Reading Dependencies from the Minimal Undirected Independence Map of a Graphoid that Satisfies Weak Transitivity

Jose M. Peña
Linköping University
Linköping, Sweden

Roland Nilsson
Linköping University
Linköping, Sweden

Johan Björkegren
Karolinska Institutet
Stockholm, Sweden

Jesper Tegnér
Linköping University
Linköping, Sweden

Abstract

We present a sound and complete graphical criterion for reading dependencies from the minimal undirected independence map of a graphoid that satisfies weak transitivity. We argue that assuming weak transitivity is not too restrictive.

1 Introduction

A minimal undirected independence map G of an independence model p is used to read independencies that hold in p. Sometimes, however, G can also be used to read dependencies holding in p. For instance, if p is a graphoid that is faithful to G then, by definition, vertex separation is a sound and complete graphical criterion for reading dependencies from G. If p is simply a graphoid, then there also exists a sound and complete graphical criterion for reading dependencies from G (Bouckaert, 1995).

In this paper, we introduce a sound and complete graphical criterion for reading dependencies from G under the assumption that p is a graphoid that satisfies weak transitivity. Our criterion allows reading more dependencies than the criterion in (Bouckaert, 1995) at the cost of assuming weak transitivity. We argue that this assumption is not too restrictive. Specifically, we show that there exist important families of probability distributions that are graphoids and satisfy weak transitivity. Before going into the details of our contribution, we review some key concepts in the following section.

2 Preliminaries

The following definitions and results can be found in most books on probabilistic graphical models, e.g. (Pearl, 1988; Studený, 2005). Let $\mathbf{U}$ denote a set of random variables. Unless otherwise stated, all the 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 p is a set of independencies of the form $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$. We represent that an independency is in p by $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ and that an independency is not in p by $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$. 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 $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W}|\mathbf{Z} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{YW}|\mathbf{Z}$, and intersection $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W}|\mathbf{ZY} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{YW}|\mathbf{Z}$. Any strictly positive probability distribution is a graphoid.

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 directed and acyclic graph (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 (i) Z does not have two parents in the path and $Z \in \mathbf{Z}$, or (ii) Z has two parents in the path and neither Z nor any of its descendants in G is in $\mathbf{Z}$. An independence model p 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 probability distribution that is faithful to some UG or DAG is a graphoid. An UG G is an undirected independence map of an independence model p

when $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ if $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$. Moreover, G is a minimal undirected independence (MUI) map of p when removing any edge from G makes it cease to be an independence map of p. A Markov boundary of $X \in \mathbf{U}$ in an independence model p 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 p is a graphoid or WT graphoid, then (i) $MB(X)$ is unique for all X, (ii) the MUI map G of p 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, θ) where G is a DAG and θ are parameters specifying a probability distribution for each $X \in \mathbf{U}$ given its parents in G, $p(X|Pa(X))$. The BN represents the probability distribution $p = \prod_{X \in \mathbf{U}} p(X|Pa(X))$. Then, G is an independence map of a probability distribution p iff p can be represented by a BN with DAG G.

3 WT Graphoids

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. We call WT graphoid to any graphoid p 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 $V \in \mathbf{U} \setminus (\mathbf{XYZ})$. We now argue that there exist important families of probability distributions that are WT graphoids and, thus, that WT graphoids are worth studying. For instance, any probability distribution that is Gaussian or faithful to some UG or DAG is a WT graphoid (Pearl, 1988; Studený, 2005). There also exist probability distributions that are WT graphoids although they are neither Gaussian nor faithful to any UG or DAG. For instance, it follows from the theorem below that the probability distribution that results from marginalizing some nodes out and instantiating some others in a probability distribution that is faithful to some DAG is a WT graphoid, although it may be neither Gaussian nor faithful to any UG or DAG.

