Composite Likelihood Data Augmentation for Within-Network Statistical Relational Learning

Joseph J. Pfeiffer III[1]
Jennifer Neville[1]
Paul Bennett[2]

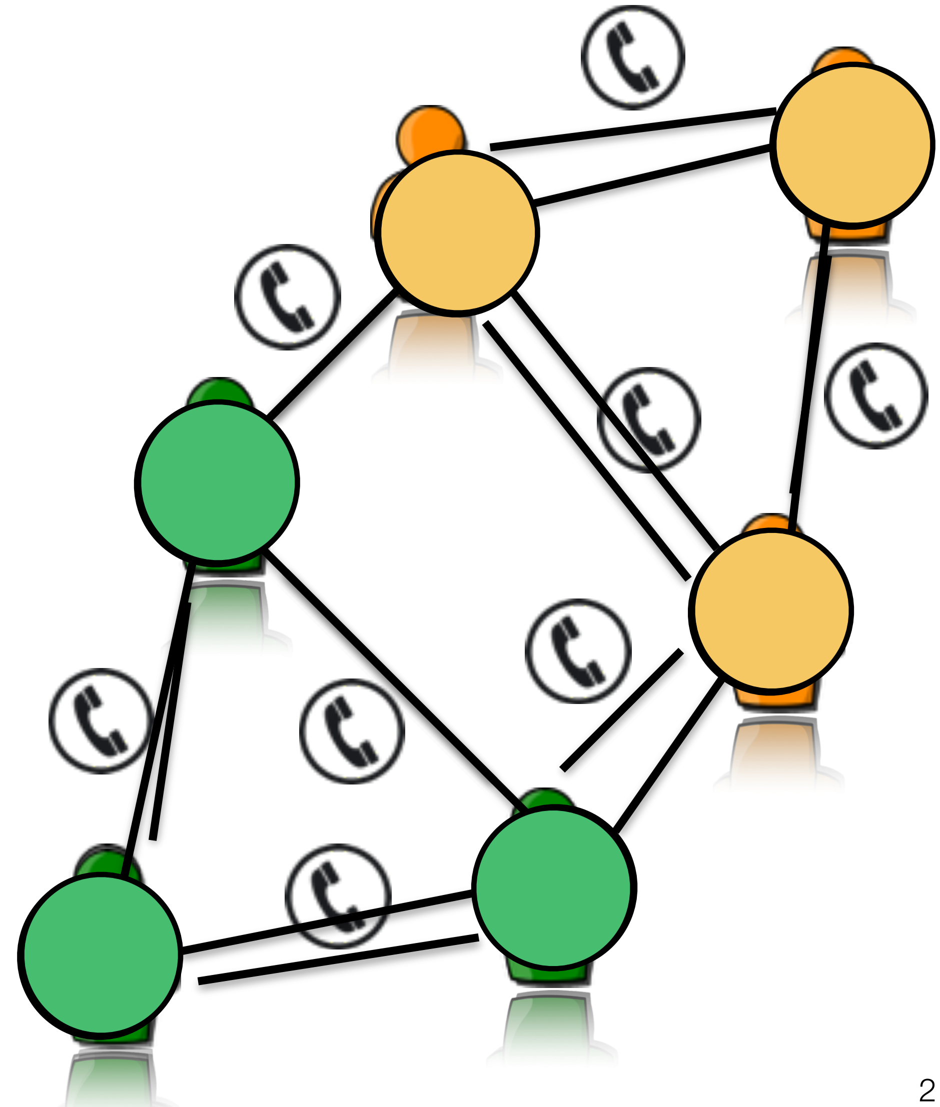
[1]Purdue University
[2]Microsoft Research

ICDM 2014, Shenzhen

Relational Machine Learning

- Items: stock brokers
  - social network users, papers, movies, products, etc.
- Label: fraudulent
  - buy product, sales rank, etc.,
- Relationships: phone calls
  - friendships, emails, citations, co-purchases, etc.
- Attributes: other observable item traits

