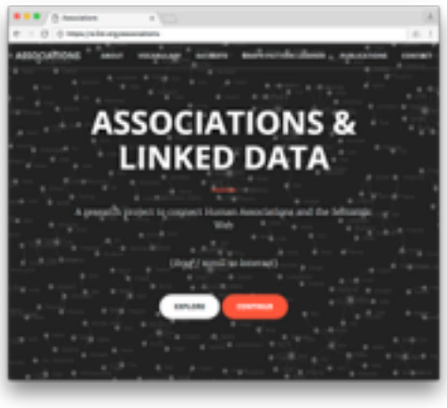
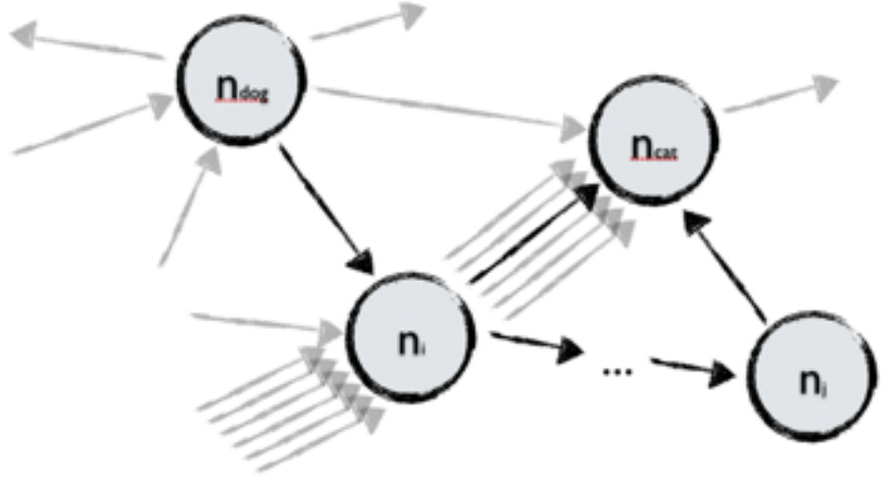
An Evolutionary Algorithm to Learn SPARQL Queries for Source-Target-Pairs

Finding Patterns for Human Associations in DBpedia





Jörn Hees

2016-11-20
EKAW 2016

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Technische Universität Kaiserslautern

Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

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?

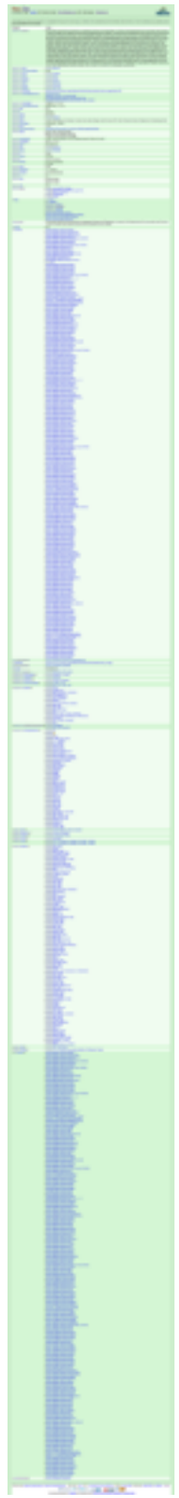
- Mental connections between concepts
- What's the first thing that comes to your mind when thinking about ... ?
- Example:
  - Dog: Cat, collar, leash, walk, fur, bark
  - House

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

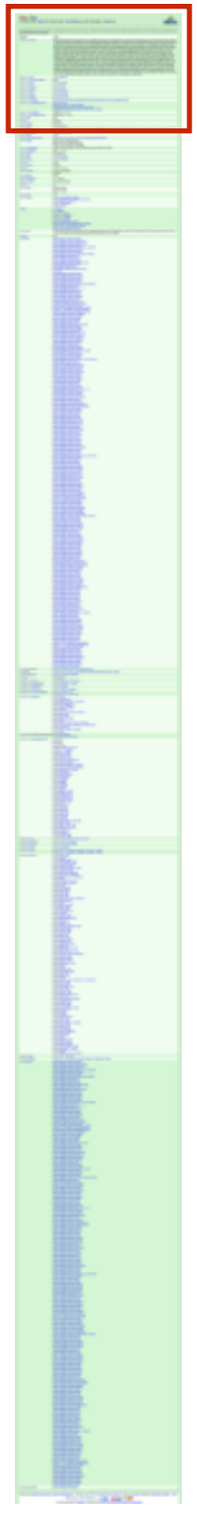
Example: DBpedia:Dog



https://w3id.org/associations

Example: DBpedia:Dog

- Basic information (description, types)



About: Dog

An Entity of Type : animal, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

The domestic dog (Canis lupus familiaris) is a subspecies of the gray wolf (Canis lupus), a member of the Canidae family of the mammalian order Carnivora. The term "domestic dog" is generally used for both domesticated and feral varieties.

Property	Value
dbpedia-owl:abstract	• The domestic dog (Canis lupus familiaris) is a subspecies of the gray wolf (Canis lupus), a member of the Canidae family of the mammalian order Carnivora. The term "domestic dog" is generally used for both domesticated and feral varieties. The dog was the first domesticated animal and has been the most widely kept working, hunting, and pet animal in human history.[citation needed] The word "dog" can also refer to the male of a canine species, as opposed to the word "bitch" which refers to the female of the species.Recent studies of "well-preserved remains of a dog-like canid from the Razboinichya Cave" in the Altai Mountains of southern Siberia concluded that a particular instance of early wolf domestication approximately 33,000 years ago did not result in modern dog lineages, possibly because of climate disruption during the Last Glacial Maximum. The authors postulate that at least several such incipient events have occurred. A study of fossil dogs and wolves in Belgium, Ukraine, and Russia tentatively dates domestication from 14,000 years ago to more than 31,700 years ago. Another recent study has found support for claims of dog domestication between 14,000 and 16,000 years ago, with a range between 9,000 and 34,000 years ago, depending on mutation rate assumptions. Dogs' value to early human hunter-gatherers led to them quickly becoming ubiquitous across world cultures. Dogs perform many roles for people, such as hunting, herding, pulling loads, protection, assisting police and military, companionship, and, more recently, aiding handicapped individuals. This impact on human society has given them the nickname "man's best friend" in the Western world. In some cultures, however, dogs are also a source of meat. In 2001, there were estimated to be 400 million dogs in the world.Most breeds of dog are at most a few hundred years old, having been artificially selected for particular morphologies and behaviors by people for specific functional roles. Through this selective breeding, the dog has developed into hundreds of varied breeds, and shows more behavioral and morphological variation than any other land mammal. For example, height measured to the withers ranges from 15.2 centimetres (6.0 in) in the Chihuahua to about 76 cm (30 in) in the Irish Wolfhound; color varies from white through grays (usually called "blue") to black, and browns from light (tan) to dark ("red" or "chocolate") in a wide variation of patterns; coats can be short or long, coarse-haired to wool-like, straight, curly, or smooth. It is common for most breeds to shed this coat.
dbpedia-owl:class	• dbpedia:Mammal
dbpedia-owl:conservationStatus	• DOM
dbpedia-owl:family	• dbpedia:Canidae
dbpedia-owl:genus	• dbpedia:Canis
dbpedia-owl:kingdom	• dbpedia:Animal
dbpedia-owl:order	• dbpedia:Carnivora
dbpedia-owl:phylum	• dbpedia:Chordate
dbpedia-owl:species	• dbpedia:Gray_wolf
dbpedia-owl:thumbnail	• http://commons.wikimedia.org/wiki/Special:FilePath/YellowLabradorLooking_new.jpg?width=300
dbpedia-owl:wikiPageExternalLink	• http://www.fci.be/ • http://www.ancient.eu.com/article/184/ • http://www.ensembl.org/Canis_familiaris/Info/Index • http://www.biodiversitylibrary.org/name/Canis_lupus_familiaris
dbpedia-owl:wikiPageID	• 4269567 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	• 606813442 (xsd:integer)

Example: DBpedia:Dog

- Basic information (description, types)
- Categorisation (categories & types)

dcterms:subject	• category:Cosmopolitan_species • category:Mammals_with_sequenced_genomes • category:Model_organisms • category:Scavengers • category:Dogs
rdf:type	• owl:Thing • dul:Organism • dbpedia-owl:Animal • dbpedia-owl:Eukaryote • dbpedia-owl:Mammal • dbpedia-owl:Species • http://umbel.org/umbel/rc/Animal • http://umbel.org/umbel/rc/BiologicalLivingObject • http://umbel.org/umbel/rc/EukaryoticCell • http://umbel.org/umbel/rc/Mammal

https://w3id.org/associations

Example: DBpedia:Dog

- Basic information (description, types)
- Categorisation (categories & types)
- Links to other datasets (also Freebase)

https://w3id.org/associations

Example: DBpedia:Dog

- Basic information (description, types)
- Categorisation (categories & types)
- Links to other datasets (also Freebase)
- Inverse links (redirects, link from other datasets)

is dbpedia-owl:wikiPageRedirects of
- dbpedia:כלב
- dbpedia:🐕
- dbpedia:🐶
- dbpedia:A_dog
- dbpedia:A_man's_best_friend
- dbpedia:C._l._familiaris
- dbpedia:C.l._familiaris
- dbpedia:Canine_lupus
- dbpedia:Canis_Canis
- dbpedia:Canis_Familaris
- dbpedia:Canis_Lupus_Familiaris
- dbpedia:Canis_familiaris
- dbpedia:Canis_familiaris_domesticus
- dbpedia:Canis_familiarus_domesticus
- dbpedia:Canis_lupus_familiaris
- dbpedia:DigitalDog
- dbpedia:Dog_(Domestic)
- dbpedia:Dog_groups
- dbpedia:Dog_hood
- dbpedia:Doggies
- dbpedia:Doggy
- dbpedia:Doghood
- dbpedia:Dogness
- dbpedia:Dogs
- dbpedia:Dogs_(Animals)

https://w3id.org/associations

Example: DBpedia:Dog

- Basic information (description, types)
- Categorisation (categories & types)
- Links to other datasets (also Freebase)
- Inverse links (redirects, link from other datasets)
- Wiki-page-links (377 for Dog)

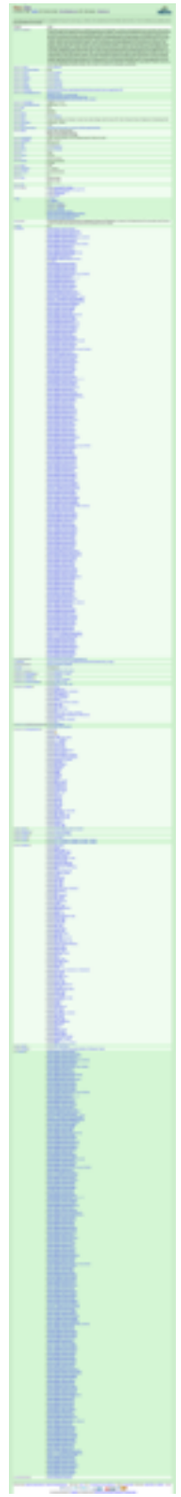
Dog

Cat

Fur

https://w3id.org/associations

Example: DBpedia:Dog



- Basic information (description, types)
- Categorisation (categories & types)
- Links to other datasets (also Freebase)
- Inverse links (redirects, link from other datasets)
- Wiki-page-links (377 for Dog)
- Associations?
  - Dog:
    Cat (✓), collar (✗), leash (✗), walk (✗), fur (✓), bark (✗)

https://w3id.org/associations

Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

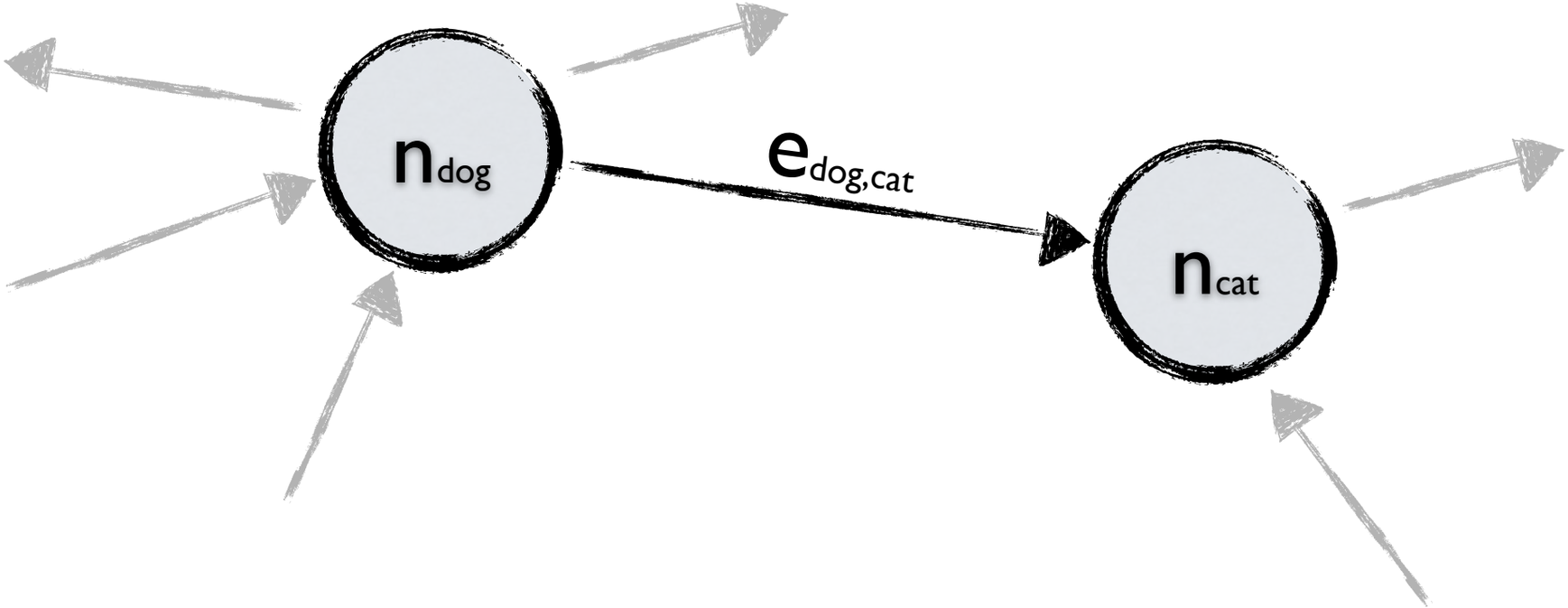
Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

My Research

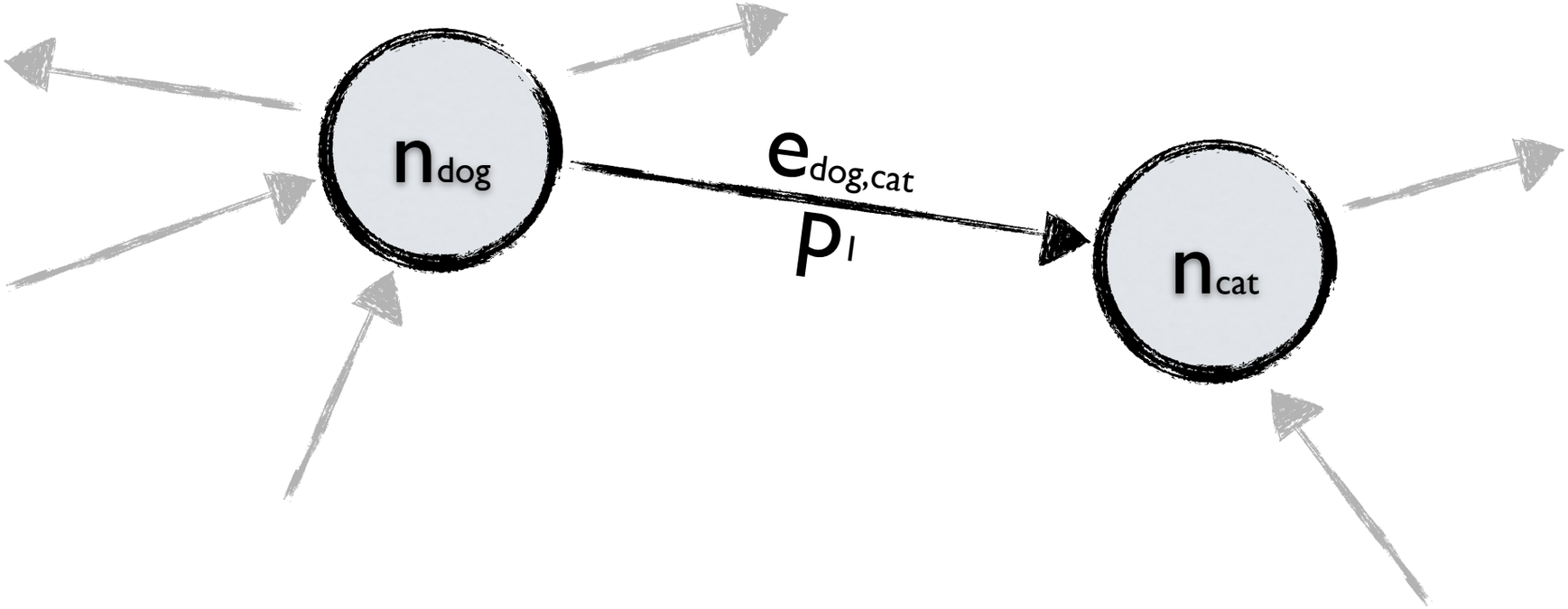
- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?



https://w3id.org/associations

My Research

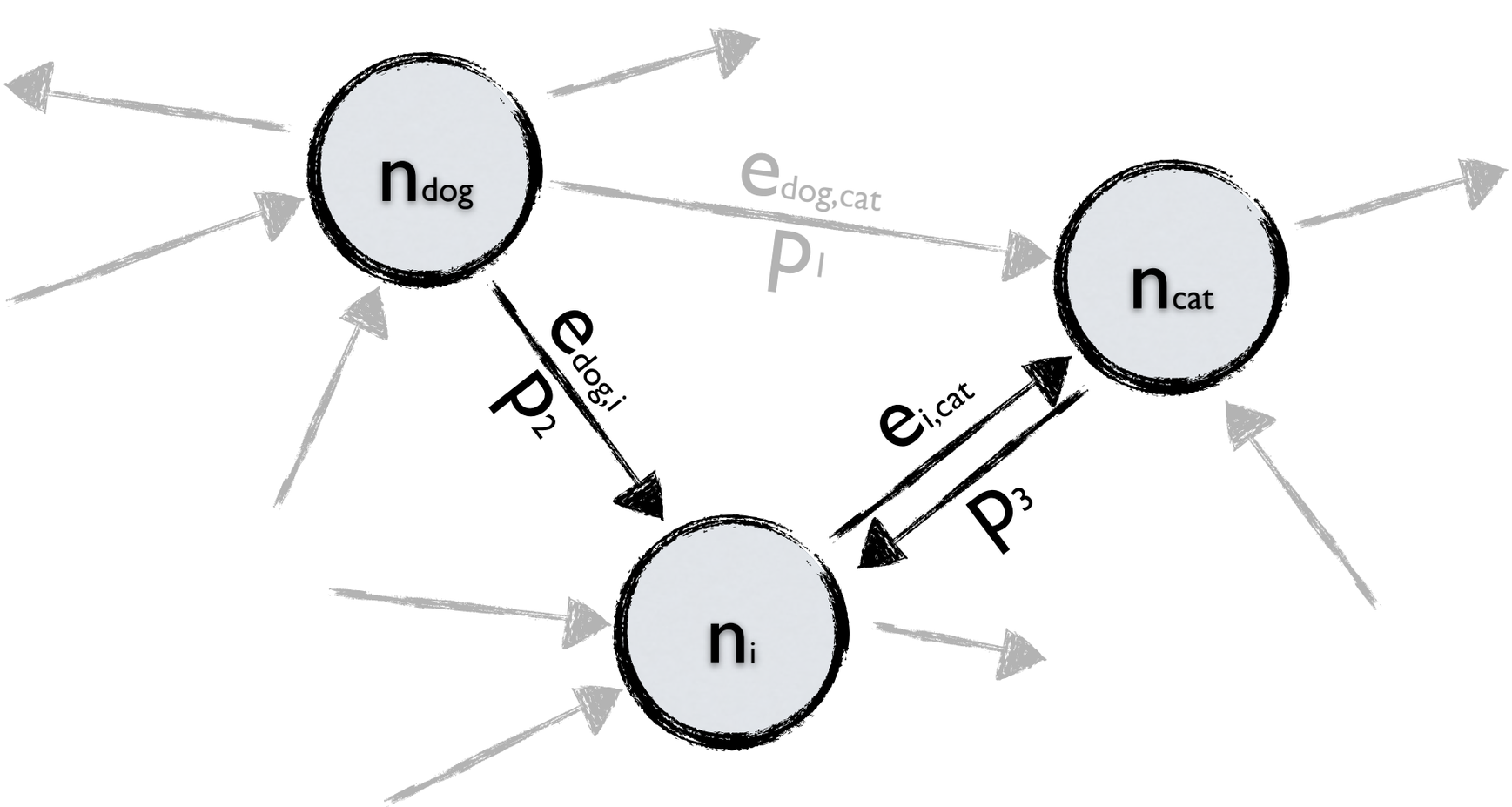
- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?



