On Identifiability of BN2A Networks

Iván Pérez[1,2] and Jiří Vomlel[1(✉)]

[1] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
{cabrera,vomlel}@utia.cas.cz

[2] Institute of Computer Science, Czech Academy of Sciences, Prague, Czechia

Abstract. In this paper, we consider two-layer Bayesian networks. The first layer consists of hidden (unobservable) variables and the second layer consists of observed variables. All variables are assumed to be binary. The variables in the second layer depend on the variables in the first layer. The dependence is characterised by conditional probability tables representing Noisy-AND or simple Noisy-AND. We will refer to this class of models as BN2A models. We found that the models known in the Bayesian network community as Noisy-AND and simple Noisy-AND are also used in the cognitive diagnostic modelling known in the psychometric community under the names of RRUM and DINA, respectively. In this domain, the hidden variables of BN2A models correspond to skills and the observed variables to students' responses to test questions. In this paper we analyse the identifiability of these models. Identifiability is an important concept because without it we cannot hope to learn correct models. We present necessary conditions for the identifiability of BN2As with Noisy-AND models. We also propose and test a numerical approach for testing identifiability.

Keywords: Bayesian networks · BN2A networks · Cognitive Diagnostic Modeling · Psychometrics · Model Identifiability

1 Introduction

Bayesian networks [10,12,13] are a popular framework for modelling probabilistic relationships between random variables. The topic of this paper is the learning of a special class of Bayesian Networks (BNs) - two-layer BNs, where the first layer consists of hidden (unobservable) variables, which are assumed to be mutually independent, and the second layer consists of observed variables. All variables are assumed to be binary. The variables in the second layer depend only on the variables in the first layer. The dependence is characterised by conditional probability tables (CPTs), which represent either Noisy-AND or simple Noisy-AND. In case the CPTs are represented by Noisy-OR models, the corresponding

This work was supported by grants 22-11101S and 21-03658S of the Czech Science Foundation.

Z. Bouraoui and S. Vesic (Eds.): ECSQARU 2023, LNAI 14294, pp. 136–148, 2024.
https://doi.org/10.1007/978-3-031-45608-4_11

BN is traditionally called BN2O [1], in case the CPTs are represented by Noisy-AND models, the corresponding BN will be called BN2A as a parallel to the BN2O models. In Fig. 1 we give an example of a directed bipartite graph that can define the structure of a BN2O or a BN2A model.

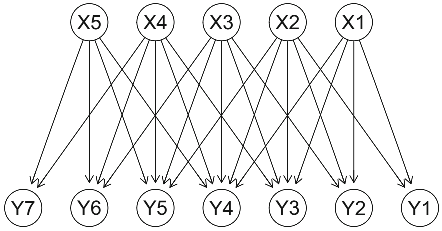

Fig. 1. An example of a directed bipartite graph.

Noisy-AND and simple Noisy-AND models are examples from the family of canonical models of CPTs [3,9]. The study of these models is motivated by practical applications. BN2O models are well suited for medical applications, where the hidden variables of the first layer correspond to diseases and the observed variables of the second layer correspond to observed symptoms. In this application, it is natural to assume that a symptom will occur if the patient has a disease that causes that symptom, unless its influence is inhibited with some probability. Therefore the CPTs are modelled using Noisy-OR models. BN2A models are used in psychometrics for cognitive diagnostic modelling of students. In this case, the hidden variables correspond to the student's skills and the observed variables correspond to the student's responses to test questions. A typical test question requires all related skills to be present, unless a missing skill is compensated by another knowledge or skill. This relationship is well represented by Noisy-AND models.

The work most closely related to ours is [5], but the main difference is that it assumes all hidden variables can be mutually dependent, whereas we assume that all hidden variables are mutually independent. The legitimacy of this assumption depends on the context of the application. Our motivation for the independence of the hidden variables is the ability to clearly distinguish between them and their effect on the observed variables. The assumption of hidden node independence has a significant impact on the identifiability of the model. In addition, BN2A with leaky Noisy-AND, corresponding to RRUM in CDM, has not been analysed in [5].

This paper is structured as follows. In Sect. 2 we formally introduce the BN2A models. First, we discuss both options for CPTs, leaky Noisy-AND and simple Noisy-AND models, but in the rest of the paper we restrict our analysis to leaky Noisy-AND. In Sect. 3, we analyse the identifiability of BN2A models, since identifiability is an important issue for models with hidden variables. Several conditions for the identifiability of these models are given in this section. Testing the identifiability condition based on the rank of the Jacobian matrix is practically non-trivial, so we propose and test a numerical approach in Sect. 4. Finally, we summarise the contribution of this paper in Sect. 5.

