Neuro-Symbolic AI for modelling Bio-Knowledge Graphs with contradictions

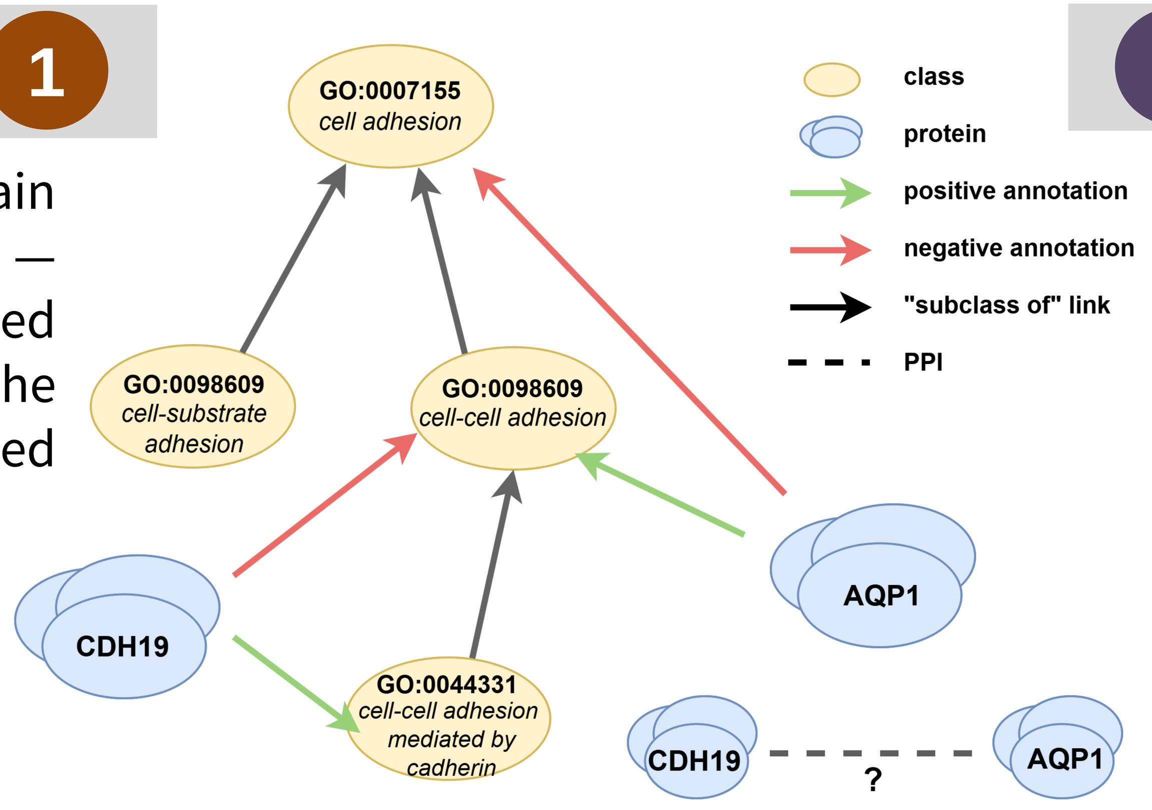
Laura Balbi, Catia Pesquita

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

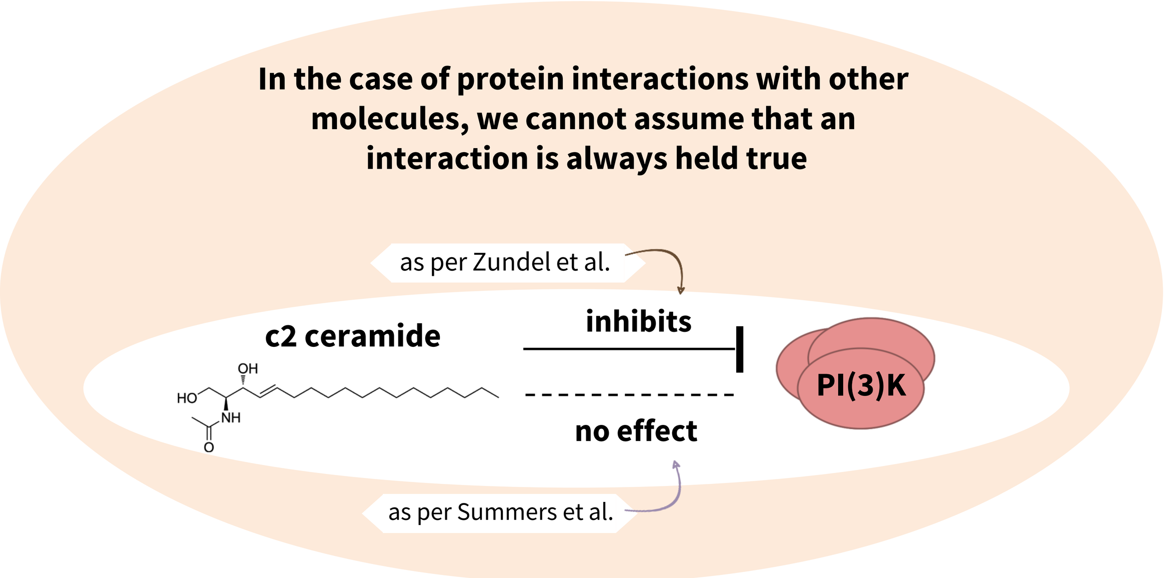
LASIGE health and biomedical informatics

1 Motivation

Knowledge Graphs (KGs) often contain contradictions that are not necessarily errors — they can represent polyvocal truths or nuanced realities that, if properly modeled, can enhance the completeness and expressiveness of KG-based machine learning.



In biomedical KGs this may be due to the universe of experimental conditions in which biological entities perform specific roles, functions and interactions.



Current KG Representation Learning is not well-equipped to handle contradictory facts within KGs, especially those that are implicit and context-dependent. There is a lack of mechanisms to explicitly model or leverage contradictions during learning.

Taxonomy of Contradictions

Explicit

Direct, logic-based conflicts that can be detected via rule violations or ontology constraints.

Implicit