***Within*-network relational machine learning**



Within-Network Relational Machine Learning

Learn model parameters $\Theta_{\mathscr{Y}}$ using the observed data

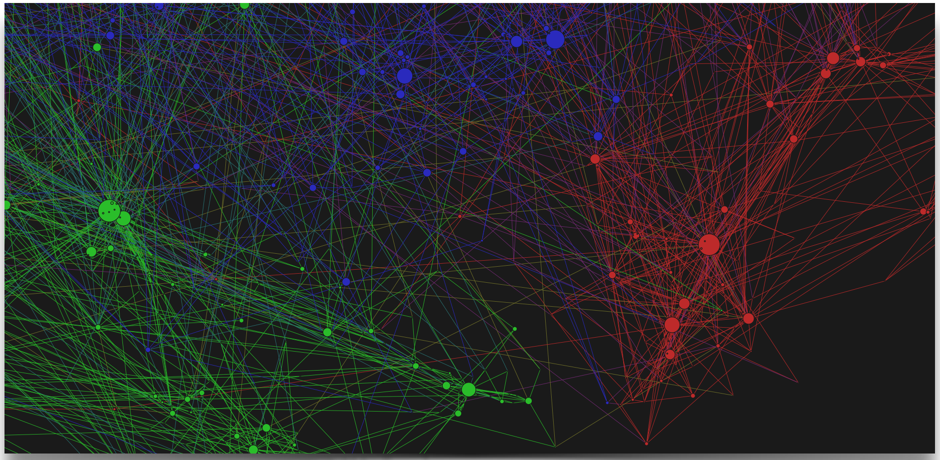
Infer (predict) remaining labels using the learned parameters $\Theta_{\mathscr{Y}}$ and available data

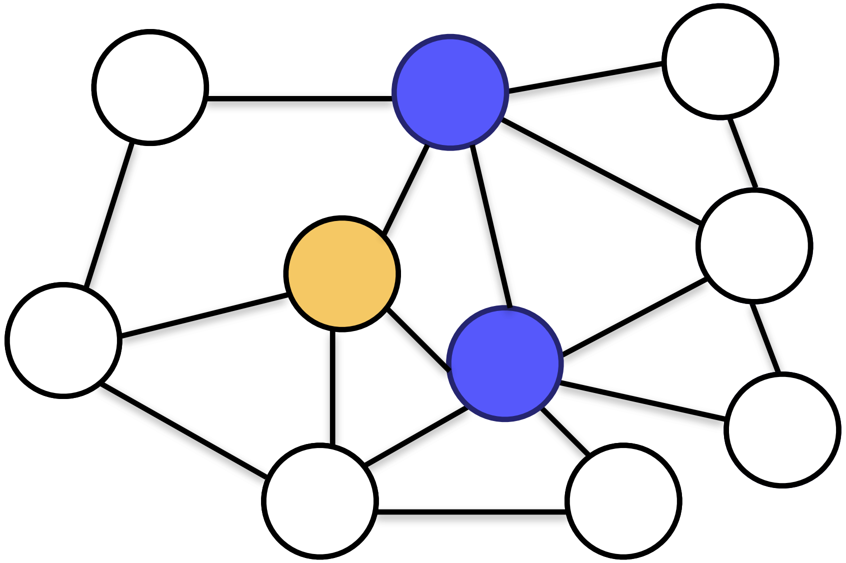
Will Show:
Learning/inference approximations prevent semi-supervised algorithms from converging

Will Develop/Demonstrate:
Methods that model parameter uncertainty overcome approximation errors

Outline

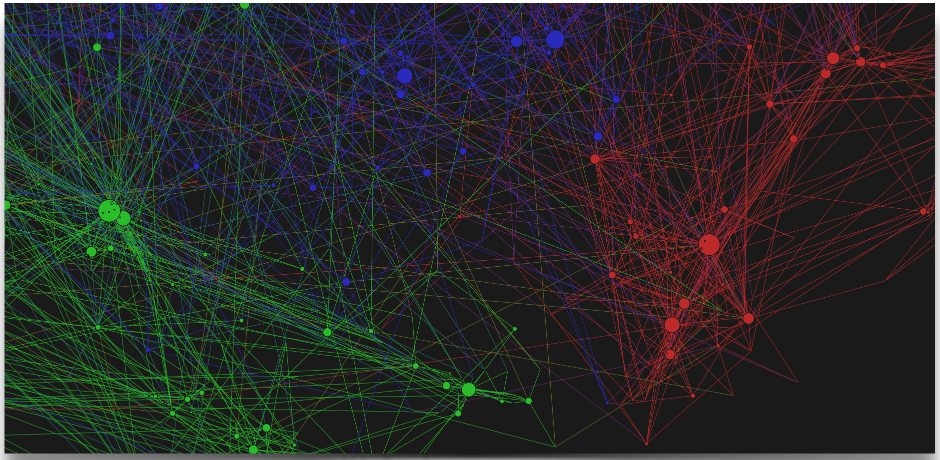
- **Introduction**
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- Conclusions