2 BN2A Models

Let $\mathbf{X}$ denote the vector $(X_1, \ldots, X_K)$ of K hidden variables, and similarly let $\mathbf{Y}$ denote the vector $(Y_1, \ldots, Y_L)$ of L observed dependent variables. The hidden variables are also called attributes or skills in the context of cognitive diagnostic models (CDMs), or diseases in the context of medical diagnostic models (MDMs). The observed dependent variables are also called items in CDMs or symptoms in MDMs. All variables are assumed to be binary, taking states from $\{0, 1\}$. The state space of the multidimensional variable $\mathbf{X}$ is denoted $\mathbb{X}$ and is equal to the Cartesian product of the state spaces of $X_k, k = 1, \ldots, K$:

$$\mathbb{X} = \times_{k=1}^{K} \mathbb{X}_k = \{0, 1\}^K \ . \tag{1}$$

Similarly, the state space of multidimensional variable $\mathbf{Y}$ is denoted $\mathbb{Y}$ and is equal to the Cartesian product of state spaces of $Y_\ell, \ell = 1, \ldots, L$:

$$\mathbb{Y} = \times_{\ell=1}^{L} \mathbb{Y}_\ell = \{0, 1\}^L \ . \tag{2}$$

The basic building blocks of a BN2A model are conditional probability tables (CPTs) specified in the form of a Noisy-AND model. Let Y_ℓ be an observed dependent variable and $pa(Y_\ell)$ be the subset of indexes of related variables from $\mathbf{X}$. They are referred to as the parents of Y_ℓ.

Definition 1 (Noisy-AND model).
A conditional probability table $P(Y_\ell|\mathbf{X}_{pa(Y_\ell)})$ represents a Noisy-AND model if

$$P(Y_\ell = y_\ell|\mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)}) = \begin{cases} q_{\ell,0} \cdot \prod_{i \in pa(Y_\ell)} (q_{\ell,i})^{(1-x_i)} & \text{if } y_\ell = 1 \\ 1 - q_{\ell,0} \cdot \prod_{i \in pa(Y_\ell)} (q_{\ell,i})^{(1-x_i)} & \text{if } y_\ell = 0. \end{cases} \tag{3}$$

Note that if $x_i = 1$ then $(q_{\ell,i})^{(1-x_i)} = 1$ and if $x_i = 0$ then $(q_{\ell,i})^{(1-x_i)} = q_{\ell,i}$. The interpretation is that if $X_i = 1$, then this variable definitely enters the AND relation with the value 1. If $X_i = 0$, then there is still a probability $q_{\ell,i}$ that it enters the AND relation with value 1. The model also contains an auxiliary parent X_0 which is always 0 and thus enters the AND relation with probability $q_{\ell,0}$ for the value 1. This probability is traditionally called *leak* probability and allows non-zero probability of $Y_\ell = 0$ even if all parents of Y_ℓ have value 1 ($Y_\ell = 1$ if and only if all parents enter the AND relation with value 1). In CDM, this model is known as the Reduced Reparametrized Unified Model (RRUM) [7] and it is a special case of the Generalized Noisy Inputs, Deterministic AND (GNIDA) gate model [2].

It is convenient to extend the vector $\mathbf{x}$ with the value 0 as its first element, i.e., we redefine $\mathbf{x} = (0, x_1, \ldots, x_K)$ so that we can write the formula (3) as

$$P(Y_\ell = y_\ell | \mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)}) = \begin{cases} \prod_{i \in \{0\} \cup pa(Y_\ell)} (q_{\ell,i})^{(1-x_i)} & \text{if } y_\ell = 1 \\ 1 - \prod_{i \in \{0\} \cup pa(Y_\ell)} (q_{\ell,i})^{(1-x_i)} & \text{if } y_\ell = 0. \end{cases} \tag{4}$$

The prior probability of the hidden attribute for $k = 1, \ldots, K$ is defined as

$$P(X_k = x_k) = (p_k)^{x_k} (1 - p_k)^{(1-x_k)} \ , \tag{5}$$

which means that if $x_k = 1$ then it is p_k and if $x_k = 0$ then it equals $1 - p_k$.

Another model of a CPT commonly known in the area of CDM as Deterministic Input Noisy AND (DINA) gate [11], corresponds to a CPT model called Simple Noisy-AND model in the context of canonical models of BNs [3].

