Relation Learning with Path Constrained Random Walks

Ni Lao

15-826 Multimedia Databases and Data Mining

School of Computer Science
Carnegie Mellon University

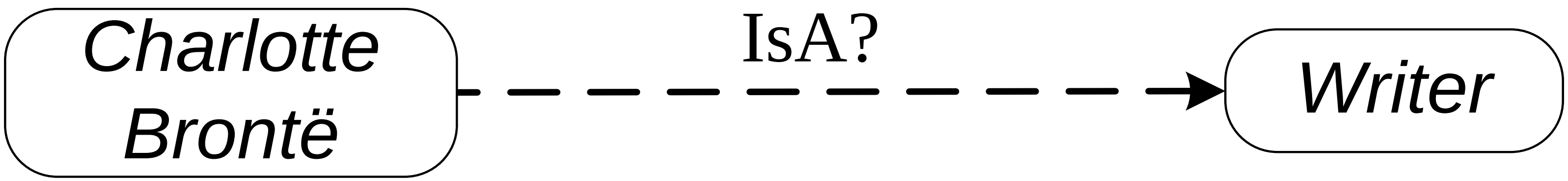
2011-09-27

Outline

- Motivation
  - Relational Learning
  - Random Walk Inference
- Tasks
  - Publication recommendation tasks
  - Inference with knowledge base
- Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Query Independent Paths
  - Popular Entity Biases
- Efficient Inference (Lao & Cohen, KDD 2010)
- Feature Selection (L. M. C., EMNLP 2011)

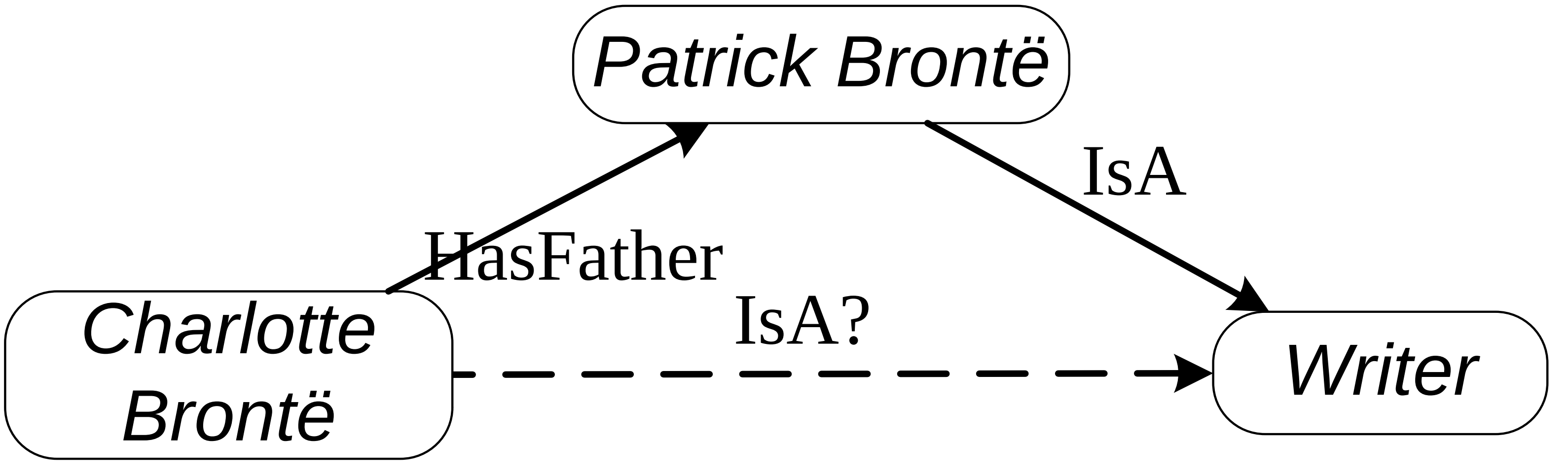
Relational Learning

- Prediction with rich meta-data has great potential and challenge, e.g.



