# Logical and Algorithmic Properties of Stable Conditional Independence

Mathias Niepert [a] Dirk Van Gucht [a] Marc Gyssens [b]

[a] *Computer Science Department, Indiana University, Lindley Hall 215, 150 S. Woodlawn Ave., Bloomington, IN 47405-7104, USA.*
Email: {mniepert,vgucht}@cs.indiana.edu

[b] *Department WNI, Hasselt University & Transnational University of Limburg, Agoralaan, Bldg. D, B-3590 Diepenbeek, Belgium.*
Email: marc.gyssens@uhasselt.be

**Abstract**

The logical and algorithmic properties of stable conditional independence (CI) as an alternative structural representation of conditional independence information are investigated. We utilize recent results concerning a complete axiomatization of stable conditional independence relative to discrete probability measures to derive perfect model properties of stable conditional independence structures. We show that stable CI can be interpreted as a generalization of Markov networks and establish a connection between sets of stable CI statements and propositional formulas in conjunctive normal form. Consequently, we derive that the implication problem for stable CI is coNP-complete. Finally, we show that Boolean satisfiability (SAT) solvers can be employed to efficiently decide the implication problem and to compute concise, non-redundant representations of stable CI, even for instances involving hundreds of random variables.

*Key words:* conditional independence, graphical models, stable conditional independence, computational complexity, concise representation

## 1 Introduction

Conditional independence is an important concept in many calculi for dealing with knowledge and uncertainty in artificial intelligence. The notion plays a fundamental role for learning and reasoning in intelligent systems. A conditional independence (CI) statement speaks to the independence of two sets of random variables relative to a third: given three mutually disjoint sets $A$, $B$, and $C$ of random variables, $A$ and $B$ are conditionally independent relative

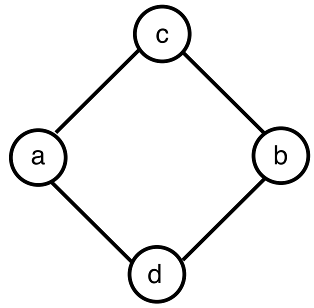


Fig. 1. An undirected graphical model over 4 variables representing the stable CI structure $\{I(a,b|cd), I(c,d|ab)\}$. Please note that we always omit symmetric and trivial CI statements.

to $C$ if any instantiation of the variables in $C$ renders the variables in $A$ and $B$ independent. In other words, if we have knowledge about the state of $C$, then knowledge about the state of $A$ does not provide additional evidence for the state of $B$ and vice versa. We use the notation $I(A, B|C)$ to specify this independence condition.

When novel information becomes available in a probabilistic system, the set of associated, relevant CI statements changes dynamically. However, some of the CI statements will continue to hold, i.e., they remain *stable* under change in the system. Technically, the notion of stability of a CI statement $I(A, B|C)$, in the context of a set of random variables $S$ and a set of CI statements $\mathcal{C}$, is defined by requiring that, for every superset $C' \supseteq C$ which is disjoint from $A$ and $B$, the CI statement $I(A, B|C')$ also holds. In other words, the independence of $A$ and $B$ relative to $C$ is unaffected by adding random variables to $C$. Clearly, this property does not hold in general. Adding variables to the set $C$ may affect the (in-)dependence of $A$ and $B$. A special case for which the stability of the CI statement $I(A, B|C)$ is guaranteed is the situation where $A \cup B \cup C = S$. (When $A \cup B \cup C = S$, the CI statement $I(A, B|C)$ is said to be *saturated*.)

Among the most frequently used models for representing conditional independence information are *graphs*, wherein the nodes correspond to random variables and the edges encode the (in-)dependence information among the variables. There are three main types of graphical models: undirected graphs, directed graphs, and chain graphs. In this paper, we focus specifically on undirected graphical models (also called Markov networks) since we will show that the class of stable CI structures is a strict generalization of the class of CI structures represented by Markov networks. Let $S = \{a, b, c, d\}$ be a set of random variables, and let $G$ be the Markov network shown in Figure 1. Then, $G$ represents the CI statements $I(a, b|cd)$ and $I(c, d|ab)$. (In this paper, we use the notation $a_1 \cdots a_n$ for the set $\{a_1, \ldots, a_n\}$.)

One of the useful properties of the existence of a stable CI statement $I(A, B|C)$ in a set of CI statements $\mathcal{C}$ is that, in a representation of $\mathcal{C}$, it is not necessary to further represent CI statements of the form $I(A, B|C')$, where $C'$ is a strict superset of $C$. This can lead to a substantial decrease in the number of CI statements that need to be maintained in the system. The importance of stable conditional independence for reducing the complexity of representation of conditional independence structures has recently been established [1].

In this paper, we approach the paradigm of stable CI as a strict generalization of Markov networks to represent and reason about conditional independence. A good understanding of its logical and algorithmic properties will lead to new theoretical insights and applications in the field of uncertain reasoning. While several results regarding these properties exist [1,2,3], no study has investigated these as comprehensively as it was done for unrestricted CI and graphical models relative to the class of discrete probability measures [4]. In this paper, we extend this study to stable conditional independence by utilizing recent results concerning a finite sound and complete axiomatization of the implication problem for stable CI statements, relative to discrete probability measures [5]. In particular, we show that (1) every stable CI structure has a perfect model, i.e., a discrete probability measure that satisfies all the CI statements in $\mathcal{C}$, but none other, (2) the number of distinct stable CI structures grows at least double-exponentially with the number of random variables, and (3) every set of CI statements represented by a Markov network is a set of stable CI statements. We establish a direct connection between sets of stable CI statements and propositional formulas in conjunctive normal form and use this connection to show that the conditional independence implication problem for stable conditional independence is coNP-complete. In light of these results, we present experimental results that show how existing SAT solvers can be employed to (1) decide instances of the stable CI implication problem and (2) compute concise, non-redundant representations of stable CI structures, even for instances involving hundreds of random variables.

## 2 Conditional Independence

Throughout this paper, $S$ will be a non-empty finite set of random variables.

**Definition 1** *The expression $I(A, B|C)$, with $A$, $B$, and $C$ pairwise disjoint subsets of $S$, is called a* conditional independence (CI) statement *over $S$.*

*The CI statement $I(A, B|C)$ over $S$ is called* saturated *if $A \cup B \cup C = S$.*

### *2.1 Markov Networks*

**Definition 2** *A Markov network over a finite set $S$ is an undirected graph $G$ with nodes corresponding to random variables in $S$. The conditional independence statement $I(A, B|C)$ is represented by $G$ if every path in $G$ between a node in $A$ and a node in $B$ contains a node in $C$, or, equivalently, if $C$ separates $A$ and $B$.*

Each Markov network $G$ over $S$ represents a set of conditional independence statements through this separation criterion. The set of CI statements *represented* by $G$ will be denoted by $\mathcal{C}(G)$. Every set of CI statements $\mathcal{C}(G)$ *represented* by a Markov network $G$ will be called a *Markov model*.

