Polynomial Constraints of Bayesian Networks with Hidden Variables

Luis David Garcia [1]

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA, lgarcia@math.vt.edu.

Abstract

Multinomial Bayesian networks with hidden variables are real algebraic varieties. Thus, they are the zeros of some polynomials in the probability simplex. These polynomials form the set of all independence and non–independence constraints on the distributions over the observable variables implied by a Bayesian network with hidden variables. We determine these contraints for Bayesian networks with three observable variables and one hidden variable. The relevance of these results for model selection is discussed.

1 Introduction

A *Bayesian network* is a family of probability distributions. In this paper, all random variables are assume to be discrete. In this case, these families are *algebraic varieties* (actually, semi–algebraic sets): they are the zeros of some polynomials in the probability simplex [3,4].

Bayesian networks can be described in two possible ways: parametrically, by an explicit mapping of a set of parameters to a set of distributions, or implicitly, by a set of independence constraints that the distributions must satisfy [9]. But Bayesian networks with hidden variables are usually defined parametrically because the independence and non–independence constraints on the distributions over the observable variables are not easily established.

Finding the independence and non–independence contraints is relevant for the problem of *model selection* [5]. Since these constraints vary from one model to another they can be used to distinguish between models. Moreover, since

[1] Partially supported by NSF grant DMS–0138323.

these constraints are over the observable variables, their fit to data can be measured directly with some specially designed statistical tests. For instance, the so–called *tetrad difference constraints* have been used for model selection and evaluation [11].

In this paper we give the constraints on the distribution over the observable variables of all Bayesian network with three observable variables and one hidden variable. These results show that all but one of these models correspond to intersections (or joins) of (higher secant varieties of) Segre varieties. Moreover, these contraints are given in a compact syntactic representation. These results were first announced in [2]. But in here, we present complete proofs of our claims.

Moreover, we also compute the *dimension* of each of these varieties. This is also relevant for model selection. Recently, the importance of using the correct dimension of a model when applying the *Bayesian Information Criterion (BIC)* for Bayesian model selection was highlighted in [10]. It was shown in [2] that the correct dimension of a model equals the dimension of the corresponding variety.

This paper is organized as follows. In Section 2, we review the algebraic statistics theory of Bayesian networks with hidden variables. In section 3, we present the main theorem that computes the set of all independence and non–independence contraints and the dimension of all Bayesian networks discussed in this paper. Most of the large–scale computations that provided enough evidence to make conjectures about the structure of these polynomial contraints were carried out in `Singular` [6].

2 Algebraic Statistics of Bayesian Networks

Let $X = \{X_1, \dots, X_n\}$ be n discrete variables. We will assume that each X_i takes values in $[d_i] = \{1, 2, \dots, d_i\}$. A Bayesian network M for variables X is a set of joint distributions for X defined by a graph G_M and a set of local (multinomial) distributions $\mathcal{F}_M$. A probability distribution $P(x)$ belongs to the model M if and only if it factors according to G_M via

$$P(X = x) = \prod_{i=1}^{n} p_i(X_i = x_i \mid \mathrm{pa}(X_i) = j), \qquad (1)$$

where x is an n–dimensional vector of values of X, $\mathrm{pa}(X_i)$ denote the parents of node X_i in G_M, j denotes the values of $\mathrm{pa}(X_i)$ in x and p_i is a conditional distribution from $\mathcal{F}_M$. We denote the *model parameters* defining the conditional probability $p_i(X_i = k \mid \mathrm{pa}(X_i) = j)$ by w_{ijk} and the *joint space parameters* $P(X = x)$ by θ_x. The mapping that relates these parameters, derived from (1), is

$$\theta_{(x_1,\ldots,x_n)} = \prod_{i=1}^{n} w_{ijk}, \tag{2}$$

where k and j denote the assignment to X_i and $\mathrm{pa}(X_i)$ as dictated by the vector of values $(x_1, \ldots, x_n)$. We can consider the model parameters and the joint space parameters as algebraic *indeterminates*. This allows us to form two rings of polynomials: $\mathbb{C}[\theta_x]$ the ring of polynomials over $\mathbb{C}$ generated by all the indeterminates θ_x and $\mathbb{C}[w_{ijk}]/J$ the ring of polynomials generated by all the indeterminates w_{ijk} modulo the ideal J generated by $\sum_{k=1}^{d_i} w_{ijk} - 1$, for each i and j. The ideal J encodes the fact that $p_i(X_i = k \mid \mathrm{pa}(X_i) = j)$ is a probability distribution, for each fixed i, j. Hence (2) induces a ring homomorphism

$$\Phi : \mathbb{C}[\theta_x] \to \mathbb{C}[w_{ijk}]/J.$$

Note that $\mathbb{C}[\theta_x]$ is a polynomial ring in $N = \prod_{i=i}^{n} d_i$ indeterminates. Recall that $\ker(\Phi)$ is a prime ideal in $\mathbb{C}[\theta_x]$. It was shown in [3] that the intersection of the variety $V(\ker(\Phi))$ with the probability simplex

$$\Delta = \{(a_1, \ldots, a_N) \mid a_i \geq 0, \sum a_i = 1\}$$

is the set of all probability distributions that factor according to G_M. Computing generators for $\ker(\Phi)$ is the so-called *implicitization* problem [1,5].

The graph G_M describes the independencies of variables in M. These independencies give an implicit description of the model M. The set of all independence relations encoded by G_M is known as the set of *global Markov relations* [9]. From this set we can construct an ideal I_M in $\mathbb{C}[\theta_x]$, see [3]. In that paper, the authors show that the variety $V(I_M) \subset \mathbb{C}[\theta_x]$ is the set of all $d_1 \times d_2 \times \cdots \times d_n$–tables with complex entries which satisfy the conditional independence statements encoded by G_M.

Note that the Factorization Theorem [9, Thm. 3.27] states that

$$V(I_M) \cap \Delta = V(\ker(\Phi)) \cap \Delta.$$

This result no longer hold if one allows complex "probabilities," see [3]. But there is a nice formula that relates both ideals in this general setting. Let $\tau \in \mathbb{C}[\theta_x]$ be the product of all the linear forms (marginals)

$$\theta_{(+,+,\ldots,+,u_{k+1},\ldots,u_n)} = \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \cdots \sum_{i_k=1}^{d_k} \theta_{(i_1,i_2,\ldots,i_k,u_{k+1},\ldots,u_n)},$$

where k is arbitrary and u_l takes all possible values in $[d_l]$ for all $k+1 \leq l \leq n$. Then we have the following theorem.

