

Finitary S5-Theories

Tran Cao Son¹, Enrico Pontelli¹, Chitta Baral², and Gregory Gelfond²

¹ Computer Science Department, New Mexico State University
{tson, epontell}@cs.nmsu.edu

² Department of Computer Science Engineering, Arizona State University
{chitta, gelfond.greg}@asu.edu

Abstract. The objective of this paper is to identify a class of epistemic logic theories with group knowledge operators which have the fundamental property of being characterized by a *finite* number of *finite* models (up to equivalence). We specifically focus on **S5**-theories. We call this class of epistemic logic theories as *finitary S5-theories*. Models of finitary **S5**-theories can be shown to be canonical in that they do not contain two worlds with the same interpretation. When the theory is *pure*, these models are minimal and differ from each other only in the actual world. The paper presents an algorithm for computing all models of a finitary **S5**-theory. Finitary **S5**-theories find applications in several contexts—in particular, the paper discusses their use in epistemic multi-agent planning.

1 Introduction and Motivation

Epistemic logics [2, 7, 8, 10] are a branch of modal logic that is concerned with representing and reasoning about the knowledge of collections of agents. These logics allow us to represent and reason about the knowledge of an agent about the world, its knowledge about other agents' knowledge, group's knowledge, common knowledge, etc. The models of an epistemic theory are commonly given by *pointed Kripke structures*. Each pointed Kripke structure consists of a set of elements named *worlds* (also known as *points*), a collection of binary relations between worlds (*accessibility relations*), a named valuation associated to each world, and an *actual world*—considered as the “*real state of the universe*”. Models of an epistemic theory can be potentially infinite. Indeed, one can easily create an infinite model of an epistemic theory from a finite one, by cloning its whole structure (including the accessibility relations, the worlds, etc.). *Bisimulation* (e.g., [2]) can be used to reduce the size of a model. It is possible to show that, given a theory that employs a single modal operator and has a finite signature, there are only finitely many models with the property that (a) all of them are finite and bisimulation-based minimal; and (b) any model of the theory is bisimilar (and, hence, equivalent) to one of those models. This is not true, however, for *multimodal* theories, i.e., theories with multiple modal operators.

In this paper, we study the questions of when a multimodal propositional epistemic theory can be characterized by *finitely many finite models* (up to equivalence), and how to compute these models. The motivation for these questions is twofold.

First, the question arises in the research on using epistemic theories in *Multi-Agent Systems (MAS)*, in particular, in the development of the *Dynamic Epistemic Logic (DEL)*

[1, 3, 6, 11] for reasoning about effects of actions in MAS. This line of research has laid the foundations for the study of the epistemic planning problem in multi-agent environments [5, 12, 14]. Yet, the majority of the research in epistemic planning assumes that the set of *initial pointed Kripke structures* is given, and it is either finite [12, 14] or recursively enumerable [5]. This creates a gap between the rich literature in theoretical investigation of epistemic planning (e.g., formalization, complexity results) and the very modest developments in automated epistemic planning systems—that can benefit from the state-of-the-art techniques developed for planning systems in single-agent environments. In particular, there is a plethora of planners for single-agent environments, that perform exceptionally well in terms of scalability and efficiency;¹ the majority of them are heuristic forward-search planners. On the other hand, to the best of our knowledge, the systems described in [12, 14] are the only epistemic multi-agent planning prototypes available, that search for solutions using breath-first search and model checking.

The second research motivation comes from the observation that the **S5**-logic is the de-facto standard logic for reasoning and planning with sensing actions in presence of incomplete information for single-agent domains. The literature is scarce on methods for computing models of **S5** multimodal epistemic theories. Works such as [15, 16] are exceptions, and they focus on the least models of a modal theory. Several papers, instead, assume that such models are, *somehow*, given. For instance, after describing the muddy-children story, the authors of [7] present a model of the theory without detailing how should one construct such model and whether or not the theory has other “interesting” models.

These observations show that, in order to be able to use epistemic logic as a specification language in practical MAS applications, such as epistemic multi-agent planning, the issue of how to compute the set of models of a theory must be addressed.

In this paper, we address this question by identifying a class of *finitary S5-theories* with group and common knowledge operators, that can be characterized by finitely many finite models. We prove that each model of a finitary **S5**-theory is equivalent to one of these canonical models, and propose an effective algorithm for computing such set of canonical models. We discuss a representation of finitary **S5**-theories suitable for use with the algorithm. We also discuss the impact of these results in epistemic multi-agent planning.

2 Preliminary: Epistemic Logic

Let us consider the epistemic logic with a set $\mathcal{AG} = \{1, 2, \dots, n\}$ of n agents; we will adopt the notation used in [2, 7]. The “physical” state of the world is described by a finite set \mathcal{P} of *propositions*. The knowledge of the world of agent i is described by a modal operator \mathbf{K}_i ; in particular, the knowledge of agents is encoded by *knowledge formulae* (or *formula*) in a logic extended with these operators, and defined as follows.

- *Atomic formulae*: an atomic formula is built using the propositions in \mathcal{P} and the traditional propositional connectives $\vee, \wedge, \rightarrow, \neg$, etc. A *literal* is either an atom $f \in \mathcal{P}$ or its negation $\neg f$. \top (resp. \perp) denotes *true* (resp. *false*).

¹ E.g., <http://ipc.icaps-conference.org/> lists 27 participants in the 2011 International Planning Competition.

- *Knowledge formulae*: a knowledge formula is a formula in one of the following forms: (i) An atomic formula; (ii) A formula of the form $\mathbf{K}_i\varphi$, where φ is a knowledge formula; (iii) A formula of the form $\varphi_1 \vee \varphi_2$, $\varphi_1 \wedge \varphi_2$, $\varphi_1 \rightarrow \varphi_2$, or $\neg\varphi_1$, where φ_1, φ_2 are knowledge formulae; (iv) A formula of the form $\mathbf{E}_\alpha\varphi$ or $\mathbf{C}_\alpha\varphi$ where φ is a formula and $\emptyset \neq \alpha \subseteq \mathcal{AG}$.