https://w3id.org/associations

My Research

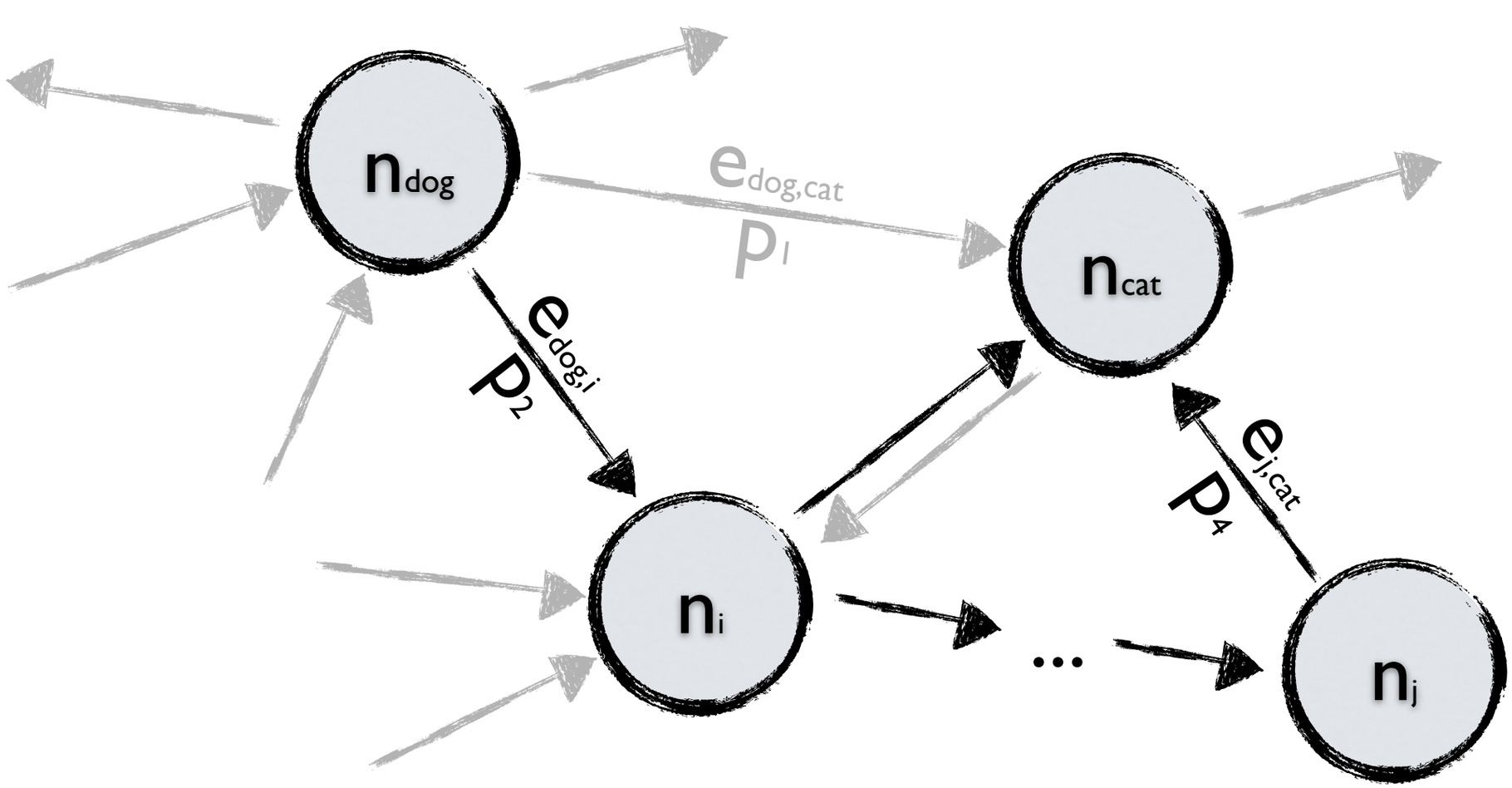
- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?



https://w3id.org/associations

My Research

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



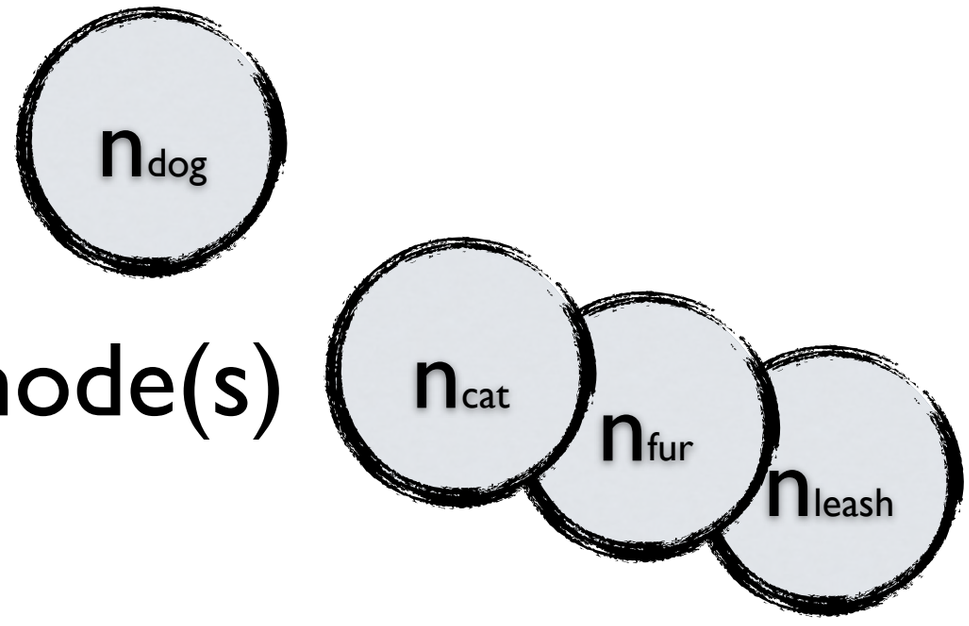
https://w3id.org/associations

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



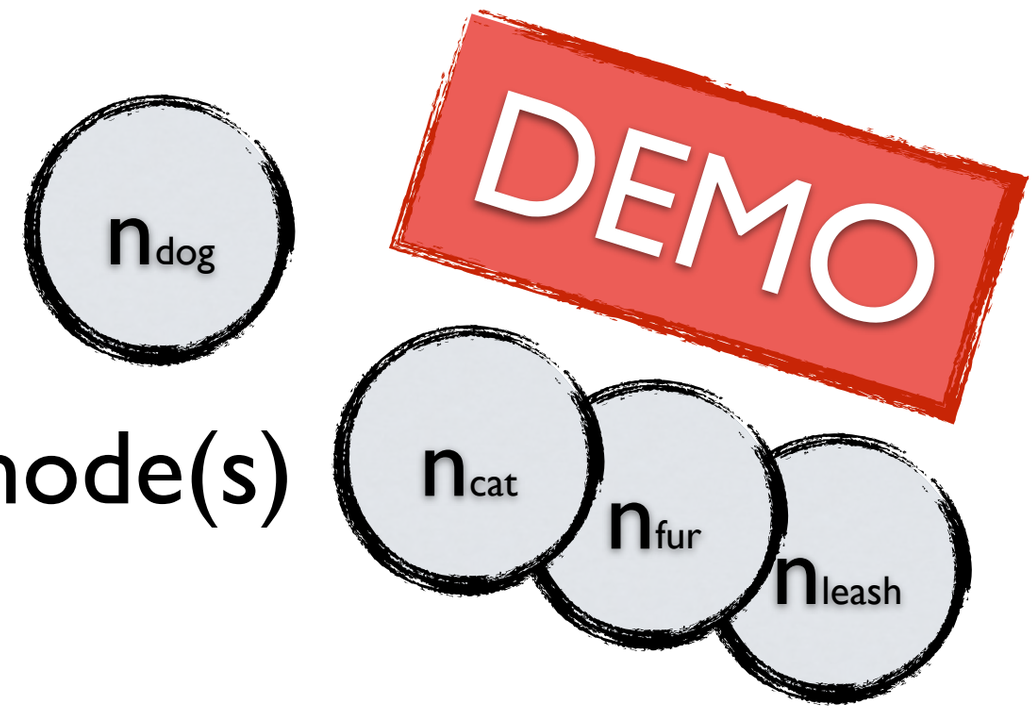
https://w3id.org/associations

My Research

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

    predict the output node(s)

    we would associate



https://w3id.org/associations

Demo

dbr:Linked_data

```
Top 10 predictions (method: target_occs)
  <http://dbpedia.org/resource/Semantic_Web> (7.000)
  <http://dbpedia.org/resource/Wikipedia> (3.000)
  <http://dbpedia.org/resource/Open_Semantic_Framework> (3.000)
  <http://dbpedia.org/resource/Tim_Berners-Lee> (2.000)
  <http://dbpedia.org/resource/Dereferenceable_Uniform_Resource_Identifier> (2.000)
  <http://dbpedia.org/resource/World_Wide_Web> (2.000)
  <http://dbpedia.org/resource/Hypertext_Transfer_Protocol> (2.000)
  <http://dbpedia.org/resource/Language> (2.000)
  <http://dbpedia.org/resource/Uniform_resource_identifier> (2.000)
  <http://dbpedia.org/resource/DBpedia> (2.000)
Top 10 predictions (method: precisions)
  <http://dbpedia.org/resource/Semantic_Web> (5.750)
  <http://dbpedia.org/resource/Open_Semantic_Framework> (2.333)
  <http://dbpedia.org/resource/Wikipedia> (1.450)
  <http://dbpedia.org/resource/Tim_Berners-Lee> (1.250)
  <http://dbpedia.org/resource/Dereferenceable_Uniform_Resource_Identifier> (1.250)
  <http://dbpedia.org/resource/Language> (1.200)
  <http://dbpedia.org/resource/Data_set> (1.000)
  <http://dbpedia.org/resource/Serialization> (1.000)
  <http://dbpedia.org/resource/DBpedia> (0.833)
  <http://dbpedia.org/resource/World_Wide_Web> (0.750)
```


https://w3id.org/associations

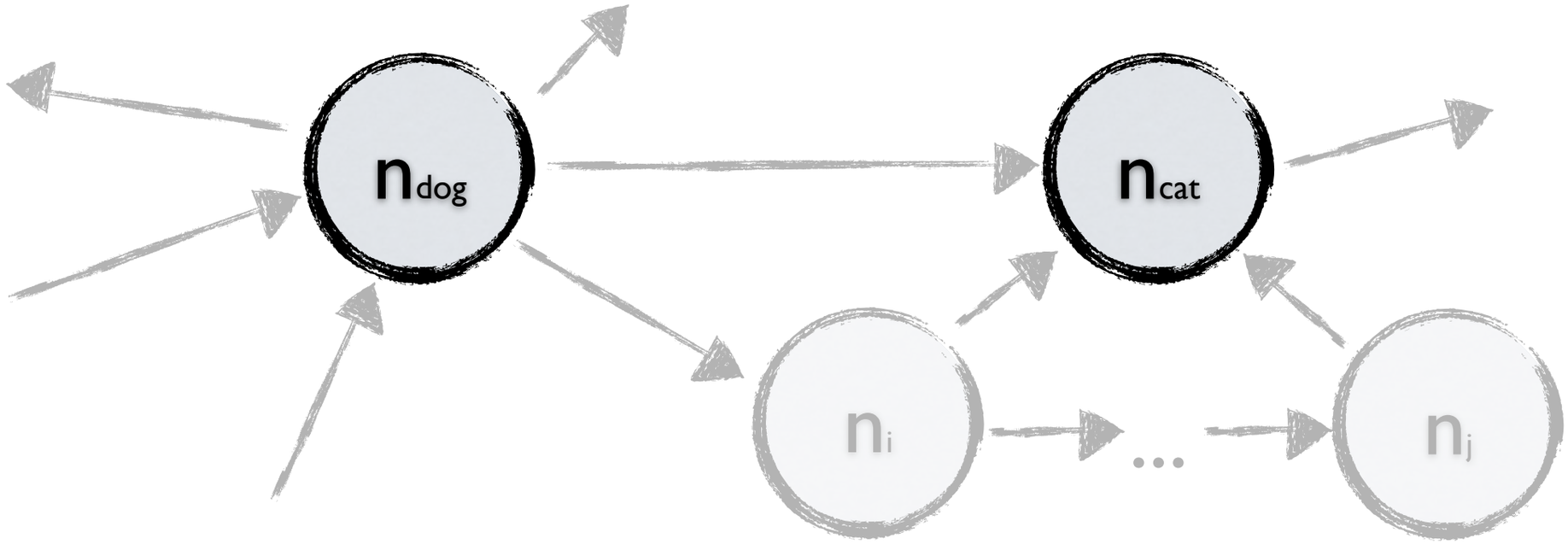
dbr:Semantic_Web

```
Top 10 predictions (method: target_occs)
  <http://dbpedia.org/resource/World_Wide_Web> (15.000)
  <http://dbpedia.org/resource/Web_search_engine> (10.000)
  <http://dbpedia.org/resource/Scalable_Vector_Graphics> (8.000)
  <http://dbpedia.org/resource/Blog> (7.000)
  <http://dbpedia.org/resource/Semantic_Web> (7.000)
  <http://dbpedia.org/resource/Web_2.0> (7.000)
  <http://dbpedia.org/resource/Artificial_intelligence> (6.000)
  <http://dbpedia.org/resource/Cascading_Style_Sheets> (6.000)
  <http://dbpedia.org/resource/Tim_Berners-Lee> (5.000)
  <http://dbpedia.org/resource/Language> (4.000)
Top 10 predictions (method: precisions)
  <http://dbpedia.org/resource/World_Wide_Web> (7.316)
  <http://dbpedia.org/resource/Blog> (4.445)
  <http://dbpedia.org/resource/Semantic_Web> (4.041)
  <http://dbpedia.org/resource/Web_search_engine> (3.051)
  <http://dbpedia.org/resource/Artificial_intelligence> (2.977)
  <http://dbpedia.org/resource/Cascading_Style_Sheets> (2.580)
  <http://dbpedia.org/resource/Web_2.0> (1.791)
  <http://dbpedia.org/resource/Tim_Berners-Lee> (1.568)
  <http://dbpedia.org/resource/Machine> (1.500)
  <http://dbpedia.org/resource/Provenance> (1.500)
```


https://w3id.org/associations

My Research

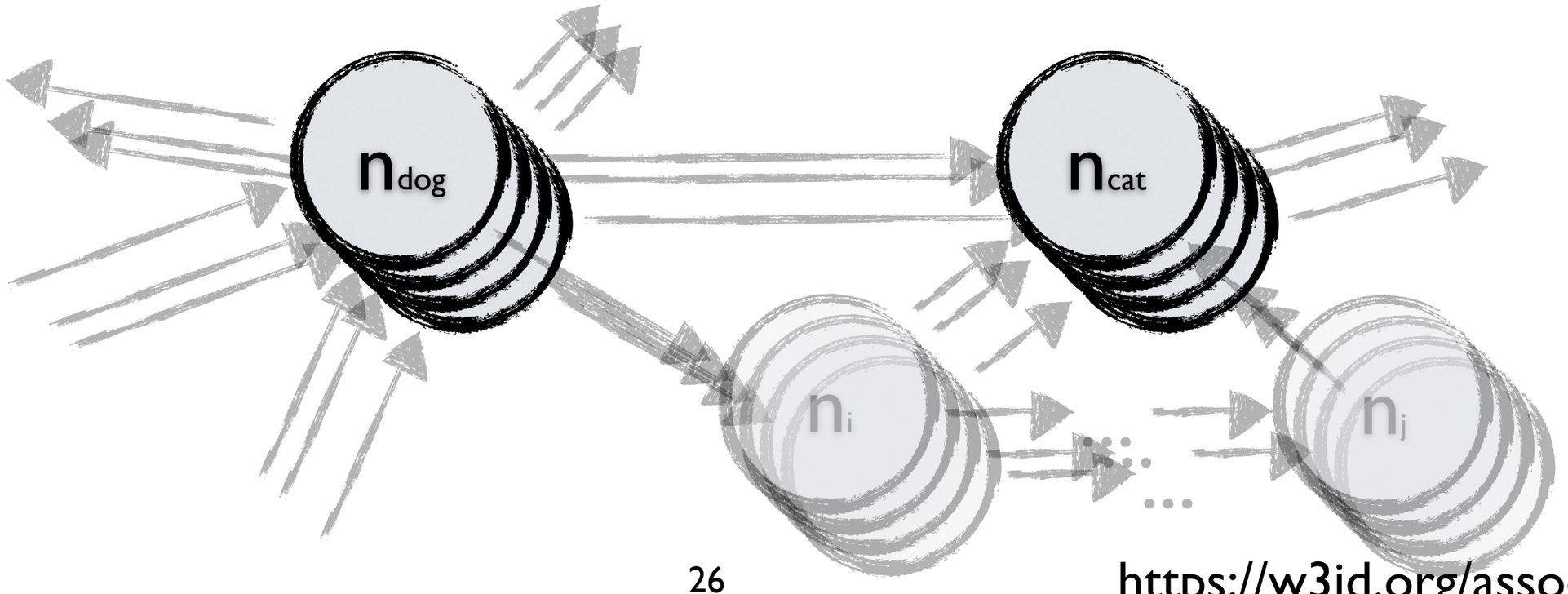
- Research Question:
  - Is it possible to learn patterns for Human Associations from Linked Data?
- Dataset of "Semantic Associations" needed



https://w3id.org/associations

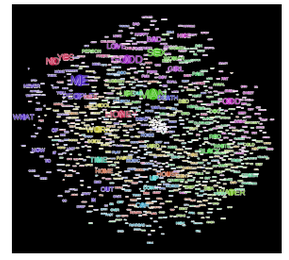
My Research

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



https://w3id.org/associations

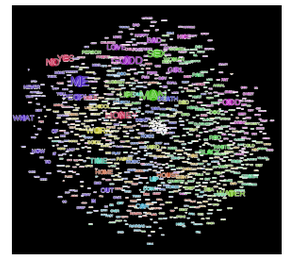
Edinburgh Associative Thesaurus



- Association corpus (1973) G. Kiss, C. Armstrong, R. Milroy, J. Piper
  - For each stimulus asked 100 ppl for a response
  - Strong responses became stimuli of next round