Markov models can be completely axiomatized using the inference system in Figure 4 [6]. The Markov network in Example 1 represents the set of CI statements $\{I(a, b|cd), I(c, d|ab)\}$. Please note that we always omit symmetric and trivial CI statements.

### *2.2 Satisfaction of a CI statement by a Probability Measure*

In this section, we recall the notion of a probability measure satisfying a CI statement.

**Definition 3** *A* probability model *over $S = \{s_1, \ldots, s_n\}$ is a pair $(dom, P)$, where dom is a domain mapping that maps $s_i$ to a finite non-empty domain $dom(s_i)$, $1 \leq i \leq n$, and $P$ is a probability measure having $dom(s_1) \times \cdots \times dom(s_n)$ as its sample space. We say that $P$ is a* binary probability measure *if, for each $s_i$, $1 \leq i \leq n$, $dom(s_i) = \{0, 1\}$.*

*For $A = \{a_1, \ldots, a_k\} \subseteq S$, we will say that $\mathbf{a}$ is a domain vector of $A$ if $\mathbf{a} \in dom(a_1) \times \cdots \times dom(a_k)$.*

In what follows, we only refer to probability measures, keeping their underlying probability models implicit.

**Definition 4** *Let $I(A, B|C)$ be a CI statement over $S$ and let $P$ be a probability measure. We say that $P$* satisfies *$I(A, B|C)$ if, for all domain vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ of $A$, $B$, and $C$, respectively, $P(\mathbf{c})P(\mathbf{a}, \mathbf{b}, \mathbf{c}) = P(\mathbf{a}, \mathbf{c})P(\mathbf{b}, \mathbf{c})$.*

**Definition 5** *Let $S$ be a finite set of random variables. A probability measure $P$ is Markovian with respect to a Markov network $G$ over $S$, if $I(A, B|C)$ is represented by $G$ implies that $P$ satisfies $I(A, B|C)$. A probability measure $P$ is perfectly Markovian with respect to $G$ if the converse implication holds as well.*

Relative to the notion of satisfaction, we can now define the *conditional independence implication problem.*

**Definition 6** *Let $\mathcal{C}$ be a set of CI statements over $S$, let $c$ be a CI statement over $S$, and let $\mathcal{P}$ be the class of discrete probability measures over $S$. We say that $\mathcal{C}$* implies *$c$ relative to $\mathcal{P}$, and write $\mathcal{C} \models c$, if each probability measure in $\mathcal{P}$ that satisfies the CI statements in $\mathcal{C}$ also satisfies the CI statement $c$. The set $\{c \mid \mathcal{C} \models c\}$ will be denoted by $\mathcal{C}^*$.*

**Definition 7** *The* conditional independence (CI) implication problem *is the problem of deciding the language*

$$\{(S, \mathcal{C}, c) \mid \mathcal{C} \text{ a set of CI statements over } S,\ c \text{ a CI statement over } S,\ \mathcal{C} \models c\}.$$

We can now define the notion of perfect models for sets of CI statements and the notion of a CI structure.

**Definition 8** *Let $\mathcal{C}$ be a set of CI statements over $S$. $\mathcal{C}$ is a* CI structure *if and only if $\mathcal{C} = \mathcal{C}^*$. Furthermore, we say that a probability measure $P$ is a perfect model for $\mathcal{C}$ if $P$ satisfies all the CI statements in $\mathcal{C}^*$ and none other.*

## 3 Inference Systems for CI Implication Problems

Given the notion of a CI implication problem, it is common place to consider inference rules and systems that are *sound* for these problems. An inference rule (an inference system) is sound relative to the class of discrete probability measures if it infers, given a set of CI statements $\mathcal{C}$, only CI statements in $\mathcal{C}^*$. When an inference system can infer all CI statements in $\mathcal{C}^*$, it is said to be complete.

The best know sound inference system for the CI implication problem relative to the class of discrete probability measures is the *semi-graphoid axiom system* [6]. We denote it by $\mathcal{G}$ and its inference rules are depicted in Figure 2. Note, however, that system $\mathcal{G}$ is not complete. In fact, it is known that there does not exist a finite set of sound inference rules that is sound and complete for the implication problem on unrestricted CI statements [7]. It is also unknown whether this implication problem is decidable.

For the implication problem for saturated CI statements, the situation is different. In Figure 3, system $\mathcal{S}$ is shown, which is a finite set of inference rules

| | |
|---|---|
| $I(A, \emptyset \| C)$ | **Triviality** |
| $I(A, B \| C) \rightarrow I(B, A \| C)$ | **Symmetry** |
| $I(A, B \cup D \| C) \rightarrow I(A, D \| C)$ | **Decomposition** |
| $I(A, B \| C \cup D)$ & $I(A, D \| C) \rightarrow I(A, B \cup D \| C)$ | **Contraction** |
| $I(A, B \cup D \| C) \rightarrow I(A, B \| C \cup D)$ | **Weak union** |

Fig. 2. System $\mathcal{G}$, the semi-graphoid axiom system, is sound, but not complete, for the implication problem for unrestricted CI statements.

| | |
|---|---|
| $I(A, \emptyset \| C)$ | **Triviality** |
| $I(A, B \| C) \rightarrow I(B, A \| C)$ | **Symmetry** |
| $I(A \cup D, B \| C)$ & $I(A, D \| B \cup C) \rightarrow I(A, B \cup D \| C)$ | **Contraction** |
| $I(A, B \cup D \| C) \rightarrow I(A, B \| C \cup D)$ | **Weak union** |

Fig. 3. System $\mathcal{S}$ is sound and complete for the CI implication problem for saturated statements. Note that the inference rule **contraction** has a slightly different form to accommodate saturated CI statements [4].

that is sound and complete for this implication problem relative to the class of discrete probability measures [6].

For sets of CI statements represented by Markov networks, the situation is yet different. Figure 4 depicts system $\mathcal{M}$, which is a finite set of inference rules that is sound and complete for the implication problem for sets of CI statements represented by Markov networks, relative to the class of discrete probability measures [6].

Let $\mathcal{I}$ be an inference system for CI statements. The derivability of a conditional independence statement $c$ from a set of conditional independence statements $\mathcal{C}$ under the inference rules of system $\mathcal{I}$ is denoted by $\mathcal{C} \vdash_{\mathcal{I}} c$.

