

# 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

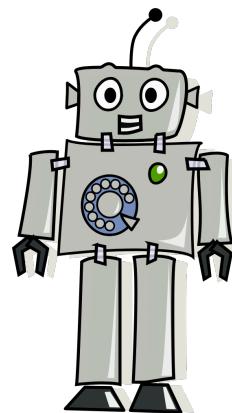


# Question Answering System

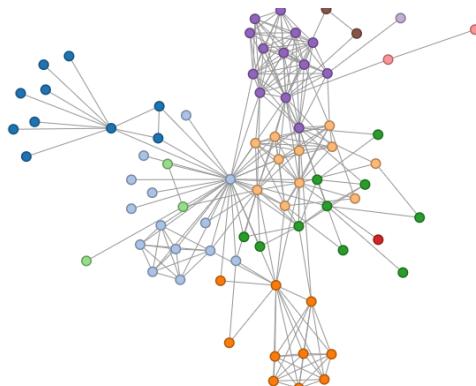
Text



Images



Knowledge Graph

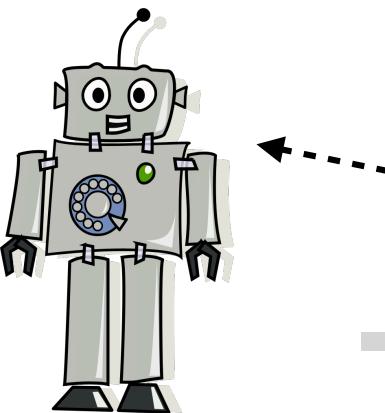
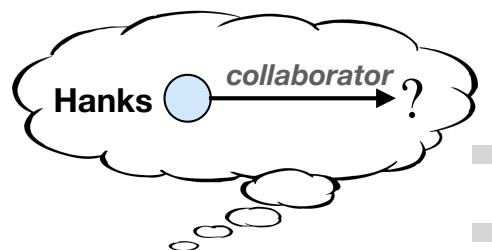
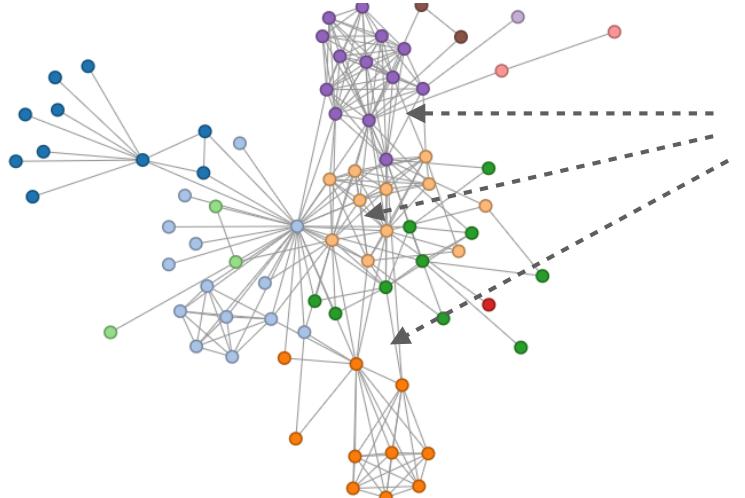


Which directors  
has Tom Hanks  
collaborated with?



# Question Answering System

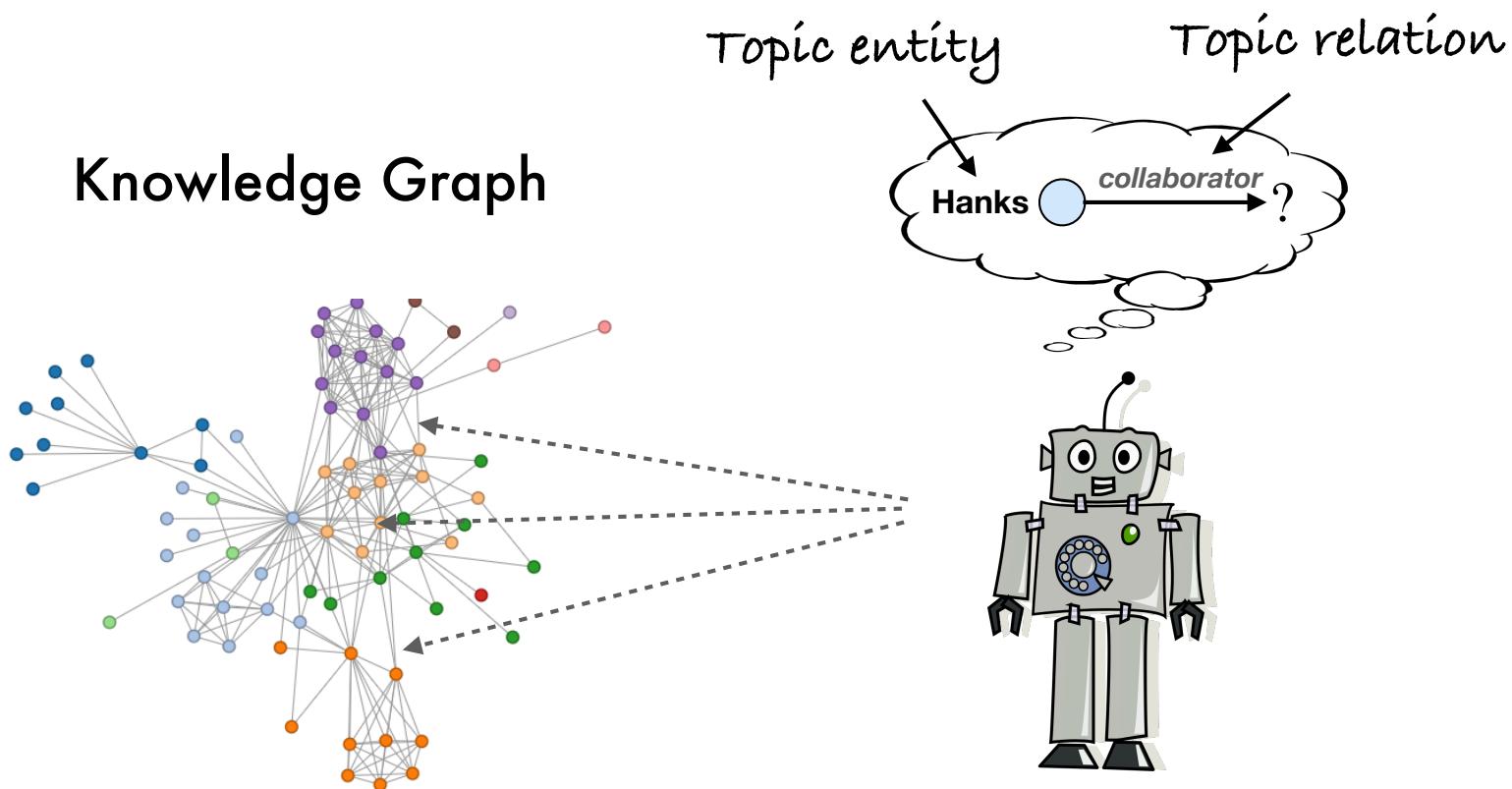
Knowledge Graph



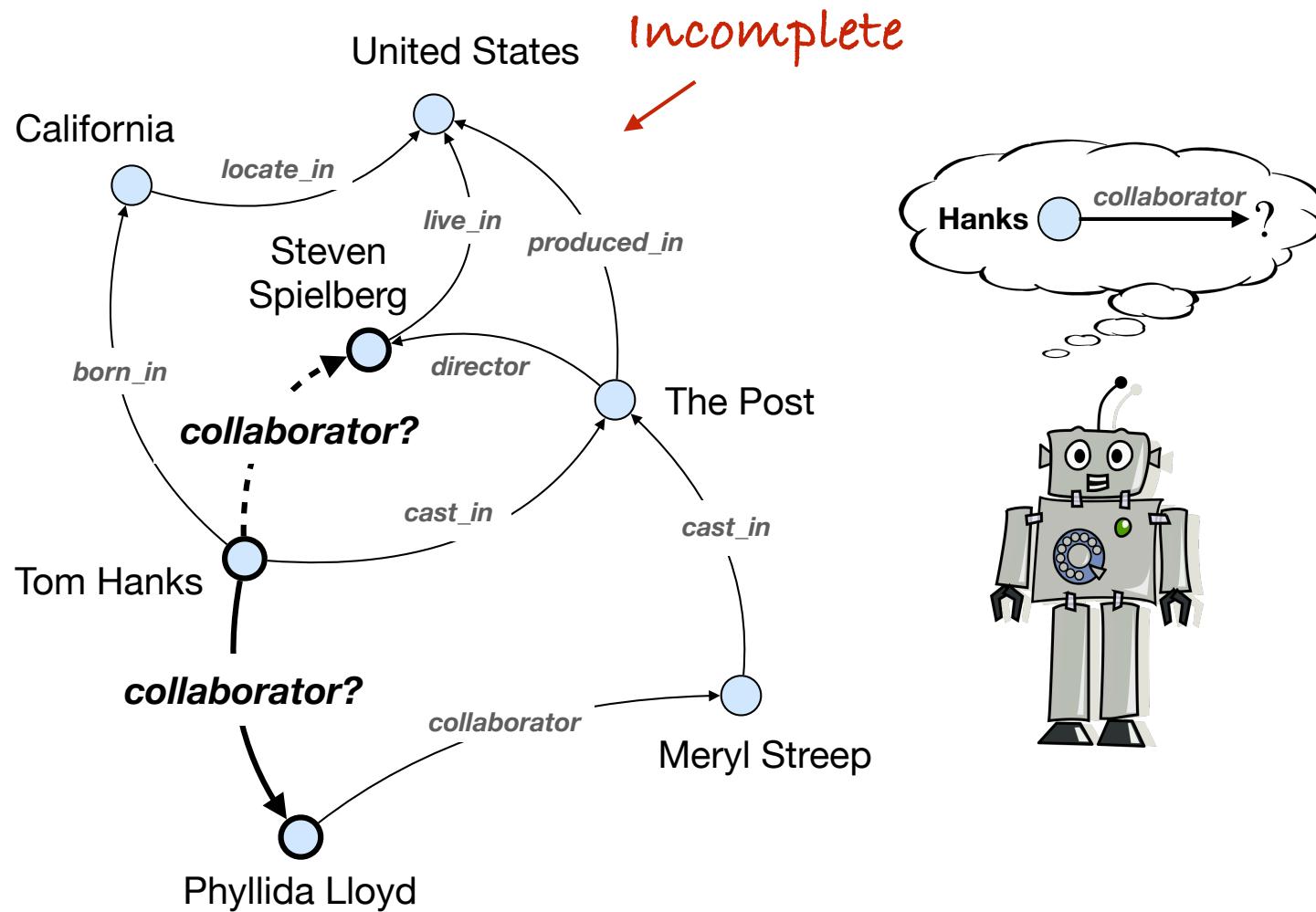
Which directors  
has **Tom Hanks**  
collaborated with?



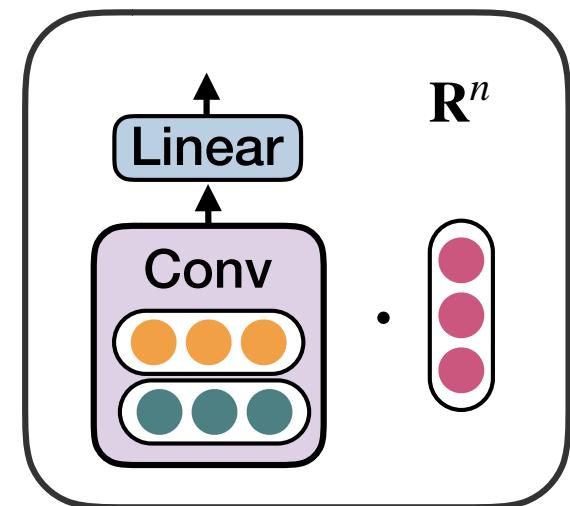
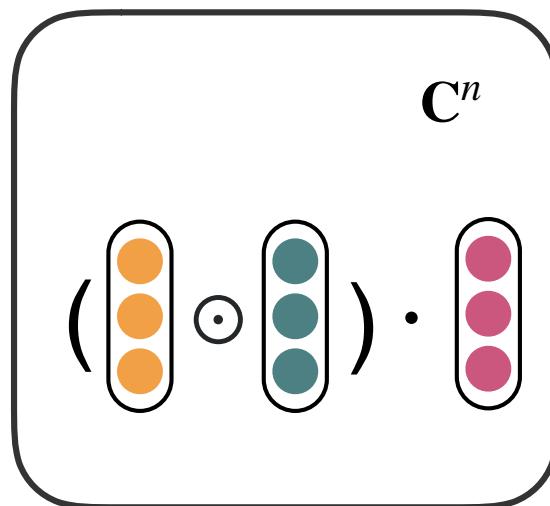
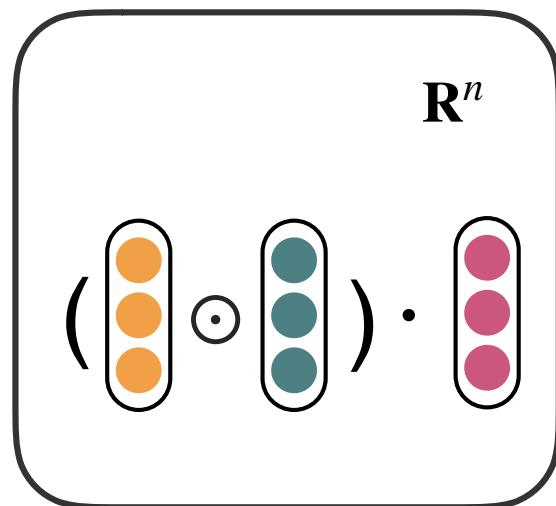
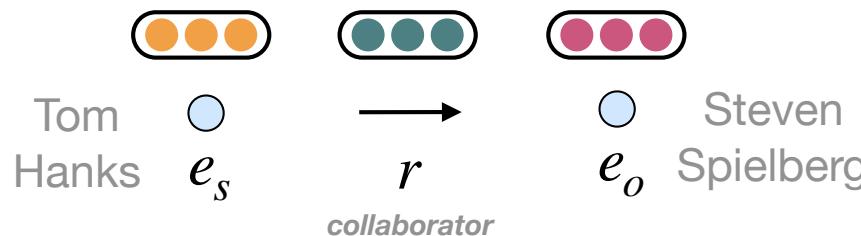
# Structured Query Answering



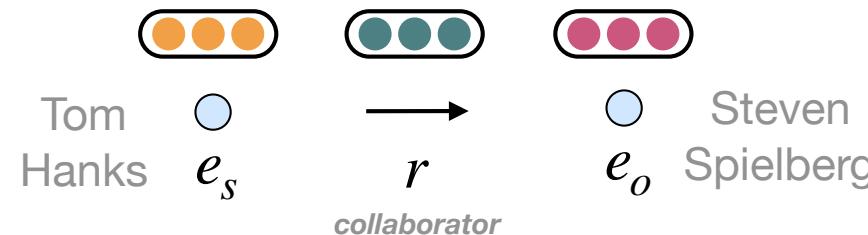
# Structured Query Answering



# Knowledge Graph Embeddings



# Knowledge Graph Embeddings



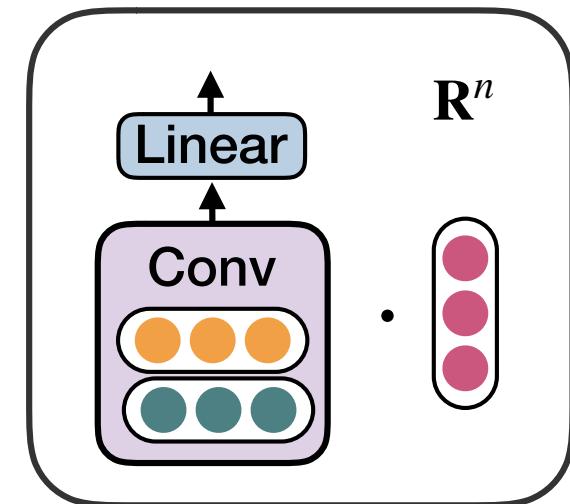
Highly accurate &  
Efficient

	MRR
ConvE	0.957

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

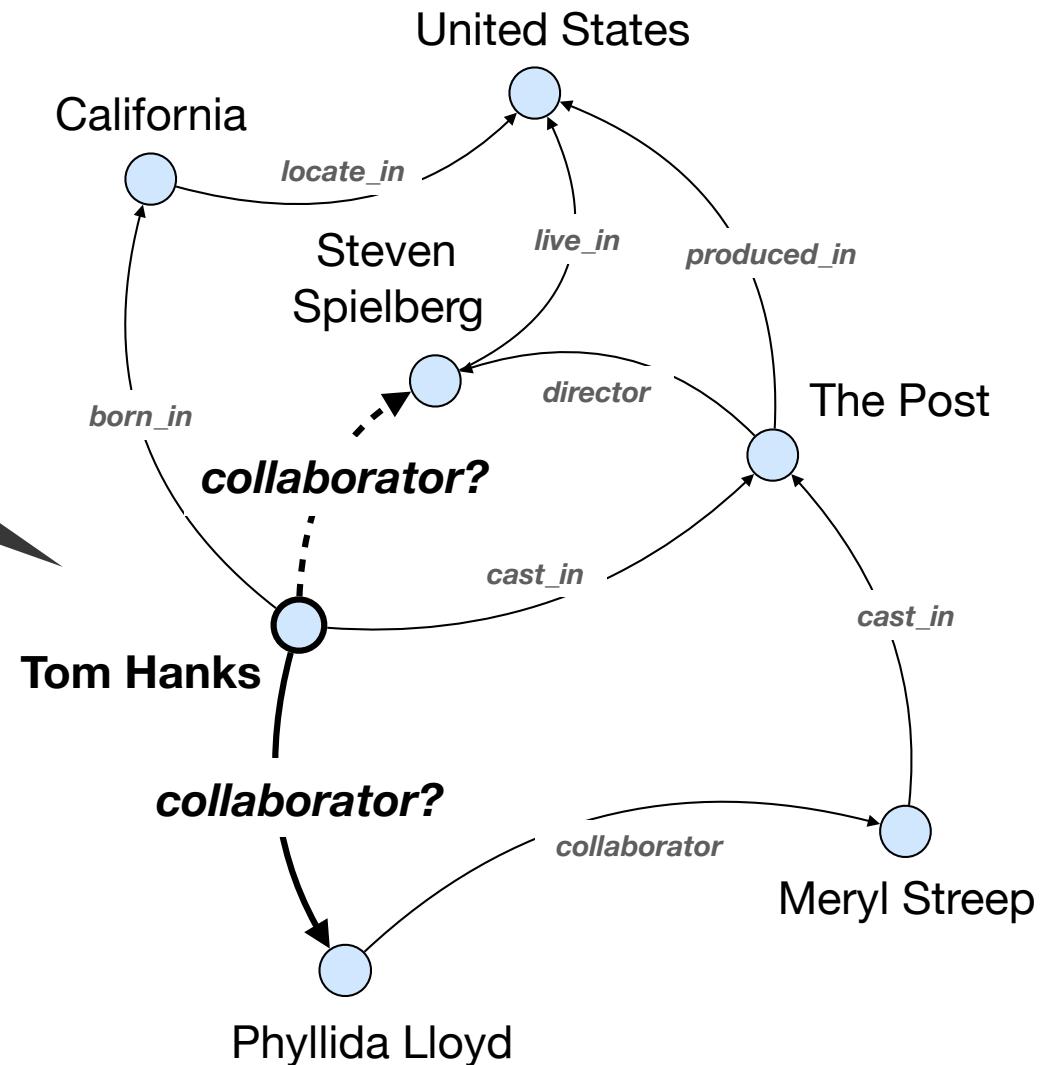
Lack  
interpretability

Why Spielberg  
is a collaborator  
of Hanks?