Definition 2 (Simple Noisy-AND model). *A conditional probability table $P(Y_\ell | \mathbf{X}_{pa(Y_\ell)})$ represents a Simple Noisy-AND model if*

$$\begin{aligned} & P(Y_\ell = y_\ell | \mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)}) \\ & = \begin{cases} (1 - s_\ell)^{\pi(\mathbf{x}, Y_\ell)} \cdot (g_\ell)^{1-\pi(\mathbf{x}, Y_\ell)} & \textit{if } y_\ell = 1 \\ 1 - (1 - s_\ell)^{\pi(\mathbf{x}, Y_\ell)} \cdot (g_\ell)^{1-\pi(\mathbf{x}, Y_\ell)} & \textit{if } y_\ell = 0 \ , \end{cases} \end{aligned} \tag{6}$$

where

$$\pi(\mathbf{x}, Y_\ell) = \prod_{i \in pa(Y_\ell)} x_i \ . \tag{7}$$

In the context of CDMs, the parameter s_ℓ represents the so-called slip probability, i.e. the probability of giving incorrect answer despite all required skills were present. The parameter g_ℓ represents guessing probability, i.e. the probability of guessing the correct answer despite the absence of a required skill. Due to space constraints, we will not analyze the simple noisy-AND model in this paper. We present its definition to show how it differs from leaky noisy-AND and link it to the existing literature in the CDM and BN communities.

Now we are ready to define a special class of Bayesian network models with hidden variables, called BN2A models.

Definition 3 (BN2A model). *A BN2A model is a pair (G, P), where G is a directed bipartite graph with its nodes divided into two layers. The nodes of the first layer correspond to the hidden variables $X_1, \ldots, X_K$ and the nodes of the second layer correspond to the observed variables $Y_1, \ldots, Y_L$. All edges are directed from a node of the first layer to a node of the second layer. The symbol P refers to the joint probability distribution over the variables corresponding to the* *nodes of the graph G. The probability distribution is parameterized by a vector of model parameters* $\mathbf{r}$*:*

$$\mathbf{r} \quad = \quad (\mathbf{p},\mathbf{q}) = \left((p_k)_{k\in\{1,\ldots,K\}},(q_{\ell,k})_{\ell\in\{1,\ldots,L\},k\in\{0\}\cup pa(Y_\ell)}\right) \quad . \tag{8}$$

We will use $E(G)$ to denote the set of edges of a bipartite graph G and $V_1(G)$ and $V_2(G)$ as the sets of nodes of the first layer and the second layer of G, respectively. The bipartite graph G can also be specified by an incidence matrix and in the context of CDM is traditionally denoted by Q. A Q-matrix is an $L \times K$ binary matrix, with entries $Q_{l,k} \in \{0,1\}$ that indicate whether or not the ℓ^{th} observed dependent variable is linked to the k^{th} hidden variable.

Definition 4 (The joint probability distribution of a BN2A model). *The joint probability distribution of a BN2A model is defined for all* $(\mathbf{x},\mathbf{y}), \mathbf{x} \in \mathbb{X}, \mathbf{y} \in \mathbb{Y}$ *as*[1]

$$P(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}) = \prod_{\ell=1}^{L} P(Y_\ell = y_\ell|\mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)}) \cdot \prod_{k=1}^{K} P(X_k = x_k). \tag{9}$$

Conditional probabilities $P(Y_\ell = y_\ell|\mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)})$ *for* $\ell = 1,\ldots,L$ *are leaky Noisy-AND models and* $P(X_k)$ *for* $k=1,\ldots,K$ *are independent prior probabilities of hidden variables.*

The joint probability distribution over the observed variables of a BN2A model for all $\mathbf{y} \in \mathbb{Y}$ is computed as

$$P(\mathbf{Y}=\mathbf{y}) = \sum_{\mathbf{x}\in\mathbb{X}} \left(\prod_{\ell=1}^{L} P(Y_\ell = y_\ell|\mathbf{X}_{pa(Y_\ell)} = \mathbf{x}_{pa(Y_\ell)}) \cdot \prod_{k=1}^{K} P(X_k = x_k)\right). \tag{10}$$

3 Identifiability of BN2A

The models we call BN2A have recently gained great interest in many areas, including psychological and educational measurement, where subjects/individuals need to be classified according to hidden variables based on their observed responses (to test items, questionnaires, etc.). For these models, identifiability affects the classification of subjects according to their hidden variables, which depends on the precision of the parameter estimates. With non-identifiable models we can lead to erroneous conclusions about subjects' classification.

A parametric statistical model is a mapping from a finite-dimensional parameter space $\Theta \subseteq \mathbb{R}^d$ to a space of probability distributions, i.e.

$$p : \Theta \to P_\Theta, \quad \theta \mapsto p_\theta \quad . \tag{11}$$

[1] Symbol $\mathbf{x}_{pa(\ell)}$ denotes the subvector of $\mathbf{x}$ whose values corresponds to variables $X_i, i \in pa(Y_\ell)$.