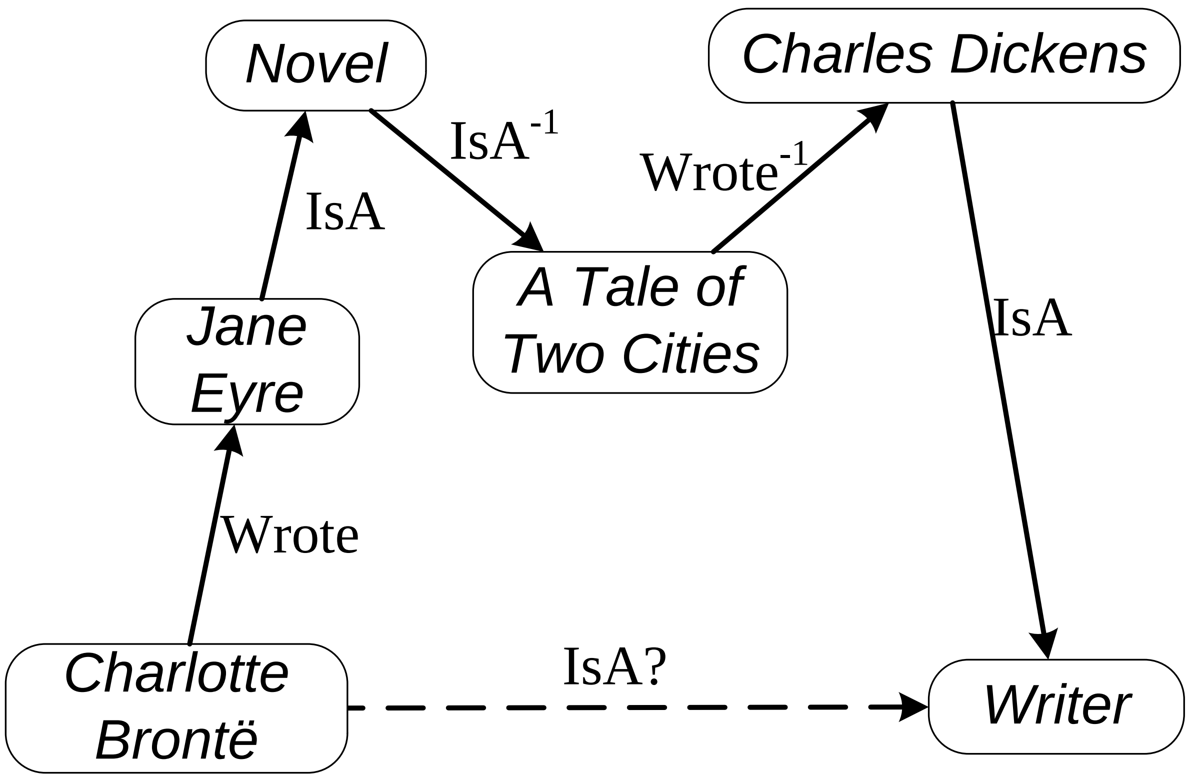
Relational Learning

- Consider friends/family



Relational Learning

- Consider people's behavior

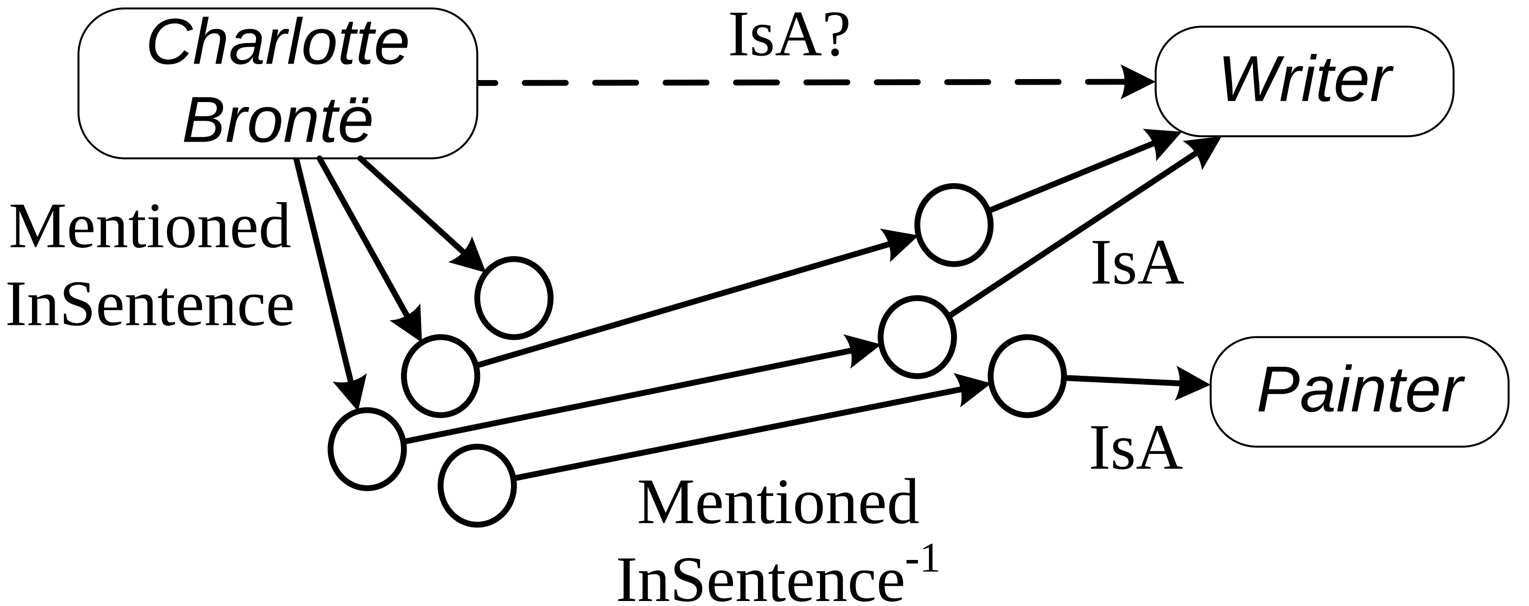


IsA^{-1} is the reverse of IsA relation

$Wrote^{-1}$ is the reverse of Wrote relation

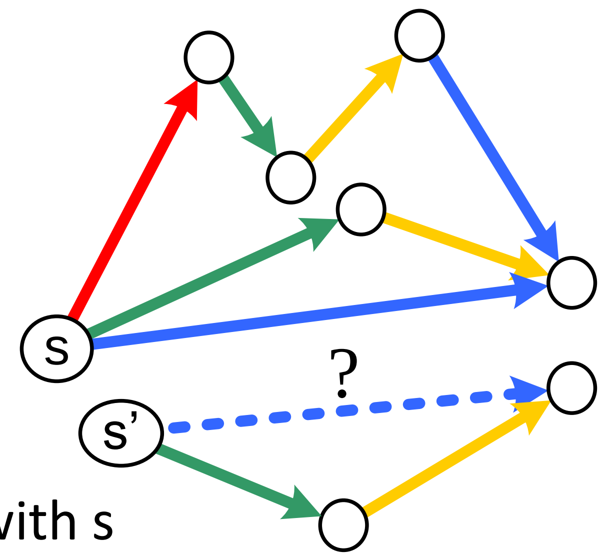
Relational Learning

- Consider literature/publication



Relational Learning

- Task
  – Given
    - a directed heterogeneous graph G
    - a starting node s
    - edge type R
  – Find
    - nodes t which should have edge R with s

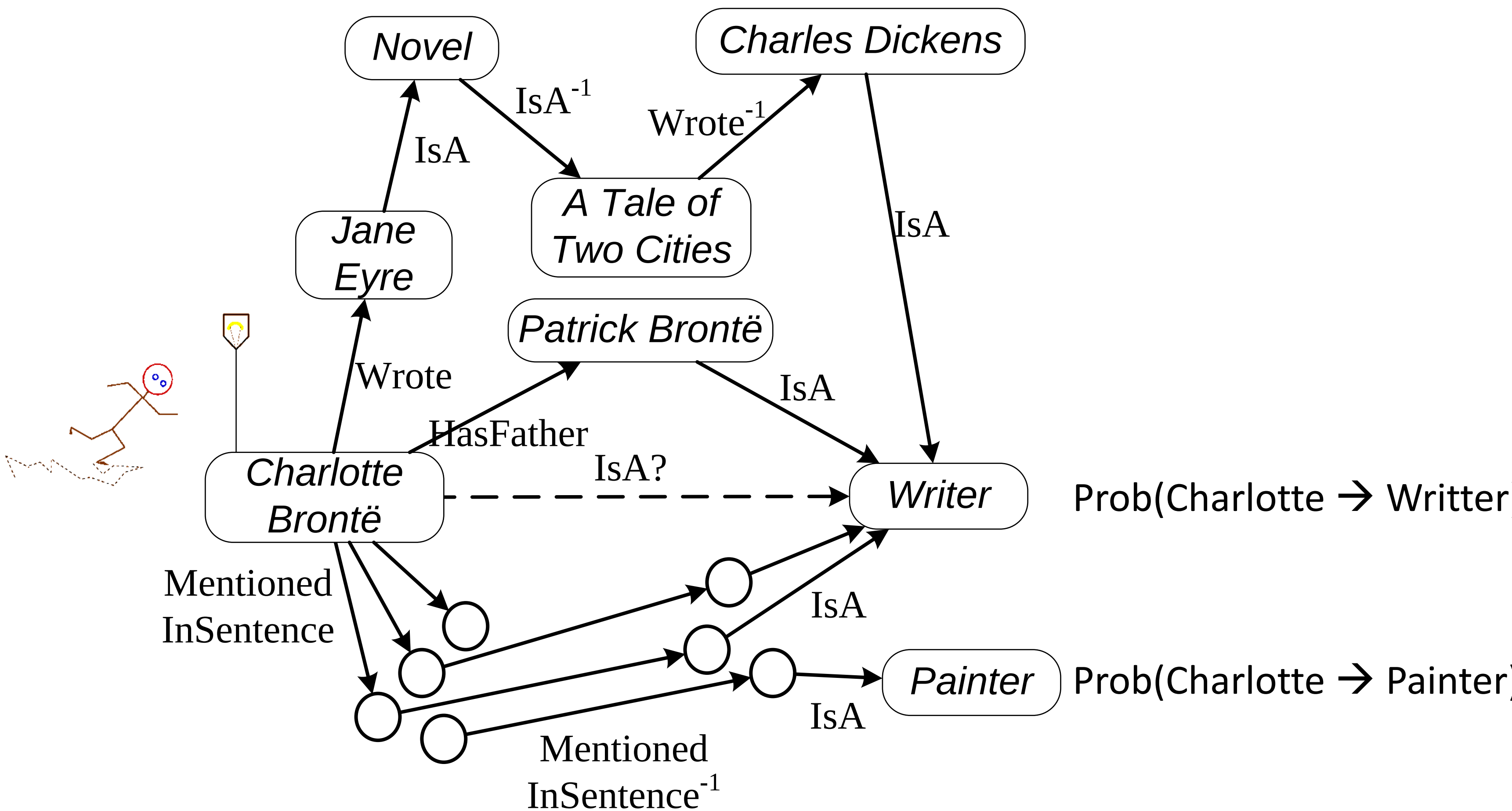


- Challenge
  – statistical learning tools (e.g. SVM) expect samples and their feature values
  – feature engineering needs domain knowledge and is not scalable to the complexity of nowadays' data

Why Not Random Walk with Restart

(Will be covered in later classes)

