

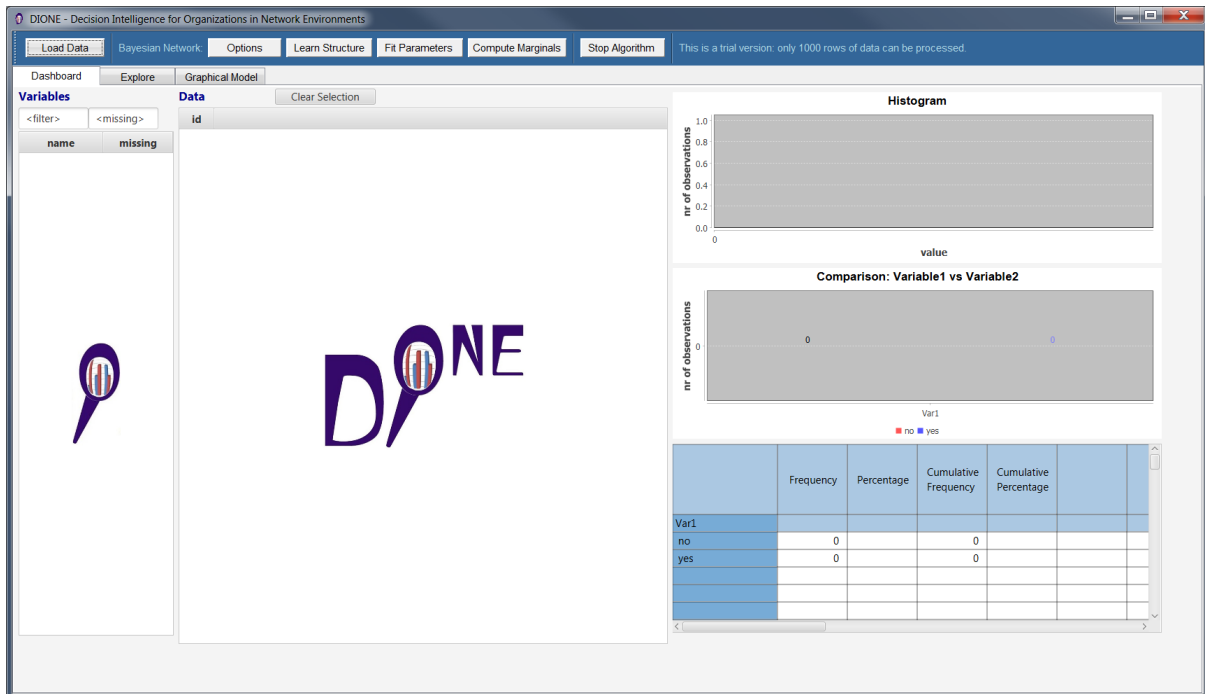


Quick Start Guide

Dr Jacob Stolk, May 2018

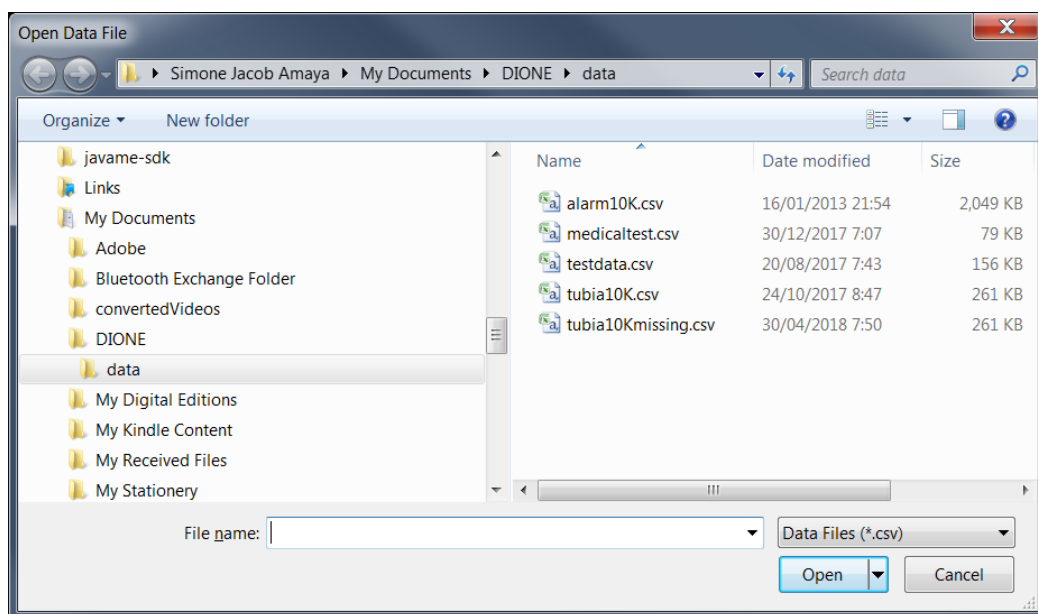
Start DIONE

To launch DIONE, double-click the DIONE desktop icon or select DIONE in the Windows Start Menu - All Programs. The start screen looks like this:

