Journal of Machine Learning Research 10 (2009) 1071-1094 Submitted 12/06; Revised 5/08; Published 5/09

# An Algorithm for Reading Dependencies from the Minimal Undirected Independence Map of a Graphoid that Satisfies Weak Transitivity

**Jose M. Peña** JOSPE@IDA.LIU.SE
*Department of Computer and Information Science*
*Linköping University*
*SE-58183 Linköping, Sweden*

**Roland Nilsson** NILSSON@CHGR.MGH.HARVARD.EDU
*Department of Systems Biology*
*Harvard Medical School*
*Boston, MA 02115, USA*

**Johan Björkegren** JOHAN.BJORKEGREN@KI.SE
*Department of Medicine*
*Karolinska Institutet*
*SE-17176 Stockholm, Sweden*

**Jesper Tegnér**[*] JESPERT@IFM.LIU.SE
*Department of Physics, Chemistry and Biology*
*Linköping University*
*SE-58183 Linköping, Sweden*

**Editor:** Marina Meila

## Abstract

We present a sound and complete graphical criterion for reading dependencies from the minimal undirected independence map $G$ of a graphoid $M$ that satisfies weak transitivity. Here, complete means that it is able to read all the dependencies in $M$ that can be derived by applying the graphoid properties and weak transitivity to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by vertex separation. We argue that assuming weak transitivity is not too restrictive. As an intermediate step in the derivation of the graphical criterion, we prove that for any undirected graph $G$ there exists a strictly positive discrete probability distribution with the prescribed sample spaces that is faithful to $G$. We also report an algorithm that implements the graphical criterion and whose running time is considered to be at most $O(n^2(e+n))$ for $n$ nodes and $e$ edges. Finally, we illustrate how the graphical criterion can be used within bioinformatics to identify biologically meaningful gene dependencies.

**Keywords:** graphical models, vertex separation, graphoids, weak transitivity, bioinformatics

## 1. Introduction

A minimal undirected independence map $G$ of an independence model $M$ is used to read independencies that hold in $M$. Sometimes, however, $G$ can also be used to read dependencies holding in $M$. For instance, if $M$ is a graphoid that is faithful to $G$ then, by definition, lack of vertex separation is a sound and complete graphical criterion for reading dependencies from $G$, where complete means

---

[*] Also at the Department of Medicine, Karolinska Institutet, SE-17176 Stockholm, Sweden.

that it is able to read all the dependencies in $M$. If $M$ is simply a graphoid, then Bouckaert (1995) proposes a sound and complete graphical criterion for reading dependencies from $G$. In this case, complete means that it is able to read all the dependencies in $M$ that can be derived by applying the graphoid properties to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by vertex separation.

In this paper, we introduce a sound and complete graphical criterion for reading dependencies from $G$ under the assumption that $M$ is a graphoid that satisfies weak transitivity. Here, complete means that it is able to read all the dependencies in $M$ that can be derived by applying the graphoid properties and weak transitivity to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by vertex separation. Our criterion allows reading more dependencies than the criterion in Bouckaert (1995) when the graphoid at hand satisfies weak transitivity. We show that there exist important families of probability distributions that are graphoids and satisfy weak transitivity. These include, for instance, the regular Gaussian probability distributions.

We think that the work presented in this paper can be of great interest for the machine learning community. Graphs are one of the most commonly used metaphors for representing knowledge because they appeal to human intuition (Pearl, 1988). Furthermore, graphs are parsimonious models because they trade off accuracy for simplicity. Consider, for instance, representing the independence model induced by a probability distribution as a graph. Though this graph is typically less accurate than the probability distribution (the graph may not represent all the (in)dependencies and those that are represented are not quantified), it also requires less space to be stored and less time to be communicated than the probability distribution, which may be desirable features in some applications. Thus, it seems sensible developing tools for reasoning with graphs. Our criterion is one such a tool: As vertex separation makes the discovery of independencies amenable to human reasoning by enabling to read independencies off $G$ without numerical calculation (Pearl, 1988), so does our criterion with respect to the discovery of dependencies. There are fields where discovering dependencies is more important than discovering independencies. It is in these fields where we believe that our criterion has greater potential. In bioinformatics, for instance, the nodes of $G$ represent (the expression levels of) some genes under study. Bioinformaticians are typically more interested in discovering gene dependencies than independencies, because the former provide contexts in which the expression level of some genes is informative about that of some other genes, which may lead to hypothesize dependencies, functional relations, causal relations, the effects of manipulation experiments, etc. As we will illustrate at the end of the paper, our criterion can be very helpful in such a scenario. Our criterion also clarifies a misconception that may exist among some bioinformaticians, namely that two genes are dependent if there exists a path in $G$ between them. We will see that there must exist exactly one path to draw such a conclusion. Hence, the importance of developing a formal criterion like ours to prevent drawing erroneous conclusions. Of course, the conclusions drawn by our criterion may be misleading if $G$ is learnt from a sample, which is most likely the case in bioinformatics. However, this has nothing to do with the correctness of our criterion, which we prove in subsequent sections, but with the fact that $G$ is an estimate.

The rest of the paper is organized as follows. We start by reviewing some concepts in Section 2. We show in Section 3 that assuming weak transitivity is not too restrictive. We prove in Section 4 that for any undirected graph $G$ there exists a strictly positive discrete probability distribution with the prescribed sample spaces that is faithful to $G$. This is an important result in itself as well as for proving the completeness of the graphical criterion that we present in Section 5. An algorithmic implementation of the graphical criterion is described in Section 6. We illustrate in Section 7 how the graphical criterion works in practice with a real world example taken from bioinformatics. Finally, we close with some discussion in Section 8.

## 2. Preliminaries

The definitions and results in this section are taken from Lauritzen (1996), Pearl (1988) and Studený (2005). We use the juxtaposition $\mathbf{XY}$ to denote $\mathbf{X} \cup \mathbf{Y}$, and $X$ to denote the singleton $\{X\}$. We use upper-case letters to denote random variables and the same letters in lower-case to denote their states. We use $Val(\mathbf{X})$ to denote the set of possible states of a random variable $\mathbf{X}$. Let $\mathbf{U}$ denote a set of random variables. Unless otherwise stated, all the probability distributions, independence models and graphs in this paper are defined over $\mathbf{U}$. Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$ denote four mutually disjoint subsets of $\mathbf{U}$. An independence model $M$ is a set of independencies of the form $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$. We represent that an independence is in $M$ by $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ and that an independence is not in $M$ by $\mathbf{X} \not\!\perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$. In the latter case, we may equivalently say that the dependence $\mathbf{X} \not\!\perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ is in $M$. An independence model is a graphoid when it satisfies the following five properties:

- Symmetry $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \Rightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}$.
- Decomposition $\mathbf{X} \perp\!\!\!\perp \mathbf{YW} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$.
- Weak union $\mathbf{X} \perp\!\!\!\perp \mathbf{YW} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{ZW}$.
- Contraction $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{ZW} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{YW} | \mathbf{Z}$.
- Intersection $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{ZW} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} | \mathbf{ZY} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{YW} | \mathbf{Z}$.

The independence model induced by any strictly positive probability distribution is a graphoid. Hereinafter, for the sake of simplicity, we do not make any distinction between a probability distribution and the independence model induced by it and, thus, we always refer to the former. For instance, instead of saying that the independence model of a probability distribution $p$ is a graphoid, we simply say that $p$ is a graphoid. In this paper, we pay particular attention to strictly positive discrete probability distributions and regular Gaussian probability distributions, that is, those whose covariance matrices are positive definite. For the strictly positive discrete probability distributions, we assume that each random variable in $\mathbf{U}$ has a finite sample space with at least two possible states. Note that if $p$ is a strictly positive discrete probability distribution, then $p(\mathbf{X})$ and $p(\mathbf{X}|\mathbf{Y}=\mathbf{y})$ are uniquely defined, strictly positive and discrete. Likewise, if $p$ is a regular Gaussian probability distribution, then $p(\mathbf{X})$ and $p(\mathbf{X}|\mathbf{Y}=\mathbf{y})$ are uniquely determined (by a convention) and regular Gaussian. Moreover, any regular Gaussian probability distribution satisfies composition $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} | \mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{YW} | \mathbf{Z}$.

A path between $X_1$ and $X_n$ in a graph $G$ is a sequence of distinct nodes $X_1, \ldots, X_n$ $(1 \leq n)$ such that there exists an edge in $G$ between every two consecutive nodes in the sequence. Given a path $X_1, \ldots, X_n$ in a directed and acyclic graph (DAG) $G$, a node $X_i$ $(1 < i < n)$ is a collider in the path if $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ in $G$. Let $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ denote that $\mathbf{X}$ is separated from $\mathbf{Y}$ given $\mathbf{Z}$ in a graph $G$. Specifically, $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ holds when every path in $G$ between $\mathbf{X}$ and $\mathbf{Y}$ is blocked by $\mathbf{Z}$. If $G$ is an undirected graph (UG), then a path in $G$ between $\mathbf{X}$ and $\mathbf{Y}$ is blocked by $\mathbf{Z}$ when there exists some $Z \in \mathbf{Z}$ in the path. If $G$ is a DAG, then a path in $G$ between $\mathbf{X}$ and $\mathbf{Y}$ is blocked by $\mathbf{Z}$ when there exists a node $Z$ in the path such that

- either $Z$ is not a collider in the path and $Z \in \mathbf{Z}$, or
- $Z$ is a collider in the path and neither $Z$ nor any of its descendants in $G$ is in $\mathbf{Z}$.

An independence model $M$ is faithful to an UG or DAG $G$ when $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ iff $sep(\mathbf{X}, \mathbf{Y} | \mathbf{Z})$. Any independence model that is faithful to some UG or DAG is a graphoid. An UG (resp. DAG) $G$ is an undirected (resp. directed) independence map of an independence model $M$ when $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ if $sep(\mathbf{X}, \mathbf{Y} | \mathbf{Z})$. Moreover, an UG $G$ is a minimal undirected independence (MUI) map of $M$ when

(i) $G$ is an undirected independence map of $M$, and

(ii) no proper subgraph of $G$ satisfies (i).

A Markov boundary of $X \in \mathbf{U}$ in an independence model $M$ is any subset $MB(X)$ of $\mathbf{U} \setminus X$ such that

(i) $X \perp\!\!\!\perp \mathbf{U} \setminus X \setminus MB(X) | MB(X)$, and

(ii) no proper subset of $MB(X)$ satisfies (i).

If $M$ is a graphoid, then

(i) $MB(X)$ is unique for each $X \in \mathbf{U}$,

(ii) the MUI map $G$ of $M$ is unique, and

(iii) two nodes $X$ and $Y$ are adjacent in $G$ iff $X \in MB(Y)$ iff $Y \in MB(X)$ iff $X \not\perp\!\!\!\perp Y | \mathbf{U} \setminus (XY)$.

A Bayesian network (BN) is a pair $(G, p)$ where $G$ is a DAG and $p$ is a discrete probability distribution that factorizes as $p = \prod_{X \in \mathbf{U}} q(X | Pa(X))$, where $q(X | Pa(X))$ denotes a conditional discrete probability distribution of $X$ given the parents of $X$ in $G$, $Pa(X)$. Recall that a node $Y$ is called a parent of $X$ if $Y \to X$ is in $G$. We denote by $\mathcal{D}(G)^+$ all the strictly positive discrete probability distributions that can be represented by a BN with DAG $G$, that is, those that factorize according to $G$ as indicated.

