Constraint-based Approach to Discovery of Inter Module Dependencies in Modular Bayesian Networks

Patrick de Oude
Intelligent Systems Laboratory Amsterdam
University of Amsterdam
The Netherlands

Gregor Pavlin
Thales Research & Technology Netherlands
Delftechpark 24, 2628 XH, Delft
The Netherlands

Abstract

This paper introduces an information theoretic approach to verification of modular causal probabilistic models. We assume systems which are gradually extended by adding new functional modules, each having a limited domain knowledge captured by a local Bayesian network. Different modules originate from independent design processes. We assume that the local models are correct, which, however does not guarantee globally coherent inference in composed systems. The introduced method supports discovery of significant inter module dependencies which are ignored in the assembled Bayesian network.

Introduction

Relevant classes of contemporary information processing challenges have been addressed by modular approaches to Bayesian modeling and inference. For example, Distributed perception networks (DPN) (Pavlin et al. 2008) support robust information fusion in domains where constellations of information sources are not known prior to the operation and can change frequently at runtime. The multi-agent MSBN (Xiang 2002), on the other hand, is particularly suitable for the diagnosis of complex systems consisting of many components, such as electronic circuits, chemical processes, etc. Both, the DPNs and multi-agent MSBNs support efficient distribution of models and inference processes over multiple networked devices. In addition, Network Fragments approach (Laskey and Mahoney 1997) supports creation of *complex monolithic* Bayesian networks out of objects representing simpler Bayesian networks.

All approaches to modular Bayesian modeling and inference exploit the locality of causal relations, which supports efficient design and inference. Typically, different modules originate from independent design processes. Since the corresponding local models are relatively small, we can assume that they correctly capture relations between the local variables. However, correctness of local models does not guarantee globally coherent inference in composed systems. Designing such systems is a multidisciplinary field where it is likely that important dependencies between variables in different modules are overlooked by the designers, which can have negative impact on the inference quality. In other words, despite correct local models the system does not support globally correct inference. Such modeling faults cannot be avoided even if advanced engineering approaches are used.

In this paper we introduce an information theoretic approach which is based on constraint-based structure discovery (Spirtes, Glymour, and Scheines 2000, Pearl 2000) and supports efficient verification of dependencies in modular systems. Note that the proofs for all propositions in this paper are given in (de Oude and Pavlin 2009). We assume that domain models are gradually constructed from modules defined through simple Bayesian networks, each correctly capturing all relations between the local variables. At each addition of a new module, the independence tests are executed between the variables of the new module and the rest of the already integrated modules. By considering the properties of *I-maps* (Pearl 1988), we can efficiently discover inadequately represented dependencies between the variables from different modules. In particular, we exploit the composition/decomposition and weak union relations which are valid in case of faithful probability distributions (Spirtes, Glymour, and Scheines 2000). We use a running example to illustrate the challenges and principles of our approach.

Causal Probabilistic Models

Often we are interested in phenomena which materialize through causal stochastic processes. Some of the phenomena influenced by a causal process can be observed while others remain hidden. By understanding the underlying causal mechanisms and by considering the observations, the hidden phenomena can be inferred. For example, by observing certain reports from chemical sensors and humans we could infer the presence of a harmful gas. Figure 1 shows a simplified representation of a causal process producing reports from chemical sensors and humans exposed to a particular gaseous substance. Directed links represent direct causal dependencies between the phenomena. The existence of $GasX$ causes certain conditions in the air $Cond$ which will trigger processes in two different sensors producing reports R_1 and R_2, respectively. Similarly, given the presence of $GasX$, the exposed people will perceive a certain smell and develop certain symptoms. More about modeling of monitoring processes can be found in (de Oude and Pavlin 2009,

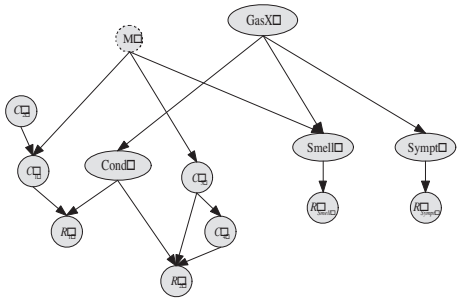

Figure 1: Causal model of a gas monitoring process.