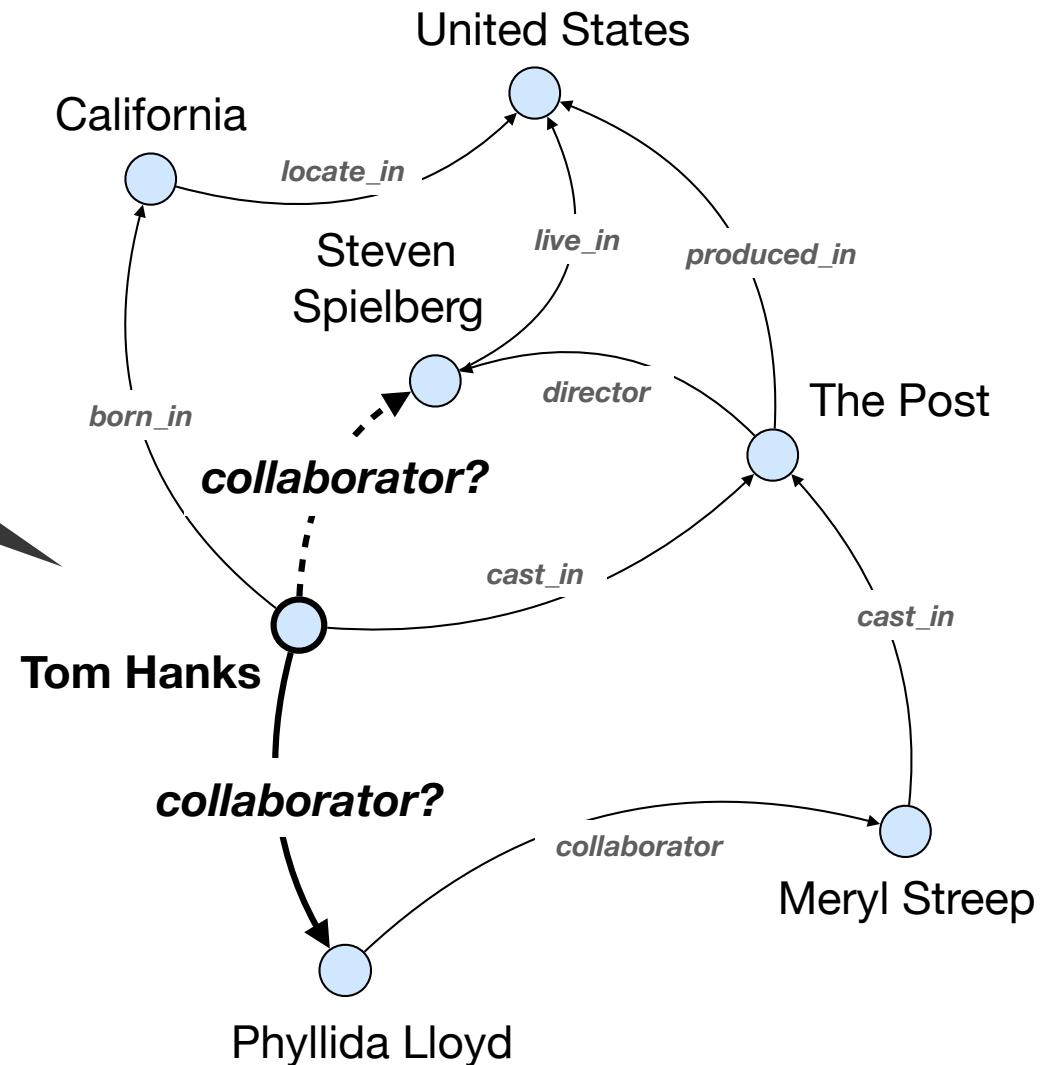
# Multi-Hop Reasoning Models

Reasoning  
over discrete  
structures



# Multi-Hop Reasoning Models

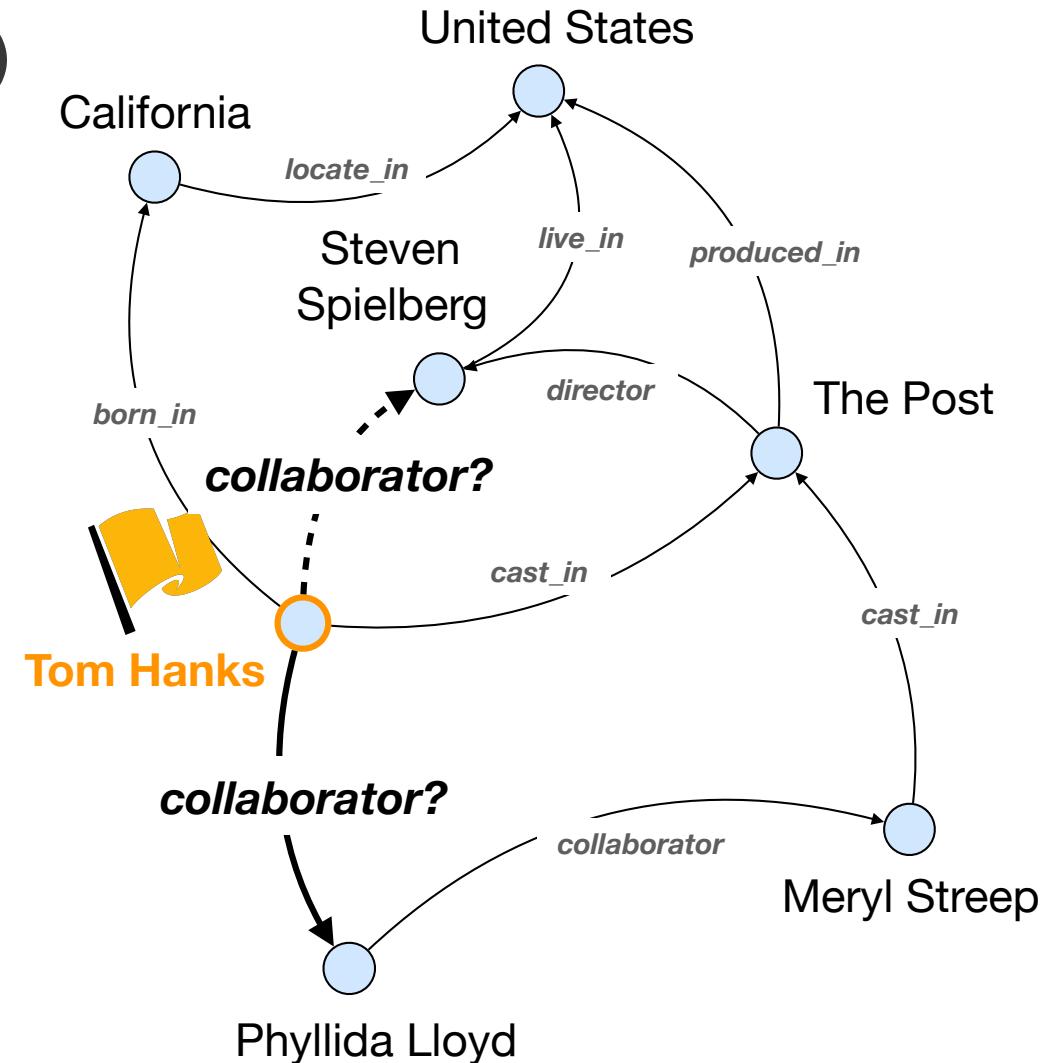
Sequential  
decision making



# Multi-Hop Reasoning Models

- Tom Hanks

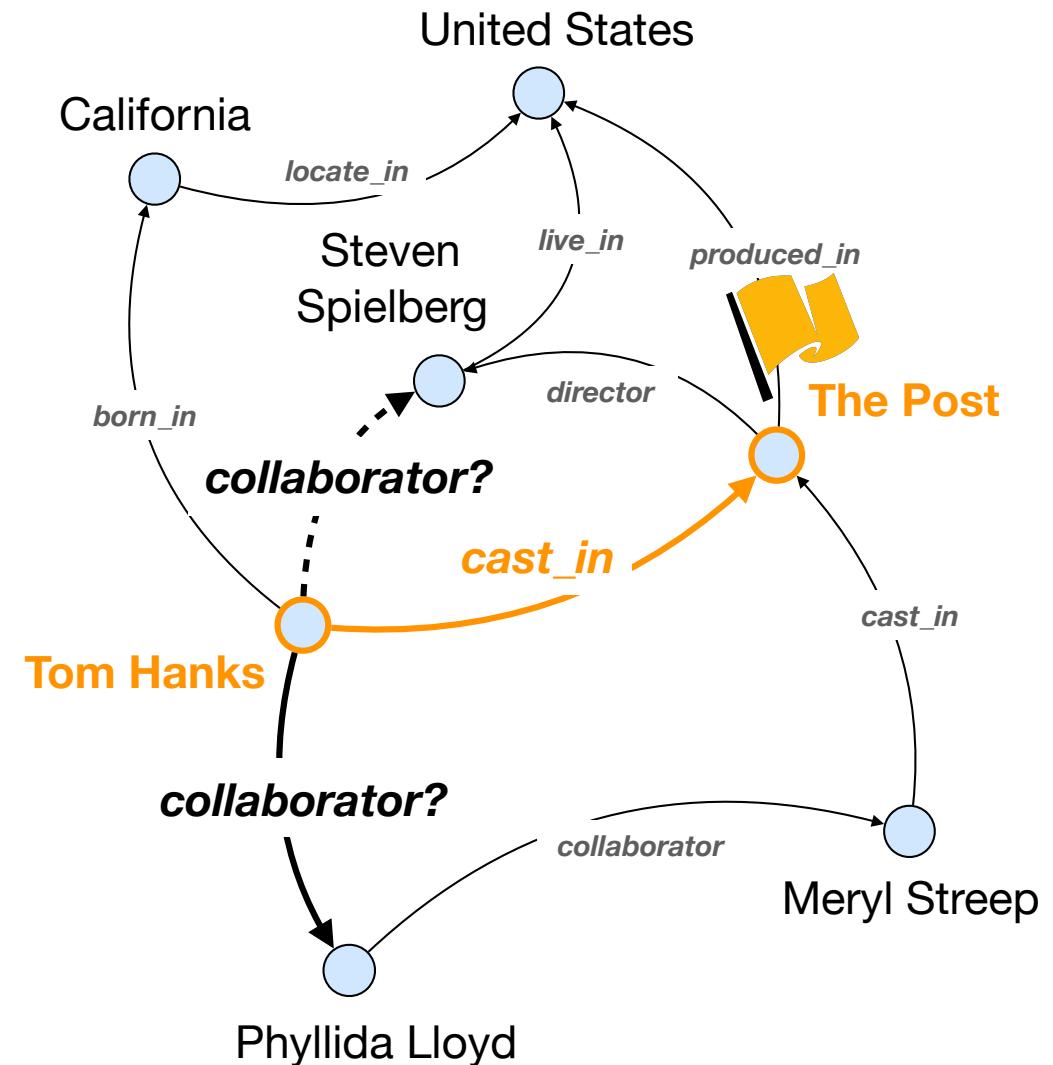
Topic entity



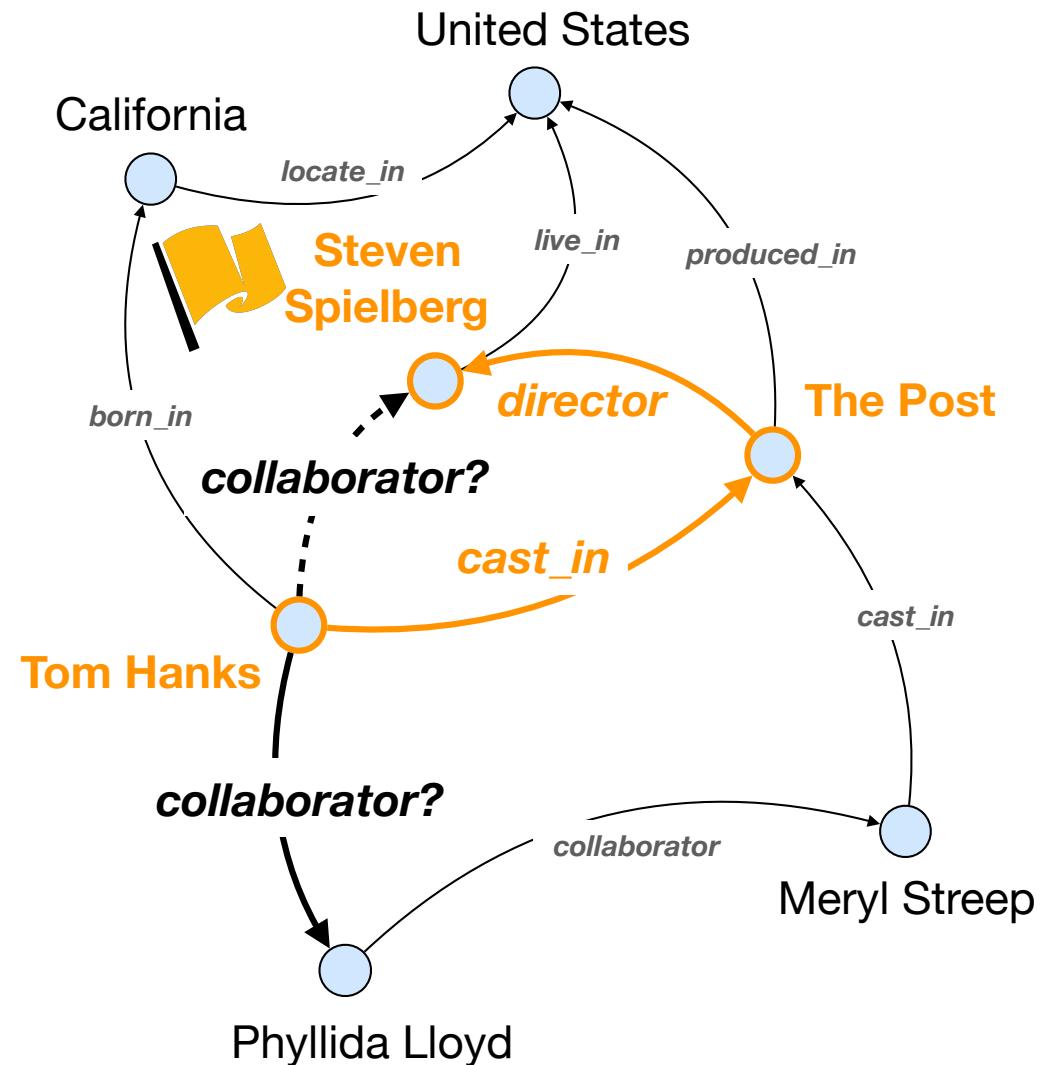
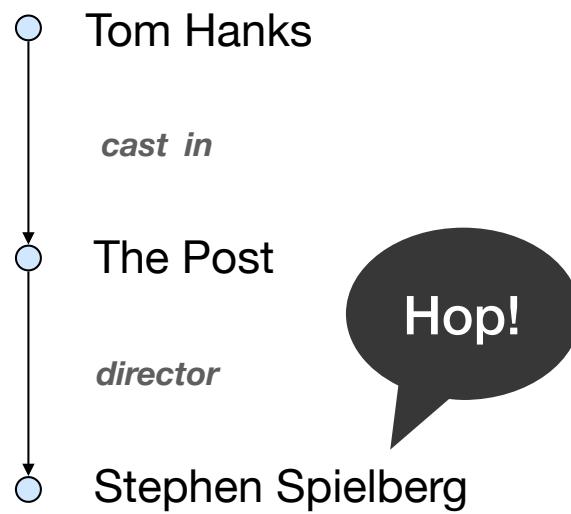
# Multi-Hop Reasoning Models



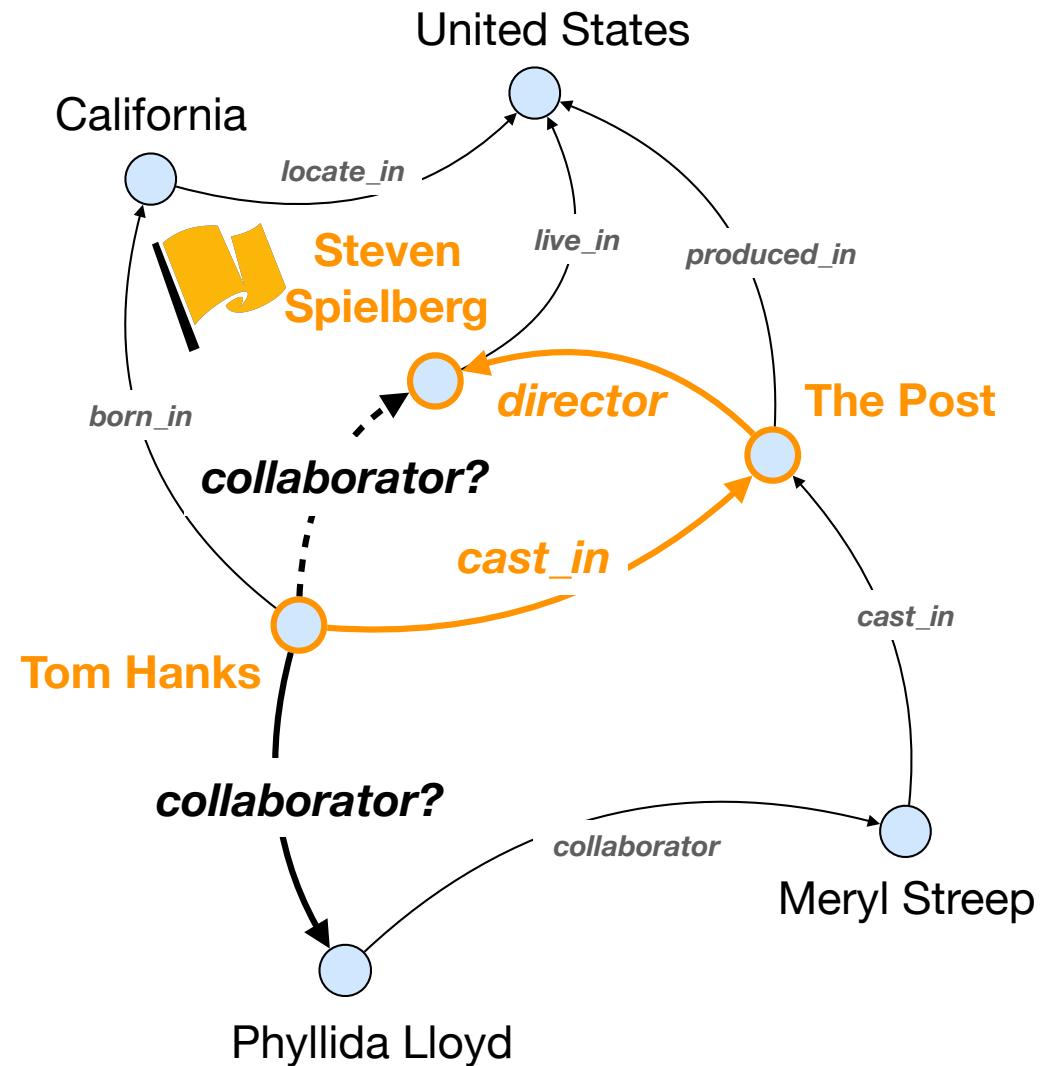
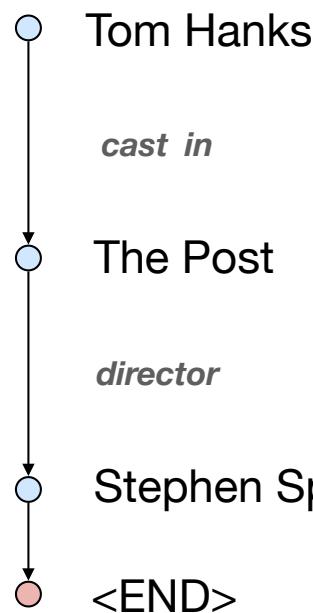
Hop!



