Uncovering Relationships using Bayesian Networks: A Case Study on Conspiracy Theories

Jiří Vomlel VOMLEL@UTIA.CAS.CZ
Institute of Information Theory and Automation, Czech Academy of Sciences, CZ
Institute of Information Studies and Librarianship, Faculty of Arts, Charles University, CZ
Aleš Kuběna VOODOOLINEK@GMAIL.COM
Institute of Information Studies and Librarianship, Faculty of Arts, Charles University, CZ
Martin Šmíd SMID@UTIA.CAS.CZ
Institute of Information Theory and Automation, Czech Academy of Sciences, CZ
Institute of Information Studies and Librarianship, Faculty of Arts, Charles University, CZ
Josefina Weinerova JOSEFINA.WEINEROVA@NOTTINGHAM.AC.UK
School of Psychology, University of Nottingham, Nottingham, UK

Editors: J.H.P. Kwisthout & S. Renooij

Abstract

Bayesian networks (BNs) represent a probabilistic model that can visualize relationships between variables. We apply various BN structure learning algorithms to a large dataset from a Czech university entrance exam. This dataset includes a test of active, open-minded thinking designed by Jonathan Baron, as well as a test of students' attitudes toward various conspiracies. Using BNs, we were able to identify the structure of the conspiracies and their relationships with active open-minded thinking. We also compared results of different BN structure learning algorithms with results of selected standard data analysis methods.

Keywords: Bayesian Networks; Data Analysis; Structural Learning of Bayesian Networks; Actively Open-minded Thinking; Conspiracy Theories.

1. Introduction

Network models, which conceptualize behavior as a complex interplay of psychological and other variables, is increasingly used in the humanities, such as sociology and psychology. Pairwise Markov Random Fields (PMRFs), which are probabilistic models based on undirected graphs, appear to be the dominant probabilistic graphical model in these application areas (Epskamp et al., 2018). Bayesian networks (BNs) represent a conceptually different probabilistic model based on directed acyclic graphs. BNs are starting to gain attention also in sociology, psychology, psychometrics (Almond et al., 2016; Alvarez-Galvez, 2016; Švorc and Vomlel, 2019; Orsoni et al., 2024). One of the goals of this paper is to illustrate their potential to uncover relationships between studied concepts. We will show the advantages that BNs can offer in addition to traditional data analysis methods. For this purpose, we use a study on conspiracies and active open-mindedness. The application of BNs to this problem has raised few theoretical challenges that have implications for the structural learning of BNs. We experimentally compare conceptually different approaches to structural learning of BNs on the studied data.

In this paper, we use data from the Czech National Comparative Examinations[1], which replace or supplement entrance examinations to dozens of universities in Czechia and Slovakia. In the tests administered in 2022, in addition to questions assessing knowledge, students were also asked questions assessing their Actively Open-Minded Thinking (Baron, 2019, 1991) and their attitudes toward various conspiracy theories. Actively Open-minded Thinking is the disposition to be fair towards different conclusions, even if they contradict one's initial favorite conclusion. Jonathan Baron (2019) has proposed an assessment which includes 11 questions listed in Appendix A. Questions on 16 conspiracies were also included in the test. They are divided into five groups and are also listed in Appendix A. For the Baron's 11 questions, a standard five-point Likert scale was used, while for the 16 conspiracies, the Likert scale had seven values. We coded the values as integers $0, 1, \ldots, 4$ and $0, 1, \ldots, 6$, respectively. The dataset we used consists of 1411 vectors, where each vector represents a single student. The vector values represent answers to 11 Baron questions and opinions on 16 conspiracy theories. All variables in all data vectors are fully observed.

The paper is organized as follows. We begin with the applications of three standard data analysis methods to the studied dataset. In Section 2 we present two regression analysis of the conspiracy score versus Baron score. In Section 3 we analyse correlation matrix of individual Baron and conspiracy variables. Another insight is offered by the Principal Component Analysis in Section 4. The BN analysis forms Section 5. In this section, we compare four different BN models using the traditional scores and by their prediction accuracy in a ten-fold cross-validation experiment. We discuss what information can be gained from graphs of BNs. We conclude the paper with a discussion of related work and possible future work in Section 6.

2. Regression Analysis

Let C be the average of all of an individual's conspiracy responses and B be the average of all of an individual's responses to the Baron questions. We begin our analysis with a linear regression of C against the variable B. The Breusch-Pagan test rejects the null hypothesis of homoscedasticity with $p = 10^{-5}$ – the residual errors are higher for lower values of B. Next, we apply the logarithmic transformation to $C + 1$ and learn the regression model

$$\log C \quad = \quad 0.52 - 0.17 \cdot B \ . \tag{1}$$

The intercept and slope are both statistically significant with p-values less than $2 \cdot 10^{-16}$. The plot on the left of Figure 1 shows the regression line along with the 95% confidence interval. Each point corresponds to an individual, and its horizontal and vertical coordinates are values of B and $C + 1$, respectively. This model passes the Breusch–Pagan test of homoscedasticity with $p = 0.42$. In the middle of Figure 1 is the absolute residual error as a function of the Baron score. The distribution of residual errors of this model does not pass the normality test – it is slightly skewed to the left, see the histogram on the right of Figure 1. However, since the mean of this distribution is practically zero ($7 \cdot 10^{-15}$), the ordinary least squares estimates in (1) are reasonable estimators of the regression coefficients.