## 3. Weak Transitivity Graphoids

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. We define a weak transitivity (WT) graphoid as a graphoid that satisfies weak transitivity $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}V \Rightarrow \mathbf{X} \perp\!\!\!\perp V | \mathbf{Z} \vee V \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ with $V \in \mathbf{U} \setminus (\mathbf{XYZ})$. There exist important families of probability distributions that are WT graphoids and, thus, we believe that WT graphoids are worth studying: For instance, any probability distribution that is regular Gaussian or faithful to some UG or DAG is a WT graphoid (Pearl, 1988; Studený, 2005). Other interesting families of probability distributions that are WT graphoids are presented in this section.

We say that a strictly positive discrete probability distribution $p$ has context-specific dependencies if there exists some $\mathbf{W} \subseteq \mathbf{U}$ such that $p(\mathbf{U} \setminus \mathbf{W} | \mathbf{W} = \mathbf{w})$ does not have the same (in)dependencies for all $\mathbf{w}$. The theorem below proves that for any DAG $G$, in a measure-theoretic sense (Halmos, 1966), almost all the probability distributions in $\mathcal{D}(G)^+$ have no context-specific dependencies. This result is not only relative to the DAG $G$, but also to the measure considered, the

Lebesgue measure in our case, as well as to the dimension of $\mathcal{D}(G)^+$. For this purpose, we discuss first how we parameterize the probability distributions in $\mathcal{D}(G)^+$. Since each probability distribution $p$ in $\mathcal{D}(G)^+$ factorizes as $p = \prod_{X \in \mathbf{U}} q(X|Pa(X))$, $p$ can be parameterized by parameterizing each probability table $q(X|Pa(X))$. Let $Val(X) = \{x_1, \ldots, x_{n(X)}\}$ denote the $n(X)$ possible states of the random variable $X$. Let $\theta_{x_i,\mathbf{pa}}$ $(1 \leq i \leq n(X))$ denote the parameter corresponding to $q(X = x_i|Pa(X) = \mathbf{pa})$ with $\mathbf{pa} \in Val(Pa(X))$. Note that the parameters are linearly dependent because $\sum_{i=1}^{n(X)} q(X = x_i|Pa(X) = \mathbf{pa}) = 1$. In order to introduce properly the parameter space for $\mathcal{D}(G)^+$ and the Lebesgue measure on it, we make the convention that $\theta_{x_{n(X)},\mathbf{pa}}$ is linearly dependent on the remaining parameters $\theta_{x_i,\mathbf{pa}}$ with $1 \leq i < n(X)$. Therefore, the number of linearly independent parameters for $p$ is $n = \sum_{X \in \mathbf{U}} (n(X) - 1)(\prod_{Y \in Pa(X)} n(Y))$. Let $\Delta_d$ denote the simplex $\{(s_1, \ldots, s_d) \in \mathbb{R}^d : s_i \geq 0 \; (1 \leq i \leq d), \sum_{i=1}^{d} s_i \leq 1\}$. The Lebesgue measure of $\Delta_d$ wrt $\mathbb{R}^d$ is $1/d!$ (Stein, 1966). Let $\Delta_d^+$ denote the set $\{(s_1, \ldots, s_d) \in \mathbb{R}^d : s_i > 0 \; (1 \leq i \leq d), \sum_{i=1}^{d} s_i < 1\}$. The Lebesgue measure of $\Delta_d^+$ wrt $\mathbb{R}^d$ is also $1/d!$, because the difference between $\Delta_d$ and $\Delta_d^+$ has Lebesgue measure zero. Then, the parameter space for $\mathcal{D}(G)^+$ is $\times_{X \in \mathbf{U}} \times_{\mathbf{pa} \in Val(Pa(X))} \Delta^+_{n(X)-1}$, whose Lebesgue measure wrt $\mathbb{R}^n$ is $\prod_{X \in \mathbf{U}} \prod_{\mathbf{pa} \in Val(Pa(X))} 1/(n(X) - 1)!$.

**Theorem 1** *Let G be a DAG. $\mathcal{D}(G)^+$ has non-zero Lebesgue measure wrt $\mathbb{R}^n$, where n is the number of linearly independent parameters in the parametrization of the probability distributions in $\mathcal{D}(G)^+$ described above. The probability distributions in $\mathcal{D}(G)^+$ that are not faithful to G or have context-specific dependencies have zero Lebesgue measure wrt $\mathbb{R}^n$.*

**Proof** First, we prove that there is a one-to-one correspondence between the elements of the parameter space for $\mathcal{D}(G)^+$ and the probability distributions in $\mathcal{D}(G)^+$. This will allow us later to compute the Lebesgue measure wrt $\mathbb{R}^n$ of a subset of $\mathcal{D}(G)^+$ as the Lebesgue measure wrt $\mathbb{R}^n$ of the corresponding subset of the parameter space for $\mathcal{D}(G)^+$. Obviously, different probability distributions in $\mathcal{D}(G)^+$ must correspond to different elements of the parameter space for $\mathcal{D}(G)^+$. On the other hand, different elements of the parameter space for $\mathcal{D}(G)^+$ correspond to different probability distributions in $\mathcal{D}(G)^+$, because the values of the parameters $\theta_{x_i,\mathbf{pa}}$ defining a probability distribution $p$ coincide with the values of the conditional probabilities $p(X = x_i|Pa(X) = \mathbf{pa})$ that are computed from $p$ (Pearl, 1988).

To prove the first statement in the theorem it suffices to note that, as discussed right before the theorem, the parameter space for $\mathcal{D}(G)^+$ has non-zero Lebesgue measure wrt $\mathbb{R}^n$. This implies that $\mathcal{D}(G)^+$ also has non-zero Lebesgue measure wrt $\mathbb{R}^n$ because, as proven above, there is a one-to-one correspondence between the elements of the parameter space for $\mathcal{D}(G)^+$ and the probability distributions in $\mathcal{D}(G)^+$.

To prove the second statement in the theorem, we first prove that the probability distributions in $\mathcal{D}(G)^+$ that have context-specific dependencies have zero Lebesgue measure wrt $\mathbb{R}^n$. The proof basically proceeds in the same way as that of Theorem 7 in Meek (1995). We start by showing that for a probability distribution in $\mathcal{D}(G)^+$ to have context-specific dependencies, some polynomials in the parameters in the parametrization of the probability distributions in $\mathcal{D}(G)^+$ must be satisfied. Specifically, these polynomials are real polynomials in real variables that we interpret as real functions on a real Euclidean space that includes the parameter space for $\mathcal{D}(G)^+$. Let $\mathbf{W} \subseteq \mathbf{U}$ and let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three disjoint subsets of $\mathbf{U} \setminus \mathbf{W}$. Since $G$ is a directed independence map of any probability distribution $p \in \mathcal{D}(G)^+$ (Neapolitan, 2003), for a constraint such as $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ to be true in $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ but false in $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w}')$, two conditions must be met. First, $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z}\mathbf{W})$

must not hold in $G$ and, second, the following equations must be satisfied:

$$p(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y},\mathbf{Z}=\mathbf{z},\mathbf{W}=\mathbf{w})p(\mathbf{Z}=\mathbf{z},\mathbf{W}=\mathbf{w})-$$

$$p(\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z},\mathbf{W}=\mathbf{w})p(\mathbf{Y}=\mathbf{y},\mathbf{Z}=\mathbf{z},\mathbf{W}=\mathbf{w})=0 \qquad (1)$$

for all $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Each equation is a polynomial in the parameters in the parametrization of the probability distributions in $\mathcal{D}(G)^+$, because each term $p(\mathbf{V}=\mathbf{v})$ in the equations is a polynomial in the parameters: $p(\mathbf{V}=\mathbf{v})=\sum_{\mathbf{v}'} p(\mathbf{V}=\mathbf{v},\mathbf{U}\setminus\mathbf{V}=\mathbf{v}')$ where each term $p(\mathbf{V}=\mathbf{v},\mathbf{U}\setminus\mathbf{V}=\mathbf{v}')$ is a polynomial in the parameters, since $p=\prod_{X\in\mathbf{U}} q(X|Pa(X))$. Let each variable in the polynomials take values in $\mathbb{R}$. Then, each polynomial in Equation (1) is non-trivial, that is, not all the values of the variables are solutions to the polynomial. To prove this, it suffices to prove that there exists a probability distribution $p'\in\mathcal{D}(G)^+$ for which the polynomial does not hold. Consider the polynomial for $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Note that there exists a probability distribution $p''\in\mathcal{D}(G)^+$ that is faithful to $G$ (Meek, 1995) and, thus, $\mathbf{X}\not\perp\mathbf{Y}|\mathbf{ZW}$ is in $p''$ because $sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$ does not hold in $G$. Then, there is some instantiation $\mathbf{x}''\mathbf{y}''\mathbf{z}''\mathbf{w}''$ of $\mathbf{XYZW}$ such that

$$p''(\mathbf{X}=\mathbf{x}'',\mathbf{Y}=\mathbf{y}'',\mathbf{Z}=\mathbf{z}'',\mathbf{W}=\mathbf{w}'')p''(\mathbf{Z}=\mathbf{z}'',\mathbf{W}=\mathbf{w}'')-$$

$$p''(\mathbf{X}=\mathbf{x}'',\mathbf{Z}=\mathbf{z}'',\mathbf{W}=\mathbf{w}'')p''(\mathbf{Y}=\mathbf{y}'',\mathbf{Z}=\mathbf{z}'',\mathbf{W}=\mathbf{w}'')\neq 0.$$

Then, by permuting the states of the random variables in $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$, we can transform $p''$ into the desired $p'$. Let $\mathbf{V}\equiv\mathbf{XYZW}$, $\mathbf{v}\equiv\mathbf{xyzw}$, and $\mathbf{v}''\equiv\mathbf{x}''\mathbf{y}''\mathbf{z}''\mathbf{w}''$. One can introduce a permutation $\pi_X$ on the set of possible states of $X$ for each $X\in\mathbf{U}$: For $X\in\mathbf{V}$, it is the transposition of the states of $X$ in $\mathbf{v}$ and $\mathbf{v}''$, and the identical mapping for $X\in\mathbf{U}\setminus\mathbf{V}$. These random variable permutations together define a permutation $\pi$ of the joint sample space of $\mathbf{U}$. Then, $p''$ can be transformed by $\pi$ to $p'\equiv p''\circ\pi$. Note that $p'\in\mathcal{D}(G)^+$ because $p''\in\mathcal{D}(G)^+$, and that the parameter values of $p'$ are obtained from the parameter values of $p''$ by local permutations. Finally, note that $p'(\mathbf{V}=\mathbf{v})=p''(\mathbf{V}=\mathbf{v}'')$ and, thus, the polynomial in Equation (1) for $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ does not hold for $p'$.

