Learning Relation Networks for Relational Retrieval

Thesis Proposal, Ni Lao

Mar 2, 2011

Committee:

William W. Cohen, (chair)

Teruko Mitamura

Geoffrey J. Gordon

C. Lee Giles, (Pennsylvania State University)

Outline

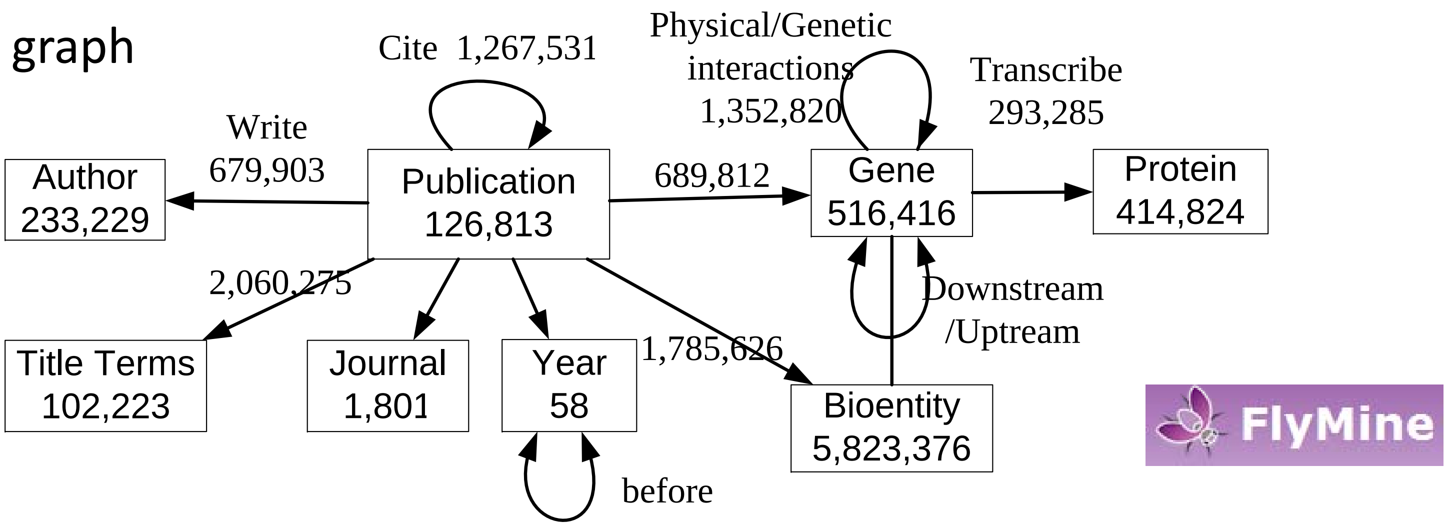
- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Relational Retrieval Problems

- Data of many retrieval/recommendation tasks can be represented as labeled directed graphs, e.g. scientific literature
  - Typed nodes: documents, terms, metadata
  - Labeled edges: citation, authorOf, datePublished
- Can support a family of *typed proximity queries*
  - ad hoc retrieval: terms → documents
  - Citation recommendation: paper → papers
- Which combination of these relations is important in answering the queries?

Biology Literature Data

- Flymine graph



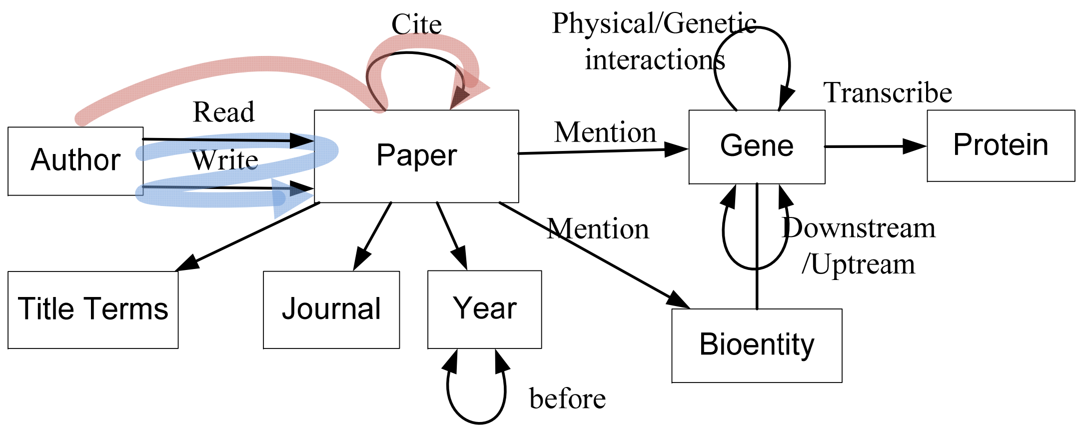
- Tasks
  - Gene recommendation: user, year→gene
  - Reference recommendation: title words,year→paper
  - Venue recommendation: genes, title words→journal
  - Expert-finding: title words, genes→author
  - Reading recommendation: user→paper

Biology Literature Data

- Example strategies for reading recommendation

$$author \xrightarrow{Write} paper^1 \xrightarrow{CitedBy} paper^2$$

$$author^1 \xrightarrow{Read} paper^1 \xrightarrow{WrittenBy} author^2 \xrightarrow{Write} paper^2$$



- Random walk with restart can be used for inference
- How to discover and combine different retrieval strategies?

Knowledgebase Data

- Graph representation of a knowledgebase
  - entities and concepts as nodes, and relations among them as edges
  - NELL@CMU (Never-Ending Language Learner) has over 242K beliefs
  - e.g. AthletePlaysSport(agassi,tennis), Generalizations(redmond,city)

- Expand knowledgebase by link prediction
  - given a node X and an relation type R, what are the nodes in the graph which should have relation R with X?
  - e.g. ?Y: AthletePlaysSport(agassi,Y)

$$athlete^{1} \xrightarrow{PlaysFor} team \xrightarrow{Plays} sport$$

$$athlete^{1} \xrightarrow{PlaysFor} team \xrightarrow{PlaysFor^{-1}} athlete^{2} \xrightarrow{Plays} sport$$

- Random walk with restart can be used for inference
- How to discover and combine different retrieval strategies?

Outline

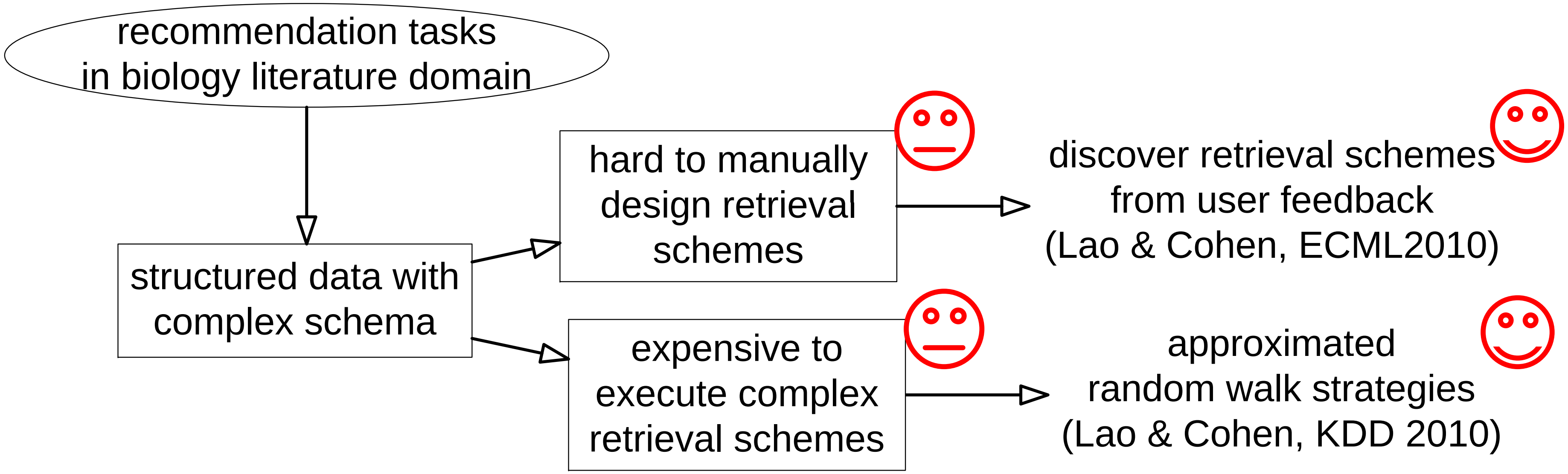
- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Thesis Goal

