ROBUSTNESS AND CONDITIONAL INDEPENDENCE IDEALS

JOHANNES RAUH[1], NIHAT AY[1,2]

Abstract. We study notions of robustness of Markov kernels and probability distribution of a system that is described by n input random variables and one output random variable. Markov kernels can be expanded in a series of potentials that allow to describe the system's behaviour after knockouts. Robustness imposes structural constraints on these potentials.

Robustness of probability distributions is defined via conditional independence statements. These statements can be studied algebraically. The corresponding conditional independence ideals are related to binary edge ideals. The set of robust probability distributions lies on an algebraic variety. We compute a Gröbner basis of this ideal and study the irreducible decomposition of the variety. These algebraic results allow to parametrize the set of all robust probability distributions.

1. Introduction

In this article we study a notion of robustness with tools from algebraic geometry. This work has been initiated in [1]. Connections to algebraic geometry have already been addressed in [6]. We consider n input nodes, denoted by $1, 2, \dots, n$, and one output node, denoted by 0. For each $i = 0, 1, \dots, n$ the state of node i is a discrete random variable X_i taking values in the finite set $\mathcal{X}_i$ of cardinality d_i. The joint state space is the set $\tilde{\mathcal{X}} = \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. For any subset $S \subseteq \{0, \dots, n\}$ write X_S for the random vector $(X_i)_{i\in S}$; then X_S is a random variable with values in $\mathcal{X}_S = \times_{i\in S}\mathcal{X}_i$. For any $x \in \tilde{\mathcal{X}}$, the *restriction* of x to a subset $S \subseteq \{0, \dots, n\}$ is the vector $x|_S \in \mathcal{X}_S$ with $(x|_S)_i = x_i$ for all $i \in S$.

We study two possible models for the computation of the output from the input: The first model is a stochastic map (Markov kernel) κ from $\mathcal{X}_{[n]}$ to $\mathcal{X}_0$, that is, κ is a function

$$\kappa : \mathcal{X}_{[n]} \times \mathcal{X}_0 \to [0, 1], \qquad (x, y) \mapsto \kappa(x; y),$$

satisfying $\sum_{y\in\mathcal{X}_0} \kappa(x; y) = 1$ for all x. The second model is a joint probability distribution p of the random vector $(X_0, X_{[n]})$. These two models are related as follows: The joint probability distribution p of $(X_0, X_{[n]})$ can be factorized as

$$p(y, x) = p(y|x)p_{\mathrm{in}}(x), \qquad \text{for all } (y, x) \in \tilde{\mathcal{X}},$$

where p_{in} is the distribution of the input nodes and $p(y|x)$ is a conditional distribution, which need not be unique. Each possible choice of this conditional distribution defines a Markov kernel $\kappa(x; y) := p(y|x)$. Conversely, when a Markov kernel κ is given, then any input distribution $p_{\mathrm{in}}(x)$ defines a joint distribution $p(x, y) = p_{\mathrm{in}}(x)\kappa(x; y)$. The result of our analysis will not depend too much on the precise form of the input distribution; it will turn out that only the *support* $\mathrm{supp}(p_{\mathrm{in}}) := \{x \in \mathcal{X} : p_{\mathrm{in}}(x) > 0\}$

Date: January 2, 2014.

is important. Similarly, in the analysis of the kernels, there will also be a set $\mathcal{S}$ of *"relevant inputs"* that will play an important role.

We study robustness with respect to knockouts of some of the input nodes $[n]$ in both models. When a subset S of the input nodes is knocked out, and only the nodes in $R = [n] \setminus S$ remain, then the behaviour of the system changes. Without further assumptions, the post-knockout function is not determined by κ and has to be specified. We therefore consider a further stochastic map $\kappa_R : \mathcal{X}_R \times \mathcal{X}_0 \to [0,1]$ as model of the post-knockout function. A complete specification of the function is given by the family $(\kappa_A)_{A\subseteq[n]}$ of all possible post-knockout functions, which we refer to as *functional modalities*. As a shorthand notation we denote functional modalities as (κ_A). The Markov kernel κ itself, which describes the normal behaviour of the system without knockouts, can be identified with $\kappa_{[n]}$.

What does it mean for a stochastic map to be robust? Assume that the input is in state x, and that we knock out a set S of inputs. Denoting the remaining set of inputs by R, we say that (κ_A) is robust in $x = (x_R, x_S)$ against knockout of S, if

$$\kappa(x_R, x_S; x_0) = \kappa_R(x_R; x_0) \qquad \text{for all } x_0 \in \mathcal{X}_0 \,. \tag{1}$$

If $\mathcal{R}$ is a collection of subsets of $[n]$ and if (κ_A) is robust in x against knockout of $[n] \setminus R$ for all $R \in \mathcal{R}$, then we say that (κ_A) is *$\mathcal{R}$-robust in x*. In Section 2, we consider Gibbs representations of functional modalities and derive structural constraints on corresponding interaction potentials that are imposed by robustness properties. These constraints do not depend on the configuration x in which the functional modalities are assumed to be robust.

Similar to the case of Markov kernels, the joint probability distribution p does not allow to predict the behaviour of a perturbed system. Nevertheless, we can ask whether it is at all possible that the behaviour of the system is robust against a given knockout of S. Let p_{in} be an input distribution, and let (κ_A) be the functional modalities of the system. If (κ_A) is robust against knockout of S in x for all $x \in \operatorname{supp}(p_{\text{in}})$, then X_0 is stochastically independent from X_S given X_R (with respect to the joint probability distribution $p(x_0, x_{\text{in}}) = p_{\text{in}}(x_{\text{in}})\kappa(x_{\text{in}}; x_0)$), where $R = [n]\setminus S$, a fact that will be denoted by $X_0 \perp\!\!\!\perp X_S \,|\, X_R$. In order to see this, assume $x = (x_R, x_S) \in \operatorname{supp}(p_{\text{in}})$. Then

$$\begin{aligned}
p(x_0 \,|\, x_R, x_S) &= \kappa(x_R, x_S; x_0) \\
&= \kappa_R(x_R; x_0) \sum_{x'_S : (x_R, x'_S) \in \operatorname{supp}(p_{\text{in}})} p(x'_S \,|\, x_R) \\
&= \sum_{x'_S : (x_R, x'_S) \in \operatorname{supp}(p_{\text{in}})} p(x'_S \,|\, x_R)\, \kappa(x_R, x'_S; x_0) \\
&= \sum_{x'_S : (x_R, x'_S) \in \operatorname{supp}(p_{\text{in}})} p(x'_S \,|\, x_R)\, p(x_0 \,|\, x_R, x'_S) \\
&= p(x_0 \,|\, x_R) \,.
\end{aligned}$$

On the other hand, if $X_0 \perp\!\!\!\perp X_S \,|\, X_R$ holds for a joint distribution p, then any family (κ_A) with the property that $\kappa_A(x_A; x_0) = p(x_0|x_A)$ whenever $p(x_A) > 0$ is robust against knockout of S for all $x \in \operatorname{supp}(p_{\text{in}})$, where p_{in} is the marginal input distribution.

Therefore, we call the joint probability distribution p *robust against knockout of S*, if $X_0 \perp\!\!\!\perp X_S \,|\, X_R$. This means that we do not lose information about the output X_0, if the subset S of the inputs is unknown or hidden (or "knocked out"). Probability distributions that are robust in this sense are studied in Section 3. Section 4 discusses the case that X_0 is a deterministic function of the input nodes. The symmetric case that p is robust against knockout of any set S of cardinality less than $n-k$ is studied in Section 5.

The results about robustness are derived from an algebraic theory of generalized binomial edge ideals, which generalize the binomial edge ideals of [6] and [9]. This theory is presented in Section 6. A Gröbner basis is constructed, and it is shown that these ideals are radical. Finally, a primary decomposition is computed. Similar CI statements have recently been studied in [11]. That work discusses what is called $(n-1)$-robustness in Section 5.