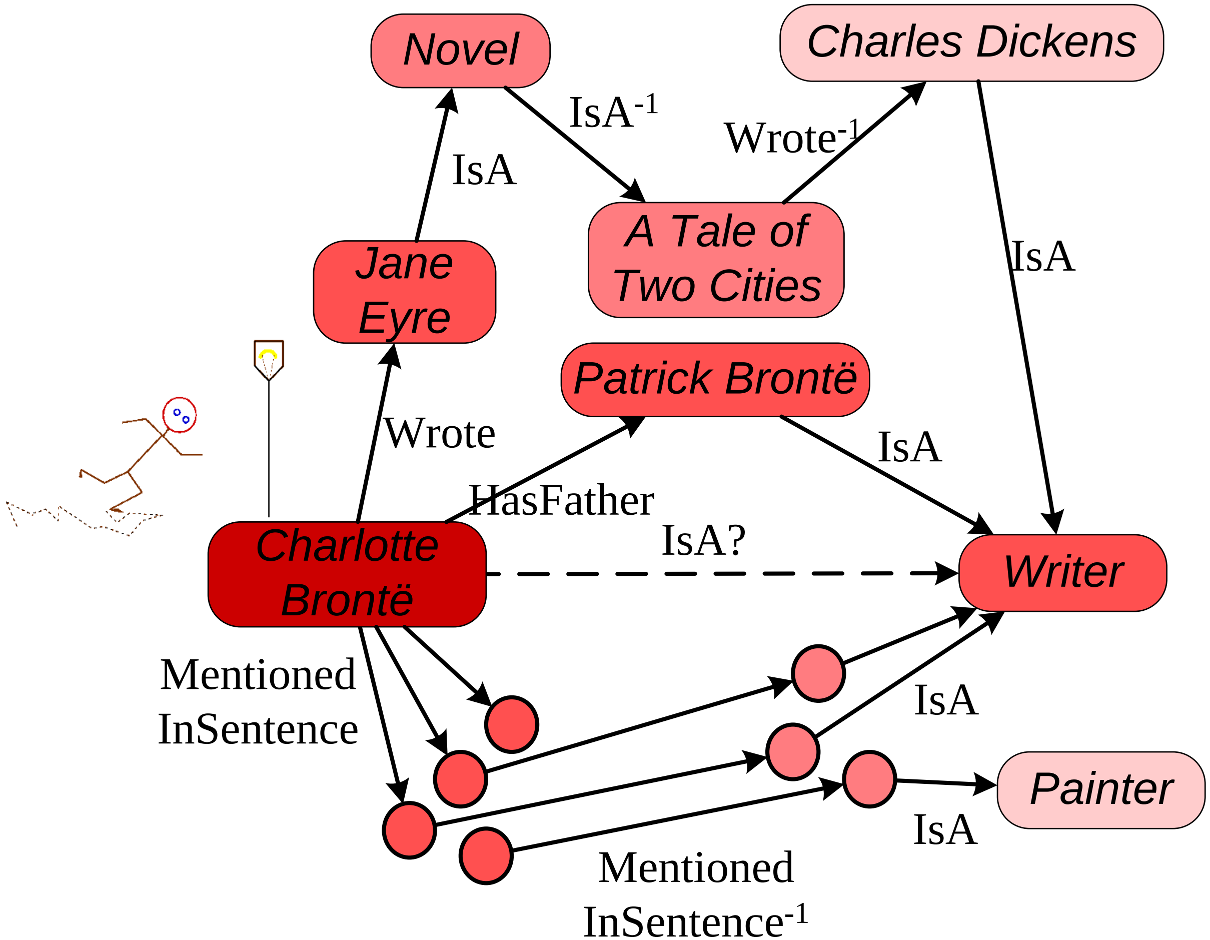
- Ignores edge types



Why Not Random Walk with Restart

(Will be covered in later classes)

- Ignores edge types



Prob(Charlotte → Writter)

Prob(Charlotte → Painter)

Why Not First Order Inductive Learner

- Learn Horn clauses in first order logic (FOIL , 1993)

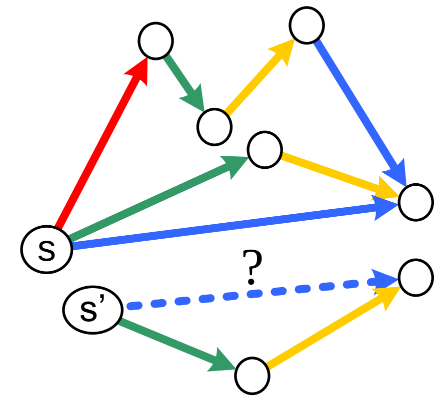
  HasFather(a, b) ^ isa(b,y) → isa(a; y) ← A low accuracy/high recall rule
  Write(a, i) ^ isa(i, x) ^ isa(j,x) ^ Write(b, j) ^ isa(b,y) → isa(a; y)
  InSentence(a, j) ^ InSentence(b, j) ^ isa(b,y) → isa(a; y)

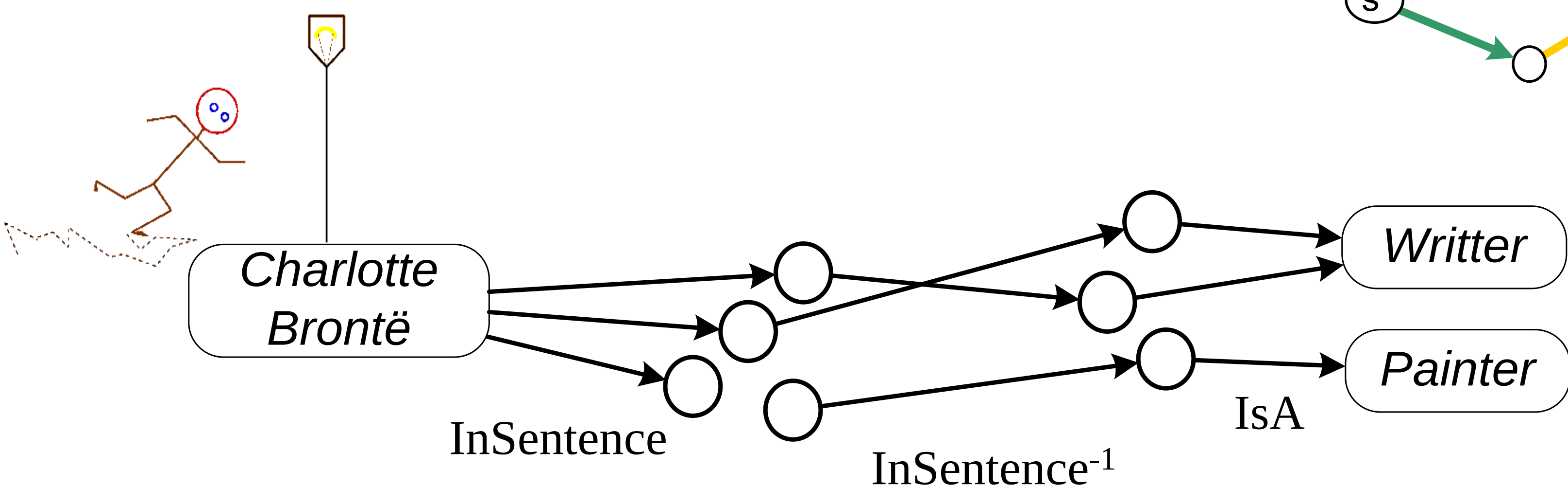
  HasFather(x, a) ^ isa(a,writer) → isa(x; writer) ← Lexicalized rule

- Horn clauses are costly to discover
- Inference is generally slow
- Cannot leverage low accuracy rules
  - Can only combine rules with disjunctions

Proposed: Random Walk Inference

- Random walk following a particular edge type sequence is very indicative





Prob(Charlotte $\rightarrow$ Writer| InSentence, InSentence^{-1}, IsA)

Random Walk Inference

- Combine features from different edge type sequences

Prob(Charlotte → Writer| HasFather, isa)
Prob(Charlotte → Writer| Write, isa, isa^{-1}, Write, isa)
Prob(Charlotte → Writer| InSentence, $InSentence^{-1}$, isa)

