

On the completeness of the semigraphoid axioms for deriving arbitrary from saturated conditional independence statements

Marc Gyssens^{a,*}, Mathias Niepert^b, Dirk Van Gucht^c

^aHasselt University and Transnational Univ. of Limburg, Faculty of Sciences, Martelarenlaan 42, B-3500 Hasselt, Belgium

^bUniv. of Washington, Dept. of Computer Science & Engineering, Paul G. Allen Center, AC101, 185 Stevens Way, Seattle, WA 98195-2350, USA

^cIndiana University, Computer Science Division, Lindley Hall, 150 S Woodlawn Ave., Bloomington, IN 47405, USA

Abstract

Conditional independence (CI) statements occur in several areas of computer science and artificial intelligence, e.g., as embedded multivalued dependencies in database theory, disjunctive association rules in data mining, and probabilistic CI statements in probability theory. Although, syntactically, such constraints can always be represented in the form $I(A, B|C)$, with A , B , and C subsets of some universe S , their semantics is very dependent on their interpretation, and, therefore, inference rules valid under one interpretation need not be valid under another. However, all aforementioned interpretations obey the so-called semigraphoid axioms. In this paper, we consider the restricted case of deriving arbitrary CI statements from so-called saturated ones, i.e., which involve all elements of S . Our main result is a necessary and sufficient condition under which the semigraphoid axioms are also complete for such derivations. Finally, we apply these results to the examples mentioned above to show that, for these semantics, the semigraphoid axioms are both sound and complete for the derivation of arbitrary CI statements from saturated ones.

Keywords: conditional independence statement, saturated conditional independence statement, axiom system, semigraphoid axioms, soundness, completeness

1. Introduction

In numerous areas of computer science, the presence of constraints allows problems to be “decomposed” into simpler ones. Often, these constraints are *conditional independence (CI) statements* of the form $I(A, B|C)$, with A , B , and C pairwise disjoint subsets of some finite universe S [1].

In database theory, S could be a relation schema, and $I(A, B|C)$ an *embedded multivalued dependency* [2], meaning that the projection of the relation onto $A \cup B \cup C$ can be losslessly decomposed into its projections onto $A \cup C$ and $B \cup C$.

In data mining, S could be a set of items, and $I(A, B|C)$ a *disjunctive association rule* [3] meaning that C can only be contained in a basket if either A or B is contained in that basket.

In probability theory, S could be a set of variables, and $I(A, B|C)$ a *probabilistic CI statement* [4–6]. A probability distribution P satisfies $I(A, B|C)$ if A and B are independent conditional upon C . Because reasoning over the full joint probability distribution is almost always intractable, the presence of probabilistic CI statements may facilitate the decomposition of joint probability distributions into smaller parts which are then processed in sophisticated ways to compute a-posteriori probabilities.

Other examples identified by Studený [1] of areas in which conditional independence constraints arise, include the theory of ordinal conditional functions [7], the Dempster-Shafer theory of belief functions [8, 9], and possibility theory [10].

Sayrafi and two of the present authors introduced measure-based constraints [11, 12], which encompass several of the above-cited examples (see also the work of Dalkilic and Robertson [13] and Lee [14]). A measure M is an increasing supermod-

*Corresponding author.

Email addresses: marc.gyssens@uhasselt.be (Marc Gyssens), mniepert@cs.washington.edu (Mathias Niepert), vgucht@cs.indiana.edu (Dirk Van Gucht)

ular function associating nonnegative real numbers¹ to the subsets of a universe S ; M satisfies the *secondary constraint* $I(A, B|C)$ if $M(A \cup B \cup C) + M(C) = M(A \cup C) + M(B \cup C)$.

In view of this wide range of applications, a deeper theoretical understanding of the mathematical and algorithmic properties of conditional independence is required. Especially Studený [15] has brought this issue to the forefront, leading to an impressive body of work on algebraic representations of conditional independence structures with links to algebraic geometry [15, 16], supermodular functions on sets, and algorithms for reasoning with conditional independencies [17–19]. A central notion in reasoning about conditional independence is the CI implication problem, i.e., deciding whether a set of CI statements implies a single CI statement relative to the semantics given to CI statements in the application under consideration. This was at the center of a study by the present authors and Sayrafi [19], in which soundness and completeness of axiom systems for CI implication was investigated.

