# A Well-founded Graph-based Summarization Framework for DLs

UMR IRISA

ENSSAT LANNION

## Motivations

***Graph summarization*** has received considerable attention in the literature for timely topics like:

- *exploration and visualization* of large graphs;
- *optimization of graph data management systems.*

➢ ABoxes can been seen as *multigraphs with typed vertices and typed edges* (Fig.1)

The ***quotient graph operation*** from *graph theory* is a well-established tool for graph summarization; it consists in fusing equivalent vertices according to an equivalence relation, e.g., bisimilarity.

➢ *Does it make sense* to summarize ABoxes with this tool as if they were merely graphs?

## Quotient operation for ABoxes

**Definition 1 (Quotient ABox).** *Let $\mathcal{A}$ be an ABox, $\equiv$ be some equivalence relation between constants, and let $a^1_\equiv, \dots, a^n_\equiv$ denote by a slight abuse of notation both the equivalence classes of the constants in $\mathcal{A}$ w.r.t. $\equiv$ and the names of these equivalence classes. The* quotient ABox of $\mathcal{A}$ w.r.t. $\equiv$ *is the ABox $\mathcal{A}_\equiv$ such that:*

- $\mathsf{C}(a^i_\equiv) \in \mathcal{A}_\equiv$ *iff there exists $a \in a^i_\equiv$ such that $\mathsf{C}(a) \in \mathcal{A}$, for $1 \leq i \leq n$,*
- $\mathsf{R}(a^i_\equiv, a^j_\equiv) \in \mathcal{A}_\equiv$ *iff there exist $a \in a^i_\equiv$ and $a' \in a^j_\equiv$ such that $\mathsf{R}(a, a') \in \mathcal{A}$, for $1 \leq i, j \leq n$.*

ABoxes *are not just graphs* because they also have a **first-order semantics**.

➢ Does there exist a *semantic relationship* between an ABox and a summary of it?