Theorem 1. *Let p be a probability distribution that is a WT graphoid and let $\mathbf{W} \subseteq \mathbf{U}$. Then, $p(\mathbf{U} \setminus \mathbf{W})$ is a WT graphoid. If $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$, then $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ for any $\mathbf{w}$ 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, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ in $p(\mathbf{U} \setminus \mathbf{W})$ iff $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ in p and, thus, $p(\mathbf{U} \setminus \mathbf{W})$ satisfies the WT graphoid properties because p satisfies them. If $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$ then, for any $\mathbf{w}$, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ in $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ iff $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW}$ in p. Then, $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ for any $\mathbf{w}$ satisfies the WT graphoid properties because p satisfies them. □

We now show that it is not too restrictive to assume in the theorem above that $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$, because there exist important families of probability distributions whose all or almost all the members satisfy such an assumption. For instance, if p is a Gaussian probability distribution, then $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$, because the independencies in $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ only depend on the variance-covariance matrix of p (Anderson, 1984). Let us now consider all the multinomial probability distributions for which a DAG G is an independence map and denote them by $M(G)$. The following theorem, which is inspired by (Meek, 1995), proves that the probability of randomly drawing from $M(G)$ a probability distribution p such that $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ does not have the same independencies for all $\mathbf{w}$ is zero (see for a proof).

Theorem 2. *The probability distributions p in $M(G)$ for which 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 independencies for all $\mathbf{w}$ have Lebesgue measure zero wrt $M(G)$.*

Proof. The proof basically proceeds in the same way as that of Theorem 7 in (Meek, 1995), so we refer the reader to that paper for more details. Let $\mathbf{W} \subseteq \mathbf{U}$ and let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three disjoint subsets of $\mathbf{U} \setminus \mathbf{W}$. 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 $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w}')$, the following equations must be satisfied: $p(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w})p(\mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w}) - p(\mathbf{X} = \mathbf{x}, \mathbf{Z} =$

$\mathbf{z}, \mathbf{W} = \mathbf{w})p(\mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w}) = 0$ for all $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Each equation is a polynomial in the BN parameters corresponding to G, because each term $p(\mathbf{V} = \mathbf{v})$ in the equations is the summation of products of BN parameters (Meek, 1995). Furthermore, each polynomial is non-trivial, i.e. not all the values of the BN parameters corresponding to G are solutions to the polynomial. To see it, it suffices to rename $\mathbf{w}$ to $\mathbf{w}'$ and $\mathbf{w}'$ to $\mathbf{w}$ because, originally, $\mathbf{X} \not\perp \mathbf{Y}|\mathbf{Z}$ in $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w}')$. Let $sol(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ denote the set of solutions to the polynomial for $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Then, $sol(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ has Lebesgue measure zero wrt $\mathbb{R}^n$, where n is the number of linearly independent BN parameters corresponding to G, because it consists of the solutions to a non-trivial polynomial (Okamoto, 1973). Then, $sol = \bigcup_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}} \bigcup_{\mathbf{w}} \bigcap_{\mathbf{x},\mathbf{y},\mathbf{z}} sol(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ has Lebesgue measure zero wrt $\mathbb{R}^n$, because the finite union and intersection of sets of Lebesgue measure zero has Lebesgue measure zero too. Consequently, the probability distributions p in $M(G)$ such that $p(\mathbf{U} \setminus \mathbf{W}|\mathbf{W} = \mathbf{w})$ does not have the same independencies for all $\mathbf{w}$ have Lebesgue measure zero wrt $\mathbb{R}^n$ because they are contained in sol. These probability distributions also have Lebesgue measure zero wrt $M(G)$, because $M(G)$ has positive Lebesgue measure wrt $\mathbb{R}^n$ (Meek, 1995). □

Finally, we argue in Section 6 that it is not unrealistic to assume that the probability distribution underlying the learning data in most projects on gene expression data analysis, one of the hottest areas of research nowadays, is a WT graphoid.

