

Learning from True Negatives in Scientific Open-World Graphs

Laura Balbi, Cátia Pesquita

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

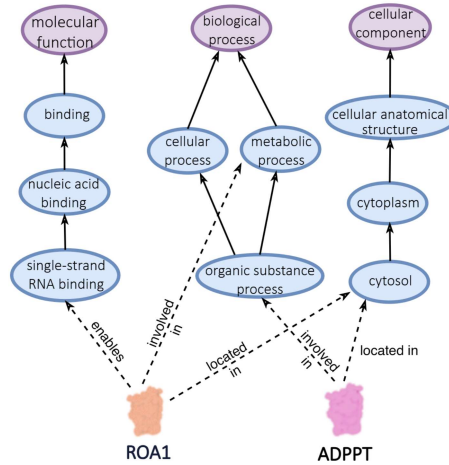
Scan here!



Background

Knowledge Graphs (KGs) capture relational facts under an **open-world assumption**, where missing statements reflect incompleteness rather than falsehood. Still, KGs are dominated by positives and often omit **negative knowledge** (verified false statements).

KG representation learning often uses training negative generation strategies that assume a **closed world** – absence equals falsehood.



Example biomedical KG with protein annotations to Gene Ontology classes.

Motivation

Negation can be a meaningful supervision signal: to learn what an entity is and what it is not.

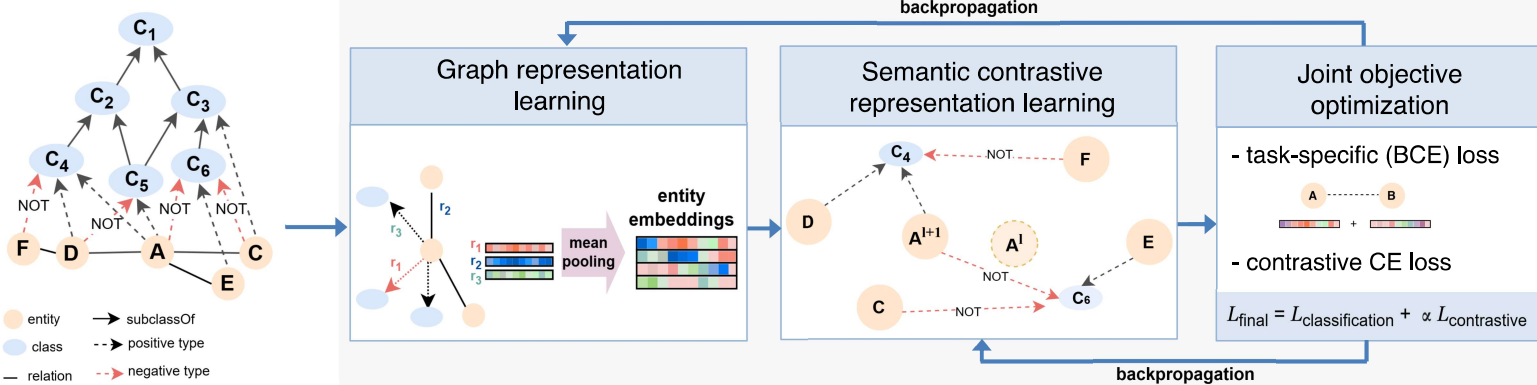
Sampling negatives from missing statements can create false negatives, semantic noise during training – and weaken the notion of negation.

Objective

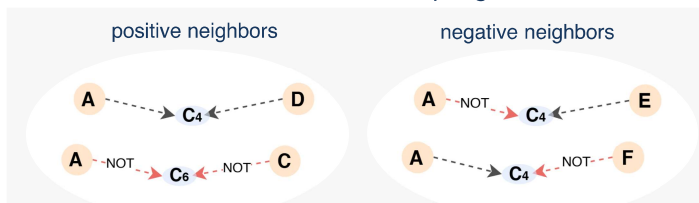
Investigate the impact of including and explicitly learning from verified negative evidence on KGRL.

Method

NegK-GNN learns entity representations by explicitly using **verified negative statements** as training negatives and supervision signals, through a **joint objective** that combines a task-specific loss with a **semantic contrastive loss**.

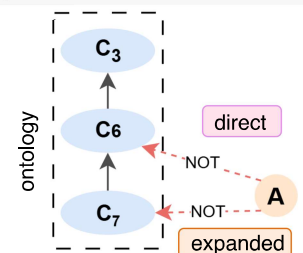


Contrastive sampling



Ontology-guided expansion

- negative statement expansion
- subclass reasoning – negation propagates downward.



Evaluation

Benchmark on Human protein–protein interaction and Gene Ontology functional annotations

Instances	12,880
Classification Triples	107,649
Positive Statements	74,901
Negative Statements	79,324

Task: Triple classification – to predict whether a candidate triple is true/false.

Results

Table 1. Mean results (%) for 10 runs.

Data	Model	Accuracy	F1	Prec (+)	Rec (+)	Roc Auc
Human	GCN	47.14	51.80	49.81	40.69	69.32
	Refined-GAE	48.76	47.84	48.37	35.84	47.29
	Signed GNN	58.90	55.26	66.43	43.38	70.11
	RGCN	56.38	52.82	63.72	29.14	65.50
	NegK-GCN	61.77	61.06	66.34	48.54	67.75
	NegK-RGCN	59.52	56.75	67.95	34.98	68.01

Acknowledgements: This work was partially supported by the CancerScan project through grant agreement No. 101186829, through the KATY project through grant agreement No. 101017453, and was supported by FCT through research fellowship 2024.01208.BD, and the LASIGE Research Unit, UID/00408/2025, DOI <https://doi.org/10.54499/UID/00408/2025>.