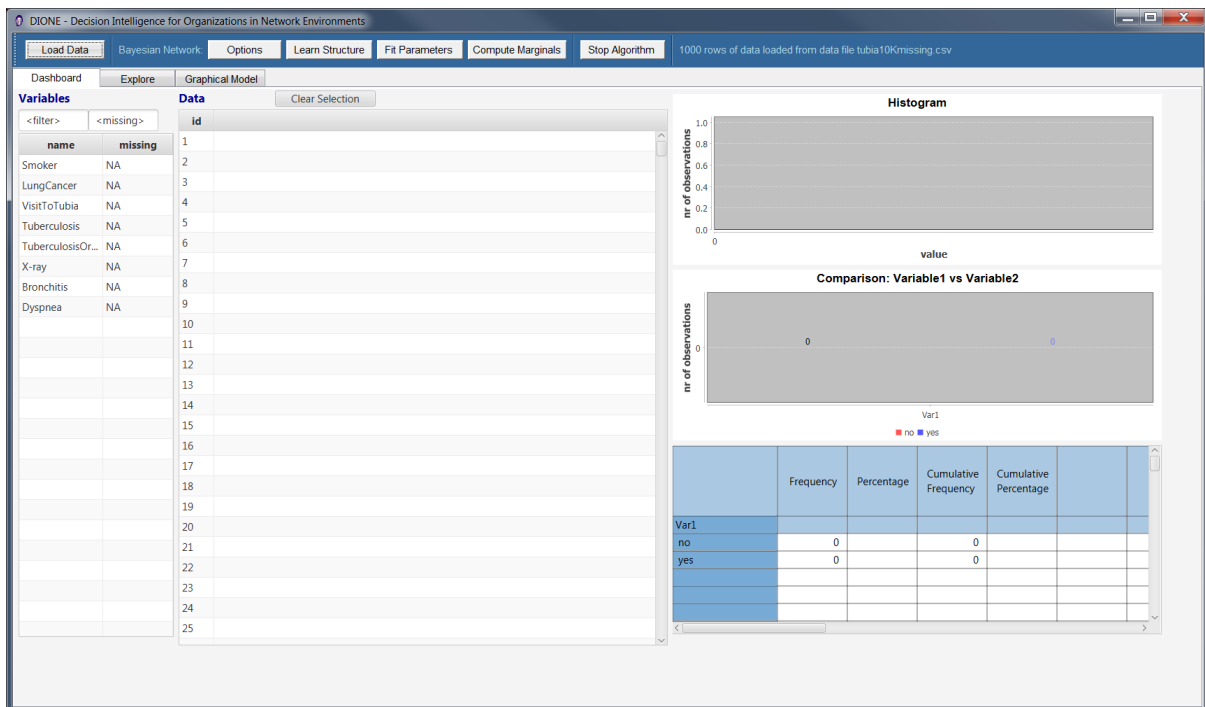


Load Data

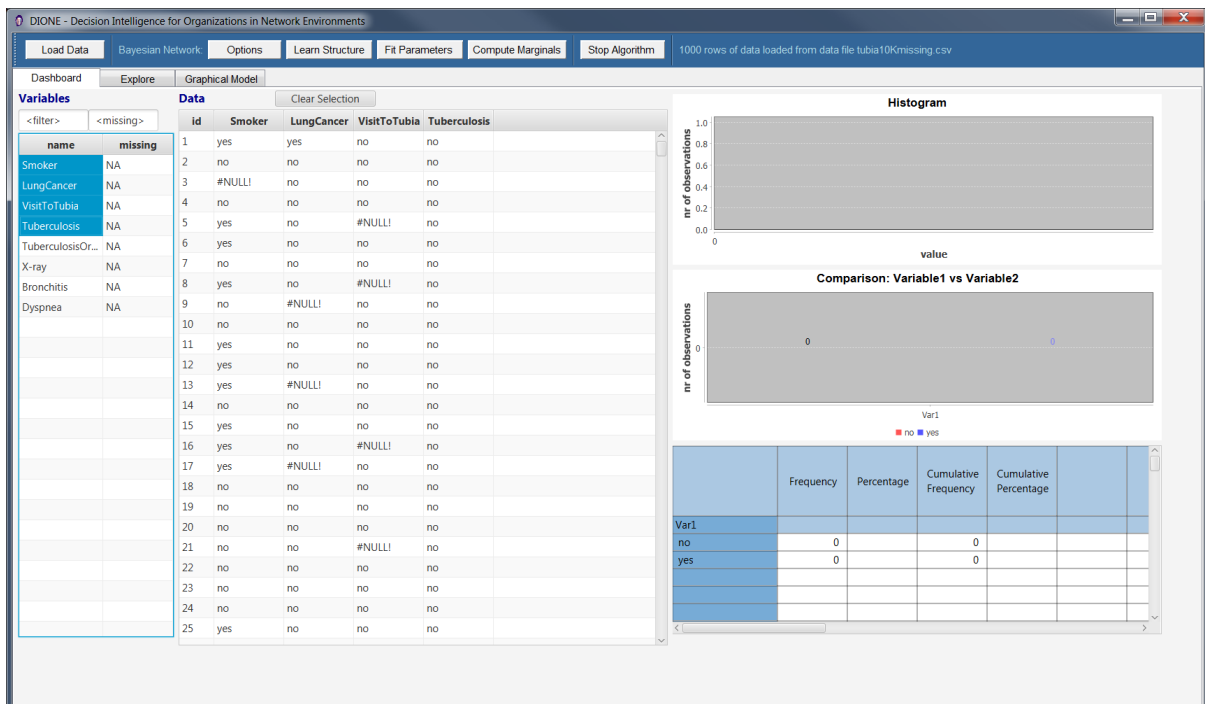
At this stage only data loading from CSV files is supported and it is assumed that column headers of the CSV data file contain variable names. To load data from a CSV file, press button Load Data and in the Open Data File dialog choose the data file you wish to analyse:



Now you see the variable names (column headers in the CSV file) in the Variables table at the left of the screen:



Select the variables you want to analyse (possibly all) by clicking and using Ctrl-click and/or Shift-click for multiple selection. You now see data corresponding to the selected variables in the Data table in the middle of the screen:



Explore Data

Select any pair of data columns, using Ctrl-click and/or Shift-click, in the Data table to see a bar chart or histogram and a frequency table of these data at the right of the screen:

