Bayesian Networks for the Analysis of Subjective Well-Being

Jan Švorc[1,2] **Jiří Vomlel**[1,2]

[1]Faculty of Management
University of Economics, Prague

[2]Institute of Information Theory and Automation
Czech Academy of Sciences

Abstract

We use Bayesian Networks to model the influence of diverse socio-economic factors on subjective well-being and their interrelations. The classical statistical analysis aims at finding significant explanatory variables, while Bayesian Networks can also help sociologists to explain and visualize the problem in its complexity. Using Bayesian Networks the sociologists may get a deeper insight into the interplay of all measured factors and their influence on the variable of a special interest. In the paper we present several Bayesian Network models – each being optimal from a different perspective. We show how important it is to pay a special attention to a local structure of conditional probability tables. Finally, we present results of an experimental evaluation of the suggested approaches based on real data from a large international survey. We believe that the suggested approach is well applicable to other sociological problems and that Bayesian Networks represent a new valuable tool for sociological research.

1 Introduction

Bayesian Networks (BNs) [13, 10] are probably the most popular representative from the class of probabilistic graphical models. In this paper we show how BNs can help social scientists to get a deeper insight into a studied problem of their interest. We will use the problem of subjective well-being throughout the paper to illustrate key benefits of the suggested approach.

Although the subjective well-being (SWB) has been researched for decades [4, 6] the debate on its association with the material living conditions still continues and many questions remain unanswered. The people mostly think their happiness and satisfaction are directly linked with the wealth. Modern researchers have also proved that material aspects of life matter, yet their findings are sometimes surprising. Better understanding of the association between the variables related to the material living conditions such as income, wealth, material deprivation and SWB is valuable.

Typical questions the classical statistical analysis can answer are questions like: what factors have a significant influence on subjective well-being? If regression models are used we may study strengths and signs of the influence of explanatory variables on the dependent variable of our interest. In addition to this analysis, probabilistic graphical models help to get a deeper insight into the interplay of all measured factors and their influence on the variable of the special interest. In BNs the relations are visualized graphically using acyclic directed graphs representing conditional independence relations among variables.

The paper is organized as follows. In Section 2 the concept of subjective well-being is introduced and briefly reviewed. In Sections 3 and 4, the hypotheses of SWB and its association with the variables of material situation are examined using appropriate statistical methods on empirical data. The main original contribution of this paper is presented in Section 5, where BNs are applied to the analysis of SWB. We present two principally different approaches to learning BN structures: one based solely on collected data and on minimization of Bayesian information criteria (BIC) and other where we use an expert version of the PC algorithm to build the model using the expert knowledge of the modeled domain. Special attention is given to learning conditional probability tables (CPTs). The general form of these CPTs, which is commonly used in diverse applications, leads to an undesired and counter-intuitive model inference despite a relatively large dataset used for learning. The main problem is a non-monotone behavior. We show that this problem can be overcome by using appropriate local structure of CPTs – we use Ordinal Logistic Regression (OLR) in Section 6. In Section 7 we evaluate models by measuring how well they fit the data and by measuring their prediction accuracy. We also provide an example of the BN model use in Section 8. We summarize our contribution in Conclusions.

2 Subjective Well-Being

Broadly speaking, SWB is the self-evaluation of one's overall life in positive terms [4]. The concept of SWB has little to do with the objective living conditions, it is determined solely by the subjective assessment.

