Answering Complex Queries on Very Large and Incomplete Knowledge Graphs

Pasquale Minervini
p.minervini@ucl.ac.uk

Gillmore Symposium, November 2021

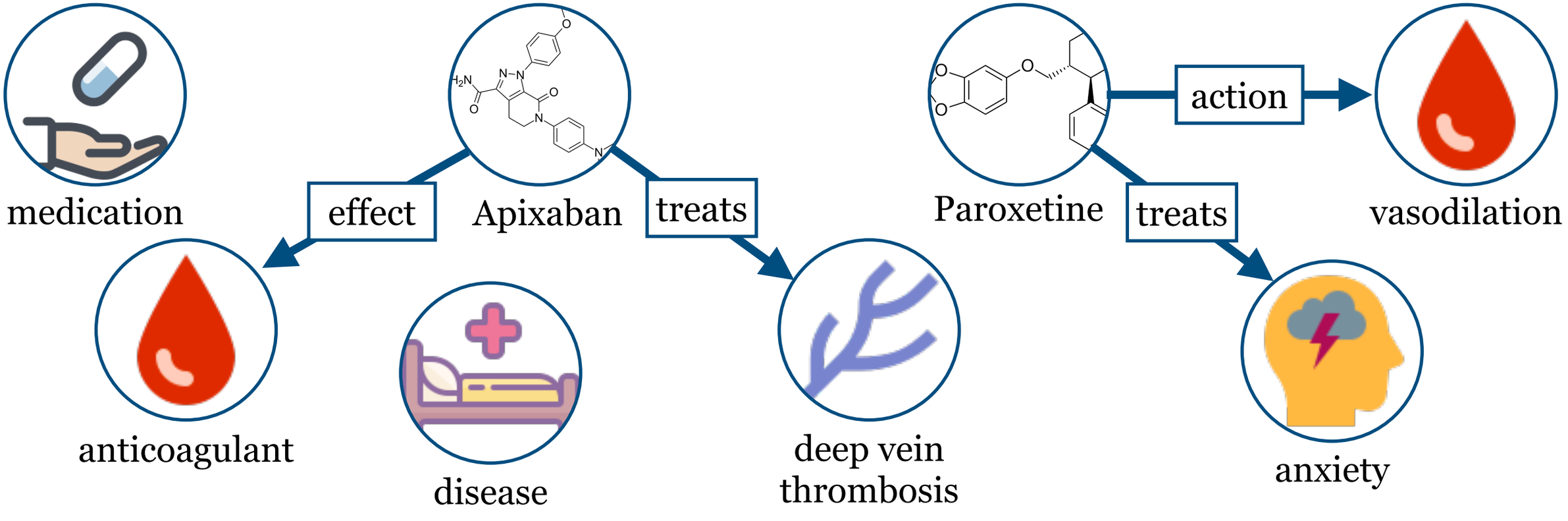
Complex Query Answering with Neural Link Predictors

Erik Arakelyan* Daniel Daza* **Pasquale Minervini*** Michael Cochez

Outstanding Paper Award 🏆 @ ICLR 2021

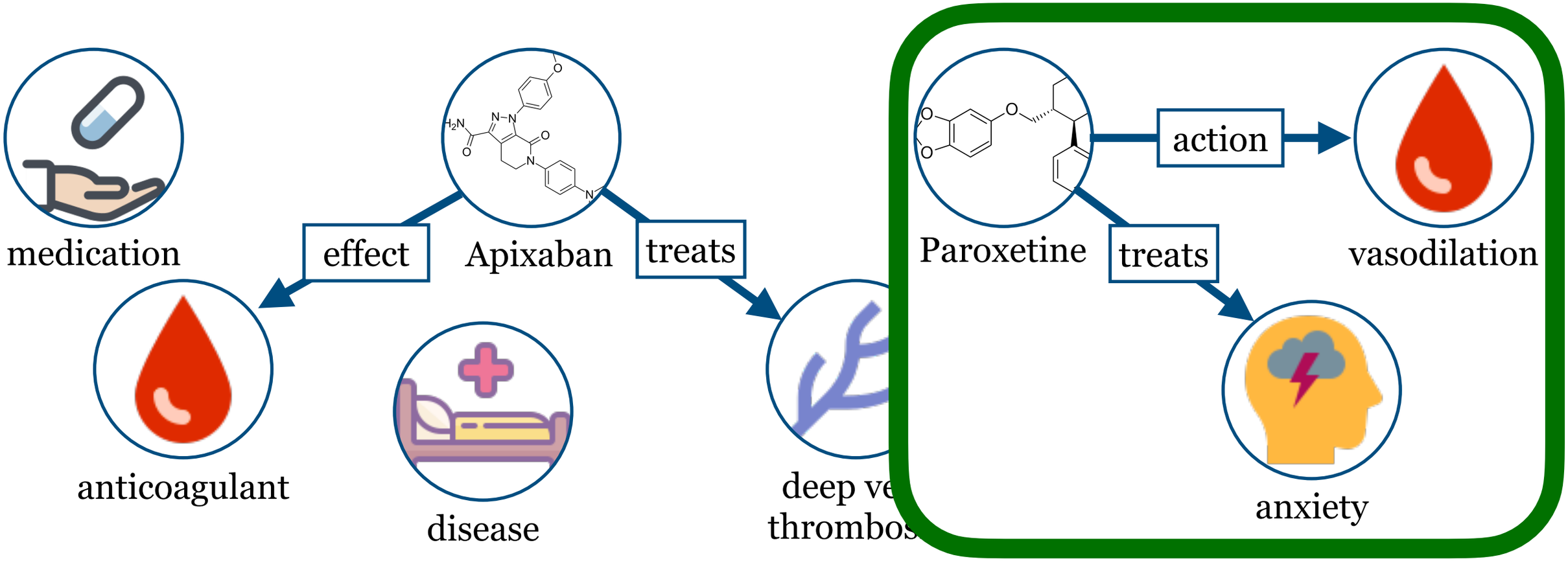
Knowledge Graphs

Knowledge Graph — graph-structured Knowledge Base, where knowledge about the world is encoded in the form of *relationships between entities*



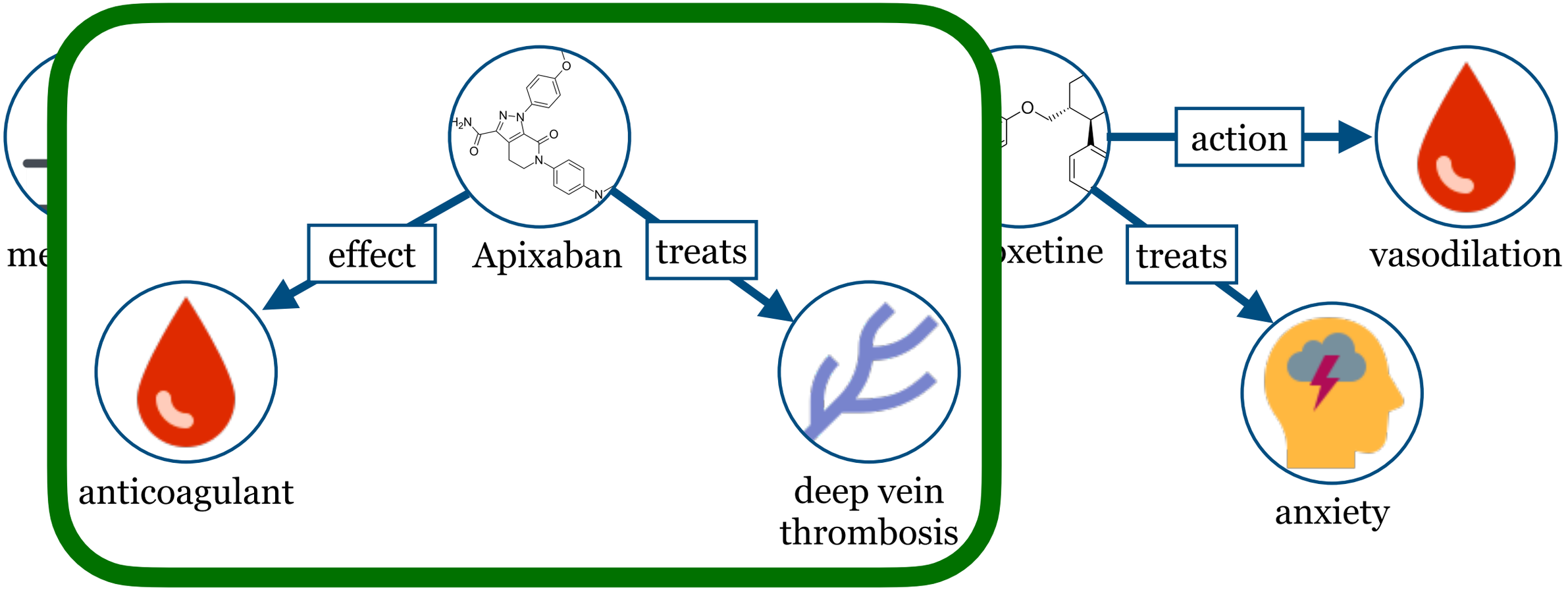
Knowledge Graphs

Knowledge Graph — graph-structured Knowledge Base, where knowledge about the world is encoded in the form of *relationships between entities*