https://w3id.org/associations

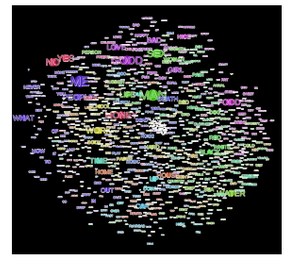
Edinburgh Associative Thesaurus



- Association corpus (1973) G. Kiss, C. Armstrong, R. Milroy, J. Piper
  - For each stimulus asked 100 ppl for a response
  - Strong responses became stimuli of next round
  - ~ 790 K associations (free text)
  - Graph: (|V| = 23 K, |E| = 325 K)

https://w3id.org/associations

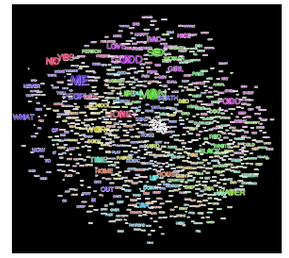
Edinburgh Associative Thesaurus



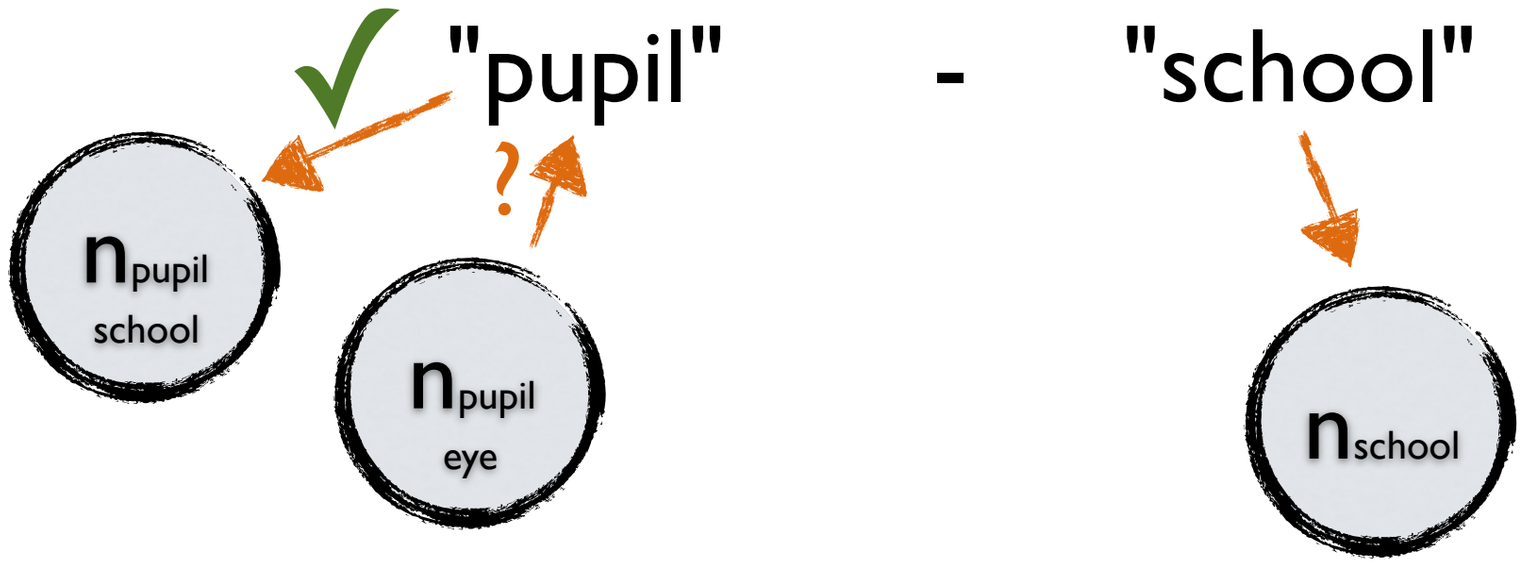
- Association corpus (1973) G. Kiss, C. Armstrong, R. Milroy, J. Piper
  - For each stimulus asked 100 ppl for a response
  - Strong responses became stimuli of next round
  - ~ 790 K associations (free text)
  - Graph: (|V| = 23 K, |E| = 325 K)
    - ~ 5000 strong associations (>19x)
      - ~167.4 K raw associations

https://w3id.org/associations

Edinburgh Associative Thesaurus



- Difficulty: Free text
  - Mapped to Semantic (DBpedia) Entities

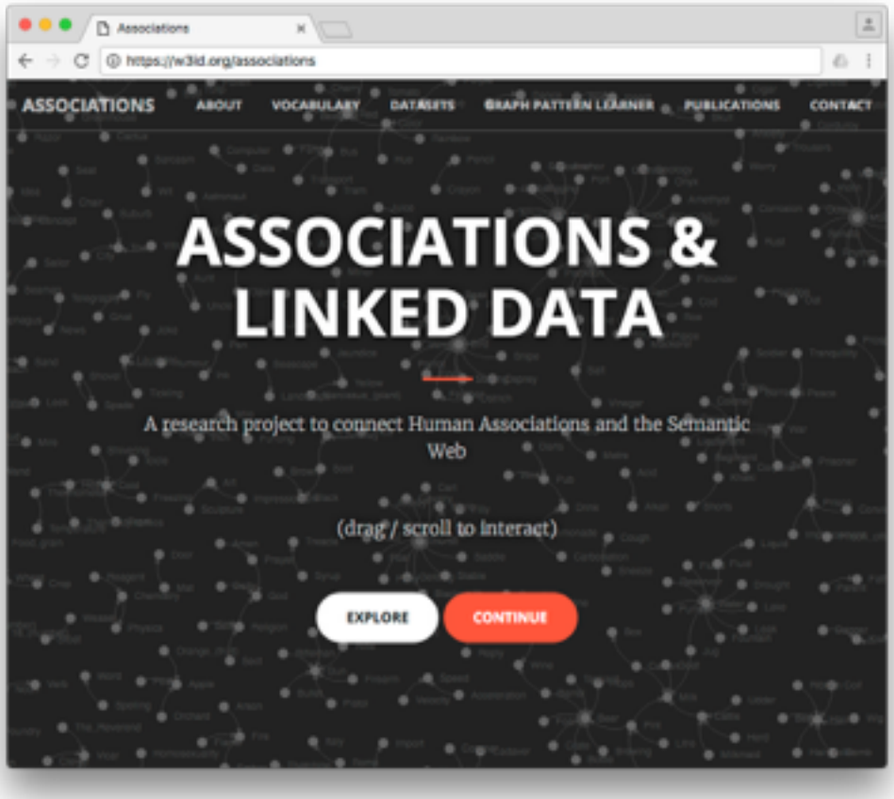


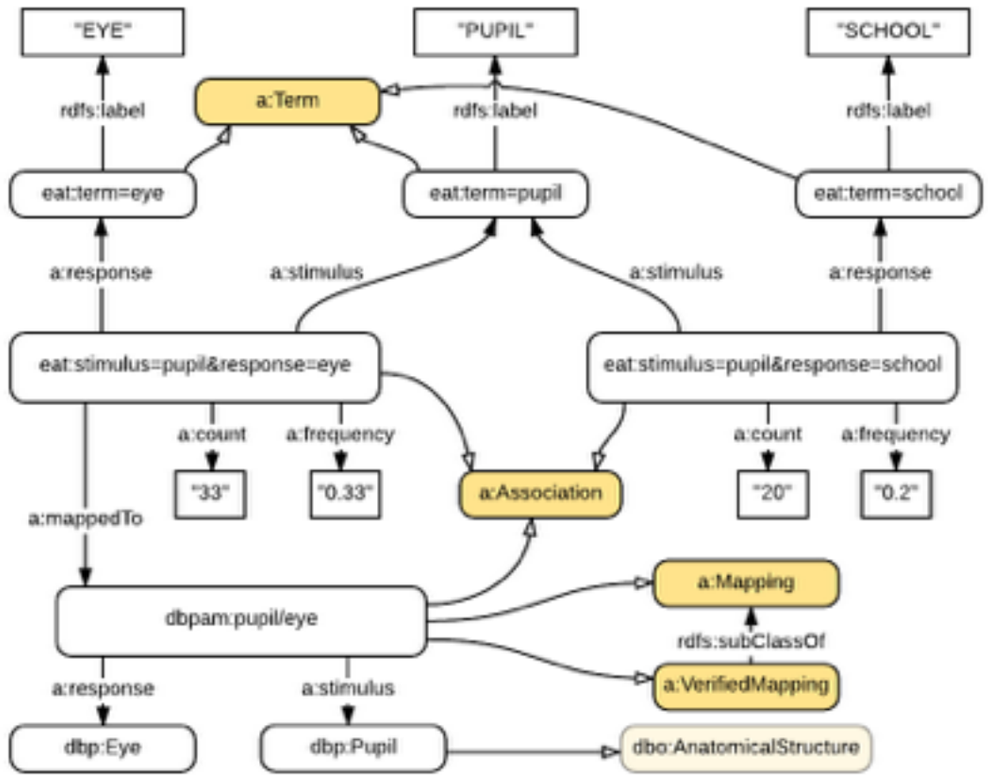
- Semi-Automatic Mapping Approach
  - 727 verified distinct "Semantic Associations"
  - ~ 25.5 K raw associations

https://w3id.org/associations

Semantic Associations Dataset

- (Raw) EAT as RDF (1.7 M triples)
- 727 verified distinct Semantic Associations





https://w3id.org/associations

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

Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

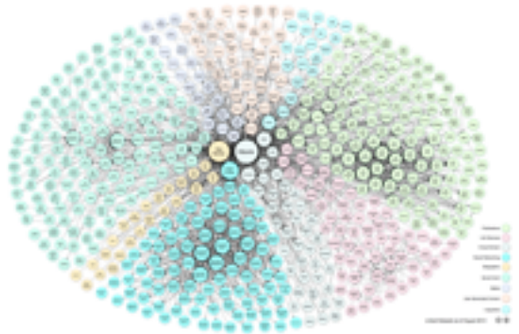
Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

Data Analysis

- Local Linked Data Endpoint:
  - Central Datasets
    - ~8 G triples
    - SPARQL Queries
- Scalability Issues



DBpedia

WIKIDATA

Freebase

Linked GeoData

https://w3id.org/associations

First Analysis

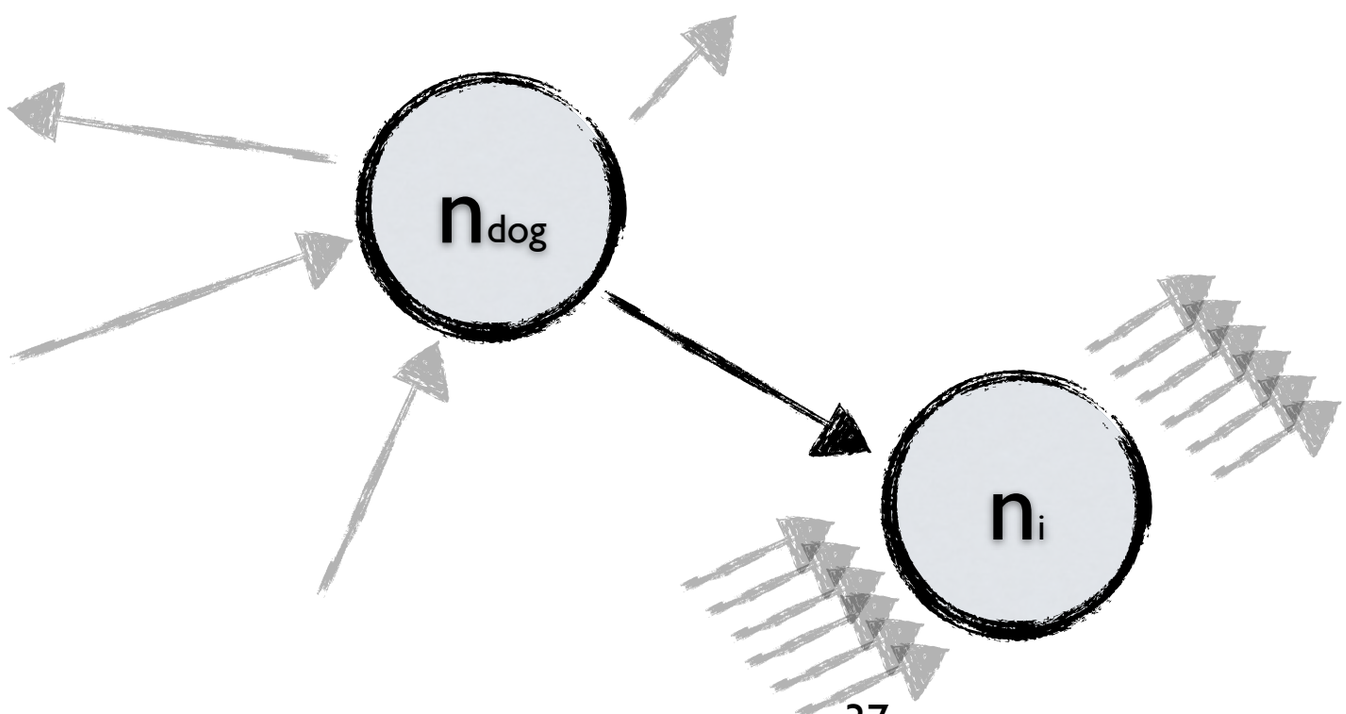
- Node Degrees:
  - Avg: 3643
- Excluding big nodes?
  - Would cause bias!
- Exclude in-links?
  - Directionality depends on modelling!

Node	Degree
dbp:Animal	400624
dbp:Insect	195058
dbp:France	190047
dbp:India	181119
dbp:Italy	132719
dbp:Village	132400
dbp:Plant	126731
dbp:Scotland	71828
dbp:Paris	64232
dbp:Switzerland	55471

https://w3id.org/associations

Graph Pattern Learning

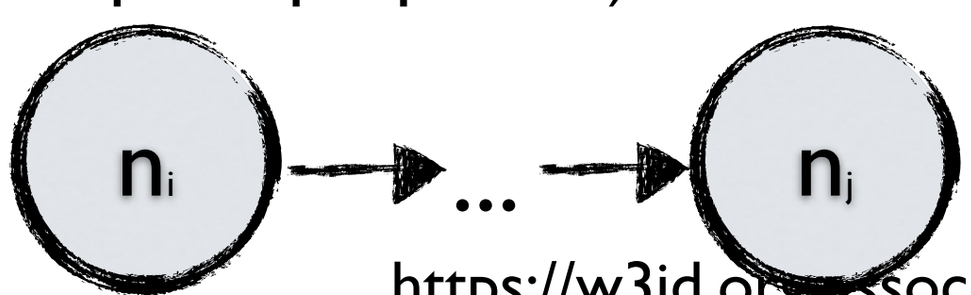
- First idea: Shortest paths
  - Many false positives (associations)
  - Problems with high degrees
    - Super-nodes (owl:Thing, Lists, countries, cities)
- Everything is connected with paths of length 2 ;)



https://w3id.org/associations

Graph Pattern Learning

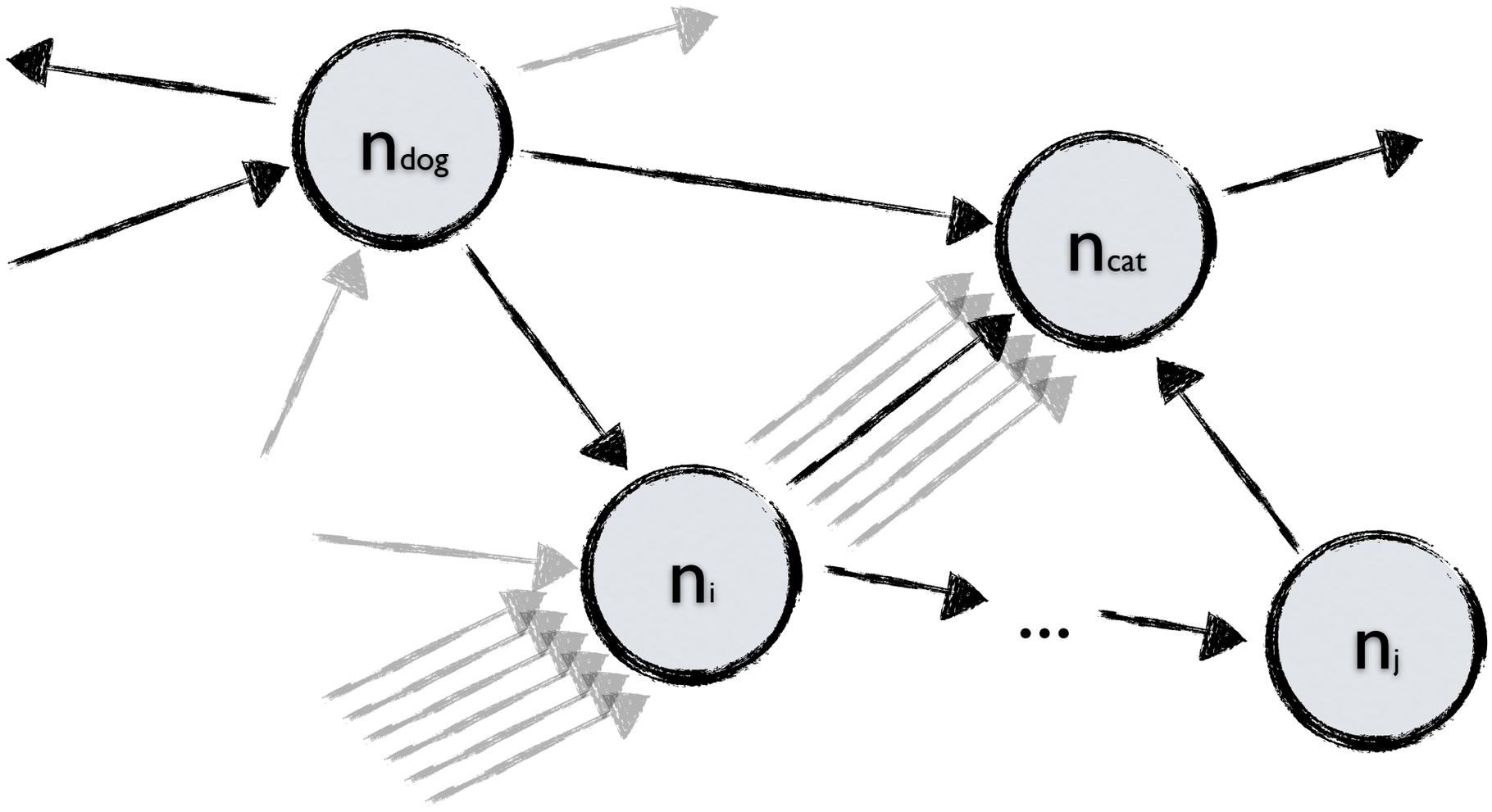
- First idea: Shortest paths
  - Many false positives (associations)
  - Problems with high degrees
    - Super-nodes (owl:Thing, Lists, countries, cities)
      - Everything is connected with paths of length 2 ;)
- Problems due to modelling
  - Nearly linear parts / chains
    - (owl:sameAs, freebase, temporal properties)



https://w3id.org/associations

Graph Pattern Learning

- Shortest ~~paths~~



https://w3id.org/associations

Good Graph Patterns?

- How often is a response reached? max

Stimulus		Response
dbr:Cow	→	dbr:Milk
dbr:Camping	→	dbr:Tent
dbr:Expense	→	dbr:Money
dbr:Bed		dbr:Sleep
dbr:Pupil	→	dbr:Eye

https://w3id.org/associations

Good Graph Patterns?

- How often is a response reached? max
- How many other nodes are reached? min

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil

Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Good Graph Patterns?

