

Multi-Hop Knowledge Graph Reasoning with Reward Shaping

Victoria Lin, Richard Socher, Caiming Xiong
[{xilin, rsocher, cxiong}@salesforce.com](mailto:{xilin,rsocher,cxiong}@salesforce.com)

EMNLP 2018

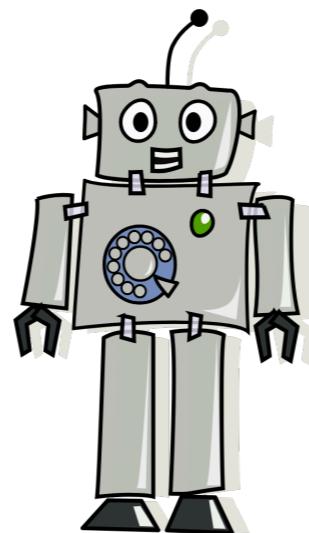


Question Answering System

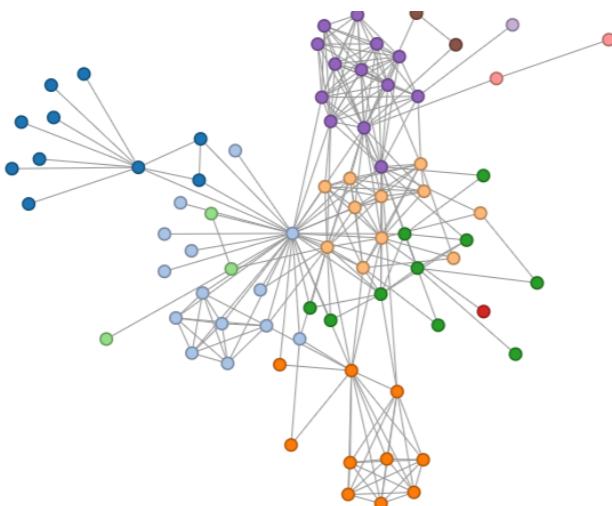
Text



Images



Knowledge Graph

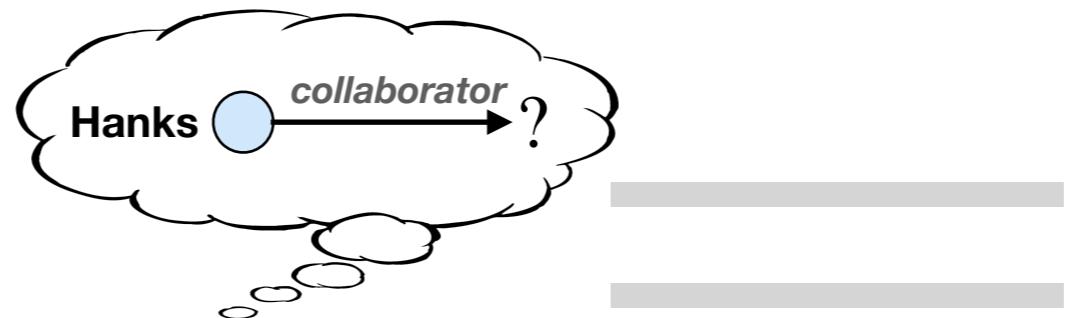
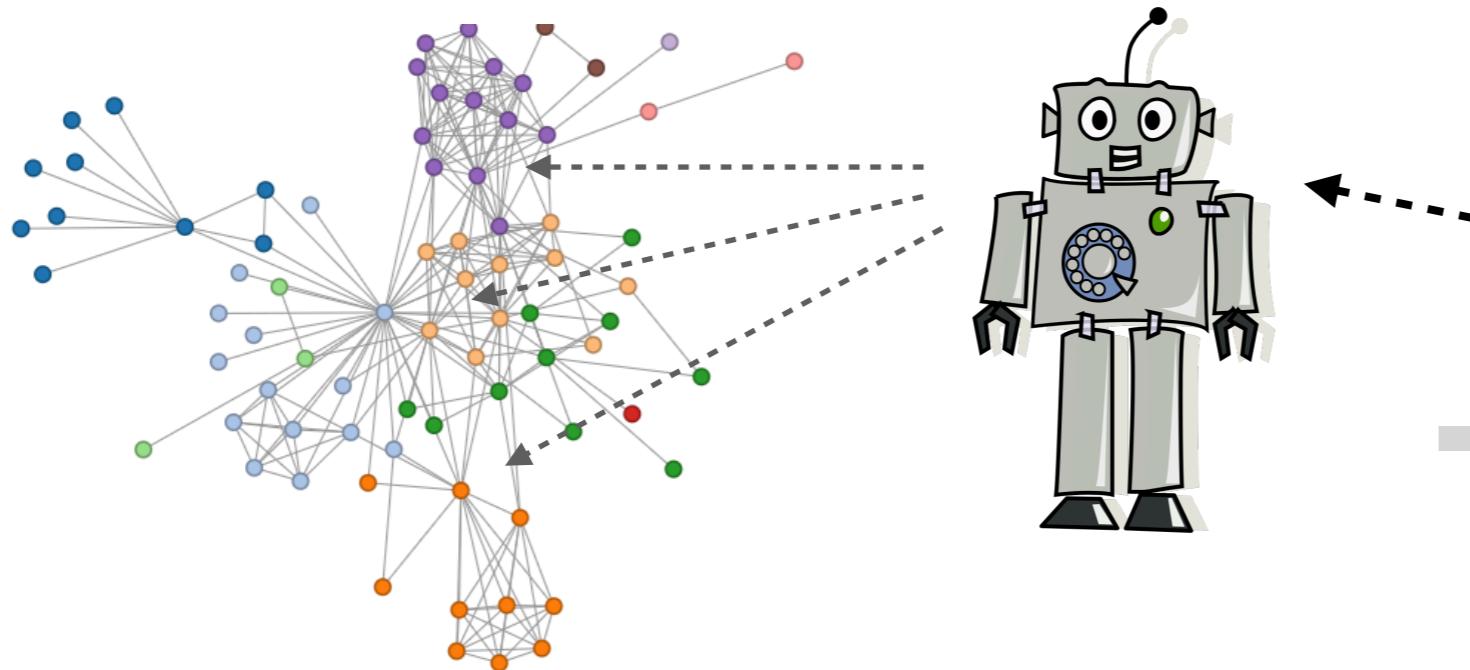


Which directors
has Tom Hanks
collaborated with?



Question Answering System

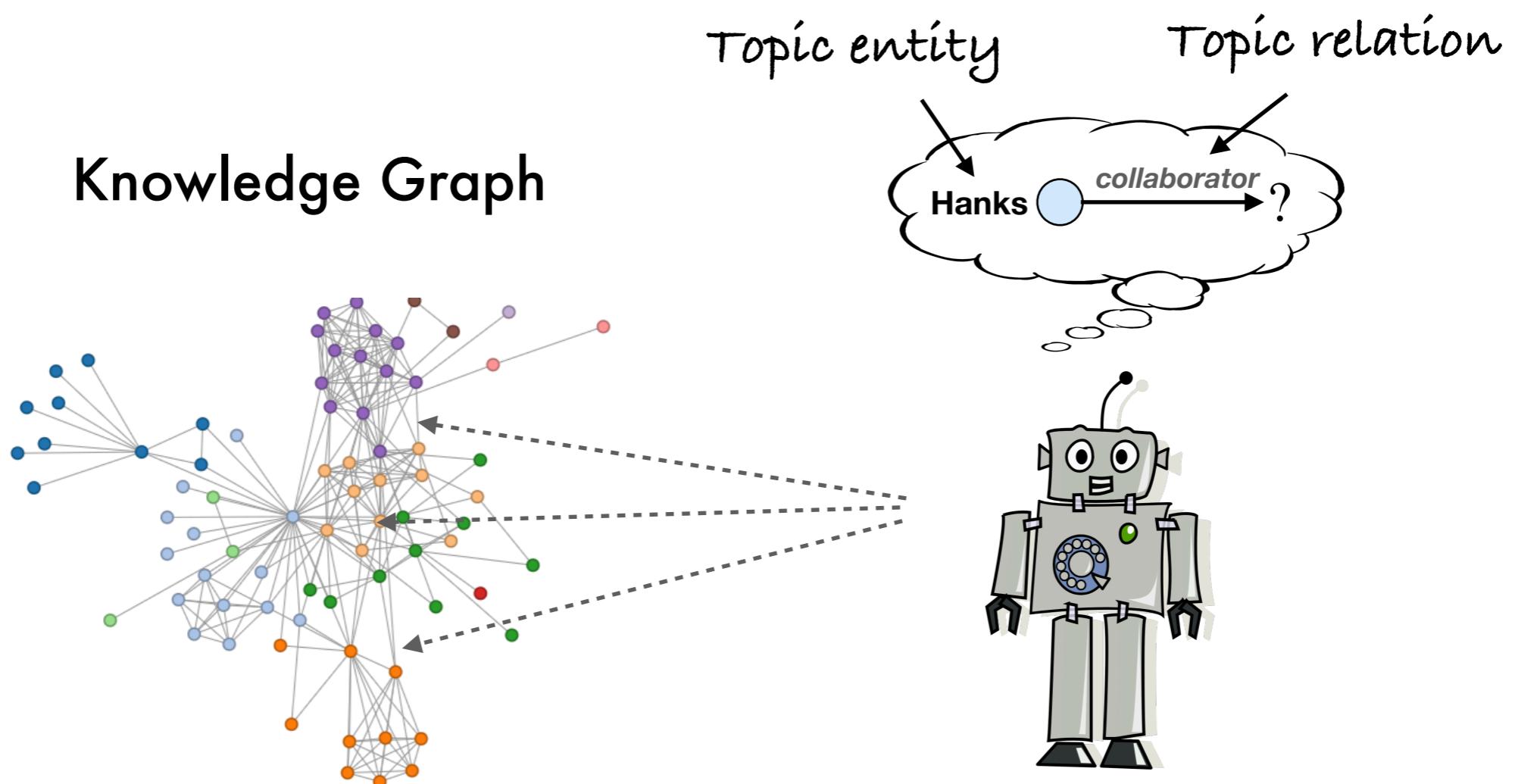
Knowledge Graph



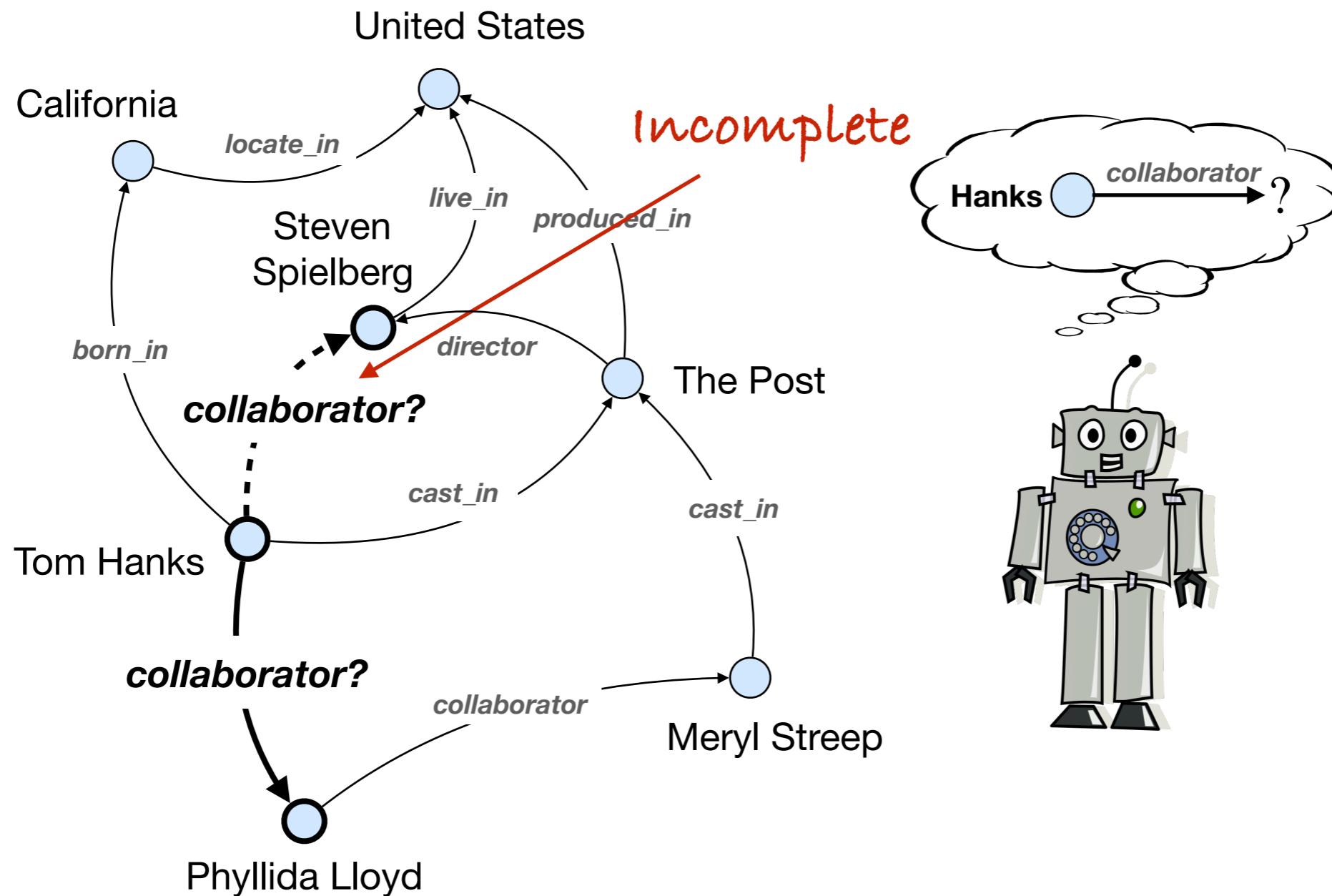
Which directors
has **Tom Hanks**
collaborated with?



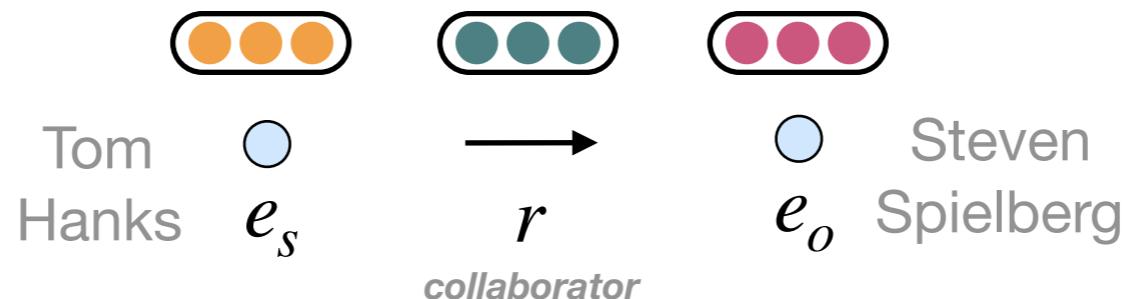
Structured Query Answering



Structured Query Answering



Knowledge Graph Embeddings



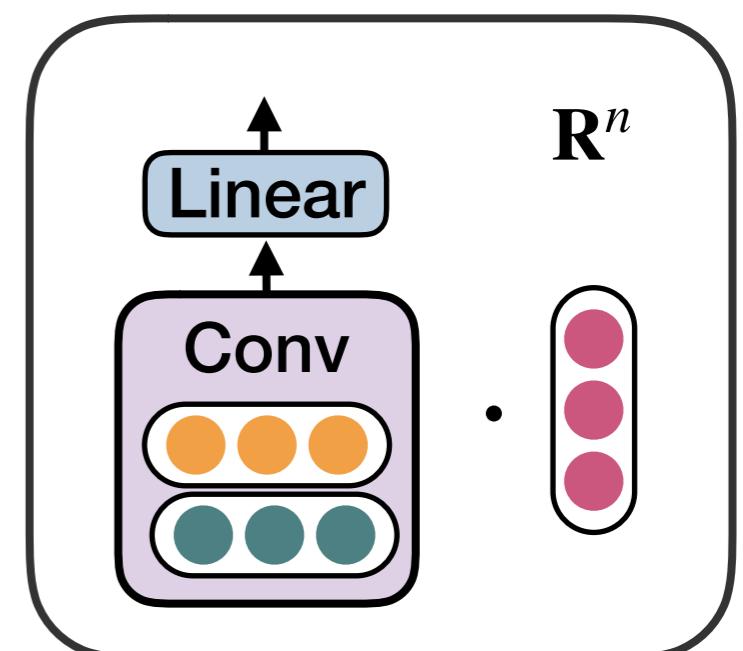
Highly accurate &
Efficient

	MRR
ConvE	0.957 (max = 1)

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

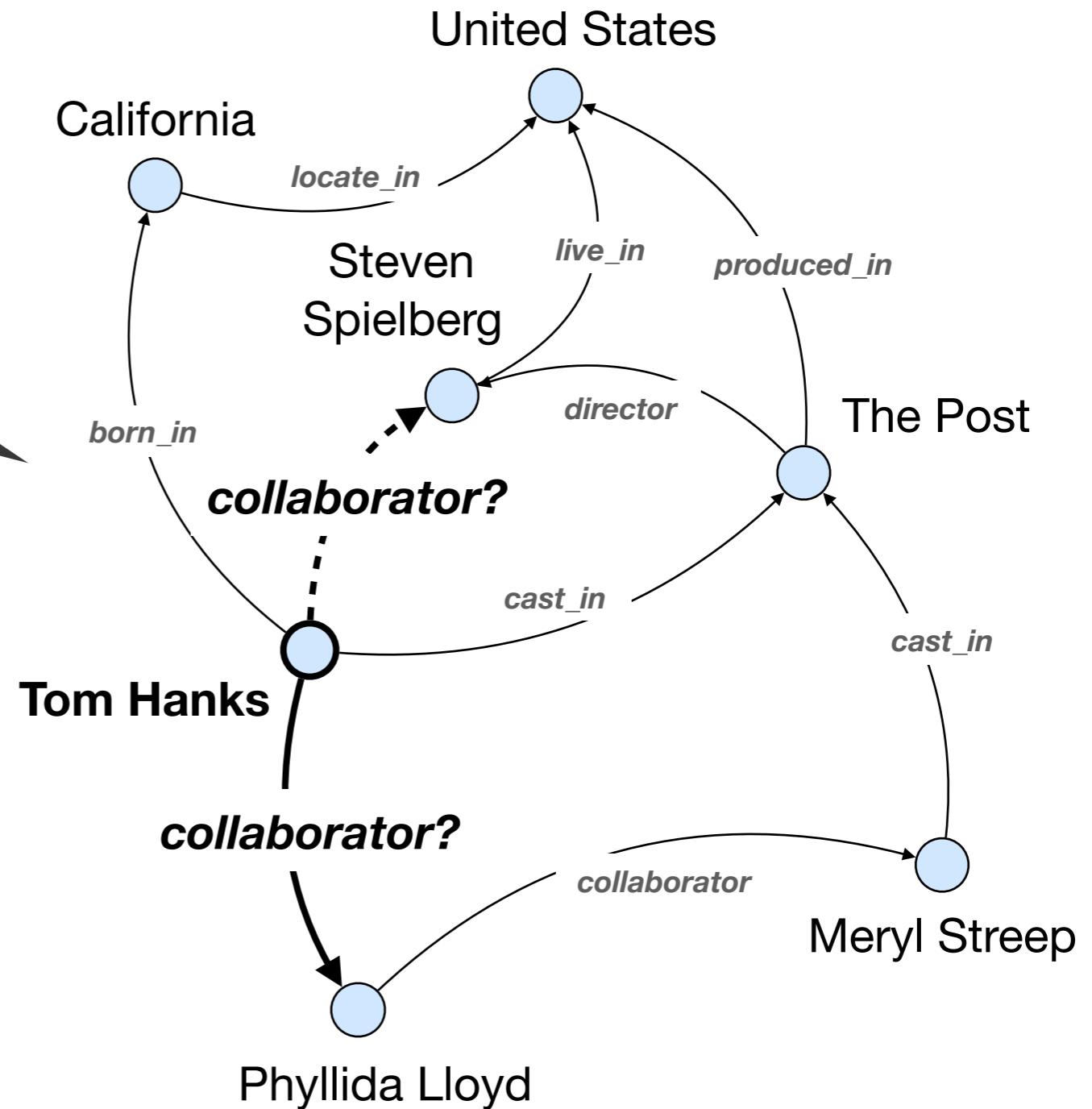
Lack
interpretability

Why Spielberg
is a collaborator
of Hanks?