- More expressive than random walk with restart

- More efficient and robust than FOIL

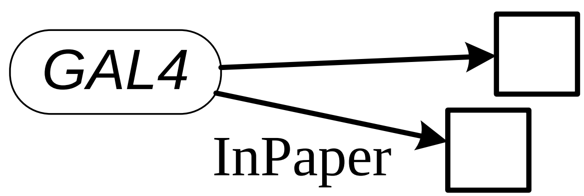
Outline

- Motivation
  - Relational Learning
  - Random Walk Inference
- Tasks
  - Publication recommendation tasks
  - Inference with knowledge base
- Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Query Independent Paths
  - Popular Entity Biases
- Efficient Inference (Lao & Cohen, KDD 2010)
- Feature Selection (L. M. C., EMNLP 2011)

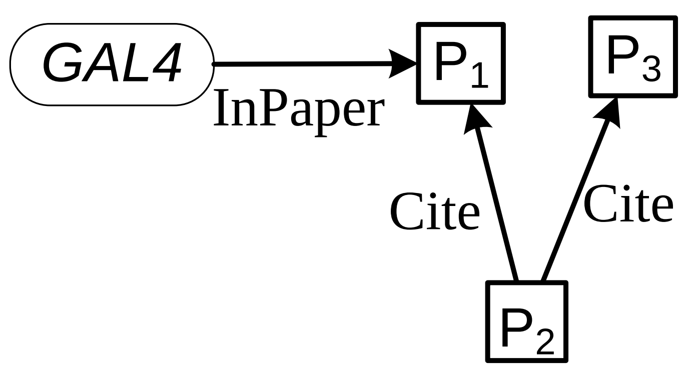
Recommendation Tasks with Biology Literature Data

- Problem
  - Given a topic e.g. “GAL4”
  - Which papers should I read?

- A simple retrieval approach (e.g. search engine)

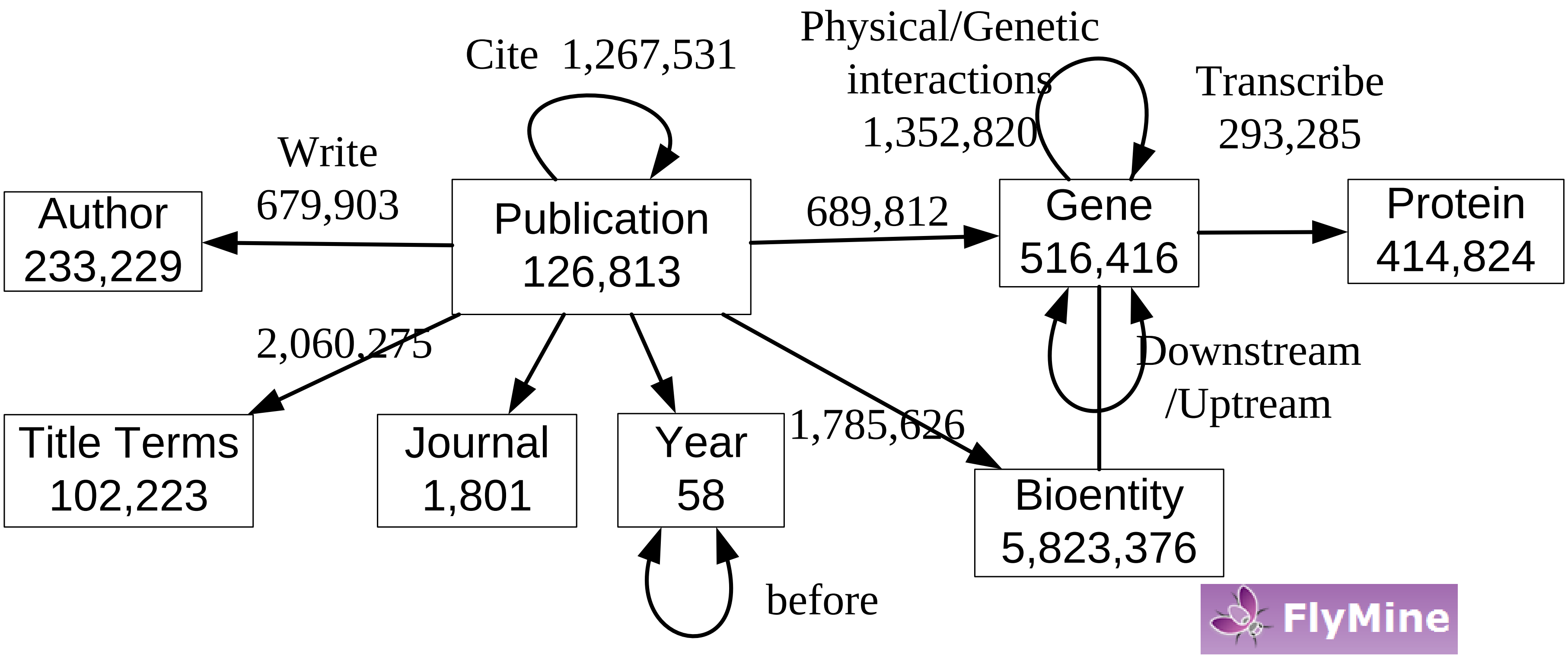


- Random walk inference find paths such as



Data sets

- Yeast: 0.2M nodes, 5.5M links
- Fly: 0.8M nodes, 3.5M links



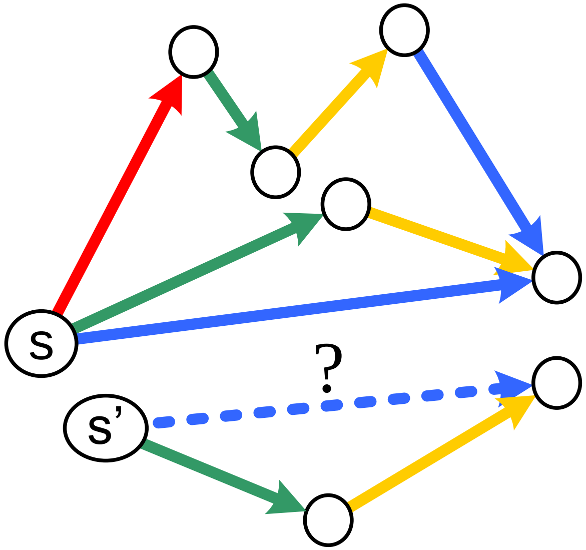
Experiment Setup

- Tasks
  - Gene recommendation: author, year→gene
  - Venue recommendation: genes, title words→journal
  - Reference recommendation: title words,year→paper
  - Expert-finding: title words, genes→author

- Data split
  - 2000 training, 2000 tuning, 2000 test

The NELL Knowledge Base

- Never-Ending Language Learning:
  - "a never-ending learning system that operates 24 hours per day, for years, to continuously improve its ability to read (extract structured facts from) the web" (Carlson et al., 2010
- Task:
  - Given
    - a knowledge base G
    - a starting node s
    - edge type R
  - Find
    - nodes t which should have edge R with s



e.g. IsA(Charlotte Brontë,?)

Experiment Setup

- We consider 96 relations for which NELL database has more than 100 instances
- Closed world assumption for training
  - The nodes y known to satisfy R(x; ?) are treated as positive examples
  - All other nodes are treated as negative examples
  - E.g.

    Training

    IsA(Charles Dickens, writter) → true

    IsA(Charles Dickens, painter) → false

    …

    Testing

    IsA(Charlotte Brontë, ??)

