Multi-Hop Knowledge Graph Reasoning with Reward Shaping

Victoria Lin, Richard Socher, Caiming Xiong

{xilin, rsocher, cxiong}@salesforce.com

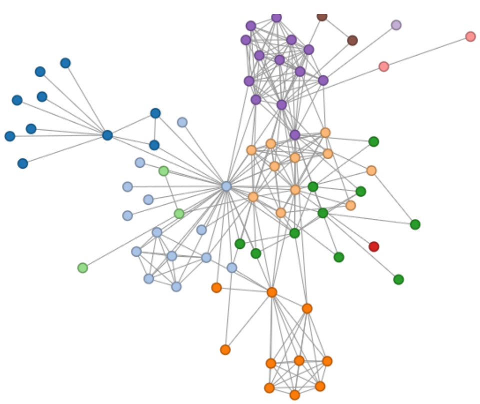
EMNLP 2018

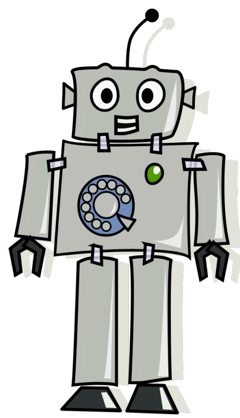
Question Answering System

Text

Images

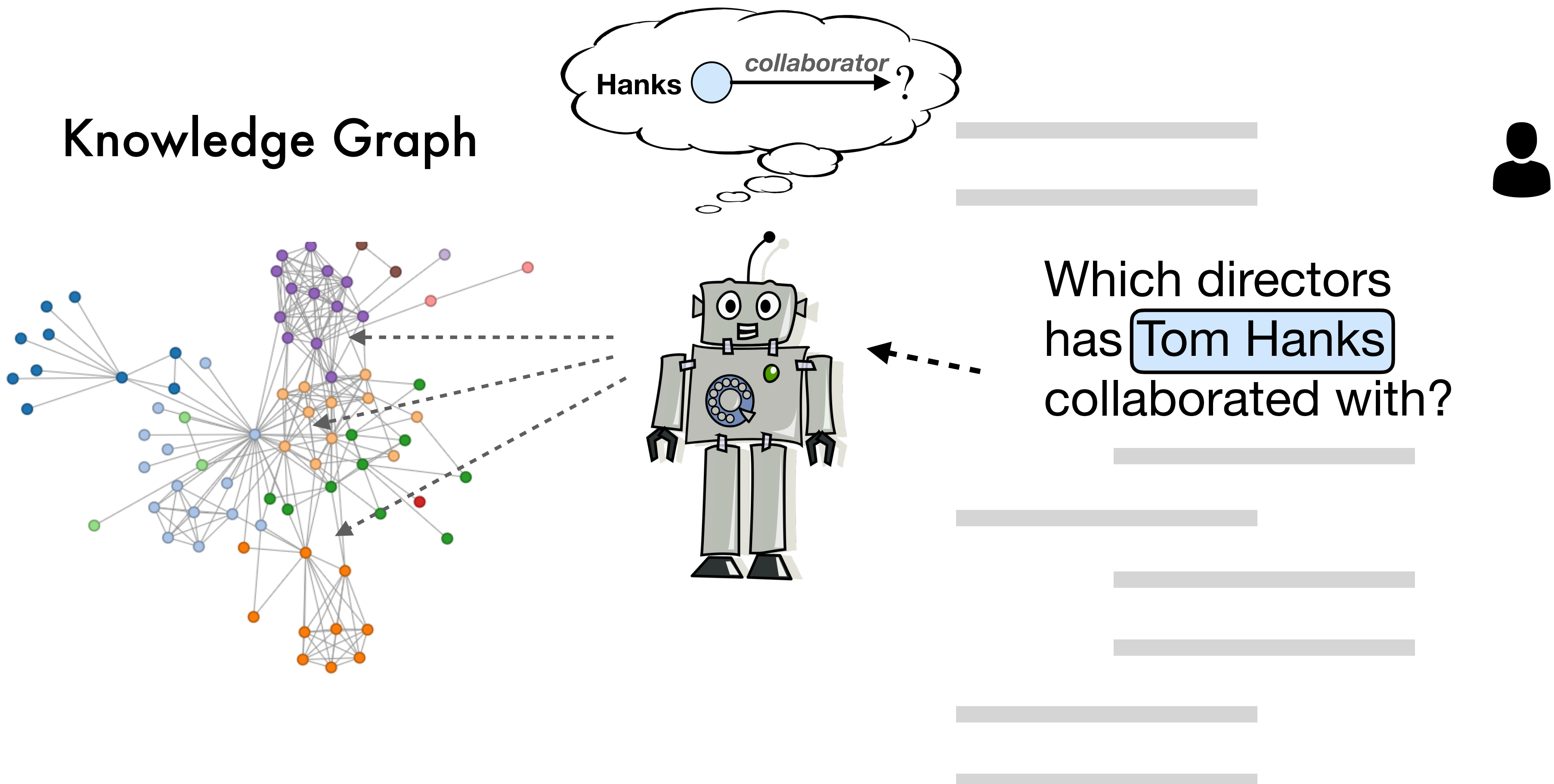
Knowledge Graph





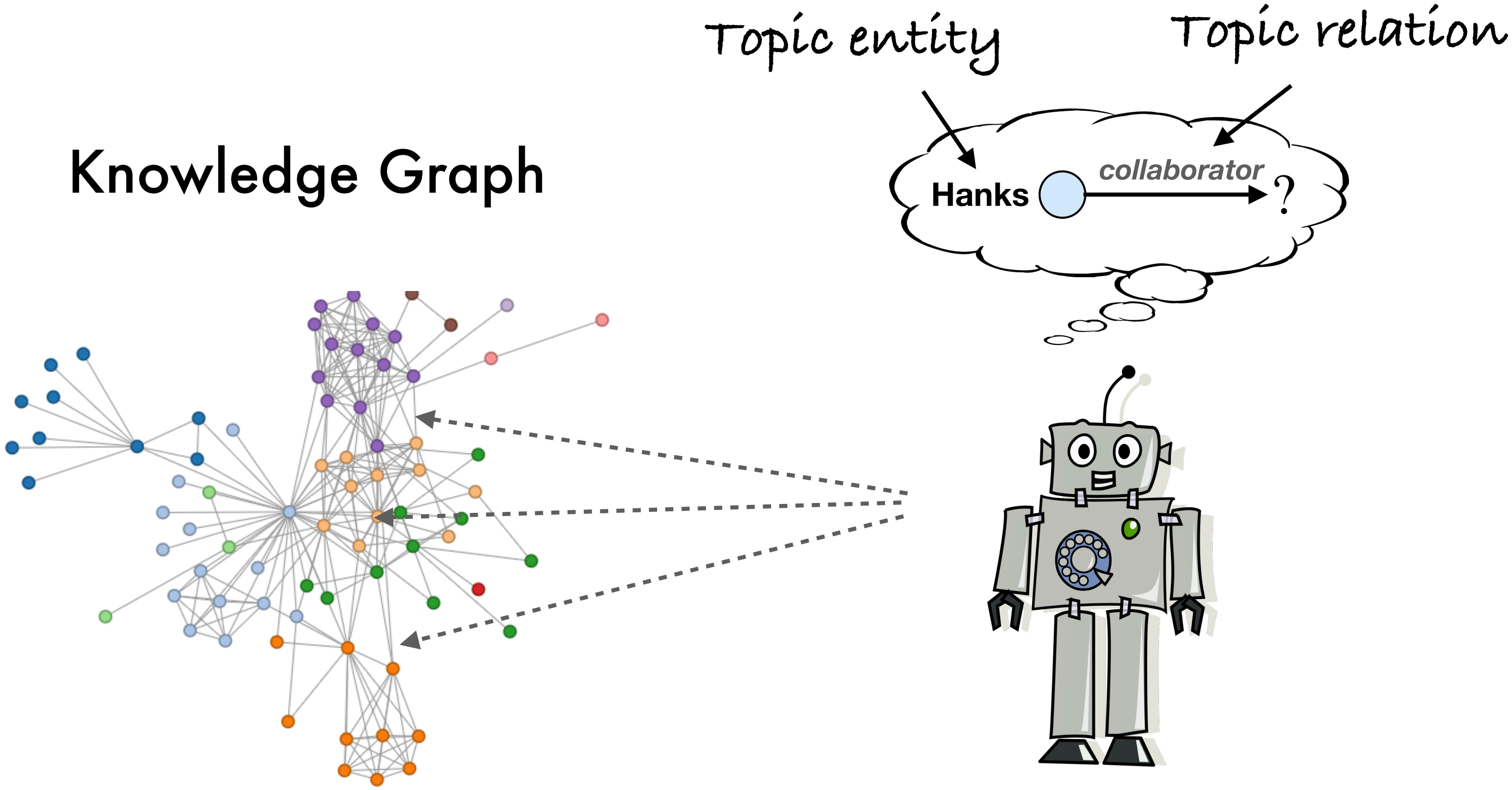
Which directors has Tom Hanks collaborated with?

Question Answering System

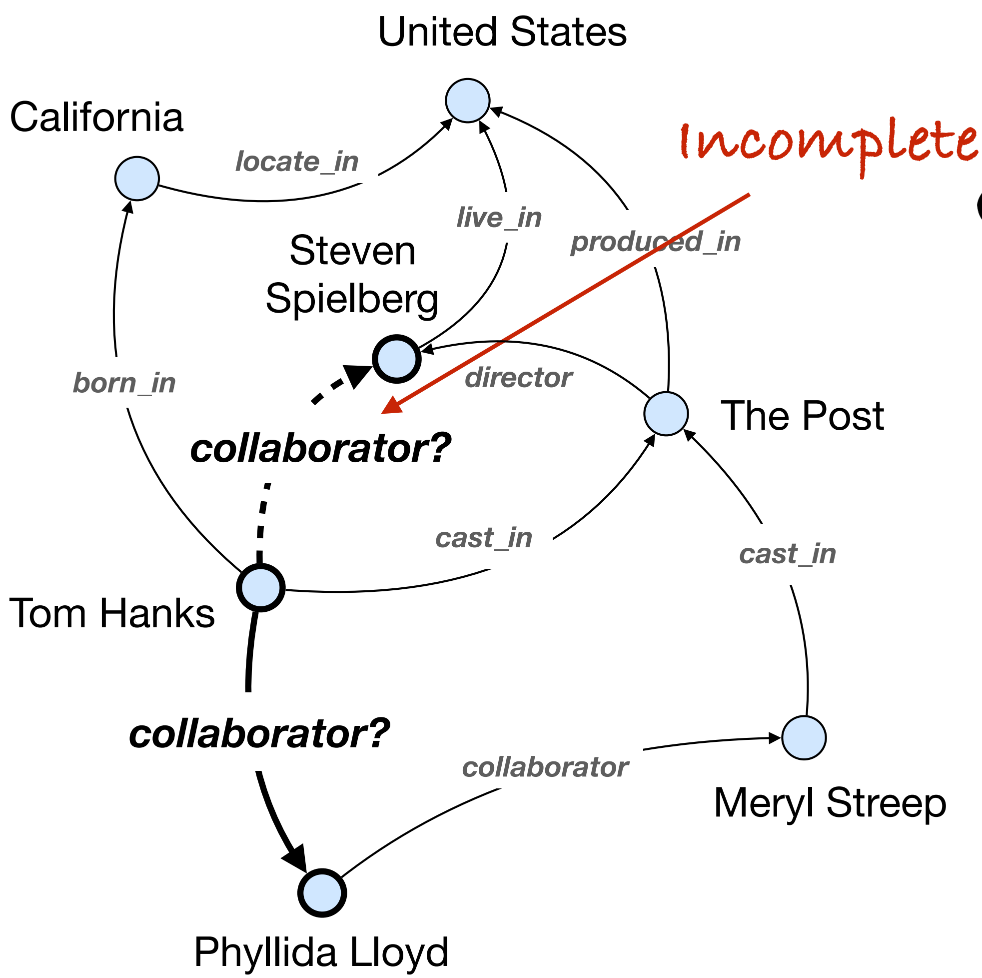


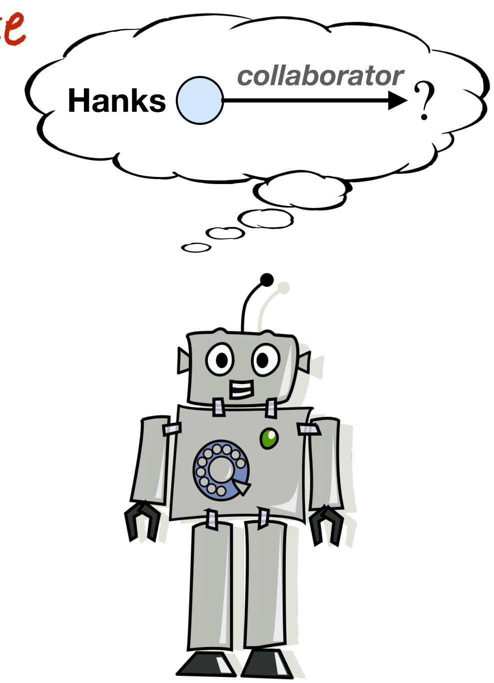
Structured Query Answering

Knowledge Graph

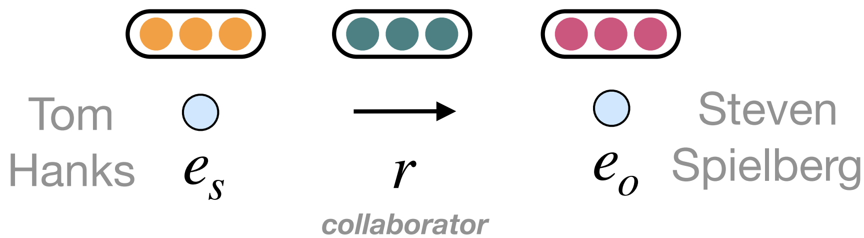


Structured Query Answering





Knowledge Graph Embeddings



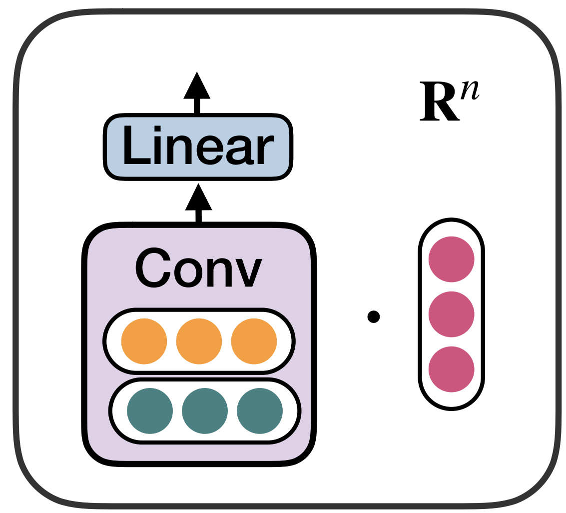
Highly accurate & Efficient

	MRR
ConvE	0.957 (max = 1)

Tab 1. ConvE query answering performance on the UMLS benchmark dataset (Kok and Domingos 2007)

Lack interpretabilit

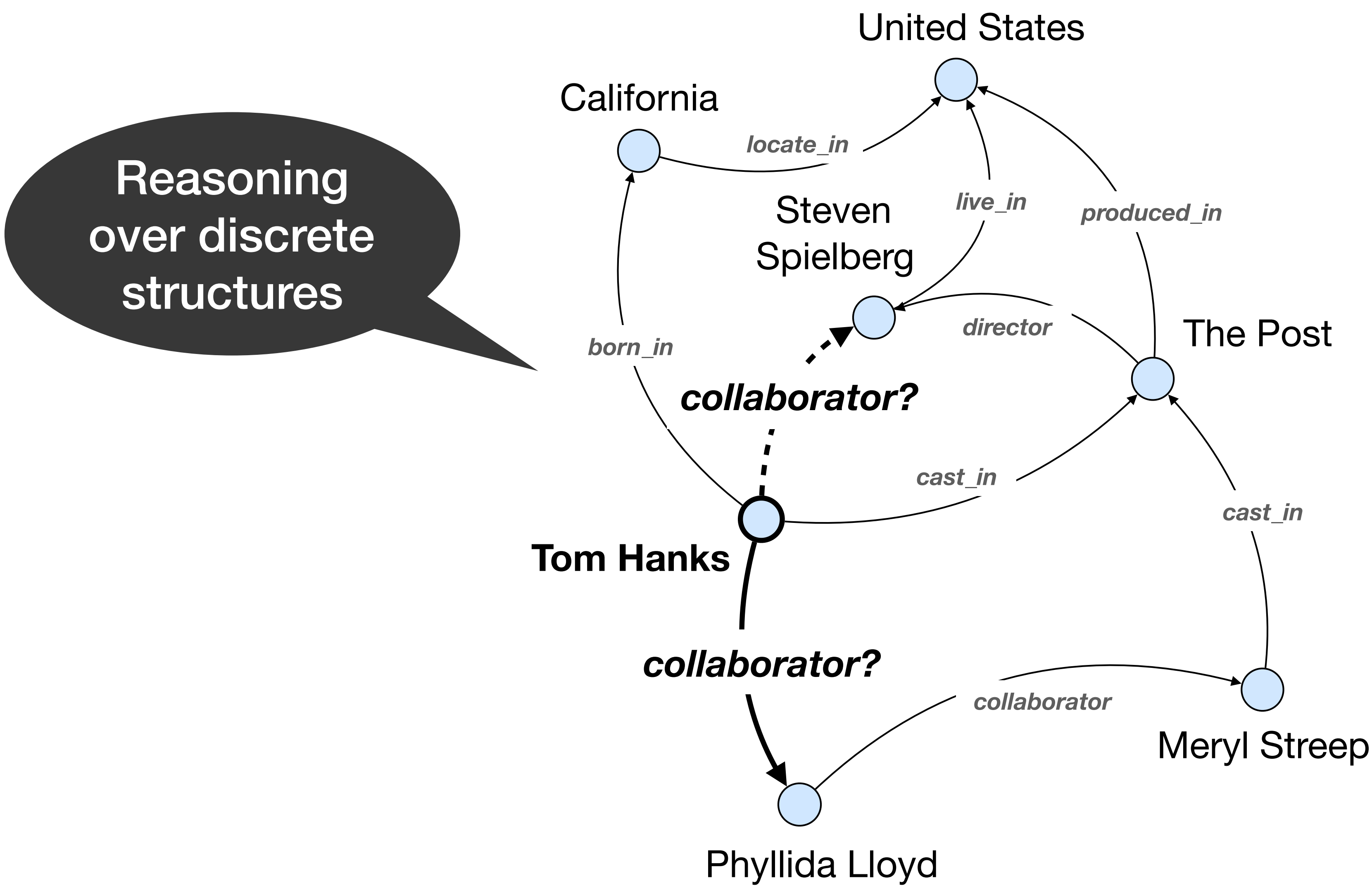
Why Spielberg is a collaborator of Hanks?



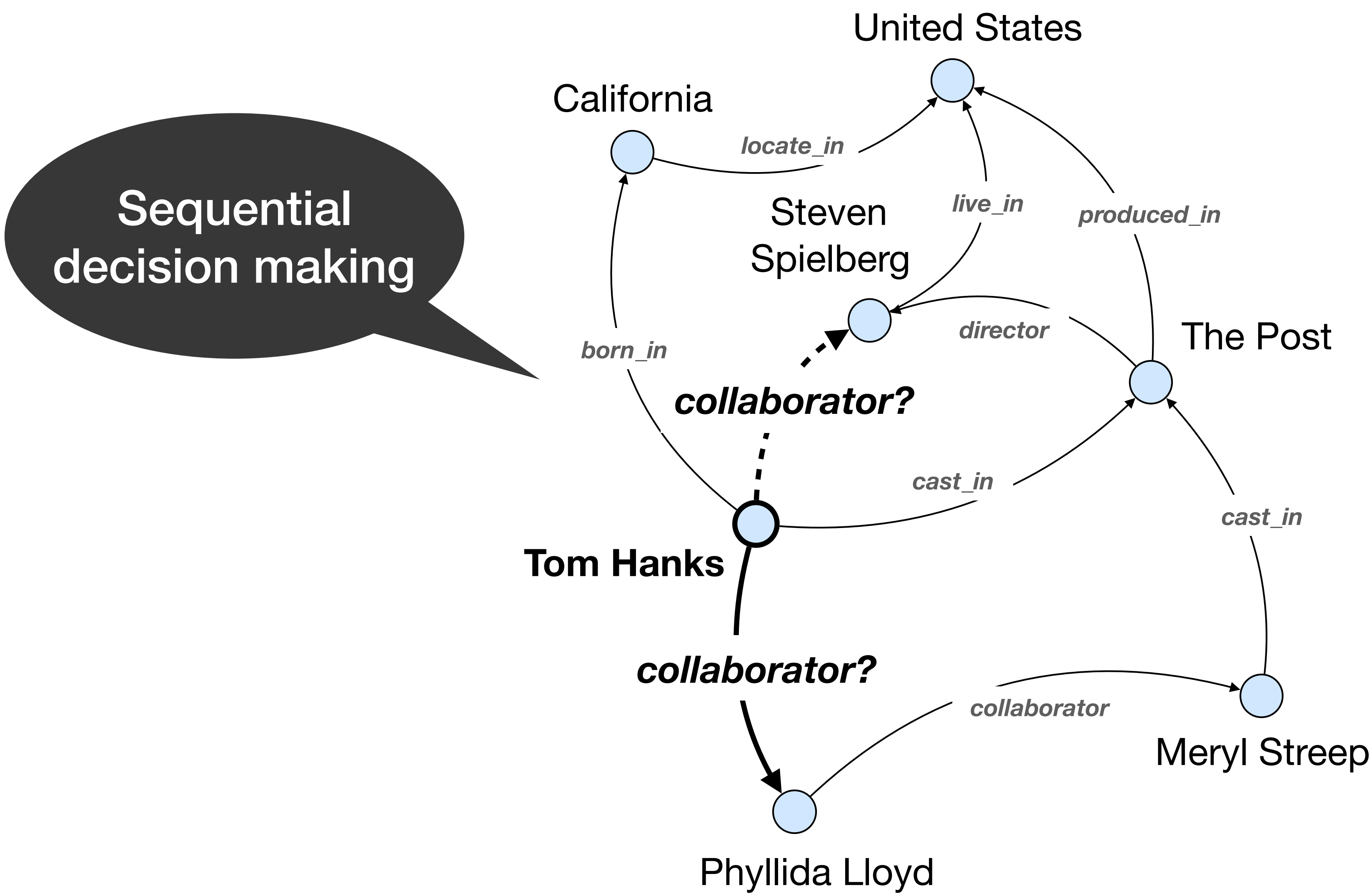
ConvE

DistMult (Yang et al. 2015), ComplEx (Trouillon et al. 2016), ConvE (Dettmers et al. 2018).