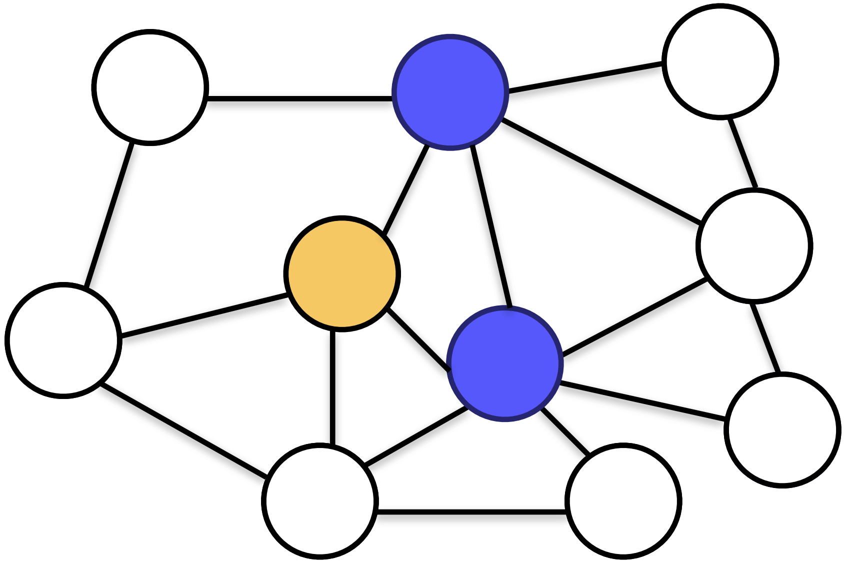




Outline

- Introduction
- **Background: Semi-supervised Relational Machine Learning**
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- Conclusions





Background - Semi-Supervised Relational Learning

Labels Edges

- Relational Machine Learning

Both learning and inference require approximations

  - Approximate Learning:

Pseudolikelihood (G_L)

M-Step: $\hat{\Theta}_{\mathscr{Y}} = \arg\max_{\Theta} \sum_{\mathbf{Y}_U} P_{\mathscr{Y}}(\mathbf{Y}_U \ldots) \sum_{v_i \in \mathbf{V}_L} \log P_{\mathscr{Y}}(y_i | \tilde{\mathbf{Y}}_{MB}(v_i), \mathbf{x}_i, \Theta_{\mathscr{Y}})$

Composite Likelihood (G_O)

How do these approximations affect semi-supervised relational learning?

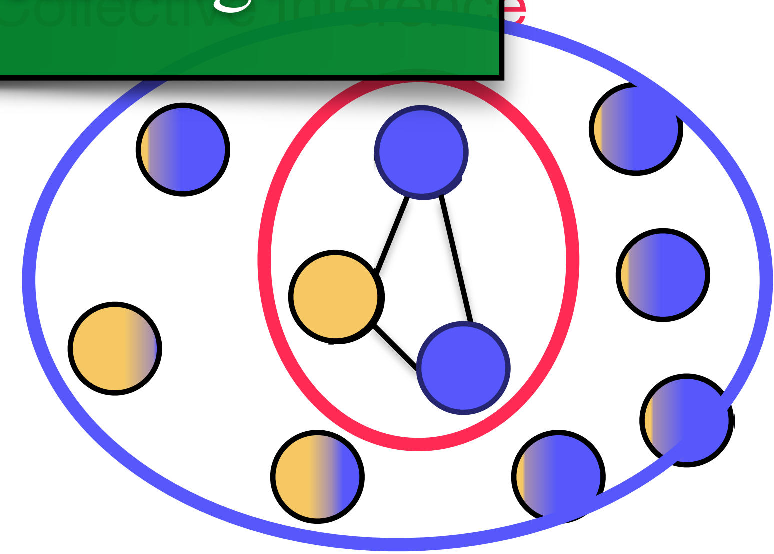
  - Approximate Inference:

E-Step: (For all unlabeled instances)

Collective Inference

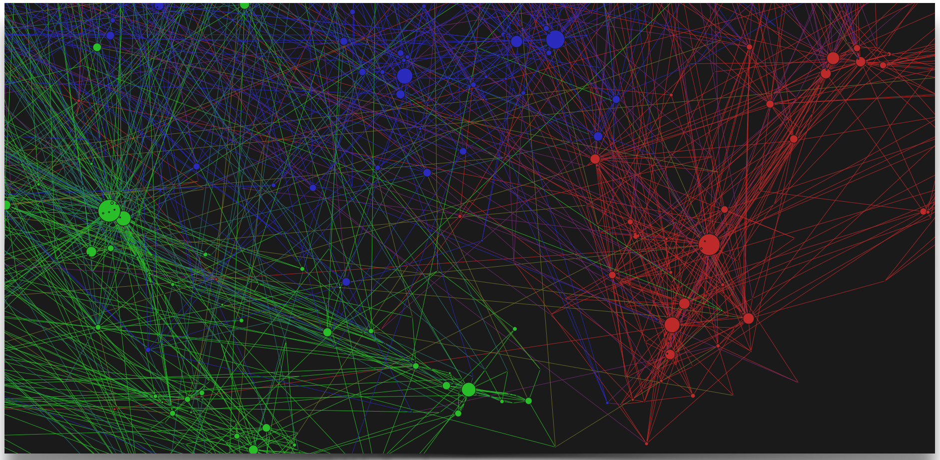
- Many conditional forms to choose from (e.g., Generative, Logistic)