4 Reading Independencies

By definition, *sep* is sound for reading independencies from the MUI map G of a WT graphoid p, i.e. it only identifies independencies in p. Now, we prove that *sep* in G is also complete in the sense that it identifies all the independencies in p that can be identified from G. Specifically, we prove that there exist multinomial and Gaussian probability distributions that are faithful to G. Such probability distributions have all and only the independencies that *sep* identifies from G. Moreover, such probability distributions must be WT graphoids because *sep* satisfies the WT graphoid properties (Pearl, 1988). The fact that *sep* in G is complete, in addition to being an important result in itself, is important for reading as many dependencies as possible from G (see Section 5).

Theorem 3. *Let G be an UG. There exist multinomial and Gaussian probability distributions that are faithful to G.*

Proof. We first prove the theorem for multinomial probability distributions. Create a copy H of G. For each connected component in H, choose a node and direct all the edges away from it. Repeat the following step while possible: Choose three nodes $X, Y, Z \in \mathbf{U}$ such that the subgraph of H induced by them is $X \to Z \leftarrow Y$ and replace it by $X \to Z \to W_{ZY} \leftarrow Y$ where $W_{ZY} \notin \mathbf{U}$ is an auxiliary node. Let $\mathbf{W}$ denote all the auxiliary nodes created in the previous step. Then, H is a DAG over $\mathbf{UW}$. Moreover, $sep(\mathbf{X}, \mathbf{Y}|\mathbf{ZW})$ in H iff $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ in G (Pearl, 1988).

The probability distributions $p(\mathbf{U}, \mathbf{W})$ in $M(H)$ that are faithful to H and satisfy that $p(\mathbf{U}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$ have positive Lebesgue measure wrt $M(H)$ because (i) $M(H)$ has positive Lebesgue measure wrt $\mathbb{R}^n$ (Meek, 1995), (ii) the probability distributions in $M(H)$ that are not faithful to H have Lebesgue measure zero wrt $M(H)$ (Meek, 1995), (iii) the probability distributions $p(\mathbf{U}, \mathbf{W})$ in $M(H)$ such that $p(\mathbf{U}|\mathbf{W} = \mathbf{w})$ does not have the same independencies for all $\mathbf{w}$ have Lebesgue measure zero wrt $M(H)$ by Theorem 2, and (iv) the union of the probability distributions in (ii) and (iii) has Lebesgue measure zero wrt $M(H)$ because the finite union of sets of Lebesgue measure zero has Lebesgue measure zero.

Let $p(\mathbf{U}, \mathbf{W})$ denote any probability distribution in $M(H)$ that is faithful to H and satisfies that $p(\mathbf{U}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$. As proven in the paragraph above, such a probability distribution exists. Fix any $\mathbf{w}$ and let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. Then, $\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}$

in $p(\mathbf{U}|\mathbf{W} = \mathbf{w})$ iff $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{ZW}$ 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.

The proof for Gaussian probability distributions is analogous. Theorem 2 is not needed in the proof because, as discussed in Section 3, any Gaussian probability distribution $p(\mathbf{U}, \mathbf{W})$ satisfies that $p(\mathbf{U}|\mathbf{W} = \mathbf{w})$ has the same independencies for all $\mathbf{w}$. □

The theorem above has previously been proven for multinomial probability distributions in (Geiger and Pearl, 1993), but the proof constrains the number of states of the random variables in $\mathbf{U}$. Our proof does not constraint such a number and applies not only to multinomial but also to Gaussian probability distributions. It has been proven in (Frydenberg, 1990) that *sep* in an UG G is complete in the sense that it identifies all the independencies holding in every Gaussian probability distribution for which G is an independence map. Our result is stronger than this because it proves the existence of a Gaussian probability distribution with exactly these independencies.

The theorem above proves that *sep* in the MUI map G of a WT graphoid p is complete in the sense that it identifies all the independencies in p that can be identified from G. However, *sep* in G is not complete in the sense that it does not identify all the independencies in p. Actually, no sound criterion for reading independencies from G alone is complete in the latter sense. An example follows.

Example 1. Let p be a multinomial probability distribution that is faithful to the DAG $X \to Z \leftarrow Y$. 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 multinomial probability distribution q that is faithful to G. 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 independency does not hold in q, and it is impossible to know whether we are dealing with p or q on the sole basis of G.