As was already observed by Studený [1], soundness and completeness of such systems critically depends on the application under consideration. Also, it is known, e.g., that, for probabilistic conditional independence, no finite, sound and complete inference system exists [20] even though it remains open whether the probabilistic CI implication problem for the class of all discrete probability measures is decidable. For the case of embedded multivalued dependencies, it has even been proven that the implication problem is undecidable [21, 22].

Despite all these differences, there is a finite inference system that is *sound* for all applications mentioned above, namely the *semigraphoid* axiom system [5], which is here referred to as System \mathcal{G} . The present authors and Sayrafi [19] investigated extensions of the System \mathcal{G} , and proposed a finite inference system referred to as System \mathcal{A} . It was shown that System \mathcal{A} , although not sound, is complete for the general probabilistic implication problem. What makes System \mathcal{A} so attractive, also beyond the case of probabilistic conditional independence, is the existence of a set-theoretical characterization of derivability under System \mathcal{A} [19], in terms of so-called meet semi-lattices [23].

¹ $M : 2^S \rightarrow \mathbb{R}^{\geq 0}$ is *increasing* if $M(A \cup B) \geq M(A)$ and *supermodular* if $M(A \cup B \cup C) + M(C) \geq M(A \cup C) + M(B \cup C)$. By swapping the inequalities, decreasing supermodular, increasing submodular, and decreasing supermodular measures are obtained, which lead to essentially the same theory.

The present authors and Sayrafi [19] have also looked to cases where some CI statements are *saturated*. Saturated CI statements involve all variables under consideration. In many cases, restricting the CI implication problem to *saturated* CI statements makes it not only decidable, but also axiomatizable, among other very desirable properties². This is, e.g., the case for saturated embedded multivalued dependencies (called “multivalued dependencies” for short) and saturated probabilistic CI statements. For the latter case, it is well known that the semigraphoid System \mathcal{G} is sound and complete for the derivation of saturated CI statements from saturated CI statements. Unfortunately, System \mathcal{G} also allows for the derivation of unsaturated CI statements from saturated ones. To circumvent this caveat, Malvestuto [26] and Geiger and Pearl [27] proposed an alternative sound and complete set of inference rules for this purpose, referred to as System \mathcal{S} , and which is subsumed by the semigraphoid rules, but which has the additional advantage that all the intermediate CI statements derived are also saturated. The present authors and Sayrafi [19] were able to generalize this result by showing that it follows from their theoretical framework that System \mathcal{A} is sound and complete for the derivation of *arbitrary* CI statements from saturated ones.

Since in all applications mentioned above, System \mathcal{G} is sound for the CI implication problem—and hence in particular for the derivation of arbitrary CI statements from saturated ones, the above result begs the question under which conditions System \mathcal{G} is also *complete* for the derivation of arbitrary CI statement from saturated ones. It is shown that a necessary and sufficient condition for this to be the case is that System \mathcal{A} is complete for the derivation of arbitrary CI statements from saturated ones, *and this regardless of the context in which the CI statements are interpreted*. In addition, Theorem 1 and Corollary 1) state an analogous characterization for soundness. These results hinge on a “normal form” result for such derivations (Proposition 2) which says that whenever an arbitrary CI statement can be derived from saturated ones under System \mathcal{A} , a “saturated version” of that CI statement can be derived under System \mathcal{S} . We show that, for our initial examples, the results of this paper yield soundness and completeness of both Systems \mathcal{A} and \mathcal{G} for deriving arbitrary CI statements from saturated ones (Proposition 3).

²See [24, 25] for examples of very recent work in this area.

2. Preliminaries

Throughout the paper, we consider a finite universe S , which we shall often leave implicit. With regard to set notation, we often write AB for the union $A \cup B$, ab for the set $\{a, b\}$, and a for the singleton set $\{a\}$ if no confusion is possible. For $A \subseteq S$, we write \bar{A} for $S - A$, the complement of A with respect to S .

We begin by defining conditional independence (CI) statements as a purely syntactic notion, without being concerned with the semantics.

Definition 1. A *conditional independence (CI) statement* is an expression $I(A, B|C)$ where A , B , and C are pairwise disjoint subsets of S . If $ABC = S$, $I(A, B|C)$ is *saturated*.

