

Two études on computer science and modal logic

Michael Zakharyashev

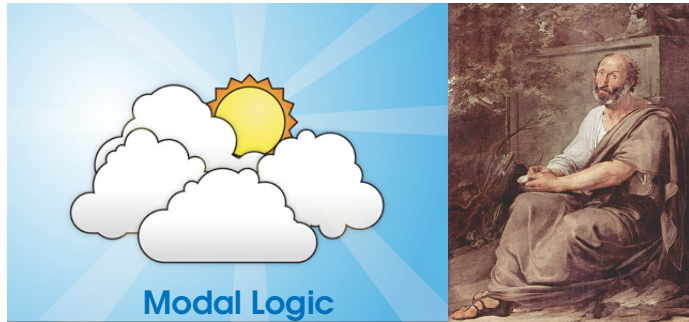
School of Computer Science, Birkbeck College, London

<http://www.dcs.bbk.ac.uk/~michael>

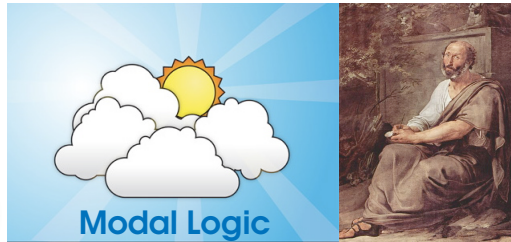
joint work with

Roman Kontchakov, Ian Pratt-Hartmann and Frank Wolter

Modal Logic (before the 1970s)



Modal Logic and Computer Science



Modal Logic

Dynamic logics

Description logics

Spatial logics

Temporal logics



Computer Science

Étude I: Description Logic

Description Logic: a family of formal languages tailored for representing knowledge about concepts and concept hierarchies

- Invented in 1980s as logic-based formalisations of semantic networks & frames
Identified in 1990s as a **'notational variant'** of modal and hybrid logics
- Has become very popular as **ontology languages** (Snomed, GO, NCI, ...) SNOMED will be a core component of the NHS Connecting for Health IT project; already worth **£6.5 billion**
- Was recognised as the **'cornerstone of the Semantic Web'** for providing a formal basis for the Web Ontology Language (OWL)
- Applications in database integration, querying via ontologies, etc.

Ontology example

- NASA's SWEET ontologies

(Semantic Web for Earth and Environmental Terminology)

<http://sweet.jpl.nasa.gov/ontology/>

- Protégé ontology editor

<http://protege.stanford.edu/>

- with DL reasoners Fact++ and Pellet

<http://owl.man.ac.uk/factplusplus/> and <http://pellet.owldl.com/>

Description logic \mathcal{ALCQI} (simplified OWL)

Vocabulary:

- individuals a_0, a_1, \dots
(e.g., john, mary)
- concept names A_0, A_1, \dots
(e.g., Person, Female)
- role names R_0, R_1, \dots
(e.g., hasChild, loves)
- roles

$R ::= R_i \mid R_i^-$

$\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ an **interpretation**

$$a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$$

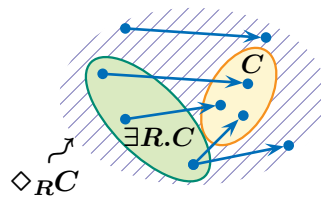
$$A_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$$

$$R_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$$

$$(R_i^-)^{\mathcal{I}} = \{(y, x) \mid (x, y) \in R_i^{\mathcal{I}}\}$$

- concepts

$C ::= A_i \mid \neg C \mid C_1 \sqcap C_2 \mid \exists R.C \mid \forall R.C \mid \geq q R.C$



$\forall R.C$

\equiv

$\neg(\exists R.\neg C)$

'there are at least q distinct R -successors that are in C '

Description logic *ALCQI* (cont.)

knowledge base \mathcal{K} = TBox \mathcal{T} + ABox \mathcal{A}

- \mathcal{T} is a set of **terminological axioms** of the form $C \sqsubseteq D$
- \mathcal{A} is a set of **assertions** of the form $C(a)$ and $R(a, b)$

Reasoning:

– satisfiability, subsumption $\mathcal{K} \models C \sqsubseteq D$

combined complexity

EXPTIME

– instance checking $\mathcal{K} \models C(a)$

data (ABox) complexity

– conjunctive query answering $\mathcal{K} \models q(\vec{a})$

coNP

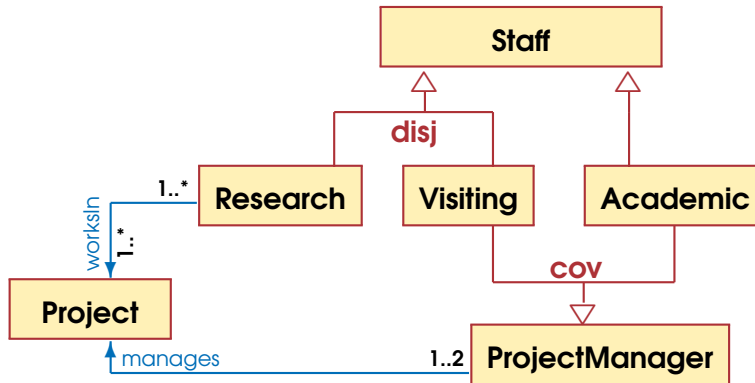
$q(\vec{a}) = \exists \vec{y} \varphi(\vec{a}, \vec{y})$, φ a conjunction of atoms

Reasoners: RACER, FaCT++, Pellet cope well with subsumption & instance checking

What about 'islands of tractability' in DL?

DL 'islands of tractability': *DL-Life* family

- Typical conceptual database schema (a fragment):



Translating into DL:

$\exists \text{manages}.\mathbf{T} \sqsubseteq \text{ProjectManager}$

$\exists \text{manages}^{\neg}.\mathbf{T} \sqsubseteq \text{Project}$

$\text{Project} \sqsubseteq \exists \text{manages}^{\neg}.\mathbf{T}$

$\geq 3 \text{manages}^{\neg}.\mathbf{T} \sqsubseteq \perp$

$\text{Research} \sqcap \text{Visiting} \sqsubseteq \perp$

$\text{Academic} \sqsubseteq \text{ProjectManager}$

$\text{ProjectManager} \sqsubseteq \text{Academic} \sqcup \text{Visiting}$

...

DL 'islands of tractability': *DL-Lite* family (cont.)

1. *DL-Lite_{bool}*

$$\begin{aligned} R &::= P \mid P^- \\ B &::= \perp \mid A \mid \geq qR \\ C &::= B \mid \neg C \mid C_1 \sqcap C_2 \end{aligned}$$

combined complexity: **NP**
data comp. instance: **LogSpace**
data comp. query: **coNP**

TBox axioms $C_1 \sqsubseteq C_2$ ABox assertions: $C(a), R(a, b)$

2. *DL-Lite_{horn}*

TBox axioms $B_1 \sqcap \dots \sqcap B_n \sqsubseteq B$

combined complexity: **P**
data comp. instance: **LogSpace**
data comp. query: **LogSpace**

3. *DL-Lite_{krom}*

TBox axioms $B_1 \sqsubseteq B_2$ $B_1 \sqsubseteq \neg B_2$ $\neg B_1 \sqsubseteq B_2$
(subclass) (disjointness)

comb. comp.: **NLOGSPACE**
d.c. instance: **LogSpace**
d.c. query: **coNP**

NB: these complexity results are closely related to
complexity of reasoning in fragments of propositional logic

Σ-inseparability

\mathcal{T}_1 :

Research \sqcap Visiting $\sqsubseteq \perp$	Academic $\sqsubseteq = 1$ teaches
\exists teaches \sqsubseteq Academic \sqcup Research	\exists writes \sqsubseteq Academic \sqcup Research
Research $\sqsubseteq \exists$ worksIn	\exists worksIn ⁻ \sqsubseteq Project
Project $\sqsubseteq \exists$ manages ⁻	\exists manages \sqsubseteq Academic \sqcup Visiting

$\mathcal{T}_2 = \mathcal{T}_1 \cup \{\text{Visiting } \sqsubseteq \geq 2 \text{ writes}\}, \quad \Sigma = \{\text{teaches}\}$

- Are \mathcal{T}_1 and \mathcal{T}_2 Σ-concept inseparable?