The screenshot shows the DIONE software interface with the 'Explore' tab active. The 'Data' table is selected, and a comparison between 'LungCancer' and 'Smoker' is displayed. The interface includes a histogram, a bar chart titled 'Comparison: LungCancer vs Smoker', and a contingency table for 'Smoker'.

name	missing	id	Smoker	LungCancer	VisitToTubia	Tuberculosis
1		1	yes	yes	no	no
2	NA	2	no	no	no	no
3	NA	3	#NULL!	no	no	no
4	NA	4	no	no	no	no
5	NA	5	yes	no	#NULL!	no
6	NA	6	yes	no	no	no
7	NA	7	no	no	no	no
8	NA	8	yes	no	#NULL!	no
9	NA	9	no	#NULL!	no	no
10		10	no	no	no	no
11		11	yes	no	no	no
12		12	yes	no	no	no
13		13	yes	#NULL!	no	no
14		14	no	no	no	no
15		15	yes	no	no	no
16		16	yes	no	#NULL!	no
17		17	yes	#NULL!	no	no
18		18	no	no	no	no
19		19	no	no	no	no
20		20	no	no	no	no
21		21	no	no	#NULL!	no
22		22	no	no	no	no
23		23	no	no	no	no
24		24	no	no	no	no
25		25	yes	no	no	no

Smoker					
LungCancer	#NULL!	no	yes		
#NULL!	0.00%	1	0.20%	2	0.39%
no	100.00%	485	98.98%	462	90.77%
yes	0.00%	4	0.82%	45	8.84%
Total	100.00%	490	100.00%	509	100.00%