Let $sol(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w})$ denote the set of solutions to the polynomial in Equation (1) for $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Then, $sol(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w})$ has zero Lebesgue measure wrt $\mathbb{R}^n$ because it consists of the solutions to a non-trivial polynomial in real variables (Okamoto, 1973). Let $sol=\bigcup_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}}\bigcup_{\mathbf{w}}\bigcap_{\mathbf{x},\mathbf{y},\mathbf{z}} sol(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w})$ and recall from above that the outer-most union only involves those cases for which $sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$ does not hold in $G$. Then, $sol$ has zero Lebesgue measure wrt $\mathbb{R}^n$, because the finite union and intersection of sets of zero Lebesgue measure has zero Lebesgue measure too. Consequently, the probability distributions in $\mathcal{D}(G)^+$ that have context-specific dependencies correspond to a set of elements of the parameter space for $\mathcal{D}(G)^+$ that has zero Lebesgue measure wrt $\mathbb{R}^n$ because it is contained in $sol$. Since, as proven above, this correspondence is one-to-one, the probability distributions in $\mathcal{D}(G)^+$ that have context-specific dependencies also have zero Lebesgue measure wrt $\mathbb{R}^n$.

To finish the proof of the second statement, it suffices to note that (i) the probability distributions in $\mathcal{D}(G)^+$ that are not faithful to $G$ have zero Lebesgue measure wrt $\mathbb{R}^n$, because they are a subset of the probability distributions that factorize according to $G$ but are not faithful to $G$, and these have zero Lebesgue measure wrt $\mathbb{R}^n$ (Meek, 1995), (ii) the probability distributions in $\mathcal{D}(G)^+$ that have context-specific dependencies have zero Lebesgue measure wrt $\mathbb{R}^n$ as proven above, and (iii) the union of the probability distributions in (i) and (ii) has zero Lebesgue measure wrt $\mathbb{R}^n$, because the finite union of sets of zero Lebesgue measure has zero Lebesgue measure too. ■

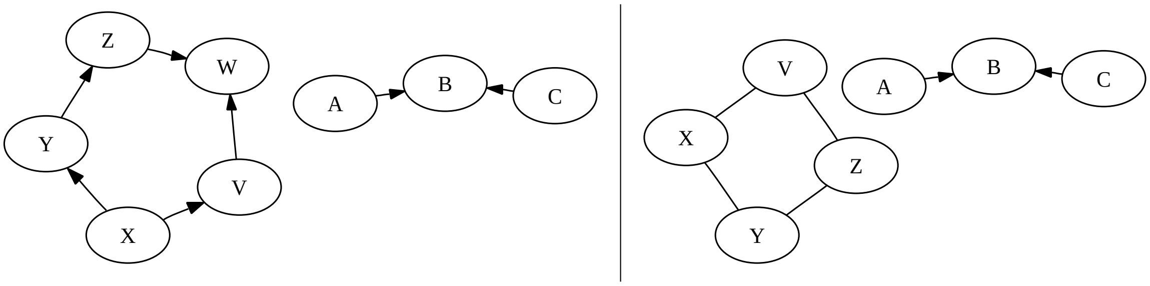


Figure 1: Chain graphs in Example 1.

The theorem below proves that the marginals and conditionals of a strictly positive discrete probability distribution that is a WT graphoid and has no context-specific dependencies are WT graphoids too.

**Theorem 2** *Let p be a strictly positive discrete probability distribution that is a WT graphoid. Then,* $p(\mathbf{U}\setminus\mathbf{W})$ *for any* $\mathbf{W}\subseteq\mathbf{U}$ *is a WT graphoid. If p has no context-specific dependencies, then* $p(\mathbf{U}\setminus\mathbf{W}|\mathbf{W}=\mathbf{w})$ *for any* $\mathbf{w}$ *and* $\mathbf{W}\subseteq\mathbf{U}$ *is a WT graphoid.*

**Proof** Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}\setminus\mathbf{W}$. Then, $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$ is in $p(\mathbf{U}\setminus\mathbf{W})$ iff $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$ is in $p$ and, thus, $p(\mathbf{U}\setminus\mathbf{W})$ satisfies the WT graphoid properties because $p$ satisfies them. Now, note that $p(\mathbf{U}\setminus\mathbf{W}|\mathbf{W}=\mathbf{w})$ is uniquely defined because $p$ is strictly positive. Furthermore, if $p$ has no context-specific dependencies, then $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}$ is in $p(\mathbf{U}\setminus\mathbf{W}|\mathbf{W}=\mathbf{w})$ iff $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{ZW}$ is in $p$. Then, $p(\mathbf{U}\setminus\mathbf{W}|\mathbf{W}=\mathbf{w})$ satisfies the WT graphoid properties because $p$ satisfies them. ■

In a nutshell, Theorem 1 proves that for any DAG $G$, in the measure-theoretic sense explained above, almost all the probability distributions in $\mathcal{D}(G)^+$ are faithful to $G$ and, thus, are WT graphoids (Pearl, 1988). On the other hand, the combination of Theorems 1 and 2 proves that, in the measure-theoretic sense explained above, all the marginals and conditionals of almost all the probability distributions in $\mathcal{D}(G)^+$ are WT graphoids. Finally, we give an example that shows that not all the probability distributions that are WT graphoids are either regular Gaussian or faithful to some UG or DAG.

**Example 1** *Let p be a strictly positive discrete probability distribution that is faithful to the DAG in the left-hand side of Figure 1 and that has no context-specific dependencies. Such a probability distribution exists due to Theorem 1 and, moreover, it is a WT graphoid (Pearl, 1988). Then,* $p(X,Y,Z,V,A,B,C|W=w)$ *for any w, which is uniquely defined because p is strictly positive, is a WT graphoid by Theorem 2. However, this conditional probability distribution is neither regular Gaussian nor faithful to any UG or DAG, because it is discrete and faithful to the chain graph in the right-hand side of Figure 1 (Chickering and Meek, 2002; Peña et al., 2006).*

## 4. Reading Independencies

By definition, *sep* is sound for reading independencies from the MUI map $G$ of a WT graphoid $M$, that is, it only identifies independencies in $M$. In the regular Gaussian case, *sep* in $G$ is also complete in the sense that it identifies all the independencies in $M$ that are shared by all the WT graphoids for which $G$ is the MUI map, because (i) it is proven in Lněnička and Matúš (2007) that there exist regular Gaussian probability distributions that are faithful to $G$, and (ii) such probability distributions are WT graphoids (Studený, 2005), $G$ is their MUI map, and they only have the independencies that *sep* identifies from $G$. In this section, we extend this result to the case where $\mathbf{U}$ is discrete. Specifically, we prove that there exist strictly positive discrete probability distributions that are faithful to $G$ for any sample spaces (with at least two possible states) of the random variables in $\mathbf{U}$. Again, such probability distributions are WT graphoids (Pearl, 1988), $G$ is their MUI map, and they only have the independencies that *sep* identifies from $G$. Therefore, when $\mathbf{U}$ is discrete *sep* in $G$ is also complete in the sense that it identifies all the independencies in $M$ that are shared by all the WT graphoids for which $G$ is the MUI map. These completeness results, in addition to being important in themselves, are crucial for reading as many dependencies as possible from $G$, as we will see in the next section.

**Theorem 3** *Let $G$ be an UG. Let us assume that the sample space of each random variable in* $\mathbf{U}$ *is prescribed and has at least two possible states. Then, there exists a strictly positive discrete probability distribution with the prescribed sample spaces for the random variables in* $\mathbf{U}$ *that is faithful to $G$.*

**Proof** Create an extended DAG $H$ of $G$ as follows. For each edge $X - Y$ in $G$, create an auxiliary discrete random variable $W_{XY}$. Let $\mathbf{W}$ denote all the auxiliary random variables created. Let $H$ be a DAG over $\mathbf{UW}$ with no edges. For each edge $X - Y$ in $G$ add $X \rightarrow W_{XY} \leftarrow Y$ to $H$. No more edges are added to $H$. It is easy to see that for any three mutually disjoint subsets $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ of $\mathbf{U}$, $sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$ in $H$ iff $sep(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ in $G$.

Let $p(\mathbf{U},\mathbf{W})$ denote any strictly positive discrete probability distribution in $\mathcal{D}(H)^+$ that is faithful to $H$ and has no context-specific dependencies. Such a probability distribution exists by Theorem 1. Now, fix any $\mathbf{w}$ and note that $p(\mathbf{U}|\mathbf{W}=\mathbf{w})$ is uniquely defined because $p(\mathbf{U},\mathbf{W})$ is strictly positive. Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. Then, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ is in $p(\mathbf{U}|\mathbf{W}=\mathbf{w})$ iff $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW}$ is in $p(\mathbf{U},\mathbf{W})$ iff $sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$ in $H$ iff $sep(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ in $G$. Then, $p(\mathbf{U}|\mathbf{W}=\mathbf{w})$ is faithful to $G$. Obviously, $p(\mathbf{U}|\mathbf{W}=\mathbf{w})$ is strictly positive and discrete. ■

Note that the theorem above proves that there exists a strictly positive discrete probability distribution that is faithful to $G$ for any sample spaces (with at least two possible states) of the random variables in $\mathbf{U}$. This result is therefore stronger than Theorem 11 in Geiger and Pearl (1993), which proves that there exists a strictly positive discrete probability distribution that is faithful to $G$ for some sample spaces (with at least two possible states) of the random variables in $\mathbf{U}$.

The theorem above proves that, when $\mathbf{U}$ is discrete, *sep* in the MUI map $G$ of a WT graphoid $M$ is complete for any sample spaces of the random variables in $\mathbf{U}$, where complete means that it is able to identify all the independencies in $M$ that are shared by all the WT graphoids for which $G$ is the MUI map. However, *sep* in $G$ is not complete if this is understood as being able to identify all the independencies in $M$. Actually, no sound criterion for reading independencies from $G$ is complete in the latter sense. An example follows.

**Example 2** *Let $p$ be a discrete probability distribution that is faithful to the DAG $\{X \rightarrow Z, Y \rightarrow Z\}$. Such a probability distribution exists (Meek, 1995). Let $G$ denote the MUI map of $p$, namely the* *complete UG. Note that $p$ is not faithful to $G$. However, by Theorem 3, there exists a discrete probability distribution $p'$ that is faithful to $G$. As proven in Pearl (1988), $p$ and $p'$ are WT graphoids. Let us assume that we are dealing with $p$. Then, no sound criterion can conclude $X \perp\!\!\!\perp Y|\emptyset$ by just studying $G$ because this independence does not hold in $p'$, and it is impossible to know whether we are dealing with $p$ or $p'$ on the sole basis of $G$.*

## 5. Reading Dependencies

In this section, we propose a sound and complete criterion for reading dependencies from the MUI map $G$ of a WT graphoid. Here, complete means that it is able to read all the dependencies that can be derived by applying the WT graphoid properties to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by *sep*.