YES

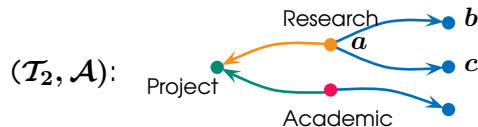
$\mathcal{T}_1 \not\models \text{Visiting } \sqsubseteq = 1 \text{ teaches}$

$\mathcal{T}_2 \models \text{Visiting } \sqsubseteq = 1 \text{ teaches}$

- Are \mathcal{T}_1 and \mathcal{T}_2 Σ-query inseparable?

$\mathcal{A} := \{\text{teaches}(a, b), \text{teaches}(a, c)\}$

NO: $q := \exists x (= 1 \text{ teaches})(x)$
 $(\mathcal{T}_1, \mathcal{A}) \not\models q$ but $(\mathcal{T}_2, \mathcal{A}) \models q$



- Σ-inseparability (both concept and query) for $DL\text{-Lite}_{bool}$ is Π_2^P -complete
- Σ-inseparability (both concept and query) for $DL\text{-Lite}_{horn}$ is coNP-complete

Off-the-shelf QBF solvers (skizzo, Quaffle, QuBE) show reasonable performance

Σ -inseparability and uniform interpolation

\mathcal{L} has uniform interpolation (logic) or admits forgetting (CS) if

$$\forall \mathcal{T} \forall \Sigma \exists \mathcal{T}'_{\Sigma} \left(\text{sig}(\mathcal{T}'_{\Sigma}) \subseteq \Sigma \wedge \mathcal{T} \text{ and } \mathcal{T}'_{\Sigma} \text{ are } \Sigma\text{-concept inseparable} \right)$$

- $DL\text{-Lite}_{bool}$ and $DL\text{-Lite}_{horn}$ have uniform interpolation

\mathcal{T} Σ -concept entails \mathcal{T}' iff $\mathcal{T} \models C_1 \sqsubseteq C_2$, for all $(C_1 \sqsubseteq C_2) \in \mathcal{T}'_{\Sigma}$
where \mathcal{T}'_{Σ} is a uniform interpolant of \mathcal{T}' w.r.t. Σ

$DL\text{-Lite}_{bool}^u$ extends $DL\text{-Lite}_{bool}$ with the **existential modality** $\exists u.C$ (or $C \neq \perp$)

- $DL\text{-Lite}_{bool}^u$ has uniform interpolation

\mathcal{T} Σ -query entails \mathcal{T}' in $DL\text{-Lite}_{bool}$ iff $\mathcal{T} \models C_1 \sqsubseteq C_2$, for all $(C_1 \sqsubseteq C_2) \in \mathcal{T}'_{\Sigma}$
where \mathcal{T}'_{Σ} is a uniform interpolant of \mathcal{T}' w.r.t. Σ in $DL\text{-Lite}_{bool}^u$

In the previous example: $\text{teaches} = 1 \text{ teaches} \neq \perp$



What is the size of uniform interpolants?

Delicate balance: either numbers restrictions or role inclusions

$DL\text{-Lite}_{core}$ with axioms of the form $B_1 \sqsubseteq B_2$ is **NLogSpace**-complete

$DL\text{-Lite}_{core}$ + role inclusions $R_1 \sqsubseteq R_2$ is **ExpTime**-complete

Example: $A_1 \sqcap A_2 \sqsubseteq C$ can be simulated by the axioms:

- $A_1 \sqsubseteq \exists R_1$ $A_2 \sqsubseteq \exists R_2$
- $R_1 \sqsubseteq R_{12}$ $R_2 \sqsubseteq R_{12} \quad \geq 2 \quad R_{12} \sqsubseteq \perp$
- $\exists R_1^- \sqsubseteq \exists R_3^-$ $\exists R_3 \sqsubseteq C$
- $R_3 \sqsubseteq R_{23}$ $R_2 \sqsubseteq R_{23} \quad \geq 2 \quad R_{23}^- \sqsubseteq \perp$

$DL\text{-Lite}$ + role inclusions - number restrictions is fine again

Étude II: Spatial Logic

Spatial logics: formal languages interpreted over various classes of geometrical structures (topological, metric, Euclidean spaces, etc.)