1. https://www.scioexams.com/

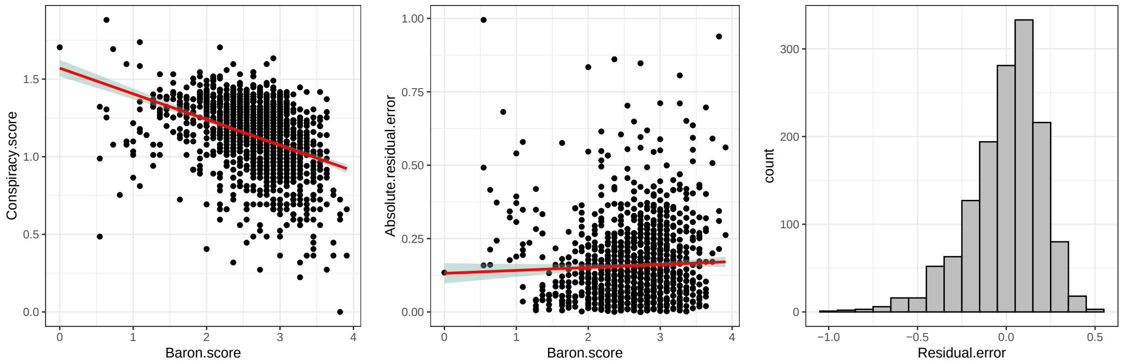


Figure 1: Conspiracy score ($\log C+1$) as a function of the Baron score B (left), the absolute residual error as a function of the Baron Score (middle), and the histogram of the residual error (right).

3. Correlation Analysis

Next, we examine pairwise correlations of individual questions. Figure 2 shows the upper triangles of the correlation matrices. The size and color intensity of the dots correspond to the correlation value. The larger the dot, the stronger the correlation. Positive correlation dots are blue, negative correlation dots are red. For the matrix on the left, the crosses indicate correlations that were not found to be significant at the 5% significance level. The more hypotheses that are tested on the same data, the higher the probability of getting false positives. This behavior is known as the multiple comparisons problem. To counteract this problem, we used the method of Holm (1979). The results are shown on the right side of Figure 2. While without Holm's correction only 95 of 351 null hypotheses (of a correlation) are rejected, with Holm's correction the number is almost doubled - there are 159 of rejected correlations. Of course, except for the crosses, the two matrices are identical. Let us take a closer look at the correlations that remain significant after the correction. We can see that all Baron variables are negatively correlated with conspiracies, all Baron variables are pairwise positively correlated with only one exception, the $\{B_9, B_{10}\}$ pair, and all conspiracies are pairwise positively correlated with a few exceptions related to C_7 (iron cannot evaporate).

Next, we apply a graph visualization method to the correlation matrix with correlations considered significant after Holm's correction. We use the algorithm of Fruchterman and Reingold (1991) to obtain a graph layout in which the distances between nodes correspond to the correlation values between these nodes, in the sense that the higher the absolute value of the correlations, the more attracted the nodes are to each other. For this we use the R package qgraph (Epskamp et al., 2012). See Figure 3. Green edges correspond to positive correlations while red edges correspond to negative correlations. The wider the edge, the higher the correlation value. In the figure, we can see that most of the Baron variables are close to each other, with the exception of B_3 and B_9. From the figure we can see that several conspiracies are not only closely related to each other (positively), but also to some Baron variables (negatively), with the strongest relationships being between $\{C_{11}, C_{14}, C_{15}\}$ and $\{B_5, B_7, B_8\}$. Some conspiracies are rather remote, e.g. C_{10}.

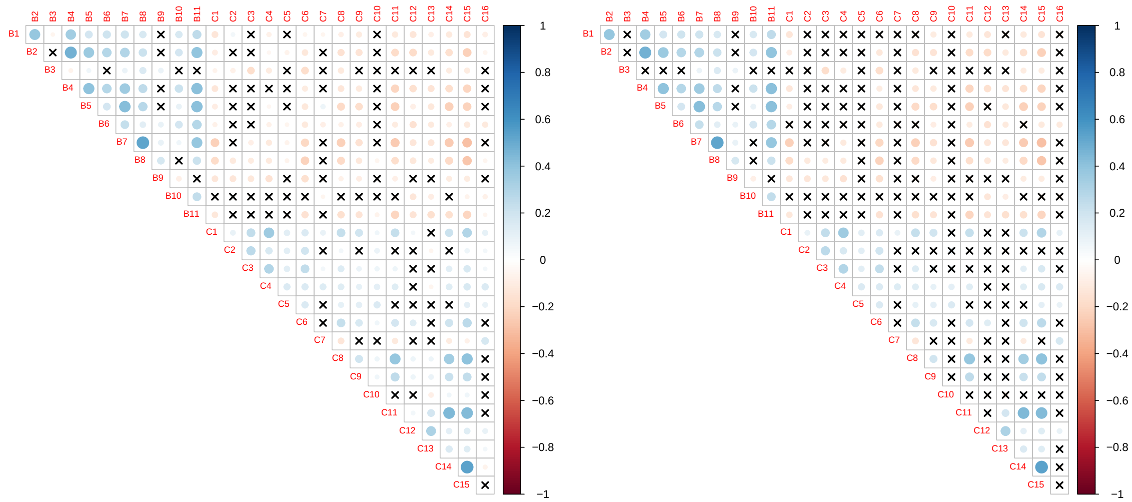


Figure 2: The upper triangles of the correlation matrices for the individual conspiracies and Baron questions. The crosses indicate correlations that are not significant at the 5% significance level – with no correction (left) and with Holm's correction (right).

