# UNINET Help

# CONTENTS

# I)      Program overview

UNINET is a stand alone program using Bayesian Belief Nets (BBNs) for stochastic modeling and for multivariate ordinal data mining. It is designed by the Risk and Environmental Modeling group at the Department of Mathematics of the Delft University of Technology, under contract with the Dutch Ministry of Transport. Chief programmer is Dan Ababei, vice programmer is Daniel Lewandowski.  Mathematical support is provided by Daniel Lewandowski, Anca Hanea, Oswaldo Morales, Dorota Kurowicka and Roger Cooke. UNINET functions in conjunction with the graphics package [UNIGRAPH](#) and the sensitivity analysis program [UNISENS](#). Certain features also work in concert with the general purpose uncertainty analysis program UNICORN. Functional influences as well as probabilistic influences are supported.  Background is found in ([Hanea et al, 2006](#), [Bedford and Cooke 2002](#), [Kurowicka and Cooke, 2006-A](#))

## 1)      Installation

The UNINET installer consists of one executable file.  Before installing a new version of UNINET, any existing old version should be uninstalled first.  Upon running the installer file the user will be prompted for a destination folder (the default is `"C:\Program Files\UNINET"`) where all the files necessary for running UNINET will be installed. Hitting the "Install" button will then completely install the program (this will take a minute or two depending on the system).

The UNINET installer actually installs three distinct programs: UNINET, UNIGRAPH and UNIsens.  UNIGRAPH and UNIsens (described at the end of chapter III) work closely with UNINET.

The minimum system requirements for installing UNINET are:

- Microsoft® Windows® XP with Service Pack 2
- 1GHz processor
- 1GB of RAM
- 20MB free disk to install
- 500MB free disk to run (this is wholly dependent of the size of the models used)

## 2)      Bayesian Belief Nets (BBNs)

A BBN is a graphical structure known as a *Directed Acyclic Graph* (DAG). A DAG is a set of nodes and arcs, or directed edges, between nodes, such that there is no directed cycle. In a BBN, nodes represent univariate random variables, and arcs represent influences. Node1 is a *parent* of Node2 if there is a directed arc going from Node1 toward Node2. Node1 is an *ancestor* of Node2 if there is a directed path from Node1 to Node2. A node with a non-empty set of parents is a *child node*. A node without parents is a *source node*, a node without children is a *sink node*.

A simple example is shown below, which is used throughout this Help to illustrate features. Note that there is an "undirected  cycle" between *truth, meaning_of_life* and *administration*, but *not* a directed cycle. Directionality in a BBN entails a (non-unique) sampling order: to sample a child node, only the current values of its parents are needed.
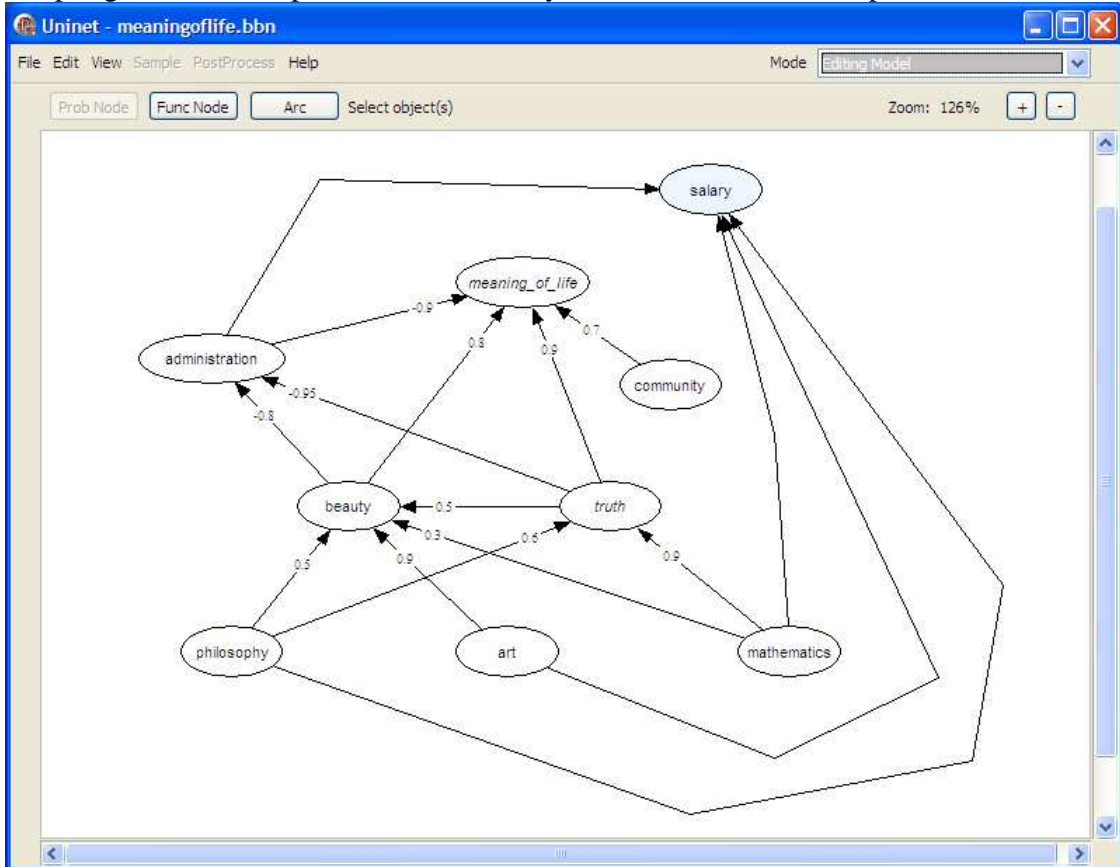


A BBN for Civil Aviation Transport Safety (CATS) involving more than 650 probabilistic nodes and 750 functional nodes is pictured below (Ale et al 2007):
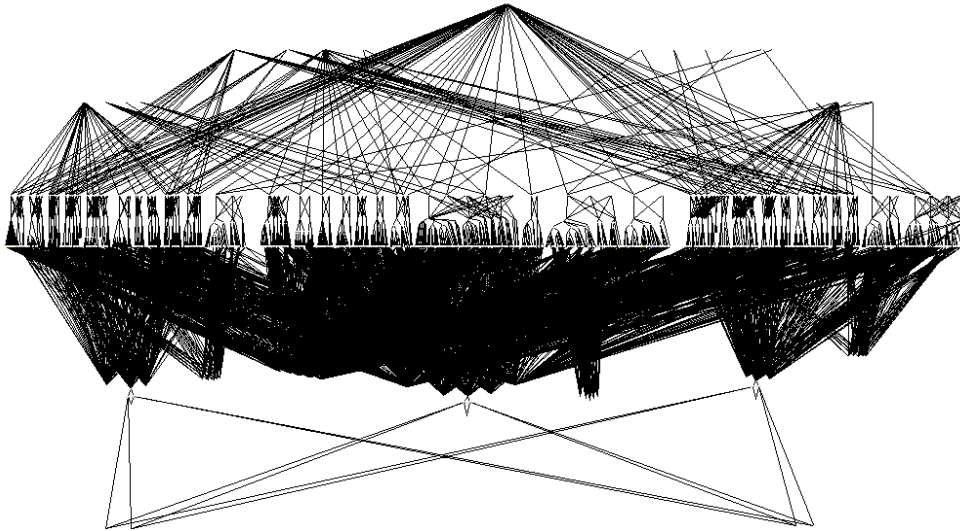
## a) Types of BBNs

Three types of BBNs may be distinguished; discrete, normal (or discrete-normal) and non parametric continuous-discrete. The first two are available in other commercial packages. All BBNs can have functional nodes (nodes which are functions of their parents); differences relate to the treatment of probabilistic nodes. The ensuing discussion refers to probabilistic nodes.

### (i) Discrete

The random variables have discrete distributions. Source nodes are assigned (marginal) distributions, each child node is assigned a conditional probability table, giving the distribution over child's possible values, conditional on each possible combination of values of the parents.` The number of probabilities that must be assessed and maintained for a child node is exponential in the number of parents. Adding a new parent node requires reassessing all children influences. If marginal distributions are known for child nodes, there is no way of entering this information directly; instead conditional probability tables must be tweaked to comply with all marginal data. Discrete BBNs have a heavy assessment/maintenance burden and are suitable for small problems only.

### (ii) Discrete normal

In this case the set of variables is assumed to follow a joint normal distribution (certain source nodes may be discrete). Each child is associated with a mean value and conditional variance. Each influence is associated with a *partial regression coefficient*. Loosely the partial regression coefficient e.g. for 5 with respect to 1 says how much the conditional mean of 5 changes as a result of a unit change in the value of parent 1. Discrete normal BBNs work well if indeed the normality assumptions hold. If not, then

The individual variables must be transformed to normal (requiring of course the marginal distributions). If F is the cumulative distribution function of random variable X (continuous, invertible), and $\Phi$ is the standard normal cumulative distribution function, then $\Phi^{-1}(F(X))$ transforms X to normal (sometimes called the normal version of X), The conditional variance *in normal units* must be constant; this is the variance with respect to the units for which the distribution is normal.  The partial regression coefficients apply to the normal units of the transformed variables, not to the original units. This places a heavy burden on any expert elicitation, and is dubious  in most real applications, unless indeed the transformation to normal is almost linear.
If a parent node is added or deleted after quantification, then the previously assessed partial regression coefficients must be re-assessed.

### *(iii) Non-parametric continuous/discrete*

UNINET supports non-parametric continuous/discrete BBNs with important distinguishing features.

## b)  Summary UNINET program features

The main distinguishing features of UNINET are:

UNINET supports non-parametric continuous/discrete BBNs, supplemented with functional nodes.

Probabilistic nodes may be assigned arbitrary continuous or discrete distributions.

Probabilistic influence is represented by *conditional rank correlation*, that is, by rank correlation in a conditional distribution, according to a protocol based on an indexing of the parents.  Influence of the first parent on the child is unconditional rank correlation, influence of the second parent on the child is the conditional rank correlation given the first parent, etc. The conditional rank correlations are algebraically independent, that is, previously chosen values do not constrain future choices. Unspecified correlations may be accommodated without confronting the matrix completion problem (Kurowicka and Cooke, 2003, 2006-B)

The rank correlation is realized using the joint normal copula. The indexing is user-editable, and parent nodes can be added without re-quantifying the previous influences.

Analytic, real time conditioning is supported for thousands of probabilistic nodes. Analytic conditioning is conditioning on a single value, and applies only for probabilistic nodes.

Sample based conditioning is supported. Sample based conditioning conditionalizes on intervals and applies both the probabilistic and functional nodes.

UNINET can model an empirical  multivariate distribution by building a joint density function, which can be analytically conditioned.

Graphics and sensitivity analysis are supported in satellite software packages UNIGRAPH and UNISENS.

## 3)      Uncertainty Analysis /  Stochastic Modeling

In uncertainty analysis/stochastic modeling, the analyst
defines a set of random variables,
defines a joint distribution for these random variables
defines other variables which are functions of the random variables
Monte Carlo samples the entire joint distributions (probabilistic and functional variables
Interprets and communicates the results.

When using UNINET to perform uncertainty analysis or stochastic modeling, the random variables are assigned marginal distributions, either chosen from a list of parametric distributions, or from a user defined distribution file. The user also specifies a DAG to capture conditional relations. Probabilistic influence between parent and child is represented as conditional rank correlation. Functional influence is defined with the help of an extended functional parser. A joint probability density for the probabilistic nodes is built using the joint normal copula to realize the dependence relations.  This density can be updated analytically, or sampled and analyzed with Monte Carlo tools.

## 4)      Ordinal Data Mining / Multivariate Density Modeling

UNINET can be used to build a joint density for ordinal multivariate data. "ordinal" means that the ordering of the values of the variables is meaningful. Ages, weights, IQ scores, income, body mass index, etc, are ordinal data. Street addresses, social security numbers, license plate numbers are not ordinal data. UNINET builds a joint distribution based on the empirical rank correlations and uses the joint normal copula. It assumes that the important dependences can be captured in dependences between the rankings of the variables.

The multivariate data set must be in the form of an XL compliant comma separated file. An example of such a file, in XL, is shown below. The variable names are on the first row, each succeeding rows contains one sample from the multivariate distribution. Note that wind direction is only an ordinal variable if it is restricted to a sector of 180 degrees.

|  | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | DC_mnthly_conc | dc_temperature | DC_windspd | DC_wndDir | nox_Richmond | so2_Richmond | so2_Phil |
| 2 | 12.5345 | 287.394 | 0.71426 | 124.2 | 80.246 | 754.229 | 30.924 |
| 3 | 12.671 | 286.908 | 0.71426 | 88.2449 | 123.45 | 1229.87 | 35.774 |
| 4 | 10.6965 | 294.807 | 0.37344 | 65.9835 | 137.128 | 1270.09 | 21.393 |
| 5 | 15.6577 | 293.649 | 0.64302 | 124.2 | 158.098 | 1296.54 | 37.101 |
| 6 | 13.1345 | 293.649 | 0.85377 | 79.7013 | 122.639 | 1028.03 | 20.404 |
| 7 | 12.8065 | 274.563 | 1.4539 | 79.7013 | 143.325 | 1118.46 | 2.474 |
| 8 | 13.0387 | 296.043 | 0.47066 | 116.145 | 107.02 | 1229.87 | 39.012 |
| 9 | 12.26 | 284.161 | 0.85377 | 56.0876 | 128.375 | 1145.06 | 1.883 |
| 10 | 14.3679 | 294.158 | 0.5823 | 124.2 | 133.703 | 1145.06 | 21.248 |
| 11 | 15.2214 | 296.043 | 0.91133 | 56.0876 | 136.496 | 1480.74 | 35.774 |
| 12 | 15.8333 | 280.974 | 1.4979 | 88.2449 | 158.098 | 1190.42 | 1.691 |
| 13 | 16.637 | 200.526 | 0.37344 | 65.9835 | 136.496 | 1118.42 | 65.724 |

UNINET models an empirical multivariate distribution with a normal copula. The marginal distributions are taken directly from the data. The adequacy of such a model is evaluated statistically, by comparing an appropriate multivariate dependence measure for the empirical and the modeled distributions. The appropriate measure in this case is the determinant of the rank correlation matrix, which is closely related to the mutual information ([Hanea et al 2008](#)). We compare the determinant of the empirical rank correlation matrix (DER) with the determinant of the rank correlation matrix under the assumption that the rank dependencies are represented by the joint normal copula (DNR). If the hypothesis that the dependences come from a joint normal copula is not rejected at an appropriate significance level, then the UNINET model is appropriate. A BBN can be built by including the most important dependence relations. The adequacy of the BBN, relative to the 'saturated' graph, is again evaluated by comparing determinant of the rank correlation matrix based on the BBN (DBBNR) with the DNR. When a suitable BBN model is found, it can be analytically conditionalized from the empirical distribution, or alternatively the BBN density can be used to generate as many samples as desired, for further Monte Carlo post processing.

## II)      Uncertainty Analysis / Stochastic Modeling

When UNINET operates in the uncertainty analysis  mode, the user defines random variables, probabilistic dependences and functional relations. This section the [simple model](#) shown above.