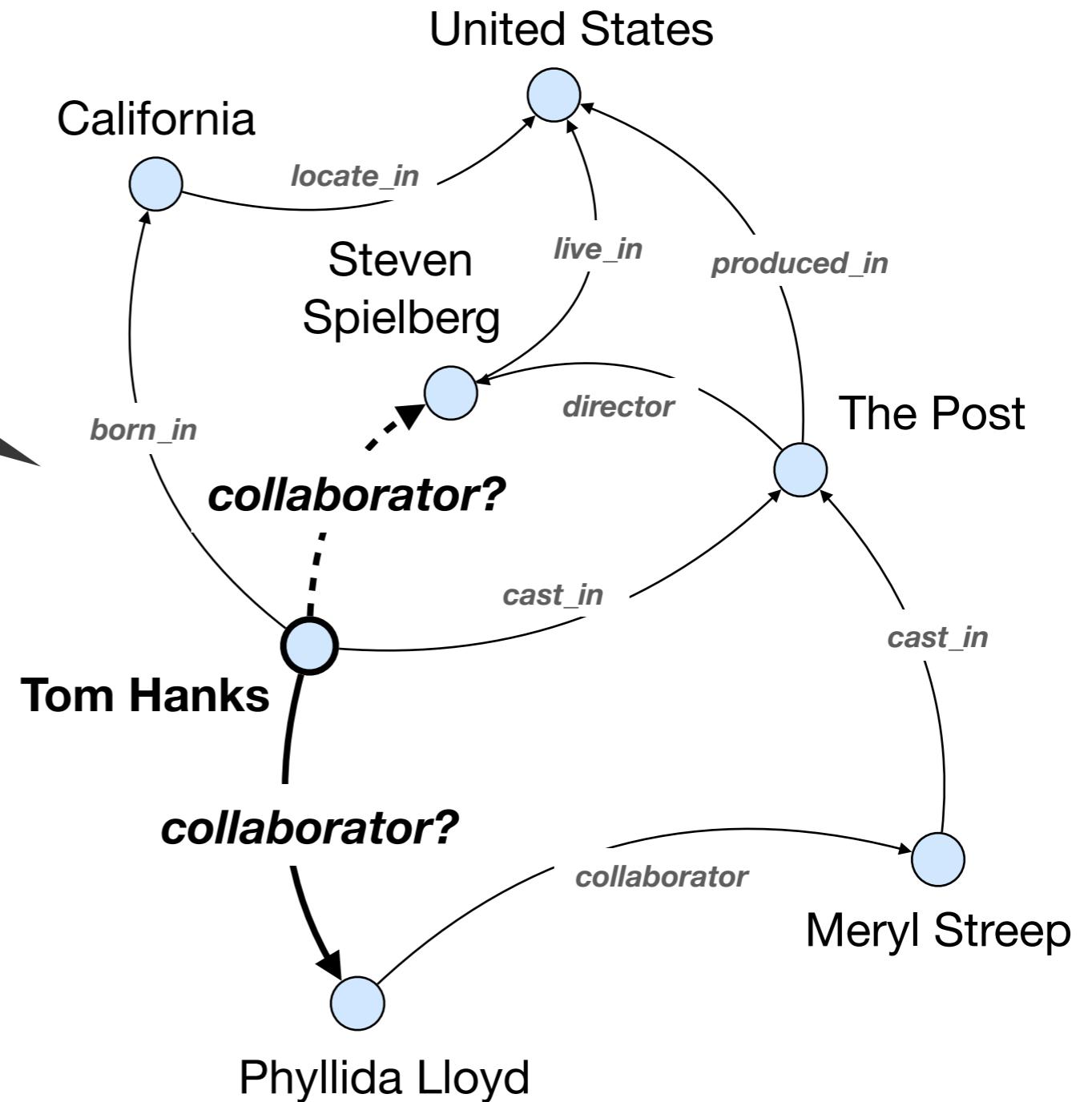
Multi-Hop Reasoning Models

Reasoning
over discrete
structures

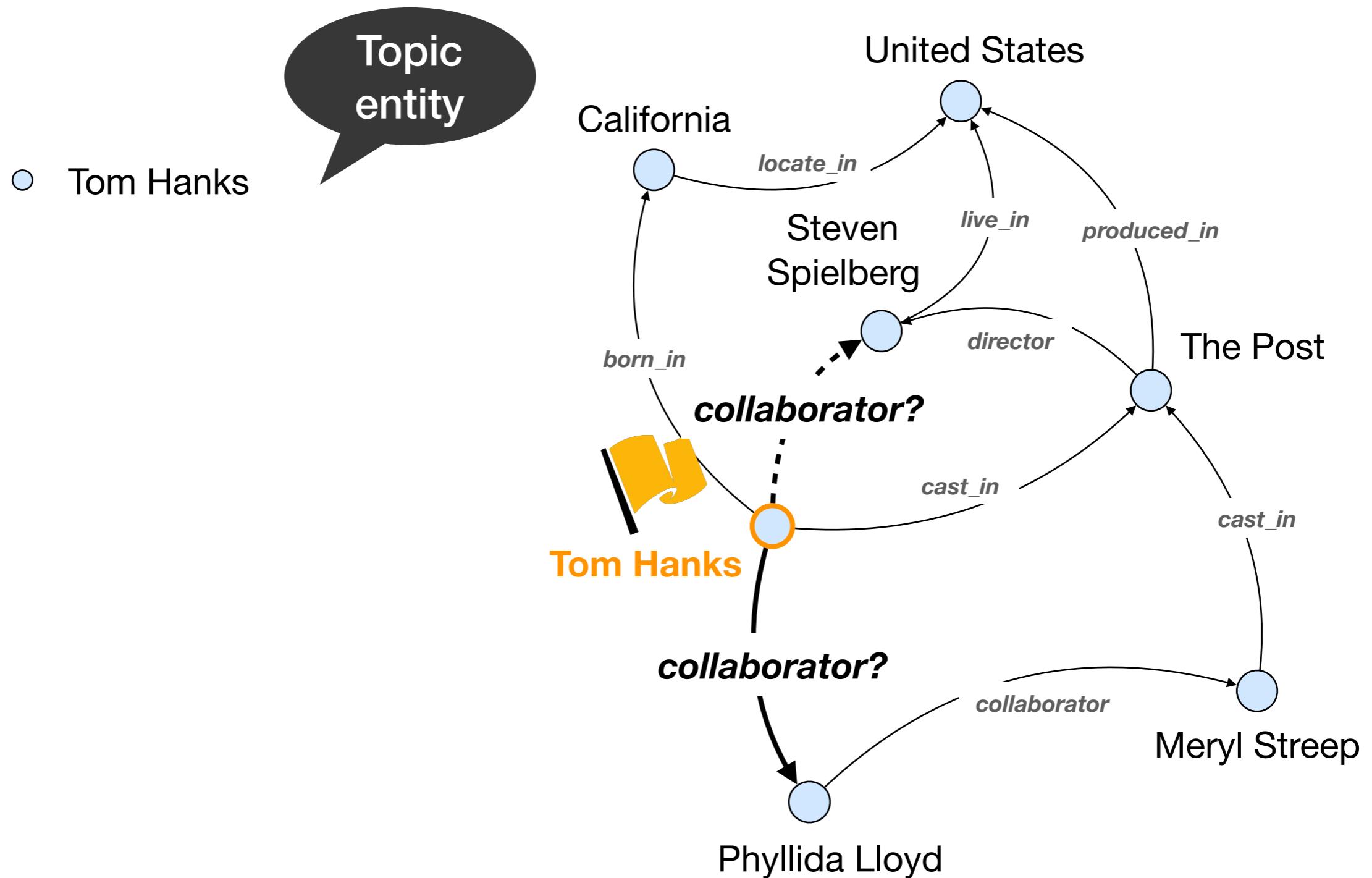


Multi-Hop Reasoning Models

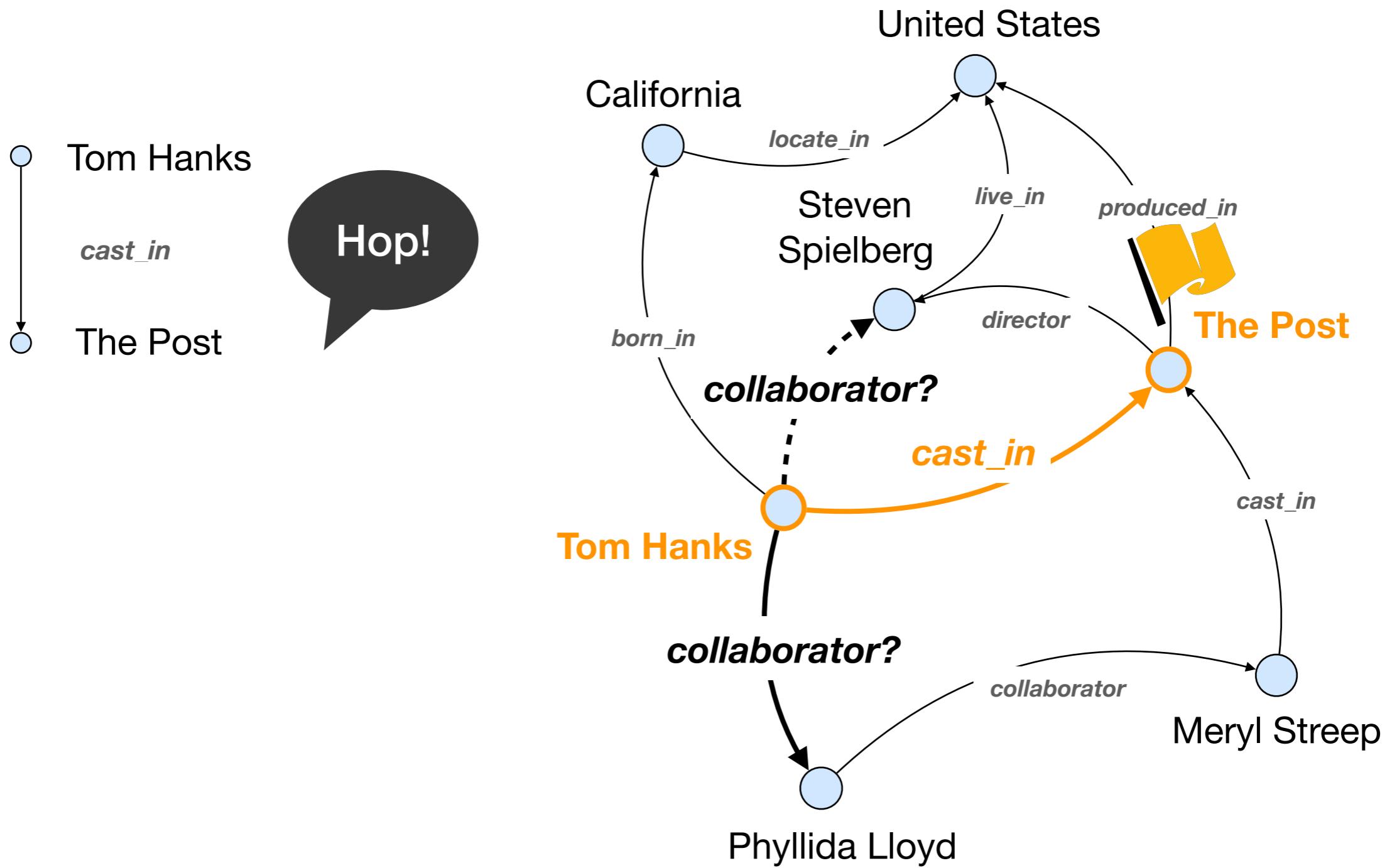
Sequential
decision making



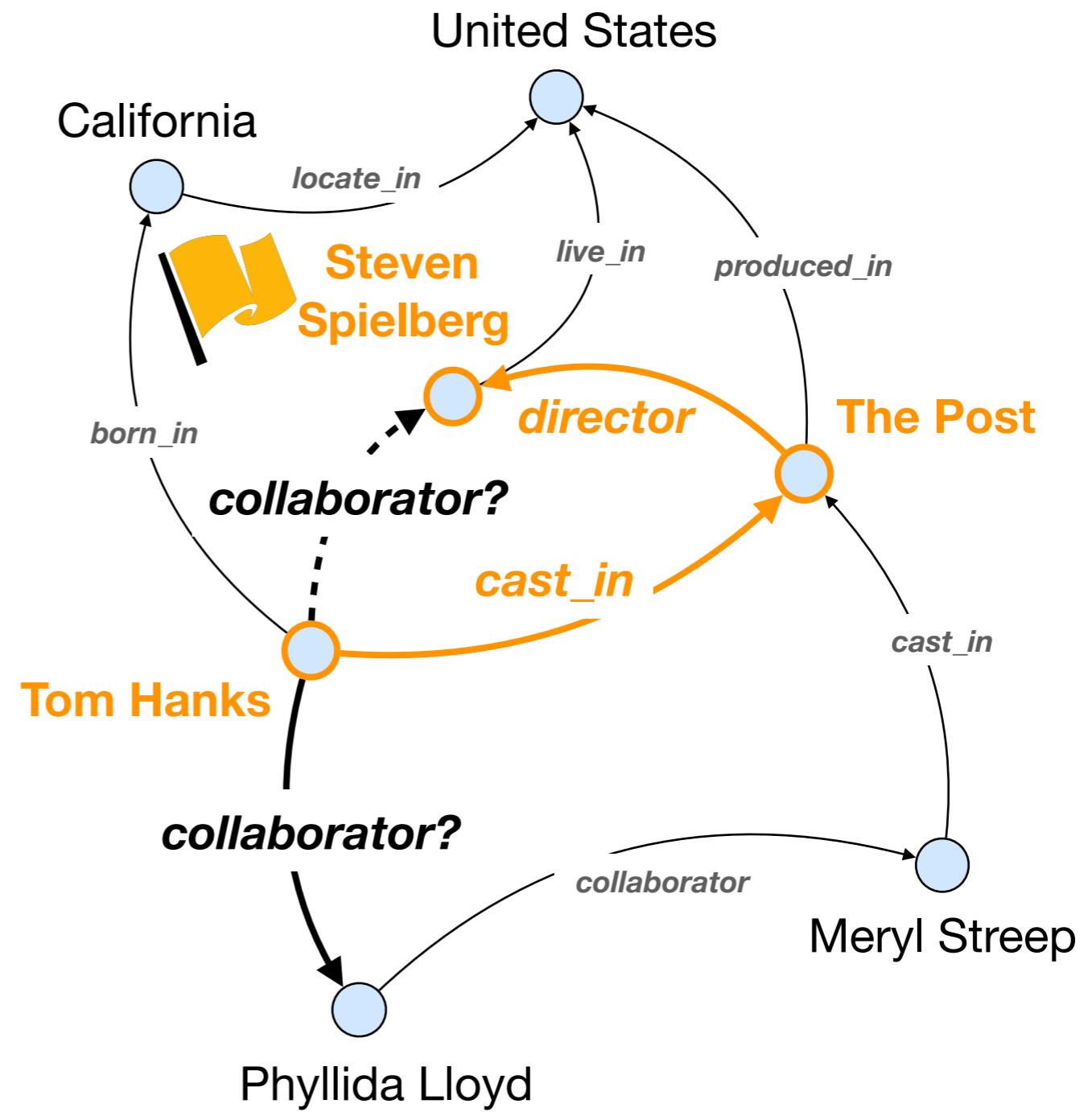
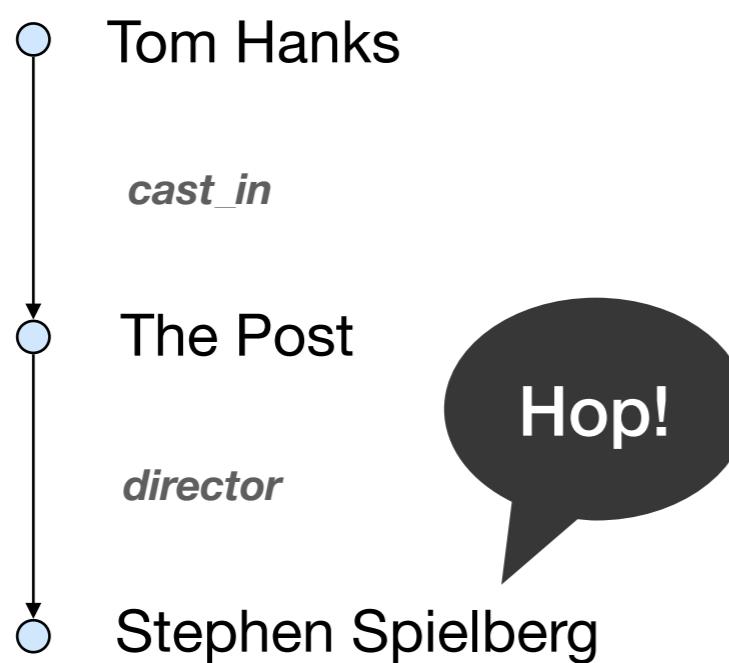
Multi-Hop Reasoning Models



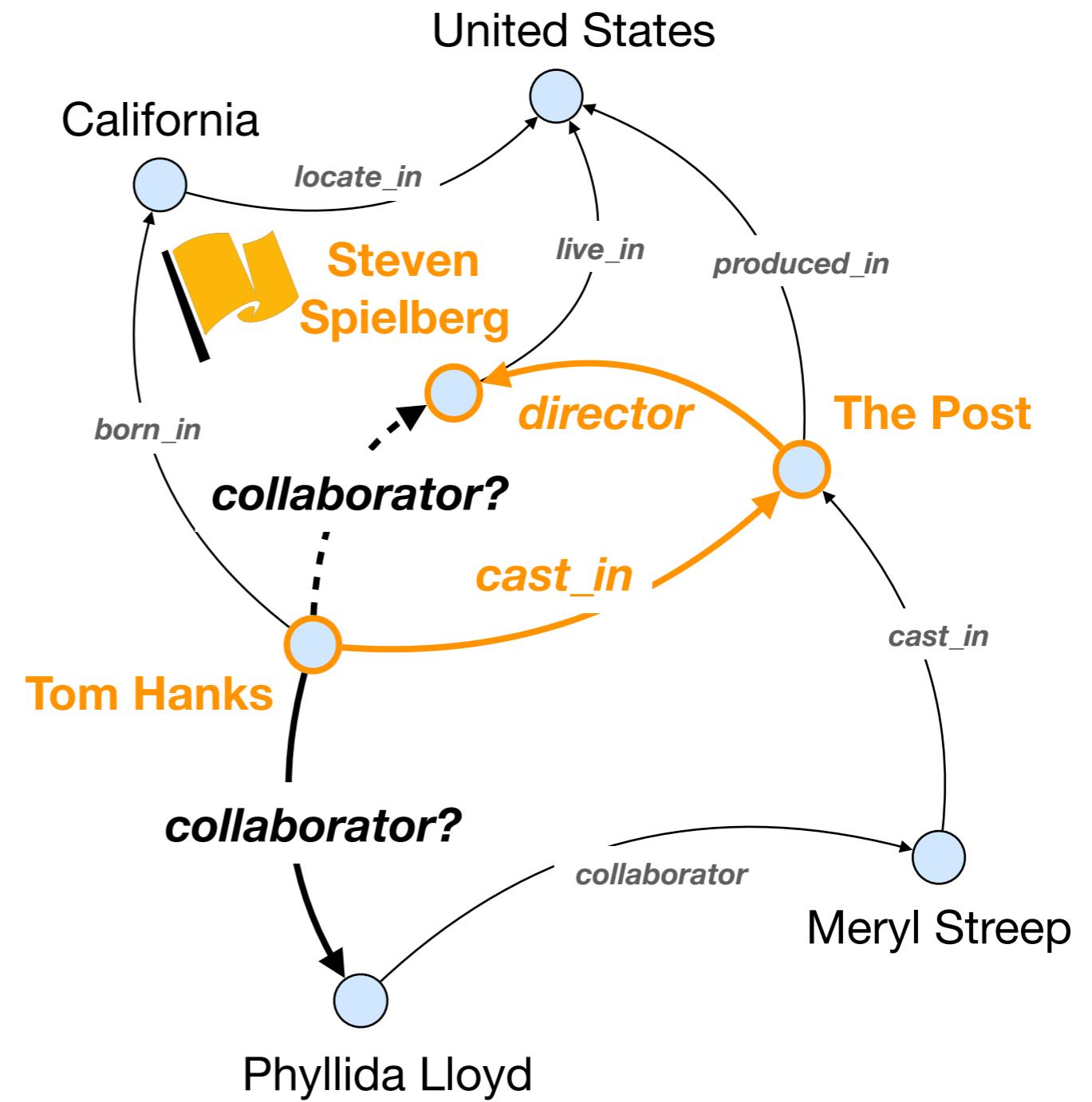
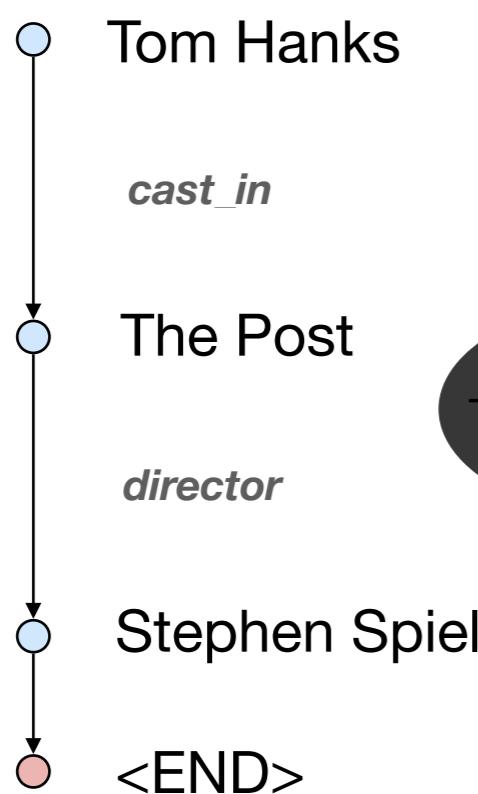
Multi-Hop Reasoning Models



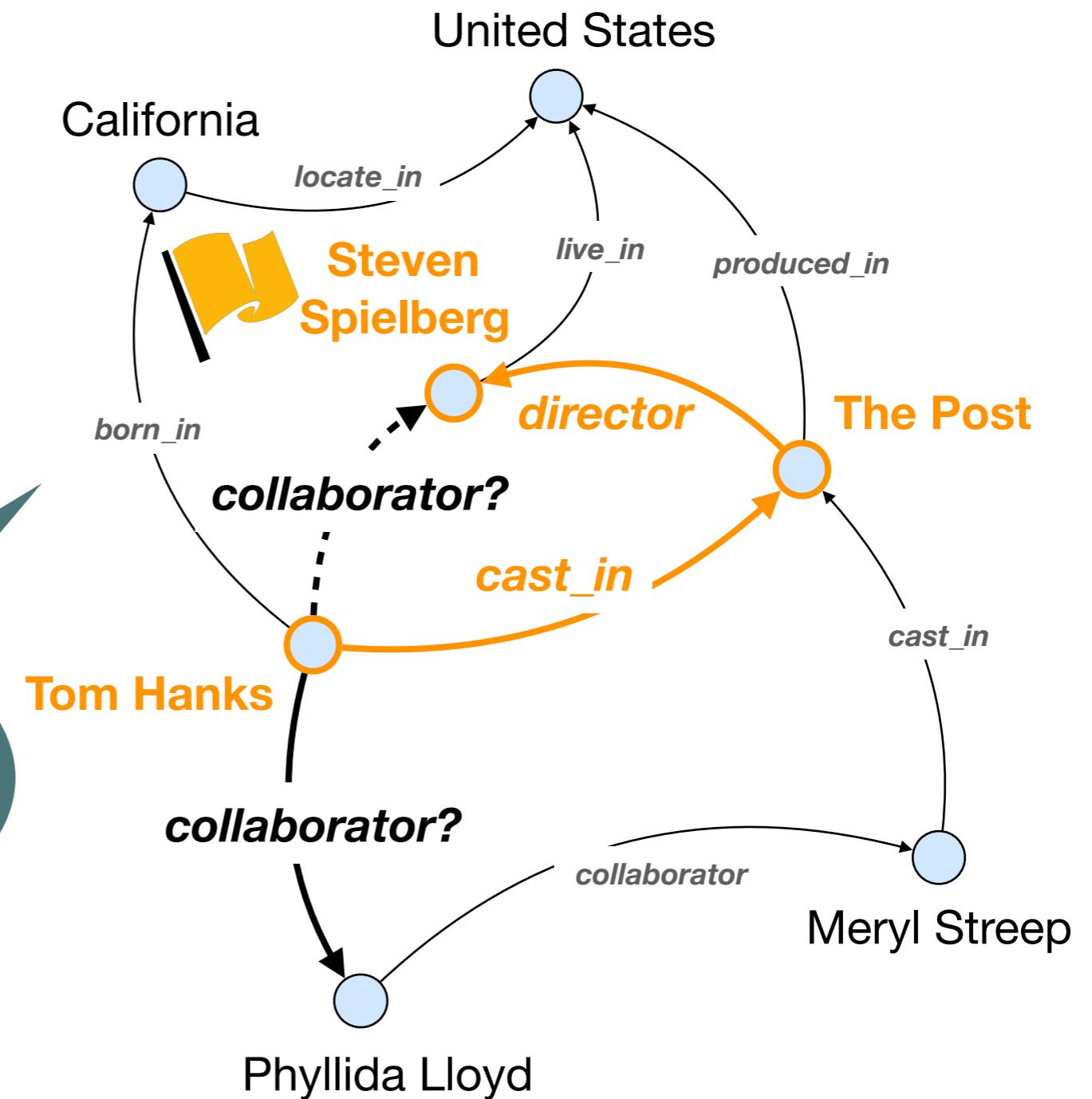
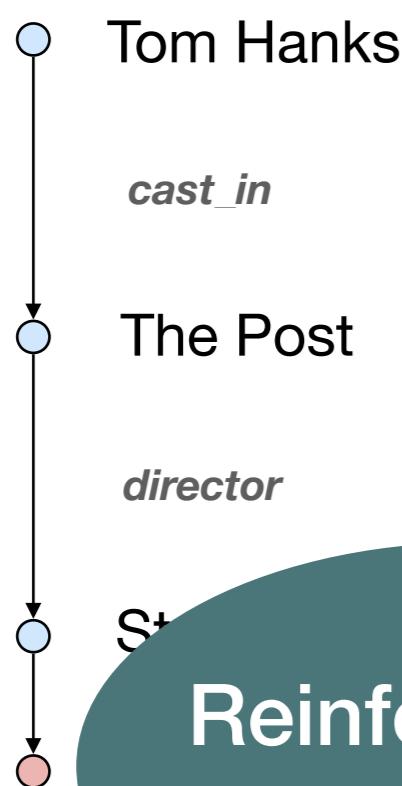
Multi-Hop Reasoning Models



