TECHNISCHE
UNIVERSITÄT
DARMSTADT
UKP Lab
Darmstadt, Germany

Graph Neural Networks for Knowledge Base Question Answering

Daniil Sorokin
@daniilmagpie 👋
https://daniilsorokin.github.io

NLP Meeting
Alan Turing Institute
December 6, 2018

Me

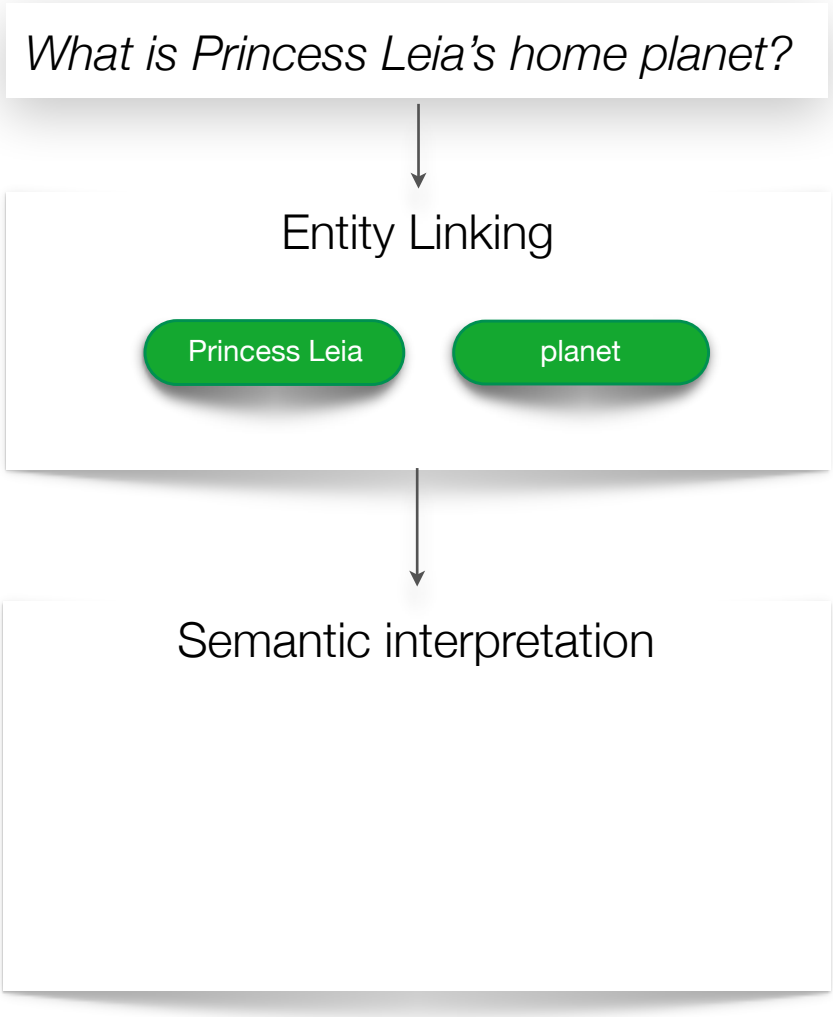
Bachelor in Applied Linguistics
at St. Petersburg State University, Russia
2011

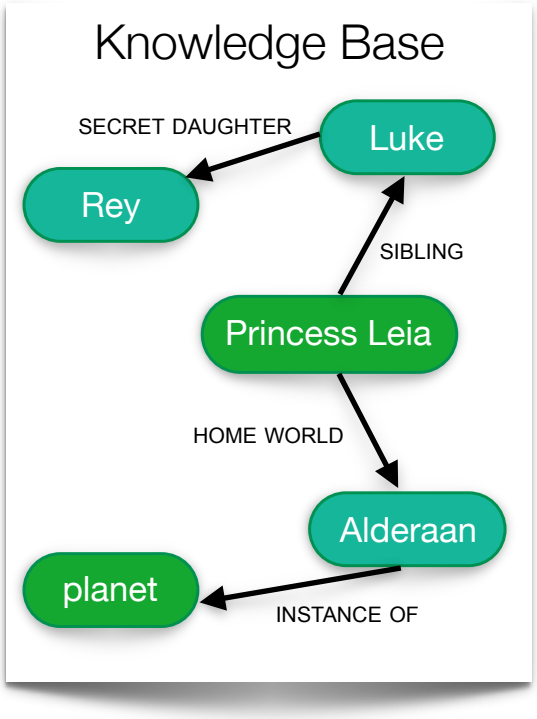
Master in Computational Linguistics
at University of Tübingen, Germany, 2014

PhD in NLP
at the UKP Lab,
Darmstadt University of Technology, Germany,
since 2015,
http://ukp.tu-darmstadt.de

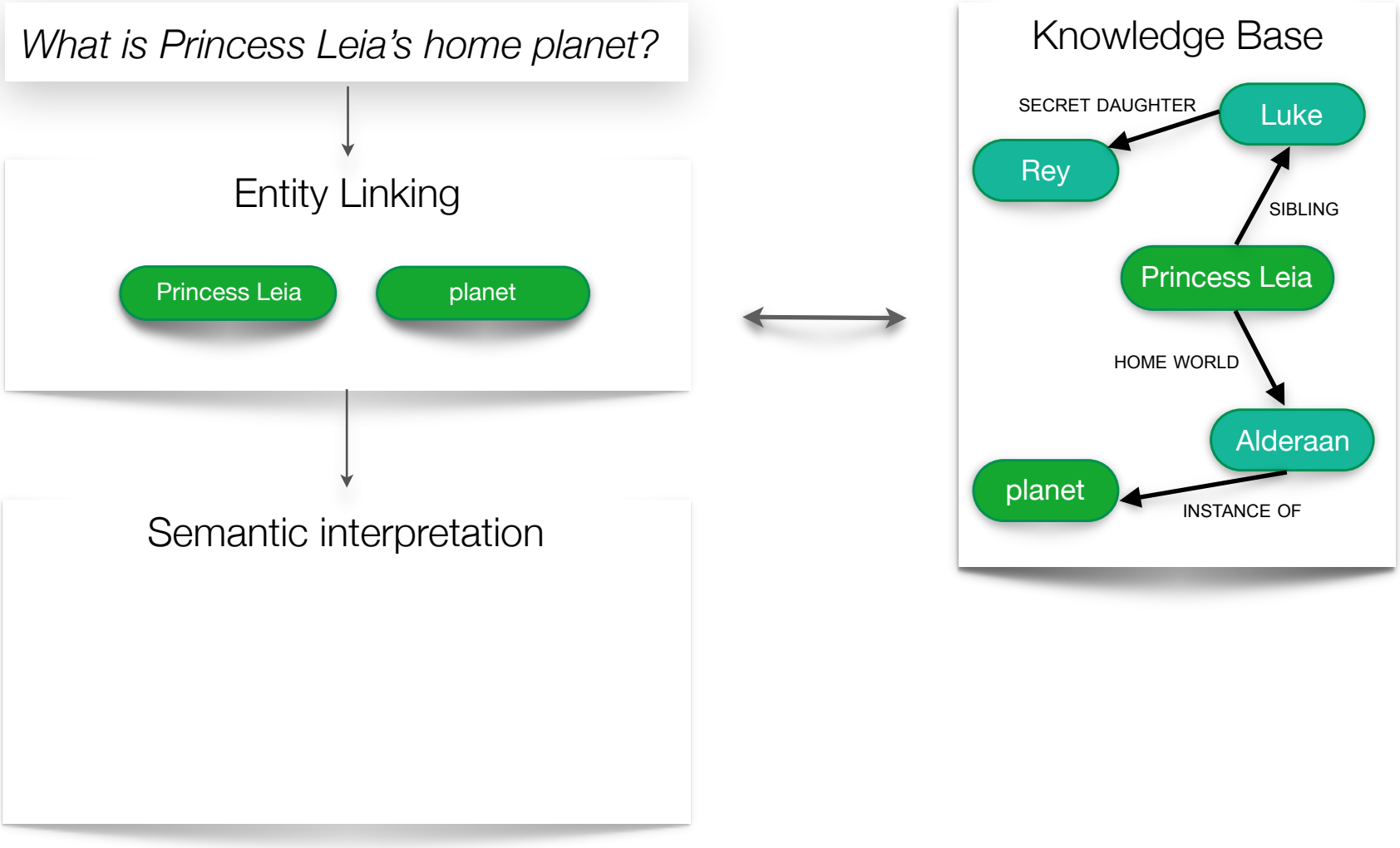
Research intern at Factmata, London
November 2018 - January 2019,
https://factmata.com