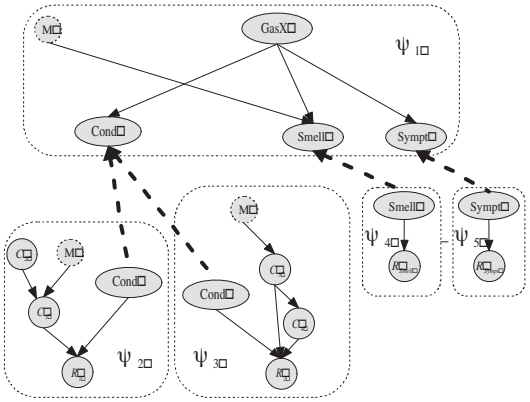

Figure 2: Modular model of the gas monitoring process in Figure 1. Dashed arrows denote the message flow between the collaborating modules.

de Oude and Pavlin 2008).

Stochastic causal processes can be modeled with the help of Bayesian networks in a rigorous and compact way (Pearl 1988). Each Bayesian network (BN) represents a joint probability distribution (JPD) $P(\mathbf{V})$ over a set of random variables defined in $\mathbf{V}$ (Pearl 1988). A BN is represented by a tuple $(G, \mathbf{P})$, where $G = (\mathbf{V}, \mathbf{E})$ is a Directed Acyclic Graph (DAG) with the variables $\mathbf{V}$ represented as nodes and directed edges $E = (V_i, V_j)$ between nodes in $\mathbf{V}$, where $E \in \mathbf{E}$. $\mathbf{P}$ is the set of conditional probabilities $P(V_i|\mathbf{Pa}(V_i))$ for all $V_i \in \mathbf{V}$, where $\mathbf{Pa}(V_i)$ represents the parent nodes of V_i in DAG G. Conditional probabilities are represented through conditional probability tables (CPTs).

BNs explicitly capture *conditional independence* between random variables. A variable X is conditionally independent of variable Y given variable Z when $P(X|Y,Z) = P(X|Z)$. Conditional independences are represented through *d-separation* (Xiang 2002, Pearl 1988) relations in DAG G.

The representational explicitness of conditional independence in BNs means that every represented d-separation in a DAG G should have a *valid* corresponding conditional independence relation in $P(\mathbf{V})$. Consequently, the following implication should hold:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_G \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_P \quad (1)$$

Whenever the implication in (1) for DAG G holds we say that G is an *I-map* of $P(\mathbf{V})$ (Pearl 1988).

Note that, the implication in (1) allows us to model superfluous dependencies between nodes in model G which are not true in $P(\mathbf{V})$. Ideally, no superfluous dependencies are captured by the model. In other words, every conditional independence in P corresponds to a certain d-separation in G and vice versa:

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_G \Leftrightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_P \quad (2)$$

Whenever Equation (2) holds we say that $P(\mathbf{V})$ is *faithful* (Spirtes, Glymour, and Scheines 2000) to DAG G (or G is a perfect map (Pearl 1988) of P). If a probability distribution is faithful then there exist a DAG G for it that faithfully represents all (conditional) dependencies and independences between the variables in this probability distribution. In this paper we only assume faithful probability distributions.

Parameters in a BN (i.e. conditional probabilities) can be estimated with the help of the maximum likelihood principle by considering relative frequencies in some data set $\mathcal{D}$ of data sample size m. A single data sample $d_i \in \mathcal{D}$, $i \leq m$ is represented as a *configuration of states* $\mathbf{v} = (x, y, \ldots, z)$ for the domain variables $\mathbf{V} = \{X, Y, \ldots, Z\}$.

In this paper we assume that the data samples $d_i = \mathbf{v}$ are sampled from a generative probability distribution $P(\mathbf{V})$ which can faithfully be represented by a BN with DAG G.

bsectionTesting conditional independence

Let's assume that a BN models a distribution $P(\mathbf{V})$ from which data set $\mathcal{D}$ was sampled. Moreover, we can use the conditional mutual information (CMI) measure between mutually exclusive sets of variables $\mathbf{X} \subset \mathbf{V}$, $\mathbf{Y} \subset \mathbf{V}$ and $\mathbf{Z} \subset \mathbf{V}$ to test conditional independence $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_P$:

$$I(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) = \sum_{\mathbf{x},\mathbf{y},\mathbf{z}} P(\mathbf{x}, \mathbf{y}, \mathbf{z}) \log \frac{P(\mathbf{x}, \mathbf{y}|\mathbf{z})}{P(\mathbf{x}|\mathbf{z})P(\mathbf{y}|\mathbf{z})} \quad (3)$$

Note that the probabilities $P(\mathbf{x}, \mathbf{y}|\mathbf{z})$, $P(\mathbf{x}|\mathbf{z})$ and $P(\mathbf{y}|\mathbf{z})$ in Equation 3 are estimated using data samples in $\mathcal{D}$.

