

# A logical approach to context-specific independence

Jukka Corander<sup>a,c</sup>, Antti Hyttinen<sup>b</sup>, Juha Kontinen<sup>a,\*</sup>, Johan Pensar<sup>a,d</sup>,  
Jouko Väänänen<sup>a,e</sup>

<sup>a</sup> Department of Mathematics and Statistics, University of Helsinki, Finland

<sup>b</sup> HIIT, Department of Computer Science, University of Helsinki, Finland

<sup>c</sup> Department of Biostatistics, University of Oslo, Norway

<sup>d</sup> Department of Mathematics and Statistics, Åbo Akademi University, Finland

<sup>e</sup> Institute for Logic, Language and Computation, University of Amsterdam, Netherlands



## ARTICLE INFO

### Article history:

Available online 10 April 2019

### MSC:

03B48

03B60

03B70

62-09

### Keywords:

Directed acyclic graph

Graphical model

Context-specific independence

Implication problem

Team semantics

## ABSTRACT

Directed acyclic graphs (DAGs) constitute a qualitative representation for conditional independence (CI) properties of a probability distribution. It is known that every CI statement implied by the topology of a DAG is witnessed over it under a graph-theoretic criterion of  $d$ -separation. Alternatively, all such implied CI statements are derivable from the local independencies encoded by a DAG using the so-called semi-graphoid axioms. We consider Labeled Directed Acyclic Graphs (LDAGs) modeling graphically scenarios exhibiting context-specific independence (CSI). Such CSI statements are modeled by labeled edges, where labels encode contexts in which the edge vanishes. We study the problem of identifying all independence statements implied by the structure and the labels of an LDAG. We show that this problem is coNP-hard for LDAGs and formulate a sound extension of the semi-graphoid axioms for the derivation of such implied independencies. Finally we connect our study to certain qualitative versions of independence ubiquitous in database theory and teams semantics.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Bayesian networks [20] are a popular tool for modeling complex multivariate systems. The basis of a Bayesian network is a directed acyclic graph (DAG) in which the nodes represent random variables and the directed edges represent direct dependencies between the variables. On the other hand, missing edges give rise to statements of conditional independence (CI) which can be verified directly from the graph using the graph-theoretic criterion called  $d$ -separation [30].

\* Corresponding author.

E-mail addresses: [jukka.corander@helsinki.fi](mailto:jukka.corander@helsinki.fi) (J. Corander), [antti.hyttinen@helsinki.fi](mailto:antti.hyttinen@helsinki.fi) (A. Hyttinen), [juha.kontinen@helsinki.fi](mailto:juha.kontinen@helsinki.fi) (J. Kontinen), [johan.pensar@helsinki.fi](mailto:johan.pensar@helsinki.fi) (J. Pensar), [jouko.vaananen@helsinki.fi](mailto:jouko.vaananen@helsinki.fi) (J. Väänänen).

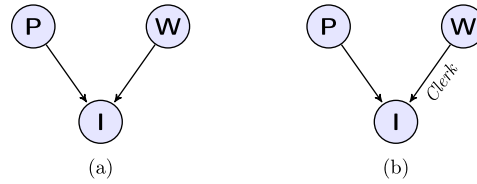


Fig. 1. (a) DAG and (b) LDAG over three variables.

In certain situations, restrictions implied by conditional independence may be too rigid to fully capture the structure of a system. As a simple example, consider the model in Fig. 1(a) which consists of three variables representing *Profession* ( $P$ ), *Weather* ( $W$ ) and *Income* ( $I$ ). Edge  $P \rightarrow I$  implies that a person's income is directly affected by his or her profession, which we for simplicity assume is either farmer or office clerk. Given that a person is a farmer, it is easy to imagine that the weather may affect the income of that person, this is represented by edge  $W \rightarrow I$ . On the other hand, if a person is a clerk by profession, then it is just as easy to imagine that the weather will not affect the income, that is, the person's income is independent of weather. This particular type of independencies arises very naturally in various real-world situations (see [33]), yet it is hidden by traditional Bayesian networks, since an edge must be either present or absent.

To remedy this, Boutilier et al. [1] introduced and formalized the notion of context-specific independence (CSI). More specifically, they showed that a certain class of local CSI statements is particularly convenient to include in the Bayesian network framework. A local CSI statement basically corresponds to the influence of an edge vanishing in a certain context. To represent general CSI-based dependence structures, Pensar et al. [31] introduced the class of labeled directed acyclic graphs (LDAGs) which capture the local CSI statements through labels assigned to the edges. For example, the scenario in our previous example can be fully captured by the LDAG in Fig. 1(b). The label on edge  $W \rightarrow I$  can be thought of as an indicator function removing the edge, or its influence, when the context is satisfied, that is, when  $P$  takes on the value *Clerk*. Analogously to the  $d$ -separation criterion, (a subset of) non-local CSI statements can be verified using a concept called CSI-separation. However, as opposed to  $d$ -separation for DAGs, CSI-separation is known not to be able to detect all CI-statements of an LDAG (see Example 3), and it is an open question to find such a complete criterion for LDAGs.

Conditional independence has also been given a qualitative characterization in terms of logical axioms (see Section 3). The semi-graphoid axioms of conditional independence are known to be sound for all distributions, and furthermore correspond exactly to  $d$ -separation in the context of Bayesian networks [11, 36]. On the other hand, the axiomatic approach for the discovery of implied independencies has also been studied in context more general than Bayesian networks (see, e.g., [12,13,26]).

Various notions of independence have a prominent role also in database theory. It is well known that the database constraint called embedded multivalued dependency (EMVD) can be viewed as a qualitative version of conditional independence. The connections between EMVDs and CI statements have been widely studied [22–24,37]. In particular, the completeness of the semi-graphoid axioms for discovering the implied independencies of a DAG remain valid if CI statements are replaced by the corresponding EMVDs and distributions by their supports [12] (see Proposition 18 and Theorem 19). Recently EMVDs have re-emerged as an important logical dependency (independence atom) in the context of team semantics [14]. Team semantics is a novel framework for the study of logics of dependence and independence for which satisfaction of a formula is defined not via single assignments but sets of assignments [35]. There is a very tight connection between EMVDs over database relations and independence atoms  $\vec{x} \perp_{\vec{z}} \vec{y}$  over teams (see, e.g., [16,18]). Informally speaking, the meaning of  $\vec{x} \perp_{\vec{z}} \vec{y}$  is that, when the value of  $\vec{z}$  is fixed, knowing the value of  $\vec{x}$  does not tell us anything new about the value of  $\vec{y}$ . The active development of logics for dependence and independence has also led to fruitful interactions between team semantics and database theory. Furthermore,

the recently defined variants based on multiteams and probabilistic teams offer a flexible framework in which qualitative and probabilistic concepts can be studied simultaneously (see, e.g., [5,6]).

In this article we study the problem of finding a complete criterion for the independence statements implied by an LDAG. For DAGs,  $d$ -separation is known to be a sound and complete method that is also decidable in polynomial time. Our first result shows that for LDAGs this problem is coNP-hard. Our proof is an adaptation of an analogous result of Koller et al. [20] for DAGs with so-called tree-structured CPDs. In Section 3 we generalize the semi-graphoid axioms to CSI-statements and define a sound extension of the semi-graphoid axioms that captures CSI-separation in LDAGs. We also address the problem of finding a complete axiomatization for the independencies implied by an LDAG structure. We show that our novel axiomatization allows us to find (some) of the independencies that are out of reach of CSI-separation but the completeness of the axiomatization is still an open question. In Section 4 we recall some basic facts about team semantics and then define CSI-atoms that are natural qualitative analogues of CSI-statements. Section 4 sets the stage for a systematic study of context specific independence in team semantics and database theory.

Regarding related work, directed graphical models incorporating context-specific independence have been studied extensively over the last few decades [1,3,8,10,31–33]. In general, previous works have mainly focused on either model learning [3,8,31,32] or the task of performing probabilistic inference for a given model [1,10,33]. To assist with model interpretation, the concept of CSI-separation was introduced already in [1], however, its incompleteness was observed later (see, e.g., [20]). There is a vast literature on EMVDs in databases and independence logic in team semantics but surprisingly context-specific independence has not yet been studied in these frameworks. On the other hand, conditional functional and inclusion dependencies have turned out to be useful for various database tasks and have been studied, e.g., in [2,7,25].

## 2. Directed graphical models for statistical independence

### 2.1. Representing conditional independence with DAGs

A Directed Acyclic Graph (DAG)  $G = (\Delta, E)$  is specified by a set of nodes  $\Delta = \{1, \dots, n\}$  and a set of directed edges  $E$  where  $(i, j) \in E$  represents a directed edge from node  $i$  to node  $j$ . The parents of a node  $j$ , denoted by  $\Pi_j$ , is defined as all nodes from which there is a directed edge to node  $j$ , that is,  $\Pi_j = \{i \in \Delta \mid (i, j) \in E\}$ . The descendants of a node  $i$  is all nodes which can be reached from node  $i$  following the direction of the edges.