Theorem 1 *The prime ideal* $\ker(\Phi)$ *is a minimal primary component of* I_M*. More precisely,*

$$I_M : \tau^\infty = \ker(\Phi). \tag{3}$$

Consider now the situation when some of the random variables in M are hidden. After relabeling we may assume that the variables $X_{k+1}, \ldots, X_n$ are hidden, while the random variables $X_1, \ldots, X_k$ are observed. Thus, the *observable probabilities* are

$$\theta_{(i_1,\ldots,i_k,+,+,\ldots,+)} = \sum_{j_{k+1}\in[d_{k+1}]} \sum_{j_{k+2}\in[d_{k+2}]} \cdots \sum_{j_n\in[d_n]} \theta_{(i_1,i_2,\ldots,i_k,j_{k+1},\ldots,j_n)}.$$

We write $\mathbb{C}[\theta_{x'}]$ for the polynomial subring of $\mathbb{C}[\theta_x]$ generated by the observable probabilities. Let $\pi : \mathbb{C}^N \to \mathbb{C}^{N'}$ denote the canonical linear epimorphism induced by the inclusion of $\mathbb{C}[\theta_{x'}]$ into $\mathbb{C}[\theta_x]$. The following result was proved in [3].

Proposition 2 *The set of all polynomial functions which vanish on the space of observable probability distributions is the prime ideal*

$$\ker(\Phi) \cap \mathbb{C}[\theta_{x'}].$$

In the next section, we use this result together with Theorem 1 as an algorithm to compute the set of polynomial constraints on the distributions over the observable variables implied by a Bayesian network with hidden variables. This method is equivalent to the implicitization method proposed by Geiger and Meek [5]. The difference is that we compute the implicitization in two steps rather than one. First, we compute the prime ideal $\ker(\Phi)$ corresponding to the model where all variables are assumed to be observed. Then, we project the variety $V(\ker(\Phi))$ into the space of observable probability distributions

$$\pi\Big(V(\ker(\Phi))\Big) \subset \mathbb{C}^{N'}.$$

This approach enabled us to find a clear syntactic structure of the constraints implied by each Bayesian network studied in this paper.

3 Polynomial Constraints and Dimension

A step towards computing the constraints of a Bayesian network with hidden variables was given in [3], where the authors conjectured that any naive Bayes model M with 2 classes and n features is generated by the 3×3–subdeterminants of any two–dimensional table obtained by flattening the n–dimensional table $\theta_{(i_1, i_2, \cdots, i_n)}$. This conjecture was proved (set–theoretically in all cases and ideal–theoretically for $n = 3$) by Landsberg and Manivel in [8]. The previous result concerning a naive Bayesian model with 2 classes and 2 ternary features obtained in [5] is one instance of this theorem.

In this section, we give the constraints on the distribution over the observable variables of all Bayesian networks with three observable variables and one hidden variable. For some networks, we had to assume that the hidden variable is binary for the results to hold. To simplify notation we will set $\theta_{i_1 \cdots i_n} = \theta_{(i_1, \ldots, i_n)}$ and $P_G = \ker(\Phi)$.

Table 1 gives all the non–isomorphic directed acyclic graphs on 4 vertices, except those arising from the complete graph.

We use Theorem 1 to compute the prime ideal $\ker(\Phi)$. Thus, we compute the ideal I_M generated by all global Markov relations and then we saturate this ideal by τ. The following theorem, proved in [3], is fundamental in this step. This theorem gives a primary decomposition of all Bayesian networks on four arbitrary random variables.

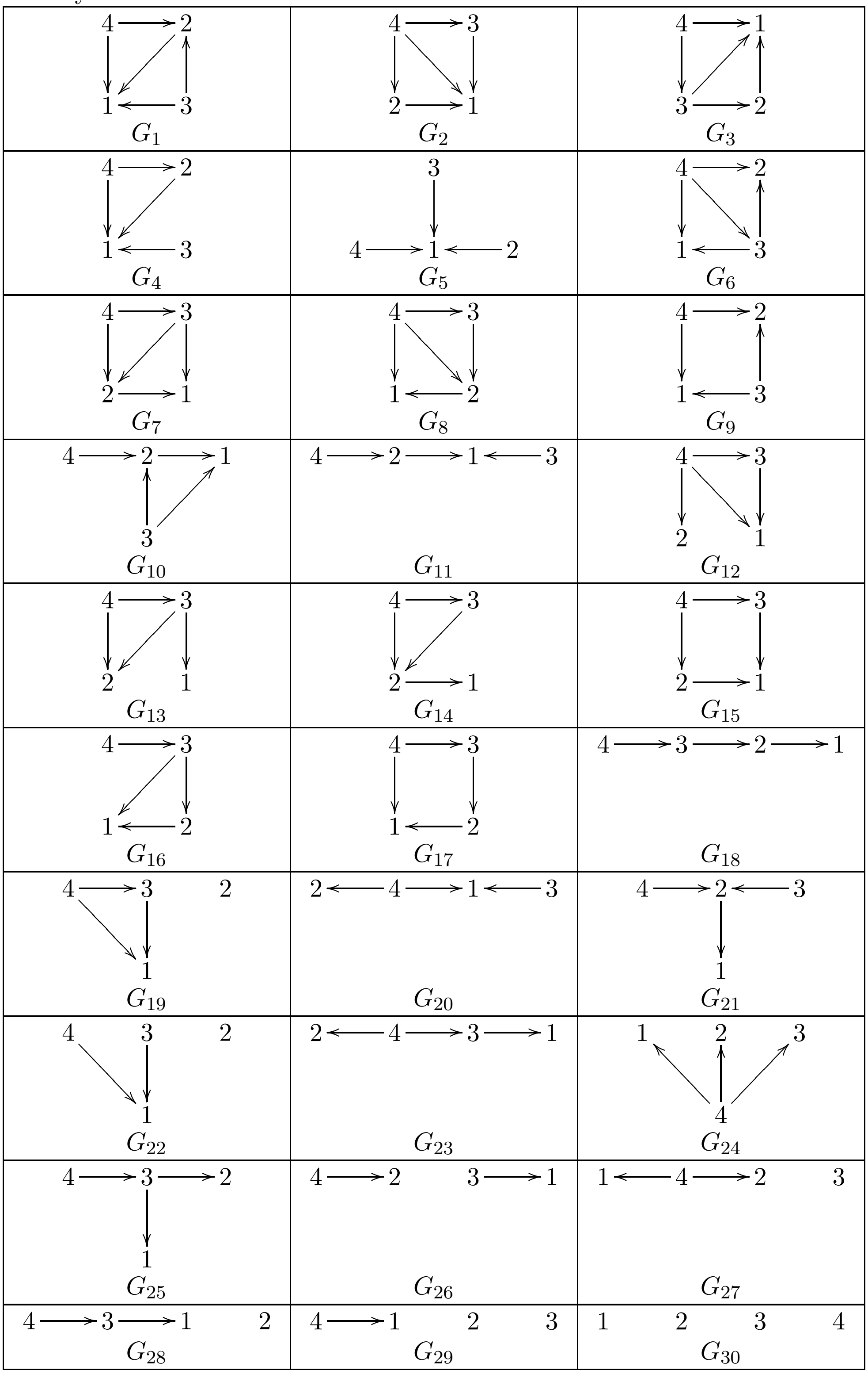
Theorem 3 *Of the* 30 *global Markov ideals on four random variables,* 26 *are always prime, one is not prime but always radical (number* 10 *in Table 1) and three are not radical (numbers* 15*,*17*,*21 *in Table 1).*

Therefore, we do not need to compute the saturation by τ for each of the 26 ideals I_M that are already prime.