Knowledge Base Question Answering

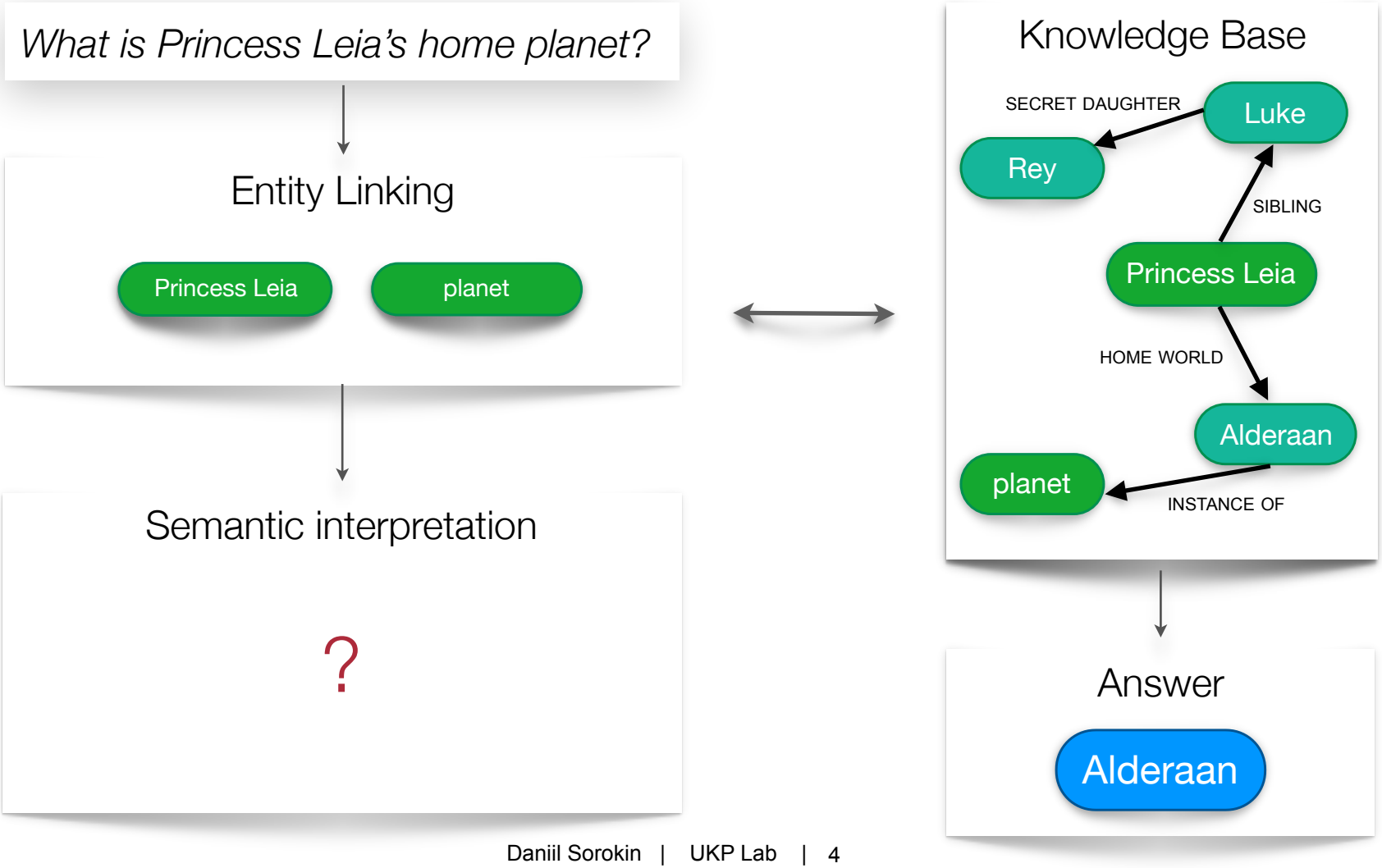




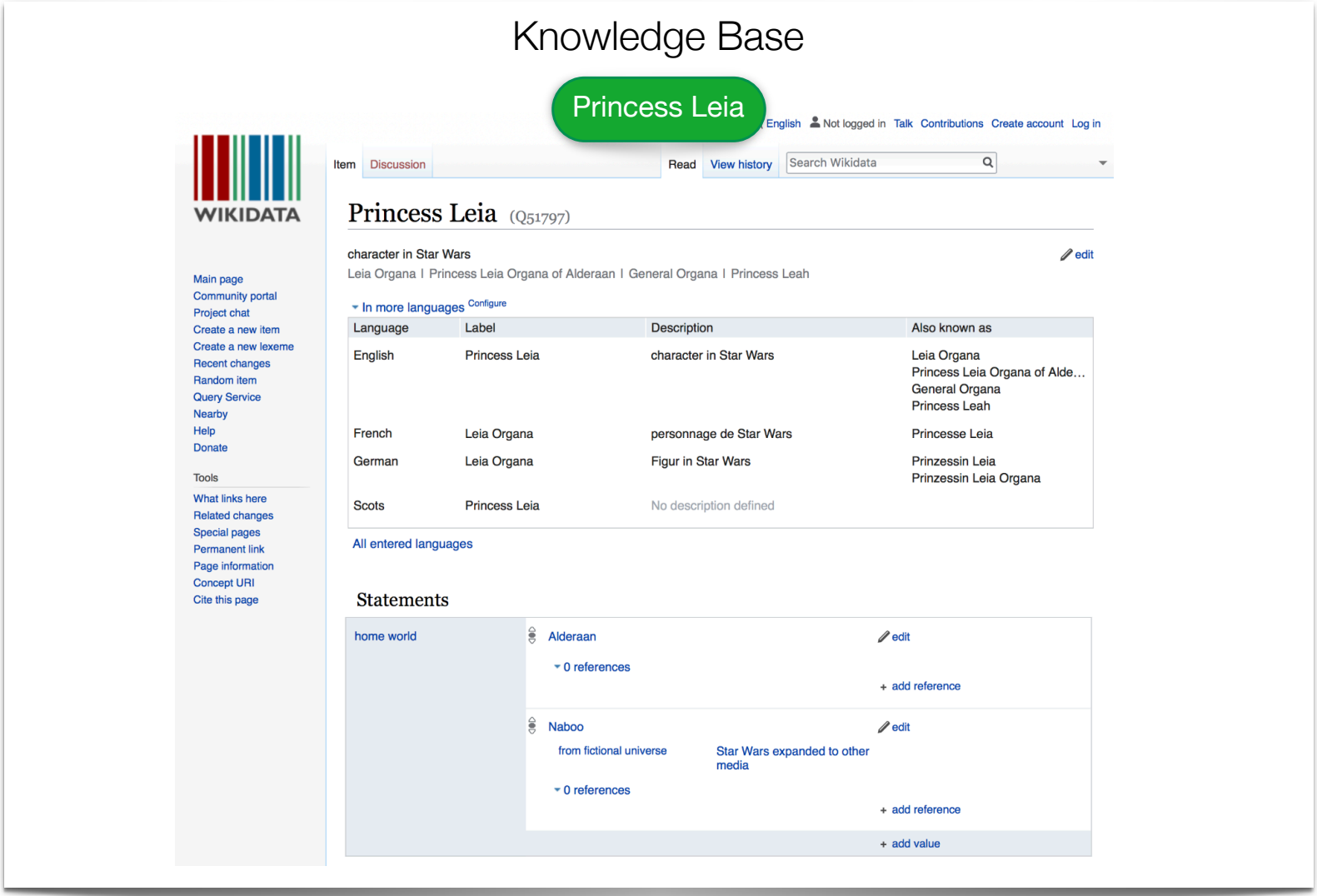
Knowledge Base Question Answering



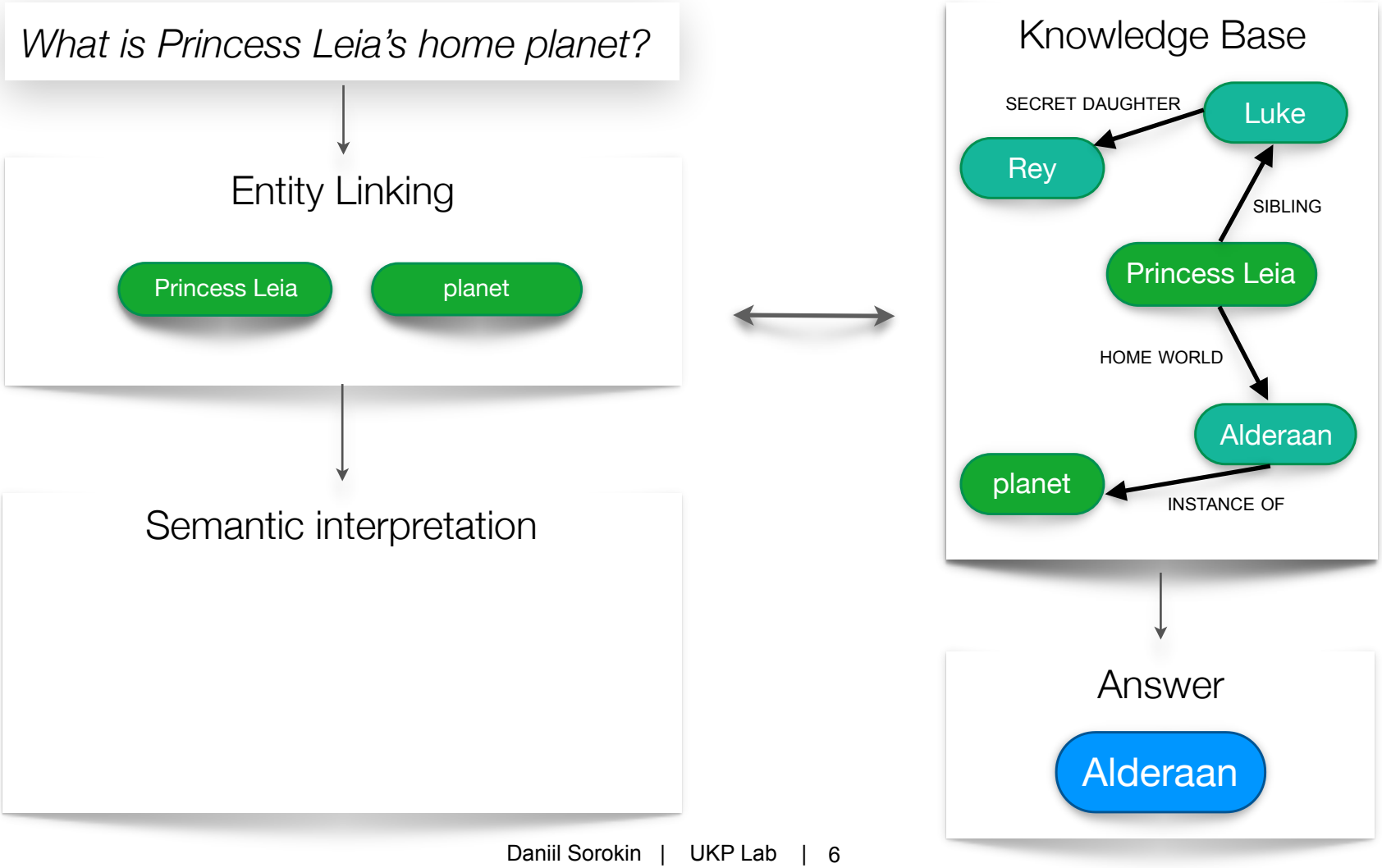
Knowledge Base Question Answering



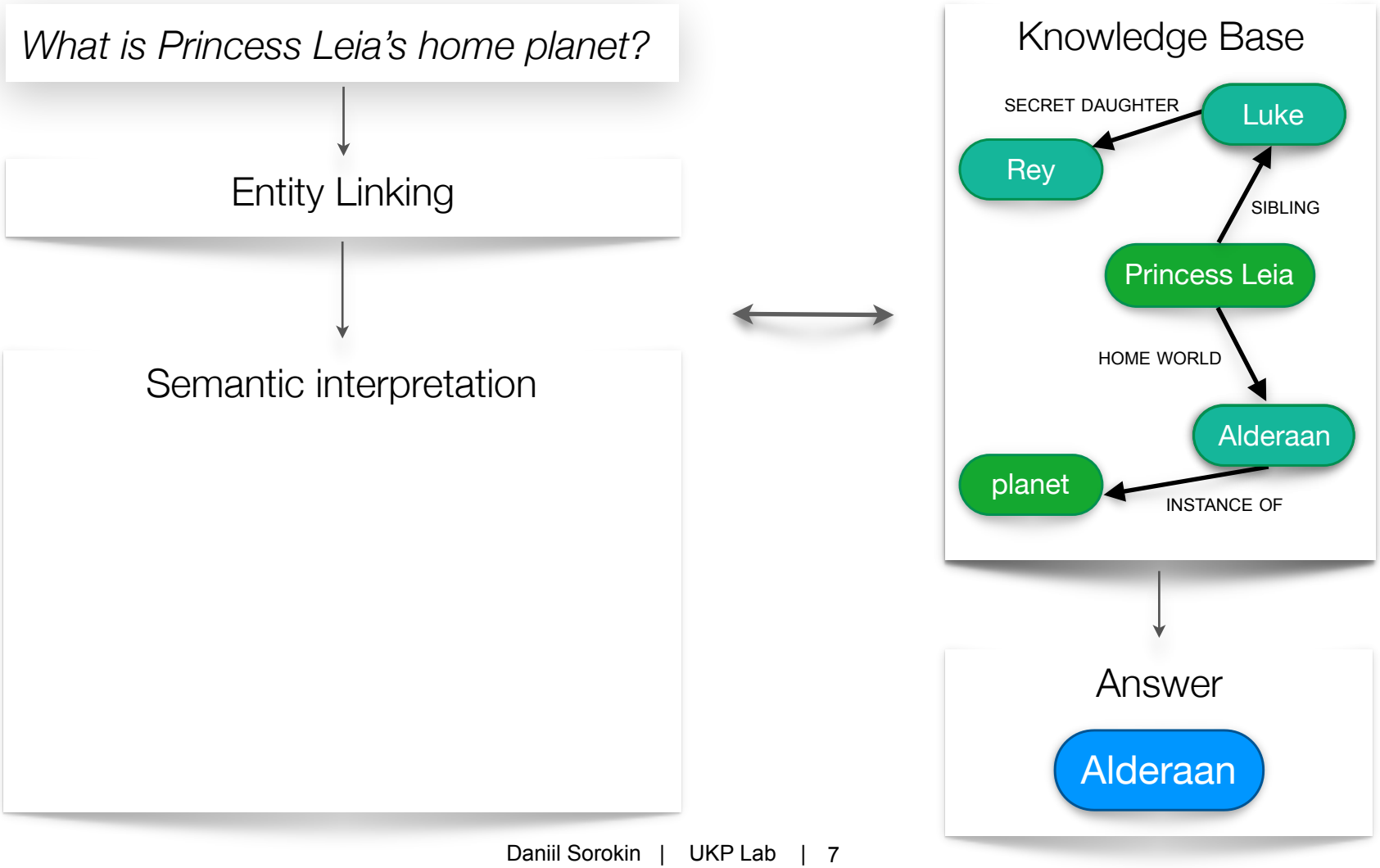
Knowledge Base



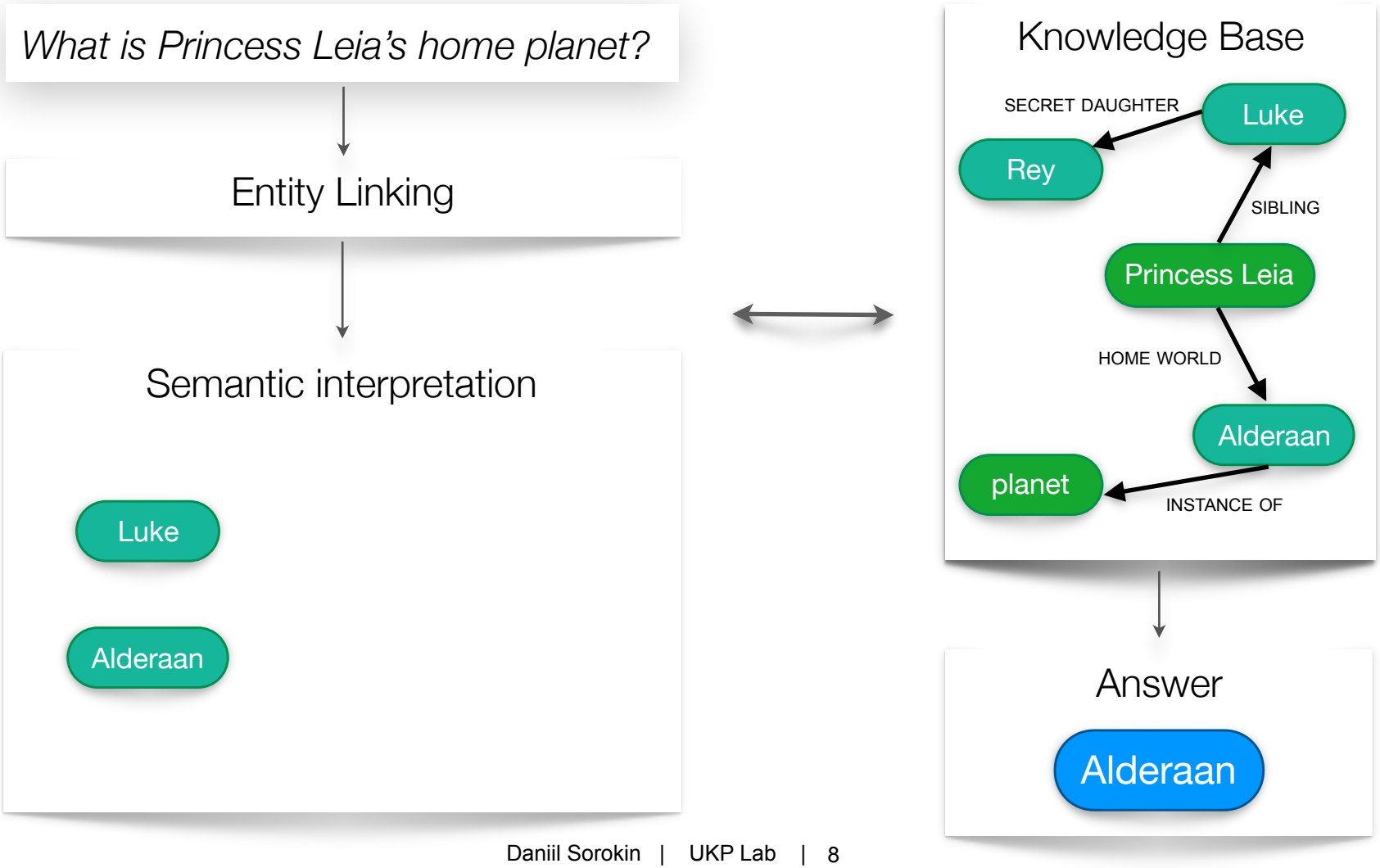
Knowledge Base Question Answering



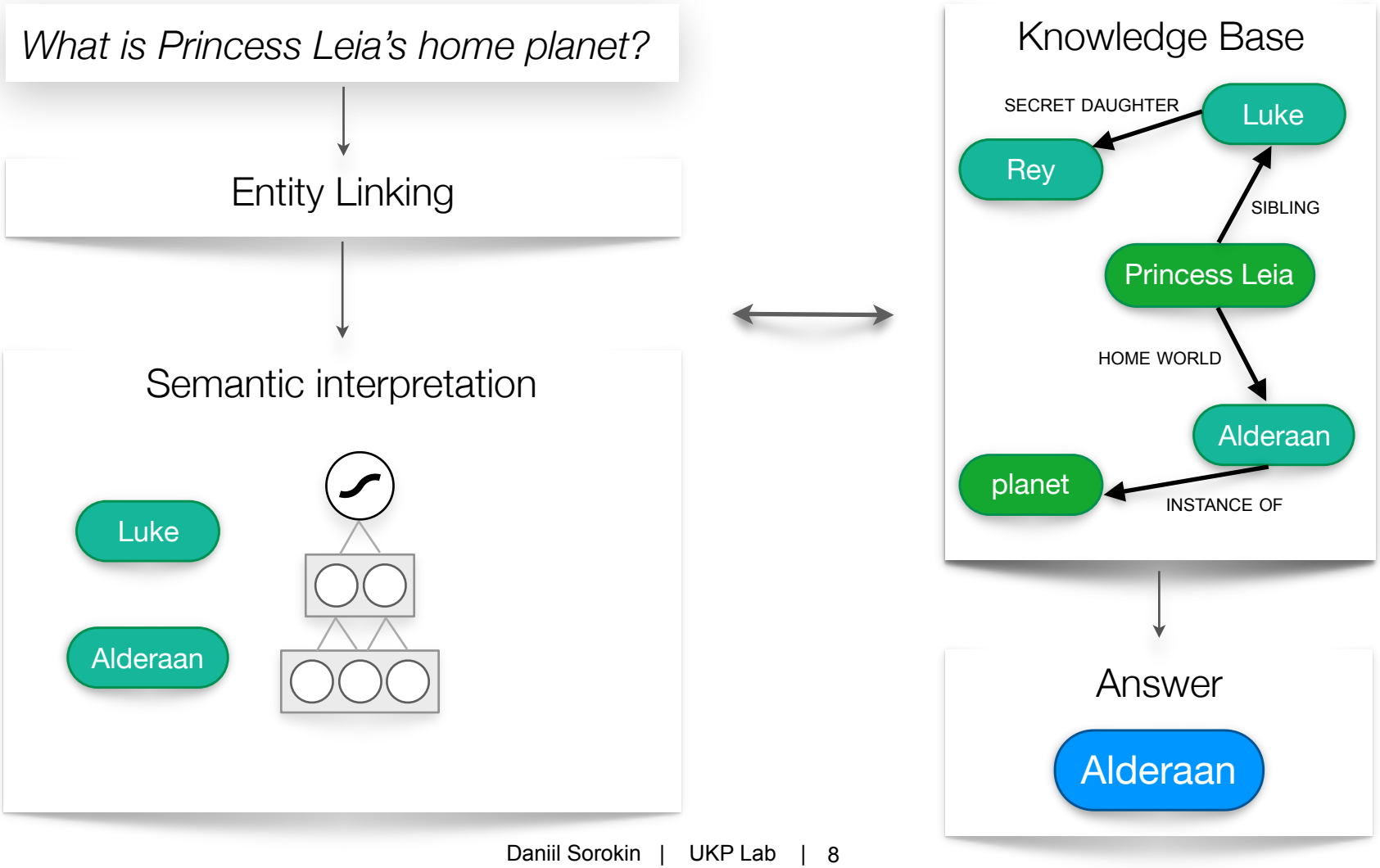
Knowledge Base Question Answering



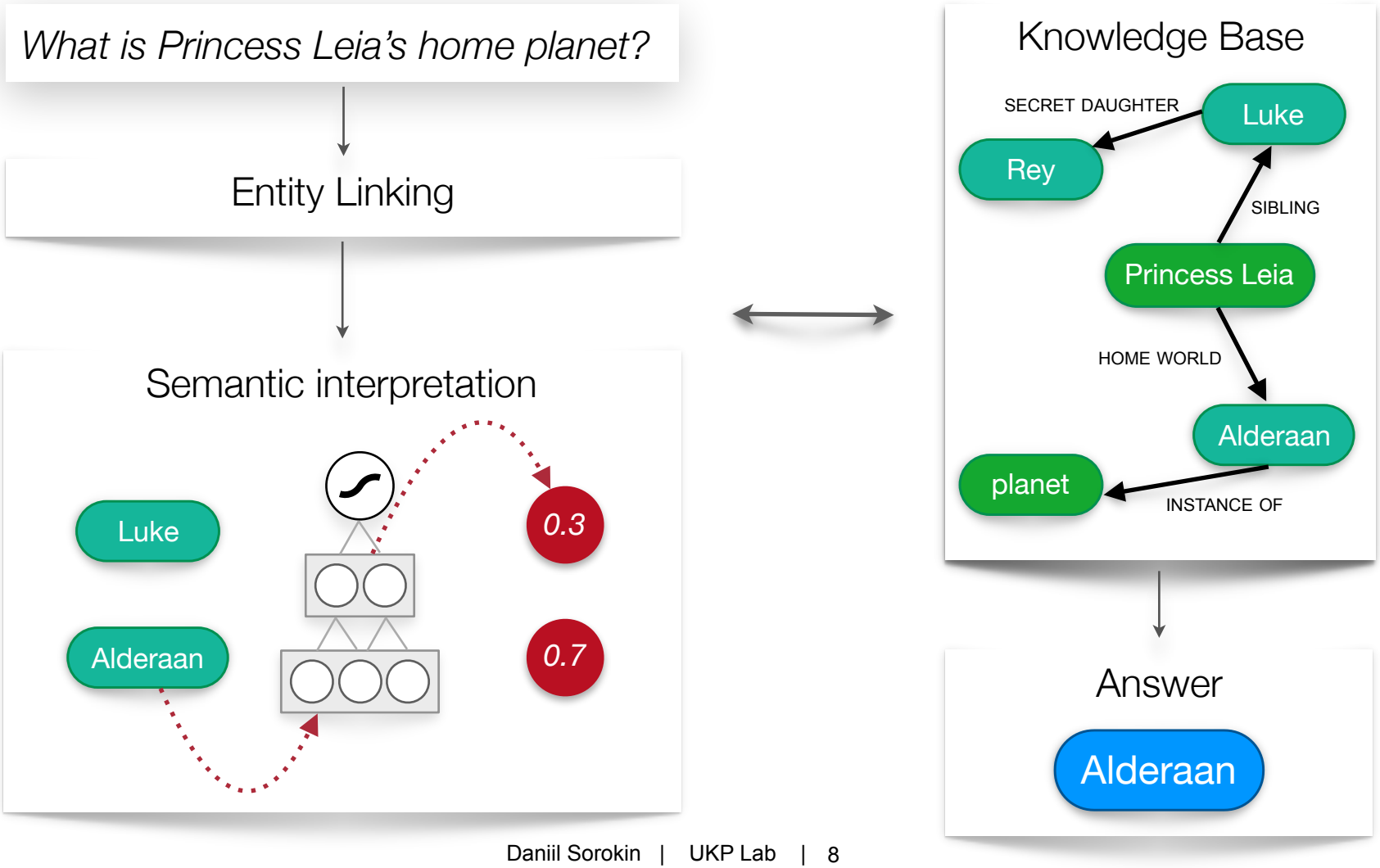
Knowledge Base Question Answering



Knowledge Base Question Answering



Knowledge Base Question Answering



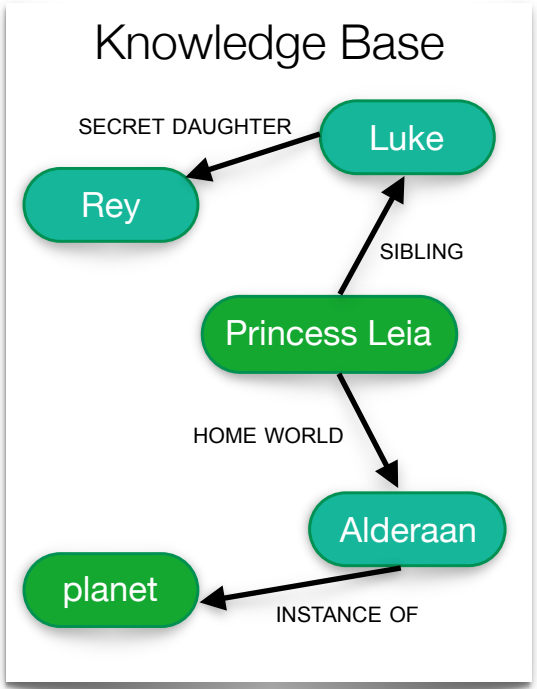
Knowledge Base Question Answering

What is Princess Leia's home planet?

Entity Linking

Semantic interpretation

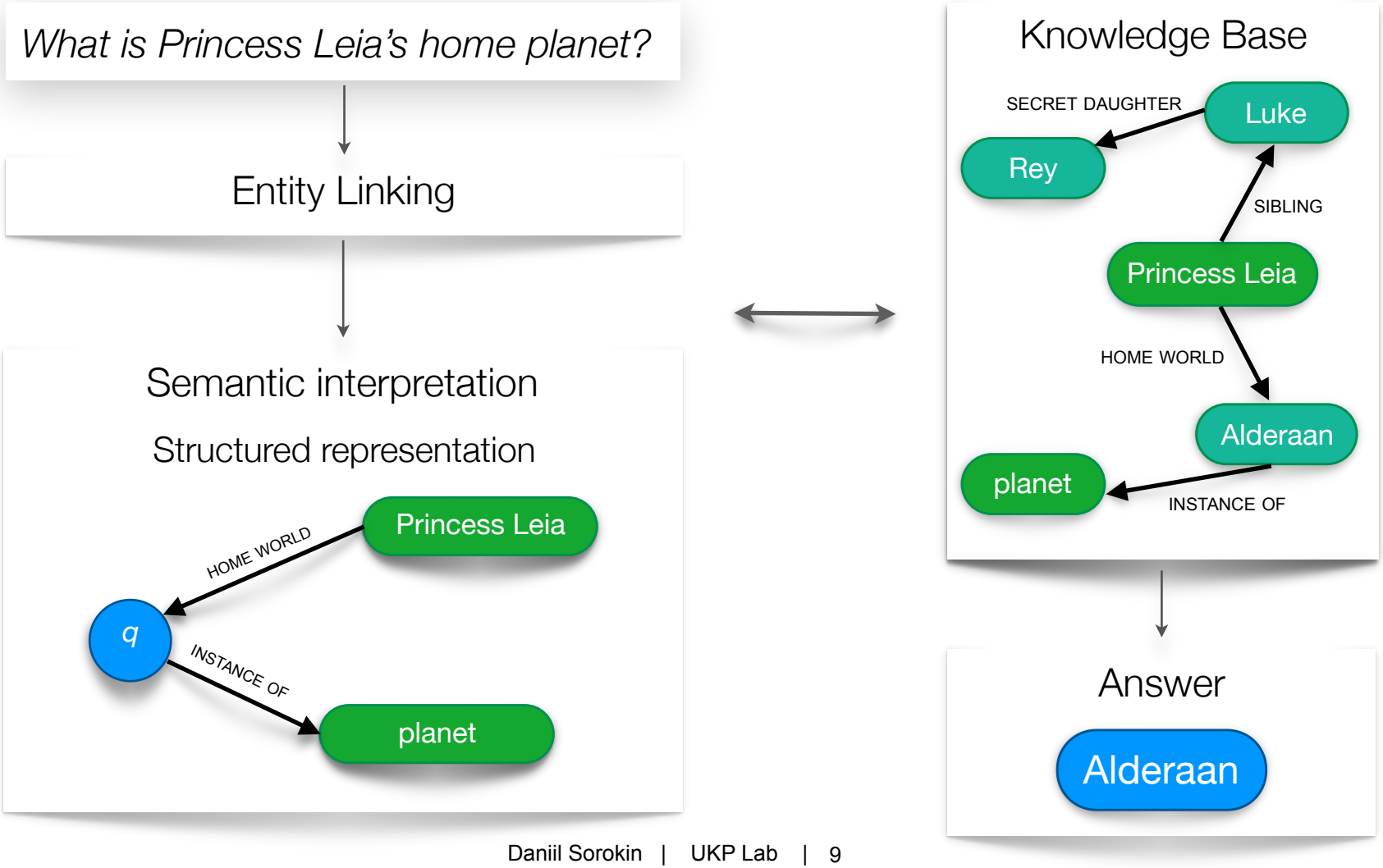
Structured representation



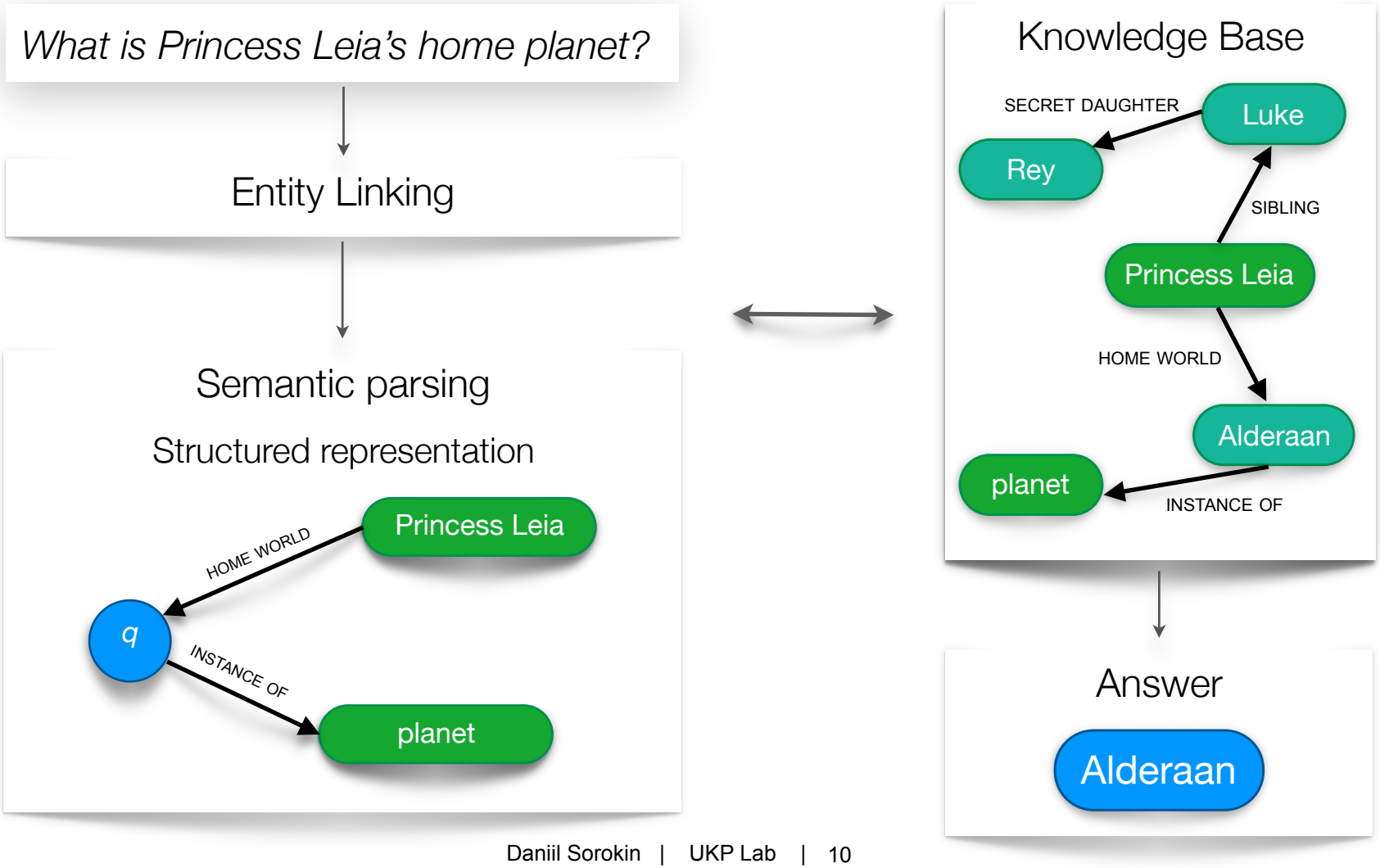
Answer

Alderaan

Knowledge Base Question Answering



Knowledge Base Question Answering



Knowledge Base Question Answering

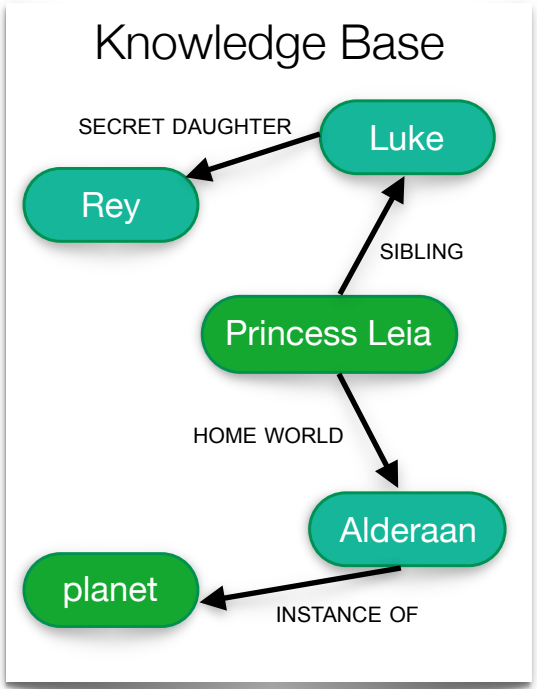
What is Princess Leia's home planet?

Entity Linking

Semantic parsing

SPARQL query

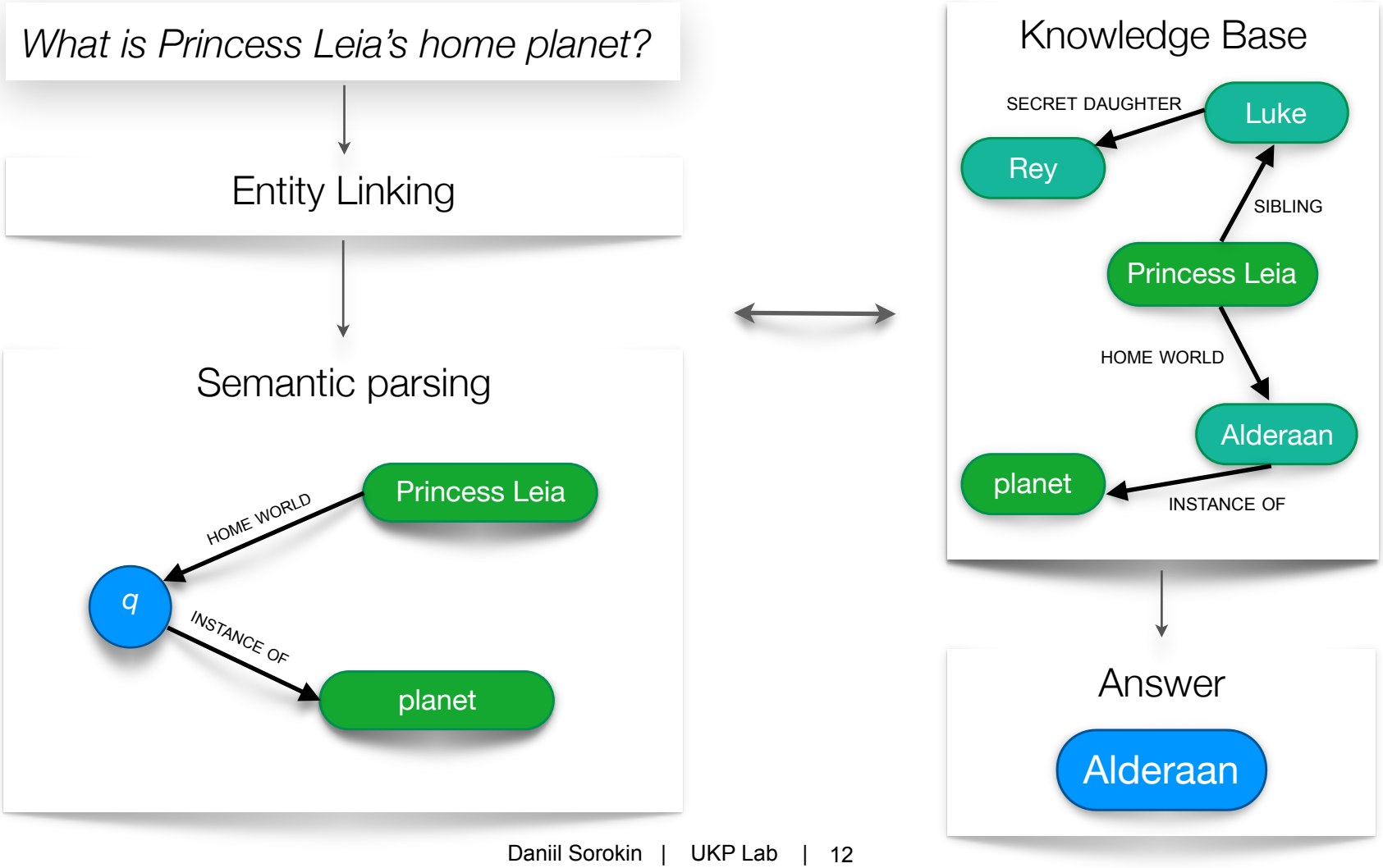
```
SELECT DISTINCT ?q WHERE {
    GRAPH <statements> {
        Princess Leia p:homeworld ?q.
        ?q rdf:type planet
    }
}
```