By considering (2) we know that a BN correctly captures the true distribution $P(\mathbf{V})$ iff for every $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_G$ the mutual information $I(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) = 0$ and vice versa. *In other words, by using the CMI measure we can use data samples from $\mathcal{D}$ to test I-mapness of causal probabilistic models.* Instead of CMI measure also other measures can be used, such as χ^2 or G^2 independence test.

Modular Bayesian Network

The causal model of the gas monitoring process given in Figure 1 can be represented by a system of collaborating modules implementing smaller BNs shown in Figure 2. Local BNs can, for example, be constructed by using the design rules described in (Pavlin et al. 2008). The modules are defined as follows:

Definition 1 (BN Module). *A* **BN module** *$\psi_i = (G_i, \mathbf{P}_i)$ is a Bayesian network with DAG $G_i = (\mathbf{V}_i, \mathbf{E}_i)$, where $\mathbf{V}_i$ are the variables in G_i and $\mathbf{E}_i$ is a set of directed edges $E = (X, Y)$, where $E \in \mathbf{E}_i$, $X \in \mathbf{V}_i$ and $Y \in \mathbf{V}_i$. $\mathbf{P}_i$ is a set of (conditional) probability distributions defined for each variable in $\mathbf{V}_i$.*

A BN module encodes probabilistic knowledge over a subset of variables $\mathbf{V}_i \subset \mathbf{U}$, where $\mathbf{U}$ represents all variables in the domain under investigation. Therefore, a BN module is a subgraph of the monolithic BN.

In this paper we assume that every BN module is a *local I-map* over the variables $\mathbf{V}_i$:

Definition 2 (Local I-mapness). *A BN module $\psi_i = (G_i, \mathbf{P}_i)$ where DAG $G_i = (\mathbf{V}_i, \mathbf{E}_i)$ is a* **local I-map** *of $P(\mathbf{V}_i) = \sum_{\mathbf{U}\setminus\mathbf{V}_i} P(\mathbf{U})$, where $\mathbf{V}_i \subset \mathbf{U}$ and $\mathbf{U}$ contains all the variables in the domain, if all d-separation relations between the variables in $\mathbf{V}_i$ correspond to valid conditional independencies in $P(\mathbf{V}_i)$.*

For example, ψ_1 in Figure 2 has the local I-map property, because all the represented d-separations, such as $(Cond \perp\!\!\!\perp Smell|GasX)_G$, $(\{GasX, Sympt\} \perp\!\!\!\perp M|\emptyset)_G$, $(Cond \perp\!\!\!\perp Sympt|\{GasX, M, Smell\})_G$ etc. correspond to valid conditional independencies in $P(\mathbf{V}_1)$ captured by the ground truth model in Figure 1. Moreover, BN modules can form a *modular Bayesian network*:

Definition 3 (Modular Bayesian networks). *A* **Modular Bayesian network** *Ω is defined as a tuple $(\mathcal{M}, \mathcal{R})$, where $\mathcal{M}$ is the set of BN modules defined in Ω. $\mathcal{R}$ is a finite set of BN module pairs $\{\langle\psi_i, \psi_j\rangle | i \geq 1, j \geq 1, i \neq j\}$ with $\psi_i = ((\mathbf{V}_i, \mathbf{E}_i), \mathbf{P}_i) \in \mathcal{M}$, $\psi_j = ((\mathbf{V}_j, \mathbf{E}_j), \mathbf{P}_j) \in \mathcal{M}$ and $\mathbf{V}_i \cap \mathbf{V}_j \neq \emptyset$. Every module pair $\langle\psi_i, \psi_j\rangle$ represents a link between the two BN modules ψ_i and ψ_j over which partial reasoning results can be shared. A modular BN must satisfy a set of properties (see de Oude and Pavlin 2009).*

For example, the monolithic gas detection fusion model shown in Figure 1 is partitioned into five different BN modules: ψ_1, ψ_2, ψ_3, ψ_4 and ψ_5 in Figure 2. The modular BN is then defined as $\mathcal{M} = \{\psi_1, \psi_2, \psi_3, \psi_4, \psi_5\}$ and $\mathcal{R} = \{\langle\psi_1, \psi_2\rangle, \langle\psi_1, \psi_3\rangle, \langle\psi_1, \psi_4\rangle, \langle\psi_1, \psi_5\rangle\}$.