5 Reading Dependencies

In (Bouckaert, 1995), the following sound and complete criterion for reading dependencies from the MUI map of a graphoid is introduced: Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$, then $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ when there exist some $X_1 \in \mathbf{X}$ and $X_2 \in \mathbf{Y}$ such that $X_1 \in MB(X_2)$ and either $MB(X_1) \setminus X_2 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_2)\mathbf{Z}$ or $MB(X_2) \setminus X_1 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_2)\mathbf{Z}$.

In this section, we propose a sound and complete criterion for reading dependencies from the MUI map of a WT graphoid. As in (Bouckaert, 1995), we define the dependence base of an independence model p as the set of dependencies $X \not\perp\!\!\!\perp Y|MB(X) \setminus Y$ for all $X, Y \in \mathbf{U}$ such that $Y \in MB(X)$. Therefore, the dependence base of p is a subset of the complement of p, and it suffices to construct the MUI map of p. If p is a WT graphoid, then additional dependencies can be derived from its dependence base via the WT graphoid properties. All such dependencies together with the dependence base of p are called the WT graphoid closure of the dependence base of p.

As done in (Bouckaert, 1995) with the graphoid properties, we rephrase the WT graphoid properties as follows in order to derive new dependencies holding in a WT graphoid p. 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 not useful 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 $\mathbf{X} \not\perp\!\!\!\perp V|\mathbf{Z} \wedge V \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}V \Rightarrow \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$. As in (Bouckaert, 1995), the independency in the left-hand side of any of the properties above holds if the corresponding *sep* statement holds in the MUI map G of p. This is the best solution we can hope for because, as discussed in Section 4, *sep* in G is sound and complete. Moreover, note that this solution does not require more information about p than what it is available, because G can be constructed from the dependence base of p.

We now introduce our criterion for reading dependencies from the MUI map of a WT graphoid. Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ denote three mutually disjoint subsets of $\mathbf{U}$. Then, $con(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ denotes 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 $(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$ blocks all the paths in G between X_1 and X_n except one. Note that there may exist several unblocked paths in G between $\mathbf{X}$ and $\mathbf{Y}$ but only one between X_1 and X_n. We now prove that *con* is sound for reading dependencies from the MUI map of a WT graphoid, i.e. it only identifies dependencies in the WT graphoid. We first introduce a result that is needed in the proof. Hereinafter, $X_{1:n}$ denotes a path $X_1, \ldots, X_n$ in an UG.

Theorem 4. *Let p be a WT graphoid and G its MUI map. If $X_{1:n}$ is the only path in G between X_1 and X_n that is not blocked by $\mathbf{U} \setminus X_{1:n}$, then $X_1 \not\perp\!\!\!\perp X_n|\mathbf{U} \setminus X_{1:n}$.*

Proof. Note that

$$X_i \not\perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_j) \quad (1)$$

iff X_i and X_j are consecutive in the sequence $X_1, \ldots, X_n$. This proves the theorem for $n = 2$. We now prove it for $n > 2$. We start by proving $X_i \not\perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k)$ for all X_i and X_j that are consecutive in the sequence $X_1, \ldots, X_n$. Let us assume $i < j < k$. The proof is analogous for $k < i < j$. By equation (1)

$$X_i \not\perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_j) \quad (2)$$

and

$$X_i \perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_k). \quad (3)$$

Let us assume to the contrary

$$X_i \perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k). \quad (4)$$

Then, $X_i \perp\!\!\!\perp X_jX_k|\mathbf{U} \setminus (X_iX_jX_k)$ due to contraction on equations (3) and (4) and, thus, $X_i \perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_j)$ due to weak union. This contradicts equation (2) and, thus,

$$X_i \not\perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k). \quad (5)$$

We now prove $X_i \not\perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_jX_k)$ for all X_i, X_j and X_k that are consecutive in the sequence $X_1, \ldots, X_n$. By equation (5), $X_i \not\perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k)$ and $X_j \not\perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_jX_k)$ and, thus, $X_i \not\perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_jX_k)$ or $X_i \not\perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_k)$ due to weak transitivity. Since the latter contradicts equation (1), we conclude

$$X_i \not\perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_iX_jX_k). \quad (6)$$