Formulae of the form $\mathbf{E}_\alpha\varphi$ and $\mathbf{C}_\alpha\varphi$ are referred to as *group formulae*. Whenever $\alpha = \mathcal{AG}$, we simply write $\mathbf{E}\varphi$ and $\mathbf{C}\varphi$ to denote $\mathbf{E}_\alpha\varphi$ and $\mathbf{C}_\alpha\varphi$, respectively. When no confusion is possible, we will talk about formula instead of knowledge formula. Let us denote with $\mathcal{L}_{\mathcal{AG}}^{\mathcal{P}}$ the language of the knowledge formulae over \mathcal{P} and \mathcal{AG} . An *epistemic theory* (or simply a *theory*) over the set of agents \mathcal{AG} and propositions \mathcal{P} is a set of knowledge formulae in $\mathcal{L}_{\mathcal{AG}}^{\mathcal{P}}$. To illustrate the language, we will use the well-known *Muddy Children* problem as a running example. For simplicity of the presentation, let us consider the case with two children.

[Muddy Children] A father says to his two children that at least one of them has mud on the forehead. He then repeatedly asks “do you know whether you are dirty?” The first time the two children answer “no.” The second time both answer “yes.” The father and the children can see and hear each other, but no child can see his own forehead.

Let $\mathcal{AG} = \{1, 2\}$. Let m_i denote that child i is muddy. Some formulae in $\mathcal{L}_{\mathcal{AG}}^{\mathcal{P}}$ are: (i) m_i (i is muddy); (ii) \mathbf{K}_1m_1 (child 1 knows he is muddy); (iii) $\mathbf{K}_1\mathbf{K}_2m_2$ (child 1 knows that child 2 knows that he is muddy); and (iv) $\mathbf{C}_{\{1,2\}}(m_1 \vee m_2)$ (it is common knowledge among the children that at least one is muddy).

The semantics of knowledge formulae relies on the notion of Kripke structures.

Definition 1 (Kripke Structure). A Kripke structure over $\mathcal{AG} = \{1, \dots, n\}$ and \mathcal{P} is a tuple $\langle S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n \rangle$, where S is a set of points, π is a function that associates an interpretation of \mathcal{P} to each element of S (i.e., $\pi : S \rightarrow 2^{\mathcal{P}}$), and $\mathcal{K}_i \subseteq S \times S$ for $1 \leq i \leq n$. A pointed Kripke structure (or, pointed structure, for short) is a pair (M, s) , where M is a Kripke structure and s , called the actual world, belongs to the set of points of M .

For readability, we use $M[S]$, $M[\pi]$, and $M[i]$, to denote the components S , π , and \mathcal{K}_i of M , respectively. Using this notation, $M[\pi](u)$ denotes the interpretation associated to the point u .

Definition 2 (Satisfaction Relation). Given a formula φ and a pointed structure (M, s) :

- $(M, s) \models \varphi$ if φ is an atomic formula and $M[\pi](s) \models \varphi$;
- $(M, s) \models \mathbf{K}_i\varphi$ if for each t such that $(s, t) \in \mathcal{K}_i$, $(M, t) \models \varphi$;
- $(M, s) \models \neg\varphi$ if $(M, s) \not\models \varphi$;
- $(M, s) \models \varphi_1 \vee \varphi_2$ if $(M, s) \models \varphi_1$ or $(M, s) \models \varphi_2$;
- $(M, s) \models \varphi_1 \wedge \varphi_2$ if $(M, s) \models \varphi_1$ and $(M, s) \models \varphi_2$;
- $(M, s) \models \mathbf{E}_\alpha\varphi$ if $(M, s) \models \mathbf{K}_i\varphi$ for every $i \in \alpha$;
- $(M, s) \models \mathbf{C}_\alpha\varphi$ if $(M, s) \models E_\alpha^k\varphi$ for every $k \geq 0$ where $E_\alpha^0\varphi = \varphi$ and $E_\alpha^{k+1} = E_\alpha(E_\alpha^k\varphi)$.

$M \models \varphi$ denotes the fact that $(M, s) \models \varphi$ for each $s \in M[S]$, while $\models \varphi$ denotes the fact that $M \models \varphi$ for all Kripke structures M . We will often depict a Kripke structure

M as a directed labeled graph, with S as the set of nodes and with edges of the form (s, i, t) iff $(s, t) \in \mathcal{K}_i$. We say that u_n is reachable from u_1 if there is a sequence of edges $(u_1, i_1, u_2), (u_2, i_2, u_3), \dots, (u_{n-1}, i_{n-1}, u_n)$ in M .

A Kripke structure denotes the possible “worlds” envisioned by the agents—and the presence of multiple worlds denotes uncertainty and presence of different knowledge. The relation $(s_1, s_2) \in \mathcal{K}_i$ indicates that the knowledge of agent i about the real state of the world is insufficient to distinguish between the state described by point s_1 and the one described by point s_2 . For example, if $(s_1, s_2) \in \mathcal{K}_i$, $M[\pi](s_1) \models \varphi$ and $M[\pi](s_2) \models \neg\varphi$, everything else being the same, then this will indicate that agent i is uncertain about the truth of φ . Figure 1 displays a possible pointed structure for the

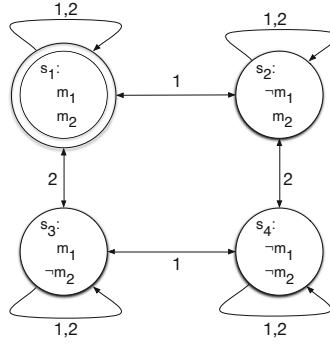


Fig. 1. A possible pointed structure for the Muddy Children Domain

Muddy Children Domain. In Figure 1, a circle represents a point. The name and interpretation of the points are written in the circle. Labeled edges between points denote the knowledge relations of the structure. A double circle identifies the actual world.