In a Bayesian network, the nodes of the graph represent random variables  $X_\Delta = \{X_1, \dots, X_n\}$ . As is typical in the graphical model literature, the terms node and variable will occasionally be used interchangeably. We denote by  $P_\Delta$  a probability distribution over the random variables  $X_\Delta$ . Each variable  $X_i$  is assumed to take values in a finite discrete set of outcomes denoted by  $\mathcal{X}_i$ . The joint outcome space of a set of variables  $X_S$ , where  $S \subseteq \Delta$ , is defined as the Cartesian product of the sets  $\mathcal{X}_i$  for  $i \in S$ .

**Definition 1** (*Conditional independence (CI)*). Let  $A$ ,  $B$ , and  $S$  be subsets of  $\Delta$ . The variables  $X_A$  are conditionally independent of  $X_B$  given  $X_S$  if

$$P(X_A = e_A | X_B = e_B, X_S = e_S) = P(X_A = e_A | X_S = e_S),$$

for all  $(e_A, e_B, e_S) \in \mathcal{X}_A \times \mathcal{X}_B \times \mathcal{X}_S$  for which  $P(X_B = e_B, X_S = e_S) > 0$ . This is denoted by  $X_A \perp X_B \mid X_S$ .

A Bayesian network is specified by a pair  $(G, P_\Delta)$  where  $G = (\Delta, E)$  is a DAG and  $P_\Delta$  is a probability distribution satisfying the CI statements encoded by  $G$ . The dependence structure of a DAG is characterized by the so-called local directed Markov property [30], which states that each variable  $X_j$  is conditionally

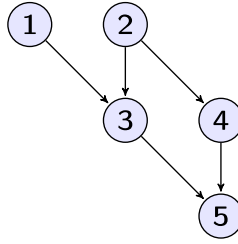


Fig. 2. DAG over five variables.

independent of its non-descendants given its parents  $X_{\Pi_j}$ . Accordingly, the joint probability distribution  $P_\Delta$  can be factorized as

$$P(X_1 = e_1, X_2 = e_2, \dots, X_n = e_n) = \prod_{j=1}^n P(X_j = e_j | X_{\Pi_j} = e_{\Pi_j}), \quad (1)$$

for any  $e_\Delta \in \mathcal{X}_\Delta$ . The joint distribution  $P_\Delta$  can hence be thought of as being constructed from node-wise conditional distributions.

The CI statements of a graph  $G$  determined by the local directed Markov property are called *local* CIs and are denoted by  $I_{loc}(G)$ . However, the set  $I_{loc}(G)$  implies also other *non-local* CIs which can be verified using a graph-theoretic criterion called *d*-separation.

**Definition 2** (*d*-separation). Let  $G = (\Delta, E)$  be a DAG and let  $A$ ,  $B$ , and  $S$  be disjoint subsets of  $\Delta$ . The set  $A$  is *d*-separated from  $B$  by  $S$  if there is no trail in  $G$  from a node in  $A$  to a node in  $B$  along which every node that delivers an arrow (i.e., tail in either direction) is outside of  $S$ , and every node with converging arrows (i.e., heads in both directions) either is or has a descendant in  $S$ .

For a CI statement  $\phi$ , we write  $I_{loc}(G) \models \phi$  if all distributions  $P_\Delta$  that satisfy  $I_{loc}(G)$  also satisfy  $\phi$ , that is,  $\phi$  is implied by  $I_{loc}(G)$ . The following result is due to [11,36].

**Theorem 3** (*Soundness and completeness of d-separation*). Let  $G = (\Delta, E)$  be a DAG and  $P_\Delta$  a distribution satisfying every CI in  $I_{loc}(G)$ . Let  $A$ ,  $B$ , and  $S$  be disjoint subsets of  $\Delta$ . Then it holds that

- if  $A$  is *d*-separated from  $B$  by  $S$  in  $G$ , then  $P_\Delta$  satisfies  $X_A \perp X_B \mid X_S$ ;
- if  $I_{loc}(G) \models X_A \perp X_B \mid X_S$ , then  $A$  is *d*-separated from  $B$  by  $S$  in  $G$ .

We will end this section with an example illustrating the use of *d*-separation.

**Example 1.** Consider the DAG  $G$  in Fig. 2. Note first that  $I_{loc}(G)$  consists of the following five CIs:  $X_1 \perp \{X_2, X_4\}$ ,  $X_2 \perp X_1$ ,  $X_3 \perp X_4 \mid \{X_1, X_2\}$ ,  $X_4 \perp \{X_1, X_3\} \mid X_2$  and  $X_5 \perp \{X_1, X_2\} \mid \{X_3, X_4\}$ , which must hold for any distribution that factorizes according to  $G$ . In addition, we can further infer that the non-local CI  $X_1 \perp X_4 \mid \{X_2, X_3, X_5\}$  must hold in such a distribution since node 1 is *d*-separated from node 4 by nodes  $\{2, 3, 5\}$ .

## 2.2. Representing context-specific independence with LDAGs

The class of Labeled Directed Acyclic Graphs (LDAGs) was recently introduced in [31] as a generalization of DAGs. The purpose of the class of LDAGs is to graphically model situations exhibiting context-specific independence [1], which cannot be captured by CI-based models (see the survey [27]).

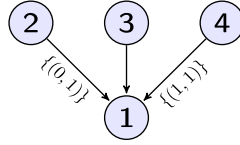


Fig. 3. LDAG over four binary variables.

**Definition 4** (*Context-specific independence (CSI)*). Let  $A$ ,  $B$ ,  $C$ , and  $S$  be disjoint subsets of  $\Delta$ . The variables  $X_A$  are contextually independent of  $X_B$  given  $X_C = e_C$  and  $X_S = e_S$  if

$$P(X_A = e_A | X_B = e_B, X_C = e_C, X_S = e_S) = P(X_A = e_A | X_C = e_C, X_S = e_S),$$

for all  $(e_A, e_B, e_S) \in \mathcal{X}_A \times \mathcal{X}_B \times \mathcal{X}_S$  for which  $P(X_B = e_B, X_C = e_C, X_S = e_S) > 0$ . This is denoted by  $X_A \perp X_B \mid X_C = e_C, X_S$ .

Local CSI statements, which address only a variable  $X_j$  and its parents  $X_{\Pi_j}$ , can naturally be included in the Bayesian network framework. More specifically, a CSI statement is defined as local with respect to a DAG if it is of the form

$$X_j \perp X_B \mid X_{\Pi_j \setminus B} = e_{\Pi_j \setminus B}, \quad (2)$$

where  $B \subset \Pi_j$ . By Definition 4, this independence statement holds if and only if

$$P(X_j = e_j | X_B = e_B, X_{\Pi_j \setminus B} = e_{\Pi_j \setminus B}) = P(X_j = e_j | X_{\Pi_j \setminus B} = e_{\Pi_j \setminus B}),$$

for all  $(e_j, e_B) \in \mathcal{X}_j \times \mathcal{X}_B$  for which  $P(X_B = e_B, X_{\Pi_j \setminus B} = e_{\Pi_j \setminus B}) > 0$ . In other words, a local CSI renders a variable conditionally independent of some of its parents given a certain context specified by the remaining parents. To capture such restrictions in the model structure, [31] proposed adding labels to the edges in the DAG.

**Definition 5** (*Labeled Directed Acyclic Graph (LDAG)*). Let  $G = (\Delta, E)$  be a DAG over random variables  $X_\Delta$  and let  $L_{(i,j)} = \Pi_j \setminus \{i\}$ . A *label* on an edge  $(i, j)$  is a subset of  $\mathcal{X}_{L_{(i,j)}}$ , denoted by  $\mathcal{L}_{(i,j)}$ , encoding a collection of local CSI statements according to

$$X_j \perp X_i \mid X_{L_{(i,j)}} = e_{L_{(i,j)}} \text{ for all } e_{L_{(i,j)}} \in \mathcal{L}_{(i,j)}.$$

An LDAG  $G_L = (\Delta, E, \mathcal{L}_E)$  is a DAG  $G = (\Delta, E)$  where the edges have been assigned labels as specified by  $\mathcal{L}_E = \{\mathcal{L}_{(i,j)}\}_{(i,j) \in E}$ .