We define the dependence base of an independence model $M$, denoted $bas(M)$, as the set of dependencies $X \not\perp\!\!\!\perp Y|\mathbf{U}\setminus(XY)$ with $X,Y \in \mathbf{U}$. Recall from Section 2 that $X$ and $Y$ are adjacent in the MUI map of $M$ iff $X \not\perp\!\!\!\perp Y|\mathbf{U}\setminus(XY)$. If $M$ is a WT graphoid, then additional dependencies in $M$ can be derived from $bas(M)$ via the WT graphoid properties. For this purpose, we rephrase the WT graphoid properties as follows. Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$ denote four mutually disjoint subsets of $\mathbf{U}$. Symmetry $\mathbf{Y} \not\perp\!\!\!\perp \mathbf{X}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$. Decomposition $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z}$. Weak union $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z}$. Contraction $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \vee \mathbf{X} \not\perp\!\!\!\perp \mathbf{W}|\mathbf{Z}$ is problematic for deriving new dependencies because it contains a disjunction in the right-hand side and, thus, it should be split into two properties: Contraction1 $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{W}|\mathbf{Z}$, and contraction2 $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW}$. Likewise, intersection $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \vee \mathbf{X} \not\perp\!\!\!\perp \mathbf{W}|\mathbf{ZY}$ gives rise to intersection1 $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{W}|\mathbf{ZY}$, and intersection2 $\mathbf{X} \not\perp\!\!\!\perp \mathbf{YW}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W}|\mathbf{ZY} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW}$. Note that intersection1 and intersection2 are equivalent and, thus, we refer to them simply as intersection. Finally, weak transitivity $\mathbf{X} \not\perp\!\!\!\perp V|\mathbf{Z} \wedge V \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \vee \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}V$ with $V \in \mathbf{U}\setminus(\mathbf{XYZ})$ gives rise to weak transitivity1 $\mathbf{X} \not\perp\!\!\!\perp V|\mathbf{Z} \wedge V \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}V$, and weak transitivity2 $\mathbf{X} \not\perp\!\!\!\perp V|\mathbf{Z} \wedge V \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}V \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$. The independence in the left-hand side of any of the properties above holds if the corresponding *sep* statement holds in the MUI map $G$ of $M$. This is the best solution we can hope for because, as discussed in Section 4, *sep* in $G$ is sound and complete in the sense that it identifies all and only the independencies in $M$ that are shared by all the WT graphoids for which $G$ is the MUI map. Moreover, this solution does not require more information about $M$ than what it is available, because $G$ can be constructed from $bas(M)$. We denote by $sep(G)$ all the *sep* statements holding in the MUI map $G$. We define the WT graphoid closure of $bas(M)$ wrt $sep(G)$, denoted $WT_{bas(M)}^{sep(G)}$, as the set of dependencies in $bas(M)$ plus those that can be derived from it and $sep(G)$ by applying the WT graphoid properties. We now introduce our criterion for reading dependencies from the MUI map of a WT graphoid.

**Definition 4** *Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. Let $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ denote that $\mathbf{X}$ is connected to $\mathbf{Y}$ given $\mathbf{Z}$ in an UG $G$. Specifically, $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds when there exist some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ such that there exists **exactly one** path in $G$ between $X_1$ and $X_n$ that is not blocked by $(\mathbf{X}\setminus X_1)(\mathbf{Y}\setminus X_n)\mathbf{Z}$.*

As an illustrative example of *con*, consider the UG in Figure 2. Some *con* statements holding in that graph are $con(A,B|CD)$, $con(AC,BD|\emptyset)$, $con(A,BCD|\emptyset)$ and $con(A,BD|\emptyset)$ because in each of these statement there exists a single path between $A$ and $B$ or $D$ that is not blocked by the rest of

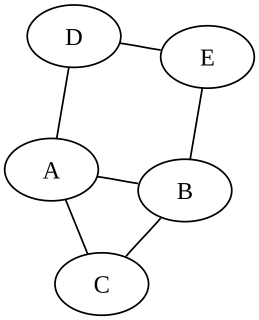


Figure 2: UG to illustrate the definition of *con*: $con(A,B|CD)$, $con(AC,BD|\emptyset)$, $con(A,BCD|\emptyset)$ and $con(A,BD|\emptyset)$ hold in the graph, whereas $con(A,B|\emptyset)$ and $con(A,B|D)$ do not hold.

the nodes in the statement. On the other hand, some *con* statements that do not hold in the graph under consideration are $con(A,B|\emptyset)$ and $con(A,B|D)$, because in both cases there exist several paths between $A$ and $B$ that are not blocked by the rest of the nodes in the statement. Note that the place the nodes take in the statement matters: For instance, $con(A,BD|\emptyset)$ holds but $con(A,B|D)$ does not.

We now prove that *con* is sound for reading dependencies from the MUI map $G$ of a WT graphoid $M$, that is, it only identifies dependencies in $M$. Actually, it only identifies dependencies in $WT^{sep(G)}_{bas(M)}$. Hereinafter, we abbreviate a path $X_1,\dots,X_n$ in an UG as $X_{1:n}$.

**Theorem 5** *Let M be a WT graphoid and G its MUI map. Then, con in G only identifies dependencies in* $WT^{sep(G)}_{bas(M)}$.

**Proof** We first prove that if $X_{1:n}$ is the only path in $G$ between $X_1$ and $X_n$ that is not blocked by $\mathbf{Y} \subseteq \mathbf{U} \setminus X_{1:n}$, then $X_1 \not\perp\!\!\!\perp X_n|\mathbf{Y}$ is in $WT^{sep(G)}_{bas(M)}$. In other words, we prove that $X_1 \not\perp\!\!\!\perp X_n|\mathbf{Y}$ can be derived from $bas(M)$ and $sep(G)$ using the WT graphoid properties. We prove it by induction over $n$. We first prove it for $n = 2$. Let $\mathbf{W}$ denote all the nodes in $\mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y}$ that are not separated from $X_1$ given $X_2\mathbf{Y}$ in $G$. Since $X_1$ and $X_2$ are adjacent in $G$, $X_1 \not\perp\!\!\!\perp X_2|\mathbf{U} \setminus X_{1:2}$ and, thus, $X_1\mathbf{W} \not\perp\!\!\!\perp X_2(\mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y} \setminus \mathbf{W})|\mathbf{Y}$ due to weak union and symmetry. This together with $sep(X_1\mathbf{W}, \mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y} \setminus \mathbf{W}|X_2\mathbf{Y})$, which follows from the definition of $\mathbf{W}$, implies $X_1\mathbf{W} \not\perp\!\!\!\perp X_2|\mathbf{Y}$ due to contraction1. Note that if $\mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y} \setminus \mathbf{W} = \emptyset$, then $X_1\mathbf{W} \not\perp\!\!\!\perp X_2(\mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y} \setminus \mathbf{W})|\mathbf{Y}$ directly implies $X_1\mathbf{W} \not\perp\!\!\!\perp X_2|\mathbf{Y}$. In any case, this dependence together with $sep(\mathbf{W}, X_2|X_1\mathbf{Y})$, because otherwise there exist several unblocked paths in $G$ between $X_1$ and $X_2$ which contradicts the definition of $\mathbf{Y}$, implies $X_1 \not\perp\!\!\!\perp X_2|\mathbf{Y}$ due to contraction1 and symmetry. Note that if $\mathbf{W} = \emptyset$, then $X_1\mathbf{W} \not\perp\!\!\!\perp X_2|\mathbf{Y}$ directly implies $X_1 \not\perp\!\!\!\perp X_2|\mathbf{Y}$.

Let us assume as induction hypothesis that the statement that we are proving holds for all $n < m$. We now prove it for $n = m$. Since the paths $X_{1:2}$ and $X_{2:m}$ contain less than $m$ nodes and $\mathbf{Y}$ blocks all the other paths in $G$ between $X_1$ and $X_2$ and between $X_2$ and $X_m$, because otherwise there exist several unblocked paths in $G$ between $X_1$ and $X_m$ which contradicts the definition of $\mathbf{Y}$, then $X_1 \not\perp\!\!\!\perp X_2|\mathbf{Y}$ and $X_2 \not\perp\!\!\!\perp X_m|\mathbf{Y}$ due to the induction hypothesis. This together with $sep(X_1, X_m|\mathbf{Y}X_2)$, which follows from the definition of $X_{1:m}$ and $\mathbf{Y}$, implies $X_1 \not\perp\!\!\!\perp X_m|\mathbf{Y}$ due to weak transitivity2.

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. If $con(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ holds in $G$, then there exist some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ such that $X_1 \not\perp\!\!\!\perp X_n|(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$ is in $WT^{sep(G)}_{bas(M)}$ due to the paragraph above and, thus, $\mathbf{X} \not\perp \mathbf{Y}|\mathbf{Z}$ is also in $WT^{sep(G)}_{bas(M)}$ due to weak union and symmetry. ■

We now prove that *con* is complete for reading dependencies from the MUI map $G$ of a WT graphoid $M$, in the sense that it identifies all the dependencies in $M$ that follow from the information about $M$ that is available, namely the dependencies in $bas(M)$, the independencies in $sep(G)$, and the fact that $M$ is a WT graphoid.

**Theorem 6** *Let $M$ be a WT graphoid and $G$ its MUI map. Then, con in $G$ identifies all the dependencies in $WT^{sep(G)}_{bas(M)}$.*

**Proof** It suffices to prove (i) that all the dependencies in $bas(M)$ are identified by *con* in $G$, and (ii) that *con* satisfies the WT graphoid properties. Since the first point is trivial, we only prove the second point. Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$ denote four mutually disjoint subsets of $\mathbf{U}$.