2. Robustness of Markov kernels

Let $(\kappa_A)_{A\subseteq[n]}$ be a collection of *functional modalities*, as defined in the introduction. Instead of providing a list of all functional modes κ_A, one can describe them in more mechanistic terms. In order to illustrate this, we first consider an example which comes from the field of neural networks. In that example, we assume that the output node receives an input $x=(x_1,\dots,x_n)\in\{-1,+1\}^n$ and generates the output $+1$ with probability

$$\kappa(x_1,\dots,x_n;+1) \;:=\; \frac{1}{1+e^{-\sum_{i=1}^n w_i\,x_i}}\,, \tag{2}$$

which implies that for an arbitrary output x_0

$$\kappa(x_1,\dots,x_n;x_0) \;:=\; \frac{e^{\frac{1}{2}\sum_{i=1}^n w_i\,x_i\,x_0}}{e^{\frac{1}{2}\sum_{i=1}^n w_i\,x_i\cdot(-1)}+e^{\frac{1}{2}\sum_{i=1}^n w_i\,x_i\cdot(+1)}}\,.$$

This representation of the stochastic map κ has a structure that allows inferring the function after a knockout of a set S of input nodes, by simply removing the contribution of all the nodes in S. In our example (2), the post-knockout function is then given as

$$\kappa_R(x_R;+1) \;:=\; \frac{1}{1+e^{-\sum_{i\in R} w_i\,x_i}},$$

where $R=[n]\setminus S$. This inference of the post-knockout function is based on the decomposition of the sum that appears in (2). Such a decomposition is referred to as a Gibbs representation of κ and contains more information than κ. More generally, we consider the following model of (κ_A)

$$\kappa_A(x_A;x_0) \;=\; \frac{e^{\sum_{B\subseteq A}\phi_B(x_B,x_0)}}{\sum_{x_0'} e^{\sum_{B\subseteq A}\phi_B(x_B,x_0')}}, \tag{3}$$

where the ϕ_B are functions on $\mathcal{X}_B\times\mathcal{X}_0$. Clearly, each κ_A is strictly positive. Using the Möbius inversion, it is easy to see that each strictly positive family (κ_A) has the representation (3). To this end, we simply set

$$\phi_A(x_A,x_0) \;:=\; \sum_{C\subseteq A}(-1)^{|A\setminus C|}\ln\kappa_C(x_C;x_0)\,. \tag{4}$$

Note that this representation is not unique: If an arbitrary function of x_A is added to the function ϕ_A, then (3) does not change.

A single robustness constraint has the following consequences for the ϕ_A.

Proposition 1. *Let $S \subseteq [n]$ and $R = [n] \setminus S$, and let (κ_A) be strictly positive functional modalities with Gibbs potentials (ϕ_A). Then (κ_A) is robust in x against knockout of S if and only if $\sum_{B\subseteq[n], B\not\subseteq R} \phi_B(x|_B, x_0)$ does not depend on x_0.*

Proof. Denote $\tilde{\phi}_A$ the potentials defined via (4). Then (1) is equivalent to

$$\sum_{B\subseteq[n]} \tilde{\phi}_B(x|_B, x_0) = \sum_{B\subseteq R} \tilde{\phi}_B(x|_B, x_0) \quad\Longleftrightarrow\quad \sum_{\substack{B\subseteq[n]\\ B\not\subseteq R}} \tilde{\phi}_B(x_B, x_0) = 0.$$

The statement follows from the fact that $\phi_B(x|_B; x_0) - \tilde{\phi}_B(x|_B; x_0)$ is independent of x_0 (for fixed x). □

Does $\mathcal{R}$-robustness in x imply any structural constraints on (κ_A)? In order to answer this question, we restrict attention to the case $\mathcal{R} = \mathcal{R}_k := \{R \subseteq [n] \ :\ |R| \geq k\}$.

If (κ_A) is $\mathcal{R}_k$-robust on a set $\mathcal{S}$, then the corresponding conditions imposed by Proposition 1 depend on $\mathcal{S}$. In this section, we are interested in conditions that are independent of $\mathcal{S}$. Such conditions allow to define sets of functional modalities that contain all $\mathcal{R}_k$-robust functional modalities for all possible sets $\mathcal{S}$. If the set $\mathcal{S}$ (which will be the support of the input distribution in Section 3) is unknown from the beginning, then the system can choose its policy within such a restricted set of functional modalities.

Denote K_k the set of all functional modalities (κ_A) such that there exist potentials ϕ_A of the form

$$\phi_A(x_A; x_0) = \sum_{\substack{B\subseteq A\\ |B|\leq k}} \Psi_{B,A}(x_B; x_0),$$

where $\Psi_{B,A}$ is an arbitrary function $\mathbb{R}^{\mathcal{X}_B\times\mathcal{X}_0} \to \mathbb{R}$. The set K_k is called the *family of k-interaction functional modalities*. It contains the subset $\tilde{K}_k$ of those functional modalities (κ_A) where the functions $\Psi_{B,A}$ additionally satisfy

$$(-1)^{|A|}\Psi_{B,A}(x_B; x_0) = (-1)^{|A'|}\Psi_{B,A'}(x_B; x_0), \qquad \text{whenever } B \subseteq A\cap A' \text{ and } |B| < k,$$

and

$$\sum_{l=0}^{|A'|-k} \frac{(-1)^{|A'|-l}}{\binom{l+k}{k}}\Psi_{B,A}(x_B; x_0) = \sum_{l=0}^{|A|-k} \frac{(-1)^{|A|-l}}{\binom{l+k}{k}}\Psi_{B,A'}(x_B; x_0), \quad \text{if } B \subseteq A\cap A' \text{ and } |B| = k,$$

for all $x_B \in \mathcal{X}_B$ and $x_0 \in \mathcal{X}_0$. Both K_k and $\tilde{K}_k$ only contain strictly positive kernels. Therefore, we are also interested in the resepective *closures* of these two families with respect to the usual real topology on the space of matrices.

The following holds:

Proposition 2. *Let $\mathcal{S}$ be a subset of $\mathcal{X}_{[n]}$ and let (κ_A) be functional modalities that are $\mathcal{R}_k$-robust in x for all $x \in \mathcal{S}$. Then there exist functional modalities $(\tilde{\kappa}_A)$ in the closure of $\tilde{K}_k$ such that $\kappa_A(x|_A) = \tilde{\kappa}_A(x|_A)$ for all A and all $x \in \mathcal{S}$. In particular, $\tilde{\kappa}_A$ belongs to the closure of the family of k-interactions.*

Proof. Assume first that κ_A is strictly positive. Define Gibbs potentials using the Möbius inversion (4). Note that

$$\begin{aligned}\sum_{\substack{C\subseteq A\\|C|\geq k}}(-1)^{|A\setminus C|}\ln\kappa_C(x_C;x_0) &= \sum_{\substack{C\subseteq A\\|C|\geq k}}(-1)^{|A\setminus C|}\frac{1}{\binom{|C|}{k}}\sum_{\substack{B\subseteq C\\|B|=k}}\ln\kappa_C(x_C;x_0)\\ &= \sum_{\substack{C\subseteq A\\C\in\mathcal{A}}}(-1)^{|A\setminus C|}\frac{1}{\binom{|C|}{k}}\sum_{\substack{B\subseteq C\\|B|=k}}\ln\kappa_B(x_B;x_0)\\ &= \sum_{\substack{B\subseteq A\\|B|=k}}\left\{\sum_{R\subseteq A\setminus B}(-1)^{|A|-|R|-k}\frac{1}{\binom{|R|+k}{k}}\right\}\ln\kappa_B(x_B;x_0)\end{aligned}$$

Together with (4) this gives

$$\phi_A(x_A,x_0) = \sum_{\substack{C\subseteq A\\|C|\leq k}}\alpha_{A,C}\,\ln\kappa_C(x_C;x_0)\,,$$

where

$$\alpha_{A,C} = \begin{cases}(-1)^{|A|-|C|}, & \text{if } |C|<k\\ \sum_{R\subseteq A\setminus C}(-1)^{|A|-|R|-k}\frac{1}{\binom{|R|+k}{k}}, & \text{if } |C|=k\end{cases}$$

depends only on the cardinalities of A and C. The statement follows with the choice $\Psi_{C,A}(x_C;x_0)=\alpha_{A,C}\ln\kappa_C(x_C;x_0)$.