SWB has two dimensions based on [6]. The affective (or emotional) dimension includes positive and negative moods and emotions (affects). They represent on-line evaluations of events occurring in one's life, whereas the happiness is the surplus of the positive affects over the negative ones. Positive and negative affects are considered to be, in essence, the independent factors. The cognitive dimension of SWB means the judgement of one's satisfaction with the life as a whole as well as with the various life domains, such as job, income, family, leisure etc. Hence, the people high in SWB experience pleasant emotions frequently and unpleasant

Table 1: Model Variables

Abbr.	Description	States
SWB	Subjective Well-Being	{unhappy, fairly unhappy, fairly happy, happy}
PAST	Income compared to own past	{better, the same, worse}
OTHR	Income compared to others	{worse, the same, better}
DEPR	Material deprivation	{none, weakly deprived, deprived}
STRS	Subjective economic strain (ability to make ends meet)	{easily, fairly easily, with some difficulty, with difficulty}
FPRO	Financial problems	{none, minor, major}
HOUS	Housing problems	{no defect, single defect, several defects}
INC	Household income	{low, fairly low, fairly high, high}
CRY	Respondent's country	{C1, C2, C3, C4}

emotions rarely and feel satisfied with the conditions of their lives [5].

Some authors strictly distinguish the happiness from the life satisfaction, where the happiness resulted from the positive experience and the life satisfaction is an outcome of an individual evaluations of discrepancy between material and social aspirations, expectations and achievements. The variable of the subjective well-being in our model incorporates both happiness and life satisfaction components.

The most frequently referred correlates of SWB can be grouped, for example, as follows: demographic factors (age; gender; marital status; religion; physical health), social factors (education; occupation; social relationships), personality factors (extraversion; neuroticism; self-esteem; optimism; purpose-in-life), and wider environmental factors (culture; governance; inflation; unemployment; climate etc.). In this study we consider only factors related to the material situation. In Table 1 we list studied model variables, their brief description, and the number of states of these variables[1].

In the analysis we use data from the third survey of the European Quality of Life Study conducted in 2011 [7]. The survey covers all persons aged 18 and more whose usual place of residence is in the territory of the surveyed countries at the time of the data collection. Only one interview per household. We used the data from four post-communist central European countries – the Czech Republic, Hungary, Poland and Slovakia. These four countries are culturally, geographically, economically, and politically similar. The total of 5,298 respondents from these four countries participated in the survey, out of whom 3,259 complete data vectors are extracted by removing respondents having answered the relevant questions incompletely (613 in the Czech Republic; 586 in Hungary; 1,428 in Poland; and 632 in Slovakia).

[1]Variables SWB and INC were transformed from the original scales using a quantile discretization. The states of variables DEPR, STRS, FPRO, and HOUS are summaries from answers of several questions on the corresponding topic.

Table 2: P-values for the hypothesis of equal means of SWB given values of each explanatory variable.

Pairs of values	1 vs. 2	1 vs. 3	2 vs. 3
INC	0.000	0.000	0.000
OTHR	0.000	0.000	0.000
PAST	0.125	0.000	0.000
DEPR	0.000	0.000	0.000
STRS	0.000	0.000	0.000
FPRO	0.000	0.000	0.629
HOUS	0.000	0.000	0.000
Pairs of values	1 vs. 4	2 vs. 4	3 vs. 4
INC	0.000	0.000	0.000
STRS	0.000	0.000	0.000

3 Basic Statistical Analysis

In [18] a basic statistical analysis of the influence of factors related to the material situation on SWB in four countries of Central Europe was performed. The null hypothesis of equal means of SWB were rejected for all variables[2]. The results are summarized in Table 2 where we present p-values of Welch t-test [19] of equal means of SWB. The results of the tests are presented for each pair of values of every factor variable.

From the table we can see that for almost all explanatory variables the hypothesis of equal means can be rejected except for PAST=1 and PAST=2 and for FPRO=2 and FPRO=3, where means are not significantly different. However, we can conclude that all explanatory variables are significant for SWB since all of them help differentiate between SWB values for at least two of their states. More details can be found in [18].

4 Ordinal Logistic Regression

A natural model for ordinal variables is Ordinal Logistic Regression (OLR) [12]. Since the variable SWB has four states the OLR model of the dependent variable Y representing SWB is defined for $i = 1, 2, 3$ using cumulative distribution functions:

$$P(Y \leq i) \quad = \quad logit^{-1}\left(\zeta_i - \sum_{j \in J} \beta_j \cdot x_j\right) \quad ,$$

[2]In this paper the SWB variable used the original ten points scale.