- How often is a response reached? max
- How many other nodes are reached? min
- How many nodes need to be expanded? min

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil

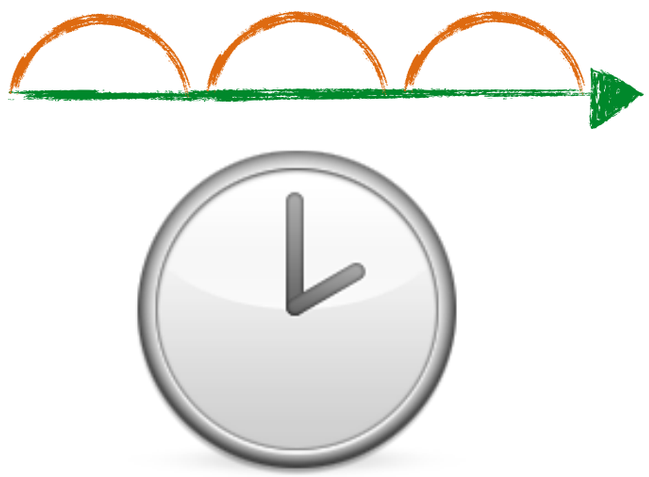
Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Good Graph Patterns?

- How often is a response reached? max
- How many other nodes are reached? min
- How many nodes need to be expanded? min
- How long does a query take? min

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil



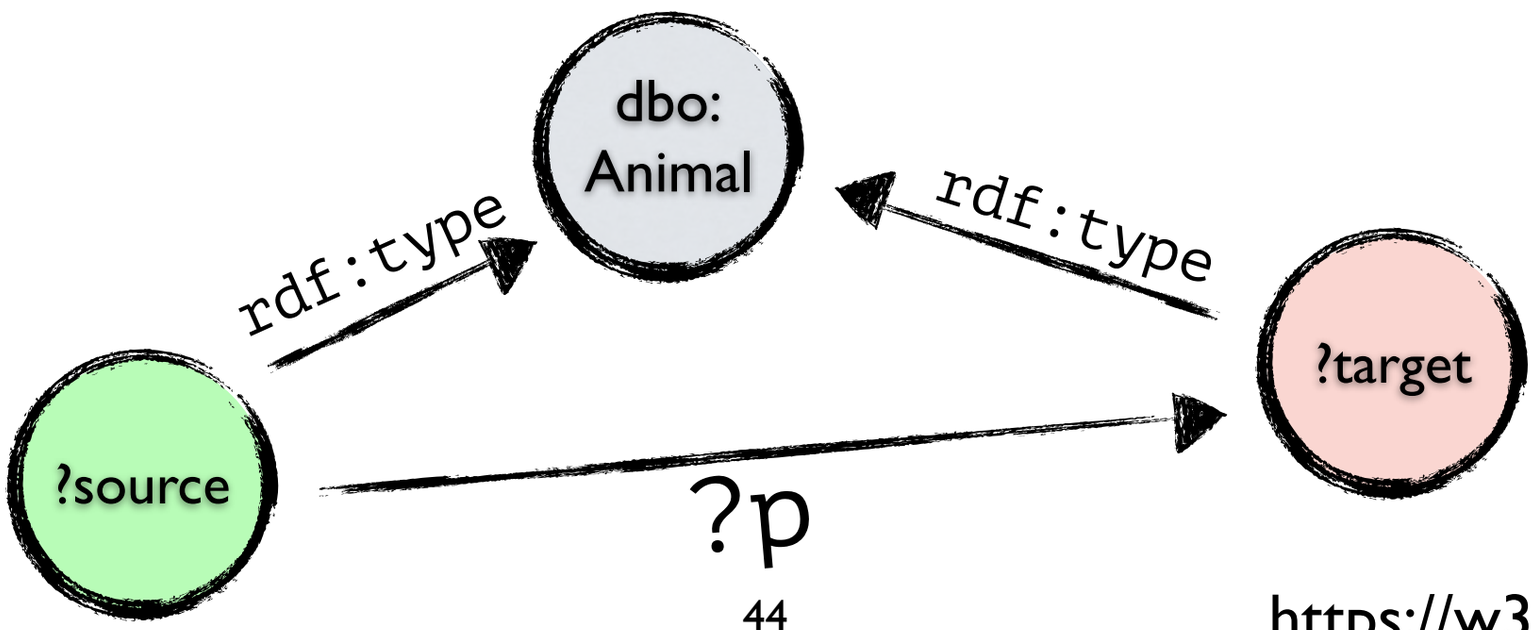
Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Evolutionary Algorithm

- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns

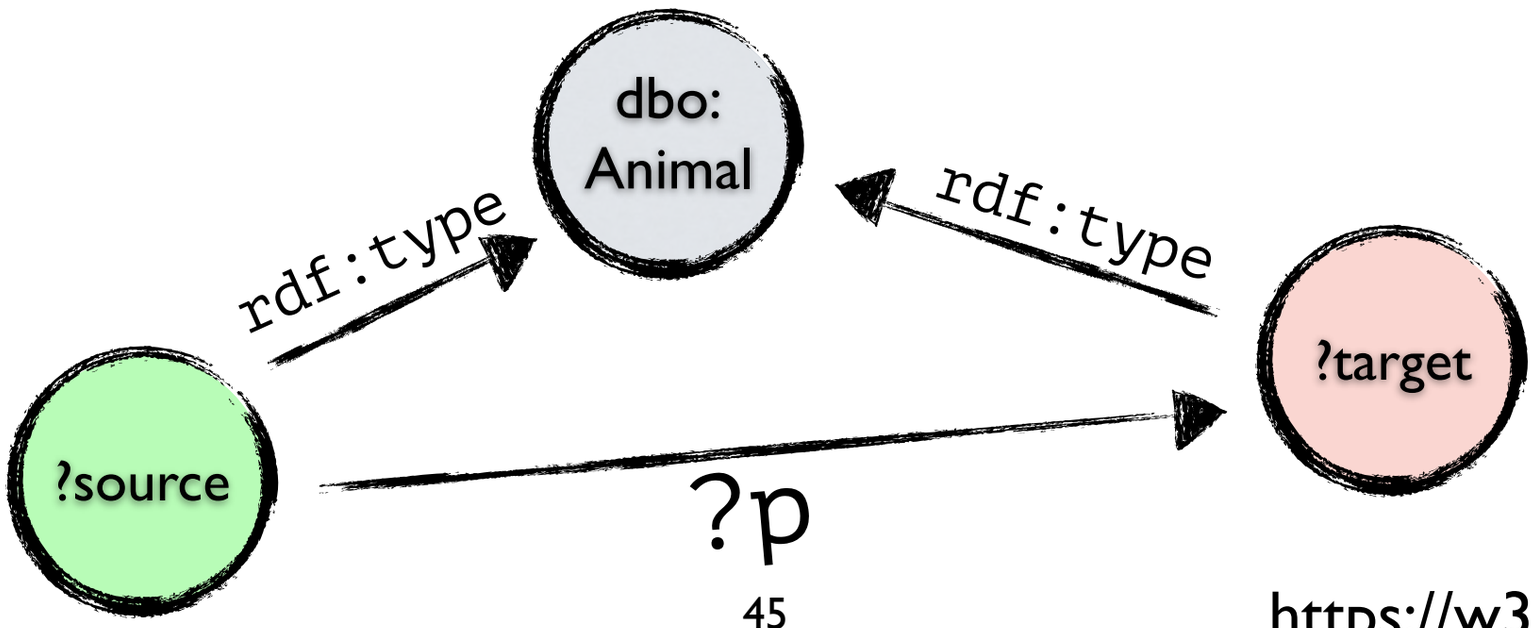
```
{
  ?source rdf:type dbo:Animal .
  ?target rdf:type dbo:Animal .
  ?source ?p ?target .
}
```



https://w3id.org/associations

Evolutionary Algorithm

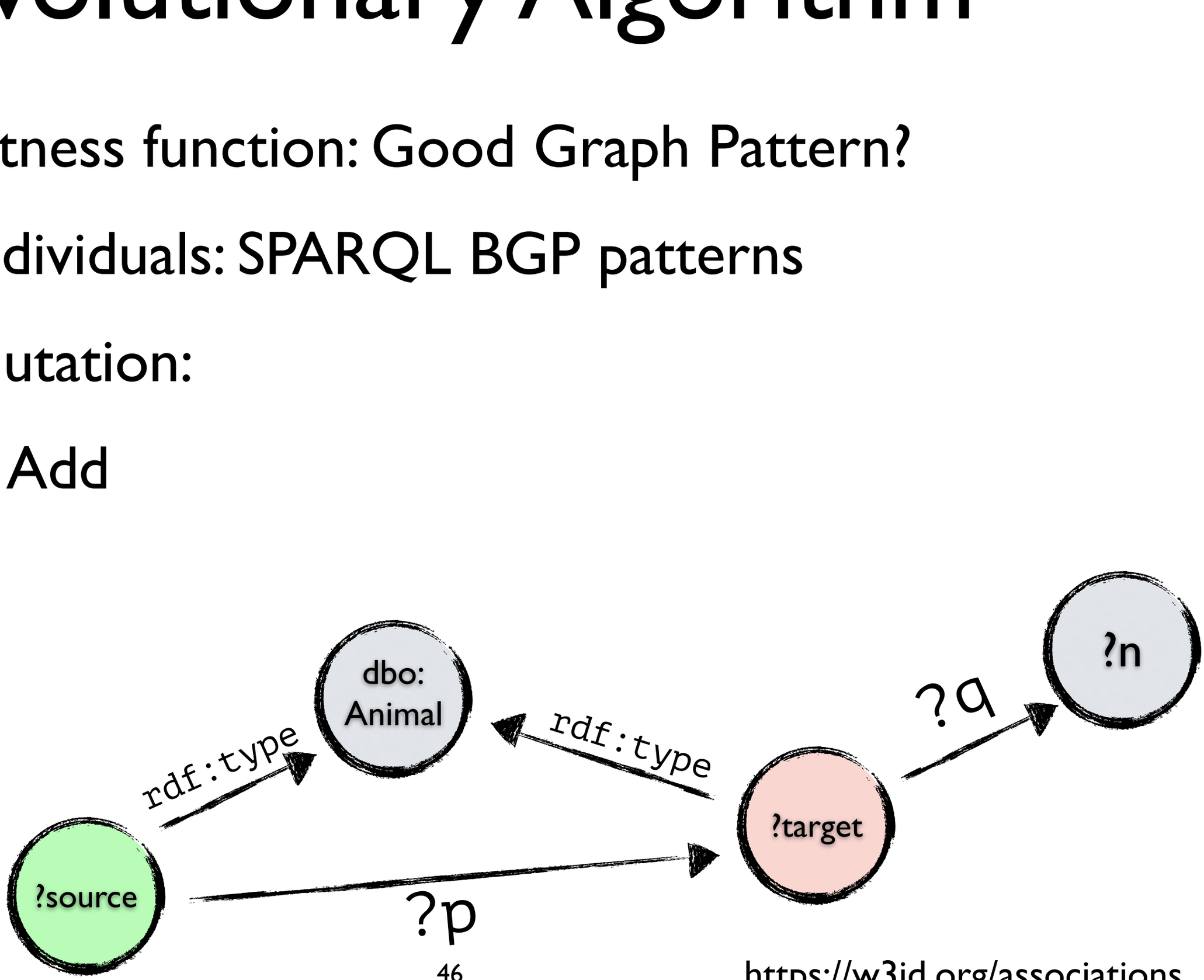
- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:



https://w3id.org/associations

Evolutionary Algorithm

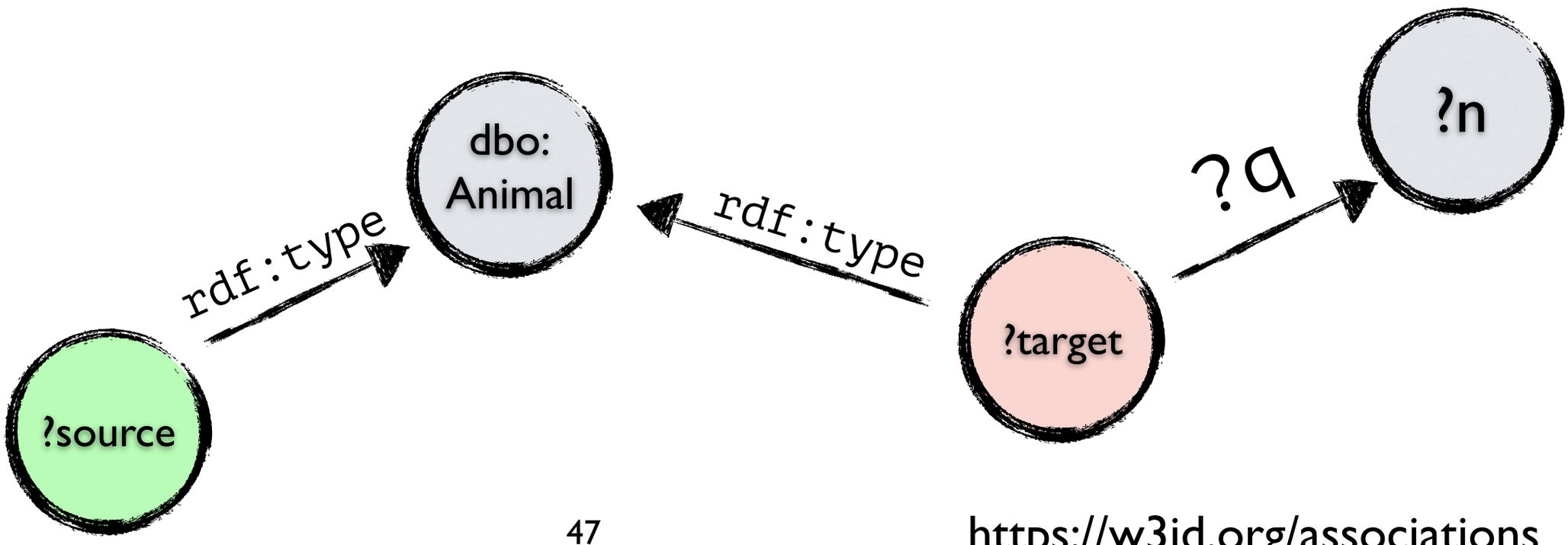
- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:
- Add



https://w3id.org/associations

Evolutionary Algorithm

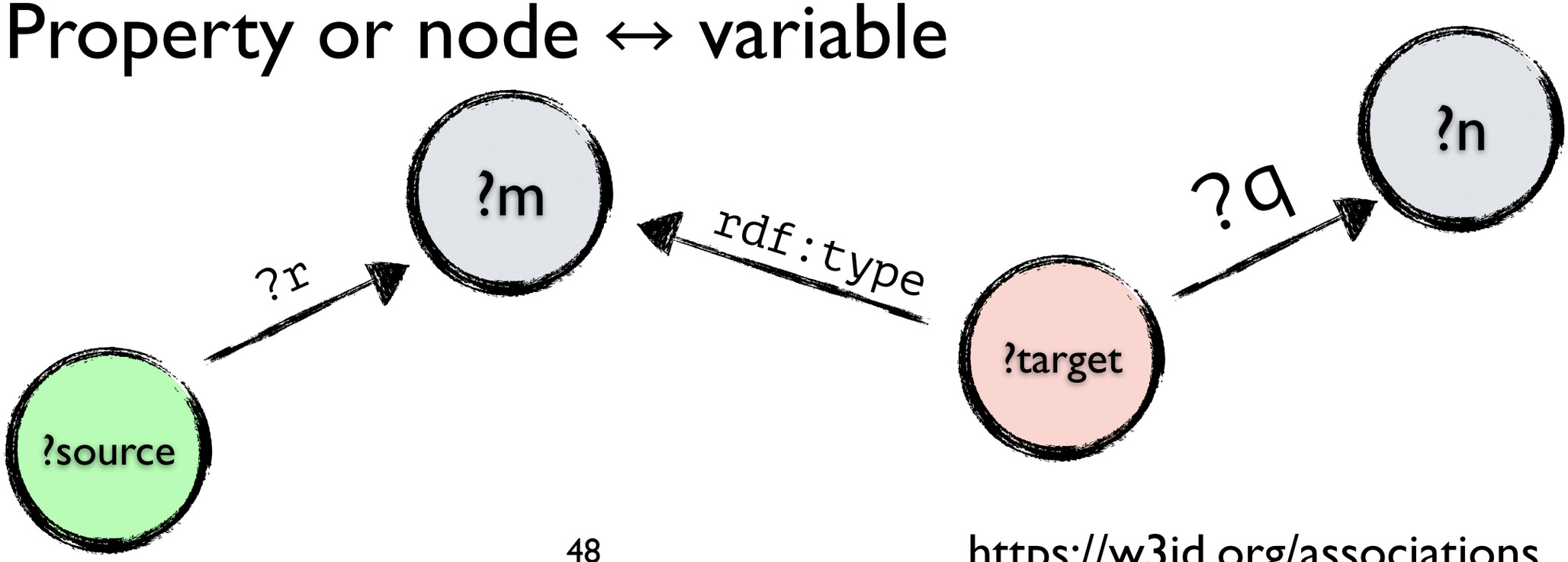
- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:
- Add / delete triples



https://w3id.org/associations

Evolutionary Algorithm

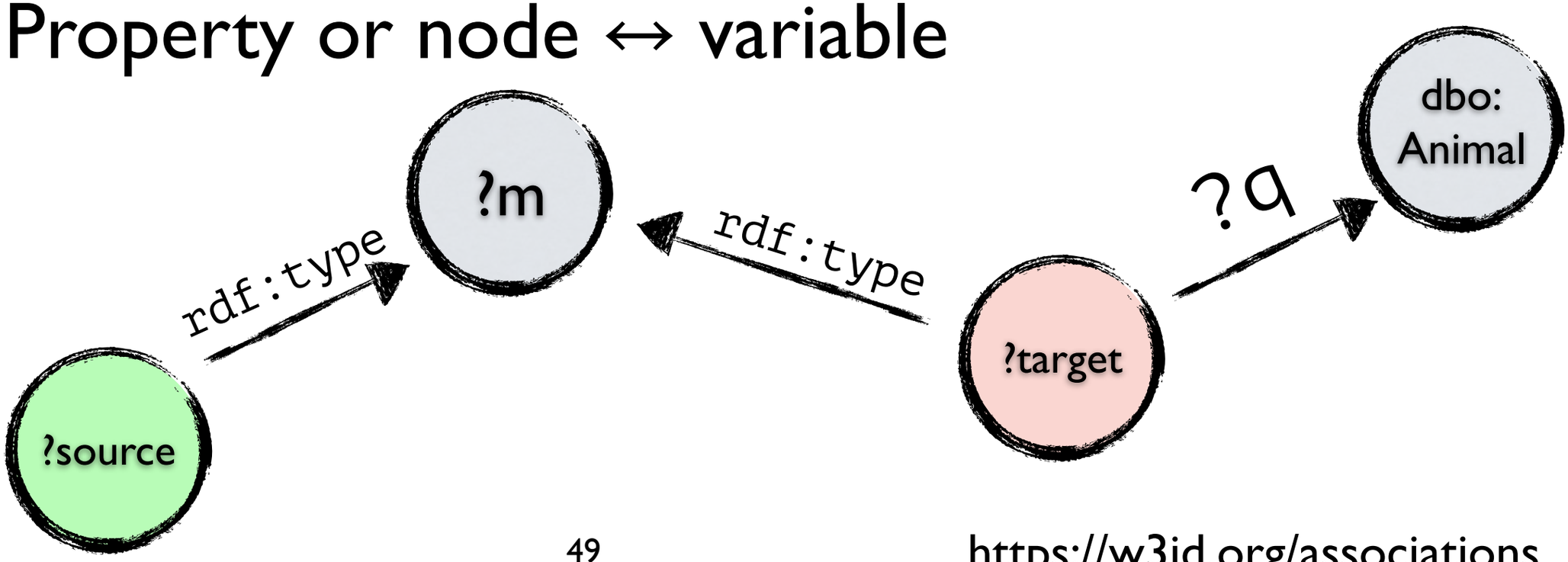
- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:
  - Add / delete triples
  - Property or node ↔ variable