- Symmetry $con(\mathbf{Y},\mathbf{X}|\mathbf{Z}) \Rightarrow con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$. Trivial.
- Decomposition $con(\mathbf{X},\mathbf{Y}|\mathbf{Z}) \Rightarrow con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$. Trivial if $\mathbf{W}$ contains no node in the path $X_{1:n}$ in $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$. If $\mathbf{W}$ contains some node in $X_{1:n}$, then let $X_m$ denote the closest node to $X_1$ that is in $X_{1:n}$ and $\mathbf{W}$. Then, the path $X_{1:m}$ ensures $con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$ because $(\mathbf{X}\setminus X_1)(\mathbf{YW}\setminus X_m)\mathbf{Z}$ blocks all the other paths in $G$ between $X_1$ and $X_m$, since $(\mathbf{X}\setminus X_1)(\mathbf{Y}\setminus X_n)\mathbf{Z}$ blocks all the paths in $G$ between $X_1$ and $X_m$ except $X_{1:m}$ because otherwise there exist several unblocked paths in $G$ between $X_1$ and $X_n$, which contradicts $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$.
- Weak union $con(\mathbf{X},\mathbf{Y}|\mathbf{ZW}) \Rightarrow con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$. Trivial because $\mathbf{W}$ contains no node in the path $X_{1:n}$ in $con(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$.
- Contraction1 $con(\mathbf{X},\mathbf{YW}|\mathbf{Z}) \wedge sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW}) \Rightarrow con(\mathbf{X},\mathbf{W}|\mathbf{Z})$. Since $\mathbf{ZW}$ blocks all the paths in $G$ between $\mathbf{X}$ and $\mathbf{Y}$, then the path $X_{1:n}$ in $con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$ must be between $\mathbf{X}$ and $\mathbf{W}$. To prove that $X_{1:n}$ ensures $con(\mathbf{X},\mathbf{W}|\mathbf{Z})$, we have to prove that this is the only path in $G$ between $X_1$ and $X_n$ that is not blocked by $(\mathbf{X}\setminus X_1)(\mathbf{W}\setminus X_n)\mathbf{Z}$. Assume to the contrary that there is a second such path in $G$. This second path cannot contain any node in $\mathbf{Y}$ for $sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$ to hold. Then, this second path is not blocked by $(\mathbf{X}\setminus X_1)\mathbf{Y}(\mathbf{W}\setminus X_n)\mathbf{Z}$ either. However, this contradicts $con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$, because we have found two paths in $G$ between $X_1$ and $X_n$ that are not blocked by $(\mathbf{X}\setminus X_1)\mathbf{Y}(\mathbf{W}\setminus X_n)\mathbf{Z}$.
- Contraction2 $con(\mathbf{X},\mathbf{YW}|\mathbf{Z}) \wedge sep(\mathbf{X},\mathbf{W}|\mathbf{Z}) \Rightarrow con(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$. Since $\mathbf{Z}$ blocks all the paths in $G$ between $\mathbf{X}$ and $\mathbf{W}$, the path $X_{1:n}$ in $con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$ must be between $\mathbf{X}$ and $\mathbf{Y}$ and, thus, it ensures $con(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$.
- Intersection $con(\mathbf{X},\mathbf{YW}|\mathbf{Z}) \wedge sep(\mathbf{X},\mathbf{Y}|\mathbf{ZW}) \Rightarrow con(\mathbf{X},\mathbf{W}|\mathbf{ZY})$. Since $\mathbf{ZW}$ blocks all the paths in $G$ between $\mathbf{X}$ and $\mathbf{Y}$, the path $X_{1:n}$ in $con(\mathbf{X},\mathbf{YW}|\mathbf{Z})$ must be between $\mathbf{X}$ and $\mathbf{W}$ and, thus, it ensures $con(\mathbf{X},\mathbf{W}|\mathbf{ZY})$.
- Weak transitivity2 $con(\mathbf{X},X_m|\mathbf{Z}) \wedge con(X_m,\mathbf{Y}|\mathbf{Z}) \wedge sep(\mathbf{X},\mathbf{Y}|\mathbf{Z}X_m) \Rightarrow con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ with $X_m \in \mathbf{U}\setminus(\mathbf{XYZ})$. Let $X_{1:m}$ and $X_{m:n}$ denote the paths in $con(\mathbf{X},X_m|\mathbf{Z})$ and $con(X_m,\mathbf{Y}|\mathbf{Z})$, respectively. For $con(\mathbf{X},X_m|\mathbf{Z})$ to hold, $X_{1:m}$ cannot contain any node in $(\mathbf{X}\setminus X_1)\mathbf{Z}$. For $con(X_m,\mathbf{Y}|\mathbf{Z})$ to hold, $X_{m:n}$ cannot contain any node in $(\mathbf{Y}\setminus X_n)\mathbf{Z}$. For $sep(\mathbf{X},\mathbf{Y}|\mathbf{Z}X_m)$ to hold, neither $X_{1:m}$

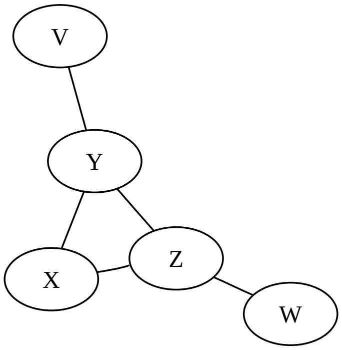


Figure 3: UG in Example 4.

can contain a node in $\mathbf{Y}$, nor $X_{m:n}$ can contain a node in $\mathbf{X}$. Consequently, neither $X_{1:m}$ nor $X_{m:n}$ is blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$.

Furthermore, $X_{1:m}$ and $X_{m:n}$ only intersect on $X_m$. To see this, assume the contrary. Let $X_l$ ($X_l \neq X_m$) denote the closest node to $X_1$ that is in $X_{1:m}$ and $X_{m:n}$. Then, it follows from the paragraph above that neither $X_{1:l}$ and $X_{l:n}$ contain a node in $\mathbf{Z}X_m$. However, this contradicts $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z}X_m)$. Consequently, $X_{1:m}$ followed by $X_{m:n}$ is a path in $G$ between $X_1$ and $X_n$ that, as shown above, is not blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. It remains to prove that this path is unique. Assume to the contrary that there exists a second such path in $G$. For $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z}X_m)$ to hold, this second path must pass through $X_m$. However, this implies that either there exists a second path in $G$ between $X_1$ and $X_m$ that is not blocked by $(\mathbf{X} \setminus X_1)\mathbf{Z}$, or there exists a second path in $G$ between $X_m$ and $X_n$ that is not blocked by $(\mathbf{Y} \setminus X_n)\mathbf{Z}$. This contradicts $con(\mathbf{X}, X_m|\mathbf{Z})$ or $con(X_m, \mathbf{Y}|\mathbf{Z})$.

- Weak transitivity1 $con(\mathbf{X}, X_m|\mathbf{Z}) \wedge con(X_m, \mathbf{Y}|\mathbf{Z}) \wedge sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) \Rightarrow con(\mathbf{X}, \mathbf{Y}|\mathbf{Z}X_m)$ with $X_m \in \mathbf{U} \setminus (\mathbf{XYZ})$. Trivial because the antecedent involves a contradiction: It follows from the proof of weak transitivity2 that $con(\mathbf{X}, X_m|\mathbf{Z})$ and $con(X_m, \mathbf{Y}|\mathbf{Z})$ imply the existence of a path in $G$ between some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ that is not blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$, which contradicts $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$.

■

We devote the rest of this section to some remarks on the two theorems above. Note that *con* in $G$ is not complete if this is understood as being able to identify all the dependencies in $M$. Actually, no sound criterion for reading dependencies from $G$ alone is complete in this sense. Example 2 illustrates this point. Let us now assume that we are dealing with $p'$ instead of with $p$. Then, no sound criterion can conclude $X \not\perp\!\!\!\perp Y|\emptyset$ by just studying $G$ because this dependence does not hold in $p$, and it is impossible to know whether we are dealing with $p$ or $p'$ on the sole basis of $G$.

It seems natural to expect that assuming further independence properties will result in more dependencies being readable from the MUI map of an independence model and, thus, in the necessity of developing a new graphical criterion that identifies them. However, this is not always the case as the following example shows. First, let us define a weak transitivity and composition (WTC) graphoid as a WT graphoid that satisfies composition.

**Example 3** *The graphical criterion con is still sound and complete for reading dependencies from the MUI map G of a WTC graphoid M. Here, complete means that it is able to read all the dependencies in M that can be derived from bas*$(M)$ *and sep*$(G)$ *by applying the WTC graphoid properties. The reason is that every dependence that follows from composition also follows from contraction1. In other words, when rephrased to derive dependencies, composition looks like* $\mathbf{X}\not\perp\!\!\!\perp\mathbf{YW}|\mathbf{Z}\wedge\mathbf{X}\perp\!\!\!\perp\mathbf{Y}|\mathbf{Z}\Rightarrow\mathbf{X}\not\perp\!\!\!\perp\mathbf{W}|\mathbf{Z}$*. As before, the independence in the left-hand side holds if the corresponding sep statement holds in G. However, if sep*$(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ *then sep*$(\mathbf{X},\mathbf{Y}|\mathbf{ZW})$*. Thus, when composition applies so does contraction1, and both imply the same consequent. Any probability distribution that is regular Gaussian or faithful to some UG or DAG is a WTC graphoid (Pearl, 1988; Studený, 2005). Moreover, similarly to Theorem 2, the marginals and conditionals of a strictly positive discrete probability distribution that is a WTC graphoid and has no context-specific dependencies are WTC graphoids (Chickering and Meek, 2002; Peña et al., 2006).*

A sensible question to ask is whether the definition of complete in Theorem 6 coincides with the definition of complete as able to identify all the dependencies shared by all the WT graphoids for which *G* is the MUI map. Currently, we do not have an answer to this question, though we incline to think that the definition in the theorem above is weaker than the alternative one. For instance, if we limit ourselves to WTC graphoids, then *con* in *G* may not identify every dependency shared by all the WTC graphoids for which *G* is the MUI map, as the following example illustrates.

**Example 4** *Consider any regular Gaussian probability distribution whose MUI map is the UG in Figure 3. Such probability distributions exist (Lněnička and Matúš, 2007). Recall that any regular Gaussian probability distribution is a WTC graphoid. Then,* $X\not\perp\!\!\!\perp Y|\emptyset$ *or* $X\not\perp\!\!\!\perp Z|\emptyset$ *because otherwise* $X\perp\!\!\!\perp YZ|\emptyset$ *by composition, which is a contradiction as con*$(X,YZ|\emptyset)$ *holds in G and thus* $X\not\perp\!\!\!\perp YZ|\emptyset$ *by Theorem 5.*

*Assume* $X\not\perp\!\!\!\perp Y|\emptyset$*. Furthermore, con*$(Y,V|\emptyset)$ *holds in G and thus* $Y\not\perp\!\!\!\perp V|\emptyset$ *by Theorem 5. Furthermore, sep*$(X,V|Y)$ *holds in G and thus* $X\perp\!\!\!\perp V|Y$*. Consequently,* $X\not\perp\!\!\!\perp V|\emptyset$ *by weak transitivity2. Likewise,* $X\not\perp\!\!\!\perp W|\emptyset$ *when assuming* $X\not\perp\!\!\!\perp Z|\emptyset$*. Then,* $X\not\perp\!\!\!\perp V|\emptyset$ *or* $X\not\perp\!\!\!\perp W|\emptyset$ *and, thus,* $X\not\perp\!\!\!\perp VW|\emptyset$ *by decomposition. However, con*$(X,VW|\emptyset)$ *does not hold in G.*

At the beginning of this section, we have defined *bas*$(M)$ as the set of dependencies $X\not\perp\!\!\!\perp Y|\mathbf{U}\setminus(XY)$ with $X,Y\in\mathbf{U}$. However, Theorems 5 and 6 remain valid if we redefine *bas*$(M)$ as the set of dependencies $X\not\perp\!\!\!\perp Y|MB(X)\setminus Y$ with $X\in\mathbf{U}$ and $Y\in MB(X)$. A proof follows. Moreover, recall from Section 2 that $X$ and $Y$ are adjacent in the MUI map of $M$ iff $Y\in MB(X)$.

**Proof** It suffices to prove that, when *bas*$(M)$ consists of the dependencies in the first definition, $WT_{bas(M)}^{sep(G)}$ includes the dependencies in the second definition, and vice versa. If $X\not\perp\!\!\!\perp Y|\mathbf{U}\setminus(XY)$, then $X\not\perp\!\!\!\perp Y(\mathbf{U}\setminus(XY)\setminus(MB(X)\setminus Y))|MB(X)\setminus Y$ due to weak union. This together with $sep(X,\mathbf{U}\setminus(XY)\setminus(MB(X)\setminus Y)|Y(MB(X)\setminus Y))$ implies $X\not\perp\!\!\!\perp Y|MB(X)\setminus Y$ due to contraction1. On the other hand, if $X\not\perp\!\!\!\perp Y|MB(X)\setminus Y$, then $X\not\perp\!\!\!\perp Y(\mathbf{U}\setminus(XY)\setminus(MB(X)\setminus Y))|MB(X)\setminus Y$ due to decomposition. This together with $sep(X,\mathbf{U}\setminus(XY)\setminus(MB(X)\setminus Y)|Y(MB(X)\setminus Y))$ implies $X\not\perp\!\!\!\perp Y|\mathbf{U}\setminus(XY)$ due to intersection. ■

It is proven in Becker et al. (2000) that if *M* is a WT graphoid whose MUI map is a forest *G*, then *M* is faithful to *G*. The soundness of *con* allows us to give an alternative proof of this result.