Table 3: Ordinal logistic regression model.

variables	β_j	std. error	t-value	p-value
CRY2	0.3834	0.1124	3.4118	0.001
CRY3	0.7840	0.0947	8.2815	0.000
CRY4	0.1367	0.1081	1.2644	0.206
INC	0.0870	0.0361	2.4108	0.016
OTHR	0.4660	0.0617	7.5470	0.000
PAST	-0.2709	0.0583	-4.6450	0.000
STRS	-0.2741	0.0451	-6.0728	0.000
DEPR	-0.4960	0.0610	-8.1378	0.000
FPRO	-0.1047	0.0496	-2.1098	0.035
HOUS	-0.2044	0.0466	-4.3879	0.000
intercepts				
ζ_1	-2.3252	0.2977	-7.8105	0.000
ζ_2	-0.8687	0.2949	-2.9460	0.003
ζ_3	0.4056	0.2949	1.3754	0.169

where $\zeta_i, i = 1, 2, 3$ are the intercepts for different values of SWB, $\beta_j, j \in J$ are coefficients of explanatory variables $X_j, j \in J$ taking states x_j. The probability distribution of the dependent variable Y is computed from the cumulative distribution functions as $P(Y = i) = P(Y \leq i) - P(Y \leq i - 1)$ for $i = 1, 2, 3, 4$, where $P(Y \leq 0) = 0$ and $P(Y \leq 4) = 1$.

The coefficients and intercepts of the OLR model learned from SWB data are presented in Table 3. We can see that income relatively higher than income of other people increases probability of higher values of SWB. On the other hand, if the respondent has had a higher income in the past than now then his/her SWB has a higher probability of being lower now. Also, problems specified by variables STRS, DEPR, FPRO, and HOUS imply lower SWB. Since the country variable is not ordinal it is transformed to 3 binary variables taking CRY=1 as the reference value.

5 Bayesian Networks for Subjective Well-Being

A main advantage of BNs is that they represent conditional independence relations graphically. Uncertain relations between variables are modeled using the conditional probability distributions. Hence, BNs enable an efficient encoding of a domain knowledge and improve understanding of complex problems. BNs provide a compact representation of the joint probability distribution.

BNs enable exact probabilistic inference assuming the structure and parameters are estimated correctly – the posterior probability distribution of any variable can be computed. BNs help answering queries under the uncertainty. As the software for modeling and learning the structure and parameters is available, the complex situations can be modelled with the help of BNs.

The BNs construction includes two main consequent phases: (1) determining the structure and (2) learning the parameters. Determining the structure includes definition of model variables and establishing of directed links among the variables in a network. The structure can be determined based on the expert knowledge or learned from the available data using a structure learning algorithm. In Section 5.1 we use the expert knowledge and in Section 5.2 we learn a BN model from data.

5.1 Expert model

In order to make the process of building the expert model structure systematic we decided to follow the scheme of the PC algorithm [16]. The major difference is that in the standard PC algorithm collected data are used to decide whether a Conditional Independence (CI) statement holds or not while in the expert version of the PC algorithm the expert knowledge is used to decide validity of CI statements for this purpose. If necessary, a detailed review of the SWB literature helped us to reach decisions[3]. The resulting model structure is presented on the left hand side in Figure 1.

5.2 BIC optimal model

Another possibility is to use data to learn a BN model structure. A class of standard model estimation methods is based on finding a model that maximizes the log-likelihood (LL) of data given the model. It is well-known that this often leads to overfitting the training data and results in complex models. Therefore criteria that penalizes complex structures are often used instead. The BIC criterion [15] subtracts from LL a penalty which is proportional to the number of parameters of the BN model $\mathcal{M}$:

$$BIC(\mathcal{M}) \quad = \quad LL(\mathcal{M}) - \frac{1}{2}\kappa \log N \ ,$$

where κ is the number of free parameters in model $\mathcal{M}$ and N is the number of data records.