Reasoning in the modular BN in Figure 2 is equivalent to reasoning in the monolithic BN in Figure 1. I.e. we can map the structure of the modular BN Ω to its corresponding monolithic graph version, which we call G^Ω. The graph in Figure 1 is the G^Ω of the modular BN in Figure 2 (for readability we will write G instead of G^Ω in the rest of the paper). As next we define the *global I-mapness* property of modular networks:

Definition 4 (Global I-mapness). *Let a modular BN Ω correspond to a monolithic DAG $G = (\mathbf{V}, \mathbf{E})$. The modular BN Ω is a* **global I-map** *if all d-separations represented in G correspond to valid conditional independencies in the true distribution $P(\mathbf{V})$.*

Constructing modular Bayesian networks

We assume that the construction of a modular BN $\Omega = (\mathcal{M}, \mathcal{R})$ is accomplished through a certain integration order of BN modules. In each *extension step* a new BN module ψ_i is added to $\Omega = (\mathcal{M}, \mathcal{R})$, which is denoted by $\Omega^* \leftarrow \Omega \sqcup \psi_i$. Module ψ_i is connected to another BN module $\psi_j \in \mathcal{M}$ to form the new modular BN $\Omega^* = (\mathcal{M} \cup \psi_i, \mathcal{R} \cup \langle\psi_j, \psi_i\rangle)$.

A modular BN Ω obtained through a sequence of additions of BN modules might correspond to a monolithic graph G which violates the global I-map property; i.e. the d-separations in graph G violate certain dependencies in the data. Consider a modular causal model in Figure 3a, where two BN modules ψ_1 and ψ_2 describe a part of the monitoring causal process. Note that ψ_2 from Figure 2 does not contain variable M. Thus, in this modular BN the direct dependence between humidity M and the sensor component C_1 is missing. This dependency is captured in the true model in Figure 1. Both BN modules in Figure 3a are local I-maps, but when connected together the global I-map property does not hold anymore. For example, say we want to extend $\Omega = (\mathcal{M} = \{\psi_1\}, \mathcal{R} = \emptyset)$ with the BN module ψ_2. In this case, ψ_2 is connected to the already integrated $\psi_1 \in \mathcal{M}$ resulting in $\Omega^* = (\mathcal{M} = \{\psi_1, \psi_2\}, \mathcal{R} = \langle\psi_2, \psi_1\rangle)$. The corresponding monolithic graph G^* is not an I-map anymore, because it encodes the d-separations $(M \perp\!\!\!\perp C_1|Cond)_{G^*}$, $(M \perp\!\!\!\perp R_1|Cond)_{G^*}$, $(Smell \perp\!\!\!\perp C_1|Cond)_{G^*}$ and $(Smell \perp\!\!\!\perp R_1|Cond)_{G^*}$. These d-separations are not valid in the ground truth model shown in Figure 1. Consequently, the implication in (1) does not hold and graph G^* is not an I-map. Thus, the modular BN Ω^* does not support correct reasoning. Namely, the dependence between M and C_1 is not captured by the BN Ω^* in Figure 3a correctly.

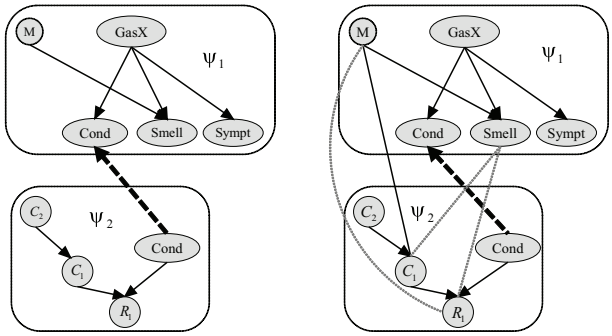

Figure 3: (a) Extending $\Omega = (\mathcal{M} = \{\psi_1\}, \mathcal{R} = \emptyset)$ with BN module ψ_2 to form the modular BN $\Omega^* = (\mathcal{M} = \{\psi_1, \psi_2\}, \mathcal{R} = \langle\psi_2, \psi_1\rangle)$. (b) Variable pairs $\mathcal{E}$ which failed the pairwise CI test are added as undirected edges to modular BN Ω^*, where the superfluous edges are shown as light gray dotted undirected edges.

Important Observation: *given that all BN modules have the local I-map property does not, in general, imply that the modular BN constructed out of these BN modules is a global I-map.*

Verification of Global I-mapness in Modular Bayesian Networks

The CMI measure (3) can be used to test global I-mapness of a modular BN; i.e. we verify whether a modular BN Ω corresponds to a monolithic BN which is an I-map of the variables contained in Ω. We facilitate further analysis by introducing the concept of *pairwise Conditional Independence* (pairwise CI) between the variables from different modules in a modular BN.