**Proof** Assume to the contrary that *M* is not faithful to *G*. Since *G* is the MUI map of *M*, the previous assumption is equivalent to assume that there exists three mutually disjoint subsets of **U**, say **X**,

```
$FirstPath(X,Y,\mathbf{Z},L)$

$CreatePointers(L)$
for each node $W$ in $L$ do
  if $W \in \mathbf{Z}$ then
    $C[W] = 0$
  else
    $C[W] = 1$
$P = \emptyset$
$Push(X,P)$
$C[X] = 0$
while $P \neq \emptyset$ and $Top(P) \neq Y$ do
  $W = NextAdj(Top(P),L)$
  if $W = \emptyset$ then
    $Pop(P)$
  else
    if $C[W] = 1$ then
      $Push(W,P)$
      $C[W] = 0$
return $P$
```

Table 1: $FirstPath(X,Y,\mathbf{Z},L)$.

$\mathbf{Y}$ and $\mathbf{Z}$, such that $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ is in $M$ but $sep(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ does not hold in $G$. However, if $sep(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ does not hold in $G$, then there must exist a path in $G$ between some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ that is not blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. Furthermore, since $G$ is a forest, that must be the only path in $G$ between $X_1$ and $X_n$ that is not blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. Consequently, $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds in $G$ and, thus, $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$ is in $M$ due to Theorem 5. This is a contradiction and, thus, $M$ is faithful to $G$. ■

Finally, we note that the following graphical criterion, denoted *bou* here, for reading dependencies from the MUI map $G$ of a graphoid $M$ is introduced in Bouckaert (1995): Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$, then $bou(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds when there exist some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ such that $X_1 \in MB(X_n)$ and either $MB(X_1) \setminus X_n \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$ or $MB(X_n) \setminus X_1 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. The criterion is proven to be sound and complete, where complete means that it is able to identify all the dependencies in $M$ that can be derived from $bas(M)$ and $sep(G)$ by applying the graphoid properties. It is clear that *con* identifies a superset of the dependencies identified by *bou* when $M$ satisfies weak transitivity. In such a case, then, *con* represents an advantage over *bou*. As discussed in Section 3, there are important families of probability distributions that are graphoids and satisfy weak transitivity.

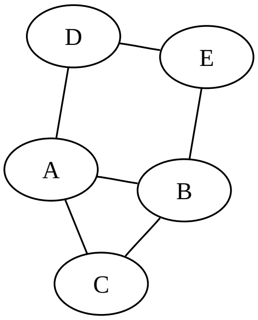
## 6. An Algorithm for Reading (In)Dependencies

In this section, we present an algorithm that jointly implements $sep$ and $con$. We first describe in Table 1 the algorithm $FirstPath(X,Y,\mathbf{Z},L)$ which returns a path in an UG $G$ between $X$ and $Y$ that is not blocked by $\mathbf{Z}$, if such a path exists. We assume that $G$ is represented as a set of adjacency lists $L$, that is, each node in $G$ has an associated ordered list containing the nodes that are adjacent to it in $G$. In the algorithm, the function $CreatePointers(L)$ creates a pointer for each adjacency list in $L$. Each pointer points to the first element of the corresponding list. If the list is empty, the pointer is NULL. The function $NextAdj(W,L)$ returns the element that is pointed by the pointer created by $CreatePointers(L)$ for the adjacency list associated to the node $W$ and, then, moves the pointer to the next element in the list. If there is no next element, the pointer takes value NULL. $NextAdj(W,L)$ returns $\emptyset$ if the pointer is NULL. In the algorithm, $P$ is a stack storing the path currently being explored. The function $Top(P)$ returns the element on the top of $P$. The function $Pop(P)$ removes the element on the top of $P$ from $P$. The function $Push(W,P)$ adds the node $W$ to the top of $P$. Finally, $C$ is an array containing a binary entry for each node in $G$ indicating whether the node should (1) or should not (0) be considered to extend $P$. The algorithm sets $C[W]$ to 0 iff $W$ is in $\mathbf{Z}$ or $W$ was considered before and the algorithm already found a path from $X$ to $W$ that is not blocked by $\mathbf{Z}$ (see lines 4 and 17). A node considered before should not be considered again because either it is still in $P$ or if it is not in $P$ then it does not lead to $Y$. To see the latter point, note that the algorithm works in a depth-first fashion: It extends $P$ with a node $W$ that is adjacent to the top element of $P$ if $W$ is not in the blocking set $\mathbf{Z}$ and has not been considered before (see lines 11, 15 and 16). When $P$ cannot be extended further, the algorithm backtracks by removing the top element of $P$ from $P$ and exploring an alternative extension (see lines 12 and 13). The algorithm ends when $P$ is empty, meaning that all the paths in $G$ starting with $X$ were explored and none of them reached $Y$ without visiting $\mathbf{Z}$, or when the top element of $P$ is $Y$, meaning that a path in $G$ between $X$ and $Y$ that is not blocked by $\mathbf{Z}$ was found and stored in $P$.

$FirstPath(X,Y,\mathbf{Z},L)$ is considered to run in at most $O(e+n)$ time where $e$ and $n$ are the number of edges and nodes in $G$, respectively. To see it, note that thanks to the use of $C$ each node is pushed into $P$ at most once. For each node $V$ pushed into $P$, the algorithm performs as many iterations of lines 10-17 as adjacent nodes $V$ has plus one last iteration when $W=\emptyset$. These iterations need not be consecutive. So, the number of iterations of lines 10-17 that the algorithm performs is bounded from above by $2e+n$. Additionally, the algorithm performs $n$ iterations of lines 2-6 to initialize $C$. As described in the paragraph above, $CreatePointers(L)$ is considered to run in $O(n)$ time whereas any other operation or function in the algorithm is considered to run in $O(1)$ time. Consequently, $FirstPath(X,Y,\mathbf{Z},L)$ is considered to require at most $O(e+n)$ time.

Recall from above that we assume that $G$ is represented as a set of adjacency lists $L$ and that each of these lists is ordered. Let us now reverse the order of the elements in each of these lists and let $L'$ denote the resulting set of adjacency lists. Our premise is that producing $L'$ from $L$ takes $O(e)$ time. Obviously, $L'$ is also an adjacency list representation of $G$. However, $FirstPath(X,Y,\mathbf{Z},L)$ and $FirstPath(X,Y,\mathbf{Z},L')$ return the same path iff that is the only path in $G$ between $X$ and $Y$ that is not blocked by $\mathbf{Z}$. We formally prove this assertion below, but first we give an example. Let $G$, $L$ and $L'$ be as shown in Table 2. Then, $FirstPath(A,B,\emptyset,L)$ returns the path $A,B$ whereas $FirstPath(A,B,\emptyset,L')$ returns the path $A,D,E,B$.

**Proof** First, we introduce a total order for the paths between $X$ and $Y$ in $G$ that are not blocked by $\mathbf{Z}$. For this purpose, we associate to each such path a sequence of length $n-1$ of natural numbers



Adjacency lists $L$

| Node | Adjacency list |
|---|---|
| A | B, C, D |
| B | A, C, E |
| C | A, B |
| D | A, E |
| E | B, D |

Adjacency lists $L'$

| Node | Adjacency list |
|---|---|
| A | D, C, B |
| B | E, C, A |
| C | B, A |
| D | E, A |
| E | D, B |

Table 2: Example where $\mathit{FirstPath}(A,B,\emptyset,L)$ and $\mathit{FirstPath}(A,B,\emptyset,L')$ return different paths.

between 0 and $n-1$, where $n$ is the number of nodes in $G$. For a path $X = X_1,\ldots,X_k = Y$, the sequence is $[o_1,\ldots,o_{k-1},0,\ldots,0]$, where $o_i$ is the position of $X_{i+1}$ in the adjacency list associated to $X_i$. We order the sequences lexicographically, which results in the desired order.

We now prove that if there exists a path between $X$ and $Y$ in $G$ that is not blocked by $\mathbf{Z}$, then $\mathit{FirstPath}(X,Y,\mathbf{Z},L)$ returns the first such a path in the order described above. Let $X = X_1,\ldots,X_k = Y$ denote the first such a path. It suffices to prove that at some point $P$ contains (from the bottom to the top) the nodes $X_1,\ldots,X_j$ for any $1 \leq j \leq k$. We prove the result by induction over $j$.

The result is immediate for $j=1$. Assume as induction hypothesis that the result holds for all $i<j$. We now prove the result for $j$. By the induction hypothesis, at some point $t$, $P$ contains (from the bottom to the top) the nodes $X_1,\ldots X_{j-1}$. Pushing into $P$ a node that is adjacent to $X_{j-1}$ and appears before $X_j$ in the adjacency list associated to $X_{j-1}$ cannot lead to a path between $X_1$ and $X_k$, because that would contradict the assumption that $X = X_1,\ldots,X_k = Y$ is the first path between $X$ and $Y$ in $G$ that is not blocked by $\mathbf{Z}$. Therefore, the algorithm will eventually pop from $P$ every element pushed after time $t$ and, then, $X_j$ will be pushed into $P$.

Finally, note that if we reverse the order of the nodes in each adjacency list in $L$, then we also reverse the order of the paths described above. Consequently, $X = X_1,\ldots,X_k = Y$ is the last path in the order considered by $\mathit{FirstPath}(X,Y,\mathbf{Z},L')$. Thus, $\mathit{FirstPath}(X,Y,\mathbf{Z},L')$ does not return $X = X_1,\ldots,X_k = Y$ unless that is the only path between $X$ and $Y$ in $G$ that is not blocked by $\mathbf{Z}$. ■

The algorithm that jointly implements *sep* and *con* is as follows. If $\mathit{FirstPath}(X,Y,(\mathbf{X}\setminus X)(\mathbf{Y}\setminus X)\mathbf{Z},L)$ and $\mathit{FirstPath}(X,Y,(\mathbf{X}\setminus X)(\mathbf{Y}\setminus X)\mathbf{Z},L')$ return the same path for some $X\in\mathbf{X}$ and $Y\in\mathbf{Y}$, then $\mathit{con}(\mathbf{X},\mathbf{Y}|\mathbf{Z})$. On the other hand, if $\mathit{FirstPath}(X,Y,(\mathbf{X}\setminus X)(\mathbf{Y}\setminus X)\mathbf{Z})$ returns no path for all $X\in\mathbf{X}$ and $Y\in\mathbf{Y}$, then $\mathit{sep}(\mathbf{X},\mathbf{Y}|\mathbf{Z})$. Finally, if neither of the two previous conditions is met, then neither $\mathit{con}(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ nor $\mathit{sep}(\mathbf{X},\mathbf{Y}|\mathbf{Z})$. The algorithm is considered to require at most $O(n^2(e+n))$ time.