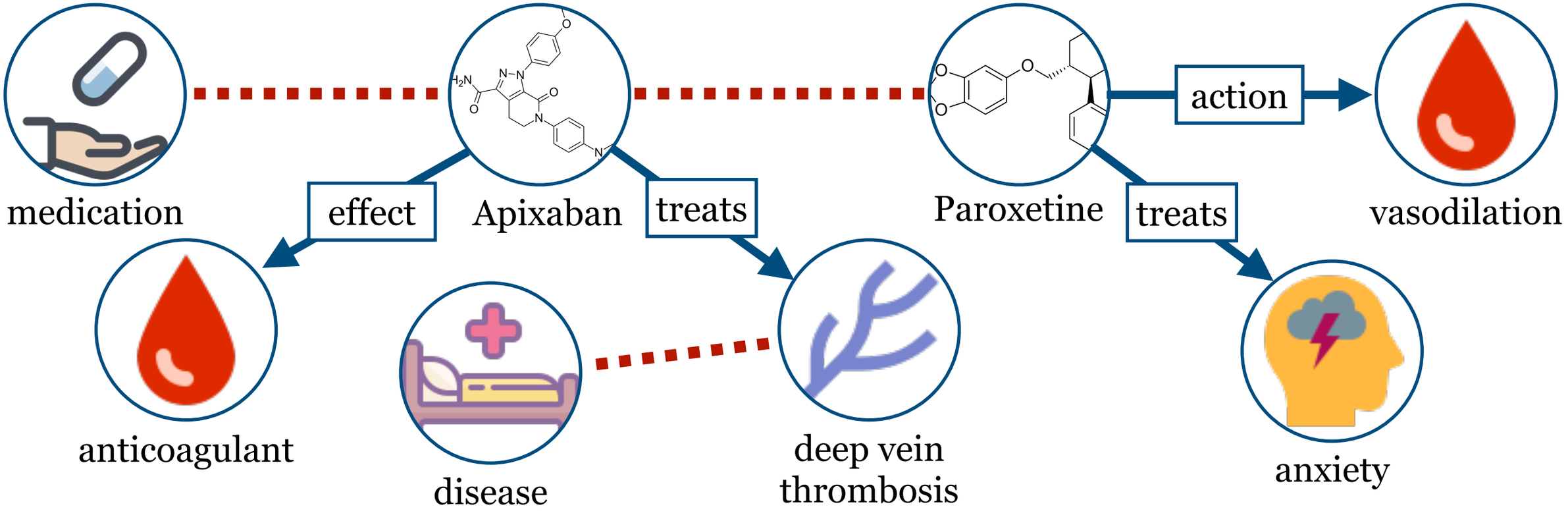
Knowledge Graphs

Knowledge Graph — graph-structured Knowledge Base, where knowledge about the world is encoded in the form of *relationships between entities*

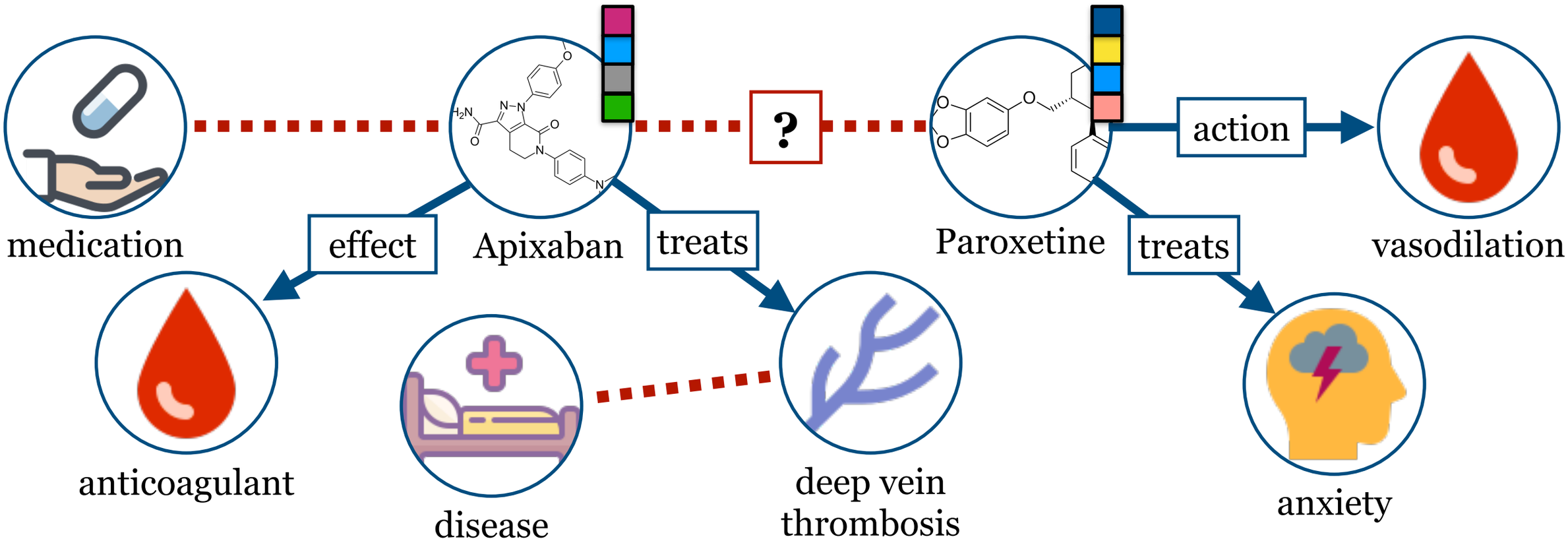


Knowledge Graphs

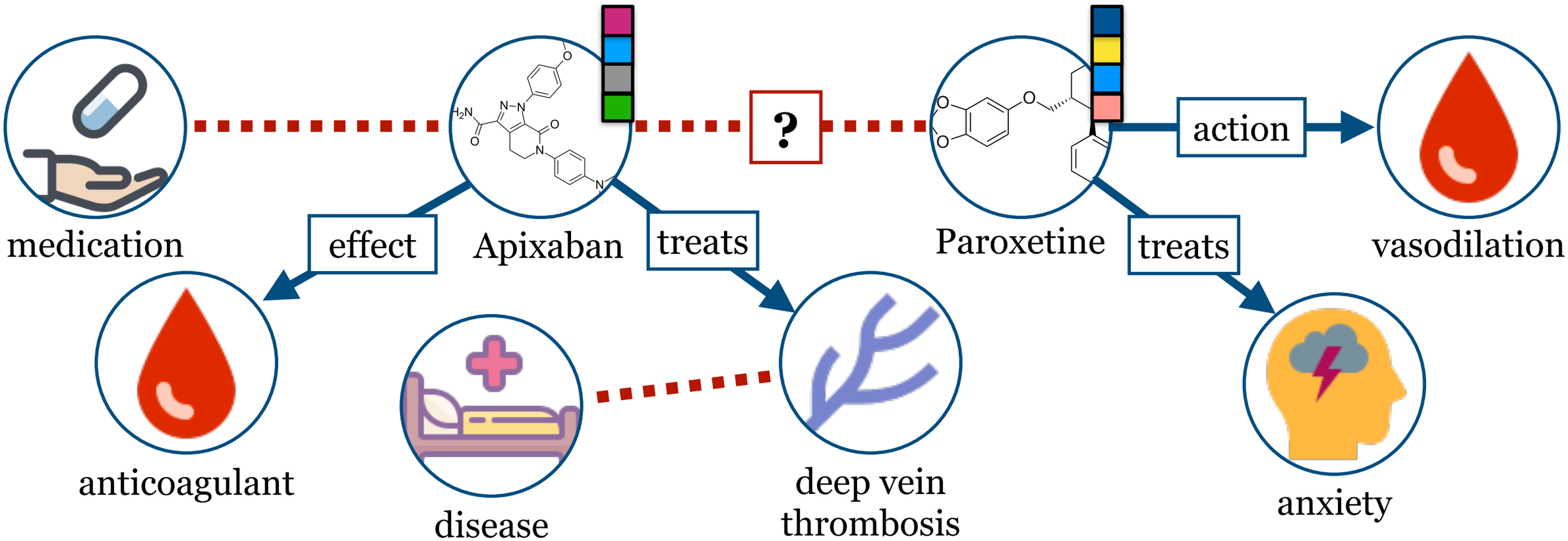
Knowledge Graph — graph-structured Knowledge Base, where knowledge about the world is encoded in the form of *relationships between entities*