### 1)      Getting started

On opening UNINET for uncertainty analysis/stochastic modeling, and choosing NEW, the user chooses the option USER DEFINED RANDOM VARIABLE MODEL.
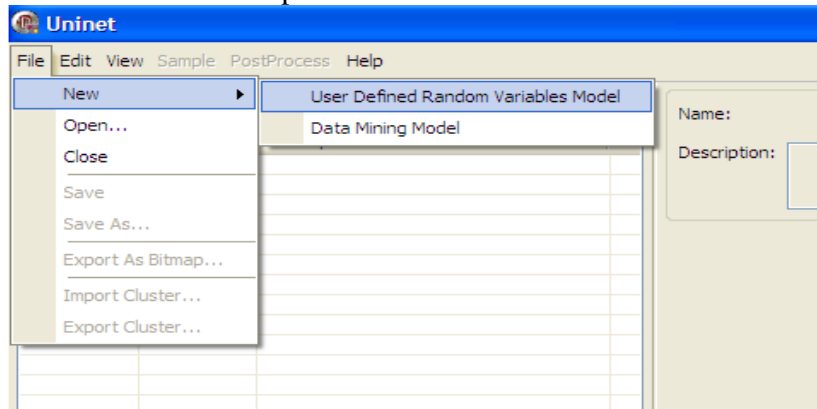


When a new case is started, and saved as NAME, UNINET creates a directory called NAME and creates three files in this directory. The file NAME.bbn is the graphical model, NAME.rvinfo contains information used in specifying the random variables, and the file NAME.tps contains sample values for the random variables (NAME.relinf acts as a header for NAME.tps)..

## 2) Specifying variables and distribution

Hit the ADD button in the random variable view to enter a new random variable. The default name is Vi, and may be edited. A description field is useful for recording an operational definition of each random variable, which can be called from the BBN. Discrete variables can be specified by entering values and probabilities. Discrete variables should have an ordinal meaning; the *meaning_of_life* shown below is measured on a 9 point scale. Dependences are correlations between rankings, so the ordering of values must be meaningful.
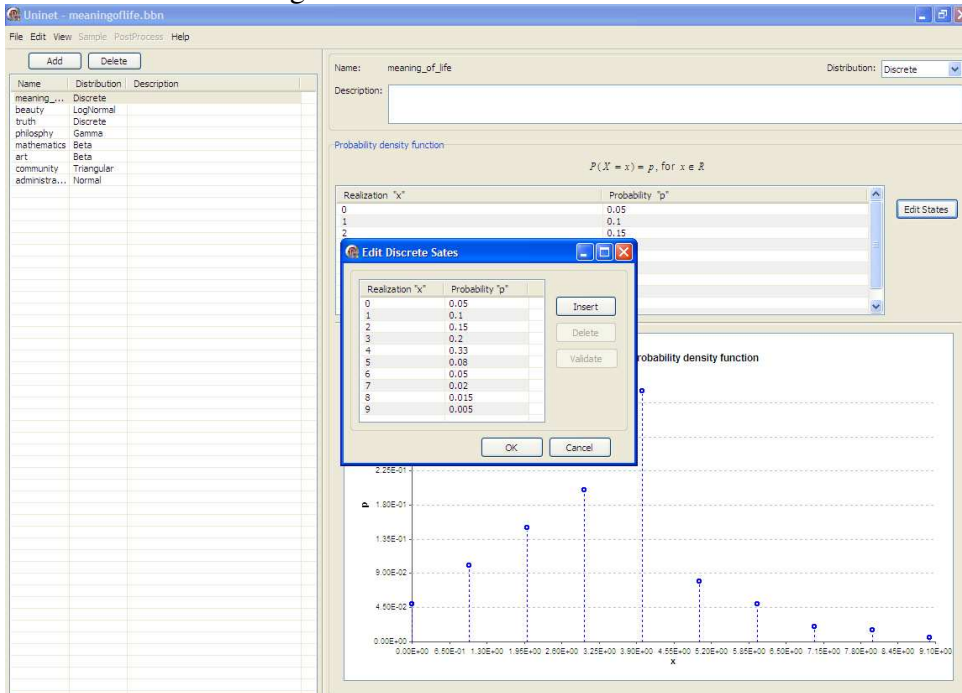


Distributions can also be assigned from file. The assignment panel is shown below.
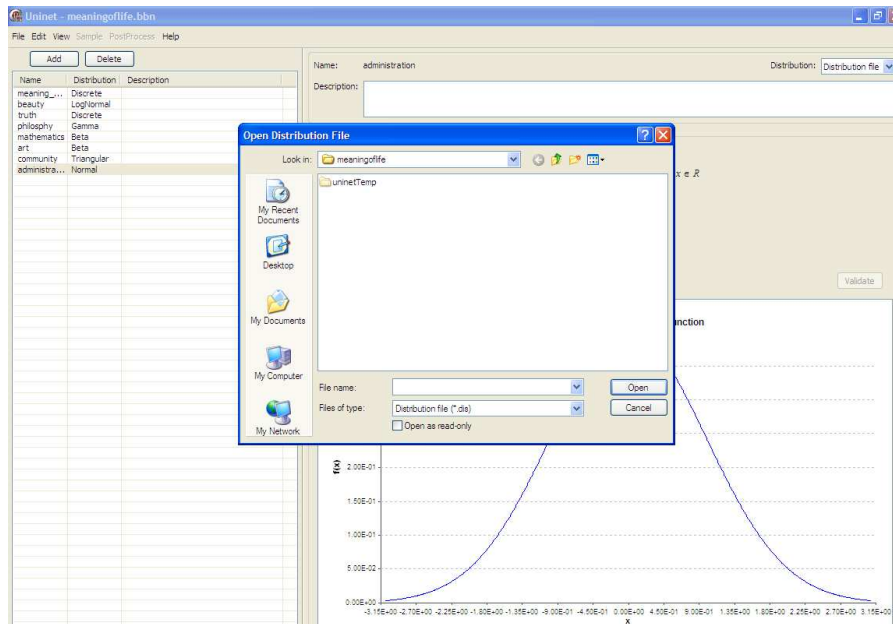
Distribution files must have the extension "*.dis". They are ascii files with two columns. The first column runs from 0 to 1, not including the endpoints; the second column gives the (cumulative) probability of being less than or equal to the value in the first column. A dis file fragment is shown below. Dis files may be created in UNINET, or in the uncertainty analysis package UNICORN, or in some third party software.



## 3)     Building a BBN

After the variables have been entered and assigned distributions, the user may switch to the Bayesian Belief Net view from the view menu. Initially all variables appear on the list to the right, and the graphical screen is empty.
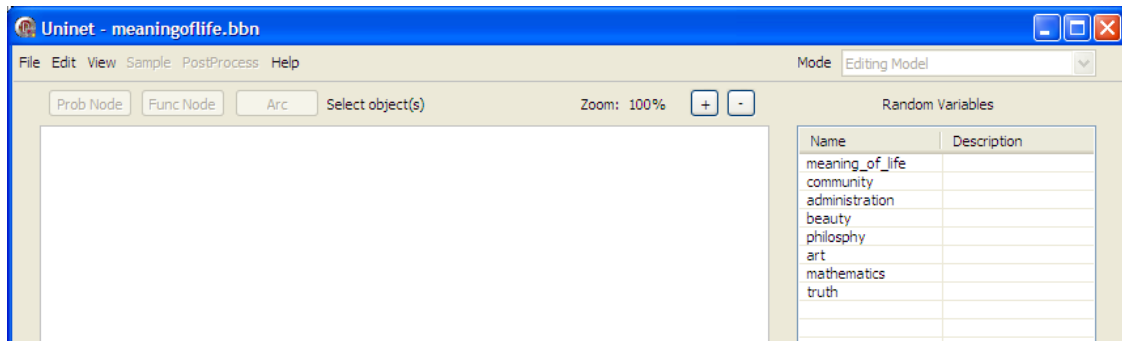
## a)    Input probabilistic nodes

Variables are dragged onto the graphical screen, where they become ellipses. The editing mode becomes active, as indicated in the mode field in the upper right corner. When all variables have been dragged, it is convenient to remove the variable list by unchecking the appropriate option in the View menu:
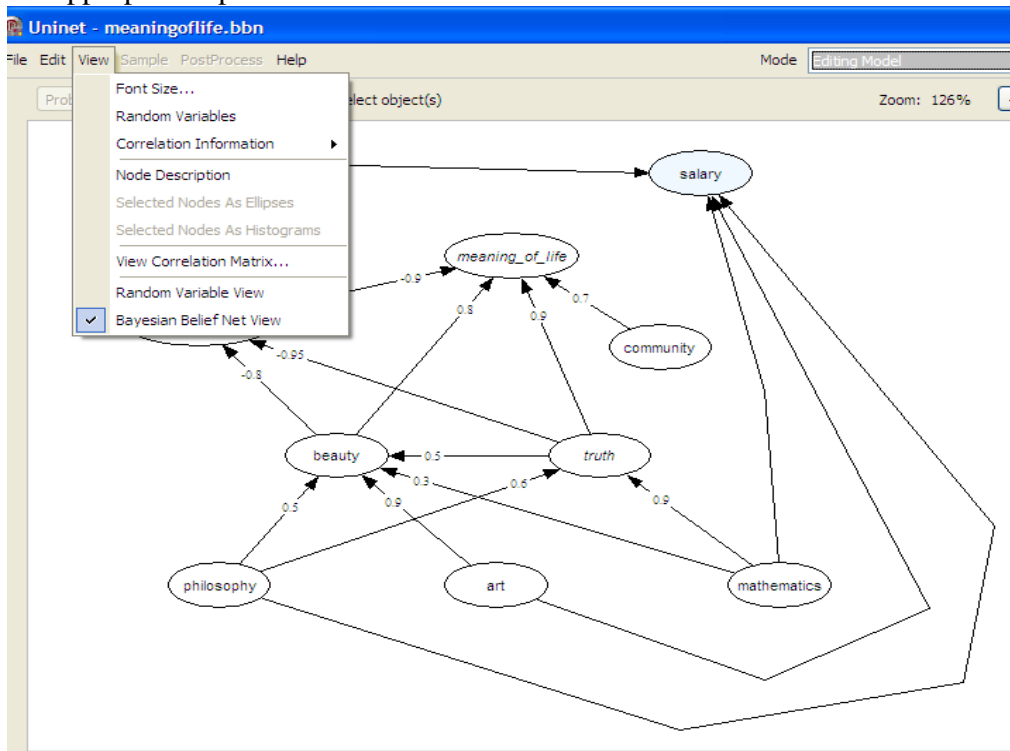


## b)       Input arcs

Arcs are added to the graph by first clicking the ADD ARCS button, then clicking on the parent and child nodes (in that order). The arc is drawn. If a new arc creates a directed cycle, the user is alerted, as shown below:
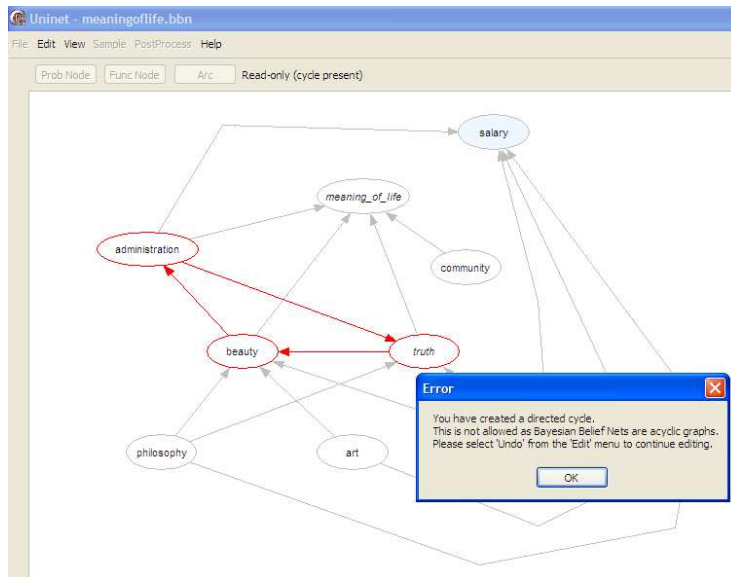
The user is instructed to go to the edit menu and undo the cycle. Arcs can be rubber-banded by clicking and dragging anywhere on the arc:
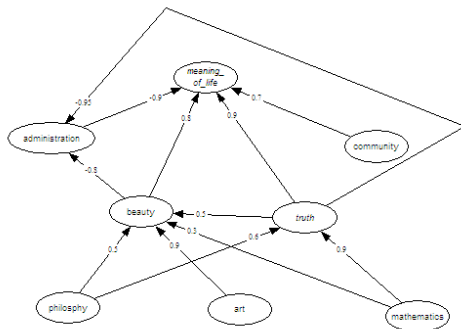


Right clicking on an arc – after unclicking the ADD ARC button, gives the options to straighten, reverse or remove.

### c) Input conditional rank correlations

Influences are represented as conditional rank correlations. After adding arcs and unclicking the ADD ARC button, right clicking on a node allows selecting the DEPENDENCE INFO option:
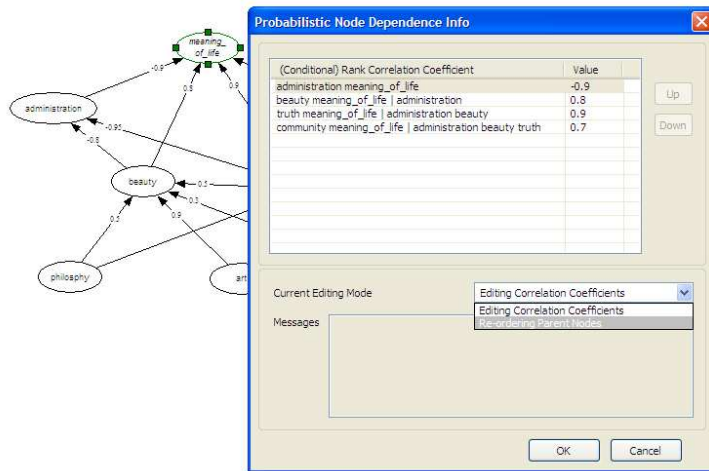
The node *meaning_of_life* has four parents, in sequence: *administration, beauty, truth* and *community.* The conditional rank correlations are sequenced in the order in which the arcs were added. The ordering can be changed from the checklist box. The sequence shown above means that the first rank correlation is *(Meaning_of_life, administration),* the second is *(Meaning_of_life, beauty | administration);* etc. Sequencing of conditional rank correlations assures that these correlations are algebraically independent: any numbers between -1 and 1 may be assigned and will be consistent, regardless of any other correlations specified elsewhere in the BBN. Moreover, the conditional rank correlations contribute in a transparent way to the Determinant of the BBN Rank correlation matrix (DBBNR). Indeed, the following approximation holds (Kurowicka and Cooke, 2006-A):

$$\text{DBBNR} = \Pi_{\text{all conditional rank correlations } \rho_r \text{ in BBN}} (1 - \rho_r^2).$$

To interpret this formula, note that if there were no arcs, then DBBNR = 1. The rank correlation ρ*(Meaning_of_life, administration)=-0.9* reduces the DBBNR by the factor $(1 - 0.81) = 0.18$. The conditional rank correlation ρ*(Meaning_of_life, beauty | administration)=0.8* reduces the DBBNR by $(1-0.64) = 0.36$, and so on.