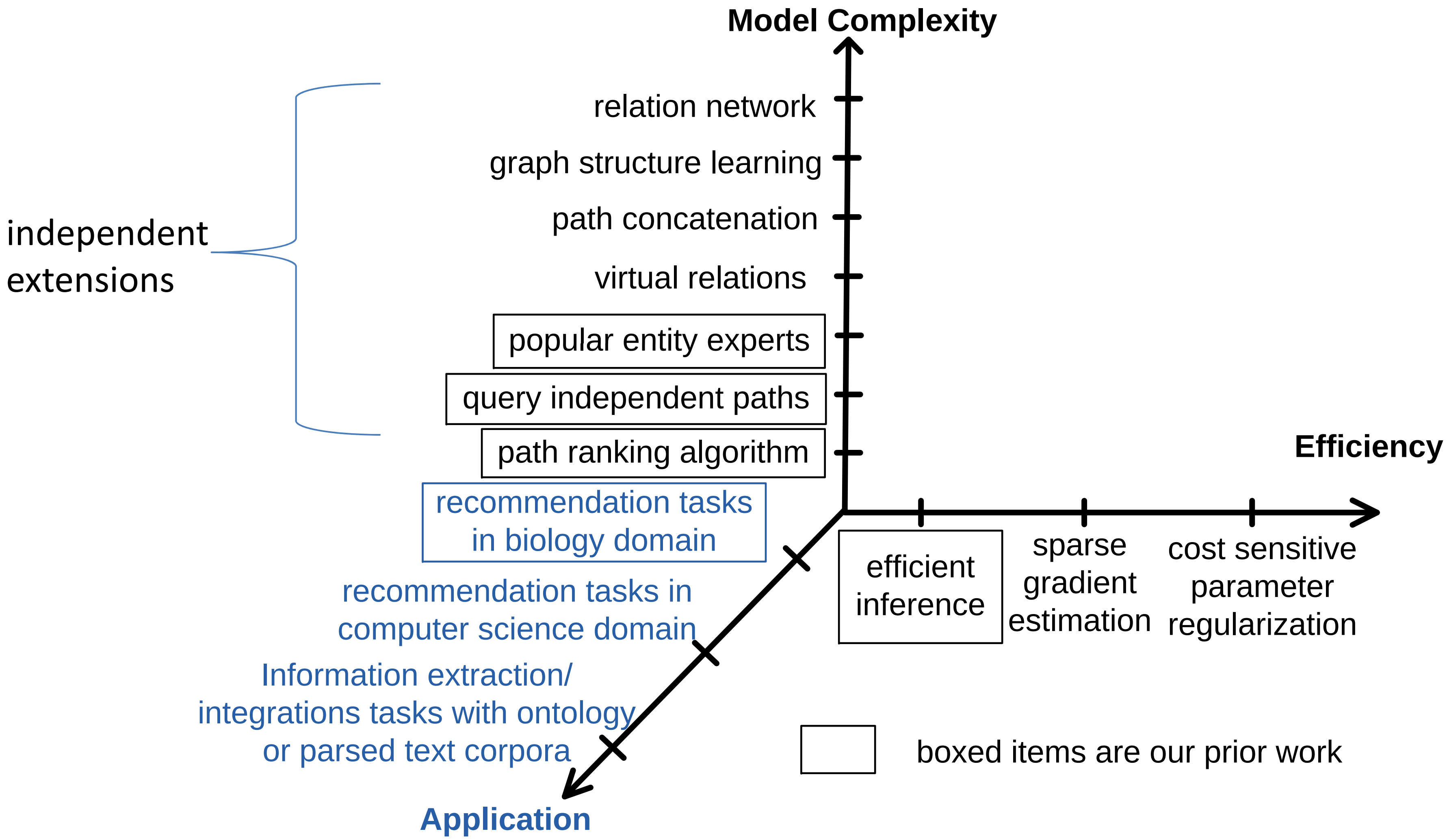
- Explore different ways of constructing the random walk models so that complex retrieval strategies on graph can be encoded

- We also develop algorithms that can efficiently discover and execute these strategies

Our Prior Work



Expected Contributions

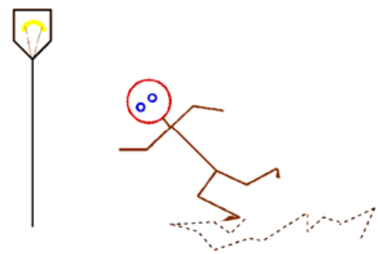


Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Random Walks with Restart (RWR)

- RWR is a commonly used similarity measure for proximity between query and target nodes
  - topic-sensitive Pagerank (Haveliwala, 2002)
  - personalized Pagerank (Jeh &. Widom, 2003)
  - ObjectRank (Balmin et al., 2004),
  - personal information management (Minkov & Cohen, 2007)



- RWR can be improved by supervised learning of edge weights
  - quadratic programming (Tsoi et al., 2003),
  - simulated annealing (Nie et al., 2005),
  - back-propagation (Diligenti et al., 2005; Minkov & Cohen, 2007),
  - limited memory Newton method (Agarwal et al., 2006)

- Document embedding and recommendation by
  - combining random walks with multiple sub-graphs which are generated form different relations (Zhou et. al, WWW2008)

The Limitation of RWR

- One-parameter-per-edge label is limited because a same edge can appears in both positive and negative paths
  - paths instead of edges encode the semantics of a retrieval strategy

Path	Comments
$user \xrightarrow{Read} paper \xrightarrow{Contain} gene \xrightarrow{ContainedBy} paper$	Don't read about **genes** which I have already studied
$user \xrightarrow{Read} paper \xrightarrow{WrittenBy} user \xrightarrow{Write} paper$	Read about my favorite **authors**

Path	Comments
$user \xrightarrow{Write} paper \xrightarrow{Contain} gene \xrightarrow{ContainedBy} paper$	Read about the **genes** that I am working on
$user \xrightarrow{Write} paper \xrightarrow{publishedBy} institute \xrightarrow{publish} paper$	Don't need to read paper from **my own lab**

Relational Learning Approaches

- To answer the query ?Y:AthletePlaysSport(agassi,Y)
- Horn clause rules
  - AthletePlaysForTeam(X,B)&TeamPlaysSport(B,Y)→AthletePlaysSport(X,Y)
- Path constrained random walk
  - $athlete^1 \xrightarrow{PlaysFor} team \xrightarrow{Plays} sport$

	softly combine rules	fast inference
First Order Inductive Learner (FOIL) (Quinlan & Cameron-Jones, 1993)	no	no
Markov Logic Network (MLN) (Richardson & Domingos, 2006)	yes	no
Path Ranking Algorithm (PRA) (Lao & Cohen, 2010 ECML)	yes	yes

Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Path Constrained Random Walk

(Lao & Cohen, ECML 2010)