On the right hand side of Figure 1 we present the structure of a BIC optimal BN model. This model was learned using the Gobnilp tool [1]. Apparently, the BIC greedy search implemented in [9] also results in a model that is equivalent to the BIC optimal one.

[3]A detailed description of the whole process exceeds the scope of this paper. This description is part of a paper currently under review in a journal.

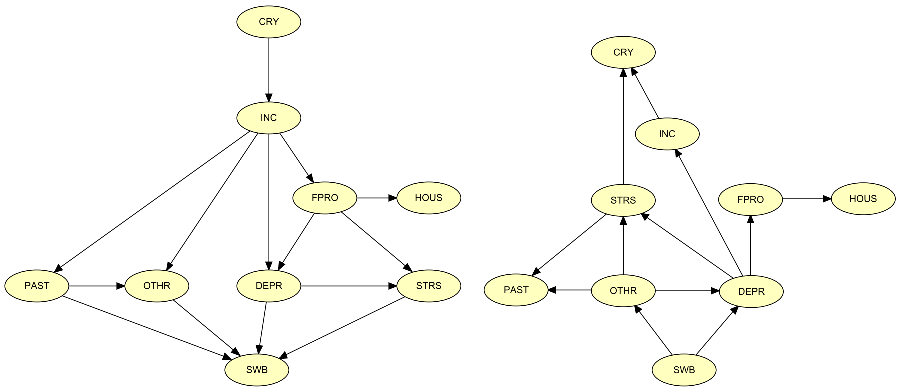

Figure 1: Expert BN (left) and a BIC-optimal BN (right).

6 Conditional probability tables

For the estimation of values of conditional probability tables (CPTs) of BNs from data the EM algorithm is commonly used [3]. If data are complete (i.e., if they contain no missing records) then this procedure reduces to computing relative frequencies from data and this is the case of all our experiments since we used complete data records only. We observed that if the general form of CPTs is used it leads to an undesired and counterintuitive inference despite a relatively large training dataset. The main problem is a non-monotone behavior.

For example, in the model learned from data we observed that $P(\text{SWB} = 1|e, \text{PAST} = 2) \leq P(\text{SWB} = 1|e, \text{PAST} = 3) \leq P(\text{SWB} = 1|e, \text{PAST} = 1)$, which means that the lowest SWB is achieved when variable PAST takes its medium value. One would rather expect the influence of PAST is monotone and SWB is the lowest when the relative income compared to one's own past has become much worse (state 3). The symbol e stands for a particular evidence on remaining parents of SWB. The undesired behaviour was observed for e = (OTHR=1, DEPR=1, STRS=1). Such non-monotone behavior can be observed if there are not enough observations for a given evidence e, which is a quite common situation. This problem should be eliminated since the users do not trust any system with such behaviour. One may believe that this problem disappears when data are large enough but we would like to stress that a mere large dataset does not guarantee that certain combinations are not rare in data. The problem can be properly solved by using a local structure of CPTs appropriate for the application.

Ordinal logistic regression models [12] have several properties that make them good candidates for CPTs in BNs for subjective well-being problem. They assume a natural ordering of the states of variables, which corresponds well to all variables in our model except the country (CRY). Also, the OLR models allow explanatory

Table 4: Comparisons of models' LL and BIC.

Model	LL	BIC std.	BIC OLR
Full	-27,328	-278,930	
OLR	-32,129	-79,419	-32,283
Expert	-29,579	-31,342	
Expert OLR	-29,938	-31,702	-30,112
BIC-optimal	**-29,195**	**-29,822**	
BIC-optimal OLR	-29,395	-30,022	**-29,706**
TAN	-29,285	-30,268	
TAN OLR	-29,525	-30,508	-29,928

variables to have either positive or negative effect on the dependent variable, which fits well the studied problem. We learned OLR models for all CPTs of our expert model except for two CPTs: $P(CRY)$ and $P(INC|CRY)$. In this way, the non-monotonicity property observed for general CPTs completely disappeared.