Example 1. Let $S = abcdefgh$ be the universe. Then $I(abcdef, bgh|c)$ and $I(a, b|c)$ are both examples of CI statements. The former is saturated, whereas the latter is not.

Finite inference systems are frequently used to decide logical entailment of such statements at the syntactic level. If \mathcal{I} is such an inference system, \mathcal{C} a set of CI statements, and c a single CI statement, we denote by $\mathcal{C} \vdash_{\mathcal{I}} c$ that c is derivable from \mathcal{C} under the inference rules of System \mathcal{I} .

In this paper, we consider System \mathcal{A} , introduced by the present authors and Sayrafi [19] and shown in Figure 1, and System \mathcal{G} , the semigraphoid axioms of Pearl [5], shown in Figure 2. We also consider System \mathcal{S} of Malvestuto [26] and Geiger and Pearl [27], shown in Figure 3, geared to the case where only *saturated* CI statements are considered. In all three systems, sets occurring within a single CI statement are considered pairwise disjoint.

We next illustrate derivations under System \mathcal{A} .

Example 2. Let $S = abcdefgh$, and let

$$\mathcal{C} = \{I(abcdef, bgh|c), I(bd, efg|ach), I(ad, efg|bch)\}.$$

Then $\mathcal{C} \vdash_{\mathcal{A}} I(d, e|ch)$. To see this, observe that, from $I(abcdef, bgh|c)$, we can derive $I(a, b|c)$ by repeated applications of *symmetry* and *decomposition*, and $I(a, b|ch)$ (1) by *strong union*. Similarly, from $I(bd, efg|ach)$, we can derive $I(d, e|ach)$ (2), and from $I(ad, efg|bch)$, we can derive $I(d, e|bch)$ (3), again by repeated applications of *symmetry* and *decomposition*. Finally, from (1), (2), and (3), we can derive $I(d, e|ch)$ by *strong contraction*.

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

Figure 1: The inference rules of System \mathcal{A} .

Triviality: $I(A, \emptyset C)$ Symmetry: $I(A, B C) \rightarrow I(B, A C)$ Decomposition: $I(A, BD C) \rightarrow I(A, B C)$ Contraction: $I(A, B CD) \ \& \ I(A, D C) \rightarrow I(A, BD C)$ Weak union: $I(A, BD C) \rightarrow I(A, B CD)$
--

Figure 2: The inference rules of the semigraphoid System \mathcal{G} .

Triviality: $I(A, \emptyset C)$ Symmetry: $I(A, B C) \rightarrow I(B, A C)$ Weak Contraction: $I(A, B CD) \ \& \ I(AB, D C) \rightarrow I(A, BD C)$ Weak union: $I(A, BD C) \rightarrow I(A, B CD)$

Figure 3: The inference rules of System \mathcal{S} .

Systems \mathcal{A} , \mathcal{G} , and \mathcal{S} are related as follows.

Proposition 1. [19, 26, 27]

1. If \mathcal{C} is a set of CI statements and c is a single CI statement, then (a) $\mathcal{C} \vdash_{\mathcal{G}} c$ implies $\mathcal{C} \vdash_{\mathcal{A}} c$ and (b) $\mathcal{C} \vdash_{\mathcal{S}} c$ implies $\mathcal{C} \vdash_{\mathcal{G}} c$.
2. If \mathcal{C} is a set of saturated CI statements and c is a single saturated CI statement, then $\mathcal{C} \vdash_{\mathcal{G}} c$ if and only if $\mathcal{C} \vdash_{\mathcal{S}} c$.

Observe that, in a derivation of saturated CI statements from saturated CI statements under System \mathcal{S} , only saturated CI statements are generated as intermediate constraints, which is not guaranteed in derivations using Systems \mathcal{A} or \mathcal{G} . The appealing property of System \mathcal{A} , however, is that derivability can be characterized in set-theoretical terms [19].

In the context of a concrete semantics for CI statements, derivations can be given a meaning. Given such a semantics \mathfrak{S} for CI statements, a set of CI statements \mathcal{C} , and a single CI statement c , we say that \mathcal{C} \mathfrak{S} -implies c , denoted $\mathcal{C} \models_{\mathfrak{S}} c$, if c holds whenever all CI statements of \mathcal{C} hold.