### d)   Elicitation

Conditional rank correlations are difficult to assess directly. Instead these can be assessed by asking for (conditional) exceedence probabilities. (Morales et al, 2007, Cooke and Goossens 2000). Roughly, one asks experts, "*Suppose Parent 1 were observed to be above its median, what is the probability that Child is also above its median. Suppose that Parent 1 AND Parent 2 are both observed to be above their medians, what is the probability that Child is also above its median, ETC".*

### e)   Input functional nodes

Functional nodes are added by clicking the Func Node button, then clicking (repeatedly for multiple nodes) on the graphical panel. After creating functional nodes, arcs to a functional node are added with the ADD ARC button. The parents of a functional node are the arguments, of which the functional node is a function. Functional nodes can have

probabilistic or functional parents, but they cannot have probabilistic children. Functional nodes appear with a bluish aspect:
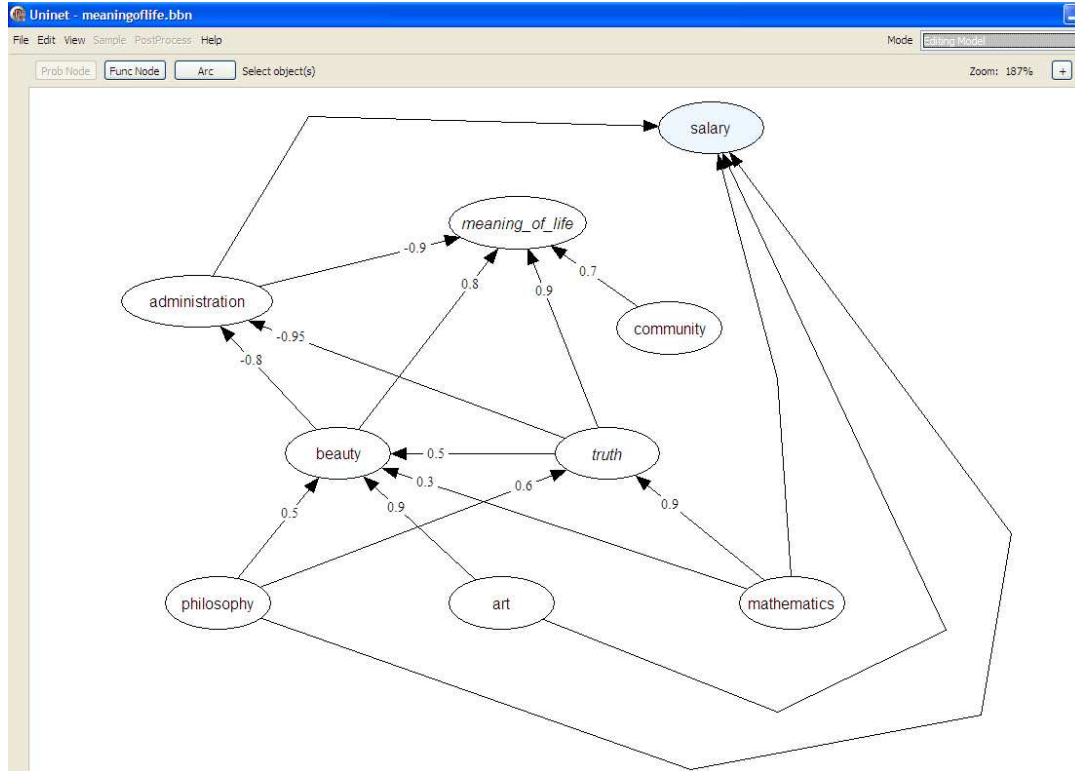


The functional relation is specified by right clicking on the functional node and selecting DEPENDENCE INFO. The following screenshot shows the functional relation between *salary* and *administration, philosophy, mathematics* and *art:*

## Formula for Functional Node "salary"

(pi*administration)^2-(philosphy*mathematics*art)

Verify

**Nodes** | Functions | Operators | Constants

| | | | |
|---|---|---|---|
| administration | | | |
| philosophy | | | |
| mathematics | | | |
| art | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Type: Probabilistic

OK    Cancel
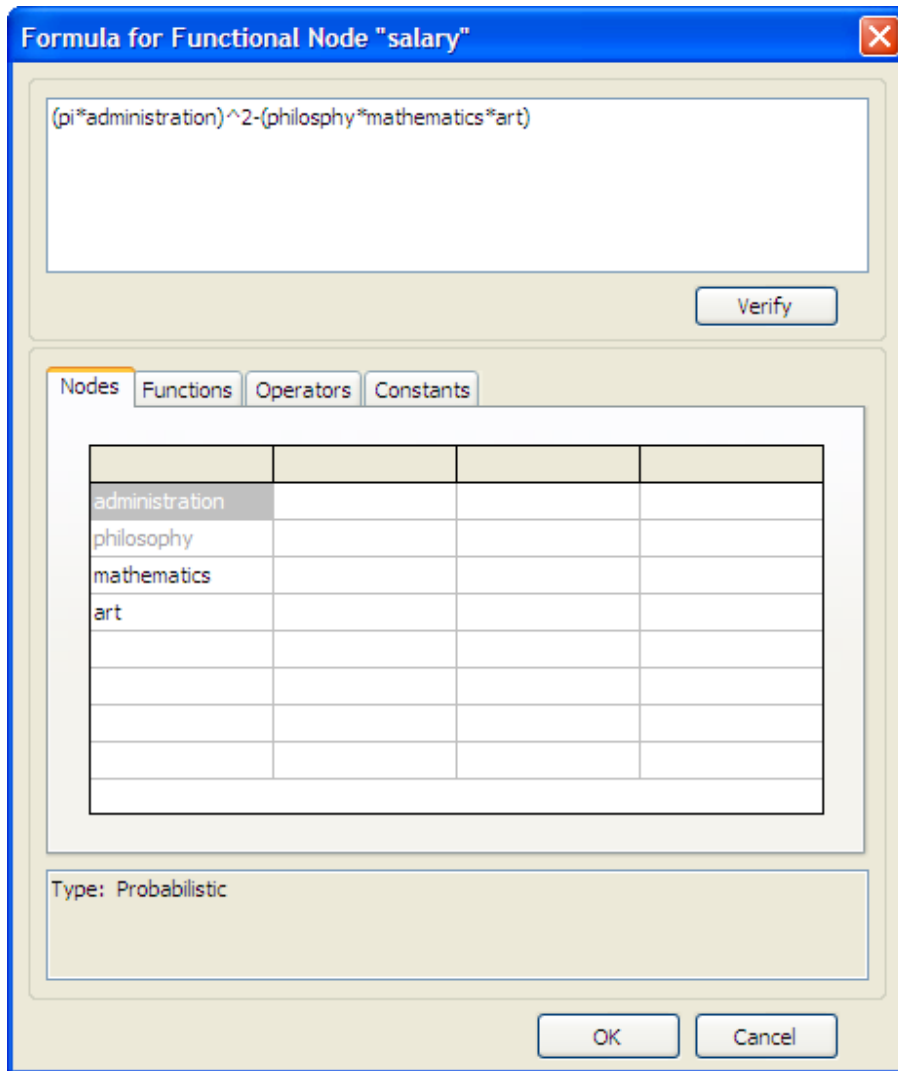
The function parser supports all standard mathematical operations.

## 4)      Configuring the BBN

Several operations are available to enhance the appearance of a BBN.

### a)   Resizing

First click the "+" button in the upper right portion beneath the mode window, then select a set of nodes by boxing them while holding the left click mouse button. Upon release, the selected portion fills the screen:
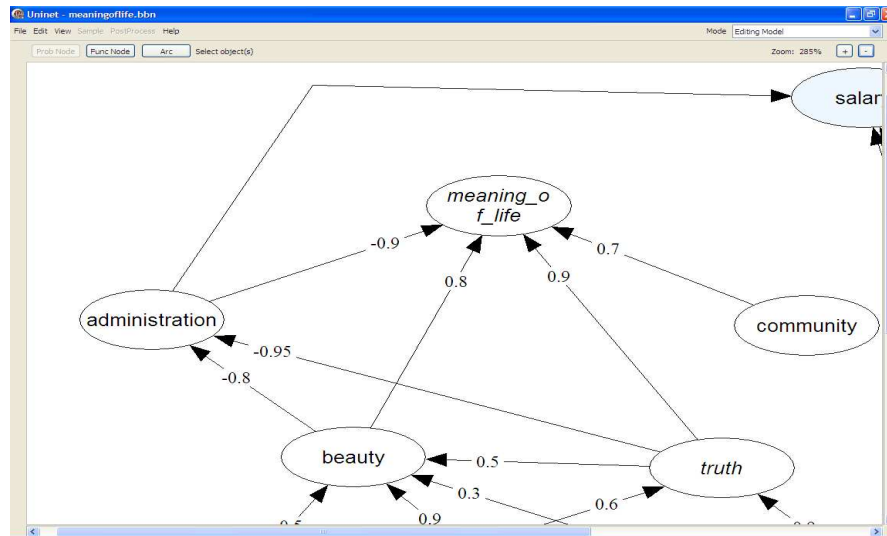
Clicking on the "-" button zooms back out.

## b) Aligning

Selecting a set of nodes and right clicking gives the option of aligning the nodes horizontally or vertically

## c) Viewing nodes

For large BBNs it may be difficult to find a given node. In the EDIT menu, choose FIND NODE, and type the first letters of the node's name. A short list appears. Selecting from this short list zooms to the requested node. Right clicking on a node provides a list of viewing options. The first option "DELETE NODE" , applied to a random variable, removes the node from the BBN and returns it to random variable to the random variable list. The NODE TEXT option enables editing the node text and node description. DEPENDENCE INFO allows entering rank correlations in the case of probabilistic nodes, or functional relations in the case of functional nodes. The final option is VIEWING HISTOGRAMS

## d) Viewing histograms

Viewing histograms is available by right clicking a node; for functional nodes histograms are available after clicking analytical conditioning. The histogram panel is shown below:
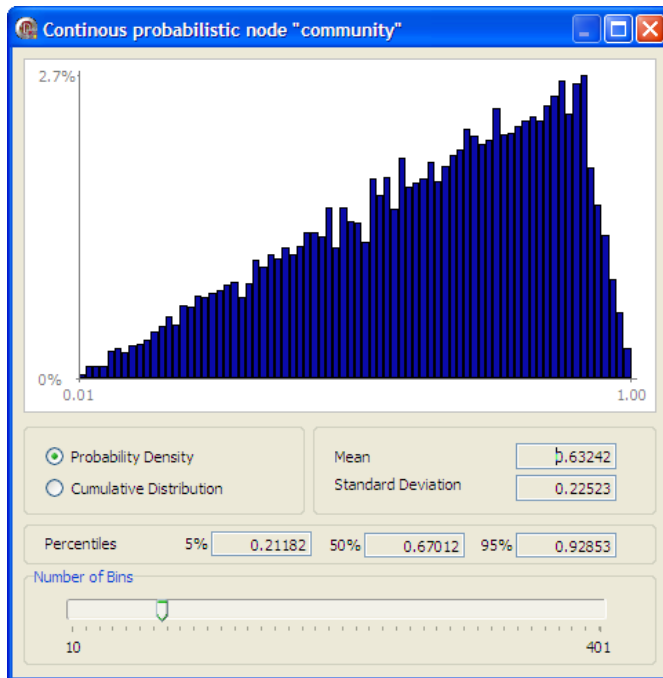
Mean, standard deviation, and 5- 50- and 95 percentiles are shown. The user may switch between the probability density function (strictly speaking a mass function) and the cumulative distribution function. The mean of random variable X is $\mu_X = \Sigma$ $x_i PROB\{X=x_i\}$; where the sum ranges over all possible values $x_i$ of X. The *standard deviation* is the square root of the mean of $(X - \mu_X)^2$. The r-th *percentile,* or *quantile* of X is the least value x for which $PROB\{X \leq x\} \geq r$. The functions may be (un)smoothed by adjusting the slider at the bottom.

### e)    Viewing correlations

Correlations may be viewed from the DEPENDENCE INFO option, or may be made visible in the BBN from the VIEW menu. The correlations may be viewed with or without labels; labels tend to become quite cluttered, even for the toy BBN below:
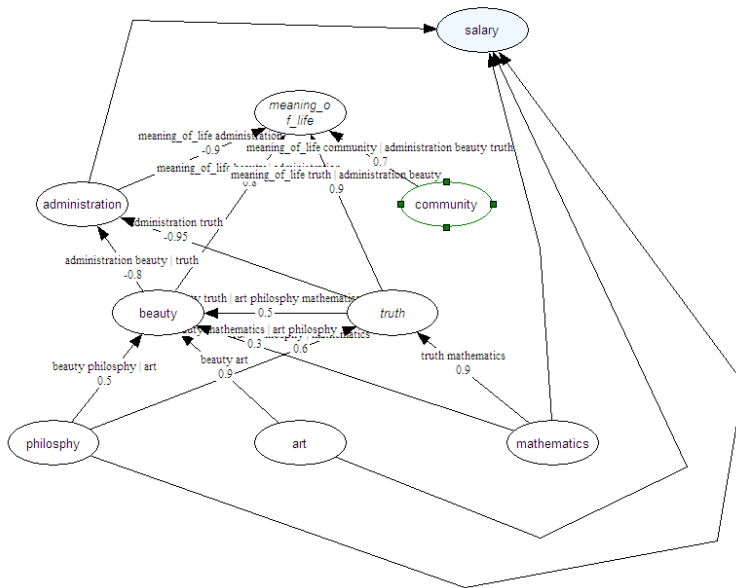
# 5) Editing the BBN

## a) Editing distributions

To edit a distribution the user returns to the random variable view from the VIEW menu. Selecting the random variable, its distributional parameters or distribution type may be changed. When returning to the BBN view, the editing mode must be active to effectuate the change.

## b) Editing node text

Node text and node descriptions may be edited either in the BBN view by right clicking the node, or in the random variable view. When editing text from the BBN, the editing mode must be active; changes are automatically made in the random variable view.

## c) Editing formulas

Editing formulas is performed by right clicking a functional node while in the editing mode. Note the four tabs in the functional specification panel. The panel "Operators" contains logical as well as arithmetic operators:

**Formula for Functional Node "salary"**

(pi*administration)^2-(philosphy*mathematics*art)

Verify

| Nodes | Functions | Operators | Constants |

| + | > | ! |
| + | < | |
| - | >= | |
| - | <= | |
| * | != | |
| / | == | |
| ^ | & | |
| % | | | |

Expression: x == y
Description: "Equal" logical operator. Returns one if x is equal to y. Returns zero otherwise.

OK     Cancel
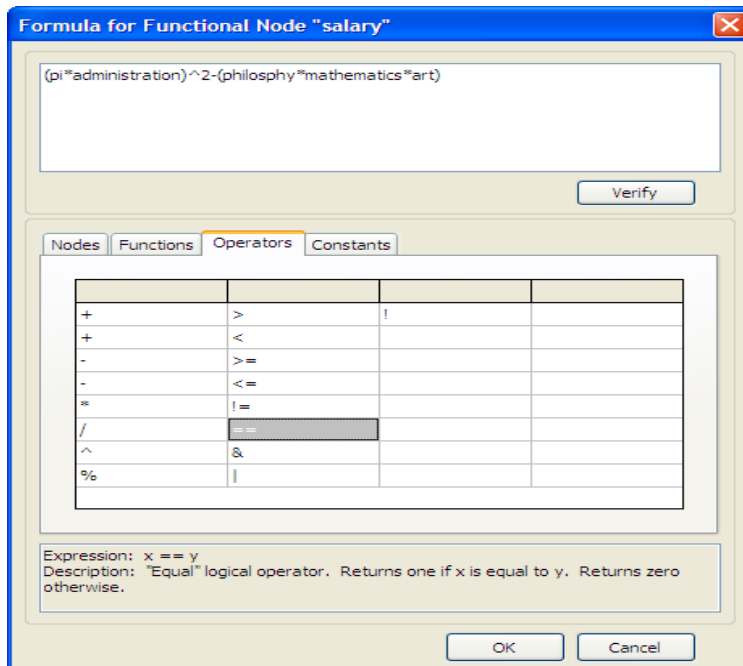
### d) Removing / deleting variables

Random variables may be removed only from the View Random Variables panel. Select a variable and hit the delete button. The variable, and all attendant information regarding the variable is permanently removed. When in the editing mode, variables may be deleted from the BBN by choosing this option after right clicking on the variable. Deleted variables are not removed, but return to the variable list in the BBN View.

### e) Moving nodes and arcs

When on the editing mode, arcs can be bent and straightened and reversed by right clicking on the arc. A node is selected by clicking on it in the editing mode. By dragging one of the hotspots, the node an be re-sized. By clicking on the center of the node and dragging, the node can be moved. A group of nodes can be selected by left clicking and inclosing it in a box. When a set of objects is selected, the set can be moved by dragging any one of the objects.