- **Mathematics:** Hilbert's geometry, Tarski's geometry, $\mathbf{Th}(\mathbb{R}^n, +, \times, \leq), \dots$
- **Philosophy:** Whitehead's & de Laguna's spatial ontologies (1920s)
region-based theories of space
- **Theoretical physics:** Logics of space-time (part of Hilbert's 6th problem)
e.g., Andréka&Németi (2007), Goldblatt (1980, 87), Shehtman&Shapirovsy (2003)
- **Computer science & AI:**
 - GIS, spatial databases and (first-order) query languages

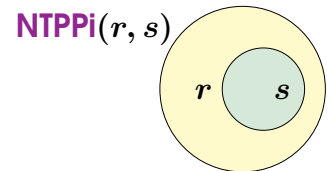
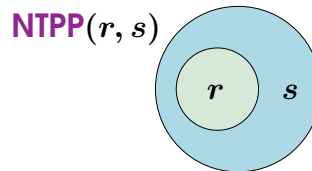
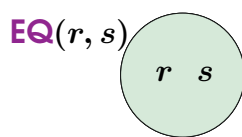
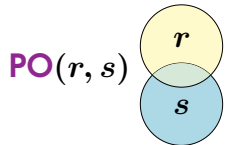
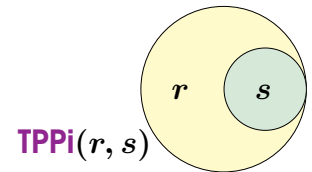
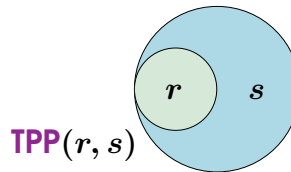
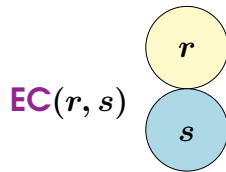
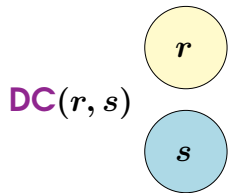
Complexity of $\mathbf{Th}(\mathbb{R}, +, \times, \leq)$ is still open: $\text{NEXPTIME} \leq ? \leq \text{EXPSpace}$

- qualitative spatial (region-based) KR&R; quantifier-free constraint systems and their algebraic counterparts

Intended models

Aim: (M. Egenhofer (GIS), A. Cohn, B. Nebel (KR&R), *et al.*, 1990s)

Effective reasoning about basic topological relations between spatial regions over Euclidean plane \mathbb{R}^2



NB: First-order theories are undecidable

Models: topological spaces, metric spaces, \mathbb{R}^2 , algebras, ...

Regions: arbitrary, connected, 'regular,' semialgebraic/linear sets, ...

Kripke frames = Aleksandrov spaces

A space is called **Aleksandrov** if arbitrary intersections of open sets are open

Aleksandrov spaces \equiv **Kripke frames** $\mathfrak{F} = (W, R)$, R is quasi-order on W ,
where the R -closed sets are open

Theorem. (McKinsey & Tarski, Shehtman, Statman, Areces *et. al.*, ...)

$\text{Sat}(\mathcal{S}4_u, \text{ALL}) = \text{Sat}(\mathcal{S}4_u, \text{ALEK}) = \text{Sat}(\mathcal{S}4_u, \text{FINALEK}) \neq \text{Sat}(\mathcal{S}4_u, \mathbb{R}^n)$, $n \geq 1$,
and these sets are **PSPACE**-complete



Reasoning too **complex**?



Strange sets are allowed as regions?



The language is **not natural**?

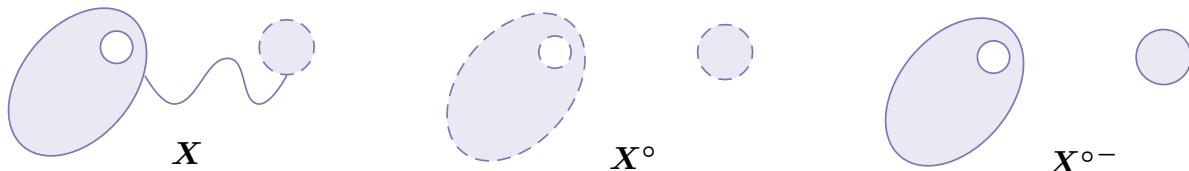


The language is too **weak**? ...