An inference system \mathcal{I} is *sound* for \mathfrak{S} , if, for each set of CI statements \mathcal{C} , and for each single CI statement c , $\mathcal{C} \vdash_{\mathcal{I}} c$ implies $\mathcal{C} \models_{\mathfrak{S}} c$; \mathcal{I} is *complete* for \mathfrak{S} , if, for each set of CI statements \mathcal{C} , and for each single CI statement c , $\mathcal{C} \models_{\mathfrak{S}} c$ implies $\mathcal{C} \vdash_{\mathcal{I}} c$.

Example 3. If CI statements are interpreted as probabilistic conditional independence constraints in the context of discrete probability measures, it is known that (1) System \mathcal{A} is complete; (2) System \mathcal{G} is sound; and (3) System \mathcal{A} , System \mathcal{G} , and System \mathcal{S} are sound and complete if all constraints under consideration are saturated. [19, 26, 27].

3. Main result

Our main result is concerned with the derivation of *arbitrary* CI statements from a set of saturated CI statements. We show that, in this context, (1) soundness of System \mathcal{A} is equivalent to soundness of System \mathcal{G} , and (2) completeness of System \mathcal{A} is equivalent to completeness of System \mathcal{G} , independent of the semantics given to the CI statements.

This result relies on Proposition 2. To state it, we first need the following notion.

Definition 2. A *saturation* of a CI statement $I(A, B|C)$ is a saturated CI statement $I(A', B'|C)$ for which $A \subseteq A'$ and $B \subseteq B'$.

Example 4. In Example 1, $S = abcdefgh$. Hence, $I(abcdef, bgh|c)$ is a saturation of $I(a, b|c)$.

Proposition 2. *Let \mathcal{C} be a set of saturated CI statements, and let c be a single, arbitrary CI statement. If $\mathcal{C} \vdash_{\mathcal{A}} c$, then there exists a saturation c' of c such that $\mathcal{C} \vdash_{\mathcal{S}} c'$.*

PROOF. The proof is by structural induction. The base case, where \mathcal{C} contains a saturation of c , trivially satisfies Proposition 2. In the inductive step, we deal with each of the six inference rules of System \mathcal{A} . In each case, the inductive hypothesis is that each of the CI statements in the left-hand side of the rule has a saturation that can be derived from \mathcal{C} under System \mathcal{S} .

For the *triviality*, *symmetry*, and *decomposition* rules, the induction step is straightforward. We now deal with the remaining three inference rules of System \mathcal{A} in further detail. For sets $X, Y \subseteq S$, X_Y is short for $X \cap Y$, and X^Y is short for $X - Y$.

- *Contraction.* Consider the *contraction* rule $I(A, B|CD) \& I(A, D|C) \rightarrow I(A, BD|C)$ of System \mathcal{A} , and let $I(AE, BF|CD)$ and $I(AG, DH|C)$ be saturations of the CI statements in the left-hand side. Hence, $ABCDEF = S = ACDGH$, so $BEF = GH$, $G = B_G E_G F_G$, and $H = B_H E_H F_H$. From $I(AE, BF|CD)$, we derive, by *symmetry* and *weak union*, $I(AE_G, B_G F_G|CDH)$, and, from this and $I(AG, DH|C)$, we derive by, *weak contraction*, $I(AE_G, B_G F_G DH|C)$, a saturation of $I(A, D|C)$.
- *Strong union.* Consider the *strong union* rule $I(A, B|C) \rightarrow I(A, B|CD)$ of System \mathcal{A} . Let $I(AE, BF|C)$ be a saturation of $I(A, B|C)$. Hence, $D \subseteq EF$. From $I(AE, BF|C)$, we derive, by *symmetry* and *weak union*, $I(AE^D, BF^D|CD)$, a saturation of $I(A, B|CD)$.
- *Strong contraction.* Consider the *strong contraction* rule $I(A, B|C) \& I(D, E|AC) \& I(D, E|BC) \rightarrow I(D, E|C)$ of System \mathcal{A} , and let (1) $I(AF, BG|C)$, (2) $I(DH, EK|AC)$, and (3) $I(DL, EM|BC)$ be saturations of the CI statements in the left-hand side. Figure 4 shows the mutual position of all sets involved in these saturations. From (3), we derive, by *weak union* and *symmetry*, $I(D_F L_A L_F, E_F M_A M_F|CBG)$ (4), and, from (4) and (1), we derive, by *weak contraction* and *symmetry*, $I(D_G E_B L_G M, D_F L_A L_F|C)$ (5). From (2), we derive, by *weak union* and *symmetry*, $I(D_G H_B H_G, E_G K_B K_G|ACF)$ (6), and from (6) and (1), we derive, by *symmetry* and *weak contraction*, $I(AD_F E_H F_K, D_G H_B H_G|C)$ (7). From (7), we derive, by *weak union*, $I(EK_B K_G M_A M_F, D_G H_B H_G|C D_F L_A L_F)$ (8), and from (8) and (5), we finally derive, by *weak contraction* and *symmetry*, $I(DH_B H_G L_A L_F, EK_B K_G M_A M_F|C)$, which is a saturation of $I(D, E|C)$.