Other methods that guarantee monotonicity in CPTs exist, e.g. [17, 11, 14]. We have decided to use OLR models since they are commonly used in sociology. However, a more detailed study considering other methods would be interesting, but we leave this task for a future research.

7 Models' Evaluation

We will compare the discussed BN models and the Tree Augmented Naive Bayes model (TAN), which is a BN model commonly used in classification problems [8]. First, we compare the models using the Log-Likelihood (LL) and the Bayesian Information Criteria (BIC). The values of these two criteria are presented in Table 4. The measures are computed with respect to the whole dataset consisting of 3259 data vectors.

From Table 4 we can see that the best model with respect to the BIC criteria is (indeed) the BIC-optimal model whose structure was learned by Gobnilp. The Expert model with unrestricted CPTs has 436 free parameters. When we restrict the conditional probability tables to have parameters of the OLR models (for all nodes except CRY and INC since CRY is not an ordinal variable) the number of free parameters drops to 43. This means the penalty reduces from 1,763 to 174, which implies the BIC value of Expert OLR drops from -31,702 to -30,112. Thus, by the OLR restriction of the CPTs we can get a significantly better BIC value. Similar observations hold for the BIC-optimal and BIC-optimal OLR models.

The primary goal of our work was to help sociologists to get a deeper insight into the problem of SWB, to explain the relation between variables, and to provide a tool for computations of marginal conditional probabilities in situations of sociologists' interest. However, this being said, we decided to test also the predictive ability of the learned models. The prediction variable is SWB.

We split data into 10 folds and used 10-fold cross-validation to evaluate models predictive abilities[4]. In order to analyze the influence of the data size we performed experiments on fractions of the whole dataset. This means that for small subsets of data we performed more cross-validation experiments. In this way we have achieved comparable results since each single respondent record was used exactly once in testing (in all considered data sizes).

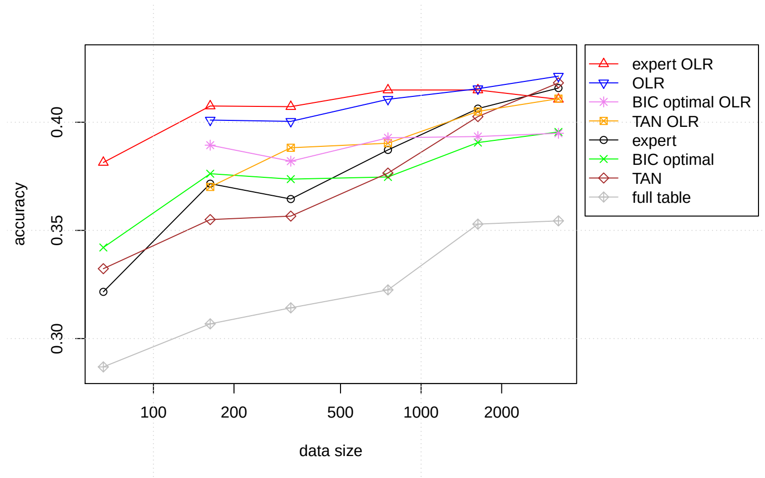

Figure 2: Accuracy with respect to the size of training data.

The models' accuracy is presented in Figure 2. It is the ratio of correctly classified instances with respect to all instances:

$$acc = \frac{1}{N} \sum_{i=1}^{4} C(i,i) \ ,$$

where symbol C denotes the confusion matrix which contains at $C(i,j)$ the number of cases predicted as SWB=i with the reference value SWB=j and N is the total number of instances, i.e.,

$$N = \sum_{i=1}^{4} \sum_{j=1}^{4} C(i,j) \ .$$

From the plot we can see that for small data sizes the OLR versions of all models have better accuracy than their standard versions. The expert OLR and standard

[4]The structure of the tested models was fixed. Each time, nine folds were used to learn model parameters only.