Finally, we prove $X_i \perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k)$ for all X_i, X_j and X_k such that neither the first two nor the last two are consecutive in the sequence $X_1, \ldots, X_n$. By equation (1), $X_i \perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_j)$ and $X_j \perp\!\!\!\perp X_k|\mathbf{U} \setminus (X_jX_k)$. Then,

$$X_i \perp\!\!\!\perp X_j|\mathbf{U} \setminus (X_iX_jX_k) \quad (7)$$

due to intersection and decomposition.

It can be seen from equations (5), (6) and (7) that the sequence $X_1, X_3, \ldots, X_n$ satisfies equation (1) replacing $\mathbf{U}$ by $\mathbf{U} \setminus X_2$: Equations (5) and (6) ensure that every two consecutive nodes are dependent, while equation (7) ensures that every two non-consecutive nodes are independent. Therefore, we can repeat the calculations above for the sequence $X_1, X_3, \ldots, X_n$ replacing $\mathbf{U}$ by $\mathbf{U} \setminus X_2$. This allows us to successively remove the nodes $X_2, \ldots, X_{n-1}$ from the sequence $X_1, \ldots, X_n$ and conclude that the sequence X_1, X_n satisfies equation (1) replacing $\mathbf{U}$ by $\mathbf{U} \setminus \{X_2, \ldots, X_{n-1}\}$. Then, $X_1 \not\perp\!\!\!\perp X_n|\mathbf{U} \setminus \{X_1, \ldots, X_n\}$. □

Theorem 5. *Let p be a WT graphoid and G its MUI map. Then, con in G only identifies dependencies in p.*

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 $X_1 \not\perp\!\!\!\perp X_n|\mathbf{Y}$. 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. Let us assume to the contrary $X_1 \perp\!\!\!\perp X_2|\mathbf{Y}$. Moreover, $\mathbf{W} \perp\!\!\!\perp X_2|X_1\mathbf{Y}$ because otherwise there exist several unblocked paths in G between X_1 and X_2, which contradicts the definition of $\mathbf{Y}$. The last two independencies imply $X_1\mathbf{W} \perp\!\!\!\perp X_2|\mathbf{Y}$ due to contraction. Moreover, $X_1\mathbf{W} \perp\!\!\!\perp \mathbf{U} \setminus X_{1:2} \setminus \mathbf{Y} \setminus \mathbf{W}|X_2\mathbf{Y}$ by definition of $\mathbf{W}$. The last two independencies imply $X_1 \perp\!\!\!\perp X_2|\mathbf{U} \setminus X_{1:2}$ due to contraction and weak union. This contradicts Theorem 4 and, thus, $X_1 \not\perp\!\!\!\perp X_2|\mathbf{Y}$. Let us assume as induction hypothesis that the theorem holds for all $n < m$. We now prove it for $n = m$. Let us assume to the contrary $X_1 \perp\!\!\!\perp X_m|\mathbf{Y}$. Since $X_{1:m}$ is the only path in G between X_1 and X_m that is not blocked by $\mathbf{Y}$, then $X_1 \perp\!\!\!\perp X_m|X_2\mathbf{Y}$. The last two independencies imply either $X_1 \perp\!\!\!\perp X_2|\mathbf{Y}$ or $X_2 \perp\!\!\!\perp X_m|\mathbf{Y}$ due to weak transitivity. Both cases contradict the induction hypothesis and, thus, $X_1 \not\perp\!\!\!\perp X_m|\mathbf{Y}$. Note that the induction hypothesis applies to both cases because (i) the paths $X_{1:2}$ and $X_{2:m}$ are shorter than m, and (ii) $\mathbf{Y}$ 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 $\mathbf{Y}$.