We next illustrate Proposition 2 and its proof.

Example 5. Let $S = abcdefgh$. The set \mathcal{C} in Example 2 consists of saturated CI statements, whereas $I(d, e|ch)$ is not. We already established in Example 2 that $\mathcal{C} \vdash_{\mathcal{A}} I(d, e|ch)$. We now show that, from \mathcal{C} , we can derive a saturation of $I(d, e|ch)$ under System \mathcal{S} .

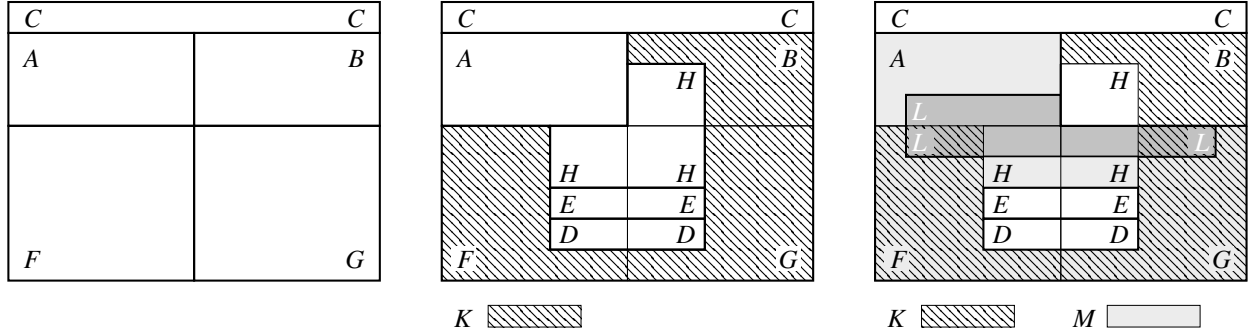


Figure 4: In the inductive proof of Proposition 2, we have in the case of the *strong contraction* rule that $S = ABCFG$ (left), $S = ACDEHK$ (middle), and $S = BCDELM$ (right). This last diagram also shows the mutual position of all these sets.

In Example 2, we derived $I(a, b|c)$ from $I(ade f, bgh|c)$, $I(d, e|ch)$ (2) from $I(bd, efg|ach)$ (2'), and $I(d, e|bch)$ (3) from $I(ad, efg|bch)$ (3'). In each of these derivations, the former CI statement is a saturation of the second one, and, hence, no further work is needed here. Next, from $I(a, b|c)$, we derived $I(a, b|ch)$ by *strong union*. In the simulation in the proof of Proposition 2, we derive from the saturation $I(ade f, bgh|c)$ of $I(a, b|c)$ the saturation $I(ade f, bg|ch)$ (1') of $I(a, b|ch)$ (1), by *weak union*. In the final step of the inference in Example 2, we derived from (1), (2), and (3) the CI statement $I(d, e|ch)$, by *strong contraction*. It remains to show that, from their saturations (1'), (2') and (3'), we can derive a saturation of $I(a, b|ch)$ under System \mathcal{S} . To this end, we closely follow the derivation exhibited in the part “Strong contraction” of the proof of Proposition 2. From (3'), we derive by *weak union* and *symmetry*, $I(ad, ef|bcgh)$ (4'), and from (4') and (1'), we derive, by *weak contraction* and *symmetry*, $I(befg, ad|ch)$ (5'). From (2'), we derive, by *weak union* and *symmetry*, $I(b, g|acdefh)$ (6'), and from (6') and (1'), we derive, by *symmetry* and *weak contraction*, $I(ade fg, b|ch)$ (7'). From (7'), we derive, by *weak union*, $I(efg, b|acdh)$ (8'), and, from (8') and (5'), we finally derive, by *weak contraction* and *symmetry*, $I(abd, efg|ch)$, which is a saturation of $I(d, e|ch)$.