Multi-Hop Reasoning Models



Multi-Hop Reasoning Models



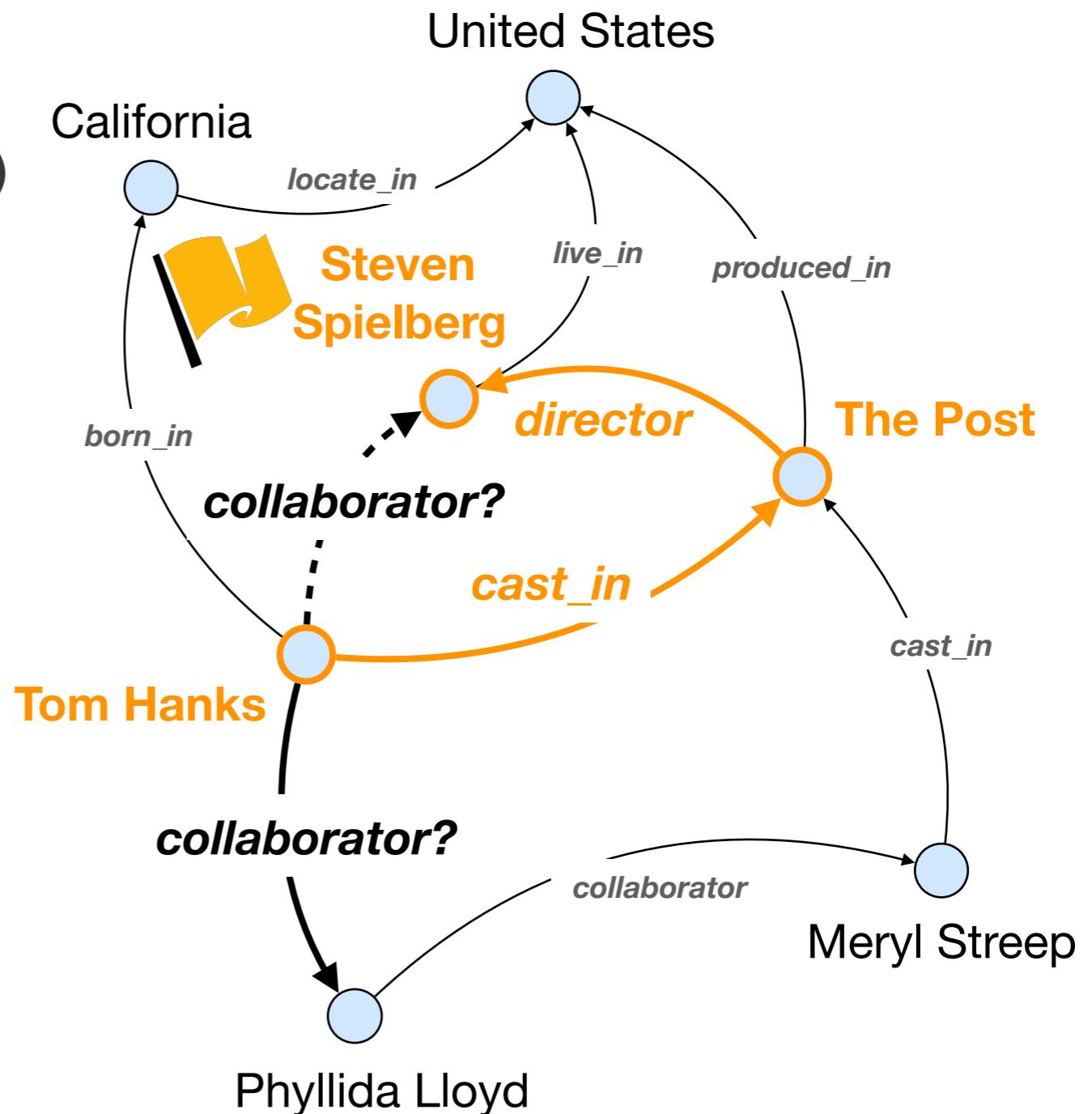
Multi-Hop Reasoning Models



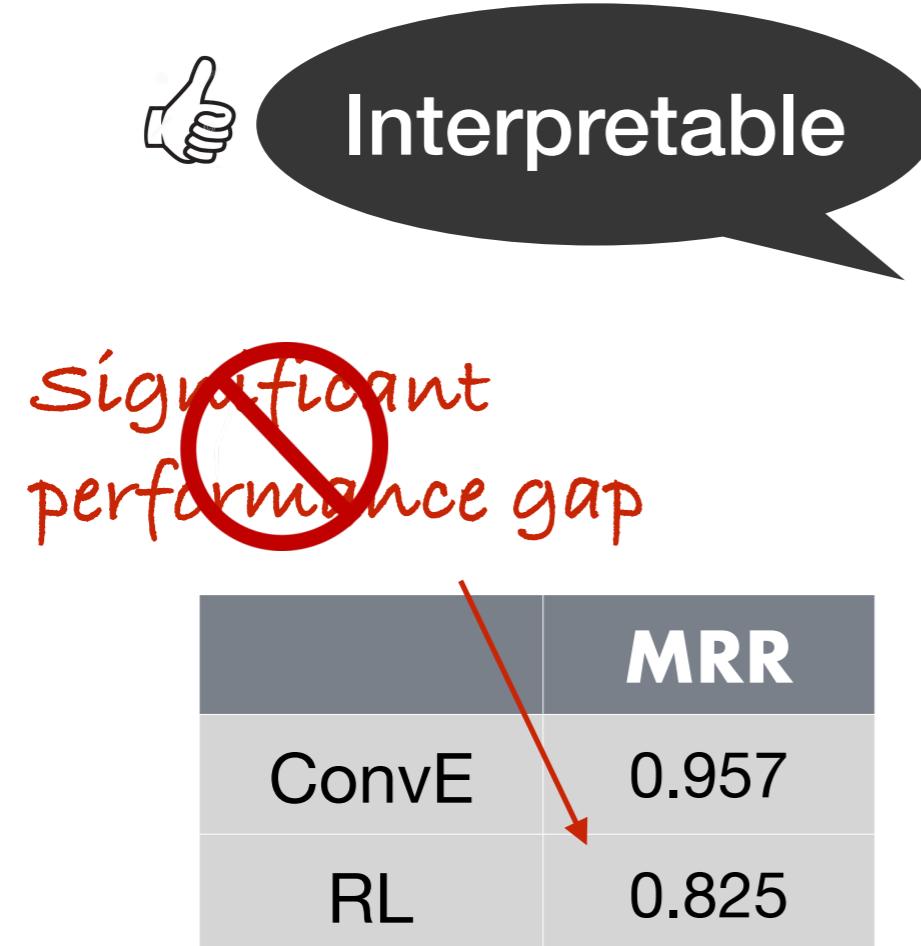
significant
performance gap

MRR	
ConvE	0.957
RL	0.825

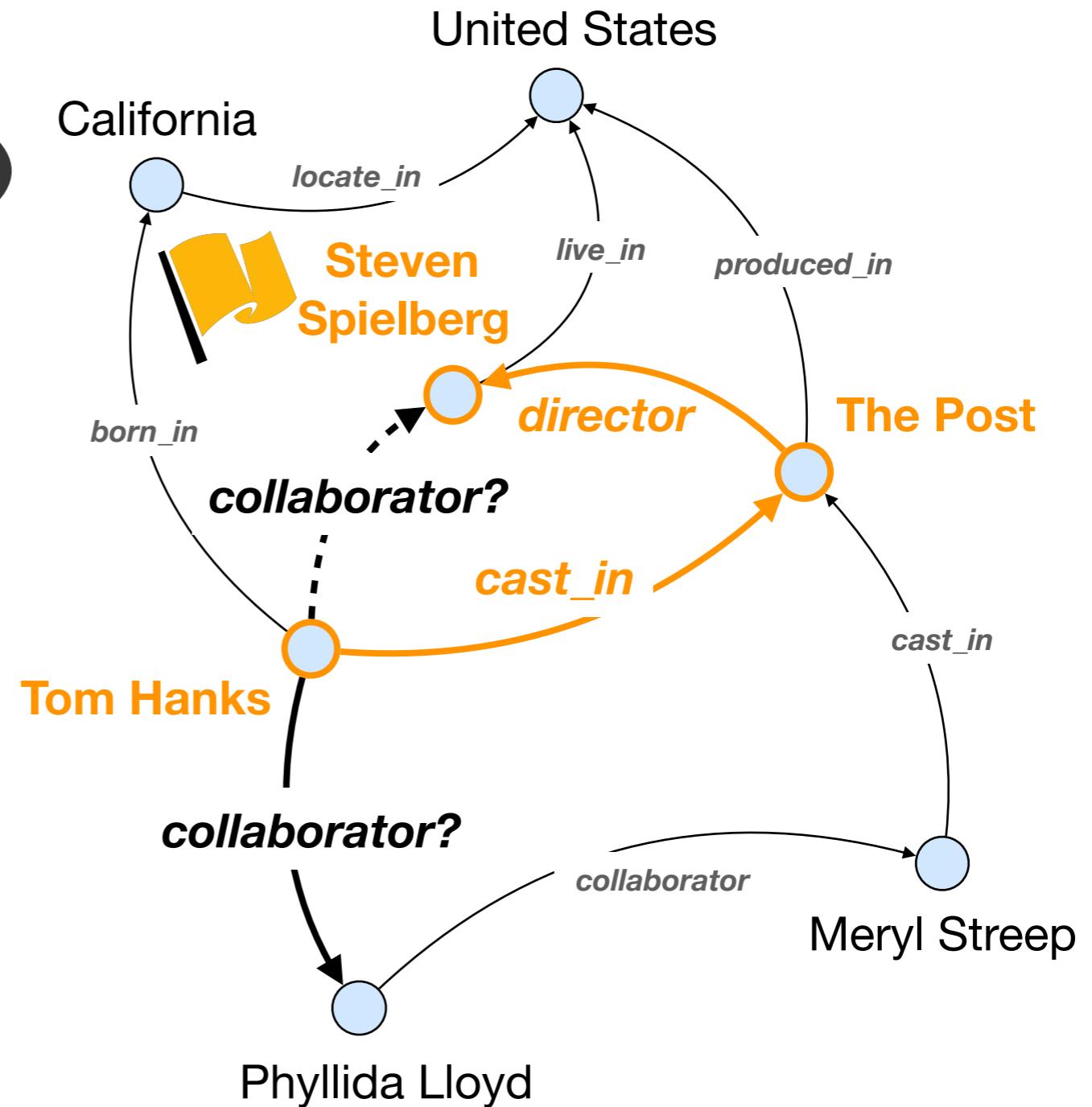
Tab 2. ConvE and RL (MINERVA) query answering performance on the UMLS benchmark dataset (Kok and Domingos 2007)



Multi-Hop Reasoning Models: Ideal Case

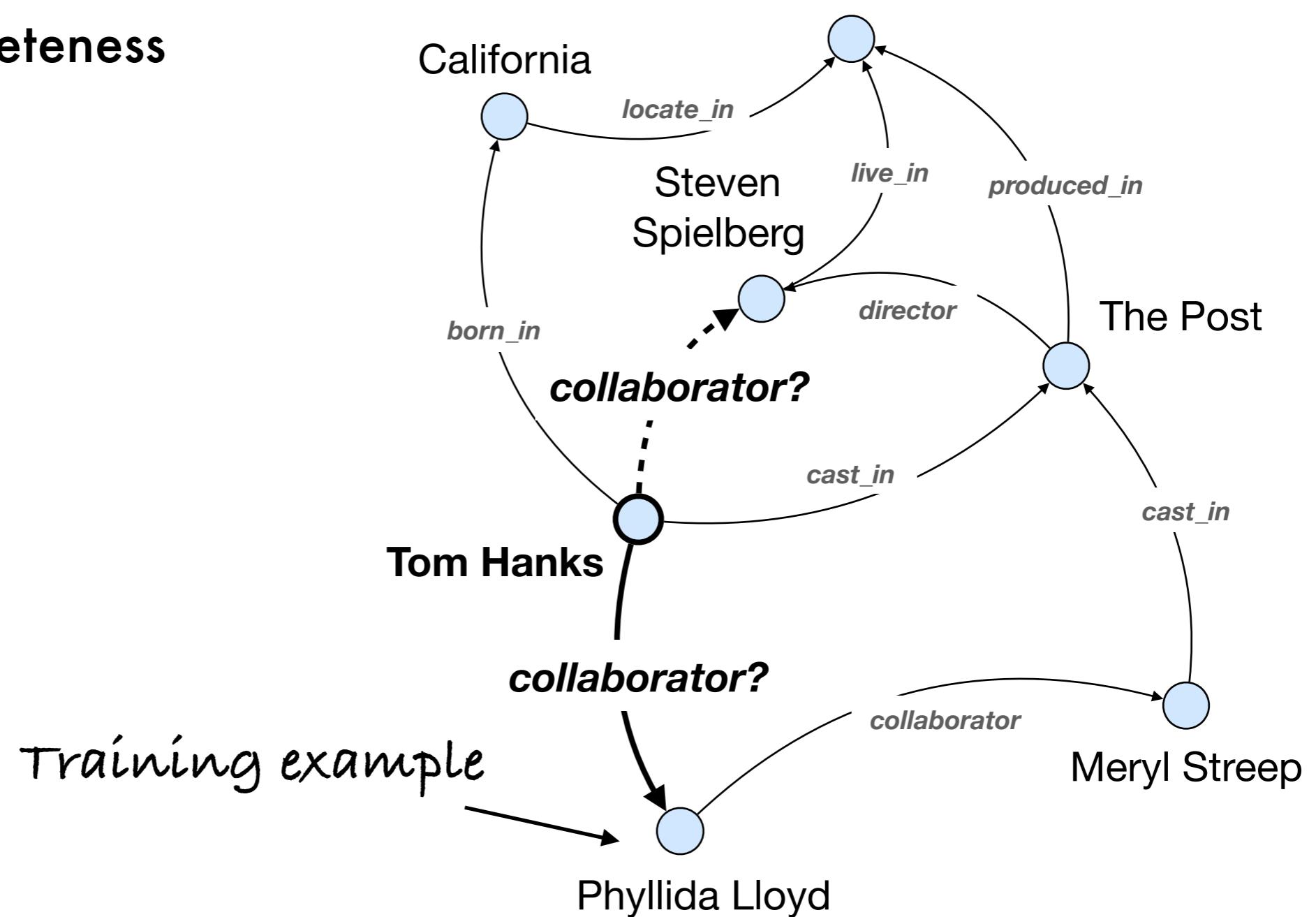


Tab 2. ConvE and RL (MINERVA) query answering performance on the UMLS benchmark dataset (Kok and Domingos 2007)



Challenges

Incompleteness



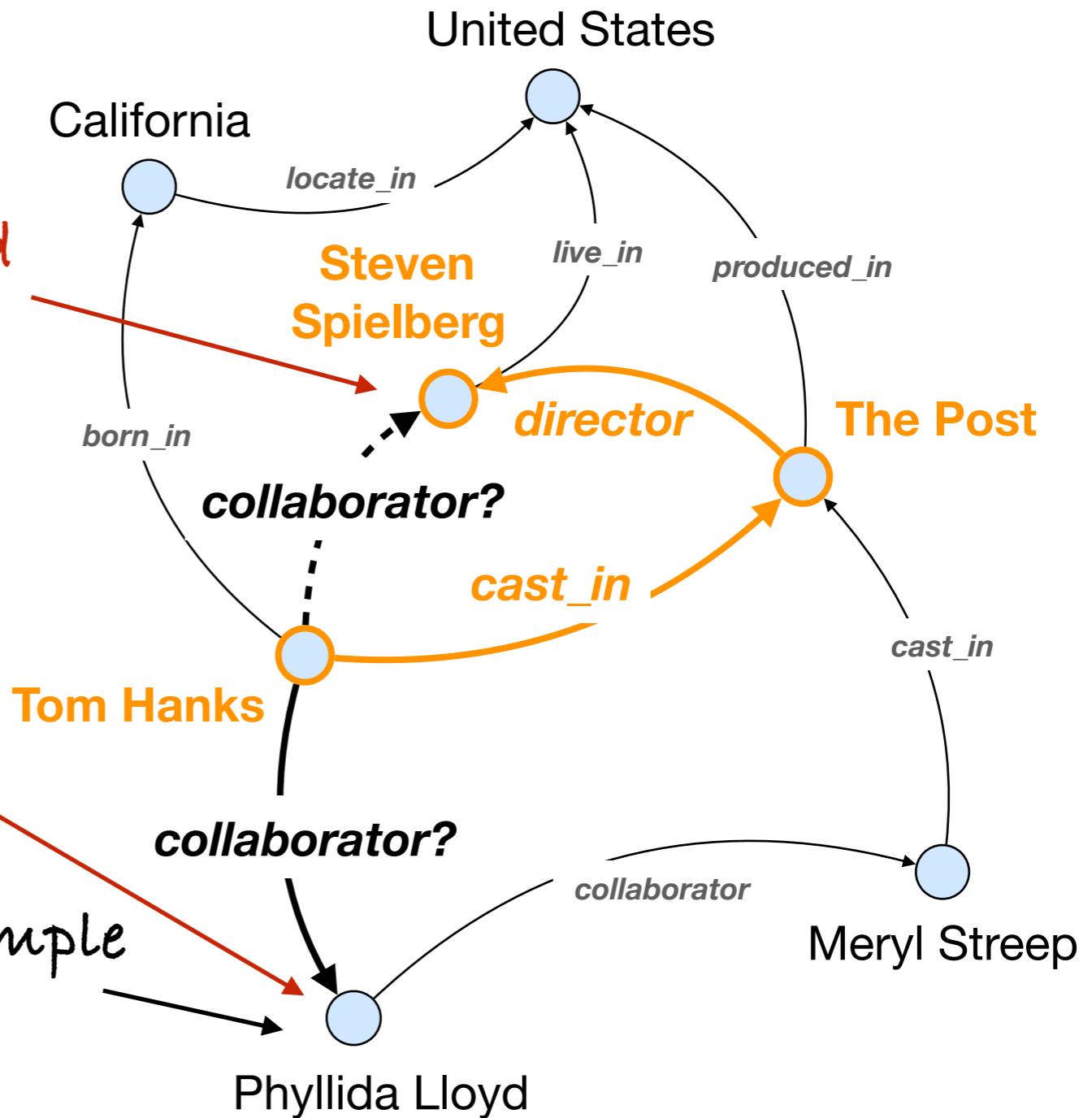
Challenges

Incompleteness

No reward

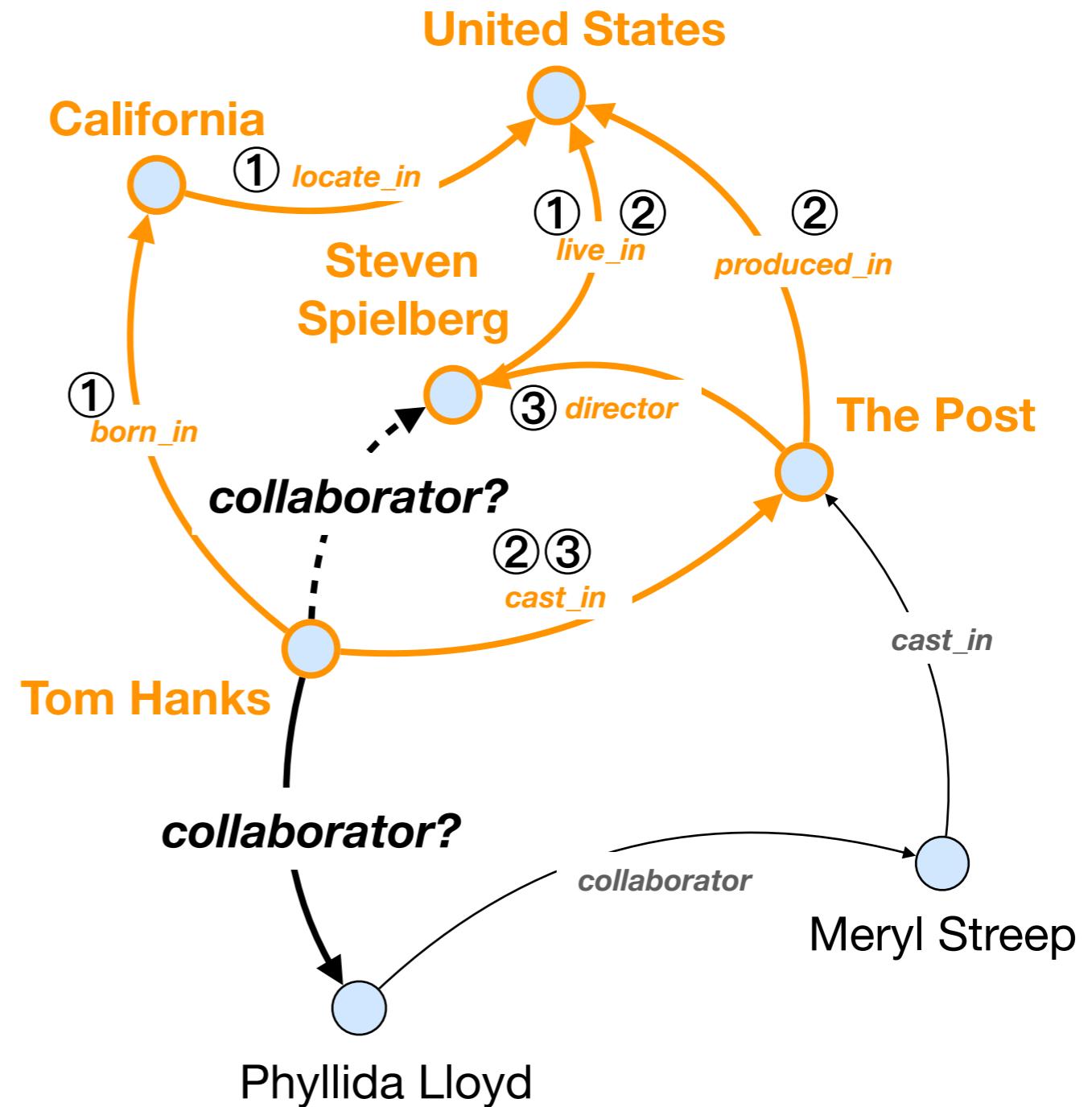
Overfit to the
observed answers

Training example



Challenges

Path Diversity

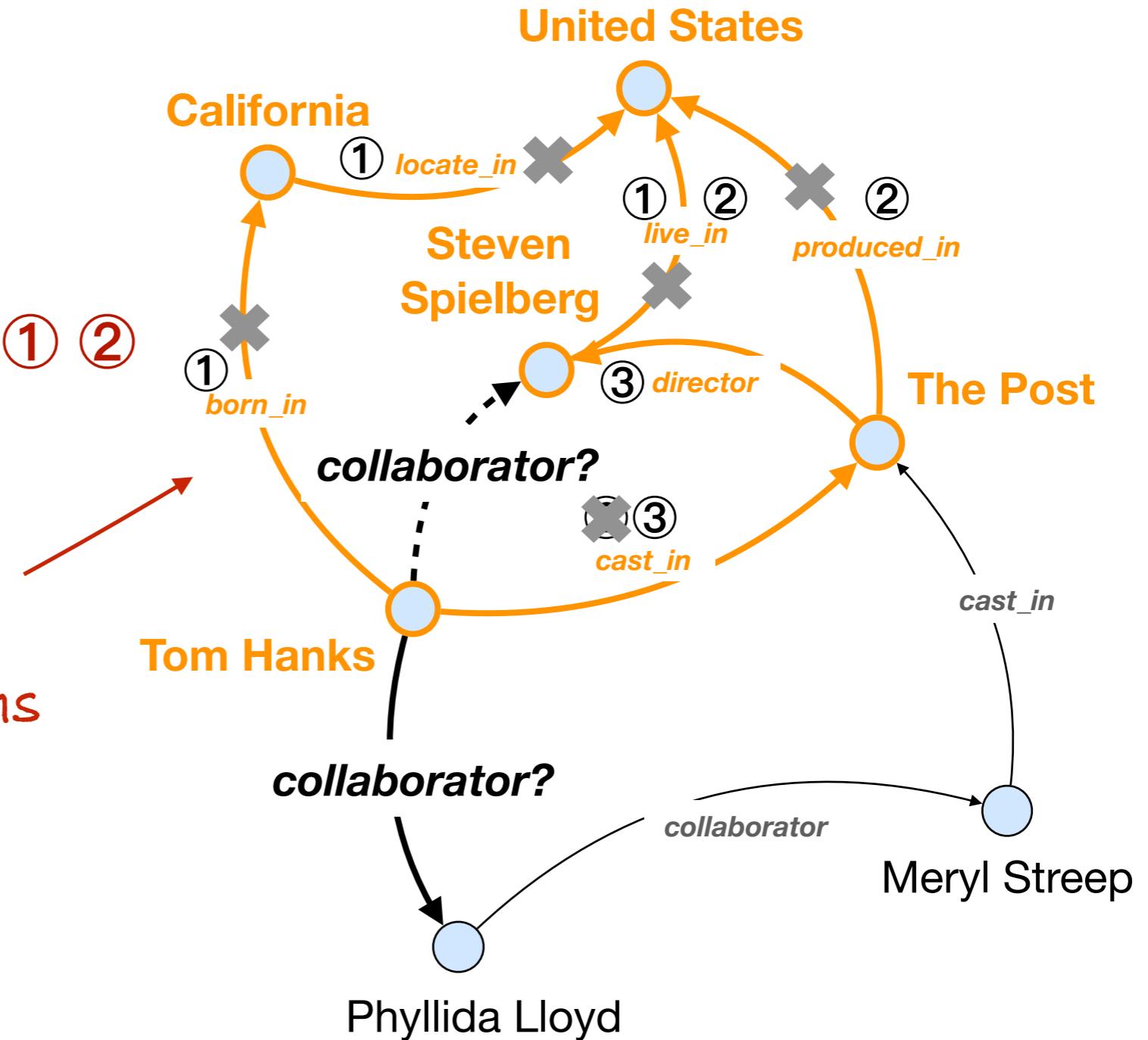


Challenges

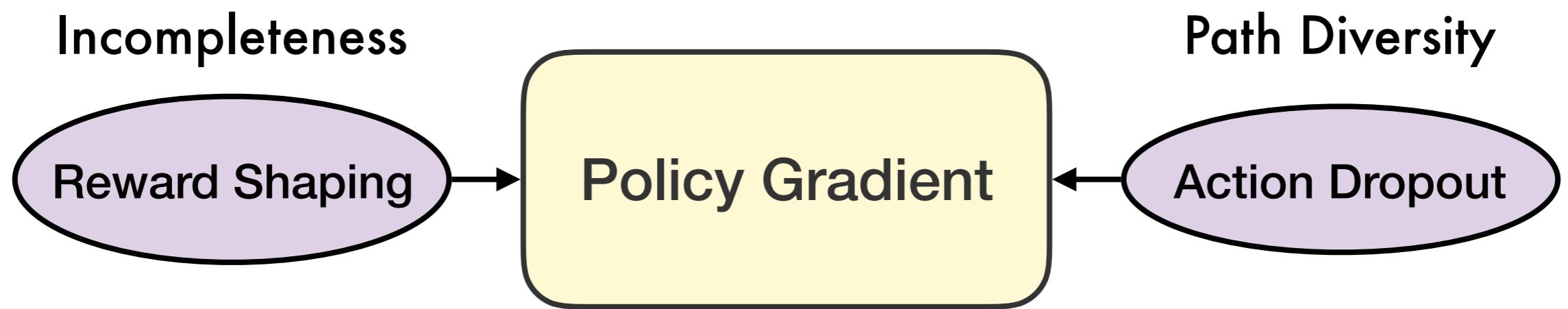
Path Diversity

False positive
(spurious) paths ① ②

Overfit to the
spurious paths



Proposed Solutions



Reinforcement Learning Framework

Environment

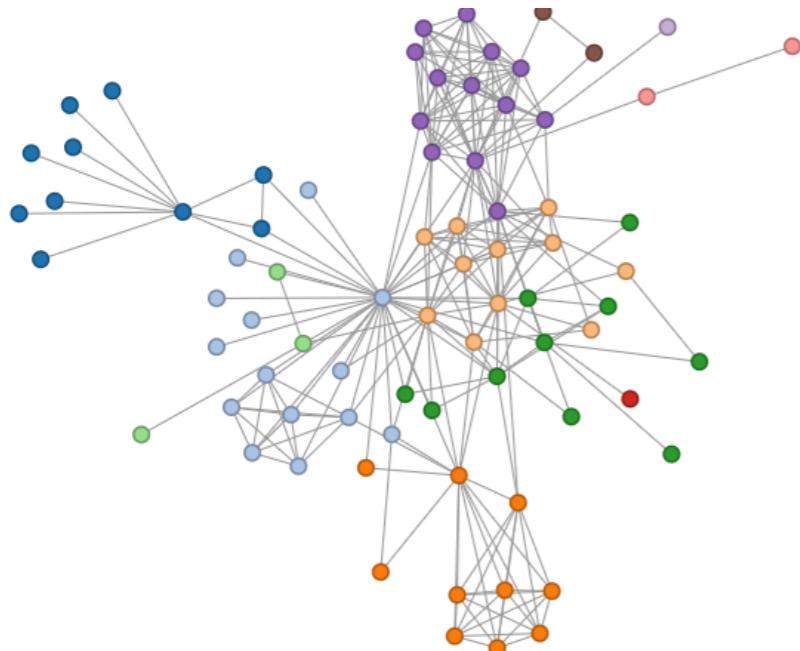
State

Action

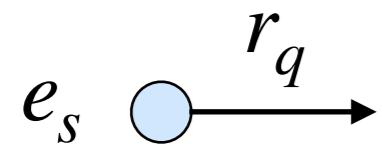
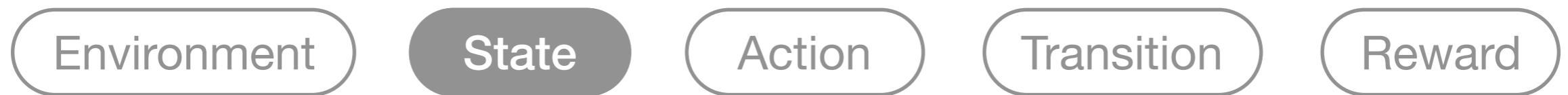
Transition

