**J. MIllan**

**SRL Project**

A Bayesian Network G, with N = 20 vertices and M = 30 edges, is given by means of its adjacency matrix A, and is reported in the file “graph”. All variables are binary. The adjacency matrix is given in the file “matrix.csv”

The joint probability distribution is, as usual:



where PAk denotes the set of parents of Xk.

1. Check whether the network has some loops, and, if yes, change the orientation of edges to eliminate them.
2. Define, for each variable Xk , either P(Xk), if Xk has no parents, or P(Xk | PAk).
3. Generate, with a method of choice, a sample D of R = 400 elements from the distribution P(**X**). How could you estimate the truthfulness of the sample to the distribution?
4. Starting from D only (*i.e.*, ignoring all the other information, except the number N of nodes in the network), learn the structure and the parameters of a Bayesian network with N = 20 binary variables.
   1. Assume that there are no missing values in D and that there are no hidden variables in the network.
   2. Delete 10% of the values in D, by replacing them with “?”. Learn the structure and parameters of the network in this case of incomplete data.
   3. Find a metric to compare the learned network with the original one.
5. Assume the following values have been observed in the network (evidence E):

X2 = X4 = 1 X1 = X9 = X12 = X15 = X17 = 0

Compute the marginal distribution of node X16  in the original network both with an exact and with an approximate method. Repeat in the learned network. Discuss the obtained results.