A more detailed view of the analysis results is shown in the Explore tab:

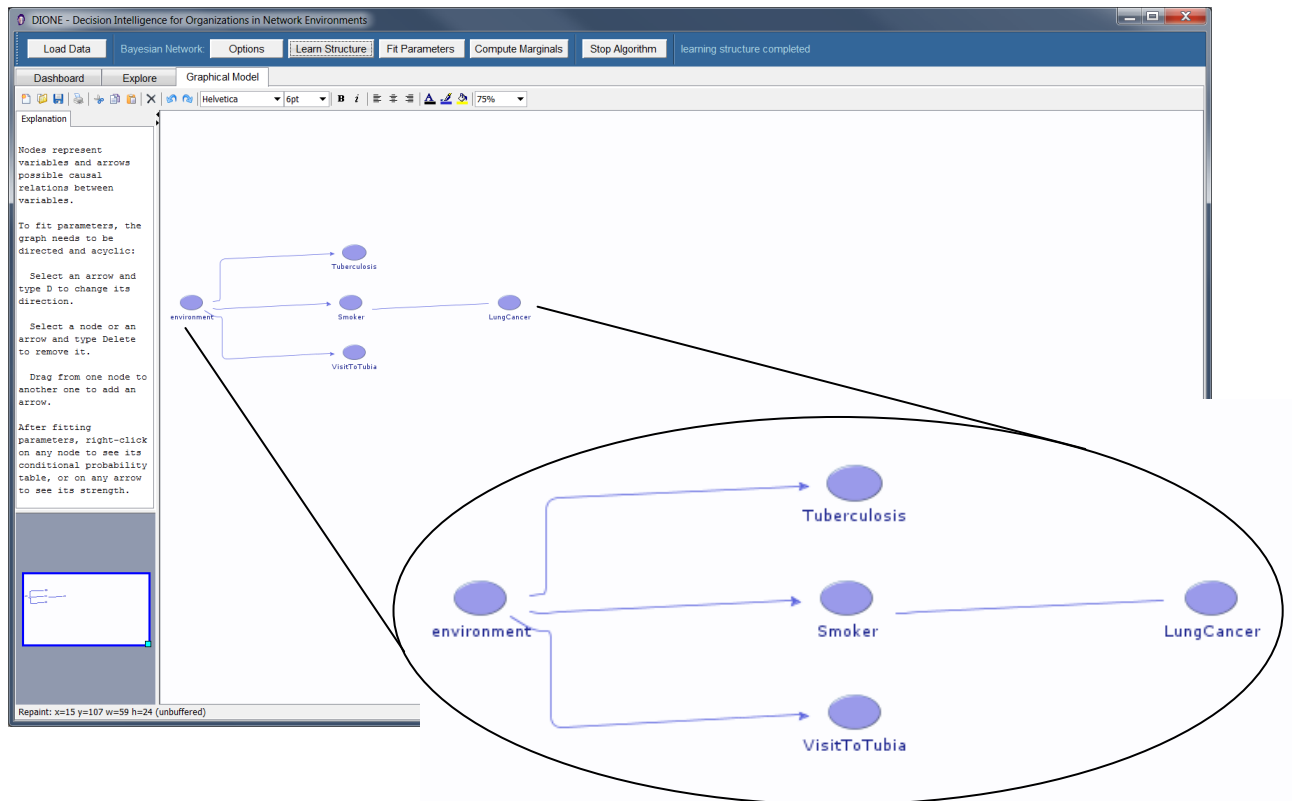
The screenshot shows the DIONE software interface with the 'Explore' tab active. The 'Comparison: LungCancer vs Smoker' bar chart and the contingency table are expanded.

Smoker						
LungCancer	#NULL!	no	yes			
#NULL!	0	0.00%	1	0.20%	2	0.39%
no	1	100.00%	485	98.98%	462	90.77%
yes	0	0.00%	4	0.82%	45	8.84%
Total	1	100.00%	490	100.00%	509	100.00%
Chi square p-value				0.00		

Analyse Data

Ensure that the variables you want to analyse are selected in the Variables table and shown in the Data table. If you want to change the selected variables, first press button Clear Selection to remove the previously selected data from the Data table.