The model is the image of the map p, and it is called *identifiable* if the parameters of the model can be recovered from the probability distributions, that is, if the mapping p is one-to-one.

Following [5] we define the joint strict identifiability of the BN2A model, which is the identifiability of the model structure (represented by a bipartite graph or equivalently by a Q-matrix) as well as the model parameters.

Definition 5 (Joint Strict Identifiability). *A BN2A model (G, P) is strictly identifiable if there is no BN2A model (G', P') with $G' \neq G$ or $P \neq P'$ or both, except for a permutation of hidden variables, such that for all $\mathbf{y} \in \mathbb{Y}$*

$$P(\mathbf{Y} = \mathbf{y}) = P'(\mathbf{Y} = \mathbf{y}) \ . \tag{12}$$

Joint identifiability may be too restrictive, for example, in cases where it considers as unidentifiable models where only very few model parameter values cause models to be unidentifiable. Therefore a weaker concept seems more practical.

Definition 6 (Joint Generic Identifiability). *A BN2A model (G, P) is generically identifiable if the set of P' of BN2A models (G', P') violating condition (12) has Lebesgue measure zero.*

Table 1. Results of the first three questions from the Mathematics Matura Exam

	$Y_2 = 0$		$Y_2 = 1$	
	$Y_3 = 0$	$Y_3 = 1$	$Y_3 = 0$	$Y_3 = 1$
$Y_1 = 0$	1517	2403	203	1121
$Y_1 = 1$	875	4482	241	3614

To illustrate the importance of the concept of identifiability, consider the data in Table 1 representing the results of the first three questions of the Mathematics Matura Exam - a national secondary school exit exam in Czechia. The table represents the results of $n = 14456$ subjects who took the exam in the spring of 2021. The values 0 and 1 correspond to a wrong and a correct answer, respectively. Next, we analyze two examples of BN2A models which are graphically represented in Fig. 2.

Example 1 (Identifiability). In this example we consider model (a) from Fig. 2. This model will be referred as model 1-3-1 in Table 2 where it corresponds to its third column. We can see the Table 1 as a $2 \times 2 \times 2$ tensor, this tensor has rank 2, then we can decompose it using Algorithm 1 from [6]. From this decomposition, we can recover the parameters of the 1-3-1 model from the system of seven equations for $\mathbf{y} \in \{0, 1\}^3 \backslash (1, 1, 1)$:

$$P(\mathbf{y}) = p_1 \prod_{i=1}^{3} q_{i0}^{y_i} (1 - q_{i0})^{1-y_i} + (1 - p_1) \prod_{i=1}^{3} (q_{i0} q_{i1})^{y_i} (1 - q_{i0} q_{i1})^{1-y_i} \ , \tag{13}$$

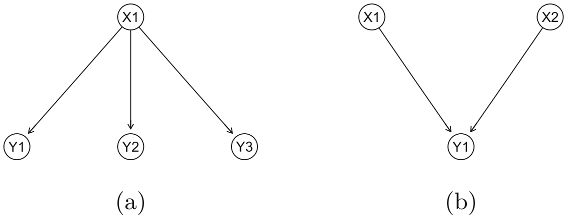

Fig. 2. BN2A models from Example 1 and Example 2

where $P(\mathbf{y})$ for $\mathbf{y} \in \{0,1\}^3 \backslash (1,1,1)$ are computed as relative frequencies from Table 1. By solving this system of equations we get:

$$\mathbf{r} = (p_1, q_{10}, q_{20}, q_{30}, q_{11}, q_{21}, q_{31}) \approx (0.317, 0.522, 0.903, 0.679, 0.086, 0.576, 0.318).$$

Since the solution is unique, then this model is identifiable, i.e., the vector $\mathbf{r}$ is uniquely determined from the data presented in Table 1.

