Type-Constrained Representation Learning in Knowledge Graphs

Denis Krompaß[1,2], Stephan Baier[1] and Volker Tresp[1,2]
[1] Department of Computer Science. Ludwig Maximilian University,
[2] Corporate Technology, Siemens AG

Outline

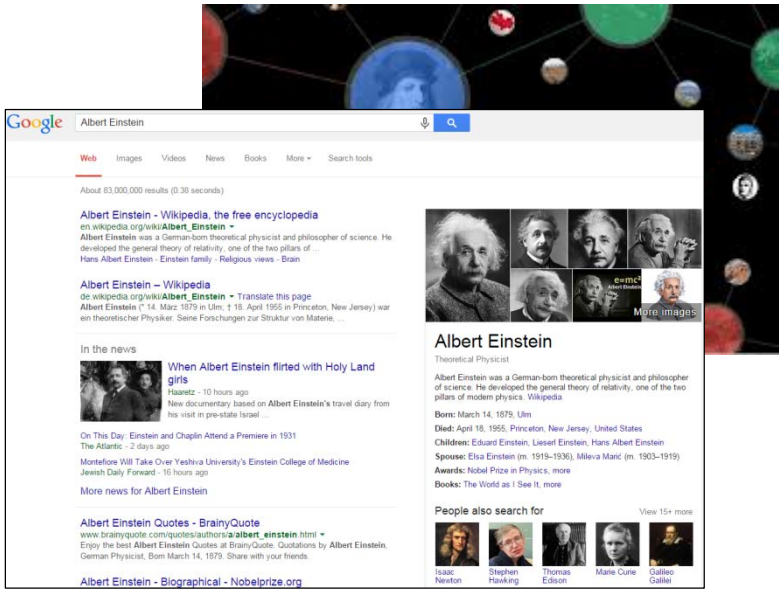
1. Knowledge graphs, what are they and what are they good for?

2. Learning in (large) knowledge graphs with Latent Variable Models

3. Exploiting prior knowledge about relation-types
    1. Type-Constraints extracted from the schema
    2. Local Closed-World assumption on relation-types

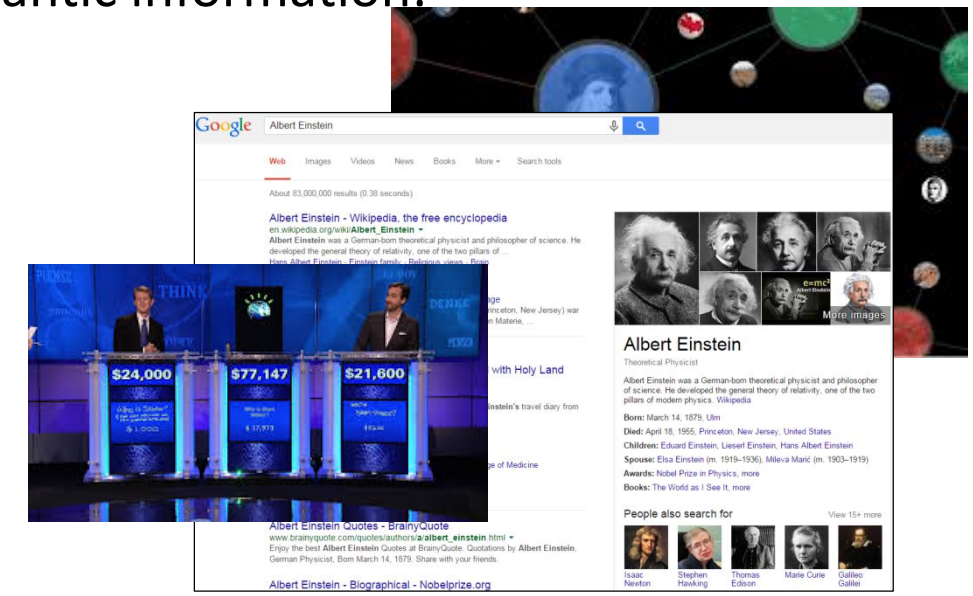
Knowledge Graphs

- Stores facts about the world as relations between entities.
- **Entities** are no longer just strings but real world objects with attributes, taxonomic information and **relations** to other objects.
  - (**AlbertEinstein**, **bornIn**, **Ulm**)



Knowledge Graphs

- Stores facts about the world as relations between entities.
- **Entities** are no longer just strings but real world objects with attributes, taxonomic information and **relations** to other objects.
  - (**AlbertEinstein**, **bornIn**, **Ulm**)
- Providing a machine with semantic information:
  - Search engines
  - Information retrieval
  - Word-sense disambiguation
  - …
- Prominent Examples:
  - Google Knowledge Graph
  - IBM Watson



Learning in Knowledge Graphs

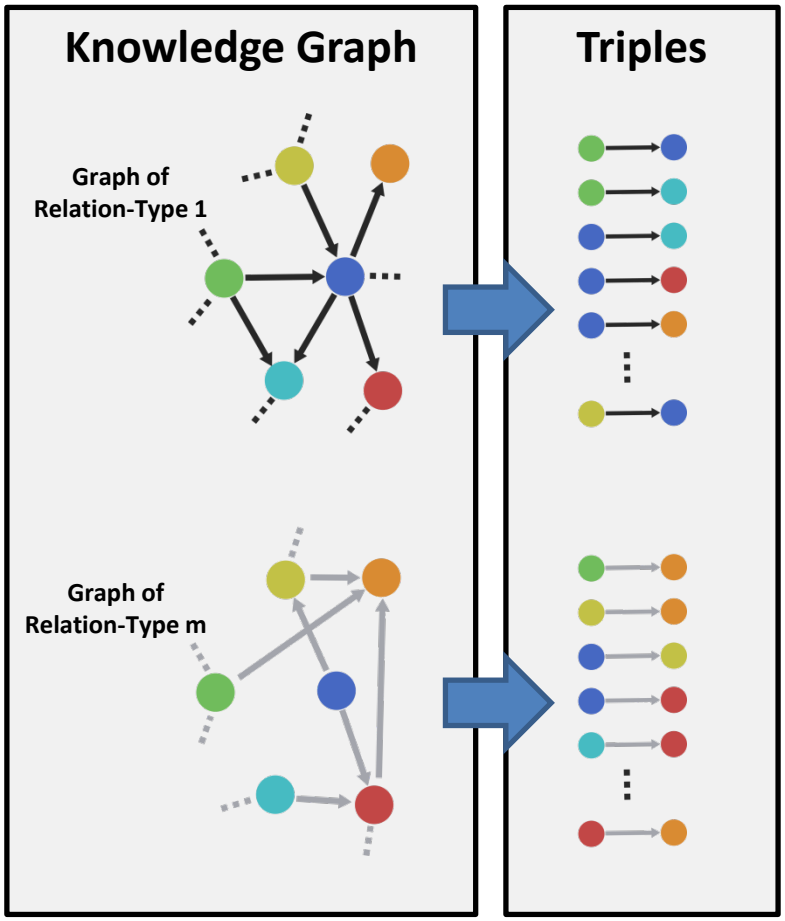
- Why do we need machine learning methods?
  - Knowledge Graphs suffer from incompleteness
    - DBpedia & Freebase ~ 60% of persons have no place of birth
  - Facts can be wrong or outdated

Learning in Knowledge Graphs

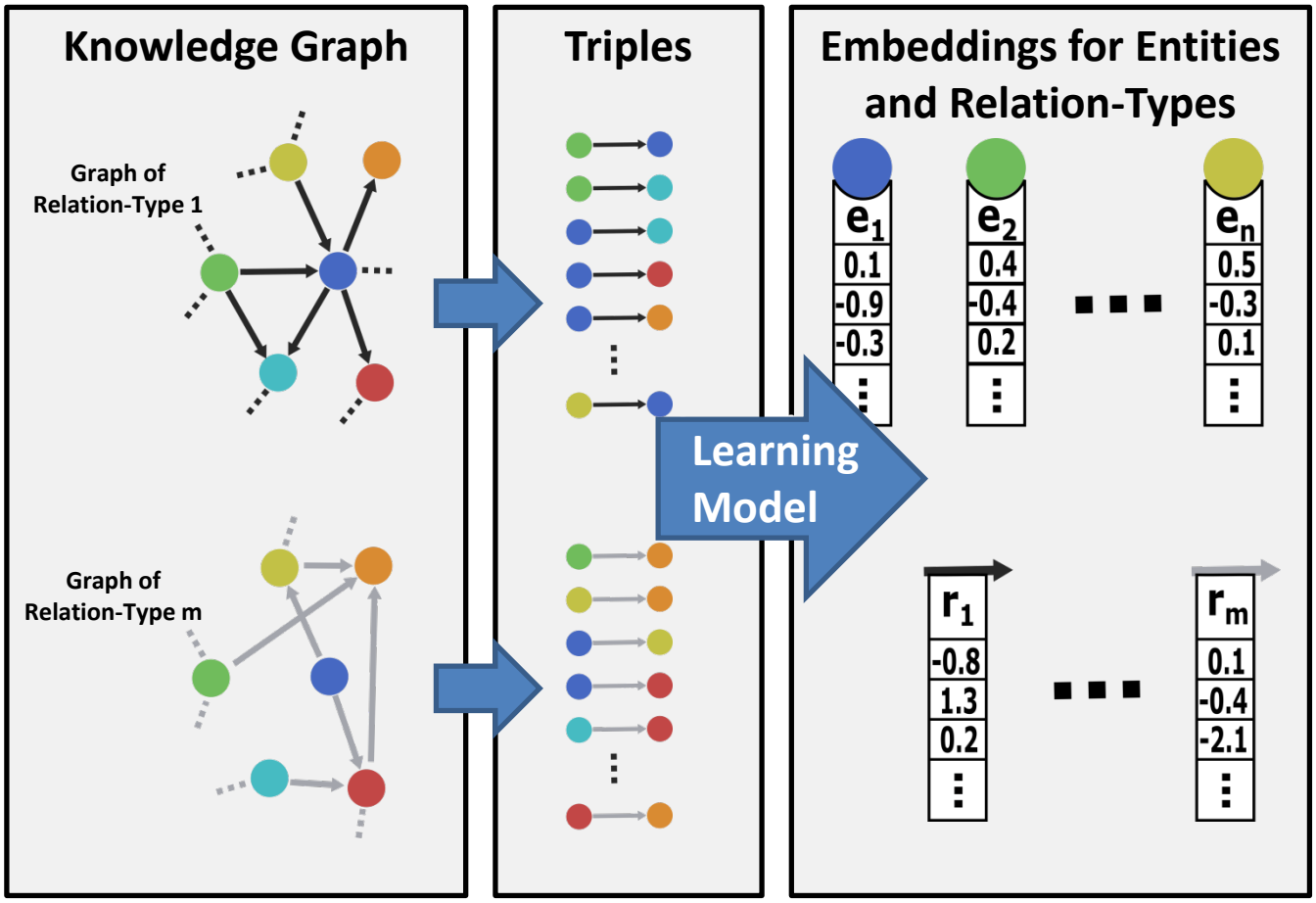
- Why do we need machine learning methods?
  - Knowledge Graphs suffer from incompleteness
    - DBpedia & Freebase ~ 60% of persons have no place of birth
  - Facts can be wrong or outdated

- Latent Variable Models work great in this domain
  - Learn latent embeddings for entities and relation-types (Feature Learning)
  - Can be used for automatic graph completion / error checking /clustering /disambiguation tasks

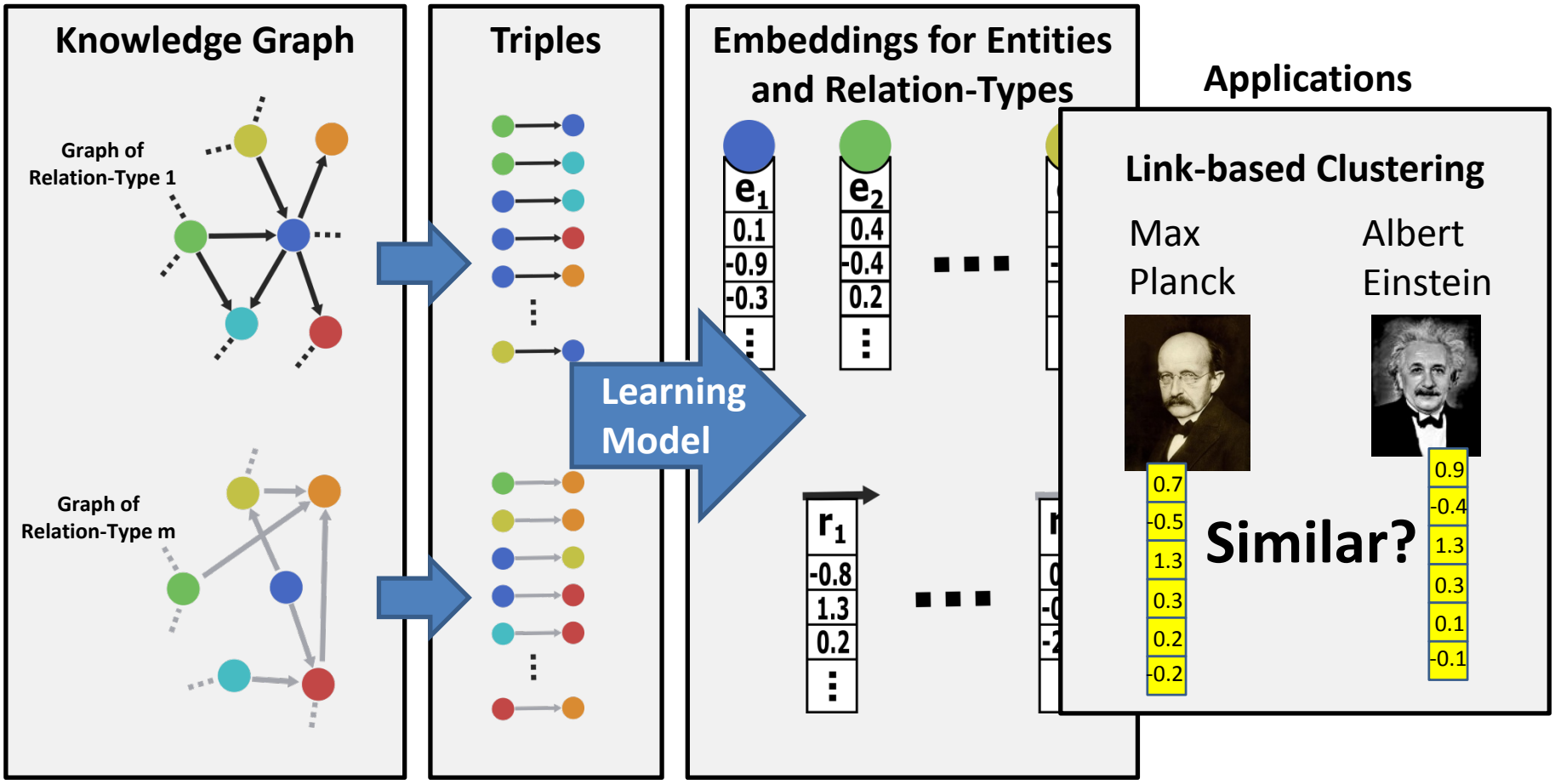
Learning in Knowledge Graphs



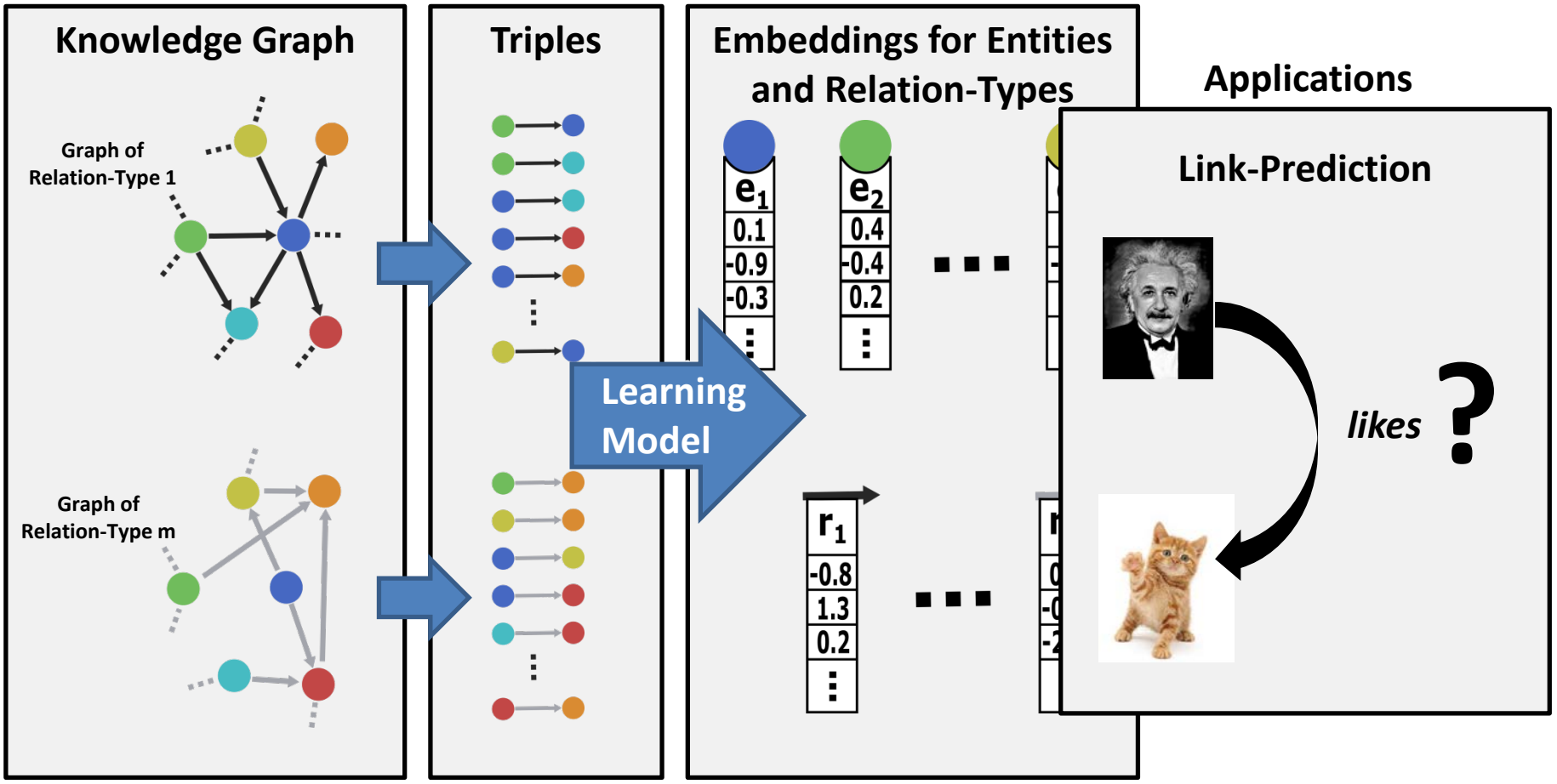
Learning in Knowledge Graphs



Learning in Knowledge Graphs

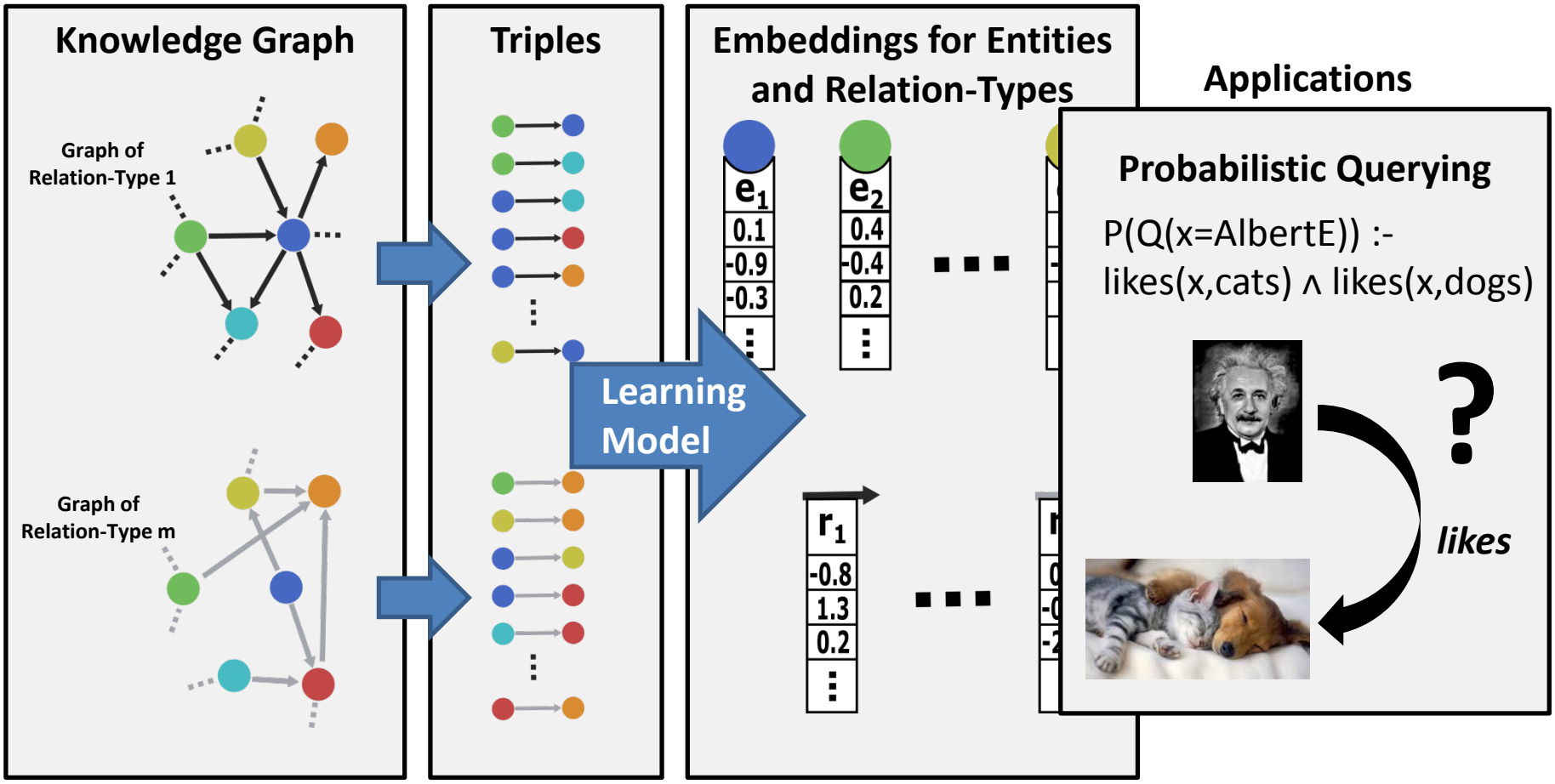


Learning in Knowledge Graphs



*Kitten image. http://www.cats.org.uk/get-involved/support-us/cat-magazine-menu/about-cat-magazine-menu /

Learning in Knowledge Graphs



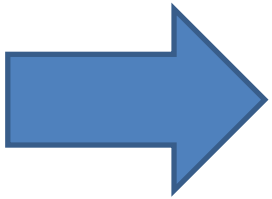
*Puppy and kitten image. http://beattiepethospitalhamilton.com/services/puppy-and-kitten-care/

Application in Large Knowledge Graphs

- Complexity (number of parameters) of latent variable models is directly dependent on the number of entities
  - Freebase ~50,000,000 entities (Topics)
  - **High dimensional embeddings are infeasible**
    - Very high **memory** requirements
    - Long **training times** of the models

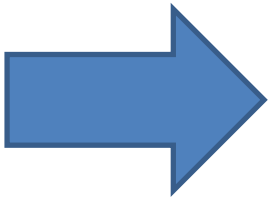
Application in Large Knowledge Graphs

- Complexity (number of parameters) of latent variable models is directly dependent on the number of entities
  - Freebase ~50,000,000 entities (Topics)
  - **High dimensional embeddings are infeasible**
    - Very high **memory** requirements
    - Long **training times** of the models

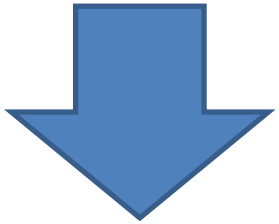


In large-scale problems we are forced to exploit less expressive models

Application in Large Knowledge Graphs



In large-scale problems we are forced to exploit less expressive models



Introduce prior knowledge about the learning task

Type-Constraints in Knowledge Graphs

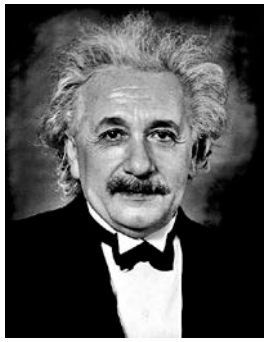
Defines subjects

knows rdfs:domain Person

Defines objects

knows rdfs:range Person

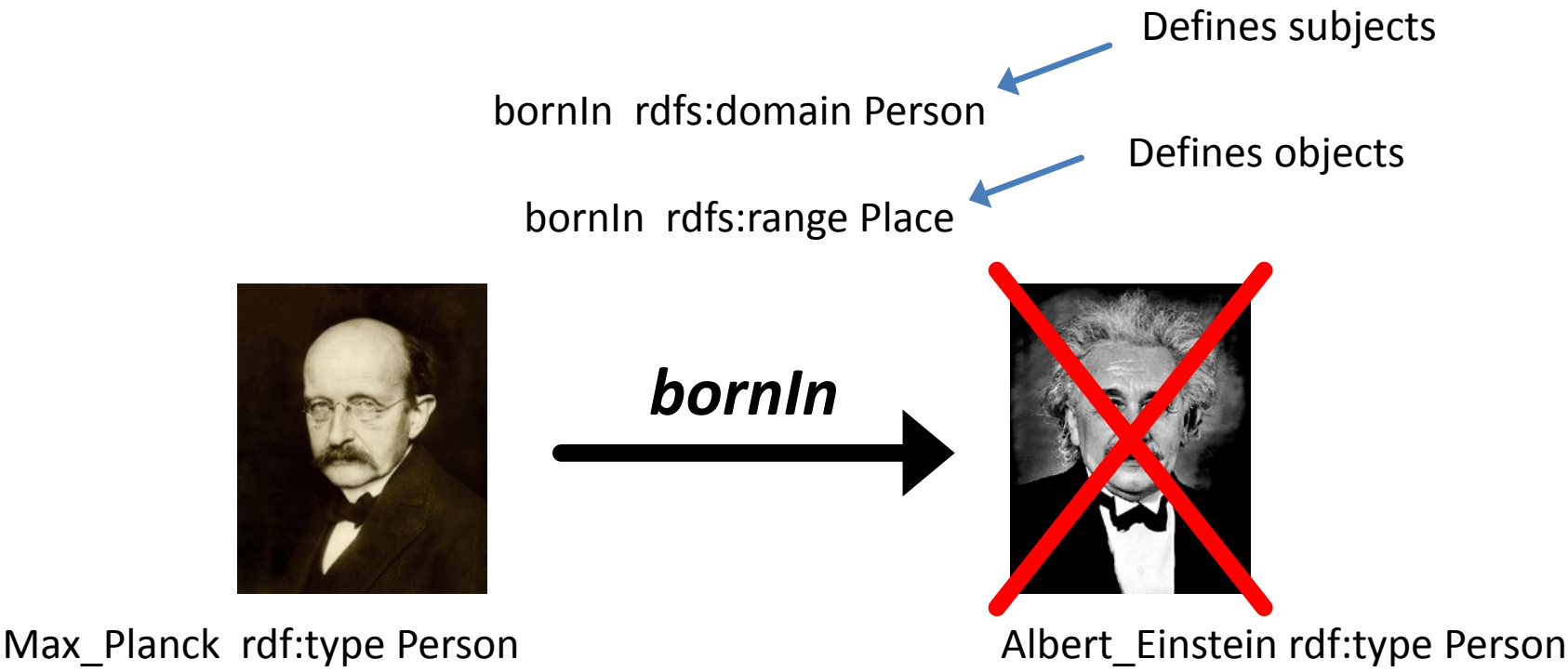
knows



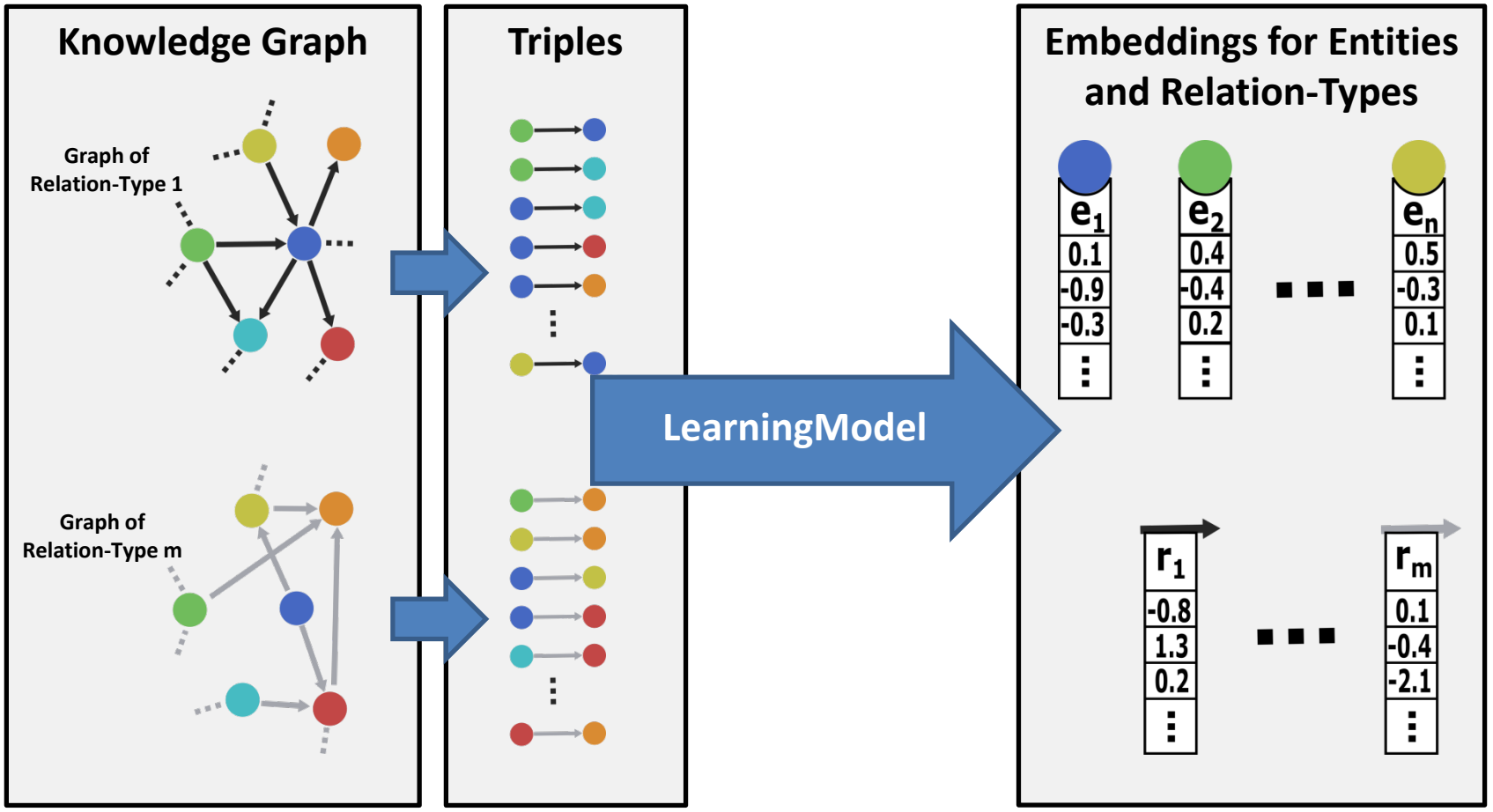
Max_Planck rdf:type Person

Albert_Einstein rdf:type Person

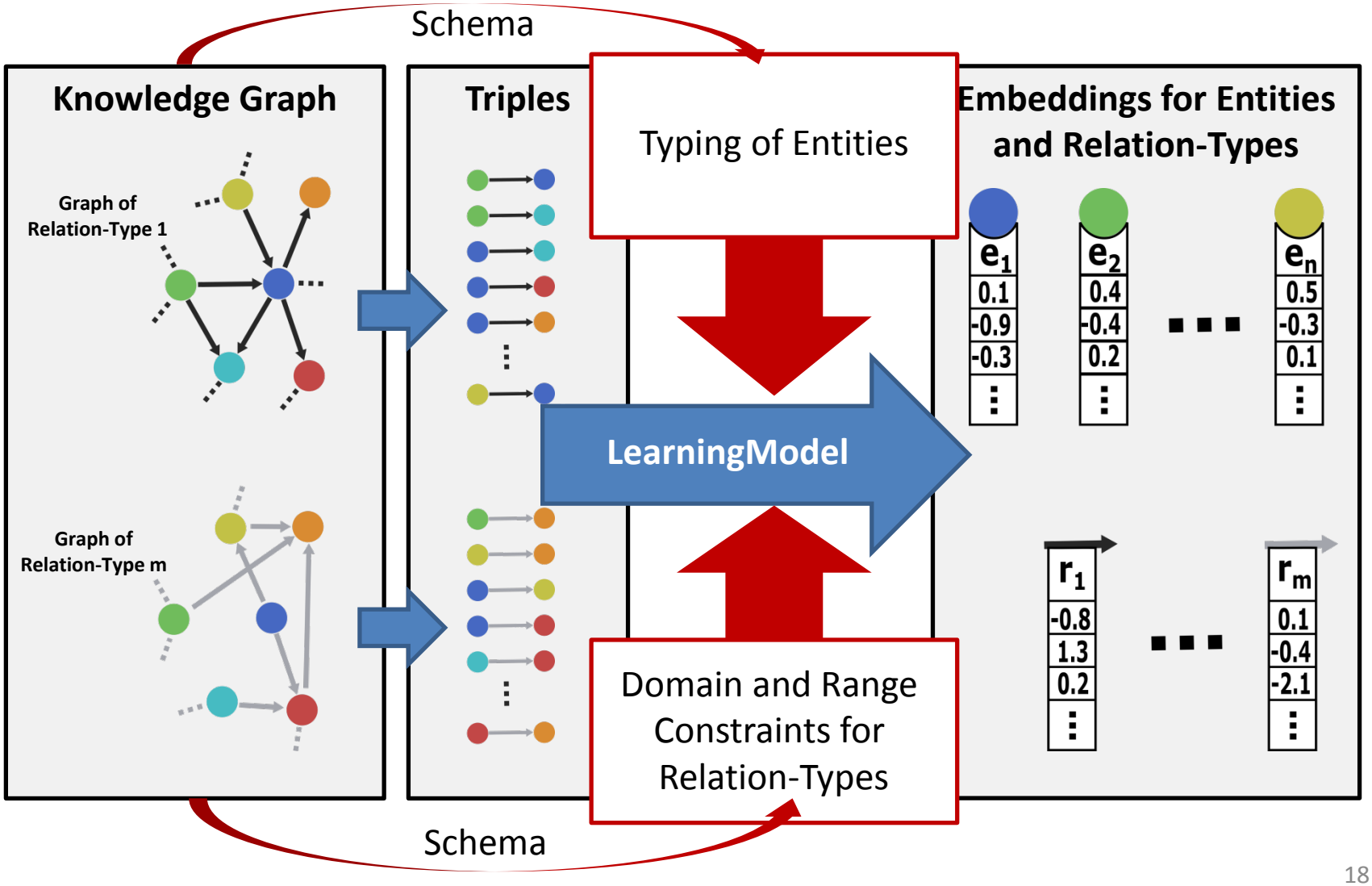
Type-Constraints in Knowledge Graphs



Learning in Knowledge Graphs with Type-Constraints



Learning in Knowledge Graphs with Type-Constraints



Experiments - Models

1. **RESCAL (Nickel et. al. 2011)**
   - Third-Order Tensor Factorization Methods
   - Least-Squares Cost Function
2. **TransE (Bordes et. al. 2013)**
   - Distance-based Method
   - Ranking Cost Function
3. **Google Knowledge Vault Model (Murphy et. al. 2014)**
   - Multi-way Neural Network (mwNN)
   - Logistic Cost Function

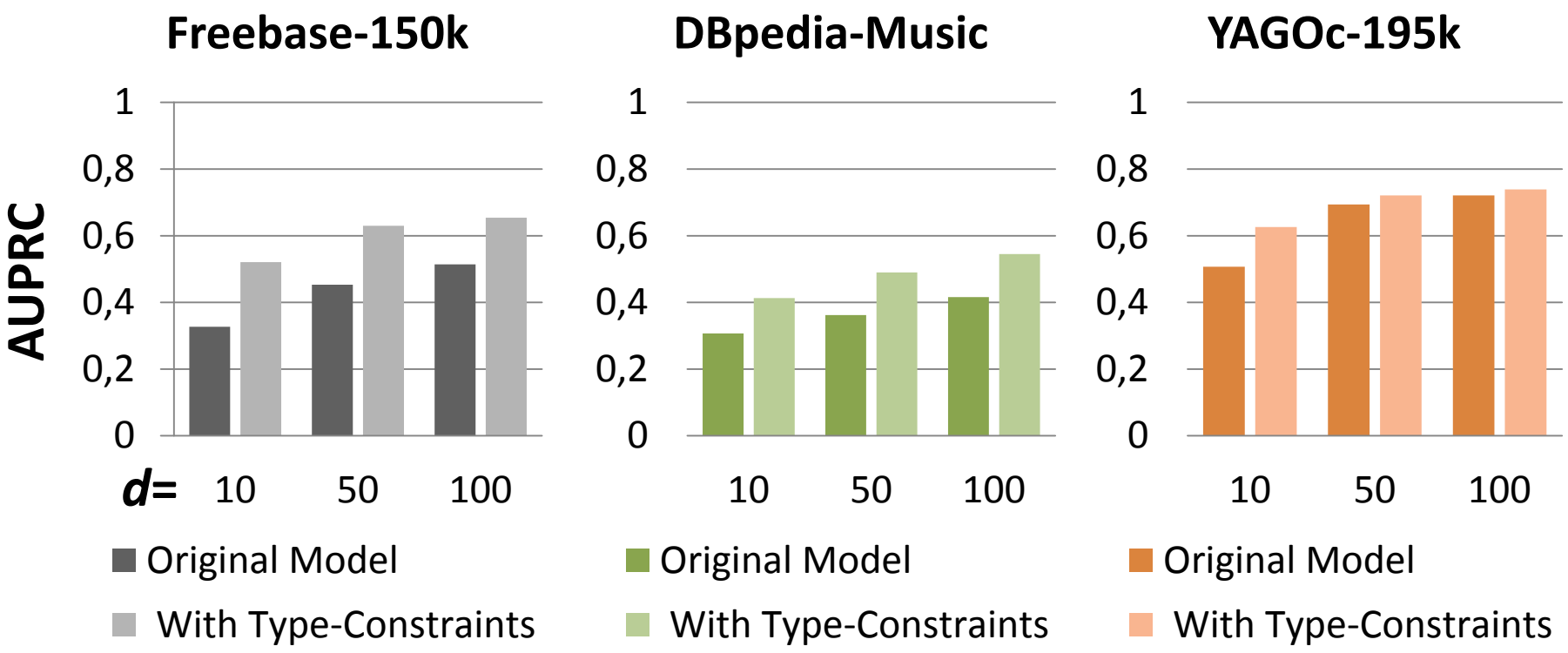
Experiments - Datasets

Dataset	Entities	Relation-Types	Triples
Freebase-150k	151.146	285	1,047,844
DBpedia-Music	321,950	15	981,333
YAGOc-195	195,639	32	1,343,684

- 70% Training Set
- 10% Validation Set (Hyperparameter Tuning)
- 20 % Test Set

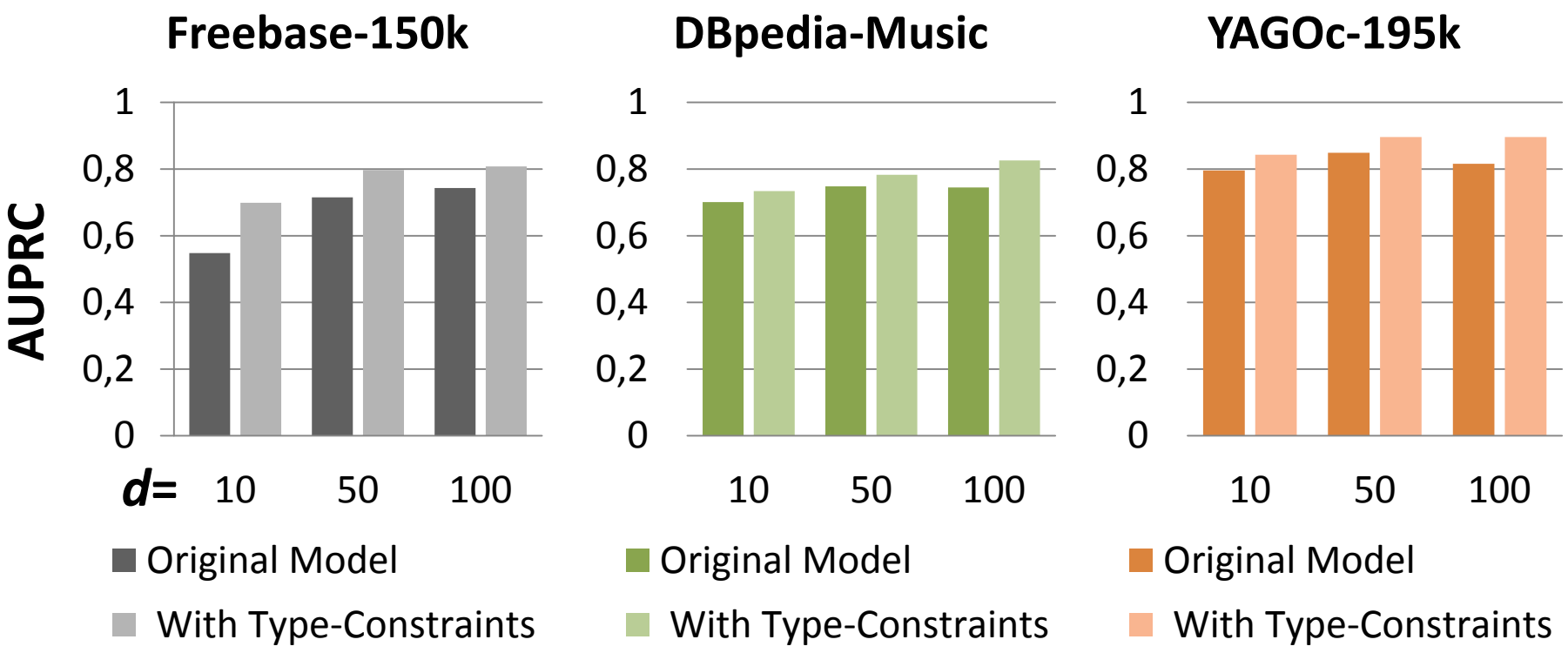
Results – Type-Constraints

RESCAL



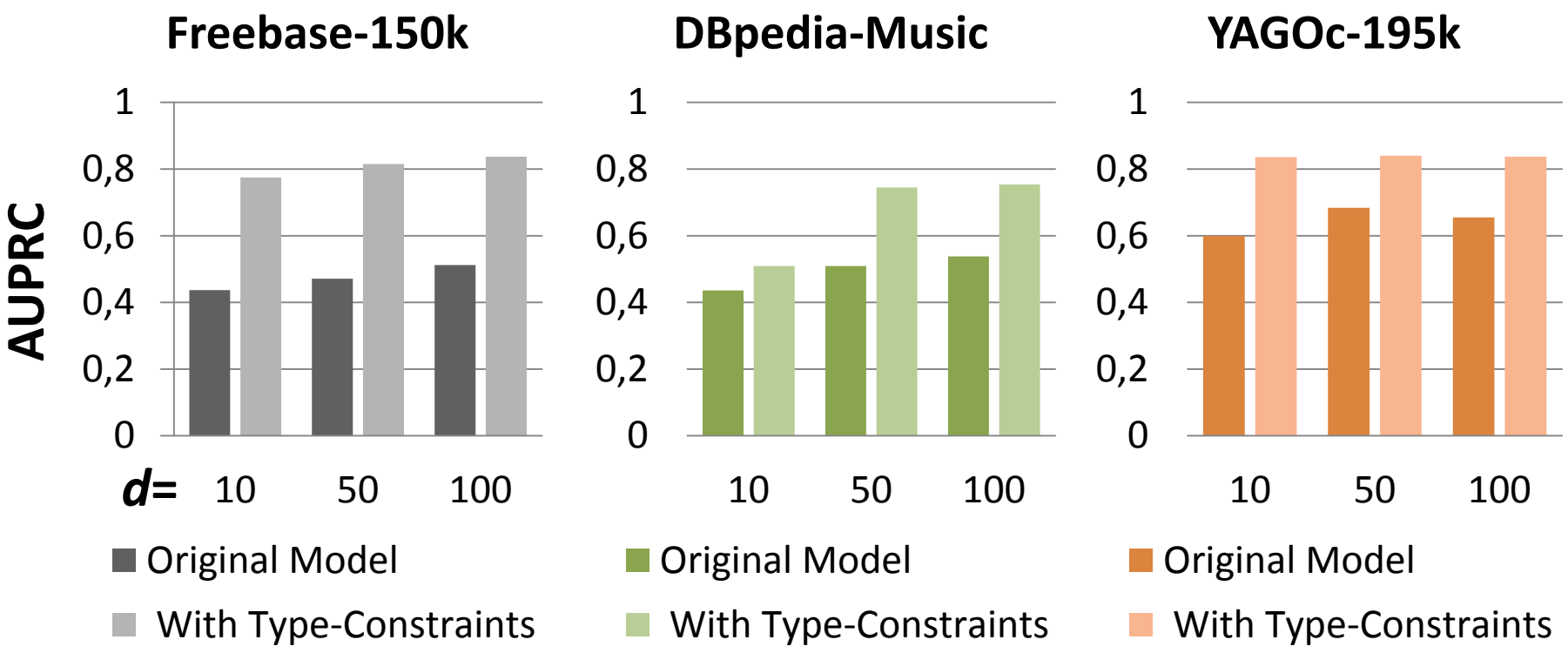
Results – Type-Constraints

TransE



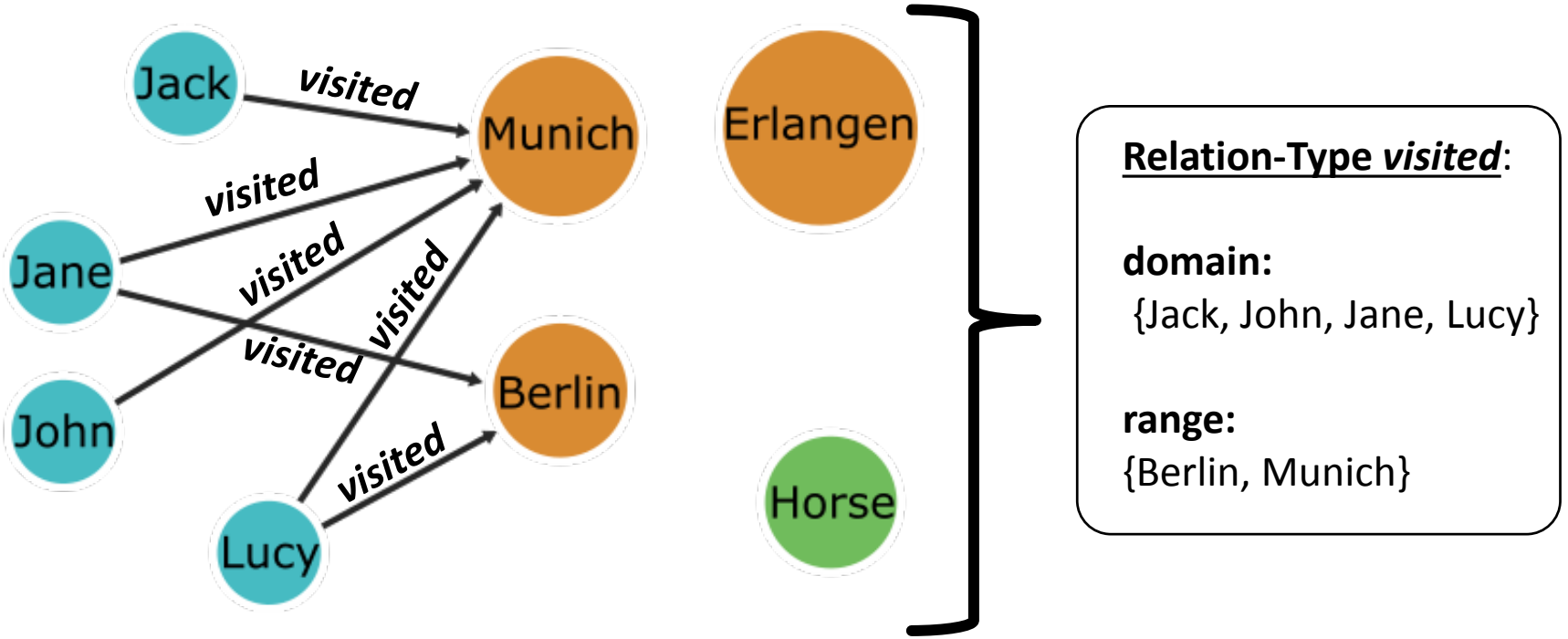
Results – Type-Constraints

Google Knowledge Vault Neural Network

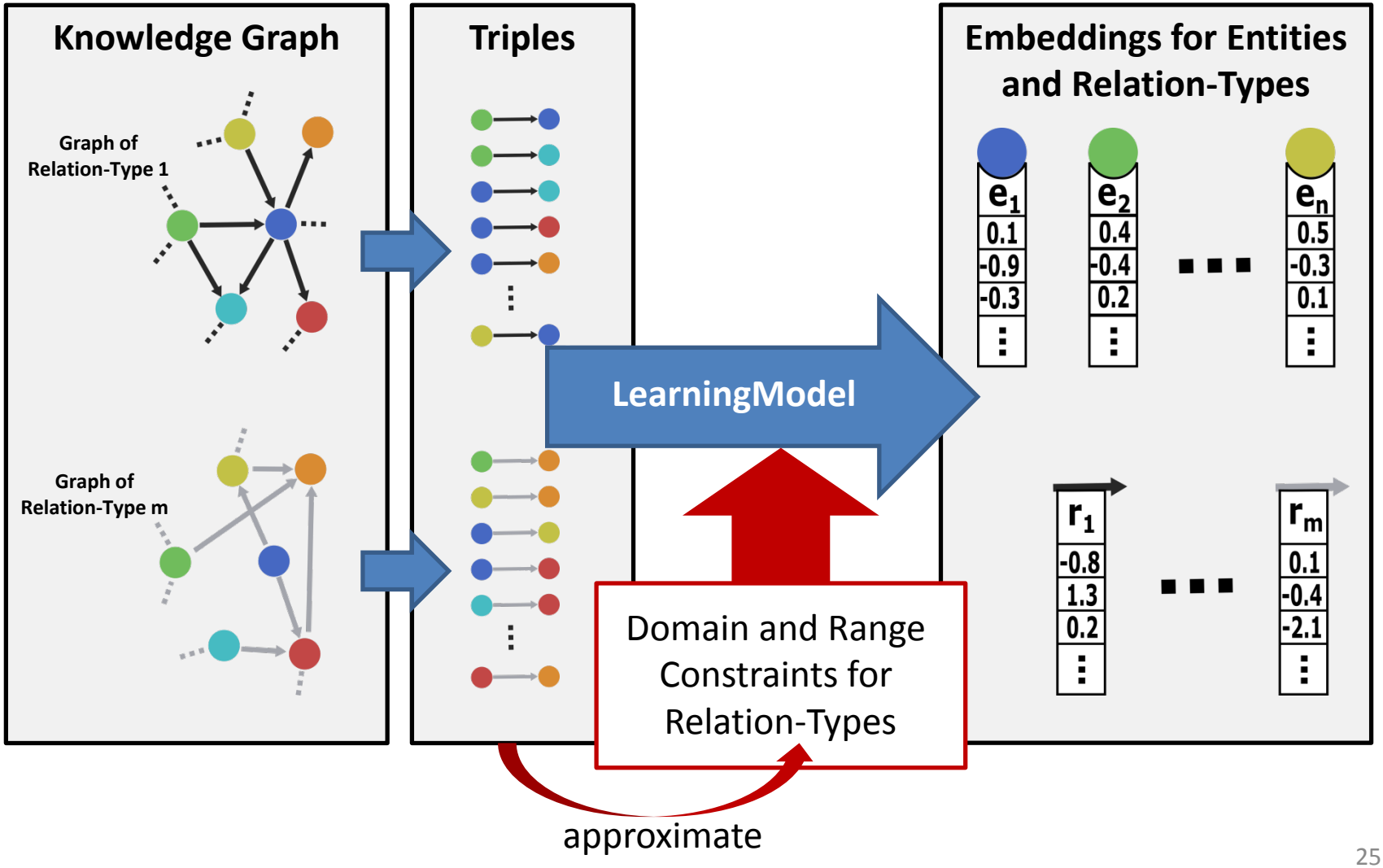


A Local Closed-World Assumption

- Approximate domain and range constraints directly from observed triples
- Can be applied to schema-less knowledge graphs

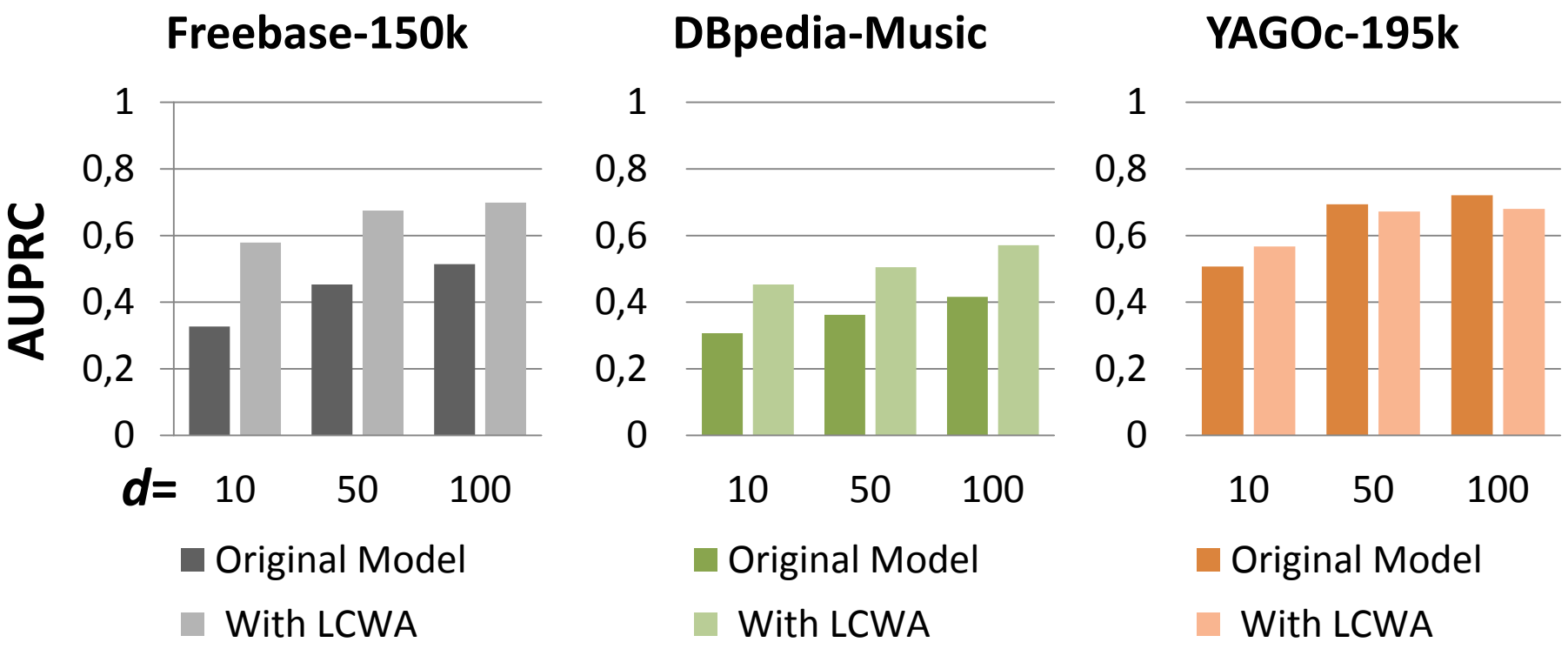


Learning in Knowledge Graphs with Local Closed-World Assumption



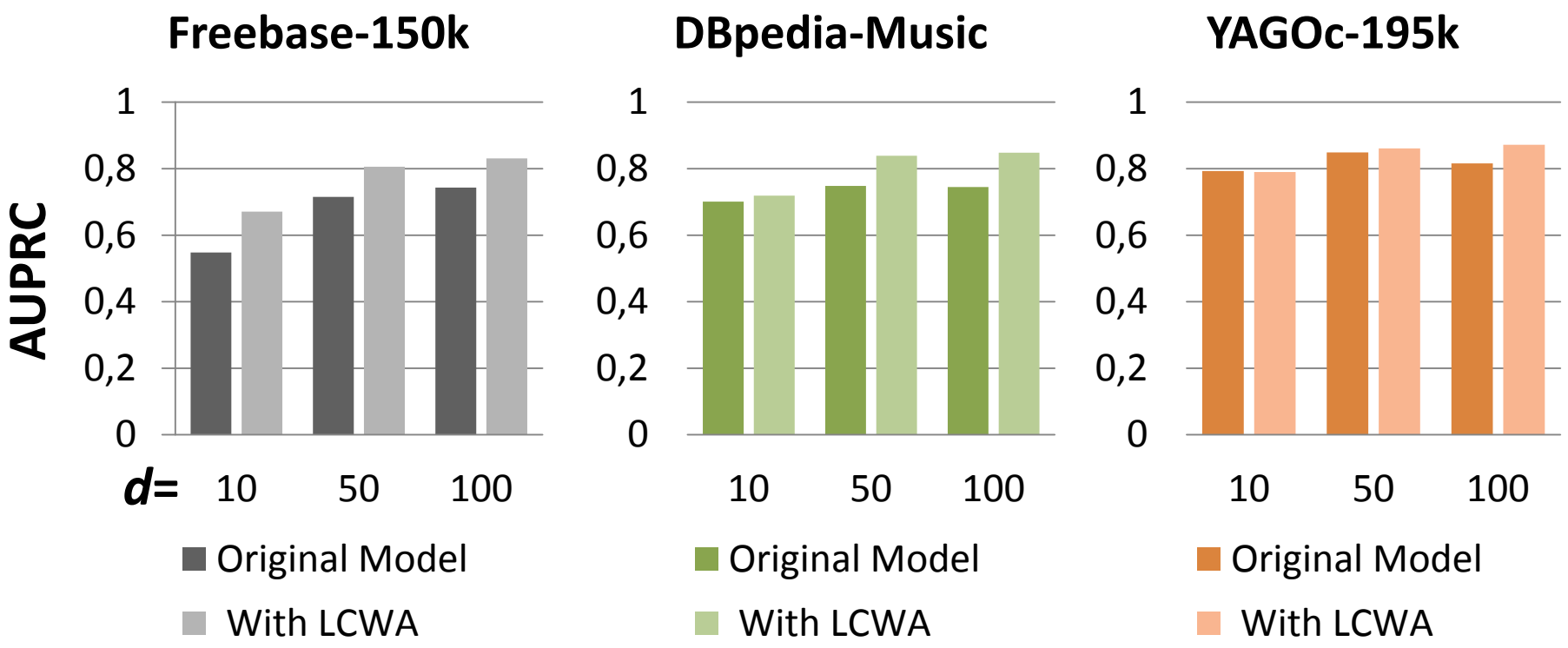
Results - LCWA

RESCAL



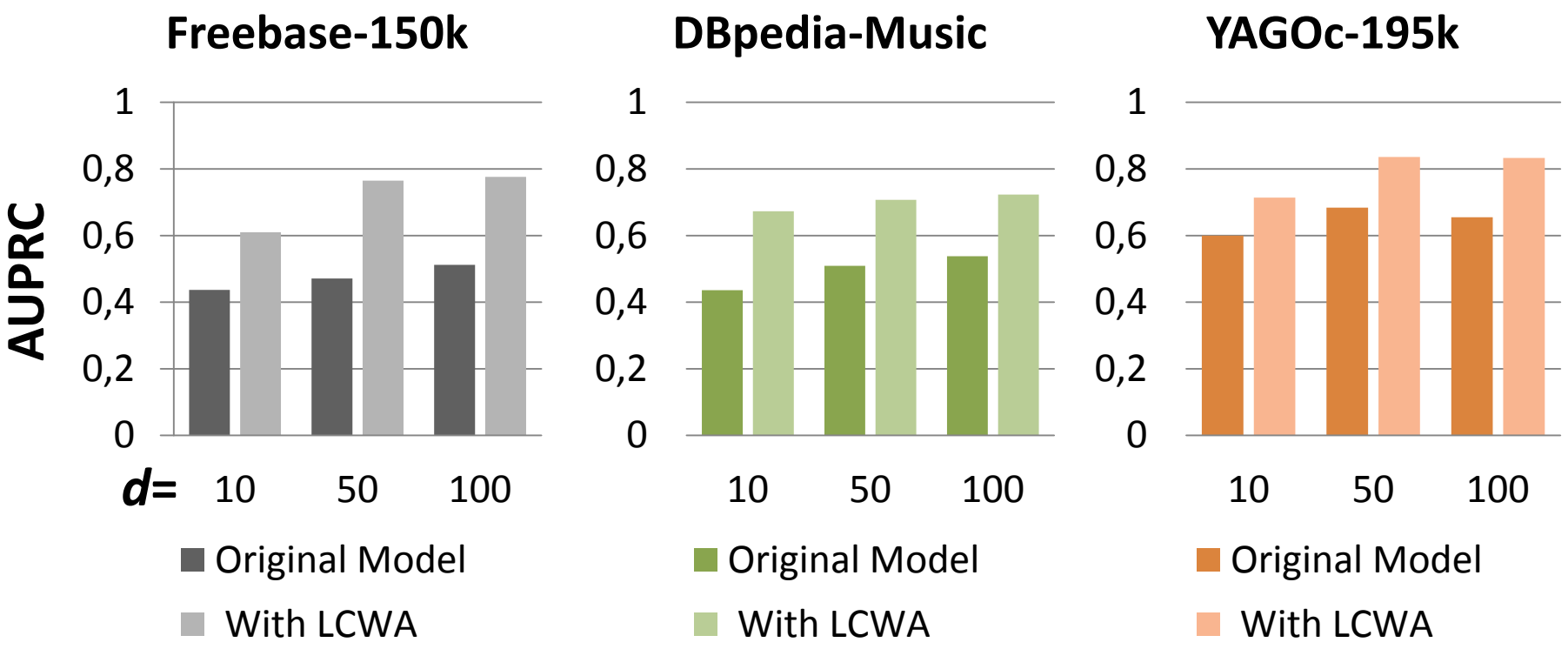
Results - LCWA

TransE



Results - LCWA

Google Knowledge Vault Neural Network



Conclusion

- Type-Constraints are essential when modeling large knowledge graphs with latent variable models
  - Up to 77% improvement in AUPRC
  - No negative impact on efficiency of algorithm
- A local closed-world assumption is a powerful but simple alternative in case type-constraints are absent or fuzzy
- Both approaches can be combined

Questions?

http://www.dbs.ifi.lmu.de/~krompass/

Python code and datasets will be available next week.

Denis.Krompass@siemens.com