### f) Clustering nodes

A subset of nodes can be selected and exported as a cluster, using the EXPORT CLUSTER option under the FILE menu. The screenshot below shows that *Meaning_of_life* and its ancestors, except *Administration* have been selected and exported, given the name *mol_cluster*.
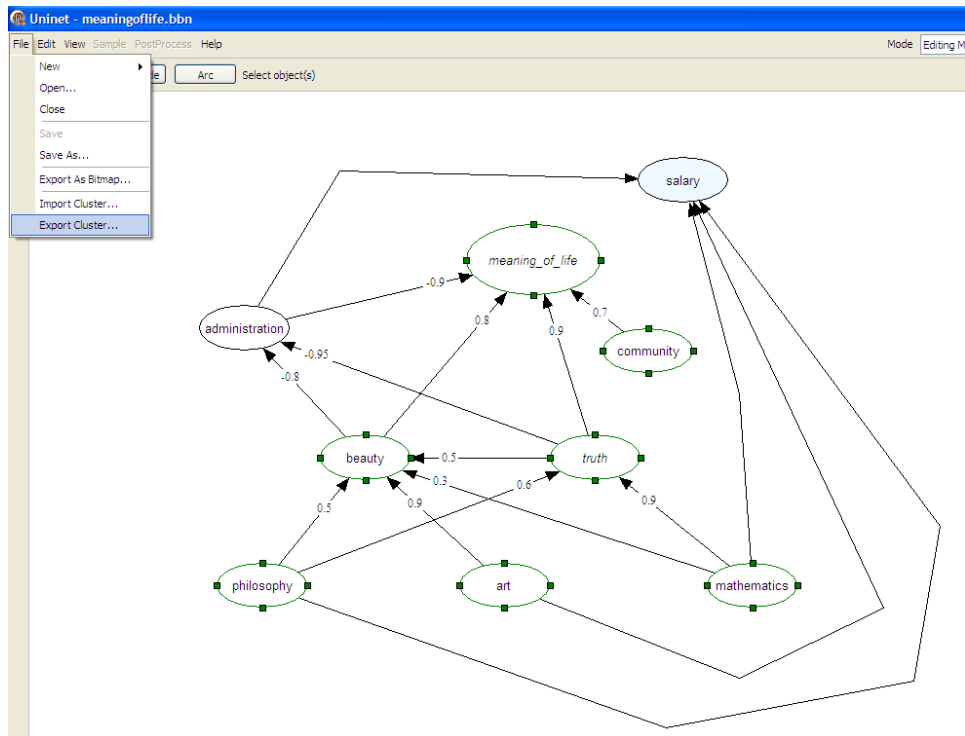
After this cluster is exported it can be imported into other BBNs, or treated as a stand alone model. When using the IMPORT CLUSTER option, UNINET checks the names for conflicts; names in the imported cluster may not duplicate names in the active BBN. If that should arise, UNINET provides the name change screen shown below:
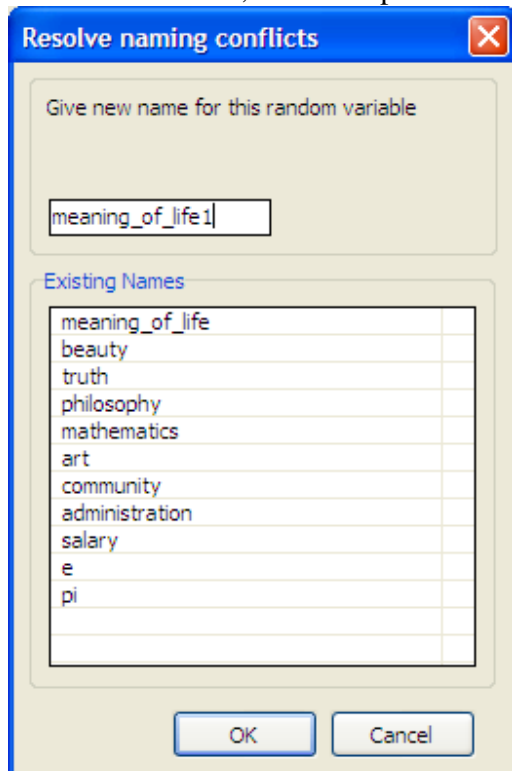
When all conflicts are removed, the cluster of nodes may be imported into the active BBN, with the marginal distributions, node descriptions and dependence information taken from the cluster.

This feature is very handy when building large BBNs with modular pieces.
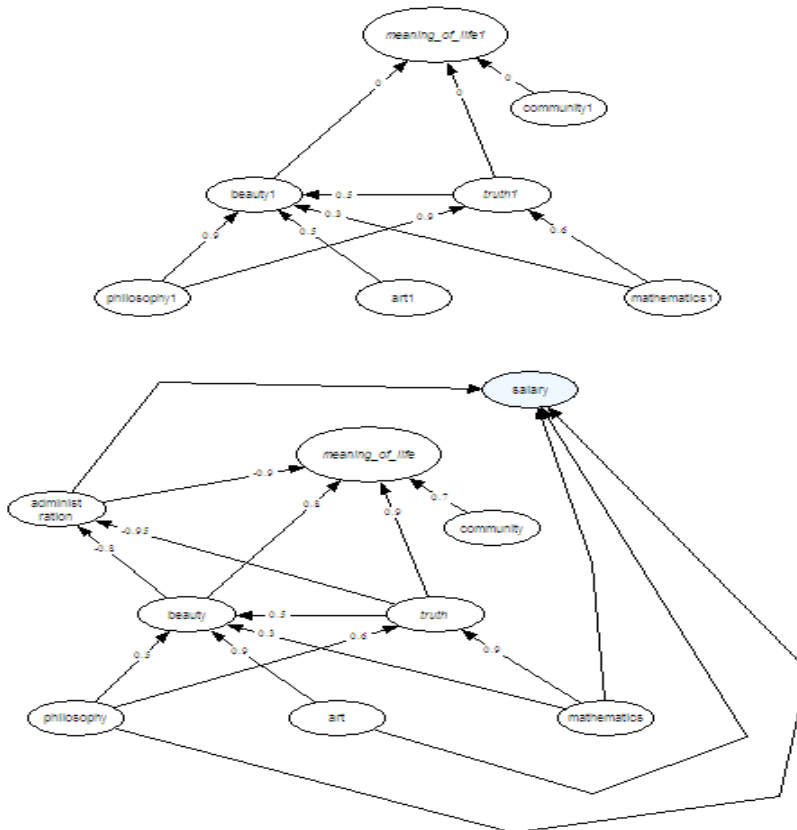


### g) Setting options

The OPTIONS screen under the EDIT menu may be used to control some program parameters, including the prompt for function names, the shadows for conditional nodes and the number of default samples in ANALYTICAL CONDITIONING.
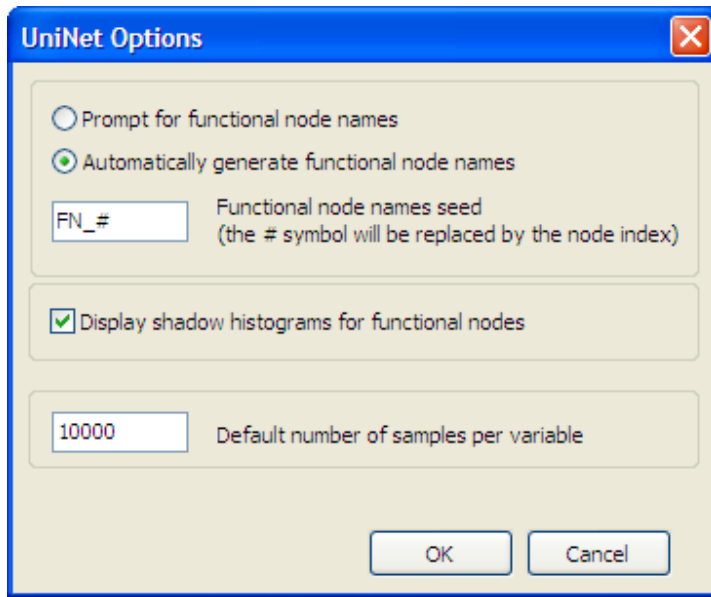
## 6) Analytic conditioning

The distinctive feature of UNINET is its analytical conditioning capability. BBNs with thousands of probabilistic nodes can be conditionalized on arbitrary values of random variables, whereby the conditional distribution is computed in a few minutes. This is only possible with the joint normal copula. Random variables are treated as transformations of joint normal variables. If $Z = (Z_1, \ldots Z_n)$ denotes a joint normal distribution of normal variables $Z_i$ with mean zero and unit variance, then random variables $X_i, \ldots X_n$ are written as

$$(X_1, \ldots X_n) = ( F_1^{-1}\Phi(Z_1), \ldots F_n^{-1}\Phi(Z_n) )$$

where $F_i$ is the cumulative distribution function of random variable $X_i$, and $\Phi$ is the cumulative distribution function of the standard normal variable. If $F_i$ is continuous invertible, this transformation is rank preserving, the rank correlations of $X_i, X_j$ will be the same as that of $Z_i, Z_j$. Complications arise for discrete variables (see Hanea, 2008) The rank correlation of normal variables is computed from the product moment correlation with the Pearson transformation. For details see (Kurowicka and Cooke, 2006-A)

Conditioning on the value $X_i = x$ entails conditioning Z on $Z_i = \Phi^{-1}(F_i(x))$. Letting $Z(X_i=x)$ denote this conditional distribution, the *conditional distribution of $X_1, \ldots X_n$, when $X_1, \ldots X_n$ are joined by the joint normal copula*, is given by

$$( X_1, \ldots X_n \mid X_i=x ) = F_1^{-1}\Phi(Z_1 (X_i=x) ), \ldots F_n^{-1}\Phi(Z_n (X_i=x)).$$

Note that $\Phi$ is the standard normal cumulative distribution function, but $Z_j(X_i=x)$ is not in general a standard normal variable, as its mean and variance are altered by the conditioning. For this reason, the *marginal distribution* of $X_j$ given $X_i = x$ is no longer equal to the unconditional marginal $F_1^{-1}\Phi(Z_j)$, but has become $F_1^{-1}\Phi(Z_j (X_i=x) )$.

### a)   Conditioning and unconditioning

When the analytical conditioning mode is entered, UNINET computes the distributions of all functional nodes and builds a joint density for all variables. Analytical conditioning may be performed on probabilistic nodes. After right clicking on a node, the CONDITIONALIZE option may be chosen.

### b)   Histogram view for conditional distributions

After selecting conditional values and hitting UPDATE, the conditional distribution of all variables is computed. To quickly view the results of conditionalization, select all nodes and choose the histogram option from the view menu (or hit hotkey *ctl + a*). The figure below shows the result of conditionalizing on high values of *philosophy* and *mathematics.* The original unconditional distributions are shown in gray, the new conditional distributions are shown in black. Some degree of rescaling is applied to compare the conditional and unconditional distributions; the histogram view should be regarded only as indicative, for better information see UNIGRAPH and UNISENS. Beneath each histogram, the mean and standard deviation of the conditional distribution are shown. If we right click on a node and select HISTOGRAM, only the current conditional distribution is seen.
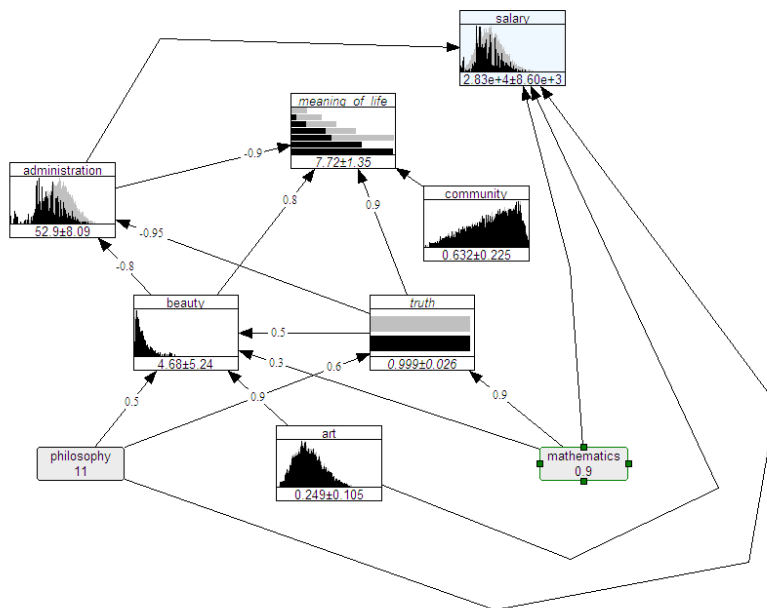


Note that histograms for variables with discrete distributions (*meaning_of_life, truth*) are displayed vertically, whereas others are displayed horizontally.

## 7)     Sampling a belief net

Analytical conditioning involves sampling the BBN; this is done automatically whenever the ANALYTICAL CONDITIONING mode is entered. Analytical conditioning draws 32,000 samples.  Under the EDIT menu, OPTIONS, the default number of samples used for creating the marginal distributions can be set. Analytical conditioning applies only to

probabilistic nodes. If we wish to conditionalize on (interval) values of functional nodes, we must use Sample based conditioning. In the ANALYTICAL CONDITIONING mode, it is also possible to sample a BBN with user chosen parameters. For sample based conditioning, the SAMPLING MODEL mode must be chosen.

## a) Sampling in ANALYTICAL CONDITIONING mode

While in the ANALYTICAL CONDITIONING mode, the user may sample a (conditional) distribution by choosing the SAMPLE menu and selecting SAMPLE BAYESIAN BELIEF NET. The user must choose one or more file type and file names. *∗.SAM* and ∗.SAE are ascii files, the former is a simple list, the latter is an XL compliant comma separated file. When conditionalization is active, the conditional sample is written to file. These may be viewed and analyzed from the POST-PROCESSING menu by choosing UNIGRAPH or UNISENS.