Various axioms are used to characterize epistemic logic systems. We will focus on the **S5**-logic that contains the following axioms for each agent i and formulae φ, ψ :

$$\models (\mathbf{K}_i\varphi \wedge \mathbf{K}_i(\varphi \Rightarrow \psi)) \Rightarrow \mathbf{K}_i\psi \quad (\mathbf{K})$$

$$\models \mathbf{K}_i\psi \Rightarrow \psi \quad (\mathbf{T})$$

$$\models \mathbf{K}_i\psi \Rightarrow \mathbf{K}_i\mathbf{K}_i\psi \quad (\mathbf{4})$$

$$\models \neg\mathbf{K}_i\psi \Rightarrow \mathbf{K}_i\neg\mathbf{K}_i\psi \quad (\mathbf{5})$$

A Kripke structure is said to be an **S5**-structure if it satisfies the properties **K**, **T**, **4**, and **5**. It can be shown that the relations \mathcal{K}_i of **S5**-structures are reflexive, transitive, and symmetric. A theory plus the **K**, **T**, **4**, and **5** axioms is often referred to as an **S5**-theory. In the rest of this paper, we will consider only **S5**-theories. A theory T is said to be *satisfiable* (or *consistent*) if there exists a Kripke structure M and a point $s \in M[S]$ such that $(M, s) \models \psi$ for every $\psi \in T$. In this case, (M, s) is referred to as a *model* of T . Two pointed structures (M, s) and (M', s') are *equivalent* if, for every formula $\varphi \in \mathcal{L}_{\mathcal{AG}}^{\mathcal{P}}$, $(M, s) \models \varphi$ iff $(M', s') \models \varphi$.

For simplicity of the presentation, we define

$$\text{state}(u) \equiv \bigwedge_{f \in \mathcal{P}, M[\pi](u)(f) = \top} f \wedge \bigwedge_{f \in \mathcal{P}, M[\pi](u)(f) = \perp} \neg f$$

for $u \in M[S]$. Intuitively, $state(u)$ is the formula representing the complete interpretation associated to the point u in the structure M , i.e., $M[\pi](u)$. We will often use $state(u)$ and $M[\pi](u)$ interchangeably. We say that (M, s) is *canonical* if $state(u) \neq state(v)$ for every $u, v \in M[S]$, $u \neq v$. By $Mod_{S5}(T)$ we denote a set of **S5**-models of a theory T such that: (a) there are no two equivalent models in $Mod_{S5}(T)$; and (b) For each **S5**-model (M, s) of T , there exists a model (M', s') in $Mod_{S5}(T)$ such that (M, s) is equivalent to (M', s') .

3 Finitary **S5**Theories: Definition and Properties

In this section, we define the notion of *finitary **S5**-theories* and show that a finitary **S5**-theory can be characterized by finitely many finite models. We start with the specification of the types of formulae that we will consider. They are, in our observation, sufficiently expressive for use in the specification of the description of the actual world and the common knowledge among the agents. The allowed types of formulae are:

$$\varphi \quad (1)$$

$$\mathbf{C}(\mathbf{K}_i\varphi) \quad (2)$$

$$\mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi) \quad (3)$$

$$\mathbf{C}(\neg\mathbf{K}_i\varphi \wedge \neg\mathbf{K}_i\neg\varphi) \quad (4)$$

where φ is an atomic formula. Intuitively, formulae of type (1) indicate properties that are true in the actual world; formulae of type (2)-(3) indicate that all agents know that agent i is aware of the truth value of φ ; formulae of type (4) indicate that all agents know that agent i is not aware of whether φ is true or false. Since our focus is on **S5**-models of epistemic theories, it is easy to see that $\mathbf{C}(\mathbf{K}_i\varphi)$ can be simplified to $\mathbf{C}(\varphi)$. We say that a formula of the form (1)-(4) is in *disjunctive form* if its formula φ is a disjunction over literals from \mathcal{P} . A *complete clause* over \mathcal{P} is a disjunction of the form $\bigvee_{p \in \mathcal{P}} p^*$ where p^* is either p or $\neg p$.

Example 1. In the muddy children story, the knowledge of the children after the father's announcement but before the children look at each other can be encoded by a theory T_0 consisting of the following formulae:

$$\begin{array}{ll} \mathbf{C}(\mathbf{K}_1(m_1 \vee m_2)) & \mathbf{C}(\mathbf{K}_2(m_1 \vee m_2)) \\ \mathbf{C}(\neg\mathbf{K}_1m_1 \wedge \neg\mathbf{K}_1\neg m_1) & \mathbf{C}(\neg\mathbf{K}_1m_2 \wedge \neg\mathbf{K}_1\neg m_2) \\ \mathbf{C}(\neg\mathbf{K}_2m_1 \wedge \neg\mathbf{K}_2\neg m_1) & \mathbf{C}(\neg\mathbf{K}_2m_2 \wedge \neg\mathbf{K}_2\neg m_2) \end{array}$$

These formulae indicate that both children are aware that at least one of them is muddy (the formulas in the first row), but they are not aware of who among them is muddy; these items are all common knowledge.

If we take into account the fact that each child can see the other, and each child knows if the other one is muddy, then we need to add to T_0 the following formulae:

$$\mathbf{C}(\mathbf{K}_1m_2 \vee \mathbf{K}_1\neg m_2) \quad \mathbf{C}(\mathbf{K}_2m_1 \vee \mathbf{K}_2\neg m_1)$$

Definition 3 (Primitive Finitary S5-Theory). A theory T is said to be primitive finitary **S5** if

- Each formula in T is of the form (1)-(4); and
- For each complete clause φ over \mathcal{P} and each agent i , T contains either (i) $\mathbf{C}(\mathbf{K}_i\varphi)$ or (ii) $\mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$ or (iii) $\mathbf{C}(\neg\mathbf{K}_i\varphi \wedge \neg\mathbf{K}_i\neg\varphi)$.

T is said to be in disjunctive form if all statements in T are in disjunctive form.

The second condition of the above definition deserves some discussion. It requires that T contains at least $|\mathcal{AG}| \times 2^{|\mathcal{P}|}$ formulae and could be unmanageable for large \mathcal{P} . This condition is introduced for simplicity of initial analysis of finitary **S5**-theories. This condition will be relaxed at the end of this section by replacing the requirement “ T contains” with “ T entails.” For example, the theory T_0 is not a primitive finitary **S5**-theory; T_0 is a finitary **S5**-theory (defined later) as it entails a primitive finitary **S5**-theory T_1 .