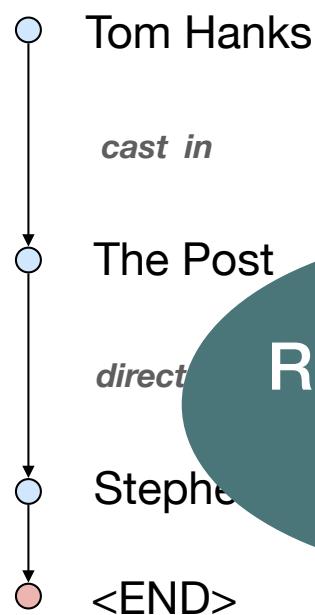
# Multi-Hop Reasoning Models



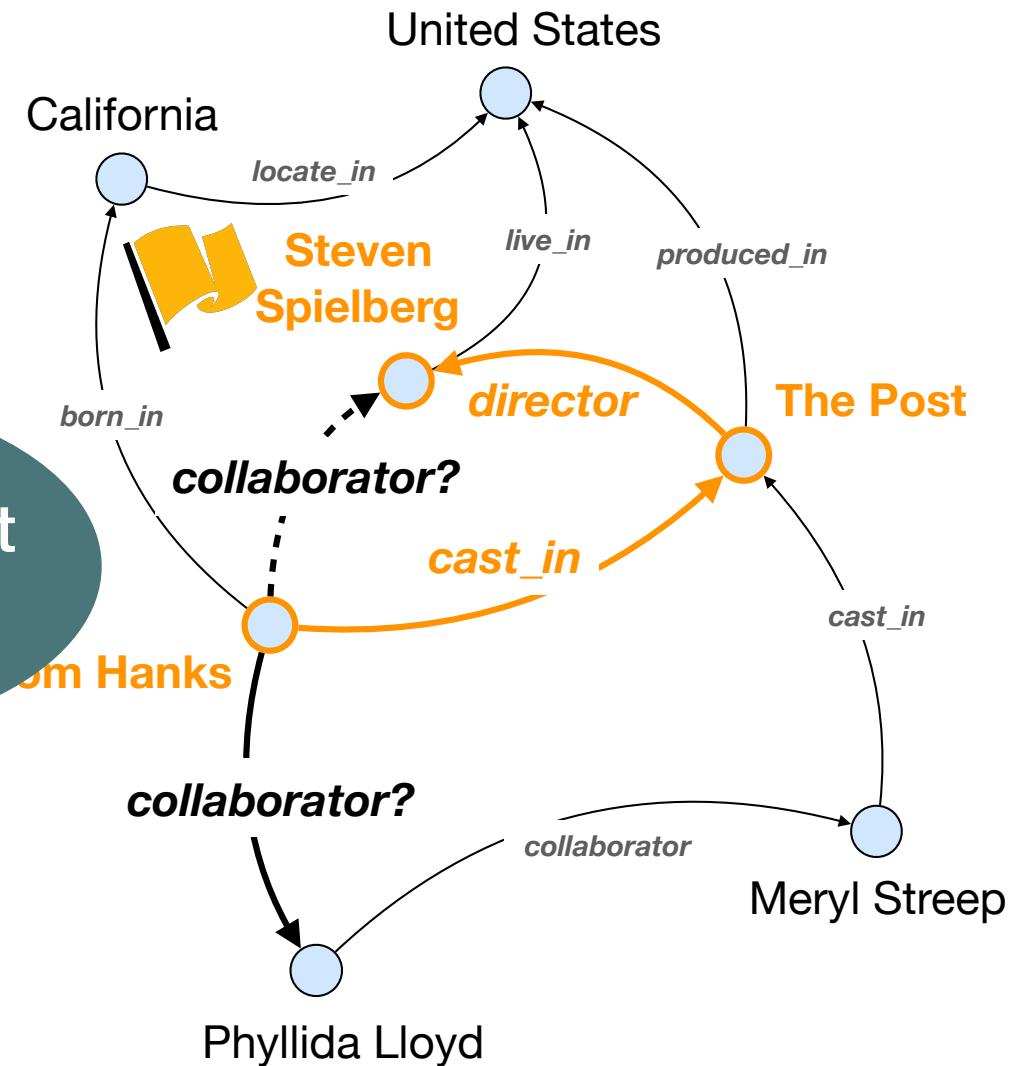
# Multi-Hop Reasoning Models



# Multi-Hop Reasoning Models



Reinforcement Learning



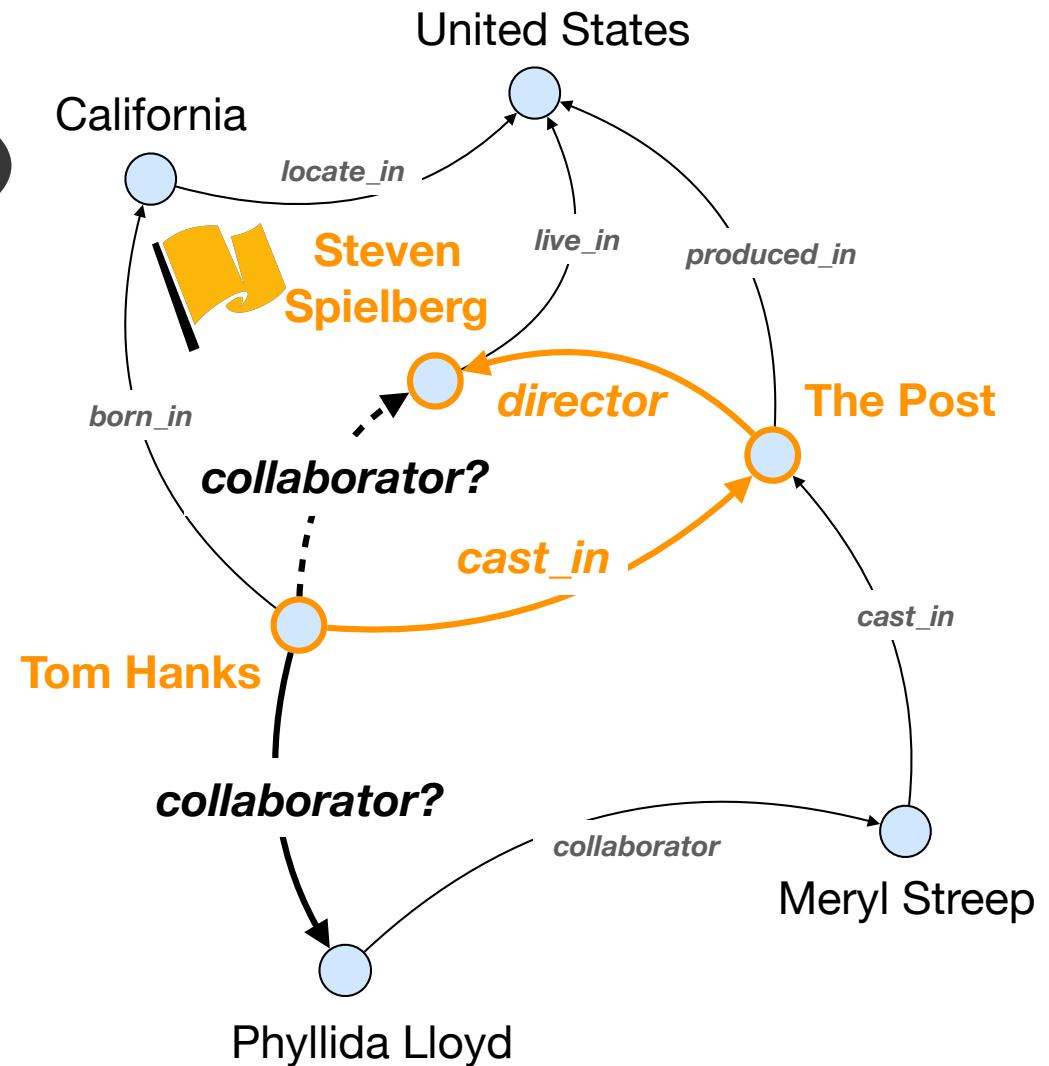
# Multi-Hop Reasoning Models

 **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)



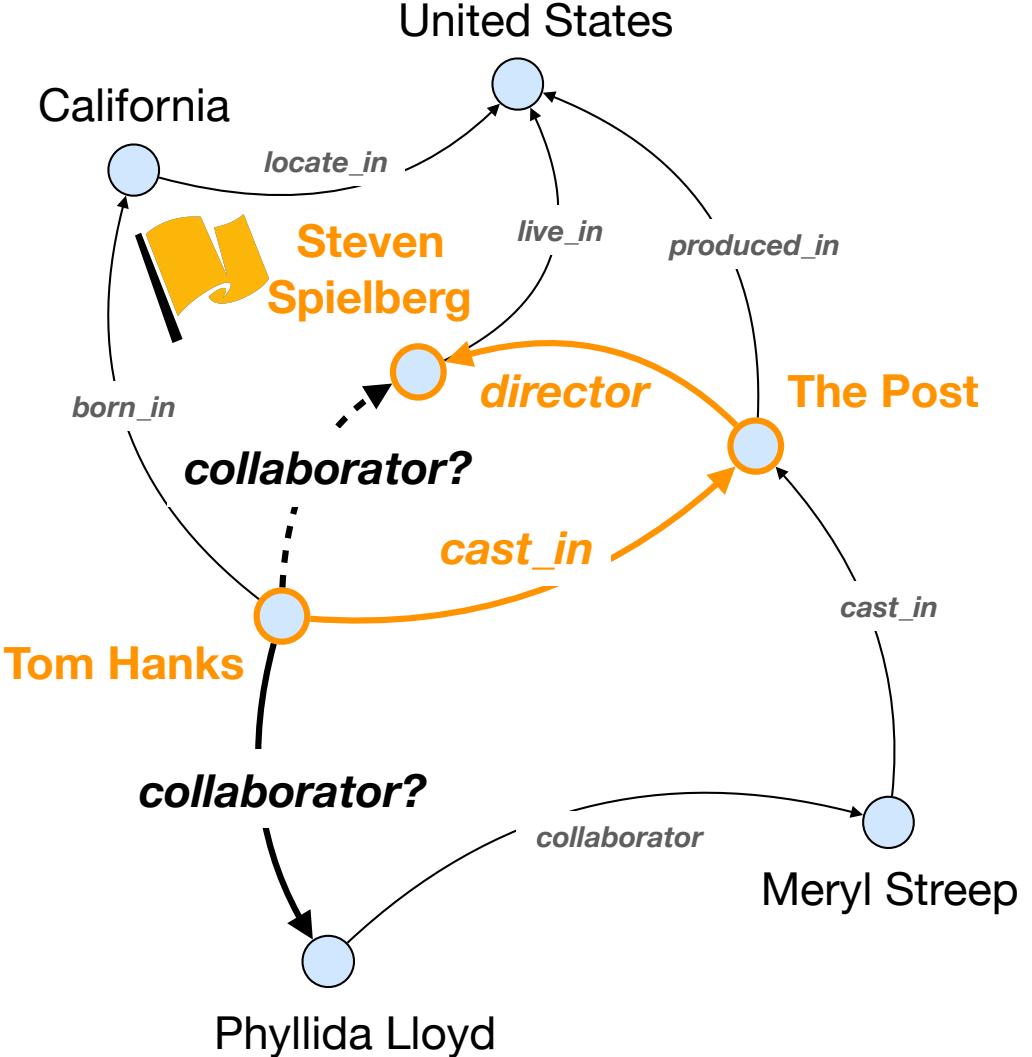
# Goal

 **Interpretable**

~~significant performance gap~~

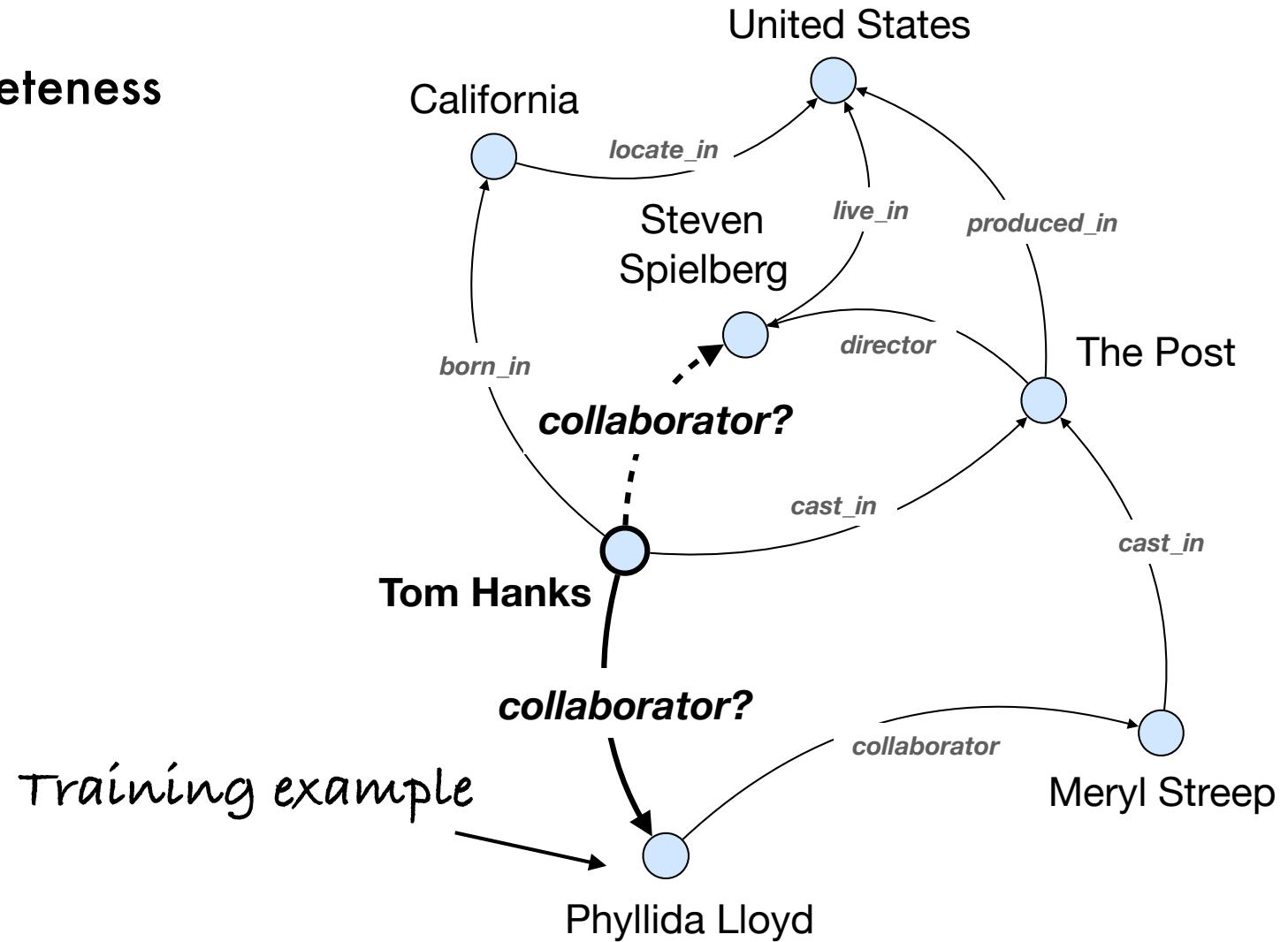
	MRR
ConvE	0.957
RL	0.825

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



# Challenges

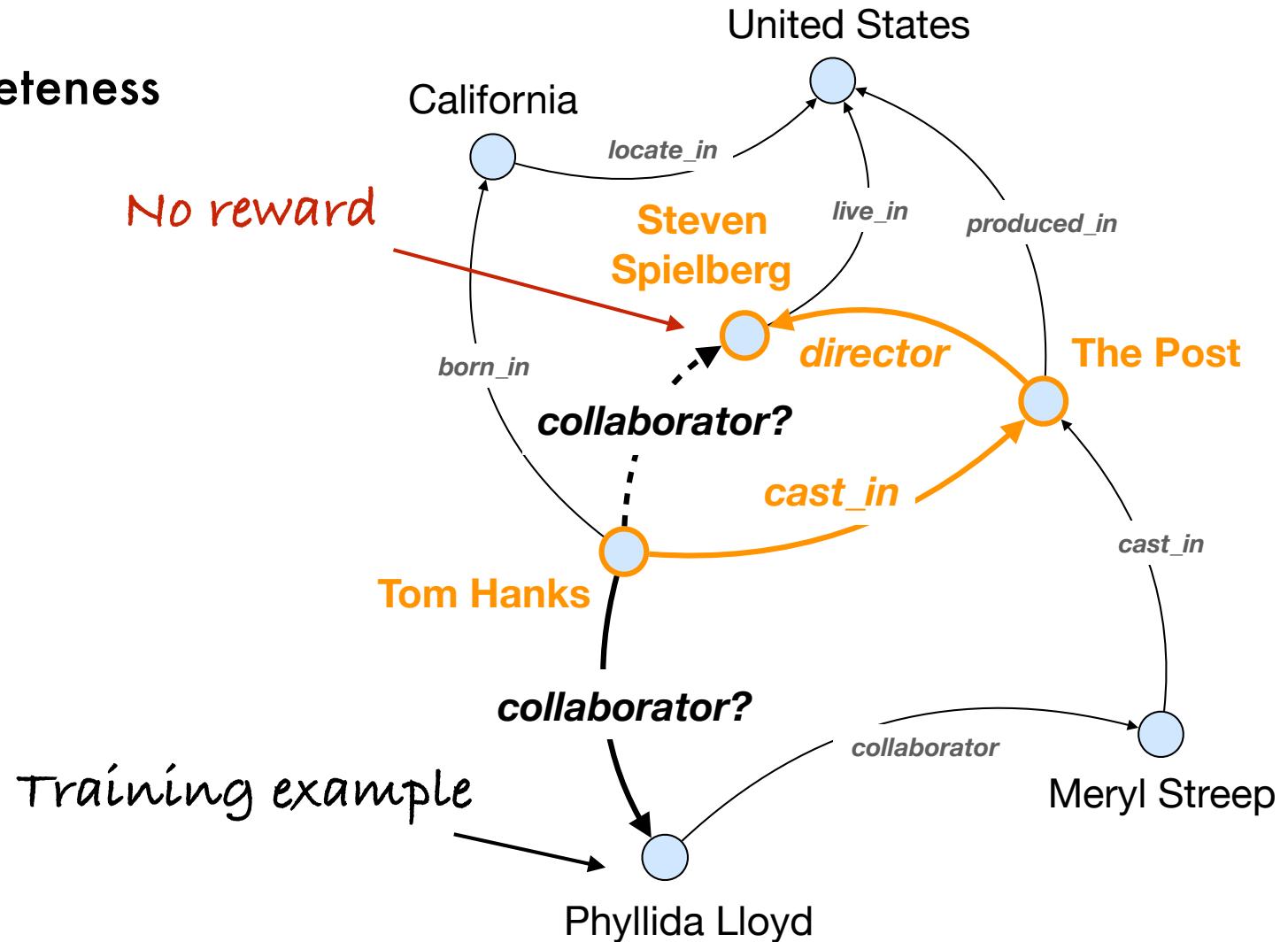
## Incompleteness



# Challenges

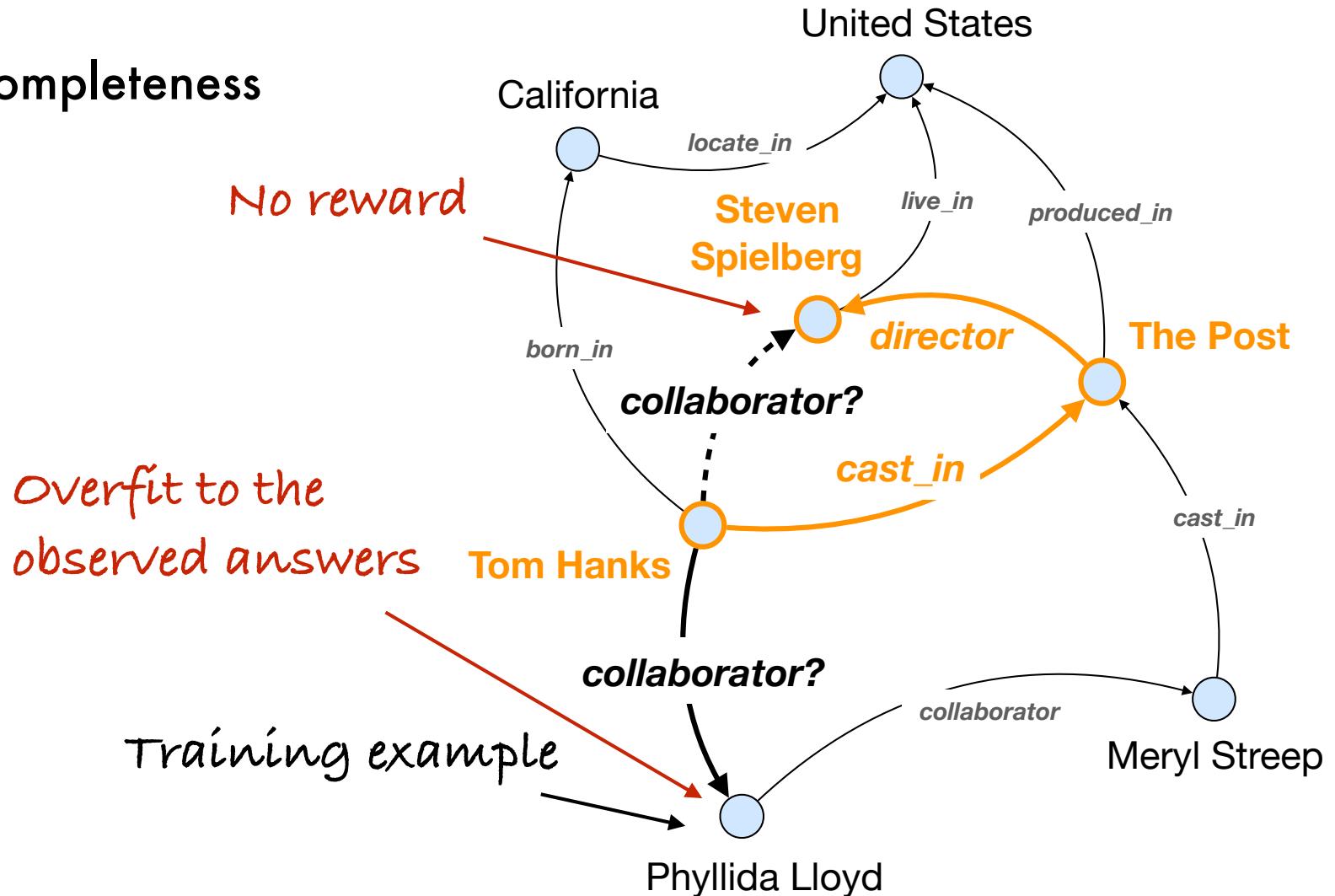
Incompleteness

No reward



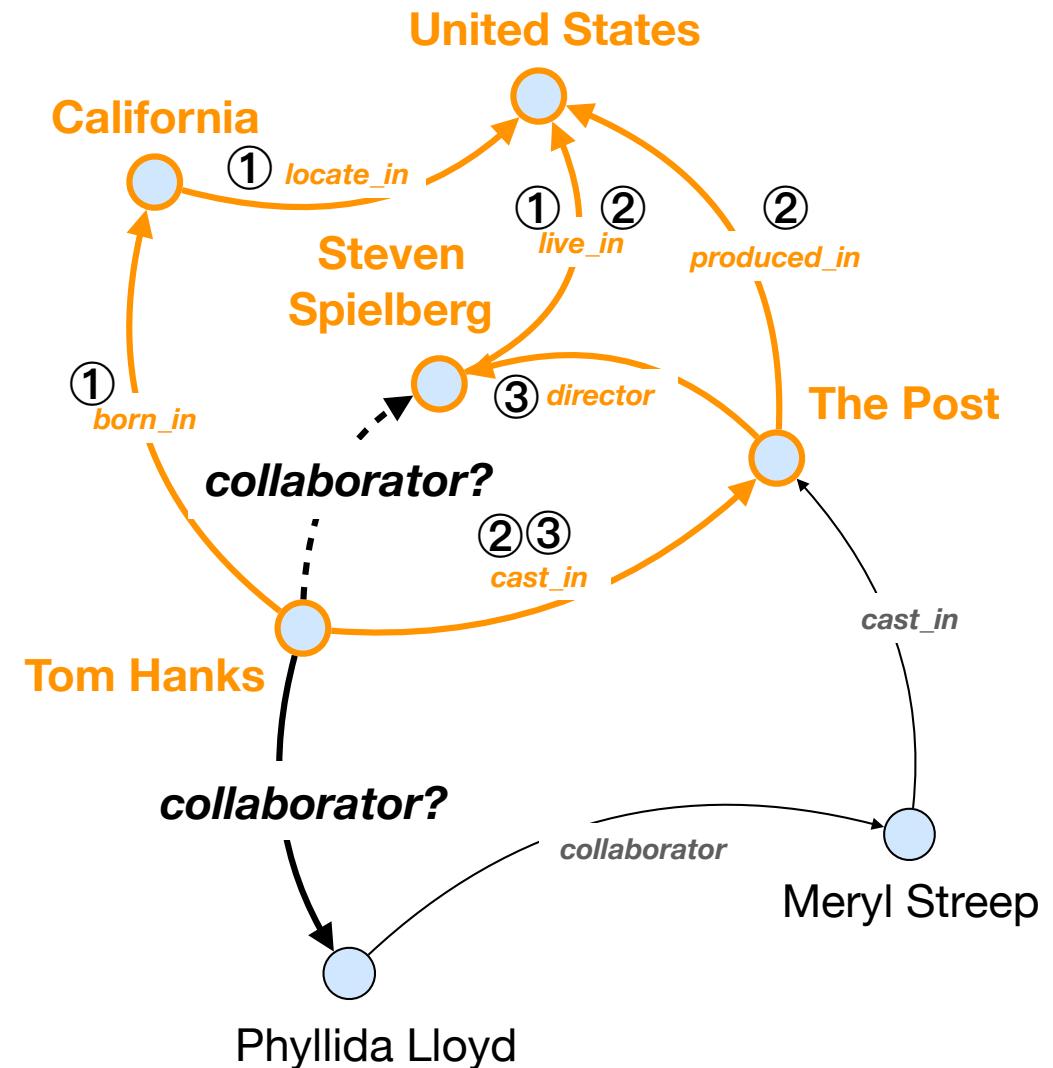
# Challenges

Incompleteness



# Challenges

## Path Diversity

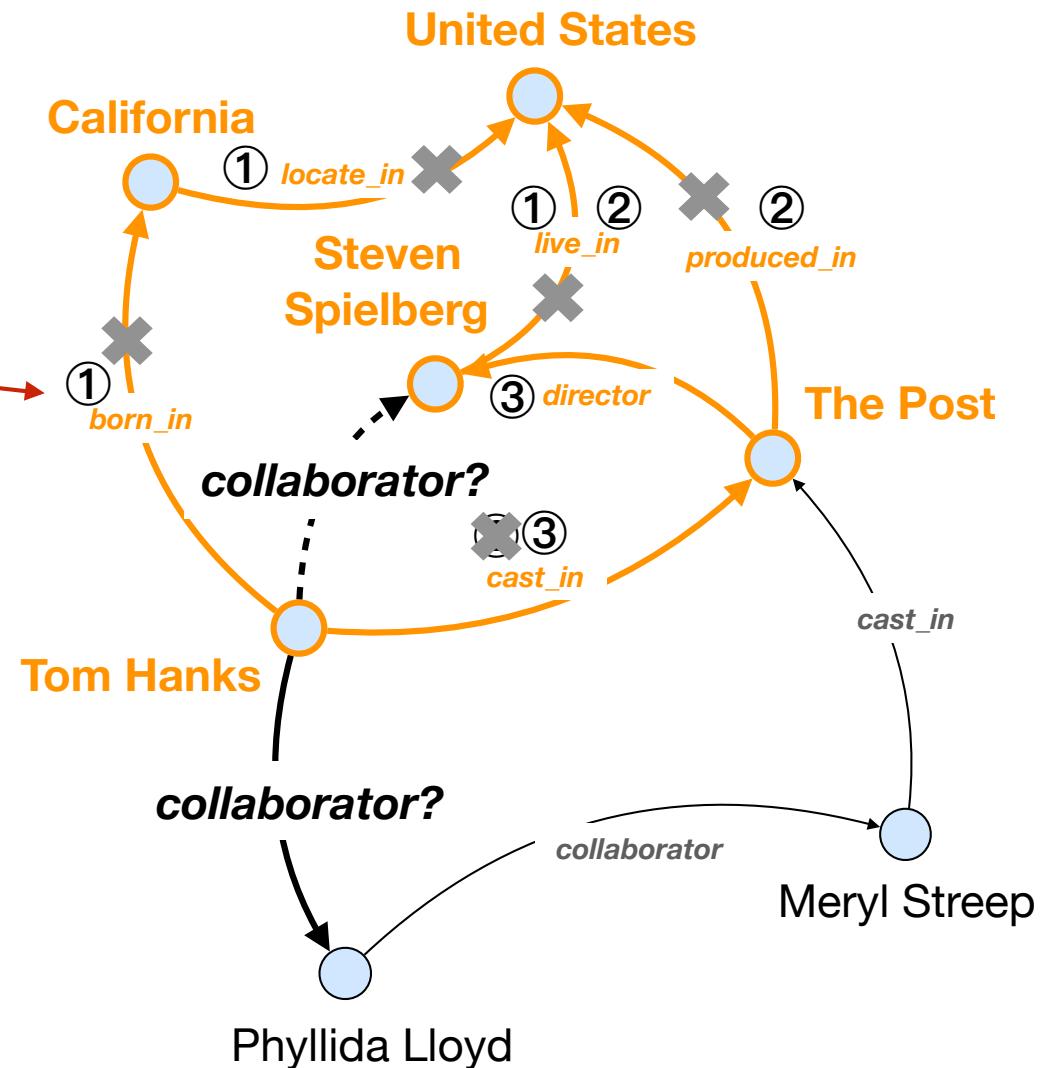


# Challenges

## Path Diversity

False positive

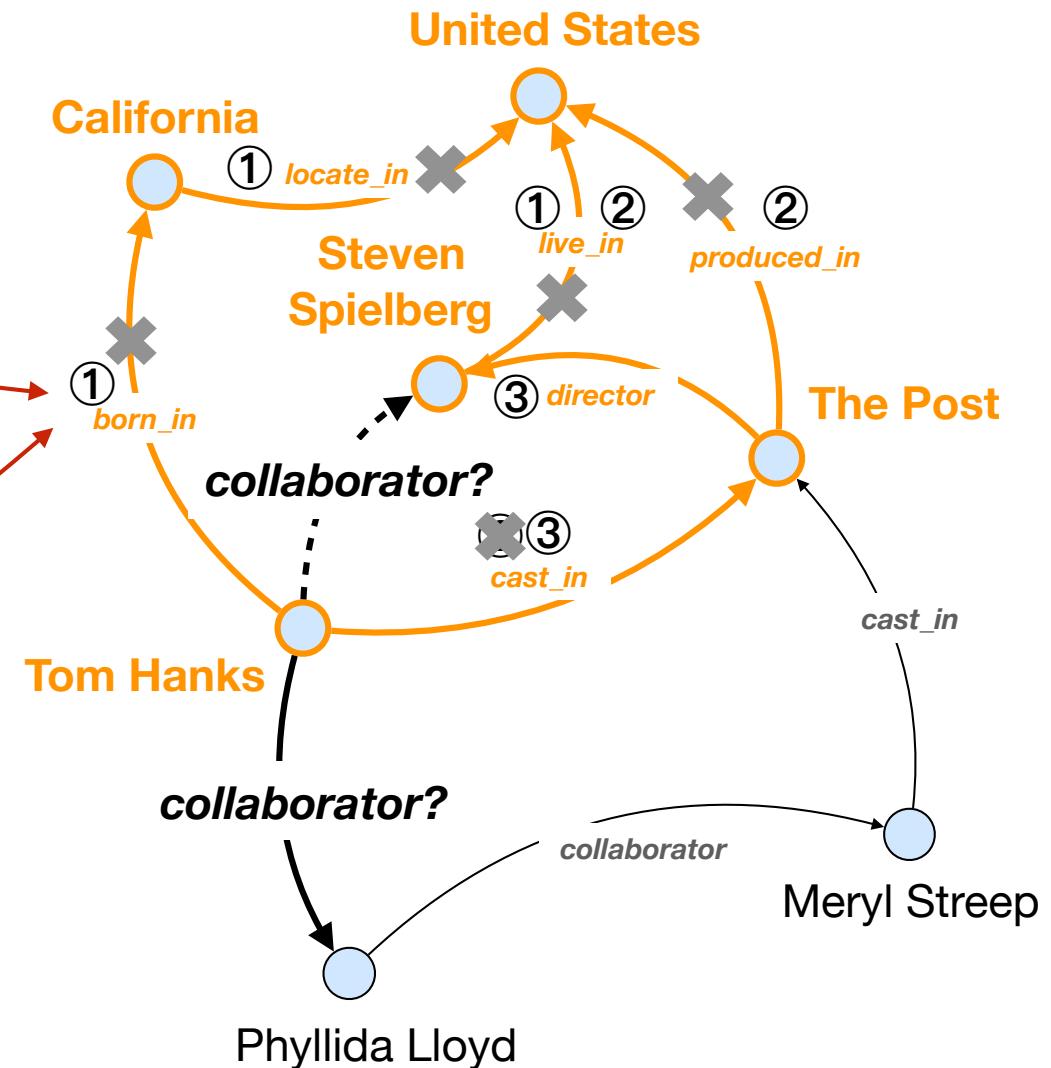