Regular closed sets and logic \mathcal{B} (mereology)

$X \subseteq T$ is **regular closed** if $X = X^{\circ-}$

$\mathbf{RC}(T)$ regular closed subsets of T



$(\mathbf{RC}(T), \cdot, -, \emptyset, T)$ is a Boolean algebra

where $X \cdot Y = (X \cap Y)^{\circ-}$ and $-X = (\overline{X})^-$

\mathcal{B} -terms: $\tau ::= r_i \mid -\tau \mid \tau_1 \cdot \tau_2$

regular closed sets!

\mathcal{B} -formulas: $\varphi ::= \tau_1 = \tau_2 \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2$

\mathcal{B} is a **fragment** of $\mathcal{S}4_u$

$\text{Sat}(\mathcal{B}, \text{REG}) = \text{Sat}(\mathcal{B}, \text{ALEKREG}) = \text{Sat}(\mathcal{B}, \text{RC}(\mathbb{R}^n))$, $n \geq 1$,

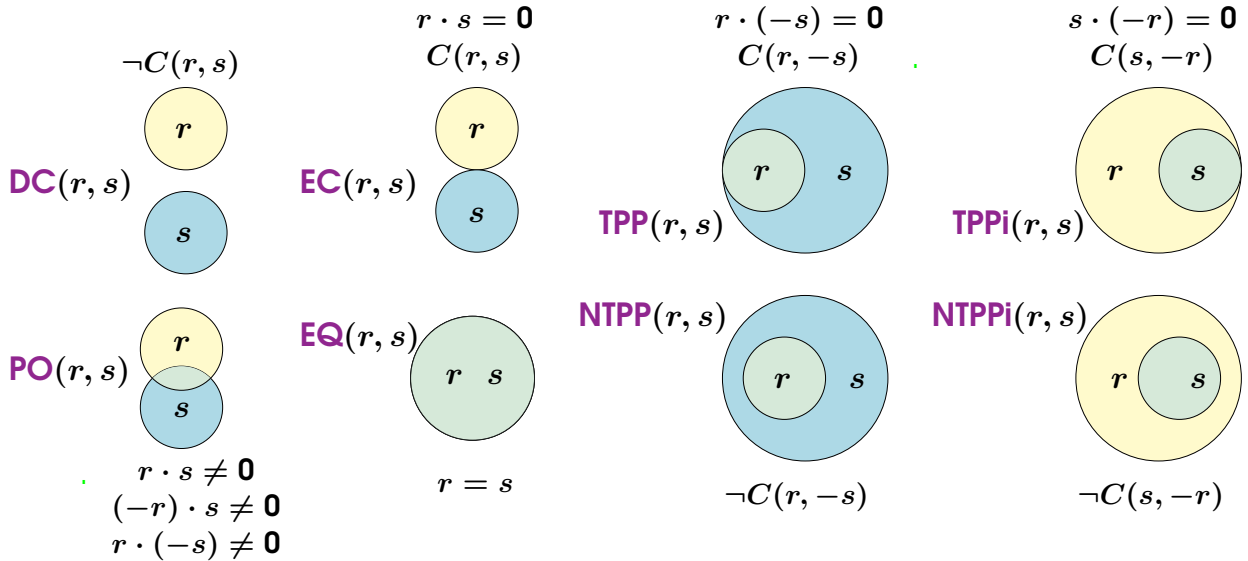
and this set is NP-complete

$\mathcal{C} = \mathcal{B} + \text{contact predicate (mereotopology)}$

↓ Whitehead's (1929) 'connection' relation

C-formulas: $\varphi ::= \tau_1 = \tau_2 \mid C(\tau_1, \tau_2) \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2$

a.k.a. **BRCC-8**, still a fragment of $\mathcal{S4}_u$



Sat(\mathcal{C}, REG) = **Sat**($\mathcal{C}, \text{ALEKREG}$) and this set is **NP**-complete; $\neq \text{Sat}(\mathcal{C}, \text{RC}(\mathbb{R}^n))$