https://w3id.org/associations

Evolutionary Algorithm

- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:
  - Add / delete triples
  - Property or node ↔ variable



https://w3id.org/associations

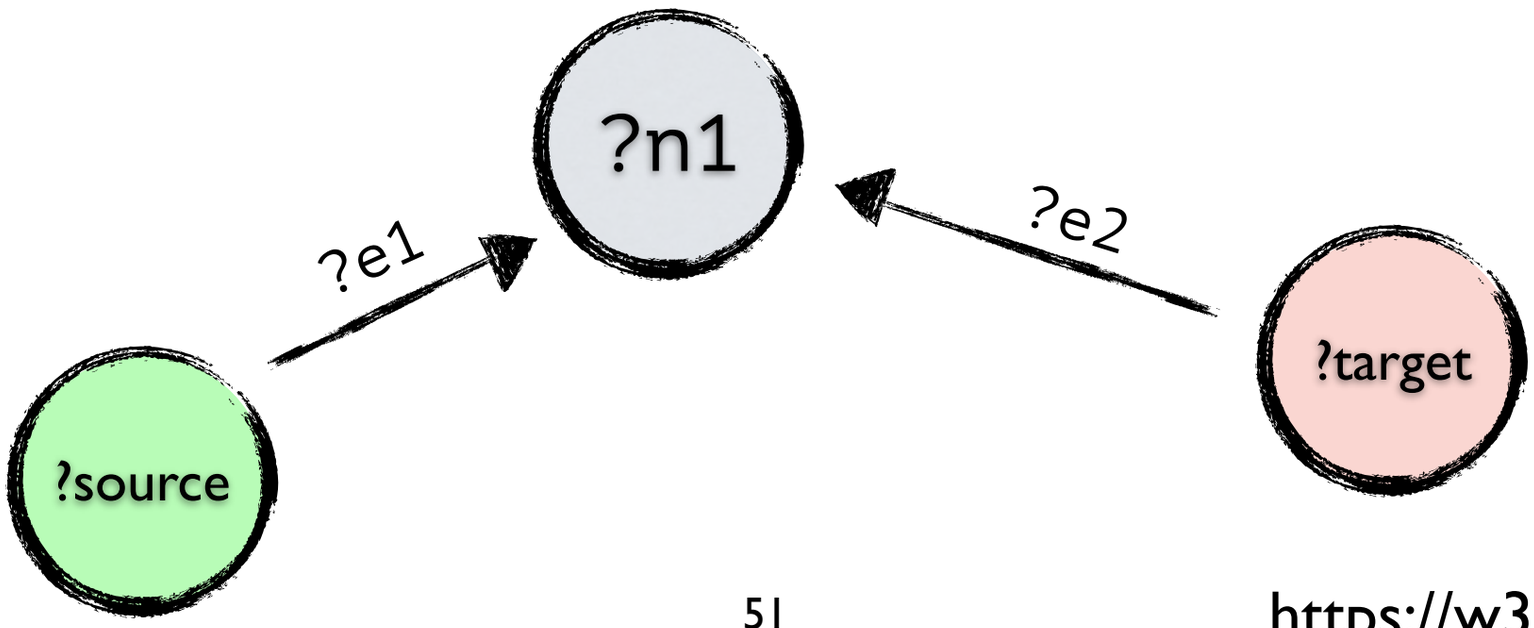
Evolutionary Algorithm

- Fitness function: Good Graph Pattern?
- Individuals: SPARQL BGP patterns
- Mutation:
  - Add / delete triples
  - Property or node ↔ variable
- Mating: Exchange of triples & unifying vars
- Selection: Tournament

https://w3id.org/associations

Evolutionary Algorithm

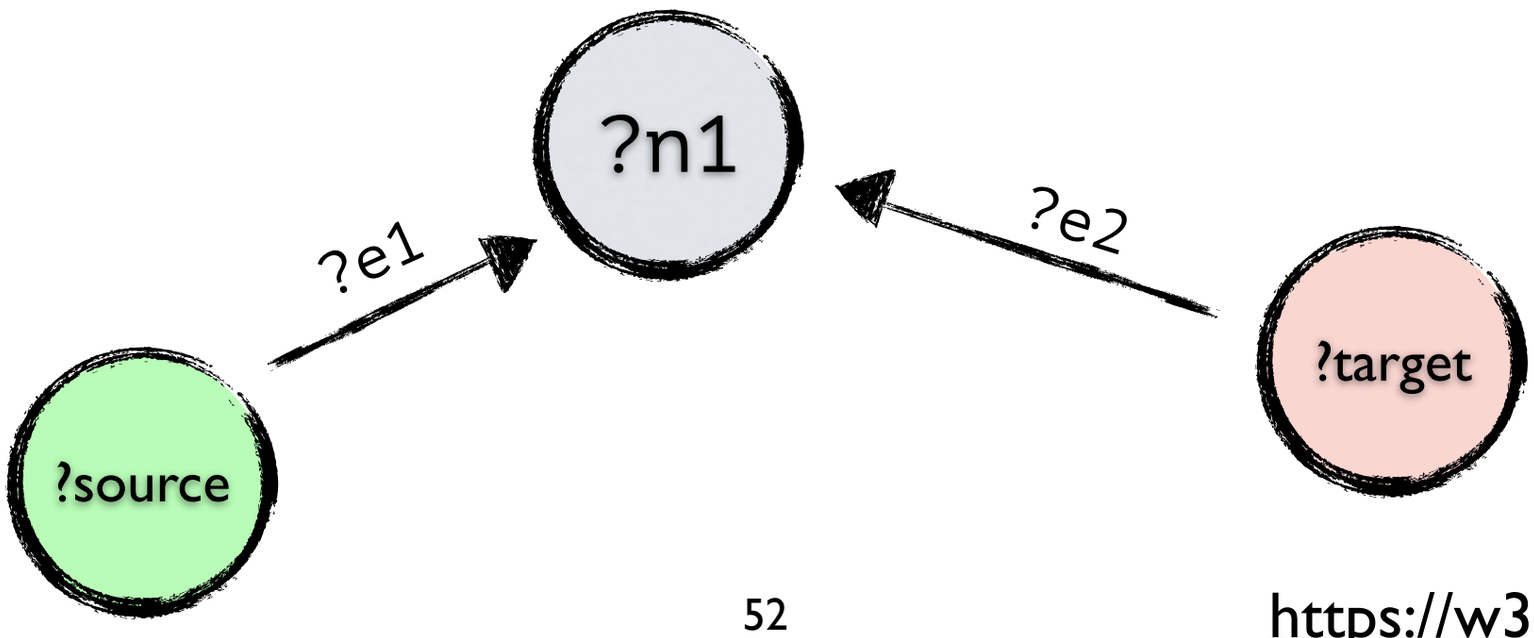
- Starting Population:
  - Randomised length paths between `?source` and `?target`



https://w3id.org/associations

Evolutionary Algorithm

- Starting Population:
  - Randomised length paths between `?source` and `?target`
- Keeping the population healthy
  - Re-introduce basic / good fit patterns



https://w3id.org/associations

Pattern Coverage

- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

Stimulus		Response
dbr:Cow	→	dbr:Milk
dbr:Camping		dbr:Tent
dbr:Expense		dbr:Money
dbr:Bed		dbr:Sleep
dbr:Pupil	→	dbr:Eye

https://w3id.org/associations

Pattern Coverage

- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil

Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Pattern Coverage

- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

Stimulus		Response
dbr:Cow		dbr:Milk
dbr:Camping	→	dbr:Tent
dbr:Expense	→	dbr:Money
dbr:Bed		dbr:Sleep
dbr:Pupil		dbr:Eye

https://w3id.org/associations

Pattern Coverage

- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil

Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Pattern Coverage

- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

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

https://w3id.org/associations

Pattern Coverage

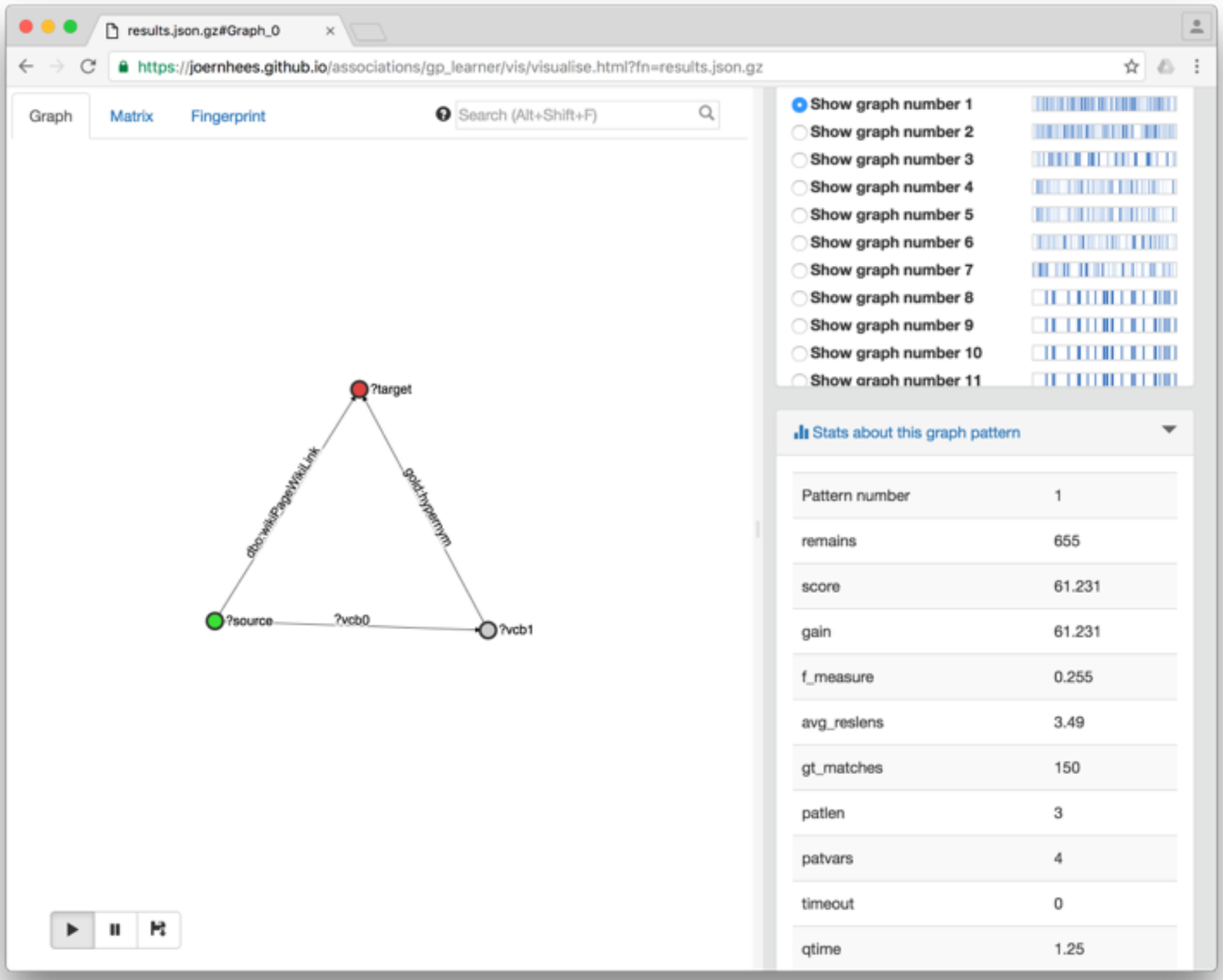
- Multiple runs of Evolutionary Algorithm
  - After good patterns are found, refocus on remaining source-target pairs

Stimulus
dbr:Cow
dbr:Camping
dbr:Expense
dbr:Bed
dbr:Pupil

Response
dbr:Milk
dbr:Tent
dbr:Money
dbr:Sleep
dbr:Eye

https://w3id.org/associations

Learned Graph Patterns



https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

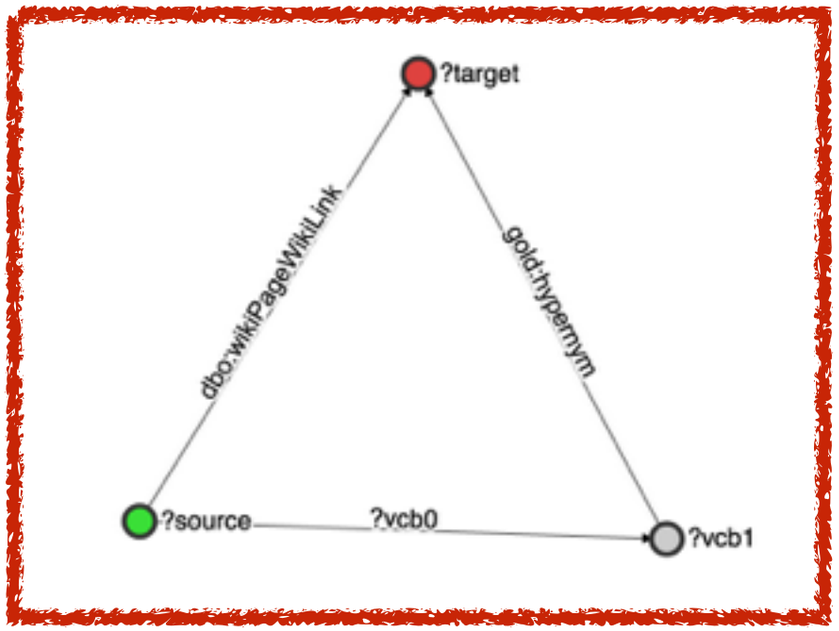
https://w3id.org/associations

results.json.gz#Graph_0

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	1
remains	655
score	61.231
gain	61.231
f_measure	0.255
avg_reslens	3.49
gt_matches	150
patlen	3
patvars	4
timeout	0
qtime	1.25

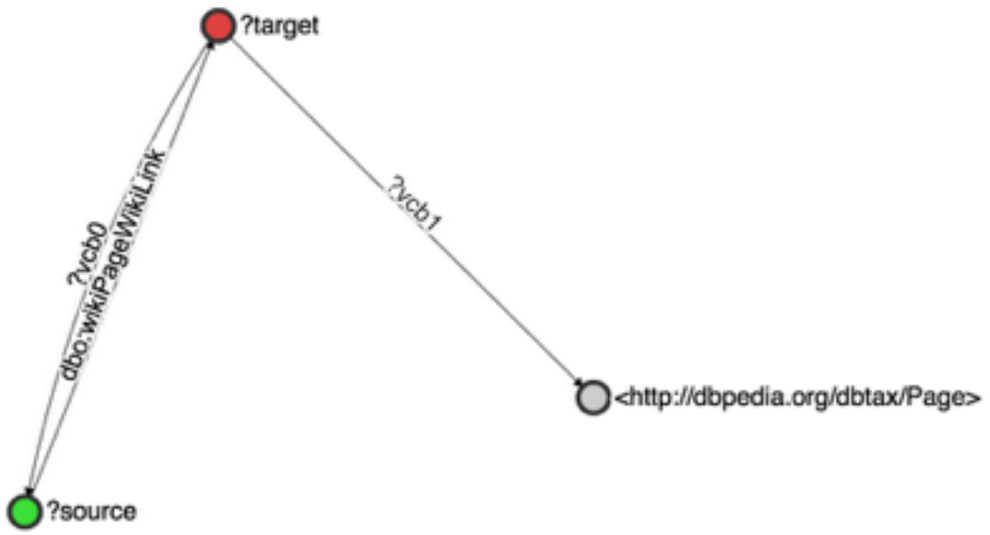
https://w3id.org/associations

results.json.gz#Graph_1

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	2
remains	655
score	52.954
gain	52.954
f_measure	0.227
avg_reslens	4.573
gt_matches	155
patlen	3
patvars	4
timeout	0
qtime	0.819

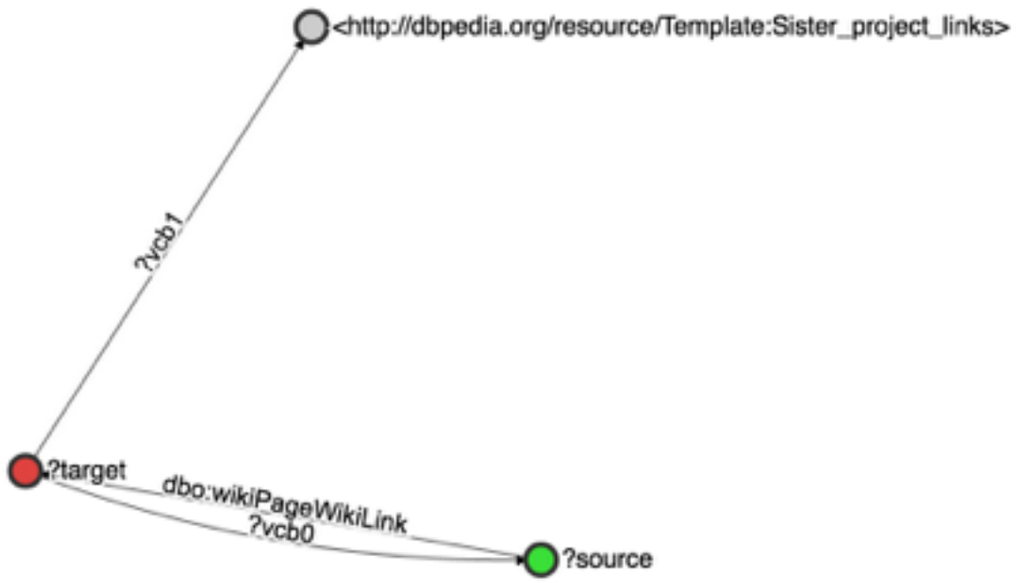
https://w3id.org/associations

results.json.gz#Graph_2

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	3
remains	655
score	30.495
gain	30.495
f_measure	0.129
avg_reslens	2.683
gt_matches	51
patlen	3
patvars	4
timeout	0
qtime	0.768

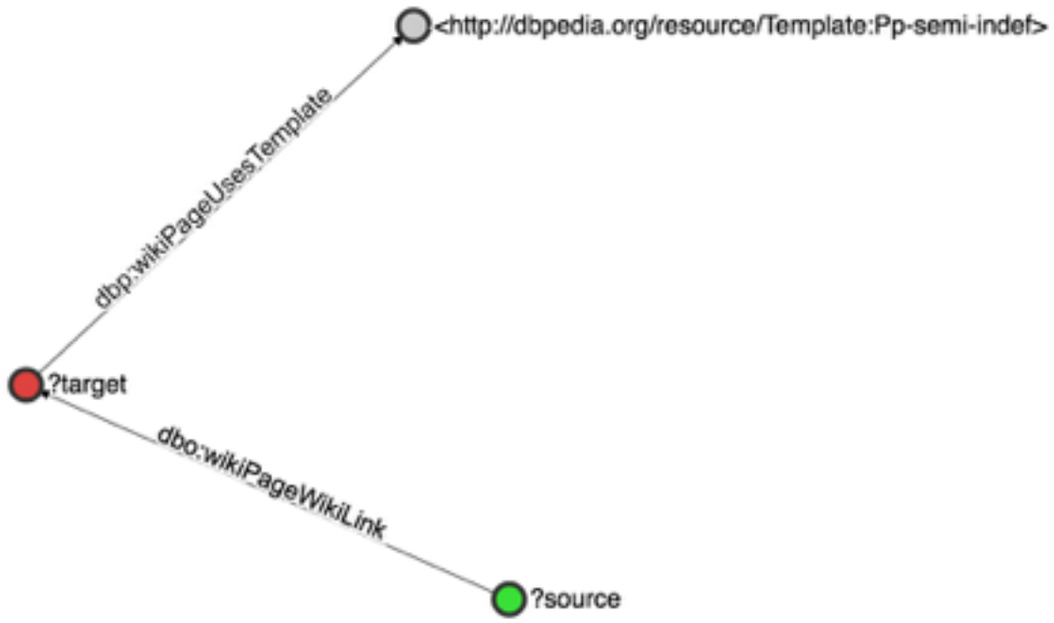
https://w3id.org/associations

results.json.gz#Graph_3

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	4
remains	655
score	28.765
gain	28.765
f_measure	0.15
avg_reslens	5.562
gt_matches	84
patlen	2
patvars	2
timeout	0
qtime	0.798