Semantically conflicting or context-sensitive facts that do not violate formal logic alone, but appear contradictory under additional assumptions or external knowledge.

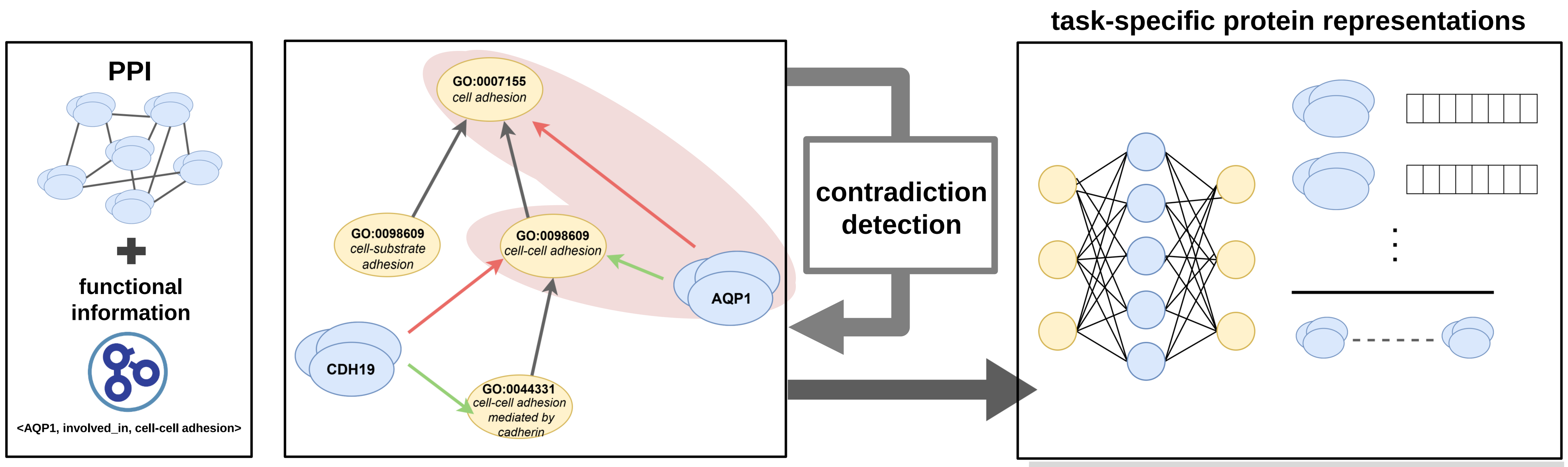
	Subtype	Description	Example
Direct	**Negated Assertion**	Presence of both an assertion and its negation.	<ProteinA, interactsWith, ProteinB> AND <ProteinA, NOT interactsWith, ProteinB>
Direct	**Disjoint Classes**	Assigns an entity to classes that are explicitly declared disjoint.	<TS T53, typeOf, Protein> AND <TS T53, typeOf, Disease> AND Protein ⊥ Disease
	Inverse Property Conflict	Conflicting values in mutually inverse properties.	<GeneA, encodes, ProteinA> AND <ProteinA, encodedBy, GeneB>
Indirect	**Cardinality/ Domain/Range Violation**	Violation of property constraints.	<ProteinA, hasCatalyticFunction, Kinase> AND <ProteinA, hasCatalyticFunction, Protease> (if hasCatalyticFunction is 1:1)
Indirect	**Hierarchically Inferred**	Conflict from inherited logical constraints.	<ProteinA, hasFunction, DNA repair> AND <ProteinA, hasFunction, Negative regulation of DNA repair> AND DNA repair ⊥ Negative regulation of DNA repair