The main result of this chapter is Theorem 4, which states that if G is a Bayesian network on four random variables ($G \neq G_{17}$) where one of the variables is a hidden binary variable, then the variety associated to G is the join (or the intersection) of several (higher secant varieties of) Segre varieties.

For networks 15, 17, 20, 21, 23, and 27, we have conjectures about the generating set based on extensive computations for particular cases and the dimension of the corresponding ideals. We remark that a proof for any of the last four networks would also yield a proof for the remaining three.

Table 1
All Bayesian networks on four random variables.



Theorem 4 *Of the 30 Bayesian networks on three random variables and one hidden variable*

(I) five networks G always give zero ideals $Q_G = P_G \cap \mathbb{C}[\theta_{x'}]$, regardless of the number of levels of the random variables. (numbers `1, 3, 7, 10, 16` *in Table 1).*

(II) Fourteen networks G always give ideals Q_G generated by quadratic polynomials arising from the 2×2 subdeterminants of certain matrices of indeterminates (numbers `4, 5, 11, 13, 14, 18, 19, 21, 22, 25, 26, 28, 29, 30` *in Table 1).*

(III) If the hidden variable is binary, ten networks G give ideals Q_G generated by quadratic and cubic polynomials arising from the 2×2 or 3×3 subdeterminants of certain matrices of indeterminates (numbers `2, 6, 8, 9, 12, 15, 20, 23, 24, 27` *in Table 1).*

(IV) The network G_{17} gives an ideal $Q_{G_{17}}$ generated by irreducible sextic polynomials and cubic polynomials, if the hidden variable is binary.

PROOF. We prove this theorem by an exhaustive case analysis of all thirty networks.

Networks `1, 3, 7, 16`*:* The ideal I_{G_1} is a prime ideal equal to $I_{3 \perp\!\!\!\perp 4}$. It is generated by the 2×2–subdeterminants of the matrix (θ_{++kl}), where the rows are indexed by $k \in [d_3]$ and the columns are indexed by $l \in [d_4]$. We claim that $I_{3 \perp\!\!\!\perp 4} \cap \mathbb{C}[\theta_{x'}] = 0$. It is enough to show that every $d_1 \times d_2 \times d_3$–table $a = (a_{ijk})$ is the projection of a table in $V(I_{3 \perp\!\!\!\perp 4})$. Let A be the $d_1 \times d_2 \times d_3 \times d_4$–table defined by $A_{ijkl} = a_{ijk}/d_4$, so $A_{++kl} = a_{++k}/d_4$. Then, the d_4 rows of $A' = (A_{++kl})$ are equal to each other, that is, A' has rank 1. So $A \in V(I_{3 \perp\!\!\!\perp 4})$ and $\pi(A) = a$. Observe that $I_{G_3} = I_{2 \perp\!\!\!\perp 4|3}$, $I_{G_7} = I_{1 \perp\!\!\!\perp 4|\{2,3\}}$ and $I_{G_{16}} = I_{\{1,2\} \perp\!\!\!\perp 4|3}$. Thus, a similar argument shows that $Q_3 = Q_7 = Q_{16} = 0$.

Network `10`*:* The ideal $I_{G_{10}}$ is a radical ideal equal to $I_{1 \perp\!\!\!\perp 4|\{2,3\}} + I_{3 \perp\!\!\!\perp 4}$. We claim that $I_{G_{10}} \cap \mathbb{C}[\theta_{x'}] = 0$. It is enough to show that every $d_1 \times d_2 \times d_3$–table $a = (a_{ijk})$ is the projection of a table in $V(I_{G_{10}})$. Let A be the $d_1 \times d_2 \times d_3 \times d_4$–table defined by $A_{ijkl} = a_{ijk}/d_4$. We saw above that $A \in V(I_{1 \perp\!\!\!\perp 4|\{2,3\}})$ and $A \in V(I_{3 \perp\!\!\!\perp 4})$, so $A \in V(I_{G_{10}})$. Theorem 11 in [3] shows that if $d_4 = 2$, then $I_{G_{10}}$ is a prime ideal. In this case, $Q_{10} = I_{G_{10}} \cap \mathbb{C}[\theta_{x'}] = 0$. If $d_4 > 2$, $I_{G_{10}}$ is the intersection of $P_{G_{10}}$ and 2^{d_3-1} prime ideals P_σ indexed by all proper subsets $\sigma \subset [d_3]$. By construction, the ideal $M_\sigma = \langle \theta_{+jkl} : j \in [d_2], k \in \sigma, l \in [d_4] \rangle$ is contained in P_σ, so $M_\sigma \cap \mathbb{C}[\theta_{x'}] = \langle \theta_{+jk+} : j \in [d_2], k \in \sigma \rangle$ is a subset of $P_\sigma \cap \mathbb{C}[\theta_{x'}]$. Then

$$\dim(P_\sigma \cap \mathbb{C}[\theta_{x'}]) \leq \dim(M_\sigma \cap \mathbb{C}[\theta_{x'}]) < d_1 d_2 d_3$$

Moreover, since $\cup_{\sigma \in [d_3]} V(P_\sigma \cap \mathbb{C}[\theta_{x'}]) \cup V(Q_{10}) = \mathbb{C}[\theta_{x'}]$, then $V(Q_{10}) = \mathbb{C}[\theta_{x'}]$.

Networks `4, 19, 26`*:* The ideal I_{G_4} is a prime ideal equal to $I_{3 \perp\!\!\!\perp \{2,4\}}$. It is generated by the 2×2–subdeterminants of the matrix $M = (\theta_{+jkl})$, where the rows are labeled by $k \in [d_3]$ and the columns are labeled by the pairs $(j,l) \in [d_2]\times[d_4]$. First, change coordinates in $\mathbb{C}[\theta_x]$ by replacing each unknown θ_{1jkl} by $\theta_{jkl} = \sum_{i=1}^{d_1} \theta_{ijkl}$. This coordinate change transforms I_{G_4} into a binomial ideal in $\mathbb{C}[\theta_x]$.