The *closure* of $\mathcal{C}$ under $\mathcal{I}$, denoted $\mathcal{C}^{\mathcal{I}^+}$, is the set $\{c \mid \mathcal{C} \vdash_{\mathcal{I}} c\}$.

$$
\begin{array}{ll}
I(A, \emptyset | C) & \textbf{Triviality} \\
I(A, B|C) \rightarrow I(B, A|C) & \textbf{Symmetry} \\
I(A, B \cup D|C) \rightarrow I(A, D|C) & \textbf{Decomposition} \\
I(A, B|C) \rightarrow I(A, B|C \cup D) & \textbf{Strong union} \\
I(A, B|C \cup D) \ \& \ I(A, D|B \cup C) \rightarrow I(A, B \cup D|C) & \textbf{Intersection} \\
I(A, B|C) \rightarrow I(A, \{d\}|C) \vee I(\{d\}, B|C) & \textbf{Transitivity}
\end{array}
$$

Fig. 4. System $\mathcal{M}$ is sound and complete for the CI implication problem for CI statements represented by Markov networks.

## 4 Stable Conditional Independence

When novel information becomes available to a probabilistic system, the set of associated, relevant CI statements changes dynamically. However, some of these CI statements will continue to be satisfied, i.e., they remain *stable*. The paradigm of stable conditional independence, and some of its the properties, were first investigated by Matúš [2], who named it *ascending* conditional independence, and later by de Waal and van der Gaag [1], who coined the term *stable* conditional independence. Every set of CI statements can be partitioned into its *stable* and *unstable* part. In this section, we will recall two different characterizations of stable CI structures, one using a finite set of inference rules, and the other using the *lattice-inclusion* property of CI statements [5]. We will harness these results to prove several important properties about stable CI. The set of inference rules in Figure 5 will be denoted by $\mathcal{A}$. The *symmetry*, *decomposition*, and *contraction* rules are part of the semi-graphoid axioms [6] (see Figure 2). *Strong union* and *strong contraction* are additional inference rules.

Stable independence can be defined *relative* to a set of CI statements [1,3]. However, we approach the paradigm of stable CI as a mechanism for the succinct representation of conditional independence information, much like graphical models are used for this purpose. Instead of assuming that every CI statement is satisfied by a probability measure which is perfectly Markovian with respect to a graphical model, we assume that every CI statement is satisfied by a probability measure which is perfectly Markovian with respect to a set of stable CI statements. Therefore, in the remainder of the paper, a set of stable conditional independence statements will be *any* set of CI statements that are *implicitly known* (i.e., assumed) to be stable. Whenever we say that a set of CI statements is stable, we implicitly assume that $\mathcal{C}^*$ satisfies the required condition. Hence, in general, a set of stable CI statements $\mathcal{C}$ can be different from the set $\mathcal{C}^*$. The motivation for this approach is to achieve a structural representation of conditional independence information which is broader than Markov networks but which still allows for efficient implication testing and storage. The next definition formalizes this approach.

**Definition 9** *Let $\mathcal{C}$ be a set of CI statements. We say that $\mathcal{C}$ is a set of stable CI statements, if for all $I(A, B|C) \in \mathcal{C}$ and for all $C' \supseteq C$ we have that $I(A, B|C') \in \mathcal{C}^*$. Equivalently, a set of stable CI statements is a set of CI statements for which the inference rule strong union (see Figure 4) is sound. A* stable CI structure *is a set of stable CI statements $\mathcal{C}$ such that $\mathcal{C} = \mathcal{C}^*$.*

The following result follows immediately from this definition.

**Proposition 10** *Let $\mathcal{C}$ be a set of saturated CI statements over $S$. Then $\mathcal{C}$ is a set of stable CI statements over $S$.*

In analogy to the definition of a (perfectly) Markovian probability measure with respect to graphical models (Definition 5), we can define the concept of a (perfectly) Markovian probability measure with respect to stable CI structures.

**Definition 11** *Let $\mathcal{C}$ be a stable CI structure. A probability measure $P$ is Markovian with respect to $\mathcal{C}$ if $I(A, B|C) \in \mathcal{C}$ implies that $P$ satisfies $I(A, B|C)$. A probability measure $P$ is perfectly Markovian with respect to $\mathcal{C}$ if the converse implications holds as well. We say that a probability measure is a perfect model for $\mathcal{C}$ if and only if it is perfectly Markovian with respect to $\mathcal{C}$.*

### *4.1 The Implication Problem for Stable Conditional Independence*

Here, we recall two characterizations of the implication problem for stable CI statements (the stable CI implication problem), one in terms of a finite system of inference rules, and another using the *lattice-inclusion property* [5]. We will use these results to show that each stable CI structure has a perfect model with respect to discrete probability measures, but not with respect to binary discrete probability measures.

A powerful tool in deriving results about the CI implication problem is the association of semi-lattices with CI statements [5]. Given subsets $A$ and $B$ of $S$ we write $[A, B]$ for the lattice $\{U \mid A \subseteq U \subseteq B\}$.

**Definition 12** *Let $I(A, B|C)$ be a CI statement over $S$. The* semi-lattice *associated with $I(A, B|C)$, and denoted by $\mathcal{L}(A, B|C)$ is the set $[C, S] - ([A, S] \cup [B, S])$.*

| | |
|---|---|
| $I(A, \emptyset \| C)$ | **Triviality** |
| $I(A, B \| C) \rightarrow I(B, A \| C)$ | **Symmetry** |
| $I(A, B \cup D \| C) \rightarrow I(A, D \| C)$ | **Decomposition** |
| $I(A, B \| C)$ & $I(A, D \| B \cup C) \rightarrow I(A, B \cup D \| C)$ | **Contraction** |
| $I(A, B \| C) \rightarrow I(A, B \| CD)$ | **Strong union** |
| $I(A, B \| C)$ & $I(D, E \| AC)$ & $I(D, E \| BC)$ $\rightarrow I(D, E \| C)$ | **Strong contraction** |

Fig. 5. The inference rules of system $\mathcal{A}$.

**Example 13** *Let $S = \{a, b, c, d\}$ and consider the CI statement $I(a, b|c)$. Then $\mathcal{L}(a, b|c) = [c, S] - ([a, S] \cup [b, S]) = \{c, cd\}$.*

We will often write $\mathcal{L}(c)$ to denote the semi-lattice associated with a CI statement $c$ and $\mathcal{L}(\mathcal{C})$ to denote the union of semi-lattices, $\bigcup_{c' \in \mathcal{C}} \mathcal{L}(c')$, associated with a set of CI statements $\mathcal{C}$. We can now state the two characterizations for the conditional independence implication problem for stable CI statements relative to the class of discrete probability measures [5].