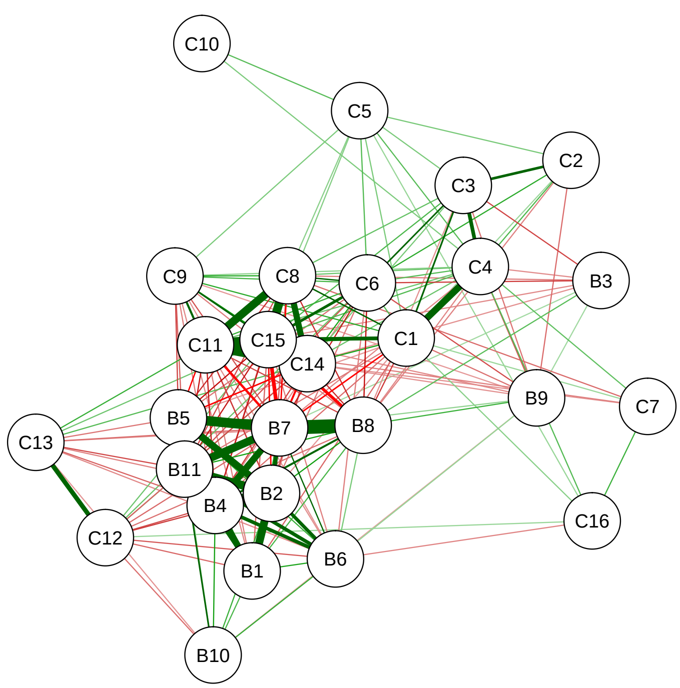


Figure 3: The undirected graph of the correlation matrix after Holm's correction.

4. Principal Component Analysis

Another exploratory data analysis technique is Principal Component Analysis (PCA) (Pearson, 1901). The data is linearly transformed into a new coordinate system so that the directions that capture the most variation in the data, called principal components, can be identified. PCA reveals that the first principal component explains more than 50% of the variance, while the second principal component explains only 9%. The plot in Figure 4 displays the individual contributions of the model variables to the first component (horizontal axis) and the second component (vertical axis). It is clearly visible that, with one exception[2] of C_7, the conspiracies constitute the first principal component with the negative value (left) and the Baron questions with the positive value (right). Note that the most negative values correspond to the hard core conspiracies, while the least negative values correspond to most of the questions from the shallow knowledge group plus C_{16} (homeopathy) from the soft conspiracies group.

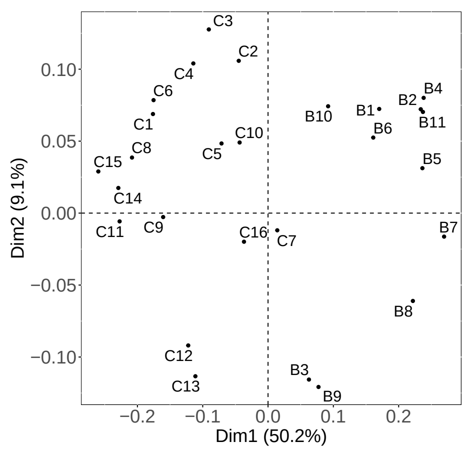


Figure 4: Contributions of individual variables to the first and second PCA components.

It is interesting to compare the relative positions of the nodes in Figures 3 and 4. The methods used to construct these plots are based on different principles – correlation analysis is based on pairwise relationships, while PCA provides a global view – but the intuition behind both methods is that nodes with similar behavior should be closer together. Therefore, it is not surprising that most of the nodes connected by an edge in Figure 3 are closer to each other than most of the other pairs in Figure 4.

5. Bayesian Networks Analysis

Bayesian networks (Pearl, 1988; Jensen and Nielsen, 2007; Koller and Friedman, 2009) are probabilistic models that use acyclic directed graphs to visualize the relationships between model variables. Specifically, the graph encodes conditional independence statements that can be derived from the graph using the d-separation criterion.

2. Note that C_7 was also an exception in the case of correlations.

5.1. Data Transformation

Recall that the individual variables are discrete with Likert scale values. To avoid overfitting problems when learning large conditional probability tables, we decided to transform the five-point Likert scale of the Baron variables and the seven-point Likert scale of the Conspiracy variables into a three-point scale. Let c denote the middle value of a Likert scale, in our case it is 2 for the five-point Likert scale and 3 for the seven-point Likert scale. We define the transformed value v as a function of the original value u as follows:

$$v = \begin{cases} -1 & \text{if } u < c\text{, meaning I disagree,} \\ 0 & \text{if } u = c\text{, meaning I am neutral, and} \\ +1 & \text{if } u > c\text{, meaning I agree.} \end{cases}$$

One question is whether this new scale should still be treated as an ordinal scale or rather as a nominal scale. We chose the latter option because the neutral value does not necessarily mean that the respondent's opinion is between agreement and disagreement, but this value can also indicate some reluctance to answer, which can be of additional value for the analysis. It also allows the learning algorithms more flexibility in learning the CPTs. The most common value for Baron variables is $+1$ with a relative frequency of 0.572. The most common value for conspiracy variables is -1 with a relative frequency of 0.547. These frequencies can be considered as lower bounds for the prediction accuracy of BN models to be considered successful.