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 $X_1 \not\perp\!\!\!\perp X_n|(\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_n)\mathbf{Z}$ due to the paragraph above and, thus, $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$ due to weak union. Then, every *con* statement in G corresponds to a dependency in p. □

We now prove that *con* is complete for reading dependencies from the MUI map of a WT graphoid p, in the sense that it identifies all the dependencies in p that follow from the information about p that is available, namely the dependence base of p and the fact that p is a WT graphoid.

Theorem 6. *Let p be a WT graphoid and G its MUI map. Then, con in G identifies all the dependencies in the WT graphoid closure of the dependence base of p.*

Proof. It suffices to prove (i) that all the dependencies in the dependence base of p are identified by *con* in G, and (ii) that *con* satisfies the WT graphoid properties. Let $X \not\perp\!\!\!\perp Y|MB(X)\setminus Y$ with $X, Y \in \mathbf{U}$ and $Y \in MB(X)$ be a dependency in the dependence base of p. Then, X and Y are adjacent in G and all the other paths in G between them are blocked by $MB(X)\setminus Y$. Then, $con(X, Y|MB(X)\setminus Y)$ holds in G.

We now prove that *con* satisfies the WT graphoid properties. Recall that the independency in the left-hand side of any of the properties corresponds to a *sep* statement in G. 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 the left-hand side. If $\mathbf{W}$ contains some node in $X_{1:n}$, then let $X_m \in X_{1:n}$ denote the closest node to X_1 such that $X_m \in \mathbf{W}$. Then, the path $X_{1:m}$ satisfies the right-hand side 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 the left-hand side.

- 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 the left-hand side.

- 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 (i) the path $X_{1:n}$ in the left-hand side must be between $\mathbf{X}$ and $\mathbf{W}$, and (ii) all the paths in G between X_1 and X_n that are blocked by $\mathbf{Y}$ are also blocked by $(\mathbf{W} \setminus X_n)\mathbf{Z}$ and, thus, $\mathbf{Y}$ is not needed to block all the paths in G between X_1 and X_n except $X_{1:n}$. Then, $X_{1:n}$ satisfies the right-hand side.

- 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 the left-hand side must be between $\mathbf{X}$ and $\mathbf{Y}$ and, thus, it satisfies the right-hand side.

- 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 the left-hand side must be between $\mathbf{X}$ and $\mathbf{W}$ and, thus, it is satisfies the right-hand side.

- 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 the first and second, respectively, *con* statements in the left-hand side. Let $X_{1:m:n}$ denote the path $X_1, \ldots, X_m, \ldots, X_n$. Then, the path $X_{1:m:n}$ satisfies the right-hand side because (i) $\mathbf{Z}$ does not block $X_{1:m:n}$, and (ii) $\mathbf{Z}$ blocks all the other paths in G between X_1 and X_n, because otherwise there exist several unblocked paths in G between X_1 and X_m or between X_m and X_n, which contradicts the left-hand side.

- 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 $X_m \in \mathbf{U} \setminus (\mathbf{XYZ})$. This property never applies because, as seen in weak transitivity2, $sep(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ never holds since $\mathbf{Z}$ does not block $X_{1:m:n}$.

□

It is worth mentioning that *con* when $n = 2$ and either $MB(X_1) \setminus X_2 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_2)\mathbf{Z}$ or $MB(X_2) \setminus X_1 \subseteq (\mathbf{X} \setminus X_1)(\mathbf{Y} \setminus X_2)\mathbf{Z}$ coincides with the criterion in (Bouckaert, 1995), which is sound and complete (in the same sense as *con*) for reading dependencies from the MUI map of a graphoid. Therefore, our criterion allows reading more dependencies than the criterion in (Bouckaert, 1995) at the cost of assuming weak transitivity which, as discussed in Section 3, is not a too restrictive assumption.

Note that the meaning of completeness in the theorem above is different from that in Theorem 3. It remains an open question whether *con* in G identifies all the dependencies in p that can be identified from G. Note also that *con* in G is not complete in the sense that it does not identify all the dependencies in p. Actually, no sound criterion for reading dependencies from G alone is complete in this sense. Example 1 illustrates this point. Let us now assume that we are dealing with q instead of with p. Then, no sound criterion can conclude $X \not\perp Y|\emptyset$ by just studying G because this dependency does not hold in p, and it is impossible to know whether we are dealing with p or q on the sole basis of G.