The proof of Proposition 2 reveals that, for each rule of System \mathcal{A} , there exists a saturation of its right-hand side which can be derived under System \mathcal{S} from saturations of its left-hand side. By Proposition 1, this saturation can also be derived under System \mathcal{A} , which can be expressed as a condition on the sets involved [19]. This observation allowed us to significantly reduce the number of possible saturations to be considered. In this regard, it

is noteworthy that, for the *strong contraction* rule, the saturation of $I(D, E|C)$ obtained in the proof of Proposition 2 is actually the *only* one that can be derived from $I(AF, BG|C)$, $I(DH, EK|AC)$, and $I(D, E|AC)$ under System \mathcal{A} , suggesting that there is a certain “tightness” to Proposition 2.

Proposition 2 also reveals that, in order to derive arbitrary CI statements from saturated ones under System \mathcal{A} or \mathcal{G} , one can stay within the framework of saturated CI statements until the very last step, which consists of removing the extraneous variables. In that sense, Proposition 2 can be seen as a *normal form* for such derivations.

We now bootstrap Proposition 2 as follows.

Theorem 1. *Let \mathcal{C} be a set of saturated CI statements, and let c be a single, arbitrary CI statement. Then $\mathcal{C} \vdash_{\mathcal{A}} c$ if and only if $\mathcal{C} \vdash_{\mathcal{G}} c$.*

PROOF. The “only if” follows from Proposition 1. To see the “if,” assume $\mathcal{C} \vdash_{\mathcal{A}} c$. By Proposition 2, there is a saturation c' of c such that $\mathcal{C} \vdash_{\mathcal{S}} c'$, and, hence, by Proposition 1, $\mathcal{C} \vdash_{\mathcal{G}} c'$. Since $\{c'\} \vdash_{\mathcal{G}} c$, by *decomposition* and *symmetry*, $\mathcal{C} \vdash_{\mathcal{G}} c$.

Example 6. Let $S = abcdefgh$. In Example 5, we showed that $\mathcal{C} \vdash_{\mathcal{S}} I(abd, efg|ch)$, with \mathcal{C} the set of CI statements first introduced in Example 2. By Proposition 1, $\mathcal{C} \vdash_{\mathcal{G}} I(abd, efg|ch)$ ³. Finally, from $I(abd, efg|ch)$ we derive $I(d, e|ch)$ by repeated applications of the *symmetry* and *decomposition* rules.

The following corollary is now immediate.

³A derivation under System \mathcal{G} can easily be constructed from a derivation under System \mathcal{S} , since the *weak contraction* rule can be simulated by the *contraction rule*, after first applying the *symmetry*, the *decomposition*, and again the *symmetry* rule to the second CI statement.

Corollary 1. *For every semantics for CI statements, we have the following:*

1. \mathcal{G} is sound for deriving arbitrary CI statements from saturated ones if and only if \mathcal{A} is;
2. \mathcal{G} is complete for deriving arbitrary CI statements from saturated ones if and only if \mathcal{A} is.

4. Applications

We now intend to show how the main results of Section 3 (notably, Theorem 1 and Corollary 1) may be applied to different semantics for CI statements to obtain that both Systems \mathcal{A} and \mathcal{G} are sound and complete for deriving arbitrary CI statements from saturated ones. We illustrate this for three examples considered in Section 1, namely embedded multivalued database dependencies, disjunctive association rules in data mining, and probabilistic conditional independence. The following two lemmas are immediate consequences of Corollary 1.

Lemma 1. *If CI statements are interpreted as embedded multivalued database dependencies, disjunctive association rules in data mining, or conditional independence statements over discrete probability measures, then System \mathcal{A} is sound for deriving arbitrary CI statements from saturated ones.*

PROOF. System \mathcal{G} is sound in all these cases [1].