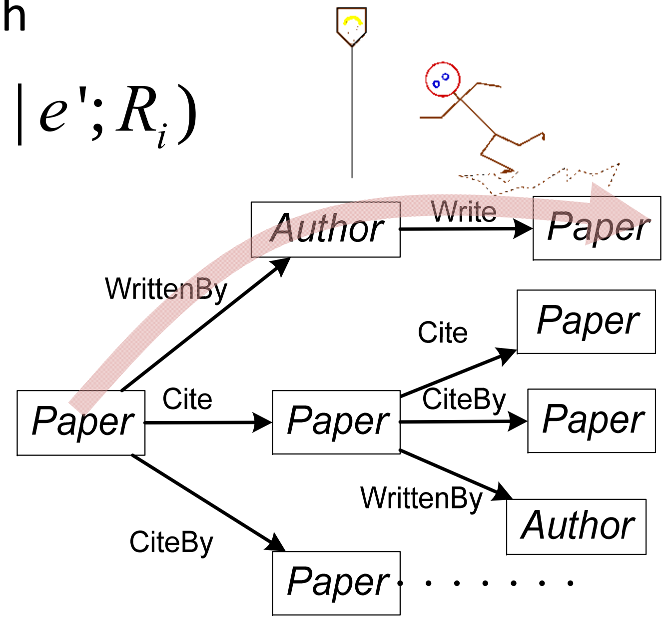
- A *Relation path* $P=(R_1, \ldots, R_n)$ *is a sequence of relations*
  - E.g. $year \xrightarrow{PublishedIn^{-1}} paper$

    $year \xrightarrow{PublishedIn^{-1}} paper \xrightarrow{Cite} paper$
- *Path Constrained Random Walk*
  - Recursively define a distribution for each path

$$h_{\mathbf{E}_s, R_1 \ldots R_{i+1}}(e) = \sum_{e' \in range(R_i)} h_{\mathbf{E}_s, R_1 \ldots R_i}(e') P(e \mid e'; R_i)$$

$$P(e \mid e'; R_i) = \begin{cases} 1/\deg(e', R_i) & e' \xrightarrow{R_i} e \\ 0 & otherwise \end{cases}$$

$$h_{\mathbf{E}_s, <empty>}(e) = \begin{cases} 1/|\mathbf{E}_s| & e \in \mathbf{E}_s \\ 0 & otherwise \end{cases}$$



Path Ranking Algorithm (PRA)

(Lao & Cohen, ECML 2010)

- Given a set of seed nodes $\mathbf{E}_q$, and a target type T_q, a PRA model can rank target entities by linearly combine the distributions of different paths

$$score(e;\theta) = \sum_{P \in \mathbf{P}(q,L)} h_P(e, \mathbf{E}_q)\theta_P$$

  – where $\mathbf{P}(q,L)=\{P\}$ is the set of all relation paths with range T_q and length $\leq L$

- or in matrix form $s=A\theta$,
  – A is the feature matrix, each column of which is a sparse distribution produced by one of the relation paths

Parameter Estimation

- Given a set of training data D={($A^{(m)}$, $y^{(m)}$)}
  - $A^{(m)}$ is the m-th feature matrix
  - $y^{(m)}$ is a binary vector representing relevance judgments
- Parameter θ can be optimized by maximizing a regularized log-likelihood

$$O(\theta) = \sum_{m=1..M} o_m(\theta) - \lambda_1 |\theta|_1 - \lambda_2 |\theta|_2 / 2$$

  - per-query objective function

$$o_m(\theta) = |P_m|^{-1} \sum_{i \in P_m} \ln p_i^{(m)} + |N_m|^{-1} \sum_{i \in N_m} \ln(1 - p_i^{(m)})$$

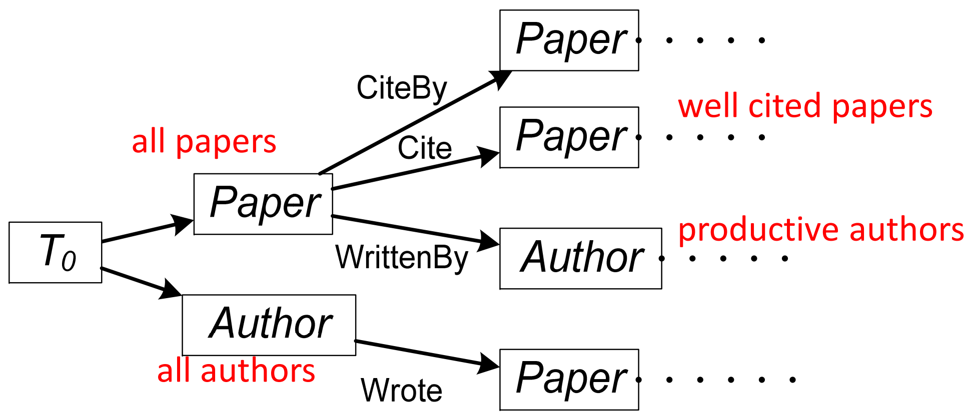
  - predicted relevance

$$p_i^{(m)} = p(y_i^{(m)} = 1 | q^{(m)}; \theta) = \frac{\exp(\theta^T A_i^{(m)})}{1 + \exp(\theta^T A_i^{(m)})}$$

  - P_m is the index set or relevant entities
  - N_m is the index set of irrelevant entities (sub-sampled)

Ext 1: Query Independent Paths

- Motivated by PageRank
  - assign a (query independent) importance score to each web page
  - later combined with a (query dependent) relevance score
- Generalize to heterogeneous graph
  - we include to each query a special entity e_0 of special type T_0
  - T_0 has relation to all other entity types, and e_0 has links to all entities
  - therefore, we have a set of query independent paths (random walks on which can be calculate offline)
- Example



Ext.2: Popular Entity Biases

- There are entity specific characteristics which cannot be captured by a general model
  - **in query log mining,** documents with lower rank to a query may be interesting to the users because of features not captured in the data
  - **in personalized search**, the same query may represent different information needs for different users

- A simple approach of adding bias terms
  - Introduce a bias θ_e for each target entity *e*
  - Introduce a bias $\theta_{e',e}$ for each query-target entity pair (*e',e*)

$$s(e;\theta) = \sum_{P.T_{last}=T_q} h_P^T(e)\theta_P + \theta_e + \sum_{e' \in \mathcal{E}_q} \theta_{e',e},$$

- Efficiency consideration
  - Only add to the model top *J* parameters (measured by $|\mathrm{d}O(\theta)/\mathrm{d}\theta_e|$) at each LBFGS iteration

Experiment Setup

- Data sources for bio-informatics
  - Saccharomyces Genome Database (SGD) a database for yeast
  - Flymine a database for fruit flies
  - PubMed on-line archive of over 18 million biological abstracts
  - PubMed Central (PMC) full-text copies of over 1 million of these papers

- 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
  - automatically generated labels
  - 2000 training, 2000 tuning, 2000 test queries

- Time variant graph (for training)
  - each edge is tagged with a time stamp (year)
  - only consider edges that are earlier than the query during random walk

Example Features

ID	Weight	Feature	
1	272.4	$word \rightarrow paper \xrightarrow{Cite^{-1}} paper \xrightarrow{Cite} paper$	1) Papers co-cited with the on-topic papers
2	156.7	$word \rightarrow paper \xrightarrow{Cite} paper$	2) Aggregated citations of the on-topic papers
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	$word \rightarrow paper$	6) Resembles an ad-hoc retrieval system
7	29.3	$year \rightarrow paper \xrightarrow{Cite} paper$	
8	13.0	$year \xrightarrow{Before^{-1}} year \rightarrow paper \xrightarrow{Cite} paper$	7,8) Papers cited during the past two years
	...		
9	3.7	$T^* \rightarrow paper \xrightarrow{Cite} paper$	9) Well cited papers
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.	
	...		10,11) (Important) early papers about specific query terms
12	-5.4	$year \xrightarrow{Before^{-1}} year \rightarrow paper$	
13	-39.1	$year \rightarrow paper$	
14	-49.0	$T^* \rightarrow year \rightarrow paper$	