Example 2. Let T_1 be the theory consisting of:

$$\begin{aligned} & \mathbf{C}(\mathbf{K}_i(m_1 \vee m_2)) \\ & \mathbf{C}(\neg\mathbf{K}_i(m_1 \vee \neg m_2) \wedge \neg\mathbf{K}_i(\neg(m_1 \vee \neg m_2))) \\ & \mathbf{C}(\neg\mathbf{K}_i(\neg m_1 \vee m_2) \wedge \neg\mathbf{K}_i(\neg(\neg m_1 \vee m_2))) \\ & \mathbf{C}(\neg\mathbf{K}_i(\neg m_1 \vee \neg m_2) \wedge \neg\mathbf{K}_i(\neg(\neg m_1 \vee \neg m_2))) \end{aligned}$$

where $i = 1, 2$. It is easy to see that T_1 is a primitive finitary **S5**-theory—and it is equivalent to T_0 from Example 1.

Primitive finitary **S5**-theories can represent interesting properties.

Example 3. Consider the statement “it is common knowledge that none of the agents knows anything.” The statement can be represented by the theory

$$T_2 = \{ \mathbf{C}(\neg\mathbf{K}_i\omega \wedge \neg\mathbf{K}_i\neg\omega) \mid i \in \mathcal{AG}, \omega \text{ is a complete clause over } \mathcal{P} \}.$$

We will show that a primitive finitary **S5**-theory can be characterized by finitely many finite **S5**-models. The proof of this property relies on a series of lemmas. We will next discuss these lemmas and provide proofs of the non-trivial ones. First, we observe that points that are unreachable from the actual world in a pointed structure can be removed.

Lemma 1. Every **S5**-pointed structure (M, s) is equivalent to an **S5**-pointed structure (M', s) such that every $u \in M'[S]$ is reachable from s .

The next lemma studies the properties of an **S5**-pointed structure satisfying a formula of the form (2) or (3).

Lemma 2. Let (M, s) be an **S5**-pointed structure such that every $u \in M[S]$ is reachable from s . Let ψ be an atomic formula. Then,

- $(M, s) \models \mathbf{C}(\psi)$ iff $M[\pi](u) \models \psi$ for every $u \in M[S]$.
- $(M, s) \models \mathbf{C}(\mathbf{K}_i\psi \vee \mathbf{K}_i\neg\psi)$ iff for every pair $(u, v) \in M[i]$ it holds that $M[\pi](u) \models \psi$ iff $M[\pi](v) \models \psi$.

Because $\mathbf{C}(\mathbf{K}_i\psi)$ implies $\mathbf{C}(\psi)$ in an **S5**-pointed structure (M, s) the first item of Lemma 2 shows that ψ is satisfied at every point in (M, s) . The second item of Lemma 2 shows that every pair of points related by \mathcal{K}_i either both satisfy or both do not satisfy the formula φ in an **S5**-pointed structure (M, s) satisfying a formula of the form (3).

The next lemma shows that an **S5**-pointed structure satisfying a formula of the form (4) must have at least one pair of points at which the value of the atomic formula mentioned in the formula differs. For a structure M and $u, v \in M[S]$, $M[\pi](u)(\psi) \neq M[\pi](v)(\psi)$ indicates that either $(M[\pi](u) \models \psi$ and $M[\pi](v) \not\models \psi$) or $(M[\pi](u) \not\models \psi$ and $M[\pi](v) \models \psi)$, i.e., the value of ψ at u is different from the value of ψ at v .

Lemma 3. *Let (M, s) be an **S5**-pointed structure such that every $u \in M[S]$ is reachable from s . Let ψ be an atomic formula. Then, $(M, s) \models \mathbf{C}(\neg\mathbf{K}_i\psi \wedge \neg\mathbf{K}_i\neg\psi)$ iff for every $u \in M[S]$ there exists some $v \in M[S]$ such that $(u, v) \in M[i]$, and $M[\pi](u)(\psi) \neq M[\pi](v)(\psi)$.*

The proofs of Lemmas 1-3 follow from the definition of the satisfaction relation \models between a pointed structure and a formula and the fact that $(M, s) \models \mathbf{C}(\psi)$ iff $(M, u) \models \psi$ for every u reachable from s . For this reason, they are omitted.

We will now focus on models of primitive finitary **S5**-theories. Let M be a Kripke structure. We define a relation \sim among points of M as follows. For each $u, v \in M[S]$, $u \sim v$ iff $\text{state}(u) \equiv \text{state}(v)$. Thus, $u \sim v$ indicates that the interpretations associated to u and v are identical. It is easy to see that \sim is an equivalence relation over $M[S]$. Let \tilde{u} denote the equivalence class of u with respect to the relation \sim (i.e., $\tilde{u} = [u]_{\sim}$).

Lemma 4. *Let (M, s) be an **S5**-model of a primitive finitary **S5**-theory such that every $u \in M[S]$ is reachable from s . Let φ be a complete clause and $i \in \mathcal{AG}$. Given $u \in M[S]$:*

- If $(M, u) \models \mathbf{K}_i\varphi$ then $(M, s) \models \mathbf{C}(\mathbf{K}_i\varphi)$ or $(M, s) \models \mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$;
- If $(M, u) \models \neg\mathbf{K}_i\varphi$ then $(M, s) \models \mathbf{C}(\neg\mathbf{K}_i\varphi \wedge \neg\mathbf{K}_i\neg\varphi)$.

The proof of Lemma 4 makes use of Lemmas 2-3 and the fact that (M, s) is an **S5**-model of a primitive finitary **S5**-theory. The next lemma states a fundamental property of models of primitive finitary **S5**-theories.

Lemma 5. *Let (M, s) be an **S5**-model of a primitive finitary **S5**-theory such that every $u \in M[S]$ is reachable from s . Let $u, v \in M[S]$ such that $u \sim v$. Then, for every $i \in \mathcal{AG}$ and $x \in M[S]$ such that $(u, x) \in M[i]$ there exists $y \in M[S]$ such that $(v, y) \in M[i]$ and $x \sim y$.*

Proof. Let $K(p, i) = \{q \mid q \in M[S], (p, q) \in M[i]\}$ —i.e., the set of points immediately related to p via $M[i]$. We consider two cases:

- **Case 1:** $K(u, i) \cap K(v, i) \neq \emptyset$. Since $M[i]$ is an equivalent relation, we can conclude that $K(u, i) = K(v, i)$ and the lemma is trivially proved (by taking $x = y$).
- **Case 2:** $K(u, i) \cap K(v, i) = \emptyset$. Let us assume that there exists some $x \in K(u, i)$ such that there exists no $y \in K(v, i)$ with $x \sim y$. This means that $(M, y) \models \neg\text{state}(x)$ for each $y \in K(v, i)$. In other words, $(M, v) \models \mathbf{K}_i(\neg\text{state}(x))$. As $\neg\text{state}(x)$ is a complete clause, this implies that (by Lemma 4):

$$(M, s) \models \mathbf{C}(\mathbf{K}_i \neg \text{state}(x)) \text{ or } (M, s) \models \mathbf{C}(\mathbf{K}_i \neg \text{state}(x) \vee \mathbf{K}_i \text{state}(x)) \quad (5)$$

On the other hand, $(M, u) \not\models \mathbf{K}_i(\neg \text{state}(x))$, since $x \in K(u, i)$ and $(M, x) \models \text{state}(x)$. This implies $(M, s) \models \mathbf{C}(\neg \mathbf{K}_i \neg \text{state}(x) \wedge \neg \mathbf{K}_i \text{state}(x))$ by Lemma 4.