## b) Sample based conditioning

Choosing the SAMPLING MODEL mode enables sample based conditioning. This can be applied to both probabilistic and functional nodes. Whereas analytical conditioning conditionalizes on single values of the conditioning (probabilistic) variables, sample based conditioning conditionalizes on intervals. After sampling, choosing the sample menu displays a new set of options:
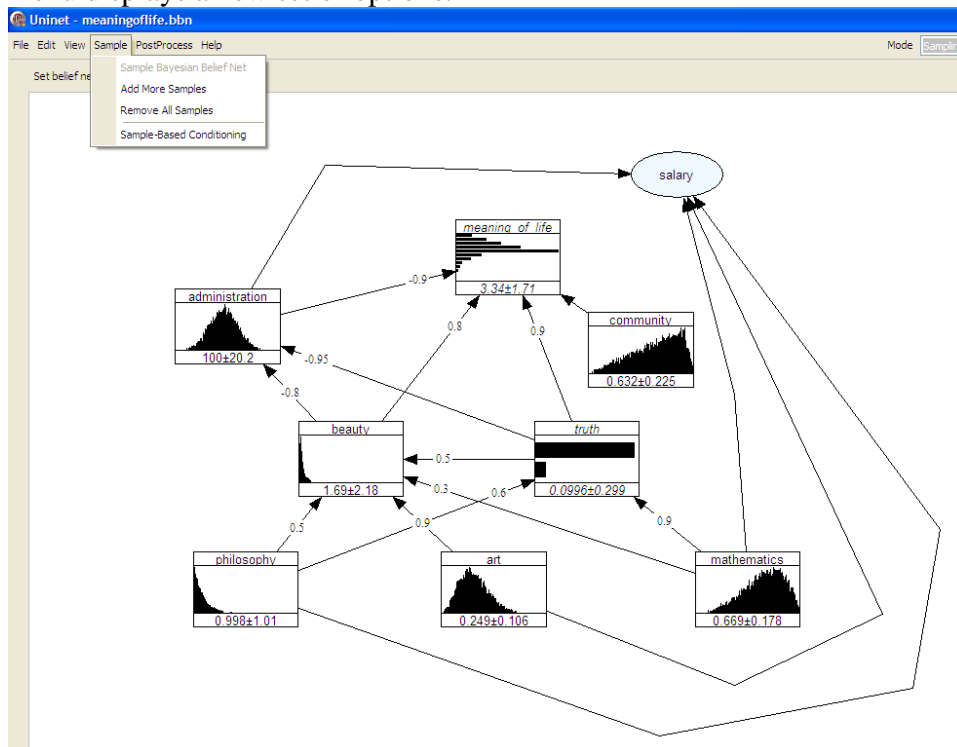


Choosing SAMPLE BASED CONDITIONING, the user is asked to specify input and output nodes. Input nodes are those on which we wish to conditionalize. Output nodes are those for which we wish to view the effects of conditionalization. If we choose to

conditionalize on the functional node *salary*, we may choose an interval; in the panel below the interval [250,000, 323,000] is selected.
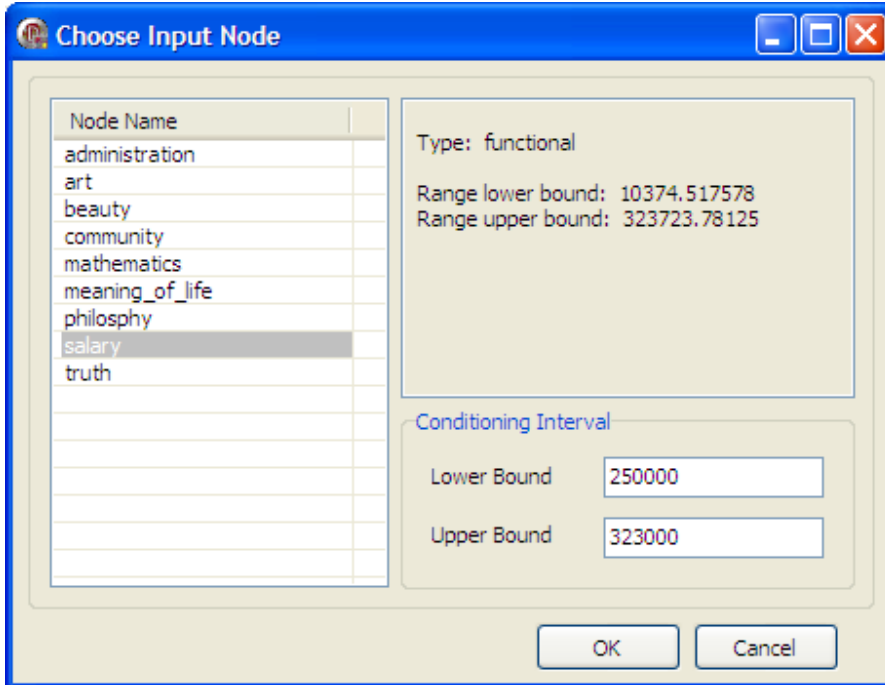


Choosing a number of output variables, we can view the effect of conditionalization by comparing the conditional and unconditional means and standard deviations of the output variables:
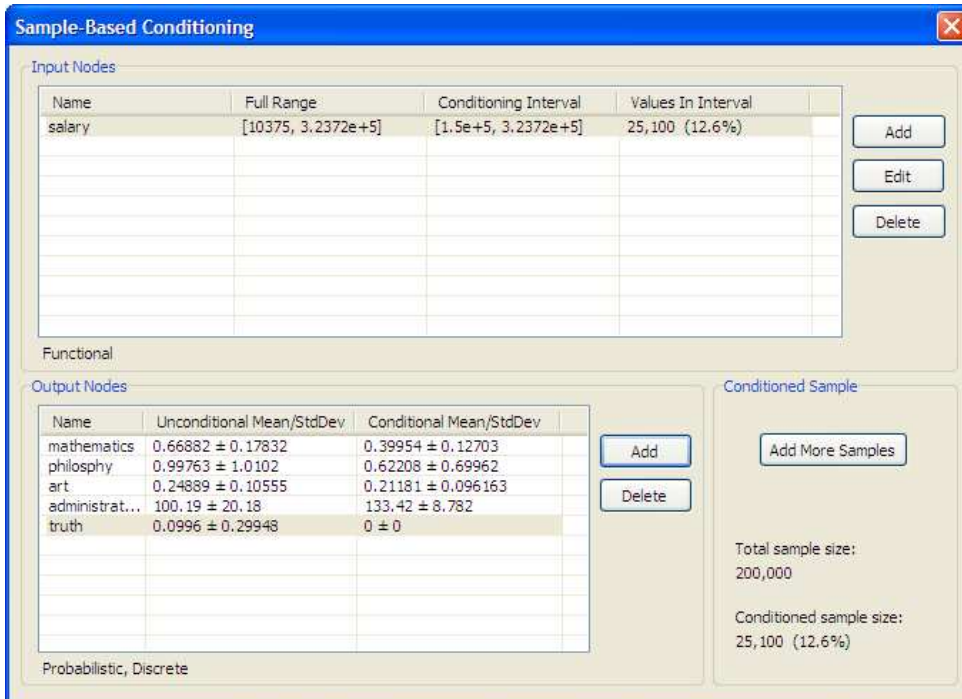
Notice that *Truth* is a discrete variable with two possible values, 0 and 1. Unconditionally the probability that *truth = 1* is 0.1. In this case the mean is the probability of the value 1. After conditionalizing on a very high salary, the conditional mean of *truth* is effectively zero. The conditioned sample size is reported as 25,100 samples (12.6%). If we wish to conditionalize on more variables, or to obtain statistically more robust results, we must increase the number of samples. This will arise if we wish to conditionalize on more input variables and/or choose narrower conditioning intervals. If we add the conditioning variable *meaning_of_life*, conditionalized on the interval [4,9] we must drastically increase the sample size, as shown below:

**Sample-Based Conditioning**

**Input Nodes**

| Name | Full Range | Conditioning Interval | Values In Interval |
|---|---|---|---|
| salary | [10373, 3.2372e+5] | [1.5e+5, 3.2372e+5] | 903,600 (12.6%) |
| meaning_of_life | [0, 9] | [4, 9] | 3,589,200 (49.9%) |

Add
Edit
Delete

Probabilistic, Discrete

**Output Nodes**

| Name | Unconditional Mean/StdDev | Conditional Mean/StdDev |
|---|---|---|
| mathematics | 0.66882 ± 0.17832 | 0.37391 ± 0.10571 |
| philosphy | 0.99763 ± 1.0102 | 0.95901 ± 0.89712 |
| art | 0.24889 ± 0.10555 | 0.47063 ± 0.094264 |
| administrat... | 100.19 ± 20.18 | 125.61 ± 2.2614 |
| truth | 0.0996 ± 0.29948 | 0 ± 0 |

Add
Delete

**Conditioned Sample**

Add More Samples

Total sample size:
7,200,000

Conditioned sample size:
1,183 (0.1%)

With 7,200,000 samples (which takes about 1 minute) there are only 1,183 samples falling within the conditioning interval of both variables. This is because of the strong negative correlation between *salary* and *meaning_of_life*. These features can be further analyzed with the satellite programs UNIGRAPH and UNISENS.

# 8)     Graphics: UNIGRAPH

UNIGRAPH is a stand-alone graphics package developed by the Risk and Environmental Modeling group at the Department of Mathematics at the Delft University of Technology. The main programmer was Valery Kritchallo.  UNIGRAPH is provided with its own help file, to which the user is referred for more detail.

UNIGRAPH supports visual analysis of multivariate distributions via cobweb plots. Variables are shown as vertical lines, each sample is a jagged line, hitting each vertical line at the sample value. 16,000 such jagged lines give a good picture of the joint distribution and dependence relations. When an analytically conditionalized BBN is sampled, the conditional values are retained in the sample.
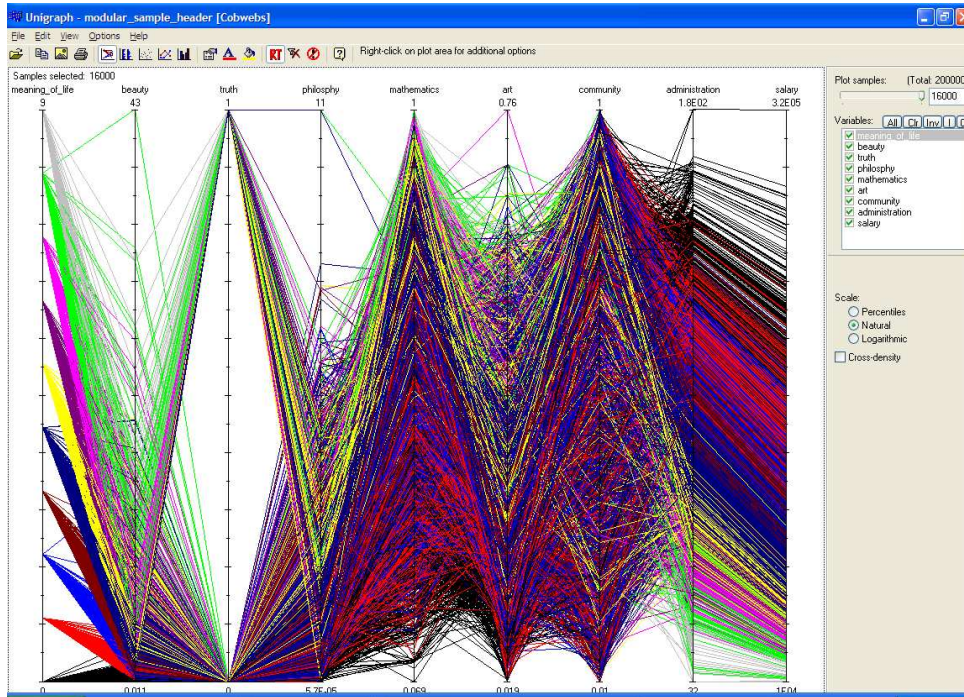
It is evident from the graph that *salary* is mostly determined by *administration*, where as high values of *meaning_of_life* are associated with low salaries. UNIGRAPH supports visual sample based conditioning, although the number of samples is currently restricted to 16,000. The picture below shows the result of conditioning on very high OR very low values of *salary*.

# 9) Sensitivity analysis UNISENS

UNISENS is a stand-alone graphics package developed by the Risk and Environmental Modeling group at the Department of Mathematics at the Delft University of Technology, designed and programmed by Daniel Lewandowski. It is provided with its own help file, to which the user is referred for more detail (see also Lewandowski et al, 2007).

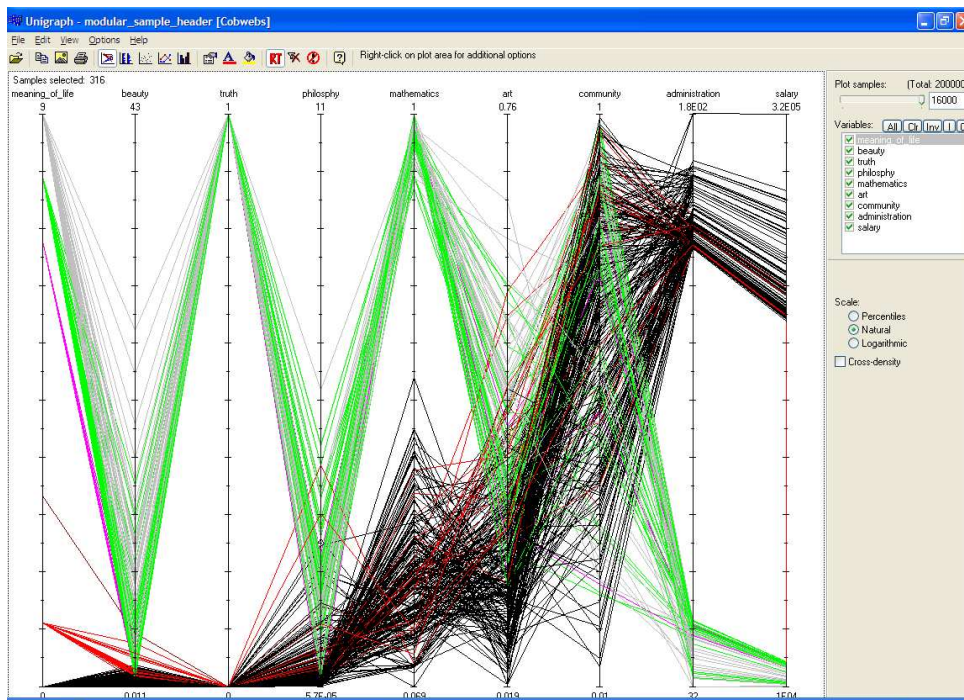After choosing UNISENS from the POST-PROCESS menu, the user selects one or more "predicted variables" and one or more "base variables" which should explain the predicted variables. A number of sensitivity measures may be chosen from TOOLS / OPTIONS, five of which are shown below. By clicking on type of sensitivity measure, variables are ordered according to that measure; the ordering based on product moment correlation is shown below:
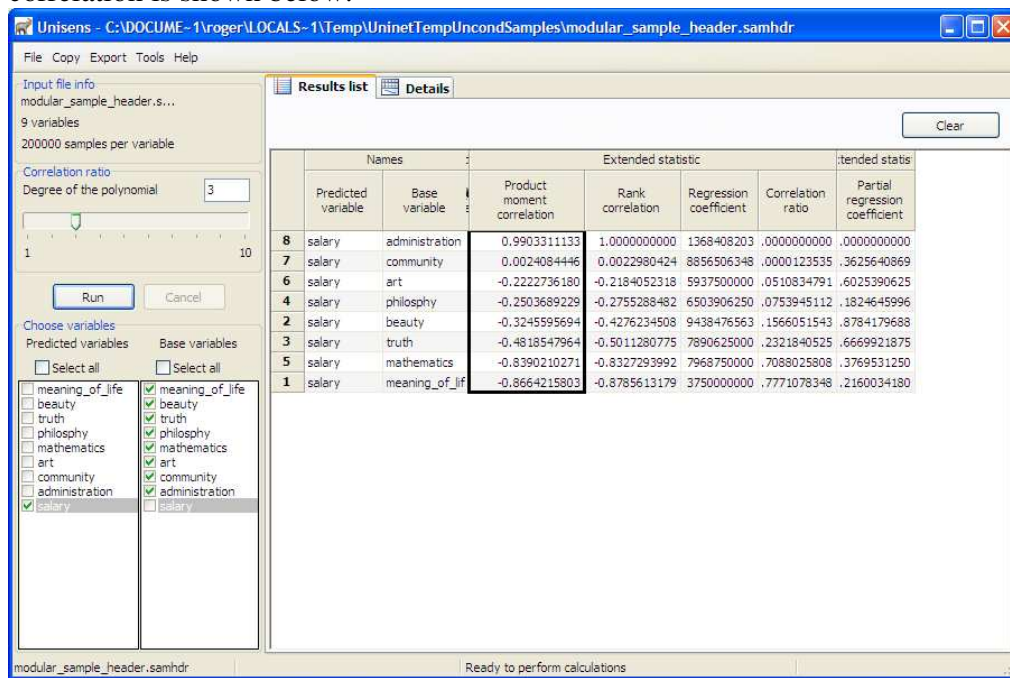