We want to show that $Q_4 = P_{G_4} \cap \mathbb{C}[\theta_{x'}]$ is equal to the ideal I generated by the 2×2–subdeterminants of the matrix $N = (\theta_{jk+})$ where the rows are indexed by $k \in [d_3]$ and the columns are indexed by $j \in [d_2]$. Each column of this matrix is obtained by taking the sum of the corresponding d_4 columns of M. A direct computation shows that $I \subset Q_4$.

The ideal Q_4 is prime as shown in Proposition 2. The ideal I is prime since it is generated by the 2×2–subdeterminants of a generic matrix. Hence, to show equality between ideals, it would suffice to show that $V(I) \subset V(Q_4)$. Let $a \in V(I)$, then a is a $d_3{\times}d_2$–matrix of rank 1 since all the 2×2–subdeterminants of a vanish. Let A be the $d_3 \times d_2 d_4$–matrix defined by $A_{jkl} = a_{jk}/d_4$. Clearly, A has rank 1, so $A \in V(P_{G_4})$. Thus $a = \pi(A) \in V(Q_4)$.

The ideal $I_{G_{19}}$ is a prime ideal equal to $I_{2 \perp\!\!\!\perp \{1,3,4\}}$. A similar argument shows that the ideal $Q_{19} = I_{G_{19}} \cap \mathbb{C}[\theta_{x'}]$ is generated by the 2×2–subdeterminants of the $d_2 \times d_1 d_3$–matrix $M_+ = (\theta_{ijk+})$ where the rows are indexed by $j \in [d_2]$ and the columns are indexed by pairs $(i,k) \in [d_1]\times[d_3]$. The prime ideal $I_{G_{26}}$ equals $I_{\{1,3\} \perp\!\!\!\perp \{2,4\}}$. A similar argument shows that $Q_{26} = Q_{19}$.

Networks `5, 11`*:* First, change coordinates in $\mathbb{C}[\theta_x]$ by replacing each indeterminate θ_{1jkl} by $\theta_{jkl} = \sum_{i=1}^{d_1} \theta_{ijkl}$. The prime ideal I_{G_5} is equal to the sum of ideals $I_{3 \perp\!\!\!\perp \{2,4\}} + I_{4 \perp\!\!\!\perp \{2,3\}}$. The first ideal equals I_{G_4} and $I_{4 \perp\!\!\!\perp \{2,3\}} \cap \mathbb{C}[\theta_{x'}] = 0$. Also $I_{3 \perp\!\!\!\perp \{2,4\}} \cap \mathbb{C}[\theta_{x'}] \subseteq I_{G_5} \cap \mathbb{C}[\theta_{x'}]$, that is, $V(Q_5) \subseteq V(Q_4)$. On the other hand, given $a \in V(Q_4)$, let A be the $d_2{\times}d_3{\times}d_4$–table defined by $A_{jkl} = a_{jk}/d_4$. Then, $A \in V(P_{G_4})$. So $A \in V(P_{G_4}) \cap V(I_{4 \perp\!\!\!\perp \{2,3\}})$ and $a = \pi(A) \in V(Q_5)$. Hence $Q_5 = Q_4$. The prime ideal $I_{G_{11}}$ is equal to the sum of ideals $I_{3 \perp\!\!\!\perp \{2,4\}} + I_{\{1,3\} \perp\!\!\!\perp 4|2}$. A similar argument shows that $Q_{11} = Q_4$.

Networks `13, 14`*:* The prime ideal $I_{G_{13}}$ equals $I_{1 \perp\!\!\!\perp \{2,4\}|3}$. A similar argument as for network G_4 shows that Q_{13} is generated by the 2×2–subdeterminants of the d_3 matrices of the form (θ_{ijk+}) where the rows are indexed by $i \in [d_1]$, the columns are indexed by $j \in [d_2]$, and k is fixed. The ideal $I_{G_{14}}$ is a prime ideal equal to $I_{1 \perp\!\!\!\perp \{3,4\}|2}$. Thus, the ideal Q_{14} is generated by the 2×2–subdeterminants of the d_2 matrices $N_j = (\theta_{ijk+})$ where the rows are indexed by $i \in [d_1]$ and the columns are indexed by $k \in [d_3]$.

Networks `18, 22, 28`*:* The prime ideal $I_{G_{18}}$ is equal to the sum $I_{1 \perp\!\!\!\perp \{3,4\}|2} +$

$I_{\{1,2\}\perp\!\!\!\perp 4|3}$. The first ideal equals $I_{G_{14}}$. Also $I_{\{1,2\}\perp\!\!\!\perp 4|3} \cap \mathbb{C}[\theta_{x'}] = 0$. Thus, a similar argument as for G_5 shows that $Q_{G_{18}} = Q_{G_{14}}$. The prime ideal $I_{G_{22}}$ equals $I_{2\perp\!\!\!\perp\{1,3,4\}} + I_{4\perp\!\!\!\perp\{2,3\}}$. We know that $I_{4\perp\!\!\!\perp\{2,3\}} \cap \mathbb{C}[\theta_{x'}] = 0$. Similarly, we conclude that $Q_{22} = Q_{19}$. The prime ideal $I_{G_{28}}$ equals $I_{2\perp\!\!\!\perp\{1,3,4\}} + I_{\{1,2\}\perp\!\!\!\perp 4|3}$. Hence $Q_{28} = I_{G_{28}} \cap \mathbb{C}[\theta_{x'}] = Q_{19}$.

Network 25*:* The prime ideal $I_{G_{25}}$ equals $I_{1\perp\!\!\!\perp\{2,4\}|3} + I_{2\perp\!\!\!\perp\{1,4\}|3} = I_{G_{13}} + J$. First, observe that $I_{G_{13}} \cap \mathbb{C}[\theta_{x'}] = J \cap \mathbb{C}[\theta_{x'}]$. Each ideal is generated by the 2×2–subdeterminants of d_3 matrices of the form (θ_{ijk+}), where the rows are indexed by $i \in [d_1]$ and the columns are indexed by $j \in [d_2]$ and k is fixed. A similar argument as for G_4 shows that $Q_{25} = Q_{13}$.

Networks 29, 30*:* The prime ideal $I_{G_{29}}$ equals $I_{2\perp\!\!\!\perp\{1,3,4\}} + I_{3\perp\!\!\!\perp\{1,2,4\}}$. Proceeding in a similar way as for G_4, we see that Q_{29} equals

$$I_{3\perp\!\!\!\perp\{1,2,4\}} \cap \mathbb{C}[\theta_{x'}] + I_{2\perp\!\!\!\perp\{1,3,4\}} \cap \mathbb{C}[\theta_{x'}] = I_{3\perp\!\!\!\perp\{1,2\}} + I_{2\perp\!\!\!\perp\{1,3\}} \subset \mathbb{C}[\theta_{x'}].$$

The prime ideal $I_{G_{30}}$ equals $I_{2\perp\!\!\!\perp\{1,3,4\}} + I_{3\perp\!\!\!\perp\{1,2,4\}} + I_{4\perp\!\!\!\perp\{1,2,3\}}$. Moreover, $I_{4\perp\!\!\!\perp\{1,2,3\}} \cap \mathbb{C}[\theta_{x'}] = 0$. This implies that $Q_{30} = Q_{29}$.