If (κ_A) is not strictly positive, then define $\lambda_A(x_A;x_0)=\frac{1}{d_0}$ for all $A\subseteq[n]$. Then the functional modalities (λ_A) are $\mathcal{R}_k$-robust for all $x\in\mathcal{S}$, and so are the strictly positive functional modalities (κ^ϵ_A) defined via $\kappa^\epsilon_A=(1-\epsilon)\kappa_A+\epsilon\lambda_A$. The statement follows from $\lim_{\epsilon\to0}\kappa^\epsilon_A=\kappa_A$. □

Example 3. Consider the case of $n=2$ binary inputs, $\mathcal{X}_1=\mathcal{X}_2=\{0,1\}$, and let $\mathcal{S}=\{(0,0),(1,1)\}$. Then $\mathcal{R}_1$-robustness on $\mathcal{S}$ means

$$\kappa_{\{1\}}(x_1;x_0)=\kappa_{\{1,2\}}(x_1,x_2;x_0)=\kappa_{\{2\}}(x_2;x_0)$$

for all x_0 whenever $x_1=x_2$. By Proposition 1 this translates into the conditions

$$\phi_{\{1,2\}}(x_1,x_2;x_0)+\phi_{\{1\}}(x_1;x_0)=0=\phi_{\{1,2\}}(x_1,x_2;x_0)+\phi_{\{2\}}(x_2;x_0) \tag{5}$$

for all x_0 whenever $x_1=x_2$ for the potentials (ϕ_A) defined via (4). This means: Assuming that (κ_A) is $\mathcal{R}_1$-robust, it suffices to specify the four functions

$$\phi_\emptyset(x_0),\phi_{\{1\}}(x_1;x_0),\phi_{\{1,2\}}(0,1;x_0),\phi_{\{1,2\}}(1,0;x_0).$$

The remaining potentials can be deduced from (5). If only the values of (κ_A) for $x\in\mathcal{S}$ are needed, then it suffices to specify $\phi_\emptyset(x_0)$ and $\phi_{\{1\}}(x_1;x_0)$.

Even though the families K_k and $\tilde{K}_k$ do not depend on the set $\mathcal{S}$, the choice of the set $\mathcal{S}$ is essential: If the set $\mathcal{S}$ is too large, then the conditions (1) imply that the output X_0 is (unconditionally) independent of all inputs. The theory developed in Sections 3 to 5 discusses the constraints on conditionals imposed by the choice of $\mathcal{S}$. In particular, Section 4 gives bounds on the strength of the interaction between the input nodes and the output node for given $\mathcal{R}$ and $\mathcal{S}$.

On the other hand, since K_k and $\tilde{K}_k$ are independent of $\mathcal{S}$, Proposition 2 shows that these two families can be used to construct robust systems, when the input distribution p_{in} is not known a priori (or may change over time) but must be learned by the system.

3. Robustness and conditional independence

We now study robustness of the joint distribution p of $(X_0, X_{[n]})$. As stated in the introduction, p is called *robust against knockout of S* if it satisfies $X_0 \perp\!\!\!\perp X_S \,|\, X_R$, where $R = [n] \setminus S$. By definition this means that

$$p(x_0, x_S, x_R)p(x_0', x_S', x_R) = p(x_0, x_S', x_R)p(x_0', x_S, x_R), \tag{6}$$

for all $x_0, x_0' \in \mathcal{X}_0, x_S, x_S' \in \mathcal{X}_S$ and $x_R \in \mathcal{X}_R$. Here, $p(x_0, x_S, x_R)$ is an abbreviation of $p(X_0 = x_0, X_S = x_S, X_R = x_R)$. It is not difficult to see that this definition is equivalent to the usual definition of conditional independence [3]. This algebraic formulation makes it possible to study conditional independence with algebraic tools.

In order to formulate the results in higher generality, we will also consider CI statements of the form $X_0 \perp\!\!\!\perp X_S \,|\, X_R = y$ for some $S \subseteq [n]$, $R = [n] \setminus S$ and $y \in \mathcal{X}_R$. By definition, this is equivalent to equations (6) for all $x_0, x_0' \in \mathcal{X}_0, x_S, x_S' \in \mathcal{X}_S$ and $x_R = y$. Such a statement models the case that, if the value of the input variables $\mathcal{X}_R$ is y, then the system does not need to know the remaining variables X_S in order to compute its output. Such CI statements naturally generalize canalizing [8] or nested canalizing functions [7], which have been studied in the context of robustness. The simpler statement $X_0 \perp\!\!\!\perp X_S \,|\, X_R$ corresponds to the special case where $X_0 \perp\!\!\!\perp X_S \,|\, X_R = y$ for all $y \in \mathcal{X}_R$.

Let $\mathcal{R}$ be a collection of pairs (R, y), where $R \subseteq [n]$ and $y \in \mathcal{X}_R$. Such a collection will be called a *robustness specification* in the following. A joint distribution is called *$\mathcal{R}$-robust* if it satisfies all conditional independence (CI) statements

$$X_0 \perp\!\!\!\perp X_{[n]\setminus R} \,\big|\, X_R = y \tag{7}$$

for all $(R, y) \in \mathcal{R}$. We denote $\mathcal{P}_{\mathcal{R}}$ the set of all $\mathcal{R}$-robust probability distributions.

Example 4. As before, let $\mathcal{R}_k$ be the set of subsets of $[n]$ of cardinality k or greater. In other words, a probability measure p is $\mathcal{R}_k$-robust, if we can knock out any $n - k$ input variables without losing information on the output.

Equations (6) are polynomial equations in the elementary probabilities. They are related to the *binomial edge ideals* introduced in [6]. The generalized binomial edge ideals will be studied in Section 6. Here, we interpret the algebraic results from the point of view of robustness.

Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. A robustness specification $\mathcal{R}$ induces a graph $G_{\mathcal{R}}$ on $\mathcal{X}$, where $x, x' \in \mathcal{X}$ are connected by an edge if and only if there exists $(R, y) \in \mathcal{R}$ such that the restrictions of x and x' to R satisfy $x|_R = x'|_R = y$.

Definition 5. Let $\mathcal{Y} \subseteq \mathcal{X}$, and denote $G_{\mathcal{R},\mathcal{Y}}$ the subgraph of $G_{\mathcal{R}}$ induced by $\mathcal{Y}$. The set $\mathcal{Y}$ is called *$\mathcal{R}$-connected* if $G_{\mathcal{R},\mathcal{Y}}$ is connected. The set of connected components of $G_{\mathcal{R},\mathcal{Y}}$ is called a *$\mathcal{R}$-robustness structure*. An $\mathcal{R}$-robustness structure $\mathbf{B}$ is *maximal* if and only if $\cup\mathbf{B} := \cup_{\mathcal{Z}\in\mathbf{B}}\mathcal{Z}$ satisfies any of the following equivalent conditions:

(1) For any $x \in \mathcal{X} \setminus \cup\mathbf{B}$ there are edges (x, y), (x, z) in $G_{\mathcal{R}}$ such that $y, z \in \cup\mathbf{B}$ are not connected in $G_{\mathcal{R},\cup\mathbf{B}}$.

(2) For any $x \in \mathcal{X} \setminus \cup\mathbf{B}$ the induced subgraph $G_{\mathcal{R},\cup\mathbf{B}\cup\{x\}}$ has less connected components than $G_{\mathcal{R},\cup\mathbf{B}}$.

For any probability distribution p on $\mathcal{X}$, $x_0 \in \mathcal{X}_0$ and $x \in \mathcal{X}$ denote $\tilde{p}_x$ the vector with components $\tilde{p}_x(x_0) = p(X_0 = x_0, X_{[n]} = x)$. Denote supp $\tilde{p} := \{x \in \mathcal{X} : \tilde{p}_x \neq 0\}$. For any family $\mathbf{B}$ of subsets of $\mathcal{X}$ let $\mathcal{P}_{\mathbf{B}}$ be the set of probability distributions p that satisfy the following two conditions:

(1) supp $\tilde{p} = \cup\mathbf{B}$,
(2) $\tilde{p}_x$ and $\tilde{p}_y$ are proportional, whenever there exists $\mathcal{Z} \in \mathbf{B}$ such that $x, y \in \mathcal{Z}$.

It follows from (10) and Theorem 23 that $\mathcal{P}_{\mathcal{R}}$ equals the disjoint union $\cup_{\mathbf{B}}\mathcal{P}_{\mathbf{B}}$, where the union is over all $\mathcal{R}$-robustness structures. Alternatively, $\mathcal{P}_{\mathcal{R}}$ equals the union $\cup_{\mathbf{B}}\overline{\mathcal{P}_{\mathbf{B}}}$, where the union is over all maximal $\mathcal{R}$-robustness structures.