Multi-Hop Reasoning Models

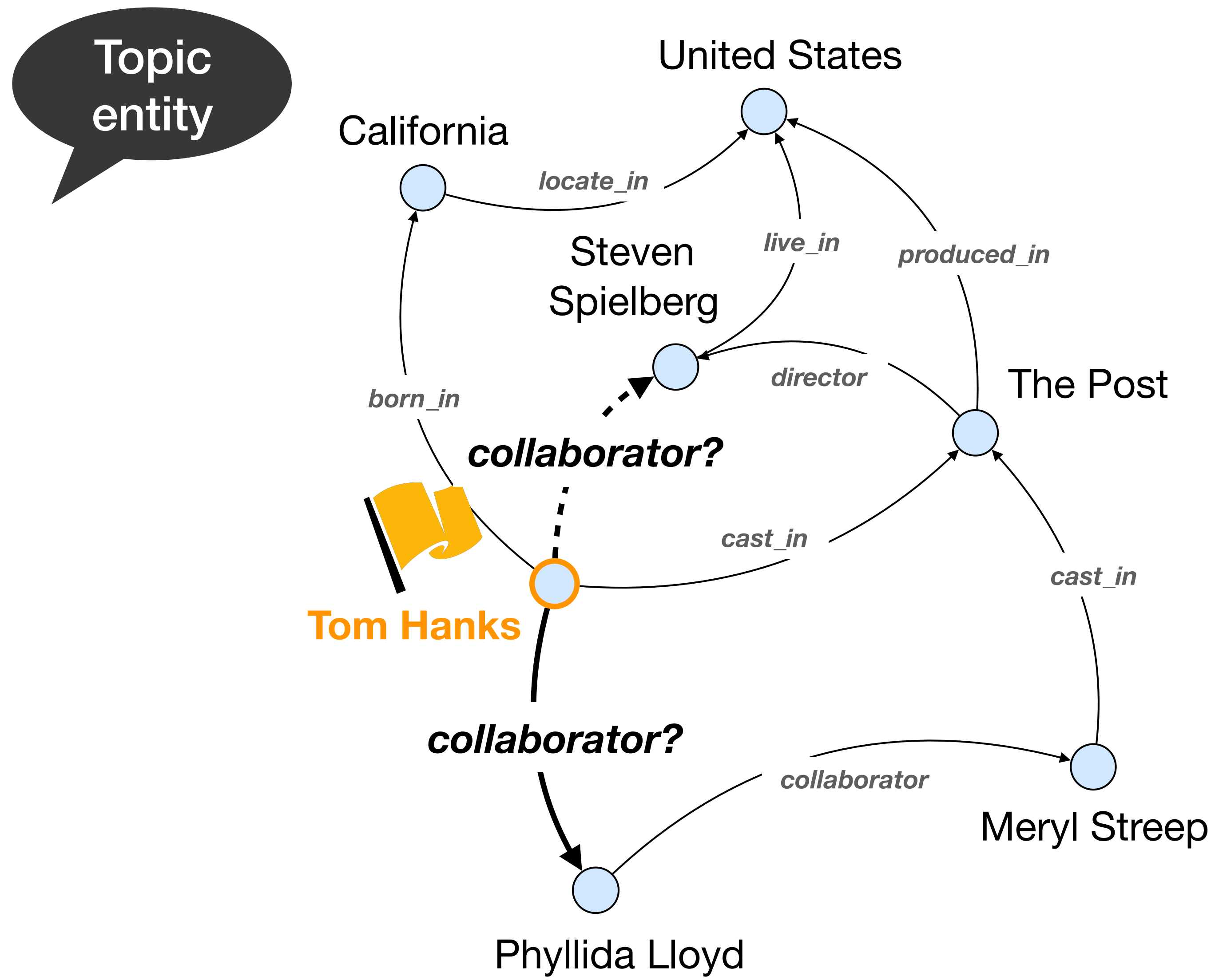


Multi-Hop Reasoning Models

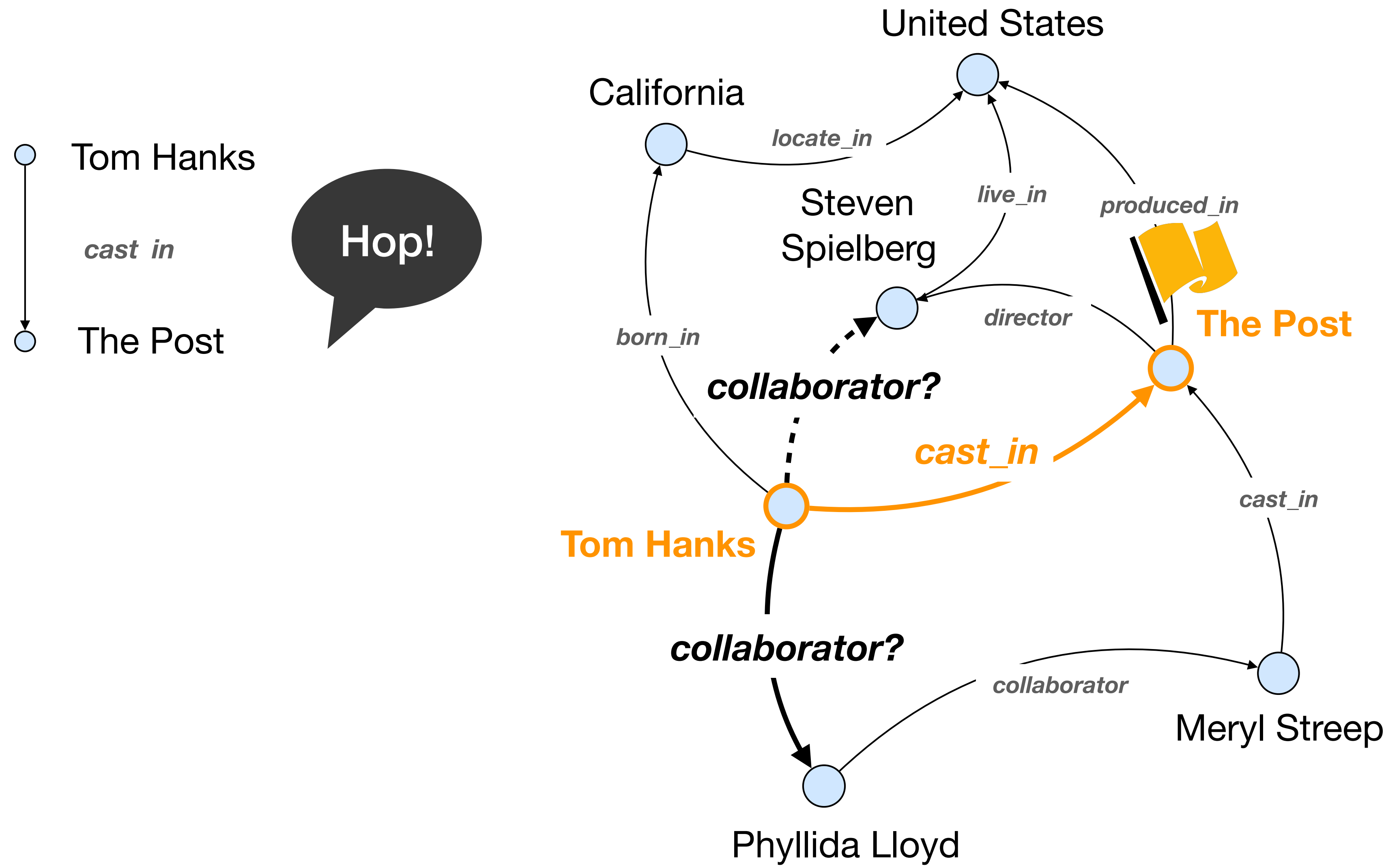


Multi-Hop Reasoning Models

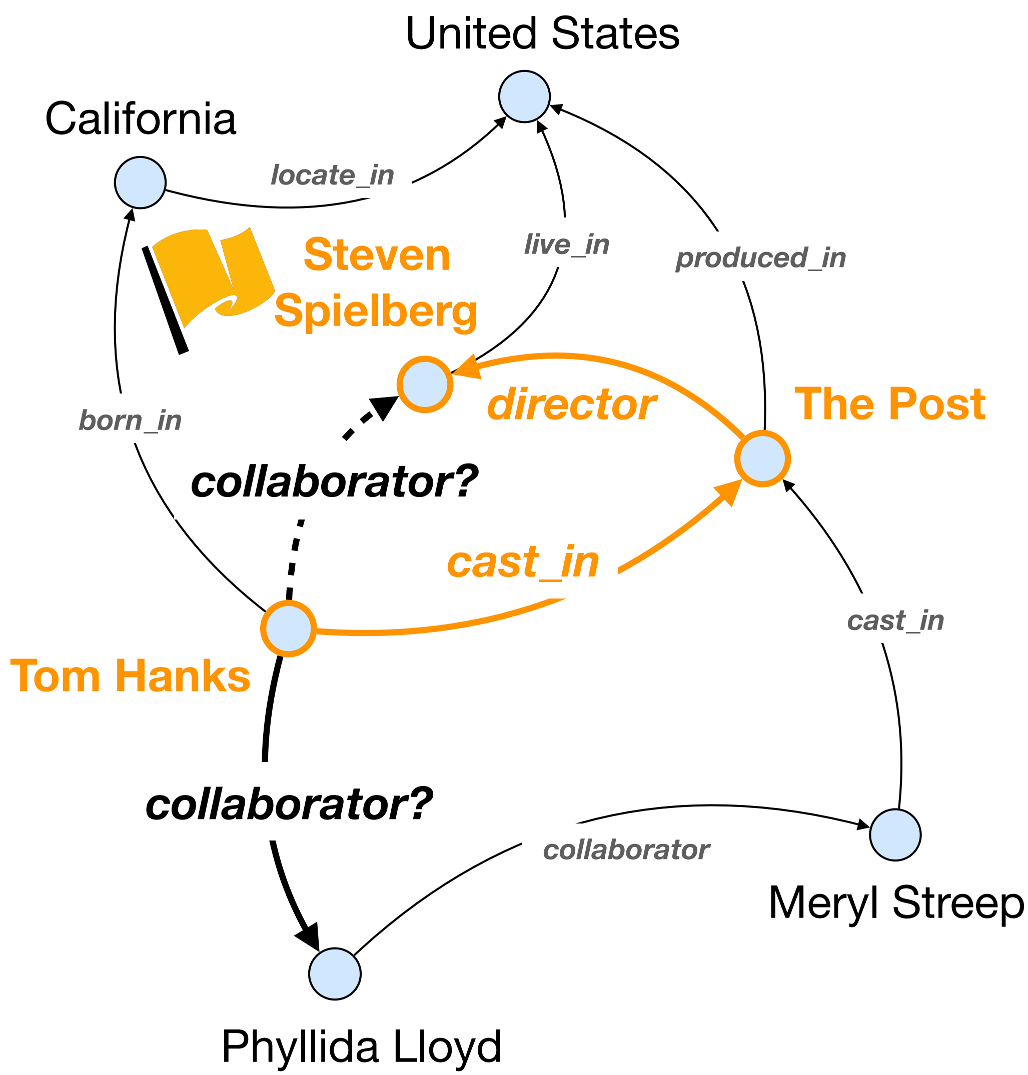
- Tom Hanks



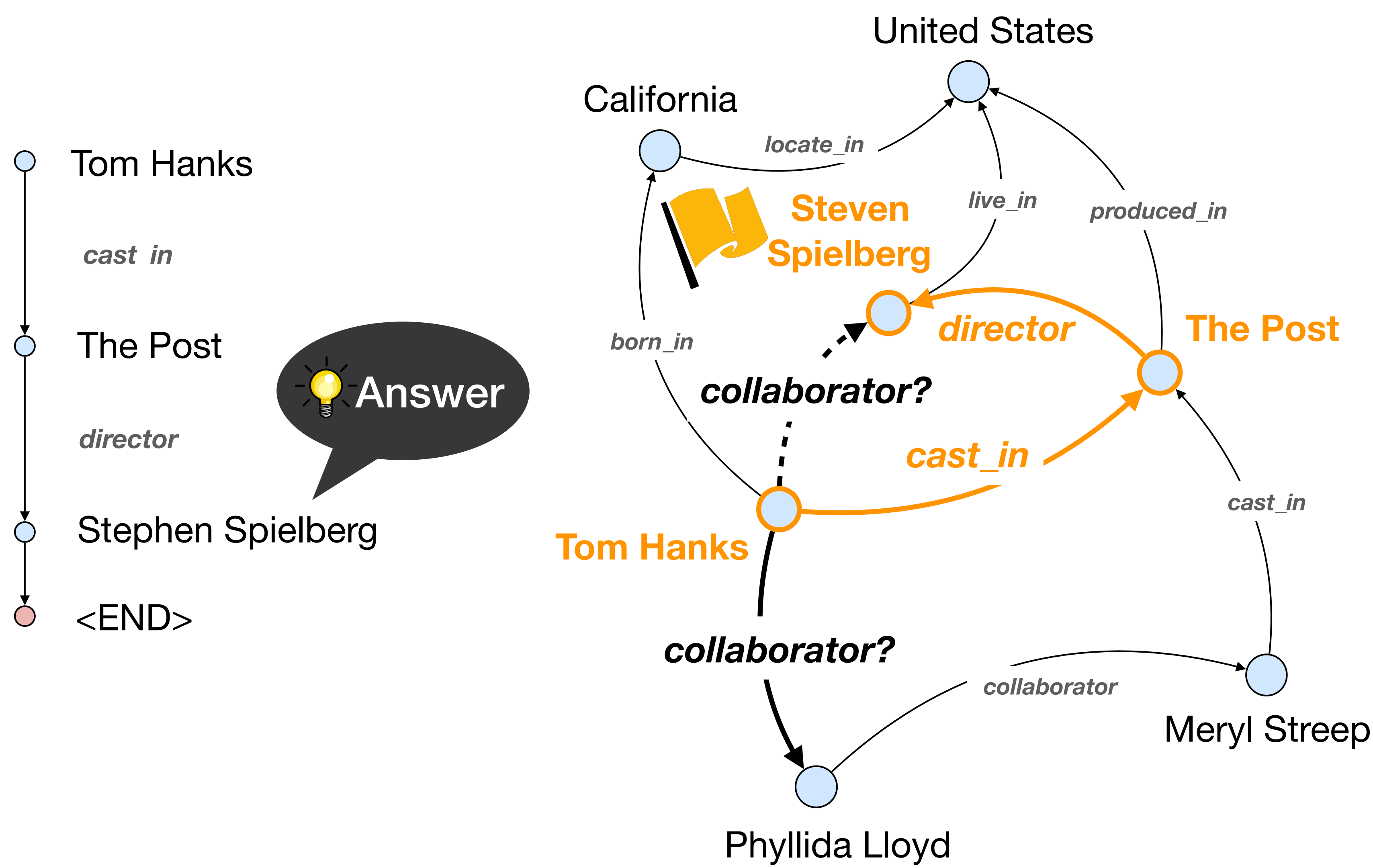
Multi-Hop Reasoning Models



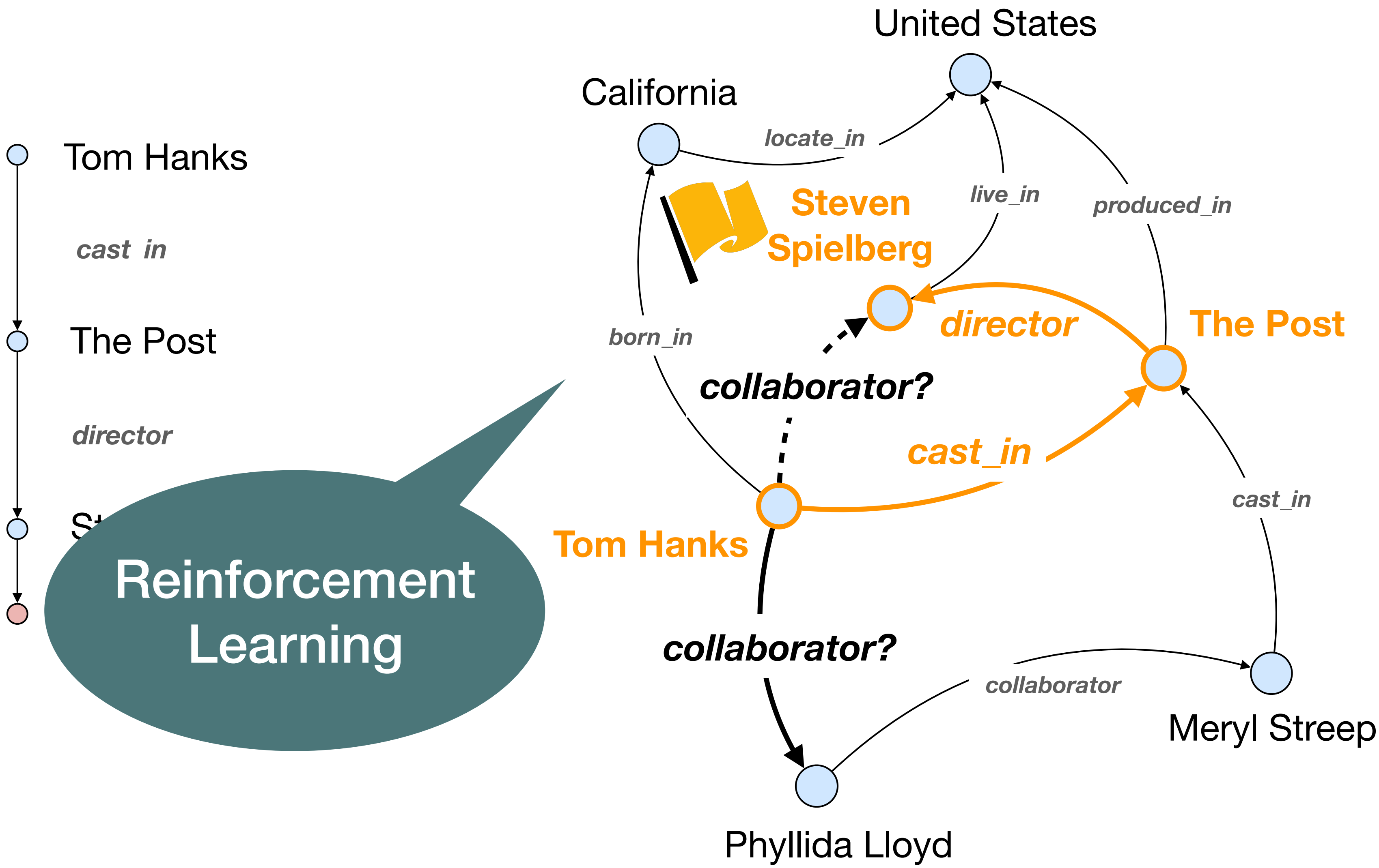
Multi-Hop Reasoning Models



Multi-Hop Reasoning Models



Multi-Hop Reasoning Models



MINERVA (Das et al. 2018)