Definition 5 (Pairwise CI). *Let's assume a modular BN Ω*

with a set of modules $\mathcal{M}$ defined over variables $\mathbf{V}$ which is extended by adding a new module ψ_i with variables $\mathbf{V}_i$, where $\mathbf{V}_i \not\subseteq \mathbf{V}$. In addition, Ω contains module $\psi_j \in \mathcal{M}$ with variables $\mathbf{V}_j$, such that $\mathcal{S} = \mathbf{V}_i \cap \mathbf{V}_j \neq \emptyset$. For any pair of variables $X \in \{\mathbf{V}_i \backslash \mathcal{S}\}$ and $Y \in \{\mathbf{V} \backslash \mathcal{S}\}$ a **pairwise conditional independence (CI)** *is defined as $(X \perp\!\!\!\perp Y | \mathcal{S})_P$.*

Note that expression $(X \perp\!\!\!\perp Y|\mathcal{S})_P$ can be tested with the help of the CMI measure in Equation (3). We illustrate pairwise CI with the help of the modular BN in Figure 2. Let's assume that we wish to extend modular BN $\Omega = (\mathcal{M}, \mathcal{R})$, where $\mathcal{M} = \{\psi_1, \psi_3, \psi_4, \psi_5\}$ and $\mathcal{R} = \{\langle\psi_1, \psi_3\rangle, \langle\psi_1, \psi_4\rangle, \langle\psi_1, \psi_5\rangle\}$), with BN module ψ_2. In this case we could identify intersection set $\mathcal{S} = \mathbf{V}_1 \cap \mathbf{V}_2 = \{Cond, M\}$. Given that the extended modular BN corresponded to the ground truth model shown in Figure 1, we could detect several pairwise CIs between variables from different modules, such as for example $(C_1 \perp\!\!\!\perp R_{Smell}|\mathcal{S})_P$, $(R_1 \perp\!\!\!\perp R_{Smell}|\mathcal{S})_P$, $(C_1 \perp\!\!\!\perp Smell|\mathcal{S})_P$, etc. Moreover, it turns out that Pairwise CI tests can be used for efficient verification of global I-mapness of modular BNs.

Proposition 1 (Pairwise CIs & I-mapness). *Assume a modular BN $\Omega = (\mathcal{M}, \mathcal{R})$ with the corresponding global I-map $G = (\mathbf{V}, \mathbf{E})$ and a BN module ψ_i with the local I-map $G_i = (\mathbf{V}_i, \mathbf{E}_i)$. Also assume that the JPD $P(\mathbf{V} \cup \mathbf{V}_i)$ is faithful. By extending Ω with ψ_i we get the modular BN Ω^*, i.e. $\Omega^* \leftarrow \Omega \sqcup \psi_i$ where Ω^* corresponds to the monolithic graph G^*. G^* is an global I-map if and only if all pairwise CIs of the sets $\{\mathbf{V} \backslash \mathcal{S}\}$ and $\{\mathbf{V}_i \backslash \mathcal{S}\}$ are valid, where $\mathcal{S} = \mathbf{V} \cap \mathbf{V}_i$. The proof of this proposition is based on the composition/decomposition axiom (Pearl 1988).*

In other words, verification based on pairwise CIs is sufficient for determining whether an extended modular BN corresponds to a global I-map.

Discovery of Missing Inter Module Dependencies

Verification of global I-mapness of modular BNs can be used as a basis for discovery of modeling faults. Whenever one of these pairwise CIs fail then we know that the modular BN is incorrect and, consequently, certain dependencies between BN modules are missing. In fact, with the help of pairwise CI tests we can discover all pairs of variables whose dependency is not correctly captured in the modular BN.

Proposition 2 (Edge Discovery). *Assume that we extend $\Omega = (\mathcal{M}, \mathcal{R})$ corresponding to global I-map $G = (\mathbf{V}, \mathbf{E})$ with module ψ_i corresponding to the local I-map $G_i = (\mathbf{V}_i, \mathbf{E}_i)$ to form the new modular BN Ω^* corresponding to graph G^*, i.e. $\Omega^* \leftarrow \Omega \sqcup \psi_i$. Moreover, the pairwise CIs for the sets $\mathbf{V}$ of Ω and $\mathbf{V}_i$ of ψ_i are tested using the conditioning set $\mathcal{S} = \mathbf{V}_i \cap \mathbf{V} \neq \emptyset$. We test each possible pair of variables. For every test failure between a pair of tested variables $X \in \mathbf{V}_i \backslash \mathcal{S}$ and $Y \in \mathbf{V} \backslash \mathcal{S}$ we add the pair (X, Y) to a set of pairs $\mathcal{E}$. The set $\mathcal{E}$ is guaranteed to contain at least all the pairs of variables which are directly dependent in the true distribution $P(\mathbf{V} \cup \mathbf{V}_i)$ but this dependence is not correctly captured by the corresponding modular BN.*

Pairs of variables correspond to edges in a graph, therefore the aforementioned procedure is called edge discovery.