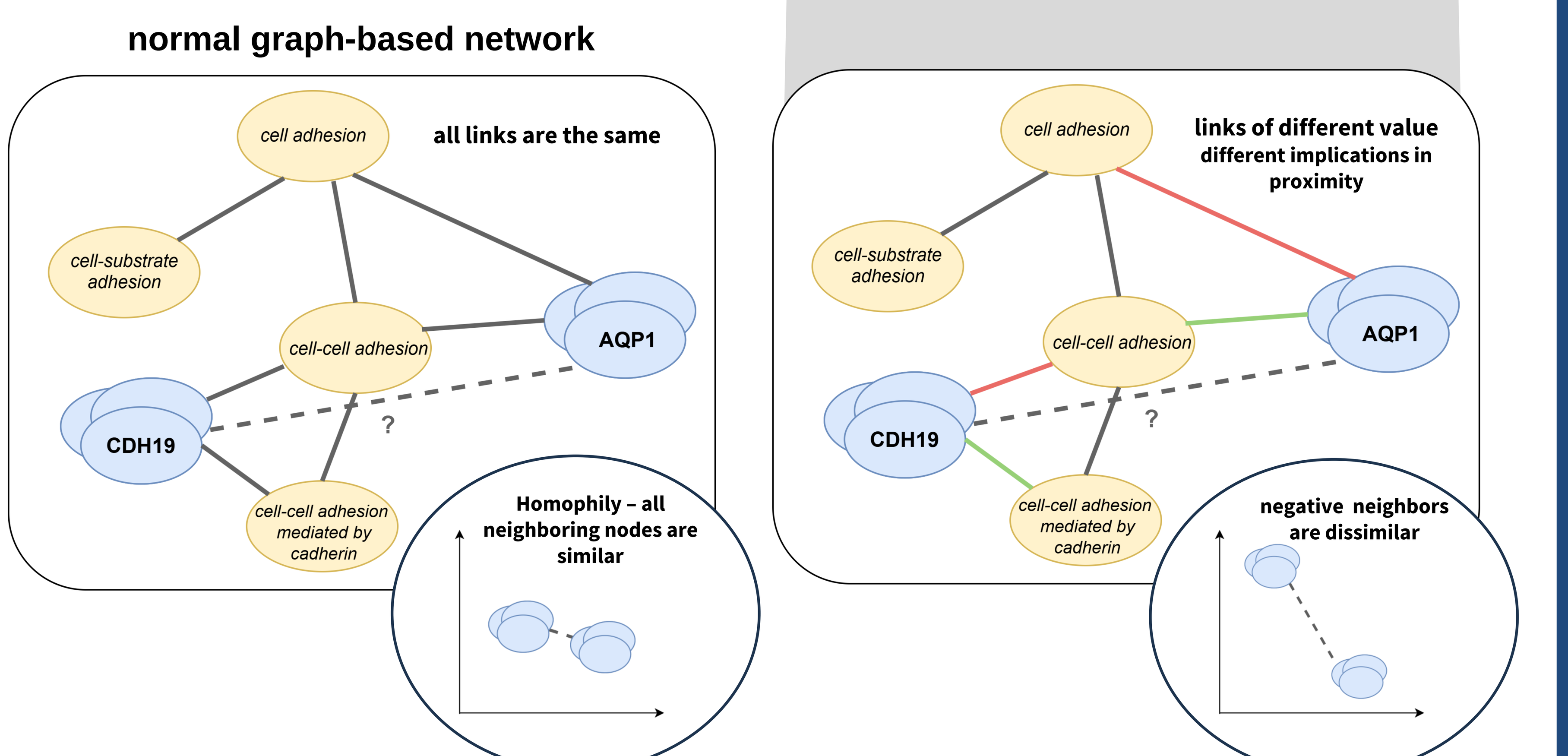
2 Hypothesis

An approach that integrates symbolic and sub-symbolic representations can bridge KGs and LLMs to answer:

- **RQ1:** How do contradictions in a KG, implicit and explicit, impact the SOTA KGRL performance for PPIs?
- **RQ2:** Does modelling contradictions onto protein representations improve ML?
- **RQ3:** Can external sources of knowledge such as LLMs be explored to detect implicit contradictions?

3 Methodology





4 Baselines

ML	KG	# Edge Types	Precision	Recall	ROC-AUC
HGCN	Full	1	0.9440	0.3193	**0.9023**
		2	**0.9574**	0.2316	0.9002
		5	0.9402	**0.3536**	0.8916
	no LC	1	0.9444	0.3153	**0.8998**
		2	**0.9546**	0.2675	0.8976
		5	0.9441	**0.3844**	0.8888
	Negative Annots	1	0.9346	0.3369	**0.8954**
		1	0.9391	0.2869	0.8909
		4	**0.9398**	**0.3640**	0.8904
	Positive Annots	1	**0.9477**	0.3045	0.9007
		2	0.9475	0.3259	**0.9010**
		4	0.9429	**0.3687**	0.8912

ML	KG	# Edge Types	Precision	Recall	ROC-AUC
HGAT	Full	1	0.9375	**0.1546**	0.6913
		2	**0.9452**	0.1438	0.6642
		5	0.9386	0.1414	**0.7030**
	no LC	1	0.9310	**0.1491**	0.6778
		2	0.8853	0.1140	0.6255
		5	**0.9391**	0.1415	**0.6985**
	Negative Annots	1	0.9269	**0.2175**	0.6777
		1	0.9105	0.1945	0.6990
		4	**0.9393**	0.1639	**0.7021**
	Positive Annots	1	0.9487	0.1419	**0.6888**
		2	0.9470	**0.1504**	0.6636
		4	**0.9599**	0.1057	0.6775

Preliminary discussion

value of including negative knowledge
only negative annotations outperform

advantage in actively modelling opposing relations and contradictions in separate
consensual increase in performance from the GCN to the HGCN

GAT weighs edges according to their importance across edge types
GAT overperformed Bi- and HGAT

Universidade de Lisboa · Ciências ULisboa · fct Fundação para a Ciência e a Tecnologia · LASIGE

Acknowledgements: This work was supported by FCT through the fellowship 2024.01208.BD, and the LASIGE Research Unit, ref. UID/00408/2025. It was also partially supported by the KATY project which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017453.