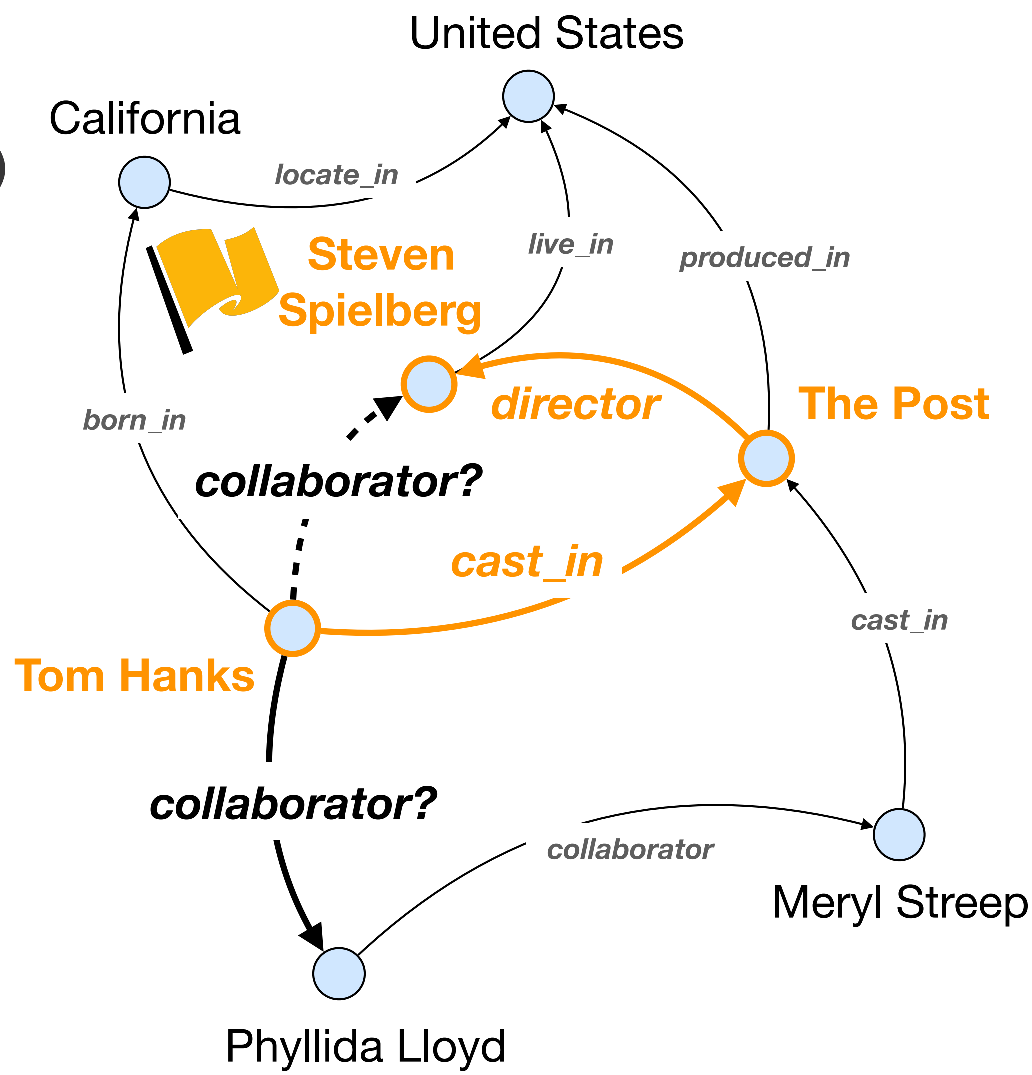
Multi-Hop Reasoning Models

Interpretable

Significant performance gap

	MRR
ConvE	0.957
RL	0.825

Tab 2. ConvE and RL (MINERVA) query answering performance on the UMLS benchmark dataset (Kok and Domingos 2007)

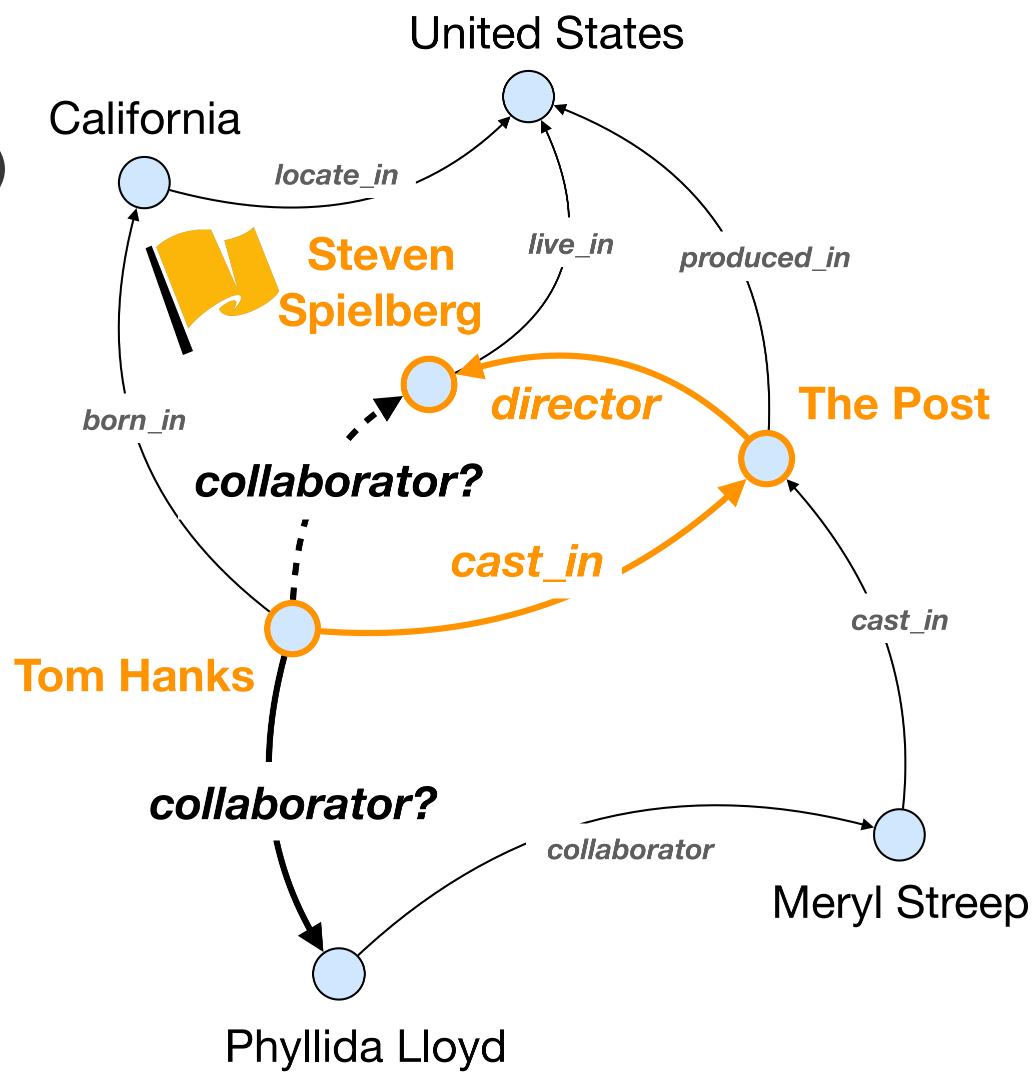


Multi-Hop Reasoning Models: Ideal Case

Significant performance gap

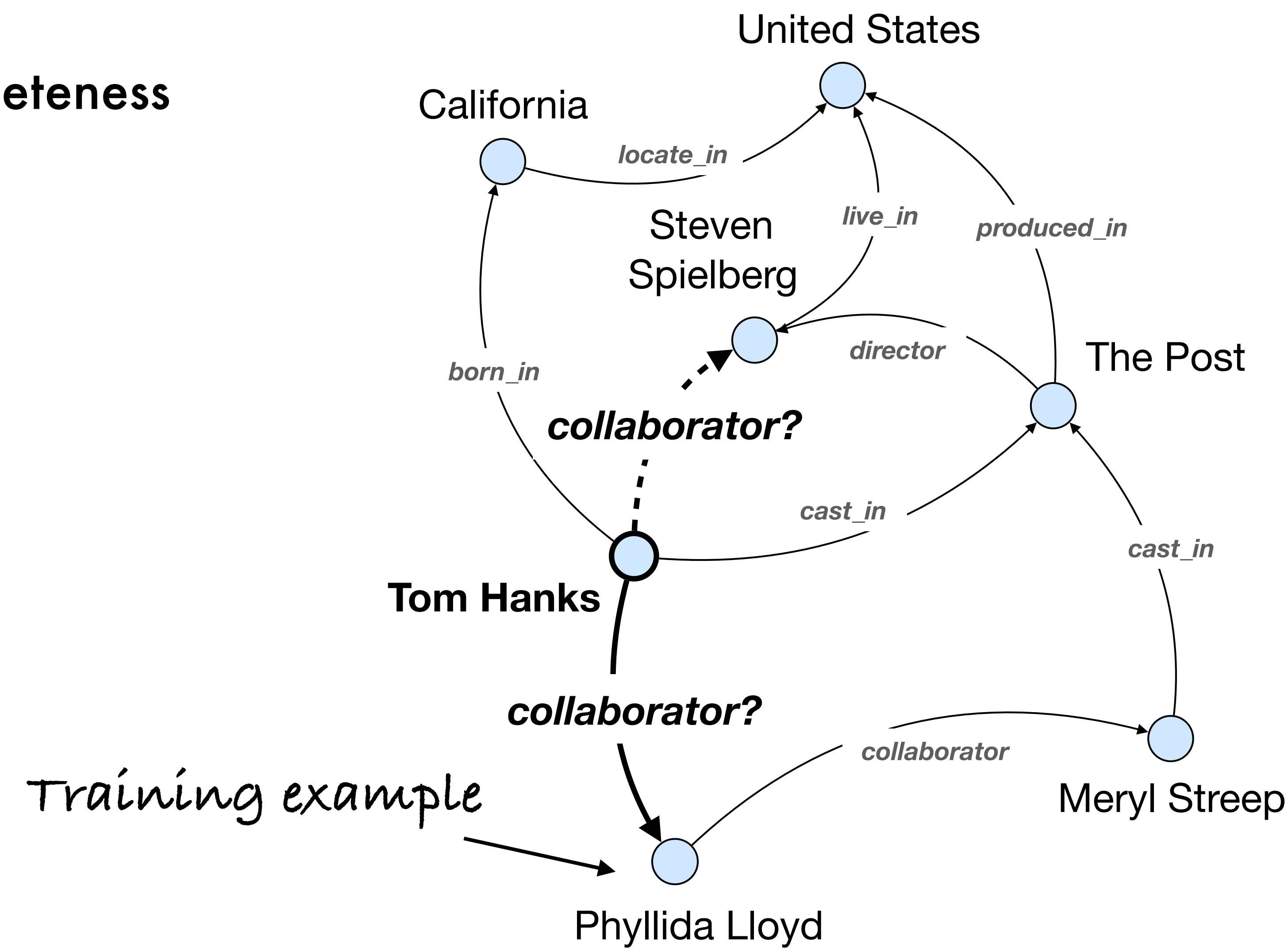
	MRR
ConvE	0.957
RL	0.825

Tab 2. ConvE and RL (MINERVA) query answering performance on the UMLS benchmark dataset (Kok and Domingos 2007)



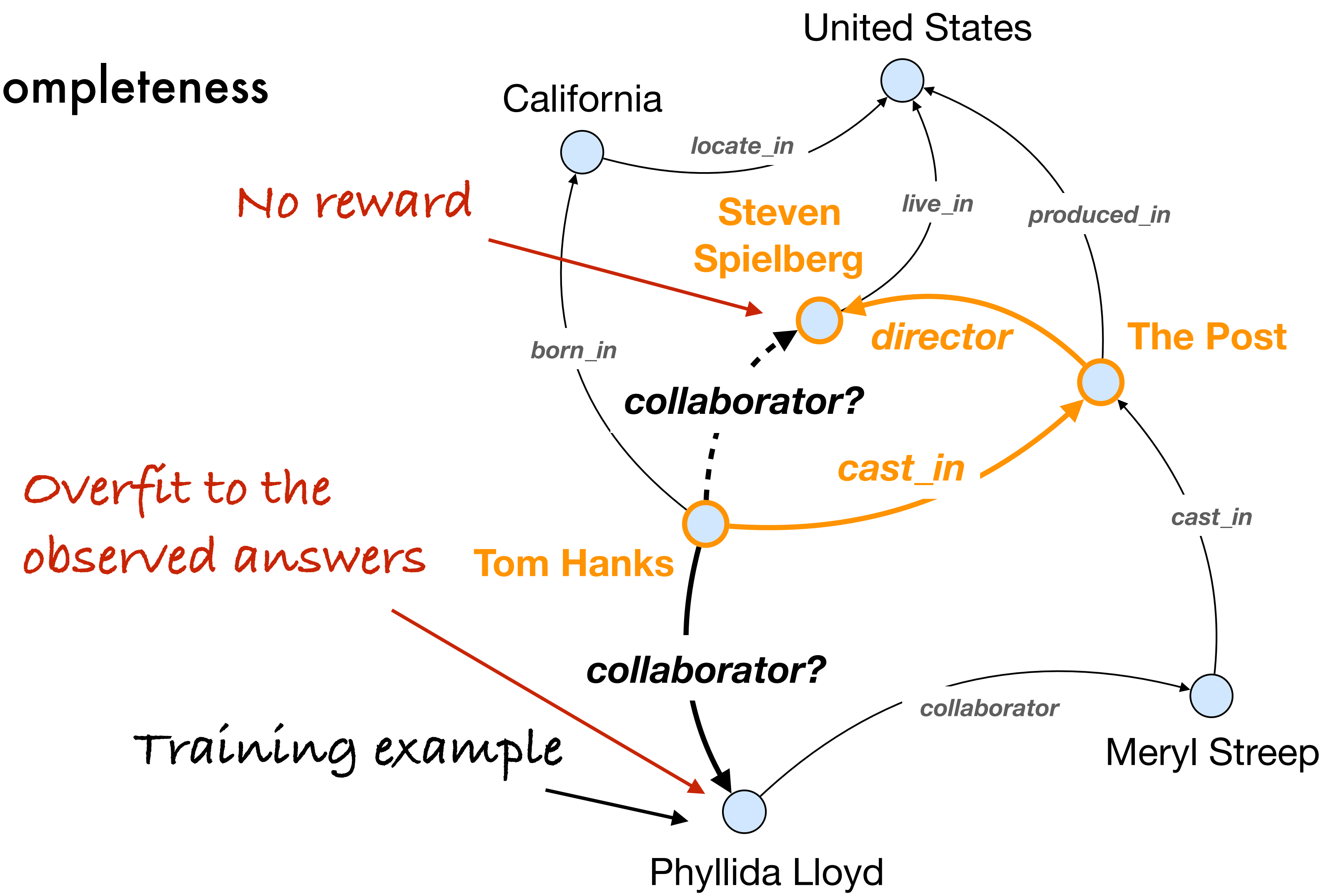
Challenges

Incompleteness



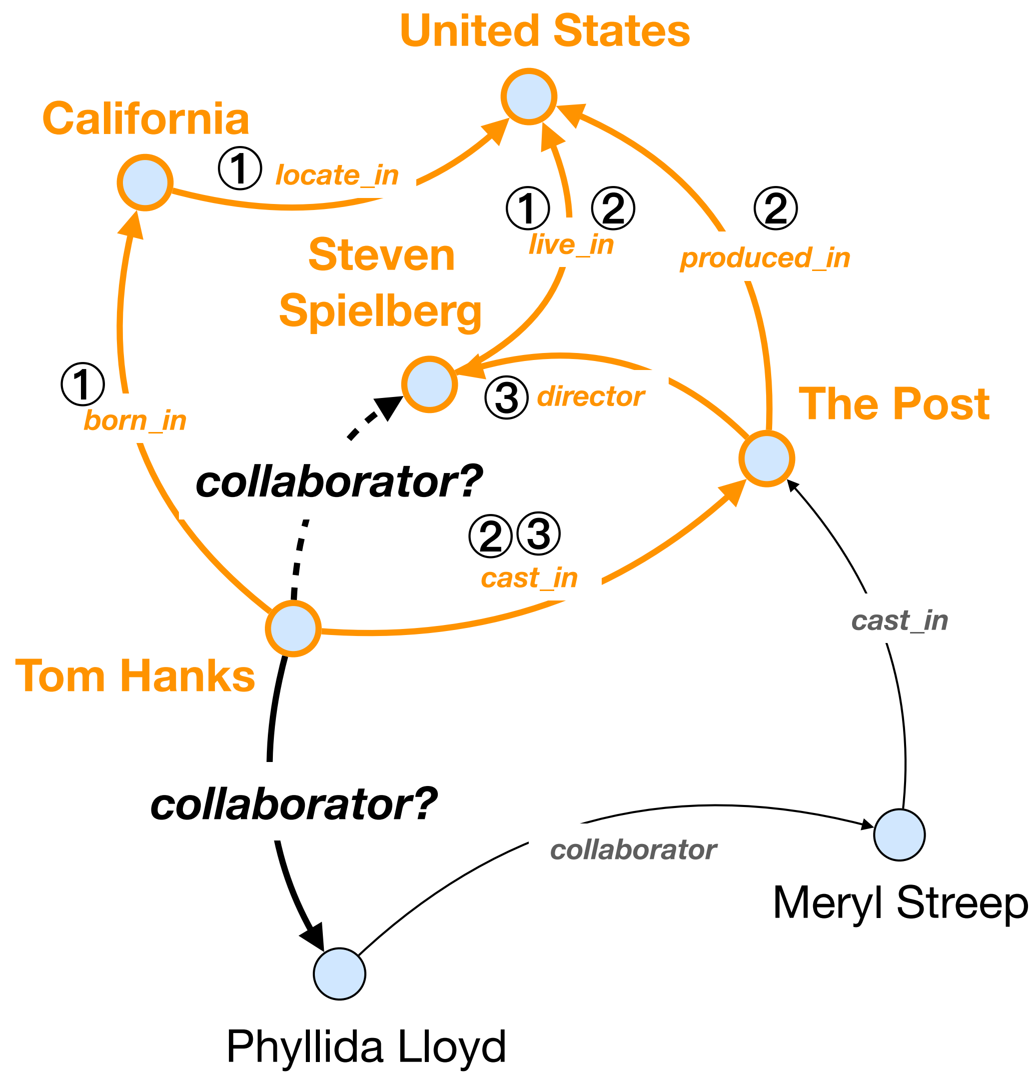
Challenges

Incompleteness



Challenges

Path Diversity

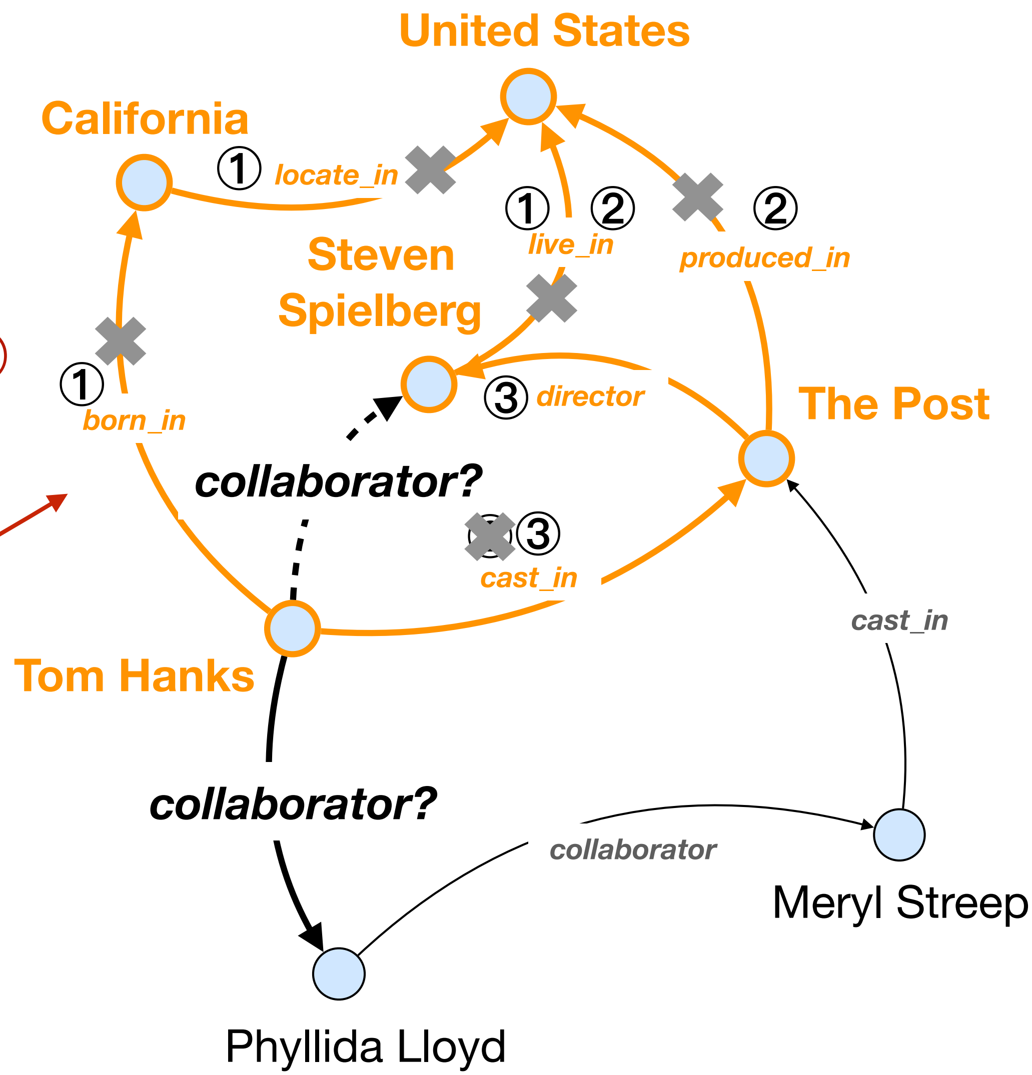


Challenges

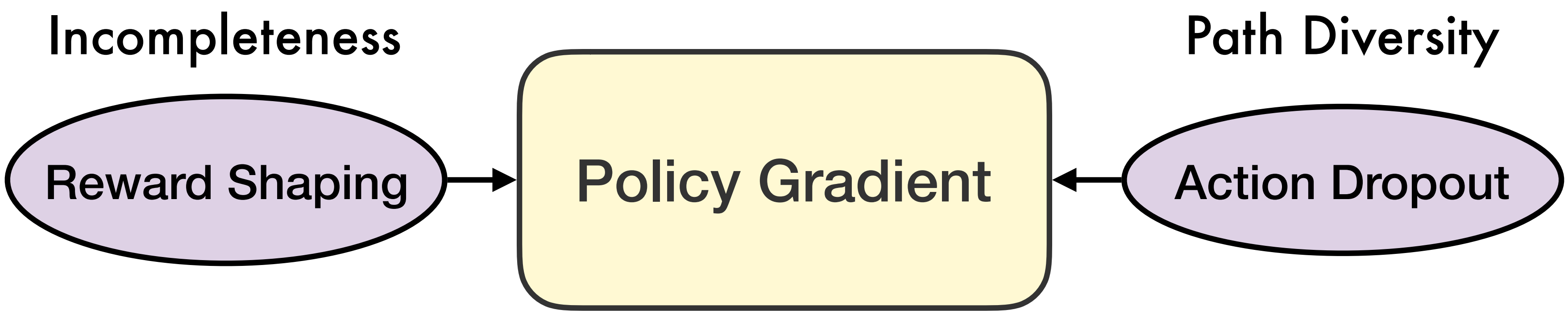
Path Diversity

False positive (spurious) paths ① ②

Overfit to the spurious paths



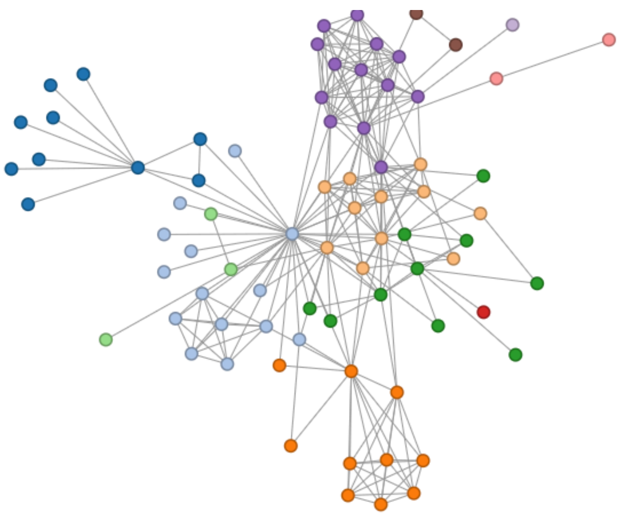
Proposed Solutions



Reinforcement Learning Framework

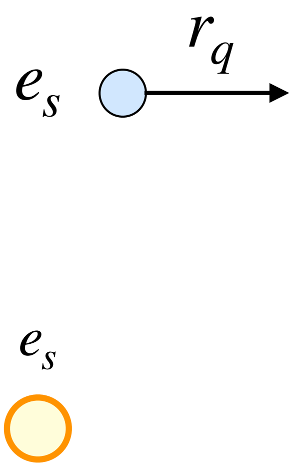
Environment | State | Action | Transition | Reward