Choosing the DETAILS tab enables a number of additional analyses, some of which are pictured below. In particular, we see the scatter plot between base and predicted variables, as well as a polynomial approximation to the regression function. The correlation matrix or the cofactor matrix may also be viewed. For additional detail the reader is referred to the UNISENS help file, and supporting documentation.
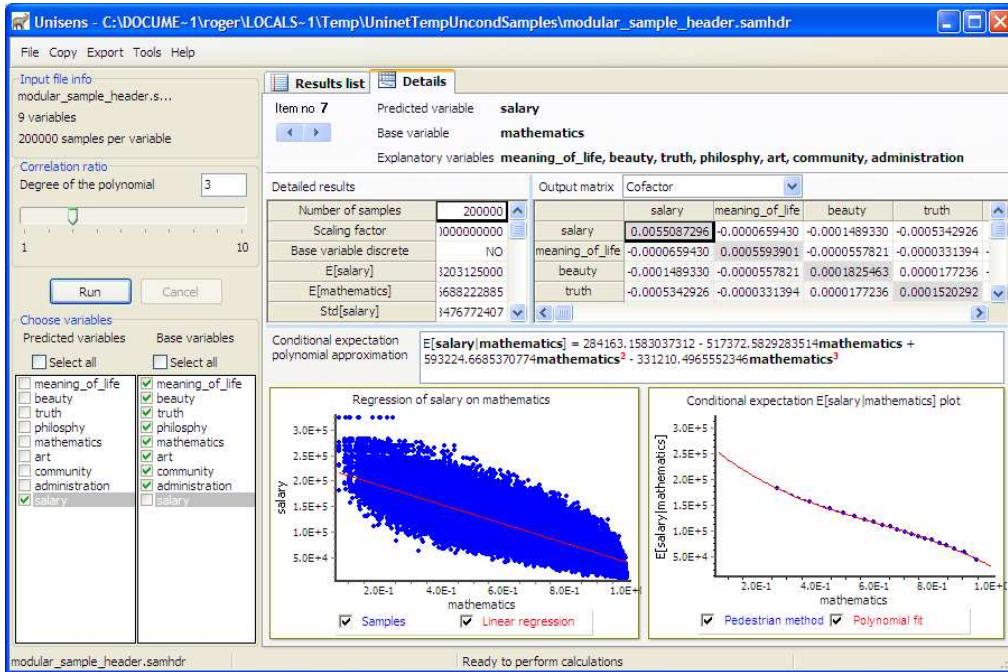
## 10) Report generation

UNINET provides a simple reporting reporting facility. Under FILE, select Write Report
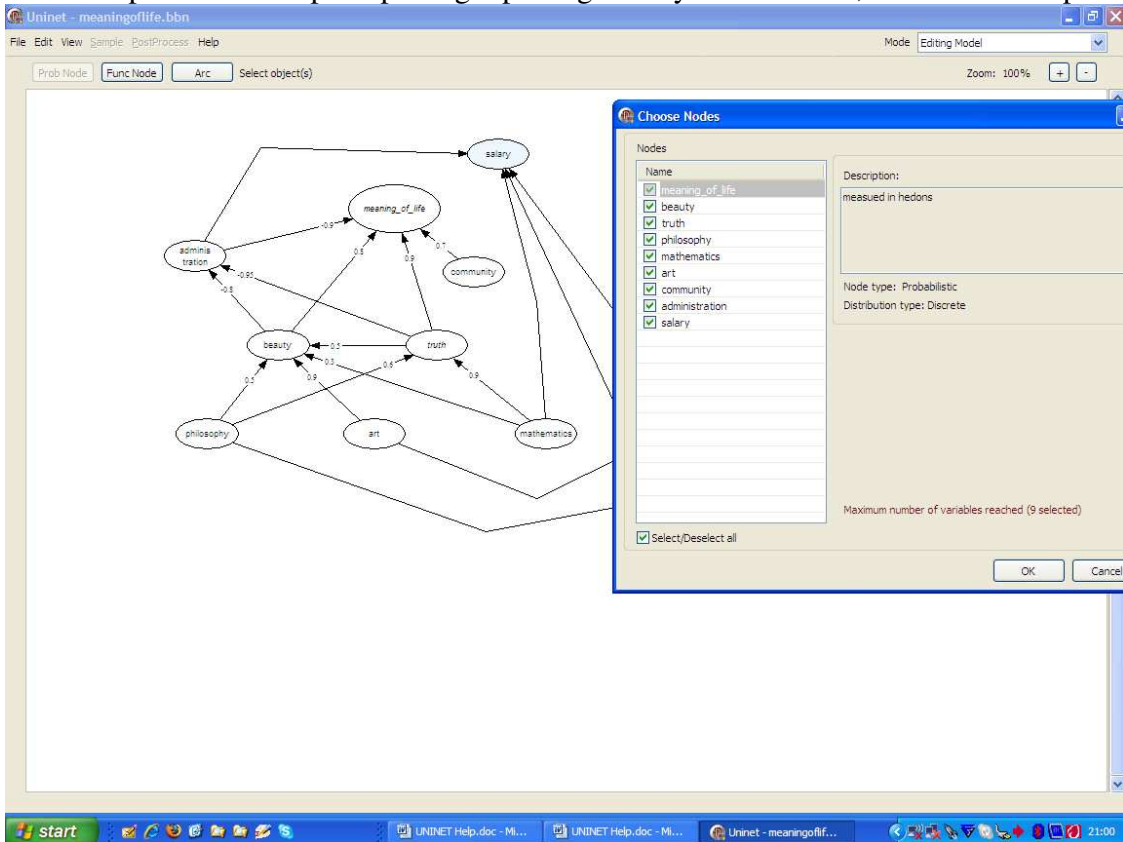
to bring up the list of variables.   After selecting the variables to be reported, an *.rtf file is created with a user chosen name. A fragment of this file is shown below:

## Uninet report

### I.  General information

1.   Bbn file name: C:\NEWUNICORN\Uninet\meaningoflife\meaningoflife.bbn.

2.   Number of samples per variable: 10000.

### II. Nodes information

1.   Node name: "meaning_of_life"
a)   Description: measued in hedons
b)   Dependence info: Parent nodes
-   "administration", probabilistic, r = -0.9, unconditional.
-   "beauty", probabilistic, r = 0.8, conditioning set: "administration".
-   "truth", probabilistic, r = 0.9, conditioning set: "administration", "beauty".
-   "community", probabilistic, r = 0.7, conditioning set: "administration", "beauty", "truth".

To output the correlation matrix, analytical conditioning must first be activated. From the View panel, selecting View Correlation matrix → more → export allows the user to create a *.txt file of the rank correlations.

## 11)    Exporting to NETICA

UNINET can also export a sample file to the commercial BBN package NETICA, by choosing the EXPORT TO NETICA option from the SAMPLE menu. This requires a choice of discretization mode and discretization number. This is useful if particular conditional probabilities are required, which are not adequately captured with conditional rank correlation. NETICA constructs conditional probability tables based on the UNINET sample file and UNINET BBN model. These can then be edited in NETICA to realize specific constraints.

# III)    Ordinal Data Mining / Multivariate Density Modeling

UNINET can be used to build a joint density for ordinal multivariate data, based on the joint normal copula. The qualification "ordinal" indicates that the rank order of the variables' values must be meaningful, as these will be used to model the rank dependence structure with the normal copula.

## 1)    File structure for multivariate ordinal data

Selecting EDIT > NEW> DATA MINING from the random variable view, we can select an sae file. A suitable sae file is the one made in sampling the *meaning_of_life* BBN discussed in the previous section.
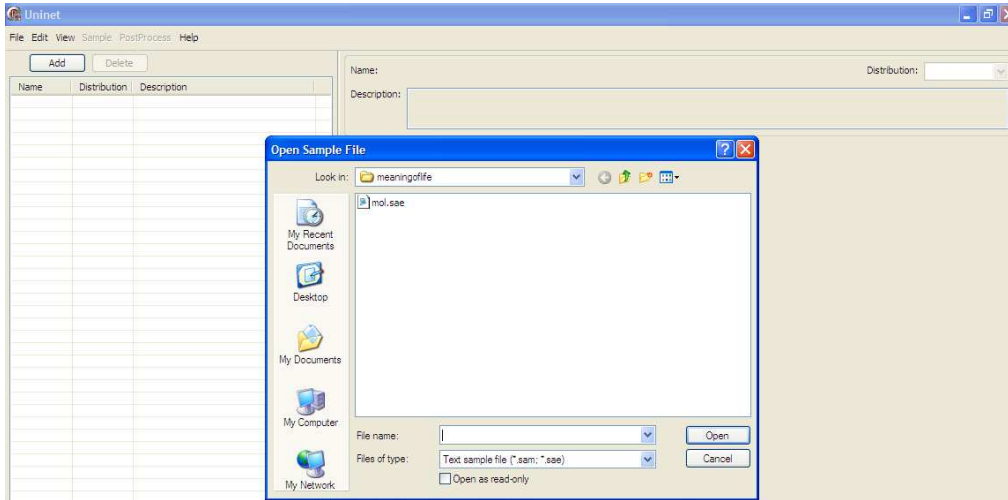
Clicking this file and switching to Bayesian Belief Net view, we find the variables listed as before. This time, however, the distribution is based on the sae file (with 10,000 samples), a fragment of which is shown below:

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | eaning_of_l | beauty | truth | philosphy | nathematic | art | community | dministratic | salary | |
| 2 | 6 | 1.56503 | 1 | 0.650838 | 0.945675 | 0.274479 | 0.646008 | 71.4718 | 50415.9 | |
| 3 | 2 | 0.265362 | 0 | 0.841412 | 0.588778 | 0.109012 | 0.628244 | 101.137 | 100953 | |
| 4 | 4 | 2.082 | 0 | 0.475581 | 0.572034 | 0.336963 | 0.529392 | 97.964 | 94717.9 | |
| 5 | 5 | 4.50105 | 0 | 1.66651 | 0.773188 | 0.3735 | 0.677908 | 85.9074 | 72838.1 | |
| 6 | 3 | 0.294747 | 0 | 0.274917 | 0.825258 | 0.151637 | 0.803872 | 96.9918 | 92847.4 | |
| 7 | 4 | 1.56636 | 0 | 0.543241 | 0.77457 | 0.280896 | 0.679652 | 93.6807 | 86616.2 | |
| 8 | 4 | 0.563141 | 0 | 0.902387 | 0.840847 | 0.176466 | 0.972672 | 95.9816 | 90923.3 | |
| 9 | 4 | 1.23077 | 0 | 1.98316 | 0.603053 | 0.239768 | 0.702372 | 90.4669 | 80775.2 | |
| 10 | 1 | 0.561562 | 0 | 0.239715 | 0.469868 | 0.207577 | 0.616046 | 121.158 | 144878 | |
| 11 | 0 | 0.510736 | 0 | 0.538407 | 0.382102 | 0.224512 | 0.877024 | 140.049 | 193580 | |
| 12 | 2 | 0.892947 | 0 | 0.907287 | 0.475338 | 0.236909 | 0.838283 | 110.65 | 120838 | |
| 13 | 1 | 0.303819 | 0 | 0.171634 | 0.561157 | 0.119359 | 0.408883 | 117.357 | 135931 | |
| 14 | 3 | 8.61782 | 0 | 1.39615 | 0.248467 | 0.518602 | 0.755469 | 127.946 | 161567 | |
| 15 | 3 | 2.31533 | 0 | 1.16348 | 0.477528 | 0.344895 | 0.208297 | 107.677 | 114432 | |
| 16 | 6 | 2.69213 | 0 | 1.7616 | 0.858772 | 0.373276 | 0.246027 | 76.6758 | 58024.5 | |
| 17 | 4 | 1.20469 | 0 | 1.07676 | 0.613183 | 0.275558 | 0.925619 | 103.521 | 105768 | |
| 18 | 2 | 0.542396 | 0 | 0.147546 | 0.708153 | 0.185777 | 0.202143 | 98.3342 | 95435.3 | |
| 19 | 6 | 3.34488 | 0 | 2.6245 | 0.74753 | 0.291046 | 0.569135 | 84.1414 | 69874 | |
| 20 | 2 | 0.934608 | 0 | 0.267261 | 0.48908 | 0.203037 | 0.329115 | 113.544 | 127242 | |
| 21 | 2 | 0.37605 | 0 | 0.555461 | 0.612617 | 0.136327 | 0.610364 | 117.173 | 135505 | |
| 22 | 3 | 1.17115 | 0 | 0.416684 | 0.503683 | 0.272773 | 0.462675 | 102.731 | 104160 | |
| 23 | 2 | 0.531577 | 0 | 1.37466 | 0.618909 | 0.187679 | 0.175416 | 115.006 | 130540 | |
| 24 | 2 | 0.225622 | 0 | 0.535254 | 0.768038 | 0.156059 | 0.628512 | 117.173 | 135505 | |
| 25 | 4 | 2.74228 | 0 | 0.367976 | 0.727356 | 0.370458 | 0.646149 | 101.359 | 101398 | |
| 26 | 4 | 0.897565 | 0 | 0.456968 | 0.758491 | 0.218176 | 0.540104 | 109.776 | 118937 | |

# 2)      Concepts for Building a BBN Model

We start by trying to explain *meaning_of_life* in terms of all the other variables. After dragging all variables onto the graphics panel, we connect each variable to *meaning_of_life.* Because we are in the data mining mode, the menu bar is slightly different, there is a new option DATA. We chose COMPUTE CORRELATIONS from the DATA option and retrieve the following screen:
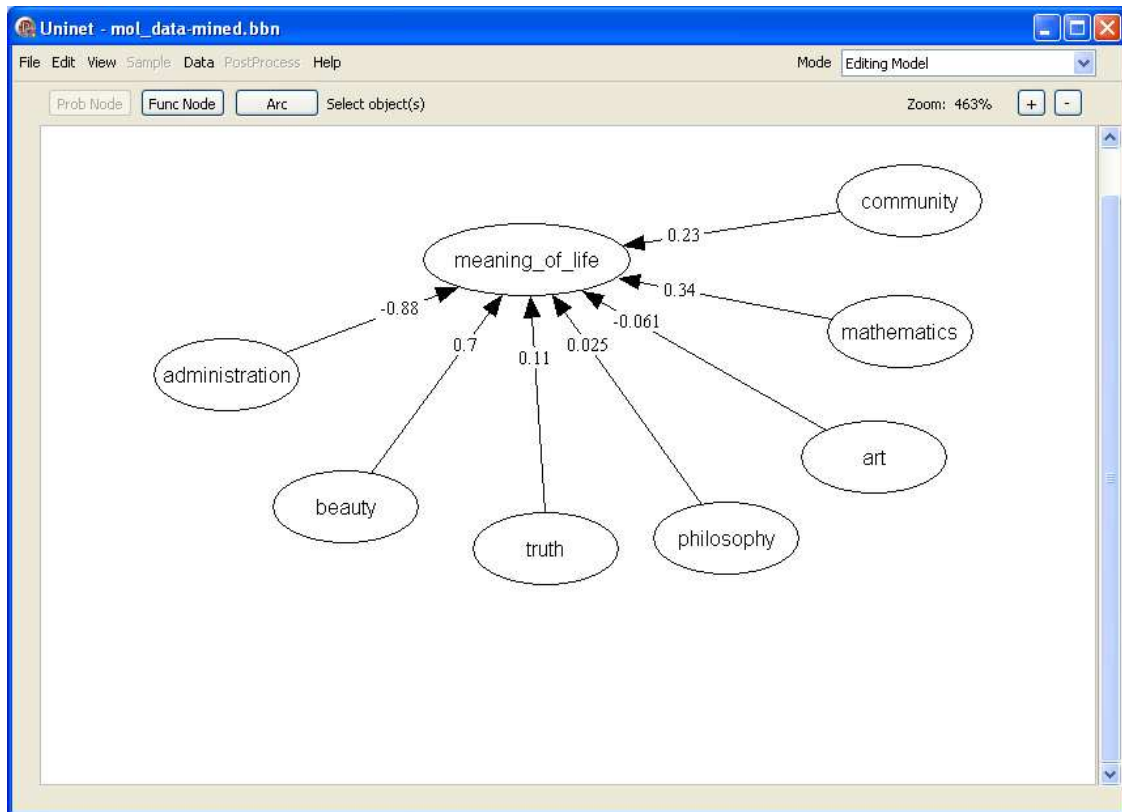


Adding the variable *salary* causes an error, as its rank correlation with *administration* is so high as to cause numerical instabilities – a sign that this should be a functional node The numbers at each arc are conditional rank correlations, according to the index sequence in which the arcs were created. To view this sequence, click on *meaning_of_life:*

## Probabilistic Node Dependence Info

| (Conditional) Rank Correlation Coefficient | Value |
|---|---|
| administration meaning_of_life | -0.880... |
| beauty meaning_of_life \| administration | 0.6982... |
| truth meaning_of_life \| administration beauty | 0.1073... |
| philosophy meaning_of_life \| administration beauty truth | 0.0249... |
| art meaning_of_life \| administration beauty truth philosophy | -0.060... |
| mathematics meaning_of_life \| administration beauty truth... | 0.3385... |
| community meaning_of_life \| administration beauty truth p... | 0.2262... |

Up
Down

Current Editing Mode     Re-ordering Parent Nodes

Messages

OK    Cancel

### a) Rank correlation

The rank correlation of two continuous random variables X,Y is the product moment correlation of their respective quantile functions:

$$\rho_r(X,Y) = \rho(F_X(X), F_Y(Y))$$

Where $F_X$, $F_Y$ are the cumulative distribution functions of X, Y respectively.