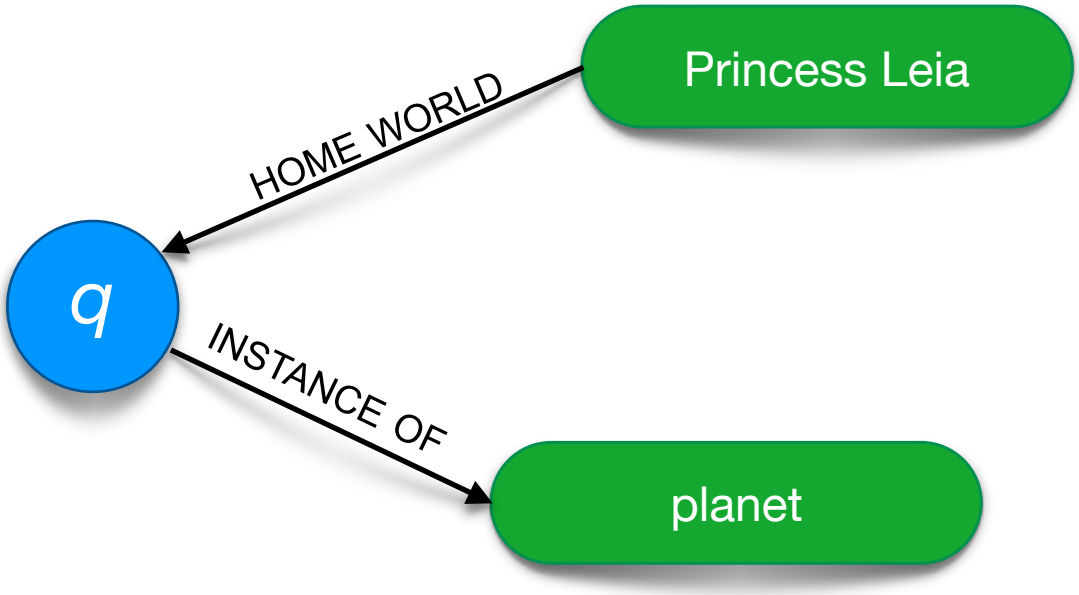
Answer

Alderaan

Knowledge Base Question Answering

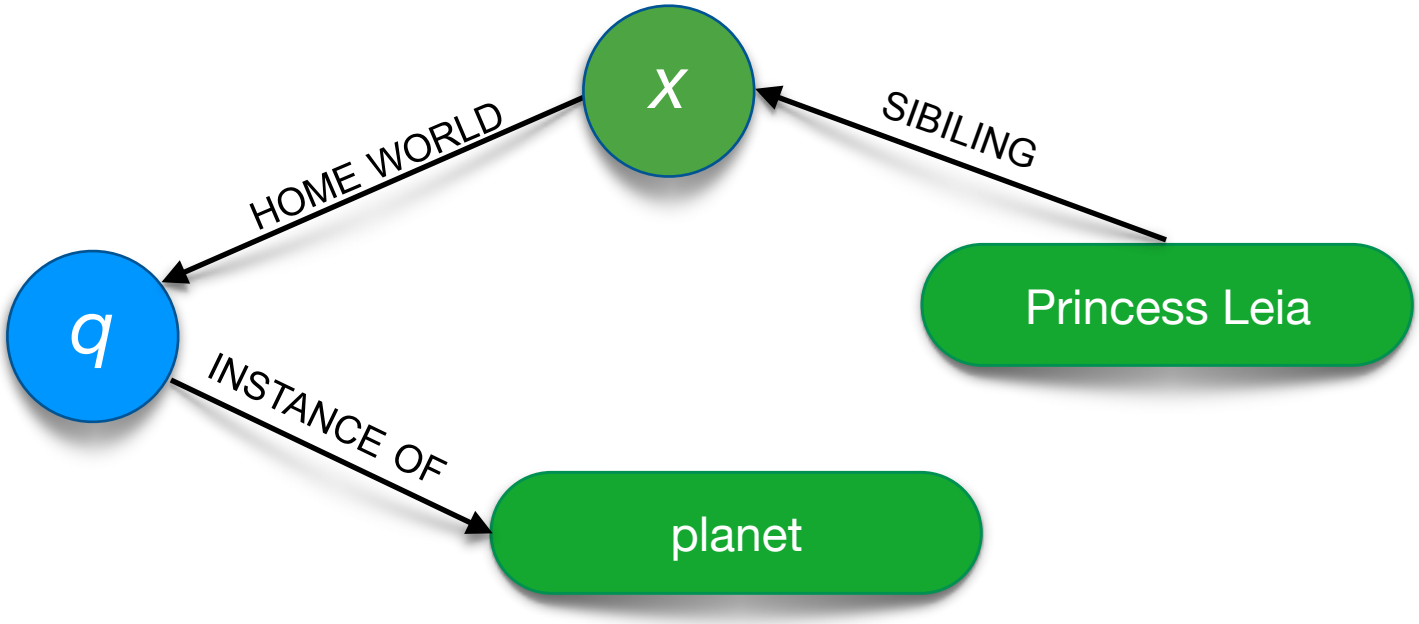


Structured Representation Approach



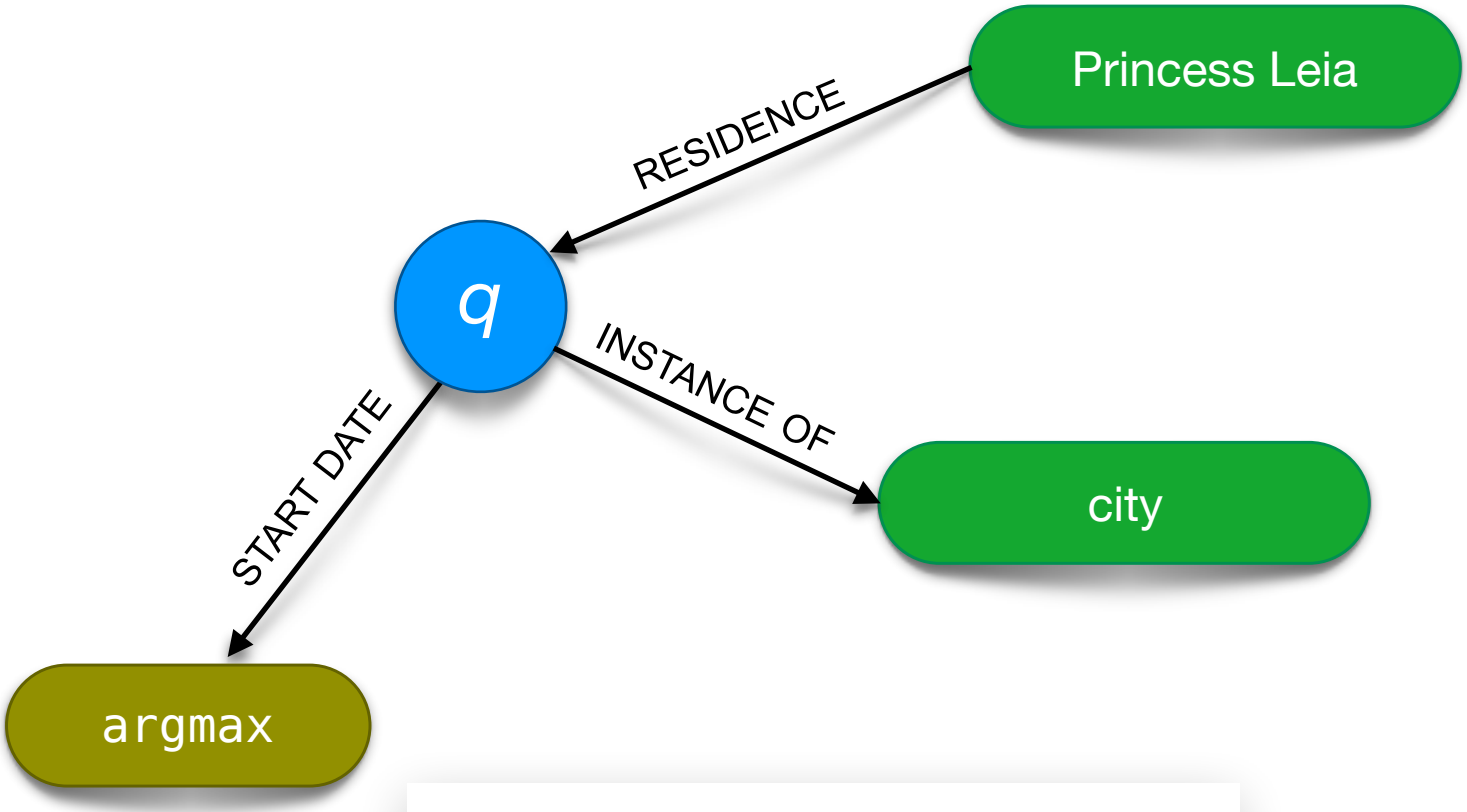
What is Princess Leia's home planet?

Structured Representation Approach



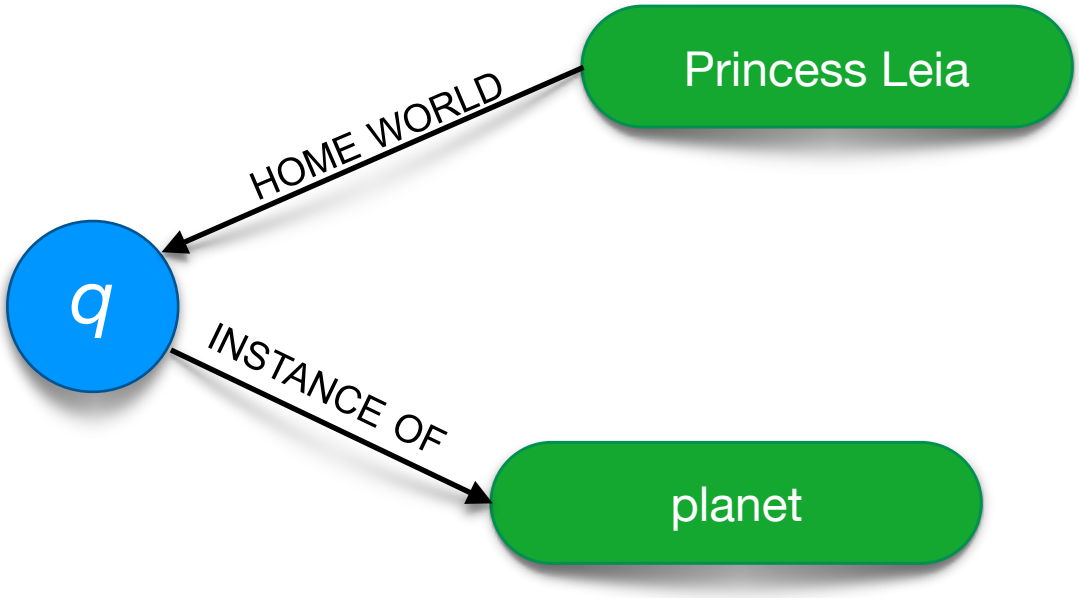
What is the home planet of Princess Leia's family?

Structured Representation Approach



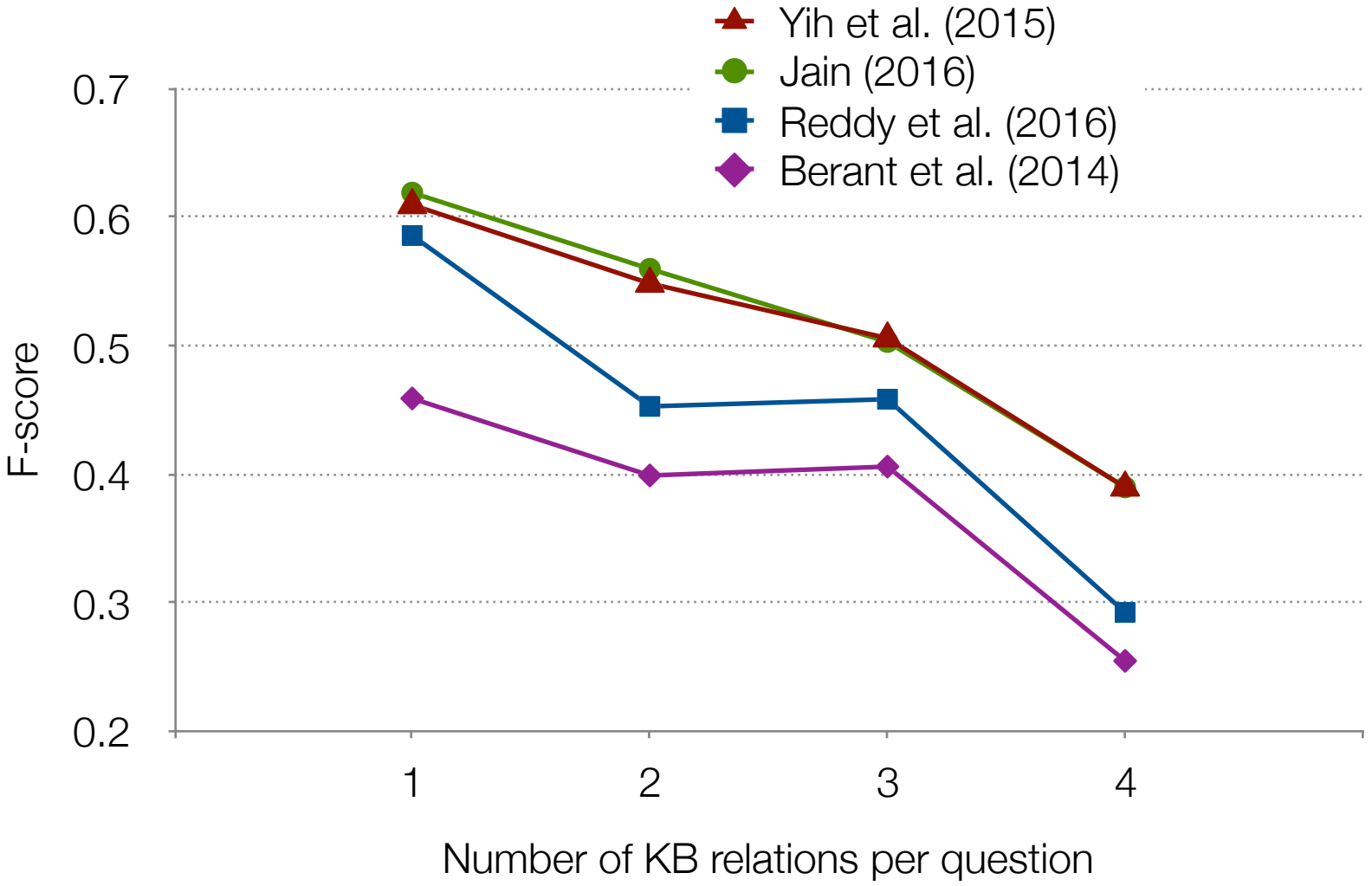
Where was Princess Leia's first home?

Structured Representation Approach



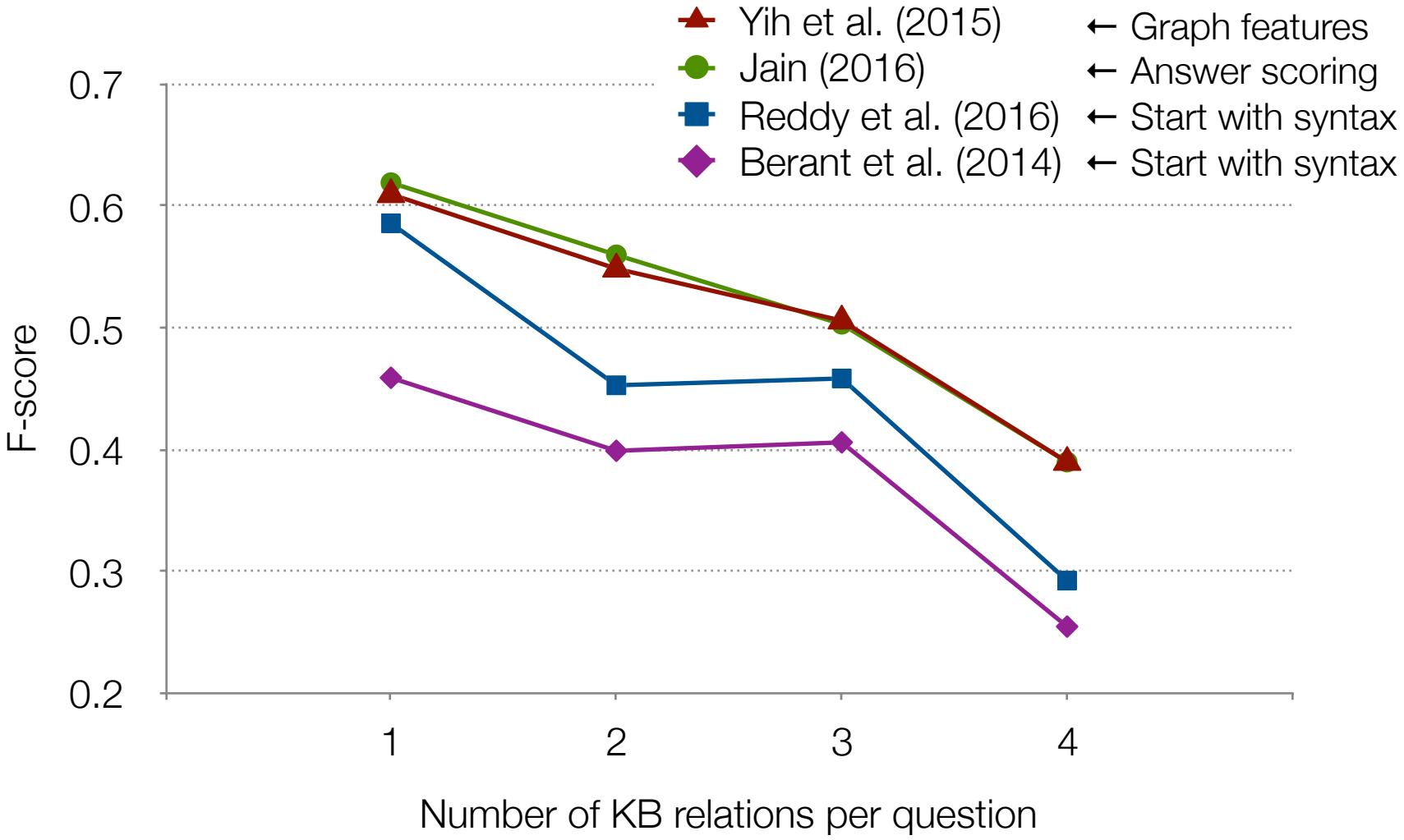
Representation structure is a graph and we model it as such

Previous results on WebQSP



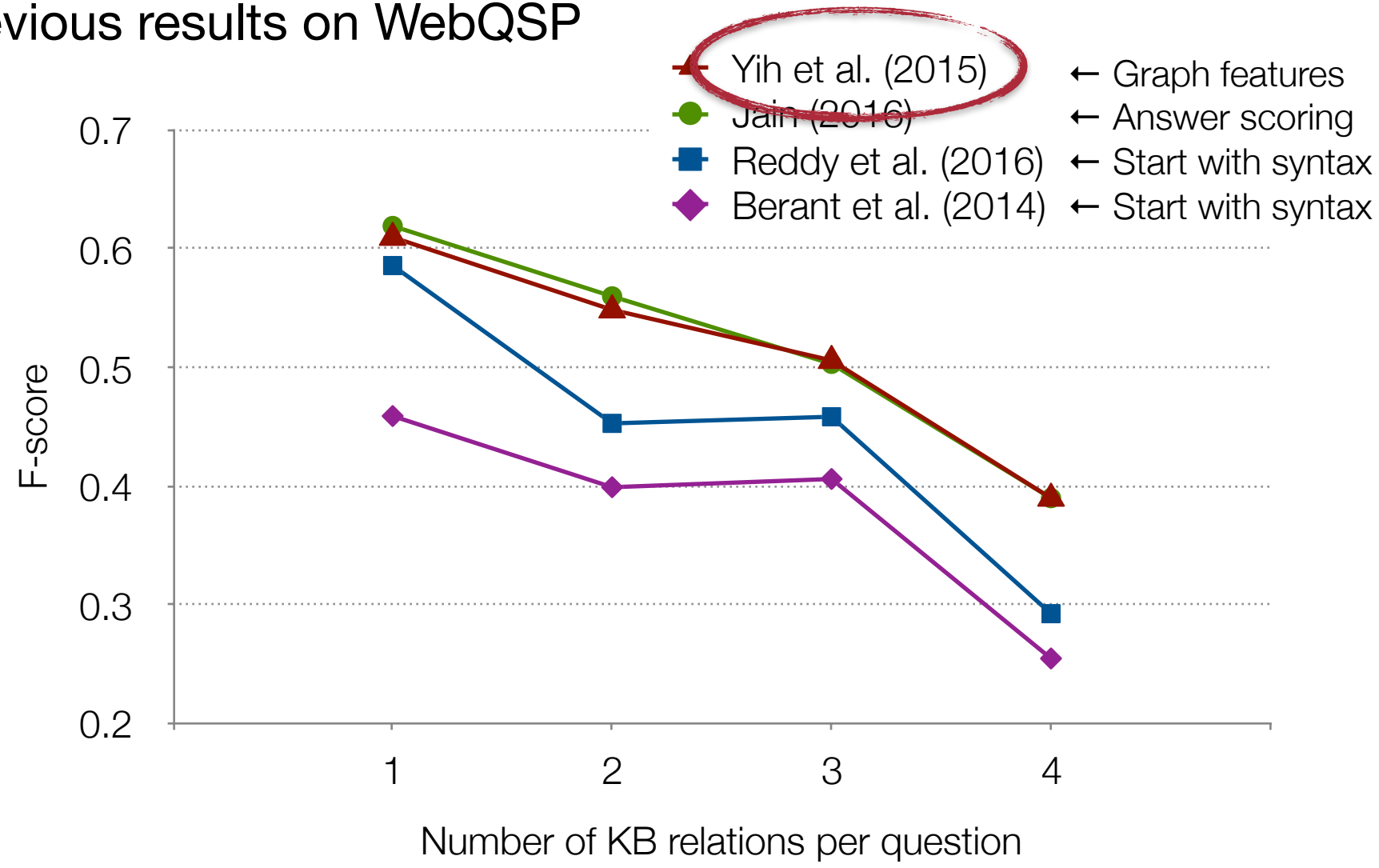
Freebase

Previous results on WebQSP



Freebase

Previous results on WebQSP



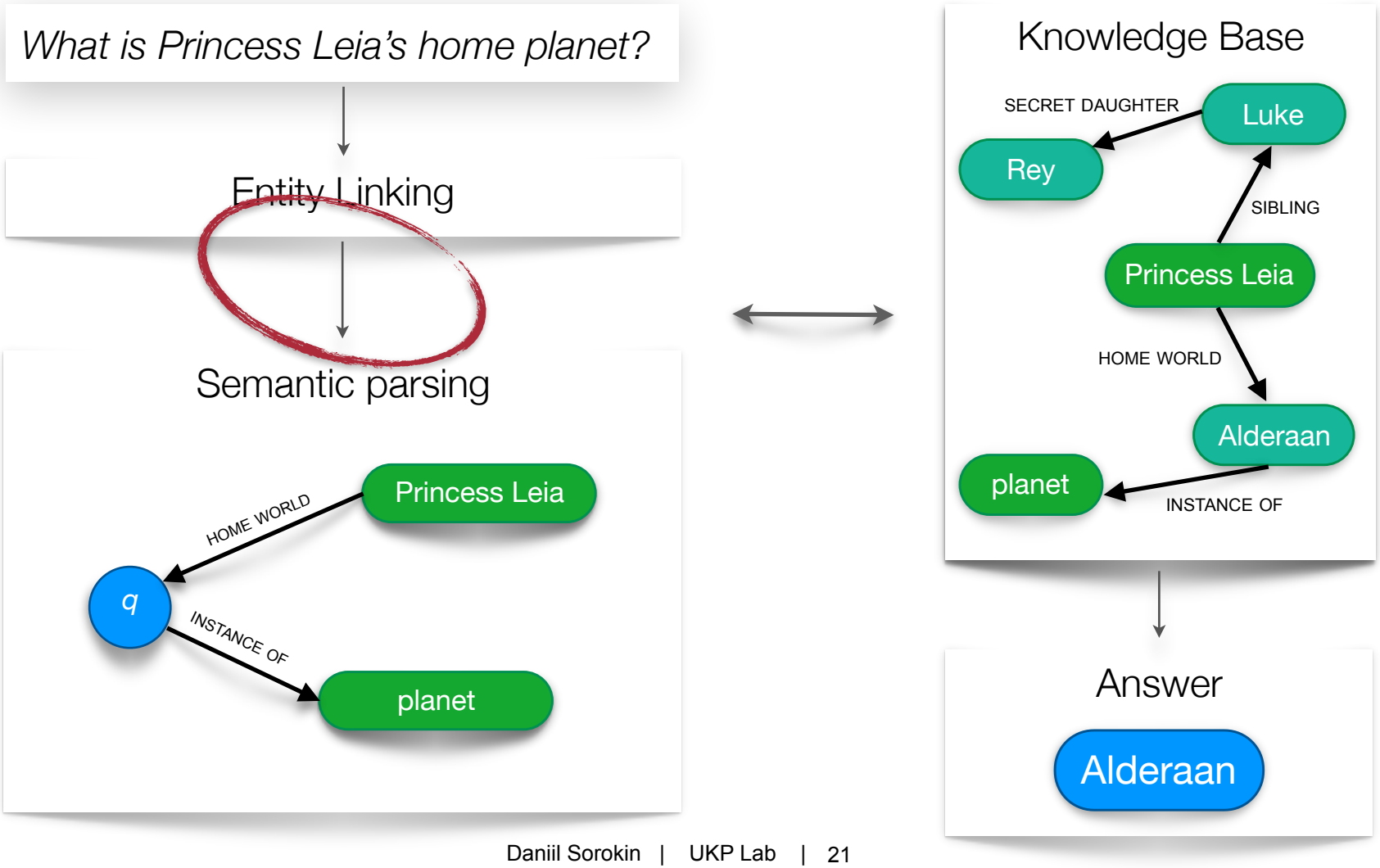
Freebase

Previous approaches do not integrate the structure of the representation into the model

Modeling the knowledge base graph and other graphs

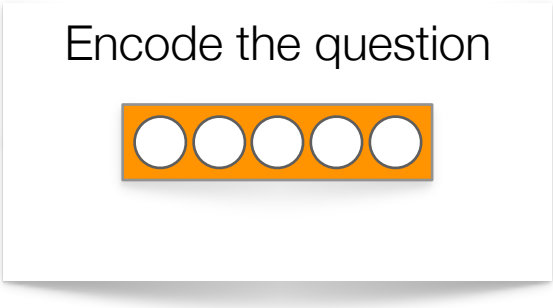
1. Graph Convolutional Networks
   - Kipf and Welling (2017) – document classification
   - Marcheggiani and Titov (2017) – encode syntax dependencies for SRL
   - Schlichtkrull et al. (2018) – KB embeddings
2. Gated Graph Neural Networks
   - Scarselli et al. (2009) – Graph Neural Networks
   - Li et al. (2016) – Gated variant
   - Chen et al. (2018) – Structured Dialogue Policy
   - Sun et al. (2018) – QA with Fusion of Knowledge Bases and Text