Example 2 (Non-Identifiability). Now we will consider model (b) from Fig. 2. This model will be referred as model 2-1-2 in Table 2 where it corresponds to its fourth column. Again, we use the data presented in Table 1 to compute the probability distribution of Y_1 as its relative frequency. In this way we get $P(Y_1 = 1) \approx 0.637$, which enters the left hand side of equation

$$P(Y_1 = 1) = \sum_{\mathbf{x} \in \mathbb{X}_1 \times \mathbb{X}_2} \left((p_k)^{x_k} (1 - p_k)^{(1-x_k)} \prod_{i \in \{0\} \cup pa(Y_\ell)} (q_{\ell,i})^{(1-x_i)} \right) . \quad (14)$$

The five parameters of this model must satisfy just this equation, therefore we can fix some parameters and find different solution vectors $\mathbf{r_1}$ and $\mathbf{r_2}$. For example, both of the following vectors satisfy (14)

$$\mathbf{r_1} = (p_1, p_2, q_{10}, q_{11}, q_{12}) \approx (\mathbf{0.716}, \mathbf{0.9}, 0.8, 0.4, 0.6)$$
$$\mathbf{r_2} = (p_1, p_2, q_{10}, q_{11}, q_{12}) \approx (\mathbf{0.842}, \mathbf{0.7}, 0.8, 0.4, 0.6) \ .$$

In this model, X_1 and X_2 represent two skills needed to correctly answer question Y_1. Both parameter vectors, $\mathbf{r_1}$ and $\mathbf{r_2}$, satisfy the model, but while in $\mathbf{r_1}$ the prior probability of X_1 is smaller than the prior probability of X_2, in $\mathbf{r_2}$ the opposite is true. In this case, the model is non-identifiable.

Remark 1. The fact that the number of hidden variables is greater than the number of observed variables, as in Example 2, is not a condition for a model to be non-identifiable. For this we can consider the 6-5-2 model, following the pattern of Table 2 we can see that the number of parameters ($R = 15$) is less than the number of free parameters of the joint probability distribution over the observed variables ($S = 31$), then, according to Theorem 1, this model could be identifiable.

Various methods have been proposed to check the identifiability – one common approach is to estimate the dimension of the image of the mapping p. This is usually done by computing the rank of the Jacobian matrix of p [14].

Now, we will specify the Jacobian matrix of a BN2A network representing a probability distribution $P(\mathbf{y}), \mathbf{y} \in \mathbb{Y}$. Each row of the Jacobian matrix corresponds to one configuration $\mathbf{y} \in \mathbb{Y}$ of the multivariable $\mathbf{Y}$. Let $S = |\mathbb{Y}|$ denote the number of configuration of $\mathbf{Y}$. Each column of the Jacobian matrix corresponds to an element of the parameter vector $\mathbf{r}$ whose number of entries is given by

$$R = K + L + \sum_{\ell=1}^{L} M_\ell \quad \text{where } M_\ell = |pa(Y_\ell)| \ . \tag{15}$$

For $k = 1, \ldots, K$

$$\frac{\partial P(\mathbf{y})}{\partial p_k} = \sum_{\mathbf{x} \in \mathbb{X}} \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{(p_k)^{x_k}(1 - p_k)^{1 - x_k}} \tag{16}$$

and for $\ell = 1, \ldots, L$ and $k \in \{0\} \cup pa(Y_\ell)$

$$\frac{\partial P(\mathbf{y})}{\partial q_{\ell,k}} = \sum_{\mathbf{x} \in \mathbb{X}} (1 - x_k) \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{(q_{\ell,k})^{y_\ell}(1 - q_{\ell,k})^{(1 - y_\ell)}} \ . \tag{17}$$

Note that the terms presented in the denominators of all the fractions in formulas (16) and (17) are also present in the corresponding numerators of these fractions. That is, they only serve to cancel the corresponding term from the numerator of the fraction. Thus, the Jacobian matrix is

$$J = \begin{pmatrix} \frac{\partial P(\mathbf{y}_1)}{\partial r_1} & \ldots & \frac{\partial P(\mathbf{y}_1)}{\partial r_R} \\ \ldots & \ldots & \ldots \\ \frac{\partial P(\mathbf{y}_S)}{\partial r_1} & \ldots & \frac{\partial P(\mathbf{y}_S)}{\partial r_R} \end{pmatrix} \ . \tag{18}$$

Since the mapping from the parameter space to the probability space over the observable variables is a polynomial, we get the following lemma as a special case of Theorem 1 from [4].

Lemma 1. *The rank of the Jacobian matrix J of a BN2A model is equal to an integer constant r almost everywhere.*[2]

The rank condition is intuitively clear but practically non-trivial to apply. As the number of variables increases, the dimension of the Jacobian matrix grows rapidly. For the second smallest model from Table 2 that could be identifiable, the dimension of J matrix is 15×14, and each entry contains a degree 13 polynomial. However, simple checks can be performed to quickly rule out identifiability can be performed. Next, we give a necessary condition for the identifiability of a BN2A model.

[2] The set of exceptions has Lebesgue measure zero.