### b) Empirical rank correlation
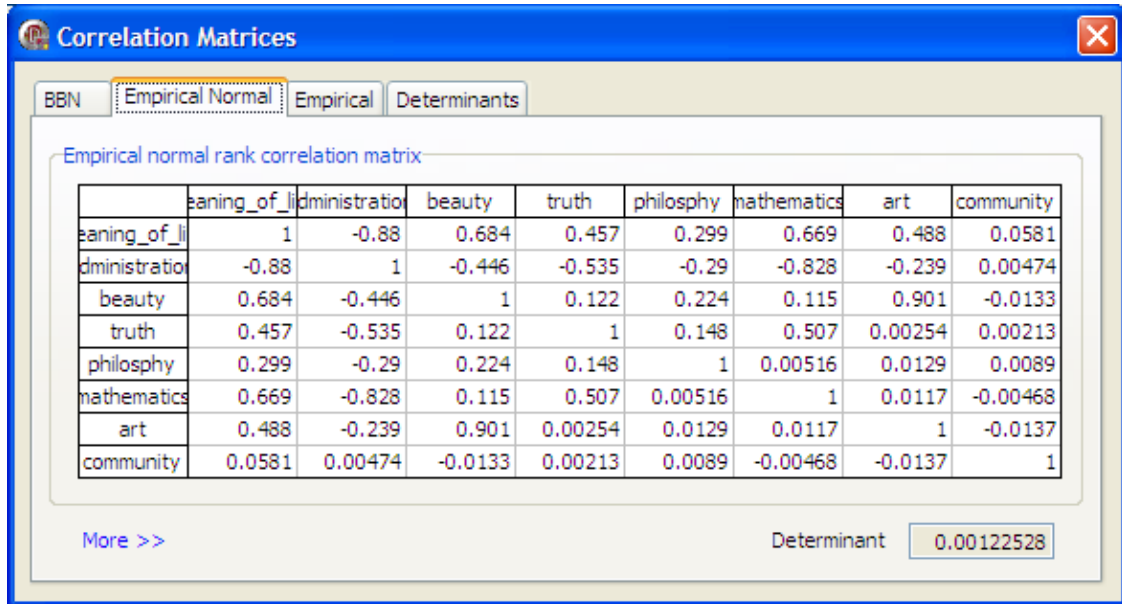
To determine the empirical rank correlation of a set $(x_1, y_1)\ldots (x_n, y_n)$ of n bivariate observations of X and Y, we first construct the *empirical distribution functions*:

$$F_X(r) = \#\{x_i \leq r\}/ (n+1); \; F_Y(r) = \#\{y_i \leq r\}/ (n+1).$$

The empirical rank correlation of X and Y is then the product moment correlation of $F_X(X)$ and $F_Y(Y)$:

$$\rho_r(X,Y) = \rho(F_X(X), F_Y(Y))$$

To view the empirical rank correlation matrix and the DER, select the DATA menu option and choose COMPARE CORRELATION MATRICES:

**Correlation Matrices**

BBN | Empirical Normal | Empirical | Determinants

Empirical normal rank correlation matrix

|  | eaning_of_li | dministratio | beauty | truth | philosphy | mathematics | art | community |
|---|---|---|---|---|---|---|---|---|
| eaning_of_li | 1 | -0.88 | 0.684 | 0.457 | 0.299 | 0.669 | 0.488 | 0.0581 |
| dministratio | -0.88 | 1 | -0.446 | -0.535 | -0.29 | -0.828 | -0.239 | 0.00474 |
| beauty | 0.684 | -0.446 | 1 | 0.122 | 0.224 | 0.115 | 0.901 | -0.0133 |
| truth | 0.457 | -0.535 | 0.122 | 1 | 0.148 | 0.507 | 0.00254 | 0.00213 |
| philosphy | 0.299 | -0.29 | 0.224 | 0.148 | 1 | 0.00516 | 0.0129 | 0.0089 |
| mathematics | 0.669 | -0.828 | 0.115 | 0.507 | 0.00516 | 1 | 0.0117 | -0.00468 |
| art | 0.488 | -0.239 | 0.901 | 0.00254 | 0.0129 | 0.0117 | 1 | -0.0137 |
| community | 0.0581 | 0.00474 | -0.0133 | 0.00213 | 0.0089 | -0.00468 | -0.0137 | 1 |

More >>                                                                 Determinant   0.00122528

### c) Joint normal copula

If $Z = (Z_1,\dots Z_n)$ denotes a joint normal distribution of normal variables $Z_i$ with mean zero and unit variance, then random variables $X_i,\dots X_n$ are connected by the *joint normal copula* if

$$(X_1,\dots X_n) = F_1^{-1}\Phi(Z_1),\dots F_n^{-1}\Phi(Z_n)$$

where $F_i$ is the cumulative distribution function of random variable $X_i$, and $\Phi$ is the cumulative distribution function of the standard normal variable. If $F_i$, $F_j$ are continuous invertible, the transformation $Z_i \rightarrow X_i$ is rank preserving, and the rank correlations of $X_i$, $X_j$ will be the same as that of $Z_i, Z_j$. If $X_i$ is discrete then rank correlation is not perfectly preserved, and the degree of discretization influences the lack of preservation (see Hanea 2008). The rank correlation of normal variables is computed from the product moment correlation with the Pearson transformation:

$$6\arcsin(\rho(Z_i, Z_j)/2)/\pi = \rho(\Phi(Z_i), \Phi(Z_j)) = \rho_r(Z_i, Z_j)$$

Where $\rho_r$ rank correlation and $\rho$ denotes the product moment correlation:

$$\rho(Z_i, Z_j) = ( E(Z_iZ_j) - E(Z_i) E(Z_j) ) / (VAR(Z_i) VAR(Z_j) )^{1/2}$$

Whereas every rank correlation between two variables can be realized as the rank correlation of a bivariate normal distribution, in higher dimensions this is NOT true. For details see ([Kurowicka and Cooke, 2006-A](#))

### d) Empirical normal rank correlation

If $X_i$ denotes a random variable with empirical distribution function $F_i$, then

$$Z_i(X_i) = \Phi^{-1} ( F_i(X_i) )$$

is the *empirical normal version* of $X_i$. Applying the Pearson transformation, we find the *empirical normal rank correlation* of $X_i$, $X_j$ as

$$6 \arcsin(\rho(Z_i(X_i), Z_j(X_j)) /2) / \pi.$$

The empirical normal rank correlation of $X_i$, $X_j$ is not in equal to the empirical rank correlation of $X_i$, $X_j$ unless $X_i$ and $X_j$ are joined by the normal copula.

### e) BBN rank correlation

When a graphical structure is specified the rank correlation of all variables is computed assuming the conditional independence implied by the graph, and assuming that any two variables have the rank correlation induced by the joint normal copula.

### f) Empirical rank determinant DER

Where $F_i$ is the empirical cumulative distribution function of variable $X_i$, $i = 1…N$; the empirical rank determinant is

$$DER = | [ \rho(F_i(X_i), F_j(X_j) ]_{i,j = 1…N} |$$

where | | denotes the determinant. The determinant of a correlation matrix measures the "amount of linear dependence" between the variables. Its value is 1 if the variables are independent, and zero if one of the variables can be written as a linear combination of the others.

### g) Empirical normal rank determinant DNR

The empirical normal rank determinant is the determinant of the rank correlation matrix of the [empirical normal version](#) of $X_i$, $i = 1,…N$:

$$DNR = | \{6 \arcsin(\rho(Z_i(X_i), Z_j(X_j)) /2) / \pi\}_{i,j = 1…N}$$

Where | | denotes the determinant. DNR is the determinant of the *saturated BBN*, that is, the BBN in which there is an arc between every pair of variables.

### h)  BBN rank determinant DBBNR

The BBN rank determinant is the determinant of the BBN rank correlation matrix. Unlike DNR, DBBNR reflects the conditional independence relations imposed by the BBN. If the BBN is not saturated, then DBBNR > DNR. DBBNR is approximated as:

$$\text{DBBNR} = \Pi_{\text{all conditional rank correlations } \rho_r \text{ in BBN}} (1\text{-}\rho_r^2).$$

### i)  Pitfalls

The user should be aware of a number of pitfalls in using the DNR or DBBNR as a test statistic for assessing the adequacy of a BBN model.

#### i)  Shrinking determinant

If the number of variables exceeds the number of samples in an empirical multivariate data set, then the DER will be zero. If some of the variables have repeated values, then this can occur even if the number of samples exceeds the number of variables. If the number of variables is large though not in excess of the number of samples, then DER can become very small.

#### ii)  Discrete variables

When rank correlations are specified for discrete variables in stochastic modeling, the rank correlation realized is the rank correlation between two uniform variables. A discrete variable is realized by $F^{-1}(U)$, where F is the cumulative distribution function of the discrete variable. $F^{-1}$ is thus a many-one transformation. This will cause empirical rank correlations involving this discrete variable to differ from the stipulated rank correlations (see Hanea 2008). In judging statistical adequacy, the user must take this into account by assuming that the discrete variable is a many-one transform of a normal variable. Otherwise, the lack of continuity will be treated as lack of fit of the joint normal copula.

#### iii)  Directionality

It is impossible to infer directionality of influence from multivariate data. Insight into the causal processes generating the data should be used, whenever possible, in constructing a BBN model. Because of this fact, there are different BBNs that are wholly equivalent, and many non-equivalent BBNs may provide statistically acceptable models of a given multivariate ordinal data set.

## 3)  Assessing model adequacy

The question of model adequacy is answered by defining a statistical measure for multivariate dependence and testing whether the hypothesis that the multivariate dependence could come from the joint normal copula should be rejected.

The UNINET approach to this question is based on the *mutual information* as a measure of multivariate dependence (Joe, 1989). The mutual information of a joint density $f(x_1, x_n)$ with marginal densities $f_1,...f_n$ is :

$$\int \ln(f(x_1,...x_n)/f_1(x_1)...f_n(x_n))\ f_1(x_1)...f_n(x_n)\ dx_1,...dx_n.$$

For a joint normal distribution the mutual information is given by $\exp(-2 \ln |C|)$ where $|C|$ is the determinant of the correlation matrix. Mutual information is invariant under monotone univariate transformations; therefore, this is also the mutual information of any distribution based on the joint normal copula. The determinant $|C|$ is closely related to the determinant $|R|$ of the normal rank correlation matrix (DNR). To evaluate the suitability of a joint normal copula it is convenient to compare DNR with the determinant of the empirical rank correlation matrix DER. We could also compare $|C|$ with the "empirical mutual information" of the multivariate data set, were it not that the computation of the "empirical mutual information" imposes obstacles which are as yet unsurmounted.

If the joint normal copula is a statistically acceptable model of the data, then the user can simplify the model by removing arcs whose conditional rank correlation is very close to zero, as these will not contribute significantly to the DNR. By considering the determinant of the rank correlation matrix based on the BBN (DBBNR) one can determine whether the BBN is a statistically acceptable model of the empirical normal rank correlation matrix.

The statistical tests are not currently supported in UNINET.

The ideas are illustrated with the BBN model created to explain the multivariate data:
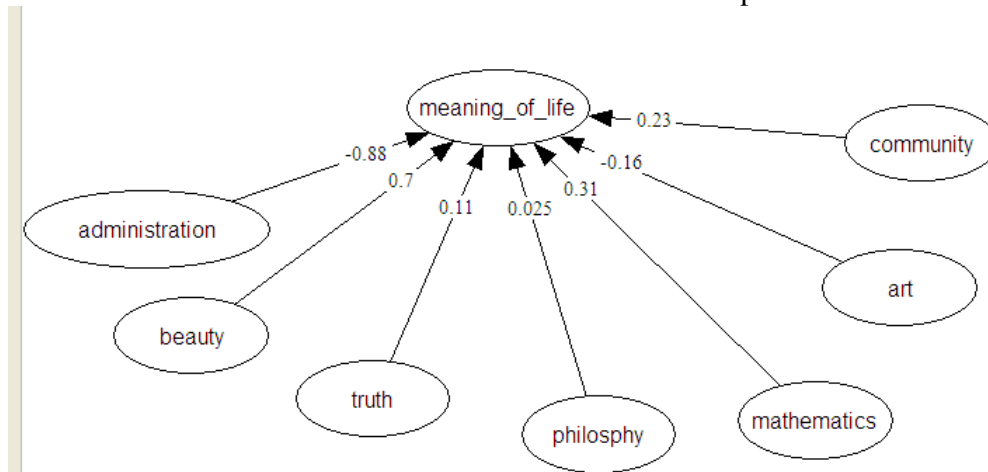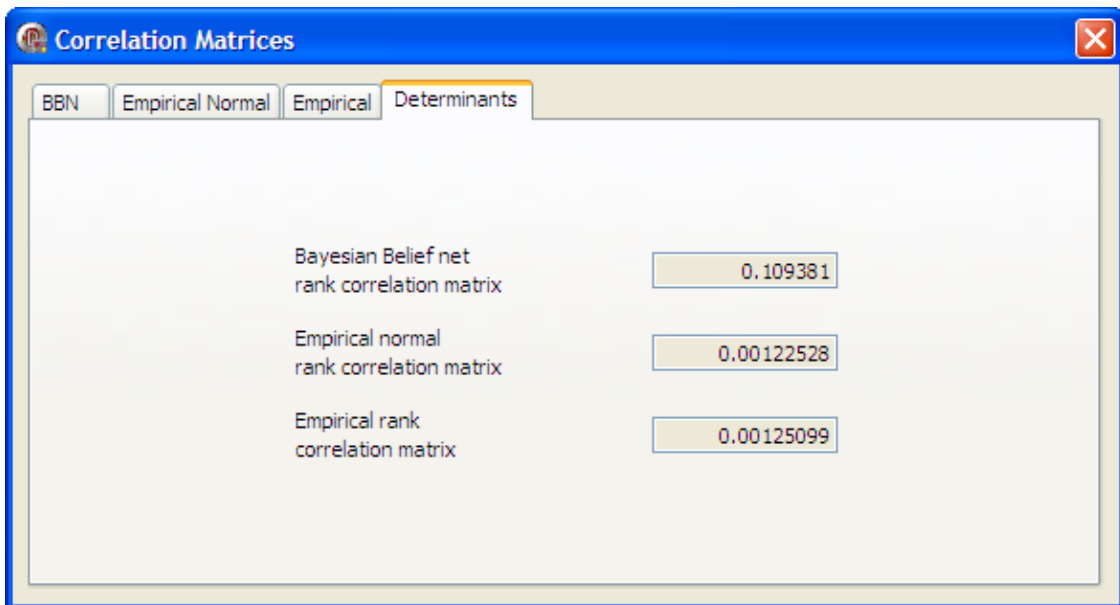


The BBN used to generate the data is shown below:
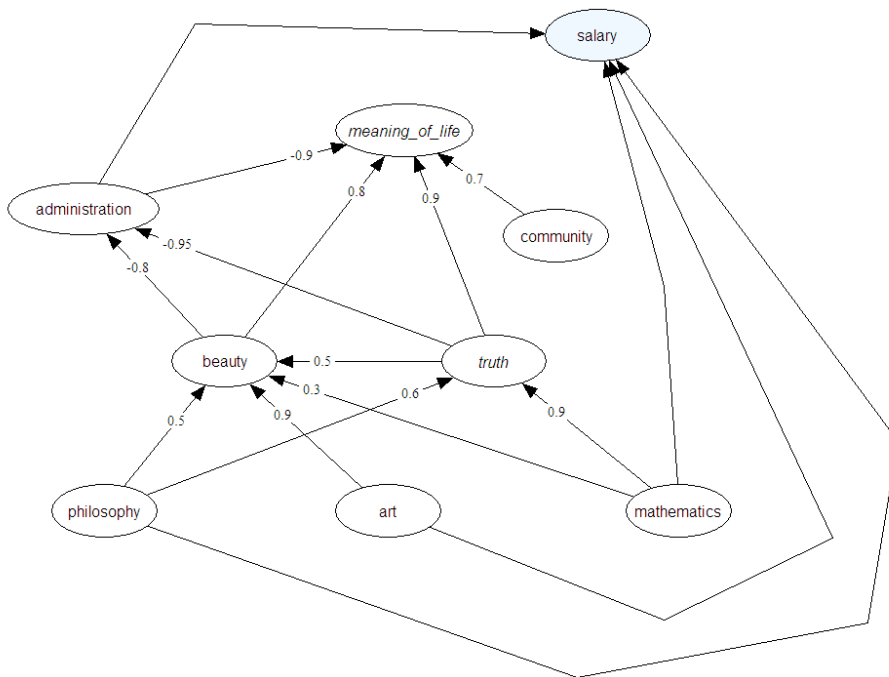
The model generating the data is very different from the model proposed to explain it.