Outline

- Motivation
  - Relational Learning
  - Random Walk Inference
- Tasks
  - Publication recommendation tasks
  - Inference with knowledge base
- Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Query Independent Paths
  - Popular Entity Biases
- Efficient Inference (Lao & Cohen, KDD 2010)
- Feature Selection (L. M. C., EMNLP 2011)

details

Path Ranking Algorithm (PRA)

(Lao & Cohen, ECML 2010)

- A relation path $P=(R_1, \ ...,R_n)$ is a sequence of relations
- A PRA model scores a source-target pair by a linear function of their path features

$$score(s,t) = \sum_{P \in \mathbf{P}} \mathrm{Prob}(s \rightarrow t; P)\theta_P$$

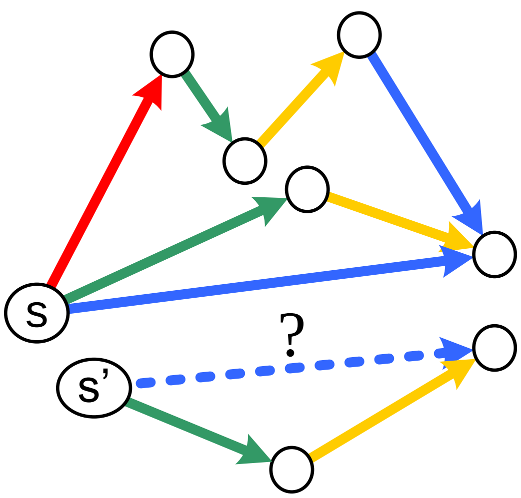
  - $\mathbf{P}$ is the set of all relation paths with length $\leq L$
  - E.g. IsA(Charlotte, ???)

Prob(Charlotte → Writer| HasFather, isa)
Prob(Charlotte → Writer| Write, isa, isa^{-1}, Write, isa)
Prob(Charlotte → Writer| InSentence, InSentence^{-1} , isa)

details

Training

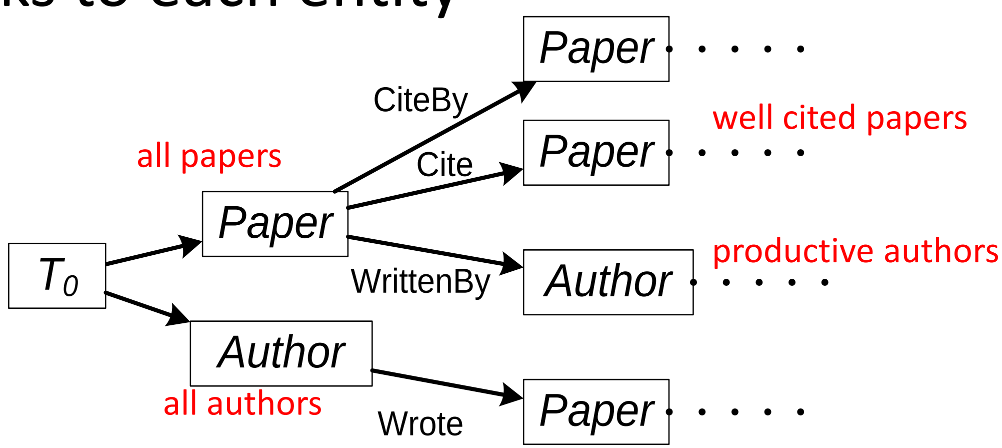
- For a relation *R* and a set of node pairs $\{(s_i, t_i)\}$, construct a training dataset D $=\{(x_i, y_i)\}$
  - x_i is a vector of all the path features for (s_i, t_i)
  - y_i indicates whether $R(s_i, t_i)$ is true or not
  - e.g. $s_i \rightarrow$Charlotte, $t_i \rightarrow$painter/writer
- θ is estimated using classifier
  - L1,L2-regularized logistic regression



more details

Extension 1: Query Independent Paths

- PageRank
  - assign an query independent score to each web page
  - later combined with query dependent score
- Generalize to multiple relation types
  - a special entity e_0 of special type T_0
  - T_0 has relation to all other entity types
  - e_0 has links to each entity



more details

Extension 2: Popular Entity Biases

- Node specific characteristics which cannot be captured by a general model
  - E.g. Certain genes have well known mile stone papers
  - E.g. Different users may have different intentions for the same query
- For a task with query type *T*, and target type *T*
  - Introduce a bias θ_e for each entity e of type *T*
  - Introduce a bias $\theta_{e',e}$ for each entity pair (e',e) where *e* is of type *T* and *e'* of type *T'*

Example Features

- A PRA+qip+pop model trained for reference recommendation task on the yeast data

ID	Weight	Feature
1	272.4	$word \rightarrow paper \xrightarrow{Cite^{-1}} paper \xrightarrow{Cite} paper$
2	156.7	$word \rightarrow paper \xrightarrow{Cite} paper$
3	100.5	$gene \rightarrow paper \xrightarrow{Cite^{-1}} paper \xrightarrow{Cite} paper$
4	83.7	$word \rightarrow paper \xrightarrow{Cite^{-1}} paper$
5	50.2	$gene \rightarrow paper \xrightarrow{Cite} paper$
6	41.4	$word \rightarrow paper$
7	29.3	$year \rightarrow paper \xrightarrow{Cite} paper$
8	13.0	$year \xrightarrow{Before^{-1}} year \rightarrow paper \xrightarrow{Cite} paper$
	...	
9	3.7	$T^* \rightarrow paper \xrightarrow{Cite} paper$
10	2.9	GAL4>Nature. 1988. GAL4-VP16 is an unusually potent transcriptional activator.
11	2.1	CYC1>Cell. 1979. Sequence of the gene for iso-1-cytochrome c in Saccharomyces cerevisiae.
	...	
12	-5.4	$year \xrightarrow{Before^{-1}} year \rightarrow paper$
13	-39.1	$year \rightarrow paper$
14	-49.0	$T^* \rightarrow year \rightarrow paper$

1) papers which are cited together with papers of this topic

6) simple retrieval stratigy

7,8) papers cited during the past two years

9) well cited papers

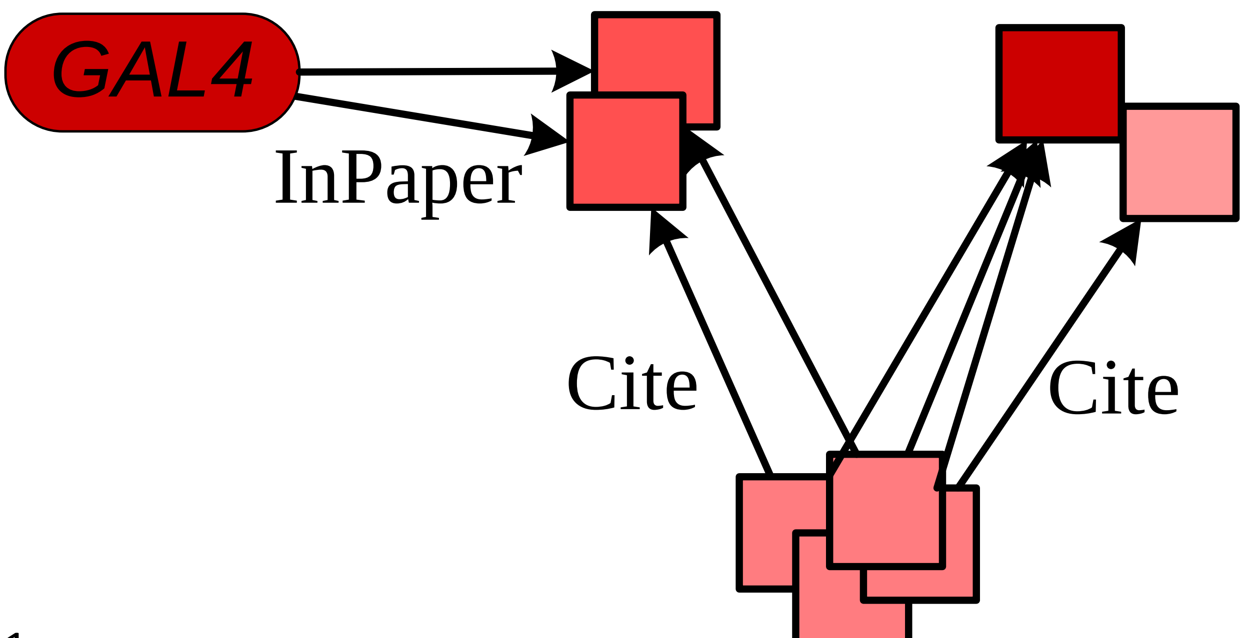
10,11) mile stone papers about specific query terms/genes