**Theorem 14** *Let $\mathcal{C}$ be a set of stable CI statements over $S$ and let $c$ be a CI statement over $S$. Then the following statements are equivalent*

*(a) $\mathcal{C} \models c$;*
*(b) $\mathcal{C} \vdash_{\mathcal{A}} c$; and*
*(c) $\mathcal{L}(\mathcal{C}) \supseteq \mathcal{L}(c)$.*

**Example 15** *Let $S = \{a, b, d, e\}$, let $\mathcal{C} = \{I(a, b|\emptyset), I(d, e|a), I(d, e|b)\}$ be a set of stable CI statements, and let $c = I(d, e|\emptyset)$. We know, by* strong contraction*, that $\mathcal{C} \vdash_{\mathcal{A}} c$ and, therefore, $\mathcal{C} \models c$ by Theorem 14. Now, $\mathcal{L}(\mathcal{C}) = \{\emptyset, d, e, de\} \cup \{a, ab\} \cup \{b, ab\} = \{\emptyset, a, b, d, e, ab, de\} \supseteq \{\emptyset, a, b, ab\} = \mathcal{L}(c)$.*

One might expect that, based on the definition of stable CI, it would be sufficient to have the inference rule *strong union* in system $\mathcal{A}$ but not *strong contraction*. However, as the following example demonstrates, system $\mathcal{A}$ without *strong contraction* is *not* complete for the stable CI implication problem.

**Example 16** *Let $S = \{a, b, c, d\}$ and consider the set of stable CI statements $\mathcal{C} = \{I(a, b|\emptyset), I(a, b|c), I(a, b|d), I(a, b|cd), I(c, d|a), I(c, d|b), I(c, d|ab)\}$. We know that $I(c, d|\emptyset)$ is implied by $\mathcal{C}$ [8]. However, one can verify that $I(c, d|\emptyset)$ cannot be derived from $\mathcal{C}$ under $\mathcal{A}$ without the rule* strong contraction.

The next result follows from the existence of discrete perfect models with respect to CI statements [4], a result which was later strengthened by [9].

**Proposition 17** *For every stable CI structure $\mathcal{C}$, there exists a discrete probability measure $P$ such that $P$ is a perfect model for $\mathcal{C}$.*

However, the previous result does not hold for the class of binary discrete probability measures.

**Proposition 18** *There exists a stable CI structure for which no binary discrete probability measure is a perfect model.*

**Proof:** Let $S = \{a, b, c\}$ and let $\mathcal{C} = \{I(a, b|\emptyset), I(a, b|c)\}$. We have that $\mathcal{L}(a, b|\emptyset) = \{\emptyset, c\}$ and $\mathcal{L}(a, b|c) = \{c\}$. Now, since $\mathcal{L}(a, c|\emptyset) = \{\emptyset, b\}$ and $\mathcal{L}(b, c|\emptyset) = \{\emptyset, a\}$, we have by Theorem 14 (c) that neither $I(a, c|\emptyset)$ nor $I(b, c|\emptyset)$ are implied by $\mathcal{C}$. Hence, $\mathcal{C}$ is a stable CI structure. However, we know that every binary probability measure that satisfies the CI statements in $\mathcal{C}$ also satisfies either $I(a, c|\emptyset)$ or $I(b, c|\emptyset)$ [4]. Thus, no binary probability measure is a perfect model for $\mathcal{C}$.

The combination of these result shows that the paradigm of stable CI has the same perfect model properties as unrestricted CI.

### *4.2 Markov Networks and Stable Conditional Independence*

The primary goal of this section is to relate stable conditional independence to Markov networks. In particular, we will show that every set of CI statements represented by a Markov network is a stable CI structure. Consequently, Markov networks are a specialization of the more general notion of stable conditional independence.

**Theorem 19** *Let $G$ be a Markov network over $S$. Then the set of CI statements represented by $G$, i.e., $\mathcal{C}(G)$, is a stable CI structure.*

**Proof:** It is well-known that *strong union* is a sound inference rule for separation in undirected graphs [10] (see Figure 4). In addition, it can be verified that the inference rule *strong contraction* is sound for separation in undirected graph. Thus, inference system $\mathcal{A}$ is sound for separation in Markov networks and the statement of the theorem follows.

**Corollary 20** *For every Markov network $G$ there exists a stable CI structure $\mathcal{C}$, and every discrete probability measure that is (perfectly) Markovian w.r.t. $G$ satisfies the elements in $\mathcal{C}$ (and none other).*

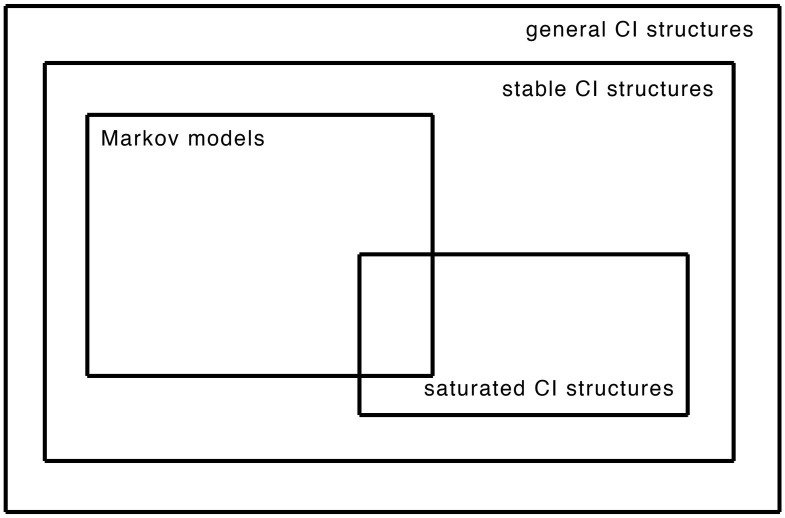


Fig. 6. Inclusion relationships between different representations of conditional independence. Every Markov model is a stable CI structure (Theorem 19). Every saturated CI structure is trivially a stable CI structure.

Theorem 19 implies that the notion of stable conditional independence is a generalization of Markov networks. In what follows, we will investigate how much broader this notion is compared to such networks. First, we show that there exists a stable CI structure that cannot be represented by a Markov network.

**Proposition 21** *There exists a stable CI structure $\mathcal{C}$ over a set $S$, such that for each Markov network $G$ over $S$, $\mathcal{C} \neq \mathcal{C}(G)$.*

**Proof:** Let $S = \{a, b, c, d\}$ and let $\mathcal{C} = \{I(a, b|cd), I(a, d|bc)\}$ be a set of stable CI statements. We have that $\mathcal{L}(a, b|cd) = \{cd\}$ and $\mathcal{L}(a, d|bc) = \{bc\}$. Hence, by Theorem 14 (c), no other CI statement is implied by $\mathcal{C}$. Thus, $\mathcal{C}$ is a stable CI structure. However, every Markov network that represents these two CI statements also represents the CI statement $I(a, bd|c)$ by the inference rule *intersection* (see Figure 4) which is sound for separation in undirected graphs [10]. Thus, the class of all CI structures represented by the class of Markov networks is a strict subclass of the class of stable CI structures.