Connectedness

A topological space is **connected** iff

it is not the union of two non-empty, disjoint, open sets

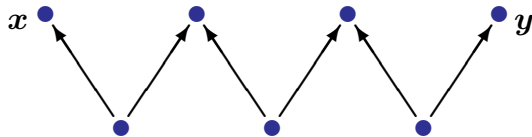
$X \subseteq T$ is **connected in T** if either it is empty,

or the topological space X (with the subspace topology) is connected

An **Aleksandrov space** induced by $\mathfrak{F} = (W, R)$ is **connected** iff

\mathfrak{F} is connected as a **non-directed graph**

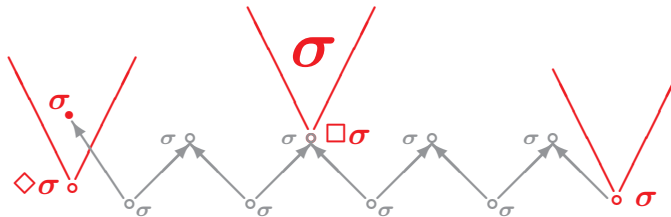
(any $x, y \in W$ are connected by a path along $R \cup R^{-1}$)



Logics with connectedness predicates

$\mathcal{S}_{4_u c^-}$, \mathcal{C}_{c^-} , \mathcal{B}_{c^-} -formulas:	$\varphi ::= \dots \mid c(\tau) \mid \dots$ τ is connected
$\mathcal{S}_{4_u c c^-}$, $\mathcal{C}_{c c^-}$, $\mathcal{B}_{c c^-}$ -formulas:	$\varphi ::= \dots \mid c^{\leq k}(\tau) \mid \dots$ τ has $\leq k$ components

How does the new 'modal operator' $c(\sigma)$ work?



- How long can such σ -paths be?
- What can be expressed by means of c ?
- What is the interaction between $c(\sigma_1), \dots, c(\sigma_n)$ and the Booleans?

$c(\sigma)$ can simulate PSpace Turing machines

Example: simulating a binary counter using predicates c and C

- Numbers: $\overbrace{(-r_n) \cdot \dots \cdot (-r_1)}^0, \overbrace{(-r_n) \cdot \dots \cdot (-r_2) \cdot r_1}^1, \dots, \overbrace{r_n \cdot \dots \cdot r_1}^{2^n - 1}$

- $c(1)$: the whole space is connected

- $-r_n \cdot \dots \cdot -r_1 \neq 0, r_n \cdot \dots \cdot r_1 \neq 0$: there exist 0 and $2^n - 1$, and so a **path connecting them**



- that all numbers occur on this path (not necessarily in proper order) is ensured by polynomially-many constraints $\neg C(l, m)$ for $m \notin \{l - 1, l, l + 1\}$

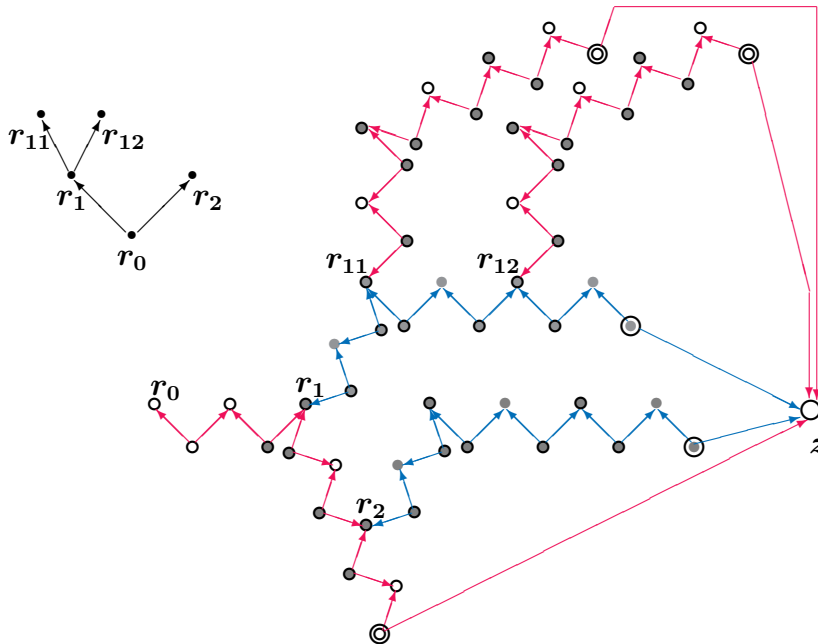
$$\neg C((-r_i) \cdot r_j \cdot (-r_k), r_i), \quad i > j > k; \quad \neg C(r_i \cdot r_{i-1}, (-r_i) \cdot r_{i-1}), \quad i > 1;$$