Neural Link Prediction

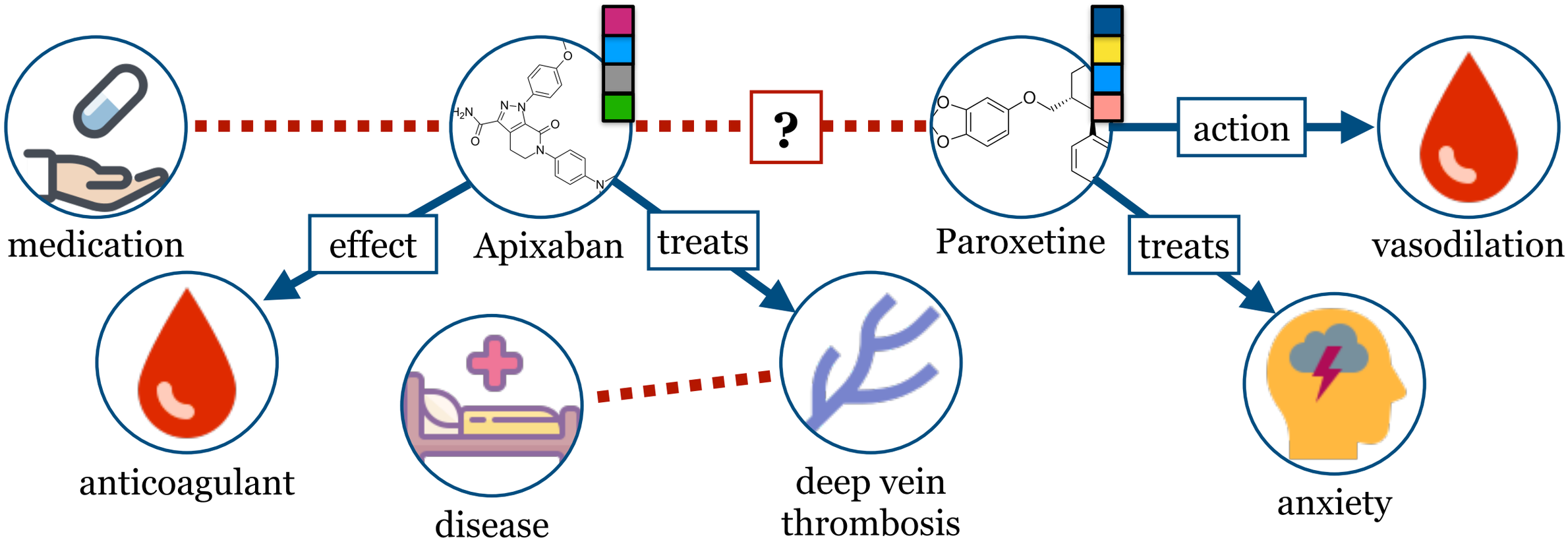


Neural Link Prediction



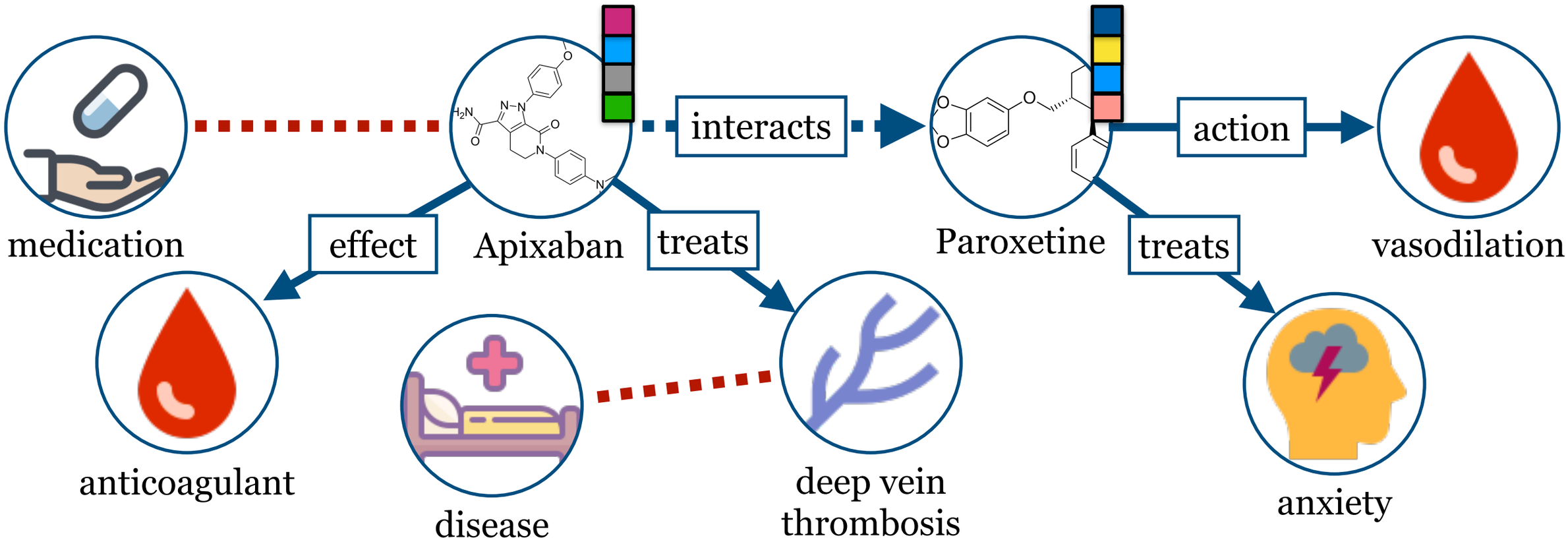
$$P(\text{Apixaban} \xrightarrow{\text{interacts}} \text{Paroxetine}) \propto$$

Neural Link Prediction



$$P(\text{Apixaban} \xrightarrow{\text{interacts}} \text{Paroxetine}) \propto \phi_{\text{interacts}}\left(\text{Apixaban}, \text{Paroxetine} \right)$$

Neural Link Prediction



$$P(\text{Apixaban} \xrightarrow{\text{interacts}} \text{Paroxetine}) \propto \phi_{\text{interacts}}\left(\text{Apixaban}, \text{Paroxetine} \right)$$

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

$$?M$$

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

$$?M : \exists D$$

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]

Process

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]

Process

1. Generate *millions* of complex query-answer pairs

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]

Process

1. Generate *millions* of complex query-answer pairs
2. Train a deep neural model to answer complex queries

Complex Queries on Incomplete Graphs

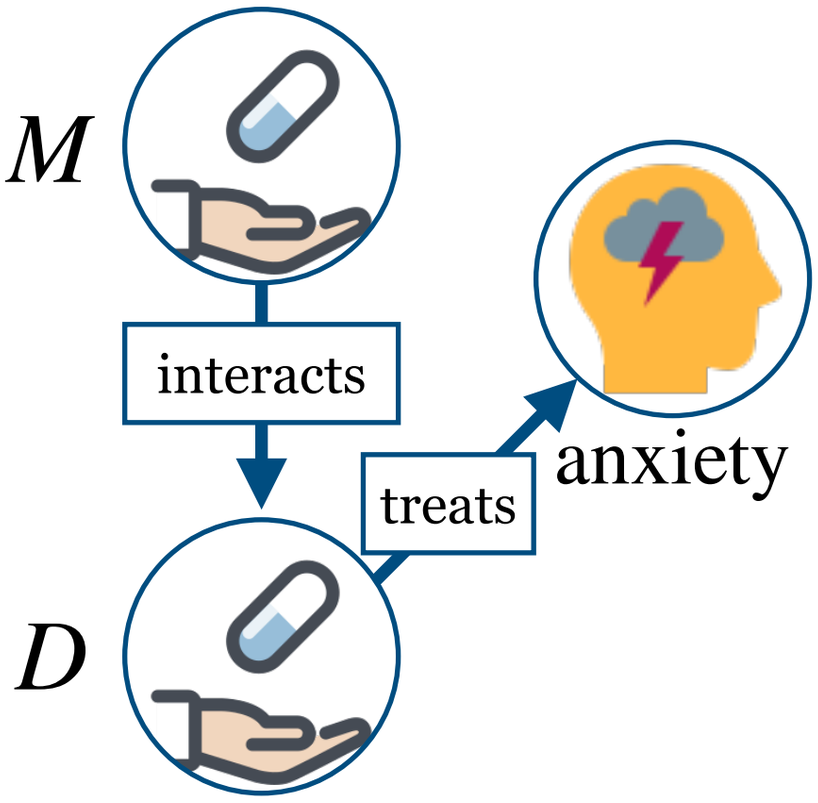
Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]



Complex Queries on Incomplete Graphs

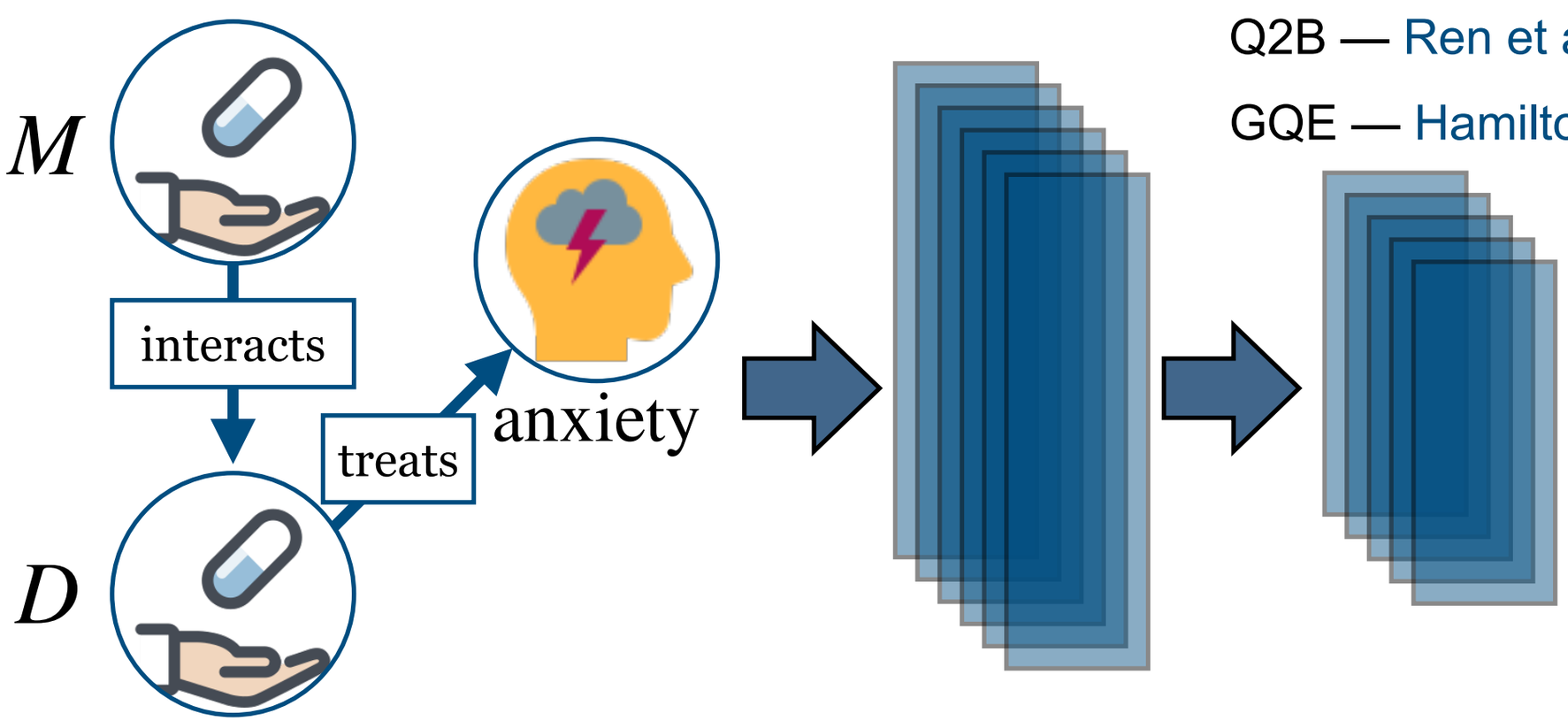
Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]