Network 21*:* The ideal $I = I_{G_{21}}$ equals $I_1 + J$, where $I_1 = I_{1\perp\!\!\!\perp\{3,4\}|2}$ and $J = I_{3\perp\!\!\!\perp 4}$. In general, this ideal is not radical, see [3, Theorem 11]. A closer look at the proof of Theorem 8 in [3] reveals that $P_{G_{21}}$ equals $(J + I_1) : \tau_1^\infty$, where τ_1 is the product of θ_{+jkl} for all $j \in [d_2], k \in [d_3]$ and $l \in [d_4]$. Hence $I_{G_{21}} = P_{G_{21}} \cap (I, \tau_1^{e_1})$ for some e_1. Thus, we have the following equalities

$$\begin{aligned} V(I) &= V(P_{G_{21}}) \cup V(I, \tau_1) \\ \overline{\pi(V(I))} &= V(Q_{21}) \cup \overline{\pi(V(I, \tau_1))} \end{aligned} \tag{4}$$

We know that $I_{3\perp\!\!\!\perp 4} \cap \mathbb{C}[\theta_{x'}] = 0$. Hence, following a similar argument as in G_9, we see that $I \cap \mathbb{C}[\theta_{x'}] = I_1 \cap \mathbb{C}[\theta_{x'}] = Q_{14}$. So, equation (4) can be rewritten as

$$V(Q_{14}) = V(Q_{21}) \cup \overline{\pi(V(I, \tau_1))}.$$

Therefore, $V(Q_{21})$ is a subvariety of the irreducible variety $V(Q_{14})$. Moreover, we conjecture that $V(Q_{21}) = V(Q_{14})$. For this we need to show that $\dim(V(Q_{14})) = \dim(V(Q_{21}))$. Since both ideals are prime, this would imply $Q_{21} = Q_{14}$.

Networks 2, 12*:* The prime ideal I_{G_2} equals $I_{2\perp\!\!\!\perp 3|4}$. It is generated by the 2×2–subdeterminants of the d_4 matrices (θ_{+jkl_0}), where the rows are indexed by $j \in [d_2]$ and the columns are indexed by $k \in [d_3]$. If $d_4 = 2$, then by [7, Exercise 11.29]

$$V(Q_2) = \overline{\pi(V(P_{G_2}))} = S(M_1) = M_2,$$

where M_k is the variety of $d_2 \times d_3$ matrices of rank at most k. Thus, the ideal Q_2 is given by the 3×3–subdeterminants of the $d_2 \times d_3$ matrix (θ_{+jk+}). The prime ideal $I_{G_{12}}$ equals $I_{2 \perp\!\!\!\perp \{1,3\}|4}$. Thus, we can proceed as for G_2 to find a set of generators for Q_{12}.

Networks `6, 8, 9`*:* The prime ideal I_{G_6} equals $I_{1 \perp\!\!\!\perp 2|\{3,4\}}$. It is generated by the 2×2–subdeterminants of the $d_3 d_4$ matrices $(\theta_{ijk_0 l_0})$, where the rows are indexed by $i \in [d_1]$ and the columns are indexed by $j \in [d_2]$. Assume $d_4 = 2$. For each $k_0 \in [d_3]$, let I_{k_0} be the ideal generated by the 2×2–subdeterminants of the 2 matrices $(\theta_{ijk_0 l})$, where l is fixed. Then just as for G_2

$$V(I_{k_0} \cap \mathbb{C}[\theta_{x'}]) = \overline{\pi(V(I_{k_0}))} = S(M_1) = M_2, \tag{5}$$

where M_k is the variety of $d_1 \times d_2$ matrices of rank at most k. Note that $I_{G_6} = \sum_{k\in[d_3]} I_k$, and the ideals I_k are defined in pairwise disjoint set of indeterminates. For each $k_0 \in [d_3]$, the fiber dimension over a general point in $V(I_{k_0} \cap \mathbb{C}[\theta_{x'}])$ is equal to 2. Thus the fiber dimension over a general point in $V(I_{G_6} \cap \mathbb{C}[\theta_{x'}])$ is equal to $2d_3$. Moreover, by [7, Proposition 12.2]

$$\operatorname{codim}(I_{G_6}) = \operatorname{codim}(\sum_{k\in[d_3]} I_k) = \sum_{k\in[d_3]} \operatorname{codim}(I_k) = \sum_{k\in[d_3]} 2(d_1-1)(d_2-1)$$

So $\dim(I_{G_6}) = 2d_1d_3 + 2d_2d_3 - 2d_3$, and

$$\operatorname{codim}\Big(\sum_{k\in[d_3]} (I_k \cap \mathbb{C}[\theta_{x'}])\Big) = \sum_{k\in[d_3]} \operatorname{codim}(I_k \cap \mathbb{C}[\theta_{x'}]) = d_3(d_1-2)(d_2-2)$$

So $\dim\Big(\sum_{k\in[d_3]}(I_k \cap \mathbb{C}[\theta_{x'}])\Big) = 2d_1d_3 + 2d_2d_3 - 4d_3$. Moreover, [7, Corollary 11.13] implies

$$\dim(Q_6) = \dim(I_{G_6}) - 2d_3 = 2d_1d_3 + 2d_2d_3 - 4d_3.$$

Therefore, $\sum_{k\in[d_3]}(I_k \cap \mathbb{C}[\theta_{x'}]) \subseteq Q_6$ and both prime ideals have the same dimension. Thus, $Q_6 = \sum_{k\in[d_3]}(I_k \cap \mathbb{C}[\theta_{x'}])$. Moreover, equation (5) gives a set of generators for this ideal.

The prime ideal I_{G_8} equals $I_{1 \perp\!\!\!\perp 3|\{2,4\}}$. So, we can proceed in a similar way as for G_6 to find a set of generators for Q_8. The prime ideal I_{G_9} equals $I_{1 \perp\!\!\!\perp 2|\{3,4\}} + I_{3 \perp\!\!\!\perp 4}$. We know that $I_{3 \perp\!\!\!\perp 4} \cap \mathbb{C}[\theta_{x'}] = 0$. Therefore, a similar argument as for G_5 shows that $Q_9 = Q_6$, if $d_4 = 2$.