14) old papers

Example Features

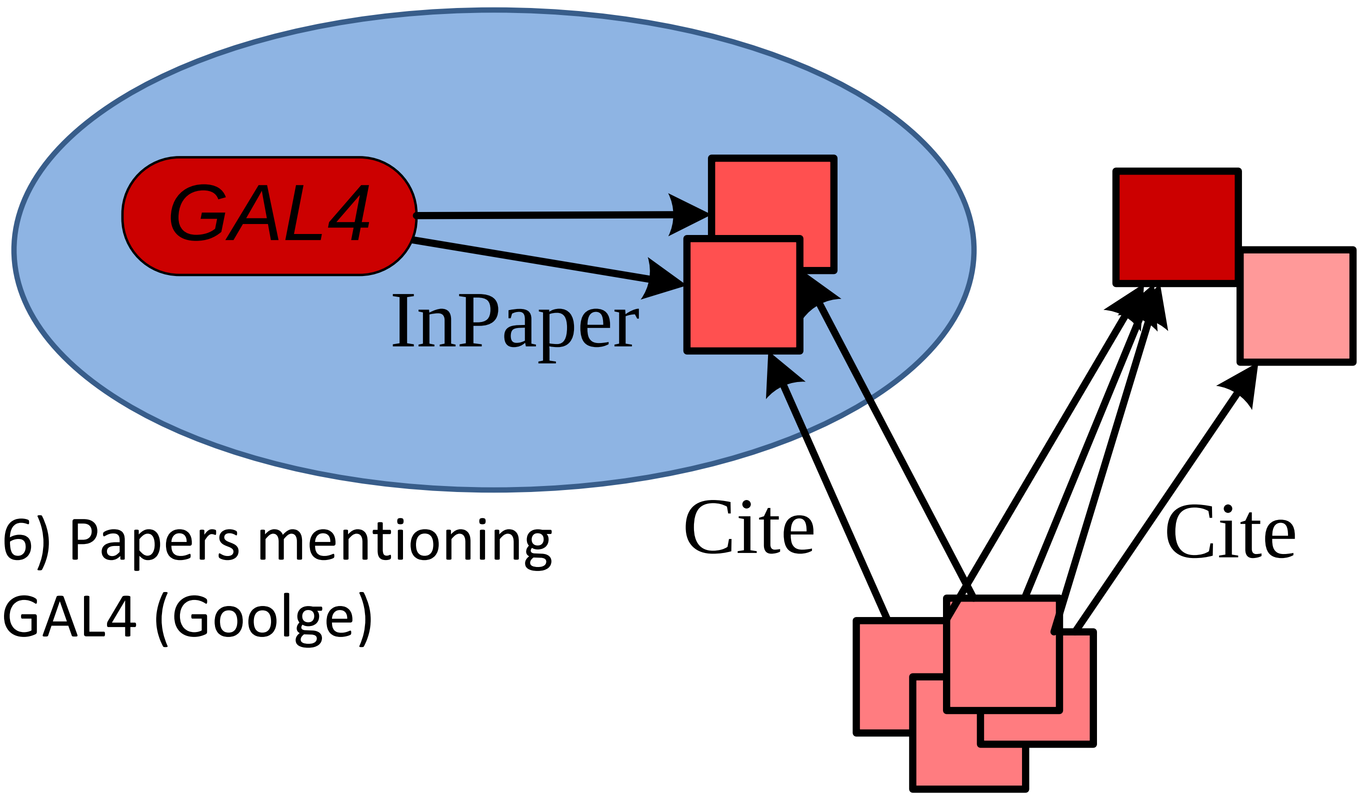
- Papers which are cited together with papers of this topic

1) Popularly cited papers



Example Features

- Papers which are cited together with papers of this topic



6) Papers mentioning GAL4 (Goolge)

Experiment Result

- Compare the MAP of PCRW to
  - Random Walk with Restart (RWR)
  - query independent paths (qip)
  - popular entity biases (pop)

Corpus	Task	RWR	PRA			
		trained	trained	+qip	+pop	+qip+pop
yeast	Ven	44.2	45.7 (+3.4)	46.4 (+5.0)	48.7 (+10.2)	49.3 (+11.5)
yeast	Ref	16.0	16.9 (+5.6)	18.3 (+14.4)	19.1 (+19.4)	19.8 (+23.8)
yeast	Exp	11.1	11.9 (+7.2)	12.4 (+11.7)	12.5 (+12.6)	12.9 (+16.2)
yeast	Gen	14.4	14.9 (+3.5)	15.1 (+4.9)	15.1 (+4.9)	15.3 (+6.3)
fly	Ven	48.3	50.4 (+4.3)	51.1 (+5.8)	50.7 (+5.0)	51.7 (+7.0)
fly	Ref	20.5	20.8 (†+1.5)	21.0 (+2.4)	21.6 (+5.4)	21.7 (+5.9)
fly	Exp	7.2	7.6 (†+5.6)	8.3 (+15.3)	7.9 (+9.7)	8.5 (+18.1)
fly	Gen	19.2	20.7 (+7.8)	21.1 (+9.9)	21.1 (+9.9)	21.0 (+9.4)

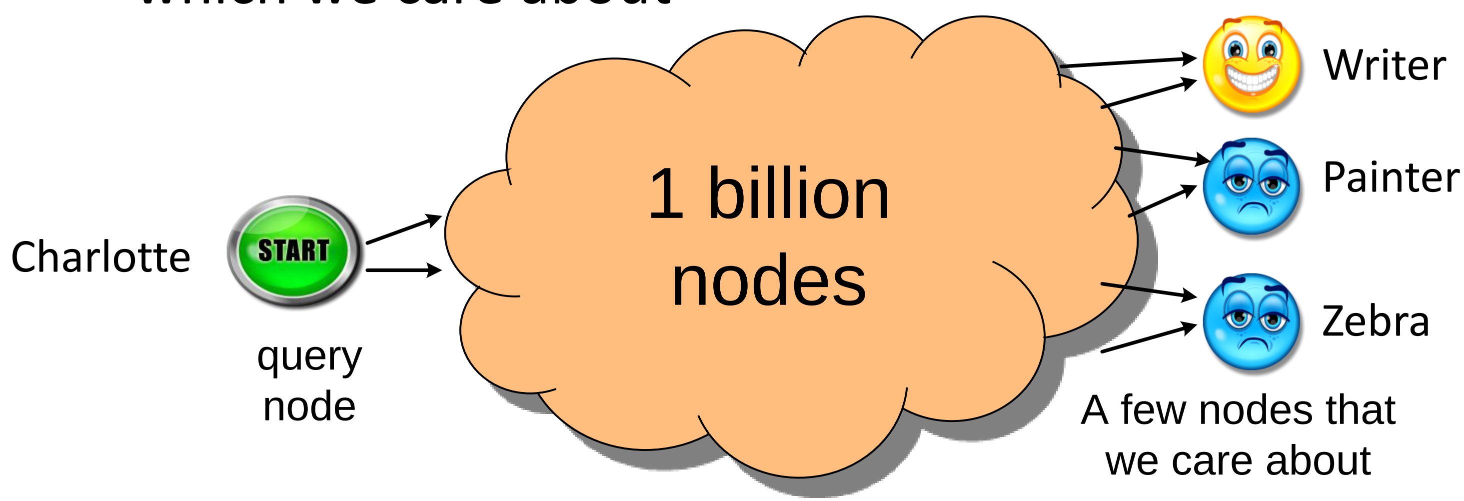
Outline

- Motivation
  - Relational Learning
  - Random Walk Inference
- Tasks
  - Publication recommendation tasks
  - Inference with knowledge base
- Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Query Independent Paths
  - Popular Entity Biases
- Efficient Inference (Lao & Cohen, KDD 2010)
- Feature Selection (L. M. C., EMNLP 2011)