$$e_s \xrightarrow{r_q}$$



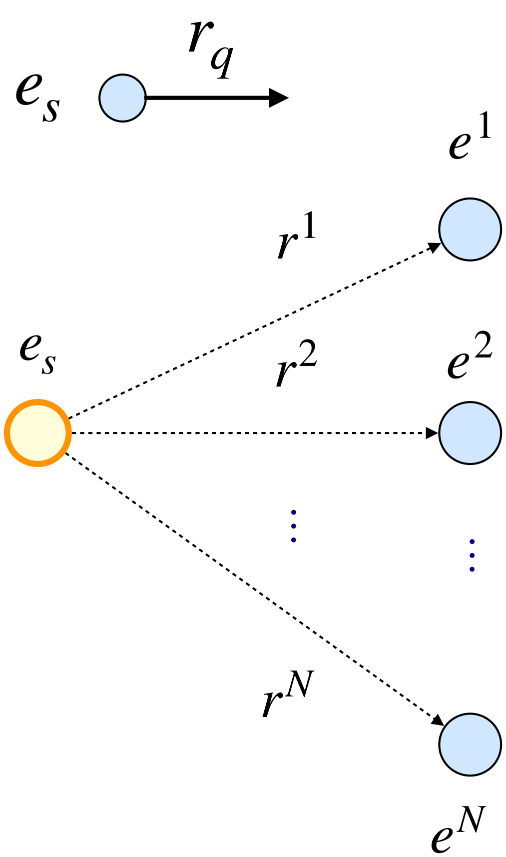
Reinforcement Learning Framework

Environment | **State** | Action | Transition | Reward



Reinforcement Learning Framework

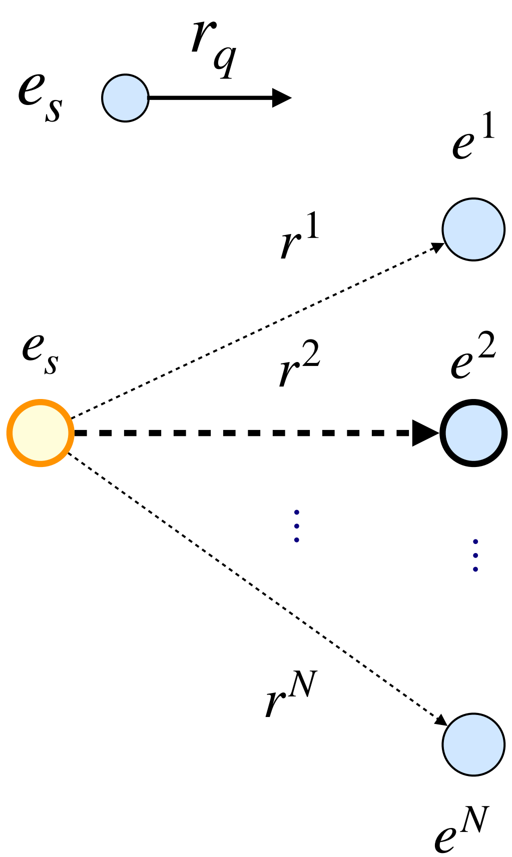
Environment | State | **Action** | Transition | Reward



A_t

Reinforcement Learning Framework

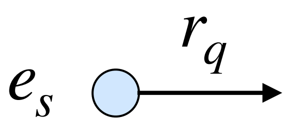
Environment State **Action** Transition Reward

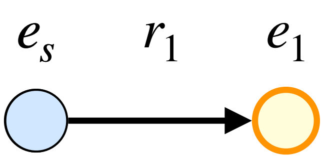


A_t

Reinforcement Learning Framework

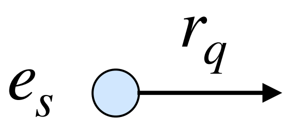
Environment | State | Action | **Transition** | Reward

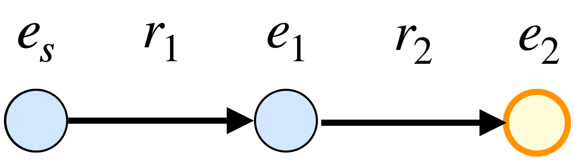




Reinforcement Learning Framework

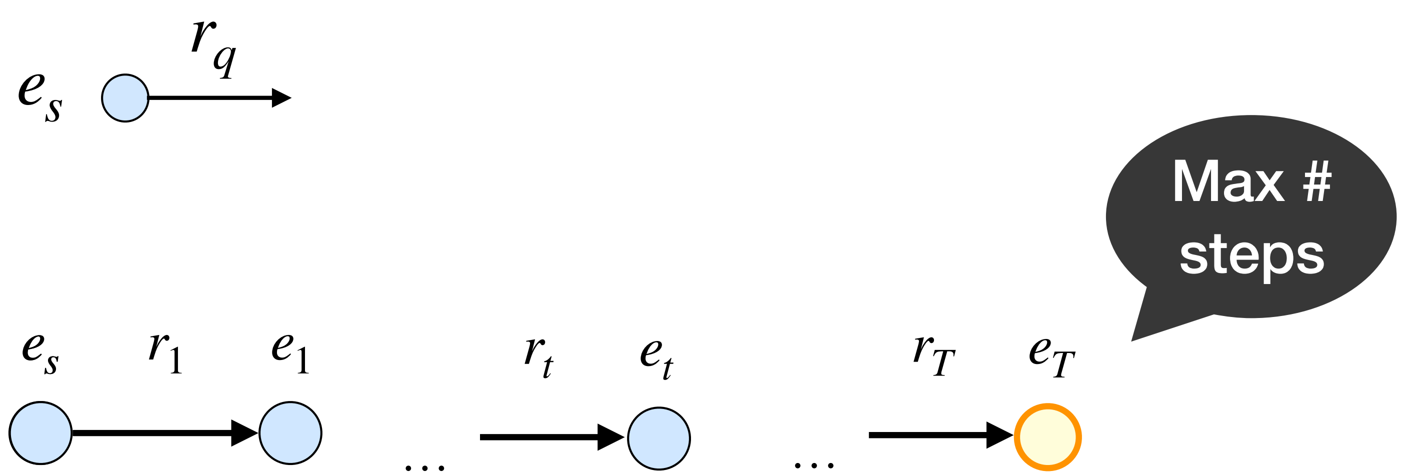
Environment | State | Action | **Transition** | Reward





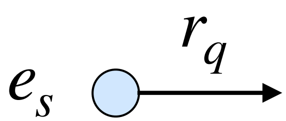
Reinforcement Learning Framework

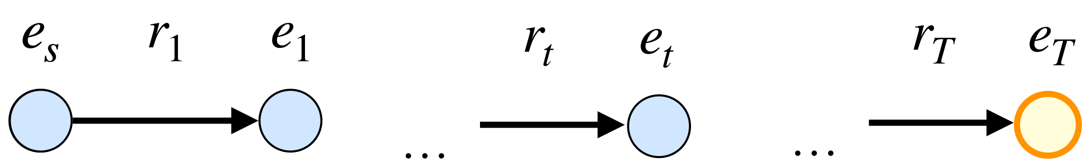
Environment | State | Action | **Transition** | Reward



Reinforcement Learning Framework

Environment | State | Action | Transition | **Reward**

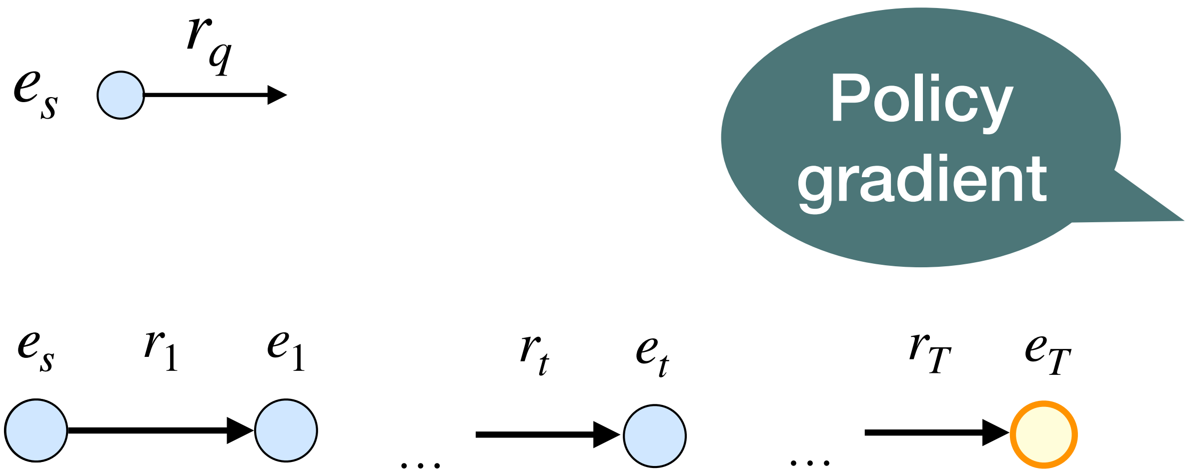




$$\boxed{R_b(s_T) = \mathbf{1}\{(e_s, r_q, e_T) \in G\}}$$

Reinforcement Learning Framework

Environment | State | Action | Transition | **Reward**



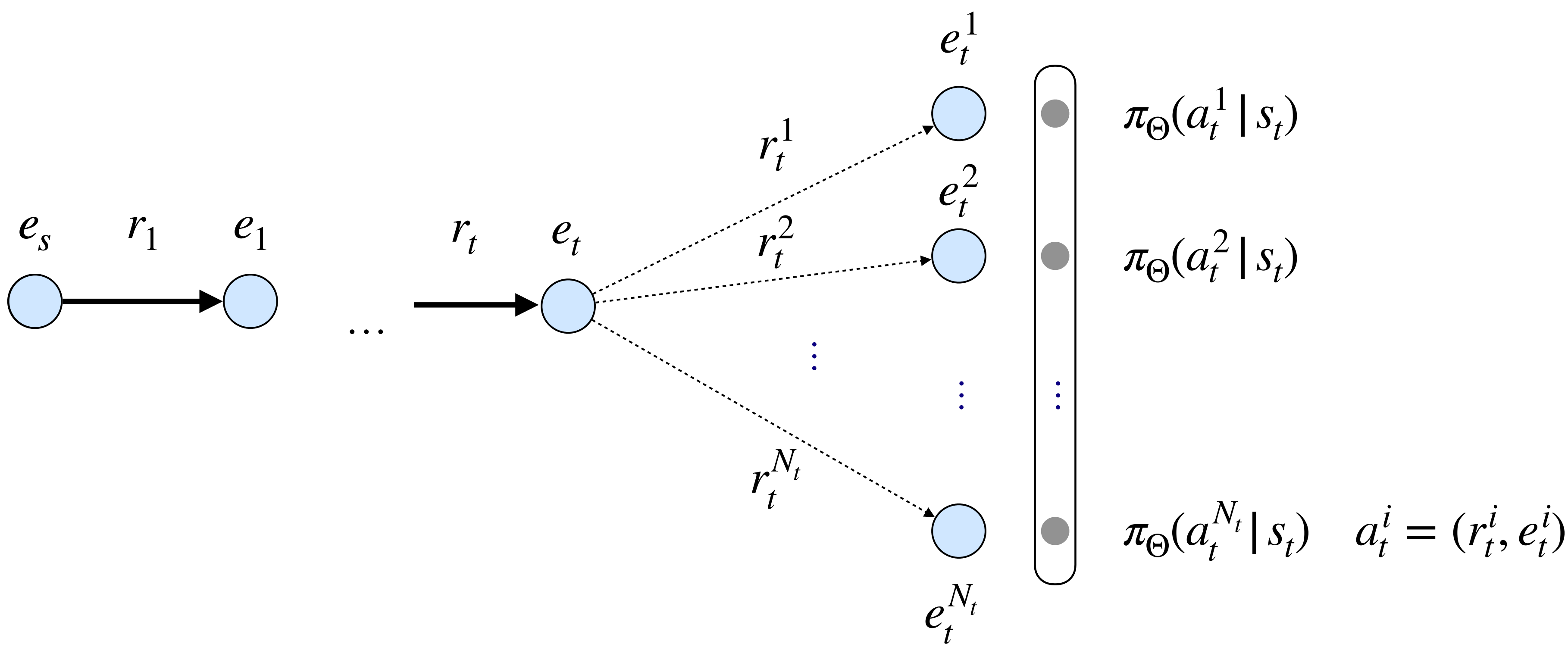
Learn ***which action to choose*** given a state

$$R_b(s_T) = \mathbf{1}\{(e_s, r_q, e_T) \in G\}$$

Policy Gradient

Policy function $\pi_{\Theta}(a_t \mid s_t)$

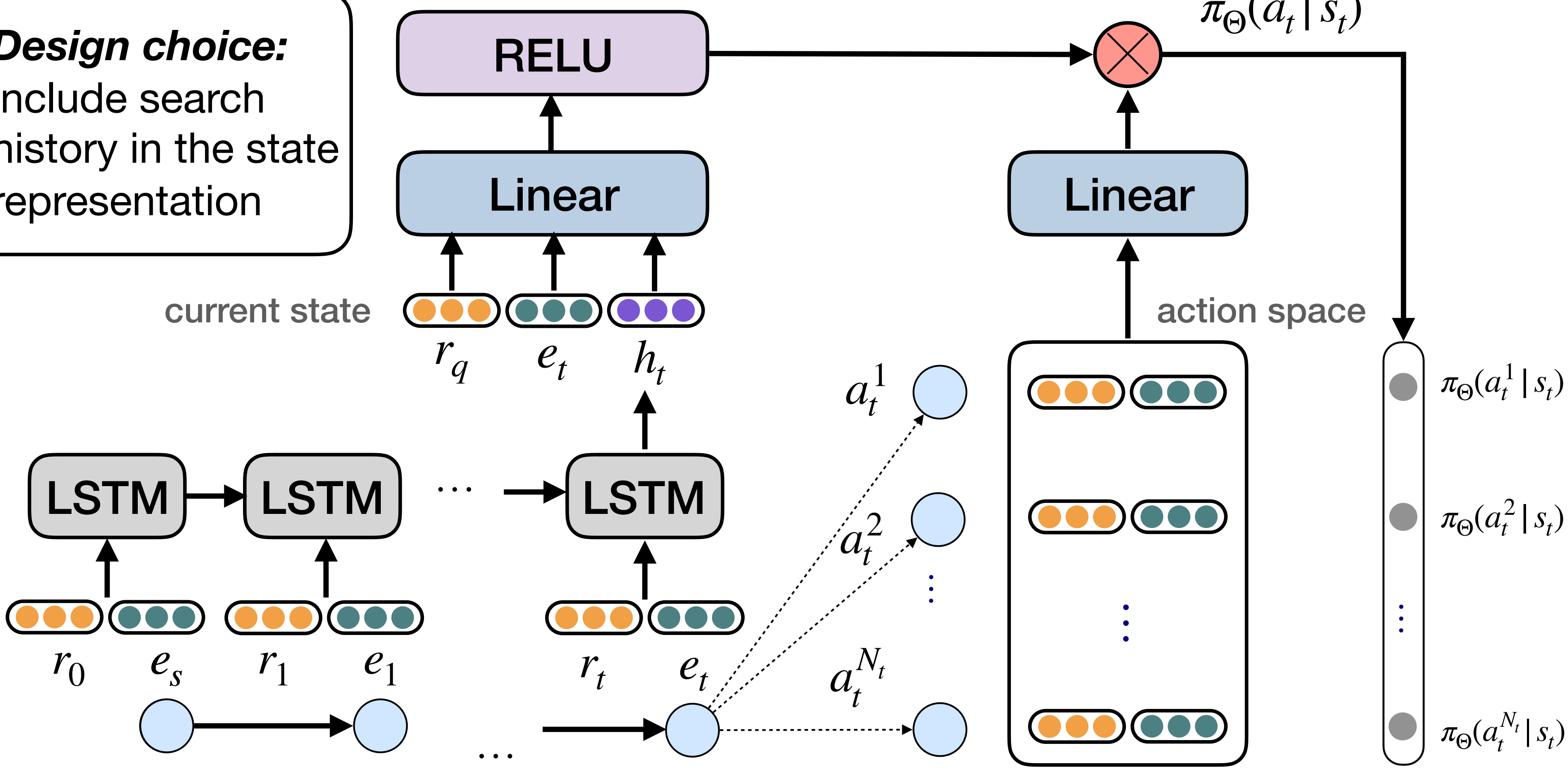
Probability of choosing an action given the current state



Policy Gradient

Policy function $\pi_\Theta(a_t | s_t)$

Design choice: include search history in the state representation



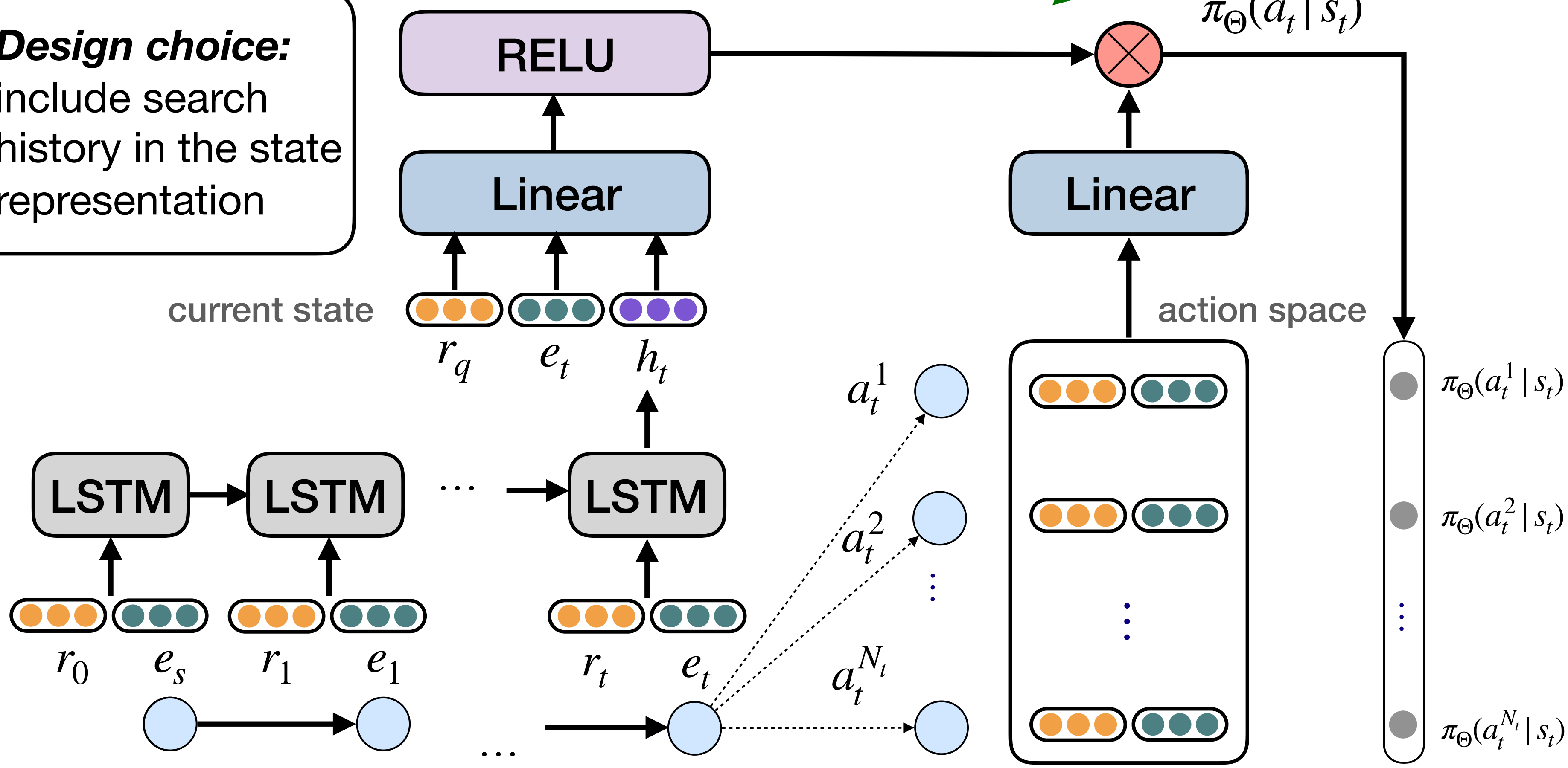
MINERVA (Das et al. 2018)