Superfluous Edges

The set $\mathcal{E}$ might contain also pairs which are not directly dependent according to $P(\mathbf{V} \cup \mathbf{V}_i)$. Such superfluous edges are not needed to restore global I-mapness and should be removed, since they introduce additional parameters which make the parameter estimation and inference unnecessarily complex.

For example, consider the situation in Figure 3a where $\Omega^* = (\{\psi_1, \psi_2\}, \{\langle\psi_2, \psi_1\rangle\})$ would be obtained through addition of ψ_2 to $\Omega = (\{\psi_1\}, \emptyset)$. Applying the edge discovery described in Proposition 2 will result in the following variable pairs $\mathcal{E} = \{(M, C_1), (M, R_1), (Smell, C_1), (Smell, R_1)\}$, because the four pairwise CIs, which assumed the conditional independencies $(M \perp\!\!\!\perp C_1|Cond)_P$, $(M \perp\!\!\!\perp R_1|Cond)_P$, $(Smell \perp\!\!\!\perp C_1|Cond)_P$ and $(Smell \perp\!\!\!\perp R_1|Cond)_P$, failed the test. The unanticipated dependence between the variables M and C_1 induces dependencies also between variable pairs (M, R_1), $(Smell, R_1)$ and $(Smell, C_1)$ corresponding to superfluous edges shown in Figure 3b (shown as light gray dotted undirected edges). This figure corresponds to a monolithic *pattern* (Pearl 2000) $\mathcal{G}^*_{\mathcal{E}}$ where the set $\mathcal{E}$ represents the undirected edges which are added to the monolithic BN G^*.

The variable pairs in $\mathcal{E}$ found by applying the *pairwise CI* test are often conditionally independent given a conditioning set that differs from the conditioning set used in the pairwise CIs. Therefore, for every variable pair in an undirected edge $\mathcal{E}_i = (X, Y)$, where $\mathcal{E}_i \in \mathcal{E}$, we try to find a set $\mathcal{S}$ of variables that renders X and Y conditionally independent. If we find such a set $\mathcal{S}$, then the edge (X, Y) is superfluous and is removed from the pattern $\mathcal{G}^*_{\mathcal{E}}$. It turns out that we can limit the search over the number of variables in the modular BN that can make the variables X and Y conditionally independent. Namely, we can find the *Markov Blanket* (Neapolitan 2003) of X or Y by only considering the pattern $\mathcal{G}^*_{\mathcal{E}}$ and try to find a minimum conditioning set $\mathcal{S}$ within this Markov blanket. We can show that a Markov blanket can be found in a pattern as follows:

Proposition 3 (Markov Blanket in Pattern). *Given a pattern $\mathcal{G} = (\mathbf{V}, \mathbf{E} \cup \mathcal{E})$ with variables $\mathbf{V}$, directed edges $\mathbf{E}$ and undirected edges $\mathcal{E}$. Let at least one BN in the set of BNs represented by this* **pattern** *be an I-map w.r.t. $P(\mathbf{V})$; then a* **Markov blanket** *for a variable $X \in \mathbf{V}$, i.e. $MB(X)$, is the union of the following sets:*

(i) All children of X denoted by $\mathbf{Ch}(X)$. If X is connected to another variable $V_i \in \mathbf{V}$ through an undirected edge $X - V_i$ then V_i is considered a child of X and is included in the set $\mathbf{Ch}(X)$;

(ii) For each $Y_k \in \mathbf{Ch}(X)$ take the parents of Y_k denoted by $\mathbf{Pa}(Y_k)$. If Y_k is connected to another variable $V_j \in \mathbf{V}$ through an undirected edge $Y_k - V_j$ then $V_j \neq X$ is considered a parent of Y_k and is included in the set $\mathbf{Pa}(Y_k)$;

(iii) All parents of X denoted by $\mathbf{Pa}(X)$. A variable Y_k is parent of X if there exist a directed link from Y_k to X.