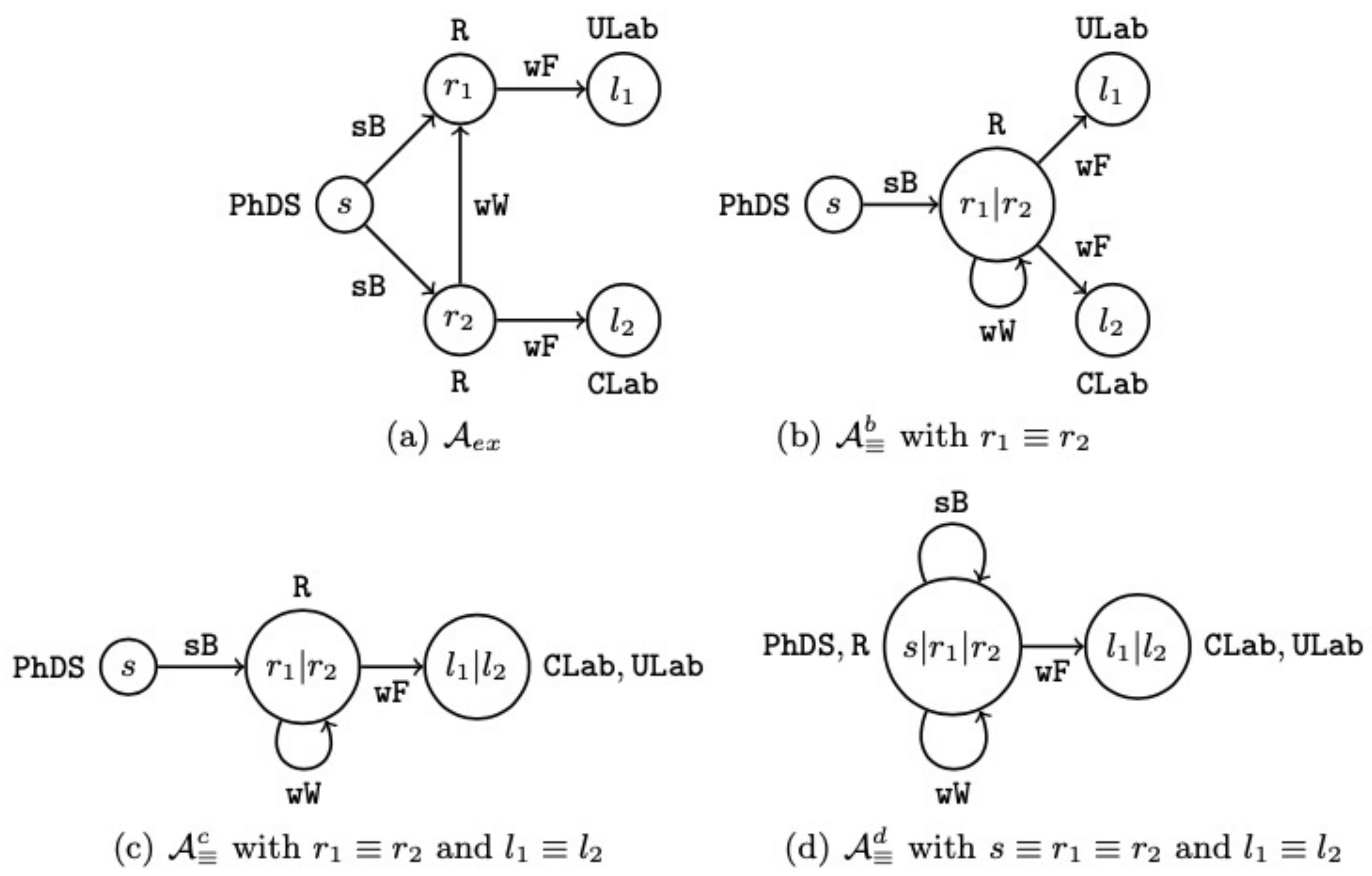


Fig. 1: The ABox $\mathcal{A}_{ex}$ and three quotient ABoxes of it: $\mathcal{A}^b_\equiv, \mathcal{A}^c_\equiv, \mathcal{A}^d_\equiv$.

## Characterization of ABox summaries

A summary is *more specific than* the summarized ABox

*Property 1.* Let $\mathcal{T}$ be a TBox, $a_1, \dots, a_m$ be the constants in an ABox $\mathcal{A}$ and $a^1_\equiv, \dots, a^n_\equiv$ be the constants in some summary $\mathcal{A}_\equiv$ of $\mathcal{A}$. If we consider these constants as existential variables, then $\exists a^1_\equiv \cdots \exists a^n_\equiv\ \mathcal{A}_\equiv \models_\mathcal{T} \exists a_1 \cdots \exists a_m\ \mathcal{A}$ holds.

➢ *To which extent* a summary is more specific than the ABox it summarizes?

For an ABox and a TBox, the most precise description of a set of constants is the *conjunction of concepts these constants are all instances of*, which is called their **most specific concept (msc)**.

**Theorem 1.** *Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be an ABox and $\mathcal{A}_\equiv$ be the summary of $\mathcal{A}$ w.r.t. the $\equiv$ equivalence relation. If $a_1, \dots, a_n$ are all the constants in $\mathcal{A}$ that belong to the equivalence class $a_\equiv$ according to $\equiv$, then the following holds:*

$$msc^{\mathcal{A}_\equiv,\mathcal{T}}(a_\equiv) \preceq_\mathcal{T} \prod_{i=1}^{n} msc^{\mathcal{A},\mathcal{T}}(a_i) \preceq_\mathcal{T} \coprod_{i=1}^{n} msc^{\mathcal{A},\mathcal{T}}(a_i) \preceq_\mathcal{T} msc^{\mathcal{A},\mathcal{T}}(a_1, \dots, a_n).$$

## Application to ontology-based data management

### Consistency-checking

*Property 2.* Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be an ABox and $\mathcal{A}_\equiv$ be some summary of $\mathcal{A}$. If $\mathcal{A}_\equiv$ is consistent w.r.t. $\mathcal{T}$, then $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$.

### Query answering

*Property 3.* Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be an ABox, $\mathcal{A}_\equiv$ be some summary of $\mathcal{A}$, and let $q$ be a UCQ. If $q^{\mathcal{A},\mathcal{T}} \neq \emptyset$ holds then $q_h^{\mathcal{A}_\equiv,\mathcal{T}} \neq \emptyset$ holds, with $q_h$ the query $q$ in which every constant $a$ is replaced by its image $h(a)$ through the $\mathcal{A}$-to-$\mathcal{A}_\equiv$ homomorphism $h$.

Property 2 and *the contraposition of* Property 3 are of practical interest to check rapidly if, *for sure*, an ABox is consistent and a query has no answer resp., by using the typically (much) smaller ABox summary.

## Perspectives

➢ Devise *DL-specific equivalence relations between constants* (i.e., which use both the TBox and the ABox) to **explore, visualize or optimize reasoning on and the management of ABoxes**.

Cheikh-Brahim EL VAIGH
Univ Rennes, France
cheikh-brahim.el-vaigh@irisa.fr

François GOASDOUÉ
Univ Rennes, France
fg@irisa.fr

### Bibliography

➢ Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook. Cambridge University Press (2003)
➢ Bienvenu, M.: Ontology-mediated query answering: Harnessing knowledge to get more from data. In: IJCAI (2016)
➢ Cebiric, S., Goasdoué, F., Kondylakis, H., Kotzinos, D., Manolescu, I., Troullinou, G., Zneika, M.: Summarizing semantic graphs: a survey. VLDB J. 28(3) (2019)
➢ El Husseini, W., El Vaigh, C.-B., Goasdoué, F., Jaudoin, H.: Ontology-mediated query answering: performance challenges and advances. In: ISWC (2021)

UNIVERSITÉ DE RENNES 1