Knowledge Base Question Answering

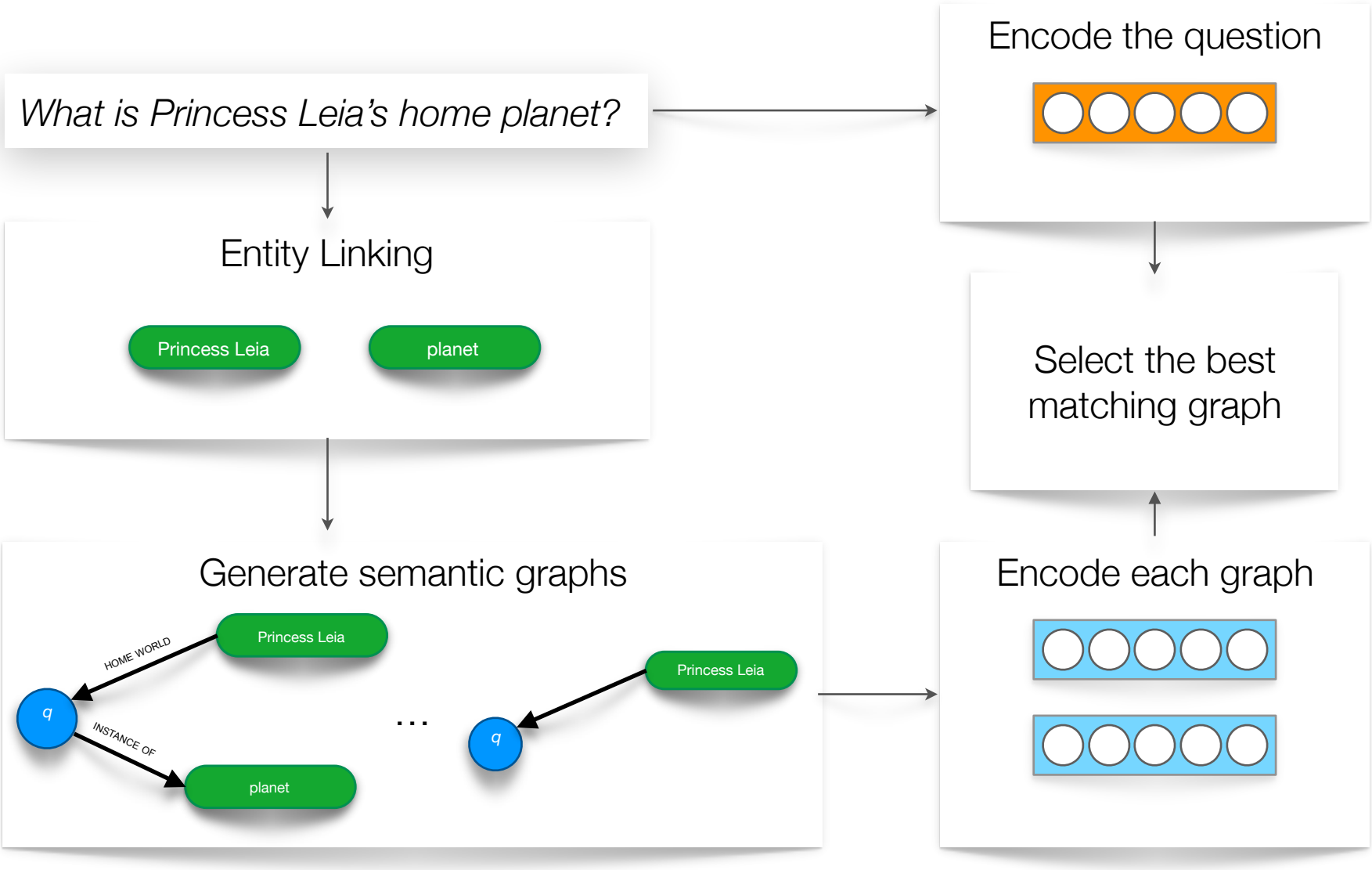


Representation Learning

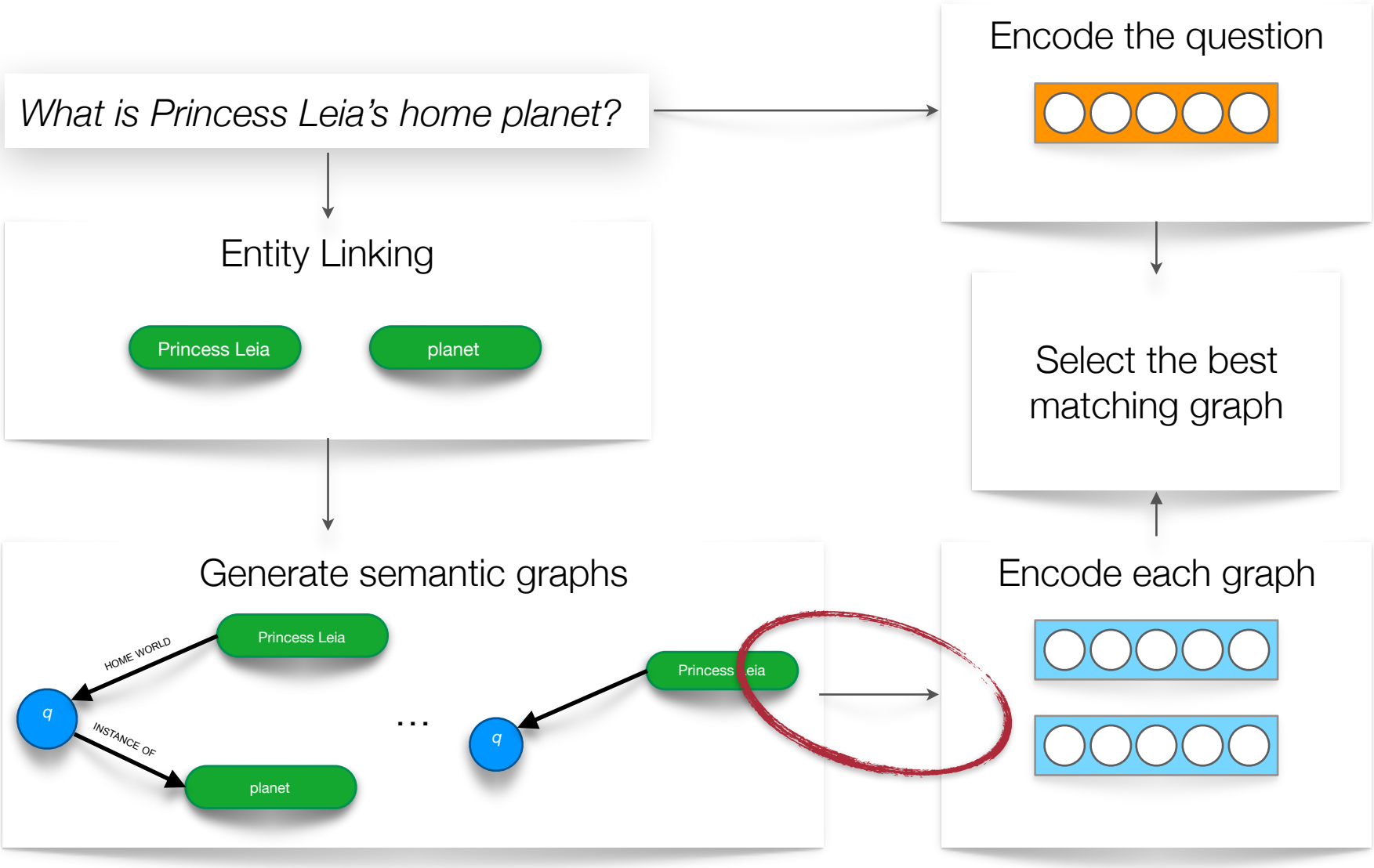
What is Princess Leia's home planet?



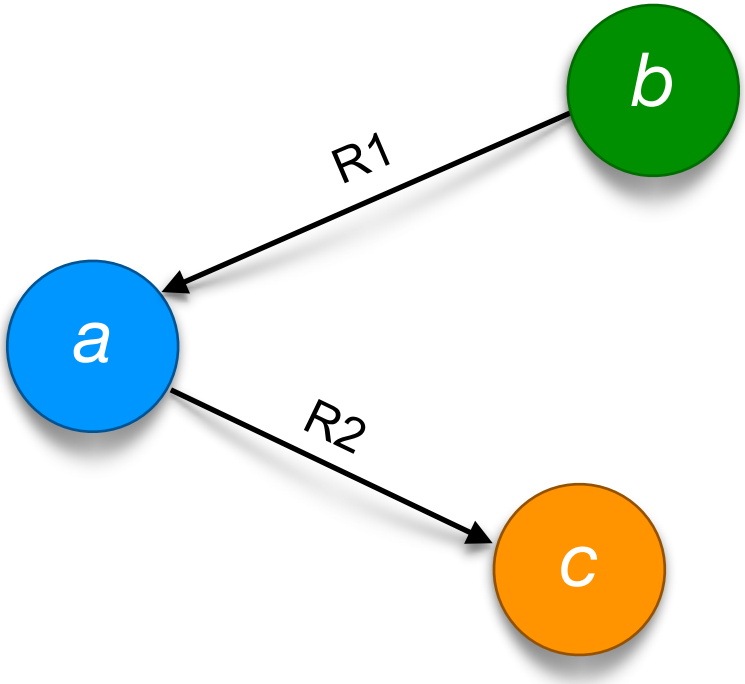
Representation Learning



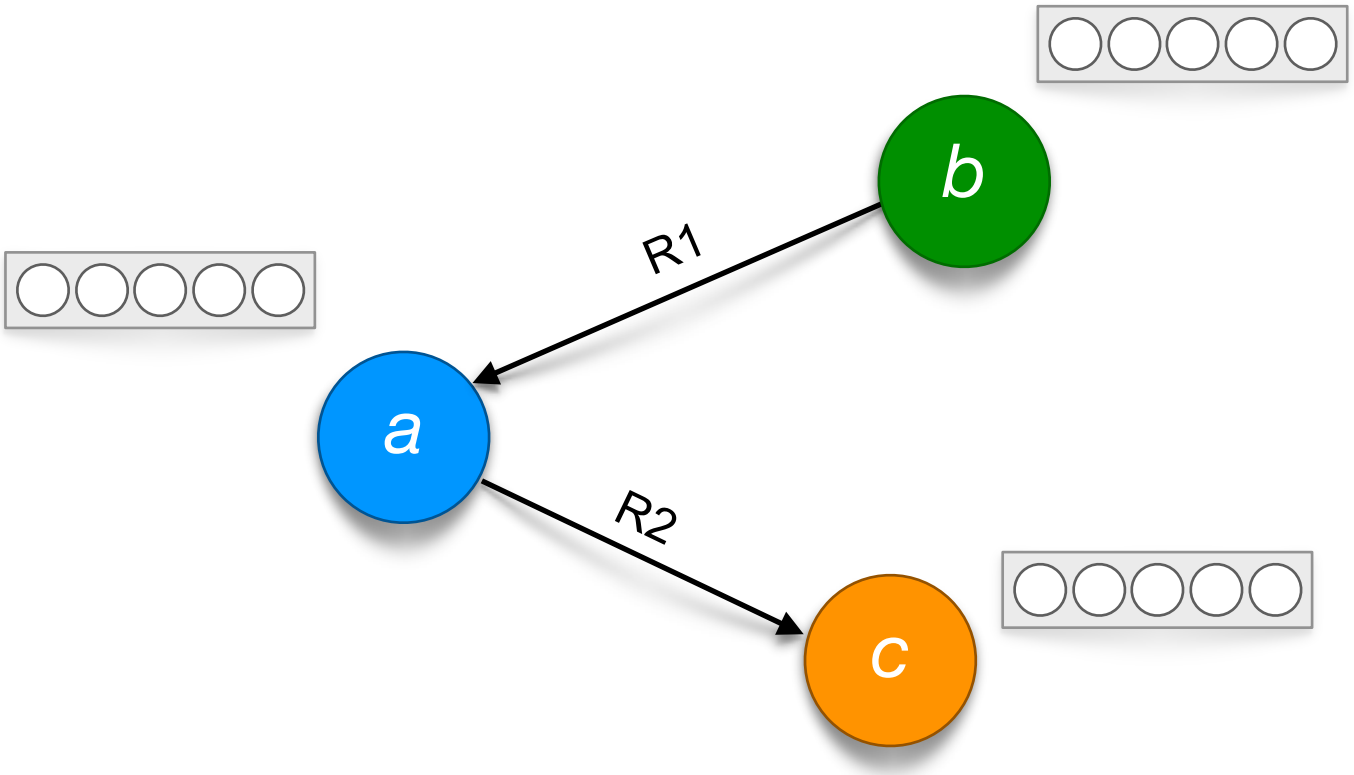
Representation Learning



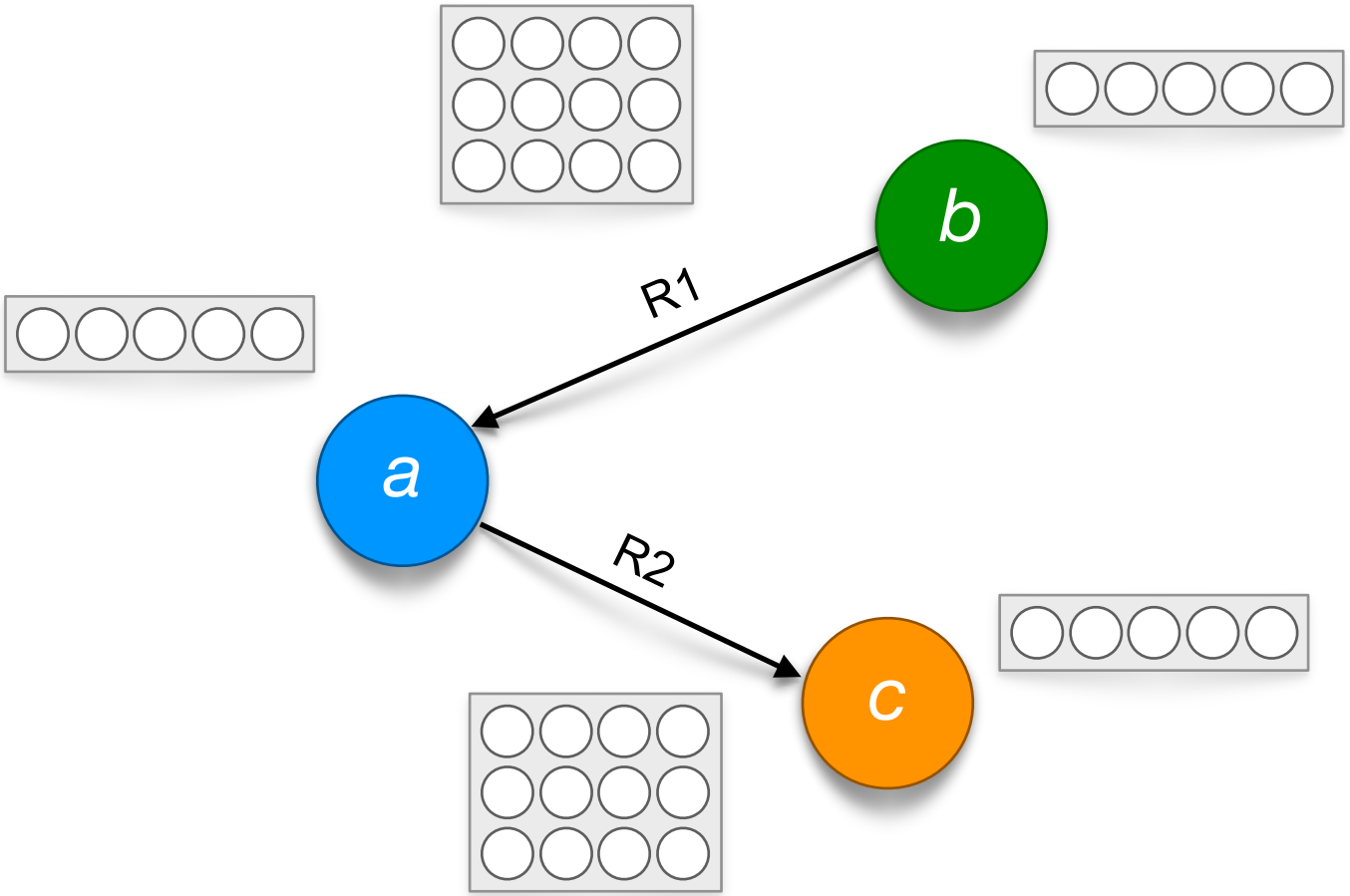
Graph Neural Networks – Scarselli et al. (2009)