In short, $MB(X) = \mathbf{Ch}(X) \cup (\bigcup_k \mathbf{Pa}(Y_k)) \cup \mathbf{Pa}(X)$ *and, consequently,* $(X \perp\!\!\!\perp \mathbf{V} \backslash (X \cup MB(X)) | MB(X))_P$ *holds. The proof is based on the weak union axiom (Pearl 1988).*

We can find a Markov blanket in a given pattern $\mathcal{G}^*_{\mathcal{E}}$ if at least one BN in the set of BNs, represented by $\mathcal{G}^*_{\mathcal{E}}$, is an I-map. It can be shown that $\mathcal{G}^*_{\mathcal{E}}$ indeed contains an I-map if $\mathcal{G}^*_{\mathcal{E}}$ is obtained by adding variable pairs $\mathcal{E}$ to G^* following Proposition 2:

Proposition 4 (I-map in Pattern). *Given a monolithic graph* $G^* = (\mathbf{V}, \mathbf{E})$ *corresponding to the modular BN* Ω^* *which is obtained by extending* Ω *with* ψ_i*, i.e.* $\Omega^* \leftarrow \Omega \sqcup \psi_i$*. Let* $\mathcal{G}^*_{\mathcal{E}}$ *be the pattern, where* $\mathcal{E}$ *is a set of variable pairs that failed the pairwise CI tests and are added to* G^* *as undirected edges. If* $\mathcal{E}$ *is obtained following Proposition 2 then* $\mathcal{G}^*_{\mathcal{E}}$ *includes at least one BN that is an I-map w.r.t.* $P(\mathbf{V})$*.*

By considering Proposition 3 and 4 we can eliminate any superfluous undirected edge corresponding to a variable pair $(X, Y) \in \mathcal{E}$ in the pattern $\mathcal{G}^*_{\mathcal{E}}$. We compute the Markov blanket either for X or Y from the pattern $\mathcal{G}^*_{\mathcal{E}\backslash(X,Y)}$ following Proposition 3; i.e. we use the pattern $\mathcal{G}^*_{\mathcal{E}}$ without the edge (X, Y). Whenever X and Y are conditionally independent w.r.t. $P(\mathbf{V})$ then there exists some minimum conditioning set $\mathcal{S} \subseteq MB(X)$ or $\mathcal{S} \subseteq MB(Y)$ that renders X conditionally independent of Y.

Moreover, we can show the following property:

Proposition 5 (Direct Dependence). *Whenever variable X and Y cannot be made conditionally independent on any subset* $\mathcal{S}$ *of the Markov blanket for X or Y then we can conclude that X and Y must be* **directly dependent** *and an edge is justified between X and Y in the modular BN* Ω^**.*

Thus, the Markov blanket provides a maximum conditioning set needed for unambiguous dependency tests. Often it makes sense to test conditional independence between variables X or Y by initially using small conditioning sets $\mathcal{S} \subseteq MB(X)$ and then gradually increase the size of $\mathcal{S}$ until the test indicates independence or $\mathcal{S} = MB(X)$. Smaller conditioning sets $\mathcal{S}$ require lower order statistics and are thus more reliable, given limited data sets.

Dealing with Limited Data Sets

Due to noise in the data the empirically determined $I(\mathbf{X}, \mathbf{Y}|\mathbf{Z})$ will never be exactly 0, even if a finite data set $\mathcal{D}$ is sampled from a distribution for which $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_P$ holds. Therefore, a threshold δ is needed for independence tests. This δ can be used as a decision threshold for deciding if two variables are (conditionally) dependent or independent.

In our approach δ is based on the CMI corresponding to pairs of variables in local models for which we know that are (conditionally) independent because of the local I-map assumption (see Definition 2): the inner BN module d-separations correspond to valid conditional independences in the underlying probability distribution. Therefore we can compute the threshold as follows: $\delta = max(\mathcal{I})$, where $\mathcal{I}$ is a list of computed CMIs based on the valid conditional independencies, represented as d-separations, in local BN modules.

Data Size	Precision gradual	Recall gradual	Precision prior	Recall prior
1000	0.25013	0.10800	0.22152	0.09900
3000	0.72400	0.68100	0.70400	0.66700
10000	0.99833	0.99900	0.99833	0.99900
50000	1.00000	1.00000	1.00000	1.00000

Table 1: Average precision and recall for GDN

Experiments

The discovery algorithm* based on Propositions 1, 2, 3, 4 and 5 was experimentally validated. We experimented with a gas detection network (GDN) and the well known ALARM network. The GDN was similar to the BN used in the running example (see Figure 1), except that it was more complex.