Theorem 1. *Let K be the number of hidden variables in the first layer of a BN2A model with CPTs represented by Noisy-AND models, L the number of observed variables in the second layer, and $M_\ell = |pa(Y_\ell)|$ for $\ell = 1, \ldots, L$. If*

$$R \;=\; K + L + \sum_{\ell=1}^{L} M_\ell > 2^L - 1 \;=\; S \tag{19}$$

then the BN2A model is NOT identifiable.

Proof. For the rank of any matrix, it holds that it is lower or equal to the minimum of the number of columns and the number of rows. Recall that S is the number of rows of the Jacobian matrix J and R is the number of model parameters and also the number of columns of J. If $R > S$ then the rank of the Jacobian matrix J is lower than the number of model parameters and the BN2A model is not identifiable. □

In Fig. 3 we visualize the necessary condition for the identifiability of BN2A models from Theorem 1 for $K = L - 1, L, L + 1$. The minimal S corresponds to models with the minimum number of parents $pa(Y_\ell)$ (which is greater than or equal to one) and the maximal S to models with the maximum number of parents $pa(Y_\ell)$ which is K. This means that the actual value of R is always between the blue and red lines. There is a threshold value of L for model identifiability so that no model with a lower L is identifiable (in Fig. 3 it is 3, 4, and 5, respectively). On the other hand, if L is greater than another threshold value (in Fig. 3 this threshold is 5, 6, and 6, respectively) then Theorem 1 does not rule out any BN2A models as unidentifiable.

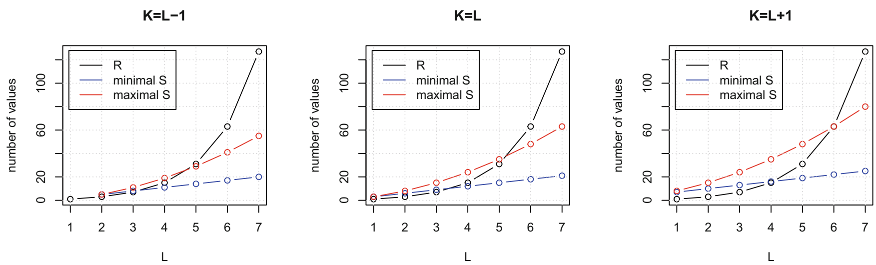

Fig. 3. Minimal S, maximal S, and R for BN2A models with $K = L - 1, L, L + 1$

In Table 2 we give examples of BN2A models, all of which have the same number of parents $|pa(Y_\ell)| = M_\ell = M$ for all $\ell \in \{1, \ldots, L\}$. Note that $|pa(Y_\ell)| \leq K$ and if it holds with equality then the BN2A model is fully connected. This table indicates the identifiability according to Theorem 1, i.e., if the number of free BN2A parameters R is greater than the number of free parameters of the joint probability distribution over the observed variables S, then BN2A is not identifiable. The columns corresponding to BN2A models that satisfy the necessary identifiability condition of Theorem 1 are printed with a gray background.

Table 2. Examples of different BN2A models. Columns printed with a gray background correspond to models for which Theorem 1 does not exclude their identifiability.

K	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3
L	1	2	3	1	2	3	4	1	2	3	4	5	3	4	5
M	1	1	1	2	2	2	2	3	3	3	3	3	2	2	2
$\|\mathbf{p}\| = K$	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3
$\|\mathbf{q}\| = L \cdot (M+1)$	2	4	6	3	6	9	12	4	8	12	16	20	9	12	15
$R = \|\mathbf{p}\| + \|\mathbf{q}\|$	3	5	7	5	8	11	14	7	11	15	19	23	12	15	18
$S = 2^L - 1$	1	3	7	1	3	7	15	1	3	7	15	31	7	15	31

Using algebraic manipulations on the smallest BN2A model from Table 2 that satisfies the necessary condition of Theorem 1 ($K = 1$, $L = 3$, $M = 1$), we observe that the corresponding Jacobian matrix has the full rank almost everywhere. This model is identifiable if its parameters satisfy the conditions of Theorem 2.

