Complex Query Answering with Neural Link Predictors

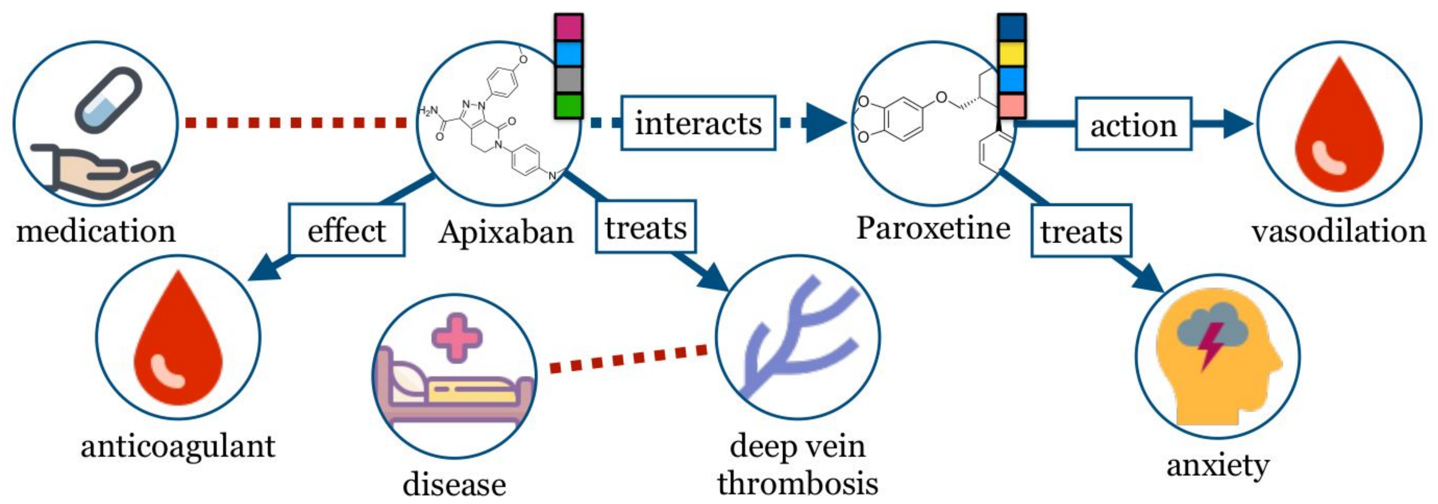
Erik Arakelyan* erik.arakelyan.18@ucl.ac.uk @_kire_kara_

Daniel Daza* d.dazacruz@vu.nl @danieldazac

Pasquale Minervini* p.minervini@ucl.ac.uk @PMinervini

Michael Cochez m.cochez@vu.nl @michaelcochez

Complex Queries over Incomplete Knowledge Graphs



Complex queries require reasoning over high order structures in a knowledge graph.
Example: *Which medications have side-effects when taken with drugs for treating anxiety?*

$$?M : \exists D \,.\, \text{interacts}\,(M, D) \wedge \text{treats}\,(D, \text{anxiety})$$

Previous works solve this by training black-box neural models on millions of generated queries.

Continuous Query Decomposition

We decompose queries into atom predicates connected by t-norms and t-conorms.

- We cast query answering as an optimisation problem. We consider **combinatorial** and **continuous** optimisation.
- Each atom is scored with a pre-trained neural link predictor.

$$?M : \exists D \,.\, \text{interacts}\,(M, D) \wedge \text{treats}\,(D, \text{anxiety})$$

Optimisation problem

Continuous relaxation of the logical AND (t-norm)

$$\arg\max_{M,D \in \mathcal{E}} \left[\phi_{\text{interacts}}\left(\mathbf{e}_M, \mathbf{e}_D\right) \top \phi_{\text{treats}}\left(\mathbf{e}_D, \mathbf{e}_{\text{anxiety}}\right) \right]$$

