Learning Rules from Incomplete KGs Using Embeddings

Vinh Thinh Ho[1], Daria Stepanova[1], Mohamed Gad-Elrab[1], Evgeny Kharlamov[2], Gerhard Weikum[1]

hvthinh@mpi-inf.mpg.de, dstepano@mpi-inf.mpg.de, gadelrab@mpi-inf.mpg.de, evgeny.kharlamov@cs.ox.ac.uk, weikum@mpi-inf.mpg.de

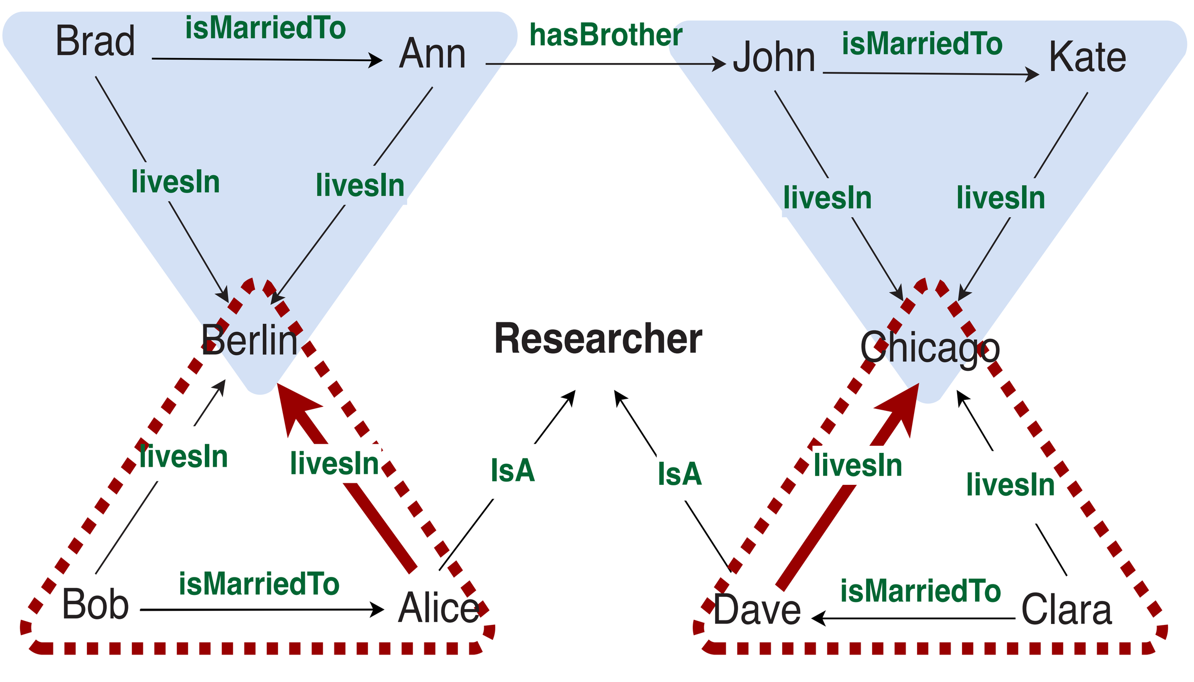
[1]Max Planck Institute for Informatics, Saarbrücken, Germany
[2]University of Oxford, Oxford, United Kingdom

1. Motivation and Contributions

Knowledge graphs: huge collections of positive unary and binary facts treated under **O**pen **W**orld **A**ssumption (e.g. $isMarriedTo(clara, dave), researcher(dave)$)

Automatically constructed, thus **incomplete** ⇒ **KG completion task**

Rule-based approach



$livesIn(Z, Y) \leftarrow livesIn(X, Y), marriedTo(X, Z)$

$$conf(r) = \frac{|\blacktriangle|}{|\blacktriangle| + |\triangle|} = 0.5$$

- \+ Interpretable
- \+ Allow for reasoning
- \- Not extendable
- \- Local patterns

- \- Hard to interpret
- \- No reasoning
- \+ Extendable (e.g., text)
- \+ Global patterns

Our approach: rule-based with embeddings support

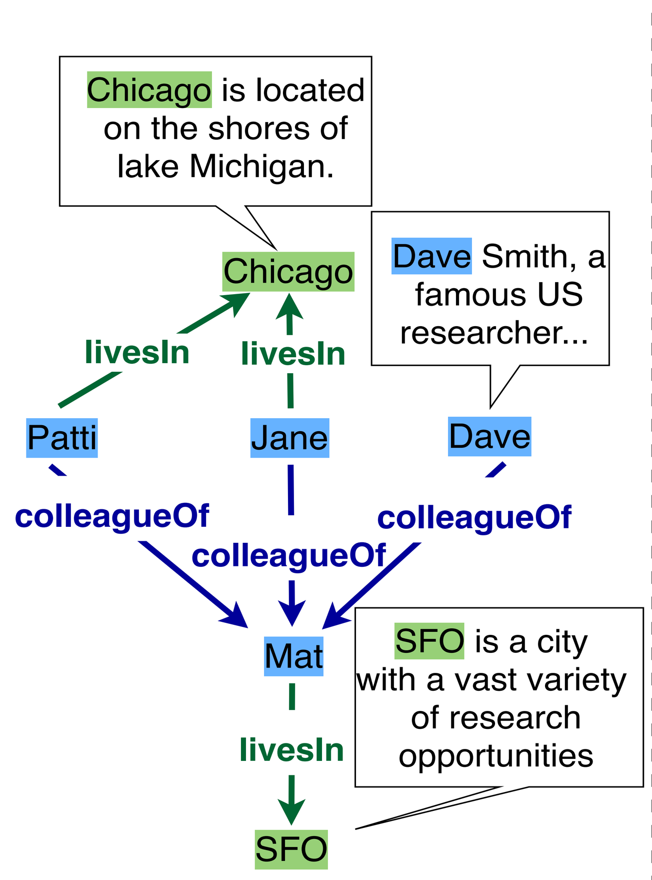
Challenges:

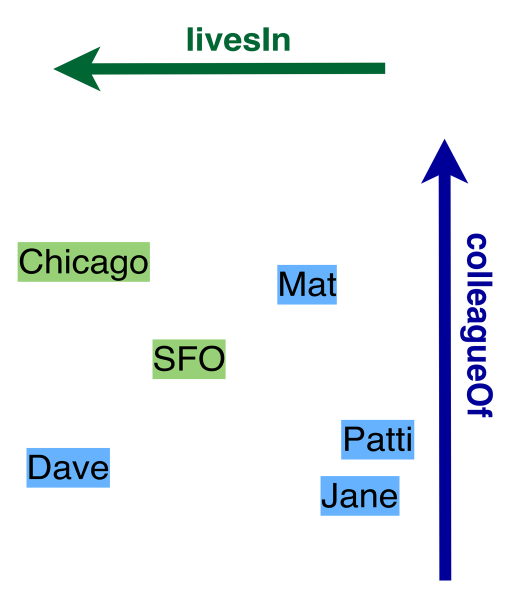
- Structurally different output
- Large embedding size
- Large rule search space

Contributions:

- Framework for rule learning with external sources
- Hybrid embedding based rule measure
- Experiments on real world KGs

Embedding-based approach





score(<Dave livesIn SFO>)=0.8
score(<Dave livesIn Chicago>)=0.4

2. Our Proposal: Rule Learning with External Sources

- **Problem statement:**

Given: $\mathcal{P} = (\mathcal{G}, f)$

- **Knowledge graph** $\mathcal{G}$
- **Probability function** f: trusfulness of $\mathcal{G}$'s missing facts

Find: Ordered set of **rules**, which

- **Describe** $\mathcal{G}$ well and **predict** highly probable facts based on f

- **Our solution:**

Hybrid rule quality function to prune search space of rules r:

$$\mu(r, \mathcal{P}) = (1 - \lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(\mathcal{G}_r, \mathcal{P})$$

- **Descriptive quality** μ_1 of rule r over $\mathcal{G}$:

$$\mu_1 : (r, \mathcal{G}) \mapsto \alpha \in [0, 1]$$

⇒ any classical rule measure, e.g., confidence

- **Predictive quality** μ_2 of r: trustfulness of predictions $\mathcal{G}_r$ made by r on $\mathcal{G}$

$$\mu_2 : (\mathcal{G}_r, \mathcal{P}) \mapsto \alpha \in [0, 1]$$

⇒ capture **information about missing facts** in $\mathcal{G}$ that are relevant for r

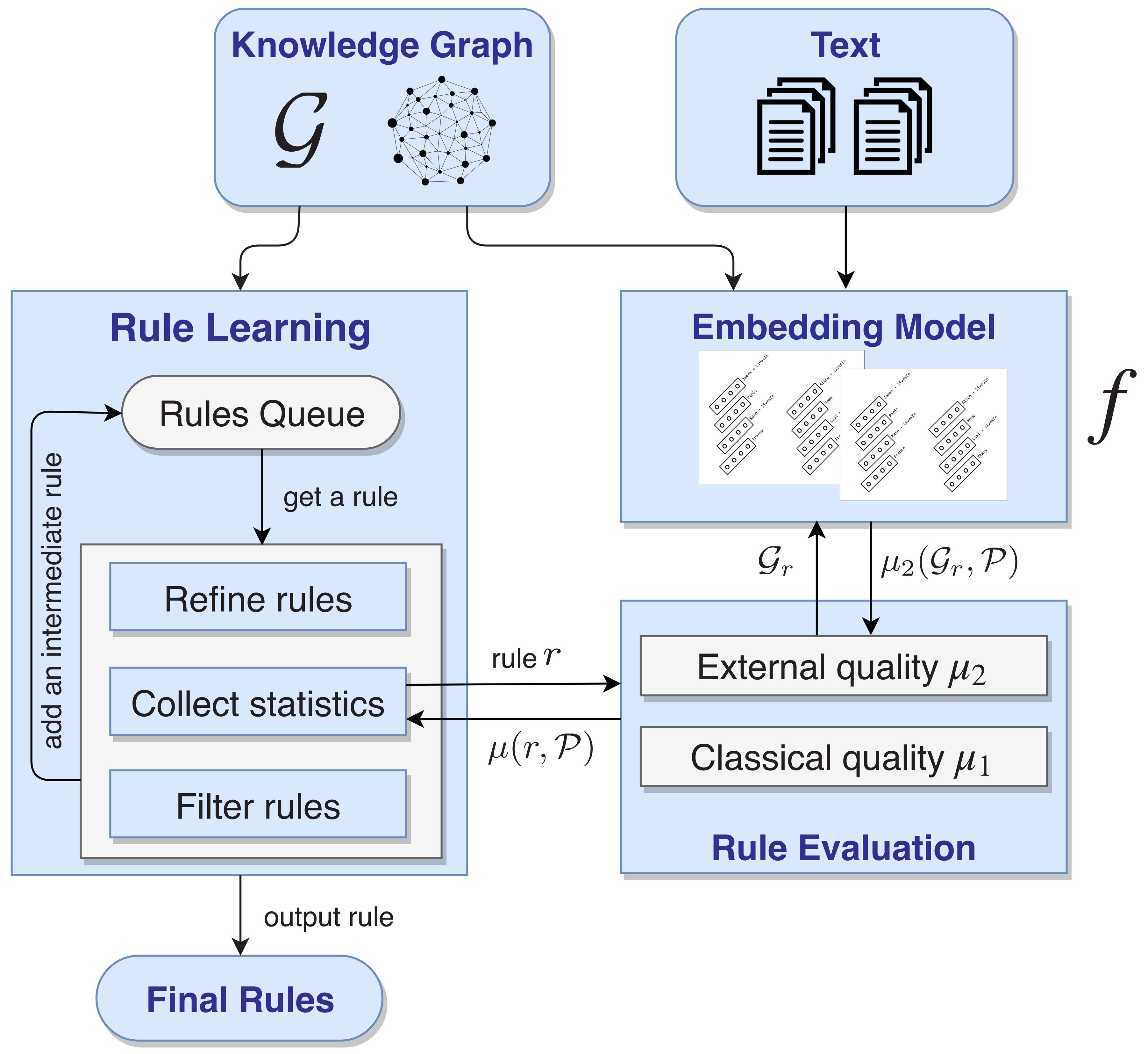
- **Weighting factor** $\lambda \in [0, 1]$ to control the distribution of μ_1 and μ_2

- **Realization of f and μ_2 relying on embeddings:**

$$f(fact) = 0.5 \times (1/subject\_rank(fact) + 1/object\_rank(fact))$$

$$\mu_2(\mathcal{G}_r, \mathcal{P}) = \frac{\Sigma_{fact \in \mathcal{G}_r \setminus \mathcal{G}} f(fact)}{|\mathcal{G}_r \setminus \mathcal{G}|}$$

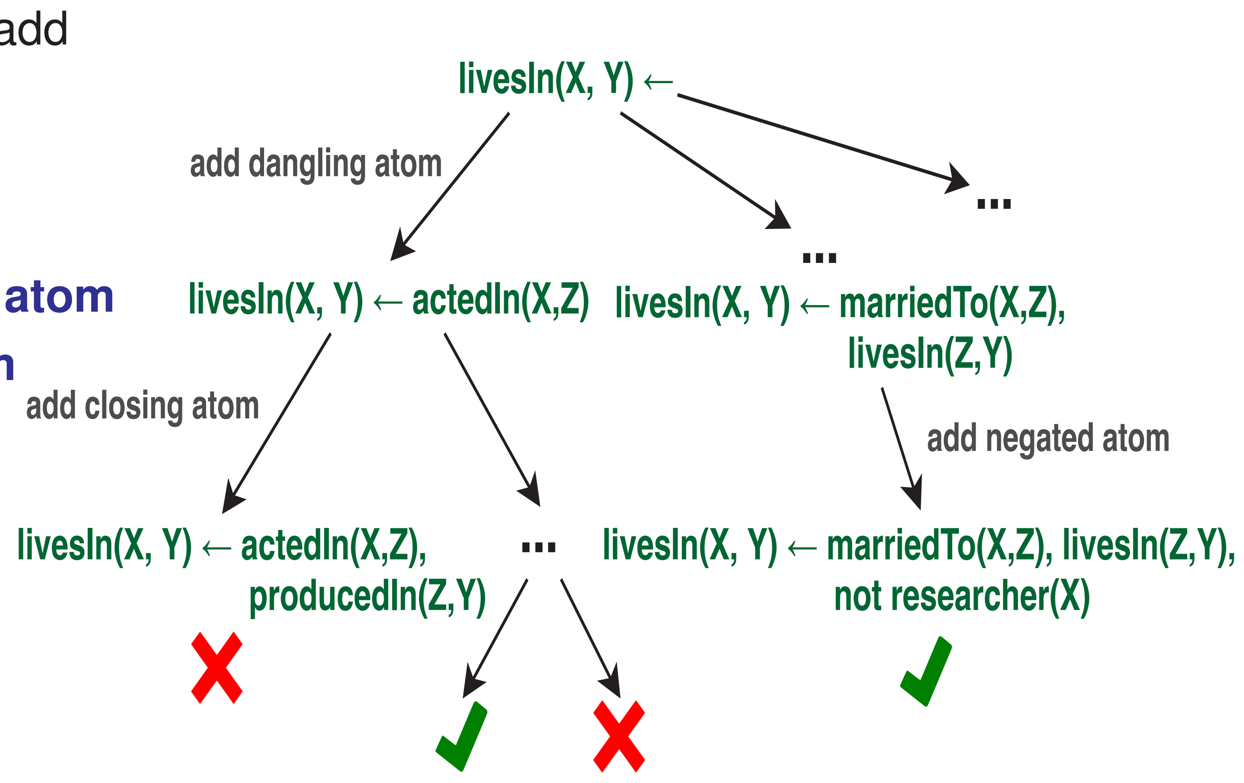
3. General Architecture



4. Rule Refinement

Extended AMIE [Galárraga, *et al*, VLDB 2015] (additions are in blue):

- **Refinement operators**: add
  - dangling atom
  - instantiated atom
  - closing atom
  - **negated instantiated atom**
  - **negated closing atom**
- **Rule filtering**:
  - language bias
  - support
  - head coverage
  - confidence
  - **embedding-based measure (μ)**
  - **exception confidence**:

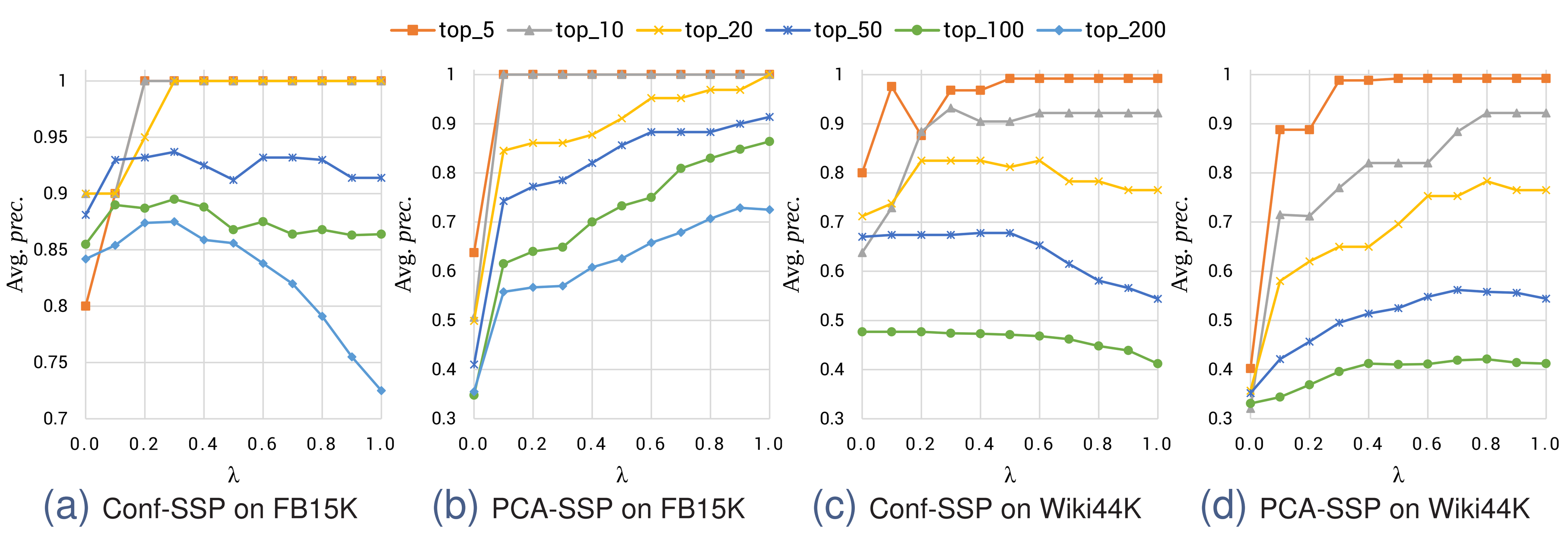


$$e\text{-}conf(r, \mathcal{G}) = conf(r', \mathcal{G})$$

where $r' : body^-(r) \leftarrow body^+(r), not\ head(r)$

5. Experiments

- **Approximation of complete KG**: original
- **Available KG**: random 80% of original KG, preserving the distribution of facts over predicates.
- **Embedding models**:
  - TransE, HolE, SSP (with text)



(a) Conf-SSP on FB15K (b) PCA-SSP on FB15K (c) Conf-SSP on Wiki44K (d) PCA-SSP on Wiki44K

Evaluation result on *closed world* setting (CW)

- **Examples of mined rules:**

r_1: $nationality(X, Y) \leftarrow graduated\_from(X, Z), in\_country(Z, Y), not\ research\_uni(Z)$

r_2: $scriptwriter\_of(X, Y) \leftarrow preceded\_by(X, Z), scriptwriter\_of(Z, Y), not\ tv\_series(Z)$