The theorem below follows from Theorems 5 and 6 by noting that the proof of the former only uses the dependence base of p and the WT graphoid properties.

Theorem 7. *Let p be a WT graphoid and G its MUI map. Then, con in G identifies all and only the dependencies in the WT graphoid closure of the dependence base of p.*

Finally, note that, following (Bouckaert, 1995), we have defined the dependence base of a WT graphoid p as the set of dependencies $X \not\perp Y|MB(X) \setminus Y$ for all $X, Y \in \mathbf{U}$ such that $Y \in MB(X)$. However, Theorems 5-7 are also valid if we define the dependence base of p as the set of dependencies $X \not\perp Y|\mathbf{U} \setminus (XY)$ for all $X, Y \in \mathbf{U}$. Actually, the theorems are valid as long as the dependence base of p is a set of dependencies such that (i) it suffices to construct the MUI map G of p, and (ii) all the dependencies in the set are identified by *con* in G.

6 Discussion

In this paper, we have introduced a sound and complete criterion for reading dependencies from the MUI map of a WT graphoid. In (Peña et al., 2006), we show how this helps to identify all the nodes that are relevant to compute all the conditional probability distributions for a given set of nodes without having to learn a BN first. We are currently working on a sound and complete criterion for reading dependencies from a minimal directed independence map, which is a more popular model than its undirected counterpart. Our end-goal is to apply these theoretical results in 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, it is a WT graphoid. 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 Bayesian network is a rather accurate model of the cell (Friedman et al., 1998; Murphy and Mian, 1999): 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. It is important that the Bayesian network 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 are a sample of the probability distribution represented by the dynamic Bayesian network after marginalizing the protein nodes out. The probability distribution with no node marginalized out is almost surely faithful to the dynamic Bayesian network (Meek, 1995) and, thus, it satisfies weak transitivity (see Section 2) and, thus, so does the probability distribution after marginalizing the protein nodes out (see Theorem 1). 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).

Acknowledgments

This work is funded by the Swedish Research Council (VR-621-2005-4202), the Swedish Foundation for Strategic Research, and Linköping Institute of Technology.

References

Theodore W. Anderson. 1984. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons.

Remco R. Bouckaert. 1995. *Bayesian Belief Networks: From Construction to Inference*. PhD Thesis, University of Utrecht.

Nir Friedman, Kevin Murphy and Stuart Russell. 1998. Learning the Structure of Dynamic Probabilistic Networks. In *14th Conference on Uncertainty in Artificial Intelligence*, pages 139–147.

Morten Frydenberg. 1990. Marginalization and Collapsability in Graphical Interaction Models. *Annals of Statistics*, 18:790–805.

Dan Geiger and Judea Pearl. 1993. Logical and Algorithmic Properties of Conditional Independence and Graphical Models. *The Annals of Statistics*, 21:2001–2021.

Christopher Meek. 1995. Strong Completeness and Faithfulness in Bayesian Networks. In *11th Conference on Uncertainty in Artificial Intelligence*, pages 411–418.

Kevin Murphy and Saira Mian. 1999. *Modelling Gene Expression Data Using Dynamic Bayesian Networks*. Technical Report, University of California.

Masashi Okamoto. 1973. Distinctness of the Eigenvalues of a Quadratic Form in a Multivariate Sample. *Annals of Statistics*, 1:763–765.

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

Jose M. Peña, Roland Nilsson, Johan Björkegren and Jesper Tegnér. 2006. Identifying the Relevant Nodes Without Learning the Model. In *22nd Conference on Uncertainty in Artificial Intelligence*.

Paola Sebastiani, Emanuela Gussoni, Isaac S. Kohane and Marco F. Ramoni. 2003. Statistical Challenges in Functional Genomics (with Discussion). *Statistical Science*, 18:33–60.

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