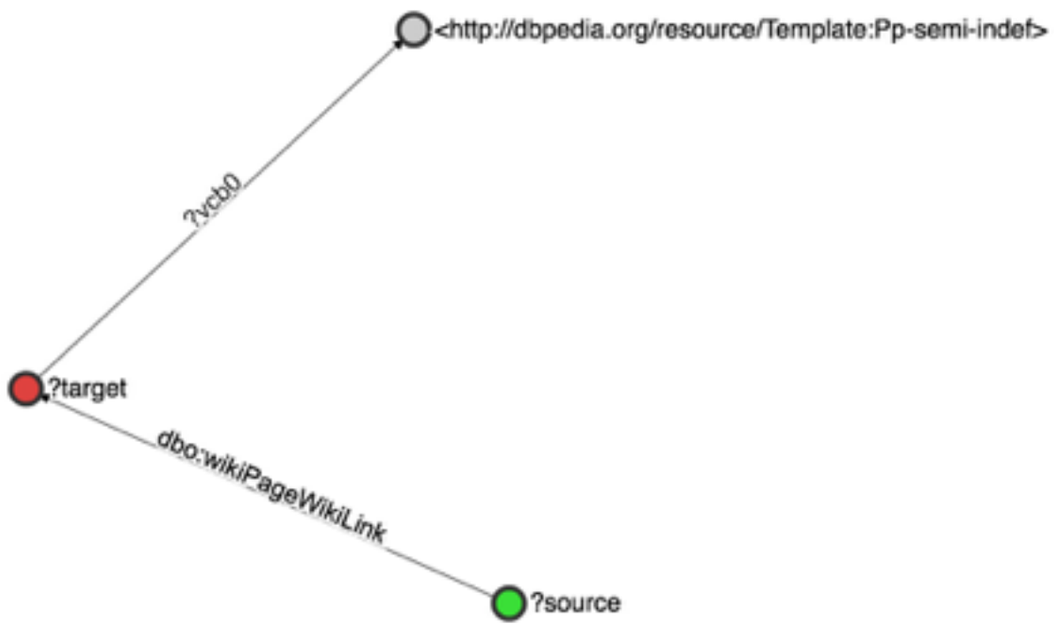
https://w3id.org/associations

results.json.gz#Graph_4

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 1
- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11

Stats about this graph pattern

Pattern number	5
remains	655
score	28.765
gain	28.765
f_measure	0.15
avg_reslens	5.562
gt_matches	84
patlen	2
patvars	3
timeout	0
qtime	0.421

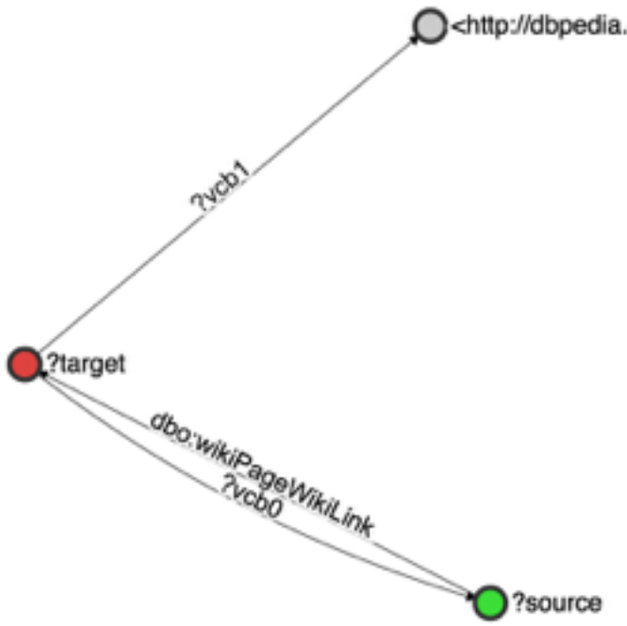
https://w3id.org/associations

results.json.gz#Graph_5

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 2
- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11
- Show graph number 12

Stats about this graph pattern

Pattern number	6
remains	655
score	28.685
gain	28.685
f_measure	0.14
avg_reslens	8.124
gt_matches	106
patlen	3
patvars	4
timeout	0
qtime	1.382

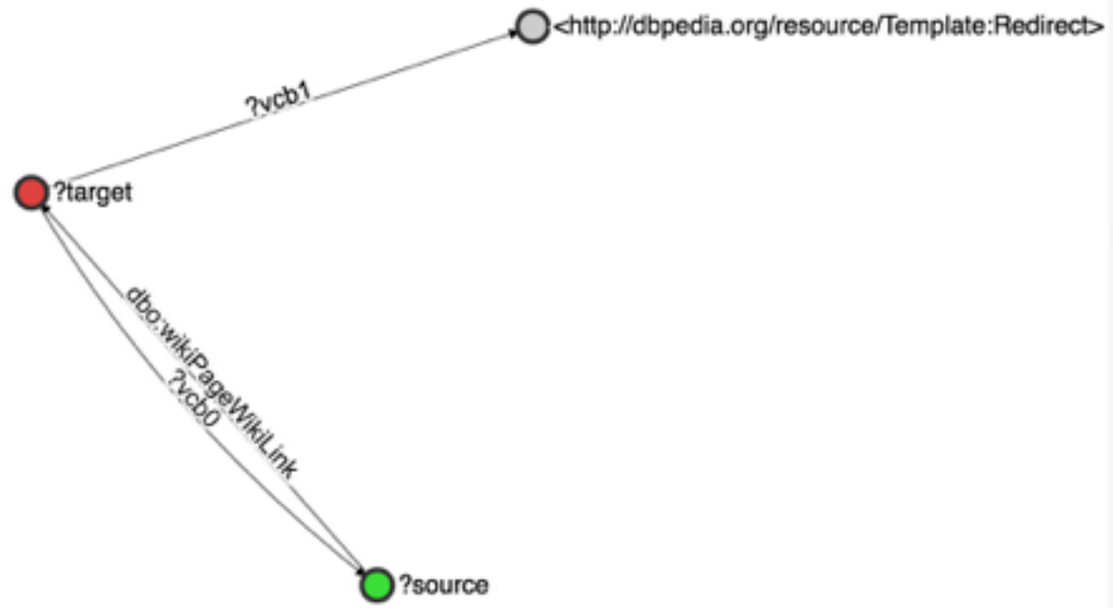
https://w3id.org/associations

results.json.gz#Graph_6

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

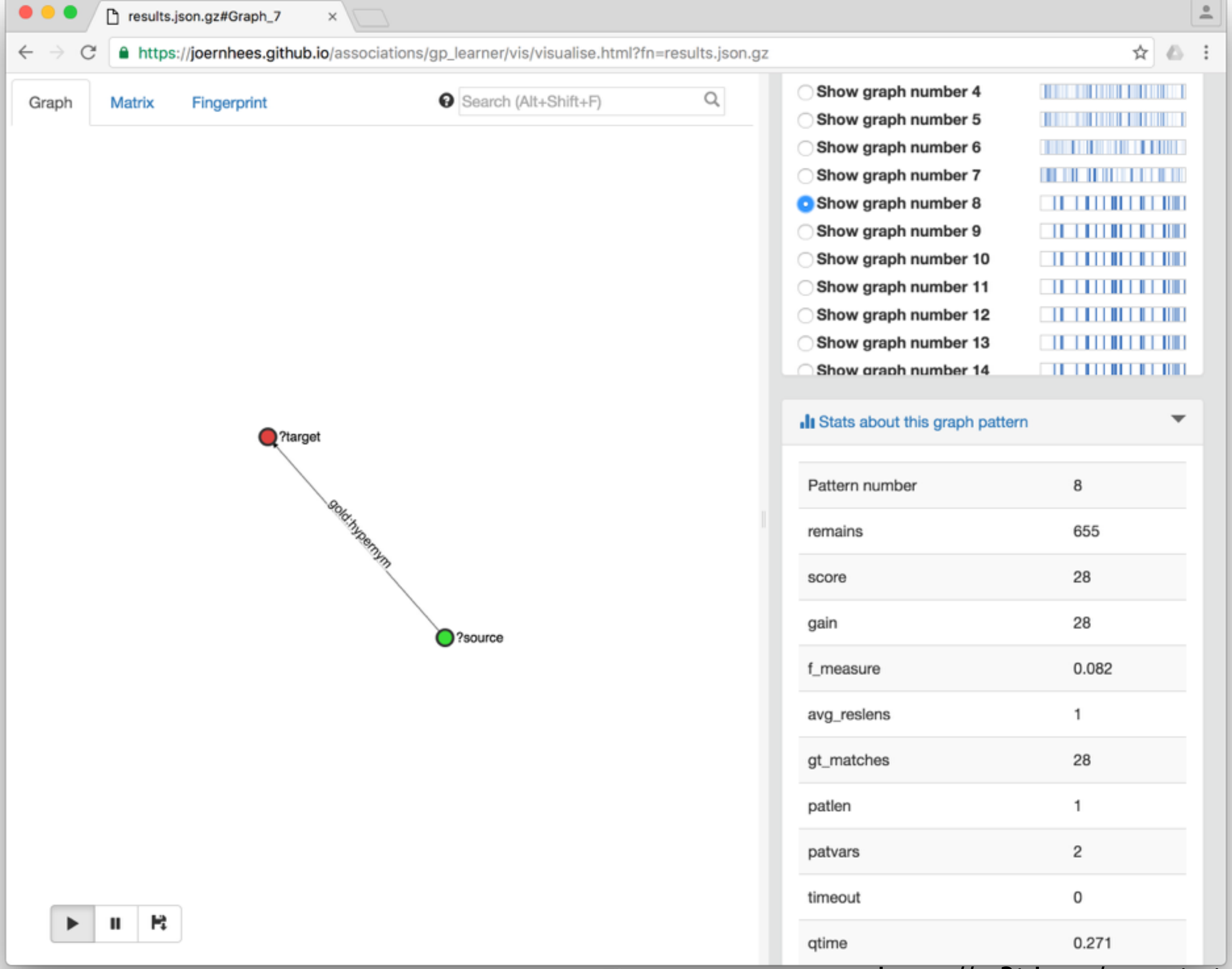


- Show graph number 3
- Show graph number 4
- Show graph number 5
- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11
- Show graph number 12
- Show graph number 13

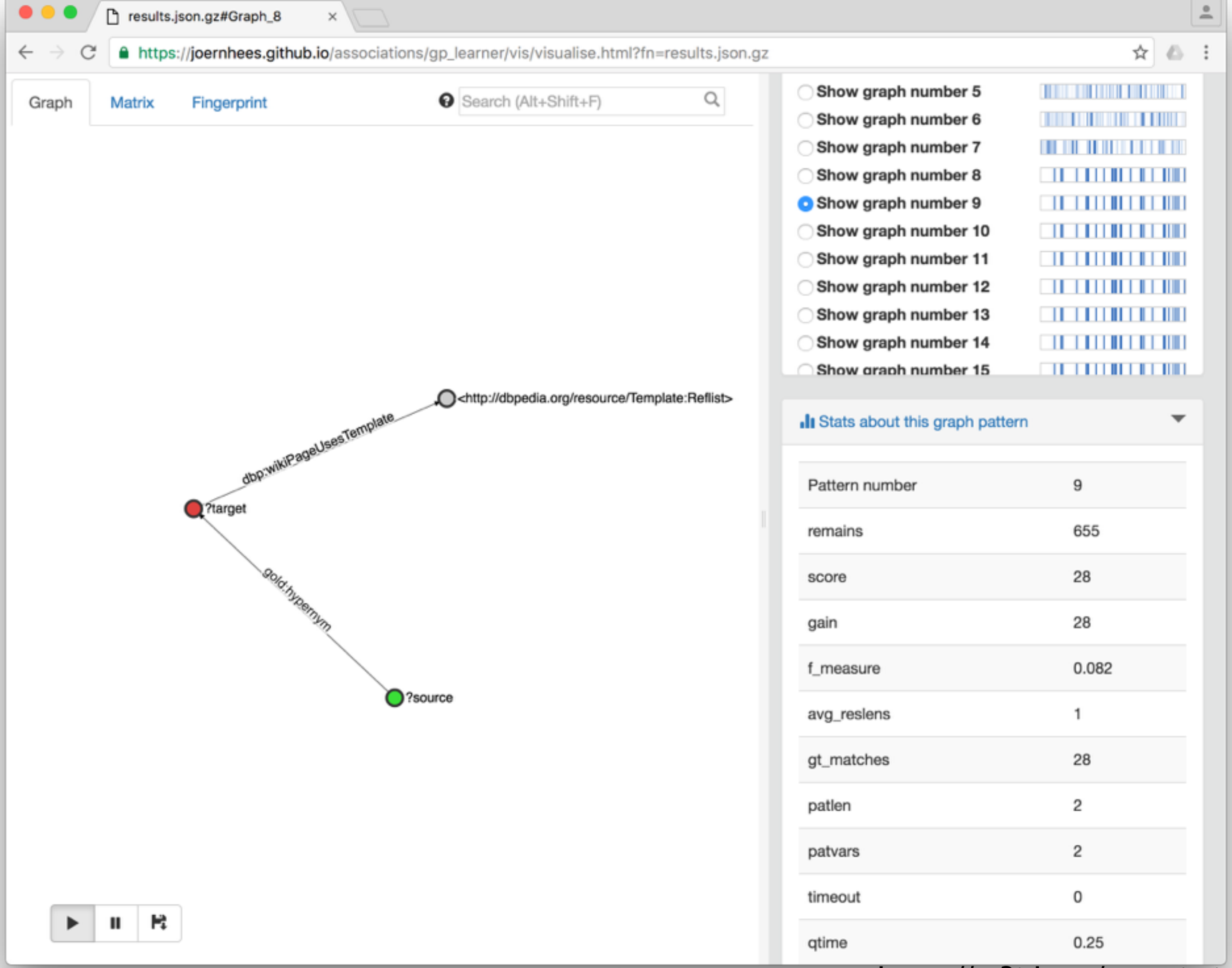
Stats about this graph pattern

Pattern number	7
remains	655
score	28.634
gain	28.634
f_measure	0.148
avg_reslens	4.198
gt_matches	70
patlen	3
patvars	4
timeout	0
qtime	0.967

https://w3id.org/associations



https://w3id.org/associations



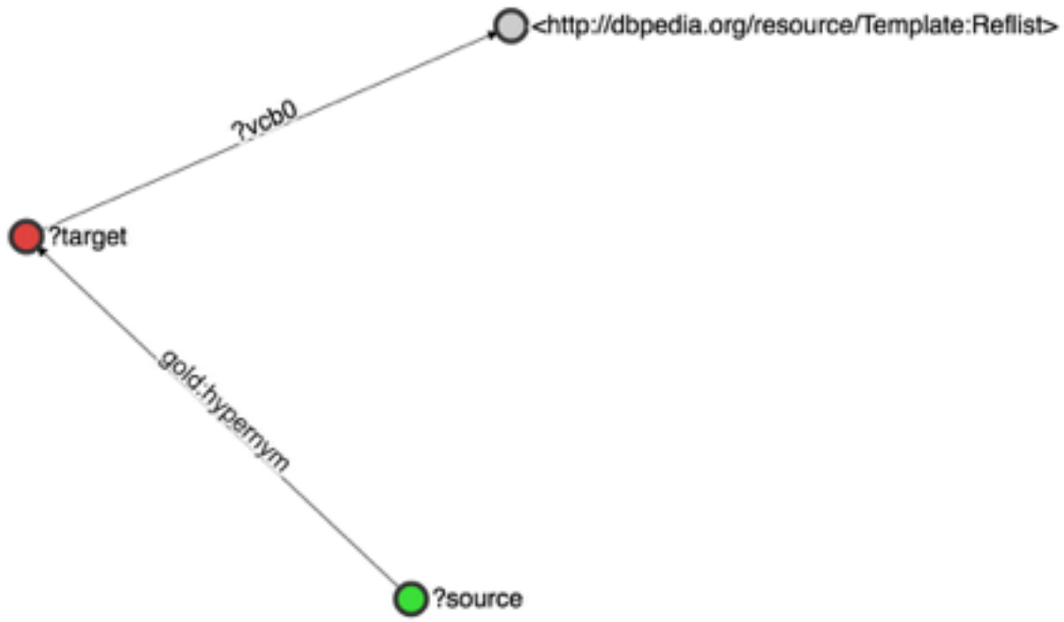
https://w3id.org/associations

results.json.gz#Graph_9

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 6
- Show graph number 7
- Show graph number 8
- Show graph number 9
- Show graph number 10
- Show graph number 11
- Show graph number 12
- Show graph number 13
- Show graph number 14
- Show graph number 15
- Show graph number 16

Stats about this graph pattern

Pattern number	10
remains	655
score	28
gain	28
f_measure	0.082
avg_reslens	1
gt_matches	28
patlen	2
patvars	3
timeout	0
qtime	0.231

https://w3id.org/associations

results.json.gz#Graph_69

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

- Show graph number 66
- Show graph number 67
- Show graph number 68
- Show graph number 69
- Show graph number 70
- Show graph number 71
- Show graph number 72
- Show graph number 73
- Show graph number 74
- Show graph number 75
- Show graph number 76

Stats about this graph pattern

Pattern number	70
remains	655
score	21
gain	21
f_measure	0.062
avg_reslens	1
gt_matches	21
patlen	3
patvars	2
timeout	0
qtime	0.308

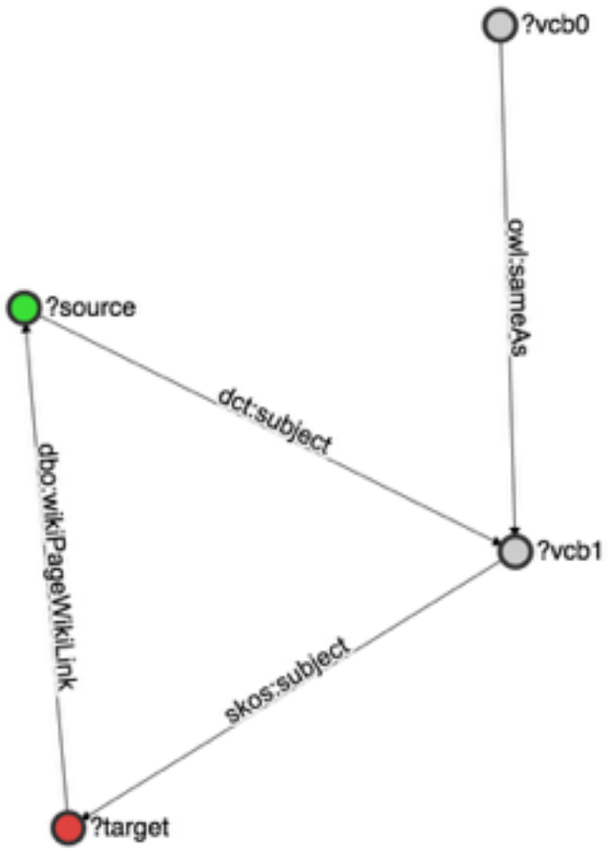
https://w3id.org/associations

results.json.gz#Graph_71

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 68
- Show graph number 69
- Show graph number 70
- Show graph number 71
- Show graph number 72
- Show graph number 73
- Show graph number 74
- Show graph number 75
- Show graph number 76
- Show graph number 77
- Show graph number 78

Stats about this graph pattern

Pattern number	72
remains	478.434
score	8.667
gain	8.667
f_measure	0.07
avg_reslens	1.214
gt_matches	24
patlen	4
patvars	4
timeout	0
qtime	0.257