**Example 2.** Consider the LDAG in Fig. 3 which represents the dependence structure over four binary variables with outcome space  $\mathcal{X}_i = \{0, 1\}$ . Note that given an ordering of the variables, the indices of the variables specifying a label need not be explicitly stated. According to Definition 5, the labels encode the local CSI statements

$$X_1 \perp X_2 \mid (X_3, X_4) = (0, 1) \text{ and } X_1 \perp X_4 \mid (X_2, X_3) = (1, 1),$$

respectively. In other words,  $X_1$  is contextually independent of  $X_2$  given  $X_3 = 0$  and  $X_4 = 1$ . Moreover,  $X_1$  is contextually independent of  $X_4$  given  $X_2 = 1$  and  $X_3 = 1$ .

**Table 1**(a) A traditional CPT and (b) a reduced CPT over variable  $X_1$  in Fig. 3.

$X_2$	$X_3$	$X_4$	$P(X_1 X_{\Pi_1})$
0	0	0	$p_1$
0	0	1	$p_2$
0	1	0	$p_3$
0	1	1	$p_4$
1	0	0	$p_5$
1	0	1	$p_2$
1	1	0	$p_6$
1	1	1	$p_6$

(a)

$X_2$	$X_3$	$X_4$	$P(X_1 X_{\Pi_1})$
0	0	0	$p_1$
*	0	1	$p_2$
0	1	0	$p_3$
0	1	1	$p_4$
1	0	0	$p_5$
1	1	*	$p_6$

(b)

One of the main motivations for including CSI in Bayesian networks was to reduce the number of parameters needed to specify the model distribution. The textbook way of defining the conditional probability distributions in (1) is through so-called conditional probability tables (CPTs) which simply list the conditional probabilities for each parent configuration. The number of parameters needed to specify a CPT of a node grows exponentially with the number of parents of the node. However, a local CSI statement implies that several distinct parent configurations induce the same conditional distribution which thereby needs only be defined once.

**Example 2 (continued).** Let us continue with the previous example concerning the LDAG in Fig. 3. A traditional CPT over variable  $X_1$  is seen in Table 1(a). Notice that there are certain regularities in the table in form of identical distributions. These regularities correspond to the local CSI statements which are encoded by the labels. Rather than defining identical distributions several times, we can construct a reduced CPT as illustrated in Table 1(b). A star means that the variable may take on any value, for example,  $(*, 0, 1) = \{(0, 0, 1), (1, 0, 1)\}$ . Ultimately, a reduced CPT represents a partition of the parent outcome space. Each row represents a class in the partition such that the conditional distribution is invariant for configurations belonging to the same class.

As in [31], without loss of generality we restrict attention to so-called *maximal* and *regular* LDAGs. Maximality states that it is not possible to add a configuration  $e_{L(i,j)}$  to the label  $\mathcal{L}_{(i,j)}$  without inducing an additional local CSI not encoded already by the original labels. Moreover, regularity simply states that each of the labels  $\mathcal{L}_{(i,j)}$  of an LDAG is a strict subset of  $\mathcal{X}_{L(i,j)}$ . Regularity and maximality together imply that a label  $\mathcal{L}_{(i,j)}$  cannot induce that  $X_j \perp X_i \mid X_{L(i,j)} = e_{L(i,j)}$  for all  $e_{L(i,j)} \in \mathcal{X}_{L(i,j)}$ , which corresponds to the CI  $X_j \perp X_i \mid X_{L(i,j)}$ .

Analogously to the case with DAGs, an important issue with LDAGs is the derivation of non-local independence statements that logically follow from the structure of the LDAG but are not explicitly encoded by it (by the labels or the DAG-structure).

**Definition 6.** Let  $G_L = (\Delta, E, \mathcal{L}_E)$  be an LDAG. We denote by  $I_{loc}(G_L)$  the set of local CI statements  $I_{loc}(G)$ , encoded by  $G = (\Delta, E)$ , together with the set of local CSI statements encoded by the labels  $\mathcal{L}_E$ .

Already in [1] a method called *CSI-separation*, which is analogous to *d-separation* for DAGs, was introduced for the purpose of verifying non-local CSI statements. Before defining CSI-separation in terms of LDAGs, we need to define the notion of *satisfied label*.

**Definition 7.** Let  $G_L = (\Delta, E, \mathcal{L}_E)$  be an LDAG, and  $X_C = e_C$  a context where  $C \subseteq \Delta$ . A label  $\mathcal{L}_{(i,j)} \subseteq \mathcal{L}_E$  is *satisfied* in the context  $X_C = e_C$  if  $L(i,j) \cap C \neq \emptyset$  and

$$\{e_{L(i,j) \cap C} \times \mathcal{X}_{L(i,j) \setminus C}\} \subseteq \mathcal{L}_{(i,j)}.$$

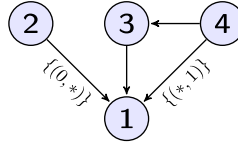


Fig. 4. LDAG over four binary variables.

For an LDAG  $G_L$  and a context  $X_C = e_C$ , we define  $G(e_C) = (\Delta, E \setminus E')$ , where  $E' = \{(i, j) \in E \mid \mathcal{L}_{(i,j)} \text{ is satisfied}\}$ . Note that  $G(e_C)$  is the subgraph of  $G$  that arises by removing edges whose labels are satisfied in the context  $X_C = e_C$ . We are now ready to define CSI-separation [31].

**Definition 8** (*CSI-separation*). Let  $G_L = (\Delta, E, \mathcal{L}_E)$  be an LDAG and let  $A, B, C, S$  be disjoint subsets of  $\Delta$ . The set  $X_A$  is CSI-separated from  $X_B$  by  $X_S$  in the context  $X_C = e_C$  in  $G_L$ , if  $X_A$  is  $d$ -separated from  $X_B$  by  $X_{C \cup S}$  in  $G(e_C)$ .

As stated by the following result (see Theorem 5.3 in [20]), CSI-separation is a sound method for verifying non-local CSIs.

**Theorem 9** (*Soundness of CSI-separation*). Let  $G_L = (\Delta, E, \mathcal{L}_E)$  be an LDAG, let  $P_\Delta$  be a distribution satisfying  $I_{loc}(G_L)$ , and let  $A, B, C, S$  be disjoint subsets of  $\Delta$ . If  $X_A$  is CSI-separated from  $X_B$  by  $X_S$  in the context  $X_C = e_C$  in  $G_L$ , then the distribution  $P_\Delta$  satisfies the CSI  $X_A \perp X_B \mid X_C = e_C, X_S$ .

However, unlike  $d$ -separation for DAG structures, CSI-separation is not a complete method for discovering non-local independencies implied by an LDAG structure. In fact,  $d$ -separation is not a complete method for discovering CI statements in an LDAG. This is illustrated in the following example.

**Example 3.** Consider the LDAG in Fig. 4, where  $\{(0, *)\}$  again is a shorthand for the set  $\{(0, 0), (0, 1)\}$ . Assume that  $P_\Delta$  is a joint distribution satisfying the independencies encoded by the LDAG. As discussed in [31], the underlying DAG structure does not allow us to infer the CI statement  $X_2 \perp X_4 \mid \{X_1, X_3\}$  through the use of  $d$ -separation. Using CSI-separation, however, we can verify that  $X_2 \perp X_4 \mid X_1, X_3 = 0$  and  $X_2 \perp X_4 \mid X_1, X_3 = 1$ . Consequently, a reasoning by cases argument allows one to conclude that  $X_2 \perp X_4 \mid \{X_1, X_3\}$  holds in  $P_\Delta$  (see [31] for more details).

### 2.3. Complexity of determining the implied CSIs by an LDAG

In this section we show that the problem of deciding whether a CI statement  $\phi$  is implied by an LDAG structure  $G_L$  is coNP-hard. It is worth noting that for DAGs this problem can be solved in polynomial time (see [20]).

**Theorem 10.** The problem of deciding whether an (context-specific) independence is implied by an LDAG structure is coNP-hard.

**Proof.** We apply a proof idea of Koller et al. (Exercise 5.10 in [20]) for DAGs with so-called tree-structured CPDs to our LDAG structures [20]. We will reduce a 3-SAT problem instance into deciding whether a CI statement is implied by an LDAG structure.

Define the corresponding LDAG to the 3-SAT instance as follows (see Fig. 5). Let binary nodes  $Z_1, \dots, Z_l$  correspond to variables in the 3-SAT instance. Let  $Y_0, Y_1, \dots, Y_k$  denote additional binary nodes of which  $Y_1, \dots, Y_k$  represent the clauses of the 3-SAT instance. Let the parents of node  $Y_i$  ( $i \geq 1$ ) be the node  $Y_{i-1}$ , and the  $Z$ -nodes appearing in the clause  $i$ , let us call them  $Z_a, Z_b, Z_c$ . The labels on the edge  $Y_{i-1} \rightarrow Y_i$



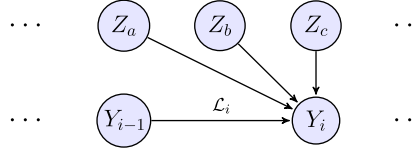


Fig. 5. LDAG for the proof of Theorem 10.

consist of assignments to the nodes  $Z_a, Z_b, Z_c$ . Let the label  $\mathcal{L}_i$  on the arc  $Y_{i-1} \rightarrow Y_i$  be exactly the set of assignments to  $Z_a, Z_b, Z_c$  that *do not* satisfy the  $i$ th clause of the 3-SAT problem.