5.2. Latent Class Analysis

The Latent Class (LCA) model assumes that each respondent can be assigned to a class such that respondents in the same class answer questions in the same way. An important task is to decide on the number of these classes so that the model best explains the observed data. As a measure of the quality of the model, we will use here and later the Bayesian Information Criterion (BIC) (Schwarz, 1978), which is defined as

$$BIC(BN, D) = LL(BN, D) - \frac{\log N}{2} \cdot d(BN) \tag{2}$$

where $LL(BN, D)$ is the log-likelihood of the data D given the BN model, N is the length of the data D (i.e., the number of data vectors), and $d(BN)$ is the number of parameters in BN. We trained the LCA models using the poLCA package (Linzer and Lewis, 2011) in R (R Core Team, 2021). Several other LCA methods could be applied to the studied problem, we decided to use the poLCA algorithm, which is a latent class analysis of polytomous nominal outcome variables with the aim of minimizing the BIC, a criterion we also use when learning other probabilistic models discussed later in this paper. The maximum BIC value is achieved by the LCA model with six classes.

In Table 1, we present the most probable values of the Baron and Conspiracy variables for six different states of the latent class variable in the LCA model (with six states). To the right of the latent class state, we show its prior probability. We can see that $LC = 1$ represents students who mostly respond positively to the Baron variables and, except for C_7, do not believe in conspiracies. The students represented by the state $LC = 2$ do not think that indecisiveness and uncertainty are OK (B_3), believe that the first answer

		Baron variables											Conspiracies															
LC	*p*	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.29	+	+	+	+	+	+	+	+	0	+	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−
2	0.41	+	+	−	+	+	+	+	+	−	+	+	−	+	+	+	+	−	+	−	−	+	−	−	−	−	−	+
3	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0.03	0	+	−	+	+	+	+	+	−	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0.17	0	0	−	0	+	0	0	0	−	0	+	−	+	0	+	+	−	+	−	−	+	−	−	−	−	−	+
6	0.07	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	+	−	+	−	−	+	−	−	−	−	−	+

Table 1: The most probable values of the Baron and Conspiracy variables given each of the six latent class states in the LCA model. To the right of the state is its prior probability p. The value $+1$ is abbreviated as $+$ and the value -1 is abbreviated as $-$.

is usually the best (B_9), and tend to accept few conspiracies (C_2, C_3, C_4, C_5, C_7, C_{10}, and C_{16}). The state $LC = 3$ corresponds to students with a neutral opinion about all variables and $LC = 4$ corresponds to students with a neutral opinion about conspiracies, both groups are rather rare. The last two states of LC correspond to students who tend to accept some conspiracies, and especially when $LC = 6$ they answer negatively most of the Baron variables. It is interesting to see that there is no state of LC for which the prevailing opinion would be positive towards any of the "Hard Core" conspiracies and towards the conspiracies of the Ignorance of New Facts (Covid-19) group. This means that the model can never predict the positive opinion about these conspiracies as the most probable, no matter what the observed state of the other variables is.

5.3. Tree Structured BNs

The second class of BN models considered in our study is the class of tree-structured BNs. These models are learned by the Chow-Liu algorithm (Chow and Liu, 1968), which greedily adds the edges with the highest mutual information among the connecting variables, unless the addition would create a cycle. These models are attractive because their structure is easy to read, they can be learned very efficiently, and the model's conditional probability tables (CPTs) can be reliably estimated. A potential weakness is that the structure may not be a good description of the problem under consideration due to its oversimplification. We will refer to these models as CLT models.

5.4. BNs Learned by Conditional Independence Tests

The third class represents BNs learned by the PC algorithm (Spirtes and Glymour, 1991). This algorithm is based on conditional independence tests and consists of two steps. First, the so-called graph skeleton is learned, and then the colliders (nodes where edges meet head-to-head) are identified. We used the NPC version of this algorithm, implemented in the Hugin tool (Madsen et al., 2003), with the conditional independence test with the default 5% significance level. We will refer to these models as NPC models.