Network `24`*:* The graph G_{24} corresponds to the *naive Bayes model* with d_4 classes and 3 features. As we mentioned earlier, Landsberg and Manivel proved in [8] that this ideal is generated by the 3×3–subdeterminants of any two–dimensional matrix obtained by flattening the 3–dimensional table $\theta_{(i_1,i_2,i_3)}$, if $d_4 = 2$. Here, we compute the dimension of this ideal.

The prime ideal $I_{G_{24}}$ equals $I_{1 \perp\!\!\!\perp \{2,3\}|4} + I_{2 \perp\!\!\!\perp \{1,3\}|4} + I_{3 \perp\!\!\!\perp \{1,2\}|4} = I_1 + I_2 + I_3$. Note that $I_{G_{24}} = I_1 + I_2 = I_1 + I_3 = I_2 + I_3$. Assume $d_4 = 2$, then by [7, Proposition 12.2] we have that $\mathrm{codim}(I_1) = 2(d_1 - 1)(d_2 d_3 - 1)$, so $\dim(I_1) = 2d_1 + 2d_2 d_3 - 2$. The ideal I_1 is generated by the 2×2–subdeterminants of the $d_1 \times d_2 d_3$–matrices $M_{l_0} = (\theta_{ijkl_0})$, where $l_0 \in \{1, 2\}$. Similarly, the ideal I_2 is generated by the 2×2–subdeterminants of the $d_2 \times d_1 d_3$–matrices $N_{l_0} = (\theta_{ijkl_0})$, where $l_0 \in \{1, 2\}$. Note that for each $(k_0, l_0) \in [d_3] \times [d_4]$, the $d_1 \times d_2$–matrix $M_{k_0 l_0} = (\theta_{ijk_0 l_0})$ is the transpose of the $d_2 \times d_1$–matrix $N_{k_0 l_0} = (\theta_{ijk_0 l_0})$. Hence, for each $k \geq 2$, the 2×2–subdeterminants of N_{kl_0} lowers the dimension of I_1 by $d_2 - 1$. Thus,

$$\begin{aligned}\dim(I_{G_{24}}) &= \dim(I_1 + I_2) = 2d_1 + 2d_2 d_3 - 2 - 2(d_2 - 1)(d_3 - 1)\\ &= 2d_1 + 2d_2 + 2d_3 - 4.\end{aligned}$$

Let $\widetilde{I}_r = I_r \cap \mathbb{C}[\theta_{x'}]$ for $r = 1, 2, 3$. Exercise 11.29 in [7] implies that $\widetilde{I}_1$ is generated by the 3×3–subdeterminants of the $d_1 \times d_2 d_3$–matrix $\widetilde{M} = (\theta_{ijk+})$ and $\dim(\widetilde{I}_1) = 2d_1 + 2d_2 d_3 - 4$. Proceeding in a similar way as for $\dim(I_1 + I_2)$, we conclude that $\widetilde{I}_2$ lowers the dimension of $\widetilde{I}_1$ by $2(d_2 - 2)(d_3 - 1)$, so $\dim(\widetilde{I}_1 + \widetilde{I}_2) = 2d_1 + 2d_2 + 4d_3 - 8$. The ideal $\widetilde{I}_3$ is generated by the 3×3–subdeterminants of the $d_3 \times d_1 d_2$–matrix $L = (\theta_{ijk+})$. Note that the k_0 row of L can be obtained by flattening the $d_1 \times d_2$–matrix $\widetilde{M_{k_0}} = (\theta_{ijk_0+})$. Hence, the ideal $\widetilde{I}_3$ lowers the dimension of $\widetilde{I}_1 + \widetilde{I}_2$ by $2(d_3 - 2)$. Thus,

$$\dim(\widetilde{I}_1 + \widetilde{I}_2 + \widetilde{I}_3) = 2d_1 + 2d_2 + 2d_3 - 4.$$

Then $\dim(\sum_{s=1}^{3} \widetilde{I}_s) = \dim(I_{\mathrm{local}\, G_{24}}) \geq \dim(Q_{24})$. But the result in [8] states that $\sum_{s=1}^{3} \widetilde{I}_s = Q_{24}$. Thus, $\dim(Q_{24}) = 2d_1 + 2d_2 + 2d_3 - 4$.

Network 20: First, change coordinates in $\mathbb{C}[\theta_x]$ by replacing each unknown θ_{1jkl} by $\theta_{jkl} = \sum_{i=1}^{d_1} \theta_{ijkl}$. The binomial prime ideal $I_{G_{20}}$ is equal to the sum of ideals $I = I_{2 \perp\!\!\!\perp \{1,3\}|4} = I_{G_{12}}$ and $J = I_{3 \perp\!\!\!\perp \{2,4\}} = I_{G_4}$. Denote by $\tilde{I} = I \cap \mathbb{C}[\theta_{x'}]$ and $\tilde{J} = J \cap \mathbb{C}[\theta_{x'}]$. Recall that the ideal I is generated by the 2×2–subdeterminants of the d_4 matrices $M_l = (\theta_{ijkl})$ where the rows are indexed by $(i, k) \in [d_1] \times [d_3]$, the columns by $j \in [d_2]$, and $l \in [d_4]$ is fixed for each matrix. The ideal J is generated by the 2×2–subdeterminants of the matrix $N = (\theta_{1jkl})$ where the rows are indexed by $k \in [d_3]$ and the columns by $(j, l) \in [d_2] \times [d_4]$.

For each l, the ideal generated by the 2×2–subdeterminants of M_l has codimension $(d_1 d_3 - 1)(d_2 - 1)$. Moreover, since the entries of each matrix are pairwise disjoint, the codimension of I equals $d_4(d_1 d_3 - 1)(d_2 - 1)$. Hence $\dim(I) = d_1 d_3 d_4 + d_2 d_4 - d_4$. Similarly, the codimension of J equals $(d_3 - 1)(d_2 d_4 - 1)$, so $\dim(J) = d_1 d_2 d_3 d_4 - d_2 d_3 d_4 + d_2 d_4 + d_3 - 1$. Let $M_{i_0 l_0}$ be the $d_3 \times d_2$–matrix

$(\theta_{i_0jkl_0})$, then

$$M_l = \begin{pmatrix} M_{1l} \\ M_{2l} \\ \vdots \\ M_{d_1l} \end{pmatrix} \qquad \text{and} \qquad N = (M_{11}M_{12}\cdots M_{1d_4}).$$

Hence, just as for G_{24}, the ideal J removes d_3-1 parameters of all but one of the matrices M_l. Thus,

$$\dim(I+J) = \dim(I) - (d_3-1)(d_4-1) = d_1d_3d_4 + d_2d_4 - d_3d_4 + d_3 - 1. \quad (6)$$

Let $d_4 = 2$. Then, the prime ideal $\tilde{I}$ is generated by the 3×3–subdeterminants of the two dimensional table $M_+ = (\theta_{ijk+})$, where the rows are indexed by $j \in [d_2]$ and the columns are indexed by pairs $(i,k) \in [d_1]\times[d_3]$. Hence $\operatorname{codim}(\tilde{I}) = (d_1d_3-2)(d_2-2)$, so $\dim(\tilde{I}) = 2d_1d_3+2d_2-4$. Similarly, since $\tilde{J}$ is generated by the 2×2–subdeterminants of the $d_2\times d_3$–matrix $N_+ = (\theta_{1jk+})$, then $\operatorname{codim}(\tilde{J}) = (d_2-1)(d_3-1)$, so $\dim(\tilde{J}) = d_1d_2d_3 - d_2d_3 + d_2 + d_3 - 1$.