Reward

$$e_s \quad r_q \rightarrow$$

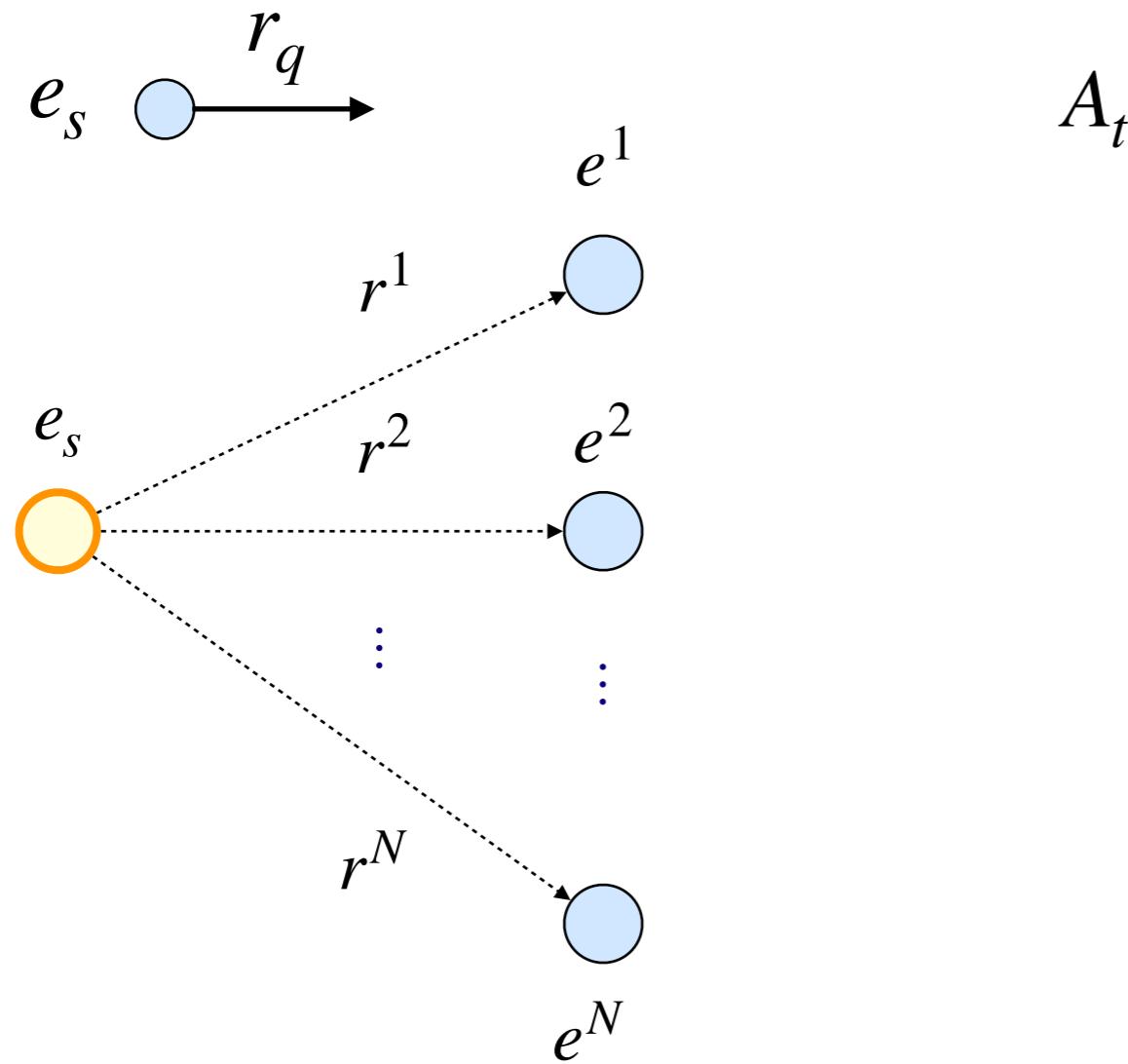


Reinforcement Learning Framework



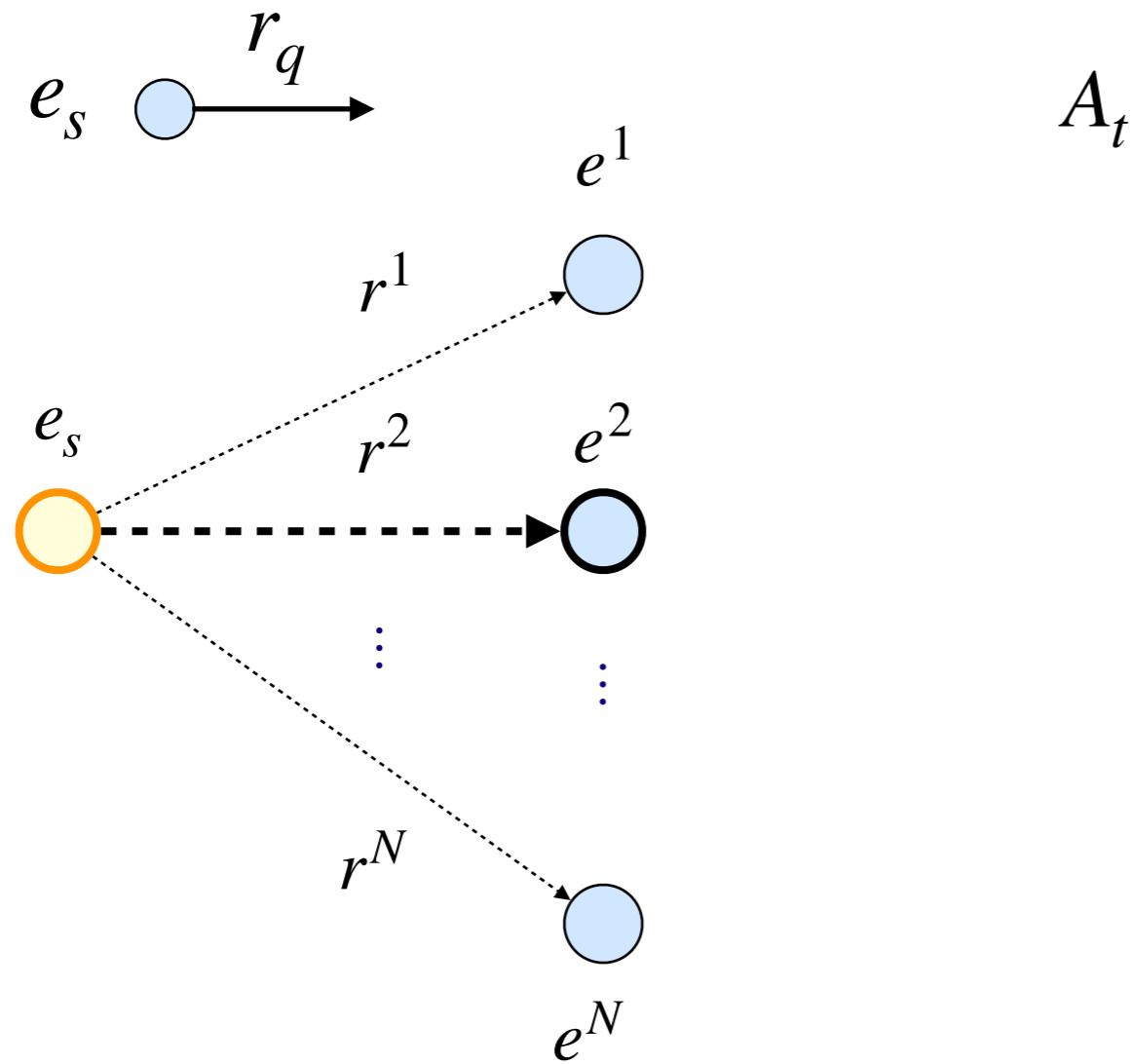
Reinforcement Learning Framework

Environment State **Action** Transition Reward

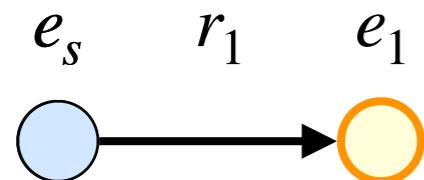
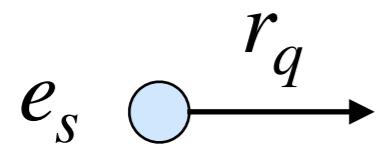
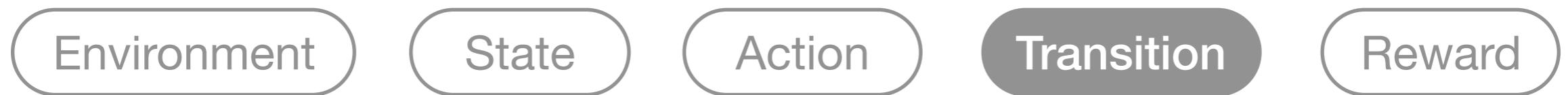


Reinforcement Learning Framework

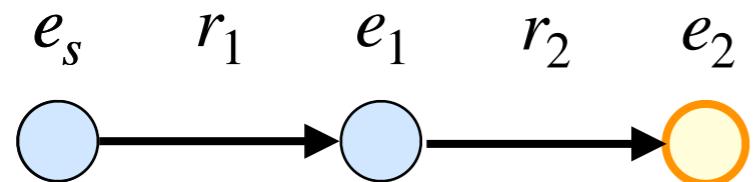
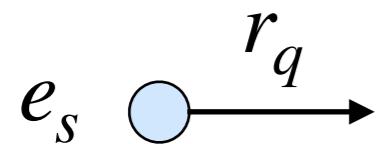
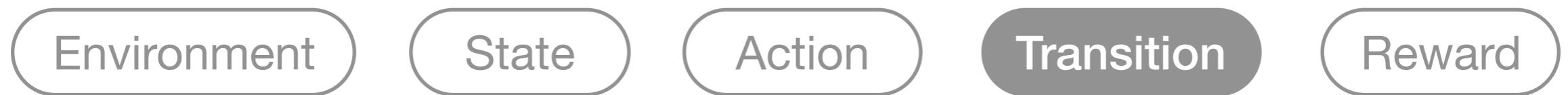
Environment State **Action** Transition Reward



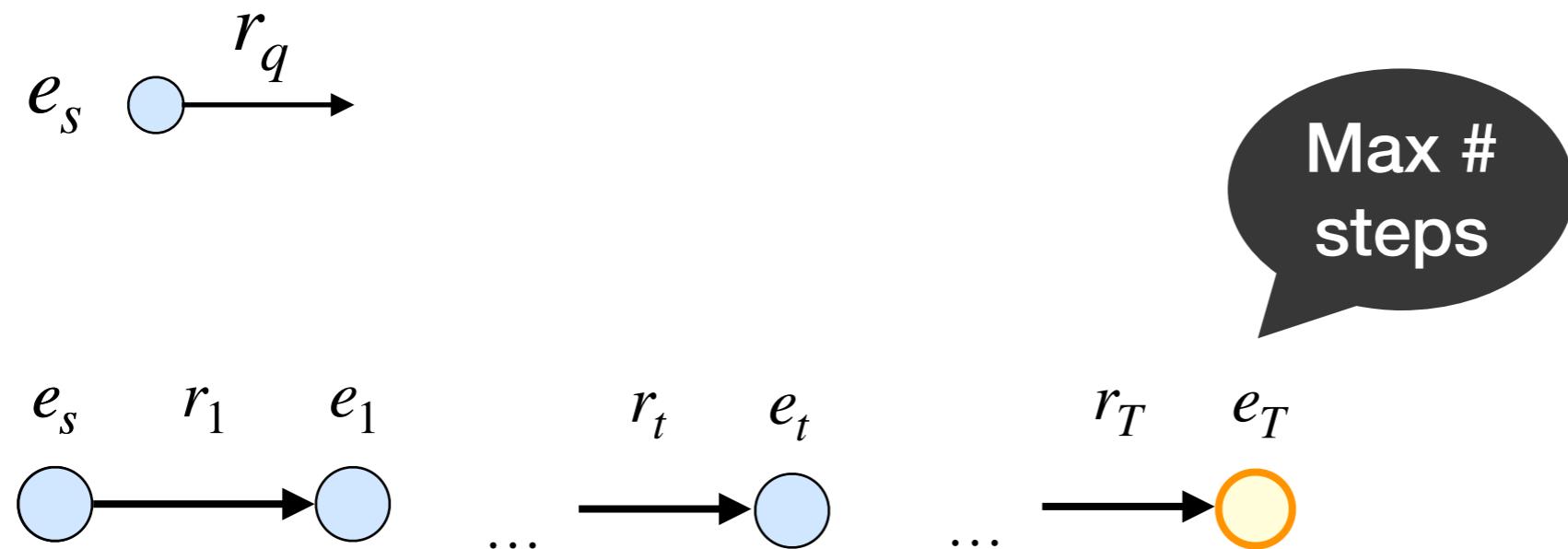
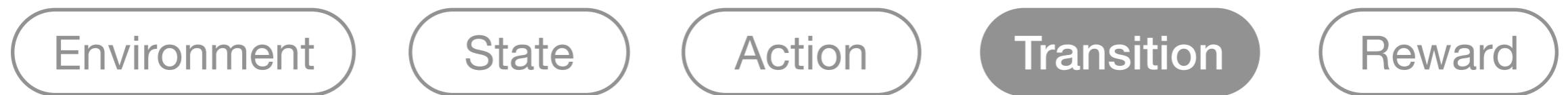
Reinforcement Learning Framework



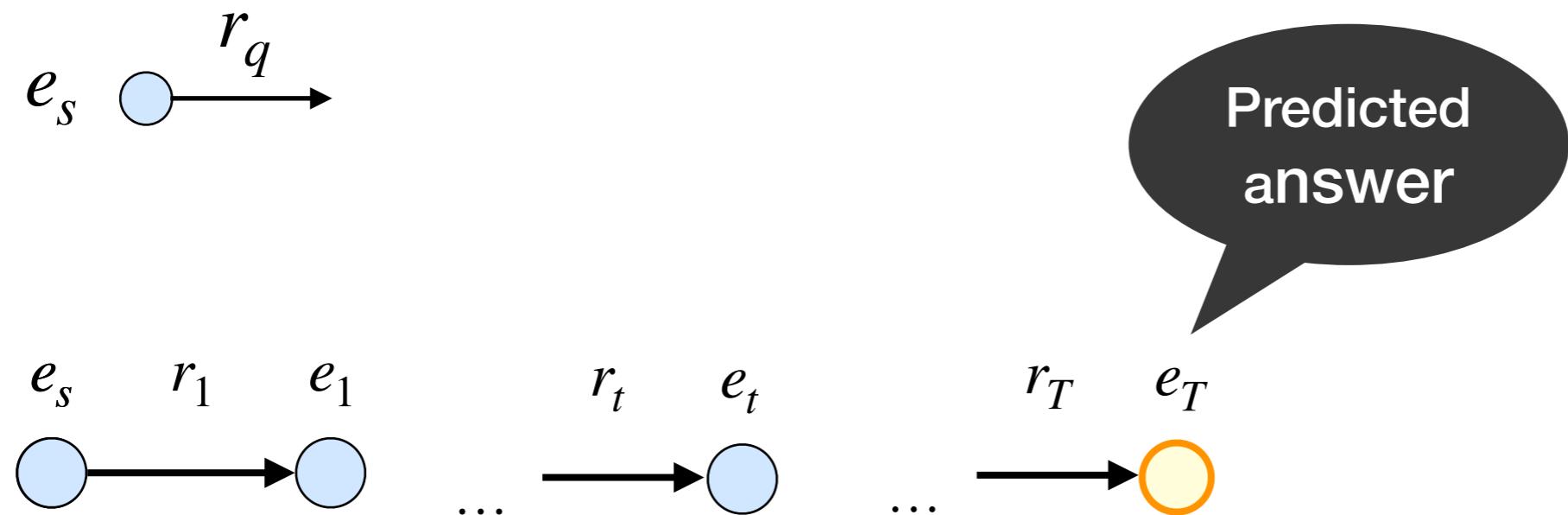
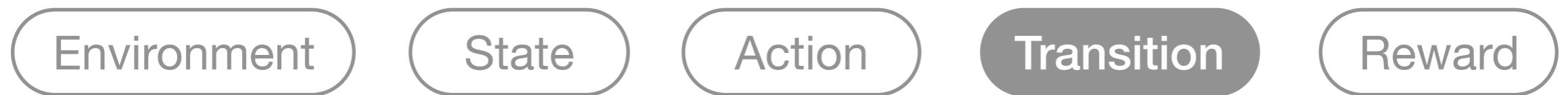
Reinforcement Learning Framework



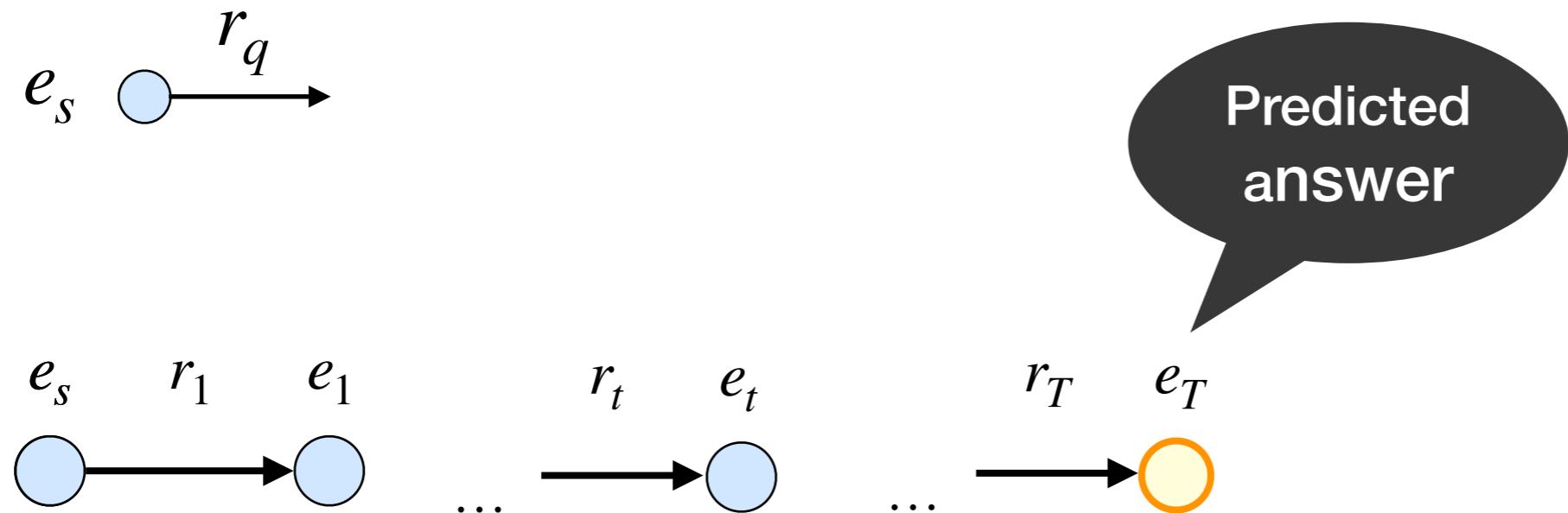
Reinforcement Learning Framework



Reinforcement Learning Framework

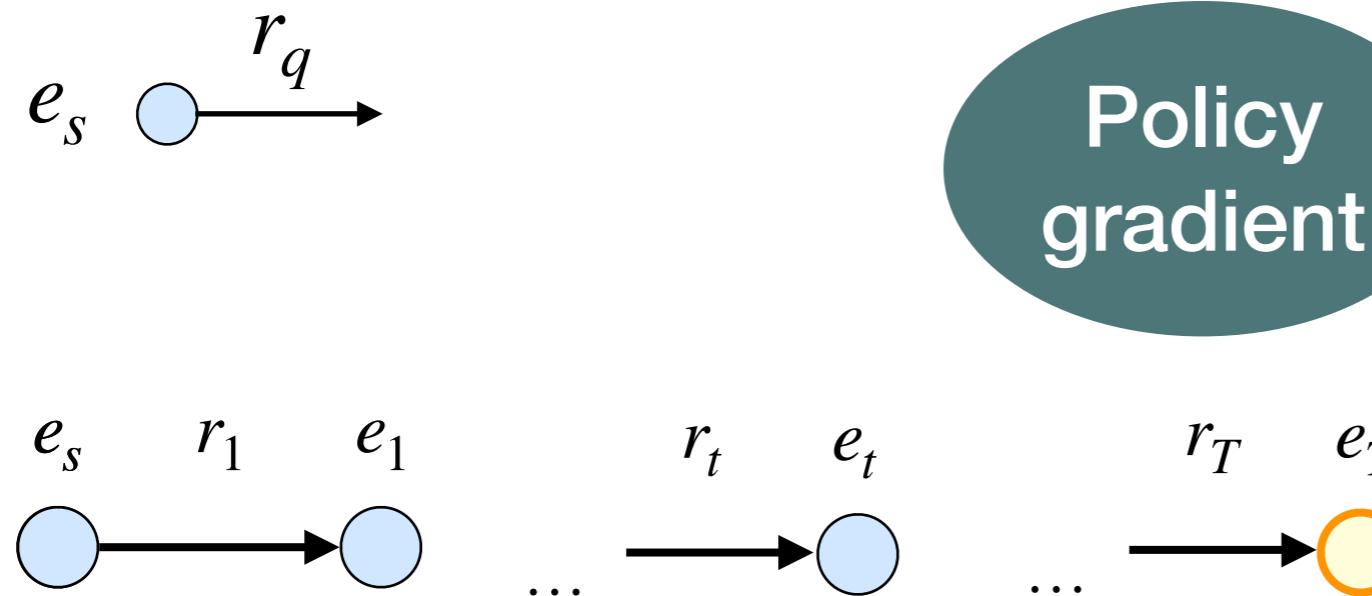
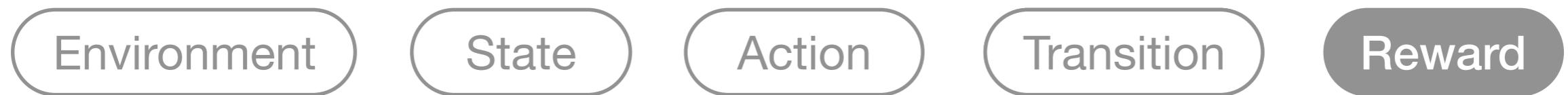