Complex Queries on Incomplete Graphs

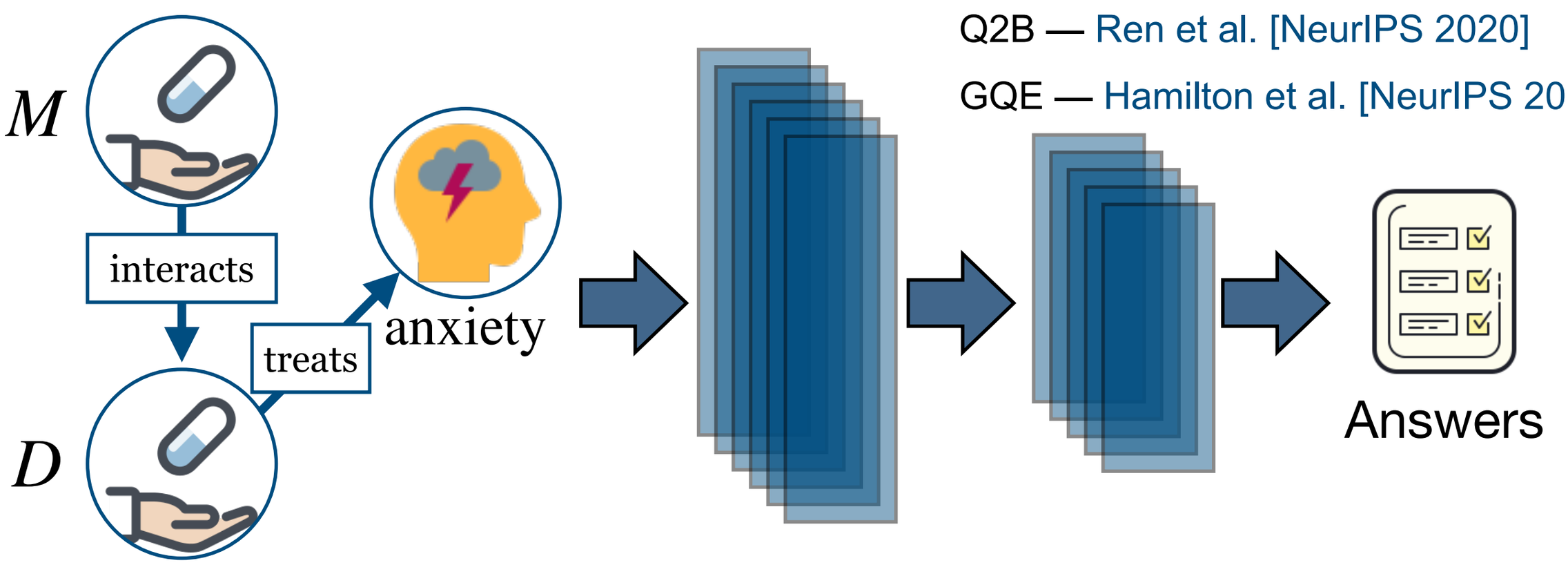
Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

GQE — Hamilton et al. [NeurIPS 2018]



Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

Problems

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

Problems

- Need to train the model on *millions* of generated queries

Complex Queries on Incomplete Graphs

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

BetaE — Ren et al. [ICLR 2020]

Q2B — Ren et al. [NeurIPS 2020]

Problems

- Need to train the model on *millions* of generated queries
- No explanation on the reasons *why* a given answer was produced by the model

Query Answering as Optimisation

Proposed solution: train a neural model ϕ for answering atomic (simple) queries (e.g. “which drugs treat Anxiety?”), and cast the query answering task as an *optimisation problem*

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

Query Answering as Optimisation

Proposed solution: train a neural model ϕ for answering atomic (simple) queries (e.g. “which drugs treat Anxiety?”), and cast the query answering task as an *optimisation problem*

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

Optimisation problem

$$\arg\max_{M,D \in \mathcal{E}}$$

Query Answering as Optimisation

Proposed solution: train a neural model ϕ for answering atomic (simple) queries (e.g. “which drugs treat Anxiety?”), and cast the query answering task as an *optimisation problem*

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

Optimisation problem

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

Likelihood that M interacts with D

Likelihood that D treats anxiety

Query Answering as Optimisation

Proposed solution: train a neural model ϕ for answering atomic (simple) queries (e.g. “which drugs treat Anxiety?”), and cast the query answering task as an *optimisation problem*

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

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

Query Answering as Optimisation

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

Query Answering as Optimisation

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

Greedy Search

Query Answering as Optimisation

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

Greedy Search

- Identify the k most likely values for D

Query Answering as Optimisation

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

Greedy Search

- Identify the k most likely values for D
- For each value of D:

Query Answering as Optimisation

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

Greedy Search

- Identify the k most likely values for D
- For each value of D:
  - Identify the k most likely values for M

Query Answering as Optimisation

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

Greedy Search

- Identify the k most likely values for D
- For each value of D:
  - Identify the k most likely values for M
- Compute the query score for all (M, D) combinations

Query Answering as Optimisation

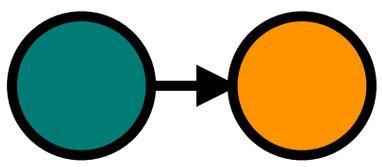
$$\arg\max_{M,D \in \mathcal{E}} \left[\phi_{\text{interacts}}\left(\mathbf{e}_M, \mathbf{e}_D\right) \top \phi_{\text{treats}}\left(\mathbf{e}_D, \mathbf{e}_{\text{anxiety}}\right) \right]$$

Greedy Search

- Identify the k most likely values for D
- For each value of D:
  - Identify the k most likely values for M
- Compute the query score for all (M, D) combinations
- Return the most likely value for (M, D)

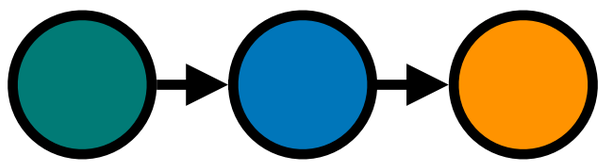
Complex Query Types 1/2

1p (atomic)



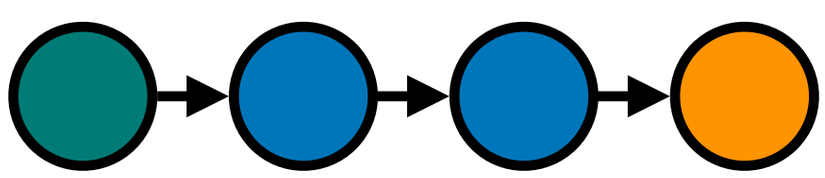
$$?T : p(a, T)$$

2p



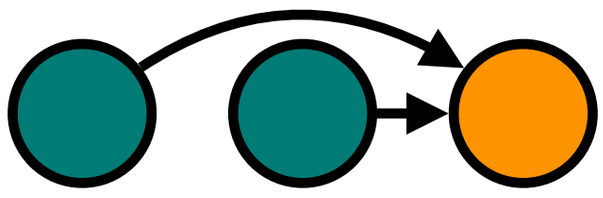
$$?T : \exists V_1 \,.\, p_1(a, V_1) \wedge p_2(V_1, T)$$

3p



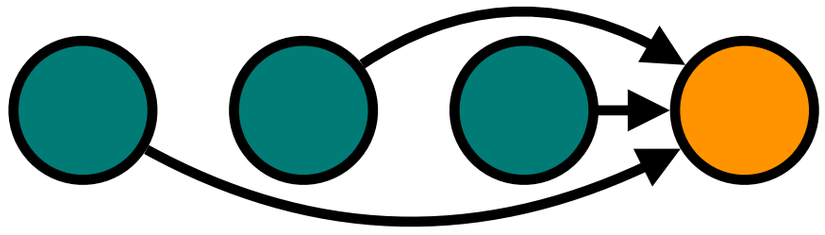
$$?T : \exists V_1, V_2 \,.\, p_1(a, V_1) \wedge p_2(V_1, V_2) \wedge p_3(V_2, T)$$