(spurious) paths



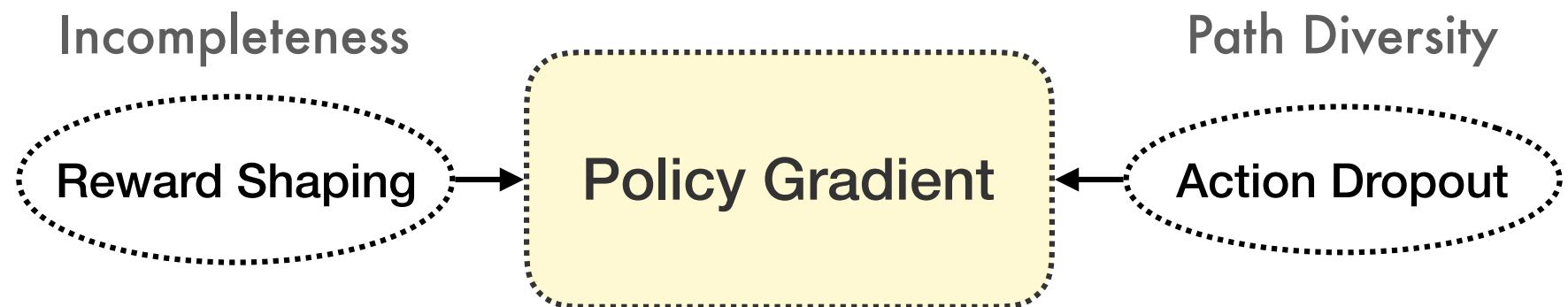
# Challenges

## Path Diversity

False positive (spurious) paths  
Overfit to the spurious paths



# Proposed Solutions



# Reinforcement Learning Framework

Environment

State

Action

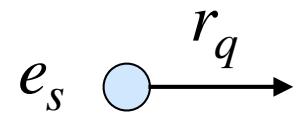
Transition

Reward

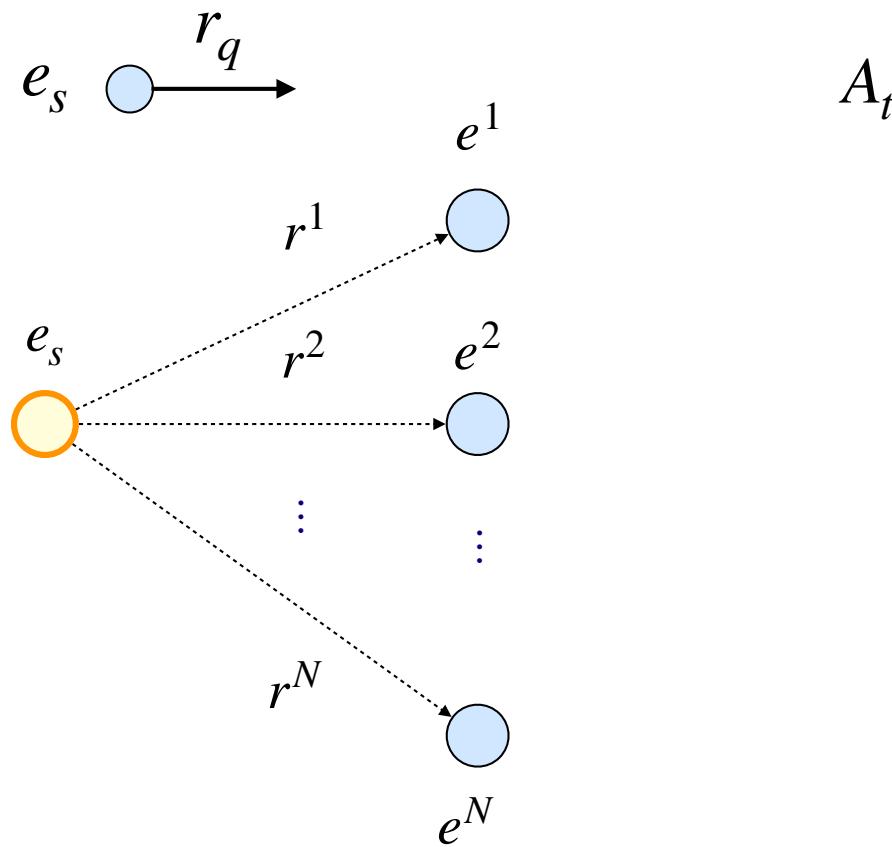
$$e_s \quad \xrightarrow{r_q}$$



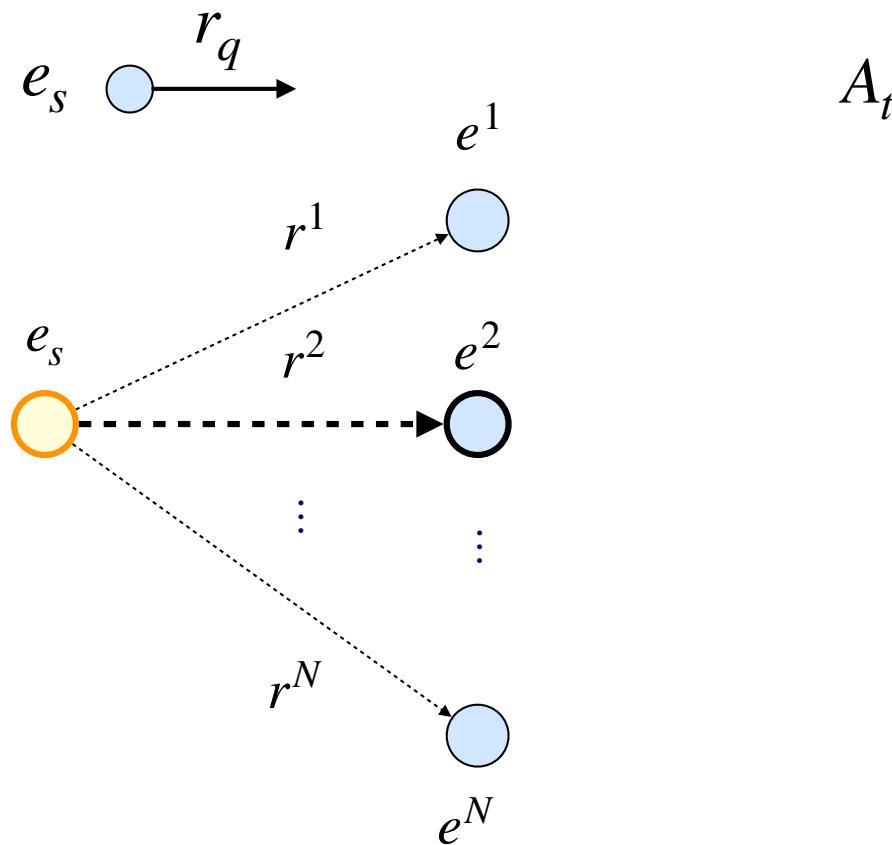
# Reinforcement Learning Framework



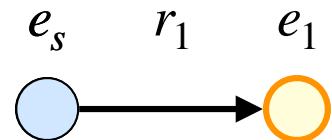
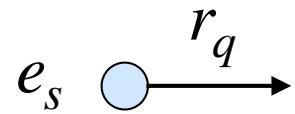
# Reinforcement Learning Framework



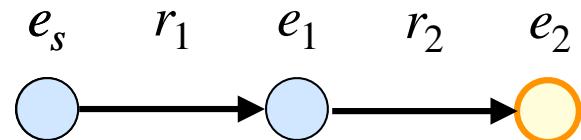
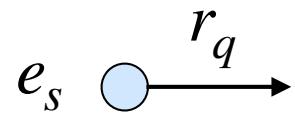
# Reinforcement Learning Framework



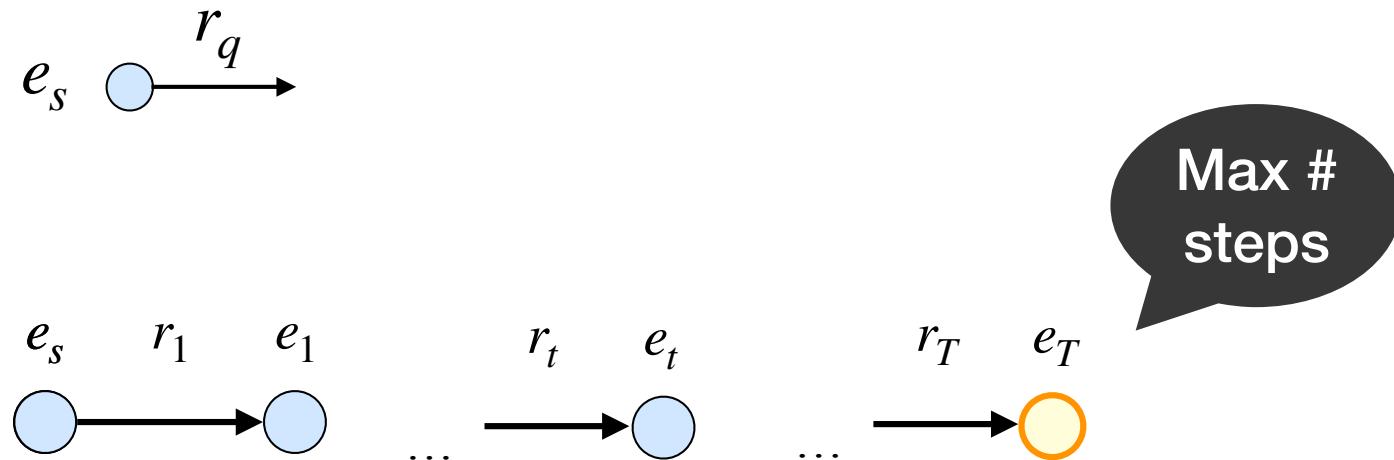
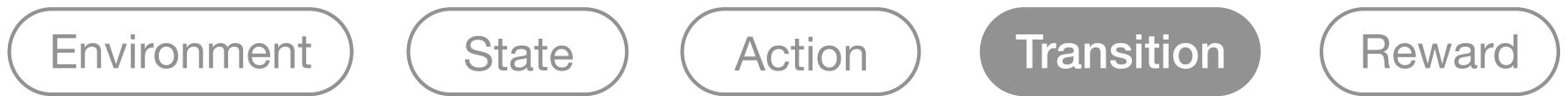
# Reinforcement Learning Framework



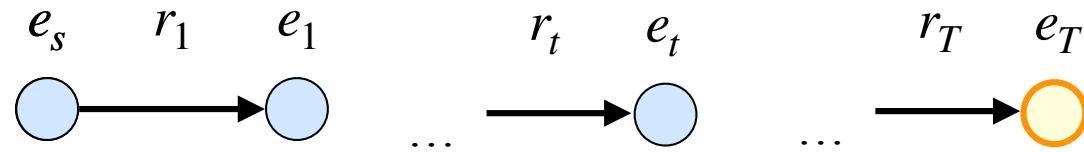
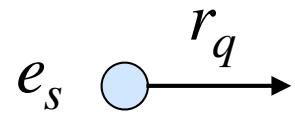
# Reinforcement Learning Framework