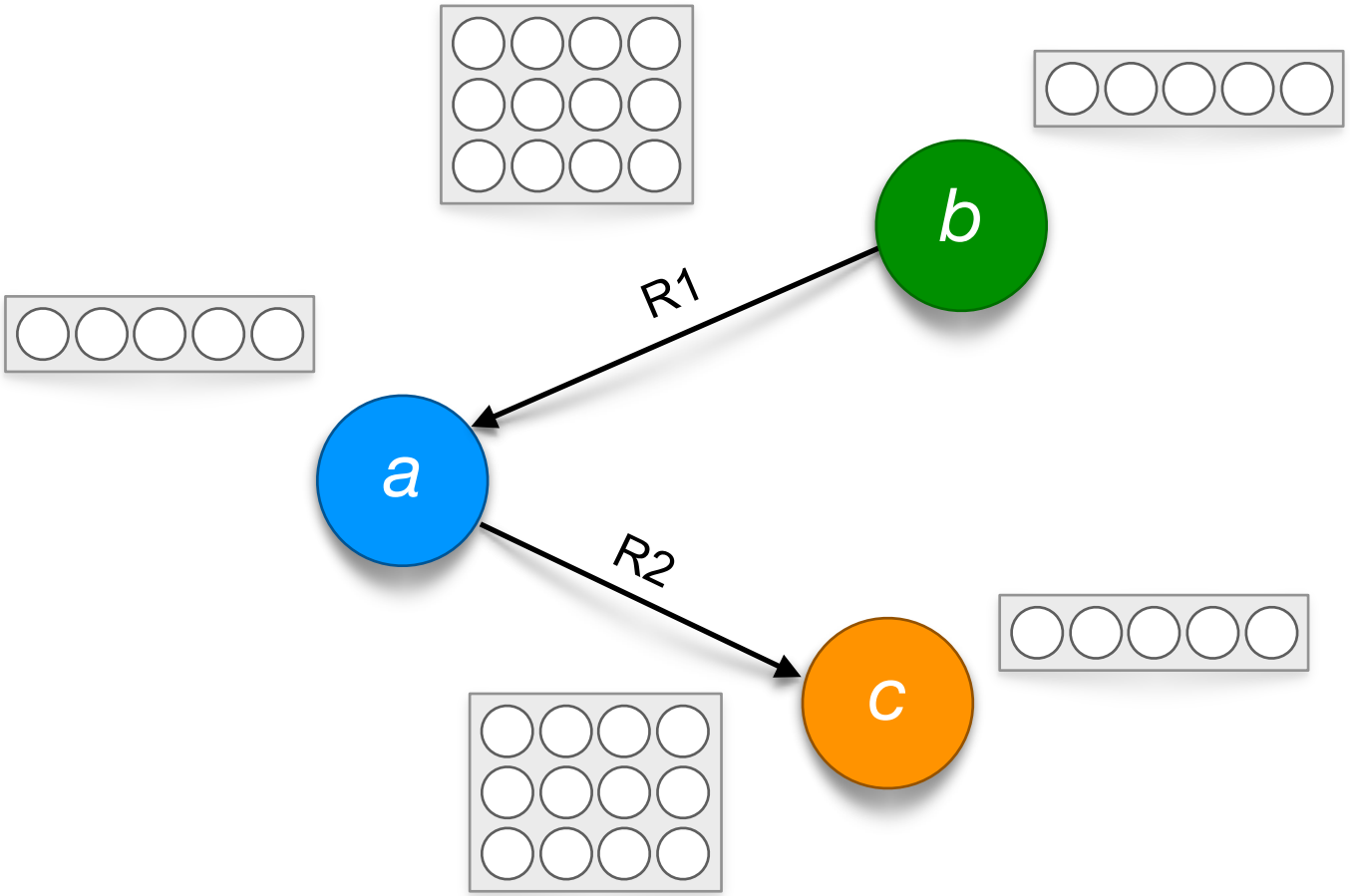
Graph Neural Networks



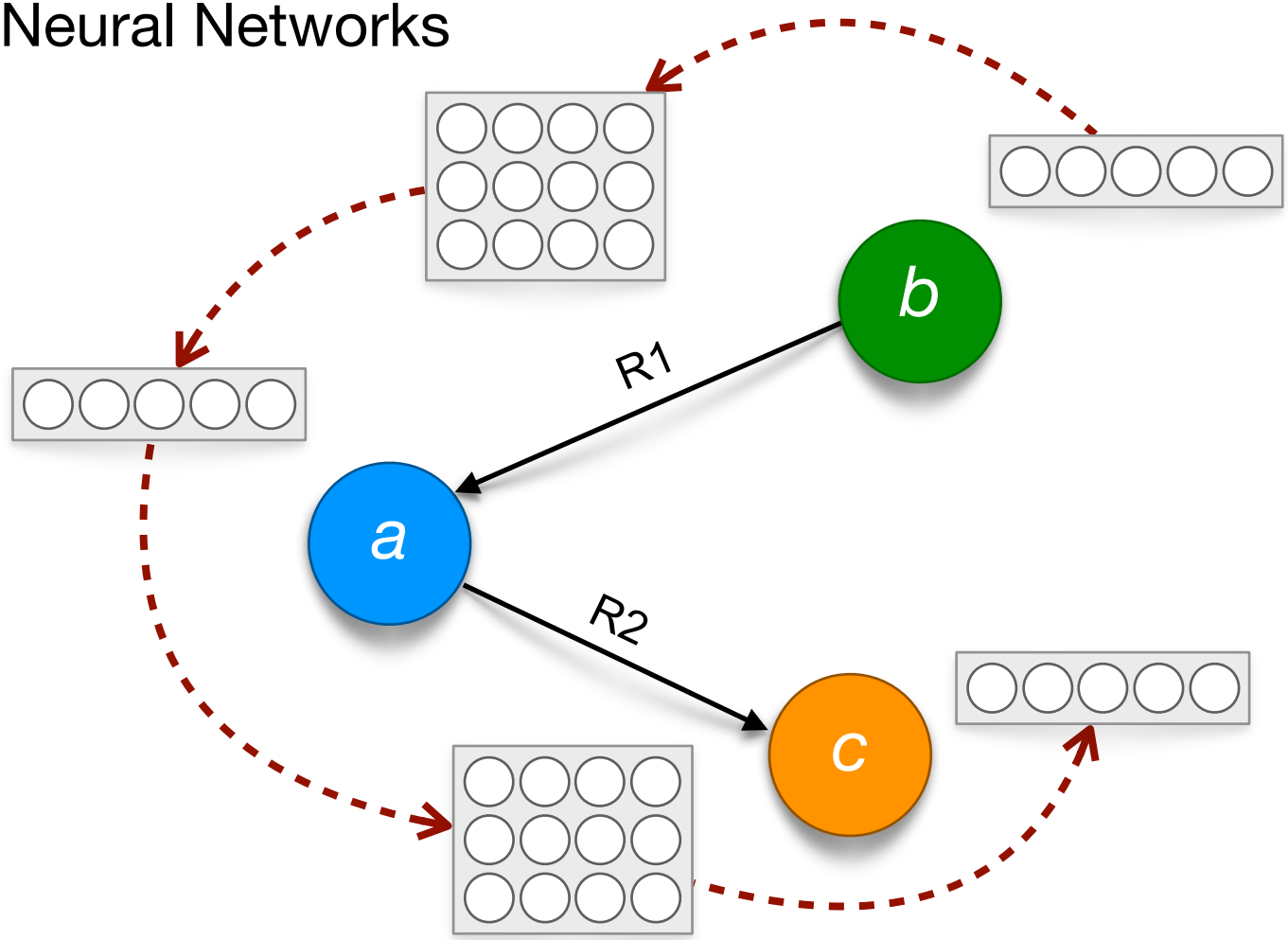
Graph Neural Networks



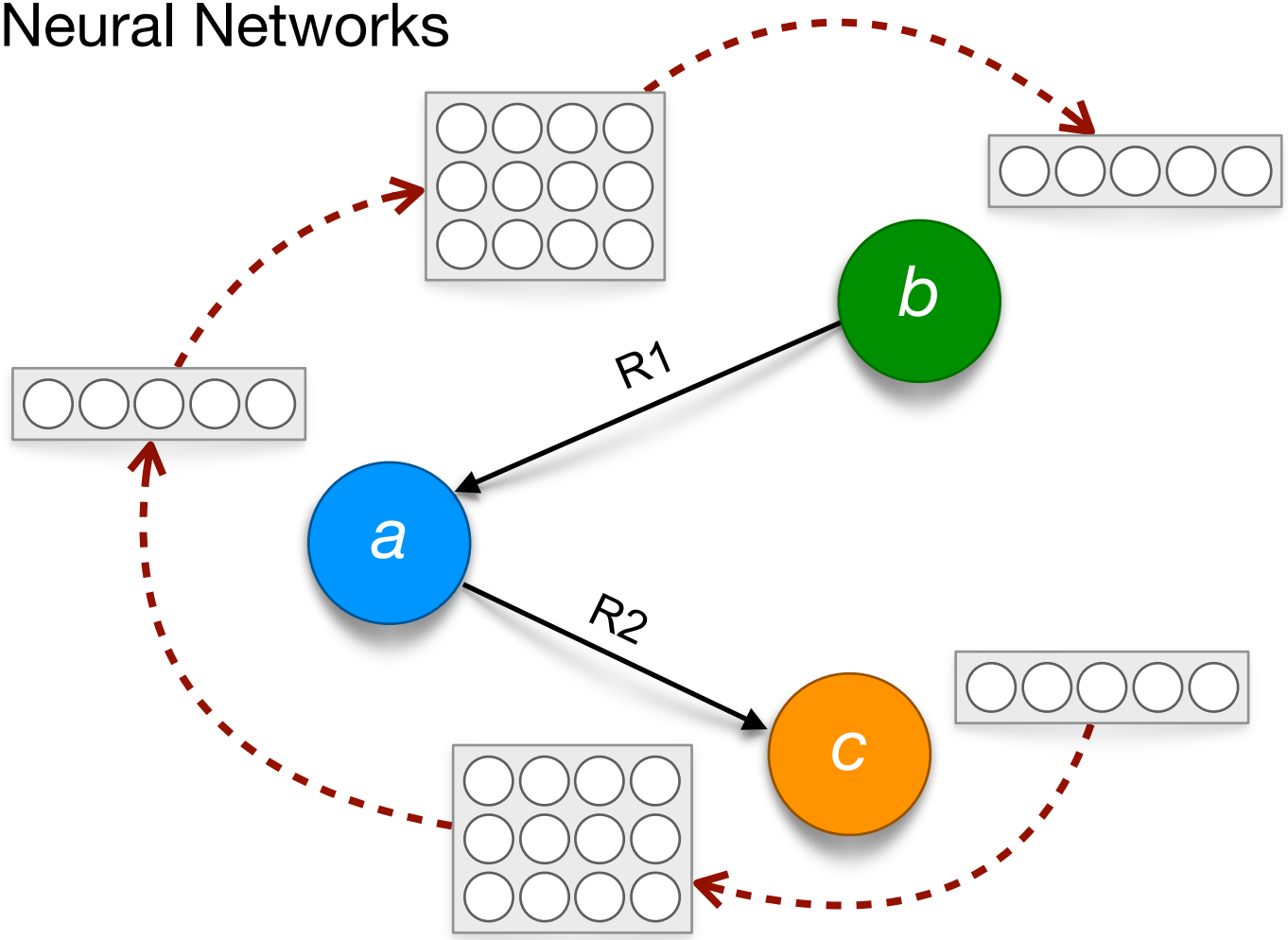
Graph Neural Networks



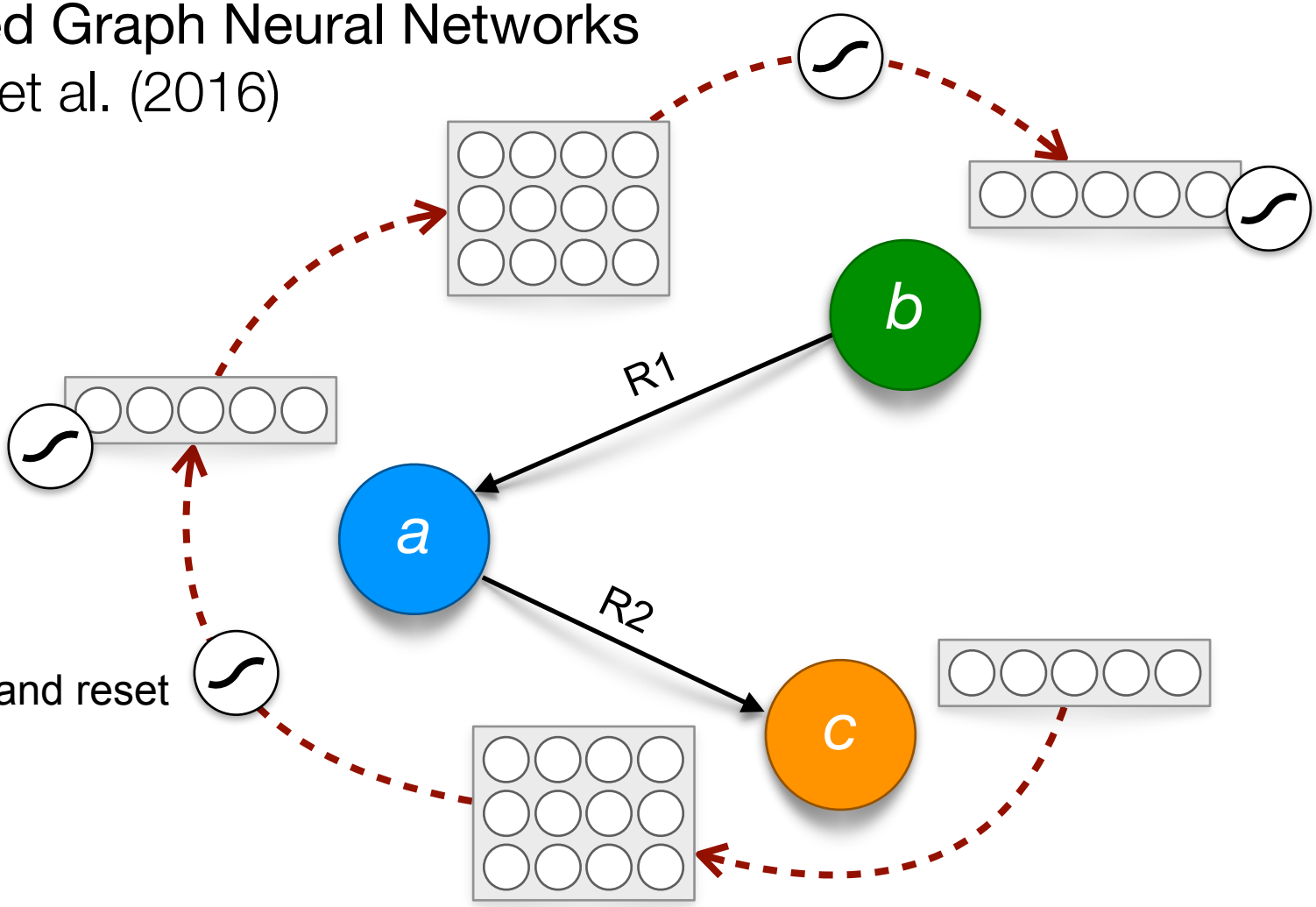
Graph Neural Networks



Graph Neural Networks

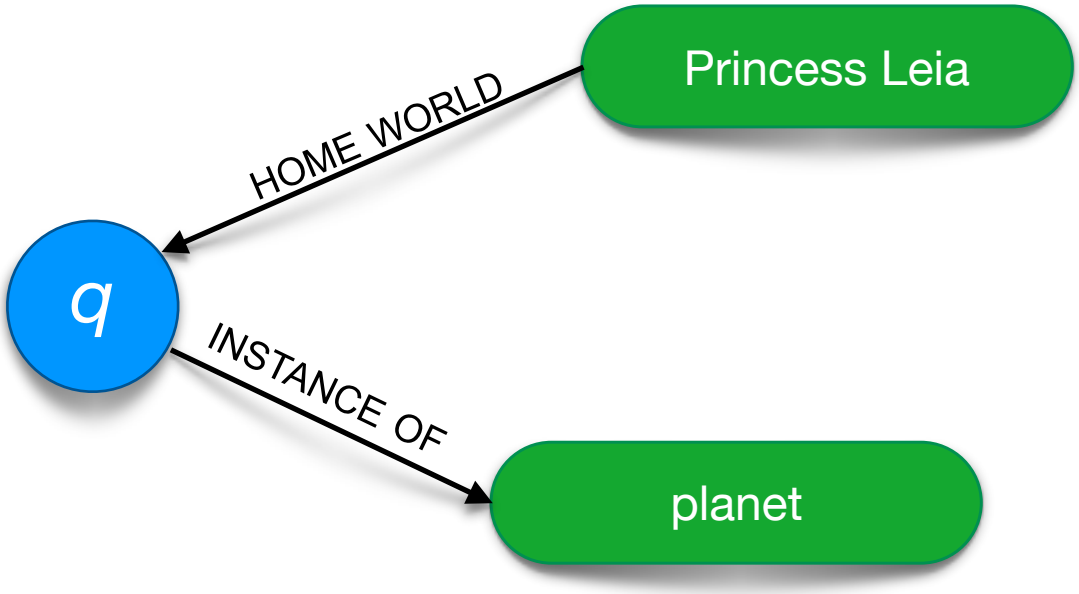


Gated Graph Neural Networks
– Li et al. (2016)

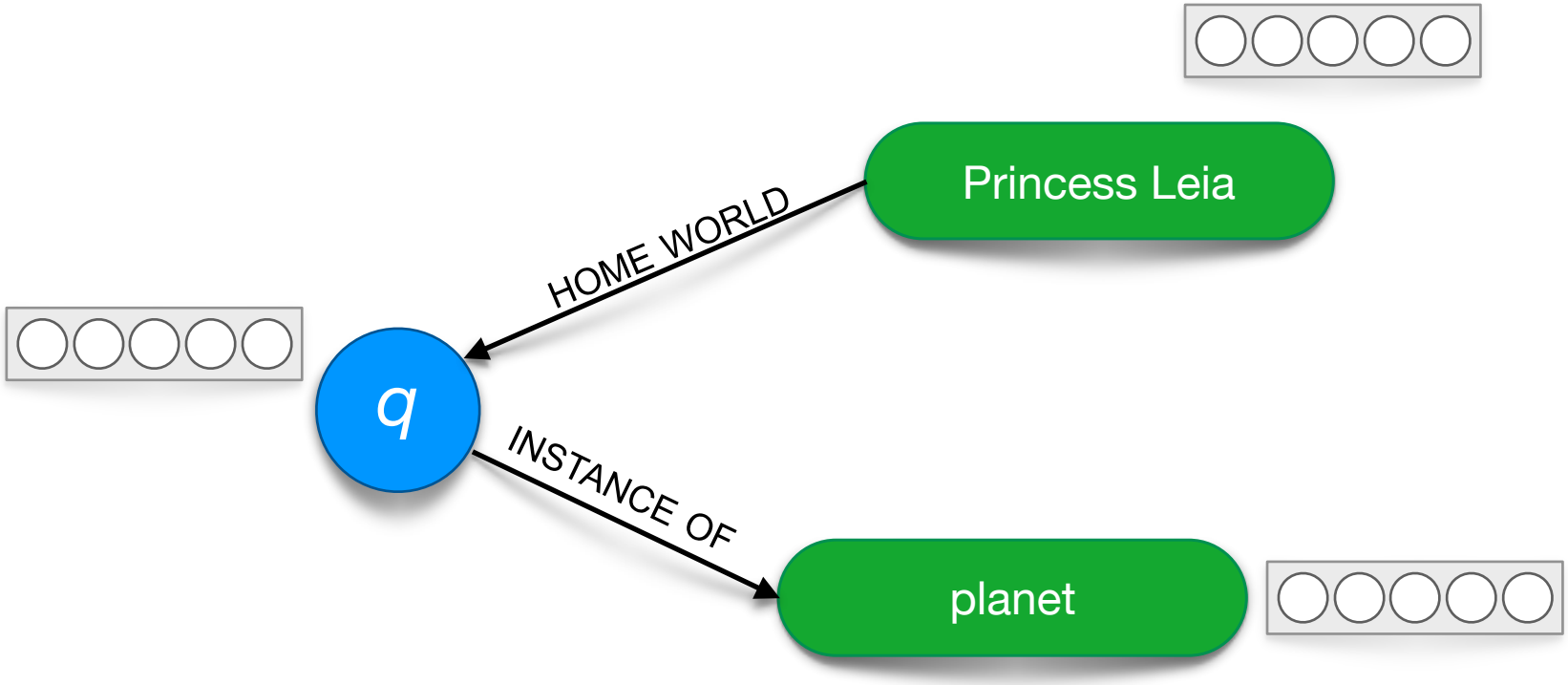


Update and reset gates

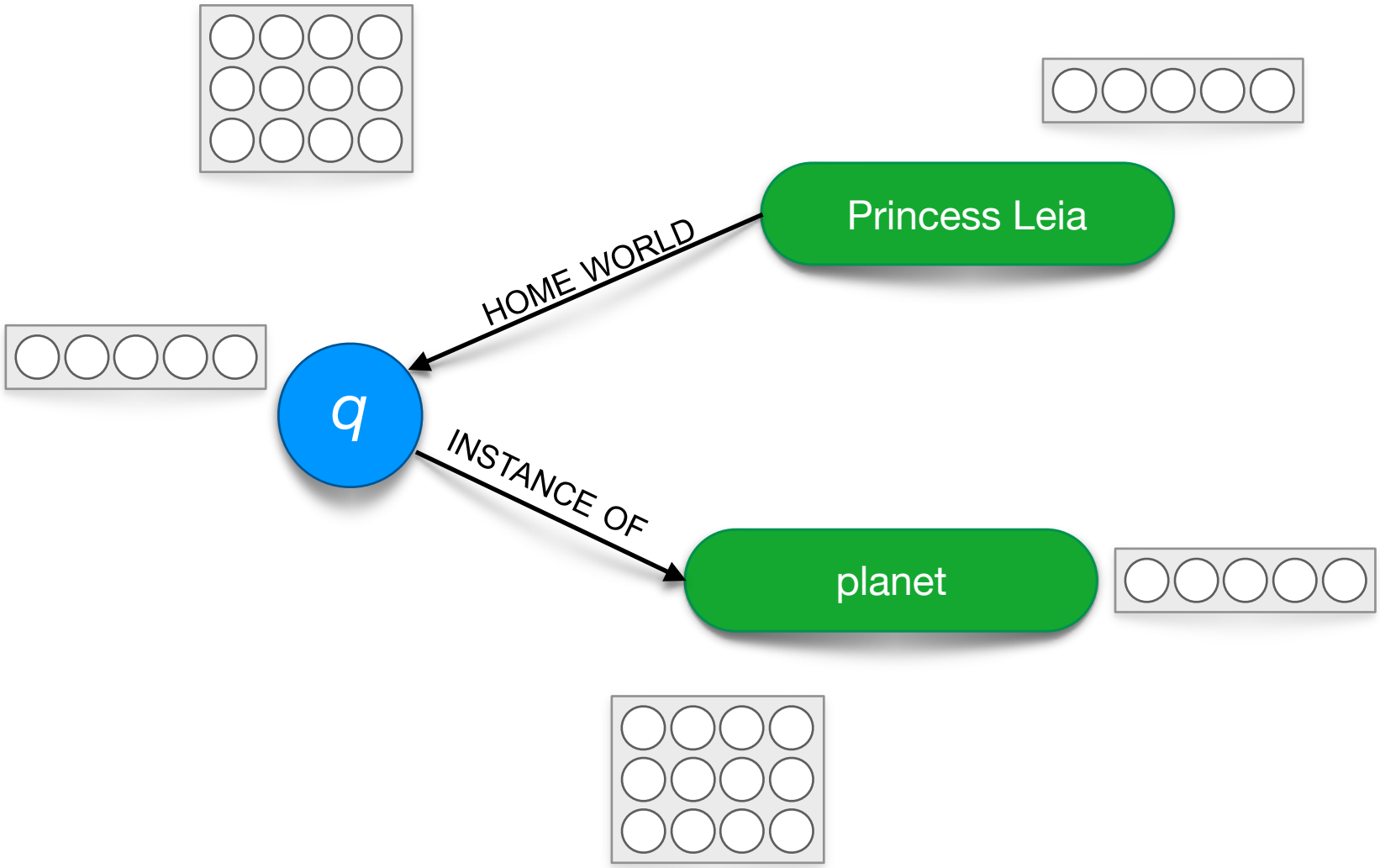
Gated Graph Neural Networks with Labels



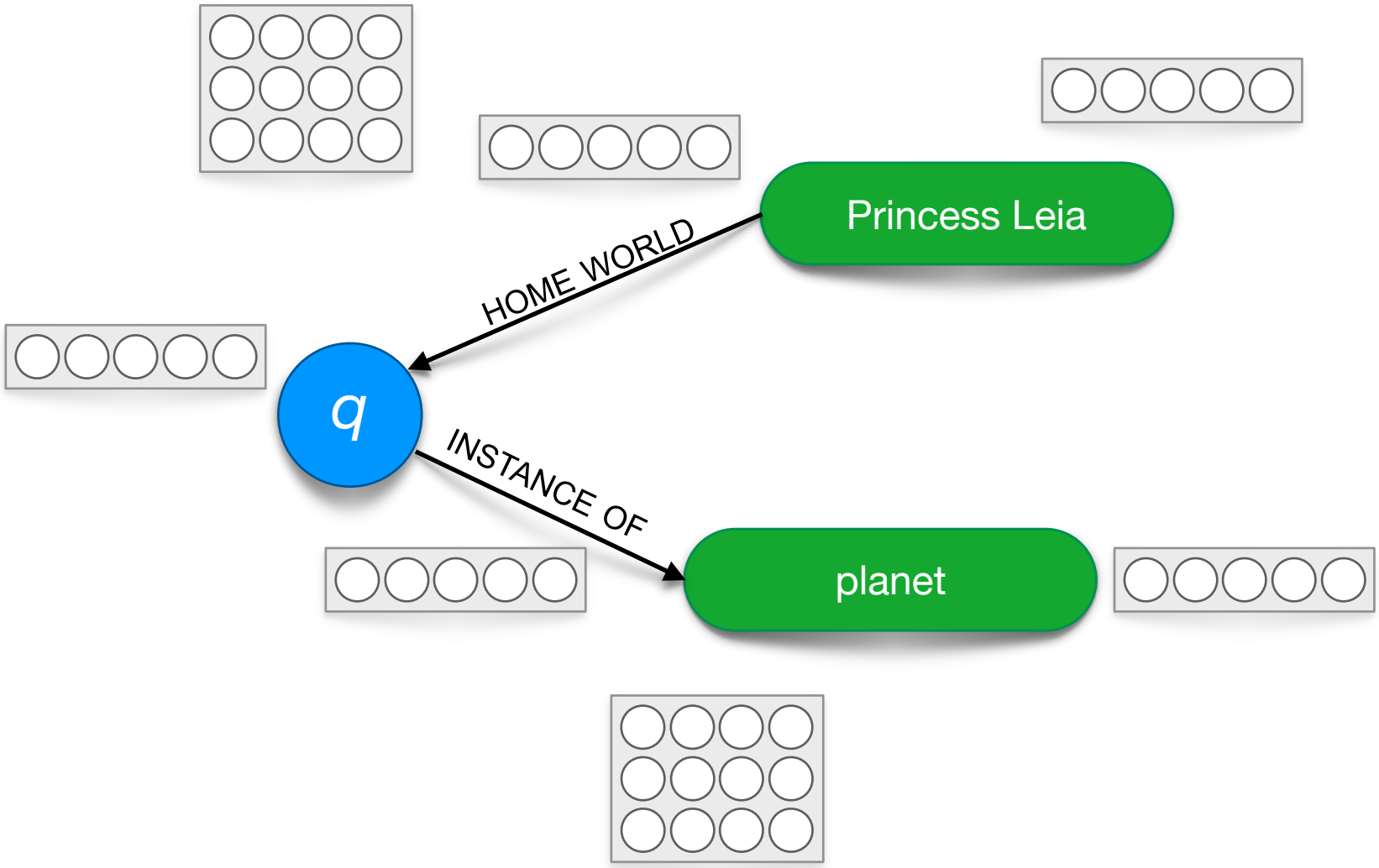
Gated Graph Neural Networks with Labels



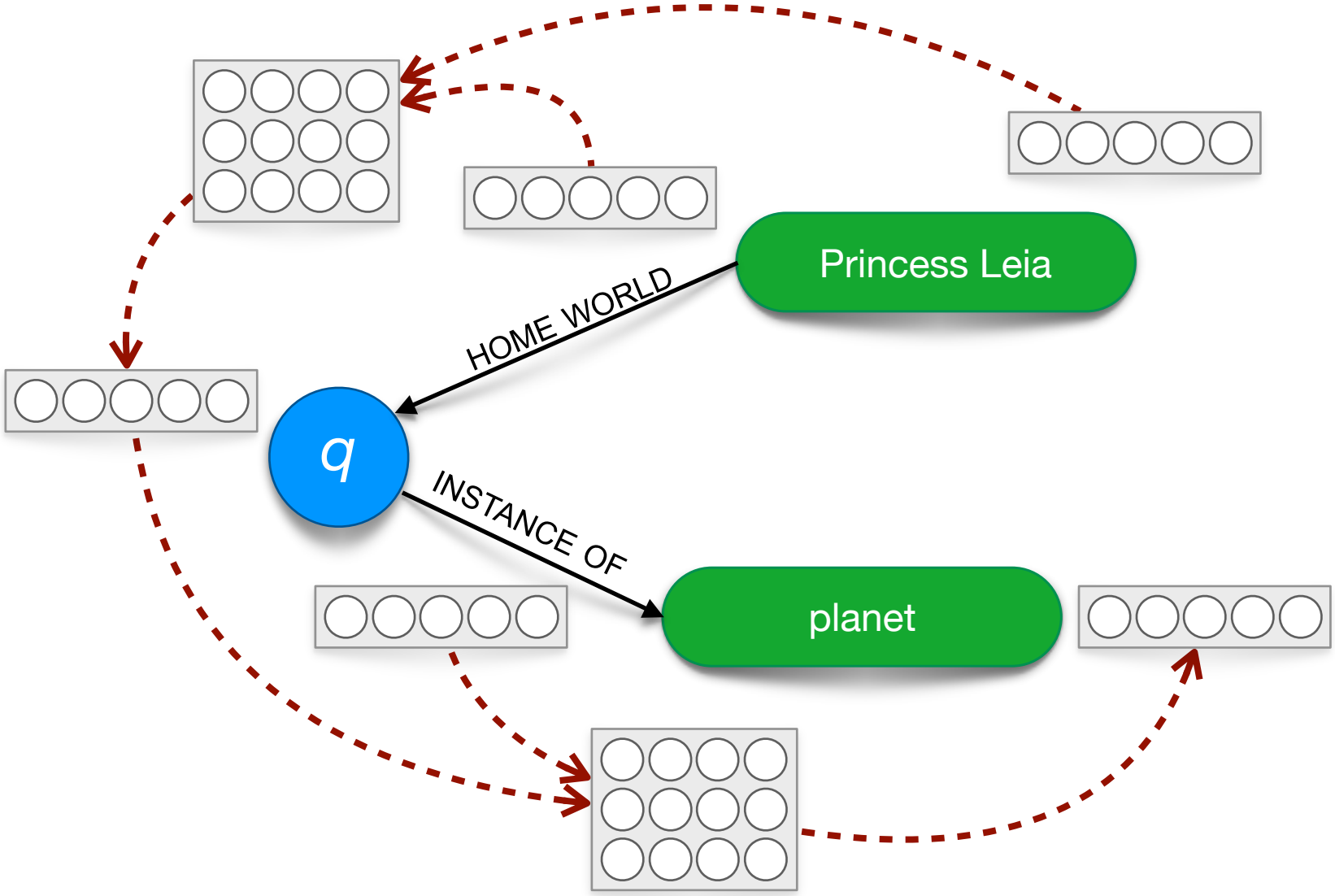
Gated Graph Neural Networks with Labels



Gated Graph Neural Networks with Labels



Gated Graph Neural Networks with Labels



Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\mathsf{T}} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\mathsf{T}}\right]$$

Propagation Model

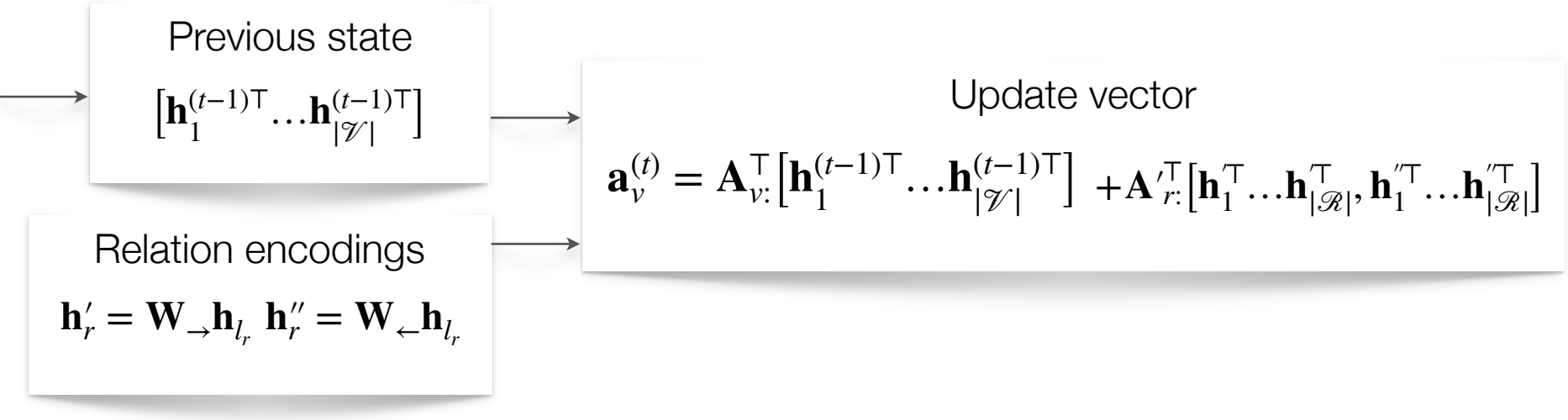
Previous state

$$\left[\mathbf{h}_1^{(t-1)\mathsf{T}} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\mathsf{T}}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Propagation Model



Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Update vector

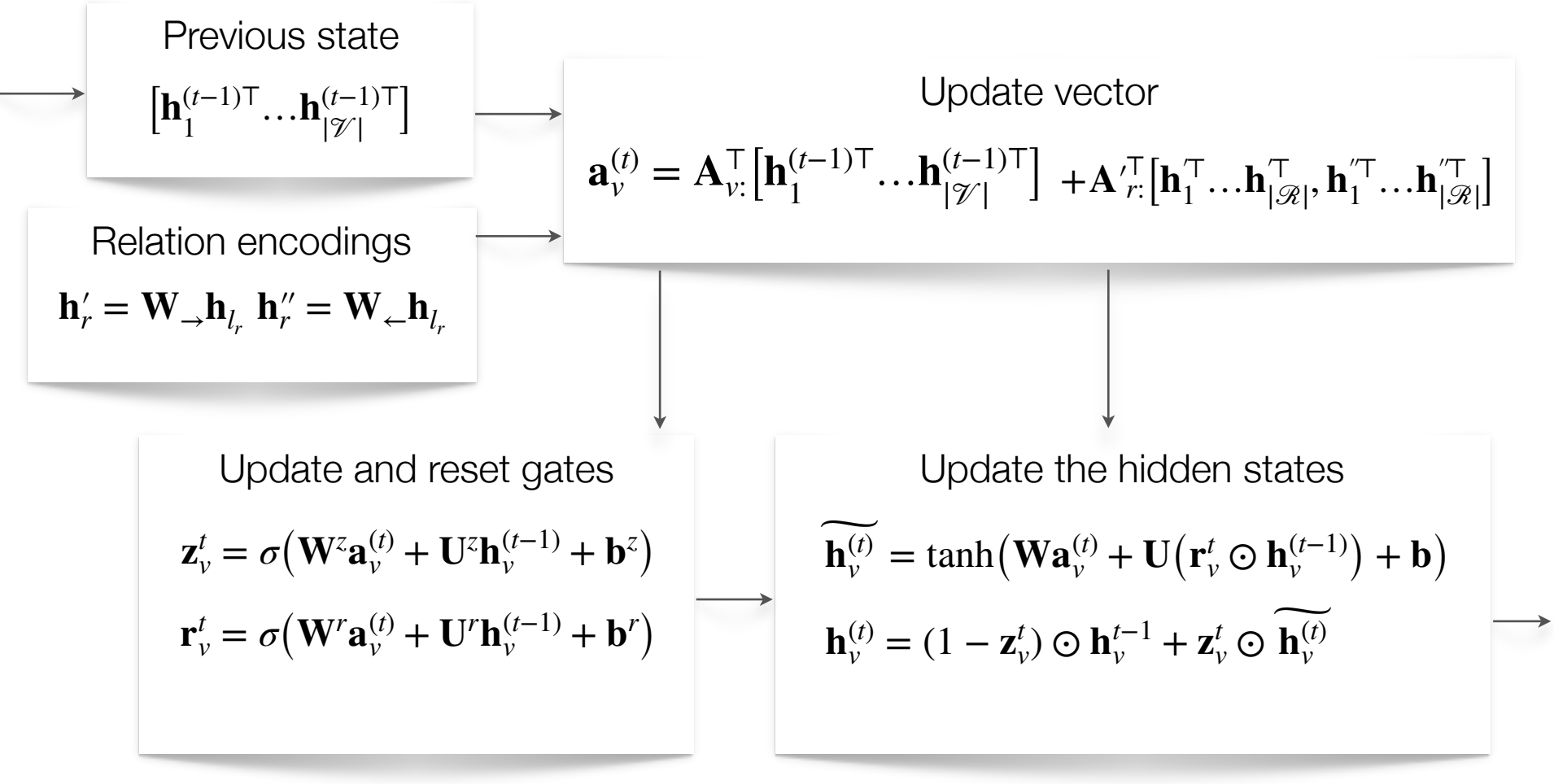
$$\mathbf{a}_v^{(t)} = \mathbf{A}_{v:}^{\top}\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right] + \mathbf{A}'^{\top}_{r:}\left[\mathbf{h}_1'^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}'^{\top}, \mathbf{h}_1''^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}''^{\top}\right]$$

Update and reset gates

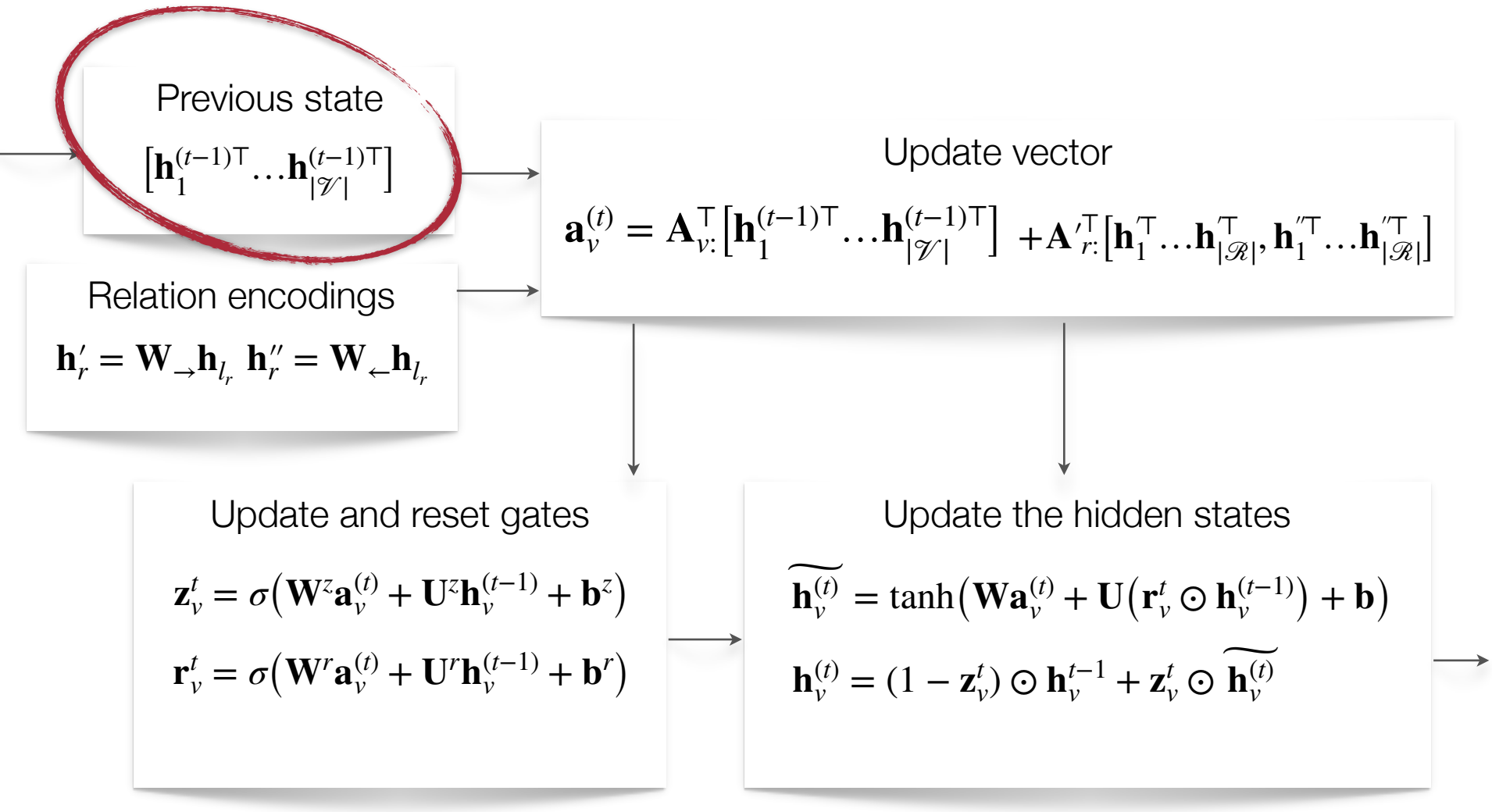
$$\mathbf{z}_v^t = \sigma\left(\mathbf{W}^z\mathbf{a}_v^{(t)} + \mathbf{U}^z\mathbf{h}_v^{(t-1)} + \mathbf{b}^z\right)$$

$$\mathbf{r}_v^t = \sigma\left(\mathbf{W}^r\mathbf{a}_v^{(t)} + \mathbf{U}^r\mathbf{h}_v^{(t-1)} + \mathbf{b}^r\right)$$

Propagation Model



Propagation Model



Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Update vector

$$\mathbf{a}_v^{(t)} = \mathbf{A}_{v:}^{\top}\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right] + \mathbf{A}'^{\top}_{r:}\left[\mathbf{h}_1'^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}'^{\top}, \mathbf{h}_1''^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}''^{\top}\right]$$

Update and reset gates

$$\mathbf{z}_v^t = \sigma\left(\mathbf{W}^z\mathbf{a}_v^{(t)} + \mathbf{U}^z\mathbf{h}_v^{(t-1)} + \mathbf{b}^z\right)$$

$$\mathbf{r}_v^t = \sigma\left(\mathbf{W}^r\mathbf{a}_v^{(t)} + \mathbf{U}^r\mathbf{h}_v^{(t-1)} + \mathbf{b}^r\right)$$

Update the hidden states

$$\widetilde{\mathbf{h}_v^{(t)}} = \tanh\left(\mathbf{W}\mathbf{a}_v^{(t)} + \mathbf{U}\left(\mathbf{r}_v^t \odot \mathbf{h}_v^{(t-1)}\right) + \mathbf{b}\right)$$

$$\mathbf{h}_v^{(t)} = (1 - \mathbf{z}_v^t) \odot \mathbf{h}_v^{t-1} + \mathbf{z}_v^t \odot \widetilde{\mathbf{h}_v^{(t)}}$$

Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Update vector

$$\mathbf{a}_v^{(t)} = \mathbf{A}_{v:}^{\top}\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right] + \mathbf{A}'^{\top}_{r:}\left[\mathbf{h}_1'^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}'^{\top}, \mathbf{h}_1''^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}''^{\top}\right]$$

Update and reset gates

$$\mathbf{z}_v^t = \sigma\left(\mathbf{W}^z\mathbf{a}_v^{(t)} + \mathbf{U}^z\mathbf{h}_v^{(t-1)} + \mathbf{b}^z\right)$$

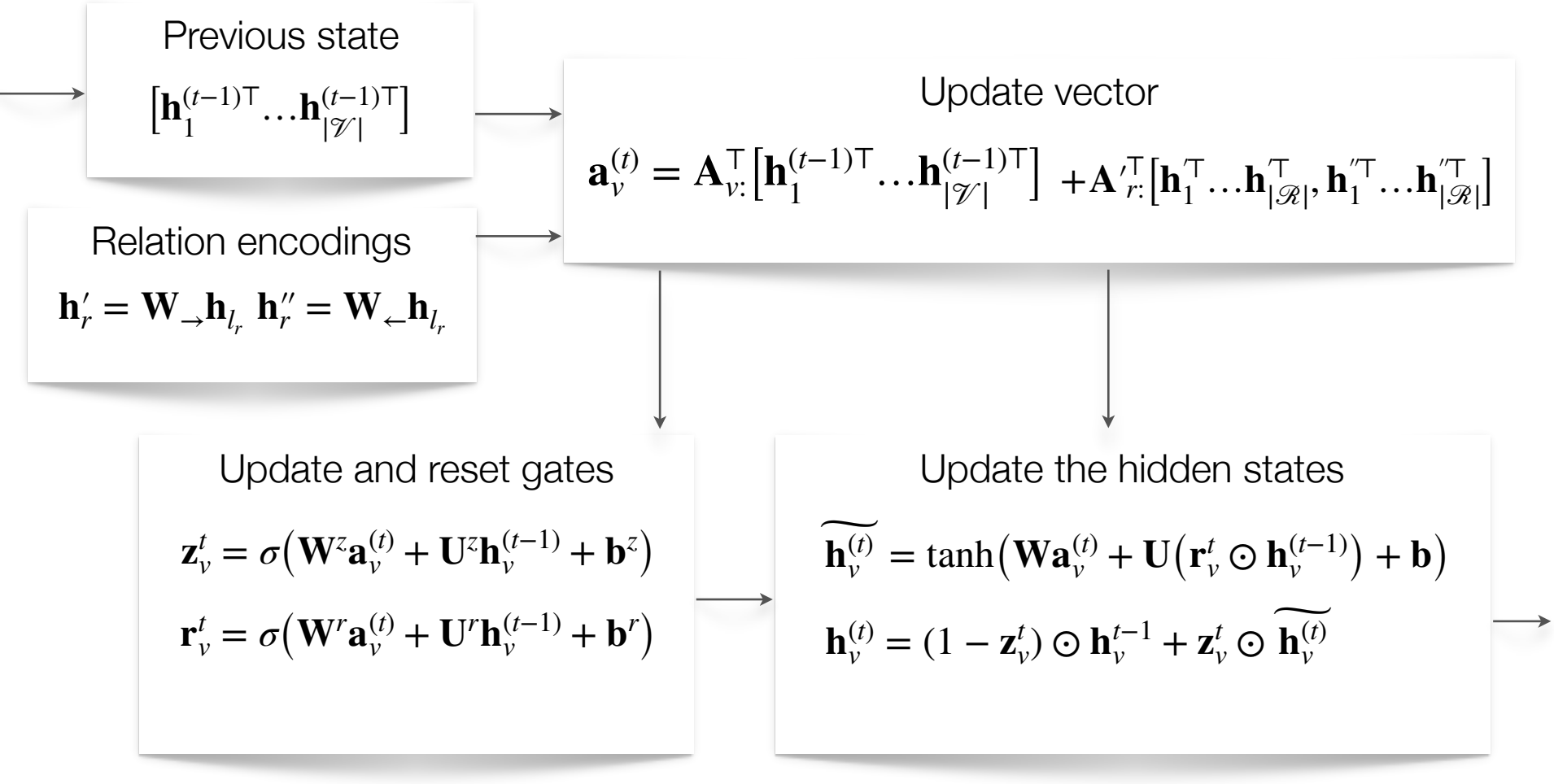
$$\mathbf{r}_v^t = \sigma\left(\mathbf{W}^r\mathbf{a}_v^{(t)} + \mathbf{U}^r\mathbf{h}_v^{(t-1)} + \mathbf{b}^r\right)$$

Update the hidden states

$$\widetilde{\mathbf{h}_v^{(t)}} = \tanh\left(\mathbf{W}\mathbf{a}_v^{(t)} + \mathbf{U}\left(\mathbf{r}_v^t \odot \mathbf{h}_v^{(t-1)}\right) + \mathbf{b}\right)$$

$$\mathbf{h}_v^{(t)} = (1 - \mathbf{z}_v^t) \odot \mathbf{h}_v^{t-1} + \mathbf{z}_v^t \odot \widetilde{\mathbf{h}_v^{(t)}}$$

Propagation Model



Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Update vector

$$\mathbf{a}_v^{(t)} = \mathbf{A}_{v:}^{\top}\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right] + \mathbf{A}'^{\top}_{r:}\left[\mathbf{h}_1'^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}'^{\top}, \mathbf{h}_1''^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}''^{\top}\right]$$

Update and reset gates

$$\mathbf{z}_v^t = \sigma\left(\mathbf{W}^z\mathbf{a}_v^{(t)} + \mathbf{U}^z\mathbf{h}_v^{(t-1)} + \mathbf{b}^z\right)$$

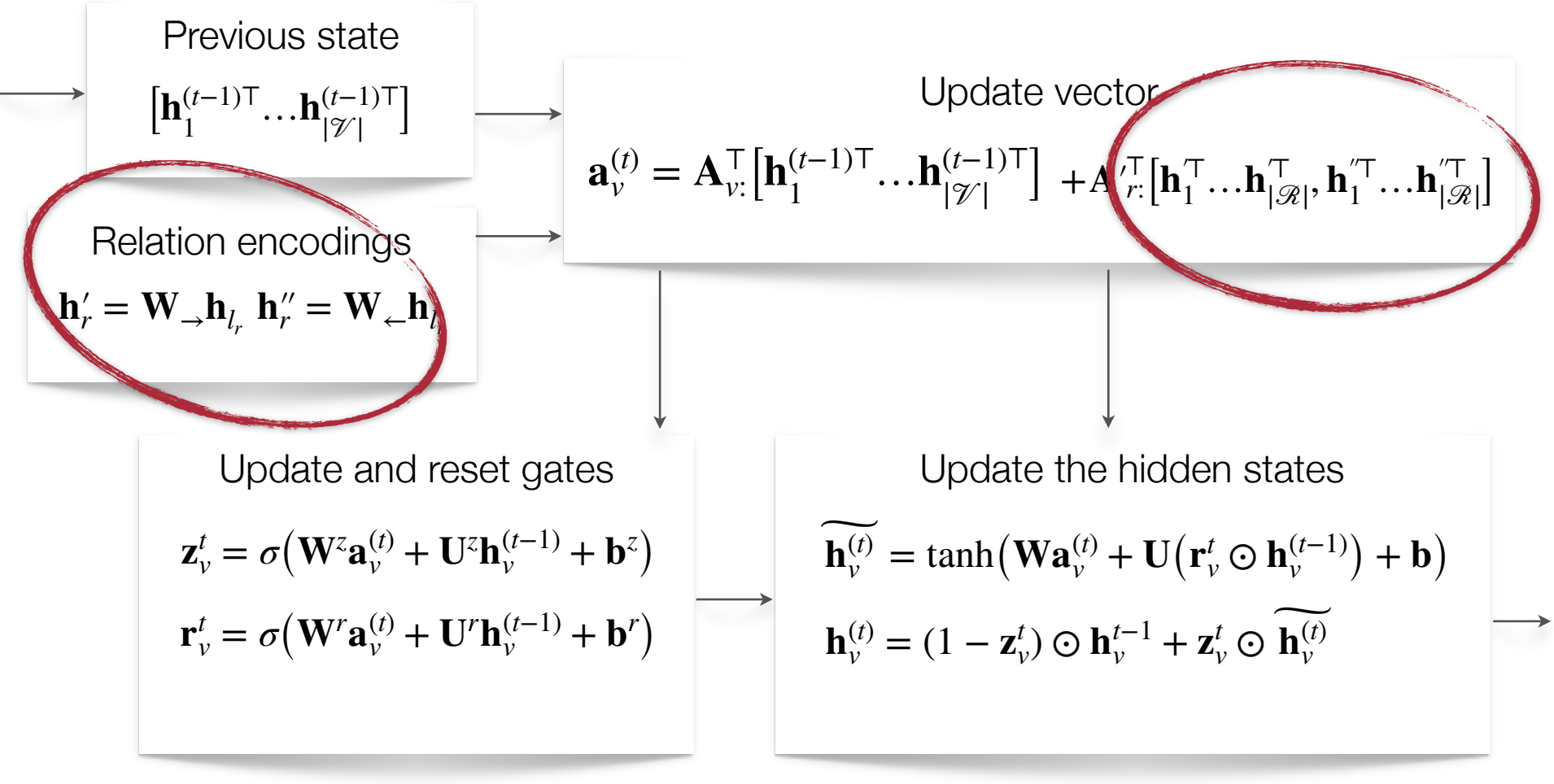
$$\mathbf{r}_v^t = \sigma\left(\mathbf{W}^r\mathbf{a}_v^{(t)} + \mathbf{U}^r\mathbf{h}_v^{(t-1)} + \mathbf{b}^r\right)$$

Update the hidden states

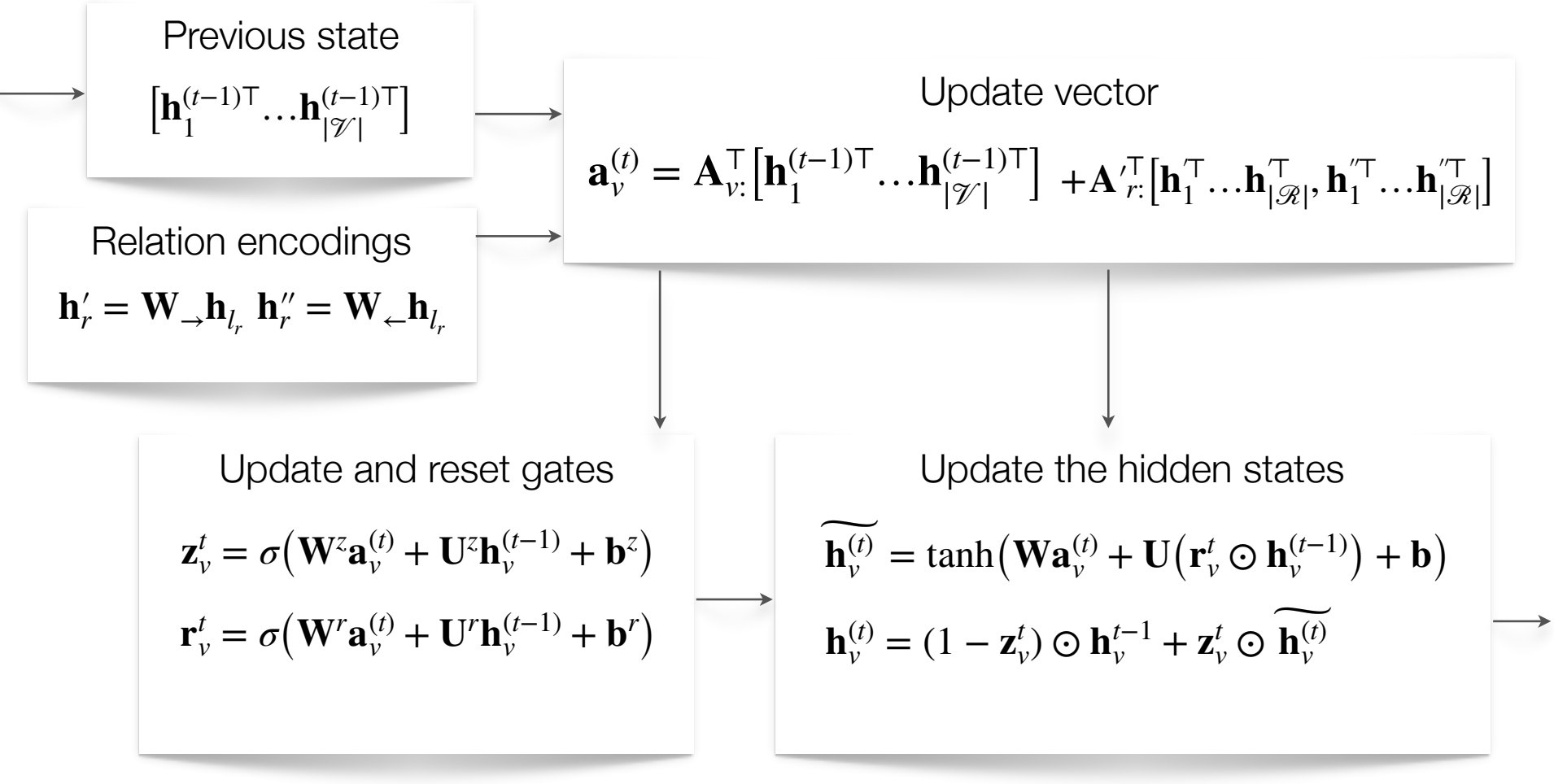
$$\widetilde{\mathbf{h}_v^{(t)}} = \tanh\left(\mathbf{W}\mathbf{a}_v^{(t)} + \mathbf{U}\left(\mathbf{r}_v^t \odot \mathbf{h}_v^{(t-1)}\right) + \mathbf{b}\right)$$

$$\mathbf{h}_v^{(t)} = (1 - \mathbf{z}_v^t) \odot \mathbf{h}_v^{t-1} + \mathbf{z}_v^t \odot \widetilde{\mathbf{h}_v^{(t)}}$$