This contradicts (5), proving the lemma. \square

Lemma 5 shows that the points with the same interpretation have the same structure in an **S5**-model of a primitive finitary **S5**-theory, i.e., the accessibility relations associated to these points are identical. This indicates that we can group all such points into a single one, producing an equivalent model that is obviously finite. Let us show that this is indeed the case. Given a structure M , let \widetilde{M} be the structure constructed as follows:

- $\widetilde{M}[S] = \{\tilde{u} \mid u \in M[S]\}$
- For every $u \in M[S]$ and $f \in \mathcal{P}$, $\widetilde{M}[\pi](\tilde{u})(f) = M[\pi](u)(f)$
- For each $i \in \mathcal{AG}$, $(\tilde{u}, \tilde{v}) \in \widetilde{M}[i]$ if there exists $(u', v') \in M[i]$ such that $u' \in \tilde{u}$ and $v' \in \tilde{v}$.

We call $(\widetilde{M}, \tilde{s})$ the *reduced pointed structure* of (M, s) and prove that it is an **S5**-pointed structure equivalent to (M, s) :

Lemma 6. *Let (M, s) be an **S5**-model of a primitive finitary **S5**-theory T such that every $u \in M[S]$ is reachable from s . Furthermore, let $(\widetilde{M}, \tilde{s})$ be the reduced pointed structure of (M, s) . Then, $(\widetilde{M}, \tilde{s})$ is a finite **S5**-model of T that is equivalent to (M, s) .*

Proof. The proof of this lemma relies on Lemmas 2-3 and 5. We prove some representative properties.

- $(\widetilde{M}, \tilde{s})$ is **S5**. Reflexivity and symmetry are obvious. Let us prove transitivity: assume that $(\tilde{u}, \tilde{v}) \in \widetilde{M}[i]$ and $(\tilde{v}, \tilde{w}) \in \widetilde{M}[i]$. The former implies that there exists $(u_1, v_1) \in M[i]$ for some $u_1 \in \tilde{u}$ and $v_1 \in \tilde{v}$. The latter implies that there exists $(x_1, w_1) \in M[i]$ for some $x_1 \in \tilde{v}$ and $w_1 \in \tilde{w}$. Since \sim is an equivalence relation, $v_1 \sim x_1$. Lemma 5 implies that there exists some $w_2 \sim w_1$ such that $(v_1, w_2) \in M[i]$ which implies that, by transitivity of $M[i]$, $(u_1, w_2) \in M[i]$, so $(\tilde{u}, \tilde{w}) \in \widetilde{M}[i]$, i.e., $\widetilde{M}[i]$ is transitive.
- $(\widetilde{M}, \tilde{s})$ is a model of T . We have that $(M, s) \models \mathbf{C}(\mathbf{K}_i \psi \vee \mathbf{K}_i \neg \psi)$ iff $\forall u, v \in M[S], (u, v) \in M[i]$ implies $M[\pi](u) \models \psi$ iff $M[\pi](v) \models \psi$ (by Lemma 2 w.r.t. (M, s)) iff $\forall \tilde{p}, \tilde{q} \in \widetilde{M}[S], u \in \tilde{p}$ and $v \in \tilde{q}, (\tilde{p}, \tilde{q}) \in \widetilde{M}[i]$ implies $\widetilde{M}[\pi](\tilde{p}) \models \psi$ iff $\widetilde{M}[\pi](\tilde{q}) \models \psi$ (construction of $(\widetilde{M}, \tilde{s})$) iff $(\widetilde{M}, \tilde{s}) \models \mathbf{C}(\mathbf{K}_i \psi \vee \mathbf{K}_i \neg \psi)$ (by Lemma 2 w.r.t. $(\widetilde{M}, \tilde{s})$).
The proof for other statements is similar.
- $(\widetilde{M}, \tilde{s})$ is equivalent to (M, s) . This is done by induction over the number of **K** operators in a formula. \square

Let

$$\mu \text{Mods}_{\mathbf{S5}}(T) = \left\{ (\widetilde{M}, \tilde{s}) \mid \begin{array}{l} (\widetilde{M}, \tilde{s}) \text{ is a reduced pointed structure} \\ \text{of a } \mathbf{S5}\text{-model } (M, s) \text{ of } T \end{array} \right\}$$

Since each **S5**-model of T is equivalent to its reduced pointed structure, which has at most $2^{|\mathcal{P}|}$ points, we have the next theorem.

Theorem 1. *For a consistent primitive finitary **S5**-theory T , $\mu\text{Mod}_{\mathbf{S5}}(T)$ is finite and such that each (M, s) in $\mu\text{Mod}_{\mathbf{S5}}(T)$ is also finite.*

This theorem shows that primitive finitary **S5**-theories have the desired properties that we are looking for. The next theorem proves interesting properties of models of primitive finitary **S5**-theories which are useful for computing $\mu\text{Mod}_{\mathbf{S5}}(T)$.

Theorem 2. *For a primitive finitary **S5**-theory T , every model (M, s) in $\mu\text{Mod}_{\mathbf{S5}}(T)$ is canonical and $|M[S]|$ is minimal among all models of T . Furthermore, for every pair of models (M, s) and (W, w) in $\mu\text{Mod}_{\mathbf{S5}}(T)$, M and W are identical, up to the names of the points.*

The first conclusion is trivial as each reduced pointed structure of a model of T is a canonical model of T . The next lemma proves the second conclusion.