Efficient Inference

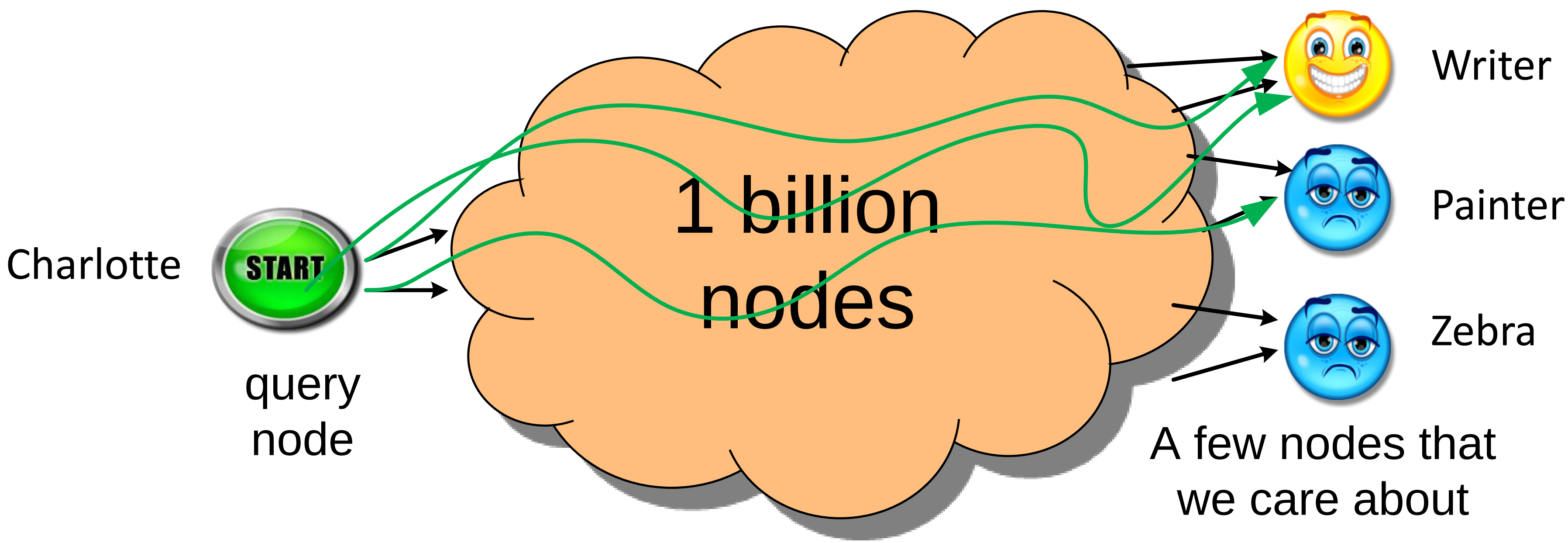
(Lao & Cohen, KDD 2010)

- Problem
  - Exact calculation of random walk distributions results in non-zero probabilities for many internal nodes in the graph
- Goal
  - Computation should be focused on the few target nodes which we care about

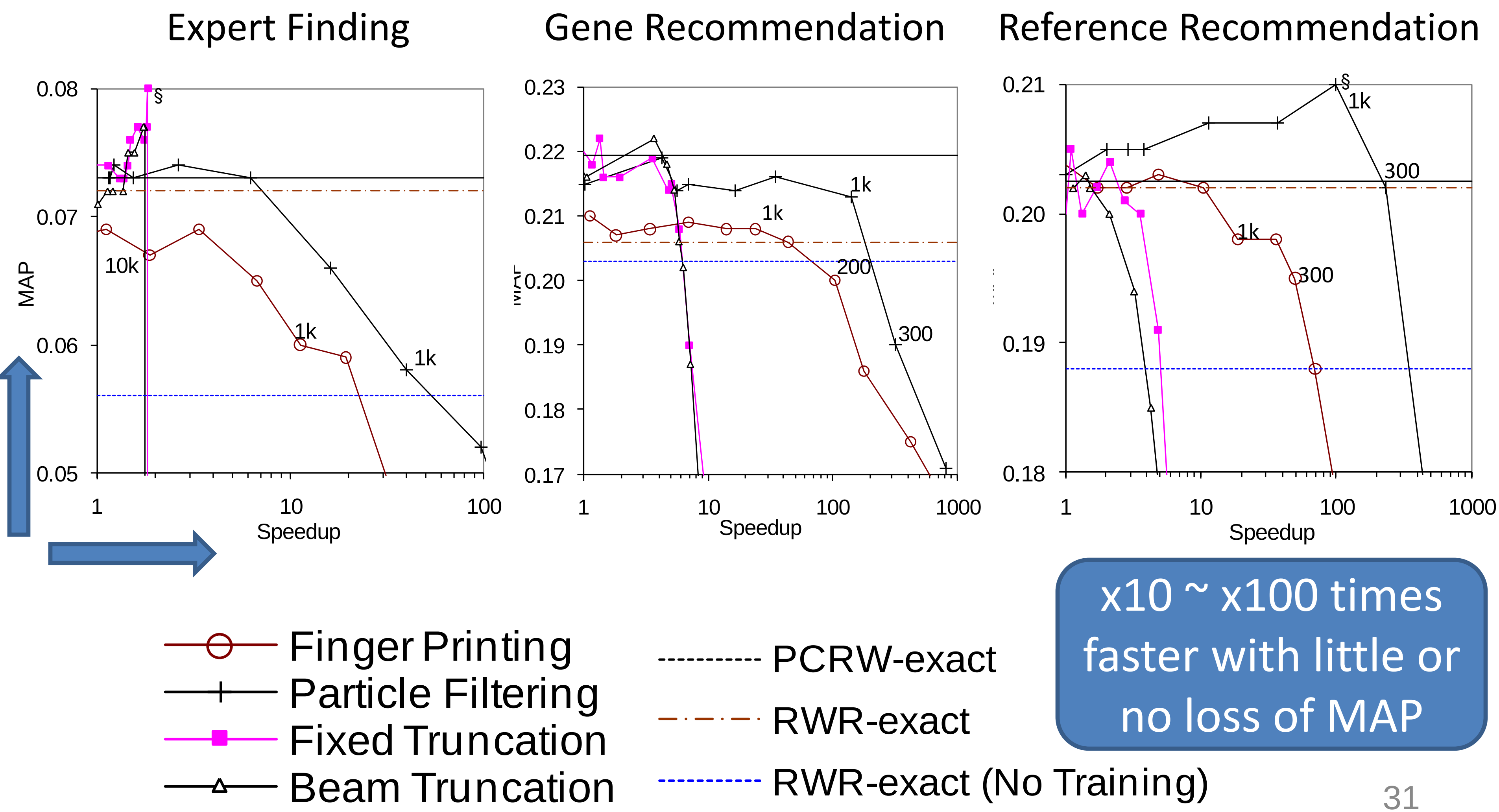


Efficient Inference

- Proposed Approach: Sampling
  - A few random walkers (or particles) are enough to distinguish good target nodes from bad ones



Results on the Fly Data



x10 ~ x100 times faster with little or no loss of MAP

Outline

- Motivation
  - Relational Learning
  - Random Walk Inference
- Tasks
  - Publication recommendation tasks
  - Inference with knowledge base
- Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Query Independent Paths
  - Popular Entity Biases
- Efficient Inference (Lao & Cohen, KDD 2010)
- Feature Selection (L. M. C., EMNLP 2011)