# Reinforcement Learning Framework



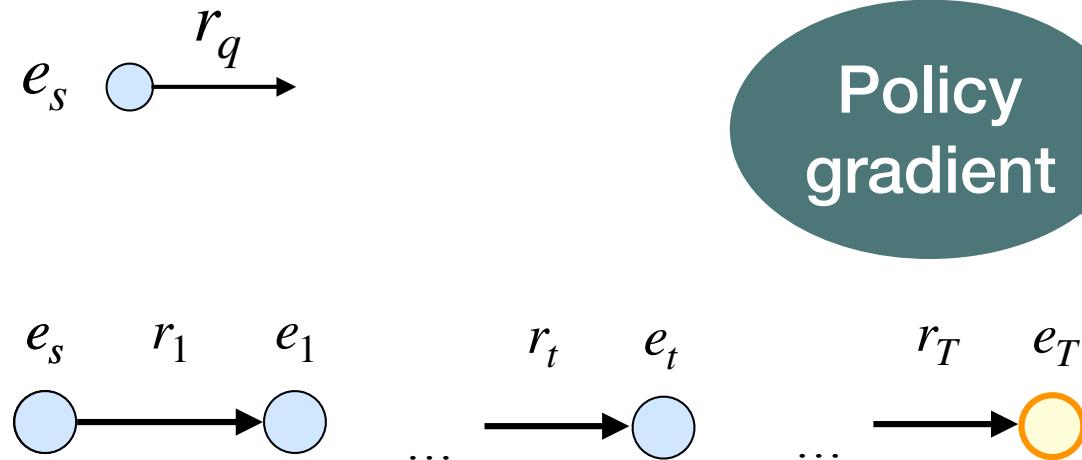
# Reinforcement Learning Framework



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

# Reinforcement Learning Framework

Environment      State      Action      Transition      Reward



Policy  
gradient

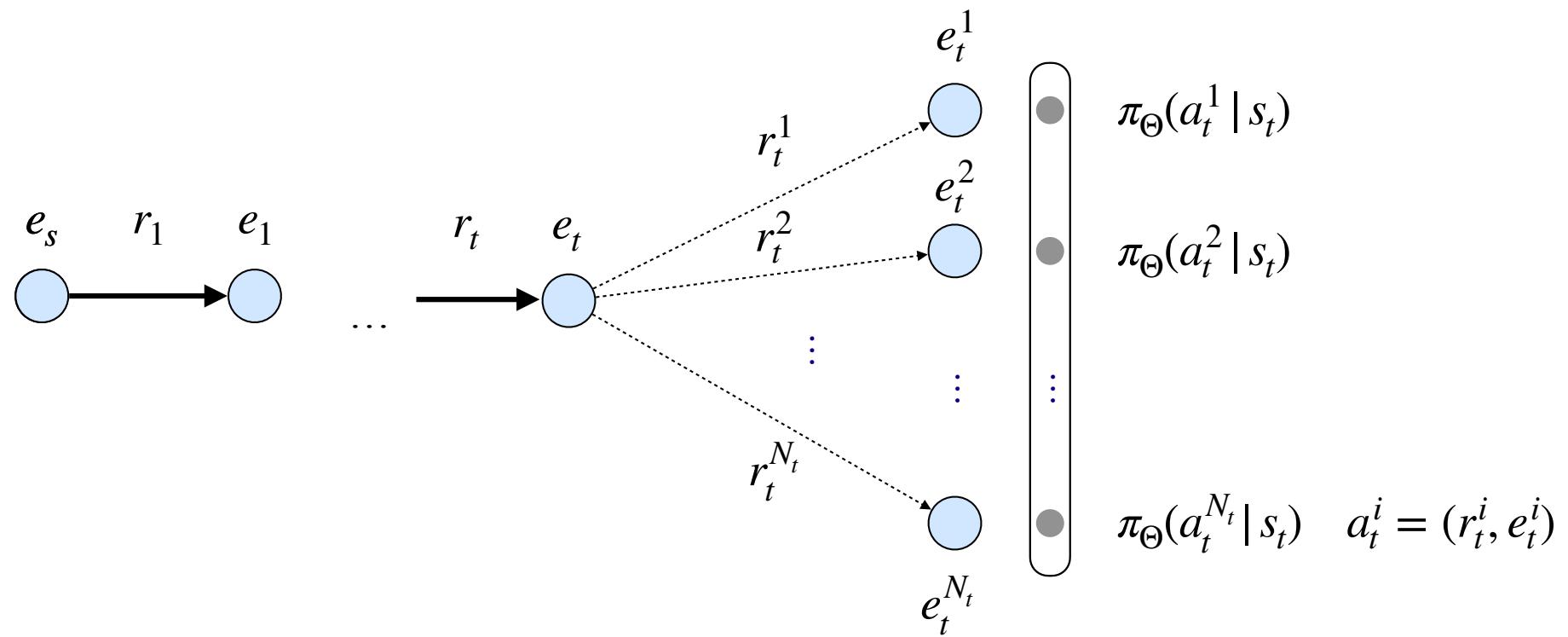
Models **which**  
**action to choose**  
at a state

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

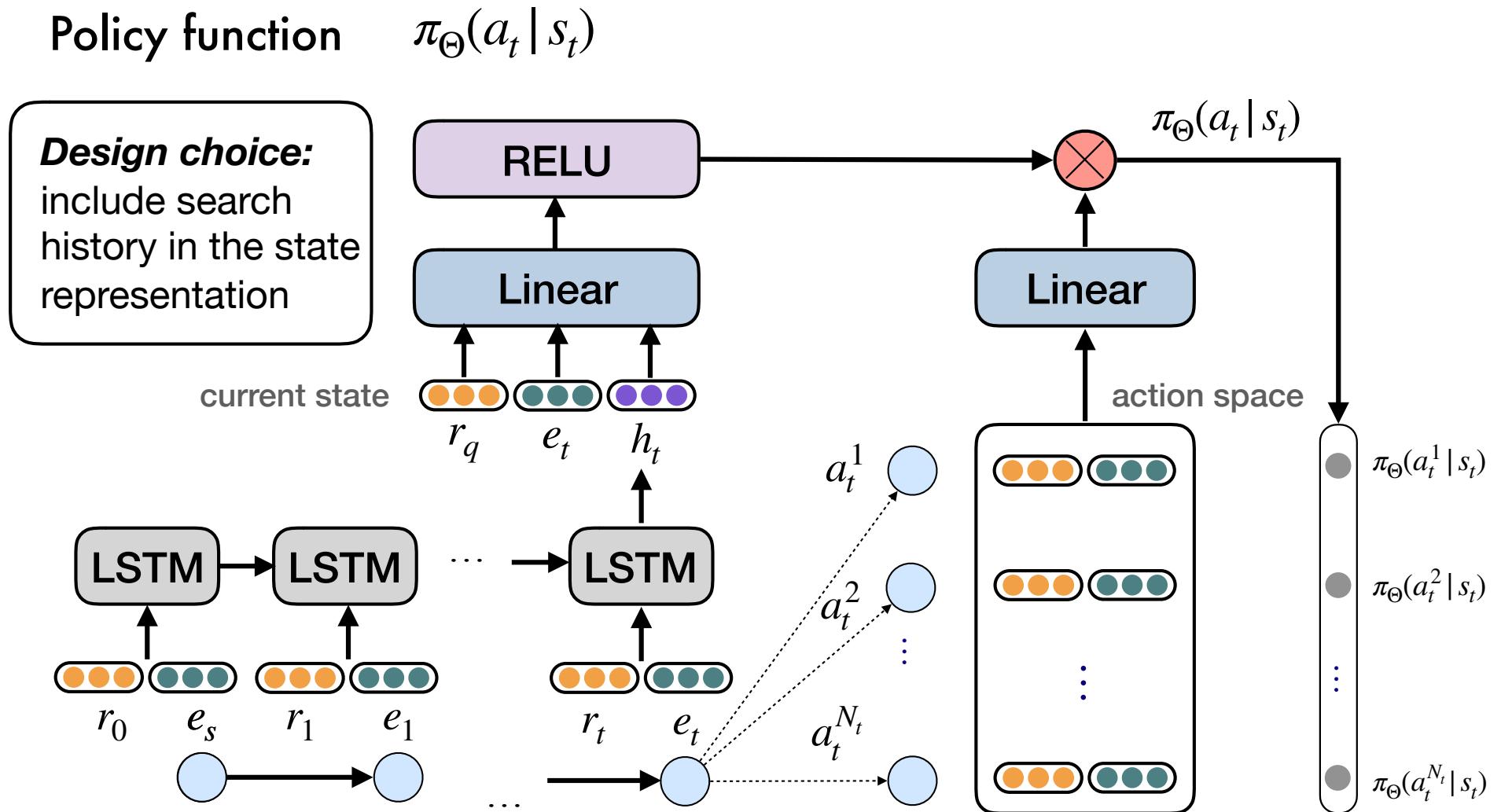
# Policy Gradient

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

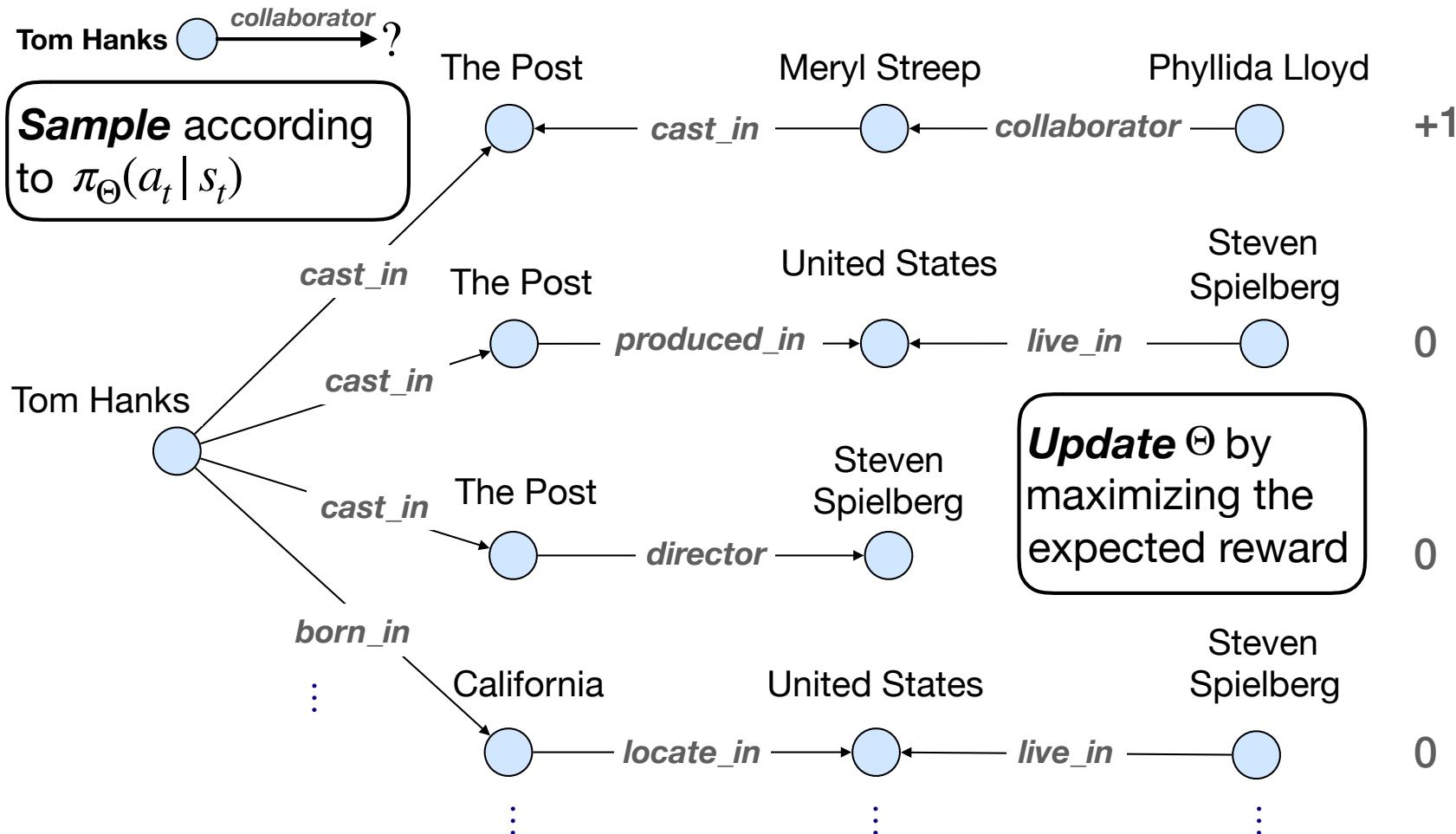
**Probability** of selecting  
an action given the  
current state