$$\neg C(r_i \cdot (-r_j) \cdot r_k, (-r_i)), \quad i > j > k; \quad \dots$$

$c(\sigma)$ can be exponentially long

Cc is PSPACE-hard
 Cc^1 is PSPACE-complete

$c(\sigma_1)$ and $c(\sigma_2)$ can simulate PSpace alternating TMs



- Two connectedness constraints $c(\sigma_0+z)$, $c(\sigma_1+z)$ and a number of $-C(\tau_1, \tau_2)$

\mathcal{C}_c is EXPTIME-hard

Contact $C(\tau_1, \tau_2)$ is expressible by means of $c(\sigma)$ 🤪

$$\mathcal{C}c \models c(\tau_1) \wedge c(\tau_2) \rightarrow [c(\tau_1 + \tau_2) \leftrightarrow C(\tau_1, \tau_2)]$$

Let φ be a $\mathcal{C}c$ -formula

- $\varphi[C(\tau_1, \tau_2)]^+$ (positive occurrence of $C(\tau_1, \tau_2)$)
eliminated using fresh variables t, t_1, t_2 :

$$\varphi^* = \varphi[t = \mathbf{0}]^+ \wedge ((t = \mathbf{0}) \rightarrow c(t_1 + t_2) \wedge \bigwedge_{i=1,2} (t_i \subseteq \tau_i) \wedge c(t_i))$$

- $\varphi[C(\tau_1, \tau_2)]^-$ (negative occurrence of $C(\tau_1, \tau_2)$)
eliminated using fresh variables s, t, t_1, t_2 :

$$\varphi^* = (\varphi[t \neq \mathbf{0}]^-)_{|s} \wedge (t \cdot s = \mathbf{0} \rightarrow \neg c(t_1 + t_2) \wedge \bigwedge_{i=1,2} c(t_i) \wedge (\tau_i \cdot s \subseteq t_i))$$

φ is satisfiable iff φ^* is satisfiable

$\mathcal{B}c$ is EXPTIME-hard

Complexity results

- $\mathcal{S4}_{uc}$ can be embedded into \mathcal{CPDL} with nominals, which is **EXPTIME**-complete [De Giacomo, 1995]

\mathcal{Bc} , \mathcal{Cc} , $\mathcal{S4}_{uc}$ are all **EXPTIME**-complete

- The component counting predicates $c^{\leq k}(\tau)$ increase complexity:

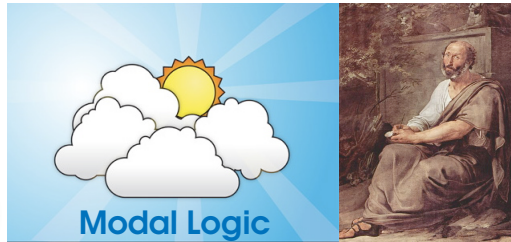
\mathcal{Bcc} , \mathcal{Ccc} , $\mathcal{S4}_{ucc}$ are all **NEXPTIME**-complete

Open problems

- 😊 Axiomatise the connectedness predicate.
- 😊 Interpretations over \mathbb{R}^2 with 'tame' regions.
- 😊 Query languages.
- 😊 ...



Modal Logic and Computer Science



Dynamic logics

Description logics

Spatial logics

Temporal logics



Computer Science