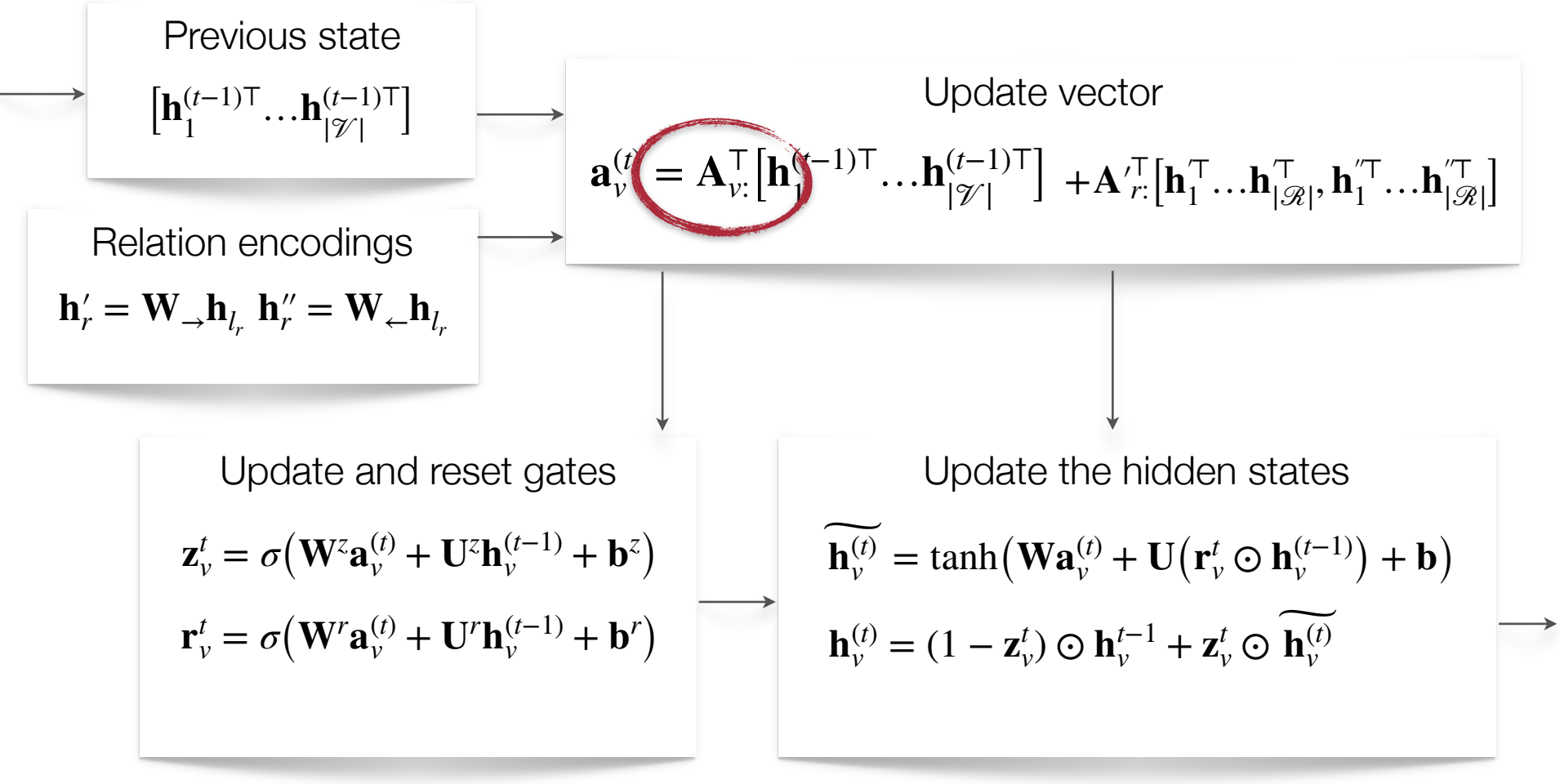
Propagation Model



Propagation Model



Propagation Model



Propagation Model

Previous state

$$\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right]$$

Relation encodings

$$\mathbf{h}'_r = \mathbf{W}_{\rightarrow}\mathbf{h}_{l_r} \quad \mathbf{h}''_r = \mathbf{W}_{\leftarrow}\mathbf{h}_{l_r}$$

Update vector

$$\mathbf{a}_v^{(t)} = \mathbf{A}_{v:}^{\top}\left[\mathbf{h}_1^{(t-1)\top} \ldots \mathbf{h}_{|\mathscr{V}|}^{(t-1)\top}\right] + \mathbf{A}'^{\top}_{r:}\left[\mathbf{h}_1'^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}'^{\top}, \mathbf{h}_1''^{\top} \ldots \mathbf{h}_{|\mathscr{R}|}''^{\top}\right]$$

Update and reset gates

$$\mathbf{z}_v^t = \sigma\left(\mathbf{W}^z\mathbf{a}_v^{(t)} + \mathbf{U}^z\mathbf{h}_v^{(t-1)} + \mathbf{b}^z\right)$$

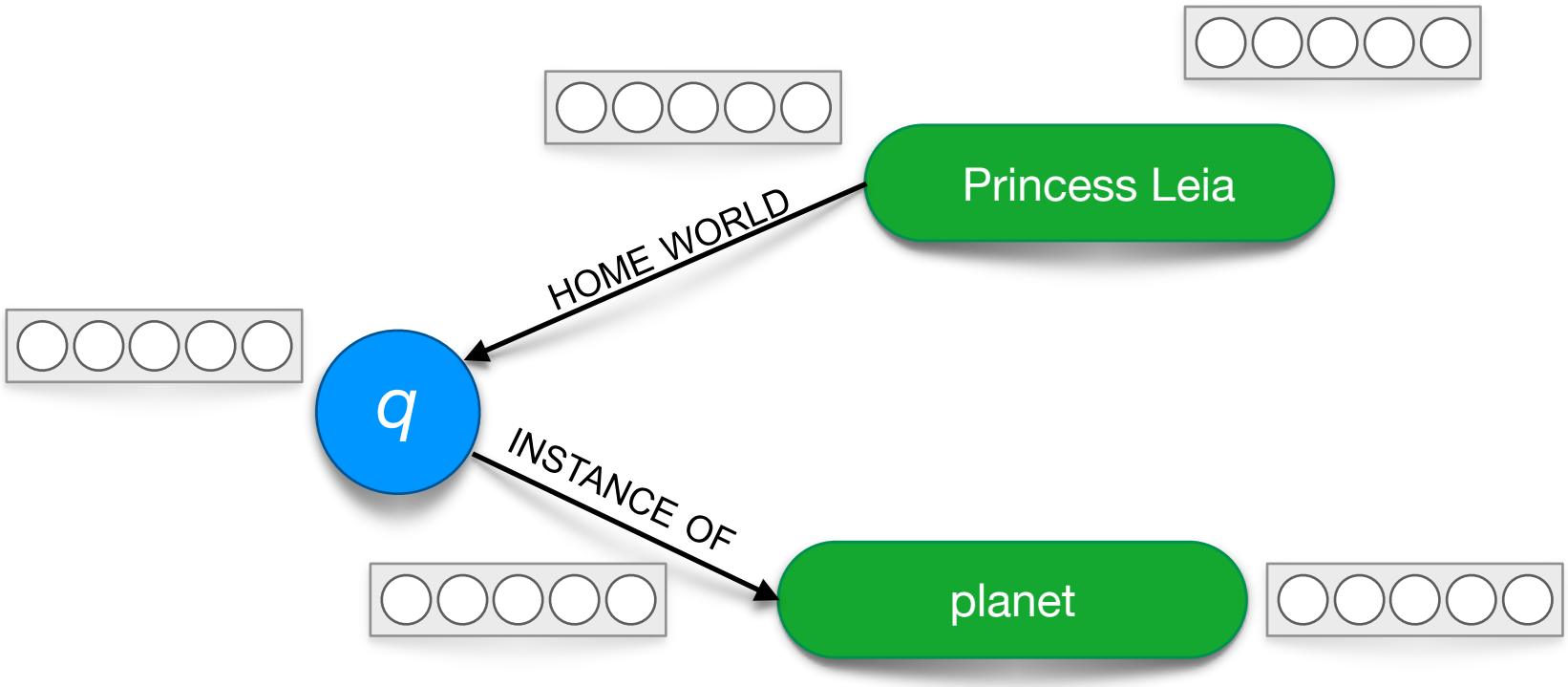
$$\mathbf{r}_v^t = \sigma\left(\mathbf{W}^r\mathbf{a}_v^{(t)} + \mathbf{U}^r\mathbf{h}_v^{(t-1)} + \mathbf{b}^r\right)$$

Update the hidden states

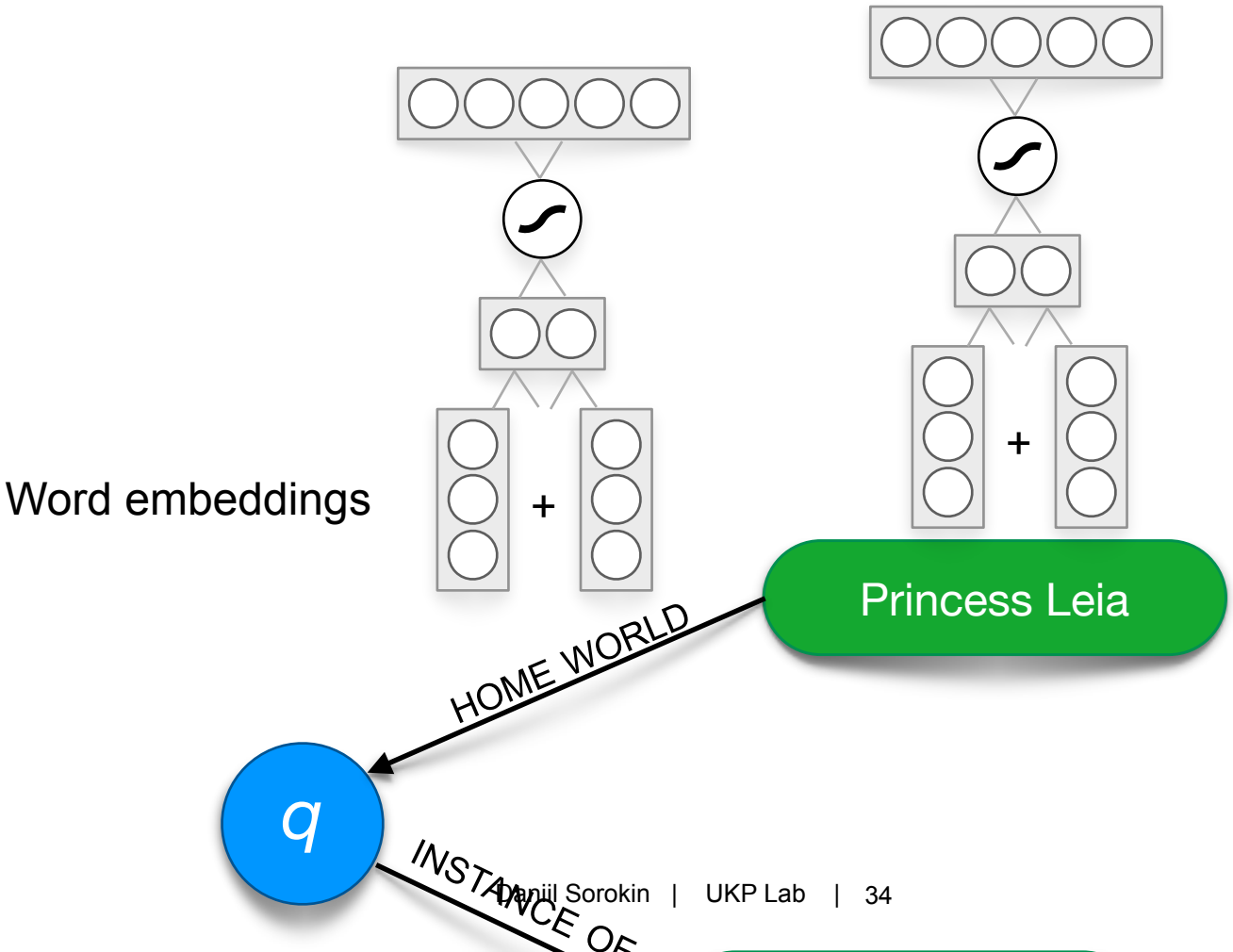
$$\widetilde{\mathbf{h}_v^{(t)}} = \tanh\left(\mathbf{W}\mathbf{a}_v^{(t)} + \mathbf{U}\left(\mathbf{r}_v^t \odot \mathbf{h}_v^{(t-1)}\right) + \mathbf{b}\right)$$

$$\mathbf{h}_v^{(t)} = (1 - \mathbf{z}_v^t) \odot \mathbf{h}_v^{t-1} + \mathbf{z}_v^t \odot \widetilde{\mathbf{h}_v^{(t)}}$$

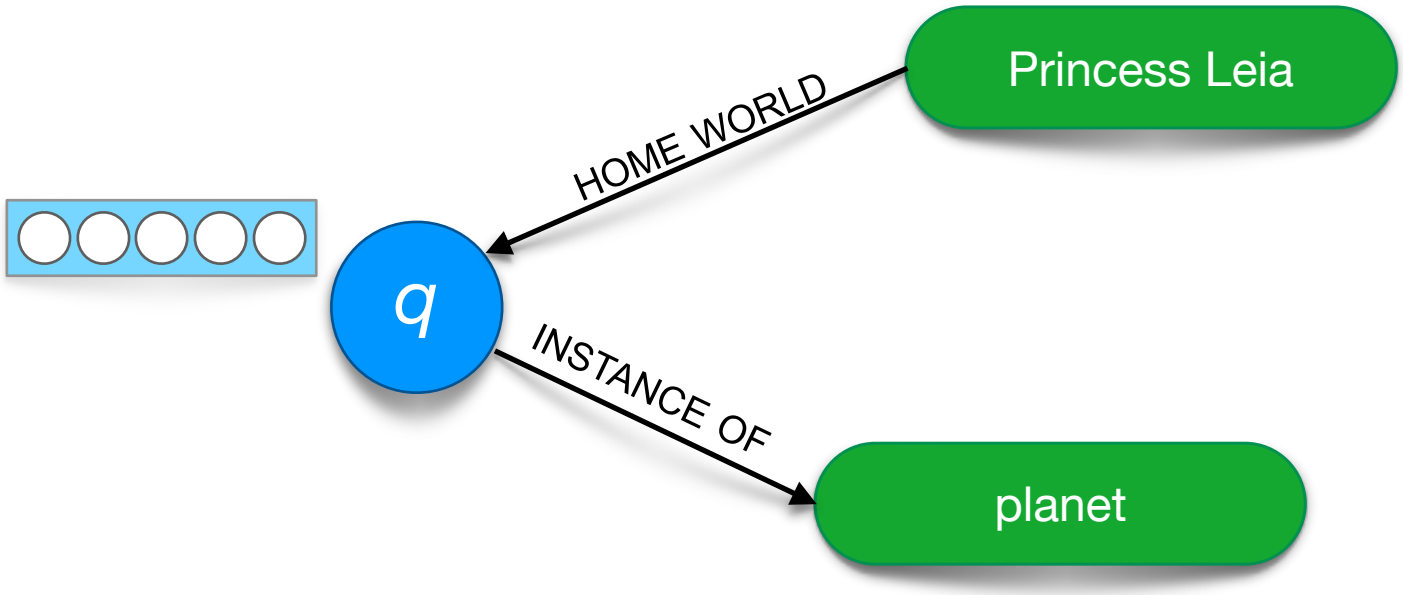
Entity and Relation Embeddings



Entity and Relation Encodings

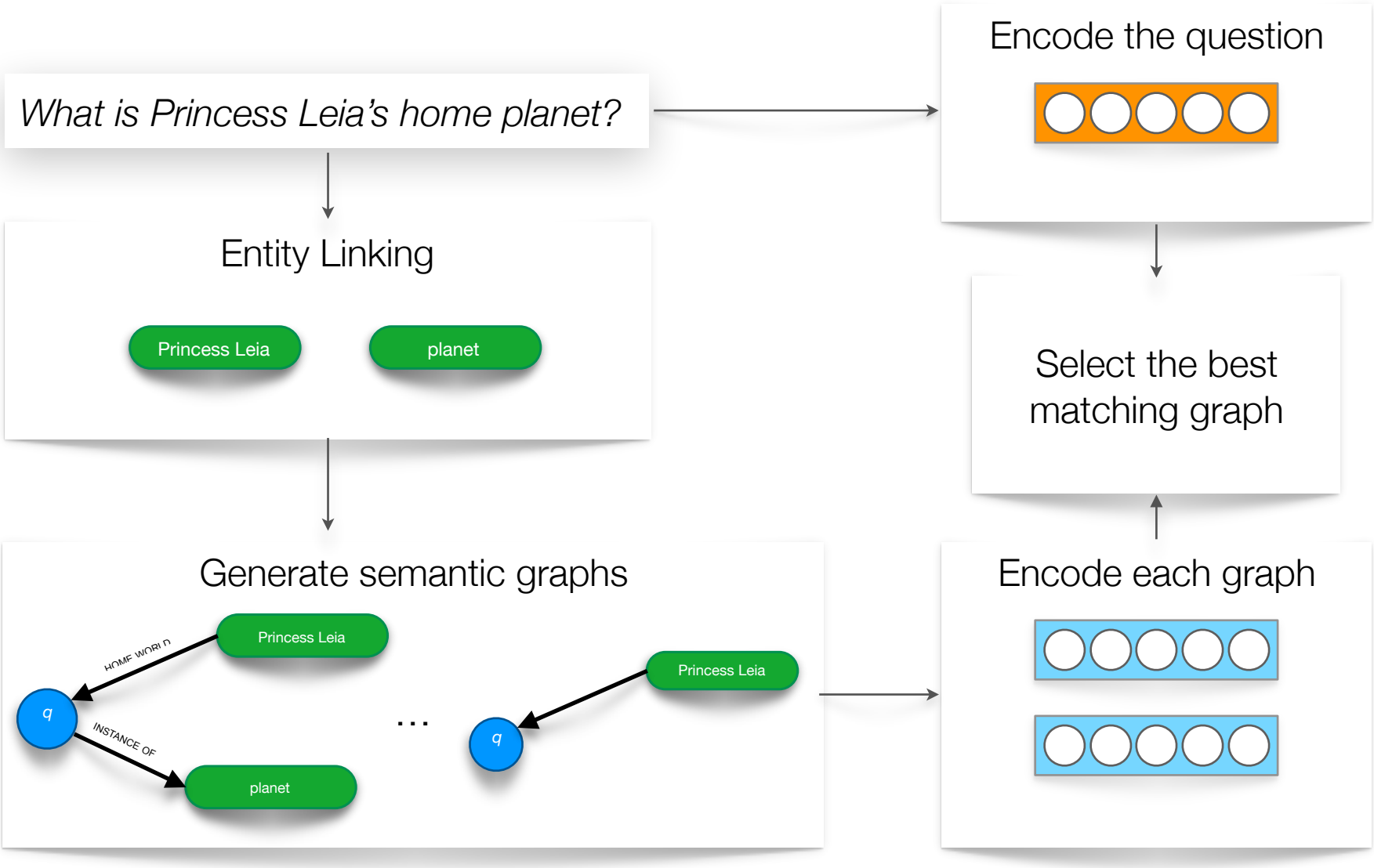


Graph Encoding

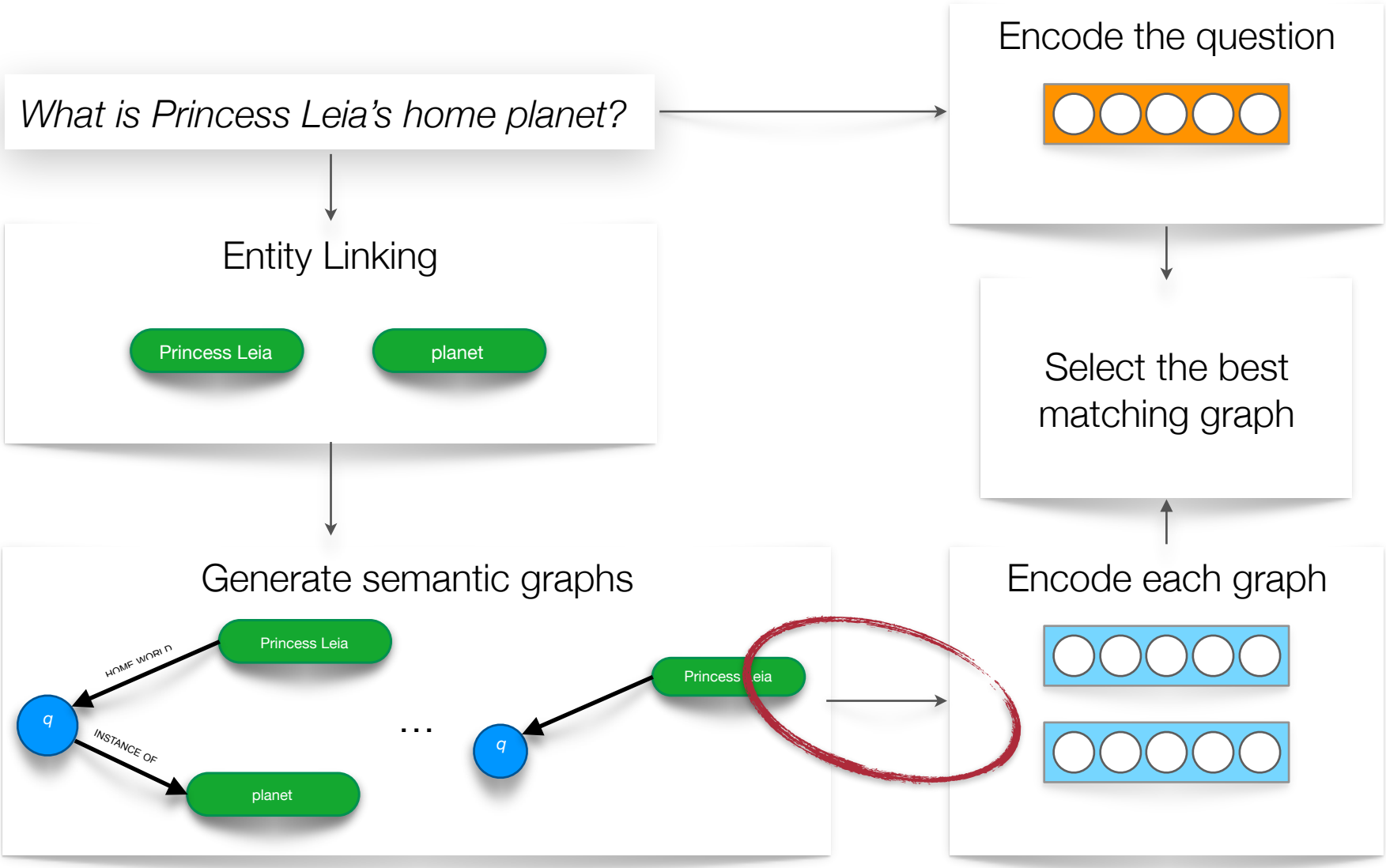


The Graph Neural Network models all aspects of the semantic graph representation

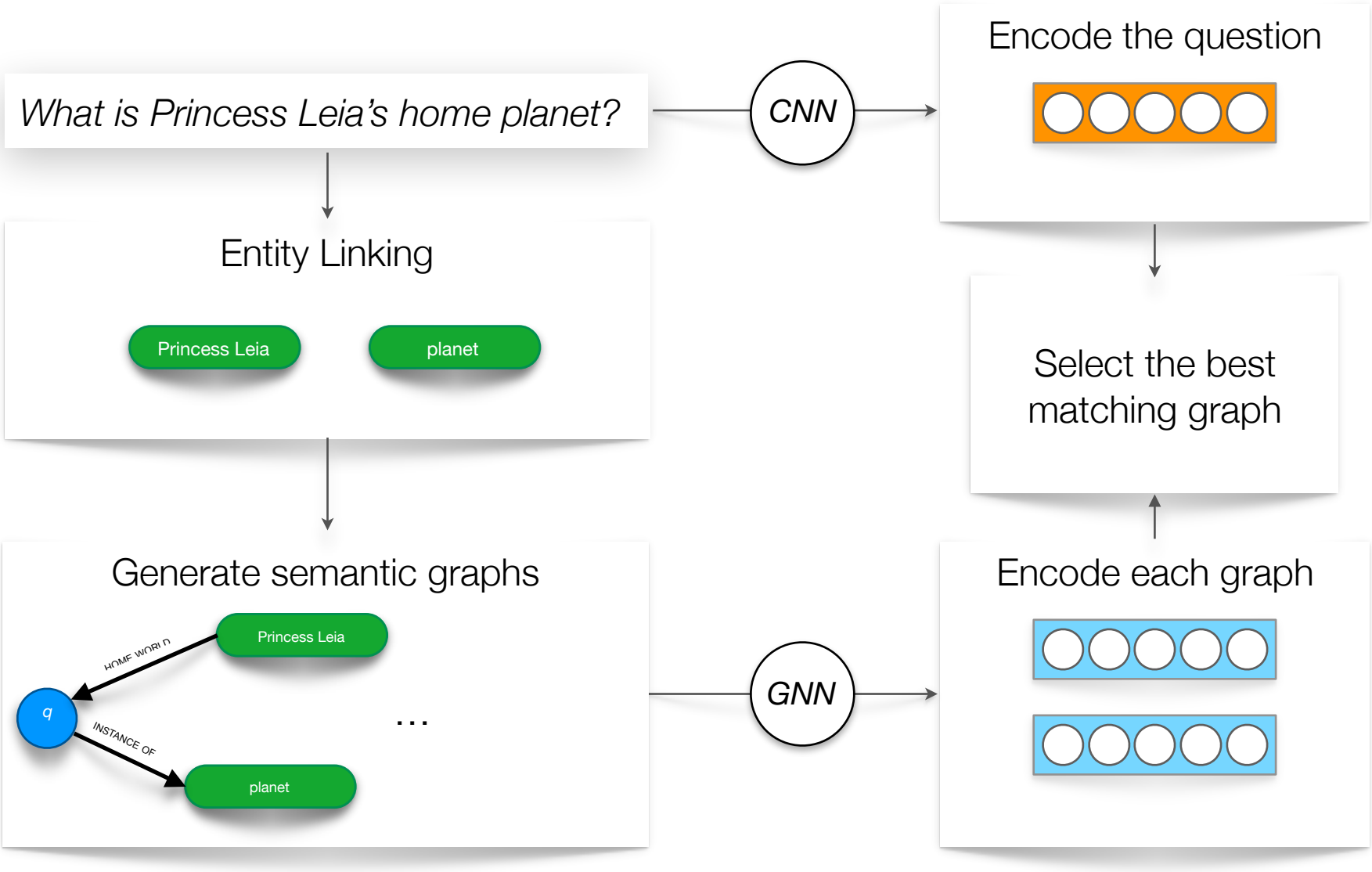
Question Answering Pipeline – Representation Learning



