1 Using Bayesian Networks

In medical expert systems one often models *risks* that cause *diseases* which in turn cause *symptoms*. For this assignment we assume that such problems may be encoded in Bayesian networks in which all variables are boolean and in which we are given a complete specification of network structure along with the corresponding conditional probability tables (CPTs). Given this, we should be able solve for the posterior probability of a *query variable* being true (or false), given any boolean setting of other network variables.

2 Assignment

In this assignment you are to build three separate Bayesian networks with hard coded structure. The network name and the conditional probability tables, along with a stream of queries, are to be read from a file. The format of this file is given in section 3. You are free to use the programming language of your choice, but you may not use any Bayesian network tools. **Note that the CPT entries must be read by your program from the file. It is not permitted that you read the file yourself and encode a joint probability distribution by hand. This must be automated.**

2.1 Network 1: Showing a symptom of a (rare) disease

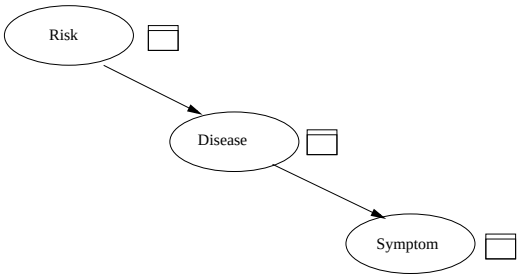

Figure 1: Network 1

An example query could be $Q(Disease|Risk)$. Given this query your program should print the probability for having the disease if we know that we are exposed to the risk. Another query could be $Q(\neg Disease|Symptom)$.

2.2 Network 2: Explaining away

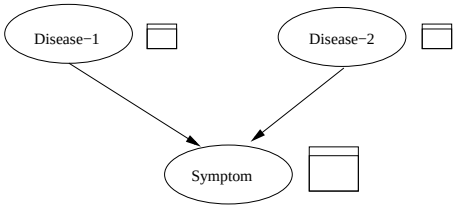

Figure 2: Network 2

An example query might be to calculate $Q(Disease-1|Symptom)$ and then to calculate $Q(Disease-1|Symptom \wedge Disease-2)$. This illustrates the property of explaining away.

2.3 Network 3: A larger network

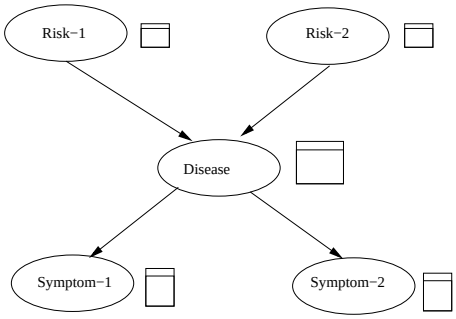

Figure 3: Network 3

There are many interesting queries to consider over network 3.

2.4 Extra Network 4: A full system (33pts extra credit)

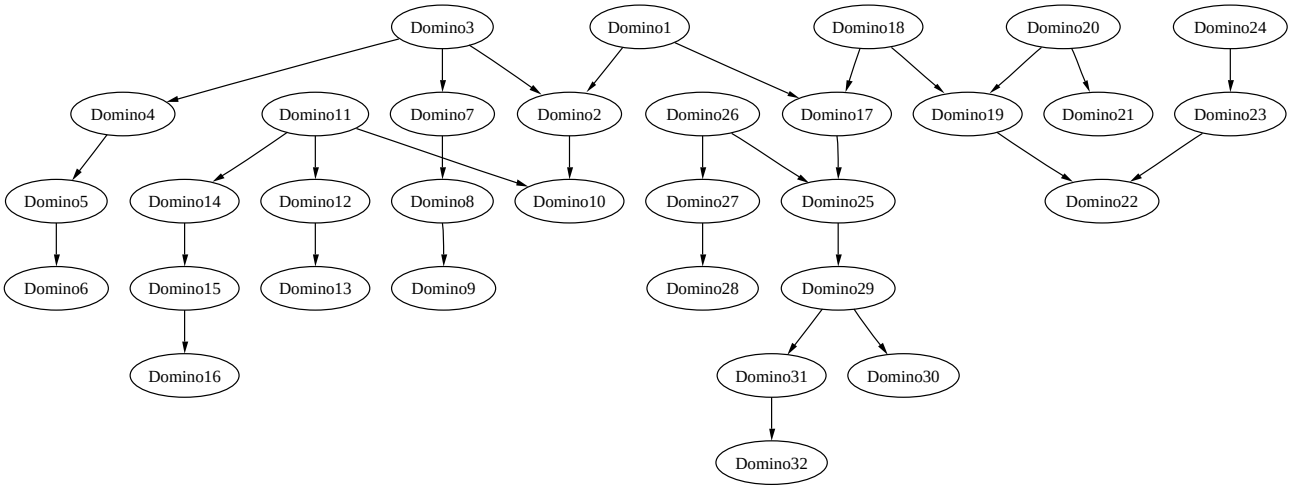

Figure 4: Network 4

Your algorithm should use either *variable elimination* or *message passing*.

3 Format

Input files are `ASCII` based. The first line of the file identifies which network it is over (`NETWORK-1`, `NETWORK-2`, `NETWORK-3`, or `NETWORK-4`). Afterwards all CPT entries are given, one per line. In doing so, commas denote $\wedge$ and exclamation marks denote $\neg$. Lines end with newlines and lines beginning with a "#" should be skipped. All whitespace in a line should be ignored. Finally a steam of queries are specified after all of the CPT entries are given. Queries are simply a line such as `Q(Symptom | !Disease-1, Risk-1)`. Your system should print the correct answer to each such query. The following file is an example of the file format.

```
NETWORK-2
```

```
  P(Disease-1) = 0.1
  P(Disease-2) = 0.004

  P(Symptom | Disease-1, Disease-2) = .999
  P(Symptom | Disease-1, !Disease-2) = 0.87
  P(Symptom | !Disease-1, Disease-2) = 0.44
  P(Symptom | !Disease-1, !Disease-2) = 0.01

# Here comes the query stream!

  Q(Symptom | !Disease-1, Risk-1)
  Q(Symptom | Risk-1)
  Q(Risk-1  | !Disease-1)
```

A series of example data files will be made available shortly. Two days before the due date we shall release the actual data files you must provide answers for. Note that the only changes between the example files and the actual files are the values of the parameters in the CPT entries and the stream of queries to be answered.

4 What to hand in

You should hand in a complete and well-written report in the box named 'Artificial Intelligence' on the fourth floor of MIT-Huset before 23:59 on the due date. Please consult `http://www.cs.umu.se/information/rapportguide.html` for guidelines on how to write this report.

Your report may be brief, but please be extra careful to turn in a test run that shows the program answering queries over the actual data files provided two days before the due date. Please place your code in `~/edu/ai/lab2/` of one of the members of your group and identify this group member in your report. You code must run under the department's LINUX system.