2i

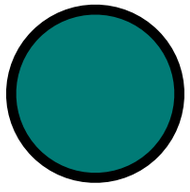


$$?T : p_1(a_1, T) \wedge p_2(a_2, T)$$

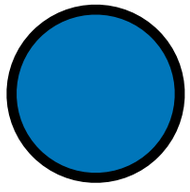
3i



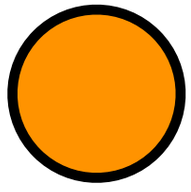
$$?T : p_1(a_1, T) \wedge p_2(a_2, T) \wedge p_3(a_3, T)$$



Anchor



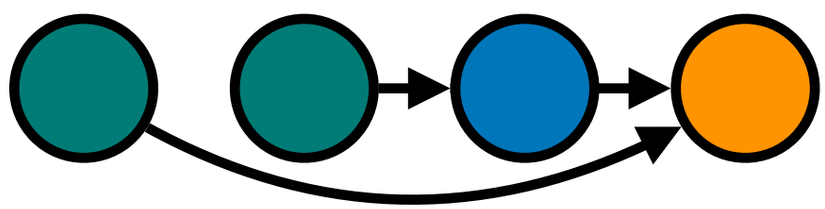
Variable



Target

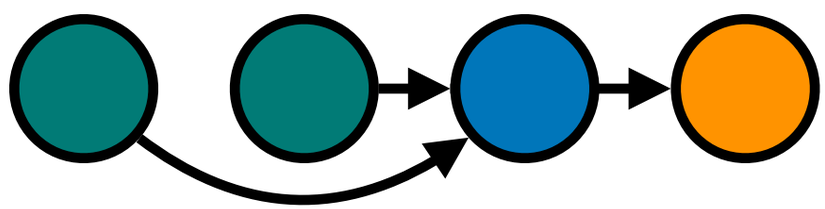
Complex Query Types 2/2

pi



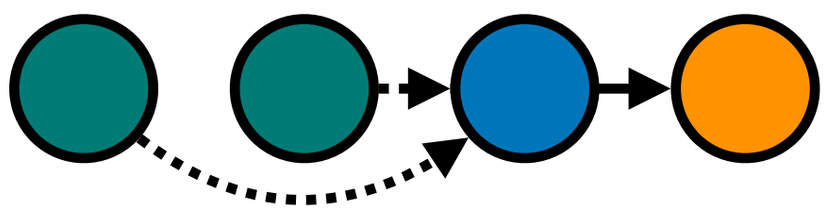
$$?T : \exists V \,.\, p_1(a_1, V) \wedge p_2(V, T) \wedge p_3(a_2, T)$$

ip



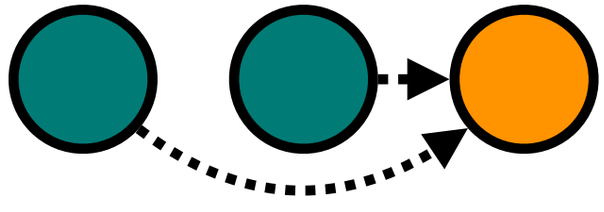
$$?T : \exists V \,.\, p_1(a_1, V) \wedge p_2(a_2, V) \wedge p_3(V, T)$$

up

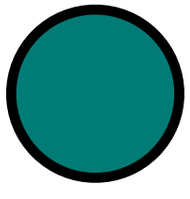


$$?T : \exists V \,.\, \left[p_1(a_1, V) \vee p_2(a_2, V)\right] \wedge p_3(V, T)$$

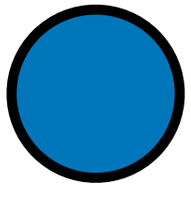
2u



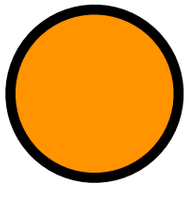
$$?T : p_1(a_1, T) \vee p_2(a_2, T)$$



Anchor



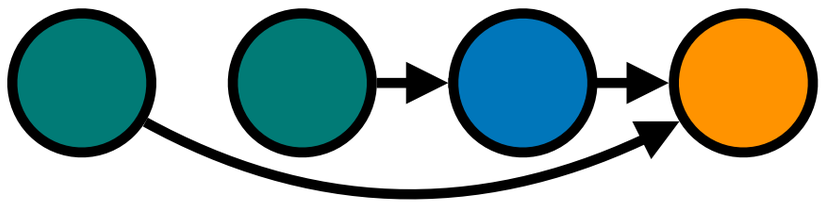
Variable



Target

Complex Queries — Examples

pi



$$?T : \exists V \,.\, p_1(a_1, V) \wedge p_2(V, T) \wedge p_3(a_2, T)$$

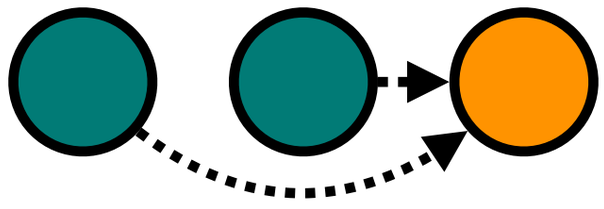
Query: Which medications used for treating fibromyalgia interact with drugs for treating Diabetes?