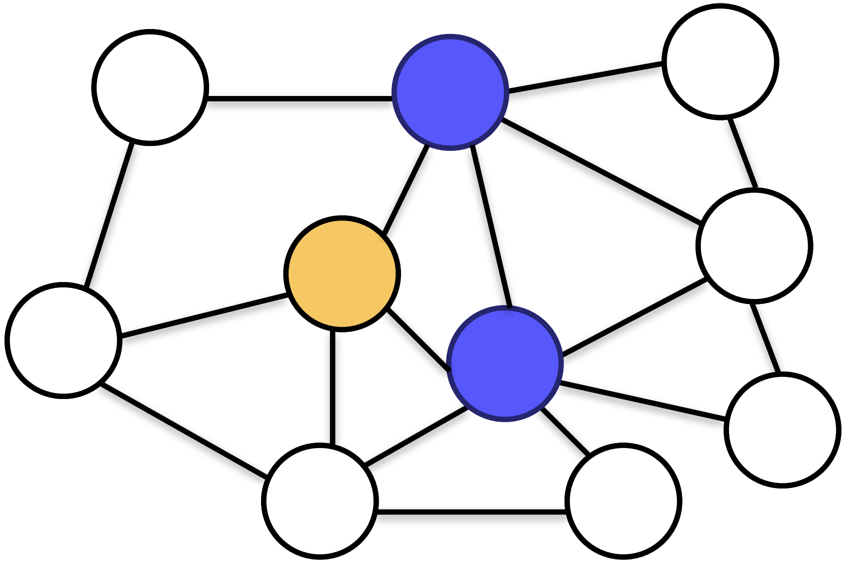
Semi-supervised Relational EM
(Xiang & Neville, 2009)



Outline

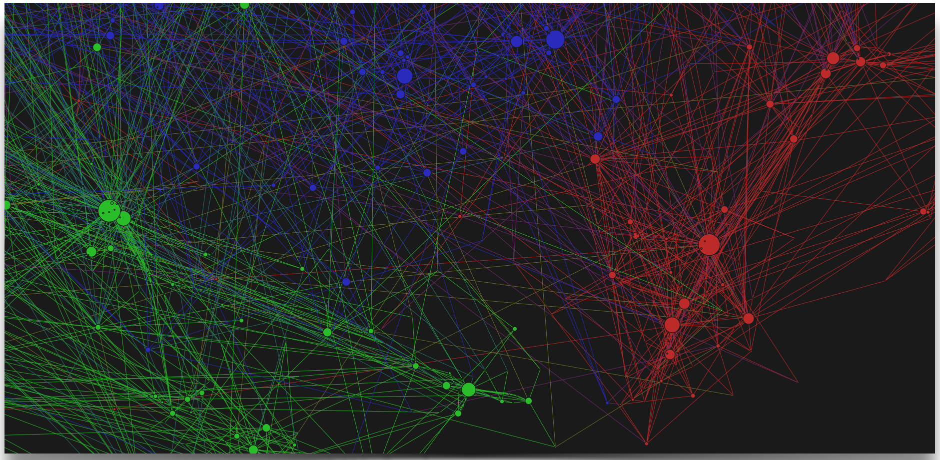
- Introduction
- **Background: Semi-supervised Relational Machine Learning**
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- Conclusions

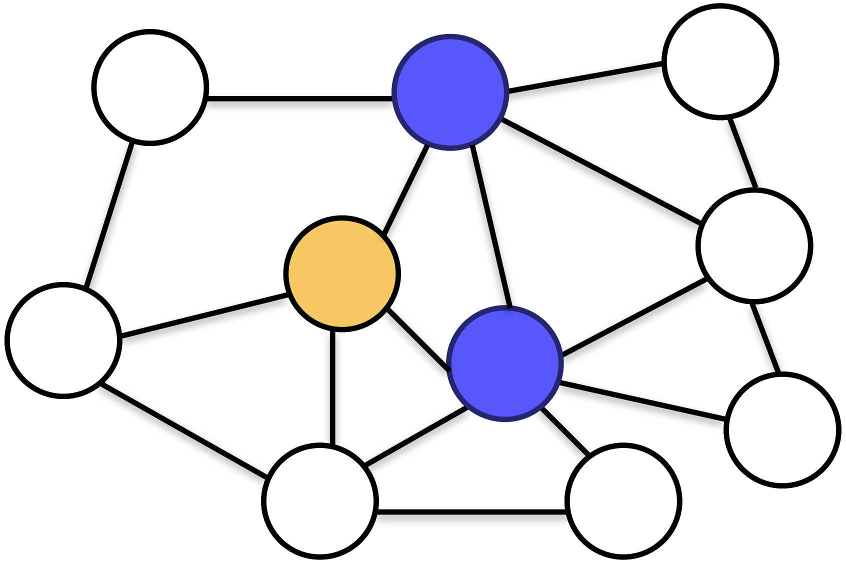




Outline

- Introduction
- Background: Semi-supervised Relational Machine Learning
- **Analysis of Relational EM convergence in real world networks**
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- Conclusions