Reinforcement Learning Framework



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

Reinforcement Learning Framework



Policy
gradient

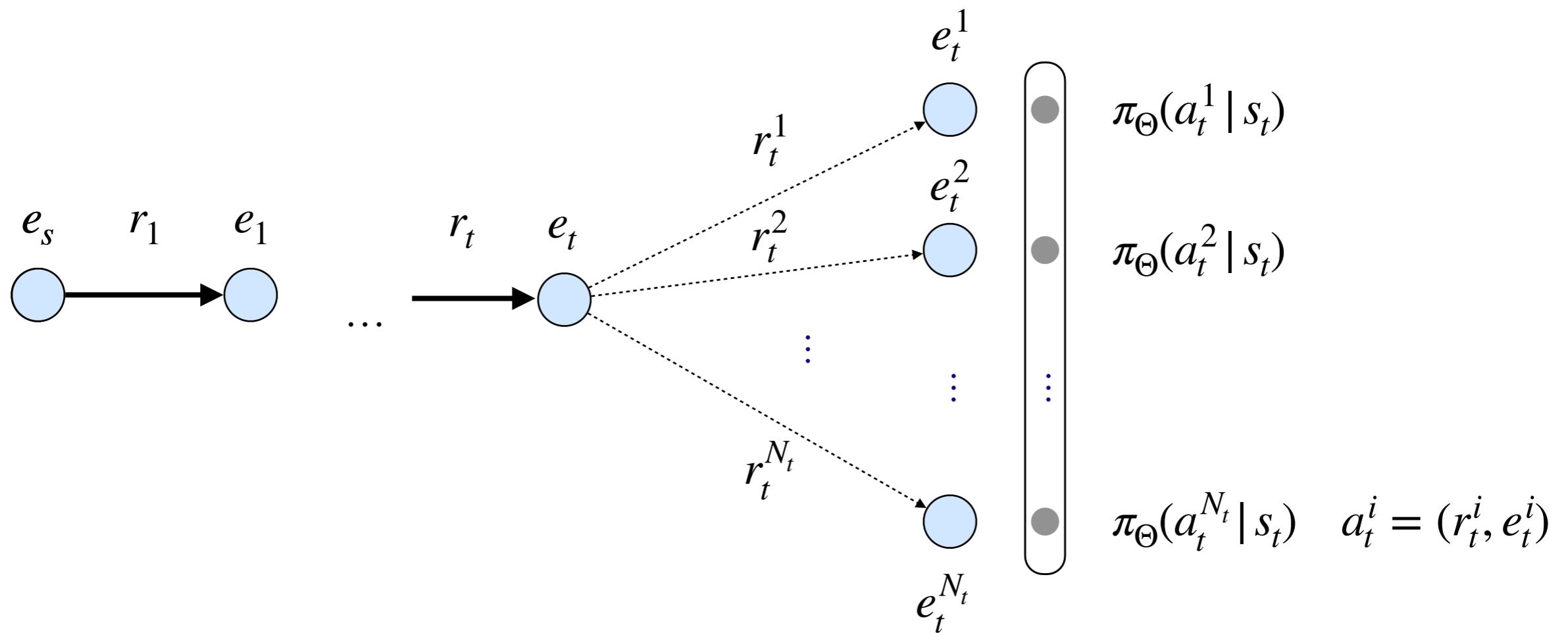
Learn **which**
action to choose
given a state

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

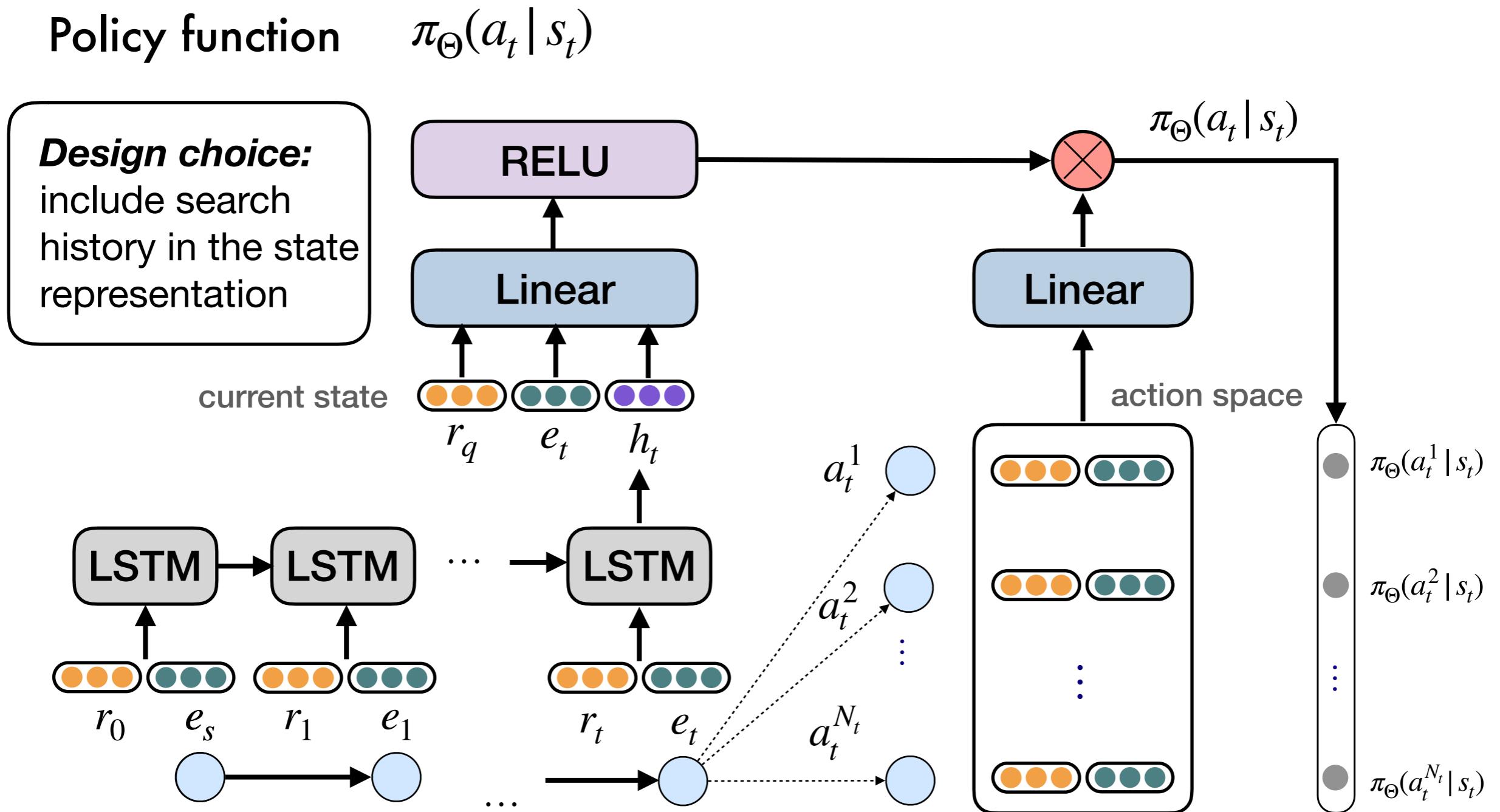
Policy Gradient

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

Probability of choosing
an action given the
current state



Policy Gradient

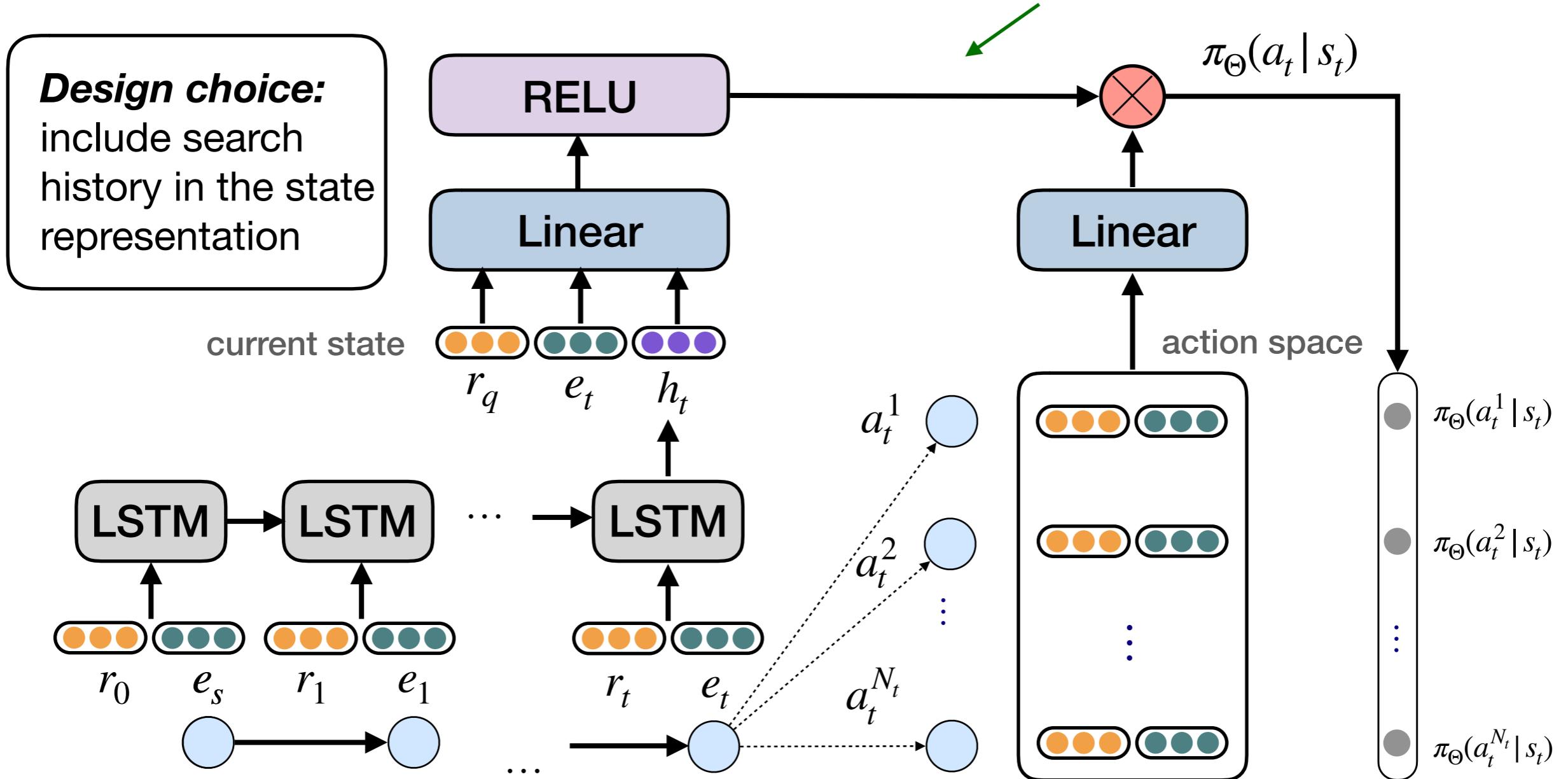


Policy Gradient

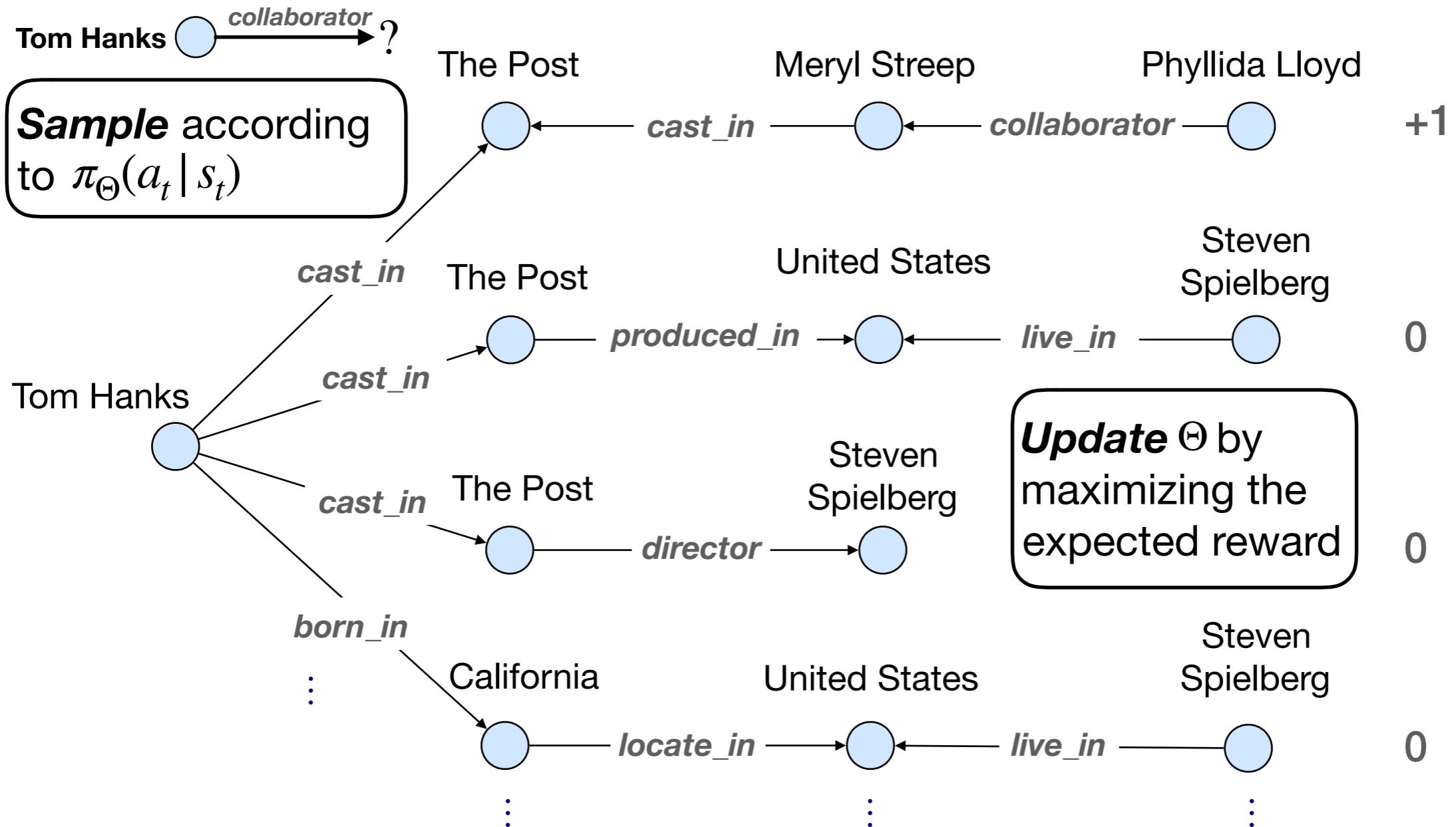
Policy function

$$\pi_{\Theta}(a_t | s_t)$$

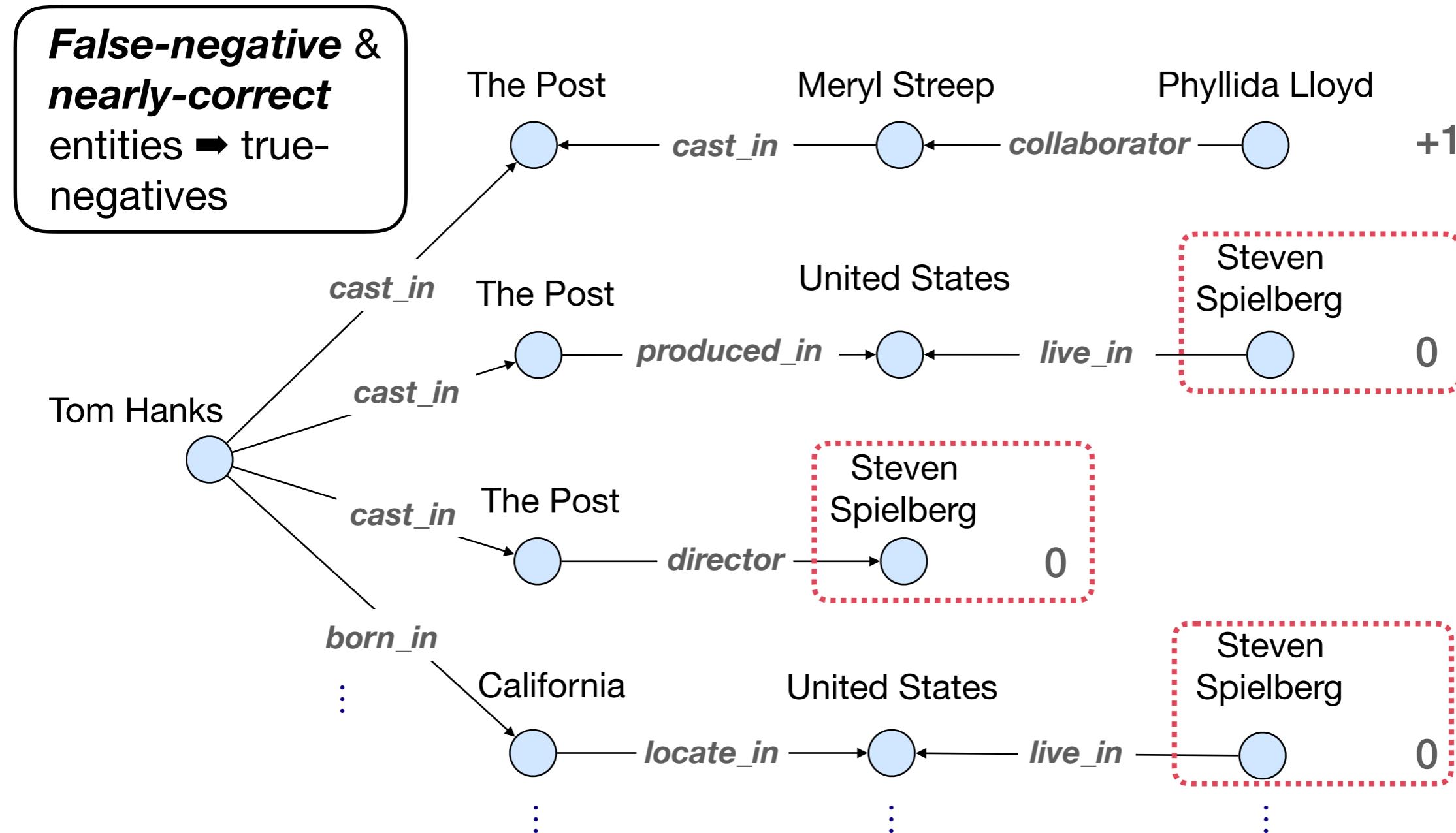
Our model extensions are applicable
to any parameterization of π_{Θ}