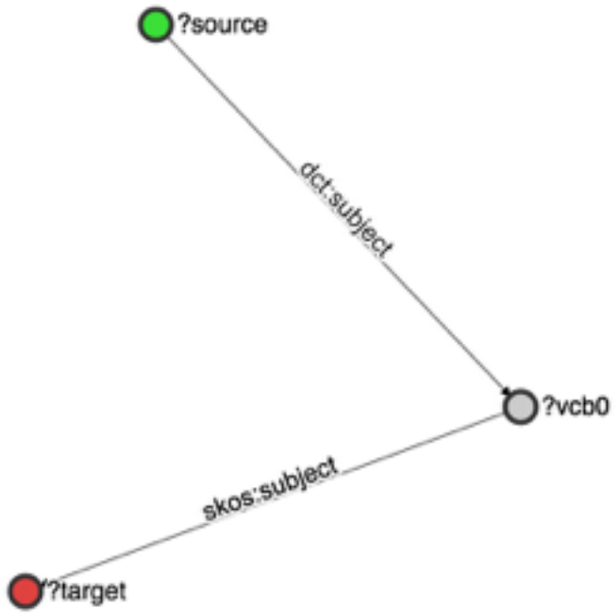
https://w3id.org/associations

results.json.gz#Graph_132

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 129
- Show graph number 130
- Show graph number 131
- Show graph number 132
- Show graph number 133
- Show graph number 134
- Show graph number 135
- Show graph number 136
- Show graph number 137
- Show graph number 138
- Show graph number 139

Stats about this graph pattern

Pattern number	133
remains	478.434
score	6.36
gain	6.36
f_measure	0.102
avg_reslens	1.868
gt_matches	37
patlen	2
patvars	3
timeout	0
qtime	0.318

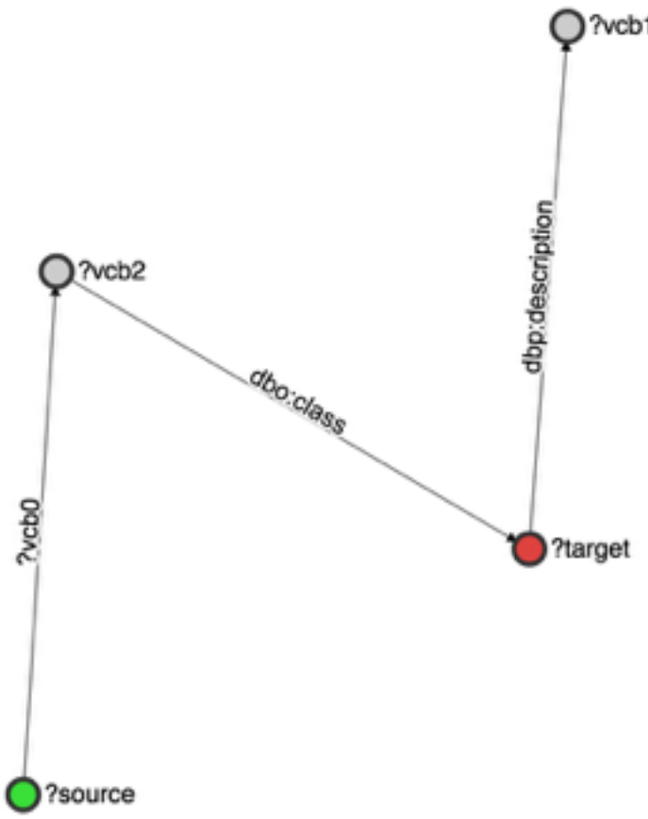
https://w3id.org/associations

results.json.gz#Graph_194

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 191
- Show graph number 192
- Show graph number 193
- Show graph number 194
- Show graph number 195
- Show graph number 196
- Show graph number 197
- Show graph number 198
- Show graph number 199
- Show graph number 200
- Show graph number 201

Stats about this graph pattern

Pattern number	195
remains	385.048
score	3.583
gain	3.583
f_measure	0.042
avg_reslens	1.198
gt_matches	14
patlen	3
patvars	5
timeout	0
qtime	0.474

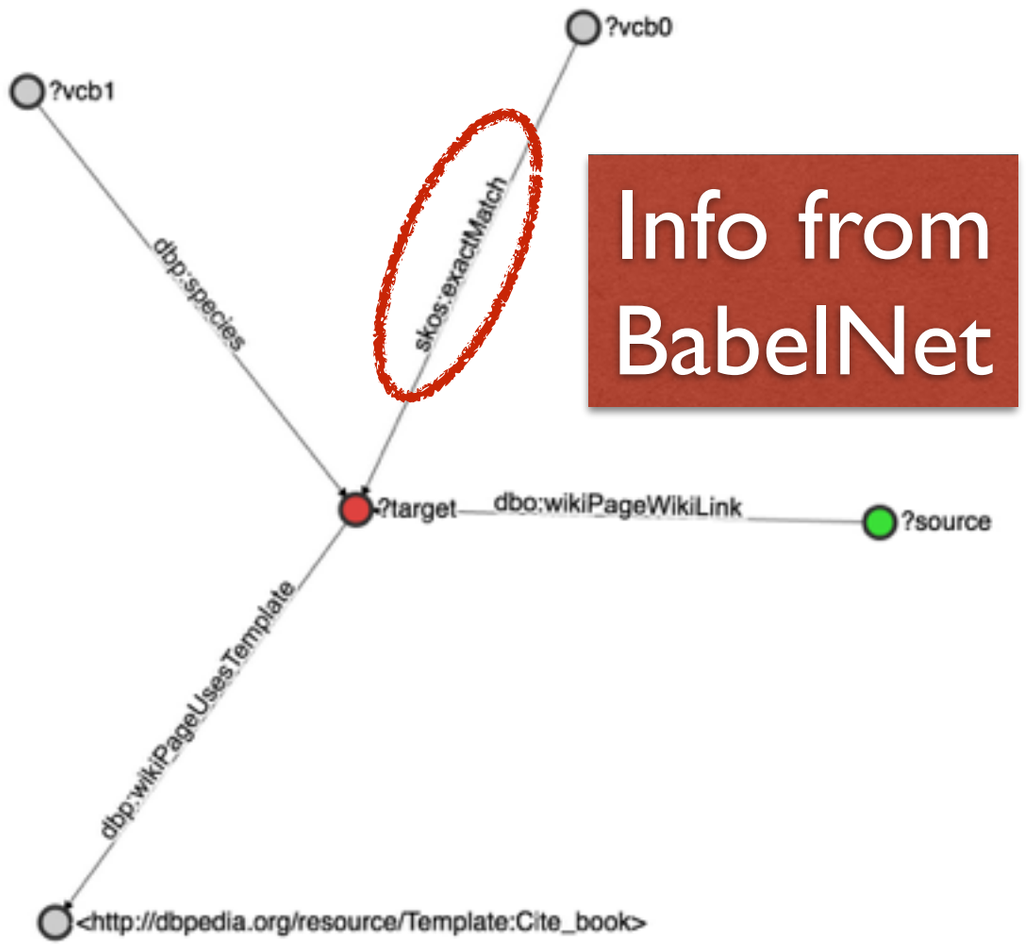
https://w3id.org/associations

results.json.gz#Graph_281

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



Info from BabelNet

- Show graph number 278
- Show graph number 279
- Show graph number 280
- Show graph number 281
- Show graph number 282
- Show graph number 283
- Show graph number 284
- Show graph number 285
- Show graph number 286
- Show graph number 287
- Show graph number 288

Stats about this graph pattern

Pattern number	282
remains	377.948
score	5.842
gain	5.842
f_measure	0.084
avg_reslens	1.919
gt_matches	30
patlen	4
patvars	4
timeout	0
qtime	0.682

https://w3id.org/associations

results.json.gz#Graph_293

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

Info from WikiData

- Show graph number 290
- Show graph number 291
- Show graph number 292
- Show graph number 293
- Show graph number 294
- Show graph number 295
- Show graph number 296
- Show graph number 297
- Show graph number 298
- Show graph number 299
- Show graph number 300

Stats about this graph pattern

Pattern number	294
remains	377.948
score	4.909
gain	4.909
f_measure	0.079
avg_reslens	1.856
gt_matches	28
patlen	4
patvars	5
timeout	0
qtime	0.52

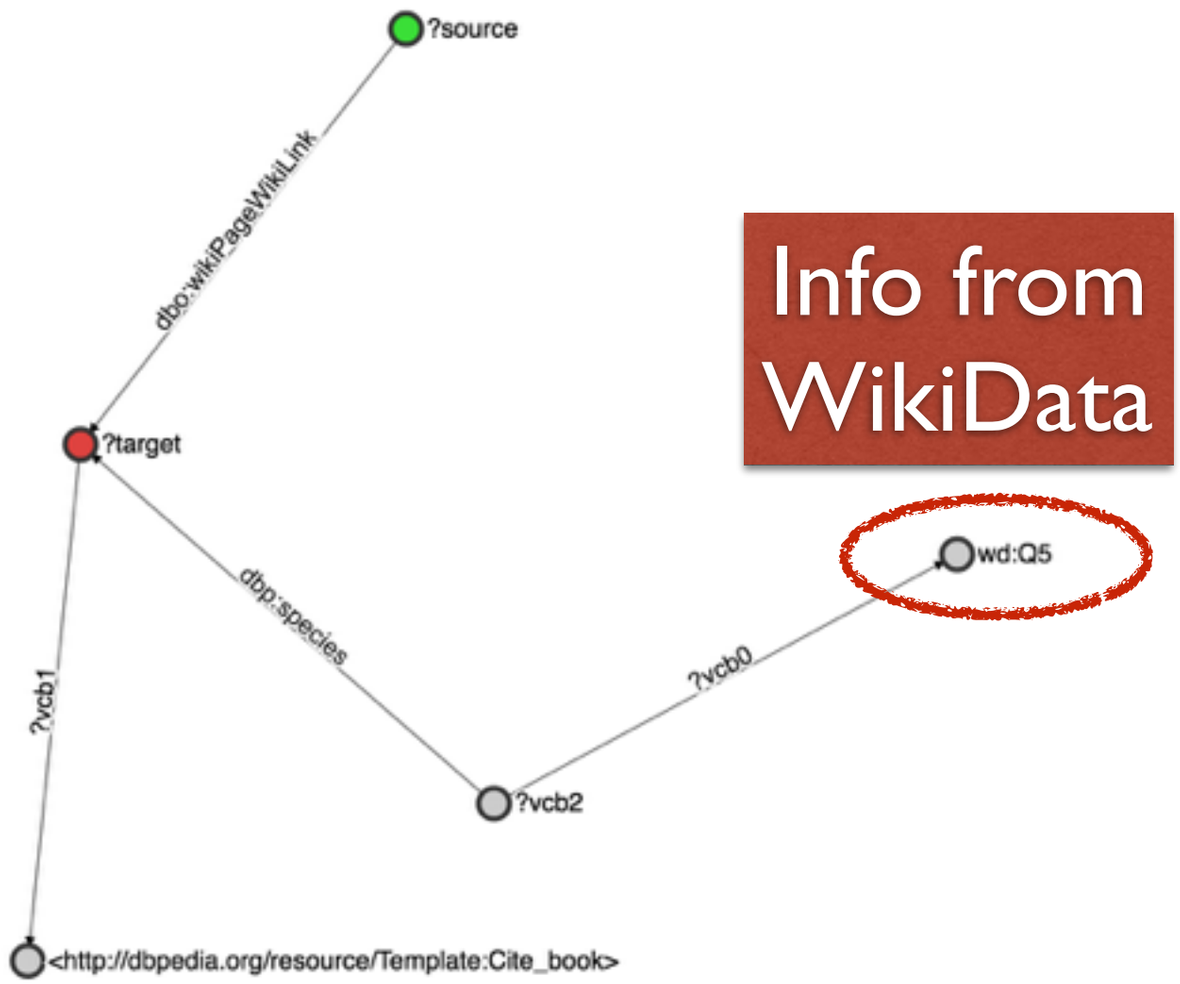
https://w3id.org/associations

results.json.gz#Graph_295

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



Info from WikiData

- Show graph number 292
- Show graph number 293
- Show graph number 294
- Show graph number 295
- Show graph number 296
- Show graph number 297
- Show graph number 298
- Show graph number 299
- Show graph number 300
- Show graph number 301
- Show graph number 302

Stats about this graph pattern

Pattern number	296
remains	377.948
score	4.909
gain	4.909
f_measure	0.079
avg_reslens	1.856
gt_matches	28
patlen	4
patvars	5
timeout	0
qtime	0.559

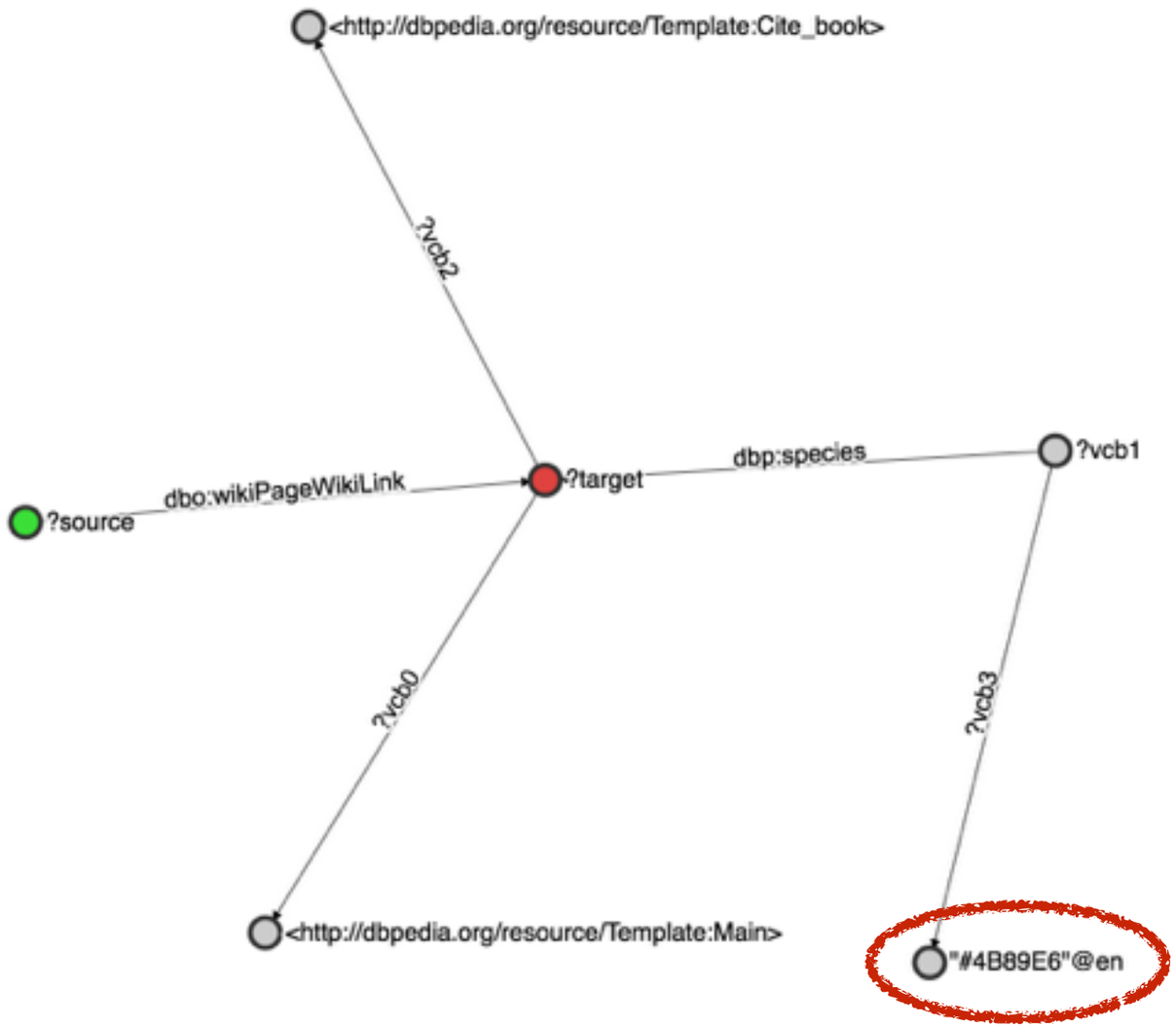
https://w3id.org/associations

results.json.gz#Graph_302

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 299
- Show graph number 300
- Show graph number 301
- Show graph number 302
- Show graph number 303
- Show graph number 304
- Show graph number 305
- Show graph number 306
- Show graph number 307
- Show graph number 308
- Show graph number 309

Stats about this graph pattern

Pattern number	303
remains	377.948
score	4.242
gain	4.242
f_measure	0.05
avg_reslens	1.422
gt_matches	17
patlen	5
patvars	6
timeout	0
qtime	0.474

https://w3id.org/associations

results.json.gz#Graph_316

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)

Info from BabelNet

- Show graph number 313
- Show graph number 314
- Show graph number 315
- Show graph number 316
- Show graph number 317
- Show graph number 318
- Show graph number 319
- Show graph number 320
- Show graph number 321
- Show graph number 322
- Show graph number 323

Stats about this graph pattern

Pattern number	317
remains	377.948
score	3.759
gain	3.759
f_measure	0.122
avg_reslens	2.746
gt_matches	48
patlen	3
patvars	4
timeout	0
qtime	0.517

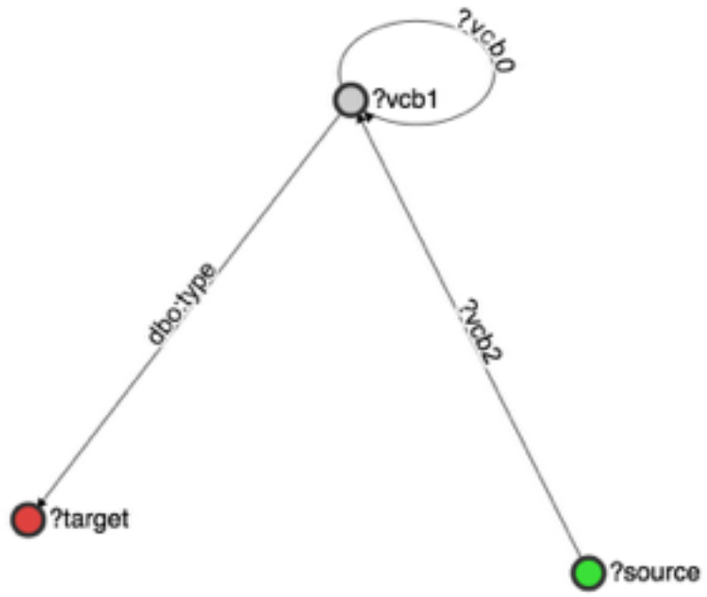
https://w3id.org/associations

results.json.gz#Graph_343

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 340
- Show graph number 341
- Show graph number 342
- Show graph number 343
- Show graph number 344
- Show graph number 345
- Show graph number 346
- Show graph number 347
- Show graph number 348
- Show graph number 349
- Show graph number 350

Stats about this graph pattern

Pattern number	344
remains	364.049
score	3.667
gain	3.667
f_measure	0.024
avg_reslens	2.806
gt_matches	8
patlen	3
patvars	5
timeout	0
qtime	0.84