Policy Gradient

Our technique is applicable to any parameterization of π_Θ

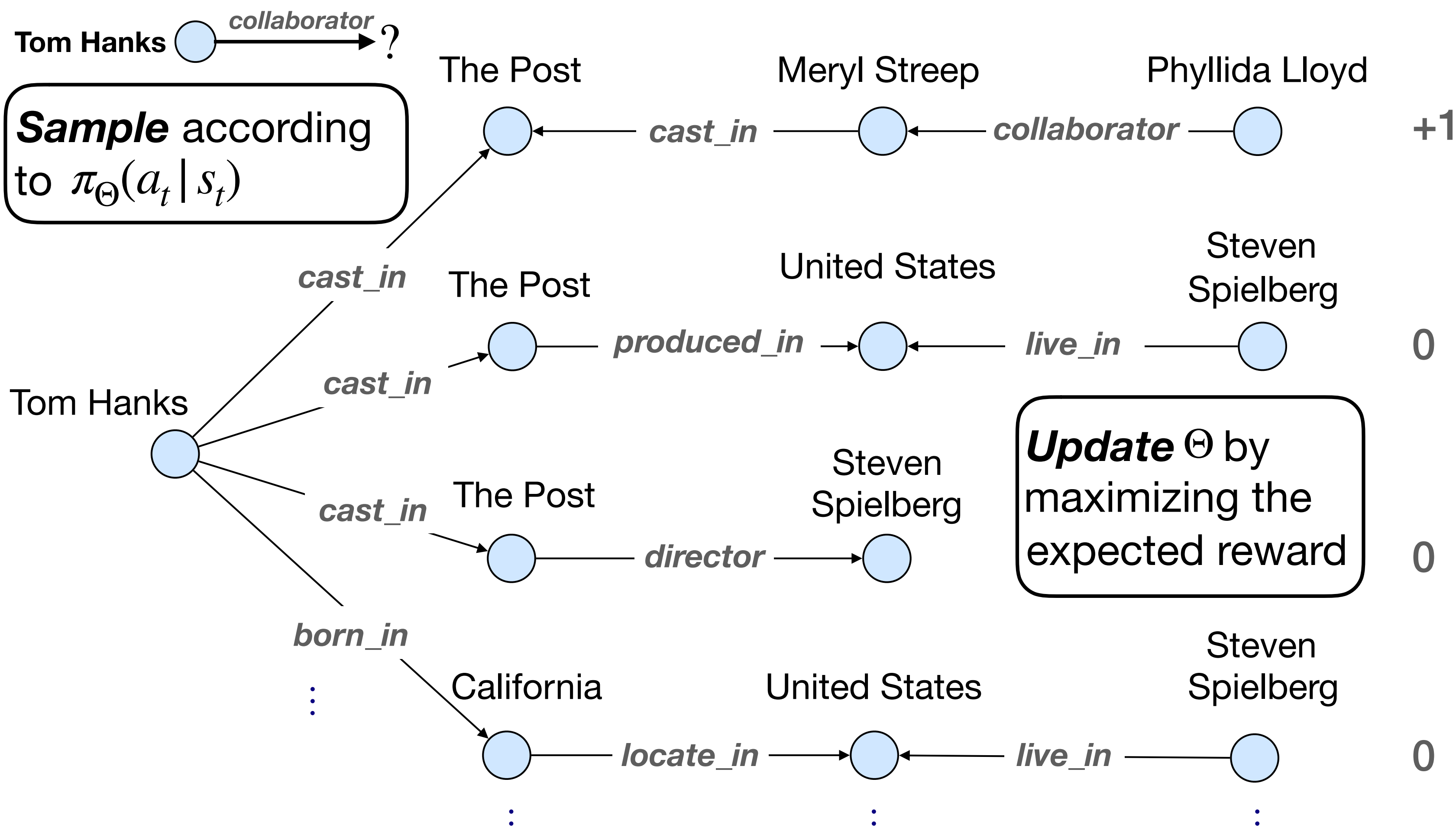
Policy function $\pi_\Theta(a_t | s_t)$

Design choice: include search history in the state representation



MINERVA (Das et al. 2018)

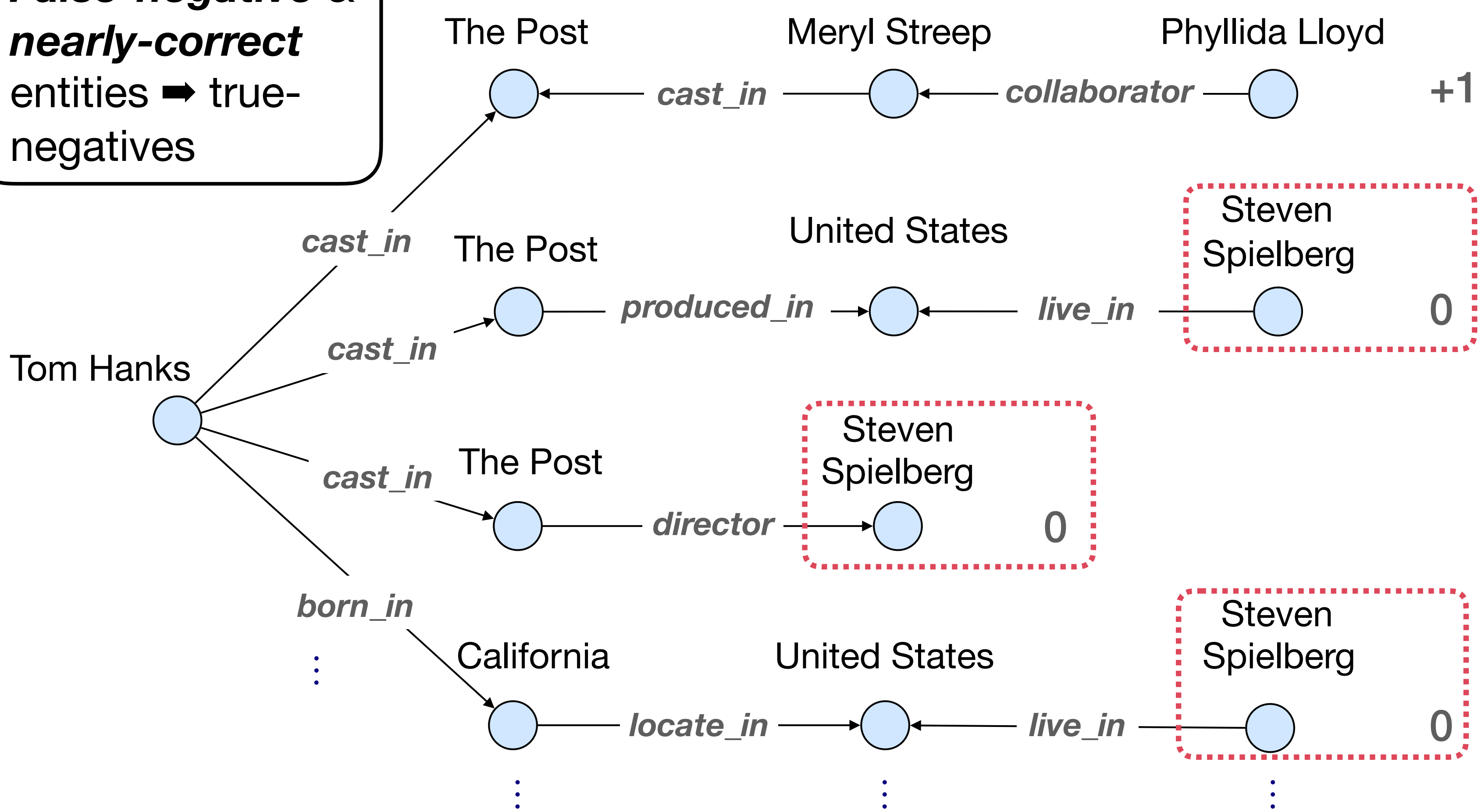
REINFORCE Training



REINFORCE (Williams, 1992)

REINFORCE Training

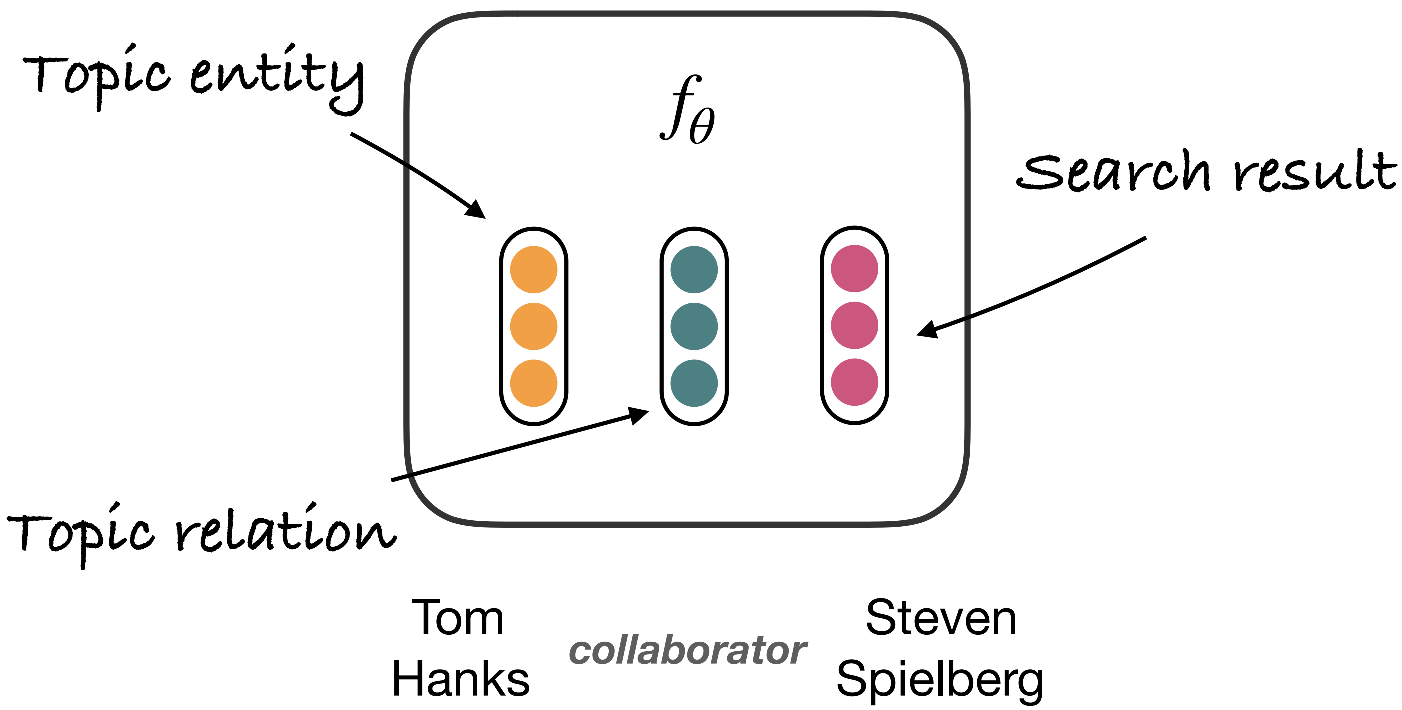
False-negative & ***nearly-correct*** entities ➡ true-negatives