Figure 6 depicts some relationships between different structural representations of conditional independence information. Please note that each saturated CI structure is trivially a stable CI structure.

### *4.3 Some Combinatorics about Stable Conditional Independence*

In this section, we will show, given a set of random variables $S$, that the number of distinct stable conditional independence structures grows at least double-exponentially. This shows analytically that stable conditional independence can represent a much broader class of CI structures compared to undirected models, since there can only be $2^{(|S|(|S|-1))/2}$ different undirected graphical models over a set of random variables $S$.

**Lemma 22** *Let $S$ be a finite set of discrete random variables. Then, the number of distinct stable CI structures $d_S$ over $S$ is at least*

$$d_S \geq \sum_{i=0}^{|S|-2} (2^{\binom{|S|}{2}\binom{|S|-2}{i}} - 1).$$

**Proof:** Let $S$ be a finite set, let $V \subseteq S$ with $|V| = |S| - 2$, and let $U \subseteq V$. For every lattice $[U, V]$, there exists a stable CI structure $\mathcal{C}$ (for instance, $\mathcal{C} = \{I(u, v|U') \mid U' \supseteq U, \{u, v\} = S - V\}$) such that $\mathcal{L}(\mathcal{C}) = [U, V]$. Consider the set $D_i^S = \{[U, V] \mid |V| = |S| - 2, |U| = |S| - 2 - i, U \subseteq V \subseteq S\}$. There are $\binom{|S|}{2}$ different subsets of $S$ of size $|S| - 2$. Each of these subsets $V$ has $\binom{|S|-2}{|S|-2-i} = \binom{|S|-2}{i}$ different subsets of size $|S| - 2 - i$. Hence, we have that $|D_i^S| = \binom{|S|}{2}\binom{|S|-2}{i}$. Now, for every $i = 0...(|S| - 2)$, each non-empty subset of the set $D_i^S$ corresponds to a set of stable CI statements whose union of semilattices is distinct from the union of semilattices of all other subsets of $D_i^S$, and, in addition, whose union of semilattices is also distinct from the union of semilattices of all non-empty subsets of every $D_j^S$ with $i \neq j$. Thus, by Theorem 14, each of the non-empty subsets of $D_i^S$ gives rise to a new stable CI structure. Hence, from each $D_i^S$ we get $2^{\binom{|S|}{2}\binom{|S|-2}{i}} - 1$ distinct stable CI structures. Since $i$ ranges from 0 to $|S|-2$ the statement of the lemma follows.

**Example 23** *For $|S| = 3$ there are 8 Markov networks, 22 unrestricted [7], and 14 stable CI structures. For $|S| = 4$ there are 64 Markov networks [7], 18,478 unrestricted [11], and at least 4,221 distinct stable CI structures. For $|S| = 5$ there are at least 2,147,485,692 distinct stable CI structures.*

Using Lemma 22, we can show that the number of stable CI structures grows double-exponentially with the size of $S$.

**Theorem 24** *The number of stable CI structures over a finite set $S$ grows at least double-exponentially with the size of $S$.*

**Proof:** Let $S$ be a finite set and assume without loss of generality that $|S|-2$ is even. It is known that $\binom{n}{k} \geq (n/k)^k$ and, therefore,

$$\sum_{i=0}^{|S|-2} 2^{\binom{|S|}{2}\binom{|S|-2}{i}} \geq 2^{\binom{|S|}{2}\binom{|S|-2}{(|S|-2)/2}} \geq 2^{\binom{|S|}{2}2^{(|S|-2)/2}}.$$

Now, by Lemma 22, we have that the number of stable CI structures is greater than or equal to $2^{\binom{|S|}{2}2^{(|S|-2)/2}} - 1$. The proof is analogous when $|S|-2$ is odd, where we use $\lfloor(|S|-2)/2\rfloor$ instead of $(|S|-2)/2$.

## 5 Computational Complexity of the Stable CI Implication Problem

Recall that the stable conditional independence implication problem, denoted here as `STABLE-IMPLICATION`, is the problem of deciding, given a set of random variables $S$, a set of stable CI statements $\mathcal{C}$, and a CI statement $c$, whether $\mathcal{C}$ implies $c$ relative to the class of discrete probability measures. In this section, we will show that `STABLE-IMPLICATION` is `coNP`-complete. Furthermore, in Section 6, we will be able to prove that a linear-time reduction exists from `STABLE-IMPLICATION` to the unsatisfiablity problem, here denoted as `UNSAT`, for propositional logic formulas over variables that correspond to the random variables in $S$. This permits the use of `SAT` solvers to decide instances of `STABLE-IMPLICATION`. In Section 7, we present experimental results which show how such instances, even with hundreds of variables, can be decided efficiently.

First, we need to introduce the notion of minterms. Minterms are certain propositional formulas associated with subsets of a set of variables.

**Definition 25** *Let $T$ be a set of propositional variables. Then, for each $X \subseteq T$, the* minterm *associated with $X$, and denoted by $\mathbf{X}$, is the propositional formula $\bigwedge_{a\in X} a \wedge \bigwedge_{b\in T-X} \neg b$.*

*Let $\Phi$ be a propositional formula over $T$. The* minset *of $\Phi$, denoted* $\text{minset}(\Phi)$*, is the set $\{X \mid \mathbf{X} \models_{prop} \Phi\}$, where $\models_{prop}$ denotes the logical implication relation for propositional logic. The* negative minset *of $\Phi$, denoted* $\text{negminset}(\Phi)$*, is the set* $\text{minset}(\neg\Phi)$.

We will now isolate a special class of propositional formulas.

**Definition 26** *Let $T$ be a set of propositional variables. Then* `3-CNFV`$(T)$ *denotes the set of all* `CNF` *propositional formulas over the variables in $T$ in which the clauses are restricted to be of the form $x \vee y$, $\neg x \vee y \vee z$, $\neg x \vee \neg y \vee z$, and $\neg x \vee \neg y \vee \neg z$, where $x, y$, and $z$ are variables in $T$.*

**Proposition 27** *Let* `3SAT-CNFV` *denote the satisfiability problem for* `3-CNFV`$(T)$ *parametrized over sets $T$ of propositional variables. Then,* `3SAT-CNFV` *is an* `NP`*-complete problem.*

**Proof:** Clearly, `3SAT-CNFV` is in `NP`. The hardness of `3SAT-CNFV` can be established via a reduction from the standard `3SAT` problem. Every clause in `3SAT` of the form $x \vee y \vee z$ is mapped to the formula $(x \vee y \vee \neg w) \wedge (z \vee w)$, where $w$ is a new variable. All other clauses in `3SAT` are retained. This reduction is possible in polynomial time and preserves satisfiability.