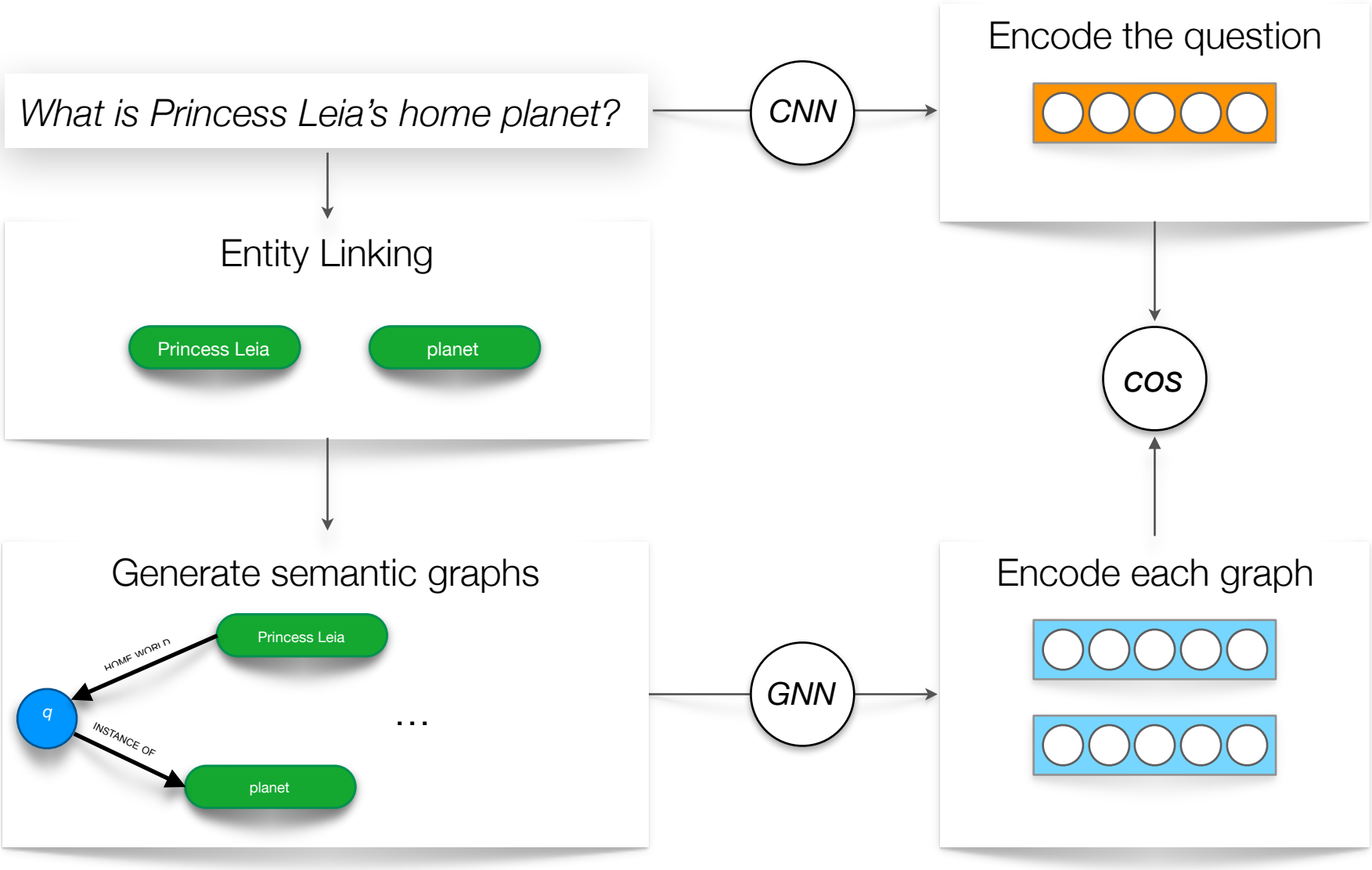
Question Answering Pipeline – Representation Learning



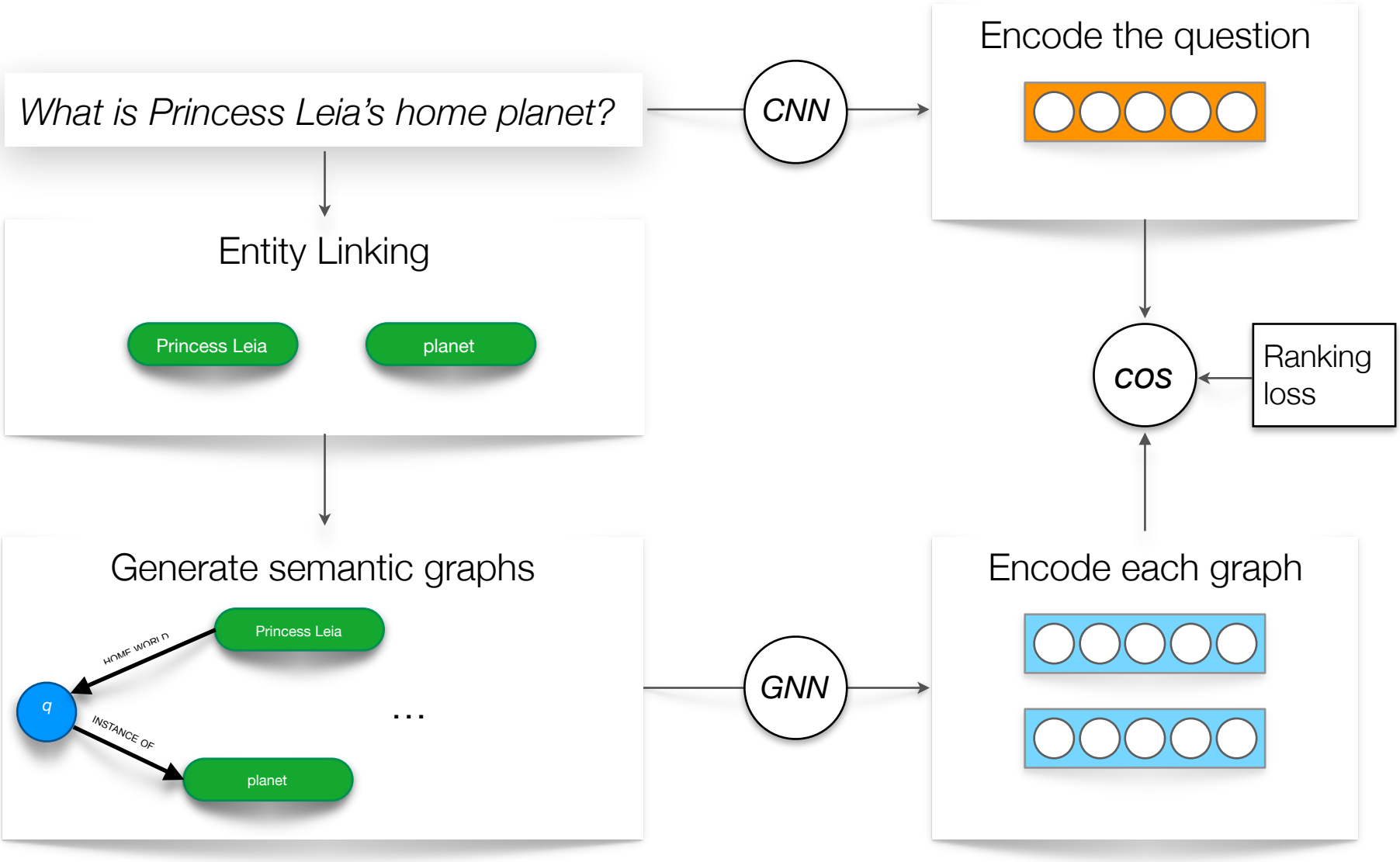
Question Answering Pipeline – Representation Learning



Question Answering Pipeline – Representation Learning



Question Answering Pipeline – Representation Learning



Experiments

Data sets

WebQSP-WD

Mapped to Wikidata

1033 test questions (293 complex)

2880 train questions (419 complex)

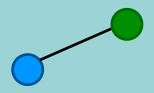
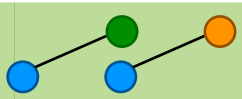
QALD-7

Manually constructed

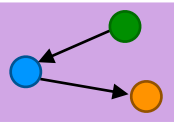
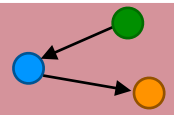
80 test questions (42 complex)

Models

Baselines

STAGG (Yih et al. 2015)	Features
Single Edge	
Pooled Edges	

Graph models

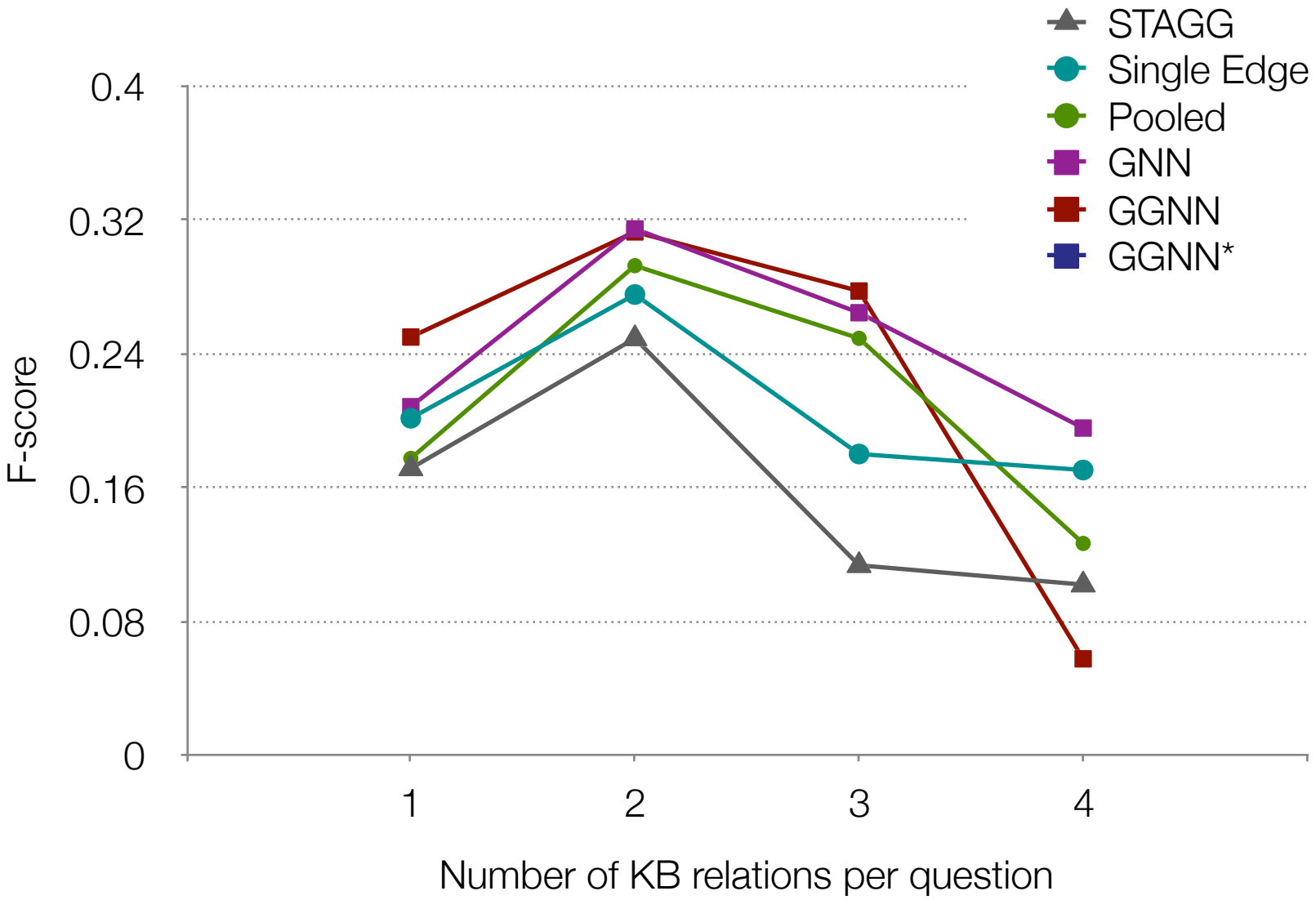
GNN	
Gated GNN	

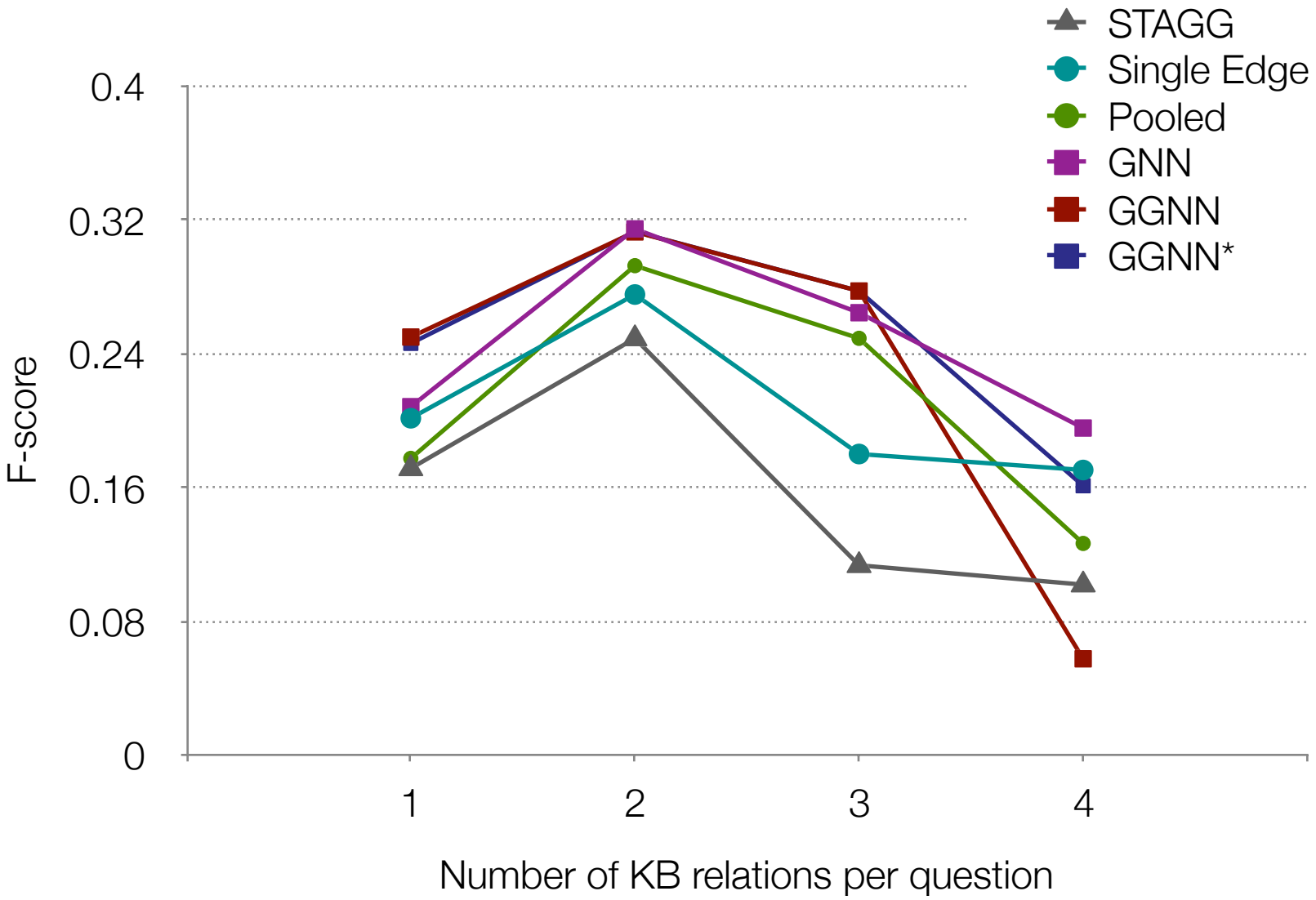
WebQSP

	Precision	Recall	F-score
STAGG	0.1911	0.2267	0.1828
Single Edge	0.2240	0.2713	0.2148
Pooled Edges	0.2094	0.2553	0.2032
GNN	0.2419	0.2890	0.2326
GGNN	0.2686	0.3179	**0.2588**

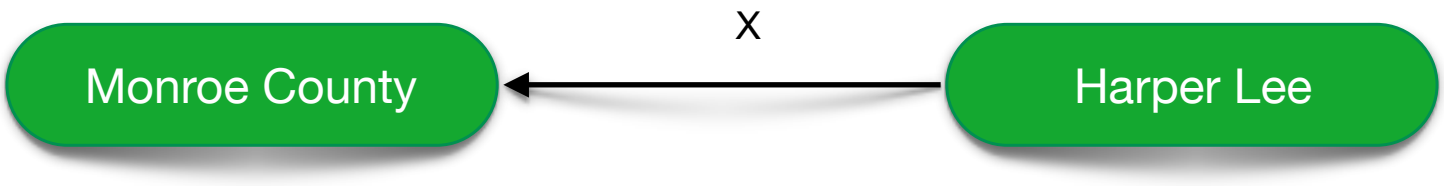
QALD-7

	Precision	Recall	F-score
STAGG	0.1934	0.2463	0.1861
Single Edge	0.2173	0.1963	0.1951
Pooled Edges	0.1904	0.1800	0.1605
GNN	0.1652	0.2072	0.1703
GGNN	0.2176	0.2751	**0.2131**


STAGG
Single Edge
Pooled
GNN
GGNN
GGNN*
0.4
0.32
0.24
0.16
0.08
0
F-score
1
2
3
4
Number of KB relations per question

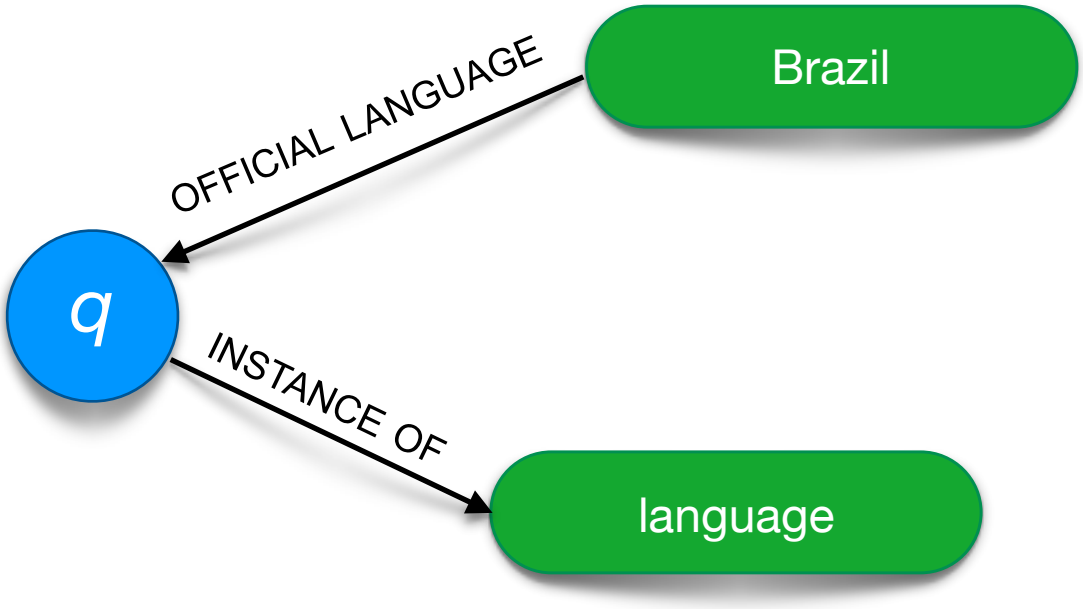

STAGG
Single Edge
Pooled
GNN
GGNN
GGNN*
0.4
0.32
0.24
0.16
0.08
0
F-score
1
2
3
4
Number of KB relations per question

Example output



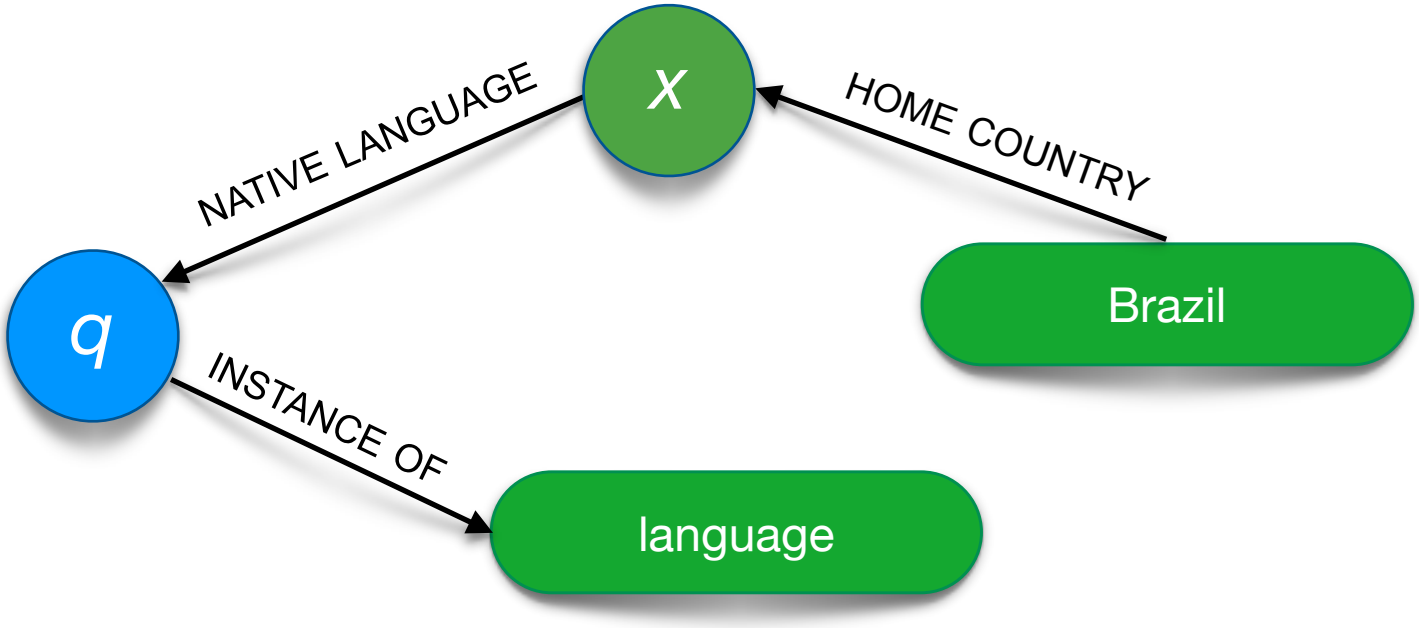
Where did Harper Lee attend high school?

Example output Pooled



What language do people speak in Brazil?

Example output GGNN

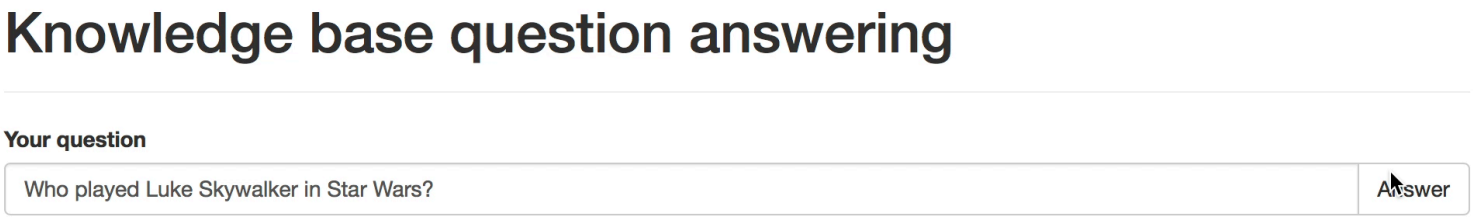


What language do people speak in Brazil?

Modelling the semantic structure
(even the most simple)
with **GNNs**
improves
the representation learning
and the knowledge base QA

Demo and code

https://github.com/UKPLab/coling2018-graph-neural-networks-question-answering



Follow up?

- https://daniilsorokin.github.io ⬅️ slides
- sorokin@ukp.informatik.tu-darmstadt.de
- @daniilmagpie
- COLING'18 Paper: http://aclweb.org/anthology/C18-1280

Manual analysis

	Pooled Edges	GGNN
F-score > 0.5	22.27%	28.37%
F-score > 0.0	29.82%	37.08%
Entity linker errors	6%	8%
Data set errors	22%	32%
Wrong prediction	72%	60%
hit@10	35.62%	44.05%