Reward Shaping

Unobserved facts

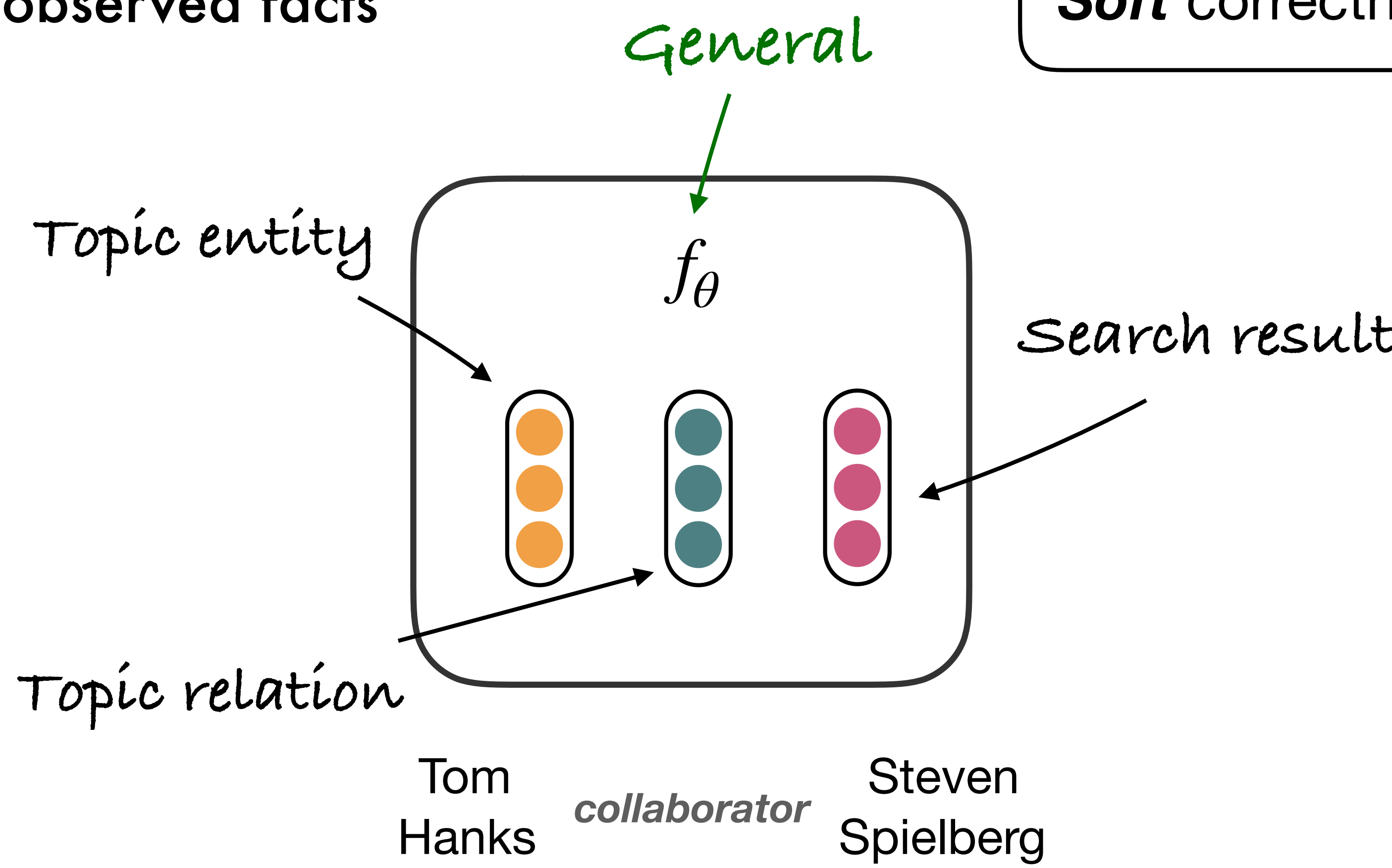
Soft **correctness**



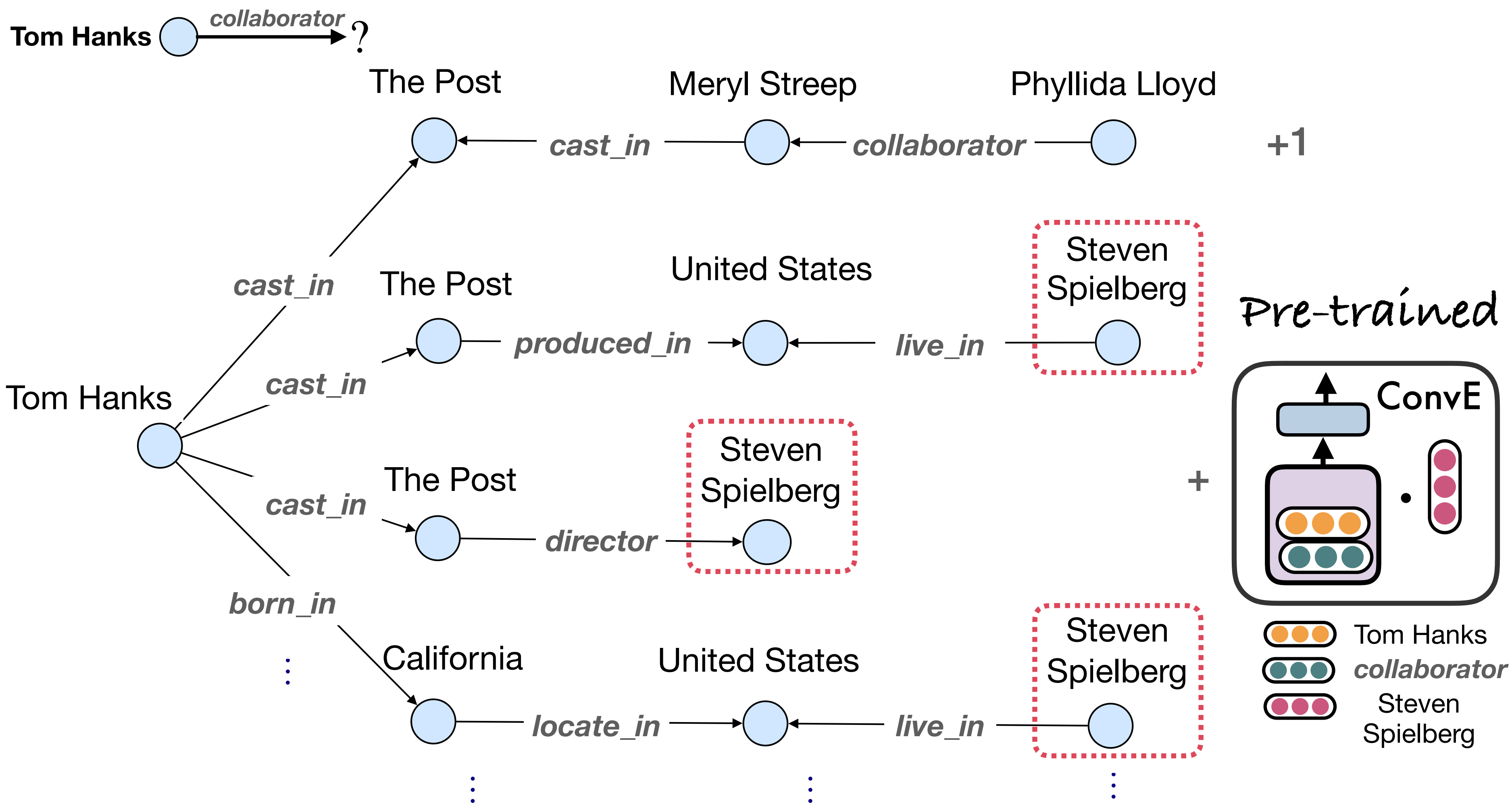
Reward Shaping

Unobserved facts

Soft **correctness**



Reward Shaping

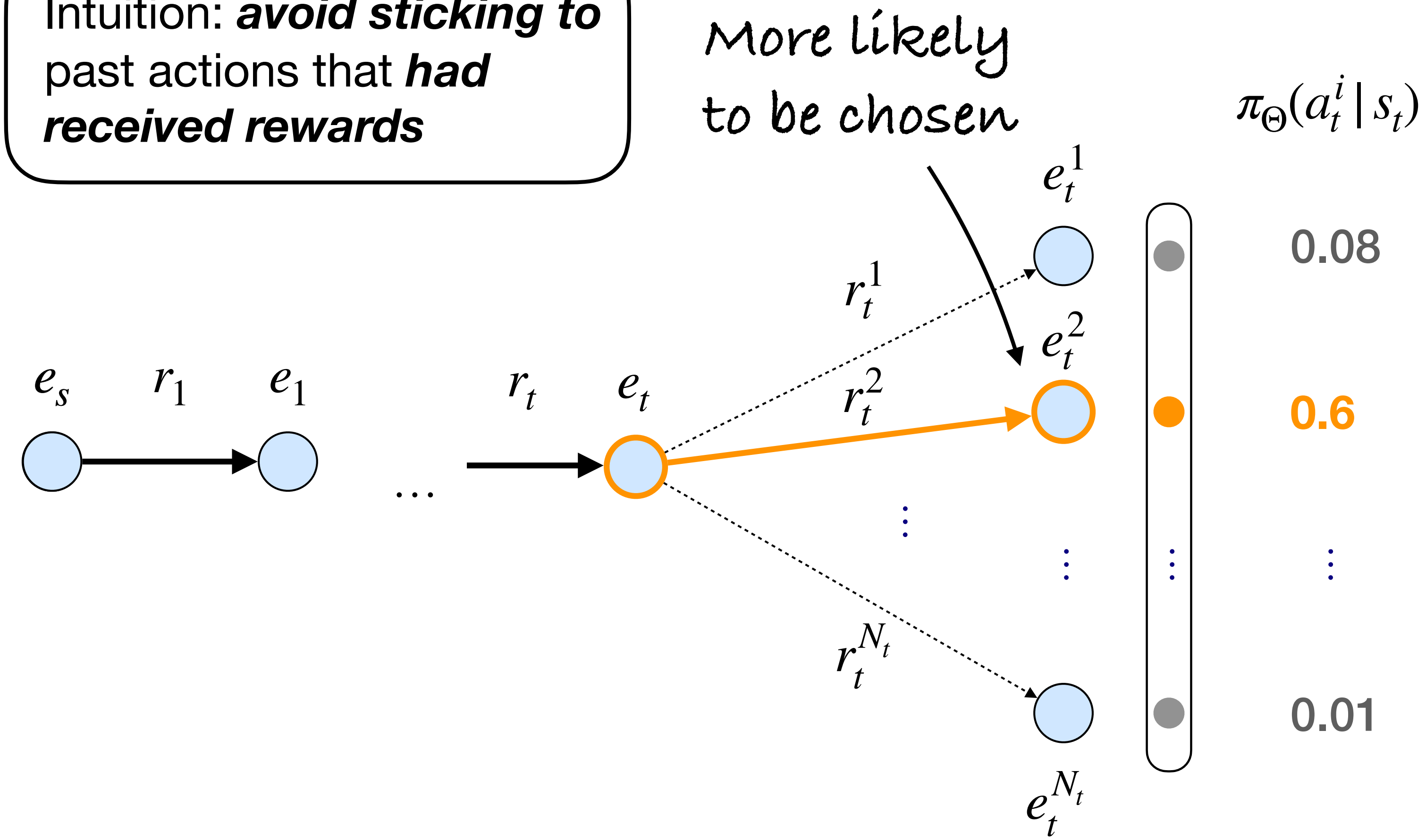


Action Dropout

Intuition: ***avoid sticking to*** past actions that ***had received rewards***

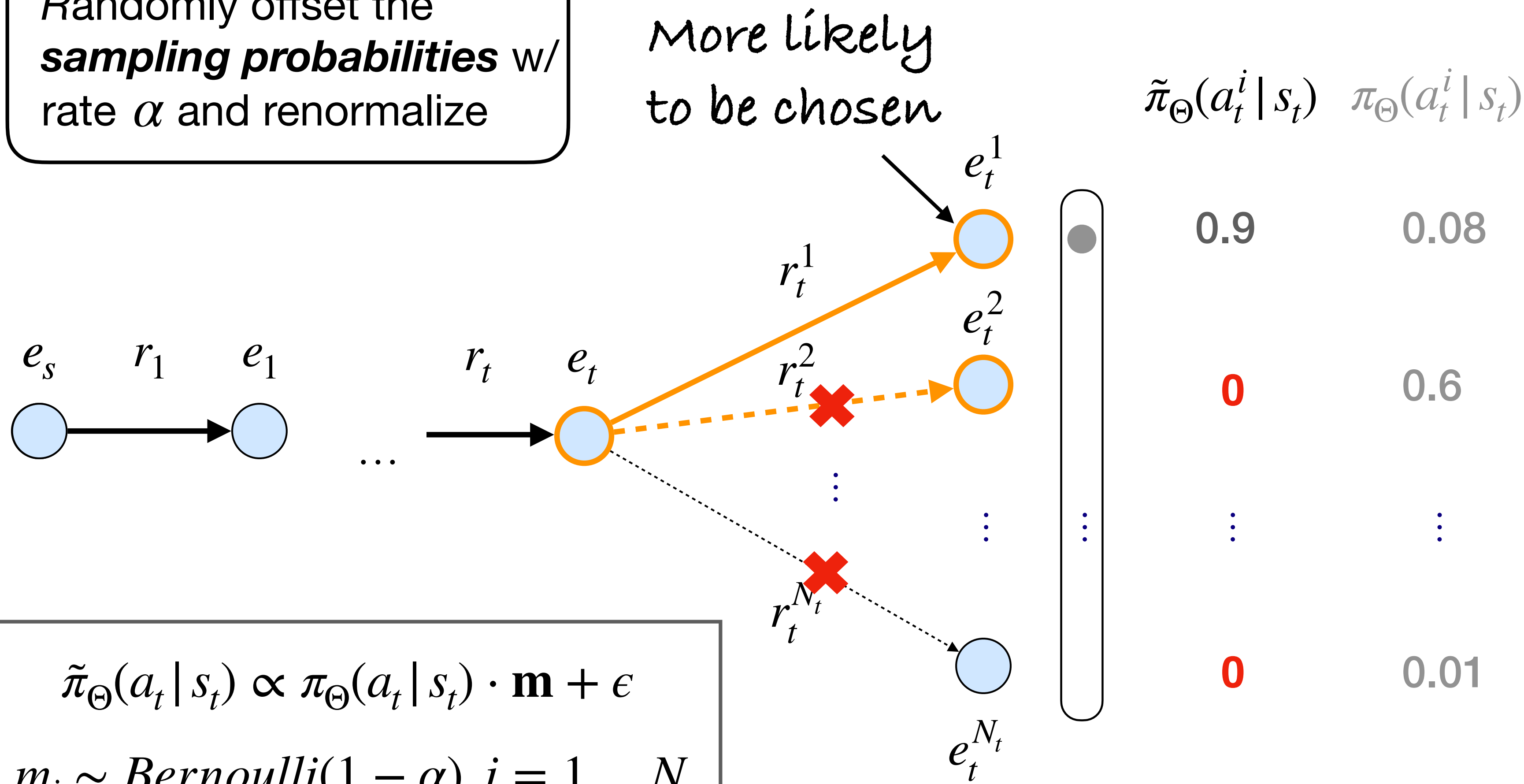
Action Dropout

Intuition: ***avoid sticking to*** past actions that ***had received rewards***



Action Dropout

Randomly offset the ***sampling probabilities*** w/ rate α and renormalize

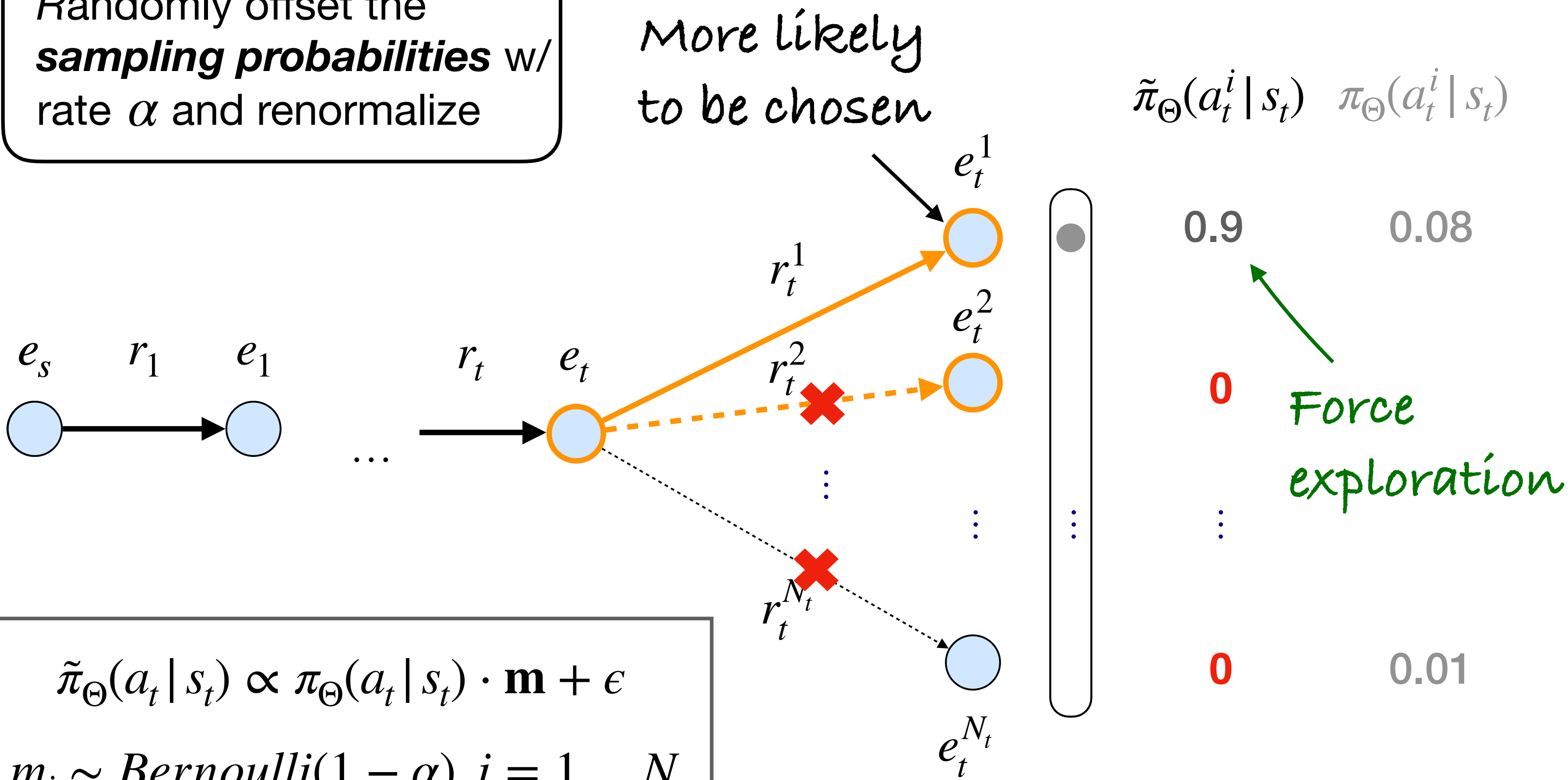


$$\tilde{\pi}_\Theta(a_t \mid s_t) \propto \pi_\Theta(a_t \mid s_t) \cdot \mathbf{m} + \epsilon$$

$$m_i \sim Bernoulli(1-\alpha), i = 1, \ldots N$$

Action Dropout

Randomly offset the ***sampling probabilities*** w/ rate α and renormalize

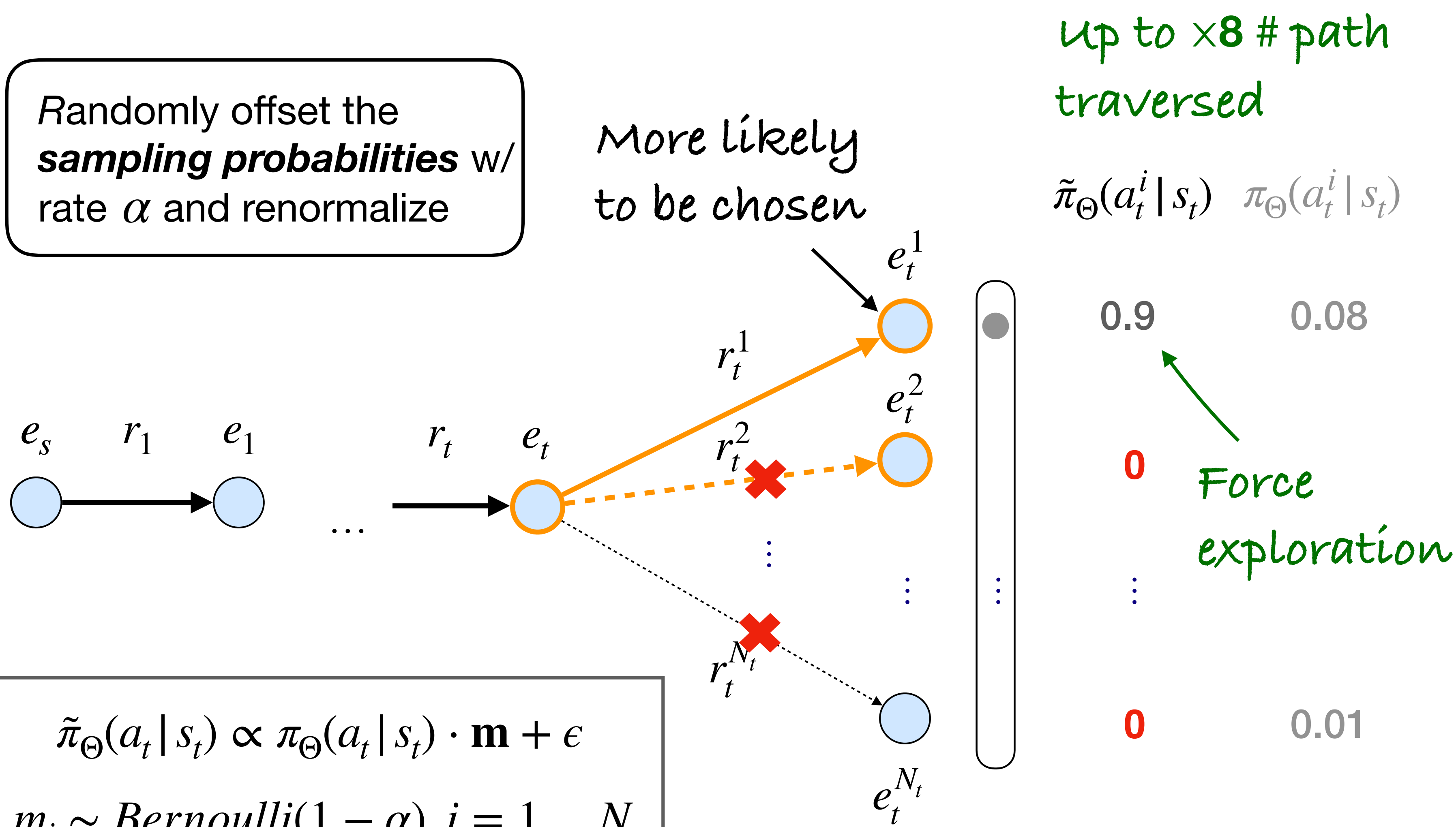


$$\tilde{\pi}_\Theta(a_t \mid s_t) \propto \pi_\Theta(a_t \mid s_t) \cdot \mathbf{m} + \epsilon$$

$$m_i \sim Bernoulli(1-\alpha), i = 1, \ldots N$$

Action Dropout

Randomly offset the ***sampling probabilities*** w/ rate α and renormalize



$$\tilde{\pi}_\Theta(a_t \mid s_t) \propto \pi_\Theta(a_t \mid s_t) \cdot \mathbf{m} + \epsilon$$

$$m_i \sim Bernoulli(1-\alpha), i = 1, \dots N$$

Experiment Setup

KG Benchmarks

Name	# Ent.	# Rel.	# Fact	# Degree Avg	# Degree Median
Kinship	104	25	8,544	85.15	82
UMLS	135	46	5,216	38.63	28
FB15k-237	14,505	237	272,115	19.74	14
WN18RR	40,945	11	86,835	2.19	2
NELL-995	75,492	200	154,213	4.07	1

Decreasing connectivity

Evaluation Protocol: MRR (Mean Reciprocal Rank)