A model trained for reference recommendation task on the yeast data

Evaluations

- Compare the MAP of PRA to
  - Random walk with restart model(RWR)
  - Query independent paths (qip)
  - Popular entity biases (pop)

PRA can improve over RWR based models

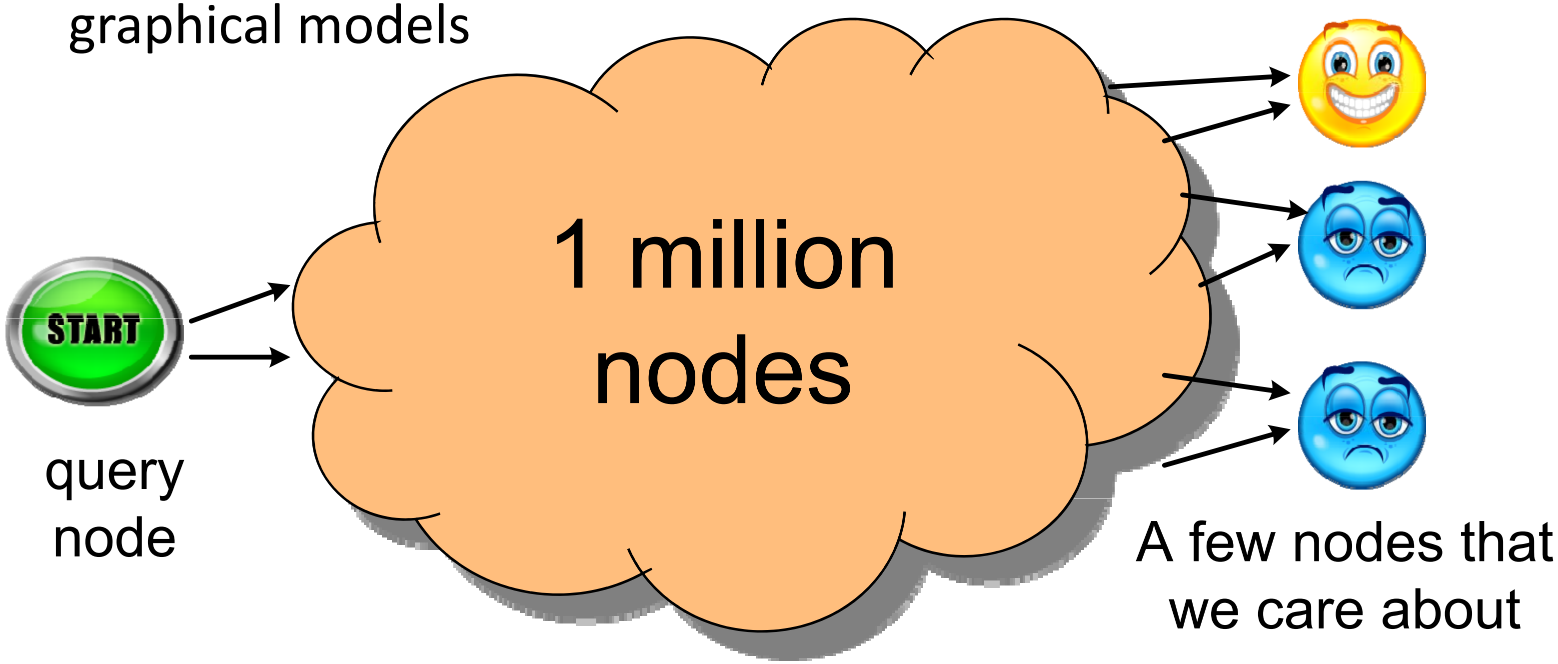
Corpus	Task	RWR	PRA+qip+pop
yeast	Venue Rec.	44.2	49.3 (+11.5)
yeast	Reference Rec.	16.0	19.8 (+23.8)
yeast	Expert Rec.	11.1	12.9 (+16.2)
yeast	Gene Rec.	14.4	15.3 (+6.3)
fly	Venue Rec.	48.3	51.7 (+7.0)
fly	Reference Rec.	20.5	21.7 (+5.9)
fly	Expert Rec.	7.2	8.5 (+18.1)
fly	Gene Rec.	19.2	21.0 (+9.4)

Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

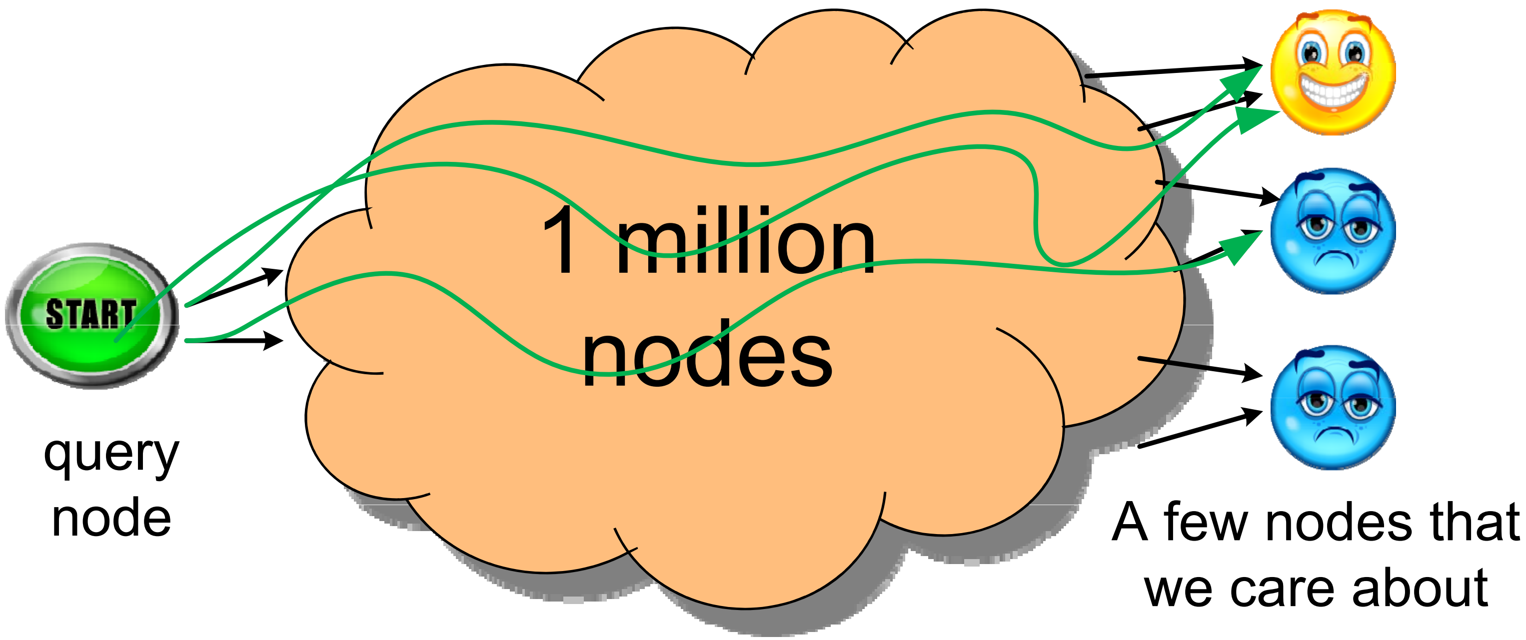
Efficient Inference

- Exact random walk results in probability to many internal nodes on the graph
  - Computation should be focused on the nodes we care about
  - e.g. Query-Specific Inference (Chechetka & Guestrin, 2010) on graphical models