Likelihood that M interacts with D

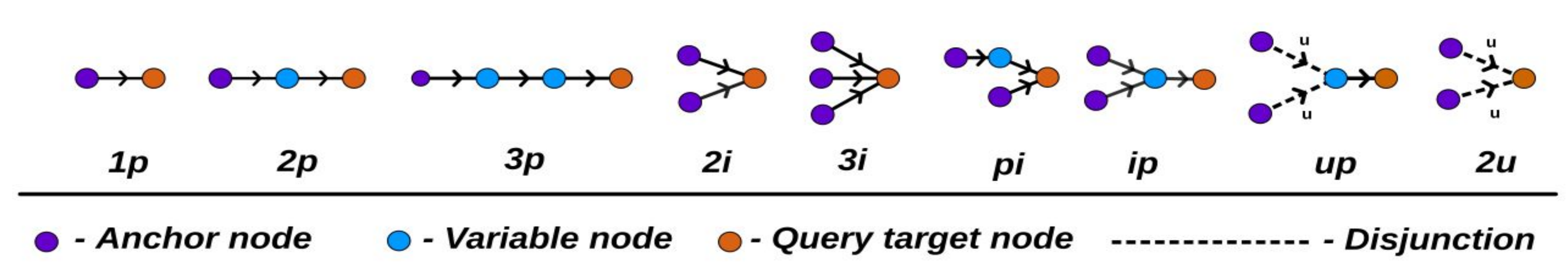
Likelihood that D treats anxiety

Optimisation in CQD

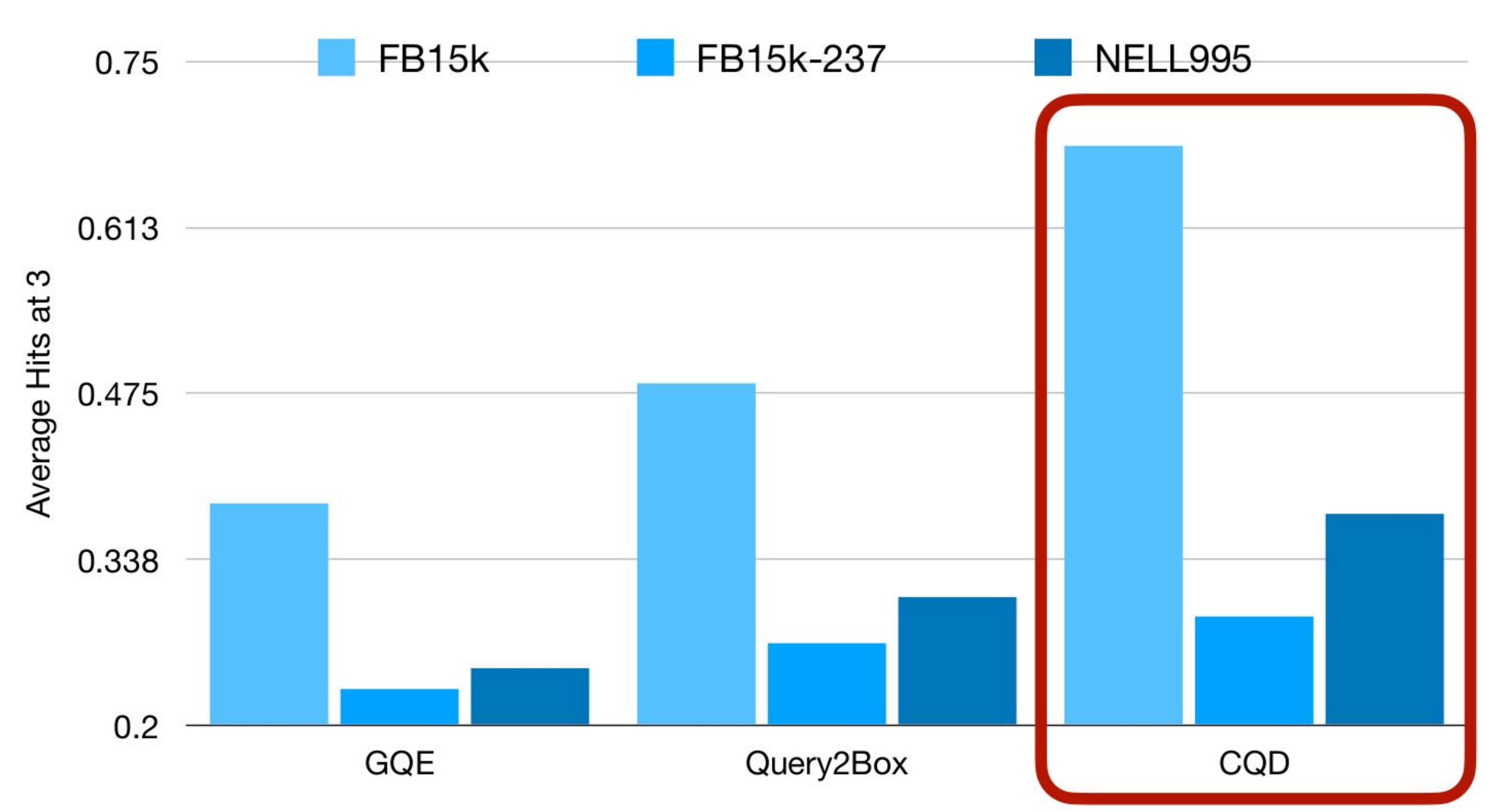
- **Greedy search:** Find the grounding of variables that maximises the score, keeping the top-k highest scoring candidates.
- **Continuous optimisation:** Initialize embeddings for variables randomly and optimise with gradient ascent.

Evaluation

We evaluate on many different query structures containing conjunctions and disjunctions.



- CQD produces more accurate results than methods explicitly trained for complex query answering, reducing training costs by orders of magnitude
- CQD generalizes to OOD instances, as it is trained for link prediction



Explaining Generated Answers

CQD allows to examine intermediate variable assignments:

$$?M : \exists D \,.\, \text{interacts}\,(M, D) \wedge \text{treats}\,(D, \text{anxiety})$$

M	*D*
Apixaban	Paroxetine
Amitriptyline	Paroxetine
Phenytoin	Paroxetine
Duloxetine	Pregabalin
Buprenorphine	Pregabalin

Summary

- CQD is an efficient method for answering complex queries that only requires training for link prediction
- Explainability in CQD allows to inspect the reasons a prediction is being made, and to identify failure modes.
- The framework can be extended with further fuzzy operators.

Source code and datasets:
https://github.com/uclnlp/cqd

References

[1] Hamilton et al. (2018). "Embedding Logical Queries on Knowledge Graphs." In: NeurIPS 2018.

[2] Ren et al. (2020). "Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings." In: ICLR 2020.

[3] Ren et al. (2020). "Beta Embeddings for Multi-Hop Logical Reasoning in Knowledge Graphs." In: NeurIPS 2020.