For any $x \in \mathcal{X}$ the vector $\tilde{p}_x$ is proportional to the conditional probability distribution $P(\cdot|X_{[n]} = x)$ of X_0 given that $X_{[n]} = x$. Hence:

Lemma 6. *Let p be a probability distribution, and let $\mathbf{B}$ be the set of connected components of $G_{\mathcal{R},\mathrm{supp}\,\tilde{p}}$. Then p is $\mathcal{R}$-robust if and only if $P(\cdot|X_{[n]} = x) = P(\cdot|X_{[n]} = y)$ whenever there exists $\mathcal{Z} \in \mathbf{B}$ such that $x, y \in \mathcal{Z}$.*

The following lemma sheds light on the structure of $\mathcal{P}_{\mathbf{B}}$:

Lemma 7. *Fix an $\mathcal{R}$-robustness structure $\mathbf{B}$. Then $\mathcal{P}_{\mathbf{B}}$ consists of all probability measures of the form*

$$p_{x_0 x} = \begin{cases} \mu(\mathcal{Z})\lambda_{\mathcal{Z}}(x)p_{\mathcal{Z}}(x_0) & \text{if } x \in \mathcal{Z} \in \mathbf{B}, \\ 0 & \text{if } x \in \mathcal{X} \setminus \cup\mathbf{B}, \end{cases} \tag{8}$$

where μ is a probability distribution on $\mathbf{B}$ and $\lambda_{\mathcal{Z}}$ is a probability distribution on $\mathcal{Z}$ for each $\mathcal{Z} \in \mathbf{B}$ and $(p_{\mathcal{Z}})_{\mathcal{Z}\in\mathbf{B}}$ is a family of probability distributions on $\mathcal{X}_0$.

Proof. It is easy to see that p is indeed a probability distribution. By Lemma 6 it belongs to $\mathcal{P}_{\mathbf{B}}$. In the other direction, any probability measure can be written as a product

$$p(x_0, x_1, \dots, x_n) = p(\mathcal{Z})p\left(x_1, \dots, x_n|(X_1, \dots, X_n) \in \mathcal{Z}\right) p(x_0|x_1, \dots, x_n),$$

if $(x_1, \dots, x_n) \in \mathcal{Z} \in \mathbf{B}$, and if p is an $\mathcal{R}$-robust probability distribution, then $p_{\mathcal{Z}}(x_0) := p(x_0|x_1, \dots, x_n)$ depends only on the block $\mathcal{Z}$ in which $(x_1, \dots, x_n)$ lies. □

4. Robust functions

The factorization in Lemma 7 admits the following interpretation:

Proposition 8. *Let $\mathbf{B}$ be an $\mathcal{R}$-robustness structure. Then the set $\overline{\mathcal{R}}_B$ is the set of probability distributions such that*

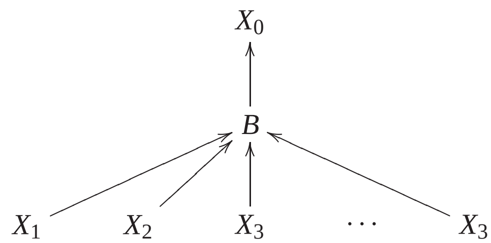
$$p(X_{[n]} \in \cup\mathbf{B}) = 1$$

and

$$X_0 \perp\!\!\!\perp X_{[n]} \,\Big|\, X_{[n]} \in \mathcal{Z} \qquad \text{for all } \mathcal{Z} \in \mathbf{B}.$$

In other words, the sets $\mathcal{Z} \in \mathbf{B}$ determine a partition of the set $\operatorname{supp} \tilde{p}$, which consists of all outcomes of $X_{[n]}$ with non-zero probability under p. Within each block $\mathcal{Z}$ the value of X_0 is independent of all inputs. Let $R \subseteq [n]$, and let $x, x' \in \mathcal{X}_{[n]}$ satisfying $(R, x|_R) \in \mathcal{R}$ and $(R, x'|_R) \in \mathcal{R}$. If x and x' belong to different blocks in $\mathbf{B}$, then $x|_R \neq x'|_R$. Therefore, the knowledge of the input variables in R is sufficient to determine in which block $\mathcal{Z} \in \mathbf{B}$ we are.

When p or $\mathbf{B}$ is fixed we can introduce an additional random variable B that takes values in $\mathbf{B}$. The situation is illustrated by the following graph:



The arrows from the input variables $X_1, \ldots, X_n$ to B are, in fact, deterministic:

$$B(x) = \mathcal{Z} \qquad \text{if } x \in \mathcal{Z} \in \mathbf{B}.$$

Note, however, that the function B is only defined uniquely on $\cup\mathbf{B}$, which is a set of measure one with respect to p. This means that in many cases it is enough to study robustness of functions on $\mathcal{X}$.

Definition 9. A function f defined on a subset $\mathcal{S} \subseteq \mathcal{X}_{[n]}$ is $\mathcal{R}$-robust if there exists an $\mathcal{R}$-robustness structure $\mathbf{B}$ such that $\mathcal{S} = \cup\mathbf{B}$ and f is constant on each $B \in \mathbf{B}$.

There are two motivations for looking at this kind of functions: First, they occur in the special case of $\mathcal{R}$-robust probability distributions $p(X_0, X_1, \ldots, X_n)$ such that all conditional probability distributions $p(X_0|x_1, \ldots, x_n)$ are Dirac measure. Second, as motivated above, we can associate to any $\mathcal{R}$-robust probability distribution p a corresponding function f characterizing the $\mathcal{R}$-robustness structure. In order to reconstruct p it is enough to specify the input distribution $p_{\text{in}}(X_1, \ldots, X_n)$ and a set of output distributions $\{p(X_0|(X_1, \ldots, X_n) \in \mathcal{Z})\}_{\mathcal{Z}\in\mathbf{B}}$ in addition to the function $f : \mathcal{S} \to \mathbf{B}$. Note that natural examples of robust functions arise from the study of canalizing functions [8, 7].

It is natural to ask the following question: Given a certain robustness structure, how much freedom is left to choose a robust function f? More precisely, how large can the image of f be? Equivalently, how many components can an $\mathcal{R}$-robustness structure $\mathbf{B}$ have?

Lemma 10. *Let f be an $\mathcal{R}$-robust function. The cardinality of the image of f is bounded from above by*

$$\min\left\{\prod_{i\in R} d_i : (R, y) \in \mathcal{R} \text{ for all } y \in \mathcal{X}_R\right\}.$$

Proof. Suppose without loss of generality that $(\{1, \ldots, r\}, y) \in \mathcal{R}$ for all $y \in \mathcal{X}_{[r]}$ and that $d_1 \ldots d_r$ equals the above minimum. The image of f cannot be larger than $d_1 \ldots d_r$, since if we knock out all X_i for $i > r$, then we can only determine $d_1 \ldots d_r$ states. □

Example 11. Suppose that $\mathcal{S} = \mathcal{X}$. This means that the $\mathcal{R}$-robustness structure satisfies $\cup\mathbf{B} = \mathcal{X}$. We first consider the case that $G_{\mathcal{R}}$ is connected. This is fulfilled, for example, if for any $k \in [n]$ there exists $R \subseteq [n]$ such that $k \notin S$ and $(R, y) \in \mathcal{R}$ for all $y \in \mathcal{X}_R$. In this case an $\mathcal{R}$-robust function f takes only one value.

Assume that $(R, y) \in \mathcal{R}$ implies $(R, y') \in \mathcal{R}$ for all $y' \in \mathcal{X}_R$. If $G_{\mathcal{R}}$ is not connected, then some input variables may never be knocked out. Let T be the set of these input variables. For every fixed value of X_T the function f must be constant. This means that f can have $\prod_{i\in[n]\setminus T} d_i$ different values.

Remark (Relation to coding theory). We can interpret $\mathcal{X}$ as a set of words over the alphabet $[d_m]$ of length n, where $d_m = \max\{d_i\}$. For simplicity assume that all d_i are equal. Consider the uniform case $\mathcal{R} = \mathcal{R}_k$. Then the task is to find a collection of subsets such that any two different subsets have Hamming distance at least k. A related problem appears in coding theory: A code is a subset $\mathcal{Y}$ of $\mathcal{X}$ and corresponds to the case that each element of $\mathbf{B}$ is a singleton. If distinct elements of the code have Hamming distance at least $n - k$, then a message can be reliably decoded even if only k letters are transmitted correctly.

5. $\mathcal{R}_k$-robustness

In this section we consider the symmetric case $\mathcal{R} = \mathcal{R}_k$. We fix n and replace any prefix or subscript $\mathcal{R}$ by k.

Let $k = 0$. Any pair (x, y) is an edge in G_0. This means that $\mathbf{B}$ can contain only one set B. There is only one maximal 0-robustness structure, namely $\overline{\mathbf{B}} = \{\mathcal{X}_{[n]}\}$. The set $\mathcal{R}_0$ is irreducible. This corresponds to the fact that $\mathcal{P}_n$ is defined by $X_0 \perp\!\!\!\perp X_{[n]}$.