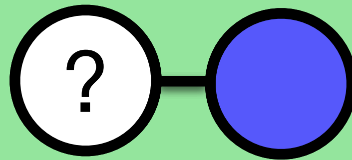




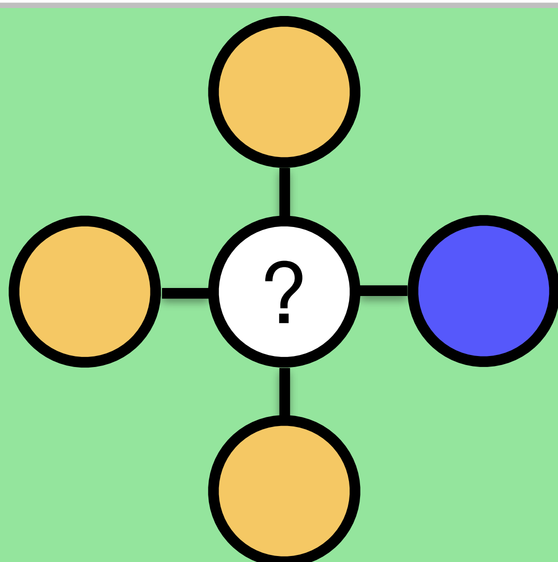
Impac

Semi-

- (Relati



$$P(\text{Purple}|\text{Purple Neighbor}) \approx 0.96$$

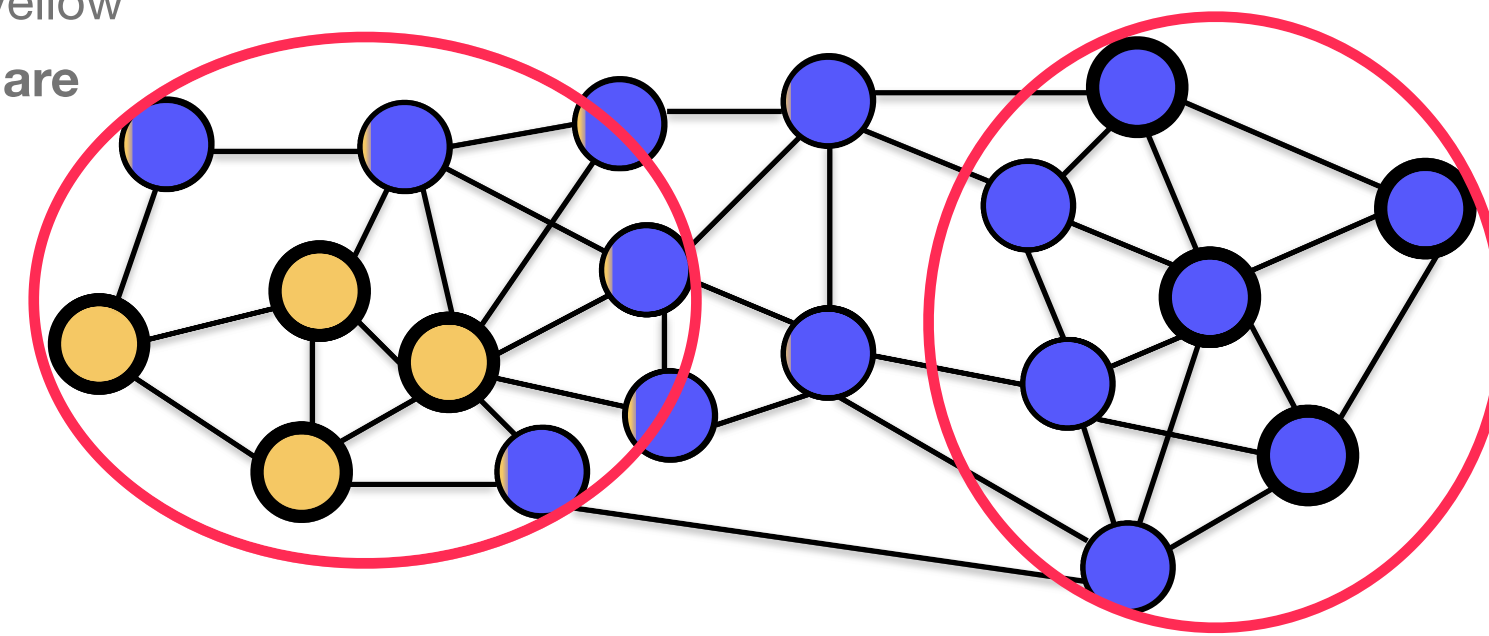


$$P(\text{Purple}|\text{Neighbors}) \approx 0.61$$

n

ion

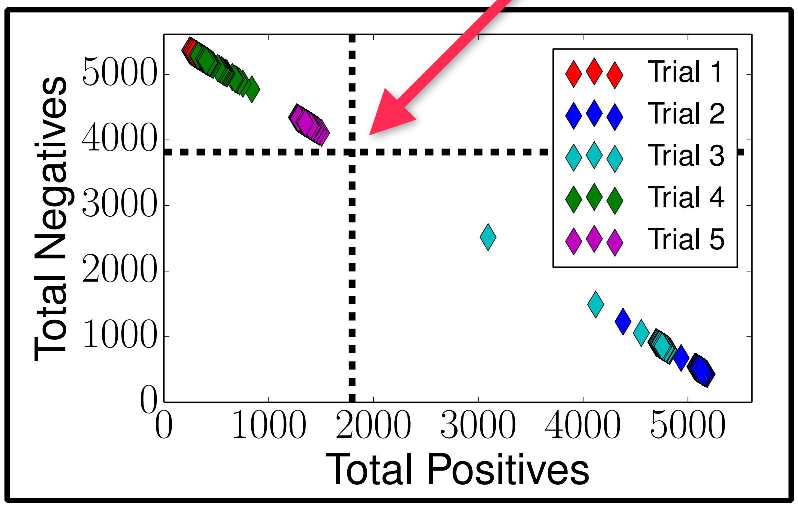