Consider different contexts  $e_z$  over variables  $Z_1, \dots, Z_l$ . If  $e_z$  does not satisfy the 3-SAT instance, there is a clause  $i$  which is unsatisfied, and thus the corresponding edge  $Y_{i-1} \rightarrow Y_i$  does not appear in  $G(e_z)$ . Thus,  $Y_0$  and  $Y_k$  are  $d$ -separated in  $G(e_z)$  and according to Theorem 9:  $Y_0 \perp Y_k | Z_1, \dots, Z_l = e_z$ .

If  $e_z$  satisfies the 3-SAT instance, all clauses are satisfied and thus all edges  $Y_{i-1} \rightarrow Y_i$  appear in  $G(e_z)$ . Thus,  $Y_0$  and  $Y_k$  are not  $d$ -separated in  $G(e_z)$ . We can define a parameterization for the LDAG under which there is a dependence. Let  $Y_0, Z_1, \dots, Z_l$  be distributed uniformly. Let  $Y_i = Y_{i-1}$  if  $Z_a, Z_b, Z_c$  satisfy the clause  $i$  and 0 otherwise. Now under a satisfying context  $e_z$ :  $Y_k = Y_{k-1} = \dots = Y_0$  hence  $Y_0 \not\perp Y_k | Z_1, \dots, Z_k = e_z$ . Thus,  $Y_0 \perp Y_k | Z_1, \dots, Z_k = e_z$  cannot follow from the LDAG structure.

If the 3-SAT problem is satisfiable there is a context  $e_z$  such that  $Y_0 \perp Y_k | Z_1, \dots, Z_k = e_z$  does not follow from the LDAG structure, hence  $Y_0 \perp Y_k | Z_1, \dots, Z_k$  does not follow from the structure either. If the 3-SAT problem is unsatisfiable we have that for all contexts  $e_z$ :  $Y_0 \perp Y_k | Z_1, \dots, Z_k = e_z$ , from which it directly follows that  $Y_0 \perp Y_k | Z_1, \dots, Z_k$ . Thus, the defined LDAG structure implies independence  $Y_0 \perp Y_k | Z_1, \dots, Z_k$  if and only if the 3-SAT problem is unsatisfiable. If we could decide whether an independence is implied by an LDAG in polynomial time, we could also solve 3-SAT in polynomial time.  $\square$

### 3. Axiomatic characterization of CSI-separation

In this section we define a sound extension of the semi-graphoid axioms that capture CSI-separation in LDAGs and study the more general problem of finding a complete axiomatization for the discovery of non-local independencies implied by an LDAG structure. We begin by recalling the semi-graphoid axioms [4,30] and their relation to  $d$ -separation in the context of Bayesian networks.

#### 3.1. The semi-graphoid axioms and the implication problem of CI statements

In the following we assume without loss of generality that the natural ordering of the nodes  $\Delta$  of a DAG  $G$  agrees with the edge relation of  $G$ . Furthermore, we redefine  $I_{loc}(G)$  as follows:

$$I_{loc}(G) = \{X_i \perp \{X_1, \dots, X_{i-1}\} \setminus X_{\Pi_i} | X_{\Pi_i} : i \in \Delta\}. \quad (3)$$

Note that  $\{1, \dots, i-1\}$  are non-descendants of the node  $i$  in  $G$ . The equivalence of the above definition of  $I_{loc}(G)$  with the previous one follows by Theorem 12.

**Definition 11** (*Semi-graphoid axioms*). The following axioms are called the semi-graphoid axioms. Below  $X$ ,  $Y$ , and  $Z$  denote sets of stochastic variables. The union of  $X$  and  $Y$  is denoted by  $XY$ .

1. **Triviality:**  $X \perp \emptyset | Z$ ,
2. **Symmetry:**  $X \perp Y | Z \Rightarrow Y \perp X | Z$ ,
3. **Decomposition:**  $X \perp YU | Z \Rightarrow X \perp Y | Z$ ,



4. **Weak Union:**  $X \perp YU|Z \Rightarrow X \perp Y|ZU$ ,
5. **Contraction:**  $X \perp Y|ZU$  and  $X \perp U|Z \Rightarrow X \perp YU|Z$ .

The semi-graphoid axioms are known to be sound for all distributions. For positive distributions the following further property holds:

6. **Intersection:**  $X \perp Y|ZU$  and  $X \perp Z|YU \Rightarrow X \perp YZ|U$

The following theorem shows that the semi-graphoid axioms correspond exactly to  $d$ -separation in the context of Bayesian networks. For a finite set  $\Sigma \cup \{\phi\}$  of CIs, we write  $\Sigma \vdash^{sg} \phi$  with the meaning that  $\phi$  can be derived from  $\Sigma$  using the semi-graphoid axioms. In other words, there exists a finite sequence  $\psi_1, \dots, \psi_k$  such that  $\psi_k = \phi$ , and  $\psi_i \in \Sigma$  or  $\psi_i$  is obtained by applying one of the semi-graphoid axioms to  $\psi_l$  and  $\psi_t$  for some  $l, t < i$ .

**Theorem 12.** [11,36] *Let  $G = (\Delta, V)$  be a DAG, and let  $A$ ,  $B$ , and  $C$  be disjoint subsets of  $\Delta$ . Then  $A$  is  $d$ -separated from  $B$  by  $C$  if and only if  $I_{loc}(G) \vdash^{sg} X_A \perp X_B|X_C$ .*

Theorems 12 and 3 together imply the following result.

**Theorem 13.** [11,36] *Let  $G = (\Delta, V)$  be a DAG, and let  $A$ ,  $B$ , and  $C$  be disjoint subsets of  $\Delta$ . Then  $I_{loc}(G) \models X_A \perp X_B|X_C$  if and only if  $I_{loc}(G) \vdash^{sg} X_A \perp X_B|X_C$ .*

This result can be viewed as a complete axiomatization of a restricted version of the *implication problem* of CI statements. The implication problem of CI statements is defined as follows. Given a finite collection  $\Sigma \cup \{\varphi\}$  of CI statements as input, determine whether for all  $P$ ,

$$P \models \Sigma \Rightarrow P \models \varphi.$$

This problem is known not to be finitely axiomatizable [34]. Despite of this negative result, the semi-graphoid axioms are also relevant for the general implication problem of conditional independence. For example, in [15] it was shown that the axioms are complete for the implication problem of conditional independence assuming  $\Sigma$  consists solely of so-called *saturated* CIs. Furthermore, in [26] the semi-graphoid axioms and a certain other set of axioms are used to approximate the CI implication problem.

### 3.2. Axioms for context specific independence

In this section we give an axiomatic characterization of CSI-separation and study the problem of finding a complete axiomatization for the discovery of non-local independencies implied by an LDAG structure.

We begin by reformulating the semi-graphoid axioms and the intersection axiom for CSI statements.

1. **Triviality:**  $X_A \perp \emptyset|X_C = e_C, X_S$
2. **Symmetry:**  $X_A \perp X_B|X_C = e_C, X_S \Rightarrow X_B \perp X_A|X_C = e_C, X_S$
3. **Decomposition:**  $X_A \perp X_{B \cup B'}|X_C = e_C, X_S \Rightarrow X_A \perp X_B|X_C = e_C, X_S$
4. **Weak Union:**  $X_A \perp X_{B \cup B'}|X_C = e_C, X_S \Rightarrow X_A \perp X_B|X_C = e_C, X_{S \cup B'}$
5. **Contraction:**  $X_A \perp X_B|X_C = e_C, X_{S \cup W}$  and  $X_A \perp X_S|X_C = e_C, X_W \Rightarrow X_A \perp X_{B \cup S}|X_C = e_C, X_W$
6. **Intersection:**  $X_A \perp X_B|X_C = e_C, X_{S \cup W}$  and  $X_A \perp X_W|X_C = e_C, X_{S \cup B} \Rightarrow X_A \perp X_{B \cup W}|X_C = e_C, X_S$

Recall that the Axioms 1-5 are sound for all distributions and Axiom 6 for all positive distributions.