REINFORCE Training



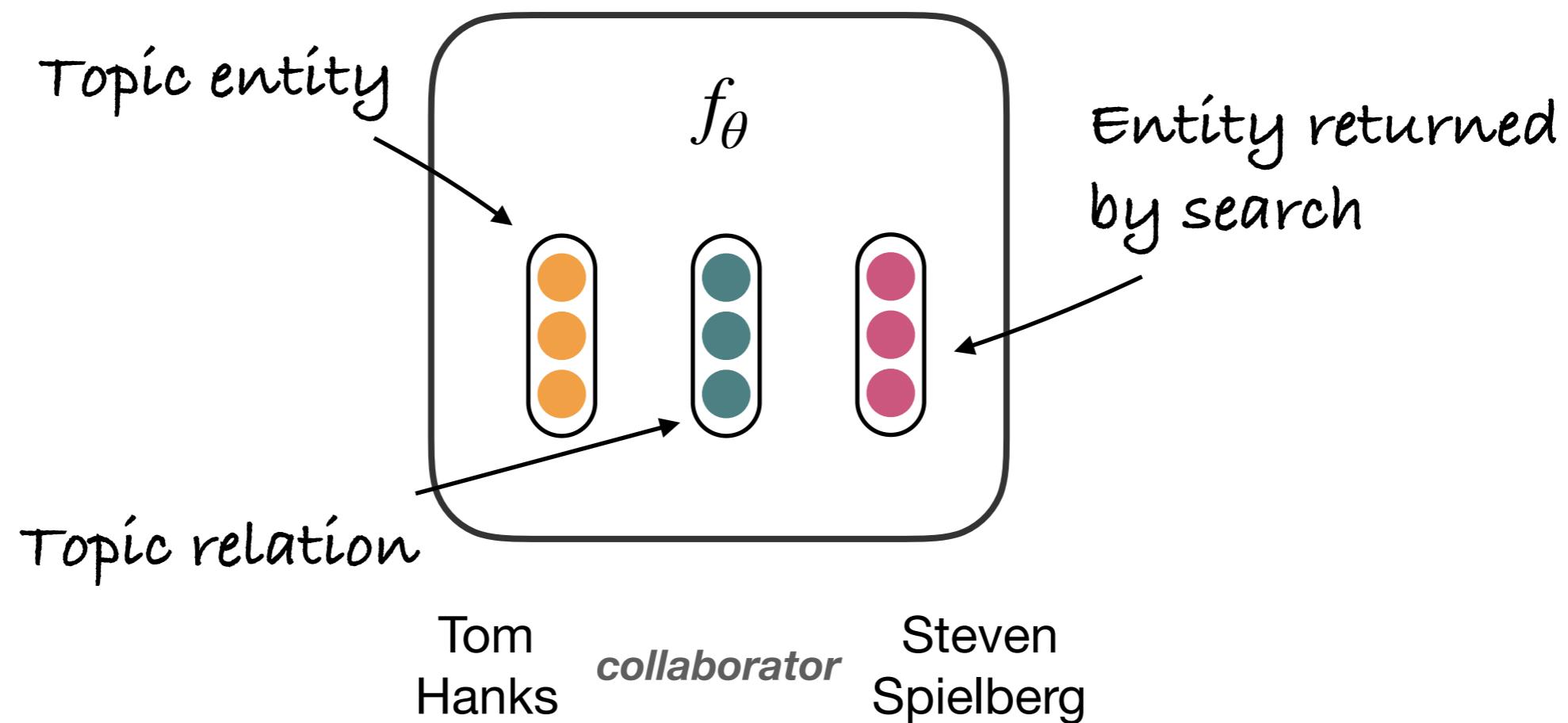
REINFORCE Training



Reward Shaping

Unobserved facts

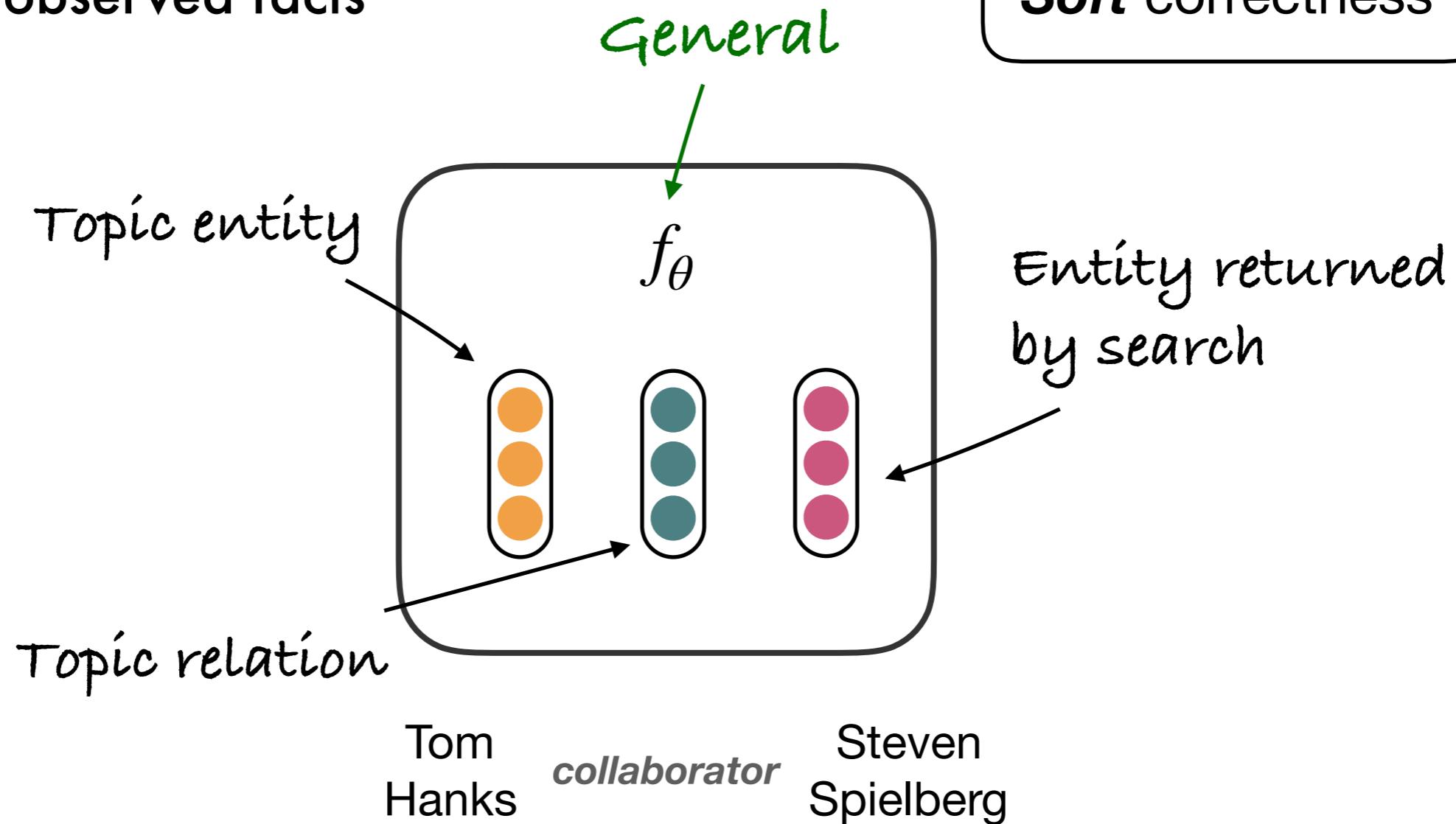
Soft correctness



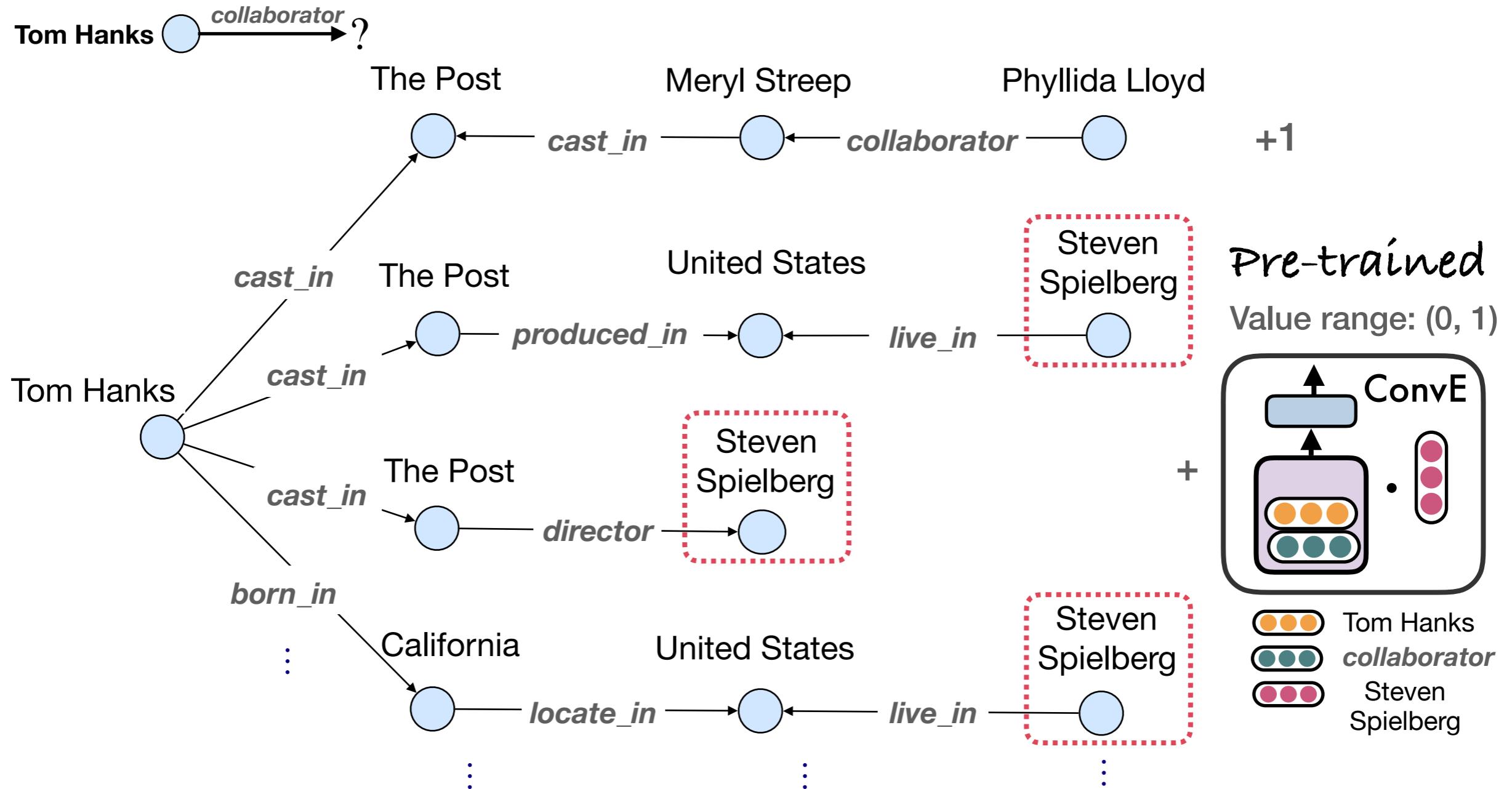
Reward Shaping

Unobserved facts

Soft correctness



Reward Shaping



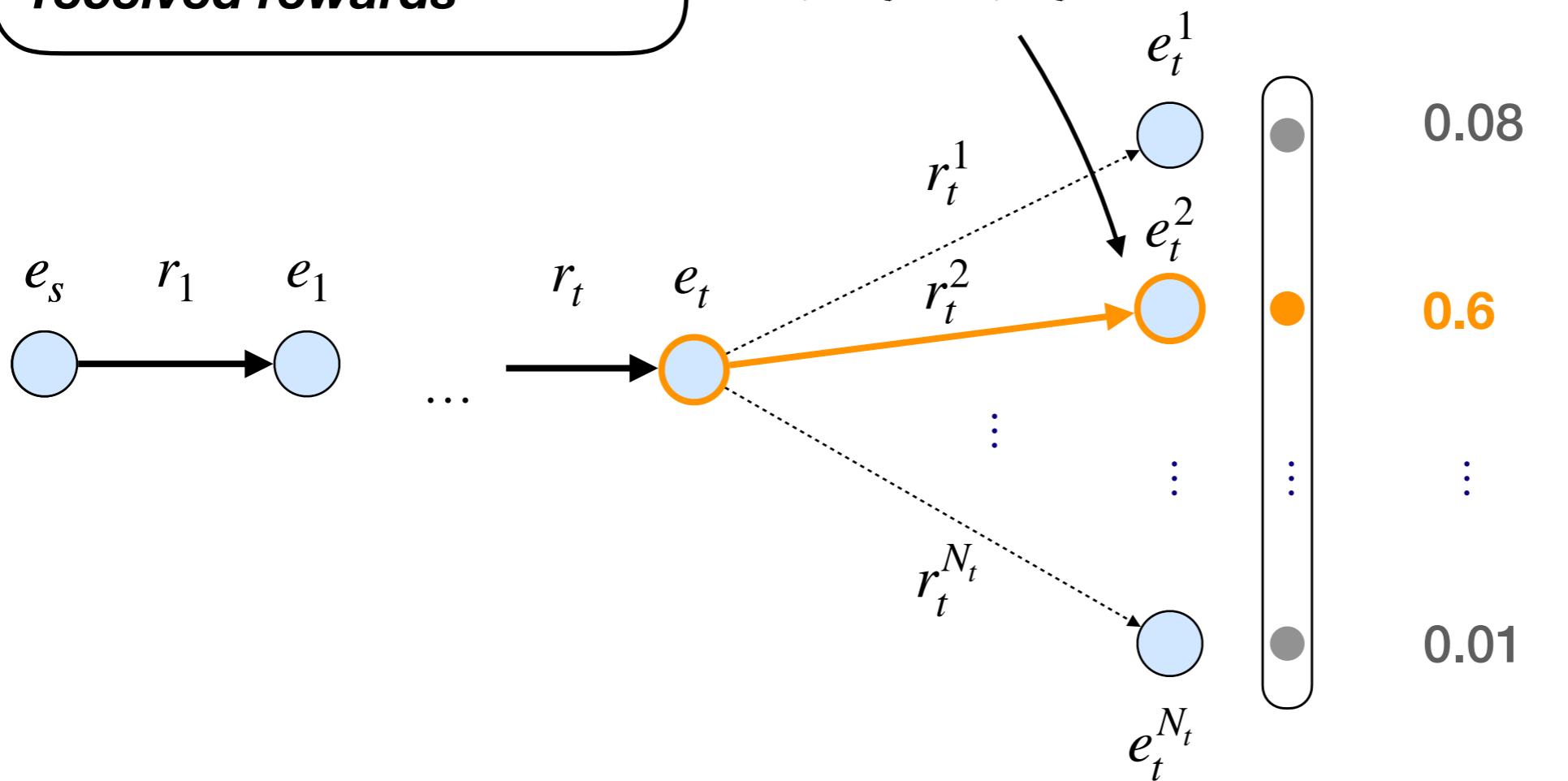
Action Dropout

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

Action Dropout

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

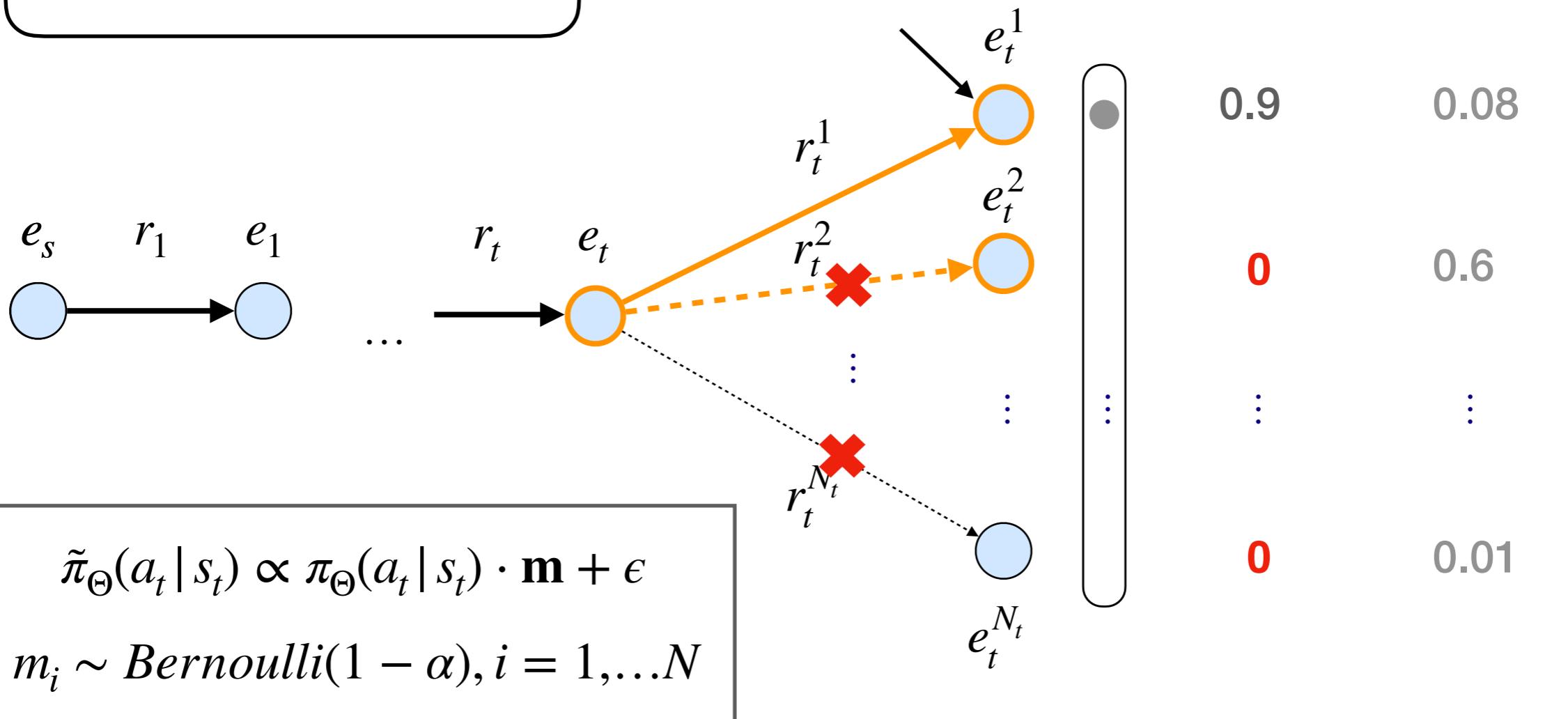
More likely
to be chosen



Action Dropout

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

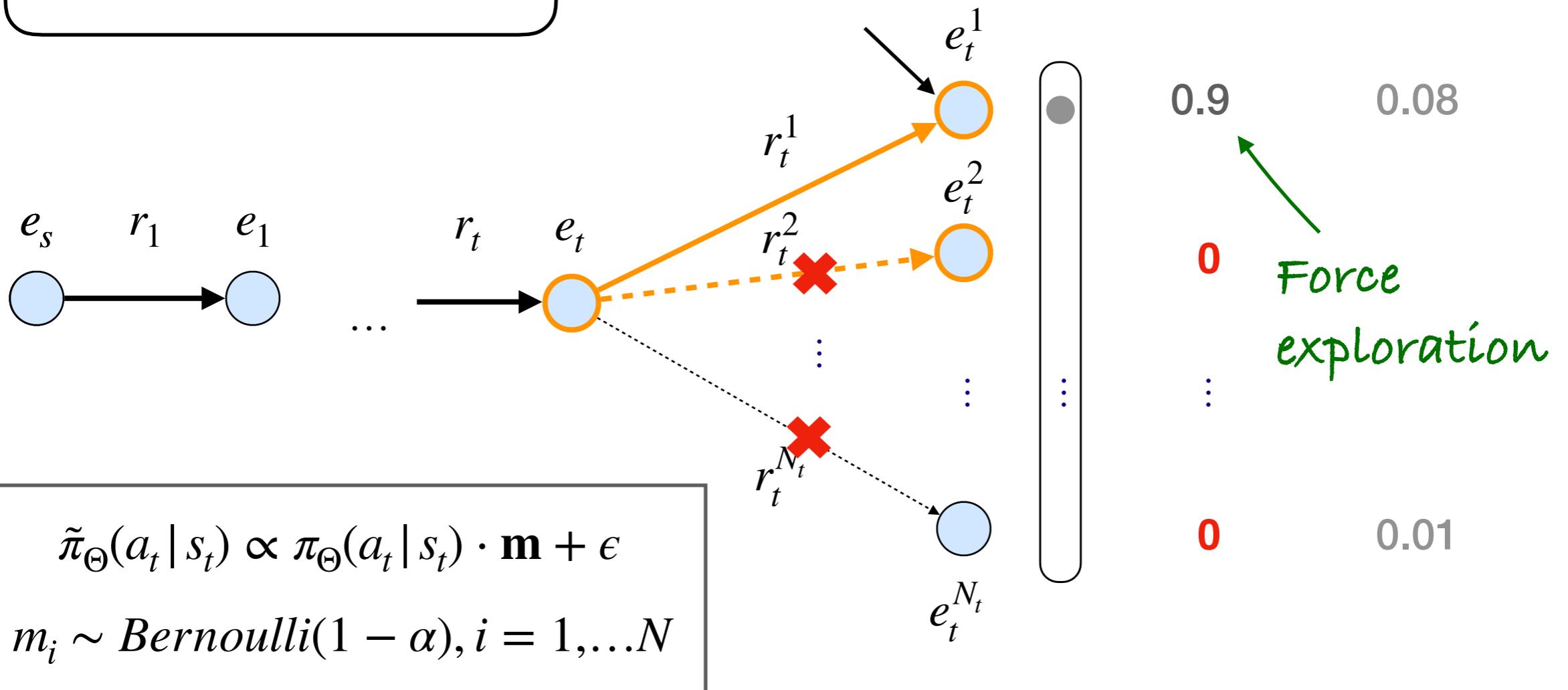
More likely
to be chosen



Action Dropout

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

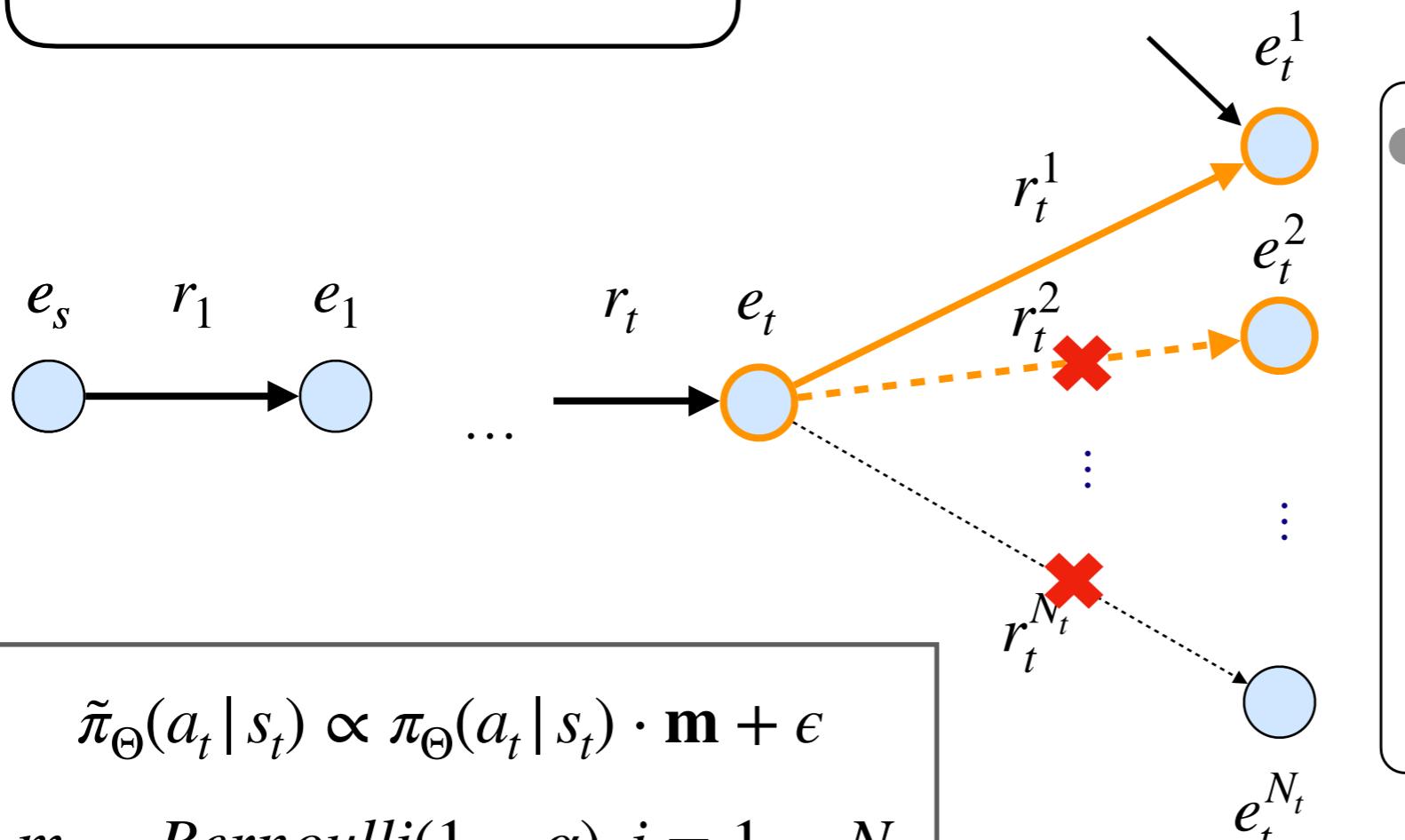
More likely
to be chosen



Action Dropout

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

More likely
to be chosen

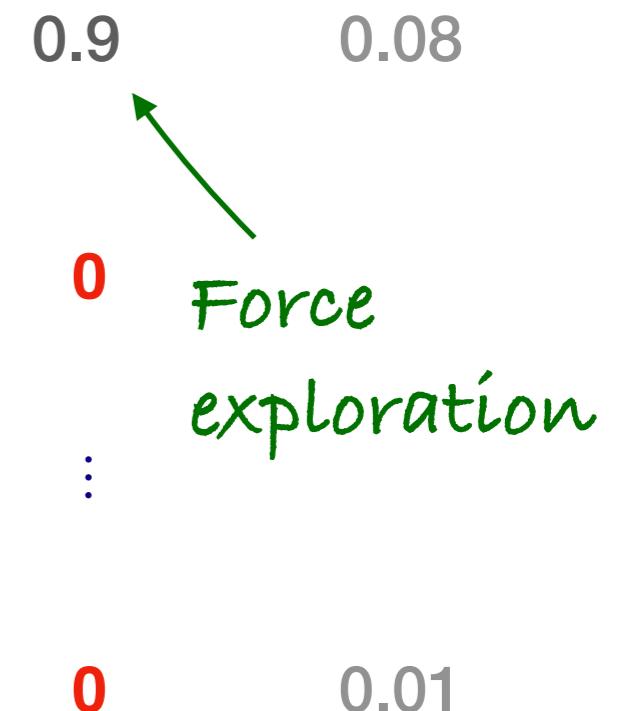


$$\tilde{\pi}_{\Theta}(a_t | s_t) \propto \pi_{\Theta}(a_t | s_t) \cdot \mathbf{m} + \epsilon$$