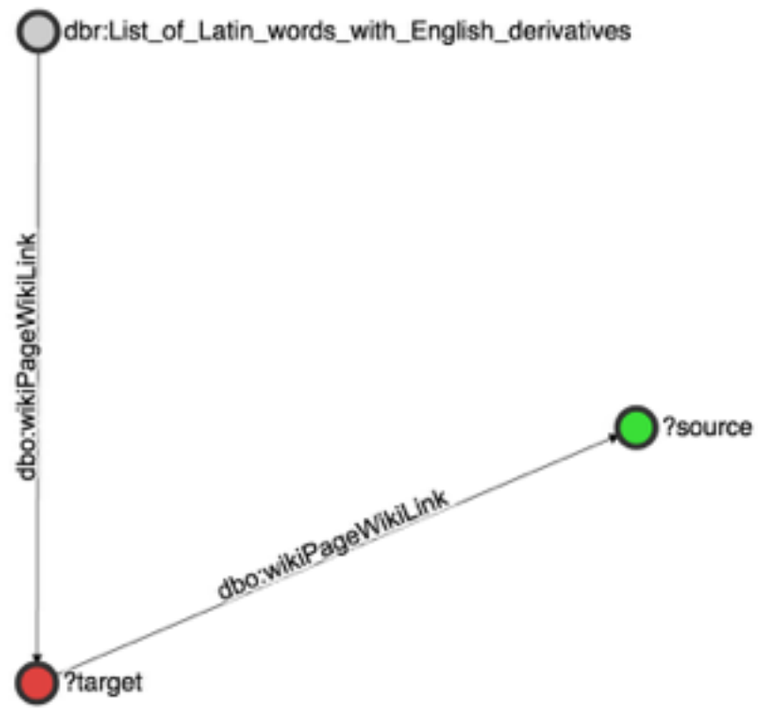
https://w3id.org/associations

results.json.gz#Graph_397

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 394
- Show graph number 395
- Show graph number 396
- Show graph number 397
- Show graph number 398
- Show graph number 399
- Show graph number 400
- Show graph number 401
- Show graph number 402
- Show graph number 403
- Show graph number 404

Stats about this graph pattern

Pattern number	398
remains	320.612
score	5.801
gain	5.801
f_measure	0.172
avg_reslens	6.511
gt_matches	128
patlen	2
patvars	2
timeout	0
qtime	1.071

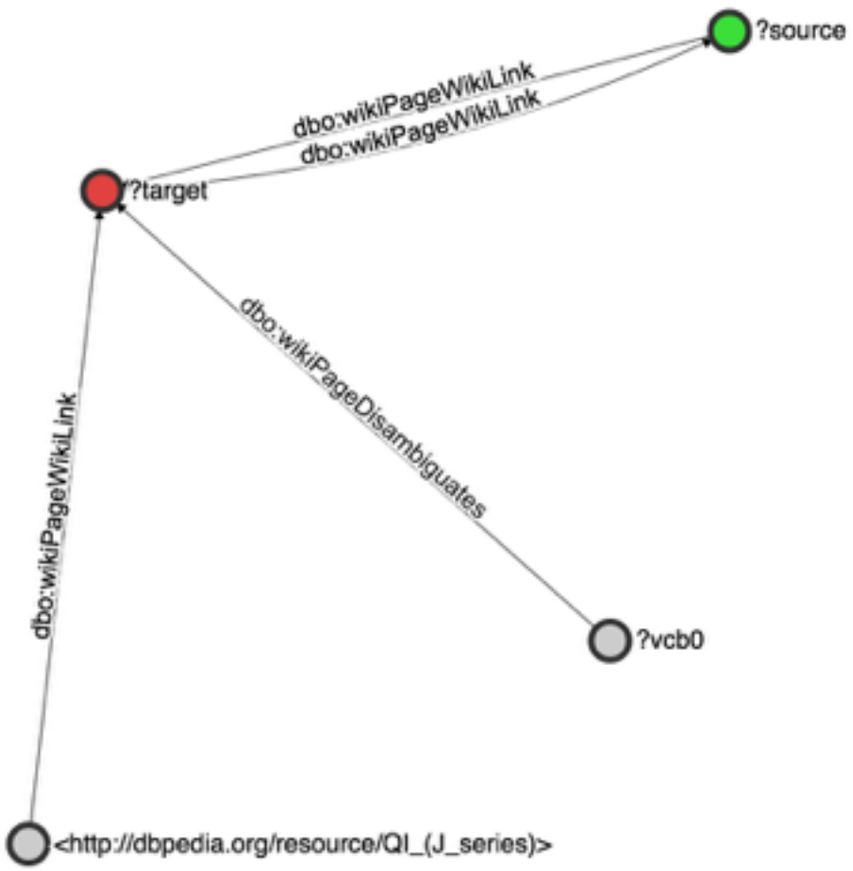
https://w3id.org/associations

results.json.gz#Graph_404

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 401
- Show graph number 402
- Show graph number 403
- Show graph number 404
- Show graph number 405
- Show graph number 406
- Show graph number 407
- Show graph number 408
- Show graph number 409
- Show graph number 410
- Show graph number 411

Stats about this graph pattern

Pattern number	405
remains	320.612
score	4.764
gain	4.764
f_measure	0.204
avg_reslens	2.006
gt_matches	84
patlen	4
patvars	3
timeout	0
qtime	0.42

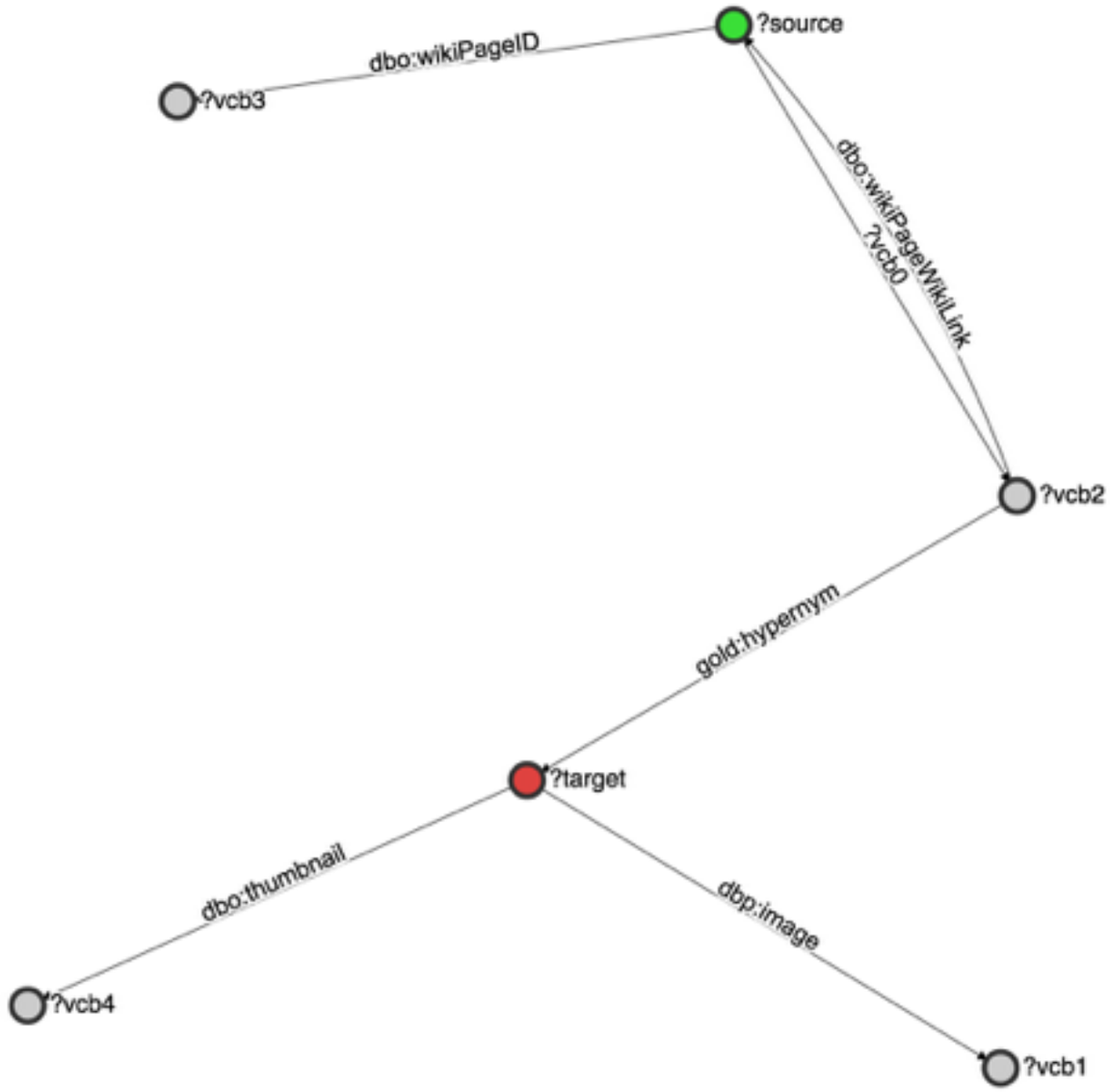
https://w3id.org/associations

results.json.gz#Graph_466

https://joernhees.github.io/associations/gp_learner/vis/visualise.html?fn=results.json.gz

Graph | Matrix | Fingerprint

Search (Alt+Shift+F)



- Show graph number 463
- Show graph number 464
- Show graph number 465
- Show graph number 466
- Show graph number 467
- Show graph number 468
- Show graph number 469
- Show graph number 470
- Show graph number 471
- Show graph number 472
- Show graph number 473

Stats about this graph pattern

Pattern number	467
remains	262.539
score	4.333
gain	4.333
f_measure	0.058
avg_reslens	1.51
gt_matches	20
patlen	6
patvars	7
timeout	0
qtime	1.463

https://w3id.org/associations

Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

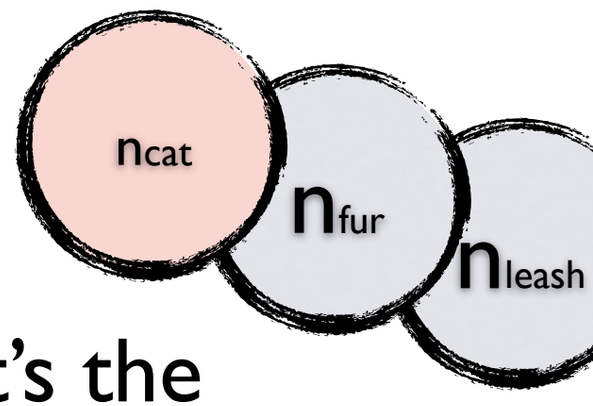
Outline

- Background
- My Research (Demo)
- Graph Pattern Learning
- Evaluation

https://w3id.org/associations

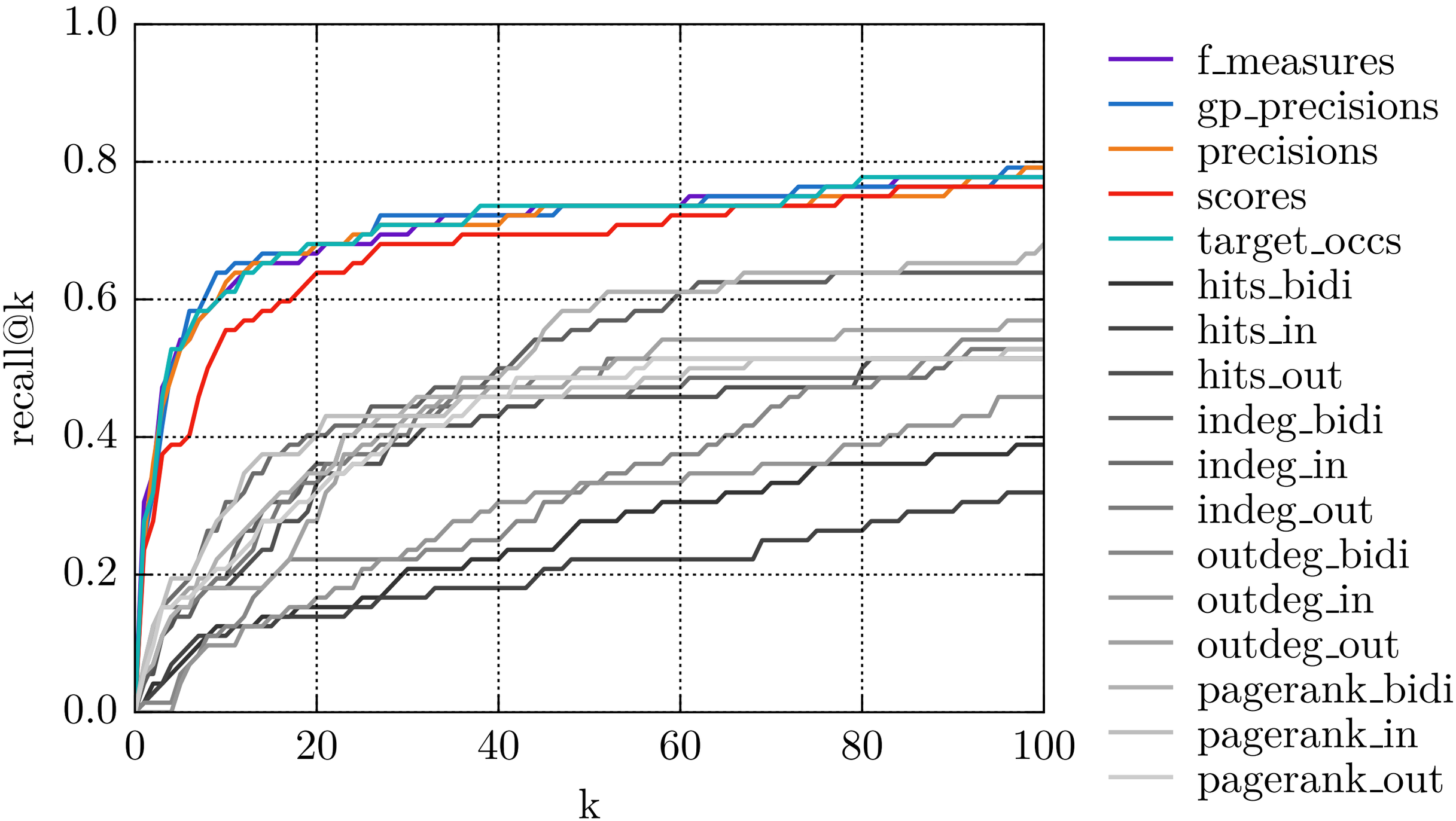
Evaluation

- How good are the learned patterns?
  - Difficult to evaluate directly
- Indirect objective approach:
  - Are they good for prediction?
- Training/Test set split
- Clustered similar GPs
- Given a stimulus from the test set, what's the rank of the true response in the prediction results?



https://w3id.org/associations

Evaluation Results



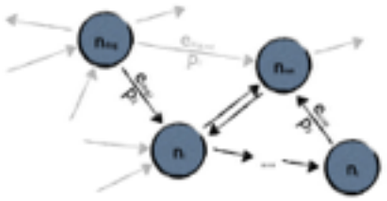
https://w3id.org/associations

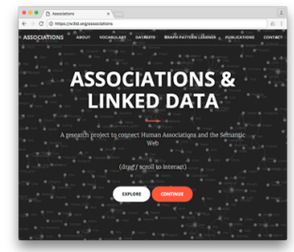
Evaluation Results

	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	Recall@10	MAP	NDCG
outdeg in	0.000	0.000	0.000	0.000	0.042	0.097	0.029	0.105
outdeg out	**0.069**	**0.125**	**0.153**	0.153	0.167	0.181	0.126	0.209
outdeg bidi	0.014	0.014	0.014	0.014	0.056	0.125	0.045	0.131
indeg in	0.056	0.111	**0.153**	0.167	0.181	**0.306**	0.129	0.207
indeg out	0.056	**0.125**	**0.153**	0.153	0.153	0.194	0.121	0.200
indeg bidi	0.042	0.069	0.111	0.139	0.139	0.194	0.104	0.205
pagerank in	**0.069**	**0.125**	**0.153**	**0.194**	**0.194**	0.292	**0.140**	**0.219**
pagerank out	0.056	0.097	**0.153**	0.153	0.167	0.208	0.117	0.195
pagerank bidi	0.056	0.069	0.111	0.139	0.153	0.236	0.113	**0.219**
hits in	0.014	0.028	0.042	0.069	0.083	0.111	0.046	0.095
hits out	0.056	0.056	0.111	0.125	0.153	0.181	0.102	0.181
hits bidi	0.014	0.042	0.042	0.056	0.069	0.125	0.050	0.110
scores	0.236	0.278	0.375	0.389	0.389	0.556	0.323	0.413
gp precisions	0.250	0.319	0.417	0.500	0.528	**0.639**	0.365	0.457
precisions	0.250	**0.361**	0.444	0.486	0.528	0.625	0.371	0.460
target occs	0.278	0.319	0.458	**0.528**	0.528	0.611	0.381	0.466
f measures	**0.306**	0.347	**0.472**	0.500	**0.542**	0.611	**0.399**	**0.479**

Summary

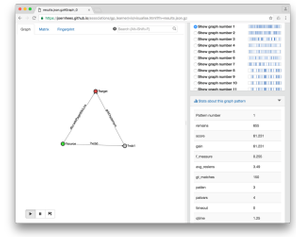
- Goal
  Learning Graph Patterns for Associations

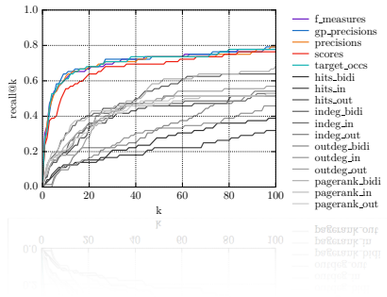




- Datasets

- Evolutionary Algorithm
  Learns SPARQL Patterns for
  Source-Target-Pairs (> 60% Top-10 Accuracy)

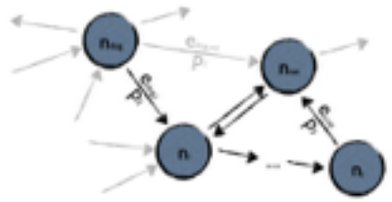




https://w3id.org/associations

Future Work

- Apply Evolutionary Algorithm
  - to other datasets
  - to other types of relations
- Extensions:
  - Work on Literals
  - Include FILTER

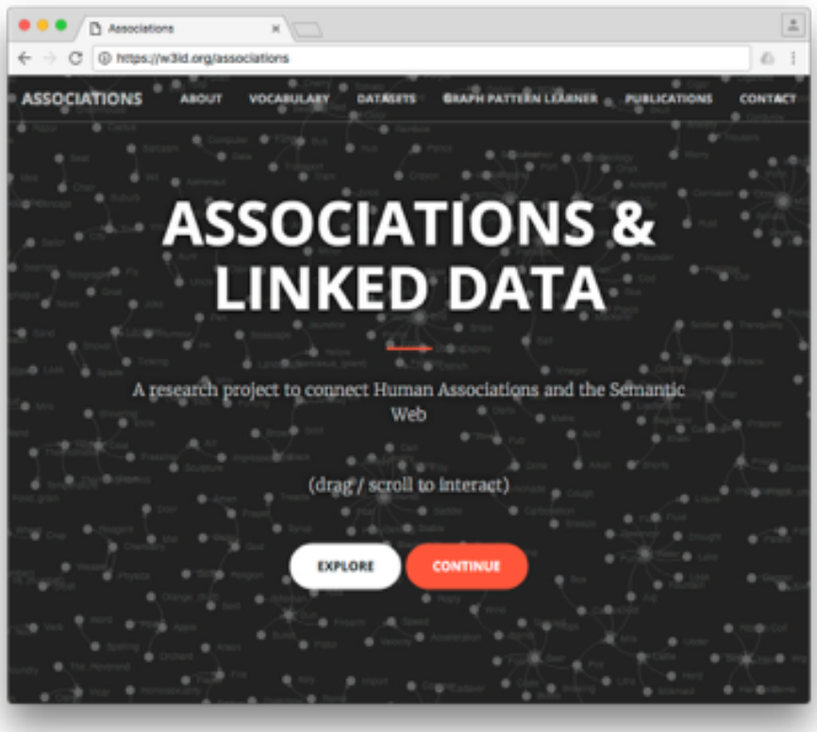


https://w3id.org/associations

Discussion

Thanks for your attention

Questions?



https://w3id.org/associations