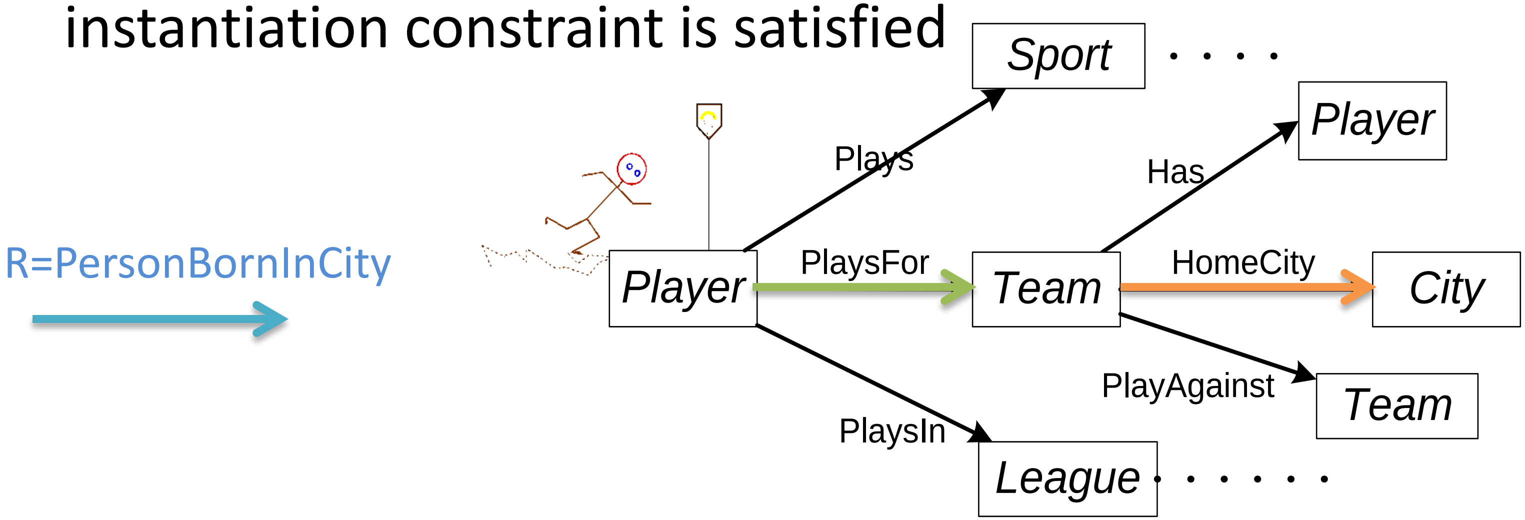
details

Path Finding & Feature Selection

(Lao, Mitchell & Cohen, EMNLP 2011)

- Impractical to enumerate all possible edge sequences $O(|V|^L)$
- How to find potentially useful paths?
  - Constraint 1: paths to instantiate in at least K(=5) training queries
  - Constraint 2: Prob(s→t| path , s→any node) > α (=0.2)

- Depth first search up to length l:
  - starts from a set of training queries, expand a relation if the instantiation constraint is satisfied



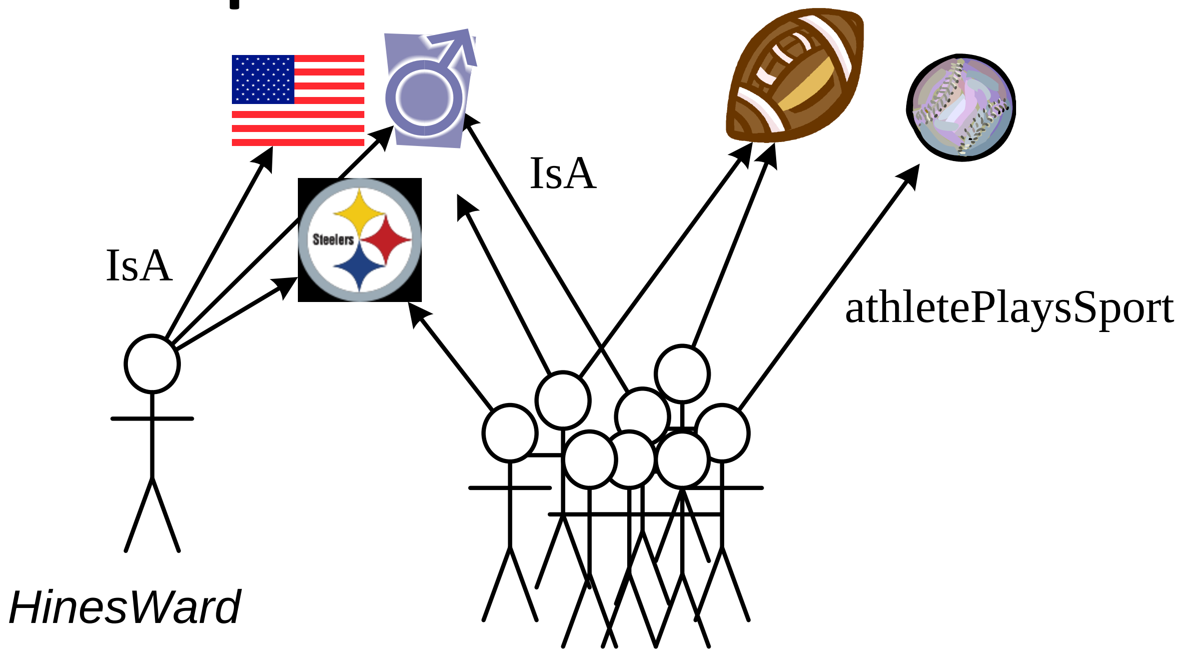
Path Finding & Feature Selection

- Dramatically reduce the number of paths

Table 1: Number of paths in PRA models of maximum path length 3 and 4. Averaged over 96 tasks.

	ℓ=3	**ℓ=4**
all paths up to length ℓ	$15,376$	$1,906,624$
+query support$\geq \alpha = 0.01$	522	5016
+ever reach a target entity	136	792
+L_1 regularization	63	271

Example Features



athletePlaysSport

$$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{athletePlaysSport}} c$$

$$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{superpartOfOrganization}} c \xrightarrow{\text{teamPlaysSport}} c$$

teamHomeStadium

$$c \xrightarrow{\text{teamPlaysInCity}} c \xrightarrow{\text{cityStadiums}} c$$

$$c \xrightarrow{\text{teamMember}} c \xrightarrow{\text{athletePlaysForTeam}} c \xrightarrow{\text{teamHomeStadium}} c$$

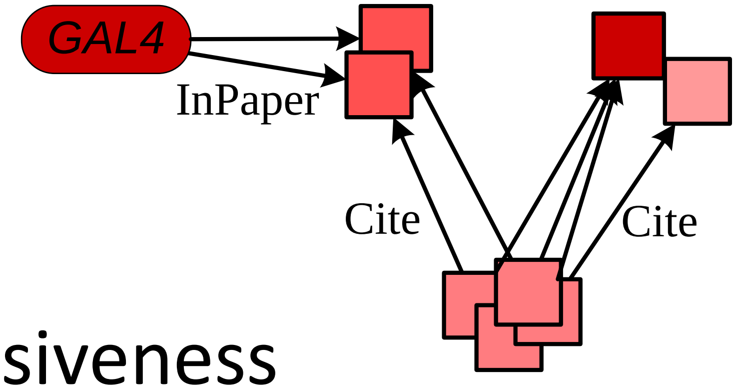
Evaluation by Mechanical Turk

- Sampled evaluation
  - only evaluate the top ranked result for each query
  - evaluate precisions at top 10, 100 and 1000 queries
- 8 functional predicates
- sampled 8 non-functional predicates

Task		#Rules	p@10	p@100	p@1000
Functional Predicates	N-FOIL	2.1(+37)	0.76	0.380	0.071
Functional Predicates	PRA	43	**0.79**	**0.668**	**0.615**
Non-functional Predicates	PRA	92	**0.65**	**0.620**	**0.615**

Conclusion

- Random walk inference for relational learning
  - Efficient
  - Robust



- Future work in model expressiveness
  - Discover lexicalized paths
  - Efficiently discover long paths

- Thank you! Questions?