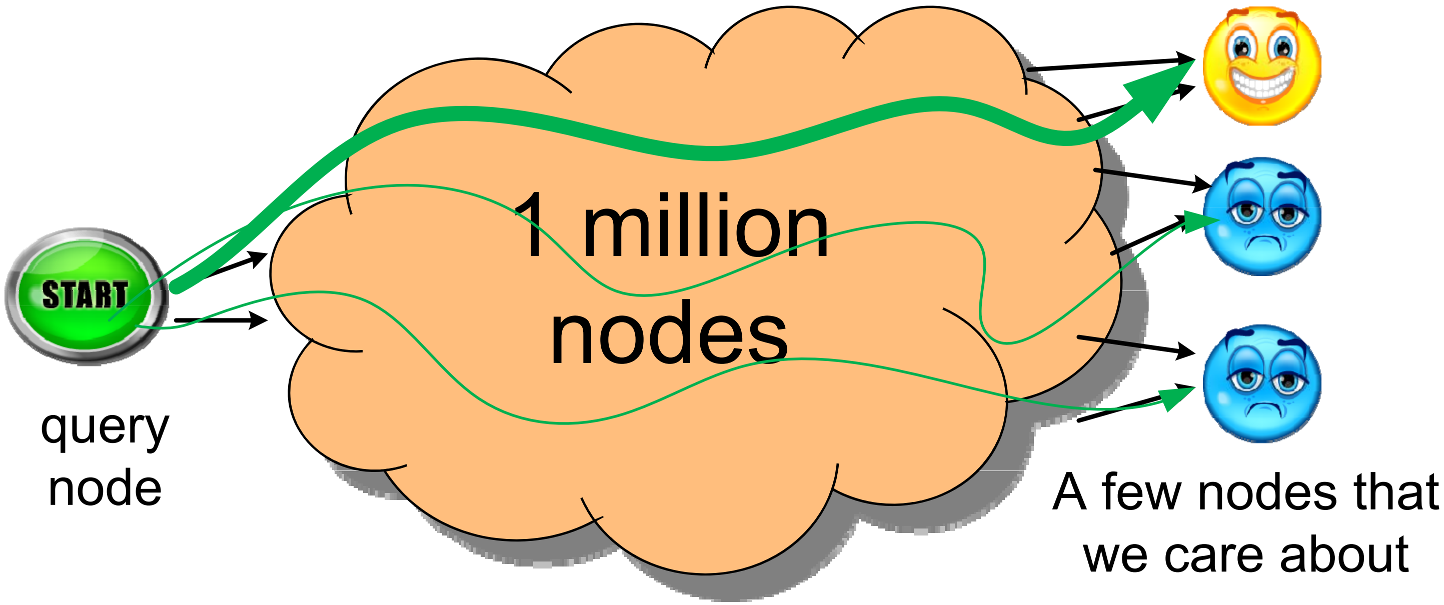
Efficient Inference

- Rational for sampling
  - a few random walkers are enough to distinguish good target nodes from bad target nodes



Efficient Inference

- Rational for truncation
  - large number of internal nodes which have small probabilities are not influential to the target nodes
  - we can safely drop them



Four Strategies for Efficient Random Walks

- Ways to keep the distribution $h(e)$ sparse at each step of random walk.
- Fixed Truncation (Chakrabarti, 2007)
  - Truncate by fixed value

$$h_{i+1}(e) = max(0, h_{i+1}(e) - \varepsilon)$$

- Beam Truncation
  - Keep top W probable entities

$$h_{i+1}(e) = max(0, h_{i+1}(e) - \varepsilon_W)$$

- Fingerprinting (sampling) (Fogaras, 2004)
  - Simulate a large number of random walkers

$$h_{i+1}(e) = \frac{\#\text{times the walkers visit } e}{\#\text{walkers}}$$

- Weighted Particle Filtering (Lao & Cohen, KDD 2010)
  - A combination of exact inference and sampling

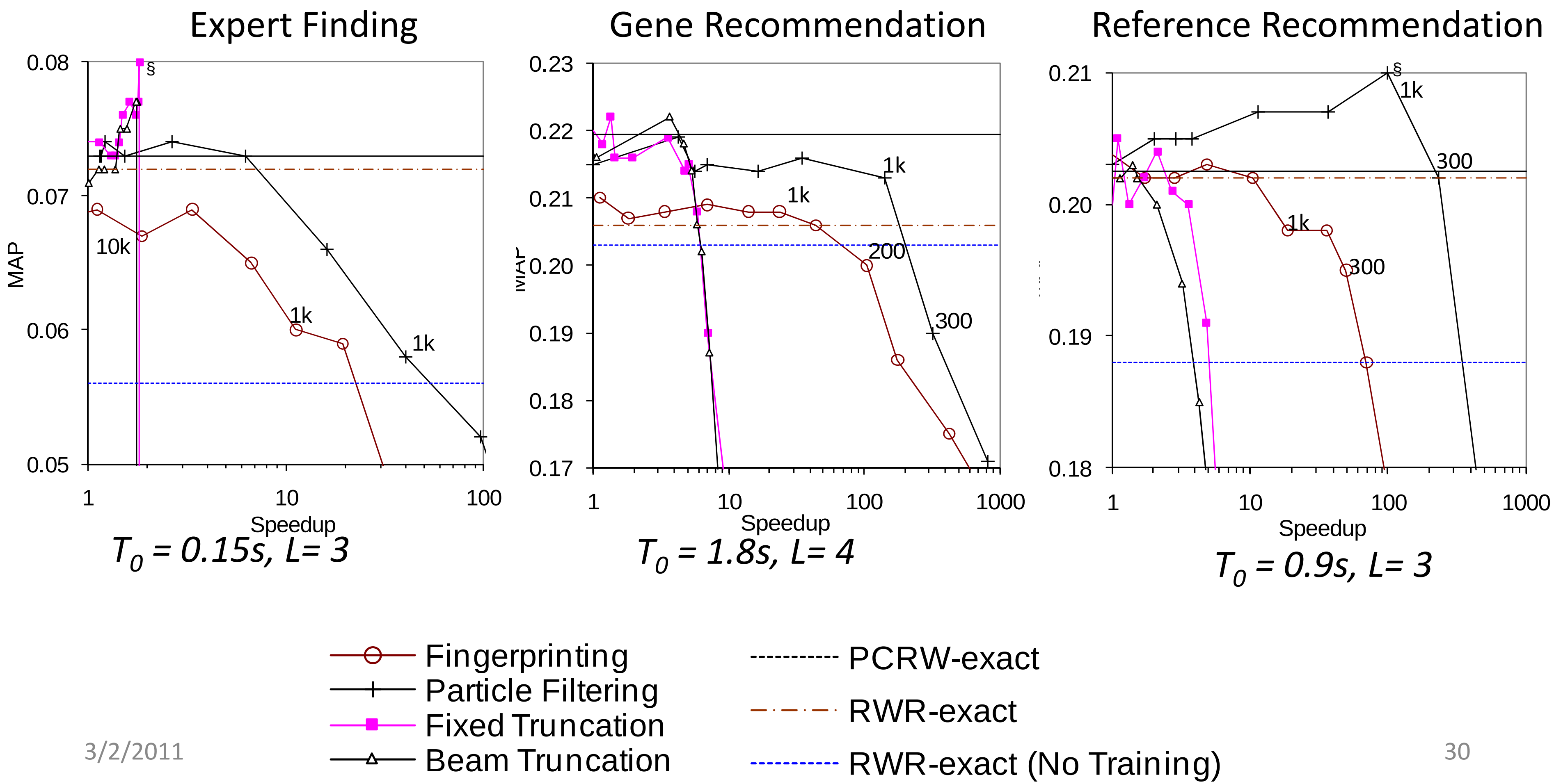
Weighted Particle Filtering

Algorithm 1 Weighted Particle Filtering

Input: distribution $h_i(e)$, relation R, threshold ε_{min}
Output: $h_{i+1}(e)$
Set $h_{i+1}(e) = 0$ (should not take any time)
for each e with $h_i(e) \neq 0$ **do**
  $size_{new} = h_i(e)/|R(e)|$
  if $size_{new} > \varepsilon_{min}$ **then** ←Start with exact inference
    for each $e' \in R(e)$ **do**
      $h_{i+1}(e') += size_{new}$
    end for
  else ← switch to sampling when the branching is high
    for k=1..$floor(h_i(e)/\varepsilon_{min})$ **do**
      randomly pick $e' \in R(e)$
      $h_{i+1}(e') += \varepsilon_{min}$
    end for
  end if
end for