Recall that $I+J$ is prime, then [7, Thm. 11.12] implies $\dim(V(I+J)) = \dim(V(Q_{20})) + \mu$. We conjecture that $\mu = 2$ which implies

$$\dim(Q_{20}) = \dim(I+J) - 2 = 2d_1d_3 + 2d_2 - d_3 - 3.$$

Let M_{i_0+} be the $d_2\times d_3$–matrix (θ_{i_0jk+}), then $M_+ = (M_{1+}M_{2+}\cdots M_{d_1+})$, and $M_{1+} = N_+$. A similar argument as for the ideal $I+J$ shows that the ideal $\tilde{J}$ lowers the dimension of $\tilde{I}$ by d_3-1. Hence

$$\dim(\tilde{I}+\tilde{J}) = 2d_1d_3 + 2d_2 - d_3 - 3 = \dim(Q_{20}).$$

Note that $\tilde{I}+\tilde{J} \subseteq I_{G_{20}} \cap \mathbb{C}[\theta_{x'}]$ and both ideals have the same dimension. One can check that $\tilde{I}+\tilde{J}$ is a radical ideal by Gröbner basis methods. In fact, if $<$ denotes the degree reverse lexicographic ordering, then the (quadratic) generators of $\tilde{J}$ and the (cubic) generators of $\tilde{I}$ form a Gröbner basis of $\tilde{I}+\tilde{J}$. Therefore, the initial ideal $\mathrm{in}_<(\tilde{I}+\tilde{J})$ is square–free, which implies that $\tilde{I}+\tilde{J}$ is a radical ideal. Moreover, a set–theoretic result as in [8] would imply that the ideal Q_{20} equals $I_{G_4} \cap \mathbb{C}[\theta_{x'}] + I_{2\perp\!\!\!\perp\{1,3\}|4} \cap \mathbb{C}[\theta_{x'}]$.

Network 23*:* The binomial prime ideal $I_{G_{23}}$ is the sum of two prime ideals $I = I_{2\perp\!\!\!\perp\{1,3\}|4}$ and $J = I_{G_{13}}$. Let $\tilde{I} = I \cap \mathbb{C}[\theta_{x'}]$. If $d_4 = 2$, the ideal $\tilde{I}$ is generated by the 3×3–subdeterminants of the $d_1d_3\times d_2$–matrix $M_+ = (\theta_{ijk+})$ obtained by flattening the 3–dimensional table (θ_{ijk+}) according to the relation $\{1,3\}\perp\!\!\!\perp 2$. Note that $\tilde{I}+\tilde{J} \subseteq I_{G_{23}} \cap \mathbb{C}[\theta_{x'}]$. Moreover, a similar argument as for G_{20} shows that the ideal Q_{23} equals $I_{G_{13}} \cap \mathbb{C}[\theta_{x'}] + I_{2\perp\!\!\!\perp\{1,3\}|4} \cap \mathbb{C}[\theta_{x'}]$.

The ideal I is generated by the 2×2–subdeterminants of the d_4 matrices $M_l = (\theta_{ijkl})$, where the rows are indexed by $(i,k) \in [d_1]\times[d_3]$, the columns by $j \in [d_2]$ and l is fixed. Recall that $\dim(I) = d_1d_3d_4 + d_2d_4 - d_4$. The ideal J is generated by the 2×2–subdeterminants of the d_3 matrices of the form $N_k = (\theta_{ijkl})$, where the rows are indexed by $i \in [d_1]$, the columns are indexed by $(j,l) \in [d_2]\times[d_4]$, and k is fixed. The codimension of J equals $d_3(d_1-1)(d_2d_4-1)$, so $\dim(J) = d_2d_3d_4 + d_1d_3 - d_3$.

For each k_0, l_0, let $M_{k_0l_0}$ be the $d_1\times d_2$–matrix $(\theta_{ijk_0l_0})$. Then

$$M_l = \begin{pmatrix} M_{1l} \\ M_{2l} \\ \vdots \\ M_{d_1l} \end{pmatrix} \quad \text{and} \quad N_k = (M_{11}M_{12}\cdots M_{1d_4}).$$

Thus, the following two $d_1d_3\times d_2d_4$–matrices are equal

$$\begin{pmatrix} M_1 & \cdots & M_{d_4} \end{pmatrix} = \begin{pmatrix} N_1 \\ \vdots \\ N_{d_3} \end{pmatrix}$$

Hence, the ideal I lowers the dimension of J by $d_4(d_2-1)(d_3-1)$. Thus, $\dim(I+J) = \dim(J) - d_4(d_2-1)(d_3-1) = d_1d_3 + d_2d_4 + d_3d_4 - d_3 - d_4$. If $d_4 = 2$, $\dim(I+J) = d_1d_3 + 2d_2 + d_3 - 2$. Theorem 11.12 in [7] implies that $\dim(V(I+J)) = \dim(V(Q_{23})) + \mu$. We conjecture that $\mu = 2$, which implies $\dim(Q_{23}) = d_1d_3 + 2d_2 + d_3 - 4$.

Recall that $\dim(\tilde{I}) = 2d_1d_3 + 2d_2 - 4$. Moreover, since $\tilde{J}$ is generated by the 2×2–subdeterminants of the d_3 matrices $N_{k+} = (\theta_{ijk+})$, then $\operatorname{codim}(\tilde{J}) = d_3(d_1-1)(d_2-1)$. So $\dim(\tilde{J}) = d_1d_3 + d_2d_3 - d_3$. Observe that $M_+ = (N_{1+}N_{2+}\cdots N_{d_3+})$. Therefore, $\tilde{I}$ lowers the dimension of $\tilde{J}$ by $(d_2-2)(d_3-2)$. Thus, $\dim(\tilde{I}+\tilde{J}) = d_1d_3 + 2d_2 + d_3 - 4 = \dim(Q_{23})$. Hence, Q_{23} equals $\tilde{I}+\tilde{J}$.