$\overline{\mathbf{B}}$ is actually a maximal k-robustness structure for any $0 \leq k \leq n$. This illustrates the fact that the single CI statement $X_0 \perp\!\!\!\perp X_{[n]}$ implies all other CI statements of the form (7). The corresponding set $\mathcal{P}_{\overline{\mathbf{B}}}$ contains all probability distributions of $\mathcal{P}_0$ of full support.

Now let $k = 1$. In the case $n = 2$, we obtain results by Alexander Fink, which can be reformulated as follows [5]: *Let $n = 2$. A 1-robustness structure $\mathbf{B}$ is maximal if and only if the following statements hold:*

- *Each $B \in \mathbf{B}$ is of the form $B = S_1 \times S_2$, where $S_1 \subseteq \mathcal{X}_1, S_2 \subseteq \mathcal{X}_2$.*
- *For every $x_1 \in \mathcal{X}_1$ there exists $B \in \mathbf{B}$ and $x_2 \in \mathcal{X}_2$ such that $(x_1, x_2) \in B$, and conversely.*

In [5] a different description is given: The block $S_1 \times S_2$ can be identified with the complete bipartite graph on S_1 and S_2. In this way, every maximal 1-robustness structure corresponds to a collection of complete bipartite subgraphs with vertices in $\mathcal{X}_1 \cup \mathcal{X}_2$ such that every vertex in $\mathcal{X}_1$ resp. $\mathcal{X}_2$ is part of one such subgraph.

This result generalizes in the following way:

Lemma 12. *A 1-robustness structure $\mathbf{B}$ is maximal if and only if the following statements hold:*

- *Each $B \in \mathbf{B}$ is of the form $B = S_1 \times \cdots \times S_n$, where $S_i \subseteq \mathcal{X}_i$.*
- *Fix $j \in [n]$ and $x_i \in \mathcal{X}_i$ for all $i \in [n]$, $i \neq j$. Then there exist $x_j \in \mathcal{X}_j$ such that $(x_1, \ldots, x_n) \in \cup_{B\in\mathbf{B}}B$. In other words, whenever $n - 1$ components of $(x_1, \ldots, x_n)$ are prescribed, there exist an n-th component such that $(x_1, \ldots, x_n) \in \cup_{B\in\mathbf{B}}B$.*

Proof. We say that a subset $\mathcal{Y}$ of $\mathcal{X}$ is connected if $G_{\mathcal{R},\mathcal{Y}}$ is connected. Suppose that $\mathbf{B}$ is maximal. Let $B \in \mathbf{B}$ and let S_i be the projection of $B \subseteq \mathcal{X}_{[n]}$ to $\mathcal{X}_i$. Let $B' = S_1 \times \cdots \times S_n$. Then $B \subseteq B'$. We claim that $(\mathbf{B} \setminus \{B\}) \cup \{B'\}$ is another coarser 1-robustness structure. By Definition 5 we need to show that B' is connected and that $A \cup B'$ is not connected for all $A \in \mathbf{B} \setminus \{B\}$. The first condition follows from the fact that B is connected. For the second condition assume to the contrary that there are $x \in B'$ and $y \in A$ such that $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ disagree in at most $n - 1$ components. Then there exists a common component $x_l = y_l$. By construction there exists $z = (z_1, \dots, z_n) \in B$ such that $z_l = y_l = x_l$, hence $A \cup B$ is connected, in contradiction to the assumptions. This shows that each B has a product structure.

Write $B = S_1^B \times \cdots \times S_n^B$ for each $B \in \mathbf{B}$. Obviously $S_i^B \cap S_i^{B'} = \emptyset$ for all $i \in [n]$ and all $B, B' \in \mathbf{B}$ if $B \neq B'$. The second assertion claims that $\cup_{B \in \mathbf{B}} S_i^B = \mathcal{X}_i$ for all $i \in [n]$: Assume to the contrary that $l \in \mathcal{X}_i$ is contained in no S_i^B. Take any B and define $B' := S_1^B \times \cdots \times (S_i^B \cup \{l\}) \times \cdots \times S_n^B$. Then $(\mathbf{B} \setminus B) \cup \{B'\}$ is a coarser 1-robustness structure.

Now assume that $\mathbf{B}$ is a 1-robustness structure satisfying the two assertions of the theorem. For any $x \in \mathcal{X} \setminus \cup\mathbf{B}$ there exists $y \in \cup\mathbf{B}$ such that $x_1 = y_1$, and hence (x, y) is an edge in G_1. This implies maximality. □

The last result can be reformulated in terms of n-partite graphs generalizing [5]: Namely, the 1-robustness structures are in one-to-one relation with the n-partite subgraphs of $K_{d_1,\dots,d_n}$ such that every connected component is itself a complete n-partite subgraph $K_{e_1,\dots,e_n}$ with $e_i > 0$ for all $i \in [n]$. Here, an n-partite subgraph is a graph which can be coloured by n colours such that no two vertices with the same colour are connected by an edge.

Unfortunately the nice product form of the maximal 1-robustness structures does not generalize to $k > 1$:

Example 13 (Binary three inputs)**.** If $n = 3$ and $d_1 = d_2 = d_3 = 2$ and $k = 2$, then the graph $G_{\mathcal{R}}$ is the graph of the cube. For a maximal 1-robustness structure $\mathbf{B}$ the set $\mathcal{X} \setminus \cup\mathbf{B}$ can be any one of the following:

- The empty set
- A set of cardinality 4 corresponding to a plane leaving two connected components of size 2
- A set of cardinality 4 containing all vertices with the same parity.
- A set of cardinality 3 cutting off a vertex.

An example for the last case is

$$\mathbf{B} := \{\{(1,1,1)\}, \{(2,2,2), (2,2,1), (2,1,2), (1,2,2)\}\} .$$

Only the isolated vertex has a product structure.

Generically, the smaller k, the easier it is to describe the structure of all k-robustness structures. We have seen above that the cases $k = 0$ and $k = 1$ are particularly nice. One might expect that all k-robustness structures are also $(k+1)$-robustness structures for all k. Unfortunately, this is not true in general:

Example 14. Consider $n = 4$ binary random variables $X_1, \dots, X_4$. Then

$$\mathbf{B} := \{\{(1,1,1,1), (2,2,1,1)\}, \{(1,2,2,2), (2,1,2,2)\}\}$$

is a maximal 2-robustness structure. Both elements of $\mathbf{B}$ are $\sim_2$-connected, but not $\sim_3$-connected.

The following two lemmas relate k-robustness to l robustness for $l > k$:

Lemma 15. *Let $\mathbf{B}$ be a k-robustness structure. For every $l > k$ there exists an l-robustness structure $\mathbf{B}'$ such that the following holds: For any $\mathcal{Y} \in \mathbf{B}$ there exists precisely one $\mathcal{Y}' \in \mathbf{B}'$ such that $\mathcal{Y} \subseteq \mathcal{Y}'$.*

Proof. The statements (7) for k imply the same statements of l, so $\overline{\mathcal{P}_\mathbf{B}}$ is a closed subset of $\mathcal{P}_l$. Thus $\overline{\mathcal{P}_\mathbf{B}}$ lies in one irreducible subset $\overline{\mathcal{P}_{\mathbf{B}'}}$ of $\mathcal{P}_l$. The statement now follows from Lemma 22. □

Lemma 16. *Assume that $d_1 = \cdots = d_n = 2$, and let $\mathbf{B}$ be a maximal k-robustness structure of binary random variables. Then each $B \in \mathbf{B}$ is connected as a subset of G_s for all $s \leq n - 2k$.*

Proof. We can identify elements of $\mathcal{X}$ with 01-strings of length n. Denote I_r the string $1\dots10\dots0$ of r ones and $n-r$ zeroes in this order. Without loss of generality assume that I_0, I_l are two elements of $B \in \mathbf{B}$, where $k \geq n - l < s$. Let $m = \lceil \frac{l}{2} \rceil$ and consider I_m. We want to prove that we can replace B by $B \cup \{I_m\}$ and obtain another, coarser k-robustness structure. By maximality this will imply that I_0 and I_l are indeed connected by a path in G_s.