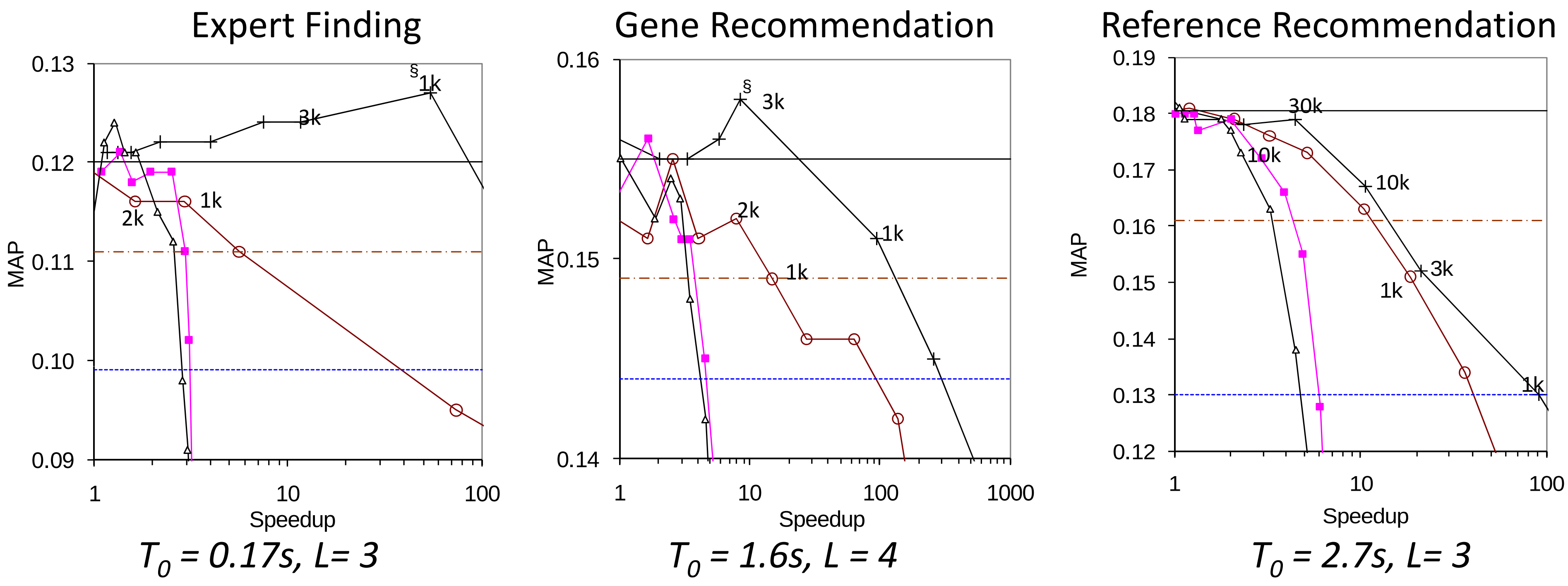
Results on The Fly Data

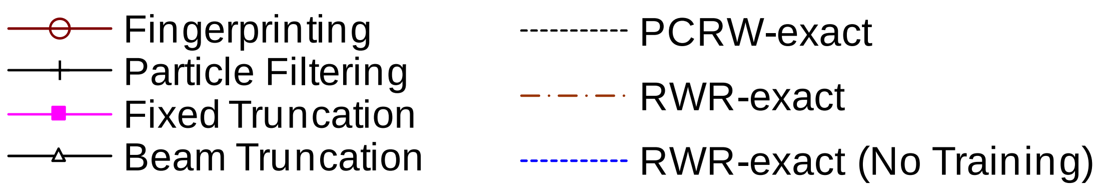
Up to 100x speedup with particle filtering



Results on The Yeast Data

Up to 100x speedup with particle filtering





Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches

- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)

- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization

- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks

- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Sparse Gradient Estimation

- How to efficiently find useful features from many?
- Observations on the decomposition of gradients
  - the gradient of a feature x_i can be approximated by considering its gradients on a subset of training examples

$$\frac{dL(\theta)}{d\theta_i} = \sum_{m=1..M} (y_i^{(m)} - p_i^{(m)}) x_i^{(m)}$$

$$= \sum_{m=1..M} (y_i^{(m)} - p_i^{(m)})(x_i^{(m)} - \bar{x}_i)$$

$$\approx \sum_{\substack{m:|y_i^{(m)} - p_i^{(m)}| > \varepsilon_1, |(x_i^{(m)} - \bar{x}_i)| > \varepsilon_2 \\ \text{or sample } m \text{ with } p \propto |y_i^{(m)} - p_i^{(m)}||(x_i^{(m)} - \bar{x}_i)|}} (y_i^{(m)} - p_i^{(m)})(x_i^{(m)} - \bar{x}_i)$$

assuming that there is a bias feature x_0, and its gradient is zero

  - a sparse gradient estimation of all the features is useful for our model complexity extensions
    - Path concatenation, Graph structure learning, Relation networks

Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Cost Sensitive Parameter Regularization

- We augment the objective function by the estimated cost of random walk on training data

$$O(\theta) = \sum_i o_i(\theta) - \mu \sum_j c_j(\pi) - \lambda|\theta|_2/2$$

  - π indicates whether or not each node on the relation tree is enabled
  - $c_j(\pi)$ is the cost of random walk on the *j*-th relation tree node
    - measured by time or number of RW operations

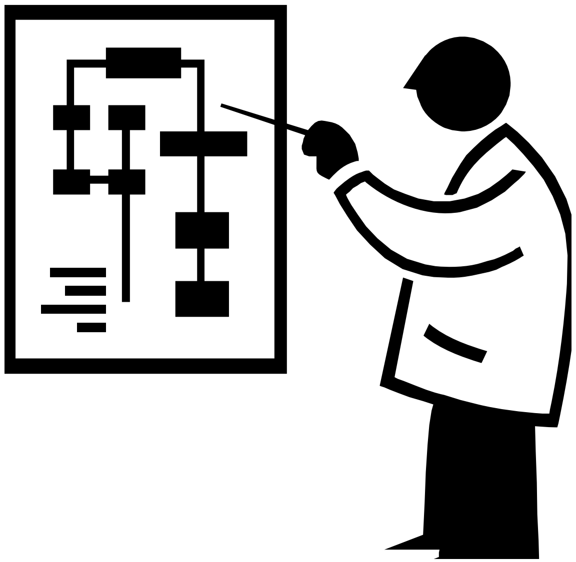
- A coordinate ascent optimization
  - θ step: optimize θ with LBFGS
    - For those nodes with $\pi_i=0$,their weights are fixed to *0*
  - π step:
    - For any node with $\pi_i=0$, if $\delta_i - \mu c_i > 0$, then set $\pi_i=1$
    - For any node with $\pi_i=1$, if $\delta_i + \mu c_i > 0$, then set $\pi_i=0$
    - δ_i is the estimated change of O(θ) as a result of adding or removing the *i*-th node from the model

Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Virtual Relation

- In order to integrate other algorithms into a random walk based model, we propose to generalize the concept of a relation to any function that projects a distribution on the graph to another distribution on the graph
  - infinite steps of random walk with restart
  - a query expansion algorithm
  - a classifier
  - a word stemmer
  - etc.