Lemma 7. *Let T be a primitive finitary **S5**-theory, (M, s) and (V, w) in $\mu\text{Mod}_{\mathbf{S5}}(T)$, and let $i \in \mathcal{AG}$.*

- *For each $u \in M[S]$ there exists some $v \in V[S]$ such that $\text{state}(u) \equiv \text{state}(v)$.*
- *If $(u, p) \in M[i]$ then there exists $(v, q) \in V[i]$ such that $\text{state}(u) \equiv \text{state}(v)$ and $\text{state}(p) \equiv \text{state}(q)$.*

Proof. (Sketch) The proof of the first property is similar to the proof of Lemma 5, with the minor modification that it refers to two structures and that both are models of T . In fact, if $u \in M[S]$ and there exists no $v \in V[S]$ such that $\text{state}(u) \equiv \text{state}(v)$ then $(V, w) \models \mathbf{C}(\mathbf{K}_k \neg \text{state}(u))$ and $(M, s) \not\models \mathbf{C}(\mathbf{K}_k \neg \text{state}(u))$ for $k \in \mathcal{AG}$, a contradiction. The proof of the second property uses a similar argument. \square

To prove that the set of points of a model in $\mu\text{Mod}_{\mathbf{S5}}(T)$ is minimal, we use the next lemma. We define:

$$F(T) = \{\varphi \mid \varphi \text{ appears in a formula of the form (2) of } T\}.$$

Lemma 8. *Let (M, s) be a canonical model of a primitive finitary **S5**-theory T . Then, the set $M[S]$ is exactly the set of interpretations of $F(T)$ and each $u \in M[S]$ is reachable.*

Proof. (Sketch) First, it follows directly from Lemma 2 and $\mathbf{C}(\mathbf{K}_i \psi) \models \mathbf{C}(\psi)$ in an **S5**-model that for each $u \in M[S]$, $\text{state}(u) \models \varphi$ for every $\varphi \in F(T)$. Second, because T is primitive finitary, if there is some interpretation I of $F(T)$ such that there exists no $u \in M[S]$ and $\text{state}(u) = I$ or there exists $u \in M[S]$ with $\text{state}(u) = I$ and u is not reachable from s then we can conclude that $(M, s) \models \mathbf{C}(\neg I)$, and because T is a primitive finitary, we have that $\neg I \in F(T)$. This implies that I cannot be an interpretation of $F(T)$, a contradiction. Both properties prove the lemma. \square

We are now ready to define the notion of a *finitary **S5**-theory* that allows for Theorem 1 to extend to epistemic theories consisting of arbitrary formulae.

Definition 4 (Finitary **S5-Theory).** *An epistemic theory T is a finitary **S5**-theory if $T \models H$ and H is a primitive finitary **S5**-theory. T is pure if T contains only formulae of the form (1)-(4).*

We have that T_0 (Example 1) is a finitary **S5**-theory, since $T_0 \models T_1$ and T_1 is a primitive finitary **S5**-theory. Since a model of T is also a model of H if $T \models H$, the following theorem holds.

Theorem 3. *Every finitary **S5**-theory T has finitely many finite canonical models, up to equivalence. If T is pure then these models are minimal and their structures are identical up to the name of the points.*

4 Computing All Models of Finitary **S5**Theories

In this section, we present an algorithm for computing $\mu\text{Mod}_{\mathbf{S5}}(T)$ for a primitive finitary **S5**-theory and discuss how this can be extended to arbitrary finitary **S5**-theories. Lemma 8 shows that $F(T)$ can be used to identify the set of points of canonical models of T . Applying this lemma to T_0 (Example 1), we know that for every canonical model (M, s) of T_0 , $M[S] = \{s_1, s_2, s_3\}$ where $\text{state}(s_1) = m_1 \wedge m_2$, $\text{state}(s_2) = m_1 \wedge \neg m_2$, and $\text{state}(s_3) = \neg m_1 \wedge m_2$.

The next step is to determine the accessibility relations of $i \in \mathcal{AG}$. We will rely on Lemmas 2-3 and the following result:

Lemma 9. *Let (M, s) be a canonical model of a consistent primitive finitary **S5**-theory T and $i \in \mathcal{AG}$. Assume that for each complete clause φ , if $T \not\models \mathbf{C}(\mathbf{K}_i\varphi)$ then $T \not\models \mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$. Then, $(u, v) \in M[i]$ for every pair $u, v \in M[S]$.*

Proof. The proof of this lemma is by contradiction and uses an idea similar to that used in the proof of Case 2 of Lemma 5. Since (M, s) is a canonical model, each $u \in M[S]$ is reachable from s . Assume that there exists a pair $u, v \in M[S]$ such that $(u, v) \notin M[i]$. We have that $(M, u) \models \mathbf{K}_i\neg\text{state}(v)$. As $\neg\text{state}(v)$ is a complete clause, by Lemma 4:

$$(M, s) \models \mathbf{C}(\mathbf{K}_i\neg\text{state}(v)) \quad \text{or} \quad (M, s) \models \mathbf{C}(\mathbf{K}_i\neg\text{state}(v) \vee \mathbf{K}_i\text{state}(v)) \quad (6)$$

On the other hand, since $(u, v) \notin M[i]$ and M is a **S5**-structure, we have that $(M, v) \not\models \mathbf{K}_i\neg\text{state}(v)$. This, together with the assumption of the lemma, contradicts (6). \square

Algorithm 1² computes all canonical minimal models of a primitive finitary **S5**-theory. Its correctness follows from the properties of an **S5**-model of primitive finitary **S5**-theories discussed in Lemmas 2-3 and 7-9. This algorithm runs in polynomial time in the size of T , which, unfortunately, is exponential in the size of \mathcal{P} .

Note that, for the theory T_2 in Example 3, Algorithm 1 returns the set of pointed structures (M, s) such that $M[S]$ is the set of all interpretations of \mathcal{P} , $M[i]$ is a complete graph on $M[S]$, and $s \in M[S]$.