Otherwise there exists $A \in \mathbf{B}$ and $x \in A$ such that x and I_m agree in at least k components. Let a be the number of zeroes in the first m components of x, let b be the number of ones in the components from $m+1$ to l and let c be the number of ones in the last $n-l$ components. Then I_m and x disagree in $a+b+c \leq n-k$ components. On the other hand, x and I_0 disagree in $(m-a)+b+c$ components, and x and I_l disagree in $a+((l-m)-b)+c \leq a+(m-b)+c$ components. Assume that $a \geq b$ (otherwise exchange I_0 and I_l). Then x and I_0 disagree in at most $m + c \leq \lceil \frac{l}{2} \rceil + n - l = n - \lfloor \frac{l}{2} \rfloor \leq n - k$ components, so $A \cup B$ is connected, in contradiction to the assumptions. □

6. Generalized binomial edge ideals

We refer to [2] for an introduction to the algebraic terminology that is used in this section.

Let $\mathcal{X}$ be a finite set, $d_0 > 1$ an integer, and denote $\tilde{\mathcal{X}} = \mathcal{X}_0 \times \mathcal{X}$. Fix a field $\mathbb{K}$. Consider the polynomial ring $R = \mathbb{K}[p_x : x \in \tilde{\mathcal{X}}]$ with $|\tilde{\mathcal{X}}|$ unknowns p_x indexed by $\tilde{\mathcal{X}}$. For all $i, j \in \mathcal{X}_0$ and all $x, y \in \mathcal{X}$ let

$$f_{xy}^{ij} = p_{ix}p_{jy} - p_{iy}p_{jx}.$$

For any graph G on $\mathcal{X}$ the ideal I_G in R generated by the binomials f_{xy}^{ij} for all $i, j \in \mathcal{X}_0$ and all edges (x, y) in G is called the d_0th *binomial edge ideal* of G over $\mathbb{K}$. This is a direct generalization of [6] and [9], where the same ideals have been considered in the special case $d_0 = 2$.

Choose a total order $>$ on $\mathcal{X}$ (e.g. choose a bijection $\mathcal{X} \cong [|\mathcal{X}|]$). This induces a lexicographic monomial order, that will also be denoted by $>$, via

$$p_{ix} > p_{jy} \quad \Longleftrightarrow \quad \begin{cases} \text{either} & i > j, \\ \text{or} & i = j \text{ and } x > y. \end{cases}$$

A Gröbner basis for I_G with respect to this order can be constructed using the following definitions:

Definition 17. A path $\pi : x = x_0, x_1, \dots, x_r = y$ from x to y in $\mathcal{X}$ is called *admissible* if

(i) $x_k \neq x_\ell$ for $k \neq \ell$, and $x < y$;
(ii) for each $k = 1, \dots, r-1$ either $x_k < x$ or $x_k > y$;
(iii) for any proper subset $\{y_1, \dots, y_s\}$ of $\{x_1, \dots, x_{r-1}\}$, the sequence $x, y_1, \dots, y_s, y$ is not a path.

A function $\kappa : \{0, \dots, r\} \to [d]$ is called *π-antitone* if it satisfies

$$x_s < x_t \Longrightarrow \kappa(s) \geq \kappa(t), \text{ for all } 1 \leq s, t \leq r. \tag{9}$$

κ is *strictly π-antitone* if it is π-antitone and satisfies $\kappa(0) > \kappa(r)$.

The notion of π-antitonicity also applies to paths which are not necessarily admissible. However, since admissible paths are *injective* (i.e. they only pass at most once at each vertex), we may write $\kappa(\ell)$ in the admissible case, instead of $\kappa(s)$, if $\ell = \pi(s)$.

For any $x < y$, any admissible path $\pi : x = x_0, x_1, \dots, x_r = y$ from x to y and any π-antitone function κ associate the monomial

$$u_\pi^\kappa = \prod_{k=1}^{r-1} p_{\kappa(k) x_k}.$$

Theorem 18. *The set of binomials*

$$\mathcal{G} = \bigcup_{i<j} \left\{ u_\pi^\kappa f_{xy}^{\kappa(y)\kappa(x)} \ : \ x < y,\ \pi \text{ is an admissible path in } G \text{ from } x \text{ to } y, \ \kappa \text{ is strictly } \pi\text{-antitone} \right\}$$

is a reduced Gröbner basis of I_G with respect to the monomial order introduced above.

The proof makes use of the following lemma, which explains π-antitonicity:

Lemma 19. *Let $\pi : x_0, \dots, x_r$ be a path in G, and let $\kappa : \{0, \dots, r\} \to [d]$ be an arbitrary function. If κ is not π-antitone, then there exists $g \in \mathcal{G}$ such that* $\mathrm{ini}_<(g)$ *divides the monomial $u_\pi^\kappa = \prod_{k=1}^{r-1} p_{\kappa(k) x_k}$.*

Proof. Let $\tau : y_0, \dots, y_s$ be a minimal subpath of π with respect to the property that the restriction of κ to τ is not τ-antitone. This means that κ is τ_0-antitone and τ_s-antitone, where $\tau_0 = y_1, \dots, y_s$ and $\tau_s = y_0, \dots, y_{s-1}$. Assume without loss of generality that $y_0 < y_s$, otherwise reverse τ. The minimality implies that $\kappa(y_0) < \kappa(y_s)$. It follows that τ is admissible: By minimality, if $y_0 < y_k < y_s$, then $\kappa(y_k) \geq \kappa(y_s) > \kappa(y_0) \geq \kappa(y_k)$, a contradiction. Define

$$\kappa(k) = \begin{cases} \kappa(s), & \text{if } k = 0, \\ \kappa(0), & \text{if } k = s, \\ \kappa(k), & \text{if } 0 < k < s. \end{cases}$$

Then κ is τ-antitone, and $\mathrm{ini}_<(u_\tau^\kappa f_{y_0 y_s}^{\kappa(y_s)\kappa(y_0)})$ divides u_π^κ. □

Proof of Theorem 18. The proof is organized in three steps.

Step 1: $\mathcal{G}$ is a subset of I_G. Let $\pi : x = x_0, x_1, \dots, x_{r-1}, x_r = y$ be an admissible path in G. We show that $u_\pi^\kappa f_{xy}^{\kappa(j)\kappa(i)}$ belongs to I_G using induction on r. Clearly the assertion is true if $r = 1$, so assume $r > 1$. Let $A = \{x_k : x_k < x\}$ and $B = \{x_\ell : x_\ell > y\}$. Then either $A \neq \emptyset$ or $B \neq \emptyset$.

Suppose $A \neq \emptyset$ and set $x_k = \max A$. The two paths $\pi_1 : x_k, x_{k-1}, \dots, x_1, x_0 = x$ and $\pi_2 : x_k, x_{k+1}, \dots, x_{r-1}, x_r = y$ in G are admissible. Let κ_1 and κ_2 be the restrictions of κ to π_1 and π_2. Let $a = \kappa(r)$, $b = \kappa(0)$ and $c = \kappa(k)$. The calculation

$$
\begin{aligned}
&(p_{by}p_{ax} - p_{bx}p_{ay})p_{cx_k} \\
&\qquad = (p_{cx}p_{bx_k} - p_{cx_k}p_{bx})p_{ay} - (p_{cx}p_{ax_k} - p_{cx_k}p_{ax})p_{by} - p_{cx}(p_{bx_k}p_{ay} - p_{by}p_{ax_k})
\end{aligned}
$$

implies that $u^\kappa_\pi f^{ab}_{xy}$ lies in the ideal generated by $u^{\kappa_1}_{\pi_1} f^{bc}_{xx_k}$, $u^{\kappa_1}_{\pi_1} f^{ac}_{xx_k}$ and $u^{\kappa_2}_{\pi_2} f^{ab}_{x_k y}$. By induction it lies in I_G.

The case $B \neq \emptyset$ can be treated similarly.

Step 2: $\mathcal{G}$ is a Gröbner basis of I_G. Let $\pi : x_0, \dots, x_r$ and $\sigma : y_0, \dots, y_s$ be admissible paths in G with $x_0 < x_r$ and $y_0 < y_s$, and let κ and μ be π- and σ-antitone. By Buchberger's criterion we need to show that the S-pairs $s := S(u^\kappa_\pi f^{\kappa(r)\kappa(0)}_{x_0x_r}, u^\mu_\sigma f^{\mu(s)\mu(0)}_{y_0y_s})$ reduces to zero.