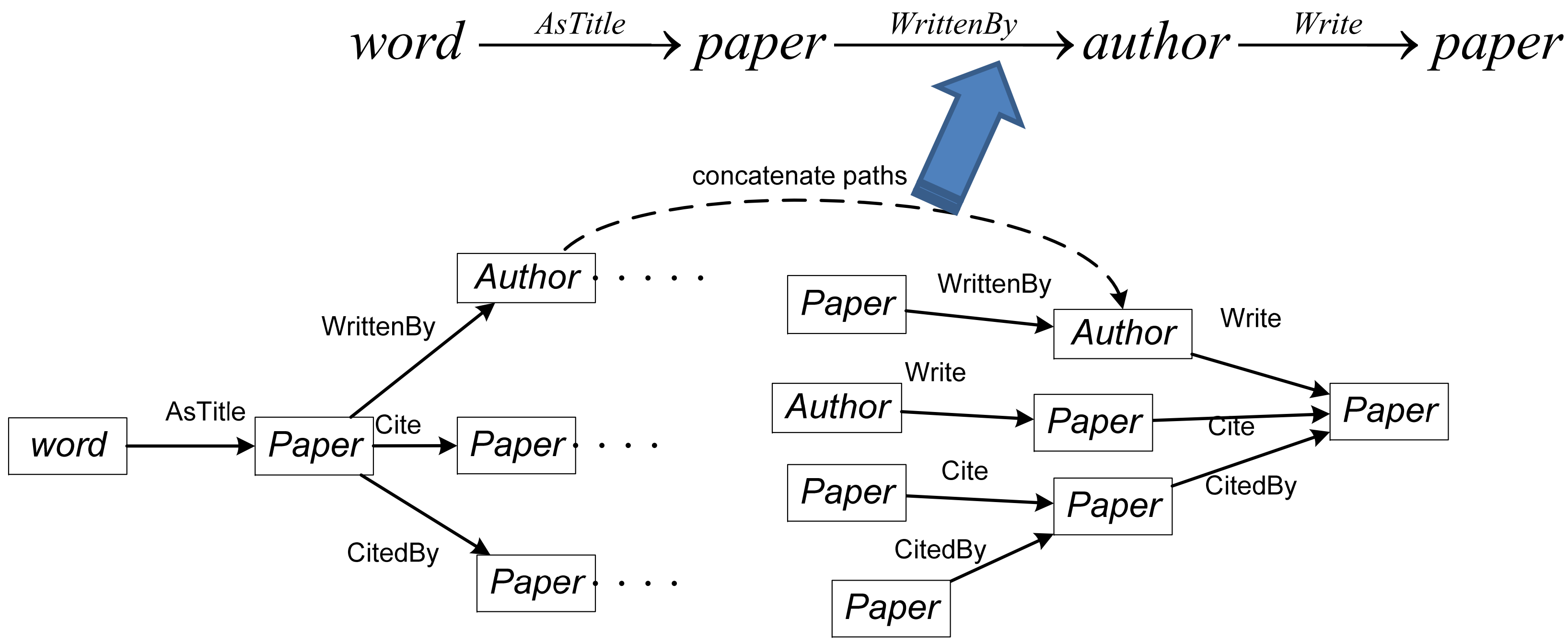


Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Path Concatenation

- We can efficiently explore large strategy space by combining existing paths
  - candidate paths' gradients can be estimated by combining forward random walk and backward propagation of gradients



Outline

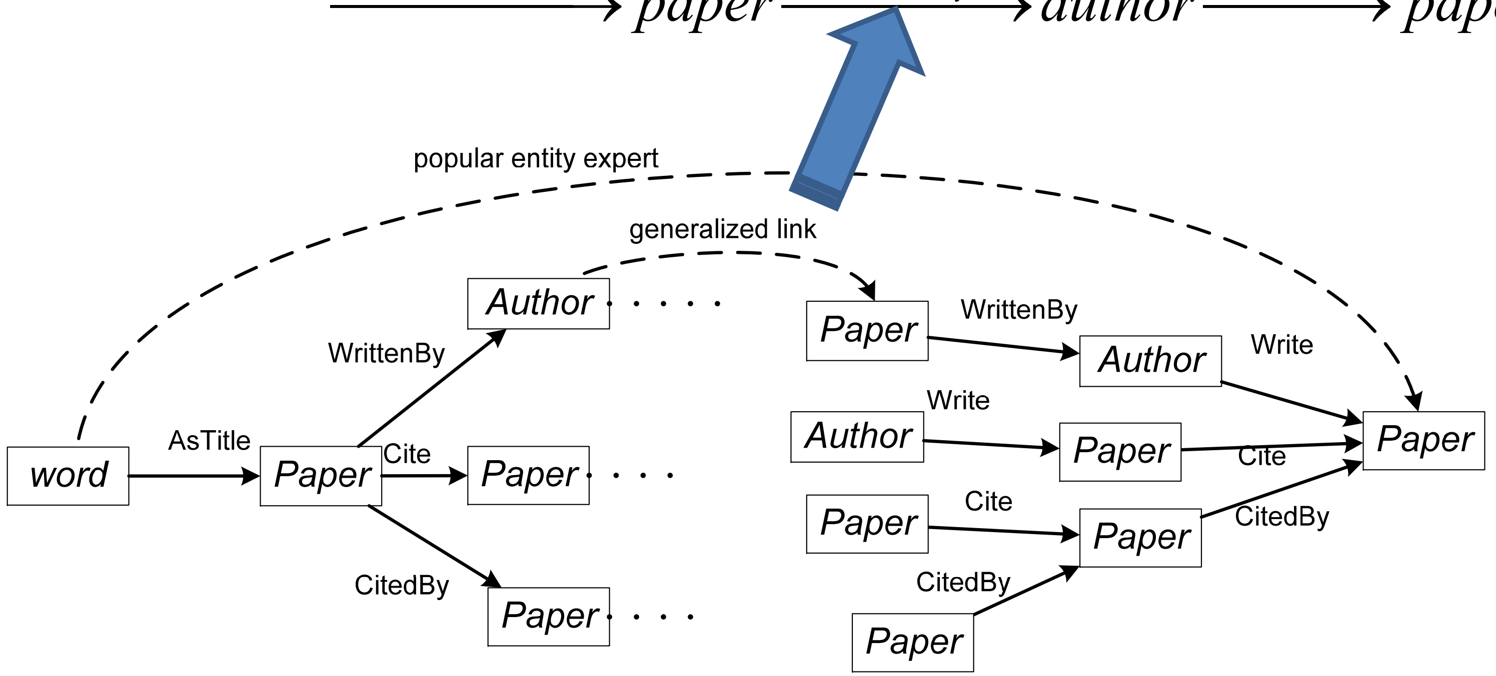
- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Graph structure learning

- Existing graph structure might not be enough to represent the correlations observed in the data.
  - Introduce new types of edges to connect any two types of entities

$$word \xrightarrow{AsTitle} paper \xrightarrow{WrittenBy} author$$

$$\xrightarrow{NewRelation} paper \xrightarrow{WrittenBy} author \xrightarrow{Write} paper$$



Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Relation networks

- Express conjunctions
  - e.g., we may want the retrieved document to be relevant to both the query words and the user who sent the query
- Switch strategies between queries
  - e.g. depending on whether a user has reading (or publication) history, the optimal weighting of paths might be different.
  - e.g. if the text query cannot retrieve satisfactory results, the relative weighting of other query expansion strategies might need to be adjusted.