Theorem 2. *The Jacobian matrix of the BN2A model with $K = 1$, $L = 3$, and $M = |pa(Y_\ell)| = 1$ for $\ell = 1, 2, 3$ has the full rank if and only if*

$$0 < p_1 < 1 \tag{20}$$
$$0 < q_{\ell,0} \leq 1 \text{ for } \ell = 1,2,3 \tag{21}$$
$$0 \leq q_{\ell,1} < 1 \text{ for } \ell = 1,2,3. \tag{22}$$

Proof. We will compute the determinant of the Jacobian matrix with seven rows corresponding to seven configurations $\mathbf{y}$ from $\{0,1\}^3 \backslash (1,1,1)$ and seven columns corresponding to seven model parameters of vector $\mathbf{r}$. The BN2A model with $K = 1$, $L = 3$, and $M = 1$ is identifiable iff $rank(J) = 7$. Using algebraic manipulations we get the determinant of the Jacobian matrix

$$\det J = -p_1^3 \cdot (1 - p_1)^3 \cdot \prod_{\ell=1}^{L} q_{\ell,0}^3 \cdot (1 - q_{\ell,1})^2 \ . \tag{23}$$

From this formula, it follows that the determinant is non-zero and, consequently, $rank(J) = 7$ if and only if the assumptions of Theorem 2 are satisfied. □

The following lemma indicates that adding an edge from X_k to Y_ℓ with $q_{\ell,k} = 1$ cannot make the model identifiable.

Lemma 2. *If the rank of the Jacobian matrix J of a BN2A model (G, P) is less than the number of its model parameters R, then the rank of the Jacobian matrix J' of a BN2A model (G', P') with $E(G)' = E(G) \cup \{X_k \to Y_\ell\}$ and with $q_{\ell,k} = 1$ is also less than the number of its model parameters.*

Proof. Note that if $q_{\ell,k} = 1$ then $(q_{\ell,k})^{(1-x_i)} = 1$ for both $x_i = 1$ and $x_i = 0$. This means that the first R columns of the new Jacobian matrix J' are equivalent to the columns of J. Only one new column is added to J, so the rank of J' is at most the rank of J plus one. The rank of J is less than the number of parameters of the BN2A model (G, P), so the rank of J' is also less than the number of parameters of the BN2A model (G', P'). $\square$

4 Computational Experiments

Lemma 1 ensures that the rank of J is a constant almost everywhere. Therefore, we can use the idea proposed in [8] to compute the rank of the Jacobian matrix numerically. We choose one hundred random points in the parameter space and compute the determinant of the Jacobian matrix (and its submatrices if necessary) at these points. In this way, one can almost certainly determine the maximum rank of the Jacobian matrix.

In the next three examples, we apply this approach to the analysis of the simplest BN2A models from Table 2 that were not ruled out as identifiable. We use computations in rational arithmetic using Mathematica software to avoid rounding errors. This arithmetic is of infinite precision, which is important when deciding whether the determinant is exactly zero.

Example 3. Let us take a closer look at the BN2A model for $K = 2$, $L = 4$, and $M = 2$. Theorem 1 does not rule out identifiability of this model. Using computations with the rational arithmetic we derive that the determinant of the Jacobian matrix is zero, which implies that the model is not identifiable, if any of the following conditions holds[3]:

- $\exists k \in \{1, 2\}$ such that $p_k \in \{0, 1\}$.
- $\exists \ell \in \{1, 2, 3, 4\}$ such that $q_{\ell,0} = 0$.
- $\exists \ell \in \{1, 2, 3, 4\}$ such that $q_{\ell,1} = 1$.
- $\exists \{\ell_1, \ell_2\} \subseteq \{1, 2, 3, 4\}$, $\ell_1 \neq \ell_2$ such that $q_{\ell_1,j} = q_{\ell_2,j}$ for all $j \in \{0, 1, 2\}$.
- $\exists \{\ell_1, \ell_2\} \subseteq \{1, 2, 3, 4\}$, $\ell_1 \neq \ell_2$ and $\{\ell_3, \ell_4\} = \{1, 2, 3, 4\} \setminus \{\ell_1, \ell_2\}$ such that $q_{\ell_1,0} = q_{\ell_2,0}$, $q_{\ell_3,0} = q_{\ell_4,0}$, $q_{\ell_1,2} = q_{\ell_2,1}$, $q_{\ell_2,2} = q_{\ell_1,2}$, $q_{\ell_3,2} = q_{\ell_4,1}$, and $q_{\ell_4,2} = q_{\ell_3,2}$.

Note that the last but one condition means that the leaky Noisy-AND models of Y_{ℓ_1} and Y_{ℓ_2} are identical. This effectively reduces this model to the BN2A model for $K = 2$, $L = 3$, and $M = 2$, for which Theorem 1 rules out identifiability. All of the above possibilities are exceptions that form a set of Lebesgue measure zero. We compute the determinant for 100 random points from the parameter space which implies that we can be almost sure that the rank of the Jacobian matrix is 14 and the model could be identified.

[3] We do not claim that this list is exclusive.

