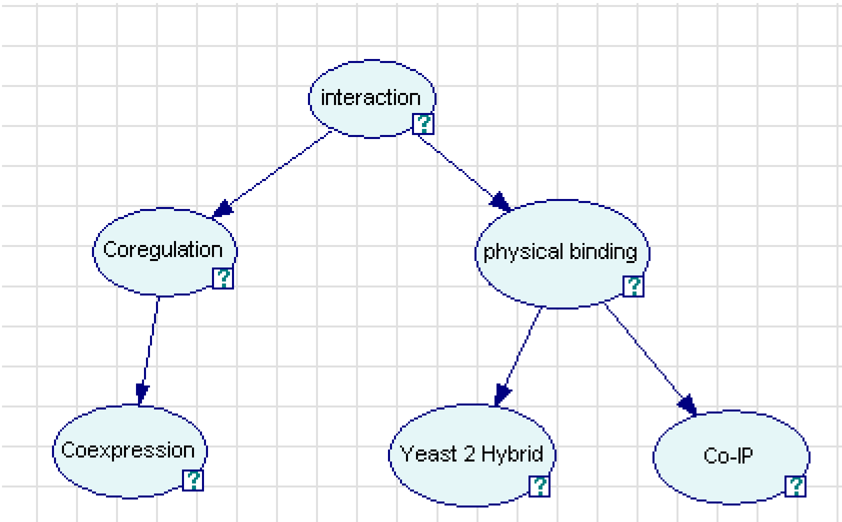
1. Write down the Master equation (with appropriate boundary conditions) that describes the dynamics the system in terms of the probability distribution function $P\left\{N_A, t \mid N_{A,0}, t=0\right\}$.
2. Solve this equation numerically at steady state for $k_1 = k_2 = 1$ and $N_A + N_B = N_{total}$. Plot the computed stationary probability distribution functions for $N_{total} = 10, 100, 1000$.
3. Implement the Gillespie's algorithm for simulating the stochastic trajectories of the system for $N_{total} = 100$. Run this algorithm starting from $N_{A,0} = 100,\ N_{B,0} = 0$. Plot a couple of representative trajectories.
4. Use the Gillespie's algorithm to construct the steady state probability distribution function. For this, run the algorithm many times for a sufficiently long time and do statistics for $N_A(t \gg (k_1 + k_2)^{-1})$.
5. Compare the stationary p.d.f. obtained with Gillespie's algorithm with the one computed using the Master equation.

**Problem 2:**

You are interested in predicting protein-protein interactions in baker's yeast based on yeast two hybrid, co-immunoprecipitation, and gene expression microarray data. You decide to use Bayesian networks to combine these methods.

1. You need a structure and parameters for the network. Luckily, in your courses in graduate school you've learned that Bayesian networks can be either expert systems or can be learnt. With this particular problem in mind, discuss what are advantages and disadvantages in learning vs. constructing based on expert opinion: A) structure of the network and B) parameters of the network. (Hint: think of properties of learning, flexibility of the algorithms, pitfalls of algorithms)

2. Either through learning or by expert opinion (depending on your answer above), you end up with the following simple structure. This network considers a pair of proteins at a time, iterating through all possible combinations of two proteins that you want to consider, and determines the confidence level of these two proteins having an interaction based on the three types of data we are using. The data includes information on whether pairs of proteins from the genome have a positive yeast two hybrid experiment, whether they have a positive co-immunoprecipitation experiment, and a gene expression dataset with expression of all proteins in the genome over a large set of conditions (50-100 experiments. Nodes in the network are discrete, and yeast two hybrid and Co-IP nodes assume binary input.




A. Discuss details of how you would process gene expression data to get it into the form that the network can accept (i.e. what exactly will the input into the Coexpression node be, and how will this input be binned into discrete categories)?

B. You decided to first use the network as an expert system. Write out conditional probability tables of each node (parameter sets). You can consider yourself the expert who provides the numbers to put in the tables (you won't be graded on the correctness of the numbers, only on the correctness of understanding how CPTs are constructed).

C. Let's say you initialized the network with input values for two proteins. Now you want to calculate the confidence level for these two proteins having an interaction. Write out formulaically (don't use the numbers from B, just symbols) how you will calculate P(interaction) based on the data you have, a parameter set, and this network structure (remember to use the structure to take into account conditional independence relationships).

3. Now you decide to learn the parameters for the network above. What data could you use for learning the parameters (i.e. what are your "answers")? Discuss how you would set up learning and evaluate how your network is doing.