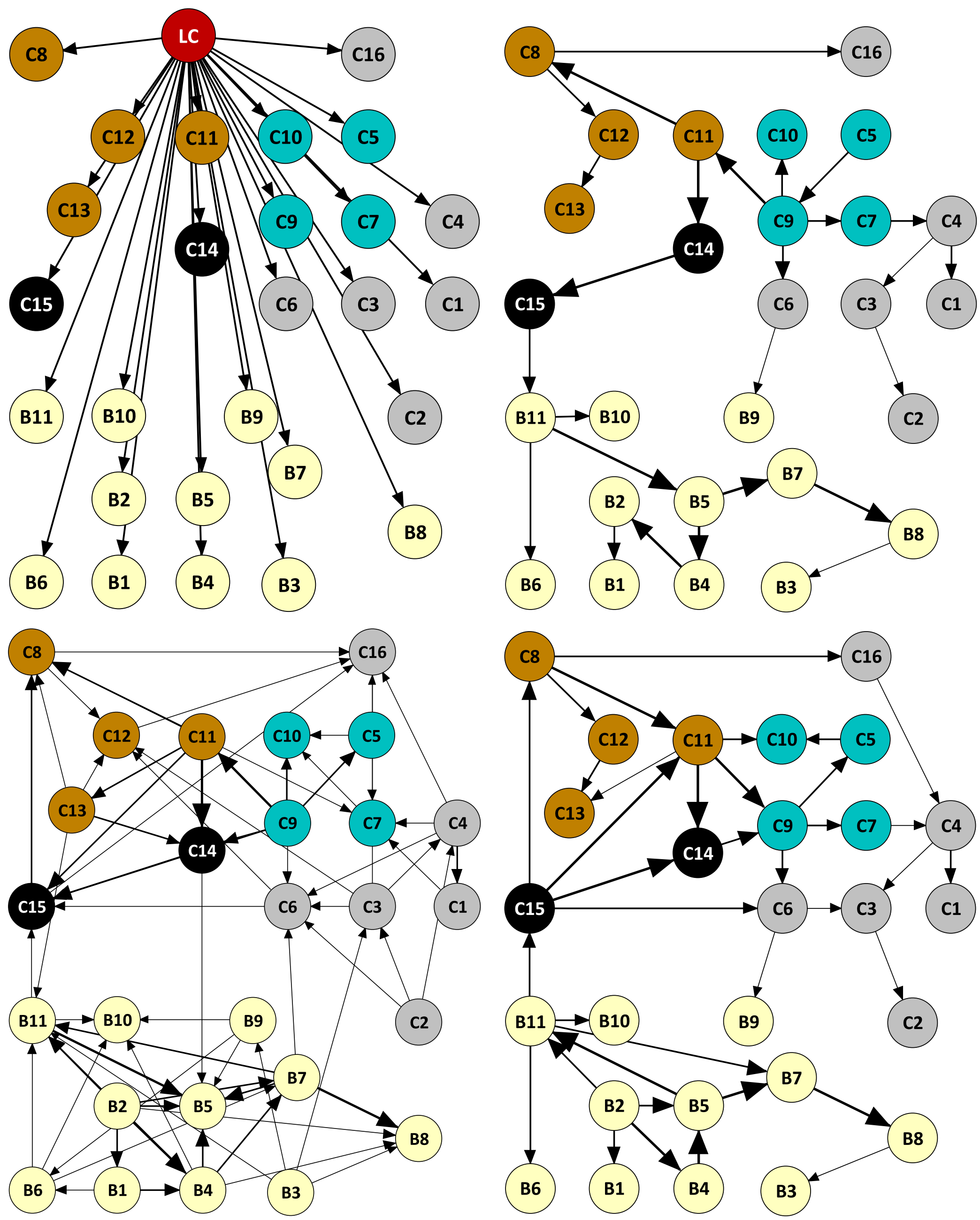


Figure 5: Comparison of graphs of BN models: LCA (top left), CLT (top right), NPC (bottom left), and BIC (bottom right).

5.5. BNs Minimizing BIC

BNs that minimize the BIC criterion (2) will be referred to as BIC models. It is known that learning BIC-optimal BNs is NP-hard. However, for smaller models[3] BIC-optimal BNs can be learned using the GOBNILP tool (Cussens and Bartlett, 2018). The class of BIC-optimal BNs is attractive because it represents a statistically sound trade-off between data fit and model complexity. These models are typically sparser, which can be seen as an advantage for their interpretability in the application domain. We also observe that direct relationships between variables are better represented in the BIC model than in the LCA model. For example, the LCA model discussed in Section 5.2 can never predict a positive opinion towards any of the "Hard Core" conspiracies and towards the Ignorance of New Facts (Covid-19) conspiracies regardless of the states of other variables, but in the BIC model $C_{15} = 1$ and $C_{11} = 1$ implies that C_{14} is most likely present.

5.6. Comparisons of Graphs and Their Layouts

In Figure 5 we compare the graph structures of the four BNs discussed above. The layout of the nodes was manually designed so that there are no crossing edges[4] and the edges are neither very long nor very short the BIC optimal BN. The layout is the same for all graphs.

The issue of BN graph layout deserves more attention, but a deeper analysis is beyond the scope of this paper. Various graph layout methods proposed in the literature could be (and are) applied to BN layout. Our concern is a layout that reveals important information from the modeled domain to the user. An open question seems to be what would be the most appropriate node distance measure for BNs, since the basic building block of BNs are families of nodes and not pairs of nodes as in the correlation graph discussed in Section 3.

In Figure 6, we compare two layouts of the BIC optimal BN – (a) our manual layout and (b) the layout of the correlation graph sfrom Figure 3. The graphs are the same, only the node positions differ. Layout (a) is more compact and easier to read, but layout (b) reveals additional information about the model variables, e.g. C_{10} and B_{10} are quite distant from the other variables, B_3 and B_9 do not have much in common with most of the other Baron variables, which otherwise form a fairly compact cluster, the conspiracies $C_{15}, C_{14}, C_{11}, C_8$ have a central position in the graph.

5.7. Uncovering Relationships Using BNs

Indeed, the BNs suggest that higher scores on Baron's questions imply a lower likelihood of conspiracy beliefs. This holds for all variables. Variables that are connected by a shorter path in the model graph tend to have a closer association. With the exception of the LCA model, which was not designed for this purpose, BNs help us gain deeper insight into the problem being modeled. For example, the graphs of the CLT and BIC models suggest that B_9: "The first answer may not be the best" directly reduces the probability of C_6: "Most problems have a clear cause and a simple solution. Also, B_{11} "Consider more than one possible answer" directly reduces the probability of C_{15}: "There are chemicals behind airplanes that affect human health. In the BIC and CLT models, these are the only links