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

Up to $\times 8$ # path
traversed

$$\tilde{\pi}_{\Theta}(a_t^i | s_t) \quad \pi_{\Theta}(a_t^i | s_t)$$



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

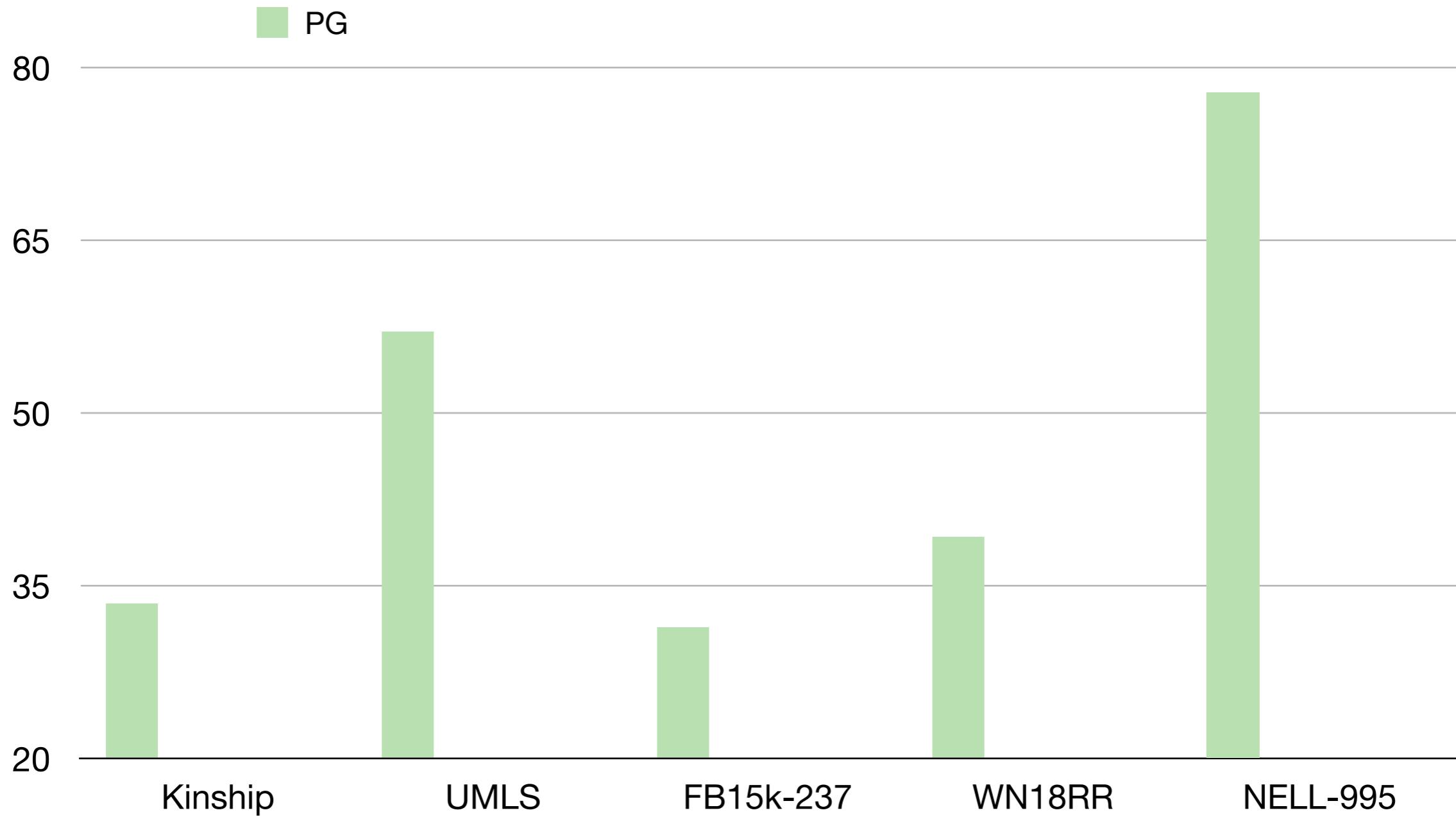


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

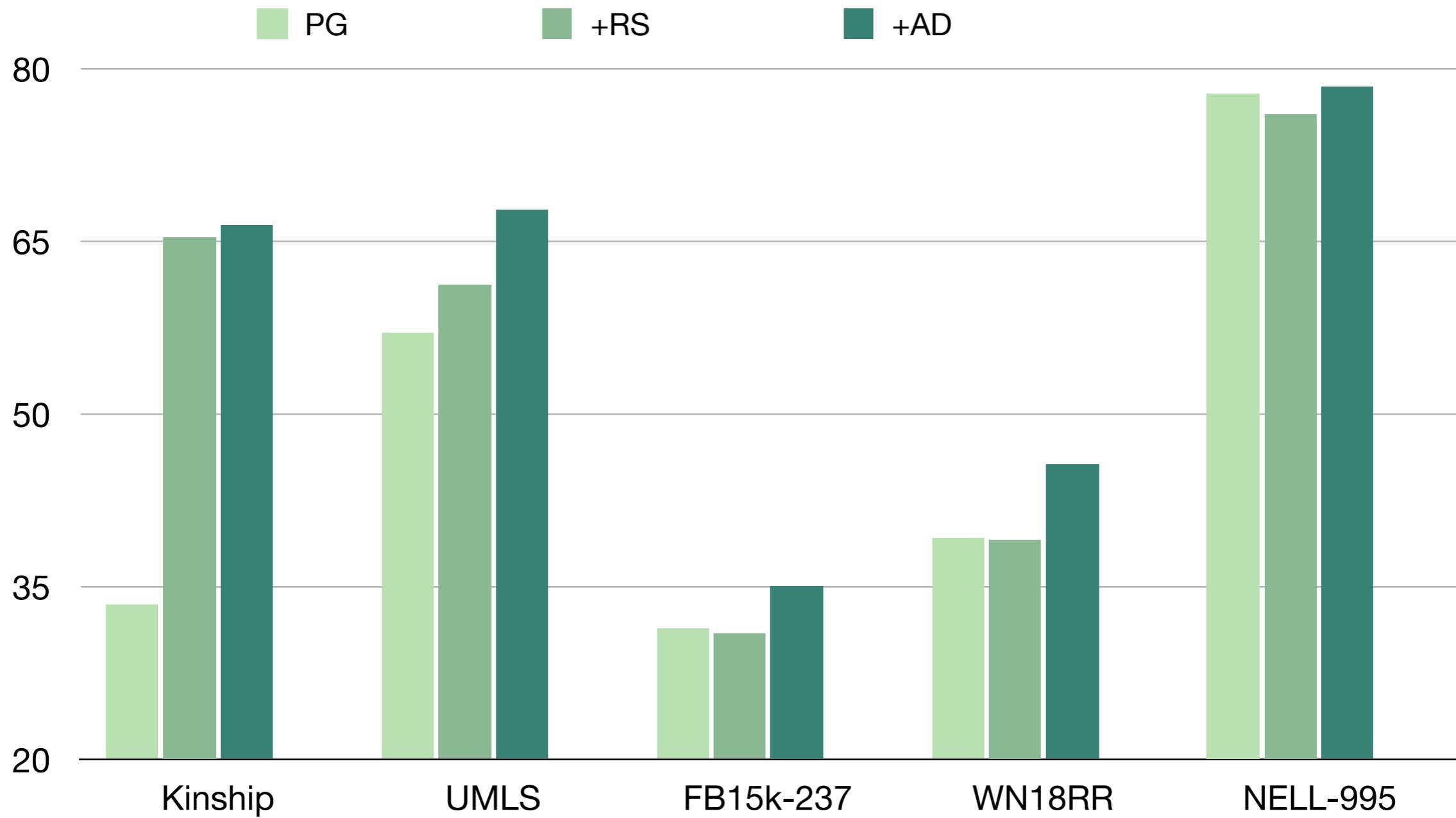


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

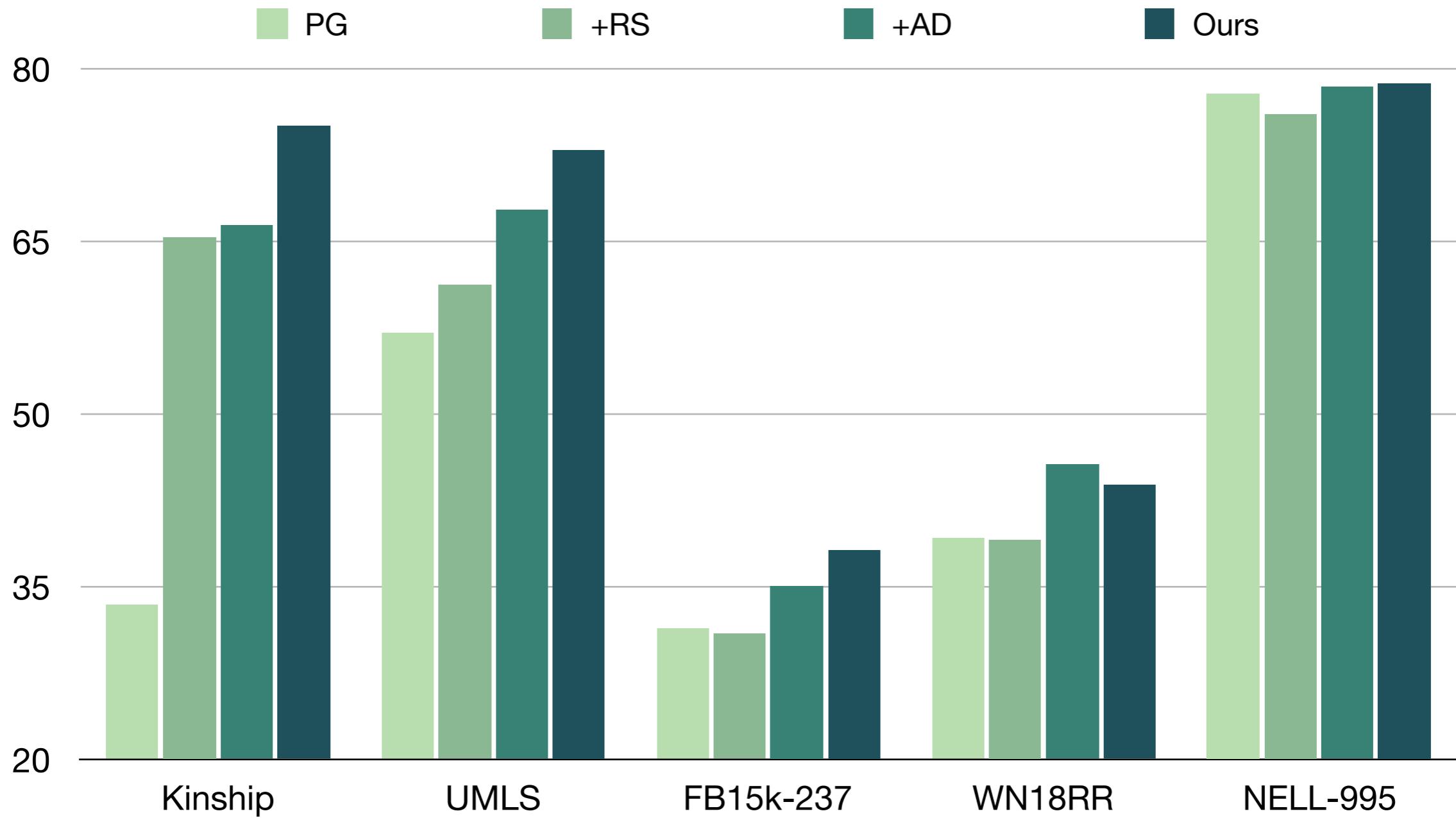


Fig 2. Dev set MRR (x100) comparison

Ablation Studies

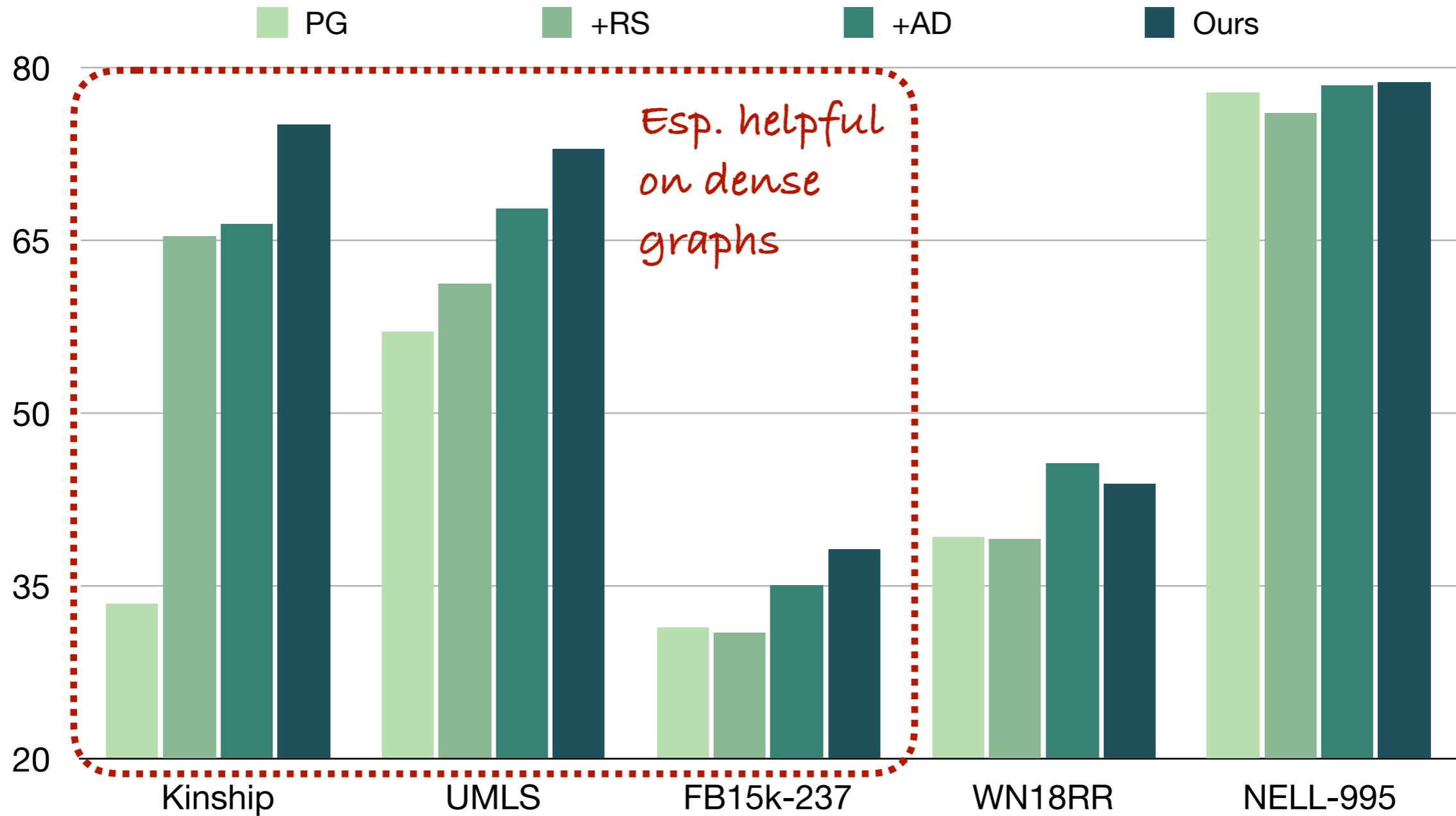


Fig 2. Dev set MRR (x100) comparison

Main Results

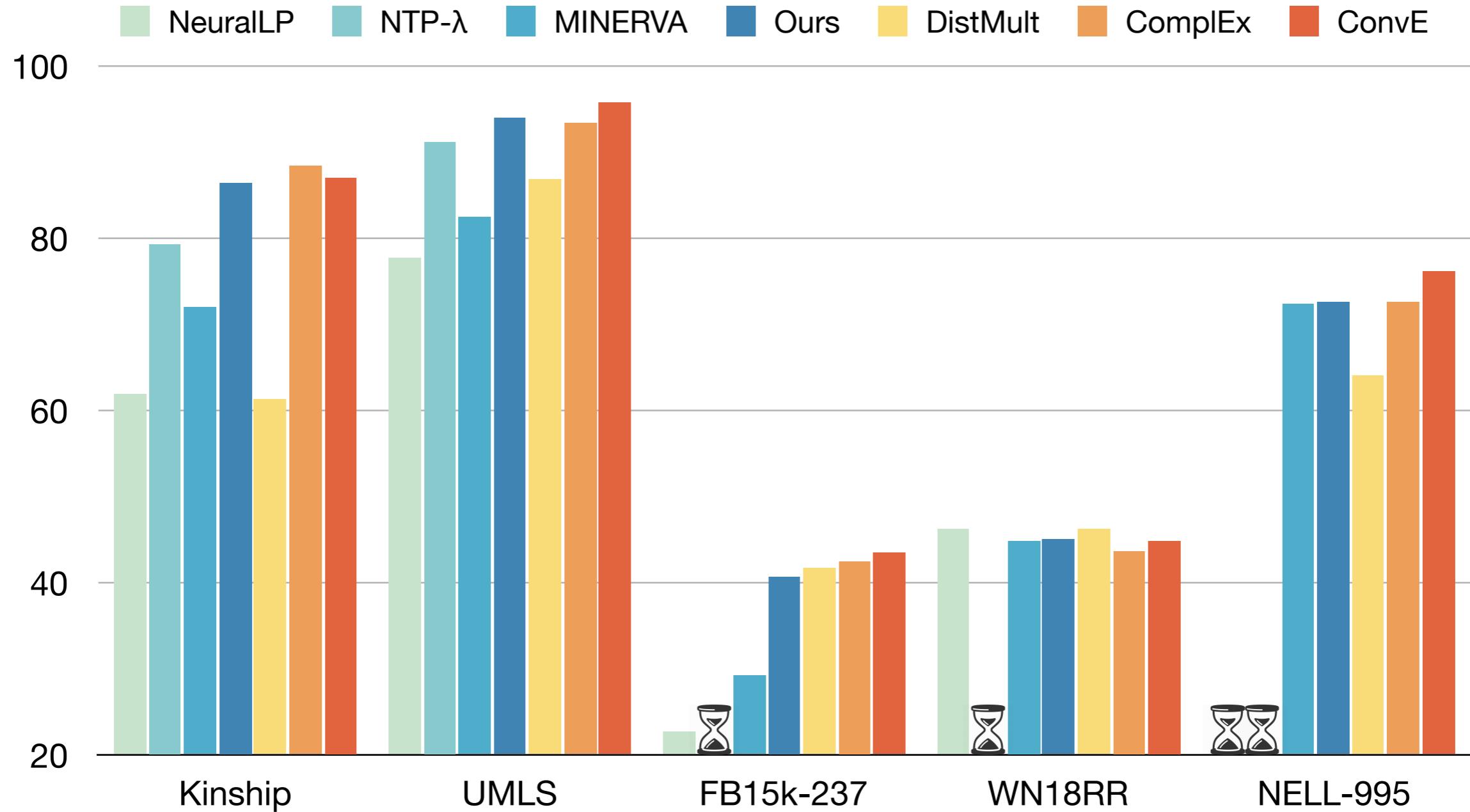


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

Main Results

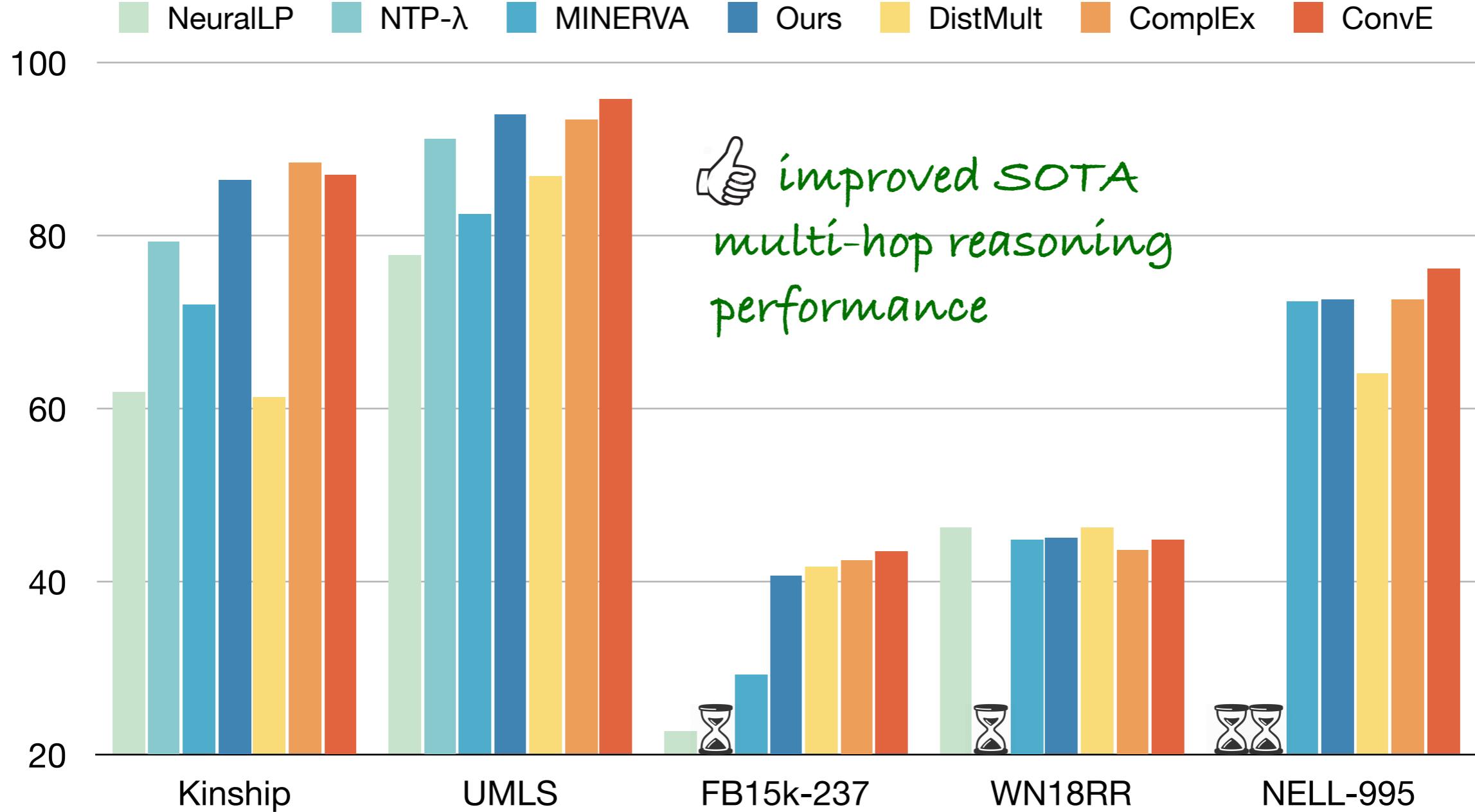


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

Main Results

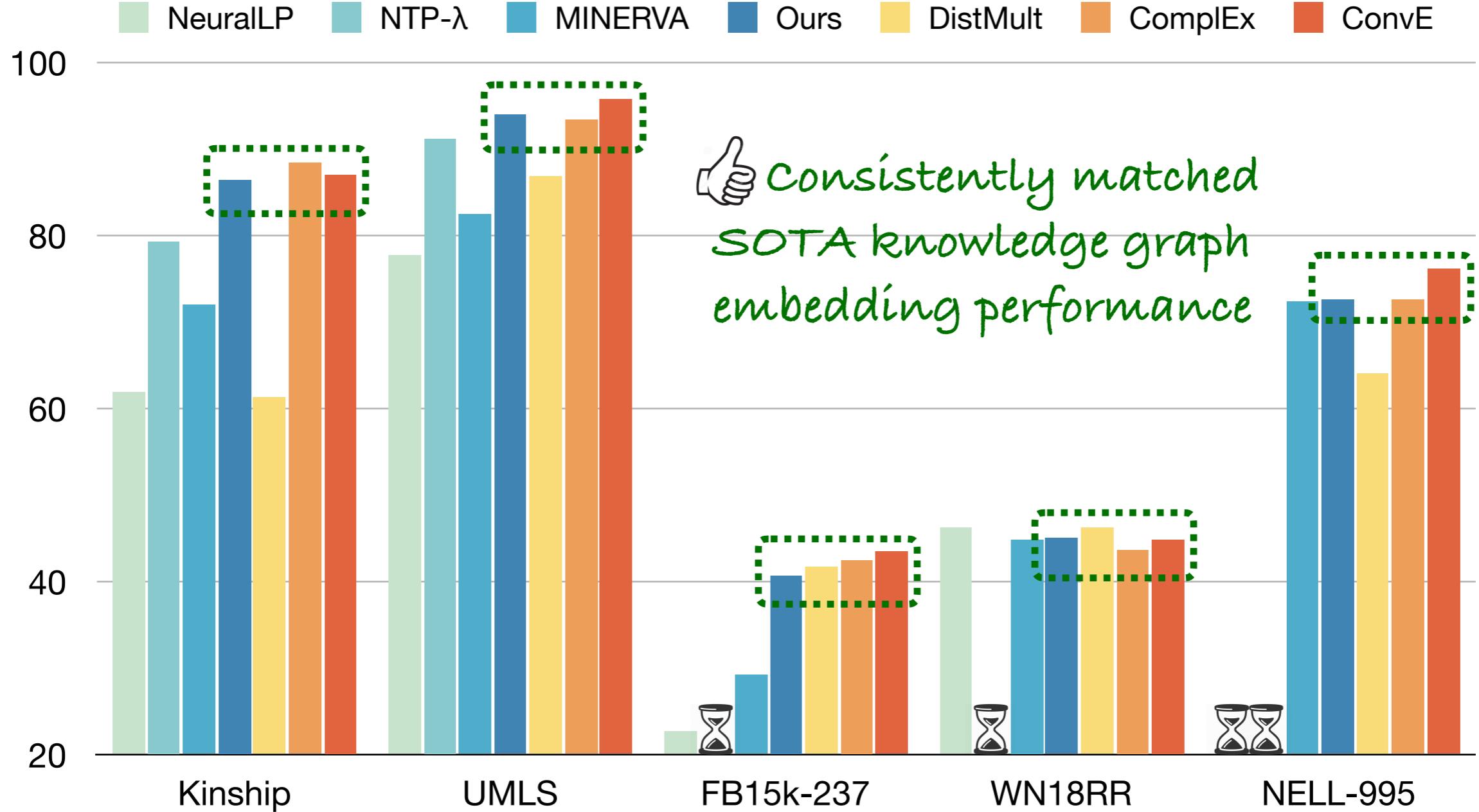
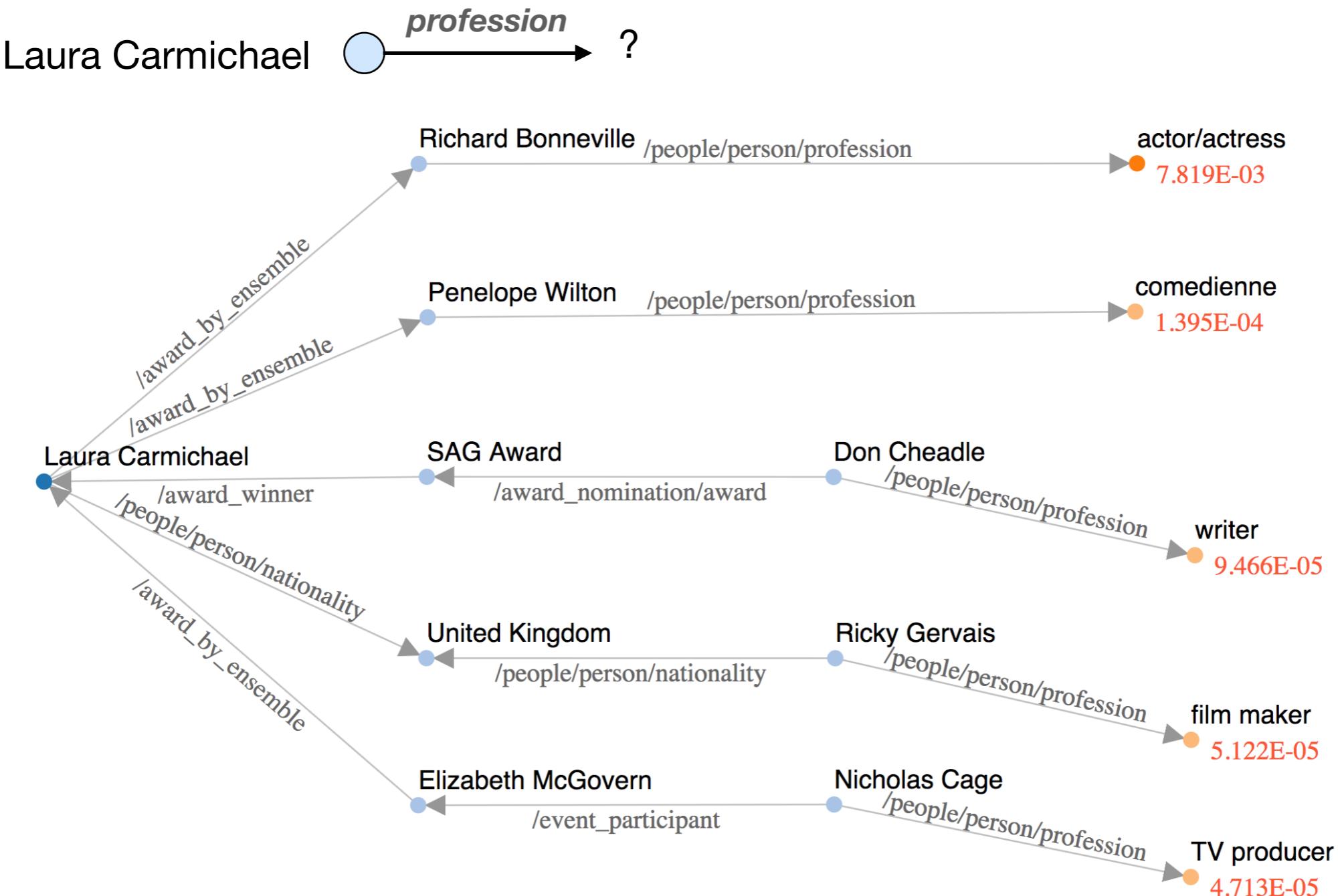
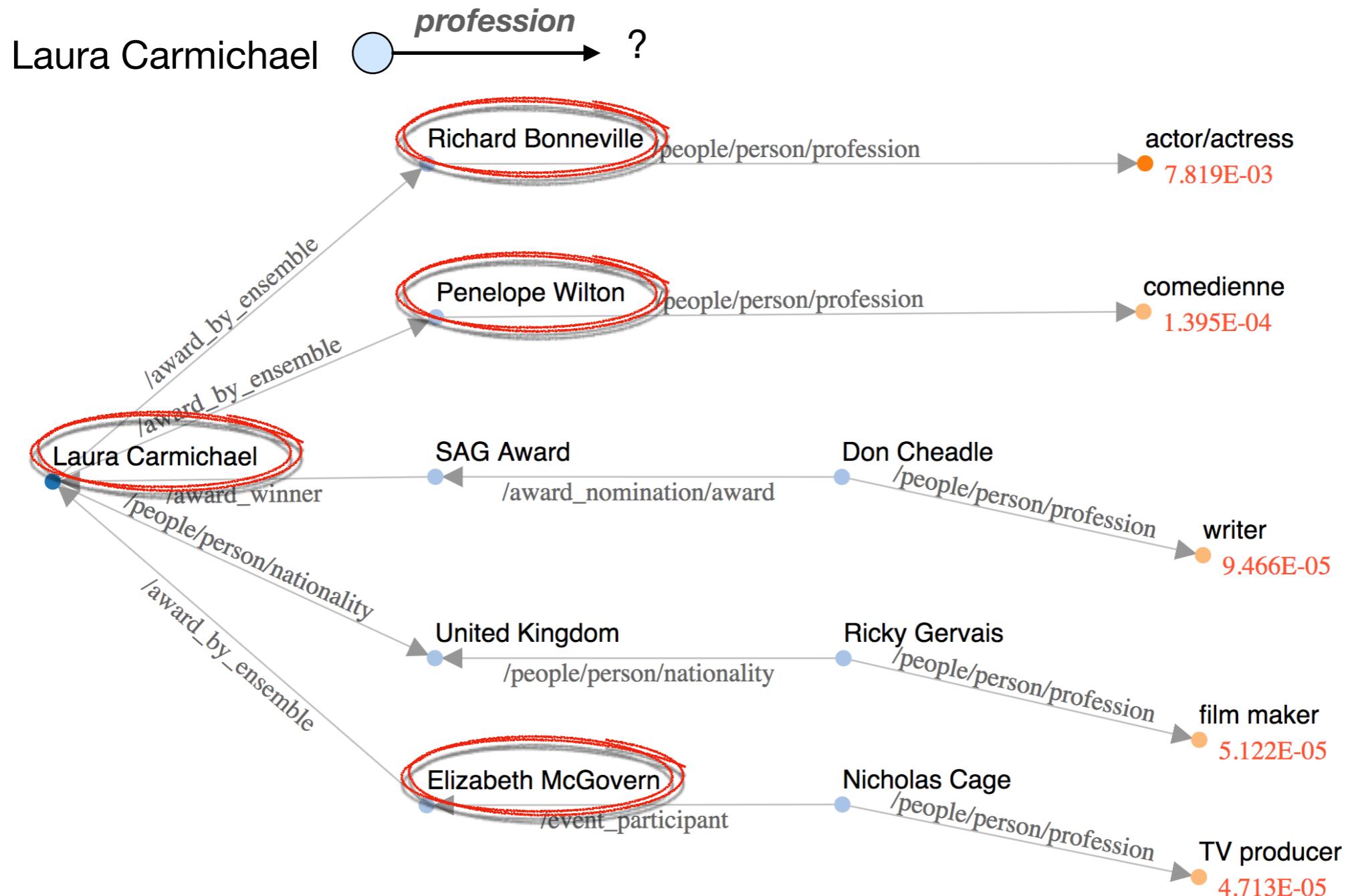


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

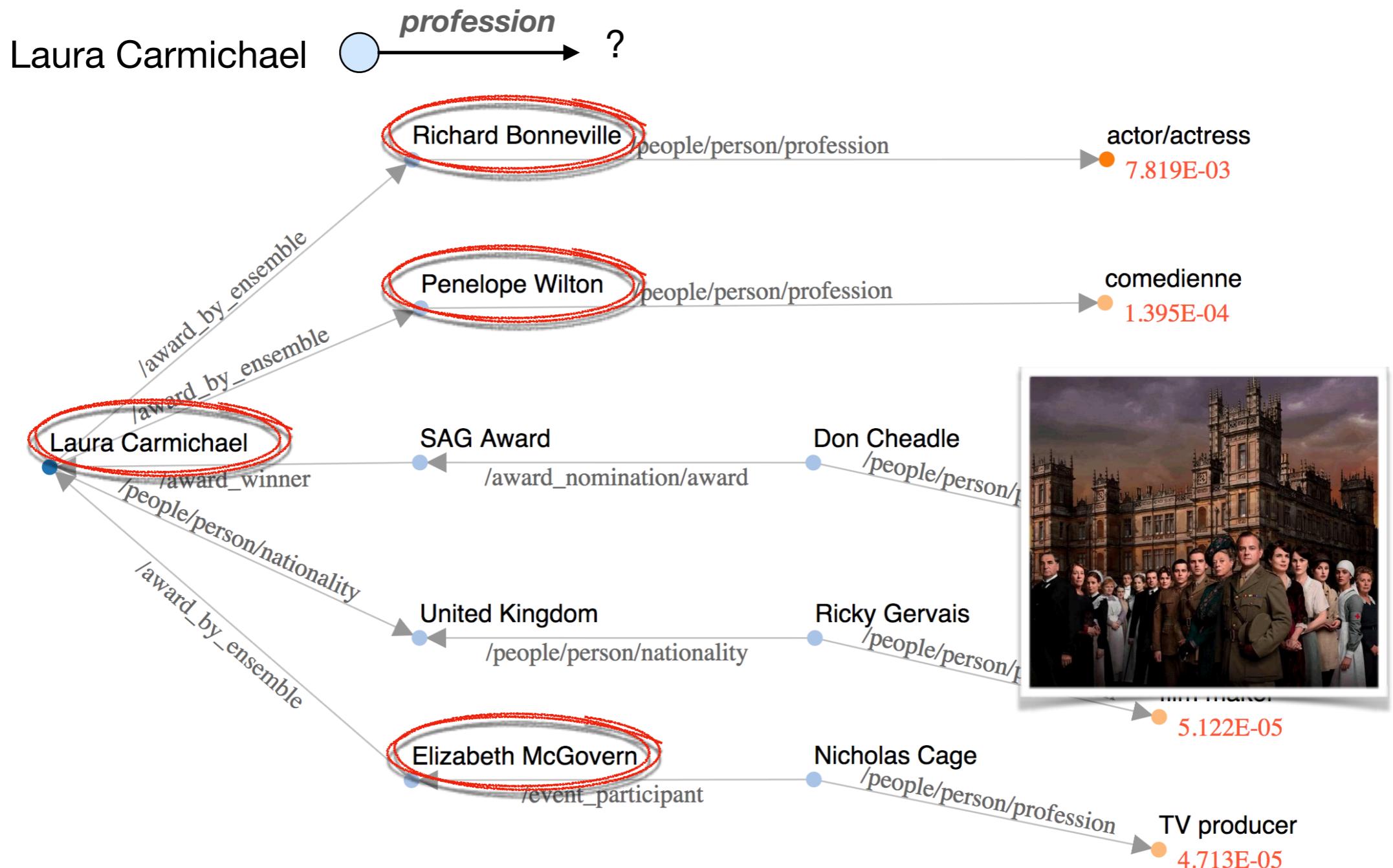
Interpretable Results



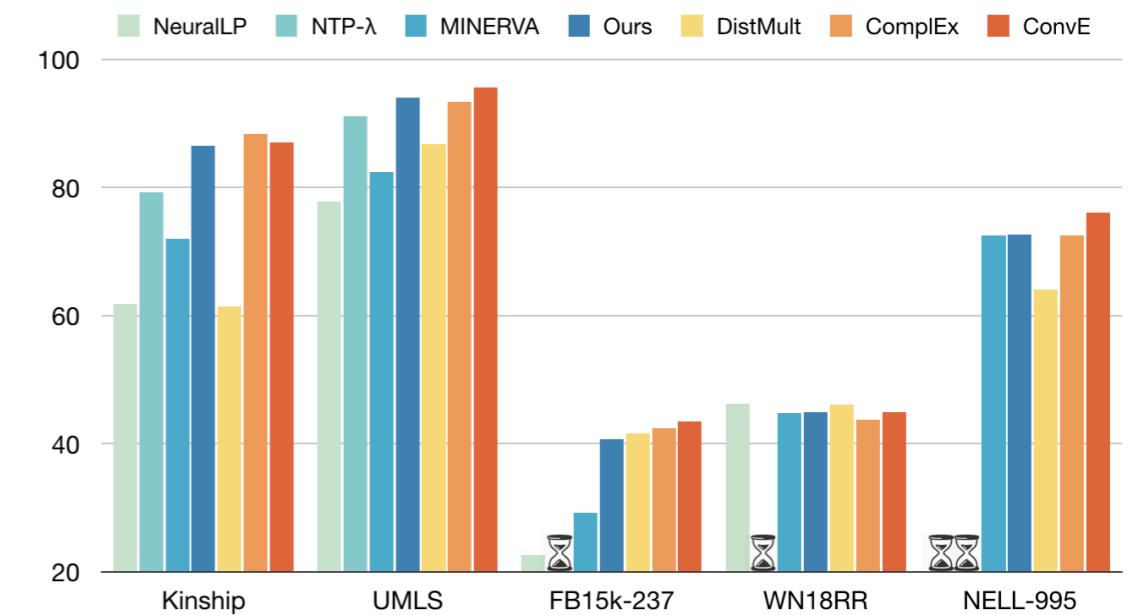
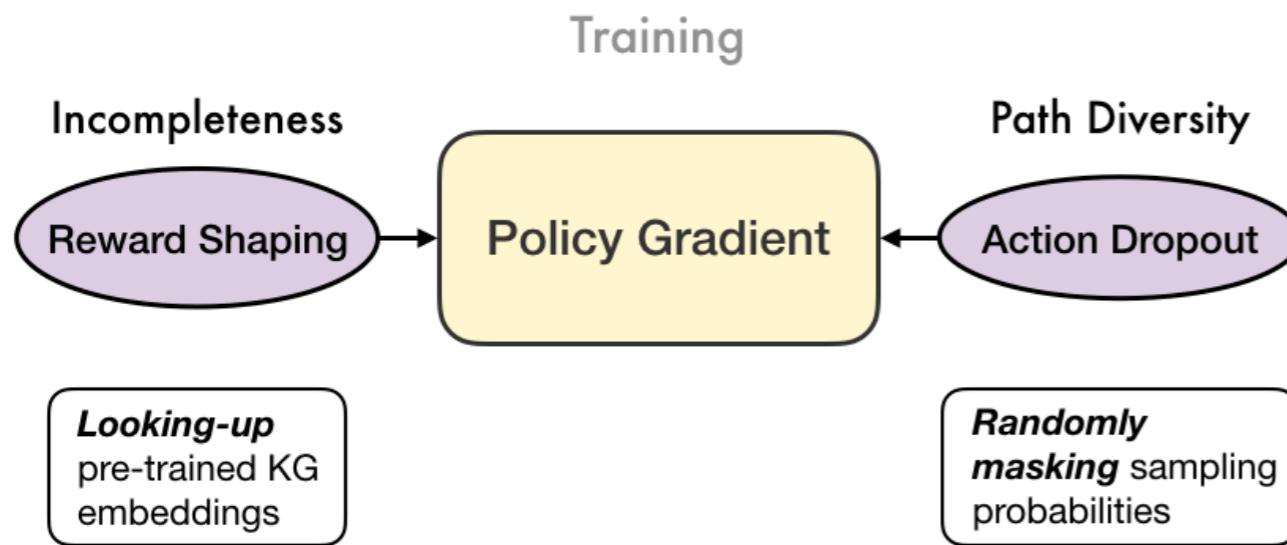
Interpretable Results



Interpretable Results



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

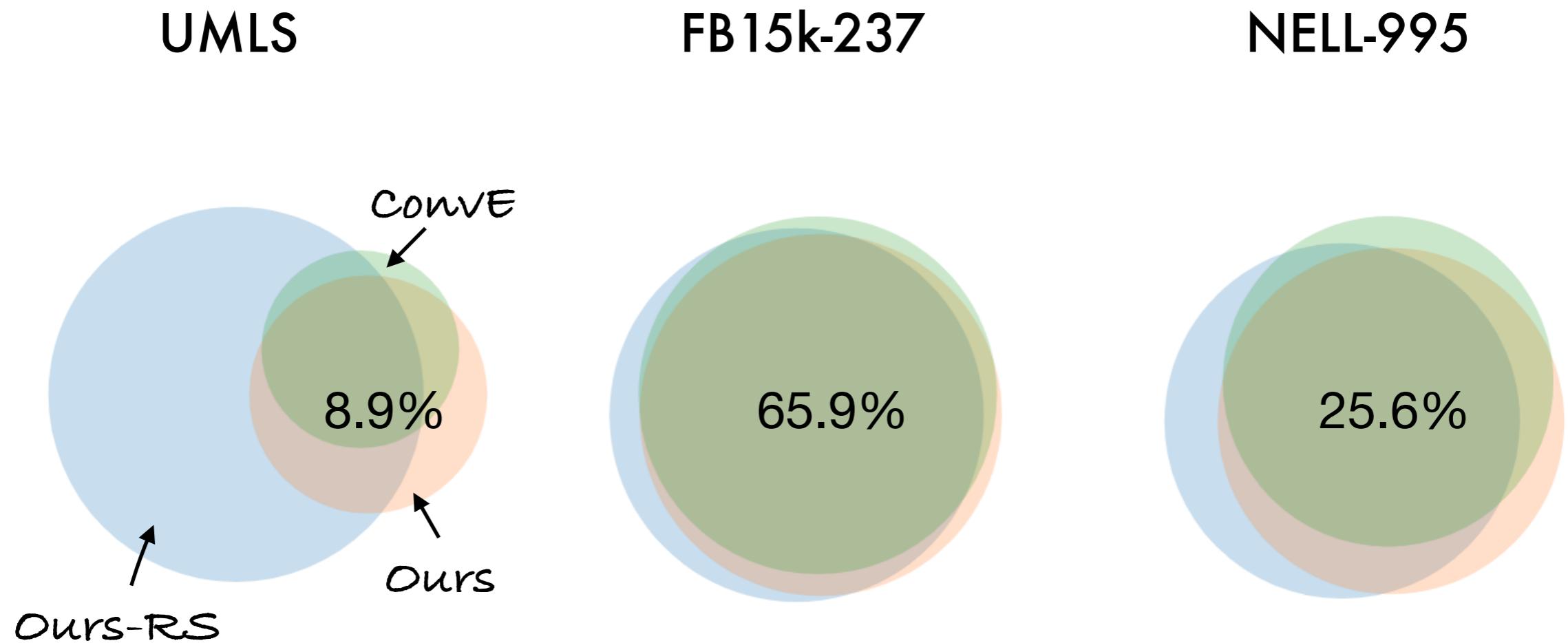


Fig 4. Dev set top-1 prediction error overlap of ConvE, Ours and Ours-RS. The absolute error rate of Ours is shown.

BK II - Challenges

Incompleteness

$\approx 30\%$ false negative feedback

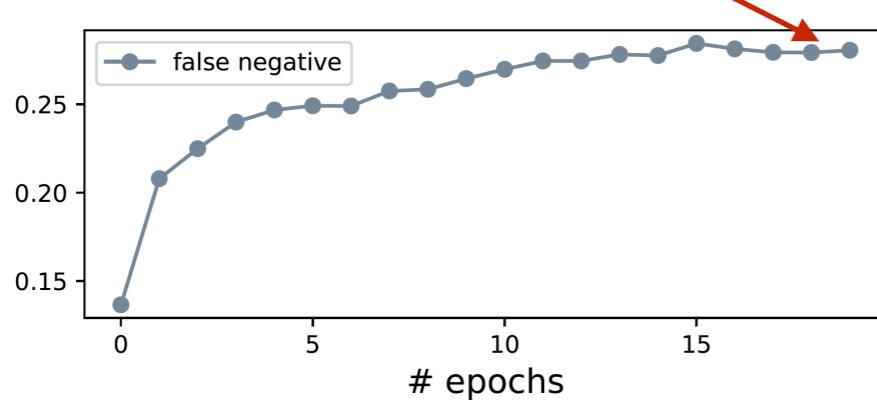
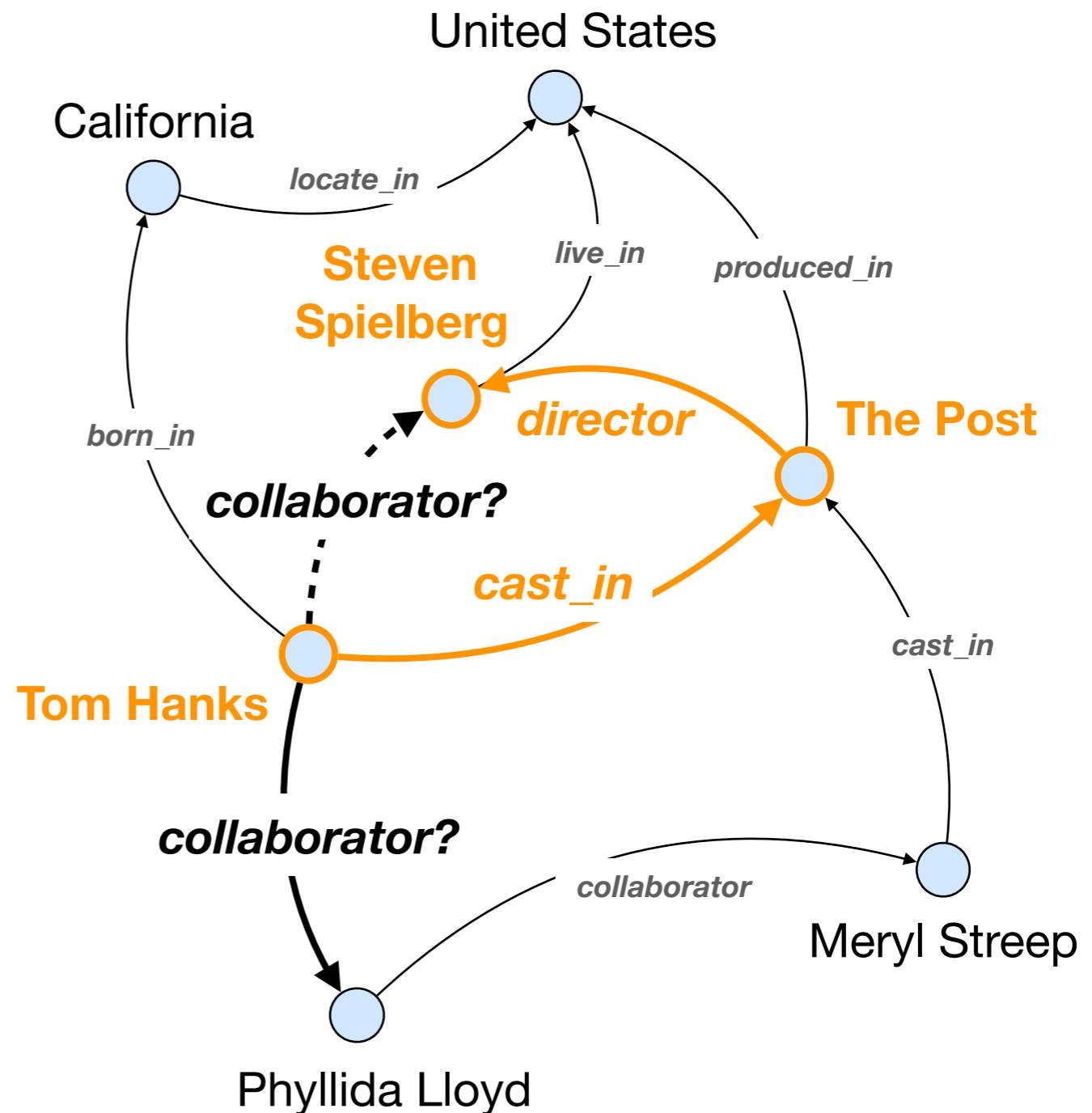


Fig 1. % of false negative hit in the first 20 epochs of RL training on the UMLS KG benchmark (Kok and Domingos 2007)



Questions for Future Research

1. One natural question to ask is why *a perceivable performance gap exists between the embedding-based (EB) model and the RL approach using the same EB model as the reward shaping module* (slide 51), *especially on FB15k-237 and NELL-995*, the two larger and more sparse KGs. Since the RL model has complete access to the EB model, why does it still lose information? A possible explanation is that for examples where the RL models make mistakes, the topic entity and the target answer cannot be reached within the specified # hops. Yet our sanity check disproved this