Ablation Studies

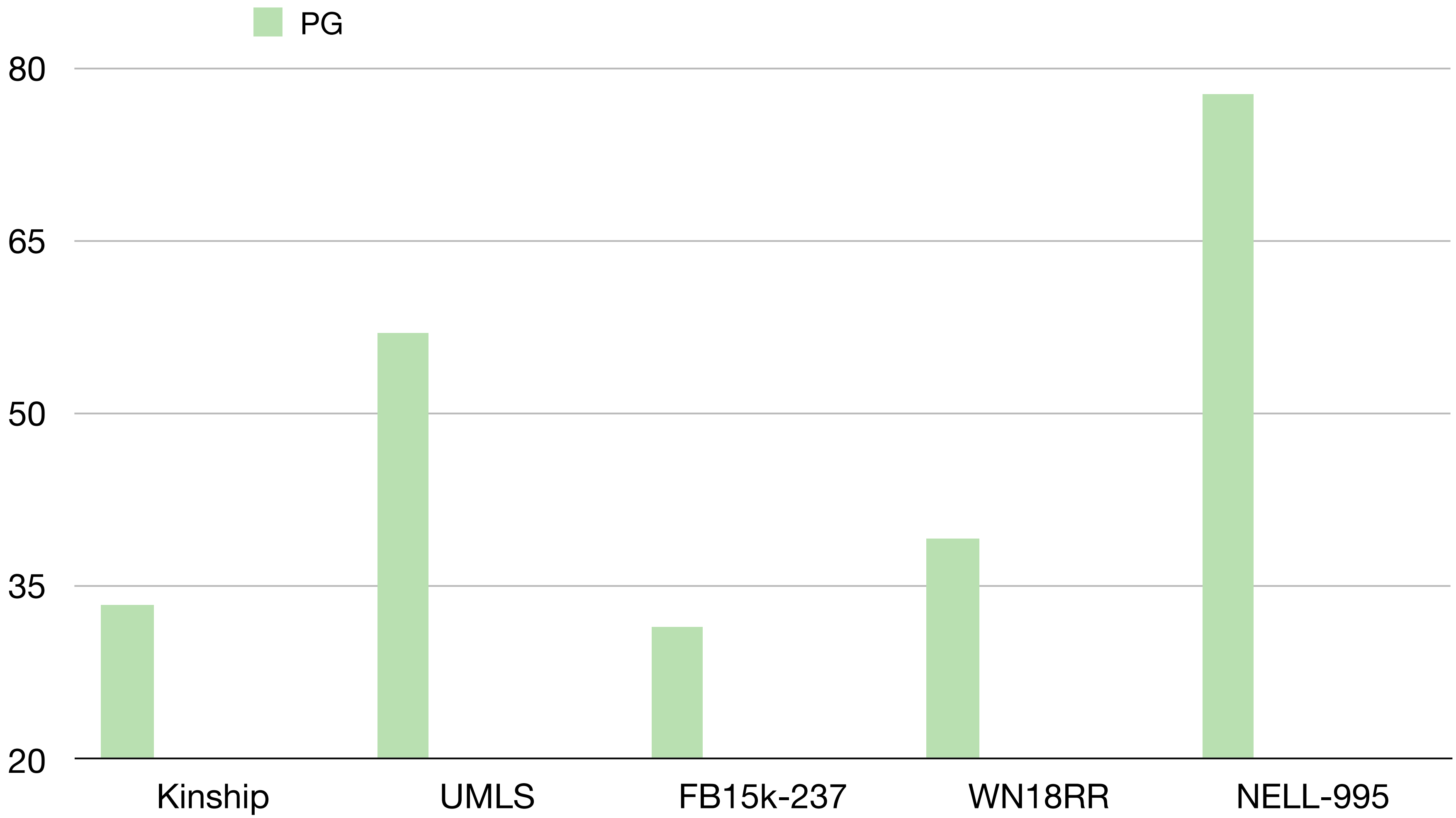


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

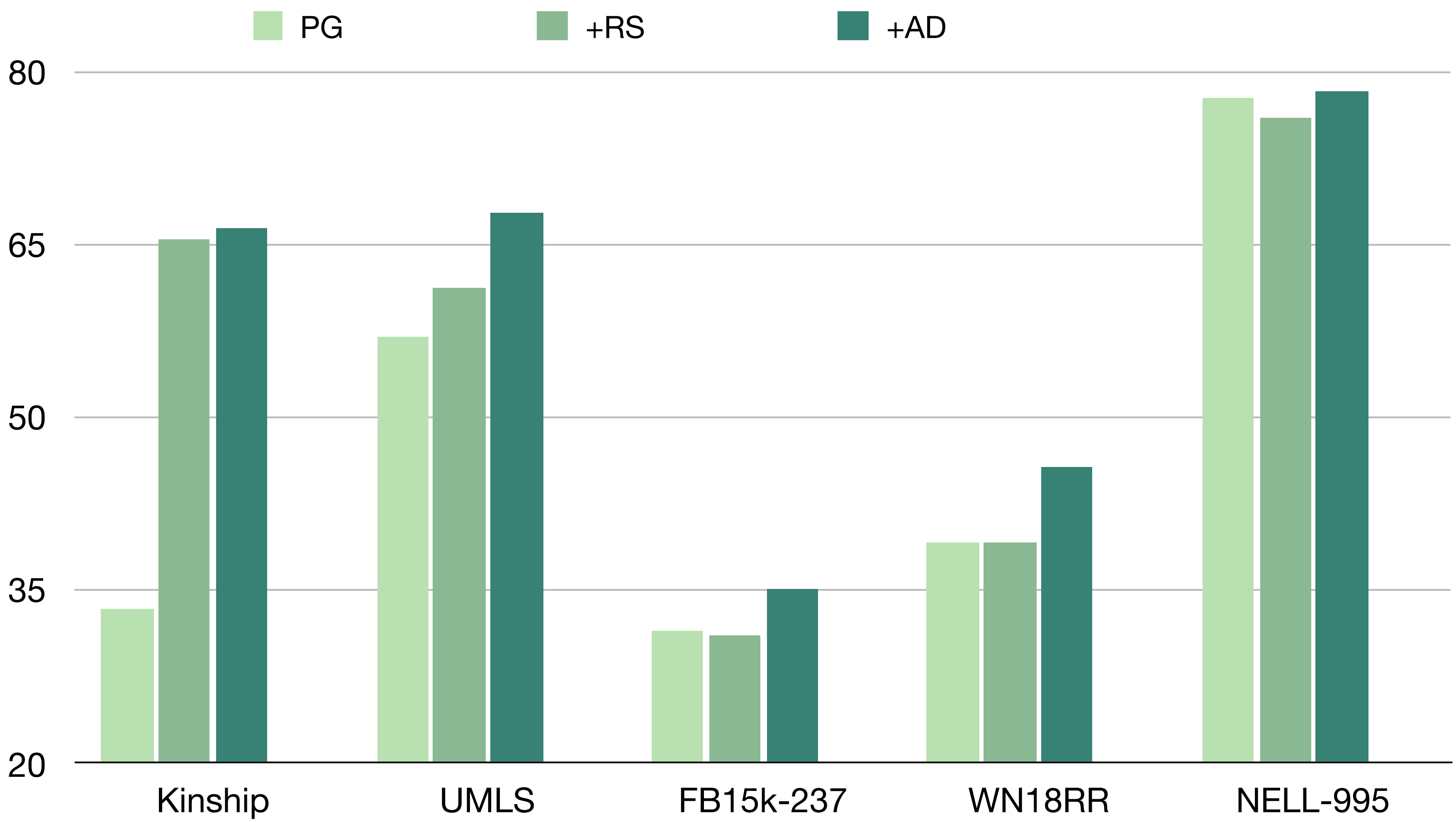


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

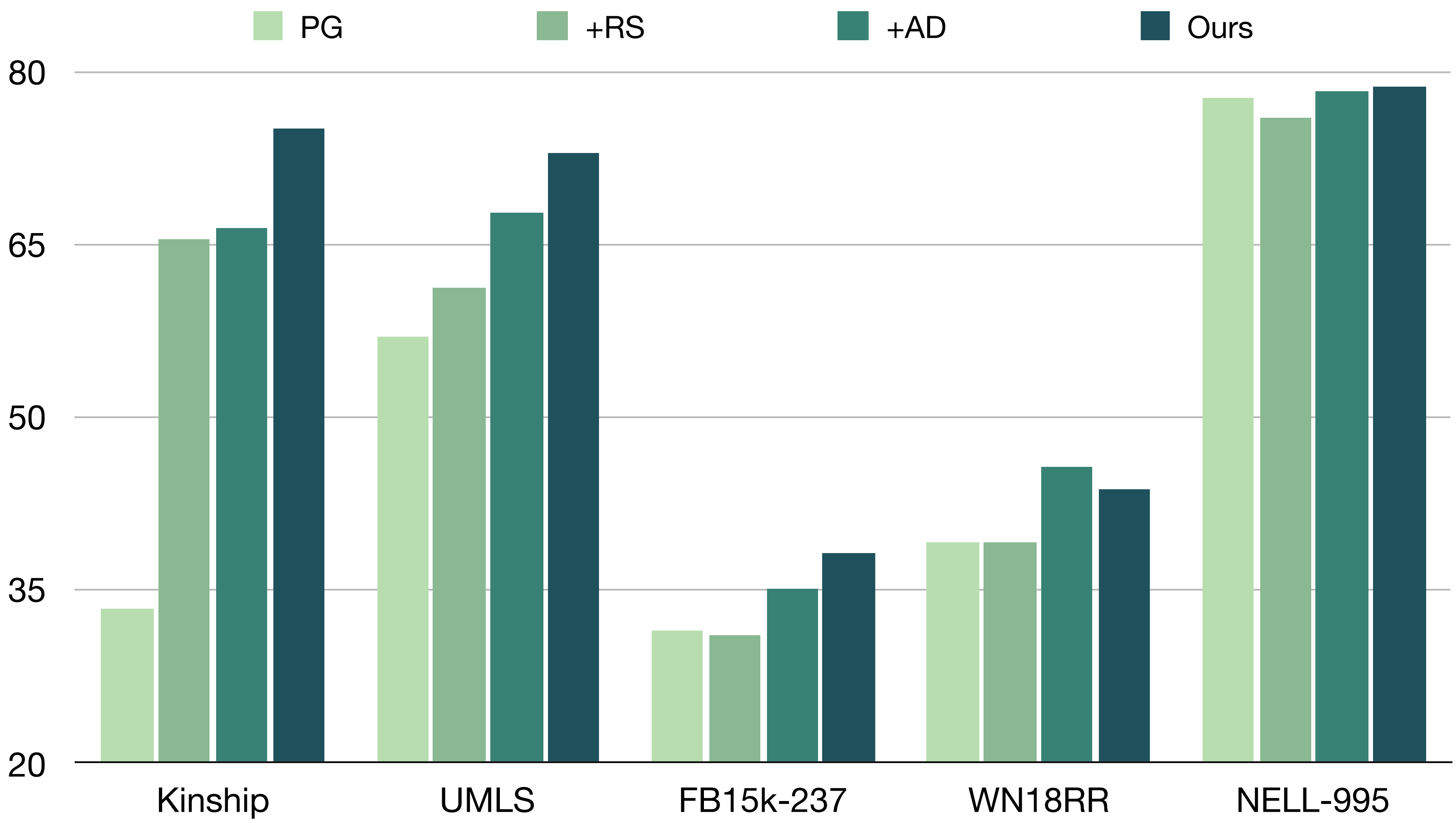


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

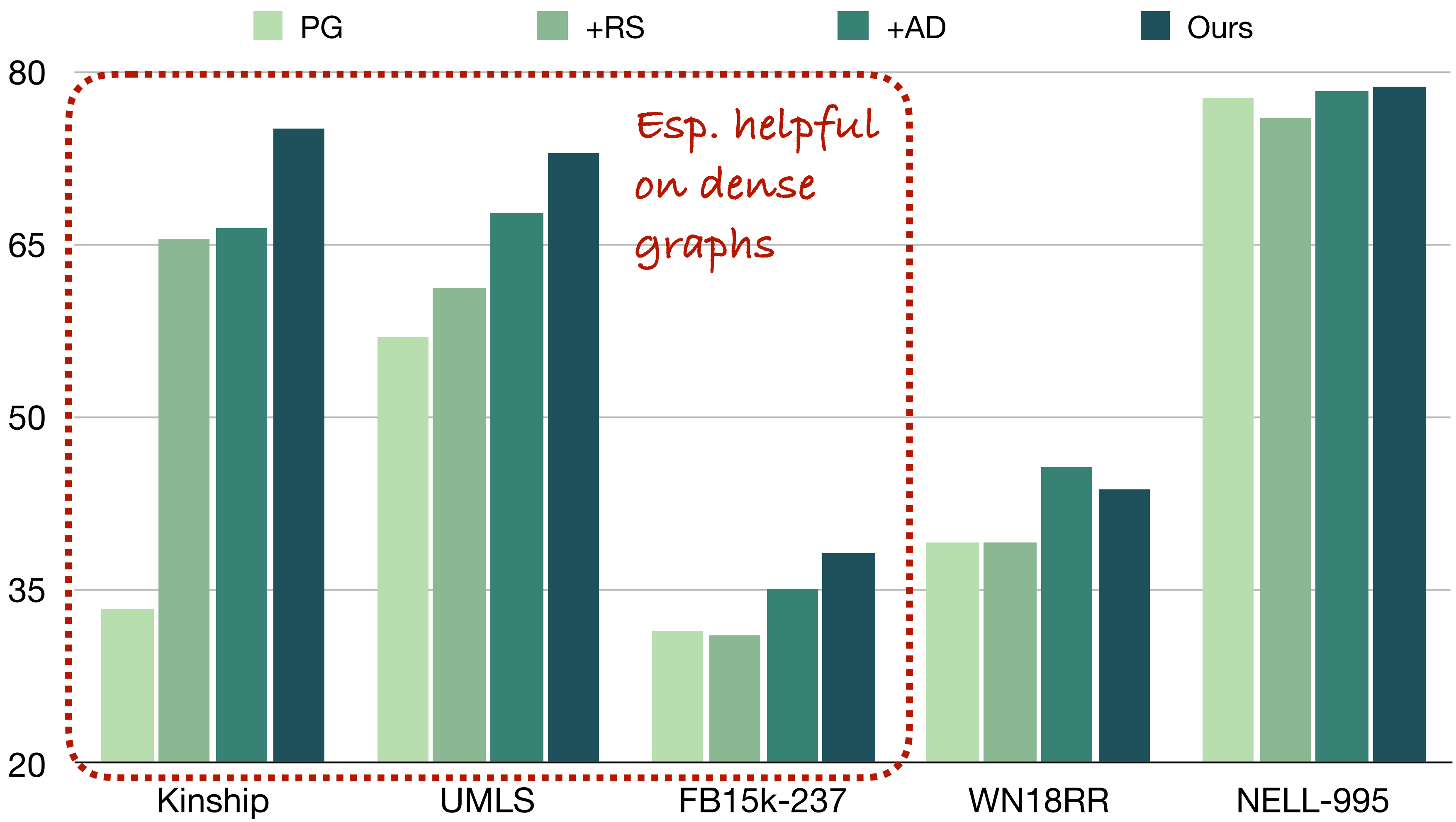


Fig 2. Dev set MRR (x100) comparison

Main Results

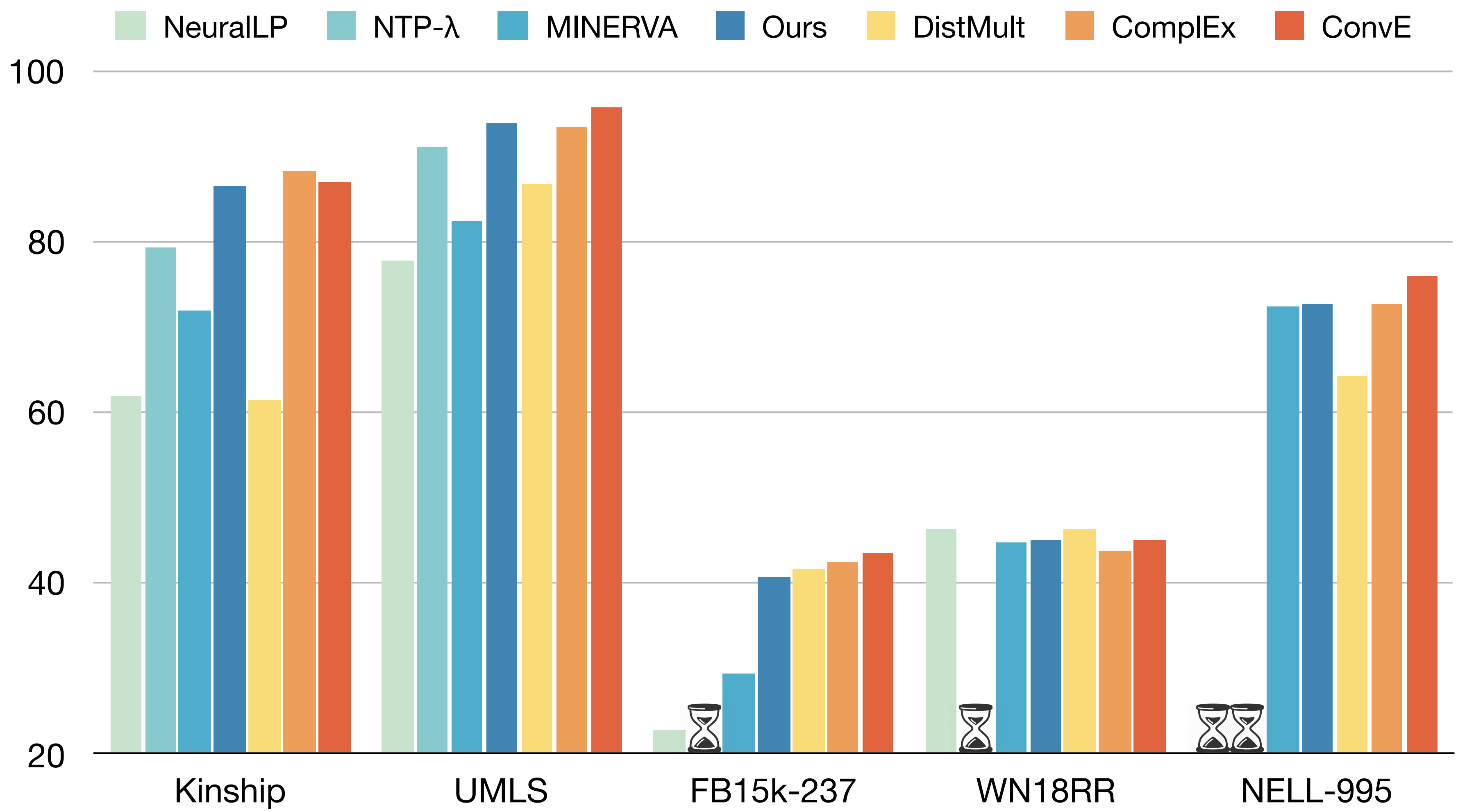


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

Main Results

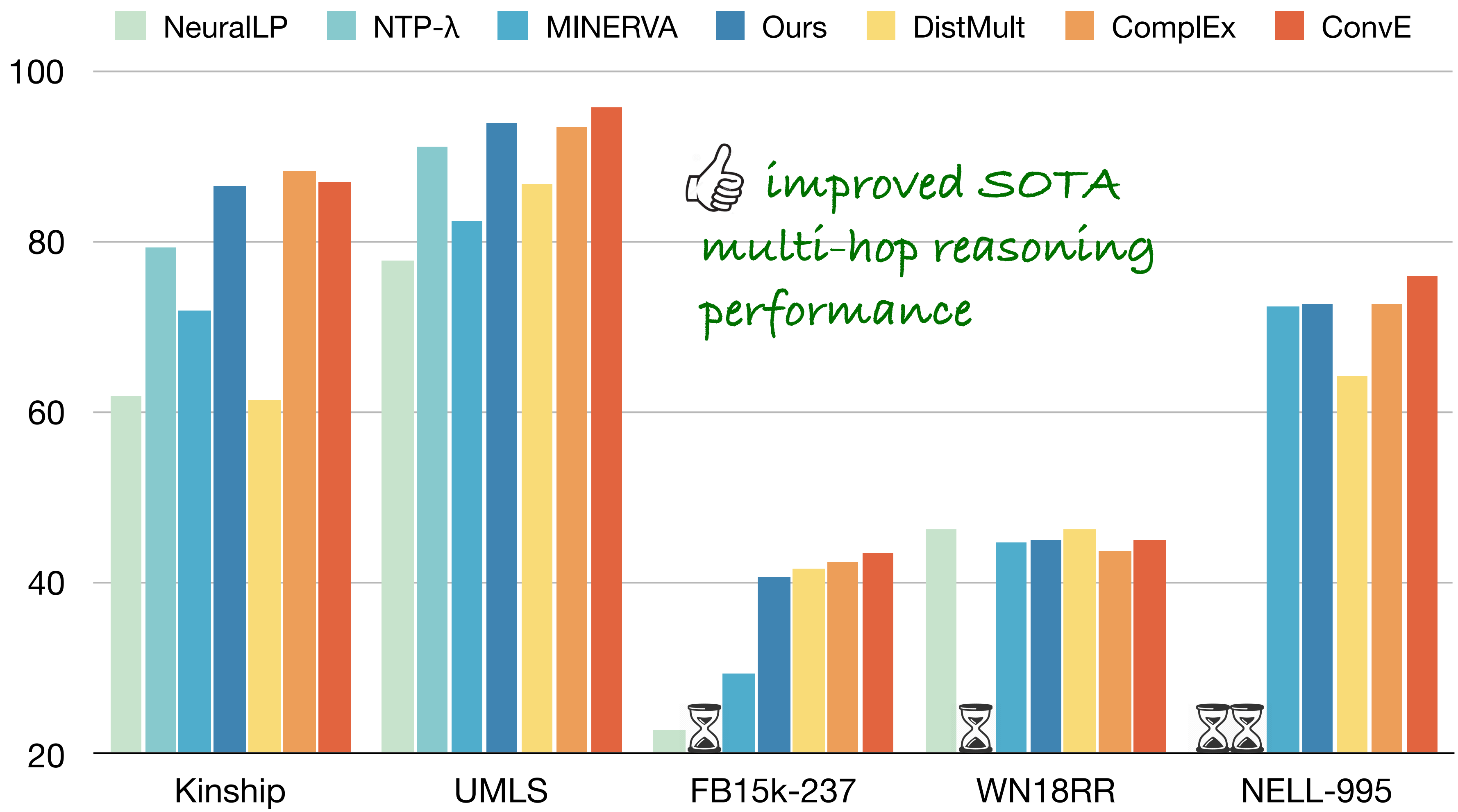


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

Main Results

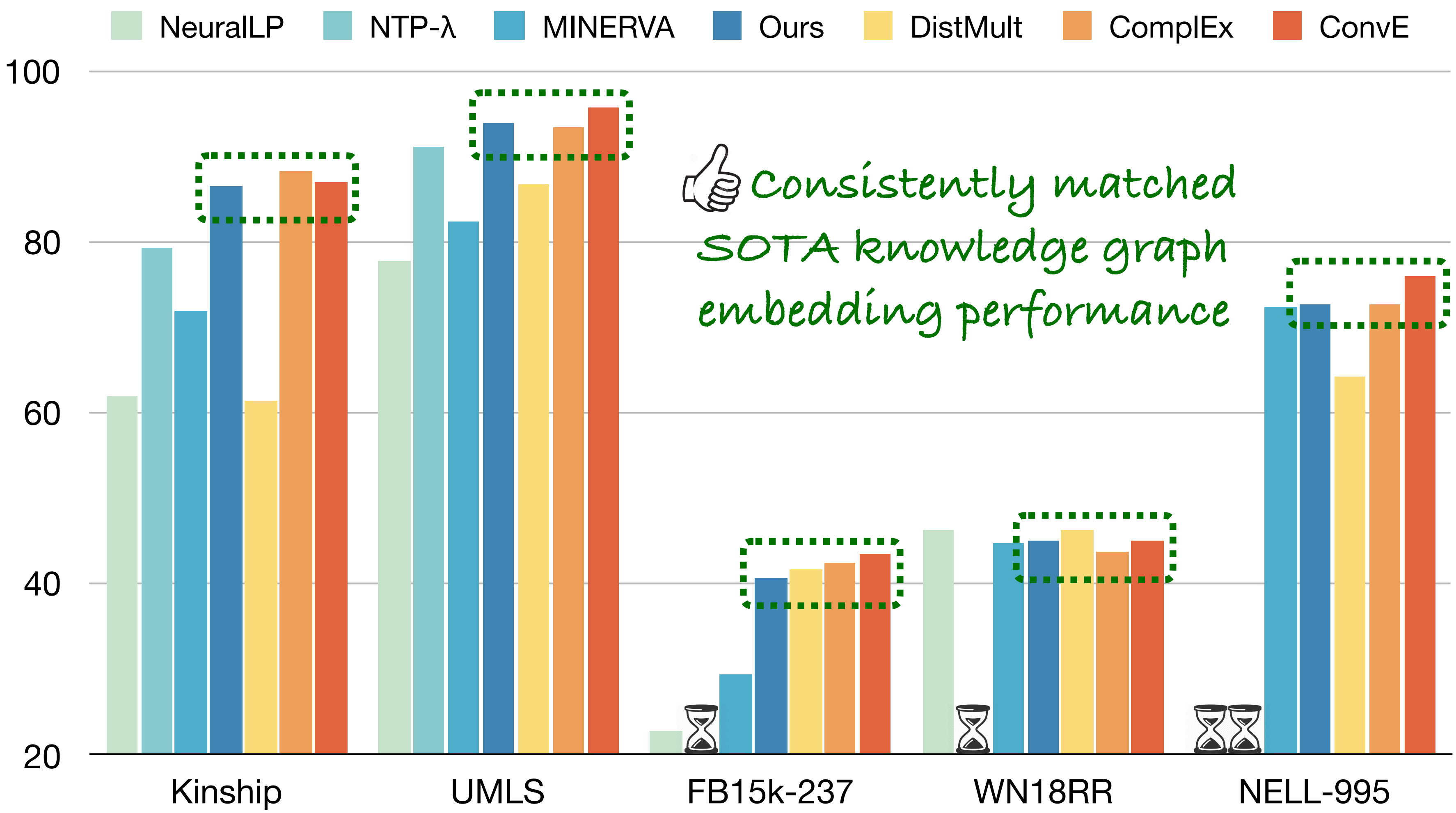
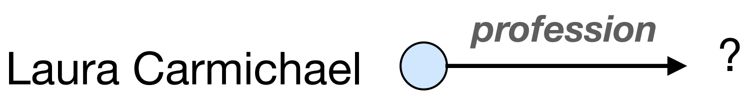
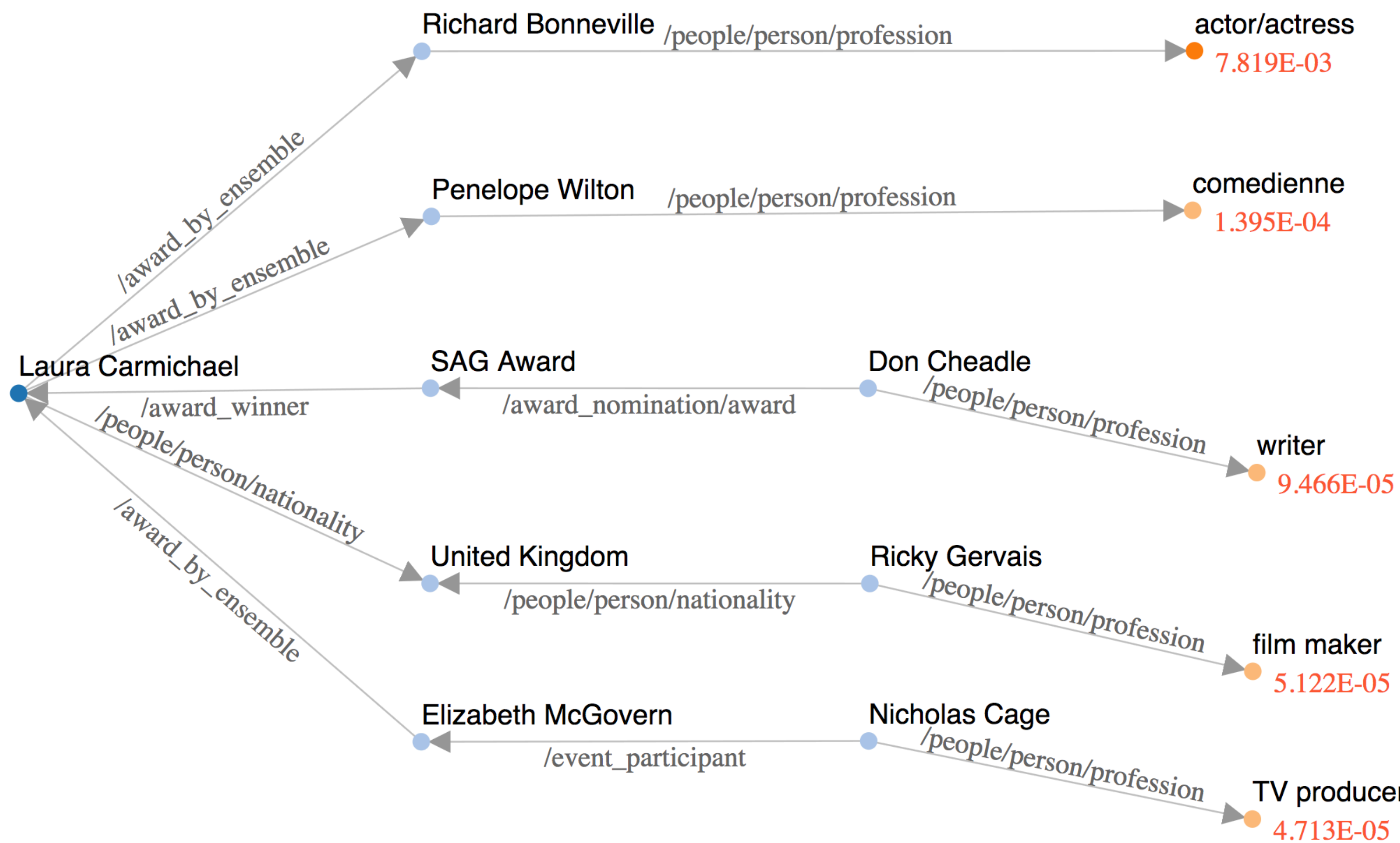