For the derivation of CSI statements, we introduce the following rule closely resembling the definition of CSI-separation. In order to apply the CSI-rule below, an input LDAG  $G_L$  as well as the outcome spaces  $\mathcal{X}_i$  of the variables  $X_i$  have to be fixed. The extra assumptions  $A(e_C)$  allowed in the subderivation of the CSI-rule:

$$A(e_C) = \{X_j \perp \{X_{j_1}, \dots, X_{j_k}\} | X_{\Pi_j - \{j_1, \dots, j_k\}} : j \in \{1, \dots, n\}\}$$

encode the information that the graph  $G(e_C)$  arises from  $G$  by removing edges  $(j_1, j), \dots, (j_k, j)$ , for  $j \in \Delta$ .

**Definition 14** (CSI-rule for context  $X_C = e_C$ ).

$$[I_{loc}(G) \cup A(e_C) \vdash^{sg} X_A \perp X_B | X_{C \cup S}] \Rightarrow X_A \perp X_B | X_C = e_C, X_S$$

The idea of the CSI-rule is that the existence of the derivation on the left-hand side (corresponding to  $d$ -separation in  $G(e_C)$ ) justifies the conclusion on the right. The auxiliary assumptions  $A(e_C)$  can only be used in the subderivation. Note that for the empty context  $C = \emptyset$  the CSI-rule warrants all derivations from the set  $I_{loc}(G)$  by the semi-graphoid axioms.

We will next show that the CSI-rule is sound and corresponds to CSI-separation in LDAGs. For an LDAG  $G_L$ , we write  $I_{loc}(G_L) \vdash^{csi} \phi$  with the meaning that a CI (or CSI) statement  $\phi$  can be derived from  $I_{loc}(G_L)$  using the CSI-rule.

**Theorem 15.** Let  $G_L = (\Delta, V, \mathcal{L}_E)$  be an LDAG, and let  $A, B, C$ , and  $S$  be disjoint subsets of  $\Delta$ . Then

$$I_{loc}(G_L) \vdash^{csi} X_A \perp X_B | X_C = e_C, X_S$$

if and only if  $X_A$  is CSI-separated from  $X_B$  by  $X_S$  in the context  $X_C = e_C$ .

**Proof.** It worth noting that for  $C = \emptyset$  the claim follows immediately from Theorem 12. Let us then assume  $C \neq \emptyset$ . Recall our assumption that  $I_{loc}(G)$  (and  $I_{loc}(G(e_C))$ ) is encoded by a set of CIs of the form (3). By Theorem 12 it suffices to show that the sets  $I_{loc}(G) \cup A(e_C)$  and  $I_{loc}(G(e_C))$  are equivalent with respect to deductions by the semi-graphoid axioms.

Note first that  $I_{loc}(G) \cup A(e_C) \vdash^{sg} \psi$ , for all  $\psi \in I_{loc}(G(e_C))$ . This holds since  $I_{loc}(G(e_C))$  consists of CIs of the form

$$X_j \perp \{X_1, \dots, X_{j-1}\} \setminus X_A | X_A, \quad (4)$$

where  $A = \Pi_j - \{j_1, \dots, j_k\}$ , and the CI in equation (4) can be derived by one application of the contraction rule applied to the CIs

$$X_j \perp \{X_{j_1}, \dots, X_{j_k}\} | X_{\Pi_j - \{j_1, \dots, j_k\}}$$

and

$$X_j \perp \{X_1, \dots, X_{j-1}\} \setminus X_{\Pi_j} | X_{\Pi_j}.$$

Let us then show  $I_{loc}(G(e_C)) \vdash^{sg} \psi$  for all  $\psi \in I_{loc}(G) \cup A(e_C)$ . Assume  $\psi$  is of the form  $X_j \perp \{X_{j_1}, \dots, X_{j_k}\} | X_{\Pi_j - \{j_1, \dots, j_k\}}$ . Then it can be derived by one application of the decomposition rule applied to the corresponding CI in (4). Analogously, if  $\psi$  is of the form  $X_j \perp \{X_1, \dots, X_{j-1}\} \setminus X_{\Pi_j} | X_{\Pi_j}$ , then one application of weak union rule applied to (4) suffices.  $\square$

It is worth noting that for positive distributions the set  $A(e_c)$  could be defined to consist of

$$X_j \perp X_{j_l} | X_{\Pi_j - \{j_l\}}$$

for each edge  $(j_l, j)$  deleted in  $G(e_c)$ . Recall that these statements correspond directly to local CSI statements as defined in (2). Namely for positive distributions a repeated use of the intersection axiom justifies the derivation of the CI

$$X_j \perp \{X_{j_1}, \dots, X_{j_k}\} | X_{\Pi_j - \{j_1, \dots, j_k\}} \quad (5)$$

from the CIs

$$\{X_j \perp X_{j_l} | X_{\Pi_j - \{j_l\}} : l \in \{1, \dots, k\}\}. \quad (6)$$

On the other hand, every CI in (6) can be derived from (5) by the weak union rule.

As discussed in Example 3, CSI-separation, and hence our axioms fail to capture some non-local CIs implied by the structure of the LDAG. The following RC-rule and its inverse are obviously sound for LDAGs and they address the problem discussed in Example 3.

**Definition 16** (*The RC-rule*).

$$X_A \perp X_B | X_{C \cup W} = e_{C \cup W}, X_S \forall e_C \in \mathcal{X}_C \Rightarrow X_A \perp X_B | X_W = e_W, X_{C \cup S} \quad (7)$$

$$X_A \perp X_B | X_W = e_W, X_{C \cup S} \Rightarrow X_A \perp X_B | X_{C \cup W} = e_{C \cup W}, X_S \quad (8)$$

It is worth noting that the size of the assumptions needed to apply the RC-rule (7) grows exponentially in the number of variables  $X_C$ . On the other hand, the question whether a CI holds in an LDAG is in general NP-hard (see Theorem 10) hence the exponential blow-up might not be avoidable.

We conjecture that the following holds. We assume below that  $G_L$  also determines the outcome spaces  $\mathcal{X}_i$  of  $X_i$  for each  $i \in \Delta$ .

**Conjecture 1.** *Let  $G_L = (\Delta, V, \mathcal{L}_E)$  be an LDAG, and let  $A, B, C$ , and  $S$  be disjoint subsets of  $\Delta$ . Then the following are equivalent:*

1. *A CSI  $X_A \perp X_B | X_C = e_c, X_S$  (CI  $X_A \perp X_B | X_C$ ) can be derived from  $I_{loc}(G_L)$  using the semi-graphoid axioms and the CSI and RC rules,*
2. *Every distribution  $P_\Delta$  satisfying  $I_{loc}(G_L)$  also satisfies  $X_A \perp X_B | X_C = e_c, X_S$  ( $X_A \perp X_B | X_C$ ).*

Note that the implication (1)  $\Rightarrow$  (2) (soundness) holds by Theorem 15 and the obvious soundness of the RC-rule. For the converse implication (completeness), it may still be possible that more general types of independence, such as the notion of partial independence [32], would enable the derivation of some further CSI statements.

In Example 3 we showed how two labels (or CSIs), when combined, imply a CI statement which cannot be inferred from the underlying graph. Instead, the CI statement was inferred by performing CSI-separation and reasoning by cases, as formalized by the RC-rule. However, applying CSI-separation in a case-wise fashion turns out to be insufficient for discovering all independencies. Therefore, this section introduces additional rules in an attempt to make our system complete. We end this section by providing some examples illustrating the need for the additional rules.

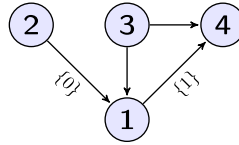


Fig. 6. LDAG over four binary variables.

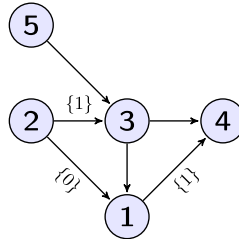


Fig. 7. LDAG over five binary variables.

**Example 4.** Consider the LDAG in Fig. 6, from which we want to infer that  $X_2 \perp X_4$ . Using CSI-separation we can conclude that  $X_2 \perp X_4 \mid X_3 = 0$  and  $X_2 \perp X_4 \mid X_3 = 1$ , which then according to the RC-rule imply that  $X_2 \perp X_4 \mid X_3$ . In addition, the local directed Markov property states that  $X_2 \perp X_3$ . At this point we need to use the SG axioms. More specifically, the contraction axiom allows us to conclude that  $X_2 \perp \{X_3, X_4\}$ , after which the decomposition axiom gives us  $X_2 \perp X_4$ . An interesting observation is that the LDAGs in Figs. 4 and 6 actually represent the same dependence structures despite that the underlying DAGs belong to different Markov equivalence classes (see [31] for more details).