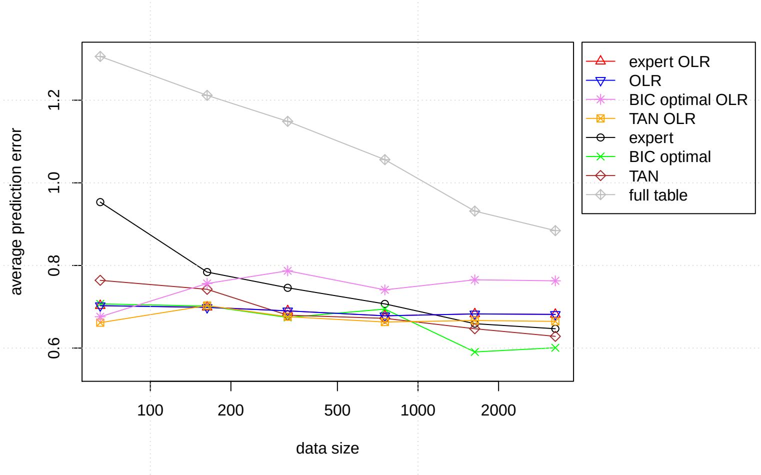

Figure 3: Prediction error with respect to the size of training data.

OLR models performed best. The standard expert model and TAN require more training data to achieve comparable performance. BIC optimal model remains significantly worse than his competitors[5]. The worst of all tested models is the full table model despite this model has the best fit of data. This model is a typical example of a model overfited to training data but performing badly on testing data. Though, the overall accuracy of 0.42 may seem to be low, we should stress that it is significantly higher than the no-information-rate, which is 0.278 (note that SWB has 4 states). The observed level of accuracy is a natural consequence of restricting the study to only factors related to the material situation. Clearly, the omitted factors also play an important role in SWB and since they are not part of our tested models we cannot hope for predictions of a very high accuracy. But, as we have already mentioned, the high accuracy predictions were not the primer goal of our work reported in this paper.

Since the SWB variable is ordinal a more appropriate measure of models' performance than accuracy might be a prediction error. Contrary to the accuracy, which does not consider the distance between the predicted and observed SWB values, the average prediction error defined below does it by means of the absolute difference between the predicted and observed SWB value:

$$err = \frac{1}{N} \sum_{i=1}^{4} \sum_{j=1}^{4} |i-j| \cdot C(i,j) \ .$$

From Figure 3 we can see that for small data sizes the full table model and the (non OLR) expert model are clear losers and TAN is also slightly worse. Other

[5] We verified the claimed differences by Wilcoxon signed-rank tests. Due to the lack of space we do not report the p-values of these tests in this paper.

models have comparable performance. When more training data are available TAN quickly gains comparable performance and latter even the (non OLR) expert model also gains comparable performance to models that were better on smaller datasets. We observe an unexpected performance deterioration for the BIC optimal OLR. We have looked more closely at this model and in its confusion matrix we can see that this model never classifies any instance as SWB=3. Since 3 is one of the middle SWB values the prediction error is affected more than the accuracy. The BIC optimal OLR model is simply not a good candidate for the classification since only two variables (OTHR and DEPR) have any influence on SWB and by restricting the already small CPTs further by the OLR requirement we worsen its performance.

When compared to accuracy the prediction error is more satisfactory, since the prediction error values imply that most testing instances have their difference between the predicted and observed SWB values at most 1. For example, for the best performing model on the largest training set, i.e. for BIC optimal, only 20% of tested instances had this difference larger than 1, while for 40% of instances this difference was equal to 1 and 40% of instances were correctly predicted.

8 An example of a BN model use

In Figure 4 we present an example of a model use. For this purpose we use the Expert OLR model whose structure was presented on the left hand side of Figure 1. This model can be used to predict most probable values of variables of interest in different life circumstances.