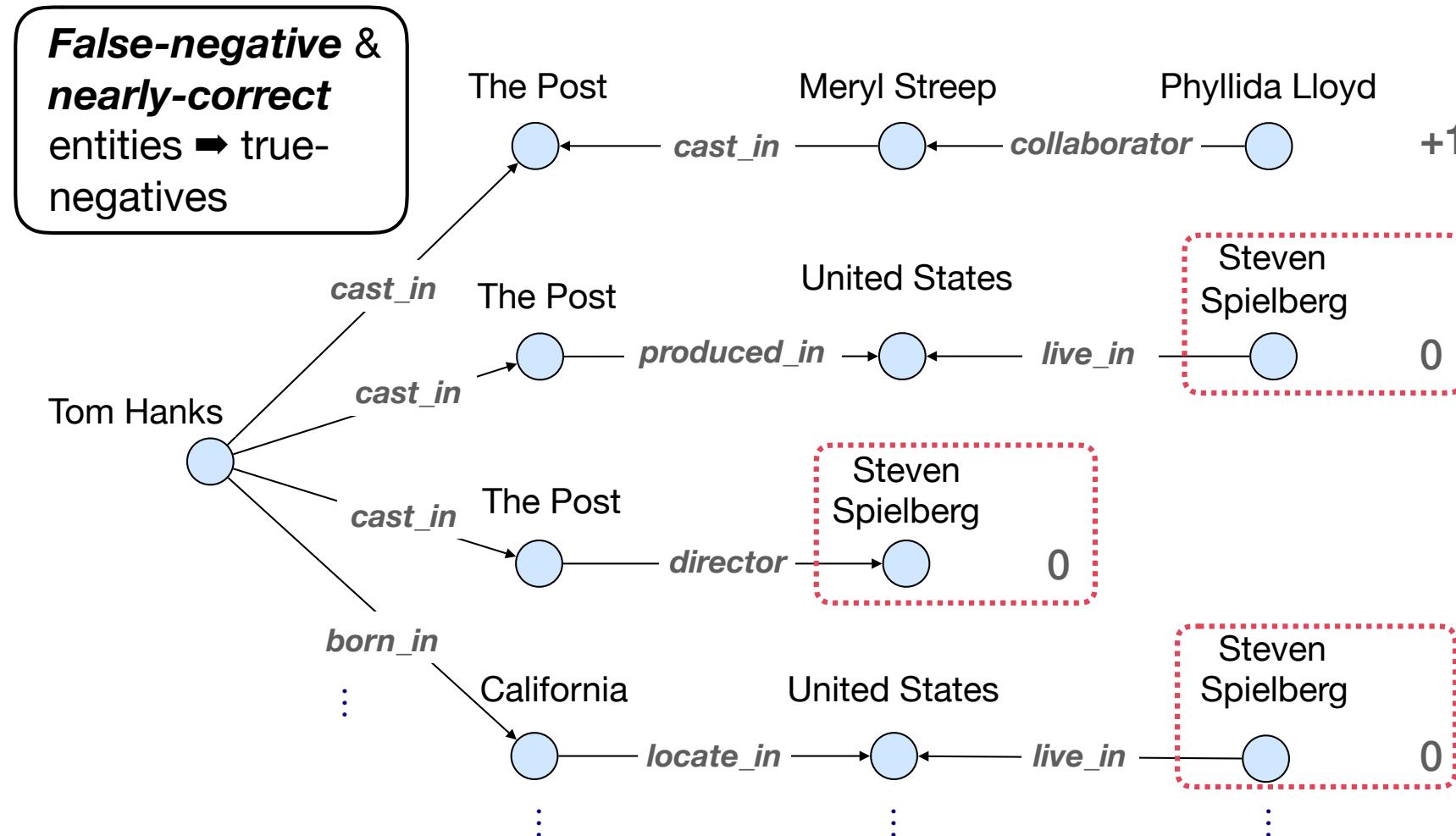
# Policy Gradient



# REINFORCE Training



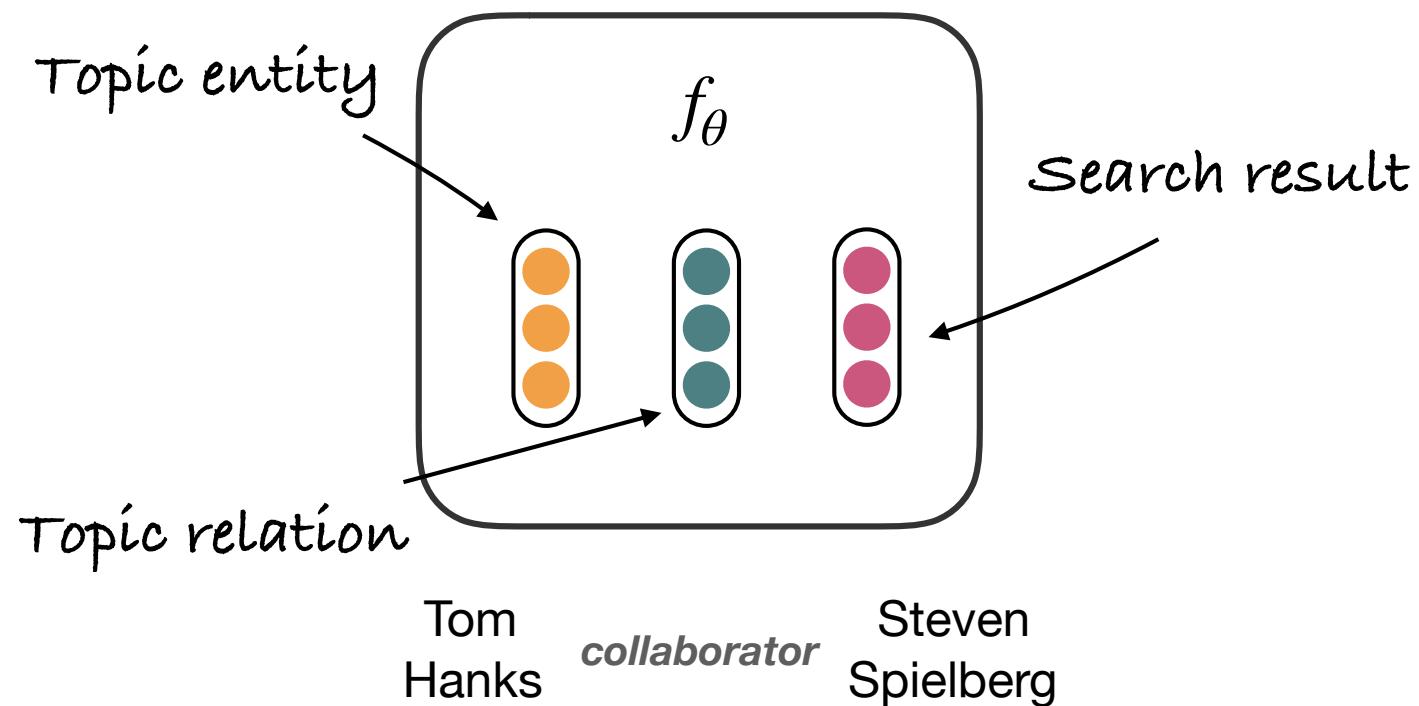
# REINFORCE Training



# Reward Shaping

Unobserved facts

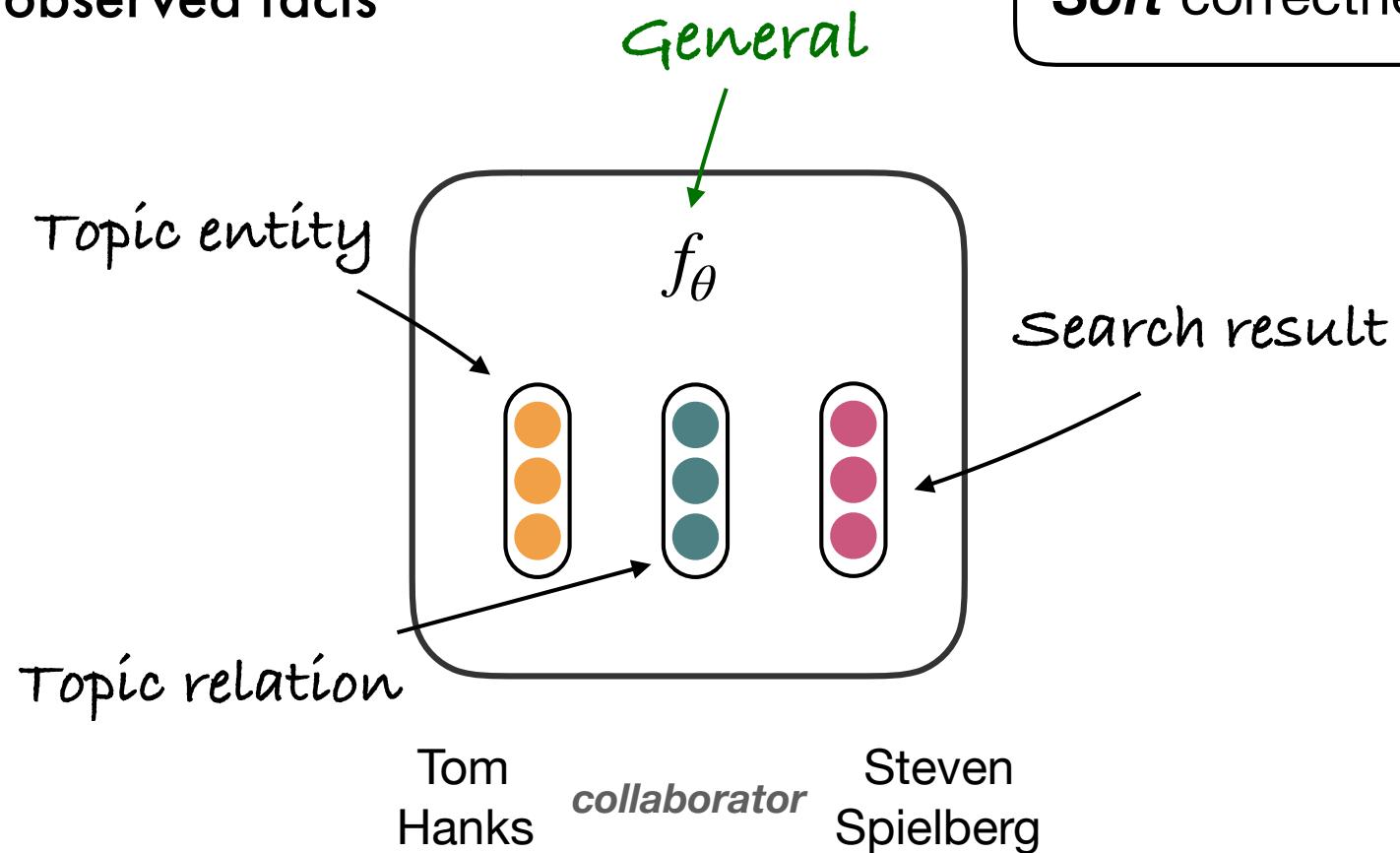
**Soft** correctness



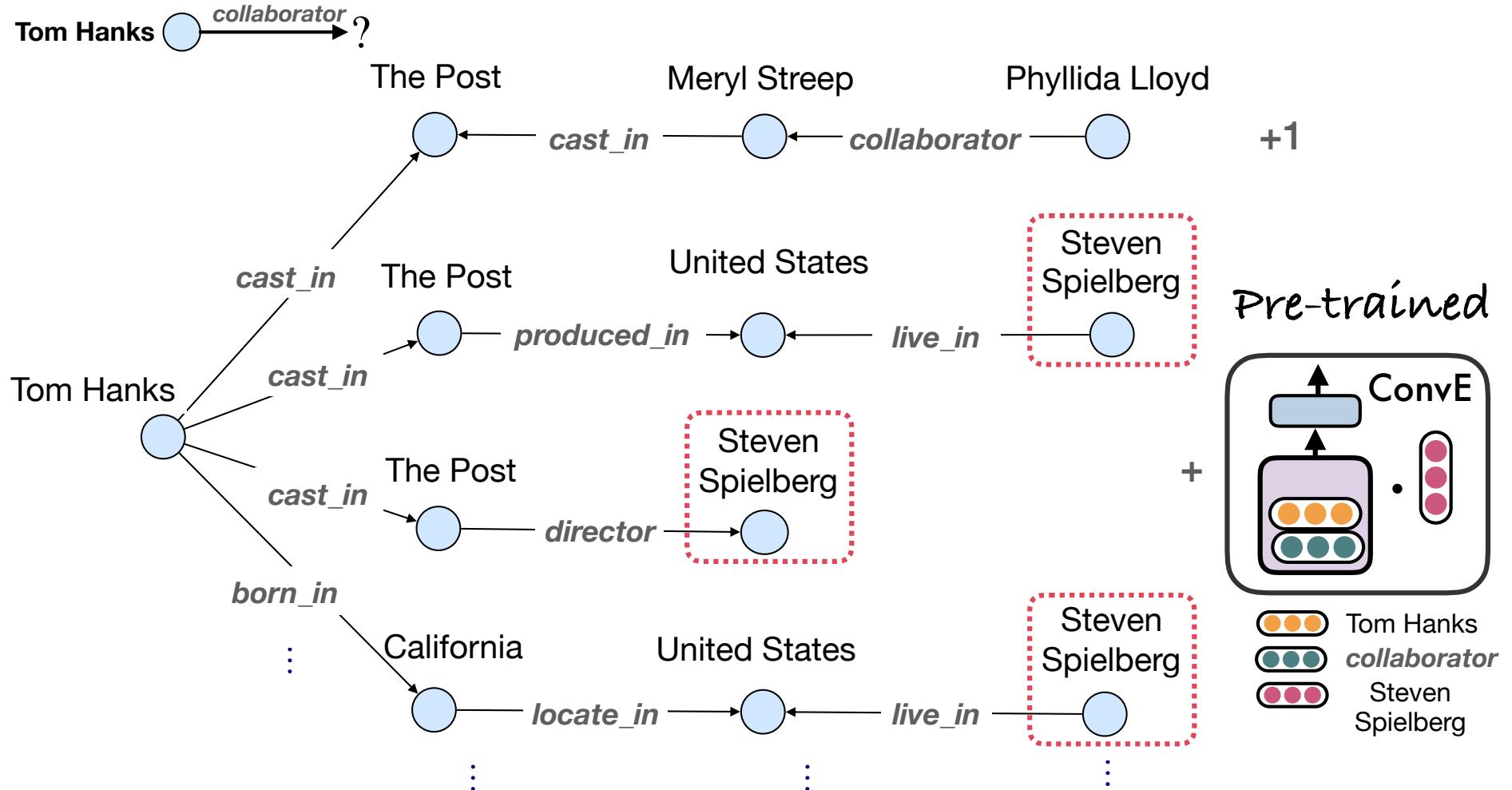
# Reward Shaping

Unobserved facts

**Soft** correctness



# Reward Shaping



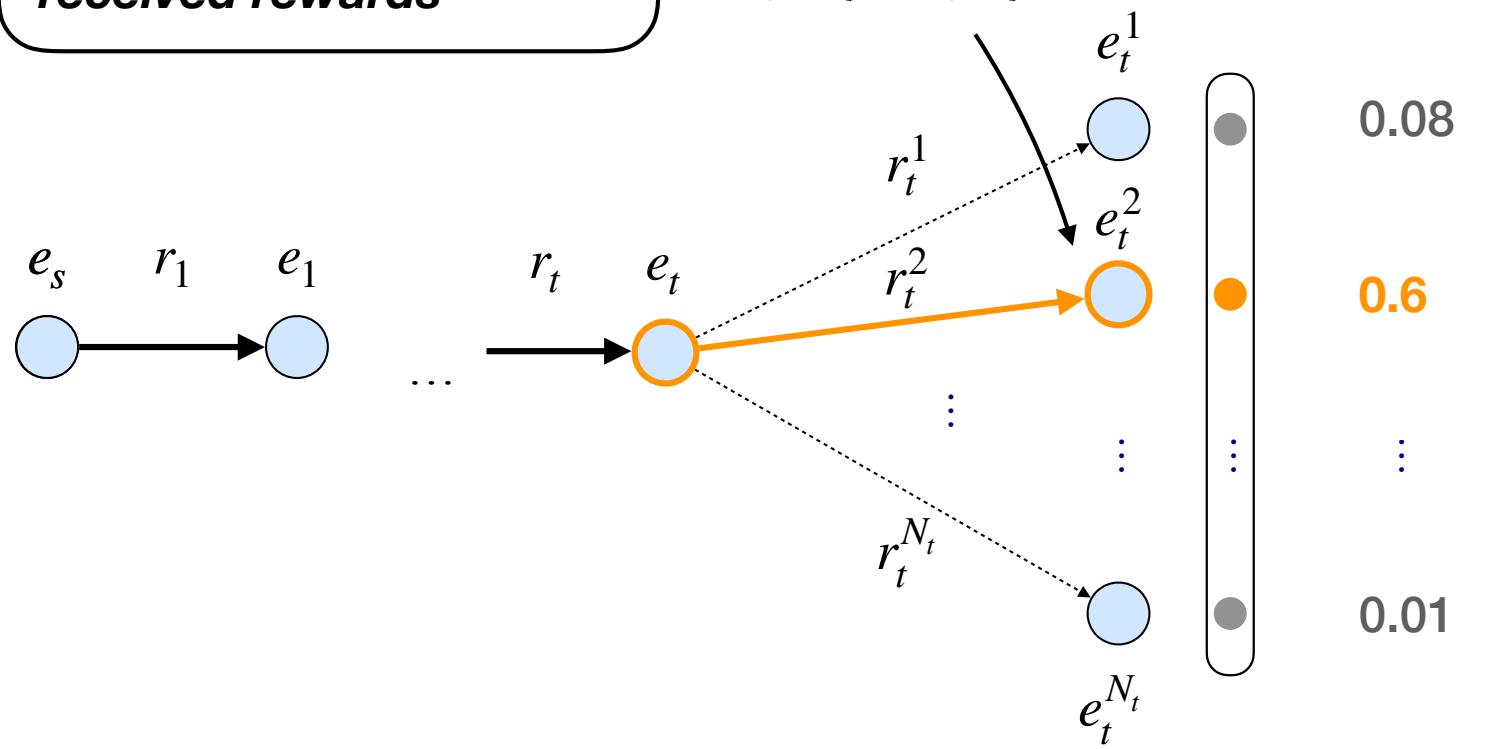
# Action Dropout

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

# Action Dropout

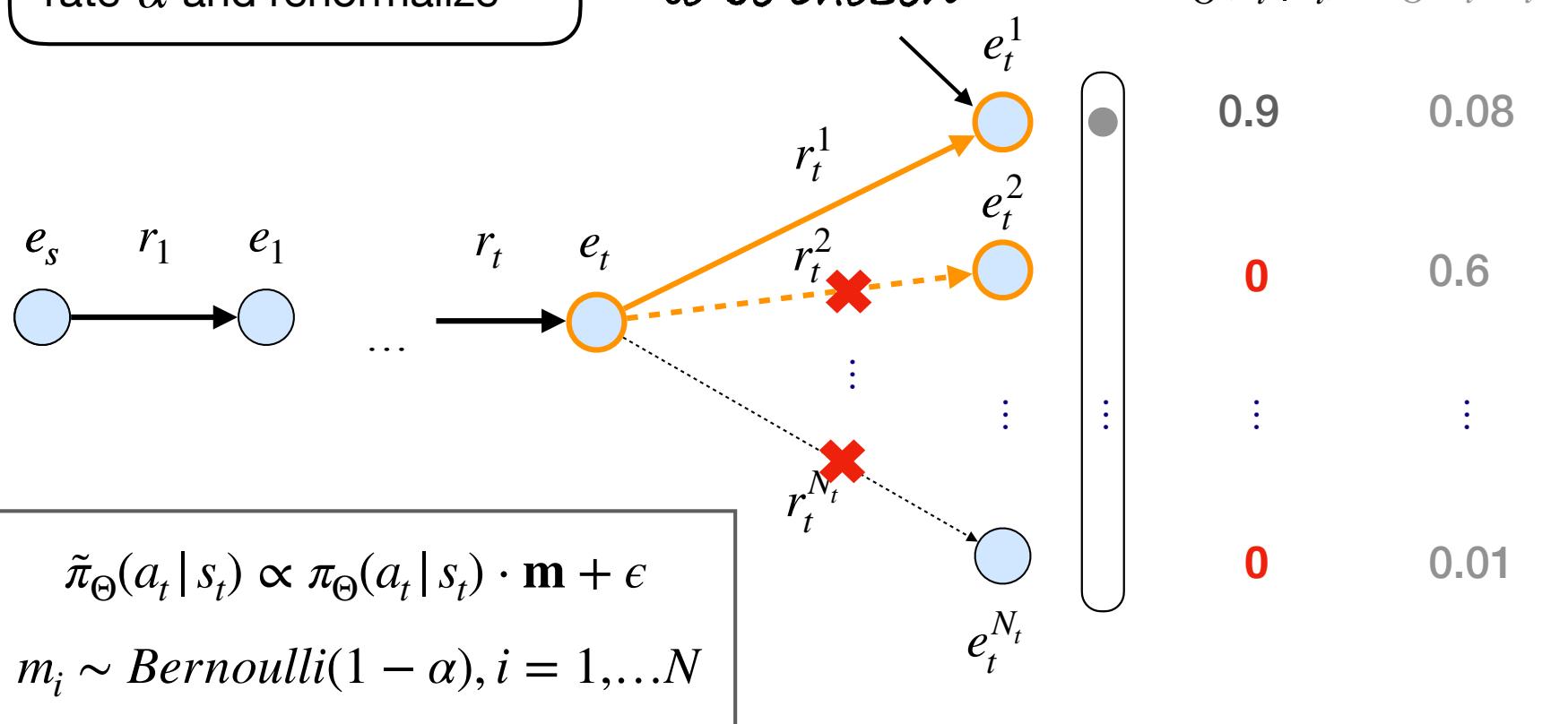
Intuition: **avoid sticking to**  
past actions that **had**  
**received rewards**

More likely  
to be chosen



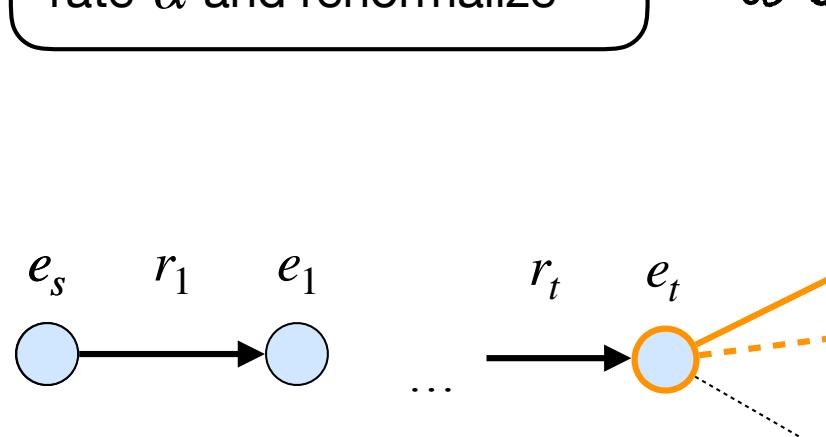
# Action Dropout

Randomly offset the **sampling probabilities** w/  
rate  $\alpha$  and renormalize



# Action Dropout

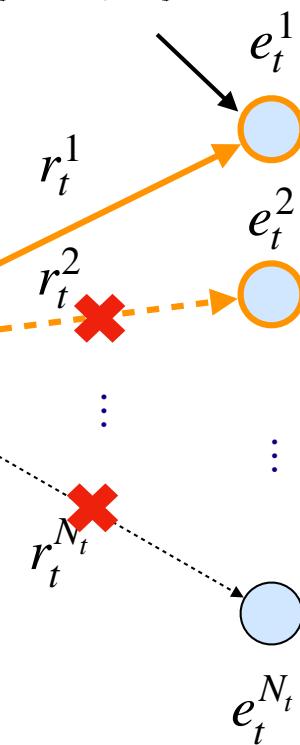
Randomly offset the **sampling probabilities** w/  
rate  $\alpha$  and renormalize



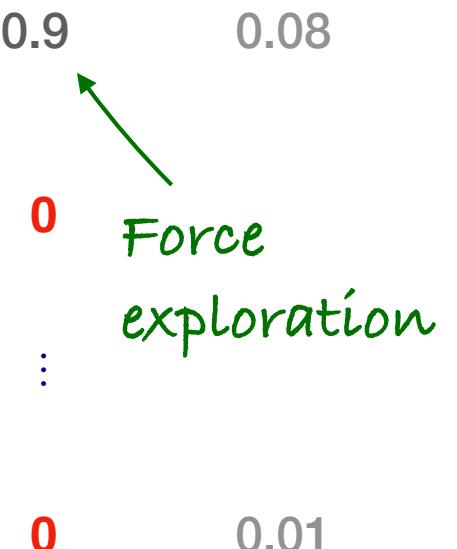
$$\tilde{\pi}_{\Theta}(a_t | s_t) \propto \pi_{\Theta}(a_t | s_t) \cdot \mathbf{m} + \epsilon$$

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

More likely  
to be chosen

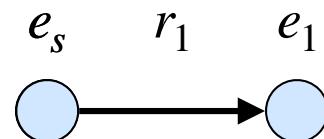


$$\tilde{\pi}_{\Theta}(a_t^i | s_t) \quad \pi_{\Theta}(a_t^i | s_t)$$



# Action Dropout

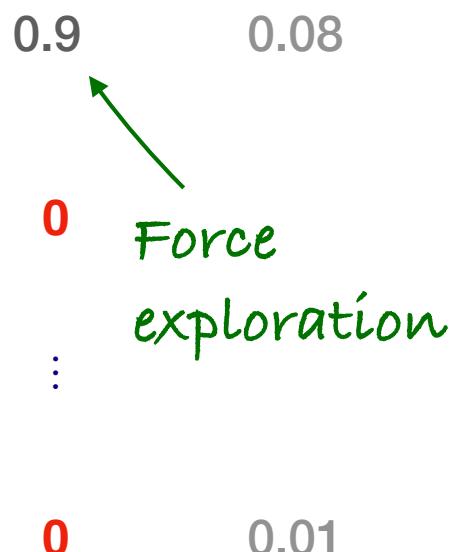
Randomly offset the **sampling probabilities** w/  
rate  $\alpha$  and renormalize



More likely  
to be chosen

up to  $\times 8$  improvement

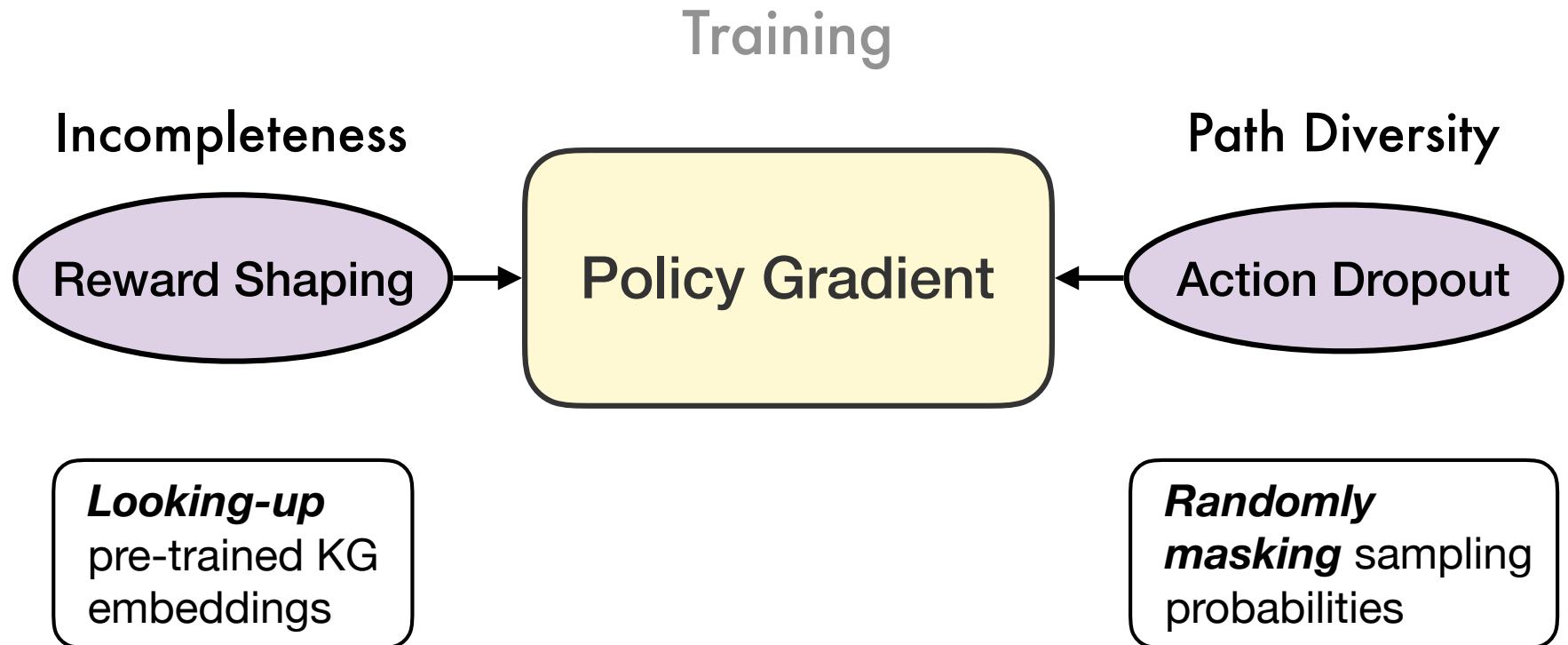
$$\tilde{\pi}_{\Theta}(a_t^i | s_t) \quad \pi_{\Theta}(a_t^i | s_t)$$



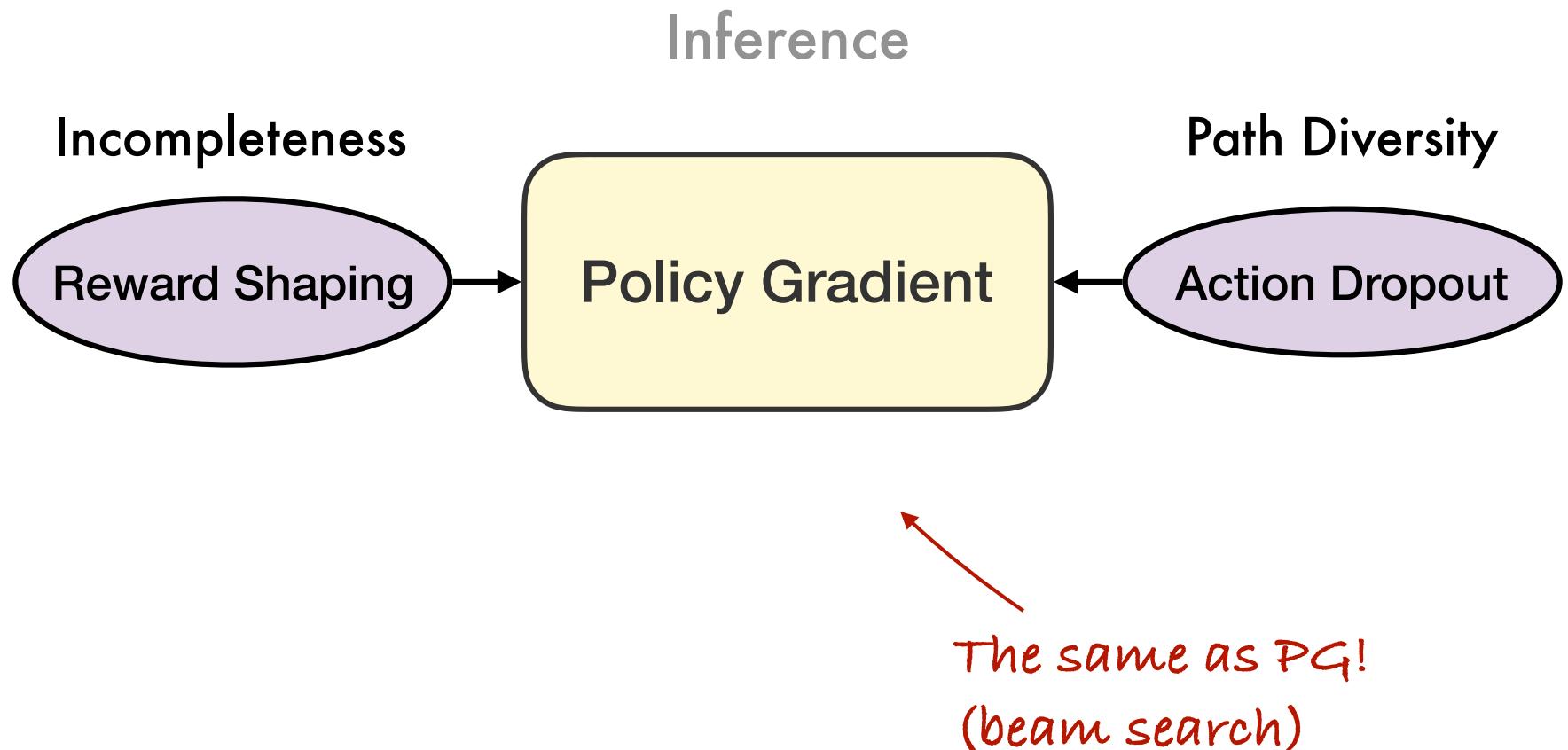
$$\tilde{\pi}_{\Theta}(a_t | s_t) \propto \pi_{\Theta}(a_t | s_t) \cdot \mathbf{m} + \epsilon$$