For the GDN network the BN modules were manually created, such that two edges were not correctly captured by the assembled modular BN. On the other hand, for the ALARM network the BN modules were automatically generated from the ALARM network form which a single randomly selected edge was removed. Each modular BN was gradually constructed out of the BN modules and we ran the discovery algorithms for each extension step. Each set of modules was assembled several times in a random order, where for each set of modules a new data set $\mathcal{D}$ of a given size was used. The data $\mathcal{D}$ used in the dependency tests was sampled from the ground truth monolithic GDN and ALARM networks.

The adaptive threshold δ used for dependency tests was computed in two different ways. Therefore for each network and data size two experiments are performed: (i) an experiment where the threshold is recomputed whenever a new local BN module is added to the modular BN by only considering the local conditional independences in the current modular BN (gradual threshold); (ii) the threshold is computed prior to the operation by using all the local conditional independences in all local BN modules before they are assembled into a modular BN (prior threshold).

The performance of the discovery algorithm was measured using the precision and recall of the missing edges. Precision is defined as the fraction of discovered missing edges divided by all discovered edges in $\mathcal{E}$. Recall is the fraction of discovered missing edges divided by the missing edges. For each modular BN precision and recall of the discovery algorithm were determined for different data sizes. For each monolithic network we created several modular BNs and computed the average precision and recall for each data size and threshold type (gradual or prior). Table 1 and Table 2 show the discovery results for the GDN and the ALARM network, respectively.

From the results it is evident that for both networks no significant difference between the gradual and the prior threshold exists. The difference in recall between ALARM and GDN for large data sizes is attributed to the difference in complexity of the networks, i.e. the ALARM network is

*More details can be found in (de Oude and Pavlin 2009).

Data Size	Precision gradual	Recall gradual	Precision prior	Recall prior
1000	0.65976	0.45556	0.65976	0.45556
3000	0.85087	0.65556	0.83937	0.65556
10000	0.76944	0.66667	0.76944	0.66667
50000	0.66839	0.77778	0.66839	0.77778

Table 2: Average precision and recall for ALARM

more complex compared to the GDN and, consequently, more data is needed for edge discovery in the ALARM network. However, given the results, it seems that for small data sizes the discovery on the ALARM performs better. This is a consequence of the strength of the missing dependencies. Namely, strong dependencies are easier to detect than the weak ones. Contrary to the GDN, the ALARM network has a significant set of dependencies which are strong. Consequently, the presence of such dependencies can have a favorable effect on the recall for small data sizes.

Discussion

The experimental results demonstrate that the introduced method supports tractable runtime discovery of missing causal dependencies in complex Bayesian models, despite the used information theoretic analysis, which is computationally demanding. This is possible in cases where complex Bayesian causal models are gradually assembled out of simpler modules which correctly capture causal relations between smaller sets of variables; i.e. prior knowledge of the relations over subsets of variables exists and global dependencies are checked for each module before it is added to the overall model. By using the theory on I-maps we show that the missing dependencies in the assembled model are detected by dependency tests on pairs of variables from different modules, which is a small fraction of conditional dependency checks between all possible variable subsets of local models. The efficiency is further improved by using Markov blankets defined over *patterns*, which describe subsets of variables that are sufficient for the discovery of missing dependencies in assembled models. Such Markov blankets limit the sets of variables in which we have to search for combinations of conditioning variables required for the reliable decision, whether an edge resulting from the pairwise dependency tests is superfluous or genuine. Our future work will focus on a systematic validation of the presented method by using a greater variety of probabilistic models and investigation of the impact the true underlying distributions have on the performance of the discovery algorithms.

References

de Oude, P., and Pavlin, G. 2008. An information theoretic approach to verification of modular bayesian fusion systems. In *The 11th International Conference on Information Fusion*.

de Oude, P., and Pavlin, G. 2009. Dependence discovery in modular bayesian networks. Technical report, Informatics Institute, University of Amsterdam, The Netherlands.

Laskey, K., and Mahoney, S. 1997. Network fragments: Representing knowledge for constructing probabilistic models. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 334–34. San Francisco, CA: Morgan Kaufmann.

Neapolitan, R. 2003. *Learning Bayesian Networks*. Prentice Hall.

Pavlin, G.; de Oude, P.; Maris, M.; Nunnink, J.; and Hood, T. 2008. A multi agent systems approach to distributed bayesian information fusion. *Internation Journal on Multisensor Multi-source Information Fusion*. To appear.

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

Pearl, J. 2000. *Causality: Models Reasoning and Inference*. Cambridge University Press.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. The MIT Press.

Xiang, Y. 2002. *Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach*. Cambridge University Press.