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

Anchor Variable Target

Complex Queries — Examples

2u

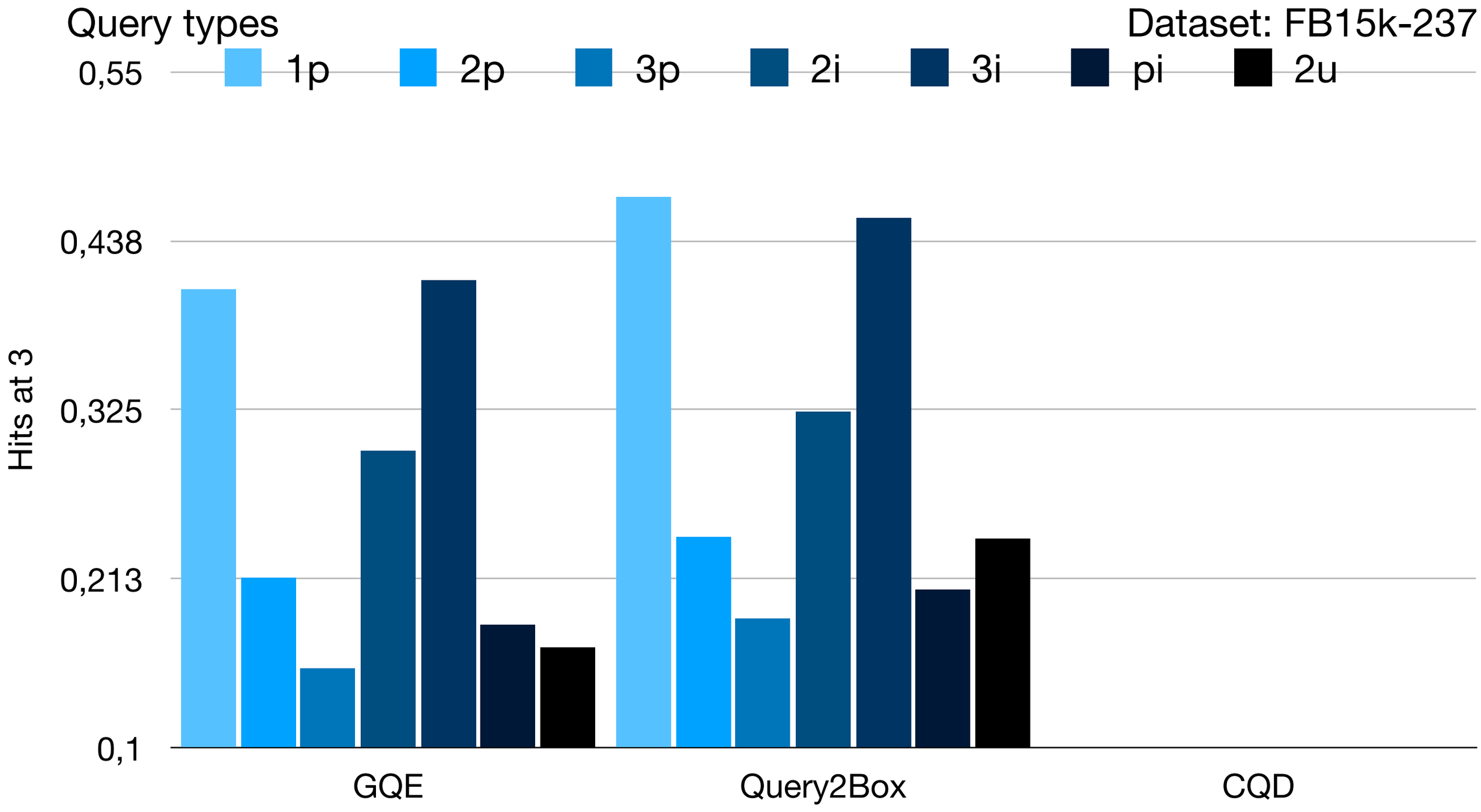


$$?T : p_1(a_1, T) \vee p_2(a_2, T)$$

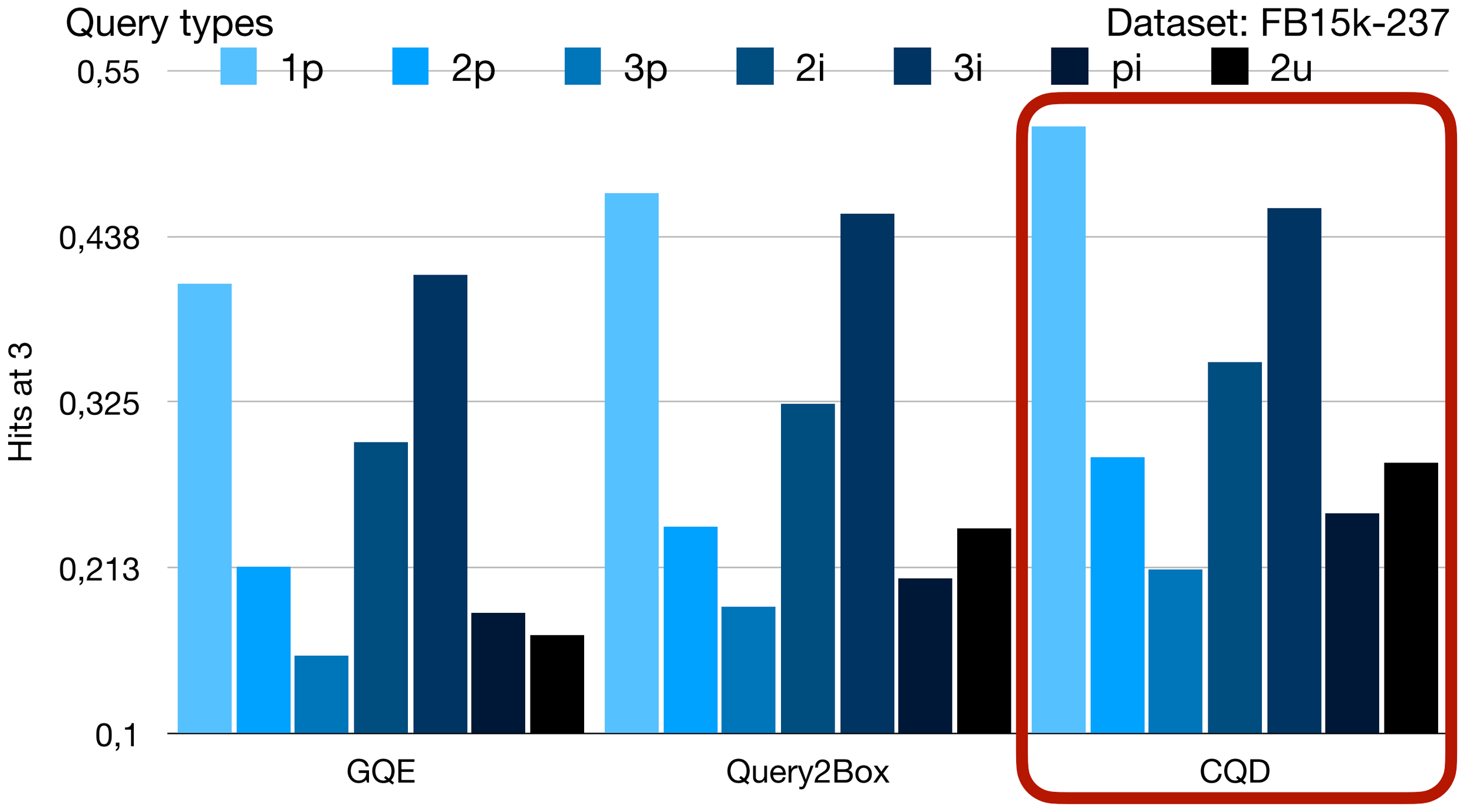
Query: Which actors played in King's Speech or in Dunkirk?

$$?A : \text{playsIn}\left(A, \text{king's speech}\right) \vee \text{playsIn}\left(A, \text{dunkirk}\right)$$