**Example 5.** Consider the LDAG in Fig. 7, from which we want to infer that  $X_2 \perp X_4 \mid X_5 = 1$ . Following a similar approach as in the previous example, we use CSI-separation to conclude that  $X_2 \perp X_4 \mid X_3 = 0, X_5 = 1$  and  $X_2 \perp X_4 \mid X_3 = 1, X_5 = 1$ , which then according to the RC-rule imply that  $X_2 \perp X_4 \mid X_3, X_5 = 1$ . In addition, the label on edge (2, 3) represents the CSI statement  $X_2 \perp X_3 \mid X_5 = 1$ . Analogously to the previous example, we can now use the introduced context-specific SG axioms to infer the sought-after independence. The contraction axiom allows us to conclude that  $X_2 \perp \{X_3, X_4\} \mid X_5 = 1$ , after which the decomposition axiom gives us  $X_2 \perp X_4 \mid X_5 = 1$ .

#### 4. Independence in databases and team semantics

In this section we review and study connections between CI statements and certain qualitative notions of independence in the areas of database theory and team semantics. It is well known that embedded multi-valued dependency from database theory can be viewed as a qualitative version of conditional independence (see, e.g., [22–24,37]). On the other hand, EMVDs have re-emerged as a central concept in the area of team semantics in which various logics of dependence and independence are studied. Subsection 4.1 gives the basics of teams and independence atoms and highlights some of the known connections to CI statements via the previous work on EMVDs. Subsection 4.2 then sets the stage for a systematic study of context specific independence in team semantics and database theory.

##### 4.1. Independence atoms and CI statements

The syntax of independence logic,  $\text{FO}(\perp_c)$ , extends the syntax of first-order logic (FO) by atomic independence formulas of the form

$$\vec{x} \perp_{\vec{z}} \vec{y}, \quad (9)$$

where  $\vec{x}$ ,  $\vec{y}$ , and  $\vec{z}$  are tuples of variables. The semantics of independence logic is defined (as in FO) in terms of first-order structures except that single variable assignments are replaced by sets  $X$  of assignments called teams.

**Definition 17.** Let  $\mathfrak{A}$  be a structure with domain  $A$ , and  $\{x_1, \dots, x_k\}$  a finite set of variables.

- A team  $X$  of  $\mathfrak{A}$  with domain  $\text{Dom}(X) = \{x_1, \dots, x_k\}$  is any set of assignments from the variables  $\{x_1, \dots, x_k\}$  into the set  $A$ .
- $\mathfrak{A} \models_X \vec{x} \perp_{\vec{z}} \vec{y}$  iff for all  $s, s' \in X$  such that  $s(\vec{z}) = s'(\vec{z})$  there is  $s^* \in X$  such that  $s^*(\vec{x}\vec{z}) = s(\vec{x}\vec{z})$ , and  $s^*(\vec{y}) = s'(\vec{y})$ .

It is worth noting that only the team and not the structure plays a role for the satisfaction of an independence atom  $\vec{x} \perp_{\vec{z}} \vec{y}$ . The definition of satisfaction for complex  $\text{FO}(\perp_c)$ -formulas is discussed in the next section.

There is an intimate connection between independence atoms and CI statements. Teams and independence atoms can be seen as qualitative analogues of probability distributions and their CI statements (so-called relational dependency models discussed, e.g., in [12]). This connection can be made explicit as follows (see [6,12]). For a set of stochastic variables  $X_A$ ,  $A \subseteq \Delta$ , we write  $x_A$  for a tuple (in any order) consisting of first-order variables  $x_i$  for  $i \in A$ .

**Proposition 18.** Let  $P_\Delta$  be a distribution. Define a team  $X$  consisting of those assignments  $s: \{x_1, \dots, x_n\} \rightarrow \cup_{1 \leq i \leq n} \mathcal{X}_i$  such that

$$X = \{s \mid P_\Delta(X_1 = s(x_1), \dots, X_n = s(x_n)) > 0\}.$$

Then if  $P$  satisfies a CI  $X_A \perp X_B | X_C$ , then  $\mathfrak{A} \models_X x_A \perp_{x_C} x_B$ , where  $\mathfrak{A} = \cup_{1 \leq i \leq n} \mathcal{X}_i$ .

**Proof.** Let  $s, s' \in X$  be such that  $s(x_C) = s'(x_C)$ . Now we must have  $P_\Delta(X_A = s(x_A) | X_C = s(x_C)) > 0$  and  $P_\Delta(X_B = s'(x_B) | X_C = s(x_C)) > 0$ . Denote these non-negative probabilities by  $c_1$  and  $c_2$ , respectively. Since  $X_A \perp X_B | X_C$  holds,

$$P_\Delta(X_A = s(x_A), X_B = s'(x_B) | X_C = s(x_C)) = c_1 c_2 > 0,$$

hence it follows that there exists  $s^* \in X$  such that  $s^*(x_A) = s(x_A)$ ,  $s^*(x_B) = s'(x_B)$ , and  $s^*(x_C) = s(x_C)$  as wanted.  $\square$

The known connections between the implication problems of EMVDs and CI statements can be also restated for independence atoms; For a finite set  $\Sigma \cup \{\phi\}$  of independence atoms, we write  $\Sigma \models \phi$  with the meaning that for all finite  $\mathfrak{A}$  and  $X$ , if  $\mathfrak{A} \models_X \psi$  for all  $\psi \in \Sigma$ , then  $\mathfrak{A} \models_X \phi$ . For a set of CIs  $\Sigma$ ,  $\Sigma^*$  denotes the corresponding set of independence atoms. The following result extending Theorem 13 is an immediate consequence of the results in [12].

**Theorem 19.** Let  $G$  be a DAG, and let  $A$ ,  $B$ , and  $C$  be disjoint subsets of  $\Delta$ . Then the following are equivalent:

1.  $I_{loc}(G) \models X_A \perp X_B | X_C$ ,
2.  $I_{loc}(G)^* \models x_A \perp_{x_C} x_B$ ,
3.  $I_{loc}(G) \vdash^{sg} X_A \perp X_B | X_C$ .

Theorem 19 does not hold for the general implication problems of independence atoms and CIs [34]. Furthermore, the implication problem of independence atoms is known to be undecidable by the result of [19], whereas for CIs the decidability of the problem is still open. It is worth noting that a version of Theorem 19 holds for marginal CIs and independence atoms of the form  $\vec{x} \perp \vec{y}$ . Furthermore, the implication problem of marginal CIs has a complete axiomatization in terms of so-called Axioms of Independence that are satisfied by various notions of independence in different areas (see [13] and [28,29]).

#### 4.2. Independence logic and CSI atoms

In this section we discuss independence logic and define a novel logical analogue of CSI atoms.

The syntax of independence logic extends the syntax of FO, defined in terms of  $\vee, \wedge, \neg, \exists$  and  $\forall$ , by atomic independence formulas of the form  $\vec{x} \perp_{\vec{z}} \vec{y}$ , where  $\vec{x}, \vec{y}$ , and  $\vec{z}$  are tuples of variables. The set  $\text{Fr}(\phi)$  of free variables of  $\phi \in \text{FO}(\perp_c)$  is defined analogously to first-order logic stipulating that all variable occurrences in independence atoms are free.

**Definition 20.** Let  $\mathfrak{A}$  be a model with domain  $A$  and  $X$  a team of  $A$  with domain  $\{x_1, \dots, x_k\}$ .

- If  $s$  is an assignment,  $x$  a variable, and  $a \in A$ , then  $s(a/x)$  denotes the assignment (with domain  $\text{Dom}(s) \cup \{x\}$ ) that agrees with  $s$  everywhere except that it maps  $x$  to  $a$ .
- For a function  $F: X \rightarrow \mathcal{P}(A) \setminus \{\emptyset\}$ , we define the operations of *Supplementation*  $X(F/x_n)$  and *Duplication*  $X(A/x_n)$  as follows:

$$\begin{aligned} X(F/x_n) &= \{s(a/x_n) : s \in X \text{ and } a \in F(s)\} \\ X(A/x_n) &= \{s(a/x_n) : s \in X \text{ and } a \in A\}. \end{aligned}$$

We are now ready to define the semantics of independence logic formulas (other than independence atoms). We restrict attention to formulas in negation normal form in which negation is allowed to appear only in front of first-order atomic formulas. Below, atomic formulas and their negations are called literals, and  $\mathfrak{A} \models_s \phi$  refers to satisfaction in first-order logic.