Lemma 2. *For every semantics for CI statements, we have the following. If System \mathcal{G} is sound and System \mathcal{A} complete for deriving arbitrary CI statements from saturated ones, then both Systems \mathcal{G} and \mathcal{A} are sound and complete for deriving arbitrary CI statements from saturated ones.*

As already pointed out in Section 1, all three semantics considered here are examples of *measure-based semantics*⁴. As a consequence, Corollary 32 in [19] holds, and we may derive the following.

Lemma 3. *If CI statements are interpreted as embedded multivalued database dependencies, disjunctive association rules in data mining, or conditional independence statements over discrete probability measures, then the following holds. If System \mathcal{A} is sound for deriving arbitrary CI statements from*

⁴We refer to earlier work of Sayrafi and the present authors for more details on measure-based constraints [12, 19].

saturated ones, and complete for deriving saturated CI statements from saturated ones, then System \mathcal{A} is also complete for deriving arbitrary CI statements from saturated ones.

Combining Lemmas 1, 2, and 3, we finally obtain the following.⁵

Proposition 3. *If CI statements are interpreted as embedded multivalued database dependencies, disjunctive association rules in data mining, or conditional independence statements over discrete probability measures, then both Systems \mathcal{A} and \mathcal{G} are sound and complete for deriving arbitrary CI statements from saturated ones.*

PROOF. By Lemmas 1, 2, and 3, it suffices to show that System \mathcal{A} is complete for deriving saturated CI statements from saturated ones. By Proposition 1, it suffices to show that System \mathcal{S} is already complete for this purpose.⁶ By a reduction argument, one can show that the implication problem for saturated disjunctive association rules coincides with that of (full) multivalued database dependencies.⁷ Finally, System \mathcal{S} is complete for the implication problem of both multivalued database dependencies and saturated conditional independence statements on discrete probability measures [26, 27].

Finally, one may wonder if our results have any bearing on the graphical representation of CI statements in the probabilistic case, and, in particular, on the separation criterion. This is not the case, however. Indeed, the closure of a set of saturated probabilistic CI statements with respect to the *symmetry*, *intersection*⁸, and *weak union* rules coincides with the separation criterion in an undirected graph representing the closure of that set of CI statements [27, 29, 30]. However, the intersection rule is not generally sound for saturated probabilistic CI statements and, hence, graph separation does not coincide with the closure under System \mathcal{G} .

⁵Proposition 3 for the case of probabilistic conditional independence is Theorem 41 in [19], which relied on the hitherto unpublished results above.

⁶As noted in Example 3, System \mathcal{A} is known to be complete for general probabilistic CI statements, so, for this case, we are already done. Here, we intend to give a more generic argument that is applicable to other semantics as well.

⁷The reduction shows that the implication problem for saturated disjunctive association rules coincides with that of the so-called *multivalued domain dependencies*, also known as *pseudo-multivalued dependencies* [12, 28]. For these constraints, the implication problem coincides with that of multivalued database dependencies [28].

⁸The rule $I(A, B|CD) \ \& \ I(A, C|BD) \rightarrow I(A, BC|D)$.

5. Conclusions and future work

In this short note, we studied the implication of arbitrary CI statements from saturated ones. Our general results hold regardless of the particular semantics given to CI statements. We also applied these results to some specific semantics, notably embedded multivalued database dependencies, disjunctive association rules in data mining, and probabilistic conditional independence, to show that, in these cases, both Systems \mathcal{A} and \mathcal{G} are sound and complete for deriving arbitrary CI statements from arbitrary ones.

All these examples are measure-based constraints [12, 19]. The present authors believe that the techniques used in Section 4 are applicable in a generic way to a much larger class of measure-based constraints. This is the subject of future research.