Relation networks

- Introduce two hyper-edges that might be useful for retrieval tasks
  - AND operator

$$h_{A\cap B}(e) = h_A(e)h_B(e)$$

  - IF-NOT operator

$$h_{A\neg B}(e) = \begin{cases} h_A(e) & h_B(e') = 0 \text{ for every } e' \\ 0 & else \end{cases}$$

- Further assume that these hyper nodes cannot be the parent to any other nodes
  - The number of potential IF-NOT nodes is $O(nm)$
  - The number of potential AND nodes is $O(m)$
  - where n is the number of nodes in G, and m is the number of nodes in G with the query target type

Outline

- Introduction
  – Tasks & motivation
  – Thesis goal & expected contributions
  – Compare to existing approaches

- Our Prior Work
  – Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  – Efficient inference (Lao & Cohen, KDD 2010)

- Proposed Model Efficiency Extensions
  – Sparse gradient estimation
  – Cost sensitive parameter regularization

- Proposed Model Complexity Extensions
  – Virtual Relations
  – Path concatenation
  – Graph structure learning
  – Relation networks

- Work in Progress
  – Reading recommendation (Lao & Cohen, ISMB)
  – Link prediction with an ontology (Lao et al., EMNLP)

Reading Recommendation

(Lao & Cohen, ISMB)

- Goal
  - a reading recommendation system based on users' past reading and publication history
- Data
  - yeast graph
  - 22 years' reading history of biologist Dr. John L. Woolford
- Experiment
  - predict for each year what Dr. Woolford is going to read

Evaluation

- Compare the MAP of PRA to
  - Random walk with restart model(RWR)
  - Query independent paths (qip)

PRA can improve over RWR based models

Tree depth	Rank by Citations	RWR (no train)	RWR	PRA	PRA+qip (L=3)
	0.080				
L=2		0.189	0.204	0.218	**0.319**
L=3		0.210	0.251	0.304	0.306
L=4		0.192	0.202	0.296	0.315

Leverage user's reading and publication history and combine with other meta-data

Model

ID	Weight	Feature	Comments
1	1712.4	$year \xrightarrow{In^{-1}} paper$	Prefer papers from the query year
2	1099.6	$author \xrightarrow{Read} paper \xrightarrow{HasMajorMQ} topic \xrightarrow{HasMajorMQ^{-1}} paper$	Prefer papers with similar mesh qualifier to what I read before
3	767.0	$author \xrightarrow{Read} paper \xrightarrow{HasTitle} word \xrightarrow{HasTitle^{-1}} paper$	Prefer papers with similar *titles* to what I have read
4	672.2	$author \xrightarrow{Read} paper \xrightarrow{Write^{-1}} author \xrightarrow{Write} paper$	Prefer paper written by my favorite *authors*
5	441.2	$author \xrightarrow{Read} paper \xrightarrow{Cite} paper \xrightarrow{Cite^{-1}} paper$	Prefer papers which *share references* with what I generally read
6	337.9	$author \xrightarrow{Write} paper \xrightarrow{HasMajorMQ} topic \xrightarrow{HasMajorMQ^{-1}} paper$	Prefer papers with the same major *Mesh qualifier* as my papers
	...		
7	131.3	$author \xrightarrow{Write} paper \xrightarrow{HasGene} gene \xrightarrow{HasGene^{-1}} paper$	Prefer the *genes* that I have been working on
8	71.4	$author \xrightarrow{Write} paper \xrightarrow{Cite^{-1}} paper \xrightarrow{Cite^{-1}} paper$	Prefer *follow up papers* to *my own paper*
9	37.2	$author \xrightarrow{Write} paper \xrightarrow{Cite^{-1}} paper \xrightarrow{Cite} paper$	Prefer papers *co-cited* with my papers
10	26.7	$year \xrightarrow{After} year \xrightarrow{Read} paper \xrightarrow{Cite^{-1}} paper$	Prefer *follow up papers* to my readings last year
11	25.6	$author \xrightarrow{Write} paper$	Prefer my own papers
	...		
12	-285.1	$e^* \xrightarrow{AnyInstitute} institute \xrightarrow{Affiliation^{-1}} paper$	Disfavor papers published by less established institutes
13	-325.6	$author \xrightarrow{Read} paper \xrightarrow{HasGene} gene \xrightarrow{HasGene^{-1}} paper$	Disfavor *genes* which I have already read about
14	-334.2	$e^* \xrightarrow{AnyJournal} journal \xrightarrow{In^{-1}} paper$	Disfavor papers published by less popular journals
15	-399.9	$author \xrightarrow{Write} paper \xrightarrow{Affiliation} institute \xrightarrow{Affiliation^{-1}} paper$	Disfavor paper from *my own lab*

Outline

- Introduction
  - Tasks & motivation
  - Thesis goal & expected contributions
  - Compare to existing approaches
- Our Prior Work
  - Path Ranking Algorithm (Lao & Cohen, ECML 2010)
  - Efficient inference (Lao & Cohen, KDD 2010)
- Proposed Model Efficiency Extensions
  - Sparse gradient estimation
  - Cost sensitive parameter regularization
- Proposed Model Complexity Extensions
  - Virtual Relations
  - Path concatenation
  - Graph structure learning
  - Relation networks
- Work in Progress
  - Reading recommendation (Lao & Cohen, ISMB)
  - Link prediction with an ontology (Lao et al., EMNLP)

Link Prediction on NELL

(Lao et al., EMNLP)

- Pick the top 50 common relations, create two tasks for each relation:
  - ?Y: AthletePlaysSport(Y,tennis)
  - ?Y: AthletePlaysSport(agassi,Y)
- A query is generated for each entity (e.g. tennis) which have such relation. Its actual links are used as relevance judgment

weight	path	comments
28.6	$athlete^1 \xrightarrow{Generalize} concept \xrightarrow{Degeneralize} athelete^2 \xrightarrow{Plays} paper$	popular sports
24.2	$athlete \xrightarrow{PlaysIn} league \xrightarrow{SubOrganization} team \xrightarrow{Plays} sport$	sports played by **other teams** in my league
14.2	$athlete^1 \xrightarrow{PlaysIn} league \xrightarrow{PlaysIn^{-1}} athlete^2 \xrightarrow{Plays} sport$	sports played by **other players** in my league
	…	

Link Prediction on NELL

- Result with 5-fold cross validation

	#Query	MRR	RW Time (sec)	Training Time (sec)
citylocatedinstate	1126	**0.622**	31	201
_citylocatedinstate	85	**0.884**	5	126
citylocatedincountry	688	**0.459**	16	66
_citylocatedincountry	295	**0.916**	5	208
hasofficeincity	398	**0.791**	27	443
_hasofficeincity	489	**0.785**	16	80
athleteplayssport	1119	**0.906**	6	21
…				

Reasonably good accuracy

Learning and inference are very efficient

Tentative Schedule

Year	Month	Research Activity
2011	Feb-Mar	Explore recommendation tasks on other data sets, and explore other information extraction/integration tasks
	Apr	Experiment with the path concatenation algorithm
	May	Experiment with the structure learning algorithm
	Jun-Jul	Experiment with the relation network algorithm
	Aug	Experiment with cost sensitive parameter regularization
	Sep	Experiment with virtual edges
	Oct-Dec	Path concatenation and structure learning for relation networks
2012	Jan-Apr	Write thesis
	May	Defend thesis

Conclusion

- We proposed to extent our prior work in three directions
- We hypothesis that more complex models can give better accuracy, and the efficiency algorithms can significantly improve speed
- Thank you for your comments