- *Over-F*
  - Both neighboring yellow
  - **Yellow neighbors are not *predictive***
- *Over-Correction*



Impact of Approximations in Semi-Supervised Relational Learning

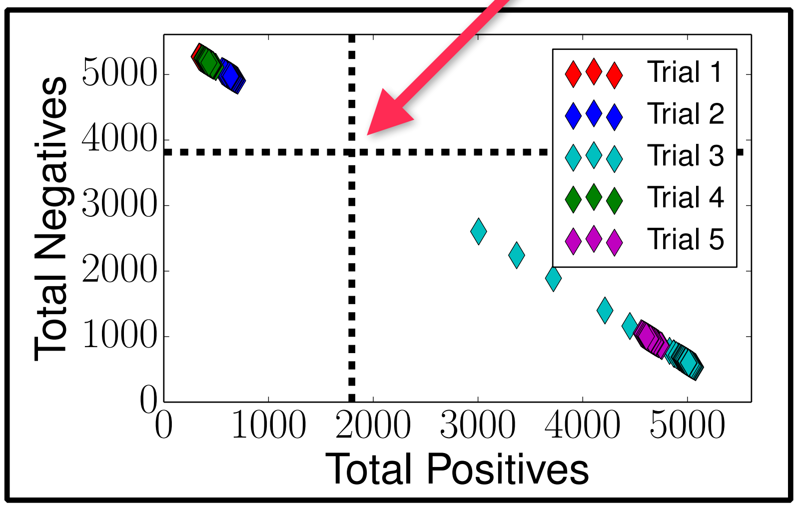
- Does **over propagation** during **prediction** happen in real world, sparsely labeled networks?