3. Our models can be considered smaller.
4. This is possible because the learned graph is planar.

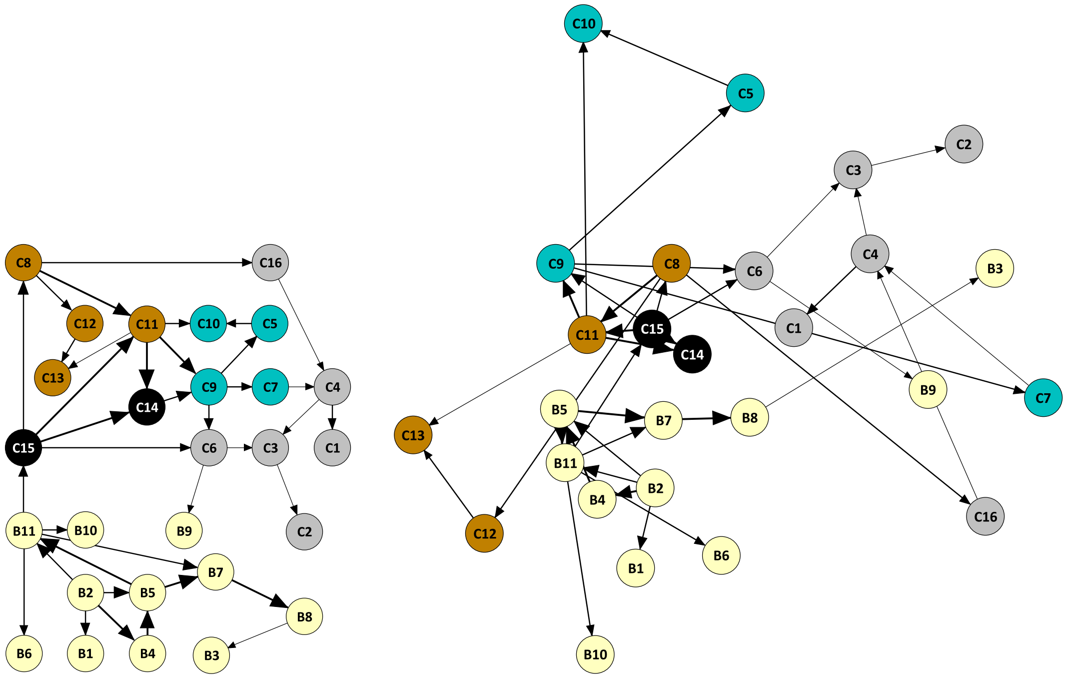


Figure 6: Comparison of two layouts of the BIC model.

between the Baron questions and conspiracies. This implies that if we know answers to B_{11} and B_9, other Baron questions are irrelevant to conspiracy beliefs.

In the BIC model, three conspiracies have the most links to other nodes: C_9: "100 million = billion" (5 edges), C_{11}: "Vaccinations cause autism" (6 edges), and C_{15}: "There are chemicals behind airplanes that affect human health" (5 edges). They can be considered as key conspiracies that characterize the tested individuals. Other measures of node centrality have been proposed in the literature (Opsahl et al., 2010). The node centrality is defined as the sum of the strengths of the edges adjacent to that node, where the edge strength is the absolute value of the edge weight. The edge weight can be defined as the correlation or mutual information of the nodes connected to the edge. Other centrality measures include closeness and betweenness based on the shortest path length between nodes in the network defined as a path minimizing the number of intermediate nodes. Betweenness of a node then measures the number of shortest paths that pass through that node.

An anonymous reviewer suggested using the size of Markov Blanket as a measure of node centrality. Markov Blanket of a node X is the set containing the parents of X, the children of X, and the other parents of all children of X. All other variables in the model are independent of X given the states of the nodes in the Markov Blanket. In Table 2 we can see that B_{11} has the largest Markov Blanket among the Baron variables and C_{11} among the conspiracies. This corresponds to their key role identified by other criteria above.

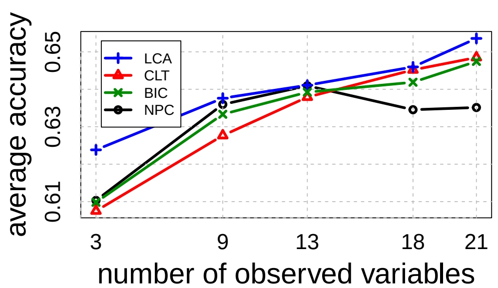
Baron variables											Conspiracies															
1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	4	1	2	4	1	3	2	1	1	**6**	1	1	3	5	3	4	1	4	5	1	**8**	3	2	3	6	2

Table 2: Size of Markov Blankets.

5.8. Experimental Comparisons of BN Models

We evaluated the ability of the models to predict states of certain variables given observations of other variables. Namely, we randomly select 3, 9, 13, 18, and 21 variables out of 26 model variables and instantiate them using values of a data vector from the test set. We then use the tested model to predict the states of the uninstantiated variables. We compare the predicted most probable state with the observed state in the data vector. We compute the average prediction accuracy over all predictions of the tested model. The process is repeated ten times, where in each experiment nine out of ten folds of data sets are used for model learning and the remaining fold is used for testing. The average results are summarized in Figure 7. The best performance[5] is achieved by the LCA model regardless of the number of observed variables. The CLT and BIC models also perform well, and their performance improves as more variables are observed. This is not the case for NPC, whose performance deteriorates with more observed variables. This is an indication that some CPTs in NPC were most likely too complex to be correctly estimated from the data.

We also compared the learned BN models by their fit to the data. All numbers reported are averages over 10 models learned from different training sets. First, we computed the average log-likelihood (LL) on the testing data[6]. Second, we computed the average Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) with respect to all data. The results are shown on the left side of Figure 7. The best values for each criterion are shown in bold. For all three criteria, the best model is LCA, followed by BIC.



	LCA	BIC	NPC	CLT
LL	**-3062**	-3110	-3518	-3112
AIC	**-30275**	-30684	-33994	-31140
BIC	**-31139**	-31446	-48053	-31555

Figure 7: Comparison of BN models prediction accuracy (left) and their fit to the data (right). LL is computed with respect to testing data, AIC and BIC with respect to all data.

5. The LCA accuracy of 0.654 is higher than 0.572 and 0.547, the average prevalence of the most common states of Baron and Conspiracy variables, respectively, see Section 5.1. It is still well below 1,000, but this is not surprising because of the natural uncertainty of the problem being modeled – two students who answered a subset of questions in the same way may naturally differ in their answers to some of the other questions.