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

# Efficient Training & Inference



# Efficient Training & Inference



# 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

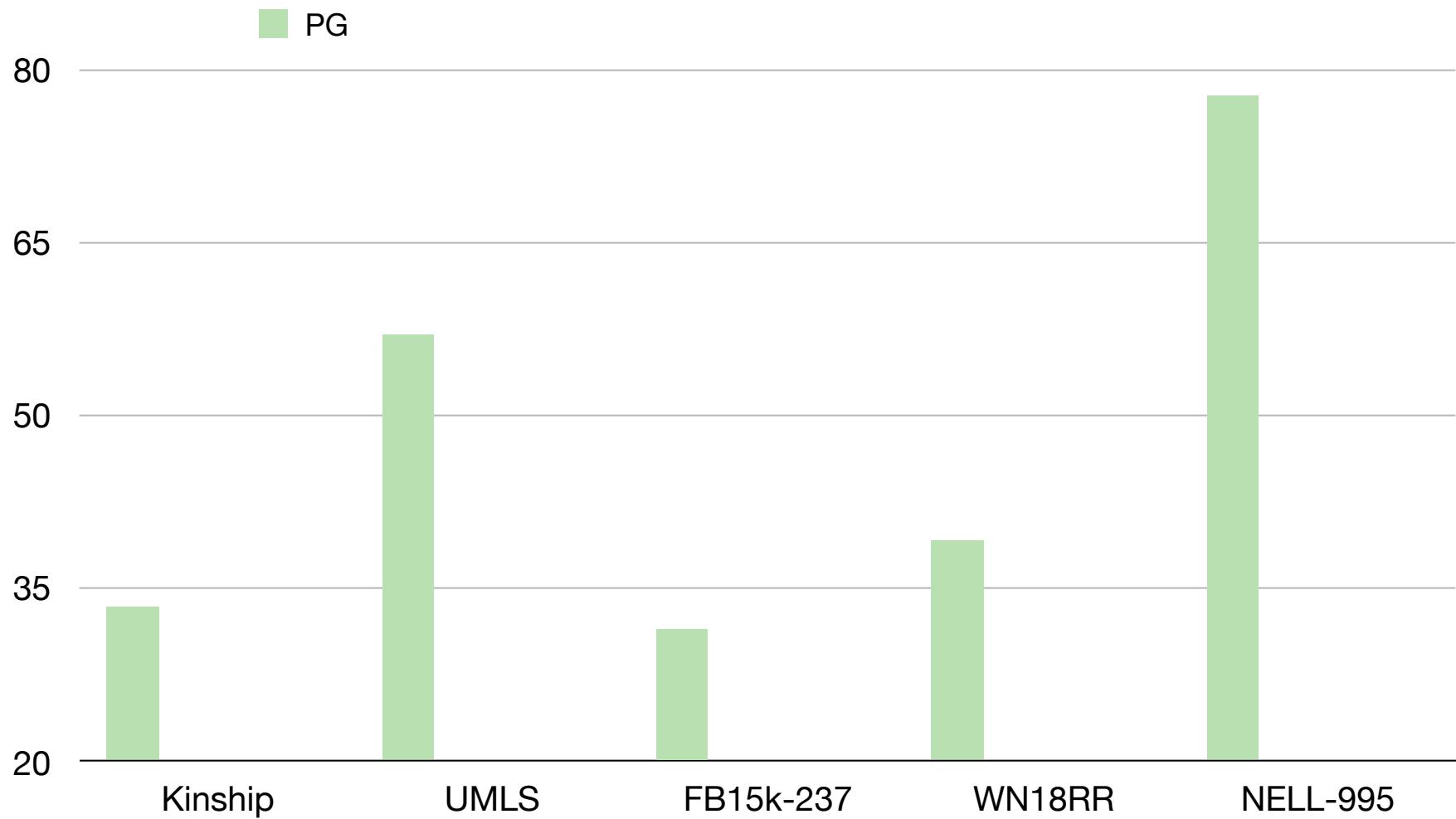


Fig 2. Dev set MRR (x100) comparison

# Ablation Studies

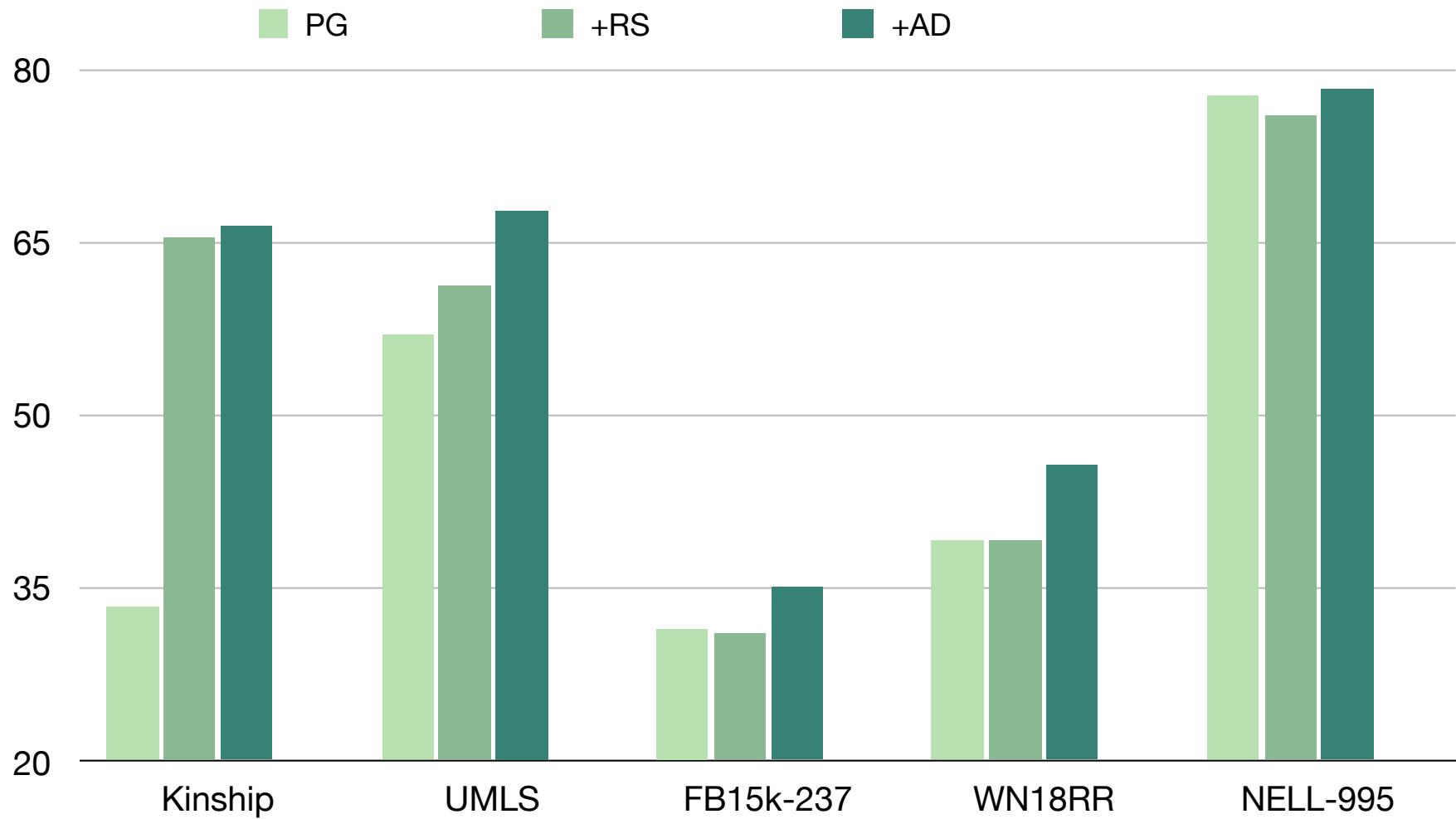


Fig 2. Dev set MRR (x100) comparison

# Ablation Studies

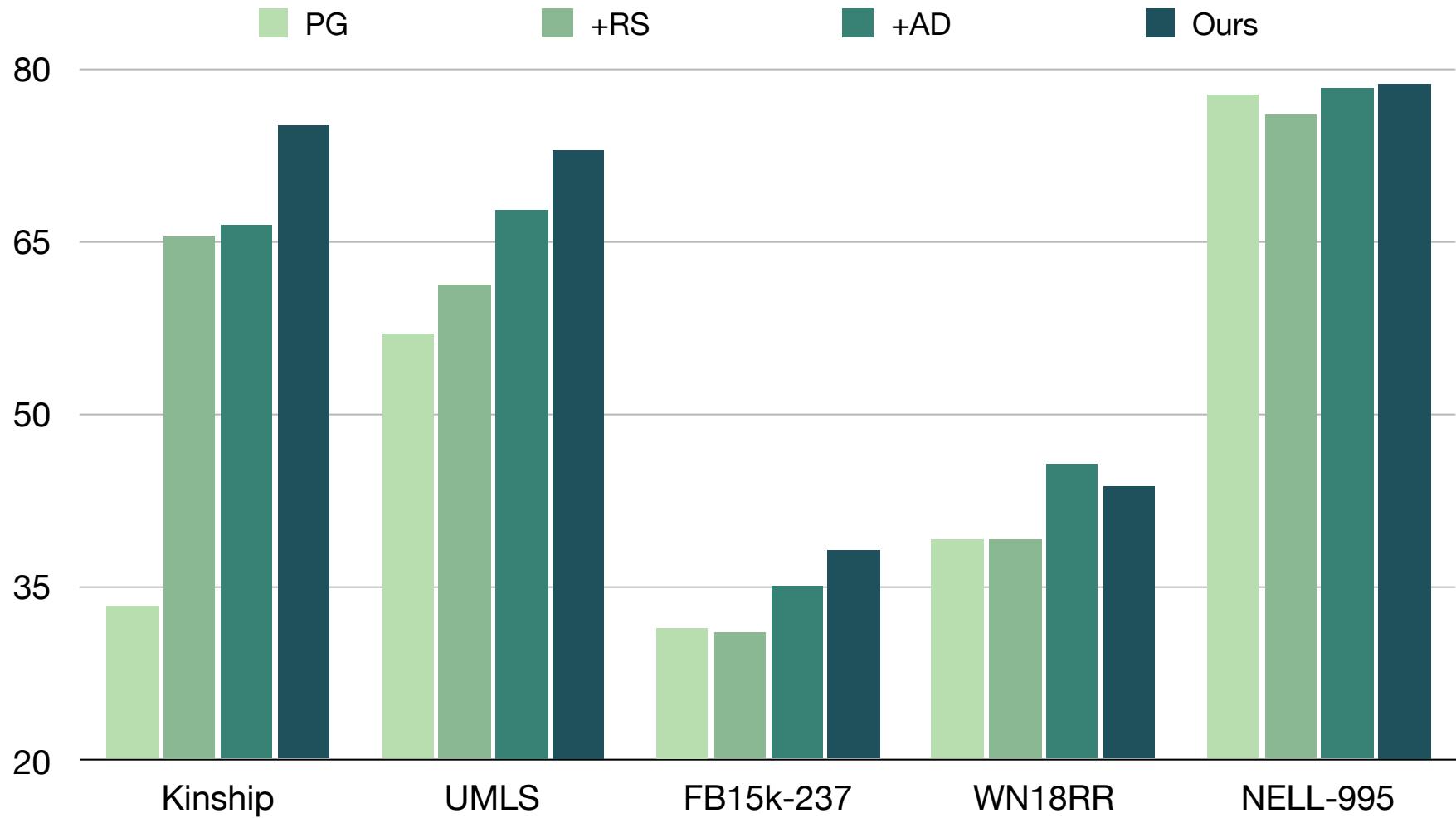


Fig 2. Dev set MRR (x100) comparison

# Ablation Studies

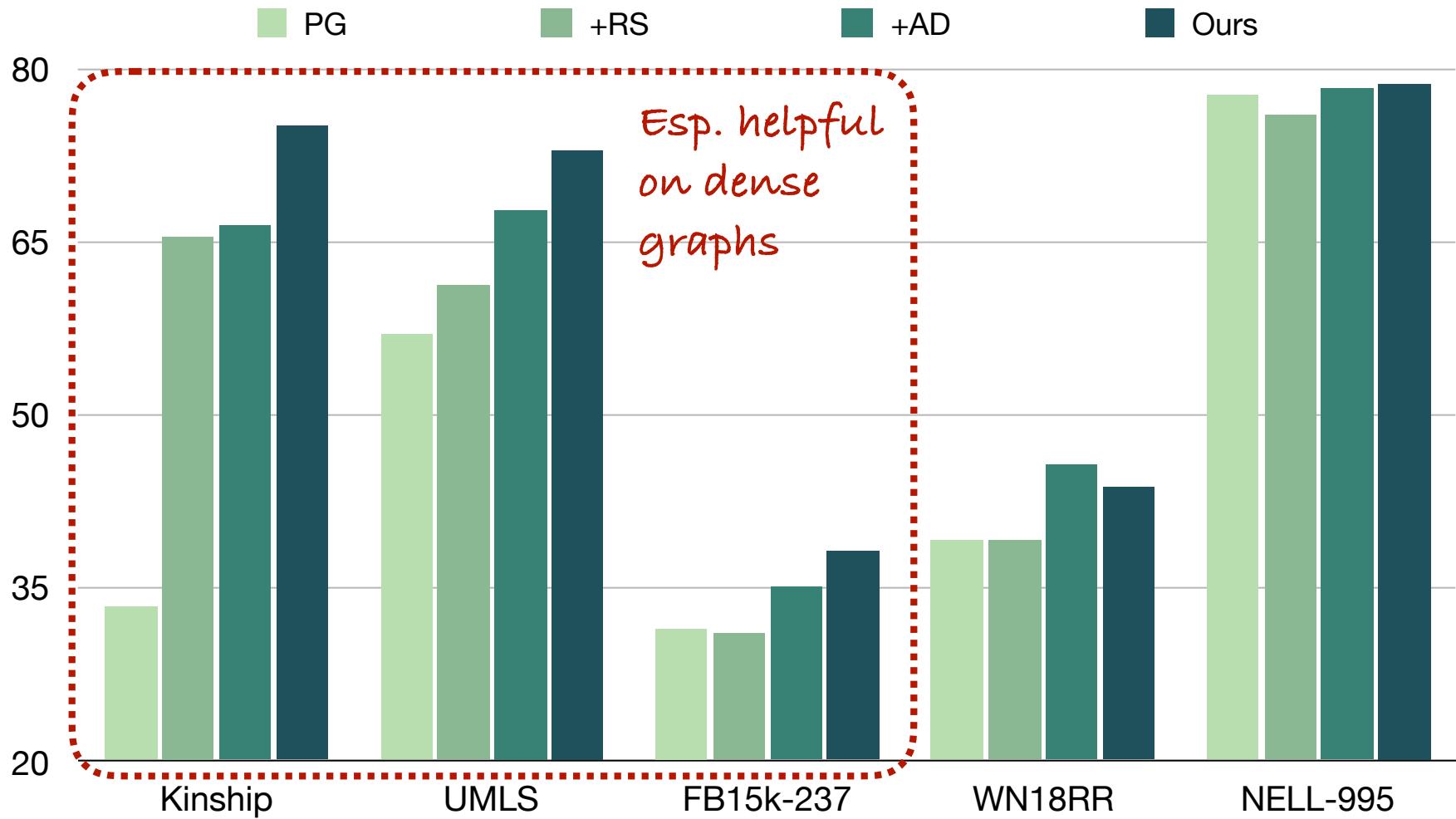


Fig 2. Dev set MRR (x100) comparison

# Main Results

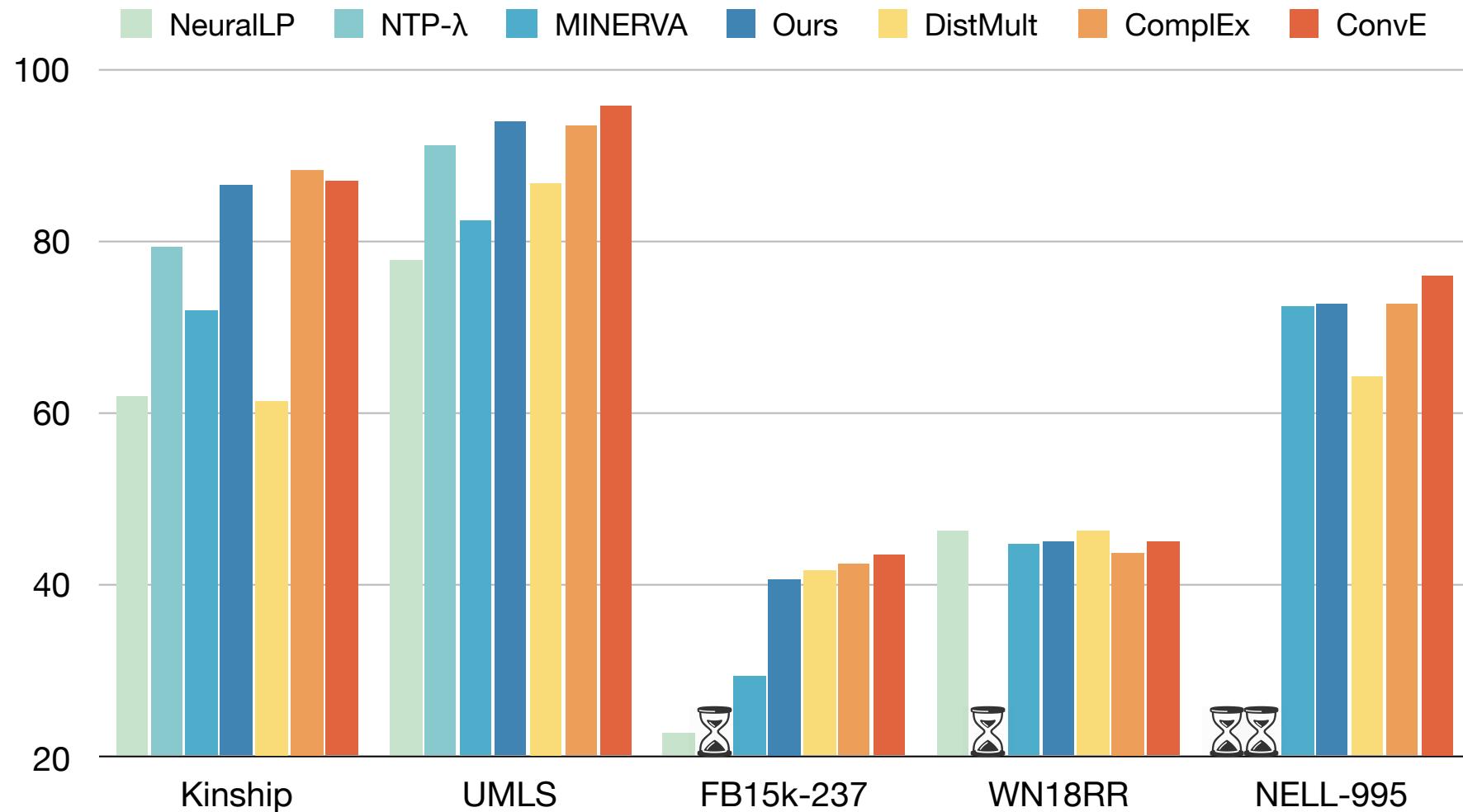


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

# Main Results

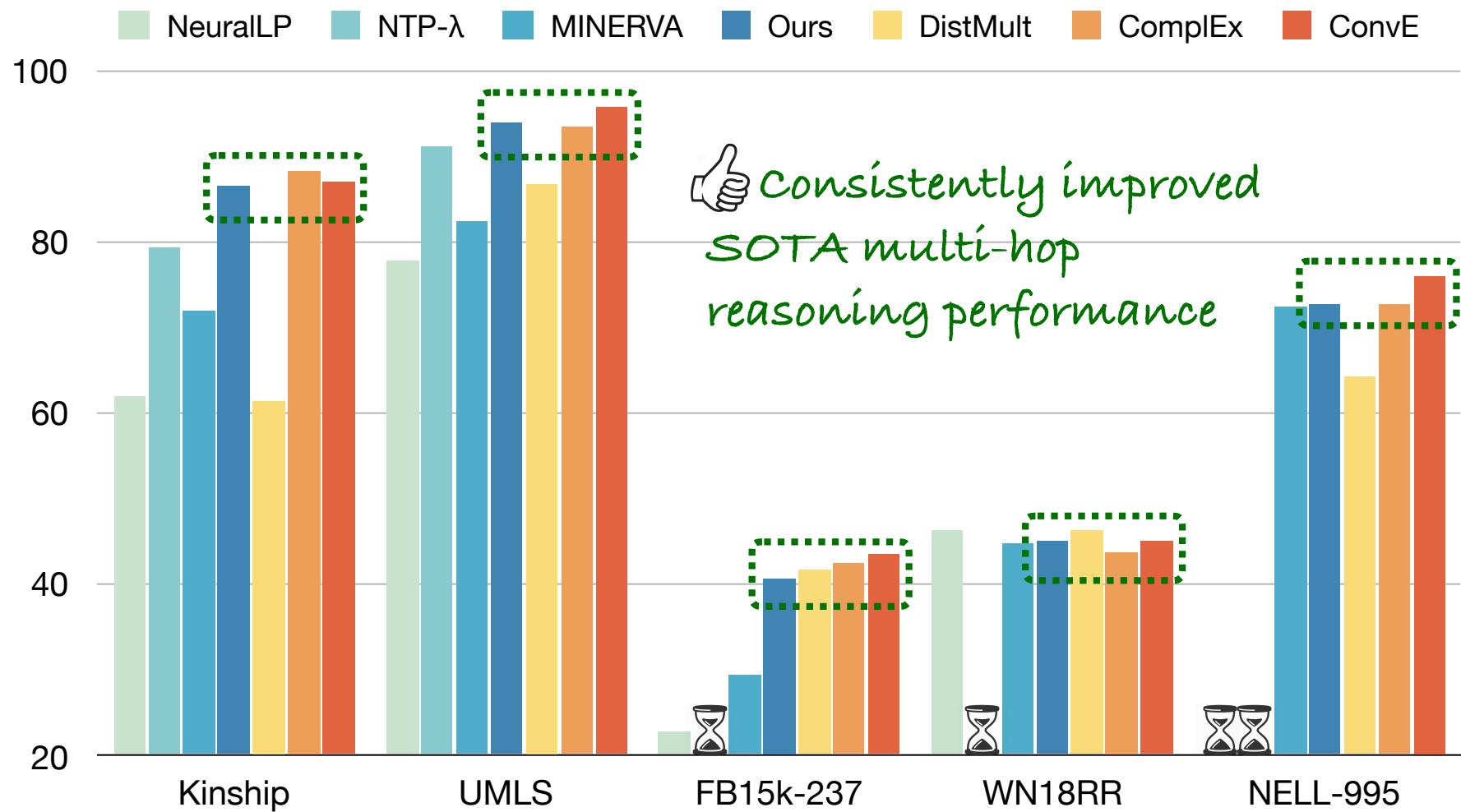
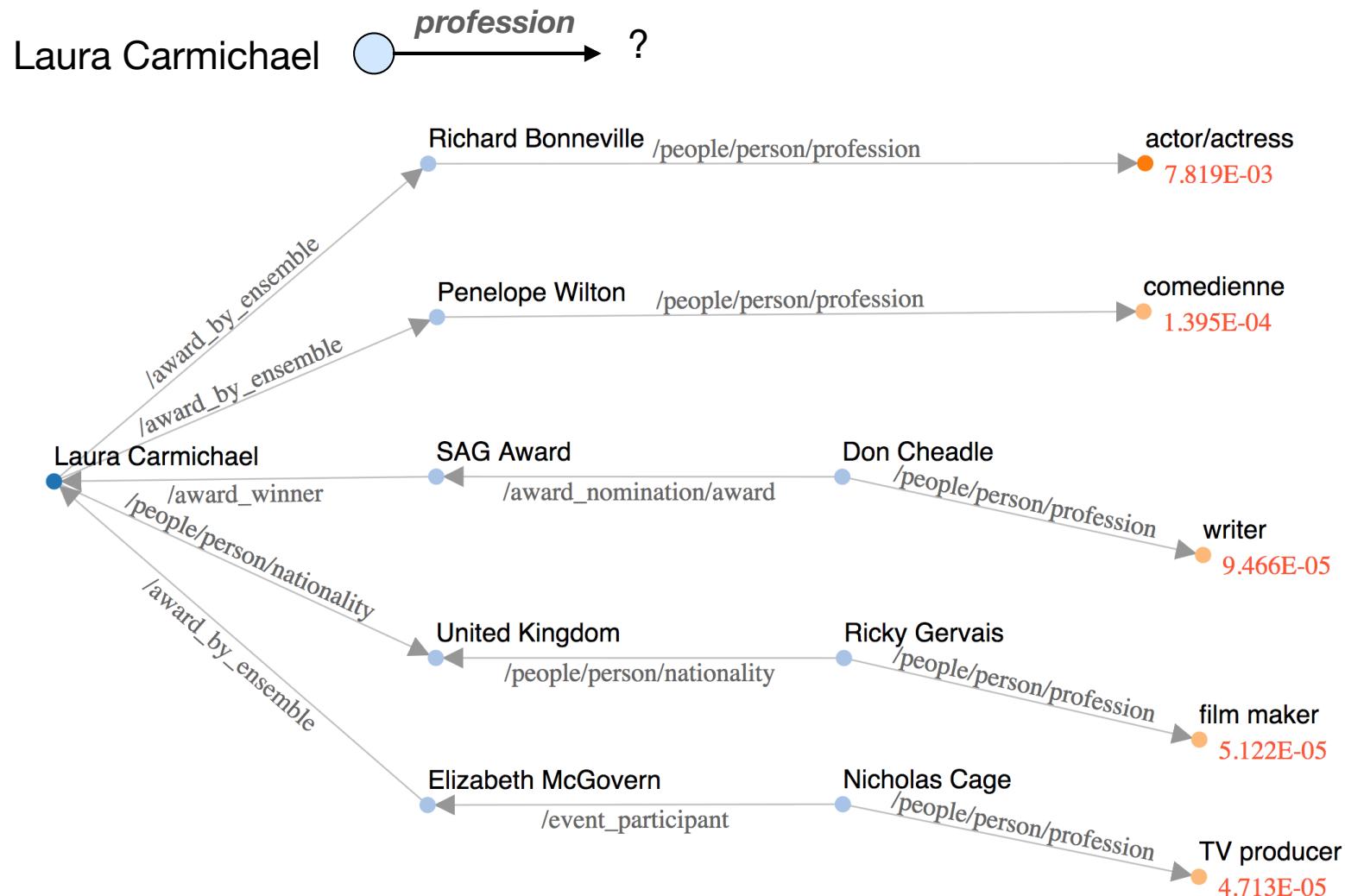