For example, assume a person with a low income, but with no subjective material deprivation and making easily ends meet (a low subjective economic strain). When we enter these conditions as evidence into the BN model we can read the conditional probability distributions of remaining variables. From Figure 4 we can see that despite a low income the person is expected to have a high SWB since a low subjective material deprivation and a low subjective economic strain overweight the negative influence of a low income.

9 Related work

Probably, the closest related work is the working paper by Ceriani and Gigliarano [2]. Our motivation is similar to their motivation but our approach differs from their approach in several aspects. Ceriani and Gigliarano do not require monotonicity in CPTs, which we believe is important, as we have shown in our paper. They also used different learning algorithms from the bnlearn R package, that, contrary to Gobnilp, do not necessarily provide BNs optimal with respect to BIC. In addition, we proposed to use an expert-based method which can a domain expert use to build a BN model.

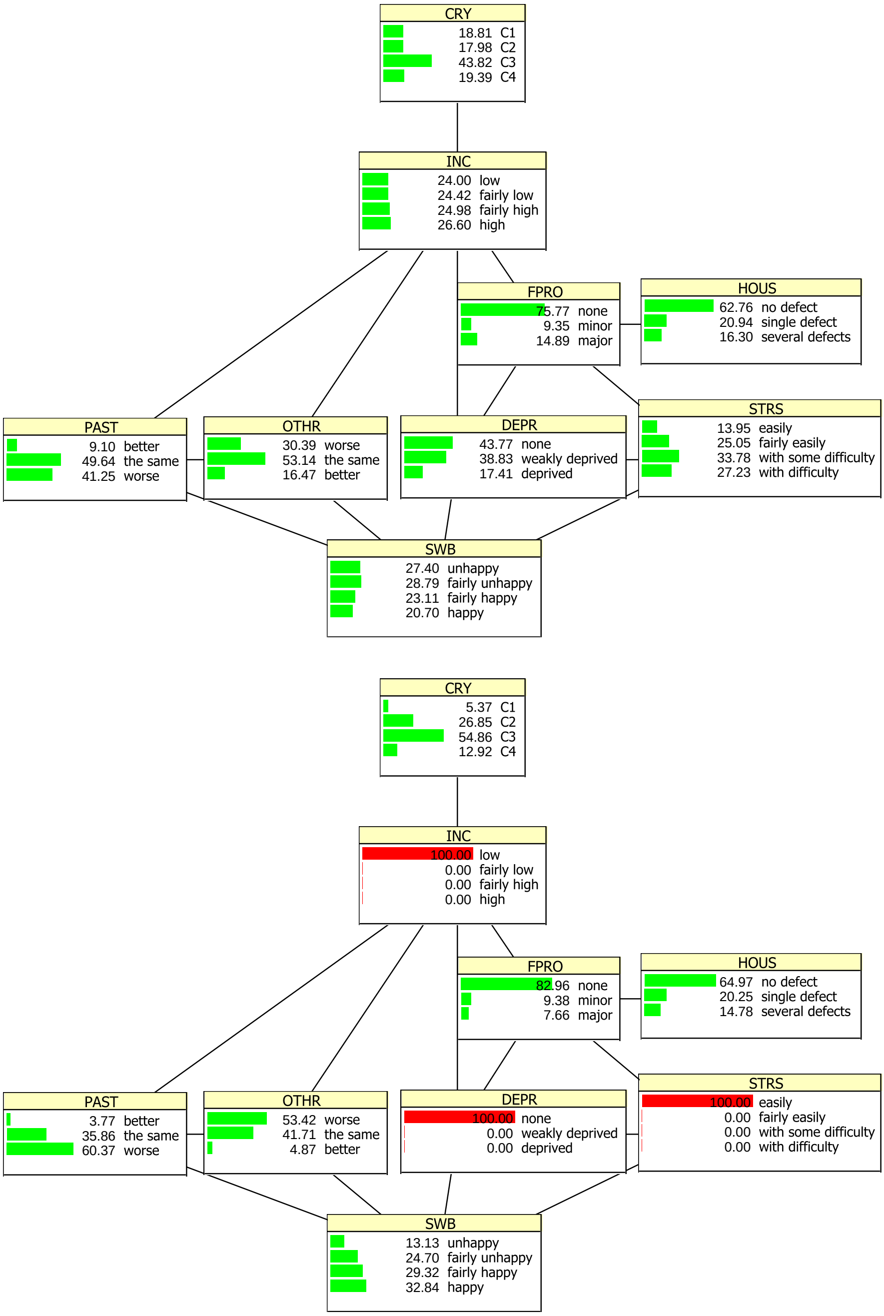

Figure 4: Prior marginal probabilities (top) and the marginals updated for the observed evidence (bottom).

10 Conclusions

In the sociology literature it is still an ongoing debate which factors are important for SWB and which are mediated through others. In the basic statistical analysis of the SWB all studied socio-economic variables were proved to be statistically significant for SWB, but this analysis cannot decide whether their influence is direct or mediated through other variables. We applied BNs to this problem since they can model complex relations between variables. The expert model constructed using an expert version of the PC algorithm can be used to resolve this debate in a systematic and mathematically rigorous way.

From the point of view of sociology, both, the Expert and the BIC-optimal BNs suggest that the objective conditions such as the income and the financial problems influence SWB only indirectly through the subjective perception of the relative income, the material deprivation and the economic stress. We were able to derive this conclusion (and few others) due to the analysis based on BNs. We believe that BNs represent a valuable tool for social scientists.

Acknowledgement

This work was supported by grants GAČR 17-08182S and GAČR 19-04579S.

References