hypothesis — 100% of the time when the RL model makes mistakes but the embedding approach was correct the topic entity and the target answer were connected. We did not check if the quality of these paths is significantly lower than those of the correct predictions. *VXL: Because the RL model & training process have a lot more hyper-parameters than the EB models, it is possible that the RL models are more difficult to optimize as we are considering a much larger space of hyper-parameter settings.*
2. In our experiments, very large action dropout rates (0.9 and 0.95) yield good performance on the dense KGs (Kinship and UMLS), but the same strategy does not work for sparser KGs. *We observed significant performance drop for FB15k-237, WN18RR and NELL-995 when using very large action dropout rates. And for WN18NN and NELL-995, action dropout rate > 0.1 hurts performance.* It is unclear why REINFORCE training on the denser KGs can tolerate a larger shift from the actual policy. *VXL: It seems that the shape of the original policy function ought to be preserved to some degree during training. For Kinship and UMLS, the average node degrees are 85 and 39, in this case on average ≥ 2 edges remain on when we randomly turned off 95% of the edges. Since other KGs have smaller average node degrees, using a large action dropout rate is equivalent to doing random exploration most of the time.*
3. Does EB models define the cap performance in the one-hop KG query answering set up? *Could the tasks of path finding and learning KG embeddings be joint together in a way s.t. they can improve each other?*
4. Our approach can be viewed as a way to explain pre-trained EB models. Are there better ways to do it?

Acknowledgement upon slides release - I

These slides benefit tremendously from the constructive feedback offered by Salesforce Research team members, including Caiming Xiong, Richard Socher, Yingbo Zhou, Alex Trott, Jin Qu, Lily Hu, Vena Li, Kazuma Hashimoto, Stephan Zheng, Jason Wu (intern) and others.

Acknowledgement upon slides release - II

I am grateful to Prof. Michael Ernst and his co-authors who released the slides of all of their paper presentations.

Especially, the **Hand Font** and the **Baloon** highlighters were borrowed from slides 1 and slides 2.

Victoria Lin
Nov. 4, 2018