6. This measures how well the probability distribution of the model fits the distribution of the testing data. The model with the best fit also maximizes predictive ability, measured as the distance of the conditional distributions given the observed states in the model and in the testing data.

6. Discussion

We applied several statistical and machine learning methods to the dataset representing students' attitudes towards active open-minded thinking and various conspiracies. Each of these methods has its own merits and has contributed to a complex view of the problem under study. This is particularly true for (1) regression analysis, which shows that active open-minded thinking, as measured by the Baron score, decreases belief in conspiracies, (2) visualization using the undirected graph of the correlation matrix after Holm's correction, where the distances correspond to the strength of the correlations, (3) the PCA method, that helps identify the key variables of the studied problem, and (4) Bayesian networks, where the LCA model helps identify the main clusters of students and has a good predictive ability, and the BIC optimal model has a good balance between simplicity and precision, so it can reveal what the key variables are and how they relate to each other. In addition, a BN GUI enables users to study the influence of variables of particular interest.

BN models can be further improved by considering the local structure of CPTs as Noisy-MIN, Noisy-MAX (Díez and Druzdzel, 2006; Díez and Galán, 2003), Noisy-Threshold (Vomlel and Tichavský, 2014), or Logistic Regression (Rijmen, 2008) models. To support this claim, we can use CPT $P(C_{11}|C_8, C_{15})$ from the BIC optimal model. For $C_{11} = 1$ to become more probable than $C_{11} = -1$, both parents must take state 1, the value 1 of one parent is not sufficient to switch to belief in the conspiracy C_{11}:

$$\begin{aligned} P(C_{11} = 1|C_8 = 1) &= 0.33, \\ P(C_{11} = 1|C_{15} = 1) &= 0.27, \quad \text{but} \\ P(C_{11} = 1|C_8 = 1, C_{15} = 1) &= 0.56 \;. \end{aligned}$$

This suggests that there is a synergy effect that may be reflected in some of the local models mentioned above. Another argument in favor of BNs with local structure of their CPTs is that they allow larger CPTs that increase the value of LL while still keeping the penalty low (Sharma et al., 2020; Vomlel et al., 2023).

In this paper, we have focused on a particular class of BN models, namely BNs based on discrete nominal data. Another natural class of BN models to consider are BNs assuming ordinal data, i.e., BNs that explicitly respect the order among the states of variables. We have already done preliminary experiments with the method of Luo et al. (2021). This method assumes that the ordinal variables arise from a marginal discretization of a set of Gaussian variables whose structural dependence in latent space follows a directed acyclic graph. The BNs learned by this method were relatively complex models with too many edges. We plan to further explore this direction by also considering an alternative method of Grzegorczyk (2024). Recently, BNs have been applied to the analysis of Likert scale responses (Orsoni et al., 2024), which are also the scales used in our study.

Acknowledgments

This work was supported by the European Regional Development Fund project "Beyond Security: Role of Conflict in Resilience-Building" (reg. no.: CZ.02.01.01/00/22_008/0004595). We are grateful to the reviewers for their insightful comments, which have greatly helped us in the improvement of our paper.

References

R. G. Almond, R. J. Mislevy, L. S. Steinberg, D. Yan, and D. M. Williamson. *Bayesian Networks in Educational Assessment*. Statistics for Social and Behavioral Sciences. Springer New York, NY, 2016. URL https://doi.org/10.1007/978-1-4939-2125-6.

J. Alvarez-Galvez. Discovering complex interrelationships between socioeconomic status and health in Europe: A case study applying Bayesian Networks. *Social Science Research*, 56:133–143, 2016. ISSN 0049-089X. URL https://doi.org/10.1016/j.ssresearch.2015.12.011.

J. Baron. Beliefs about thinking. In *Informal Reasoning and Education*, pages 169–186. Routledge, 1991. URL https://doi.org/10.4324/9780203052228.

J. Baron. Actively open-minded thinking in politics. *Cognition*, 188:8–18, 2019. ISSN 0010-0277. URL https://doi.org/10.1016/j.cognition.2018.10.004.

C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467, 1968. URL https://doi.org/10.1109/TIT.1968.1054142.

J. Cussens and M. Bartlett. GOBNILP, 2018. Version 1.6.3, https://www.cs.york.ac.uk/aig/sw/gobnilp/.

F. J. Díez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006. URL http://62.204.199.55/wordpress/wp-content/papercite-data/pdf/techreports/diez2006a.pdf.

F. J. Díez and S. F. Galán. An efficient factorization for the noisy MAX. *International Journal of Intelligent Systems*, 18:165–177, 2003. URL https://doi.org/10.1002/int.10080.

S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4):1–18, 2012. URL http://www.jstatsoft.org/v48/i04/.

S. Epskamp, G. K. J. Maris, L. J. Waldorp, and D. Borsboom. Network psychometrics, 2018. URL https://arxiv.org/abs/1609.02818.

T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991. URL https://doi.org/10.1002/spe.4380211102.

M. Grzegorczyk. Being Bayesian about learning Bayesian networks from ordinal data. *International Journal of Approximate Reasoning*, 170:109205, 2024. URL https://doi.org/10.1016/j.ijar.2024.109205.

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65—70, 1979. URL http://www.jstor.org/stable/4615733.

F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer New York, NY, 2 edition, 2007. URL https://doi.org/10.1007/978-0-387-68282-2.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

D. A. Linzer and J. B. Lewis. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1—-29, 2011. URL https://doi.org/10.18637/jss.v042.i10.