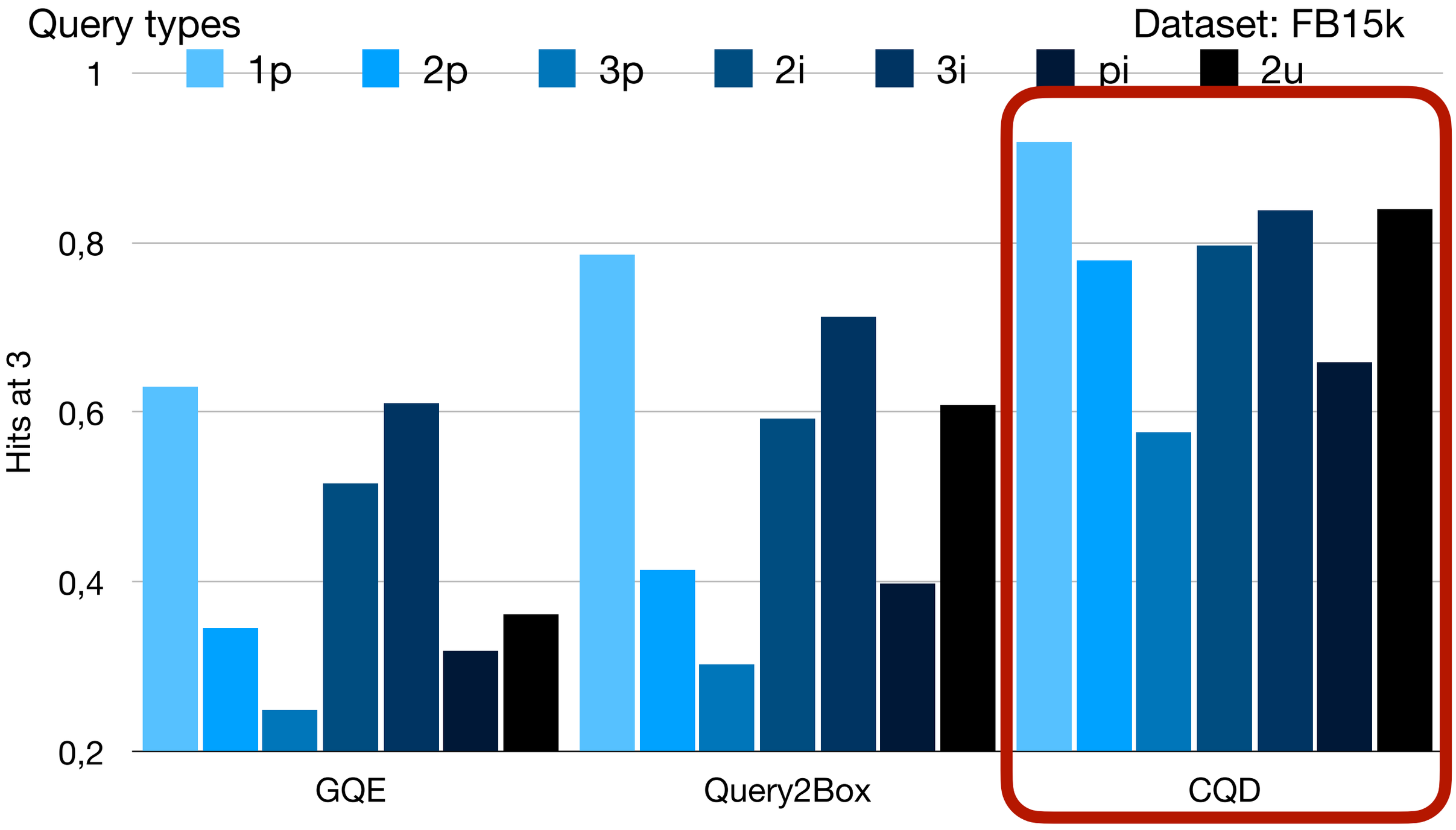
Experiments



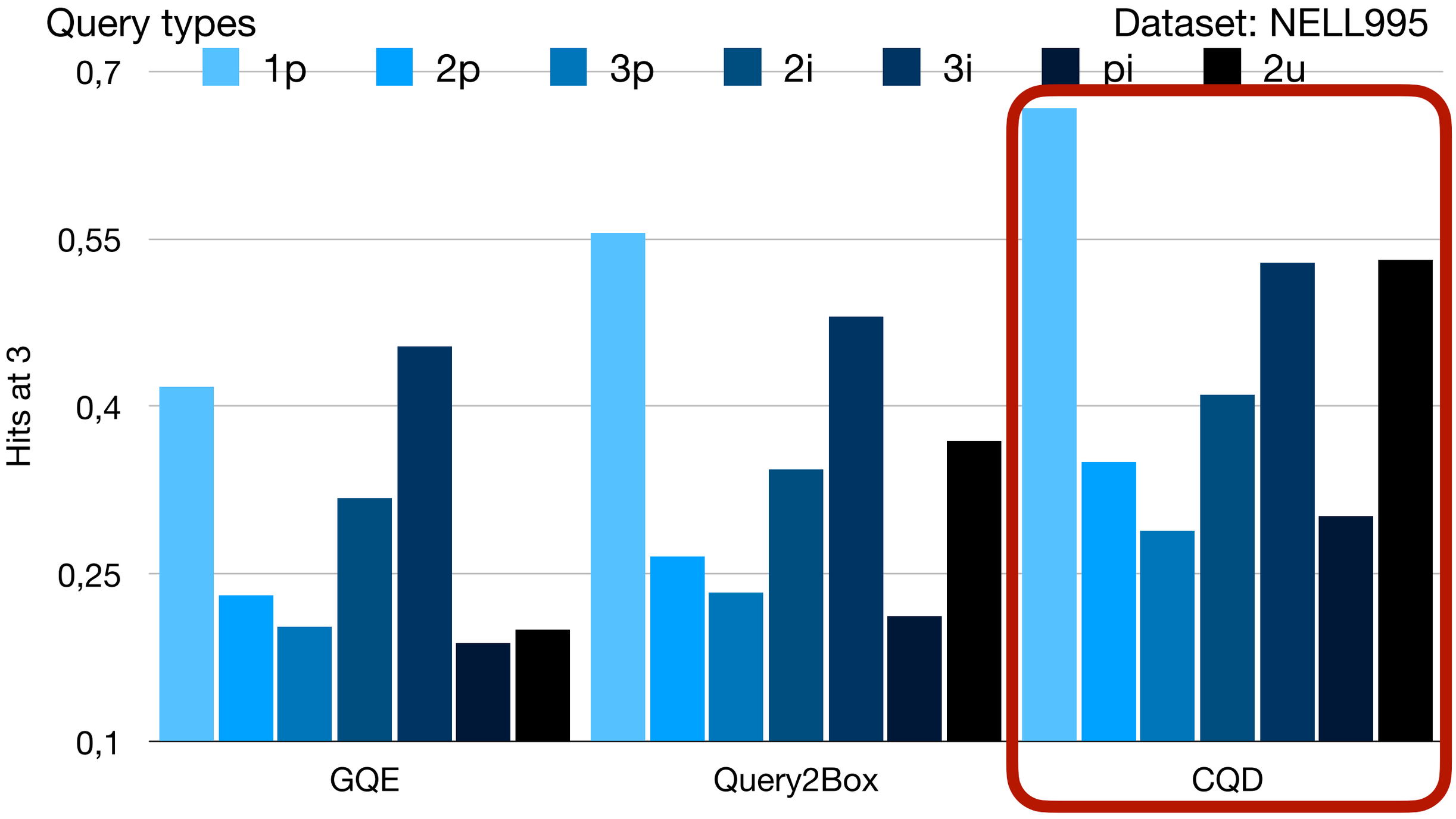
Experiments



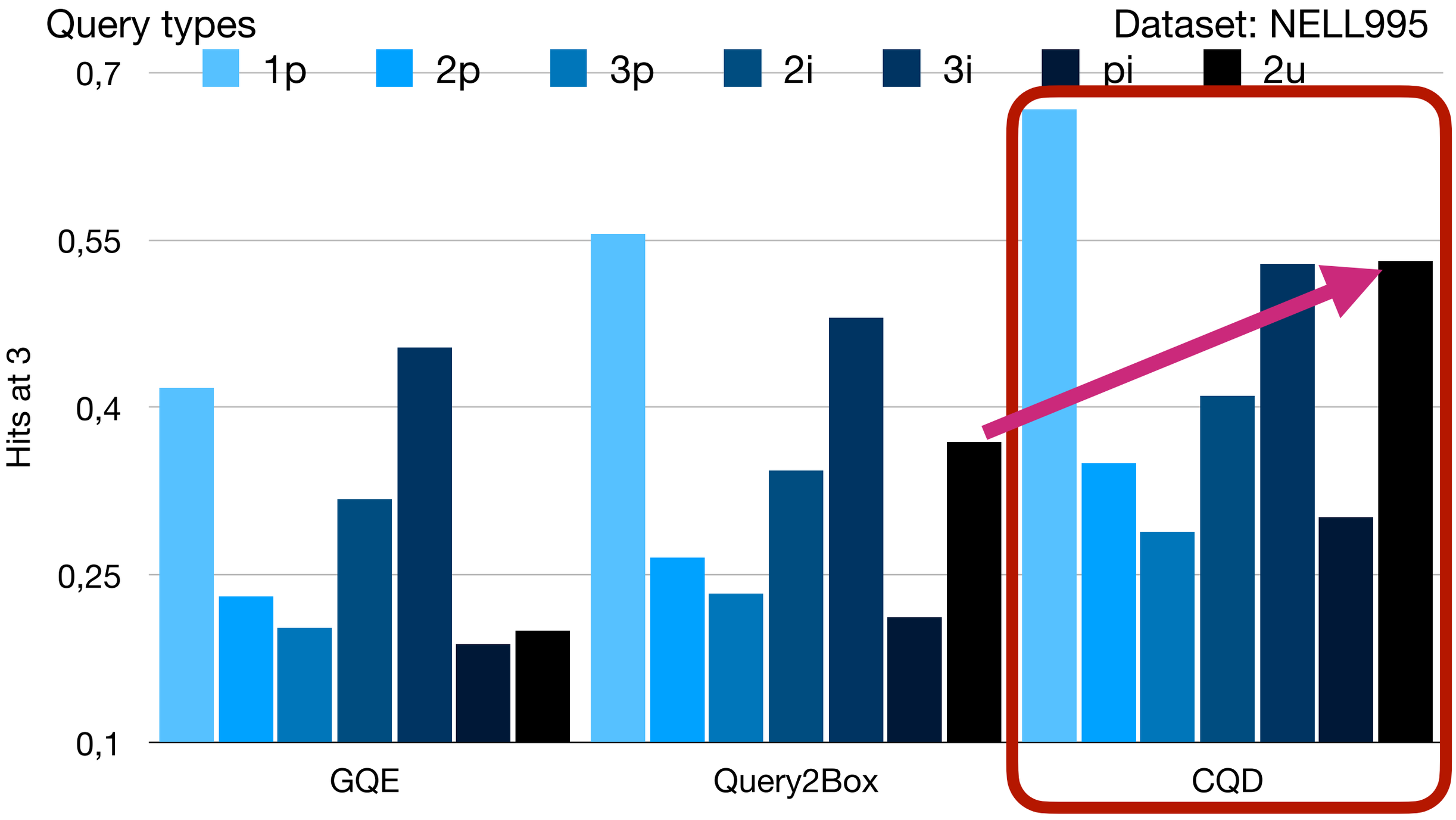
Experiments



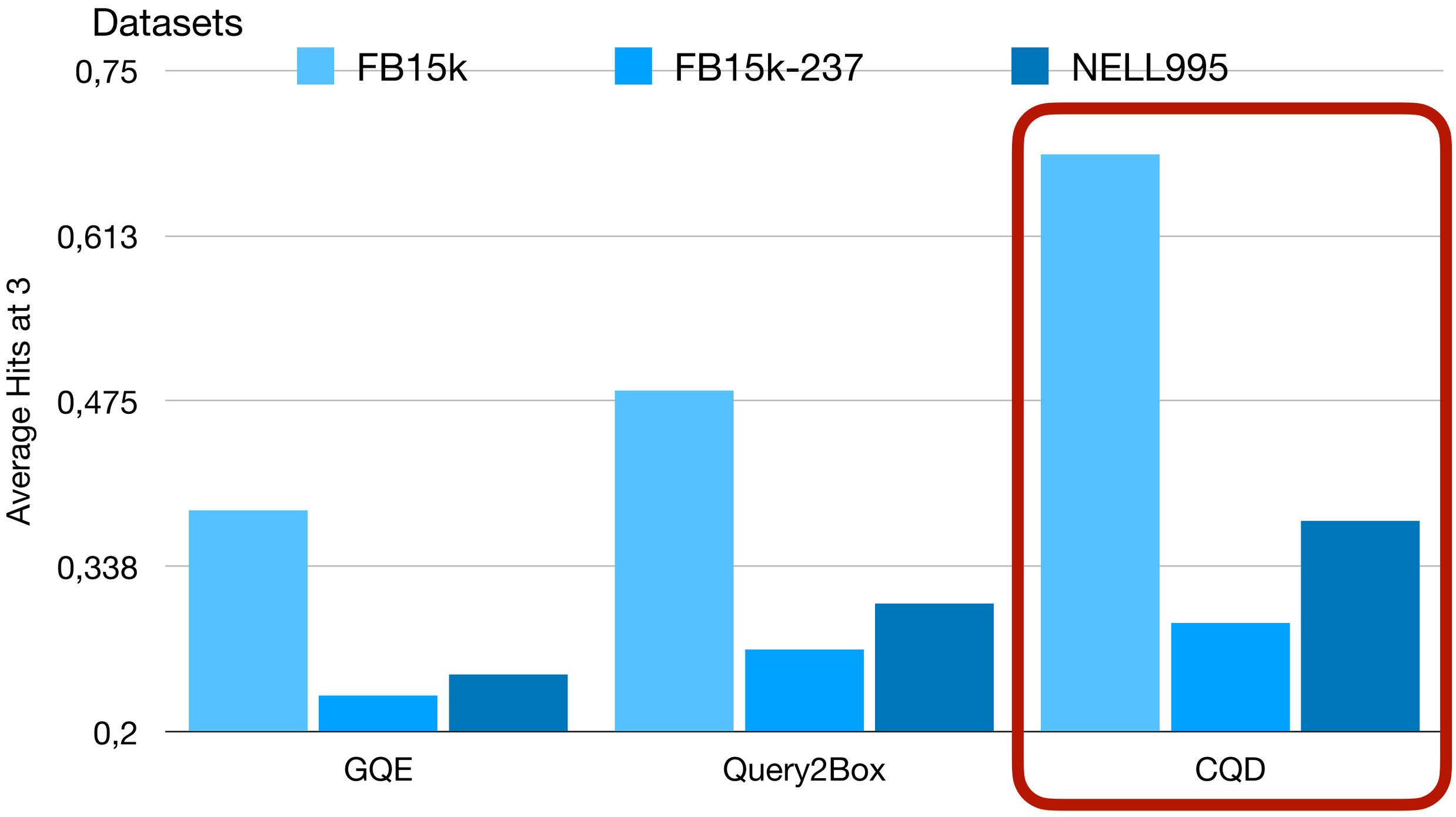
Experiments



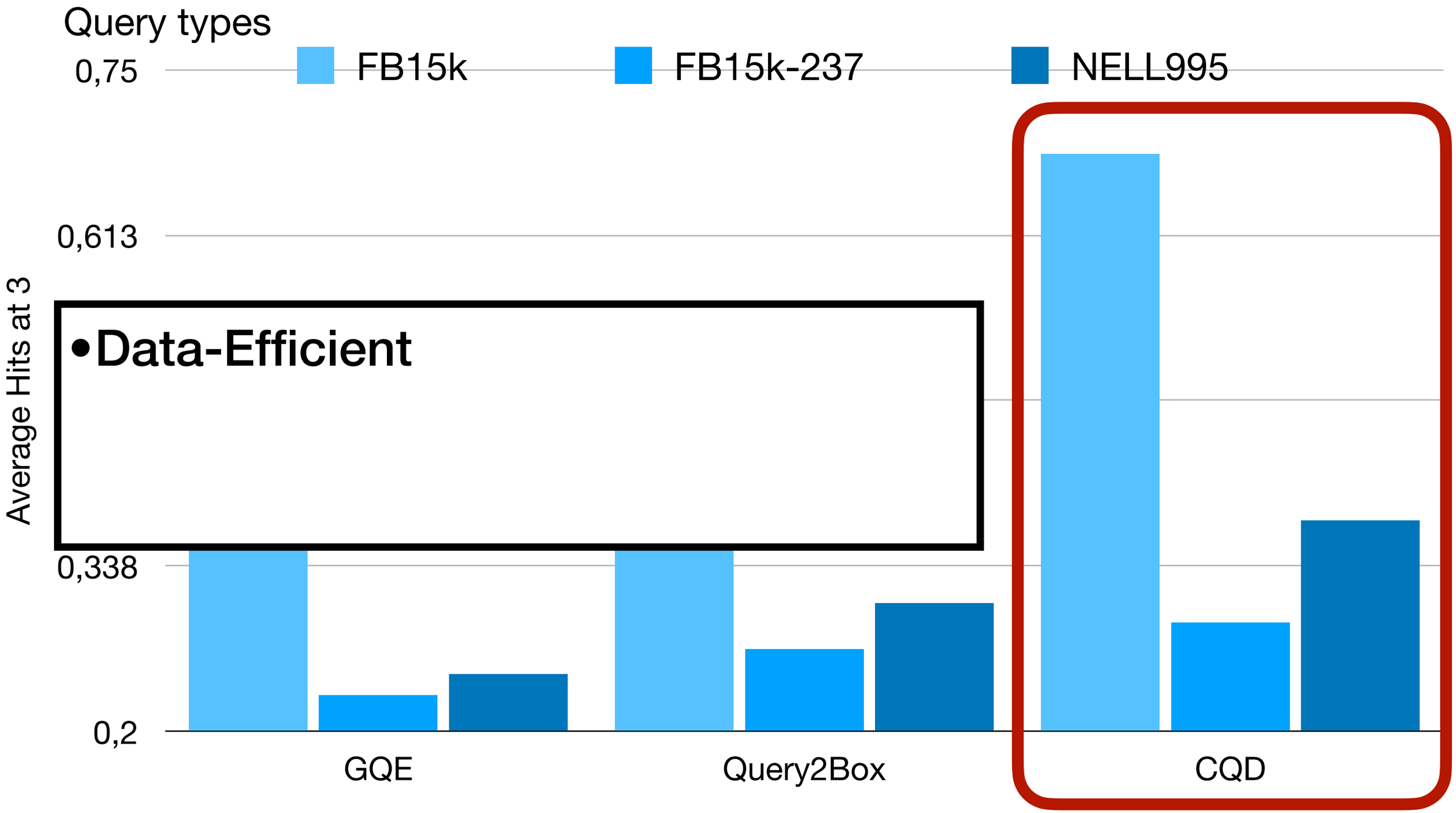
Experiments



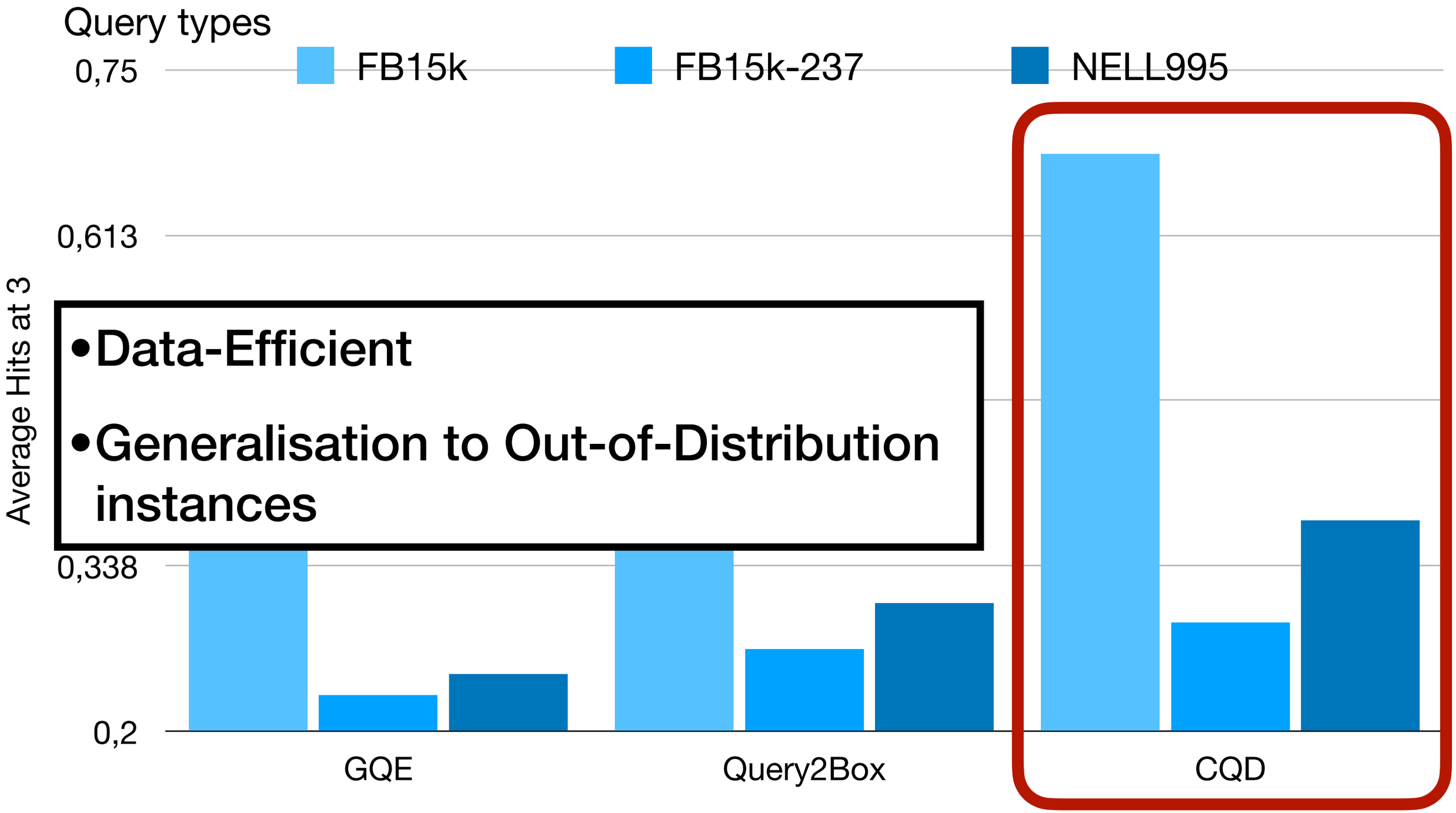
Experiments



Experiments



Experiments



Explainability

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

M
Apixaban
Amitriptyline
Phenytoin
Duloxetine
Buprenorphine

Explainability

Query: Which medications have side-effects when taken with drugs for treating Anxiety?

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

- **Explainable**

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

Summary

Summary

Novel approach to answering Complex Queries on large-scale incomplete Knowledge Graphs:

Summary

Novel approach to answering Complex Queries on large-scale incomplete Knowledge Graphs:

- Train a neural link predictor on atomic queries

Summary

Novel approach to answering Complex Queries on large-scale incomplete Knowledge Graphs:

- Train a neural link predictor on atomic queries
- Answer complex queries by formulating the task as an optimisation problem

Summary

Novel approach to answering Complex Queries on large-scale incomplete Knowledge Graphs:

- Train a neural link predictor on atomic queries
- Answer complex queries by formulating the task as an optimisation problem

Generalises extremely well to complex queries, despite not being trained on them

Summary

Novel approach to answering Complex Queries on large-scale incomplete Knowledge Graphs:

- Train a neural link predictor on atomic queries
- Answer complex queries by formulating the task as an optimisation problem

Generalises extremely well to complex queries, despite not being trained on them

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