Class Prior



Relational Naive Bayes

Class Prior

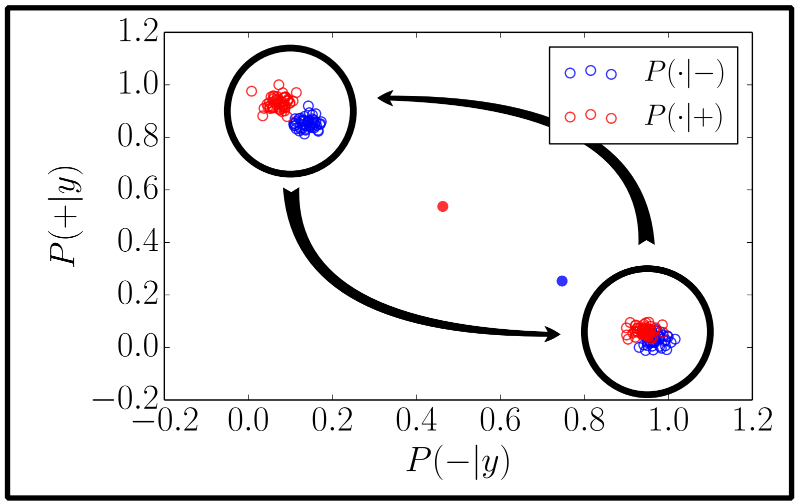


Relational Logistic Regression

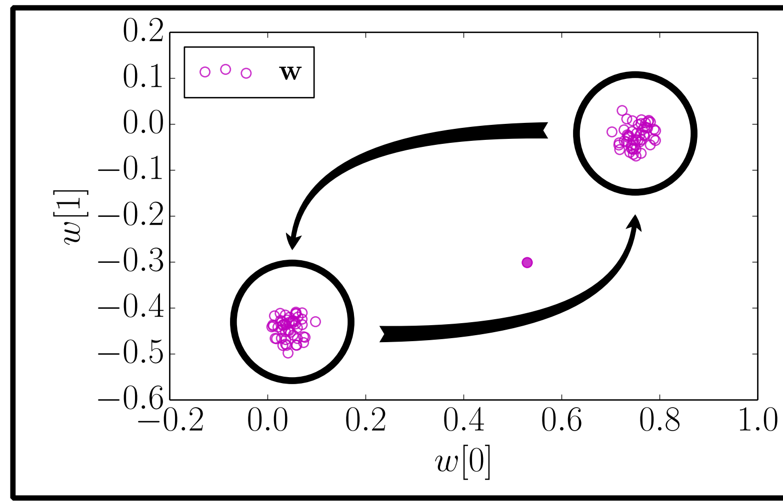
Yes

Impact of Approximations in Semi-Supervised Relational Learning

- Does **over correction** happen during **parameter estimation** for semi-supervised relational learning for real world, sparsely labeled networks?



Relational Naive Bayes

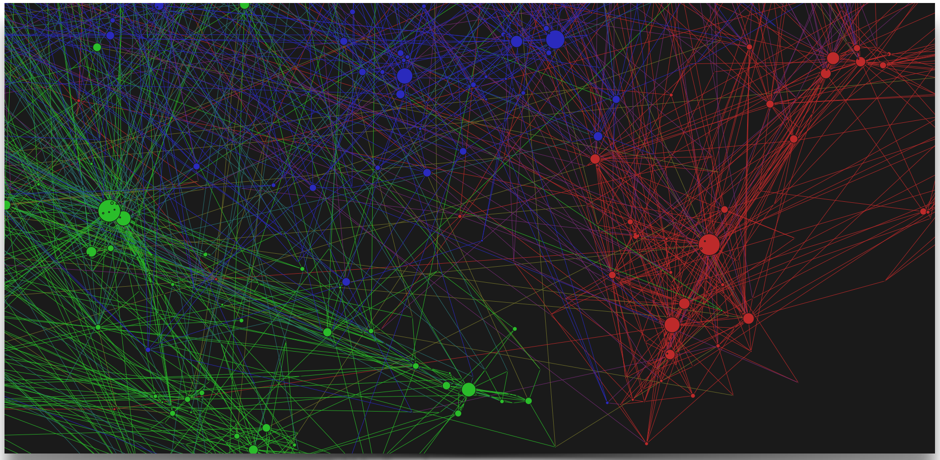


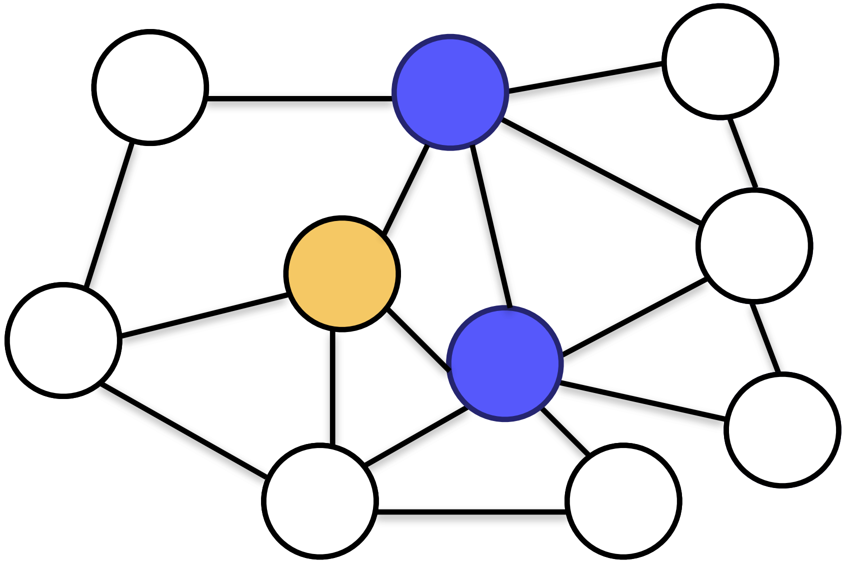
Relational Logistic Regression

Yes

Outline

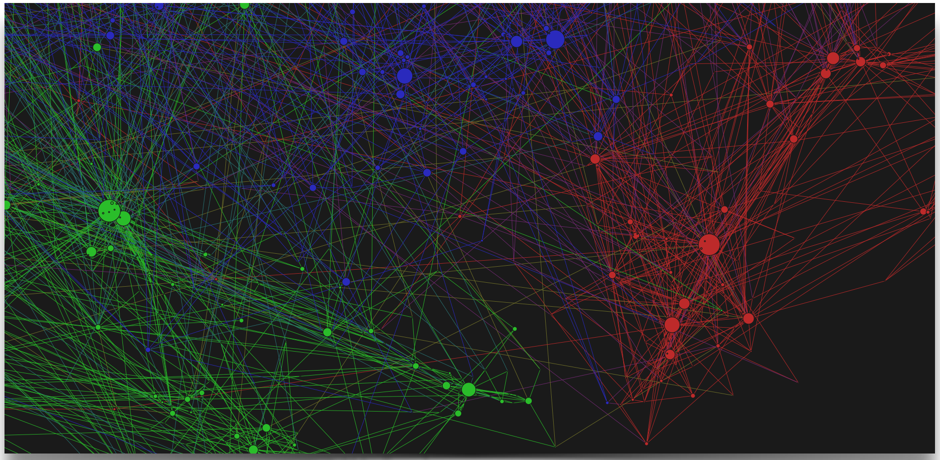
- Introduction
- Background: Semi-supervised Relational Machine Learning
- **Analysis of Relational EM convergence in real world networks**
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- Conclusions