From the menu DATA, we can compare DER, DNR and DBBNR:



The DNR = 0.00122528, which is very close to DER = 0.00125009.  DNR is the determinant of the saturated graph. The small difference is caused by the fact that *truth* is a discrete variable, whose empirical rank correlation with other variables may not correspond with the values which the user specified in creating this data set (see pitfalls) .

The [DBBNR](#) = 0.109381, which is way off, indicating that this model does not adequately represent the data (the statistical tests are currently not supported in UNINET, but the statement is true).

We can look for missing interactions by choosing COMPARE CORRELATION MATRICES from the DATA menu. Checking the appropriate boxes, we can highlight those correlations between variables connected by arcs in the graph (the matrix shows rank correlations, these are not the conditional rank correlations shown in the graph). We can also choose to view the k highest correlations between pairs NOT connected in the graph.
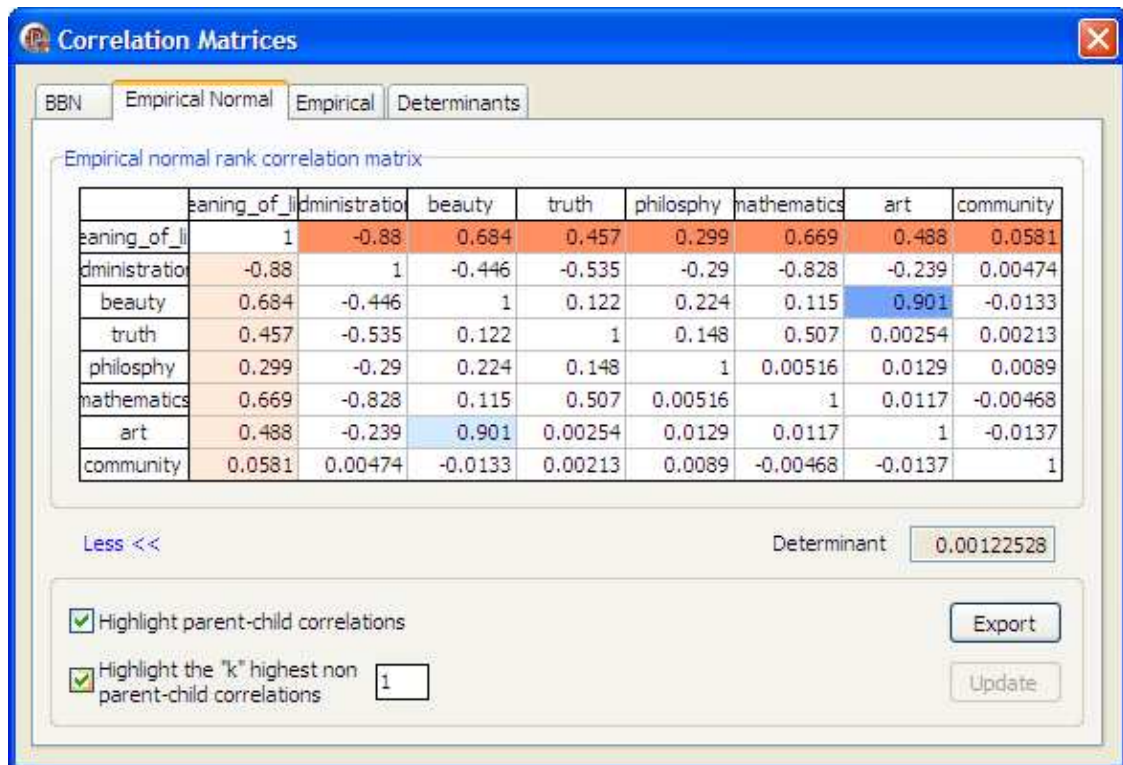


We see above that art and beauty have rank correlation in the empirical normal rank correlation matrix of 0.901. Adding this arc to the BBN:
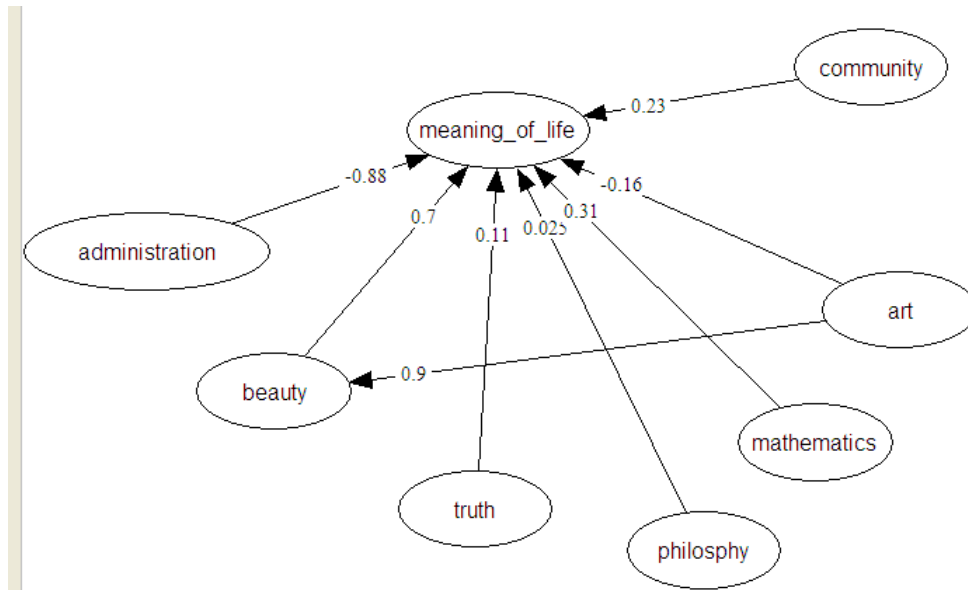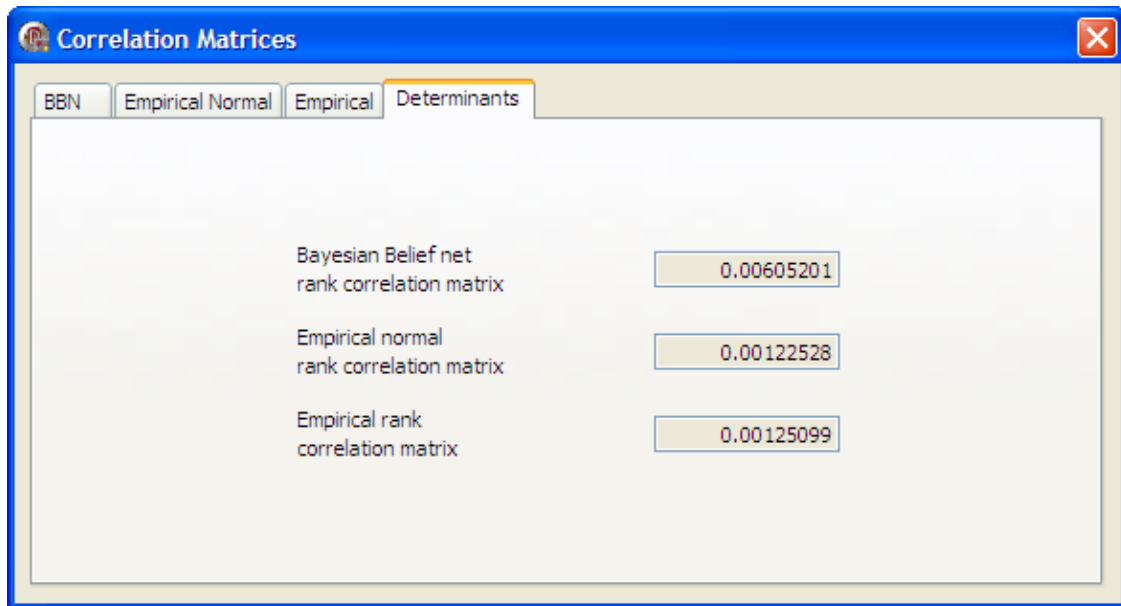
we can again compute DBBNR:
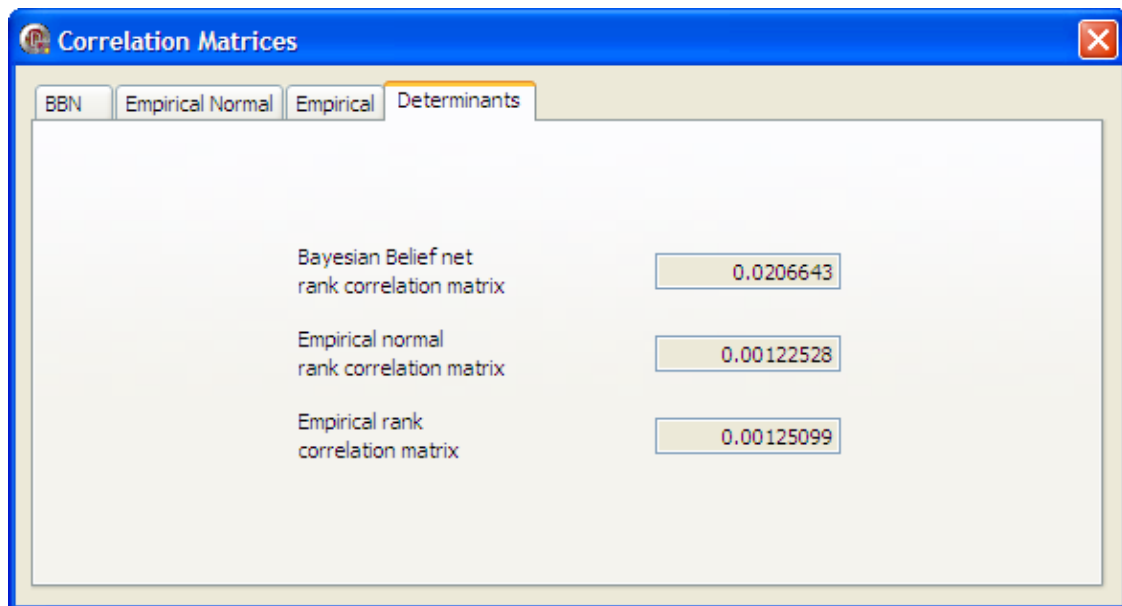


The result DBBNR= 0.0206643, would be very well predicted by the approximation:

$$(1 - (0.901)^2) \times 0.109381 \; = 0.020585395.$$

If we reproduce exactly the graph that was used to construct the data we find the following comparison of determinants.

The difference between DBBNR and DNR is now entirely due to sampling correlations that, although theoretically zero, are nonetheless non-zero in the sample.

## 4) Adding functional nodes

Functional nodes may be added to a multivariate ordinal data model. The procedure is identical to that in the stochastic modeling mode.

## 5) Conditionalizing empirical distribution

If we do not sample a BBN based on multivariate data, but simply shift to the ANALYTICAL CONDITIONING mode, then UNINET allows conditionalization while preserving the original marginal distributions.

## 6) Report generation

**** 2b written ****

# V) References

Ale. B.J.M., Bellamy, L.J., Boom, R. van der, Cooper, J. Cooke, R.M. Goossens, L.H.J., Hale, A.R. , Kurowicka, D. Morales, O. , Roelen, A.L.C. , Spouge, J.   " Further development of a Causal model for Air Transport Safety (CATS); Building the mathematical heart. " Appearing in special issue of the Reliability Engineering and System Safety Journal on selected papers from the ESREL 2007 Association 84 (405) 157-164, 1989.

Bedford, T.J. , and Cooke, R.M. , "Vines - a new graphical model for dependent Random Variables"- in  Annals of Statistics, vol. 30, No.4, August 2002.

Bedford, T.J. and Cooke, R.M. "Probability density decomposition for conditionally dependent random variables modelled by Vines, "  Annals of Mathematics and Artificial Intelligence, Volume 32 pp 245-268, 2001.

Cooke, R.M. , and Goossens, L.H.J.. "Procedures Guide for Structured Expert Judgement in Accidents Consequence Modeling", Radiation Protection Dosimetry Vol. 90 No. 3.2000, pp 303-311

Hanea, A.M., Kurowicka, D. and Cooke, R.M. " Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets" (2006) *Quality and Reliability Engineering International* 22,709-729, 2006.

Hanea, A.M., Kurowicka, D. Cooke, R.M. and Ababei, D.A.A. "Ordinal data mining with non-parametric continuous Bayesian belief nets", Dept. of Mathematics, T.U. Delft, 2008.

Hanea, A.M. "Algorithms for non-parametric Bayesian Belief Nets", PhD thesis, Dept. of Mathematics, Delft University of Technology, 2008.

Joe, H. , Relative entropy measures of multivariate dependence, Journal of the American Statistical

Kurowicka D., and Cooke R.M., A Parametrization of Positive Definite Matrices in Terms of Partial Correlation Vines", in Linear Algebra and its Application 372, 225-251, 2003.

Kurowicka D., and Cooke R.M., Completion Problem with Partial Correlation Vines, Linear Algebra and Its Applications,, vol. 418, nr 1, pp 188-200, 2006_B.

Kurowicka, D. , and Cooke, R.M. Uncertainty Analysis and High Dimensional Dependence Modeling Wiley, New York, 2006-A.

Kurowicka, D. , Cooke, R.M. , Sampling algorithms for generating joint uniform distributions using the vine-copula method, , Comp. Stat. & Data Analysis, Volume 51 , Issue 6 March 2007.

Lewandowski, D. Cooke, R.M. and Duintjer Tebbens, R.J. "Sample Based Estimation of Correlation Ratio with Polynomial Approximation" Transactions on Modeling and Computer Simulation, vol. 18 no.1, article 3, 2007.

Morales, O., Kurowicka, D. and Roelen, A., Eliciting Conditional and Unconditional Rank Correlations from Conditional Probabilities, Special issue on expert judgment Reliability Engineering & System Safety, 93, pp***, Available online 12 March 2007.