## 7. An Application to Bioinformatics

Our end-goal is to apply the results in this paper to our project on atherosclerosis gene expression data analysis in order to learn dependencies between genes. We believe that it is not unrealistic to assume that the probability distribution underlying our data satisfies strict positivity and weak transitivity and, thus, that it is a WT graphoid. We base this belief on the following argument. The cell is the functional unit of all the organisms and includes all the information necessary to regulate its function. This information is encoded in the DNA of the cell, which is divided into a set of genes, each coding for one or more proteins. Proteins are required for practically all the functions in the cell. The amount of protein produced depends on the expression level of the coding gene which, in turn, depends on the amount of proteins produced by other genes. Therefore, a dynamic BN seems to be a relatively accurate model of the cell: The nodes represent the genes and proteins, and the edges and parameters represent the causal relations between the gene expression levels and the protein amounts. As a matter of fact, dynamic BNs have become very popular models of gene networks for the last few years (Bernard and Hartemink, 2005; Friedman et al., 1998; Husmeier, 2003; Kim et al., 2003; Murphy and Mian, 1999; Ong et al., 2002; Perrin et al., 2003; Zou and Conzen, 2005). It is important that the BN is dynamic because a gene can regulate some of its regulators and even itself with some time delay. Since the technology for measuring the state of the protein nodes is not widely available yet, the data in most projects on gene expression data analysis is a finite sample of the probability distribution represented by the dynamic BN after marginalizing the protein nodes out. The probability distribution with no node marginalized out is, in the measure-theoretic sense discussed in Section 3, almost surely faithful to the dynamic BN (Meek, 1995) and, thus, it satisfies weak transitivity (Pearl, 1988) and, thus, so does the probability distribution after marginalizing the protein nodes out (see Theorem 2). The assumption that the probability distribution sampled is strictly positive is justified because measuring the state of the gene nodes involves a series of complex wet-lab and computer-assisted steps that introduces noise in the measurements (Sebastiani et al., 2003). Obviously, the reasoning above can be extended to include any other molecules that, in addition to proteins, regulate gene expression but are not measured.

In the rest of this section we focus on Gaussian graphical models (GGMs) of gene networks, which have received increasing attention from the bioinformatics community as a means to gain insight into gene networks (Castelo and Roverato, 2006; Dobra et al., 2004; Kishino and Waddell, 2000; Li and Gui, 2006; Schäfer and Strimmer, 2005a,b; Toh and Horimoto, 2002; Waddell and Kishino, 2000; Wang et al., 2003; Wu et al., 2003). Assume that each random variable in $\mathbf{U}$ represents (the expression level of) a gene in the network under study. Assume also that $\mathbf{U}$ follows a regular Gaussian probability distribution $\mathcal{N}(\mu,\Sigma)$. This is a ubiquitous assumption in bioinformatics. The GGM of the gene network is nothing else but the MUI map of $\mathcal{N}(\mu,\Sigma)$ (Lauritzen, 1996; Whittaker, 1990). Therefore, two genes $X$ and $Y$ are adjacent in the GGM iff $X \not\perp Y | \mathbf{U} \setminus (XY)$ or, equivalently, iff $(\Sigma^{-1})_{XY} \neq 0$ (Lauritzen, 1996). In practice, $\Sigma$ is unknown and the only information about it that is available is a finite sample from $\mathcal{N}(\mu,\Sigma)$. The usual way of proceeding in practice consists of two steps: First, estimating $\Sigma$ from the sample and, then, making two genes adjacent in the GGM iff the corresponding entry in the inverse of the estimate of $\Sigma$ significantly differs from zero. We refer the interested reader to the works cited above for different solutions to these two steps. Recall that regular Gaussian probability distributions are WT graphoids and, thus, that the results obtained in the previous sections of this paper apply to GGMs of gene networks.

The GGM of a gene network is a powerful tool for discovering gene independencies, because *sep* is sound and complete (in the sense discussed in Section 4) for reading independencies from it. However, bioinformaticians are typically more interested in discovering gene dependencies, because these provide contexts in which the expression level of some genes is informative about that of some other genes which, in some cases, can lead to hypothesize dependencies, functional

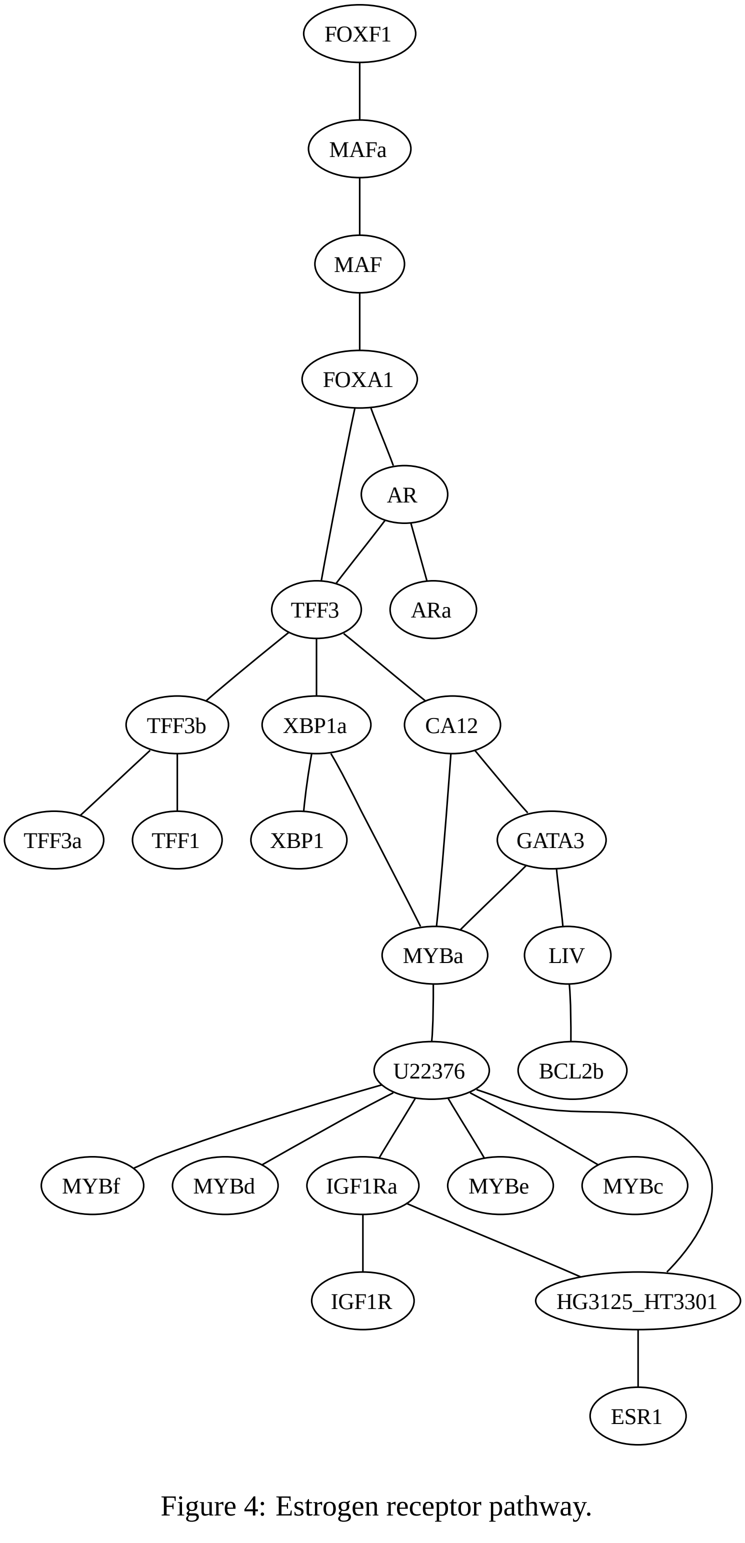


Figure 4: Estrogen receptor pathway.

relations, causal relations, the effects of manipulation experiments, etc. Thanks to *con*, which is sound and complete (in the sense discussed in Section 5) for reading dependencies from the GGM of a gene network, such a model is now a powerful tool for discovering gene dependencies too. We illustrate this with a real world example. In Dobra et al. (2004), a GGM over 12558 nodes is learnt from 158 breast cancer samples. Unfortunately, the GGM is neither reported in the paper nor available from the authors. However, the authors do report in the paper the subgraph of the GGM that is induced by some genes that are known to be related to the estrogen receptor (ER) and TFF1 genes. ER is a transcription factor that plays a key role in breast cancer, and TFF1 is a target of ER. This subgraph, depicted in Figure 4, suffices for our illustrative purposes. Note that the graph in the figure is not necessarily the GGM over the nodes in the graph. It is also worth mentioning that the nodes in the subgraph do not correspond to genes but to probe sets. A probe set is a collection of probes designed to measure the abundance of a particular DNA sequence. Since this sequence is for technical reasons usually shorter than that of a gene, a gene may have several probe sets associated with it, each measuring the abundance of a different subsequence of the gene. This aims at measuring gene expression more accurately. The following genes have multiple probe sets (nodes) in Figure 4: ER (ESR1, HG3125-HT3301), MYB (U22376, MYBa, MYBc, MYBd, MYBe, MYBf), AR (AR, ARa), c-MAF (MAF, MAFa), TFF3 (TFF3, TFF3a, TFF3b), XBP (XBP1, XBP1a), and IGF1R (IGF1R, IGF1Ra). According to Dobra et al. (2004), it is known that ER regulates TFF1, FOXA1 regulates TFF1, GATA3 possibly regulates TFF1, AR regulates ER, MAF inhibits MYB, FOXF1 possibly interacts with ER, and AR regulates IGF1R. Had these relations been unknown, we could have obtained principled clues about them by just applying *con* to the subgraph in Figure 4. For instance, if $\mathbf{U}$ denotes all the 12558 probe sets in the GGM, then *con* enables us to conclude that the following gene dependencies hold in the underlying probability distribution: ER is conditionally dependent on TFF1 since *con*(HG3125-HT3301, TFF1$|\mathbf{U}\setminus\{$HG3125-HT3301, U22376, MYBa, XBP1a, TFF3, TFF3b, TFF1$\}$), FOXA1 is conditionally dependent on TFF1 since *con*(FOXA1, TFF1$|\mathbf{U}\setminus\{$FOXA1, TFF3, TFF3b, TFF1$\}$), GATA3 is conditionally dependent on TFF1 since *con*(GATA3, TFF1$|\mathbf{U}\setminus\{$GATA3, CA12, TFF3, TFF3b, TFF1$\}$), AR is conditionally dependent on ER since *con*(AR, HG3125-HT3301$|\mathbf{U}\setminus\{$AR, TFF3, CA12, MYBa, U22376, HG3125-HT3301$\}$), MAF is conditionally dependent on MYB since *con*(MAF, MYBa$|\mathbf{U}\setminus\{$MAF, FOXA1, TFF3, CA12, MYBa$\}$), FOXF1 is conditionally dependent on ER since *con*(FOXF1, HG3125-HT3301$|\mathbf{U}\setminus\{$FOXF1, MAFa, MAF, FOXA1, TFF3, CA12, MYBa, U22376, HG3125-HT3301$\}$), and AR is conditionally dependent on IGF1R since *con*(AR, IGF1Ra$|\mathbf{U}\setminus\{$AR, TFF3, CA12, MYBa, U22376, IGF1Ra$\}$). Furthermore, *con* also enables us to conclude that $X \not\perp Y|\mathbf{U}\setminus\{X, \text{U22376}, Y\}$ with $X, Y \in \{$MYBa, MYBc, MYBd, MYBe, MYBf$\}$ and $X \neq Y$, and that TFF3 $\not\perp$ TFF3a $|\mathbf{U}\setminus\{$TFF3, TFF3b, TFF3a$\}$. These dependencies make sense as they are between different probe sets of the same gene. Note that none of the dependencies discussed above was used in the construction of the GGM. Note also that each of the dependencies discussed above involves a conditioning set of maximum size. It is very likely that these dependencies also hold for smaller conditioning sets, but we cannot confirm this point without seeing the complete GGM the subgraph in Figure 4 is part of which, as discussed above, is not available. In any case, it is clear that *con* improves the current interpretation of GGMs of gene networks by allowing reading biologically meaningful gene dependencies.