Network 27*:* The prime ideal $I_{G_{27}}$ equals $I_{3\amalg\{1,2,4\}} + I_{2\amalg\{1,3\}|4}$. Observe that $\tilde{I} = I_{3\amalg\{1,2,4\}} \cap \mathbb{C}[\theta_{x'}]$ is generated by the 2×2–subdeterminants of the matrix (θ_{ijk+}), where the rows are indexed by $k \in [d_3]$ and the columns are indexed by $(i,j) \in [d_1]\times[d_2]$. If $d_4 = 2$, the prime ideal $\tilde{J} = I_{2\amalg\{1,3\}|4} \cap \mathbb{C}[\theta_{x'}]$ is generated by the 3×3–subdeterminants of the two dimensional table (θ_{ijk+}), where the rows are indexed by $j \in [d_2]$ and the columns are indexed by the pairs $(i,k) \in [d_1]\times[d_3]$. Moreover, following a similar procedure as for G_{20}, we conjecture that Q_{27} equals $\tilde{I}+\tilde{J}$.

Network 15*:* The ideal $I_{G_{15}}$ equals $I+J$, where $I = I_{1 \amalg 4|\{2,3\}}$ and $J = I_{2 \amalg 3|4}$. This ideal is not radical, in general. Hence $I_{G_{15}} = P_{G_{15}} \cap L$ for some ideal L. Thus,

$$\overline{\pi(V(I_{G_{15}}))} = V(Q_{15}) \cup \overline{\pi(V(L))}.$$

Note that $I \cap \mathbb{C}[\theta_{x'}] = 0$. Moreover, if $d_4 = 2$, then a similar argument as for G_{11} shows that $I_{G_{15}} \cap \mathbb{C}[\theta_{x'}] = J \cap \mathbb{C}[\theta_{x'}] = Q_2$. Hence, $V(Q_2) = V(Q_{15}) \cup \overline{\pi(V(L))}$. Similar to G_{21}, we conjecture that $Q_{15} = Q_2$.

Network 17*:* The ideal $I_{G_{17}}$ equals $I+J$, where $I = I_{1 \amalg 3|\{2,4\}}$ and $J = I_{2 \amalg 4|3}$. This ideal is not radical in general. Hence $I_{G_{17}} = P_{G_{17}} \cap L$ for some ideal L. So, we have the following equality of varieties

$$\overline{\pi(V(I_{G_{17}}))} = V(Q_{17}) \cup \overline{\pi(V(L))}.$$

Note that $J \cap \mathbb{C}[\theta_{x'}] = 0$. Also $I = I_{G_8}$ and we have given a generating set for Q_8 for the case $d_4 = 2$. Moreover, $I_{G_{17}} \cap \mathbb{C}[\theta_{x'}] = Q_8$. Hence $V(Q_{17})$ is an irreducible subvariety of the irreducible variety Q_8. But, opposed to all the previous varieties, in general $V(Q_{17})$ will be a proper subvariety of $V(Q_8)$. We have a conjecture for the case $d_1 = d_4 = 2$. Note that for this case $Q_8 = 0$, that is, $V(Q_8) = \mathbb{C}[\theta_{x'}]$. To simplify notation, let $\theta_{ijk} = \theta_{ijk+}$.

The ideal Q_{17} is generated by $\binom{d_2}{2}\binom{d_3}{3}$ *sextic* polynomials constructed as follows. For each $j_0 \in [d_2]$, let M_{j_0} be the $d_1 \times d_3$–matrix $M_{j_0} = (\theta_{ij_0k})$. Each $j_1, j_2 \in [d_2]$, $j_1 \neq j_2$ specify two matrices M_{j_1} and M_{j_2}. Also, each triplet k_1, k_2, k_3 of distinct elements in $[d_3]$ specify three columns on each $2 \times d_3$–matrix M_{j_1} and M_{j_2}. So we get two 2×3 submatrices N_{j_1} and N_{j_2}

$$\begin{pmatrix} \theta_{1j_1k_1} & \theta_{1j_1k_2} & \theta_{1j_1k_3} \\ \theta_{2j_1k_1} & \theta_{2j_1k_2} & \theta_{2j_1k_3} \end{pmatrix} \text{ and } \begin{pmatrix} \theta_{1j_2k_1} & \theta_{1j_2k_2} & \theta_{1j_2k_3} \\ \theta_{2j_2k_1} & \theta_{2j_2k_2} & \theta_{2j_2k_3} \end{pmatrix}$$

The irreducible sextic polynomial arising from these two submatrices is given by the following alternating sum

$$\theta_{+j_1k_1}U_1V_1 - \theta_{+j_1k_2}U_2V_2 + \theta_{+j_1k_3}U_3V_3.$$

Where U_s is the determinant of the 2×2–submatrix of N_{j_1} obtained by eliminating the s-th column. And V_s is the determinant of the 2×2–matrix N'_{j_2} where the first column of N'_{j_2} equals the s-th column of N_{j_2} and the second column of N'_{j_2} is the product of the remaining two columns of N_{j_2}.

References

[1] D. Cox, J. Little and D. O'Shea: *Ideals, Varieties and Algorithms*, Springer Undergraduate Texts in Mathematics, Second Edition, 1997.

[2] L. D. Garcia: Algebraic Statistics in Model Selection, *Proceedings of the Twentieth Annual Conference on Uncertainty in Artificial Intelligence (UAI–2004)*, accepted.

[3] L. D. Garcia, M. Stillman and B. Sturmfels: Algebraic Geometry of Bayesian Networks, *Journal of Symbolic Computation, Special Issue Méthodes Effectives en Géometrie Algébrique* (MEGA 2004).

[4] D. Geiger, D. Heckerman, H. King and C. Meek: Stratified exponential families: graphical models and model selection, *Annals of Statistics* **29** (2001) 505–529.

[5] D. Geiger and C. Meek: Graphical models and exponential families, *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–98)* 156–165.

[6] G.-M. Greuel, G. Pfister and H. Schönemann: Singular 2.0: A computer algebra system for polynomial computations, University of Kaiserslautern, 2001, `http://www.singular.uni-kl.de`.

[7] J. Harris: *Algebraic Geometry: A First Course*, Springer Graduate Texts in Mathematics, 1992.

[8] J. M. Landsberg and L. Manivel: On the ideals of secant varieties of Segre varieties, sumbitted.

[9] S. L. Lauritzen: *Graphical Models*, Oxford University Press, 1996.

[10] D. Rusakov and D. Geiger: Asymptotic model selection for naive Bayesian networks, *Proceedings of the Eighteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–02).*

[11] P. Spirtes, C. Glymour and R. Scheines: *Causation, Prediction, and Search.* Springer–Verlag, 1993.

[12] B. Sturmfels, *Solving Systems of Polynomial Equations*, CBMS Lectures Series, American Mathematical Society, 2002.