**Definition 21.** Let  $\mathfrak{A}$  be a model and  $X$  a team of  $A$ . The satisfaction relation  $\mathfrak{A} \models_X \phi$  is defined as follows:

- If  $\phi$  is a first-order literal, then  $\mathfrak{A} \models_X \phi$  iff for all  $s \in X$ :  $\mathfrak{A} \models_s \phi$ .
- $\mathfrak{A} \models_X \psi \wedge \phi$  iff  $\mathfrak{A} \models_X \psi$  and  $\mathfrak{A} \models_X \phi$ .
- $\mathfrak{A} \models_X \psi \vee \phi$  iff  $X = Y \cup Z$  such that  $\mathfrak{A} \models_Y \psi$  and  $\mathfrak{A} \models_Z \phi$ .
- $\mathfrak{A} \models_X \exists x_n \psi$  iff  $\mathfrak{A} \models_{X(F/x_n)} \psi$  for some  $F: X \rightarrow \mathcal{P}(A) \setminus \{\emptyset\}$ .
- $\mathfrak{A} \models_X \forall x_n \psi$  iff  $\mathfrak{A} \models_{X(A/x_n)} \psi$ .

Above, we assume that the domain of  $X$  contains the variables free in  $\phi$ . Finally, a sentence  $\phi$  is true in a model  $\mathfrak{A}$  (abbreviated  $\mathfrak{A} \models \phi$ ) if  $\mathfrak{A} \models_{\{\emptyset\}} \phi$ .

One of the most basic observations about team semantics is the so-called flatness property of FO-formulas.

**Theorem 22.** [35] Let  $\phi \in \text{FO}$ . Then for all  $\mathfrak{A}$  and  $X$  it holds that

$$\mathfrak{A} \models_X \phi \Leftrightarrow \mathfrak{A} \models_s \phi \text{ for all } s \in X.$$

Another important property of all independence logic formulas is the following locality property. For a team  $X$  and  $V \subseteq \text{Dom}(X)$ , we define  $X \upharpoonright V := \{s \upharpoonright V \mid s \in X\}$ .

**Theorem 23.** [9] Let  $\phi$  be an  $\text{FO}(\perp_c)$ -formula. Then for all  $\mathfrak{A}$  and  $X$  it holds that

$$\mathfrak{A} \models_X \phi \Leftrightarrow \mathfrak{A} \models_{X \upharpoonright \text{Fr}(\phi)} \phi.$$

Both dependence and independence logic are equi-expressive with existential second-order logic and are hence both non-axiomatizable. On the other hand, by restricting attention to syntactic fragments of these logics, complete axiomatization is possible [17,21].

The next definition formulates a natural logical analogue of CSI statements.

**Definition 24** (*CSI-atom*). A context-specific independence atom (CSI-atom) is a formula of the form

$$\vec{x} \perp_{\phi(\vec{v}), \vec{u}} \vec{y},$$

where  $\phi$  is an FO-formula. Satisfaction for CSI-atoms is defined as follows:  $\mathfrak{A} \models_X \vec{x} \perp_{\phi(\vec{v}), \vec{u}} \vec{y}$  iff for all  $s, s' \in X$  such that  $s(\vec{u}) = s'(\vec{u})$ ,  $\mathfrak{A} \models_s \phi(\vec{v})$ , and  $\mathfrak{A} \models_{s'} \phi(\vec{v})$  there is  $s^* \in X$  such that  $\mathfrak{A} \models_{s^*} \phi(\vec{v})$ ,  $s^*(\vec{x}\vec{u}) = s(\vec{x}\vec{u})$ , and  $s^*(\vec{y}) = s'(\vec{y})$ .

Denote the extension of FO by CSI-atoms by  $\text{FO}(\perp_{\text{CSI}})$ . The following observations are straightforward to prove. First of all, the locality property holds also for  $\text{FO}(\perp_{\text{CSI}})$ .

**Theorem 25.** Let  $\phi$  be an  $\text{FO}(\perp_{\text{CSI}})$ -formula. Then for all  $\mathfrak{A}$  and  $X$  it holds that

$$\mathfrak{A} \models_X \phi \Leftrightarrow \mathfrak{A} \models_{X \upharpoonright \text{Fr}(\phi)} \phi.$$

The next lemma shows that CSI-atoms can be expressed as  $\text{FO}(\perp_c)$ -formulas. Below, we denote by  $\phi^d$  the dual of a first-order formula  $\phi$  that is obtained by transforming  $\neg\phi$  to negation normal form.

**Lemma 26.** Let  $\phi \in \text{FO}$ . Then the formula  $\vec{x} \perp_{\phi(\vec{v}), \vec{u}} \vec{y}$  is logically equivalent to the independence logic formula

$$\phi^d \vee (\phi \wedge \vec{x} \perp_{\vec{u}} \vec{y}).$$

**Proof.** Let  $\mathfrak{A}$  be a structure and  $X$  a team such that

$$\mathfrak{A} \models_X \vec{x} \perp_{\phi(\vec{v}), \vec{u}} \vec{y}. \quad (10)$$

Let  $Y_1 = \{s \in X \mid \mathfrak{A} \models_s \phi\}$  and  $Y_2 = \{s \in X \mid \mathfrak{A} \models_s \phi^d\}$ . Now since  $\phi \in \text{FO}$  it holds that  $X = Y_1 \cup Y_2$  and  $Y_1 \cap Y_2 = \emptyset$ . Furthermore, by Theorem 22 it holds that  $\mathfrak{A} \models_{Y_2} \phi^d$ , and  $\mathfrak{A} \models_{Y_1} \phi$ . In order to show  $\mathfrak{A} \models_{Y_1} \vec{x} \perp_{\vec{u}} \vec{y}$ , let  $s, s' \in Y_1$  be such that  $s(\vec{u}) = s'(\vec{u})$ . As both  $s$  and  $s'$  satisfy  $\phi$ , by (10) there exists  $s^* \in X$  such that  $\mathfrak{A} \models_{s^*} \phi$  (i.e.,  $s^* \in Y_1$ ),  $s^*(\vec{x}\vec{u}) = s(\vec{x}\vec{u})$ , and  $s^*(\vec{y}) = s'(\vec{y})$ . This shows that  $\mathfrak{A} \models_{Y_1} \vec{x} \perp_{\vec{u}} \vec{y}$ , and hence

$$\mathfrak{A} \models_X \phi^d \vee (\phi \wedge \vec{x} \perp_{\vec{u}} \vec{y})$$

as wanted. The converse implication is proved analogously.  $\square$

Lemma 26 implies that  $\text{FO}(\perp_c)$  and  $\text{FO}(\perp_{\text{CSI}})$  are equi-expressive.

**Theorem 27.**  $\text{FO}(\perp_{\text{CSI}}) \equiv \text{FO}(\perp_c)$ .



It is worth noting that Theorem 27 can be also obtained as a corollary of the result of [9] showing that independence logic can define all properties of teams definable in existential second-order logic. Lemma 26 shows, additionally, that CSI-atoms can be given quantifier-free translations in independence logic.

We end this section by showing that, analogously to CIs and independence atoms, certain CSI-atoms can be viewed as qualitative analogues of CSI statements.

**Proposition 28.** *Let  $P_\Delta$  be a distribution. Define a team  $X$  consisting of those assignments  $s: \{x_1, \dots, x_n\} \rightarrow \cup_{1 \leq i \leq n} \mathcal{X}_i$  such that*

$$X = \{s \mid P_\Delta(X_1 = s(x_1), \dots, X_n = s(x_n)) > 0\}.$$

*If  $P_\Delta$  satisfies the CSI*

$$X_A \perp X_B \mid X_C = e_C, X_S,$$

*then  $\mathfrak{A} \models_X x_A \perp_{\phi(x_C), x_S} x_B$ , where  $\mathfrak{A} = \cup_{1 \leq i \leq n} \mathcal{X}_i$  and  $\phi(x_C) = \bigwedge_{i \in C} (x_i = e_i)$  ( $e_i$  appears as a constant symbol in  $\phi(x_C)$ ).*

**Proof.** Analogous to the proof of Proposition 18.  $\square$

## 5. Conclusion

We considered the problem of identifying all independence statements implied by the structure and the labels of an LDAG. We showed the task is coNP-hard, similarly as for other models capturing context specific independence. Then we defined a sound extension of the semi-graphoid axioms for deriving independence relations. We showed that our axiomatization finds independence relations not captured by CSI-separation, but the completeness of the axiomatization is still an open question. However, the more complete set of implied independencies are already useful in various applications,<sup>1</sup> including structure discovery and probabilistic inference for graphical models.

In the second part of the article, we discussed the connections between conditional independence and qualitative notions of independence prominent in database theory and team semantics. As opposed to the so-called conditional functional and inclusion dependencies, context specific independence has not yet been studied in these frameworks. In order to initiate such study, we defined a novel version  $\text{FO}(\perp_{\text{CSI}})$  of independence logic suitable for formalizing qualitative versions of CSI statements. Team semantics has already been successfully used to axiomatize dependencies in the database theory framework (see, e.g., [16,18]) and it is an interesting open question to formulate general axioms for the logic  $\text{FO}(\perp_{\text{CSI}})$ . In particular, it is an interesting task to identify subclasses of CSI-atoms for which the implication problem is axiomatizable or decidable.