Example 4. Using computations with the rational arithmetic[4], we derive that for 100 randomly selected points from the parameter space of the BN2A model for $K = 3$, $L = 5$, and $M = 3$ from Table 2 the rank of the Jacobian matrix is 23, suggesting that the model can be generically identifiable.

Example 5. Using rational arithmetic computations, we observed that for all 100 randomly selected points from the parameter space of the BN2A model for $K = 3$, $L = 4$, and $M = 2$ from Table 2, the determinant of the Jacobian matrix was zero. We decided to perform symbolic computations that revealed that the determinant is zero regardless of the parameter values. This implies the model is not identifiable.

The presented examples illustrate different roles that the proposed numerical computations can play in deciding the identifiability of a BN2A model. The source code used in the examples is available as Mathematica notebooks and PDF files at: https://www.vomlel.cz/publications#h.w2xm776ugu54.

5 Discussion

In this paper, we analyzed the identifiability of BN2A networks, i.e., Bayesian networks where CPTs are represented by Noisy-AND models having the structure of a bipartite graph where all nodes from the first layer are hidden. Corresponding results also hold for BN2O networks, where CPTs are represented by Noisy-OR models, since it is easy to transform one class into the other by simply relabeling the states of the observed variables (state 0 to 1 and vice versa). Due to space limitations, we only present results for BN2A networks. The reason for our preference of BN2A over BN2O is that BN2A models have not been widely studied in the BN community, although models similar to BN2A are widely used as cognitive diagnostic models in psychometrics.

Perhaps the most important practical observation is that many small-sized BN2A models are unidentifiable, but as their size increases, the proportion of models ruled out as unidentifiable decreases. It should also be noted that the BN2A and BN2O models require a number of parameters proportional to $K \cdot L$. This is significantly less than the number of parameters of bipartite Bayesian networks with general CPTs, which can be exponential in K. This implies, especially for models with a higher number of parents, that the class of identifiable BN structures is substantially larger for the BN2A and BN2O networks.

The study of the identifiability of statistical models has a long history, but it is still a topic of current research. For example, the so-called Jacobian conjecture, which relates identifiability to the determinant of the Jacobian matrix, is still considered an open problem. In this paper, we have not presented any new deep theoretical results, but rather we have shown how the question of identifiability of a popular class of BN models can be addressed practically.

[4] We emphasize that there is no hope of getting correct results with finite-precision real arithmetic since, e.g., in one run, the absolute values of the computed determinants were in the interval $[10^{-37}, 10^{-72}]$ for this model.

References

1. D'Ambrosio, B.: Symbolic probabilistic inference in large BN2O networks. In: de Mantaras, R.L., Poole, D. (eds.) Uncertainty in Artificial Intelligence (UAI'94) Proceedings, pp. 128–135. Morgan Kaufmann, San Francisco (CA) (1994). https://doi.org/10.1016/B978-1-55860-332-5.50022-5
2. de la Torre, J.: The generalized DINA model framework. Psychometrika **76**, 179–199 (2011)
3. Díez, F.J., Druzdzel, M.J.: Canonical probabilistic models for knowledge engineering. Technical report CISIAD-06-01, UNED, Madrid, Spain (2006)
4. Geiger, D., Heckerman, D., Meek, C.: Asymptotic model selection for directed networks with hidden variables. In: Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96). pp. 283–290 (1996)
5. Gu, Y., Xu, G.: Sufficient and necessary conditions for the identifiability of the Q-matrix. Stat. Sin. **31**, 449–472 (2021). https://doi.org/10.5705/ss.202018.0410
6. Halpern, Y., Sontag, D.: Unsupervised learning of noisy-or Bayesian networks. In: Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence, pp. 272–281. UAI 2013, AUAI Press, Arlington, Virginia, USA (2013)
7. Hartz, S.M.: A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. Ph.D. thesis, University of Illinois at Urbana-Champaign (1996)
8. Heller, J.: Identifiability in probabilistic knowledge structures. J. Math. Psychol. **77**, 46–57 (2017). https://doi.org/10.1016/j.jmp.2016.07.008
9. Henrion, M.: Some practical issues in constructing belief networks. In: Proceedings of the Third Conference on Uncertainty in Artificial Intelligence (UAI-87), pp. 161–173. Elsevier Science Publishers B.V. (North Holland) (1987)
10. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. Information Science and Statistics, 2nd edn. Springer, New York (2007). https://doi.org/10.1007/978-0-387-68282-2
11. Junker, B.W., Sijtsma, K.: Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Appl. Psychol. Meas. **25**, 258–272 (2001)
12. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press, Cambridge (2009)
13. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco (1988)
14. Sullivant, S.: Algebraic Statistics. American Mathematical Society, Providence (2018)