Next, we defined a polynomial-time computable reduction from formulas in `3-CNFV` to sets of non-trivial CI statements.

**Definition 28** *Let $T$ be a set of propositional variables and let $S = T \cup \{r, s\}$ with $r \notin T$ and $s \notin T$. Let $\mathcal{T}(S)$ denote the set of all CI statements over $S$. For a formula $\Phi$ in `3-CNFV`$(T)$, let `clauses`$(\Phi)$ denote the set of clauses in $\Phi$. Then $f$ : `3-CNFV`$(T) \to 2^{\mathcal{T}(S)}$ is defined as follows. For formula $\Phi$,*

$$f(\Phi) = \bigcup_{c \in \texttt{clauses}(\Phi)} f(c),$$

*with*[1]

$$\begin{aligned}
f(x) &= \{I(x, v|\emptyset) \mid v \in S - \{x\}\};\\
f(\neg x) &= \{I(u, w|x) \mid u, w \in S - \{x\}, u \neq w\};\\
f(x \vee y) &= \{I(x, y|\emptyset)\};\\
f(\neg x \vee y) &= \{I(y, v|x) \mid v \in S - \{x, y\}\};\\
f(\neg x \vee \neg y) &= \{I(v, w|xy) \mid v, w \in S - \{x, y\}, v \neq w\}\\
f(\neg x \vee y \vee z) &= \{I(y, z|x)\};\\
f(\neg x \vee \neg y \vee z) &= \{I(z, v|xy) \mid v \in S - \{x, y, z\}\};\\
f(\neg x \vee \neg y \vee \neg z) &= \{I(v, w|xyz) \mid v, w \in S - \{x, y, z\}, v \neq w\}.
\end{aligned}$$

Notice that the mapping $f$ can be computed in polynomial time and that for each formula $\Phi$, for each clause $\mathbf{c} \in$ `clauses`$(\Phi)$, and for each $X \subseteq T$, we have that $X \in \mathcal{L}(f(c))$ if and only if $\mathbf{X} \models_{prop} \neg\mathbf{c}$.

**Example 29** *Let $T = \{a, b, c\}$, let $S = T \cup \{d, e\}$ and let $\Phi = (a \vee c) \wedge (\neg a \vee \neg b \vee c)$. Then*

[1] To simplify the mapping, we assume that every formula in `3-CNFV`$(T)$ is preprocessed to transform clauses with duplicate literals (e.g., $\ell \vee \ell$ or $\neg\ell \vee \neg\ell \vee \neg\ell$) into their simplified forms (here: $\ell$ and $\neg\ell$). Of course, this preprocessing step is computable in polynomial time.

$$
\begin{aligned}
f(\Phi) &= f(a \vee c) \cup f(\neg a \vee \neg b \vee c) \\
&= \{I(a, c|\emptyset)\} \cup \{I(c, d|ab), I(c, e|ab)\} \\
&= \{I(a, c|\emptyset), I(c, d|ab), I(c, e|ab)\}.
\end{aligned}
$$

*Furthermore,*

$$
\begin{aligned}
\mathcal{L}(f(\Phi)) &= \{\emptyset, b, d, e, bd, be, de, bde, ab, abe, abd\} \text{ and} \\
\text{negminset}(\Phi) &= \{\mathbf{X} \mid X = \emptyset \vee X = \{b\} \vee X = ab\} = \{\emptyset, b, ab\}.
\end{aligned}
$$

We can now state the main result of this section.

**Theorem 30** `STABLE-IMPLICATION` *is* `coNP`*-complete.*

**Proof:** We first show that the co-problem of `STABLE-IMPLICATION` is in `NP`. Let $\mathcal{C}$ be a set of stable CI statements over $S$ and let $c$ be a CI statement over $S$. Since, by Theorem 14, $\mathcal{C} \not\models c$ if and only if $\mathcal{L}(\mathcal{C}) \not\supseteq \mathcal{L}(c)$, it is sufficient to guess $X \in \mathcal{L}(c) - \mathcal{L}(\mathcal{C})$ and then verify in polynomial time that, for all $I(A, B|C) \in \mathcal{C}$, one has that $(X \supseteq A) \vee (X \supseteq B) \vee \ (X \not\supseteq C)$.

To show the hardness of `STABLE-IMPLICATION` we use a reduction from `3SAT-CNFV`. Let $T$ be a set of propositional variables, let $S = T \cup \{r, s\}$ with $r \notin T$, $s \notin T$, let $f$ be the function from Definition 28, and let $\Phi \in$ `3SAT-CNFV`$(T)$. Then we have the following:

(1) $\mathit{negminset}(\Phi) \subseteq \mathcal{L}(f(\Phi))$; and
(2) $\Phi$ is a contradiction if and only if $\mathcal{L}(I(r, s|\emptyset)) \subseteq \mathcal{L}(f(\Phi))$.

To show (1), let $X \in \mathit{negminset}(\Phi)$. Then, there exists a clause $\mathbf{c}$ in `clauses`$(\Phi)$ such that $\mathbf{X} \models_{prop} \neg\mathbf{c}$. But then, there exists $I(x, y|U) \in f(\mathbf{c})$ such that $X \supseteq U$, $x \notin X$ and $y \notin X$ because otherwise $\mathbf{X} \models_{prop} \mathbf{c}$. It follows that $X \in \mathcal{L}(f(c))$ and, therefore, $X \in \mathcal{L}(f(\Phi))$. To show (2), let $\Phi$ be a contradiction. Notice that $\Phi$ is a contradiction if and only if $\mathit{negminset}(\Phi) = 2^T$. Now, $\mathcal{L}(I(r, s|\emptyset)) = 2^T = \mathit{negminset}(\Phi) \subseteq \mathcal{L}(f(\Phi))$, where the last inclusion follows from (1). But then, by Theorem 14, $\Phi$ is a contradiction if and only if $f(\Phi) \models I(r, s|\emptyset)$. Since $f$ is computable in polynomial time, the result follows.

The logical and algorithmic properties of unrestricted CI, stable CI, saturated CI, and Markov models are summarized in Figure 7.

| **Property of CI** | **Un-restricted** | **Stable** | **Saturated** | **Markov models** |
|---|---|---|---|---|
| Finite, complete axiomatization | No [12] | Yes | Yes [4] | Yes [10] |
| Implication problem | ? | coNP-com. | P [4] | ? |
| Perfect models | Yes [10] | Yes | Yes [4] | Yes [10] |
| Perfect binary models | No [4] | No | ? | ? |

Fig. 7. Summary of properties of conditional independence.