X. G. Luo, G. Moffa, and J. Kuipers. Learning Bayesian networks from ordinal data. *Journal of Machine Learning Research*, 22(266):1–44, 2021. URL http://jmlr.org/papers/v22/20-1338.html.

A. L. Madsen, M. Lang, U. B. Kjærulff, and F. Jensen. The Hugin tool for learning Bayesian networks. In T. D. Nielsen and N. L. Zhang, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 594–605, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. URL https://doi.org/10.1007/978-3-540-45062-7_49.

T. Opsahl, F. Agneessens, and J. Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010. URL https://doi.org/10.1016/j.socnet.2010.03.006.

M. Orsoni, M. Benassi, and M. Scutari. Information theory, machine learning, and Bayesian networks in the analysis of dichotomous and Likert responses for questionnaire psychometric validation. Technical report, OSF Preprints, January 2024. URL https://doi.org/10.31234/osf.io/r4y68.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. URL https://dl.acm.org/doi/10.5555/534975.

K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11): 559–572, 1901. URL https://doi.org/10.1080/14786440109462720.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

F. Rijmen. Bayesian networks with a logistic regression model for the conditional probabilities. *International Journal of Approximate Reasoning*, 48(2):659–666, 2008. URL https://doi.org/10.1016/j.ijar.2008.01.001. In memory of Philippe Smets (1938–2005).

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. URL http://www.jstor.org/stable/2958889.

C. Sharma, Z. A. Liao, J. Cussens, and P. van Beek. A score-and-search approach to learning Bayesian networks with noisy-or relations. In *Proceedings of the 10th International*

Conference on Probabilistic Graphical Models (PGM 2020), volume 138 of *Proceedings of Machine Learning Research*, pages 413–424, 2020. URL http://proceedings.mlr.press/v138/sharma20a/sharma20a.pdf.

P. Spirtes and C. Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9(1):62–72, 1991. URL https://doi.org/10.1177/089443939100900106.

J. Švorc and J. Vomlel. Bayesian networks for the analysis of subjective well-being. In *Proceedings of the 22nd Czech-Japan Seminar on Data Analysis and Decision Making (CJS'19)*, pages 175–188, 2019. URL http://cjs.utia.cas.cz/proceedings.pdf#section.0.20.

J. Vomlel and P. Tichavský. Probabilistic inference with noisy-threshold models based on a CP tensor decomposition. *International Journal of Approximate Reasoning*, 55(4): 1072–1092, 2014. URL https://doi.org/10.1016/j.ijar.2013.12.002.

J. Vomlel, V. Kratochvíl, and F. Kratochvíl. Structural learning of mixed noisy-OR Bayesian networks. *International Journal of Approximate Reasoning*, 161:108990, 2023. URL https://doi.org/10.1016/j.ijar.2023.108990.

Appendix A. Baron Questions

We use (R) to indicate reverse scoring.

(B_1) Allowing oneself to be convinced by a solid opposing argument is a sign of good character.
(B_2) People should take into consideration evidence that goes against conclusions they favor.
(B_3) Being undecided or unsure is the result of muddled thinking. (R)
(B_4) People should revise their conclusions in response to relevant new information.
(B_5) Changing your mind is a sign of weakness. (R)
(B_6) People should search actively for reasons why they might be wrong.
(B_7) It is OK to ignore evidence against your established beliefs. (R)
(B_8) It is important to be loyal to your beliefs even when evidence is brought to bear against them. (R)
(B_9) When we are faced with a new question, the first answer that occurs to us is usually best. (R)
(B_{10}) Good thinking leads to uncertainty when there are good arguments on both sides.
(B_{11}) When faced with a new question, we should consider more than one possible answer before reaching a conclusion.

Appendix B. Conspiracy Theories

The following is our translation of the original Czech text. We have classified the conspiracies into groups. You can see that some of the statements are not conspiracies, but rather lack of knowledge, etc. For the sake of simplicity, we will refer to all these statements as conspiracies. Again, we use (R) to indicate reverse evaluation.

“Hard core” Conspiracies

(C_{15}) The trail behind the jets is made up of chemicals released to affect human health or natural processes.
(C_{14}) Humans never actually landed on the moon, it was all recorded in TV studios.

Shallow Knowledge

(C_5) We are warmer in summer than in winter because the distance from the Earth to the Sun is smaller in the summer than in the winter.
(C_7) At very high temperatures, iron can evaporate. (R)
(C_{10}) An astronaut in a spacecraft is not affected by gravity because it’s far from Earth.
(C_9) 100 million is the same as one billion.

Ignorance of New Facts (Covid-19)

(C_{11}) Vaccinations cause autism.
(C_{12}) Obese people with diabetes are at high risk of complications from coronavirus infection. (R)
(C_{13}) A common complication of coronavirus infection is pneumonia. (R)
(C_8) Gargling salt water or lemon juice reduces the risk of contracting Covid 19.

Not so “Hard Core” (Soft) Conspiracies

(C_1) 5G networks may be more risky to human health or the environment than proven, longer-running wireless networks.
(C_2) Modern technology makes it possible to derive a near-complete diagnosis and find an effective treatment from analyzing a drop of a patient’s blood under an optical microscope.
(C_3) The cause of most of the so-called diseases of civilization is a toxic burden on the body or unrecognized multicellular parasites.
(C_4) Genetically modified foods (GMOs) threaten human health.
(C_{16}) Homeopathic drugs contain virtually no active ingredients. (R)

Simple Mind

(C_6) Most social problems have a clear cause and a simple solution.