[1] M. Bartlett and J. Cussens. Integer linear programming for the Bayesian network structure learning problem. *Artificial Intelligence*, 24:258–271, 3 2017.

[2] L. Ceriani and C. Gigliarano. Multidimensional well-being: A Bayesian Networks approach. Working Papers 399, ECINEQ, Society for the Study of Economic Inequality, 2016.
https://EconPapers.repec.org/RePEc:inq:inqwps:ecineq2016-399.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[4] E. Diener. Subjective Well-Being. *Psychological Bulletin*, 95(3):542–575, 1984.

[5] E. Diener and R. Biswas-Diener. Will Money Increase Subjective Well-Being? *Social indicators research*, 57(2):119–169, 2002.

[6] E. Diener, E. M. Suh, R. E. Lucas, and H. L. Smith. Subjective Well-Being: Three Decades of Progress. *Psychological Bulletin*, 125(2):276—-302, 1999.

[7] European Quality of Life Survey. Computer file, European Foundation for the Improvement of Living and Working Conditions, Colchester, Essex: UK Data Archive, 2011–2012. http://dx.doi.org/10.5255/UKDA-SN-7316-2.

[8] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2):131–163, 1997.

[9] Hugin. API Reference Manual, version 8.2, 2015. `http://www.hugin.com/`.

[10] F. Jensen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, New York, 2001.

[11] A. R. Masegosa, A. J. Feelders, and L. van der Gaag. Learning from incomplete data in Bayesian networks with qualitative influences. *International Journal of Approximate Reasoning*, 69:18–34, 2016.

[12] P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142, 1980.

[13] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[14] M. Plajner and J. Vomlel. Monotonicity in Bayesian networks for computerized adaptive testing. In *The Proceedings of 14th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2017)*, pages 125–134. Springer International Publishing, 2017.

[15] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[16] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.

[17] L. C. van der Gaag, H. L. Bodlaender, and A. Feelders. Monotonicity in Bayesian networks. In *Proceedings of the 20th Conference in Uncertainty in Artificial Intelligence (UAI 2004)*, pages 569–576, 2004.

[18] J. Švorc. Subjective well-being and individual material situation in four countries of Central Europe. *Sociológia*, 50:727–759, 2018.

[19] B. L. Welch. The generalisation of Student's problem when several different population variances are involved. *Biometrika*, 34(1/2):23—35, 1947.