If $S \neq 0$, then S is a binomial. Write $S = S_1 - S_2$, where $S_1 = \mathrm{ini}_<(S)$. S is homogeneous with respect to the multidegrees given by

$$
\deg(p_{zm})_b = \delta_{zb} = \begin{cases} 1, & \text{if } z = b, \\ 0, & \text{else.} \end{cases}
$$

and

$$
\deg(p_{zm})_n = \delta_{mn} = \begin{cases} 1, & \text{if } m = n, \\ 0, & \text{else.} \end{cases}
$$

If π and σ are disjoint, then $S = 0$, since $u^\kappa_\pi f^{\kappa(r)\kappa(0)}_{x_0x_r}$ and $u^\mu_\sigma f^{\mu(s)\mu(0)}_{y_0y_s}$ contain different variables. The same happens if the intersection of π and σ does not involve the starting or end points of π and σ, since in this case S is proportional to the S-pair of the two monomials u^κ_π and u^μ_σ.

Assume that π and σ meet and that $S \neq 0$. Then S_1 and S_2 are monomials, and the unknowns p_{ix} occurring in S_1 and S_2 satisfy $x \in \pi \cup \sigma$. Assume that there are $x < y$ such that $D_x := \min\{i \in \mathcal{X}_0 : p_{ix} \,|\, S_1\} < \max\{i \in \mathcal{X}_0 : p_{iy} \,|\, S_1\} =: D_y$. Since $\pi \cup \sigma$ is connected there is an injective path $\tau : z_0, \dots, z_s$ from $x = z_0$ to $y = z_s$ in $\pi \cup \sigma$. Choose a map $\lambda : \{0, \dots, s\}$ such that $\lambda(0) = D_x$, $\lambda(s) = D_y$ and $p_{\lambda(a)a} \,|\, S_1$ for all $0 \leq a \leq s$. Then u^λ_τ divides S_1, and λ is not τ-antitone. So we can apply Lemma 19 in order to reduce S to a smaller binomial.

Let S' be the reduction of S modulo $\mathcal{G}$. If $S' \neq 0$, then let $S'_1 = \mathrm{ini}_<(S')$. The above argument shows that $\min\{i \in \mathcal{X}_0 : p_{ix} \,|\, S'_1\} \geq \max\{i \in \mathcal{X}_0 : p_{iy} \,|\, S'_1\}$ for all $x < y$. This property characterizes S'_1 as the unique minimal monomial in R with multidegree $\deg(S'_1) = \deg(S)$. But since the reduction algorithm turns binomials into binomials, $S' - S'_1$ is also a monomial of multidegree $\deg(S)$, and smaller than $\deg(S'_1)$. This contradiction shows $S' = 0$.

Step 3: $\mathcal{G}$ is reduced. Let $\pi : x_0, \dots, x_r$ and $\sigma : y_0, \dots, y_s$ be admissible paths in G with $x_0 < x_r$ and $y_0 < y_s$, and let κ and μ be π- and σ-antitone. Let $u = \kappa(r), v = \kappa(0), w = \mu(s), t = \mu(0)$, and suppose that $u^\kappa_\pi p_{ux_0} p_{vx_r}$ divides either $u^\mu_\sigma p_{wy_0} p_{ty_s}$ or $u^\mu_\sigma p_{wy_s} p_{ty_0}$. Then $\{x_0, \dots, x_r\}$ is a subset of $\{y_0, \dots, y_s\}$, and $\kappa(b) = \mu(\sigma^{-1}(x_b))$ for $0 < b < r$.

If $x_0 = y_0$ and $x_r = y_s$, then π is a sub-path of σ. By Definition 17, π equals σ (up to a possible change of direction). Hence $u^\kappa_\pi f^{\kappa(r)\kappa(0)}_{x_0x_r}$ and $u^\mu_\sigma f^{\mu(s)\mu(0)}_{y_0y_s}$ have the same (total) degree, hence they agree.

If $x_0 = y_0$ and $x_r \neq y_s$, then p_{vx_r} divides u^μ_σ, and so $x_r = y_t$ for some $t < s$ such that $v = \mu(t)$. Then $y_t = x_r > x_0 = y_0$, and hence $v \leq \mu(0) = \kappa(0) < \kappa(r)$, in contradiction to $x_0 < x_r$. A similar argument applies if $x_0 \neq y_0$ and $x_r = y_s$. Finally, if $x_0 \neq y_0$ and $x_r \neq y_s$, then $p_{ux_0}p_{vx_r}$ divides u^μ_σ. This implies $u = \kappa(0) = \kappa(j) = v$, a contradiction. ◻

Corollary 20. *I_G is a radical ideal.*

Proof. The assertion follows from Theorem 18 the following general fact: A graded ideal that has a Gröbner basis with square-free initial terms is radical. See the proof of [6, Corollary 2.2] for the details. ◻

Since I_G is radical, in order to compute the primary decomposition of the ideal it is enough to compute the minimal primes. We are mainly interested in the irreducible decomposition of the variety V_G of I_G in the case of characteristic zero. While the basic arguments remain true for finite base fields there is no relation between the primary decomposition of an ideal and the irreducible decomposition of its variety, since the irreducible decomposition consists of all closed points in this case. The following definition is needed: Two vectors v, w (living in the same $\mathbb{R}$-vector space) are *proportional* whenever $v = \lambda w$ or $w = \lambda v$ for some $\lambda \in \mathbb{R}$. A set of vectors is *proportional* if each pair is proportional. Since $\lambda = 0$ is allowed, proportionality is not transitive: If v and w are proportional and if u and v are proportional, then we can conclude that u and w must be proportional only if $v \neq 0$.

We now study the solution variety V_G of I_G, which is a subset of $\mathbb{R}^{\mathcal{X}_0\times\mathcal{X}}$. As usual, elements of $\mathbb{R}^{\mathcal{X}_0\times\mathcal{X}}$ will be denoted with the same symbol $p = (p_{ix})_{i\in\mathcal{X}_0,x\in\mathcal{X}}$ as the unknowns in the polynomial ring $R = \mathbb{R}[p_{ix} : (i, x) \in \mathcal{X}_0 \times \mathcal{X}]$. Such a p can be written as a $d_0 \times |\mathcal{X}|$-matrix. Each binomial equation in I_G imposes conditions on this matrix saying that certain submatrices have rank 1. For a fixed edge (x, y) in G the equations $f^{ij}_{xy} = 0$ for all $i, j \in \mathcal{X}_0$ require that the submatrix $(p_{kz})_{k\in\mathcal{X}_0,z\in\{x,y\}}$ has rank one. More generally, if $K \subseteq \mathcal{X}$ is a clique (i.e. a complete subgraph), then the submatrix $(p_{kz})_{k\in\mathcal{X}_0,z\in K}$ has rank one. This means that all columns of this submatrix are proportional. The columns of p will be denoted by $\tilde{p}_x$, $x \in \mathcal{X}$. A point p lies in V_G if and only if $\tilde{p}_x$ and $\tilde{p}_y$ are proportional for all edges (x, y) of G.

Even if the graph G is connected, not all columns $\tilde{p}_x$ must be proportional to each other, since proportionality is not a transitive relation. Instead, there are "blocks" of columns such that all columns within one block are proportional.

For any $p \in \mathbb{R}^{\mathcal{X}_0\times\mathcal{X}}$ let G_p be the subgraph of G induced by $\operatorname{supp}\tilde{p} := \{x \in \mathcal{X} : \tilde{p}_x \neq 0\}$. We have shown:

- A point p lies in V_G if and only if $\tilde{p}_x$ and $\tilde{p}_y$ are proportional whenever $x, y \in \operatorname{supp}\tilde{p}$ lie in the same connected component of G_p.

For any subset $\mathcal{Y} \subseteq \mathcal{X}$ denote $G_\mathcal{Y}$ the subgraph of G induced by $\mathcal{Y}$. Let $V_{G,\mathcal{Y}}$ be the set of all $p \in \mathbb{R}^{\mathcal{X}_0\times\mathcal{X}}$ for which $\tilde{p}_x = 0$ for all $x \in \mathcal{X} \setminus \mathcal{Y}$ and for which $\tilde{p}_x$ and $\tilde{p}_y$ are proportional whenever $x, y \in \mathcal{X}$ lie in the same connected component of $G_\mathcal{Y}$. Then

$$V_G = \cup_{\mathcal{Y}\subseteq\mathcal{X}} V_{G,\mathcal{Y}}. \tag{10}$$

The sets $V_{G,\mathcal{Y}}$ are irreducible algebraic varieties:

Lemma 21. *For any $\mathcal{Y} \subseteq \mathcal{X}$ the set $V_{G,\mathcal{Y}}$ is the variety of the ideal $I_{G,\mathcal{Y}}$ generated by the monomials*

$$p_{ix} \qquad \textit{for all } x \in \mathcal{X} \setminus \mathcal{Y} \textit{ and } i \in \mathcal{X}_0, \tag{11}$$

and the binomials f^{ij}_{xy} for all $i, j \in \mathcal{X}_0$ and all $x, y \in \mathcal{Y}$ that lie in the same connected component of $G_{\mathcal{Y}}$. The ideal $I_{G,\mathcal{Y}}$ is prime.