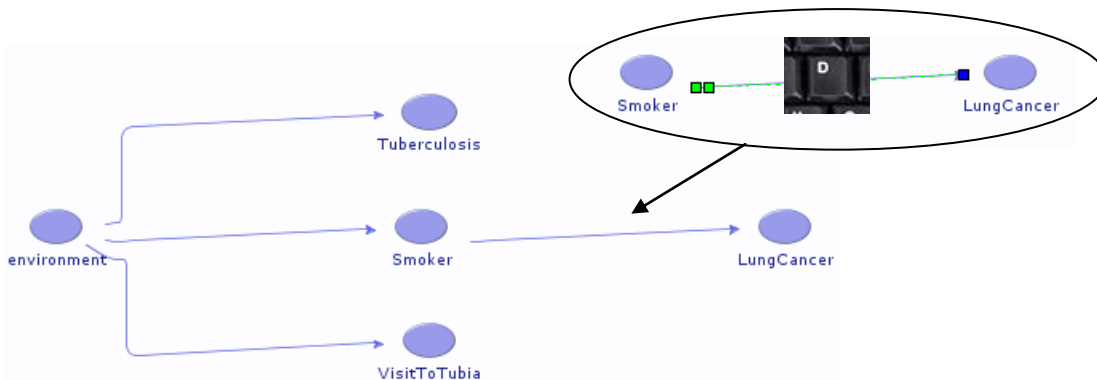
Now press button Learn Structure to learn a Bayesian network from the selected data. The result should look like this:



In the learned Bayesian network, nodes represent variables and arrows represent possible causal relations between variables. Directed lines, with arrows, represent possible causal relations including the direction of causality, while undirected lines, without arrows, represent possible causal relations with unspecified causal direction.

More information about these relations can be gained by fitting parameters of the network to obtain conditional probability tables, showing the probability of values of a node given values of its parent nodes. The parameter fitting algorithm requires that the graph is directed, i.e. no undirected arrows are present.

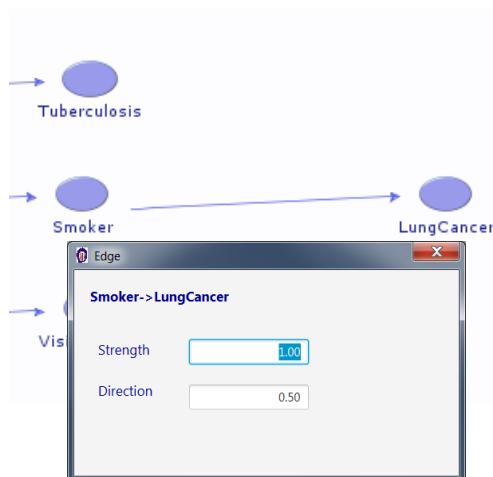
If undirected arrows are present in the learned network, you can specify a direction by selecting an arrow and pushing the **d** (lowercase) key on your keyboard, possibly several times until the arrow has the direction you want:



In this way you can also change the direction of arrows if their direction does not make sense to you.

Now to fit parameters, push button Fit Parameters.

When the algorithm has finished, right-click on any node to see its conditional probability table and on any arrow to see its strength:



Node

LungCancer

Conditional Probabilities

Parent Node	LungCancer=#NU...	LungCancer=no	LungCancer=yes
Smoker=#NULL!	0.00	1.00	0.00
Smoker=no	0.00	0.99	0.01
Smoker=yes	0.00	0.91	0.09

Marginal Probabilities

LungCancer=#NU...	LungCancer=no	LungCancer=yes	Evidence
0.003	0.948	0.049	<input type="checkbox"/>

Now you can perform inference on the Bayesian network by setting evidence on one or more nodes, then recomputing marginal probabilities to see the effect of the evidence on the other nodes:

Node

Smoker

Conditional Probabilities

Parent Node	Smoker=#NULL!	Smoker=no	Smoker=yes
env=e	0.00	0.49	0.51

Marginal Probabilities

Smoker=#NULL!	Smoker=no	Smoker=yes	Evidence
0	0	1	<input checked="" type="checkbox"/>

Node

LungCancer

Conditional Probabilities

Parent Node	LungCancer=#NU...	LungCancer=no	LungCancer=yes
Smoker=#NULL!	0.00	1.00	0.00
Smoker=no	0.00	0.99	0.01
Smoker=yes	0.00	0.91	0.09

Marginal Probabilities

LungCancer=#NU...	LungCancer=no	LungCancer=yes	Evidence
0.004	0.908	0.088	<input type="checkbox"/>