## Conflict of interest statement

No conflict of interest.

## Acknowledgements

We thank the referees for useful comments. The second author was supported by grants 251170 and 295673 of the Academy of Finland. The third author was supported by grants 292767 and 308712 of the

<sup>1</sup> A preliminary implementation of the axiomatization is available at: <https://cs.helsinki.fi/u/ajhyttin/csi/>.

Academy of Finland. The fourth author was supported by FDPSS via grant 141318 of the Academy of Finland.

## References

- [1] C. Boutilier, N. Friedman, M. Goldszmidt, D. Koller, Context-specific independence in Bayesian networks, in: *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence, UAI'96*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996, pp. 115–123, <http://dl.acm.org/citation.cfm?id=2074284.2074298>.
- [2] L. Bravo, W. Fan, S. Ma, Extending dependencies with conditions, in: C. Koch, J. Gehrke, M.N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C.Y. Chan, V. Ganti, C. Kanne, W. Klas, E.J. Neuhold (Eds.), *Proceedings of the 33rd International Conference on Very Large Data Bases*, University of Vienna, Austria, September 23–27, 2007, ACM, 2007, pp. 243–254.
- [3] D.M. Chickering, D. Heckerman, C. Meek, A Bayesian approach to learning Bayesian networks with local structure, in: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 1997, pp. 80–89.
- [4] A.P. Dawid, Conditional independence in statistical theory, *J. Roy. Statist. Soc. Ser. B (Methodological)* 41 (1) (1979) 1–31, <https://doi.org/10.2307/2984718>.
- [5] A. Durand, M. Hannula, J. Kontinen, A. Meier, J. Virtema, Probabilistic team semantics, in: *Foundations of Information and Knowledge Systems – 10th International Symposium, FoIKS 2018, Proceedings*, in: *Lecture Notes in Computer Science*, Springer, 2018.
- [6] A. Durand, M. Hannula, J. Kontinen, A. Meier, J. Virtema, Approximation and dependence via multiteam semantics, *Annals of Mathematics and Artificial Intelligence*, <https://doi.org/10.1007/s10472-017-9568-4>.
- [7] W. Fan, F. Geerts, X. Jia, A. Kementsietsidis, Conditional functional dependencies for capturing data inconsistencies, *ACM Trans. Database Syst.* 33 (2) (2008) 6:1–6:48, <https://doi.org/10.1145/1366102.1366103>.
- [8] N. Friedman, M. Goldszmidt, *Learning Bayesian Networks With Local Structure*, *Learning in Graphical Models*, vol. 89, Springer, 1998.
- [9] P. Galliani, Inclusion and exclusion dependencies in team semantics – on some logics of imperfect information, *Ann. Pure Appl. Logic* 163 (1) (2012) 68–84.
- [10] D. Geiger, D. Heckerman, Knowledge representation and inference in similarity networks and Bayesian multinets, *Artificial Intelligence* 82 (1) (1996) 45–74.
- [11] D. Geiger, J. Pearl, On the logic of causal models, in: *Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence, UAI '88*, North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands, 1990, pp. 3–14, <http://dl.acm.org/citation.cfm?id=647231.719429>.
- [12] D. Geiger, T. Verma, J. Pearl, Identifying independence in Bayesian networks, *Networks* 20 (5) (1990) 507–534.
- [13] D. Geiger, A. Paz, J. Pearl, Axioms and algorithms for inferences involving probabilistic independence, *Inform. and Comput.* 91 (1) (1991) 128–141.
- [14] E. Grädel, J.A. Väänänen, Dependence and independence, *Studia Logica* 101 (2) (2013) 399–410.
- [15] M. Gyssens, M. Niepert, D.V. Gucht, On the completeness of the semigraphoid axioms for deriving arbitrary from saturated conditional independence statements, *Inform. Process. Lett.* 114 (11) (2014) 628–633, <https://doi.org/10.1016/j.ipl.2014.05.010>, <http://www.sciencedirect.com/science/article/pii/S0020019014001057>.
- [16] M. Hannula, Reasoning about embedded dependencies using inclusion dependencies, in: M. Davis, A. Fehnker, A. McIver, A. Voronkov (Eds.), *Logic for Programming, Artificial Intelligence, and Reasoning – 20th International Conference, LPAR-20 2015*, Suva, Fiji, November 24–28, 2015, *Proceedings*, in: *Lecture Notes in Computer Science*, vol. 9450, Springer, 2015, pp. 16–30.
- [17] M. Hannula, Axiomatizing first-order consequences in independence logic, *Ann. Pure Appl. Logic* 166 (1) (2015) 61–91, <https://doi.org/10.1016/j.apal.2014.09.002>.
- [18] M. Hannula, J. Kontinen, A finite axiomatization of conditional independence and inclusion dependencies, *Inform. and Comput.* 249 (2016) 121–137.
- [19] C. Herrmann, On the undecidability of implications between embedded multivalued database dependencies, *Inform. and Comput.* 122 (2) (1995) 221–235.
- [20] D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- [21] J. Kontinen, J.A. Väänänen, Axiomatizing first-order consequences in dependence logic, *Ann. Pure Appl. Logic* 164 (11) (2013) 1101–1117.
- [22] S. Link, Reasoning about saturated conditional independence under uncertainty: axioms, algorithms, and levesque's situations to the rescue, in: *Proc. AAAI*, AAAI Press, 2013.
- [23] S. Link, Sound approximate reasoning about saturated conditional probabilistic independence under controlled uncertainty, *J. Appl. Log.* 11 (3) (2013) 309–327.
- [24] S. Link, Frontiers for propositional reasoning about fragments of probabilistic conditional independence and hierarchical database decompositions, *Theoret. Comput. Sci.* 603 (2015) 111–131.
- [25] S. Ma, W. Fan, L. Bravo, Extending inclusion dependencies with conditions, *Theoret. Comput. Sci.* 515 (2014) 64–95, <https://doi.org/10.1016/j.tcs.2013.11.002>.
- [26] M. Niepert, M. Gyssens, B. Sayrafi, D.V. Gucht, On the conditional independence implication problem: a lattice-theoretic approach, *Artificial Intelligence* 202 (2013) 29–51, <https://doi.org/10.1016/j.artint.2013.06.005>.
- [27] H. Nyman, J. Pensar, J. Corander, Context-specific and local independence in Markovian dependence structures, in: S. Abramsky, J. Kontinen, J. Väänänen, H. Vollmer (Eds.), *Dependence Logic: Theory and Applications*, Springer International Publishing, Cham, 2016, pp. 219–234.

- [28] G. Paolini, Independence logic and abstract independence relations, *Math. Logic Q.* 61 (3) (2015) 202–216, <https://doi.org/10.1002/malq.201400031>.
- [29] G. Paolini, J. Väänänen, Dependence logic in pregeometries and  $\omega$ -stable theories, *J. Symbolic Logic* 81 (1) (2016) 32–55, <https://doi.org/10.1017/jsl.2015.16>.
- [30] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, CA, 1988.
- [31] J. Pensar, H.J. Nyman, T. Koski, J. Corander, Labeled directed acyclic graphs: a generalization of context-specific independence in directed graphical models, *Data Min. Knowl. Discov.* 29 (2) (2015) 503–533, <https://doi.org/10.1007/s10618-014-0355-0>.
- [32] J. Pensar, H. Nyman, J. Lintusaari, J. Corander, The role of local partial independence in learning of Bayesian networks, *Internat. J. Approx. Reason.* 69 (2016) 91–105, <https://doi.org/10.1016/j.ijar.2015.11.008>.
- [33] D. Poole, N.L. Zhang, Exploiting contextual independence in probabilistic inference, *J. Artificial Intelligence Res.* 18 (2003) 263–313.
- [34] M. Studeny, Conditional independence relations have no finite complete characterization, in: S. Kubik, J. Visek (Eds.), *Transactions of the 11th Prague Conference vol. B, Information Theory, Statistical Decision Functions and Random Processes*, Kluwer, Dordrecht–Boston–London, 1992, pp. 377–396.
- [35] J. Väänänen, *Dependence Logic: A New Approach to Independence Friendly Logic*, London Mathematical Society Student Texts, vol. 70, Cambridge University Press, Cambridge, 2007.
- [36] T. Verma, J. Pearl, Causal networks: semantics and expressiveness, in: R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Eds.), *UAI '88: Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence*, Minneapolis, MN, USA, July 10–12, 1988, 1988, pp. 69–78, North-Holland.
- [37] S. Wong, C. Butz, D. Wu, On the implication problem for probabilistic conditional independency, *Systems, Man and Cybernetics, Part A: Systems and Humans*, *IEEE Transactions on* 30 (6) (2000) 785–805, <https://doi.org/10.1109/3468.895901>.