Proof. The first statement follows from the definition of $V_{G,\mathcal{Y}}$. Write $I^1_{G,\mathcal{Y}}$ for the ideal generated by all monomials (11), and for any $\mathcal{Z} \subseteq \mathcal{Y}$ write $I^2_{\mathcal{Z}}$ for the ideal generated by the binomials f^{ij}_{xy}, with $i, j \in \mathcal{X}_0$ and $x, y \in \mathcal{Z}$. Then $I^1_{G,\mathcal{Y}}$ is obviously prime. Each of the $I^2_{\mathcal{Z}}$ is a 2×2 determinantal ideal. It is a classical (but difficult) result that this ideal is the defining ideal of a Segre embedding, and that it is prime (see [10] for a rather modern proof). The ideal $I_{G,\mathcal{Y}}$ is the sum of the prime ideal $I^1_{G,\mathcal{Y}}$ and the prime ideals $I^2_{\mathcal{Z}}$ for all connected components $\mathcal{Z}$ of $G_{\mathcal{Y}}$, and since the defining equations of all these ideals involve disjoint sets of unknowns, $I_{G,\mathcal{Y}}$ itself is prime. □

The decomposition (10) is not the irreducible decomposition of V_G, because the union is redundant. Let $\mathcal{Y}, \mathcal{Z} \subseteq \mathcal{X}$. Using Lemma 21 it is easy to remove the redundant components:

Lemma 22. *Let $\mathcal{Y}, \mathcal{Z} \subseteq \mathcal{X}$. Then $V_{G,\mathcal{Y}}$ contains $V_{G,\mathcal{Z}}$ if and only if the following two conditions are satisfied:*

- *$\mathcal{Z} \subseteq \mathcal{Y}$.*
- *If $x, y \in \mathcal{Z}$ are connected in $G_{\mathcal{Y}}$, then they are connected in $G_{\mathcal{Z}}$.*

Proof. Assume that $V_{G,\mathcal{Y}} \subseteq V_{G,\mathcal{Z}}$. Then $I_{G,\mathcal{Y}} \supseteq I_{G,\mathcal{Z}}$. For any $x \in \mathcal{X} \setminus \mathcal{Z}$ and any $i \in \mathcal{X}_0$ this implies $p_{ix} \in I_{G,\mathcal{Y}}$. On the other hand, Lemma 21 shows that the point with coordinates

$$p_{iy} = \begin{cases} 1, & \text{if } y \in \mathcal{Y}, \\ 0, & \text{else,} \end{cases}$$

lies in $V_{G,\mathcal{Y}}$, and hence in $V_{G,\mathcal{Z}}$. This implies $x \in \mathcal{Y}$.

Let $x \in \mathcal{Z}$. Choose two linearly independent non-zero vectors $v, w \in \mathbb{R}^{d_0}$. By Lemma 21 the matrix with columns

$$\tilde{p}_y = \begin{cases} v, & \text{if } y \text{ is connected to } x \text{ in } G_{\mathcal{Y}}, \\ w, & \text{if } y \in \mathcal{Y} \text{ is not connected to } x \text{ in } G_{\mathcal{Y}}, \\ 0, & \text{else,} \end{cases}$$

is contained in $V_{G,\mathcal{Y}}$ and hence in $V_{G,\mathcal{Z}}$. Therefore, if z is connected to x in $G_{\mathcal{Y}}$, then it is connected to x in $G_{\mathcal{Z}}$.

Conversely, if the two conditions are satisfied, then all defining equations of $I_{G,\mathcal{Z}}$ lie in $I_{G,\mathcal{Y}}$. □

Theorem 23. *The primary decomposition of V_G is*

$$I_G = \cap_{\mathcal{Y}} I_{G,\mathcal{Y}},$$

where the intersection is over all $\mathcal{Y} \subseteq \mathcal{X}$ such that the following holds: For any $x \in \mathcal{X} \setminus \mathcal{Y}$ there are edges (x, y), (x, z) in G such that $y, z \in \mathcal{Y}$ are not connected in $G_{\mathcal{Y}}$.

Equivalently, for any $x \in \mathcal{X} \setminus \mathcal{Y}$ the induced subgraph $G_{\mathcal{Y}\cup\{x\}}$ has less connected components than $G_{\mathcal{Y}}$.

Proof. First, assume that $\mathbb{R}$ is algebraically closed. By (10) and Lemma 21 it suffices to show that the condition on $\mathcal{Y}$ stated in the theorem characterizes the maximal sets $V_{G,\mathcal{Y}}$ in the union (10) (with respect to inclusion). This follows from Lemma 22.

If $\mathbb{R}$ is not algebraically closed, then one can argue as follows: By [4] a binomial ideal has a binomial primary decomposition over some extension field $\hat{\mathbb{R}} = \mathbb{R}[\alpha_1, \dots, \alpha_k]$. The algebraic numbers $\alpha_1, \dots, \alpha_k$ are coefficients of the defining equations of the primary components. Let $\mathbb{C}$ be the algebraic closure of $\mathbb{R}$. Since the ideals $I_{G,\mathcal{Y}}$ are defined by pure differences and since the ideals $\mathbb{C} \otimes I_{G,\mathcal{Y}}$ are the primary components of $\mathbb{C} \otimes I_{G,\mathcal{Y}}$ in $\mathbb{C} \otimes R$ it follows that the ideals $I_{G,\mathcal{Y}}$ are already the primary components of I_G (in other words, the primary decomposition is independent of the base field). □

Remark (Comparison to [6]). Theorems 18 and 23 are generalizations of Theorems 2.1 and 3.2 from [6]. While Theorem 2.1 in [6] was proved with a case by case analysis, the proof of Theorem 18 is much more conceptual. The proof of Theorem 23 relied on the irreducible decomposition of the corresponding variety. On the other hand, the proof of Theorem 3.2 in [6] directly proves the equality of the two ideals.

Acknowledgement. This work has been supported by the Volkswagen Foundation and the Santa Fe Institute. Nihat Ay thanks David Krakauer and Jessica Flack for many stimulating discussions on robustness.

References

[1] Ay, N., and Krakauer, D. C., "Geometric robustness theory and biological networks," *Theory in Biosciences*, vol. 125, no. 2, pp. 93 – 121, 2007.

[2] Cox, D. A., Little, J., and O'Shea, D., *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd ed. Springer, 2008.

[3] Drton, M., Sturmfels, B., and Sullivant, S., *Lectures on Algebraic Statistics*, 1st ed., ser. Oberwolfach Seminars. Birkhäuser, Basel, 2009, vol. 39.

[4] Eisenbud, D., and Sturmfels, B., "Binomial ideals," *Duke Mathematical Journal*, vol. 84, no. 1, pp. 1–45, 1996.

[5] Fink, A., "The binomial ideal of the intersection axiom for conditional probabilities," *Journal of Algebraic Combinatorics*, vol. 33, no. 3, pp. 455–463, 2011.

[6] Herzog, J., Hibi, T., Hreinsdóttir, F., Kahle, T., and Rauh, J., "Binomial edge ideals and conditional independence statements," *Advances in Applied Mathematics*, vol. 45, no. 3, pp. 317 – 333, 2010.

[7] Jarrah, A. S., Raposa, B., and Laubenbacher, R., "Nested canalyzing, unate cascade, and polynomial functions," *Physica D*, vol. 233, no. 2, pp. 167 – 174, 2007.

[8] Kauffman, S. A., *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[9] Ohtani, M., "Graphs and ideals generated by some 2-minors." *Commun. Algebra*, vol. 39, no. 3, pp. 905–917, 2011.

[10] Sturmfels, B., "Gröbner bases of toric varieties," *Tohoku Mathematical Journal*, vol. 43, pp. 249 – 261, 1991.

[11] Swanson, I., and Taylor, A., "Minimal primes of ideals arising from conditional independence statements," *Preprint: arXiv:1107.5604v3*, 2011.

E-mail address: {rauh, nay}@mis.mpg.de

[1]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig,Germany

[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA