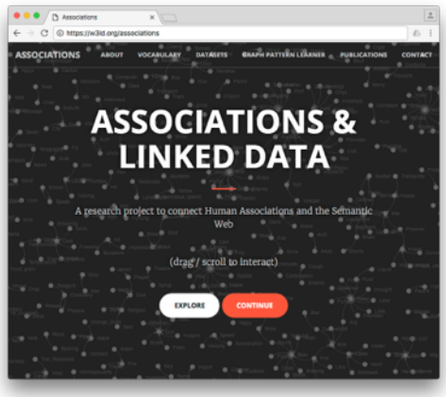
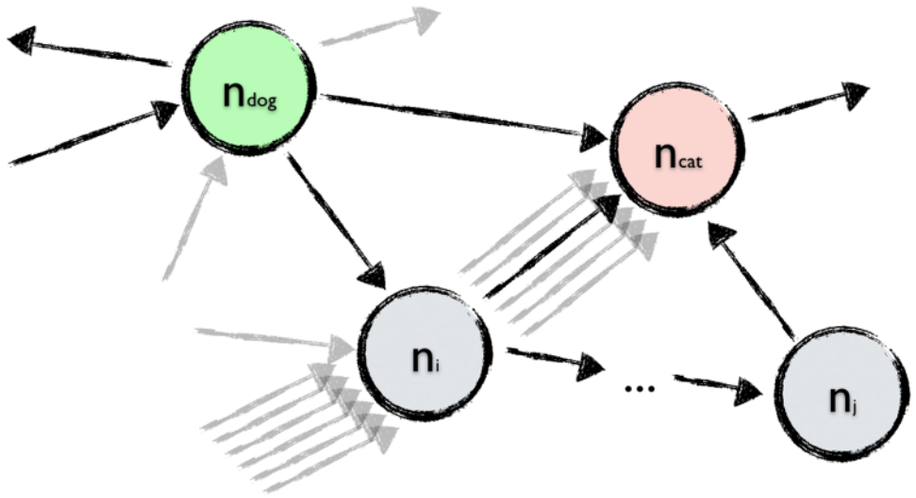
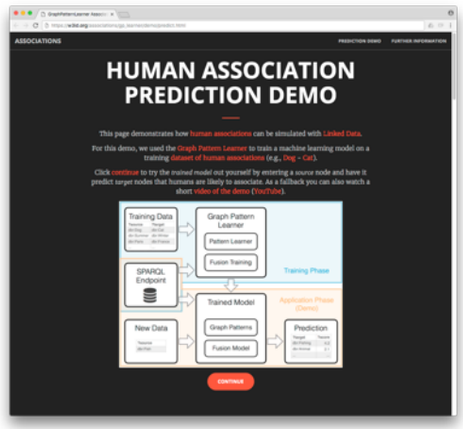
# Learning to Associate DBpedia Entities like Humans




Jörn Hees

2017-09-14
DBpedia Meetup Amsterdam

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Technische Universität Kaiserslautern

# Jörn Hees

- Researcher at DFKI & TU Kaiserslautern
- Linked Data, Machine Learning, AI
- An RDFLib maintainer (Python)
- http://joernhees.de , @joernhees
- https://w3id.org/associations

# What are Human Associations?

- Mental connections between concepts
- What's the first thing that comes to your mind when thinking about ... ?
- Example:
  - Dog



https://w3id.org/associations

# What are Human Associations?

- Mental connections between concepts
- What's the first thing that comes to your mind when thinking about ... ?
- Example:
  - Dog: Cat, collar, leash, walk, fur, bark

# What are Human Associations?

- Mental connections between concepts
- What's the first thing that comes to your mind when thinking about ... ?
- Example:
  - Dog: Cat, collar, leash, walk, fur, bark
  - House



https://w3id.org/associations

# What are Human Associations?

- Mental connections between concepts
- What's the first thing that comes to your mind when thinking about ... ?
- Example:
  - Dog: Cat, collar, leash, walk, fur, bark
  - House: Roof, door, window, flat, live

# Associations vs. Similarity

- Partially overlapping, but ≠
- Strongly Associated but not Similar:
  - Baby - Crying
- Similar but not Strongly Associated:
  - Dog - Terrier (100 ppl top answers: Cat (57 %), Collar (5 %), bark (2 %))



https://w3id.org/associations

# Outline

- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

# Outline

- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

# Motivation

- Associations are important for thinking:
  - Navigate from one thought to another
    - "Closeness of concepts in our mind"

Chris Welty's First Lady "Nixon" example

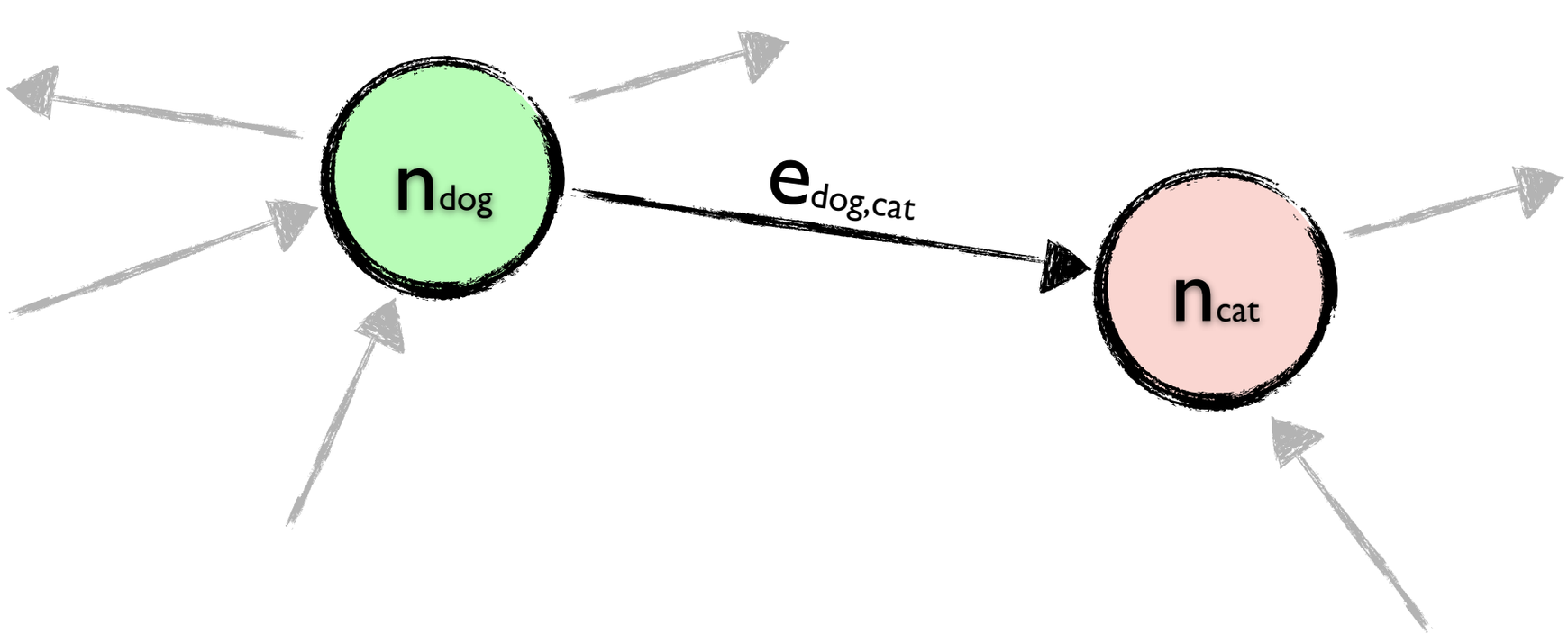
- Can we teach machines to do the same?
  - Using their Knowledge?
    - Linked Data



https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?

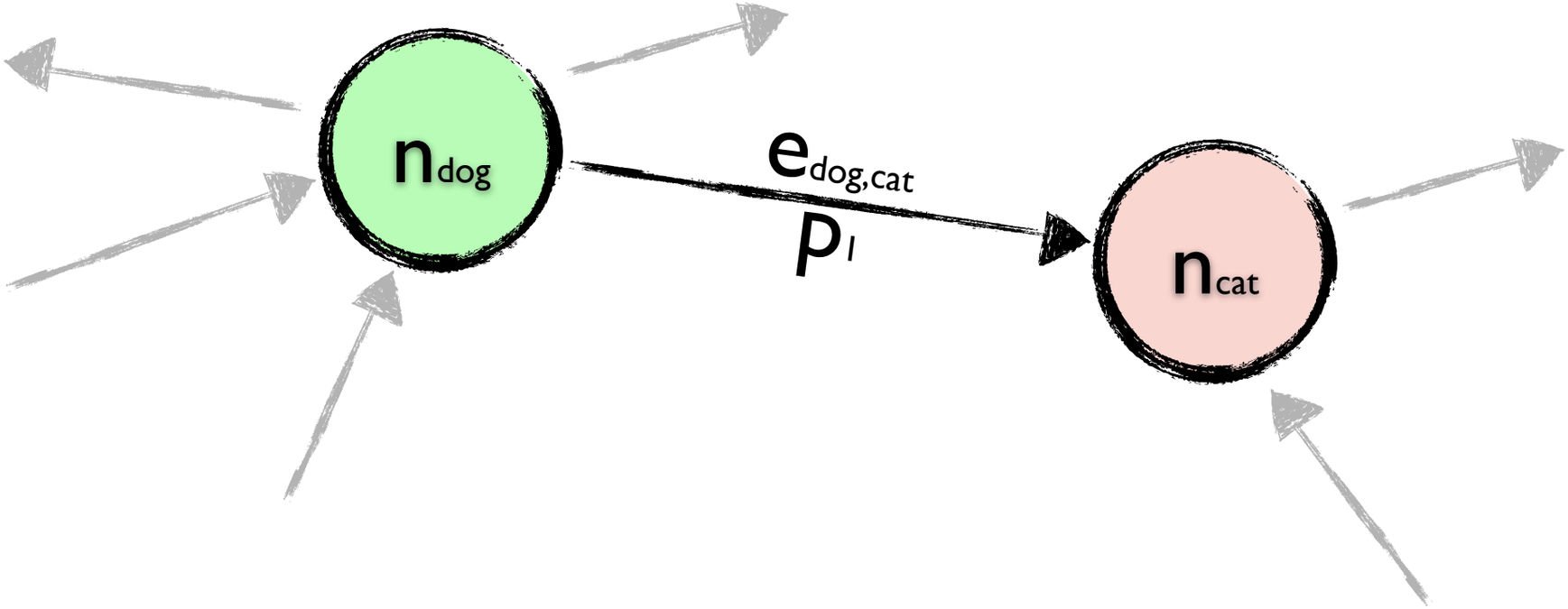





https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?

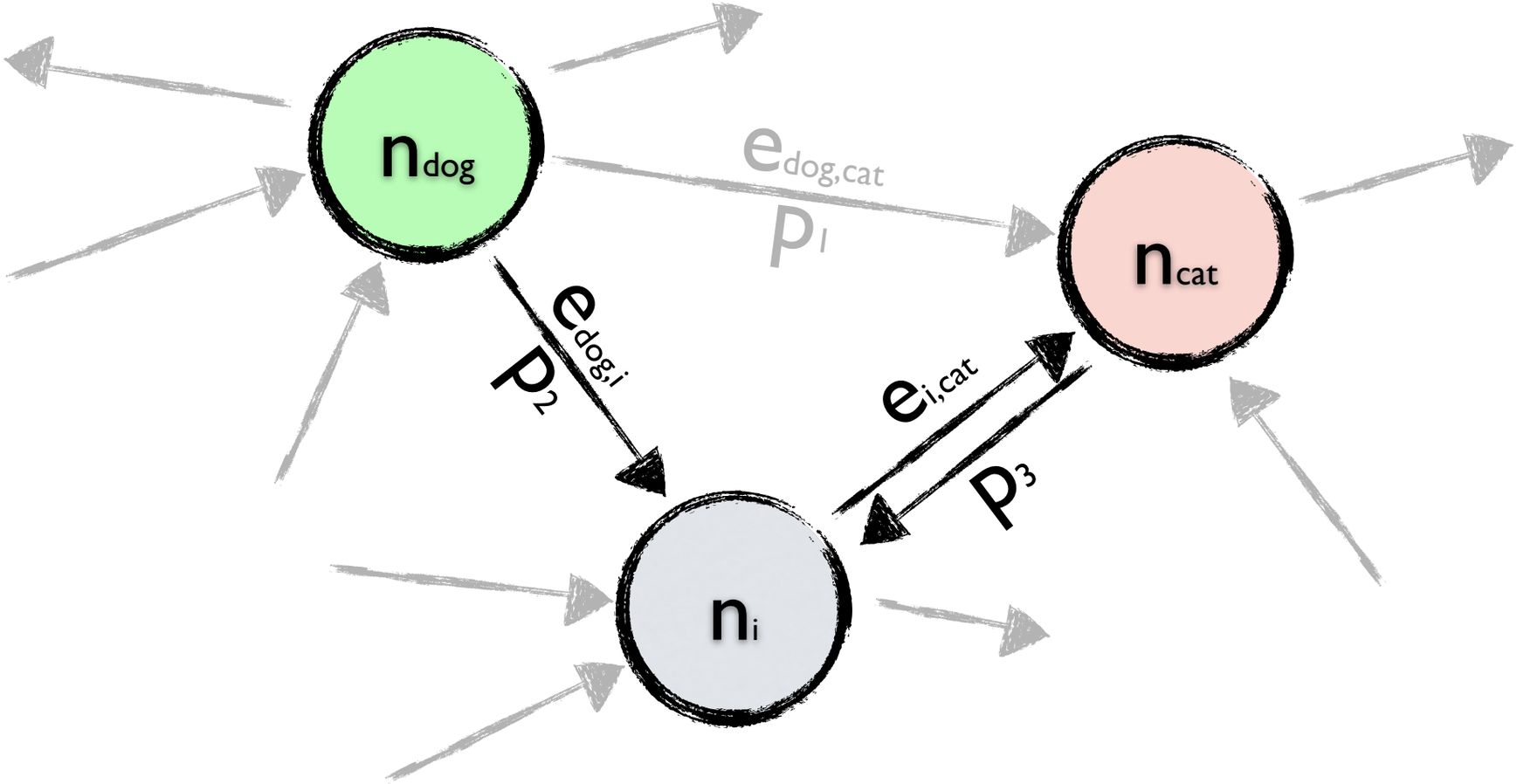





https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?

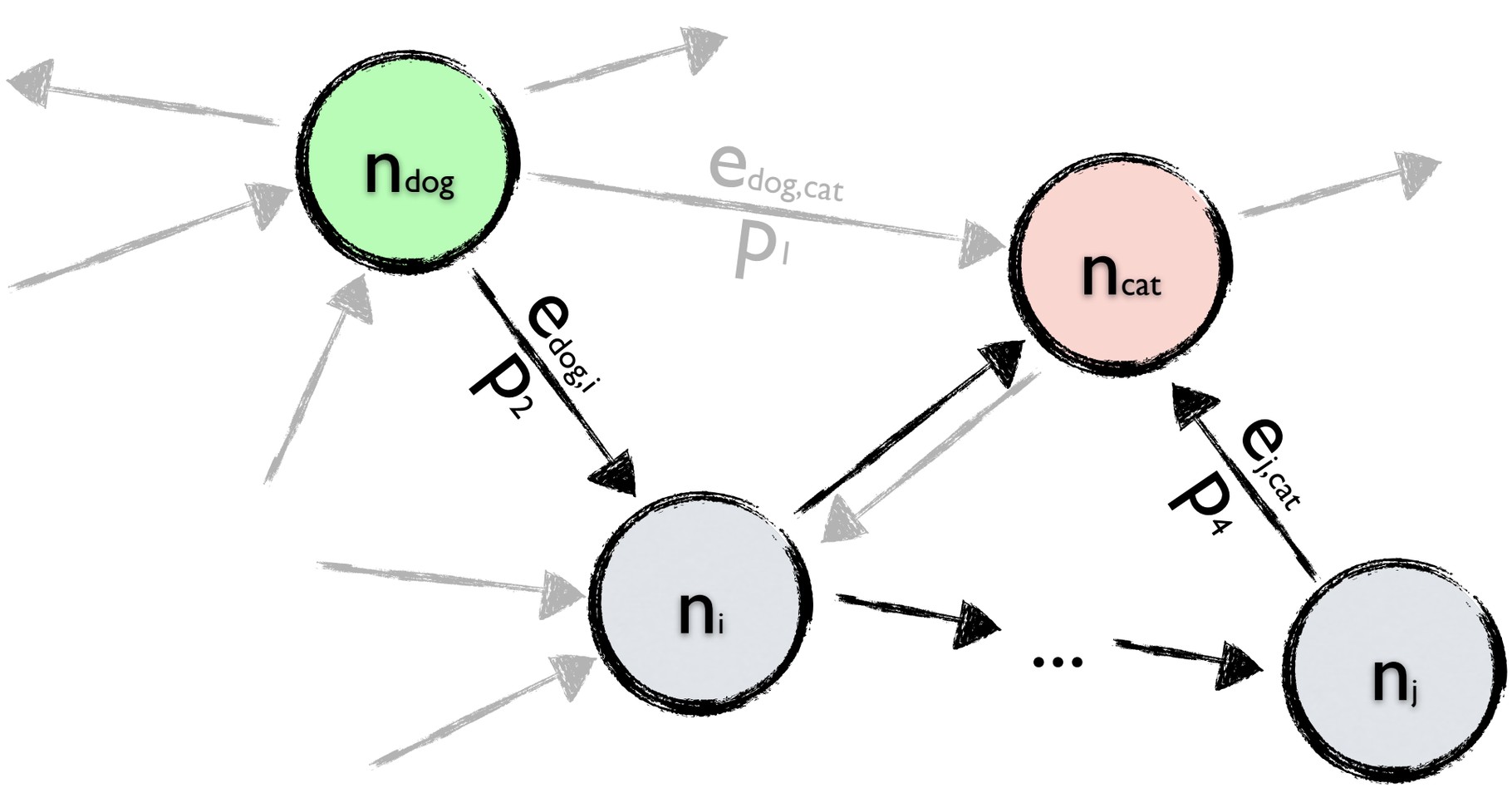





https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?






https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?
- Goal:
  - Given an input node

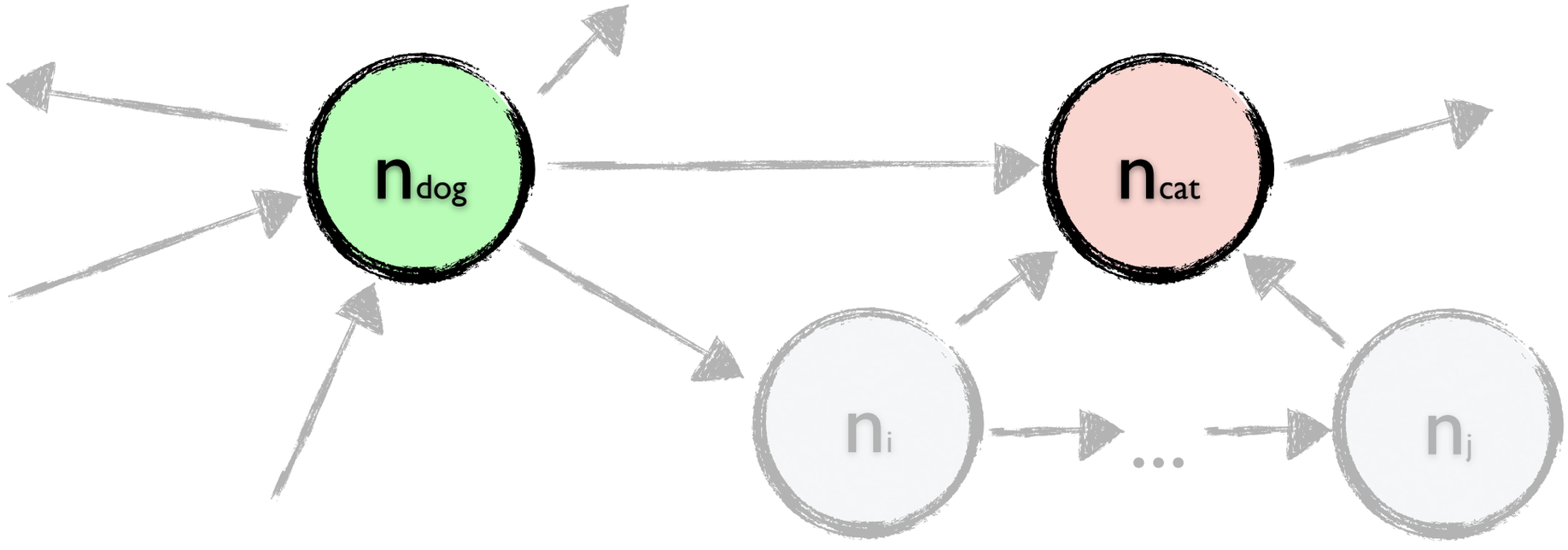

predict the output node(s) we would associate




https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?
- Dataset of "Semantic Associations" needed

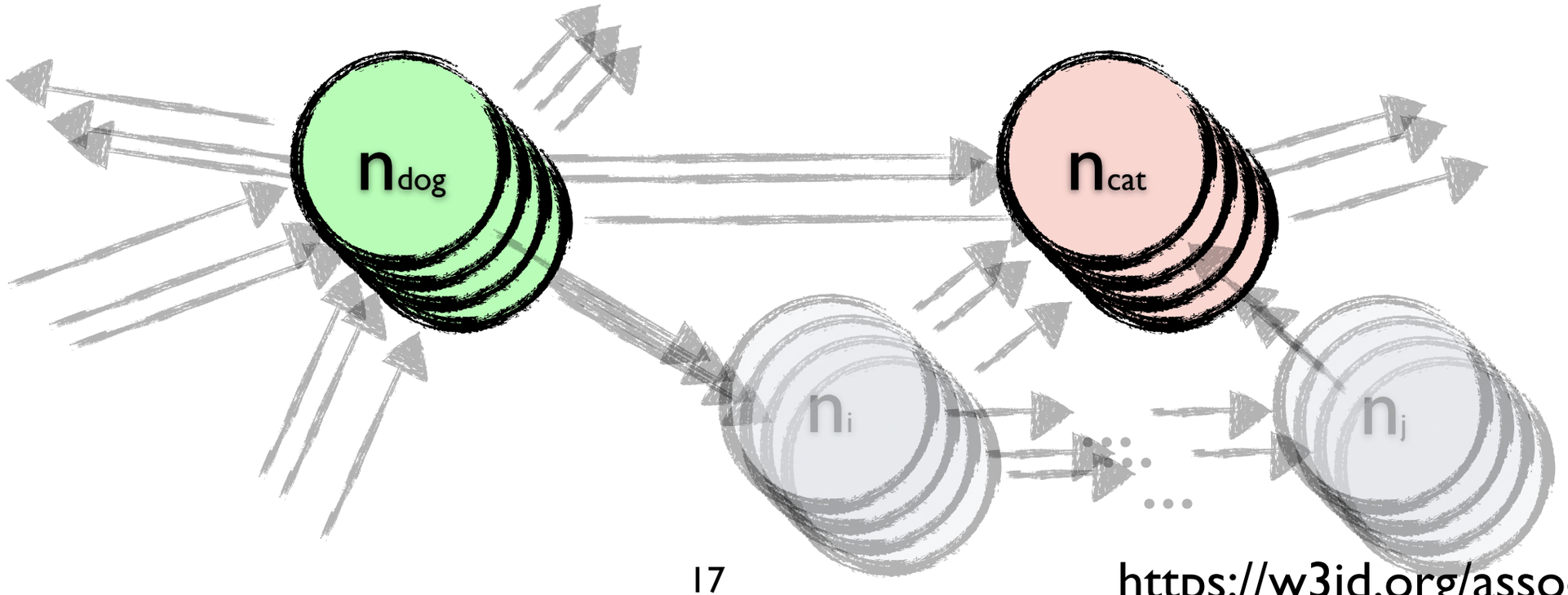





https://w3id.org/associations

# My Research

- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?
- Dataset of "Semantic Associations" needed



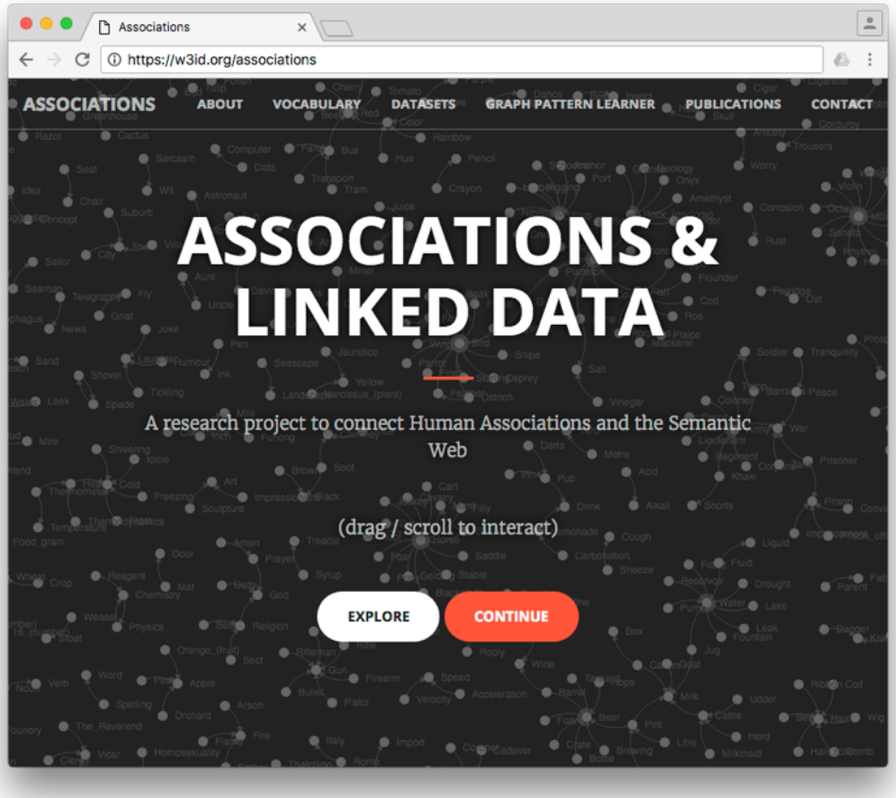


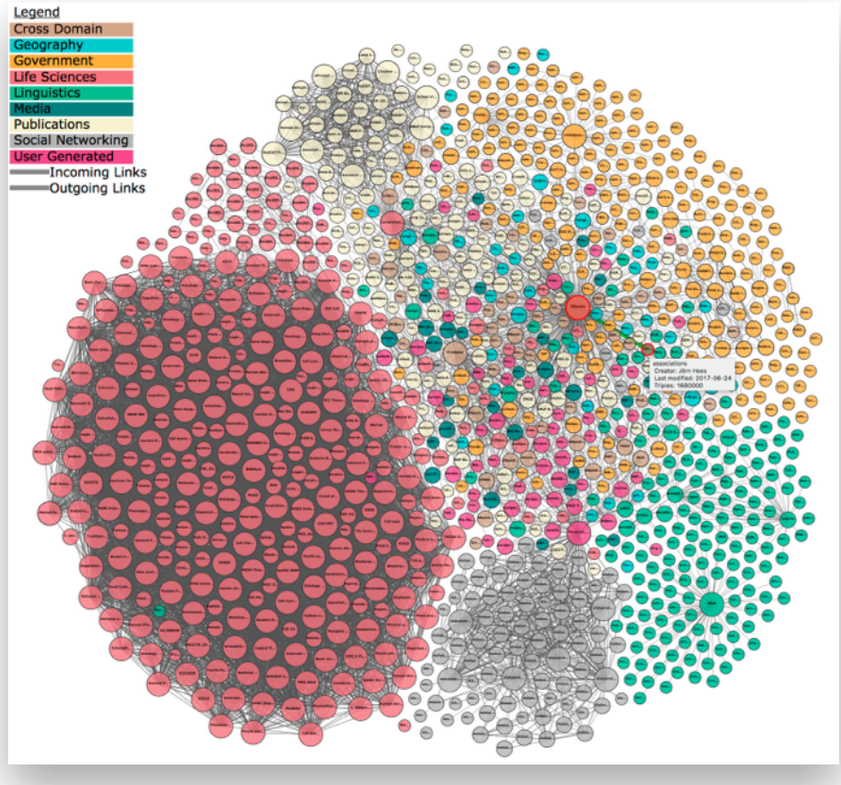

https://w3id.org/associations

# Semantic Associations Dataset

- (Raw) Edinburgh Associative Thesaurus (EAT) as RDF (1.7 M triples)

- 727 verified distinct Semantic Associations






https://w3id.org/associations

Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
Incoming Links
Outgoing Links

associations
Creator: Jörn Hees
Last modified: 2017-06-24
Triples: 1680000

Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
Incoming Links
Outgoing Links

associations
Creator: Jörn Hees
Last modified: 2017-06-24
Triples: 1680000

# Semantic Associations Dataset

- 727 verified distinct Semantic Associations

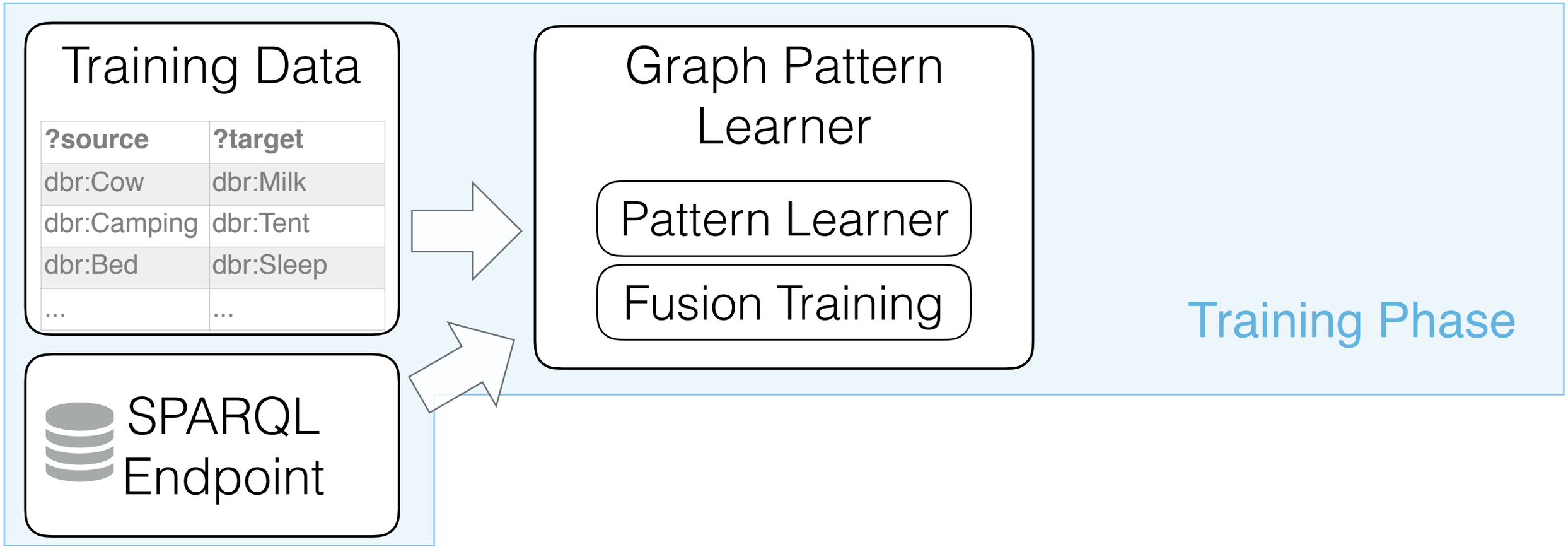
| Stimulus | Response |
|---|---|
| dbr:Cow | dbr:Milk |
| dbr:Camping | dbr:Tent |
| dbr:Expense | dbr:Money |
| dbr:Bed | dbr:Sleep |
| dbr:Pupil | dbr:Eye |
| … | … |

- Not readily modelled in DBpedia!
- Not one property!



https://w3id.org/associations

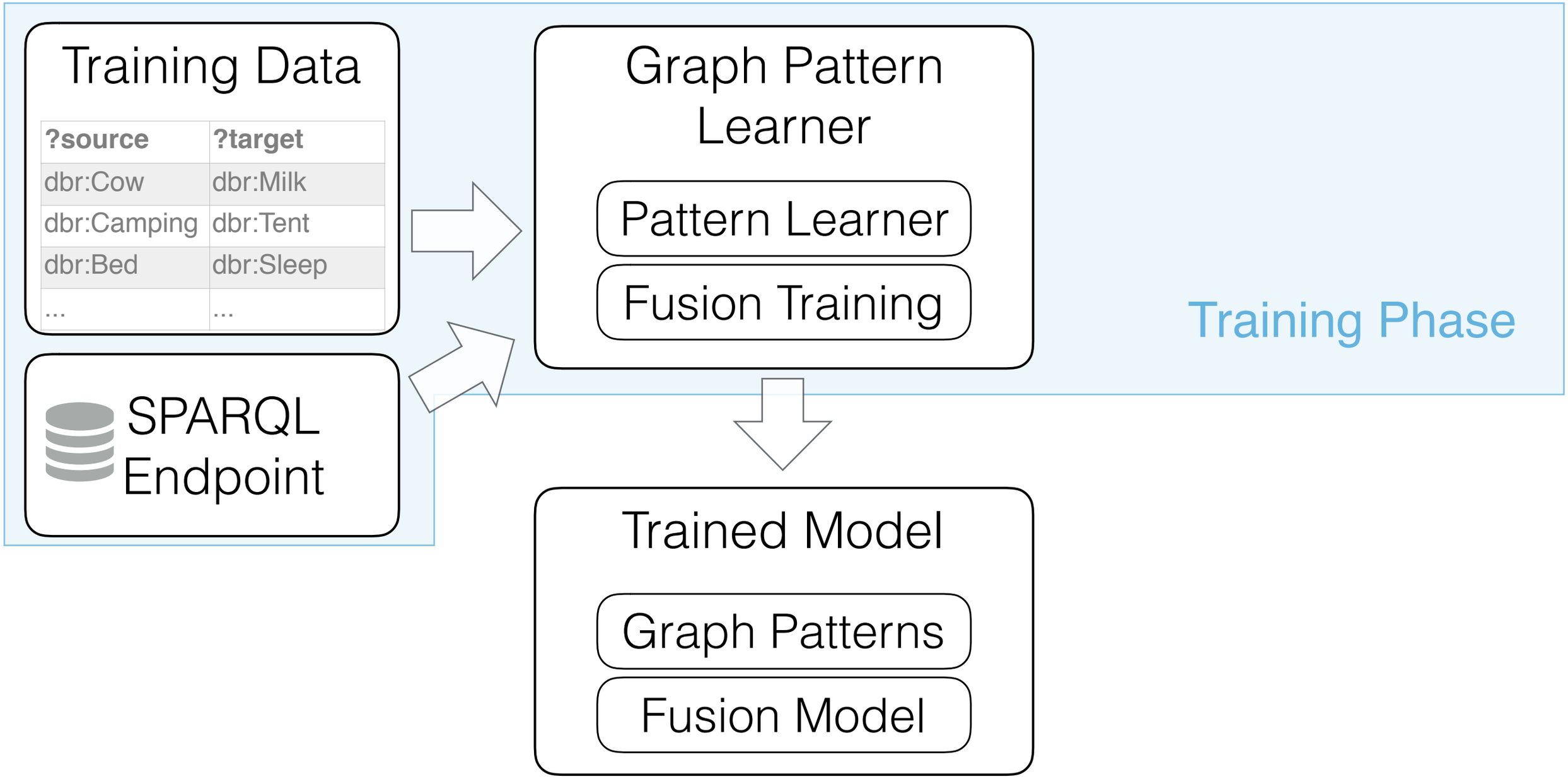
# Machine Learning Outline



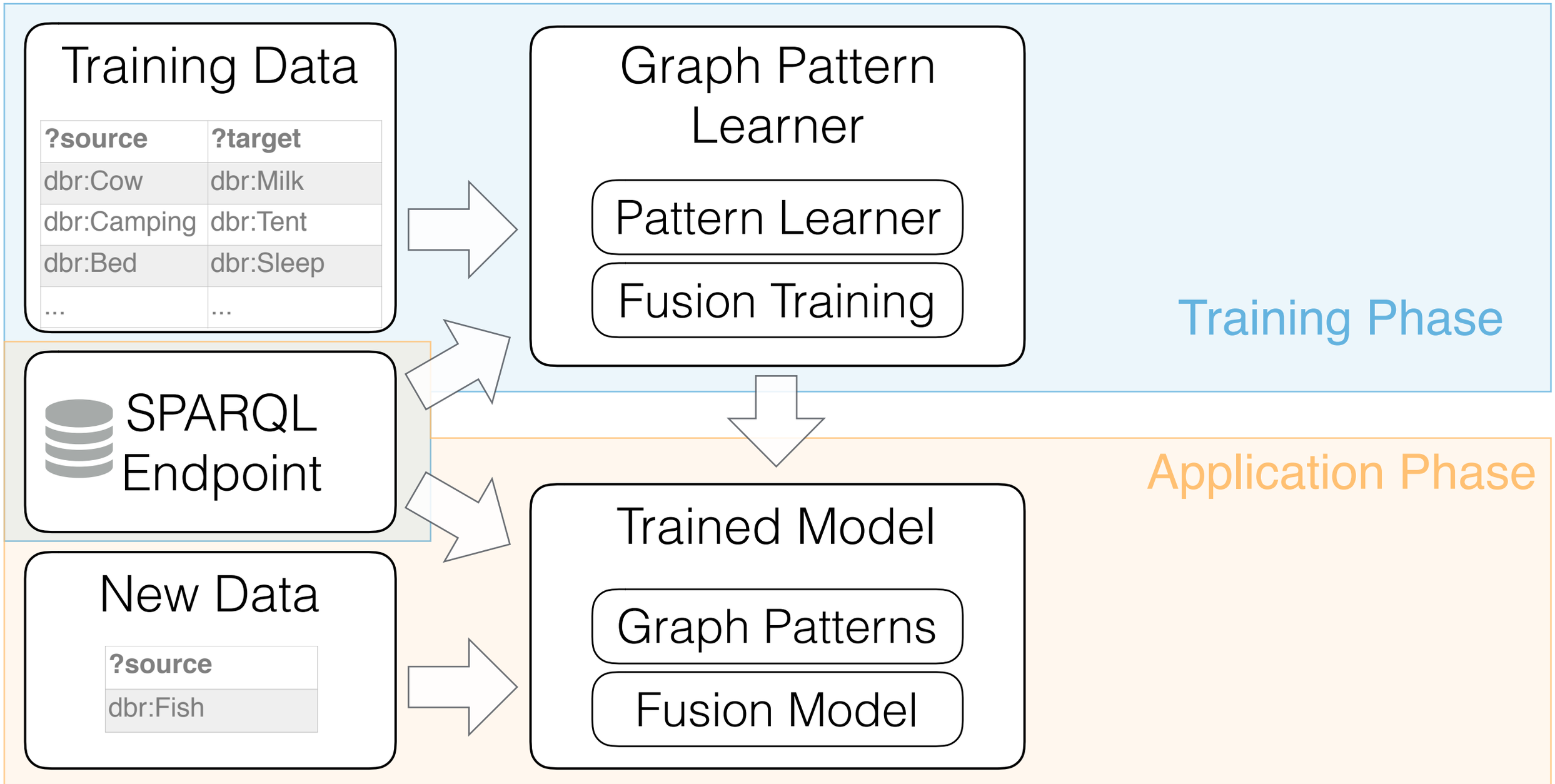



https://w3id.org/associations

# Machine Learning Outline



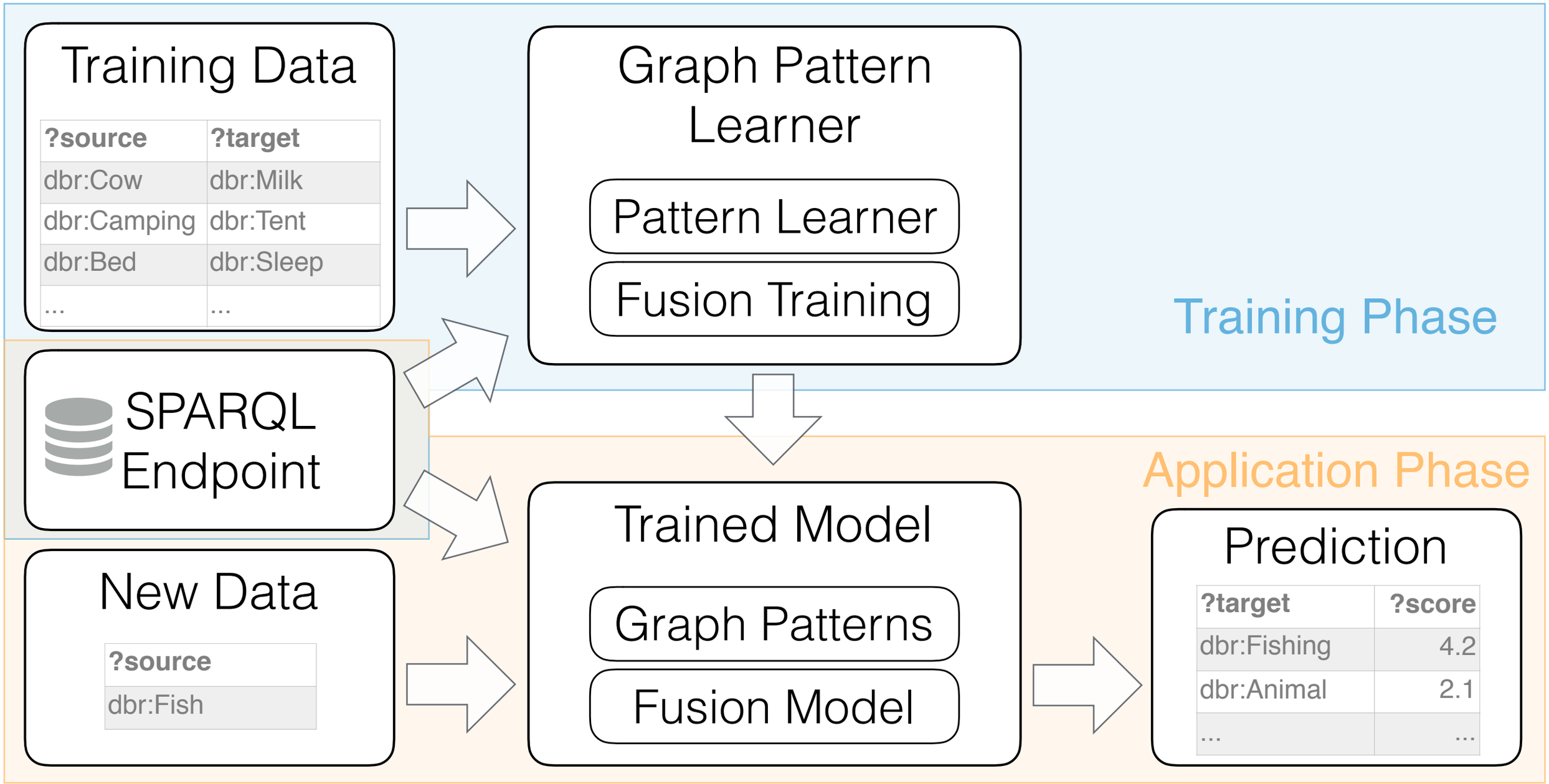
# Machine Learning Outline






https://w3id.org/associations

# Machine Learning Outline






https://w3id.org/associations

# Outline

- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

# Outline

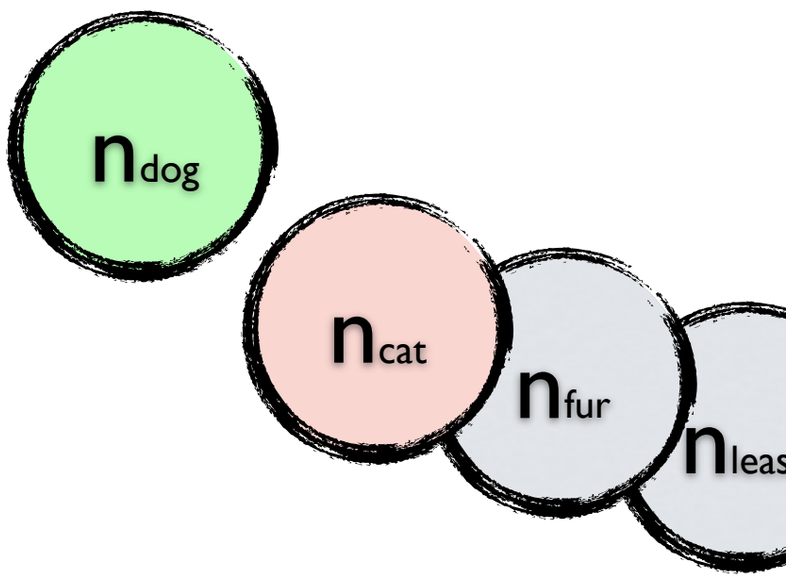
- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

# Evaluation

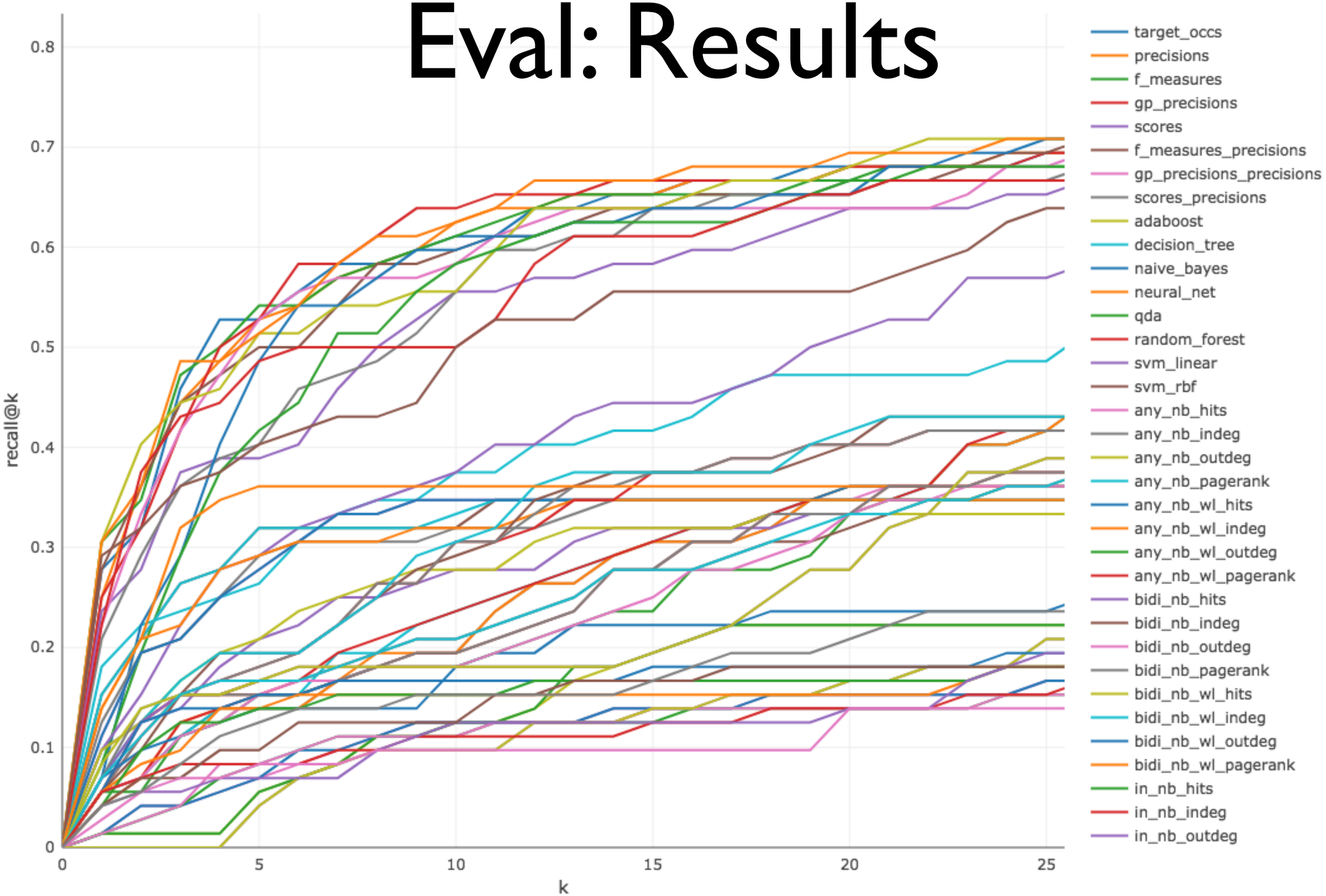
- How good are the predictions?
  - Training/Test set split
  - Given a stimulus from the test set, what's the rank of the true response in the prediction results?






https://w3id.org/associations

# Eval: Results

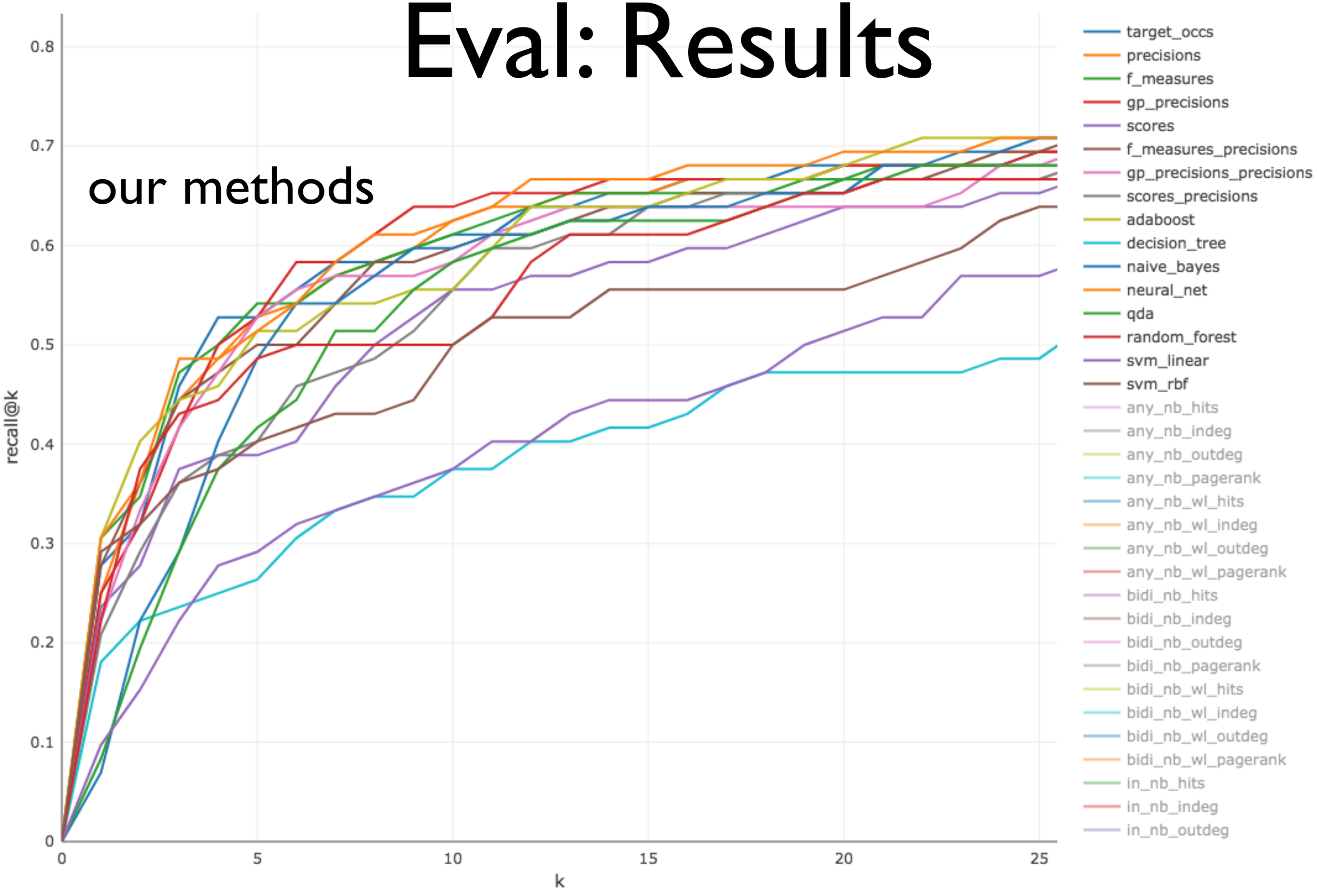





https://w3id.org/associations

# Eval: Results

our methods

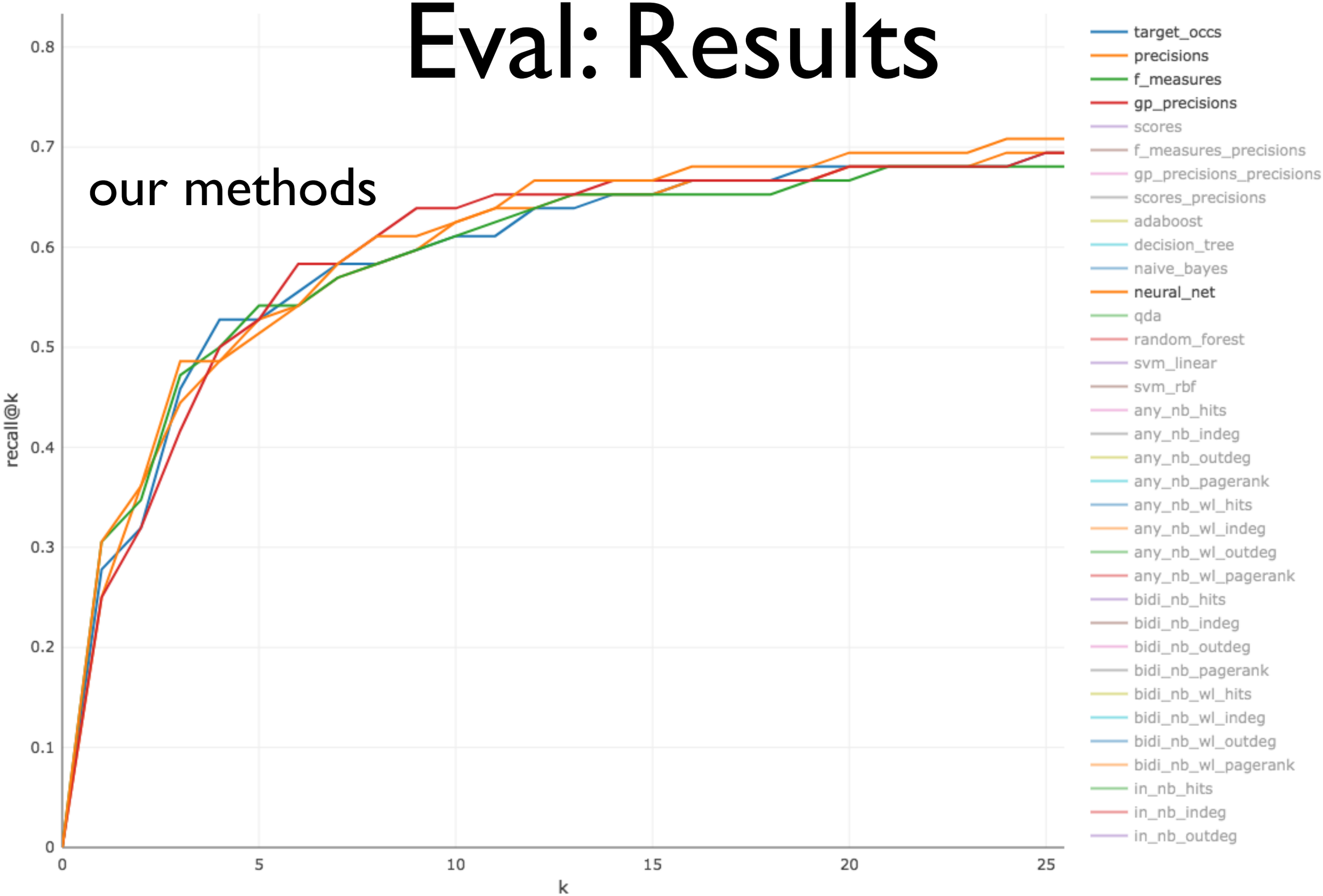





https://w3id.org/associations

# Eval: Results

our methods



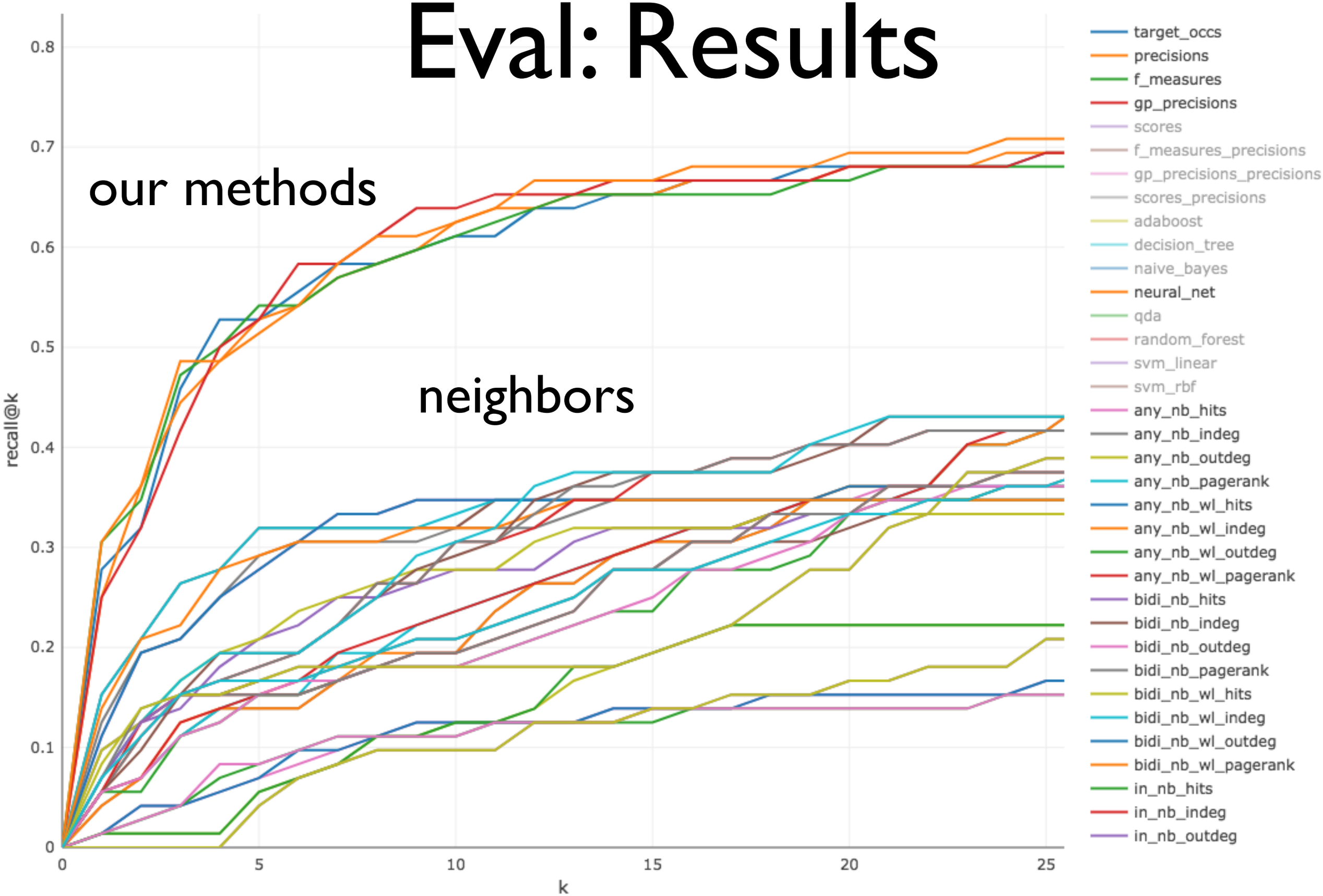



https://w3id.org/associations

# Eval: Results

our methods

neighbors



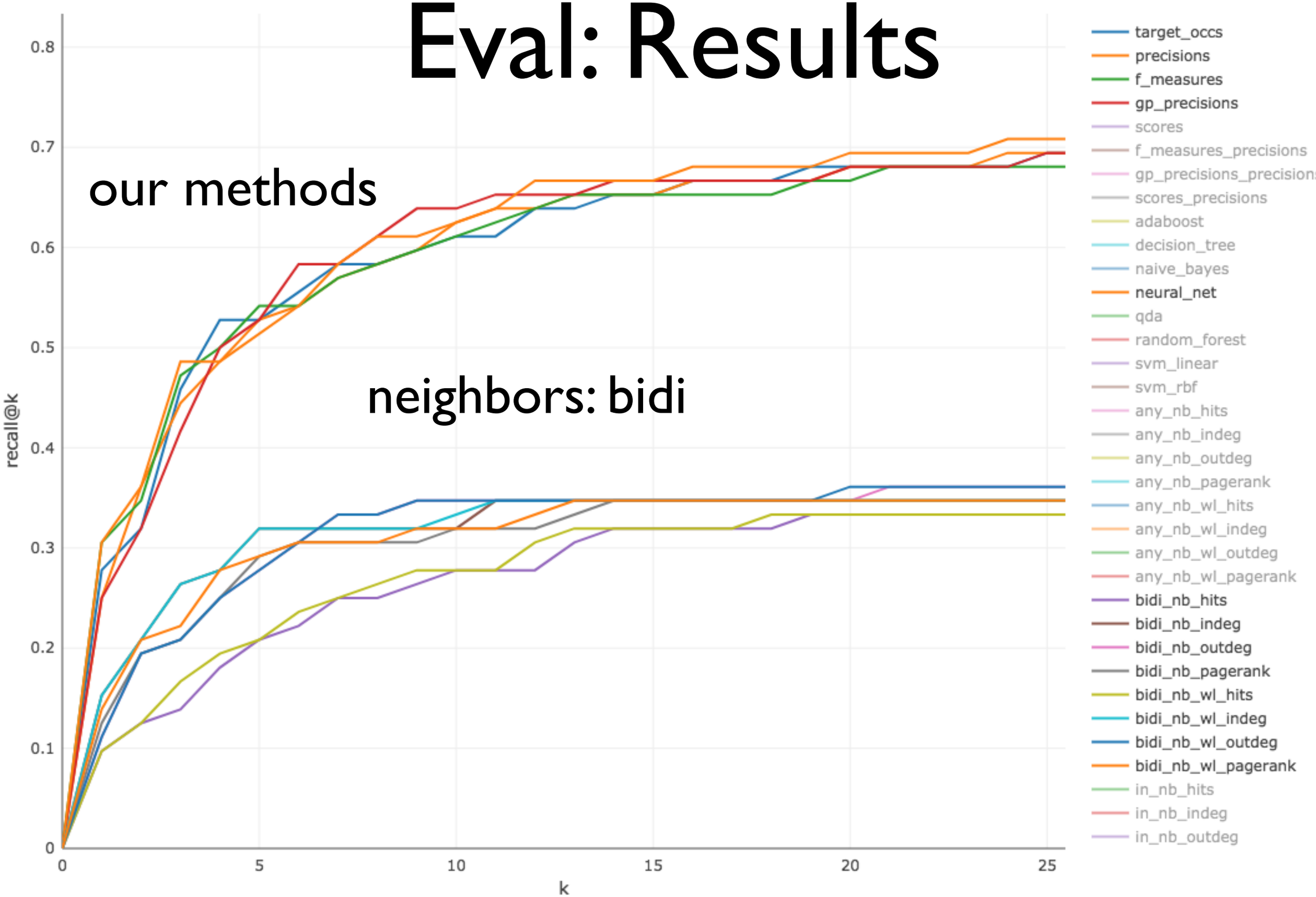



https://w3id.org/associations

# Eval: Results

our methods

neighbors: bidi



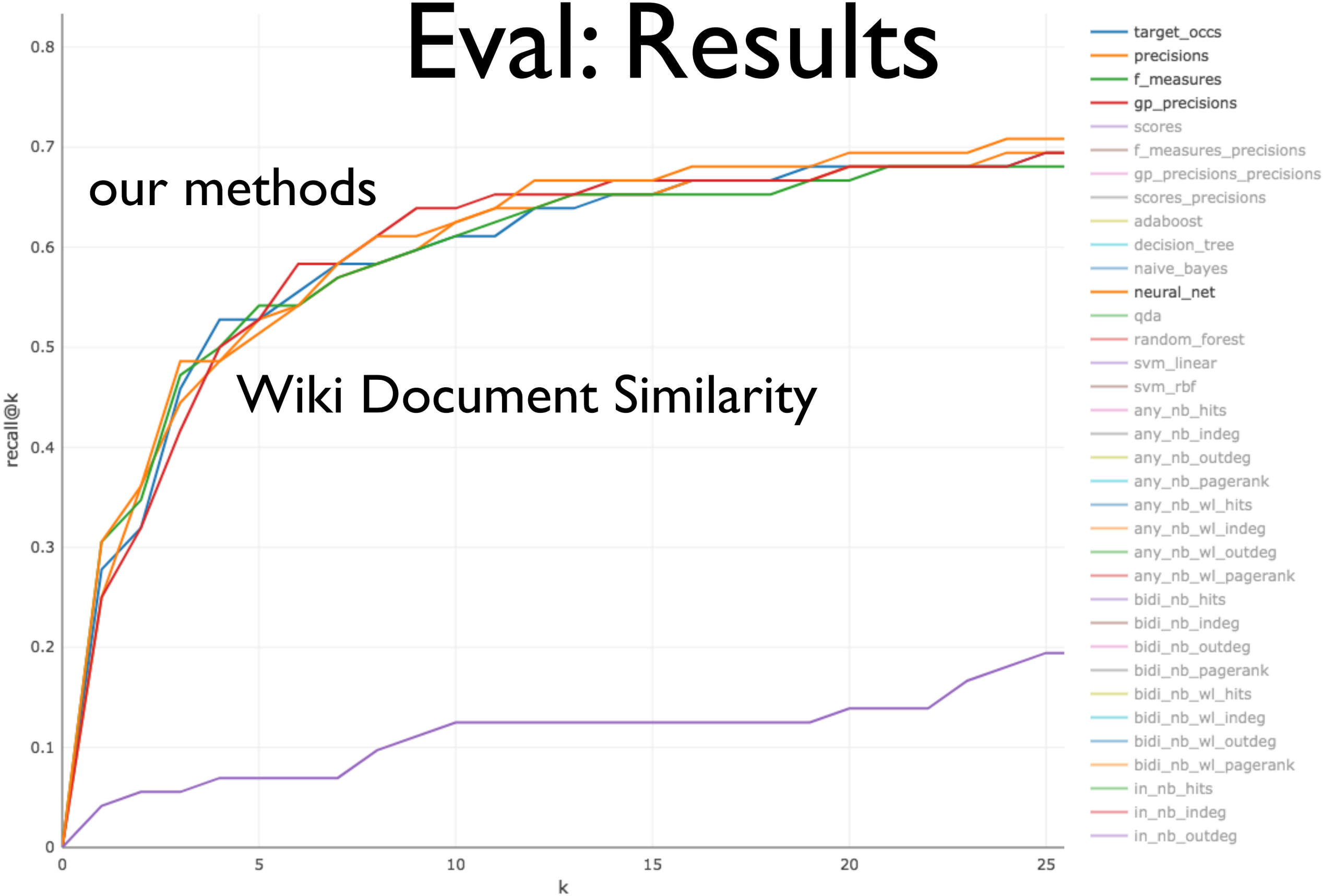



https://w3id.org/associations

# Eval: Results

our methods

Wiki Document Similarity






https://w3id.org/associations

# Eval: Results

our methods

Milne-Witten Similarity

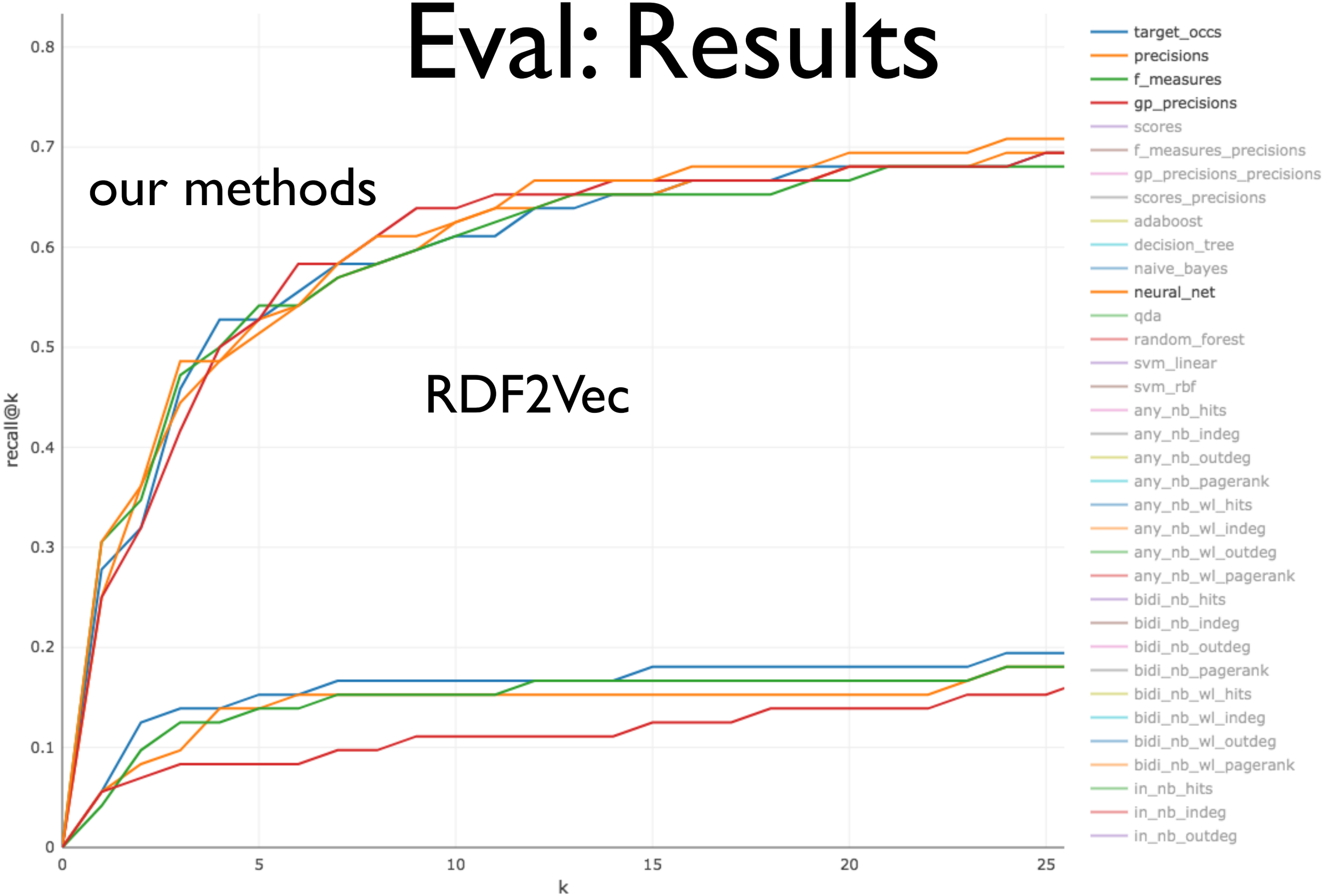



https://w3id.org/associations

# Eval: Results

our methods

RDF2Vec



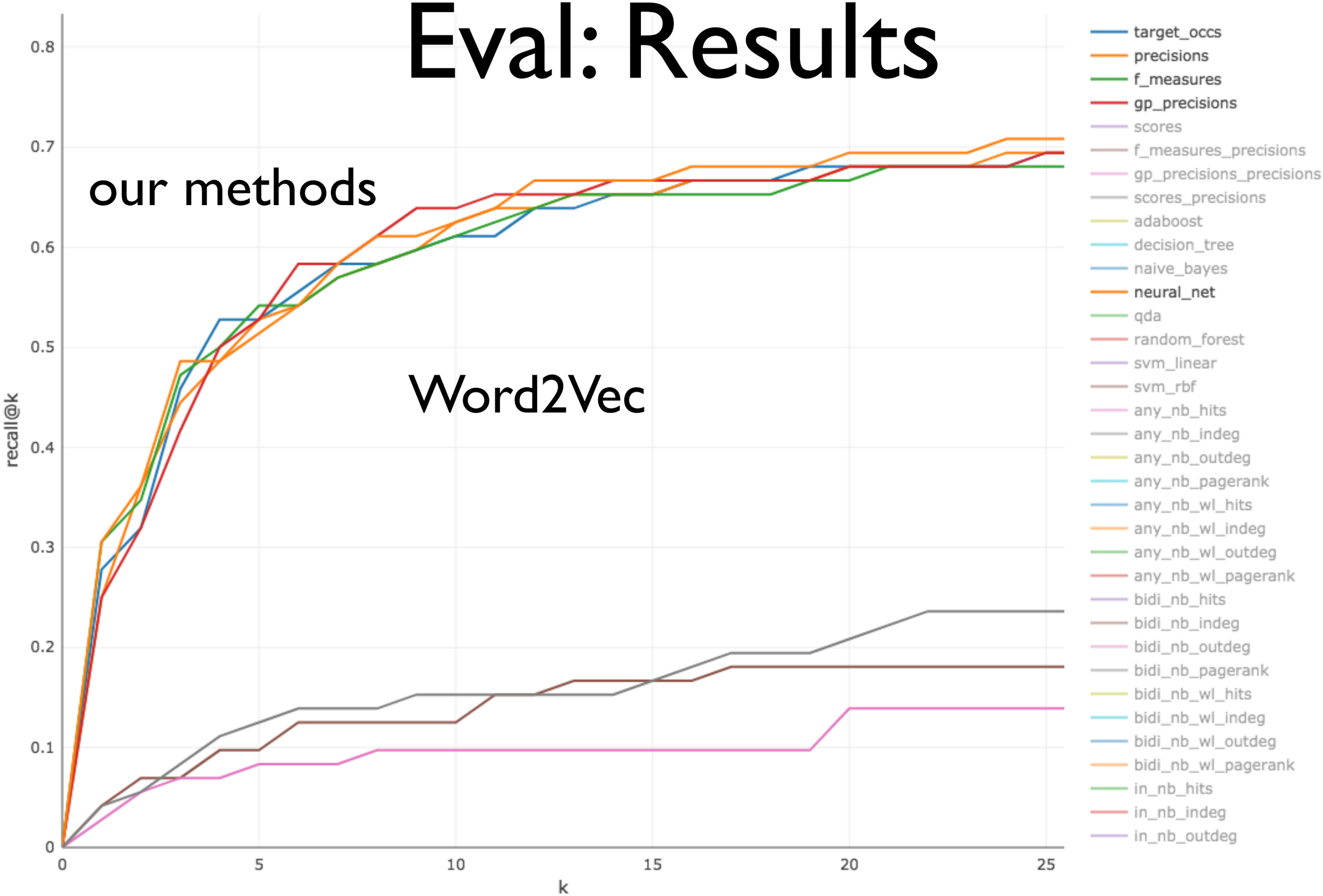



https://w3id.org/associations

# Eval: Results

our methods

Word2Vec






https://w3id.org/associations

# Evaluation Results

| Method | Rec@1 | Rec@2 | Rec@3 | Rec@5 | Rec@10 | MAP | NDCG |
|---|---|---|---|---|---|---|---|
| DocSim | 4.2% | 5.6% | 5.6% | 6.9% | 12.5% | 6.6% | 12.5% |
| Word2Vec | 4.2% | 5.6% | 8.3% | 12.5% | 15.3% | 7.8% | 12.5% |
| RDF2Vec | 5.6% | 12.5% | 13.9% | 15.3% | 16.7% | 10.3% | 14.4% |
| MW | 6.9% | 9.7% | 11.1% | 13.9% | 18.1% | 11.0% | 17.9% |
| NB Bidi WL PR | 13.9% | 20.8% | 22.2% | 29.2% | 31.9% | 20.2% | 23.8% |
| NB Bidi WL InDeg | **15.3%** | **20.8%** | **26.4%** | **31.9%** | **33.3%** | **21.4%** | **24.8%** |
| gpl + precisions | 25.0% | 36.1% | 44.4% | 52.8% | 62.5% | 37.1% | 46.0% |
| gpl + neural net | **30.6%** | **36.1%** | **48.6%** | **51.4%** | **62.5%** | **40.3%** | **48.3%** |



https://w3id.org/associations

# Evaluation Results

| Method | Rec@1 | Rec@2 | Rec@3 | Rec@5 | Rec@10 | MAP | NDCG |
|---|---|---|---|---|---|---|---|
| DocSim | 4.2% | 5.6% | 5.6% | 6.9% | 12.5% | 6.6% | 12.5% |
| Word2Vec | 4.2% | 5.6% | 8.3% | 12.5% | 15.3% | 7.8% | 12.5% |
| RDF2Vec | 5.6% | 12.5% | 13.9% | 15.3% | 16.7% | 10.3% | 14.4% |
| MW | 6.9% | 9.7% | 11.1% | 13.9% | 18.1% | 11.0% | 17.9% |
| NB Bidi WL PR | 13.9% | 20.8% | 22.2% | 29.2% | 31.9% | 20.2% | 23.8% |
| NB Bidi WL InDeg | **15.3%** | **20.8%** | **26.4%** | **31.9%** | **33.3%** | **21.4%** | **24.8%** |
| gpl + precisions | 25.0% | 36.1% | 44.4% | 52.8% | 62.5% | 37.1% | 46.0% |
| gpl + neural net | **30.6%** | **36.1%** | **48.6%** | **51.4%** | **62.5%** | **40.3%** | **48.3%** |

- Avg. Inter-Human Agreement: ~ 32 %



https://w3id.org/associations

# Outline

- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

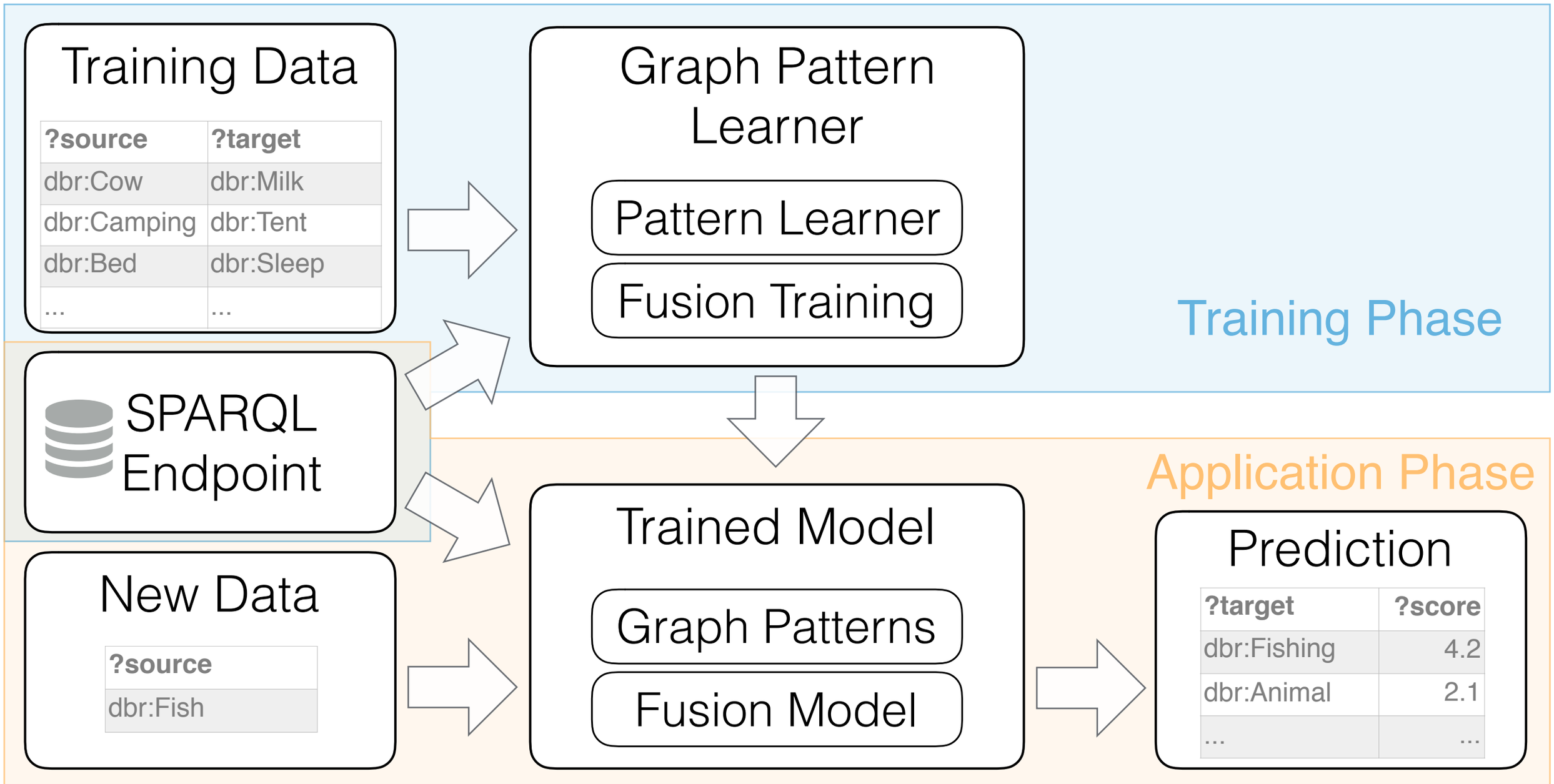
# Outline

- Background
- My Research
- Evaluation
- Demo



https://w3id.org/associations

# Machine Learning Outline

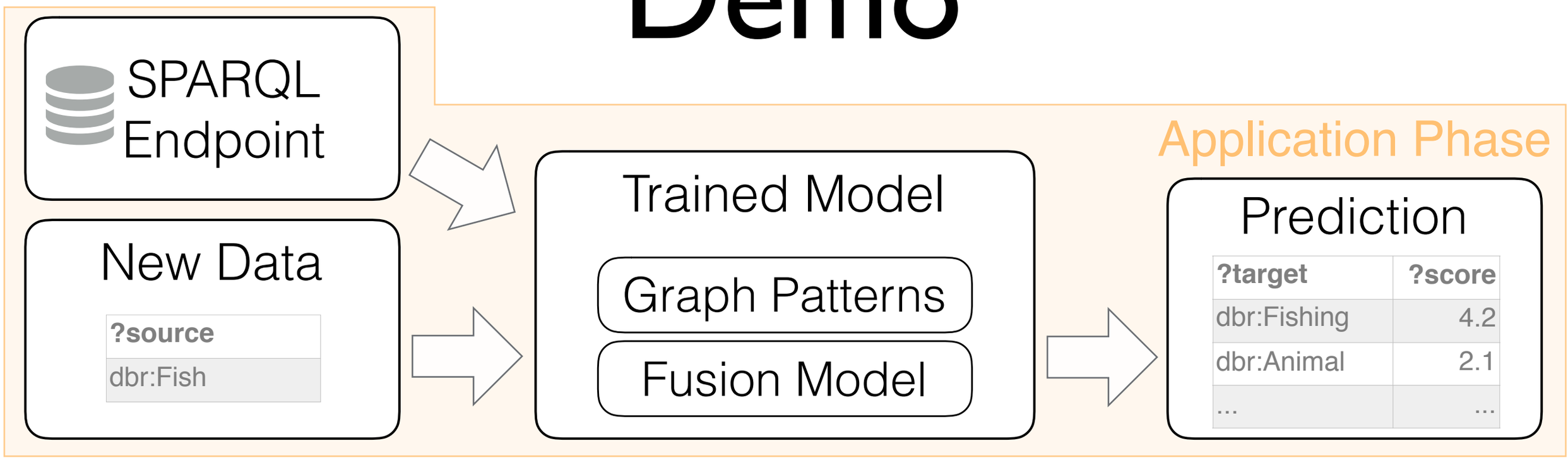





https://w3id.org/associations

# Machine Learning Outline

## Demo






https://w3id.org/associations

# Demo

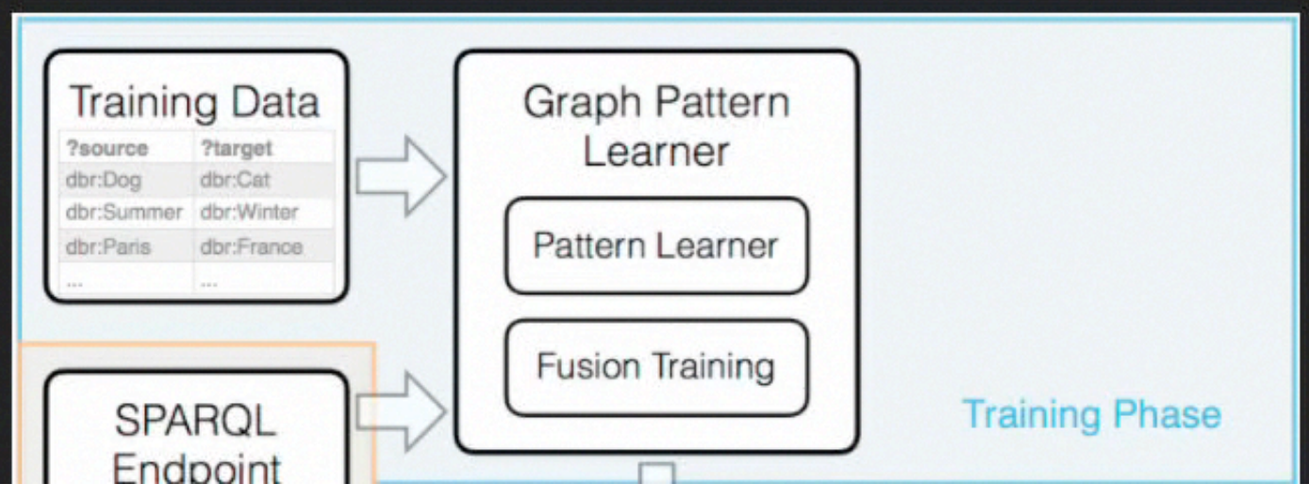
ASSOCIATIONS PREDICTION DEMO FURTHER INFORMATION

## HUMAN ASSOCIATION PREDICTION DEMO

This page demonstrates how human associations can be simulated with Linked Data.

For this demo, we used the Graph Pattern Learner to train a machine learning model on a training dataset of human associations (e.g., Dog – Cat).

Click continue to try the *trained model* out yourself by entering a *source* node and have it predict *target* nodes that humans are likely to associate. As a fallback you can also watch a short video of the demo (YouTube).




https://w3id.org/associations#demos

# Demo

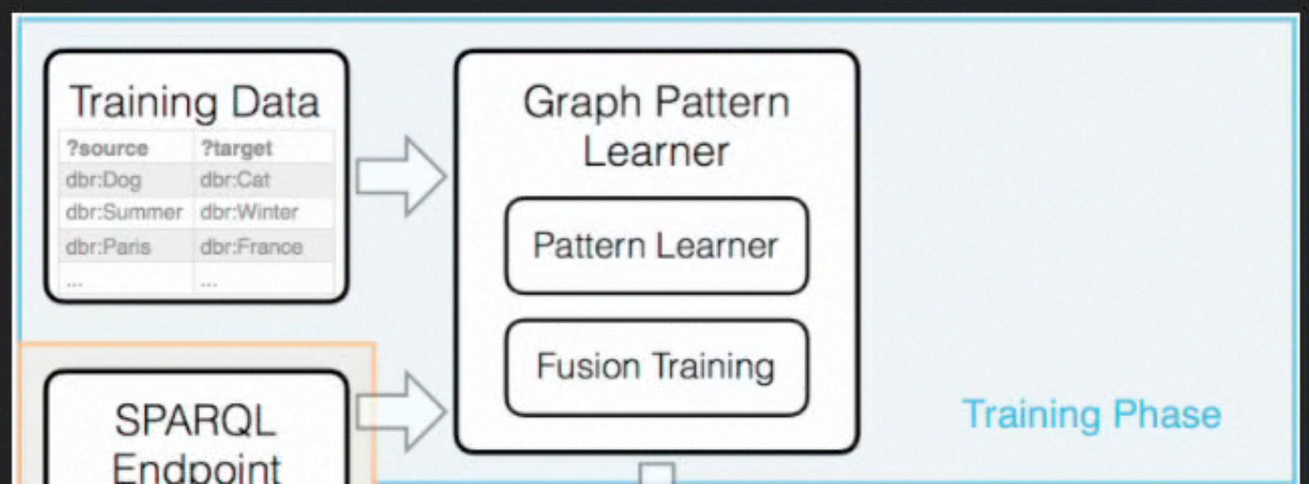
ASSOCIATIONS PREDICTION DEMO FURTHER INFORMATION

## HUMAN ASSOCIATION PREDICTION DEMO

This page demonstrates how human associations can be simulated with Linked Data.

For this demo, we used the Graph Pattern Learner to train a machine learning model on a training dataset of human associations (e.g., Dog – Cat).

Click continue to try the *trained model* out yourself by entering a *source* node and have it predict *target* nodes that humans are likely to associate. As a fallback you can also watch a short video of the demo (YouTube).




https://w3id.org/associations#demos

# Other Applications

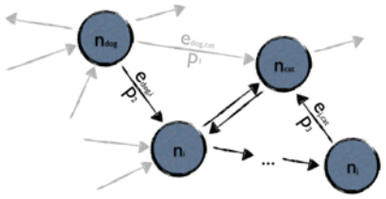
- TasteDive (Recommendation Engine) Books
  - ~ 50 % Recall@10
- DBpediaNYD
  - ~ 63 % Recall@10



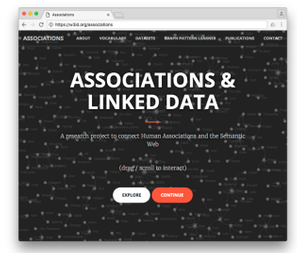
https://w3id.org/associations

# Summary

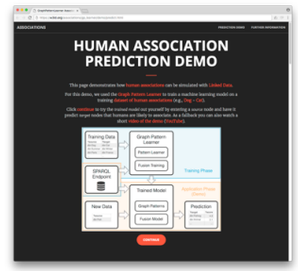
- Goal
  Learning Graph Patterns for Associations

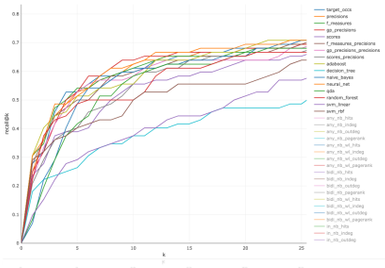



- Semantic Association Dataset




- Graph Pattern Learner
  Learns SPARQL Patterns for Source-Target-Pairs
  Demo



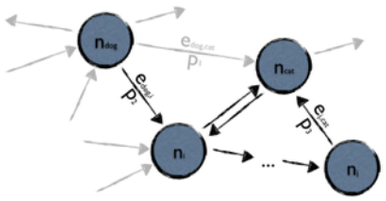



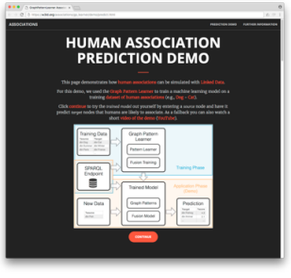


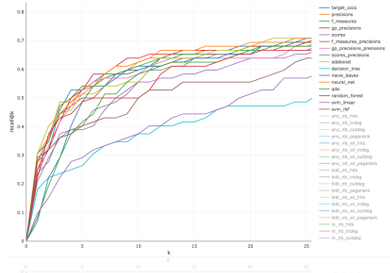

https://w3id.org/associations

# Future Work

- Apply Evolutionary Algorithm
  - to other datasets
  - to other problems
- Extensions:
  - Literals
  - LOD-a-lot & #LD



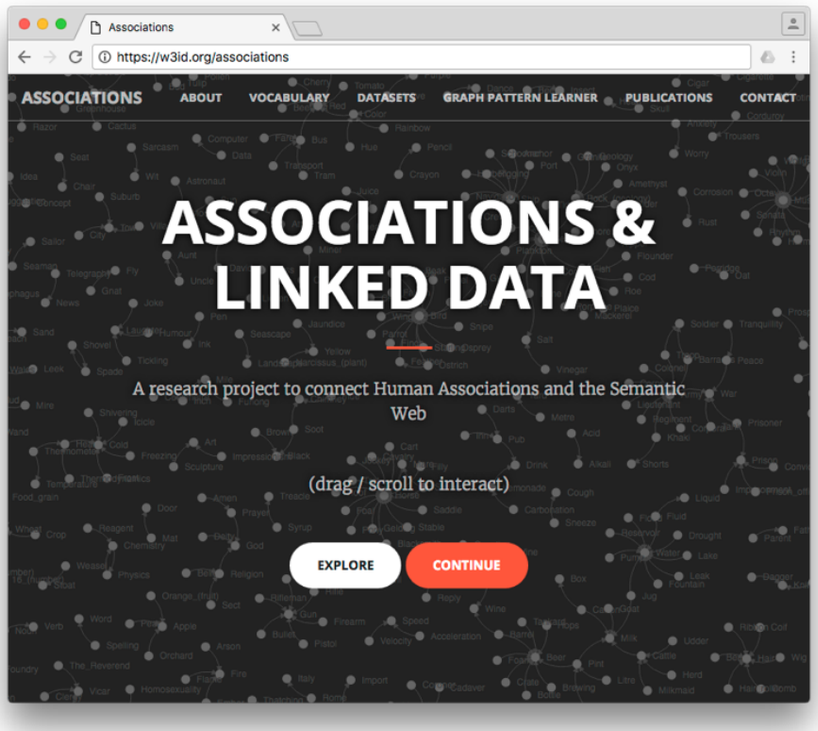


https://w3id.org/associations

# Discussion

Thanks for your attention

Questions?




https://w3id.org/associations