References

- [1] M. Studený, Formal properties of conditional independence in different calculi of AI, in: *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, in: *Lecture Notes in Computer Science*, vol. 747, 1993, pp. 341–348.
- [2] R. Fagin, Multivalued dependencies and a new normal form for relational databases, *ACM T. Database Syst.* 2 (1977) 262–278.
- [3] A. Bykowski, C. Rigotti, A condensed representation to find frequent patterns, in: *Proc. of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ACM Press, New York, 2001, pp. 267–273.
- [4] A. Dawid, Conditional independence in statistical theory, *J. R. Stat. Soc.* 41B (1979) 1–31.
- [5] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers Inc., 1988.
- [6] W. Spohn, Stochastic independence, casual independence and shieldability, *J. Philos. Log.* 9 (1980) 73–99.
- [7] W. Spohn, Ordinal conditional functions: a dynamic theory of epistemic states, in: W. L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, vol. II, Kluwer Academic Publishers, Dordrecht, 1988, pp. 105–134.
- [8] A. Dempster, A generalization of Bayesian inference, *J. R. Stat. Soc.* 30B (1968) 205–247.
- [9] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [10] L. Zadeh, Fuzzy sets as the basis for a theory of possibility, *Fuzzy Sets and Systems* 1 (1978) 3–28.
- [11] B. Sayrafi, D. Van Gucht, Differential constraints, in: *Proc. of the 24th ACM Symposium on Principles of Database Systems*, 2005, pp. 348–357.
- [12] B. Sayrafi, D. Van Gucht, M. Gyssens, The implication problem for measure-based constraints, *Inf. Syst.* 33 (2008) 221–239.
- [13] M. M. Dalkilic and E. L. Robertson, Information dependencies, in: *Proc. of the 19th ACM Symposium on Principles of Database Systems*, 2000, pp. 245–253.
- [14] T. T. Lee, An information-theoretic analysis of relational databases. Part I: data dependencies and information metric, *IEEE Trans. Software Eng.* SE-13 (1987) 1049–1061.
- [15] M. Studený, *Probabilistic Conditional Independence Structures*, Springer-Verlag, 2005.
- [16] R. Hemmecke, J. Morton, A. Shiu, B. Sturmfels, O. Wienand, Three counter-examples on semi-graphoids, *Combin. Probab. Comput.* 17 (2008) 239–257.
- [17] R. Bouckaert, M. Studený, Racing algorithms for conditional independence inference, *Int. J. Approx. Reason.* 45 (2007) 386–401.
- [18] R. Bouckaert, R. Hemmecke, S. Lindner, M. Studený, Efficient algorithms for conditional independence inference, *J. Mach. Learn. Res.* 11 (2010) 3453–3479.
- [19] M. Niepert, M. Gyssens, B. Sayrafi, D. Van Gucht, On the conditional independence implication problem: a lattice-theoretic approach, *Artif. Intell.* 202 (2013) 29–51.
- [20] M. Studený, Conditional independence statements have no complete characterization, in: *Transactions of the 11th Prague Conference on Information Theory*, 1992, pp. 377–396.
- [21] C. Herrmann, On the undecidability of implications between embedded multivalued dependencies, *Inf. Comput.* 122 (1995) 221–235.
- [22] C. Herrmann, Corrigendum to “On the undecidability of implications between embedded multivalued database dependencies” [*Inf. Comput.* 122 (1995) 221–235], *Inf. Comput.* 204 (2006) 1847–1851.
- [23] G. A. Grätzer, *General Lattice Theory*, 2nd Edition, Birkhäuser-Verlag, 1998.
- [24] S. Link, Sound approximate reasoning about saturated conditional probabilistic independence under controlled uncertainty, *J. Appl. Log.* 11 (2013) 309–327.
- [25] S. Link, Reasoning about saturated conditional independence under uncertainty: axioms, algorithms, and Levesque’s situations to the rescue, in: *Proc. of the AAAI Conference on Artificial Intelligence*, 2013, pp. 612–618.
- [26] F. M. Malvestuto, A unique formal system for binary decompositions of database relations, probability distributions, and graphs, *Inform. Sci.* 59 (1992) 21–52.
- [27] D. Geiger, J. Pearl, Logical and algorithmic properties of conditional independence and graphical models, *Ann. Statist.* 21 (1993) 2001–2021.
- [28] M. Gyssens, J. Paredaens, Another view of functional and multivalued dependencies in the relational database model, *Int. J. Parallel Prog.* 12 (1983) 247–267.
- [29] J. Pearl and T. Verma, The logic of representing dependencies by directed graphs, in: *Proc. of the AAAI Conference on Artificial Intelligence*, 1987, pp. 374–379.
- [30] S. M. M. S. Renooij, H. J. M. Tabachneck-Schijf (Ed.), *Proc. of the 6th UAI Bayesian Modelling Applications Workshop*, 2008.