## 6 Implication Testing Using Satisfiability Solvers

In this section, we show that every set of CI statements can be reduced to a propositional formula in linear time. This, together with the results from the previous section, allows us to employ SAT solvers to decide `STABLE-IMPLICATION`. Furthermore, we will show experimentally that numerous instances of the stable CI implication problem can be decided efficiently, even if several hundreds of random variables are involved.

**Definition 31** *Let $\mathcal{C}$ be a set of CI statements over $S$, and let $\texttt{PROP}(S)$ be the set of propositional formulas over variables in $S$. Let $\mathcal{T}(S)$ denotes the set of all CI statements over $S$. The mapping $g : 2^{\mathcal{T}(S)} \rightarrow \texttt{PROP}(S)$ is defined by $g(\mathcal{C}) = \bigwedge_{c \in \mathcal{C}} g(c)$, and $g(I(A, B|C)) = \bigwedge_{a \in A} a \vee \bigwedge_{b \in B} b \vee \bigvee_{c \in C} \neg c$, for each CI statement $I(A, B|C)$ in $\mathcal{C}$.*

The mapping $g$ can be computed in linear time in the size of $\mathcal{C}$. Now, using this mapping we can state the following theorem.

**Theorem 32** *Let $\mathcal{C}$ be a set of stable CI statements over $S$ and let $c$ be a CI statement over $S$. Then $\mathcal{C} \models c$ if and only if $g(\mathcal{C}) \models_{prop} g(c)$.*

**Proof:** We will again use the concepts *minset* and *negminset* introduced in Definition 25. Let $\mathcal{C}$ be a set of CI statements and let $c$ be a CI statement. One can verify that $\mathcal{L}(\mathcal{C}) = \mathit{negminset}(g(\mathcal{C}))$ and $\mathcal{L}(c) = \mathit{negminset}(g(c))$. By Theorem 14, we have that $\mathcal{C} \models c$ if and only if $\mathcal{L}(\mathcal{C}) \supseteq \mathcal{L}(c)$. Now, the statement of the theorem follows.

**Example 33** *Let $S = \{a, b, d, e\}$, let $\mathcal{C} = \{I(a, b|\emptyset), I(d, e|a), I(d, e|b)\}$, and let $c = I(d, e|\emptyset)$. We have $g(\mathcal{C}) = (a \vee b) \wedge (d \vee e \vee \neg a) \wedge (d \vee e \vee \neg b)$ and $g(c) = d \vee e$. We also have $g(\mathcal{C}) \models_{prop} g(c)$ if and only if $g(\mathcal{C}) \wedge \neg g(c)$ is not satisfiable. Now, $g(\mathcal{C}) \wedge \neg g(c) = (a \vee b) \wedge (d \vee e \vee \neg a) \wedge (d \vee e \vee \neg b) \wedge \neg d \wedge \neg e$. This formula is a contradiction. Hence, $\mathcal{C} \models c$ by Theorem 32.*

In this section, we study the notion of an *irredundant equivalent subset* of a set of stable CI statements. We will use this notion to represent a stable CI structure more concisely. For this purpose, we will adopt terminology which was recently introduced in the context of redundancy of propositional formulas in conjunctive normal form [13].

**Definition 34** *A set of CI statements $\mathcal{C}$ over $S$ is* irredundant *if $\mathcal{C} - \{c\} \not\models c$, for all $c \in \mathcal{C}$. Otherwise, it is* redundant*.*

A related definition is that of an irredundant equivalent subset. Note that a set of stable CI statements may have several different irredundant equivalent subsets and that the cardinality of these sets can differ.

**Definition 35** *Let $\mathcal{C}$ be a set of stable CI statements over $S$. A set of stable CI statements $\mathcal{C}'$ is an* irredundant equivalent subset *of $\mathcal{C}$ if and only if*

*(1) $\mathcal{C}' \subseteq \mathcal{C}$;*
*(2) $\mathcal{C}' \models c$ for all $c \in \mathcal{C}$; and*
*(3) $\mathcal{C}'$ is irredundant.*

**Example 36** *Let $S = \{a, b, c\}$ and let $\mathcal{C} = \{I(a, b|\emptyset), I(a, b|c)\}$. Then, $\mathcal{C}' = \{I(a, b|\emptyset)\}$ is an irredundant equivalent subset of $\mathcal{C}$.*

We now have the following property.

**Proposition 37** *Let $\mathcal{C}$ be a set of CI statements over $S$. Then $\mathcal{C}$ is irredundant if and only if for all $c$ in $\mathcal{C}$ we have that $g(\mathcal{C} - \{c\}) \wedge \neg g(c)$ is satisfiable, where $g$ is the mapping defined in Definition 31.*

## 7 Experimental Results

In a first experiment, we randomly generated instances of the stable CI implication problem with up to 400 variables. We then used the mapping $g$ from Definition 31 to transform instances of the implication problem for stable CI into instances of the unsatisfiability problem of propositional logic (UNSAT), to which we applied a SAT solver. We used MiniSat [2] by Niklas Eén and Niklas Sörensson on a Pentium4 dual-core Linux system for the experiments. The performance of the SAT solver is quite remarkable. Figure 8 shows the

[2] http://minisat.se

| **variables** | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|
| **time** [ms] | 740 | 1523 | 3362 | 5627 | 7076 |

Fig. 8. Average time (in milliseconds) needed to decide the implication problem for different numbers of variables and 100,000 antecedents.

**irredundant-subset** ($\mathcal{C}$ : set) $\mathcal{C}'$ : set

$\mathcal{C}' := \mathcal{C}$

**for each** $c \in \mathcal{C}'$

  **begin**

    **if** $g(\mathcal{C}' - \{c\}) \wedge \neg g(c)$ *not* satisfiable

      **then** $\mathcal{C}' := \mathcal{C}' - \{c\}$

  **end**

**return** $\mathcal{C}'$

Fig. 9. A function to compute an irredundant equivalent subset.

average time (out of 10 tests) needed to decide the implication problem $\mathcal{C} \models c$ for $|\mathcal{C}| = 100,000$ and different numbers of variables.

The goal of the second experiment was to determine the average size of irredundant equivalent subsets of a randomly generated set of stable CI statements. The algorithm in Figure 9 is based on Corollary 37. It takes as input a set of stable CI statements $\mathcal{C}$ and returns an irredundant equivalent subset of $\mathcal{C}$ based on several satisfiability tests. For each number of variables from 5 to 25 we randomly created sets of 500 CI statements and determined the size of the irredundant equivalent subsets using the algorithm. Figure 10 shows the average size of 1000 different runs. As one can expect, the fewer variables there are, the smaller is the irredundant equivalent subset. For the 500 satisfiability tests made to compute an irredundant equivalent subset, the algorithm took at most 1100 ms, where the majority of the time was spent on unsatisfiable instances of the problem. This amounts on average to 2ms per satisfiability test for sets of 500 CI statements.