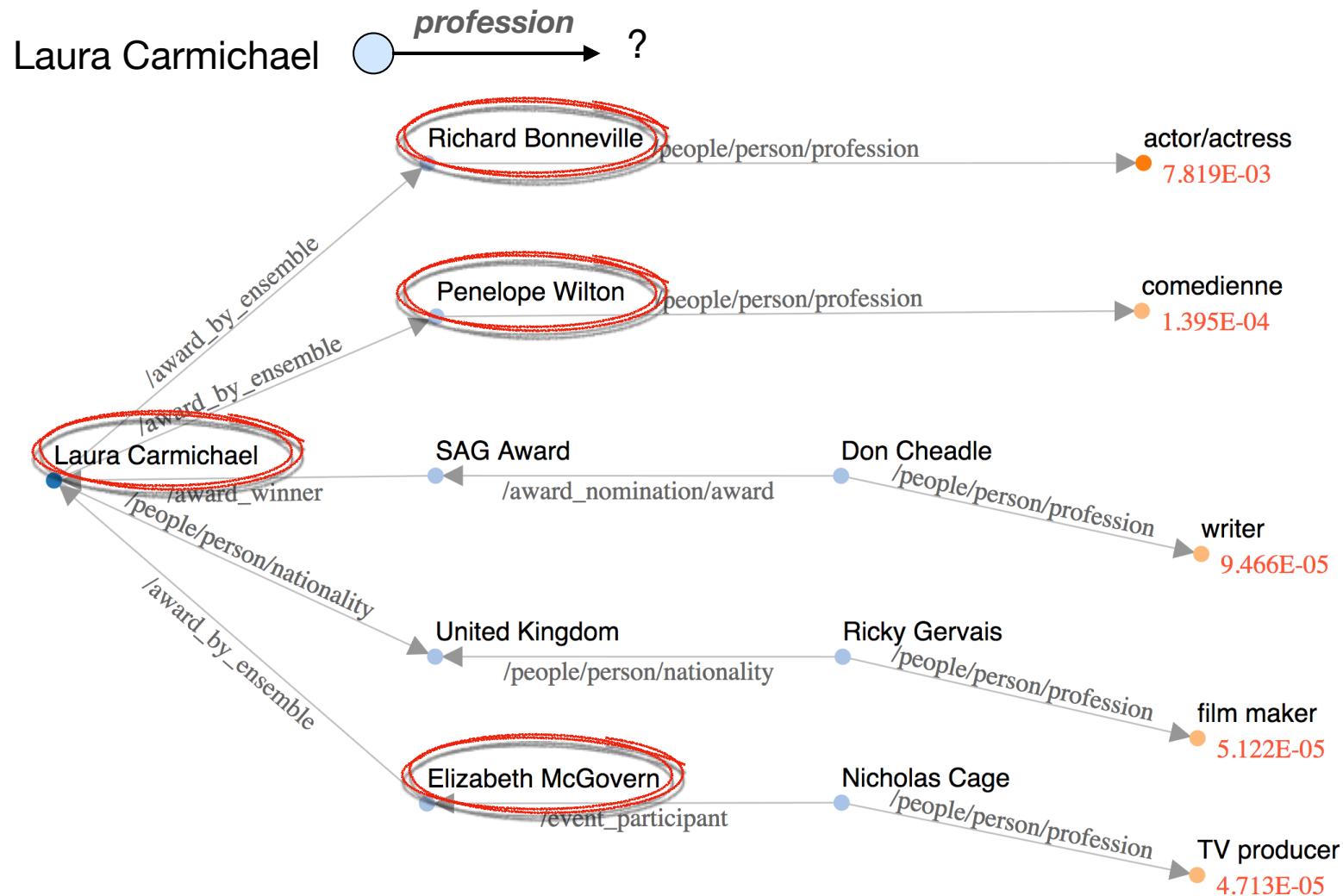


Fig 3. Test set MRR (x100) compared to SOTA multi-hop reasoning and embedding-based approaches

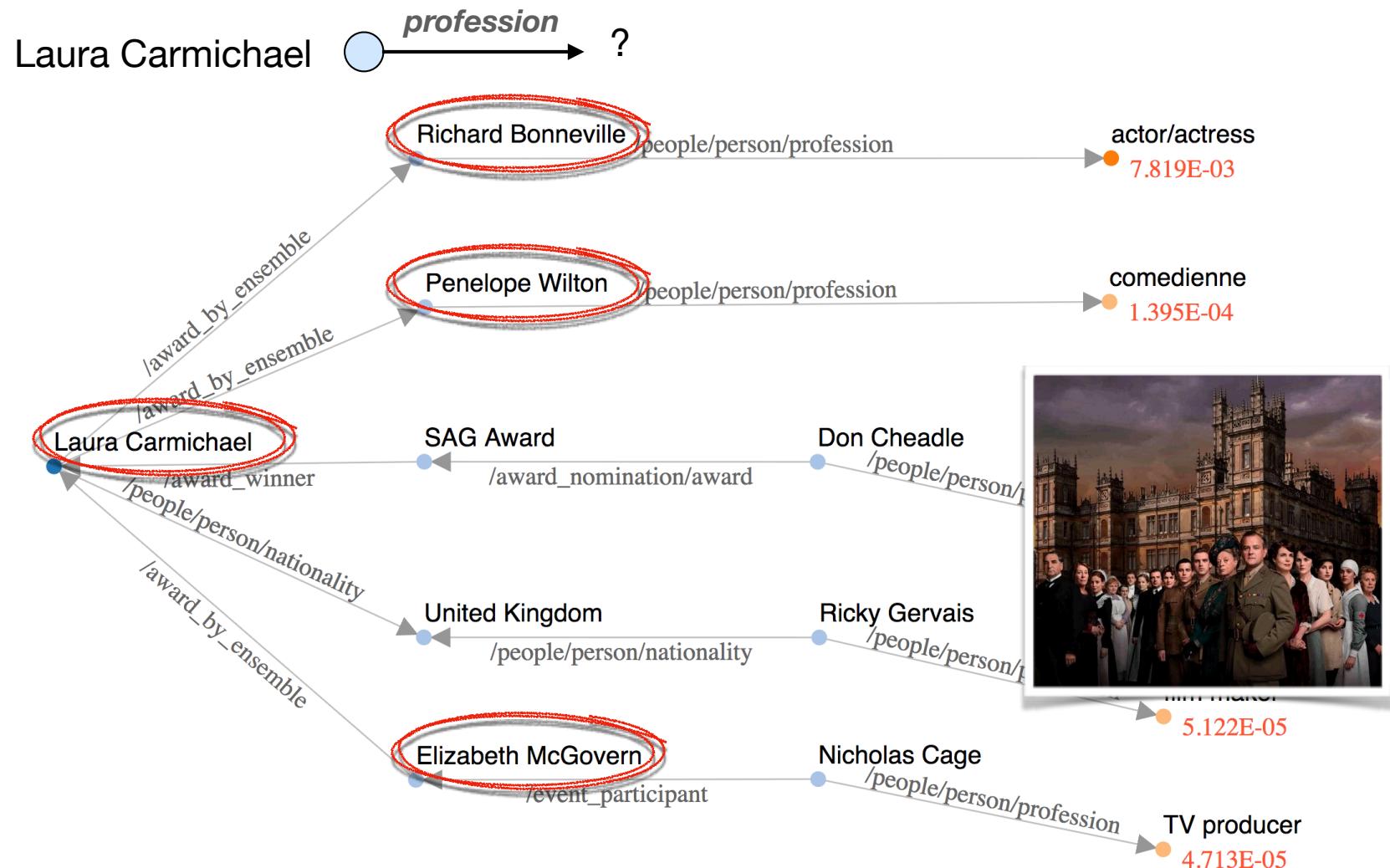
# Interpretable Results



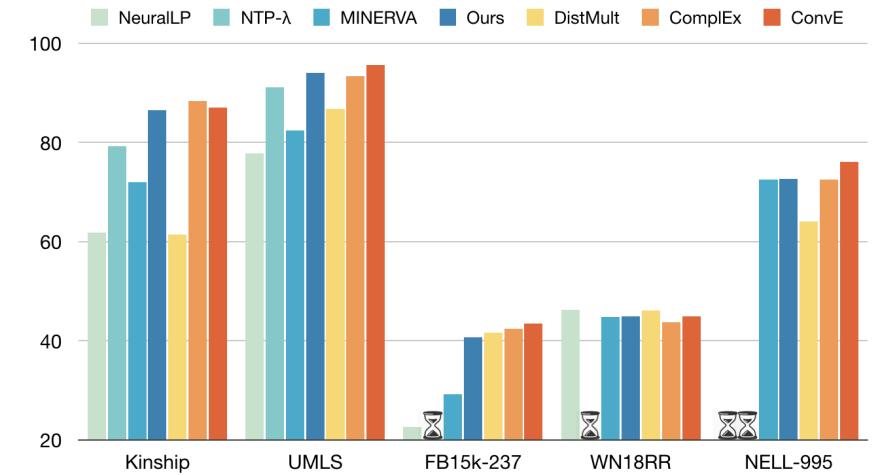
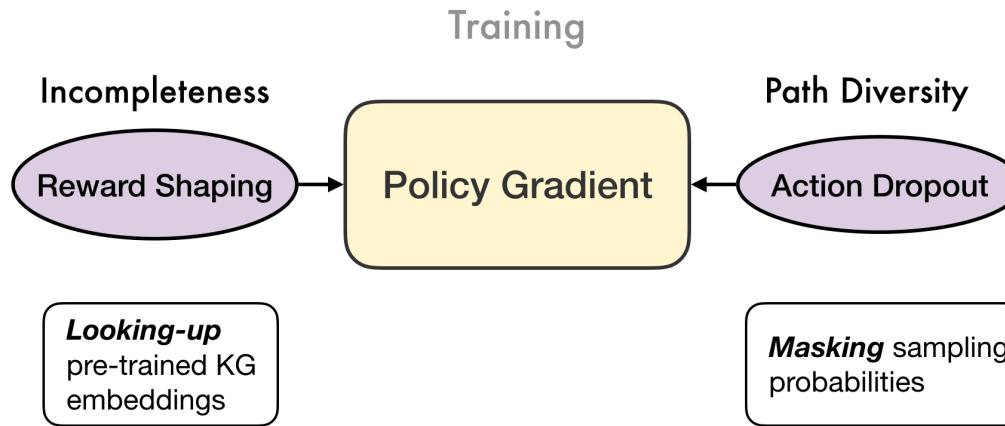
# Interpretable Results



# 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

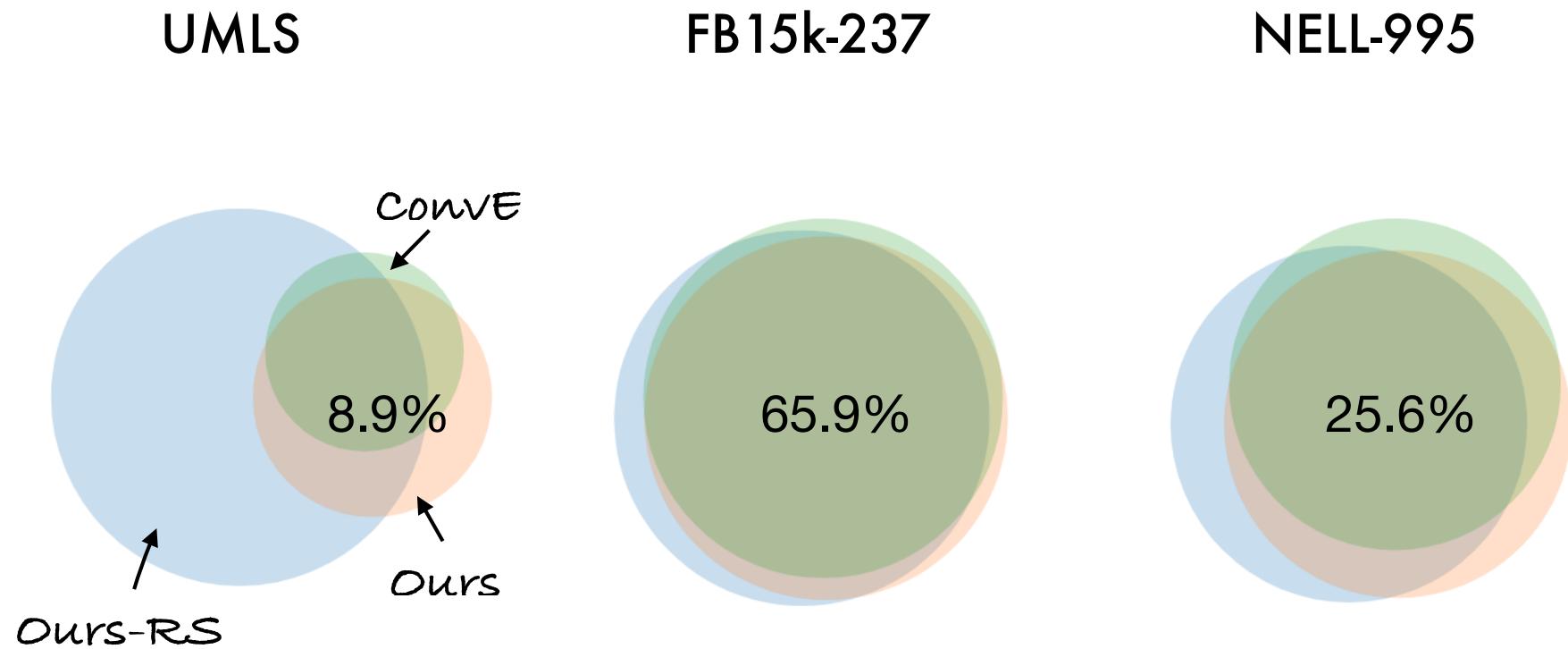


Fig 4. Dev set top-1 prediction error overlap of ConvE, Ours and Ours-RS. The absolute error rate of Ours is shown.

# BK II - Challenges

## Incompleteness

$\approx 30\%$  false negative feedback

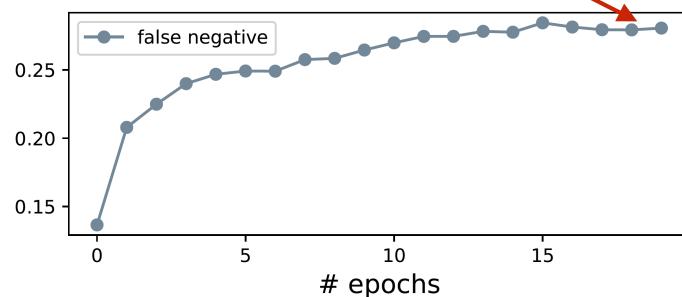


Fig 1. % of false negative hit in the first 20 epochs of RL training on the UMLS KG benchmark (Kok and Domingos 2007)