Fig. 2 shows one model of T_1 returned by Algorithm 1. Since $\mathbf{C}(\mathbf{K}_i(l_1 \vee l_2) \vee \mathbf{K}_i\neg(l_1 \vee l_2)) \notin T_1$ for every complete clause over $\{m_1, m_2\}$ that is different from $m_1 \vee m_2$, there is a link labeled i between every pair of worlds of the model. Since $I(T_1)$ is empty, $\mu\text{Mod}_{\mathbf{S5}}(T_1)$ contains three models, which differ from each other only in the actual world.

² We assume that the theory is consistent.

Algorithm 1. $Model(T)$

1. **Input:** A primitive finitary **S5**-theory T
 2. **Output:** $\mu Mod_{S5}(T)$
 3. Compute $I(T) = \{\varphi \mid \varphi \text{ appears in some (1) of } T\}$
 4. Compute $F(T) = \{\varphi \mid \varphi \text{ appears in some (2) of } T\}$
 5. $\Sigma = \{u \mid u \text{ is an interpretation satisfying } F(T)\}$
 6. Let $M[S] = \Sigma$, $M[\pi](u) = u$, and $M[i] = \{(u, v) \mid u, v \in \Sigma\}$
 7. **for each** $C(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$ in T **do**
 8. remove $(u, v) \in M[i]$ such that $M[\pi](u)(\varphi) \neq M[\pi](v)(\varphi)$
 9. **end for**
 10. **return** $\{(M, s) \mid s \text{ satisfies } I(T)\}$
-

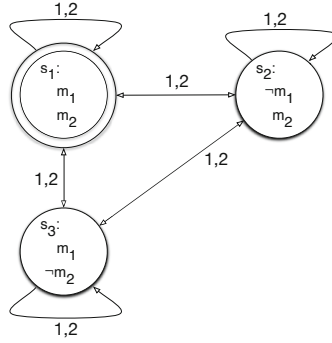


Fig. 2. A model of the theory T_1 in Example 2

The application of Algorithm 1 to an arbitrary finitary **S5**-theory T , where $T \models H$ for some primitive finitary **S5**-theory H , can be done in two steps: (a) Compute $\mu Mod_{S5}(H)$; and (b) Eliminate models from $\mu Mod_{S5}(H)$ which are not a model of T . Step (b) is necessary, since T can contain other formulae that are not entailed by H .³ To accomplish (a), the following tasks need to be performed: (i) Verify that T is finitary; (ii) Compute $I(T) = \{\psi \mid \psi \text{ is an atomic formula and } T \models \psi\}$ (Line 3) and $F(T) = \{\varphi \mid T \models C(\varphi)\}$ (Line 4); (iii) Test for entailment (Line 7); (iv) Eliminate pointed structures that are not models of T (Line 10). Since these tasks are generally computational expensive, it is naturally to seek ways to improve performance. In the next section, we discuss a possible way to deal with (iv). We will next show that when T is pure and in disjunctive form then the computation required in (i)-(iii) can be done in polynomial time in the size of T .

Given a pure theory T in disjunctive form, Task (ii) can be done as described in Lines 3 and 4 and does not require any additional computation. Given a pair (i, φ) of an agent i and a complete clause φ , we would like to efficiently determine whether $T \models C(\mathbf{K}_i\varphi)$, $T \models C(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$, or $T \models C(\neg\mathbf{K}_i\varphi \wedge \neg\mathbf{K}_i\neg\varphi)$ hold. This task can be accomplished via a test for coverage defined as follows. We say that φ is *covered* by a set W of disjunctions over \mathcal{P} if $\varphi \equiv \bigvee_{\psi \in W} \psi$. A pair (i, φ) is *covered* by T if

³ A consequence of this elimination is that canonical models of a non-pure finitary **S5**-theory may not have the same structure and/or different set of worlds.

- T contains some statement $\mathbf{C}(\mathbf{K}_k\psi)$ (for some $k \in \mathcal{AG}$) such that $\psi \models \varphi$; or
- φ is covered by some consistent set of disjunctions
 $W \subseteq \{\psi \mid \mathbf{C}(\mathbf{K}_i\psi \vee \mathbf{K}_i\neg\psi) \in T\}$.

Intuitively, if (i, φ) is covered by T then it is common knowledge that i knows the truth value of φ . The first item implies that $T \models \mathbf{C}(\mathbf{K}_i\varphi)$, i.e., everyone knows φ . The second item states that everyone knows that i knows φ —because (i) φ is covered by a set of disjunctions that are known by i , (ii) this is common knowledge, and (iii) the axiom $\models (\mathbf{K}\psi_1 \vee \mathbf{K}\neg\psi_1) \wedge (\mathbf{K}\psi_2 \vee \mathbf{K}\neg\psi_2) \Rightarrow \mathbf{K}((\psi_1 \vee \psi_2) \vee \mathbf{K}\neg(\psi_1 \vee \psi_2))$. Thus, if φ is a complete clause and (i, φ) is not covered by T then $T \not\models \mathbf{C}(\mathbf{K}_i\varphi)$ and $T \not\models \mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)$. It is easy to see that checking whether (i, φ) is covered by T can be done in polynomial time in the size of T when T is pure and in disjunctive form.

The above discussion shows that, when T is pure and in disjunctive form, Algorithm 1 can compute all models of T , if it is finitary, without significant additional cost. For example, T_0 is pure and in disjunctive form and Algorithm 1 will return the same set of models as if T_1 is used as input.

5 Discussion

The previous sections focused on the development of the notion of a finitary **S5**-theory and the computation of its models. We now discuss a potential use of finitary **S5**-theories as a specification language. Specifically, we consider their use in the specification of the initial set of pointed structures for epistemic multi-agent planning. Let us consider a simple example concerning the Muddy Children Domain: “The father sees that his two children are muddy. The children can see each other, hear the father, and truthfully answer questions from the father but cannot talk to each other. They also know that none of the children knows whether he is muddy or not. How can the father inform his children that both of them are muddy without telling them the fact?” As we have mentioned earlier, previous works in epistemic multi-agent planning assume that the set of initial pointed structures is given, and these are assumed to be finite or enumerable. However, a way to specify the set of initial pointed structures is not offered. Clearly, finitary **S5**-theories can fill this need. Let us discuss some considerations in the use of finitary **S5**-theories as a specification language.