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

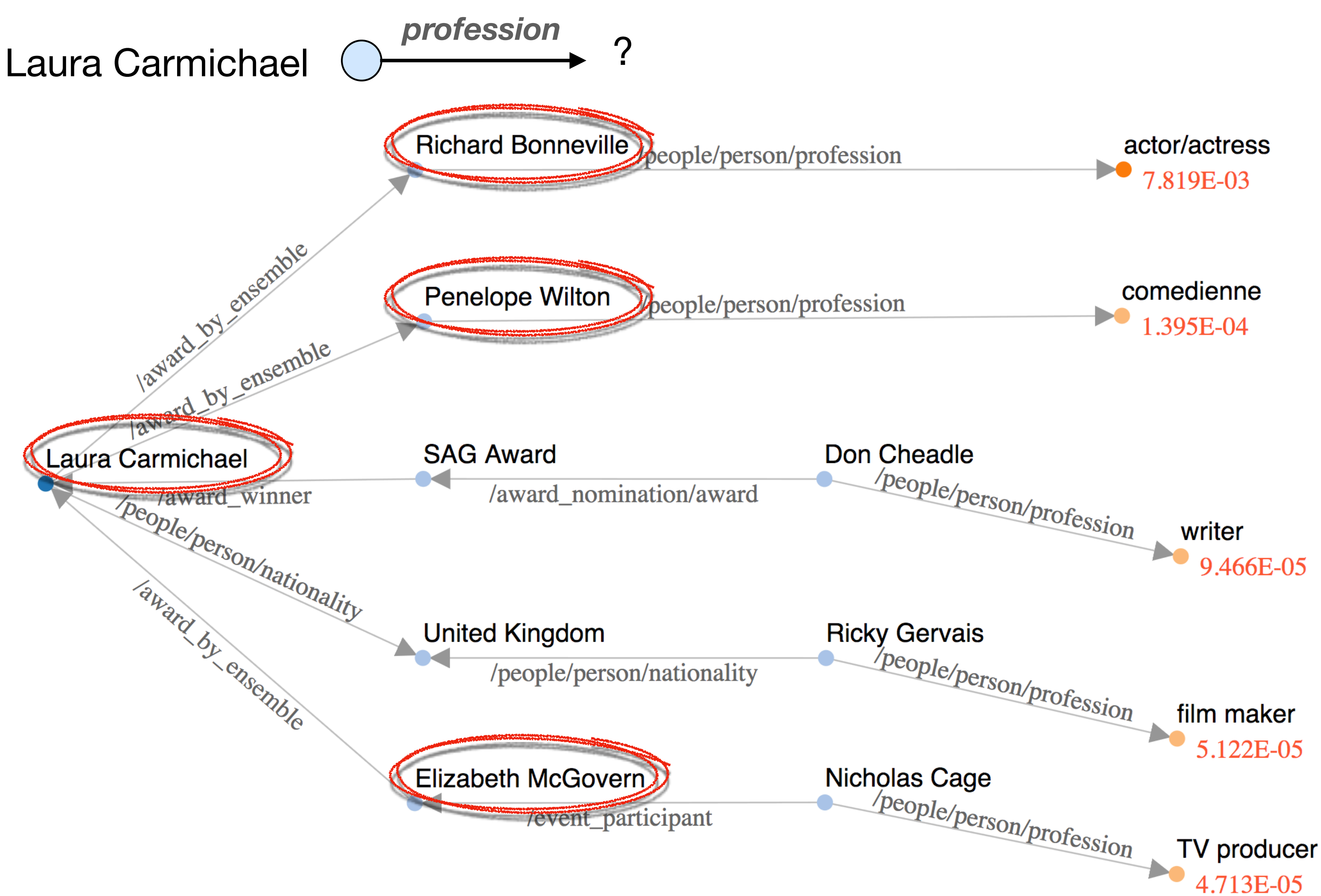
Interpretable Results





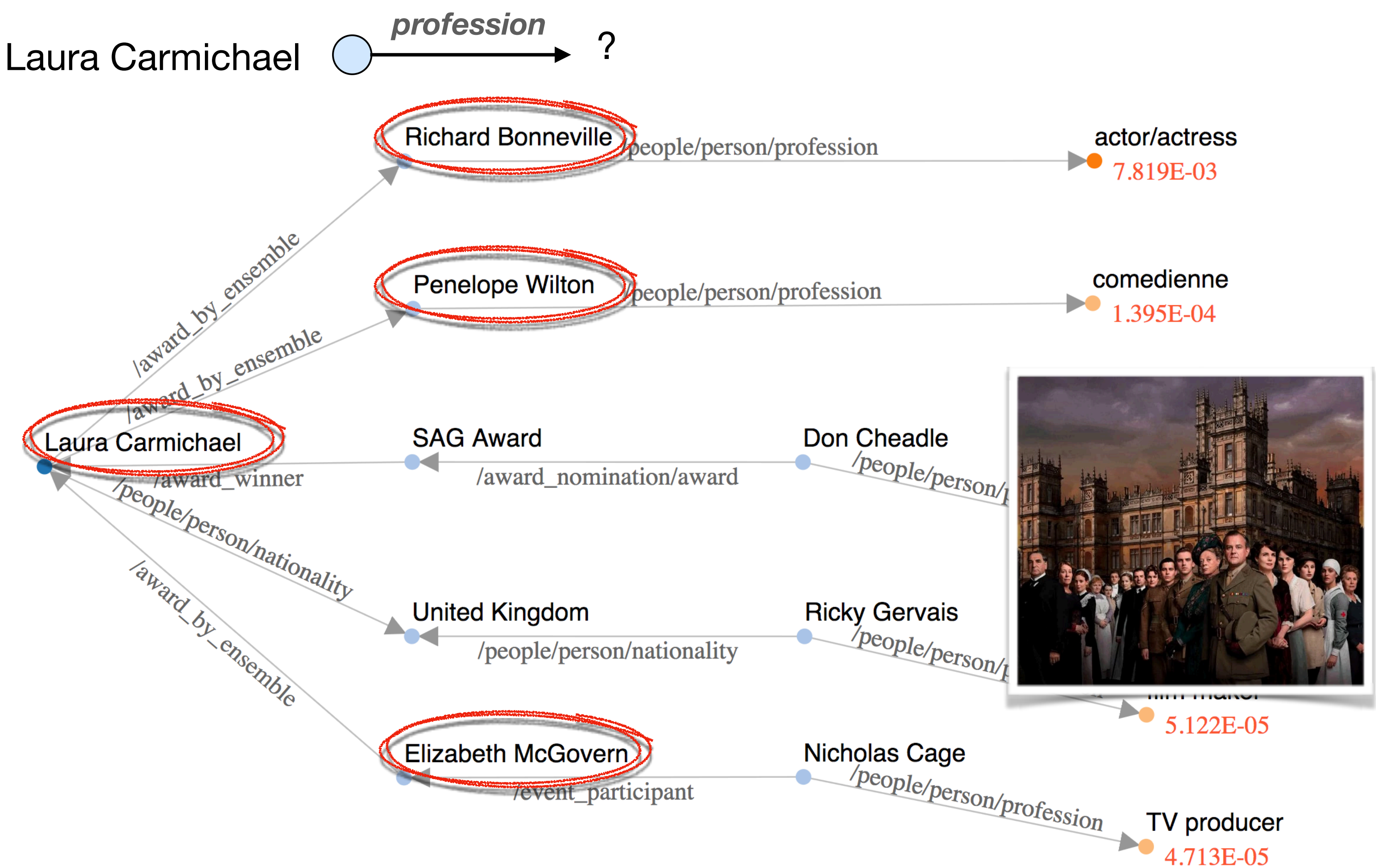
FB15k-237 (Toutanova and Chen 2016)

Interpretable Results



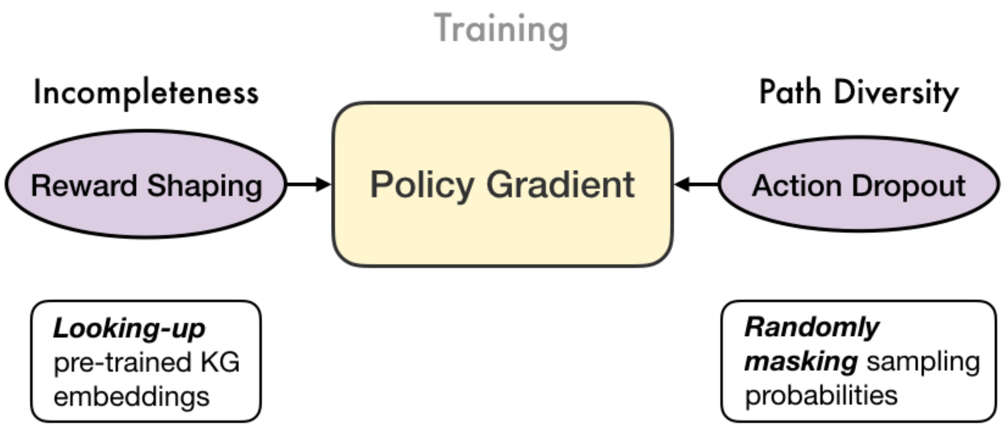
FB15k-237 (Toutanova and Chen 2016)

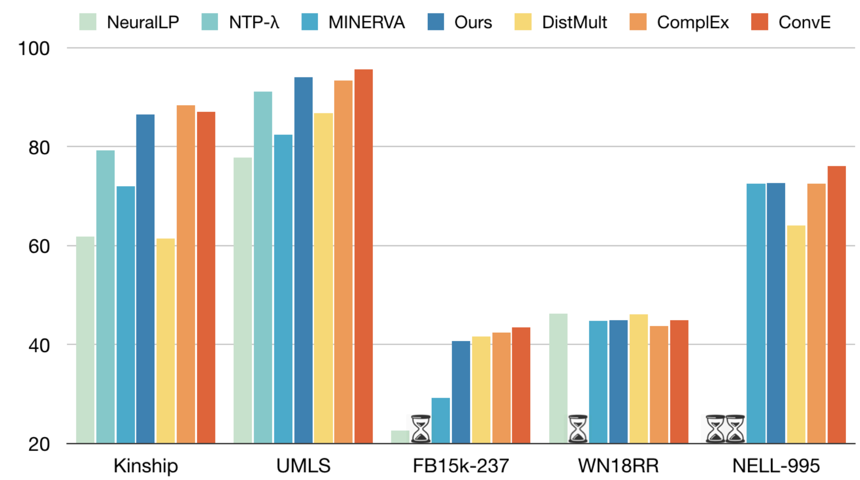
Interpretable Results



FB15k-237 (Toutanova and Chen 2016)

Code: https://github.com/salesforce/MultiHopKG





Future directions

- Learn better reward shaping functions
- Investigate similar techniques for other RL paradigms (e.g. Q-learning)
- Extend to more complicated structured queries (e.g. more than one topic entities)
- Extend to natural language QA

BKI - Error Analysis

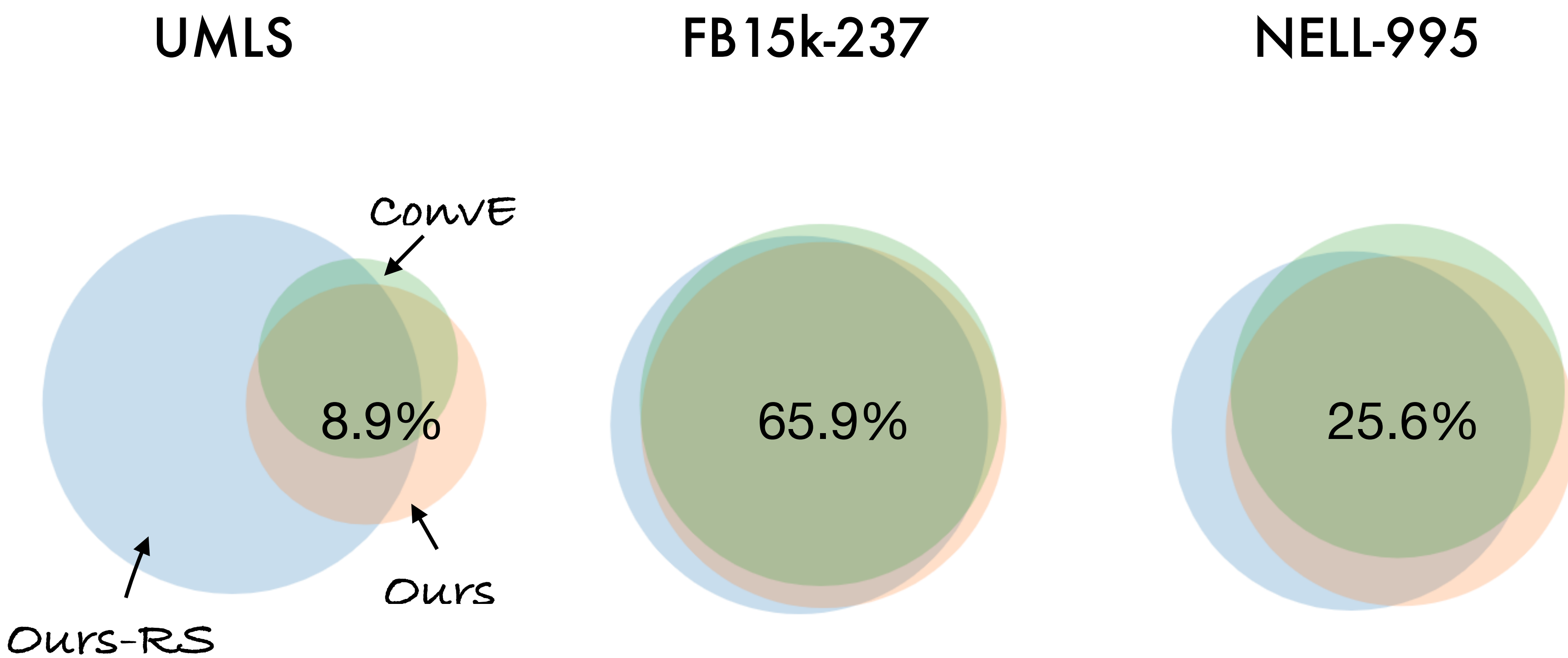


Fig 4. Dev set top-1 prediction error overlap of ConvE, Ours and Ours-RS. The absolute error rate of Ours is shown.

BK II - Challenges

Incompleteness

≈ 30% false negative feedback

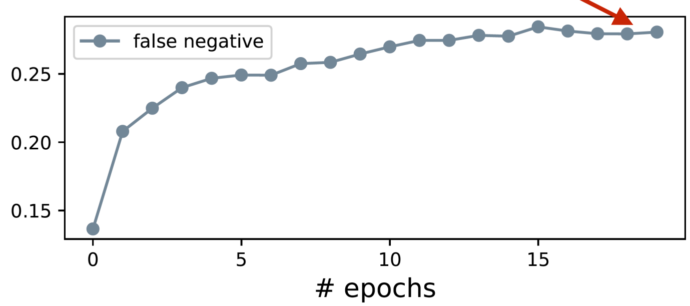
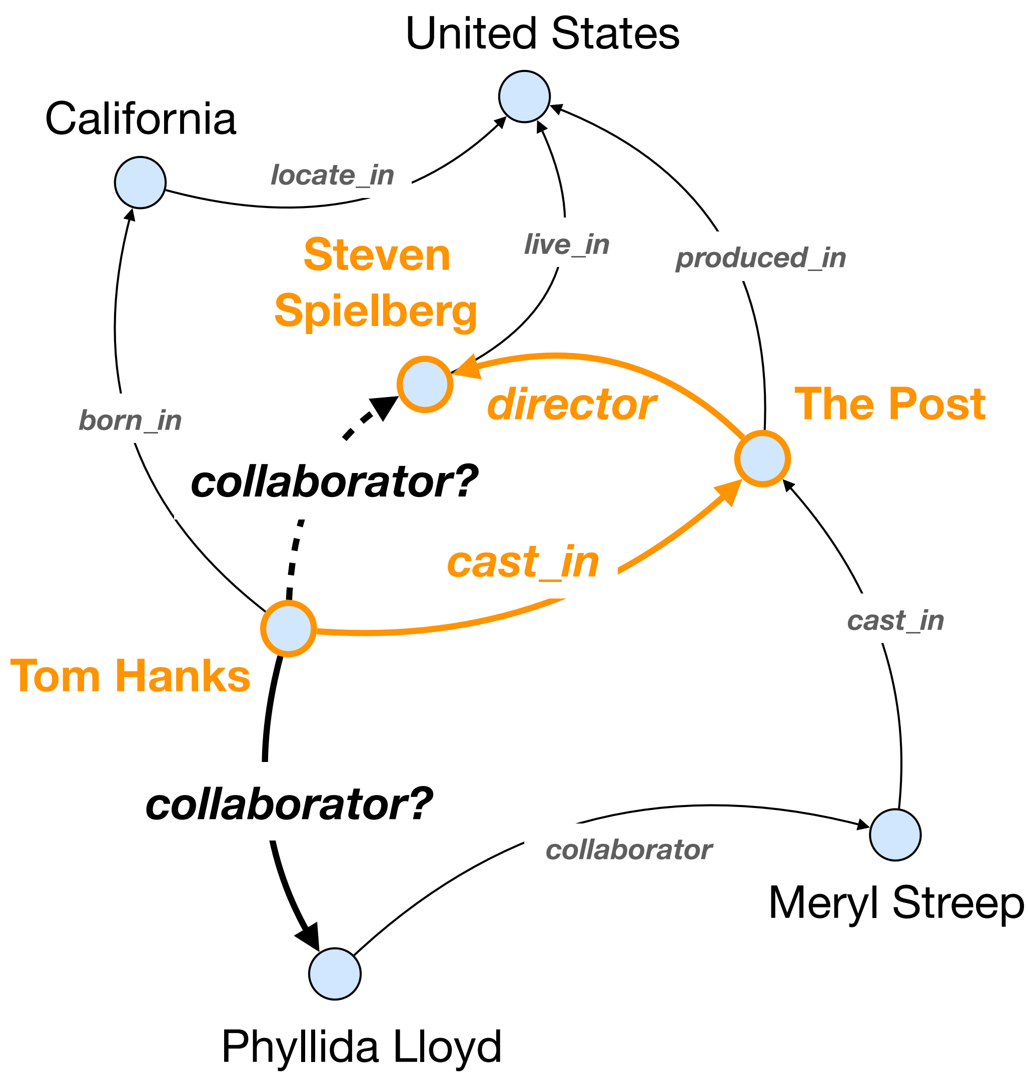


Fig 1. % of false negative hit in the first 20 epochs of RL training on the UMLS KG benchmark (Kok and Domingos 2007)



BKIII - Efficient Training & Inference