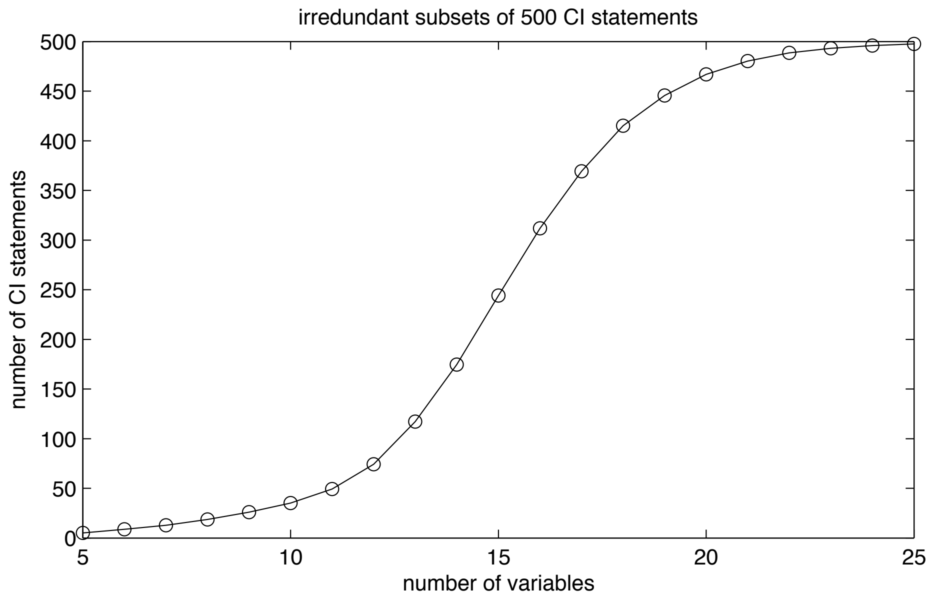


Fig. 10. Size of irredundant equivalent subset of a set of initially 500 stable CI statements for different numbers of attributes.

## 8 Discussion and Future Work

We used a finite complete axiomatization of stable conditional independence to show that the class of stable conditional independence structures has the same perfect model properties as the class of unrestricted conditional independence structures. In addition, we proved that stable conditional independence can be interpreted as a generalization of Markov models in that the class of stable CI structures is a strict superset of the class of CI structures represented by Markov networks.

Many procedures that learn graphical models are based on the *data faithfulness assumption* (see for example [7]). The data faithfulness assumption states that data are "generated" by a probability measure which is perfectly Markovian with respect to an instance of the class of Markov networks under consideration. Now, learning methods based on these procedures are only safely applicable if the data faithfulness assumption is guaranteed. While the data faithfulness assumption is *not* guaranteed for the class of stable CI structures, we have, by Theorem 24, that the number of stable CI structures grows double-exponentially with the size of $S$. Therefore, *more* probability measures (as compared to Markov models) are perfectly Markovian with respect to some stable CI structure. On one hand, this implies that a reasonable graphical representation of stable CI is unlikely, using arguments similar to those made in [7] (page 63). On the other hand, it shows that the class of stable CI structures is the broadest and only double-exponentially growing class of CI structures for which a complete finite axiomatization using inference rules and an implication algorithm are known. We also know that this class of CI structures includes the class of all CI structures represented by Markov networks and that there exists an interesting, direct connection to propositional logic. Furthermore, we have demonstrated that SAT solvers can be used to efficiently decide the implication problem for stable conditional independence, even for instances involving large numbers of random variables. These results allow for a non-redundant, concise representation of conditional independence information using irredundant equivalent subsets of stable CI structures.

In addition to the aforementioned possible applications, stable CI can also be used as part of a probabilistic system that keeps track of both the stable and non-stable part as in [1]. Our results can be leveraged to store information about conditional independencies more efficiently, using irredundant equivalent subsets computed by the algorithm in Figure 9.

Future research should be concerned with the development of algorithms that can learn stable CI structures from data. We believe that methods similar to those used to learn the structure of Markov networks, harnessing independence tests and inference rules[14], can be applied to this problem. An additional interesting research challenge would be *probabilistic inference* in the context of stable conditional independence.

## References


[1] P. R. de Waal, L. C. van der Gaag, Stable independence and complexity of representation, in: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, 2004, pp. 112–119.

[2] F. Matúš, Ascending and descending conditional independence relations, in: Transactions of the 11th Prague Conference on Information Theory, 1992, pp. 189–200.

[3] P. R. de Waal, L. C. van der Gaag, Stable independence in perfect maps, in: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, 2005, pp. 161–168.

[4] D. Geiger, J. Pearl, Logical and algorithmic properties of conditional independence and graphical models, The Annals of Statistics 21 (4) (1993) 2001–2021.

[5] M. Niepert, D. V. Gucht, M. Gyssens, On the conditional independence implication problem: A lattice-theoretic approach, in: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, 2008, pp. 435–443.

[6] J. Pearl, A. Paz, Graphoids: Graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z?, in:

Advances in Artificial Intelligence II, Seventh European Conference on Artificial Intelligence, 1986, pp. 357–363.

[7] M. Studený, Probabilistic Conditional Independence Structures, Springer-Verlag, 2005.

[8] M. Studený, Multiinformation and the problem of characterization of conditional independence relations, Problems of Control and Information Theory 18 (1) (1989) 3–16.

[9] J. Peña, R. Nilsson, J. Björkegren, J. Tegnér, Reading dependencies from the minimal undirected independence map of a graphoid that satisfies weak transitivity, in: Proccedings of the European Workshop on Probabilistic Graphical Models, 2006, pp. 247–254.

[10] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann Publishers Inc., 1988.

[11] P. Šimeček, A short note on discrete representability of independence models, in: Proccedings of the European Workshop on Probabilistic Graphical Models, 2006, pp. 287–292.

[12] M. Studený, Conditional independence statements have no complete characterization, in: Proceedings of the 11th Prague Conference on Information Theory, 1993, pp. 377–396.

[13] P. Liberatore, Redundancy in logic i: Cnf propositional formulae, Artif. Intell. 163 (2) (2005) 203–232.

[14] P. Gandhi, F. Bromberg, D. Margaritis, Learning markov network structure using few independence tests, in: SIAM International Conference on Data Mining, 2008, pp. 680–691.