The definition of a finitary **S5**-theory, by Definitions 3-4, calls for the test of entailment (or the specification) of $|\mathcal{AG}| \times 2^{|\mathcal{P}|}$ formulae. Clearly, this is not desirable. To address this issue, let us observe that Algorithm 1 makes use of formulae of the form (4) implicitly (Line 6), by assuming that all but those complete clauses entailed by $F(T)$ are unknown to agent i and that is common knowledge. This means that we could reduce the task of specifying T by assuming that its set of statements of the form (4) is given implicitly, i.e., by representing the information that the agents do not know *implicitly*. This idea is similar to the use of the *Closed World Assumption* to represent incomplete information. This can be realized as follows. For a theory T and an agent $i \in \mathcal{AG}$, let

$$C(T, i) = \{\varphi \mid \varphi \text{ is a complete clause, } T \not\models \mathbf{C}(\mathbf{K}_i\varphi), T \not\models \mathbf{C}(\mathbf{K}_i\varphi \vee \mathbf{K}_i\neg\varphi)\}$$

Let $neg(T) = \bigcup_{i \in \mathcal{AG}} \{\mathbf{C}(\neg\mathbf{K}_i\varphi \wedge \neg\mathbf{K}_i\neg\varphi) \mid \varphi \in C(T, i)\}$. The *completion* of T is $comp(T) = T \cup neg(T)$.

Given an arbitrary theory T , $comp(T)$ is a finitary **S5**-theory; as such, it could be used as the specification of a finitary **S5**-theory. If $comp(T)$ is used and T is pure and in disjunctive form, then the specification (of T) only requires statements of the form (1)-(3). As such, finitary **S5**-theories can be used in a manner similar to how the conventional PDDL problem specification describes the initial states—for epistemic multi-agent planning. We expect that this can help bridging the gap between the development of epistemic multi-agent planning systems and the research in reasoning about the effects of actions in multi-agent domains mentioned earlier since several approaches to reasoning about actions and changes in multi-agent domains (e.g., [1, 3, 6, 11]) facilitate the implementation of a forward search planner in multi-agent domains.

We close the section with a brief discussion on other potential uses of finitary **S5**-theories. Finitary **S5**-theories are useful in applications where knowing that a property is true/false is insufficient, e.g., knowing that a theorem is correct is good but knowing the proof of the theorem (its witness) is necessary; knowing that a component of a system malfunctions is a good step in diagnosis but knowing why this is the case is better; knowing that a plan exists does not help if the sequence of actions is missing; etc.

6 Conclusion and Future Work

In this paper, we proposed the notion of *finitary S5-theories* and showed that a finitary **S5**-theory has finitely many finite **S5**-models. We proved that models of primitive finitary **S5**-theories share the same structure and have minimal size in terms of the number of worlds. We presented an algorithm for computing all canonical **S5**-models of a finitary **S5**-theory. We also argued that the algorithm runs in polynomial time in the size of a pure finitary **S5**-theory in disjunctive form. We proposed the use of completion of finitary **S5**-theories, enabling the implicit representation of negative knowledge, as a specification language in applications like epistemic multi-agent planning.

As future work, we plan to expand this research in four directions. First, we will experiment with the development of an epistemic multi-agent planner. Second, we will investigate possible ways to relax the conditions imposed on finitary **S5**-theories, while still maintaining its finiteness property. Third, we intend to investigate the relationships between the notion of completion of finitary **S5**-theory and the logic of only knowing for multi-agent systems developed by others (e.g., [13, 9]). Finally, we would like to identify situations in which the **S5**-requirements can be lifted.

Acknowledgments. The research has been partially supported by NSF grants HRD-1345232 and DGE-0947465. The authors would like to thank Guram Bezhanishvili and Nguyen L. A for the useful discussions and suggestions.

References

- [1] Baltag, A., Moss, L.: Logics for epistemic programs. Synthese (2004)
- [2] van Benthem, J.: Modal Logic for Open Minds. Center for the Study of Language and Information (2010)

- [3] van Benthem, J., van Eijck, J., Kooi, B.P.: Logics of communication and change. *Inf. Comput.* 204(11), 1620–1662 (2006)
- [4] Blackburn, P., Van Benthem, J., Wolter, F. (eds.): *Handbook of Modal Logic*. Elsevier (2007)
- [5] Bolander, T., Andersen, M.: Epistemic Planning for Single and Multi-Agent Systems. *Journal of Applied Non-Classical Logics* 21(1) (2011)
- [6] van Ditmarsch, H., van der Hoek, W., Kooi, B.: *Dynamic Epistemic Logic*. Springer (2007)
- [7] Fagin, R., Halpern, J., Moses, Y., Vardi, M.: *Reasoning about Knowledge*. MIT Press (1995)
- [8] Gabbay, D., Kurucz, A., Wolter, F., Zakharyashev, M.: *Many-Dimensional Modal Logics: Theory and Application*. Elsevier (2003)
- [9] Halpern, J.Y., Lakemeyer, G.: Multi-agent only knowing. In: Shoham, Y. (ed.) *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, De Zeeuwse Stromen, The Netherlands, pp. 251–265. Morgan Kaufmann (1996)
- [10] Halpern, J., Moses, Y.: A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence* 54, 319–379 (1992)
- [11] Herzig, A., Lang, J., Marquis, P.: Action Progression and Revision in Multiagent Belief Structures. In: *Sixth Workshop on Nonmonotonic Reasoning, Action, and Change, NRAC* (2005)
- [12] van der Hoek, W., Wooldridge, M.: Tractable multiagent planning for epistemic goals. In: *Proceedings of The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002, Bologna, Italy*, pp. 1167–1174. ACM (2002)
- [13] Lakemeyer, G., Levesque, H.J.: Only-knowing meets nonmonotonic modal logic. In: Brewka, G., Eiter, T., McIlraith, S.A. (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14*. AAAI Press (2012)
- [14] Löwe, B., Pacuit, E., Witzel, A.: DEL planning and some tractable cases. In: van Ditmarsch, H., Lang, J., Ju, S. (eds.) *LORI 2011*. LNCS, vol. 6953, pp. 179–192. Springer, Heidelberg (2011)
- [15] Nguyen, L.A.: Constructing the least models for positive modal logic programs. *Fundam. Inform.* 42(1), 29–60 (2000)
- [16] Nguyen, L.A.: Constructing finite least kripke models for positive logic programs in serial regular grammar logics. *Logic Journal of the IGPL* 16(2), 175–193 (2008)