## 8. Discussion

The MUI map $G$ of an independence model $M$ is typically used to identify independencies that hold in $M$ via vertex separation. However, lack of vertex separation in $G$ does not necessarily imply dependency in $M$. In this paper, we have studied when lack of vertex separation does imply dependency. This should be relevant for those interested in graphical models, as it allows to infer from MUI maps (i.e., without numerical calculation) not only independencies but also dependencies. Specifically, in this paper we have introduced a graphical criterion called *con* for reading dependencies from $G$ when $M$ is a WT graphoid, that is, a graphoid that satisfies weak transitivity. Specifically, $con(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds when there exist some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ such that there exists exactly one path in $G$ between $X_1$ and $X_n$ that is not blocked by $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. We have proven that the criterion is sound and complete, where complete means that it is able to read all the dependencies in $M$ that can be derived by applying the WT graphoid properties to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by vertex separation. Note that our criterion remains inconclusive if there are several unblocked paths between any node in $\mathbf{X}$ and any node in $\mathbf{Y}$, though $\mathbf{X}$ and $\mathbf{Y}$ may be dependent given $\mathbf{Z}$ in $M$. However, we have shown in Section 5 that neither our criterion nor any other sound criterion can identify all the dependencies in $M$.

Note that our criterion is antimonotone in the following sense: If some edges are added to $G$ then some *con* statements may not longer hold, whereas if some edges are removed from $G$ then some new *con* statements may hold. This antimonotone property should be taken into account if one is to remove ”weak” edges from $G$ to make it sparser, because this may result in false dependencies being identified. However, this has nothing to do with the correctness of our criterion, something that we have proved, but with the fact that after removing ”weak” edges $G$ is an approximation to the true MUI map. One may consider to extend our work with a measure of confidence in the dependencies identified by our criterion. Such a measure could be a function of the confidence in the dependencies used in the construction of $G$. We have not pursued this idea further.

A work that is closely related to ours is Bouckaert (1995), which introduces the following graphical criterion, denoted *bou* here, for reading dependencies from the MUI map $G$ of a graphoid $M$: $bou(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds when there exist some $X_1 \in \mathbf{X}$ and $X_n \in \mathbf{Y}$ such that $X_1 \in MB(X_n)$ and either $MB(X_1) \setminus X_n \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$ or $MB(X_n) \setminus X_1 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$. In other words, $bou(\mathbf{X},\mathbf{Y}|\mathbf{Z})$ holds when there exist two nodes, one in $\mathbf{X}$ and the other in $\mathbf{Y}$, that are neighbors in $G$ and, moreover, the rest of the neighbors of one of them are among the rest of the nodes in the statement. The criterion is proven to be sound and complete, where complete means that it is able to identify all the dependencies in $M$ that can be derived by applying the graphoid properties to the dependencies used in the construction of $G$ and the independencies obtained from $G$ by vertex separation.

Although *bou* is tailored to the case where $M$ is a graphoid, it can obviously be applied when $M$ is a WT graphoid. However, there are two main differences between *bou* and our criterion that make the latter more powerful when $M$ is a WT graphoid. First, our criterion does not require that $X_1$ is adjacent to $X_n$ in $G$, that is, there can be a path of length greater than one between $X_1$ and $X_n$. Second, our criterion does not require than all the nodes adjacent to either $X_1$ or $X_n$ in $G$ are among $\mathbf{XYZ}$, that is, all the paths between $X_1$ and $X_n$ but one can be blocked by nodes that are neither adjacent to $X_1$ nor $X_n$. Consequently, our criterion represents an advantage over *bou* when $M$ is a WT graphoid, as it allows us to identify a superset of the dependencies identified by *bou*. Interestingly, WT graphoids are a rich subclass of graphoids, including any regular Gaussian distribution, any probability distribution that is faithful to some UG or DAG, and all the marginals and conditionals of almost all the strictly positive discrete probability distributions that factorize according to a DAG. We believe that these probability distributions are encountered in many applications and, thus, that the work presented in this paper is of interest to the machine learning community. For instance, regular Gaussian distributions are ubiquitous. Likewise, it is rather usual to assume that the probability distribution underlying the domain at hand is strictly positive and faithful to some DAG, though one may be forced to work with a marginal of it because only a subset of the nodes in the DAG are observable. The strict positivity assumptions is usually justified by measurement errors, whereas the DAG faithfulness assumption is usually justified by the fact that many domains have a causality structure. For a more concrete case example, recall our discussion on bioinformatics applications in Section 7.

A problem that remains open is whether our criterion is complete in the sense that it identifies all the dependencies shared by all the WT graphoids for which $G$ is the MUI map. This is a problem that we are currently studying though, as we have argued in Section 5, we incline to think that our criterion only identifies a proper subset of those dependencies. To the best of our knowledge, there has not been any attempt to solve this problem. As a matter of fact, one can think of an analogous problem for *bou* and the graphoids for which $G$ is the MUI map. However, such a problem is not even mentioned in Bouckaert (1995).

Another problem we are currently working on is the development of a graphical criterion for reading dependencies from the minimal directed independence maps of WT graphoids. As a first step, we have derived such a criterion for the case where the minimal directed independence maps are polytrees (Peña, 2007).

Finally, it is worth recalling that, as an intermediate step in the derivation of our criterion, we have proved that for any UG $G$ there exists a strictly positive discrete probability distribution that is faithful to $G$ for any sample spaces (with at least two possible states) of the random variables in $\mathbf{U}$. For any DAG, an analogous result follows from Meek (1995). Both results are subsumed by our recent work in Peña (2009), which proves an analogous result for chain graphs. Note that such a result is stronger than Theorem 7.2 in Studený and Bouckaert (1998), which proves the result for some sample spaces (with at least two possible states) of the random variables in $\mathbf{U}$ but not for any such sample spaces.

## Acknowledgments

This work is funded by the Swedish Research Council (ref. VR-621-2005-4202), the Wallenberg Foundation, the Ph.D. Programme in Medical Bioinformatics, the Swedish Foundation for Strategic Research, Clinical Gene Networks AB, and Linköping University. We thank the editor and the anonymous referees for their insightful comments.

## References

Ann Becker, Dan Geiger and Christopher Meek. Perfect tree-like Markovian distributions. In *Proceedings on the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 19-23, 2000.

Allister Bernard and Alexander J. Hartemink. Informative structure priors: Joint learning of dynamic regulatory networks from multiple types of data. In *Pacific Symposium on Biocomputing 2005*, pages 459-470, 2005.

Remco R. Bouckaert. *Bayesian belief networks: From construction to inference*. PhD Thesis, University of Utrecht, 1995.

Robert Castelo and Alberto Roverato. A robust procedure for Gaussian graphical model search from microarray data with p larger than n. *Journal of Machine Learning Research*, 7:2621-2650, 2006.

David M. Chickering and Christopher Meek. Finding optimal Bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pages 94-102, 2002.

Adrian Dobra, Chris Hans, Beatrix Jones, Joseph R. Nevins, Guang Yao and Mike West. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90:196-212, 2004.

Nir Friedman, Kevin Murphy and Stuart Russell. Learning the structure of dynamic probabilistic networks. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 139-147, 1998.

Dan Geiger and Judea Pearl. Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics*, 21:2001-2021, 1993.

Paul R. Halmos. *Measure Theory*. Van Nostrand, 1966.

Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19:2271-2282, 2003.

Sun Yong Kim, Seiya Imoto and Satoru Miyano. Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. In *Proceedings of the First International Workshop on Computational Methods in Systems Biology*, pages 104-113, 2003.

Hirohisa Kishino and Peter J. Waddell. Correspondence analysis of genes and tissue types and finding genetic links from microarray data. *Genome Informatics*, 11:83-95, 2000.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

Hongzhe Li and Jiang Gui. Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics*, 7:302-317, 2006.

Radim Lněnička and František Matúš. On Gaussian conditional independence structures. *Kybernetika*, 43:327-342, 2007.

Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 411-418, 1995.

Kevin Murphy and Saira Mian. Modelling gene expression data using dynamic Bayesian networks. Technical Report, University of California, 1999.

Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 2003.

Masashi Okamoto. Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *The Annals of Statistics*, 1:763-765, 1973.

Irene M. Ong, Jeremy D. Glasner and David Page. Modelling regulatory pathways in E. Coli from time series expression profiles. *Bioinformatics*, 18:S241-S248, 2002.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

Jose M. Peña. Faithfulness in chain graphs: The discrete case. *International Journal of Approximate Reasoning*. Under review, 2009.

Jose M. Peña, Roland Nilsson, Johan Björkegren and Jesper Tegnér. Identifying the relevant nodes without learning the model. In *Proceedings of the Twentysecond Conference on Uncertainy in Artificial Intelligence*, pages 367-374, 2006.

Jose M. Peña. Reading dependencies from polytree-like Bayesian networks. In *Proceedings of the Twentythird Conference on Uncertainty in Artificial Intelligence*, pages 303-309, 2007.

Bruno-Edouard Perrin, Liva Ralaivola, Aurélien Mazurie, Samuele Bottani, Jacques Mallet and Florence d'Alché-Buc. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, 19:ii138-ii148, 2003.

Juliane Schäfer and Korbinian Strimmer. Learning large-scale graphical Gaussian models from genomic data. In *Science of Complex Networks: From Biology to the Internet and WWW*, 2005a.

Juliane Schäfer and Korbinian Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21:754-764, 2005b.

Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4, 2005c.

Paola Sebastiani, Emanuela Gussoni, Isaac S. Kohane and Marco F. Ramoni. Statistical challenges in functional genomics (with discussion). *Statistical Science*, 18:33-60, 2003.

P. Stein. A note on the volume of a simplex. *The American Mathematical Monthly*, 73:299-301, 1966.

Milan Studený. *Probabilistic Conditional Independence Structures*. Springer, 2005.

Milan Studený and Remco R. Bouckaert. On chain graph models for description of conditional independence structures. *The Annals of Statistics*, 26:1434-1495, 1998.

Hiroyuki Toh and Katsuhisa Horimoto. Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling. *Bioinformatics*, 18:287-297, 2002.

Peter J. Waddell and Hirohisa Kishino. Cluster inference methods and graphical models evaluated on NCI60 microarray gene expression data. *Genome Informatics*, 11:129-140, 2000.

Junbai Wang, Ola Myklebost and Eivind Hovig. MGraph: Graphical models for microarray analysis. *Bioinformatics*, 19:2210-2211, 2003.

Joe Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, 1990.

Xintao Wu, Yong Ye and Kalpathi R. Subramanian. Interactive analysis of gene interactions using graphical Gaussian model. In *Proceedings of the Third ACM SIGKDD Workshop on Data Mining in Bioinformatics*, pages 63-69, 2003.

Min Zou and Suzanne D. Conzen. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21:71-79, 2005.