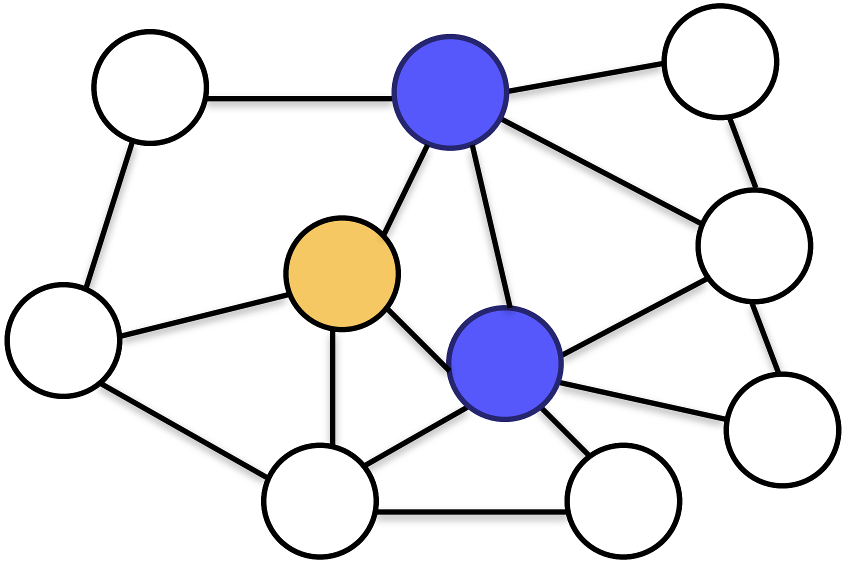




Outline

- Introduction
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- **Relational Stochastic EM and Relational Data Augmentation**
- Experiments
- Conclusions





Relational Stochastic EM and Relational Data Augmentation

Parameters	Predictions	
	Fixed Point	***Stochastic***
Fixed Point	*Relational EM*	—
Stochastic	***Relational Stochastic EM***	***Relational Data Augmentation***

Relational Stochastic EM

- Stochastic approximation to relational EM algorithm
  - *Sample from the joint conditional distribution of labels*
  - *Maximize the composite likelihood*
- Contrasts with Relational EM, which utilizes *expectations* of unlabeled items
- Average over the parameters to reduce the variance (Celex *et al.*, 2001)
- Inference is still performed using a single, *fixed point* set of parameters

Parameters	*Predictions*	
	Fixed Point	*Stochastic*
Fixed Point	*Relational EM*	–
Stochastic	***Relational Stochastic EM***	***Relational Data Augmentation***

```
Alternate Between:
    Gibbs sample of labels
```
$$\tilde{\mathbf{Y}}_U^t \sim P_{\mathscr{Y}}^t(\mathbf{Y}_U | \mathbf{Y}_L, \mathbf{X}, \mathbf{E}, \tilde{\Theta}_{\mathscr{Y}}^{t-1})$$
```
    Maximizing Parameters
```
$$\tilde{\Theta}_{\mathscr{Y}}^t = \arg\max_{\Theta_{\mathscr{Y}}} \sum_{v_i \in \mathbf{V}_L} \log P_{\mathscr{Y}}(y_i | \tilde{\mathbf{Y}}_{\mathcal{MB}(v_i)}^t, \mathbf{x}_i, \Theta_{\mathscr{Y}})$$
```
Aggregate Parameters:
```
$$\hat{\Theta}_{\mathscr{Y}} = \frac{1}{T} \sum_t \tilde{\Theta}_{\mathscr{Y}}^t$$
```
Final Inference:
```
$$P_{\mathscr{Y}}(\mathbf{Y}_U | \mathbf{Y}_L, \mathbf{X}, \mathbf{E}, \hat{\Theta}_{\mathscr{Y}})$$

Relational Data Augmentation

- Data Augmentation is a Bayesian viewpoint of EM
  - Parameters are random variables.
  - Computes *posterior* predictive distribution (Tanner&Wong,1987)
- Developed a version for relational semi-supervised learning
- Final inference is over a *distribution* of parameter values
- Requires prior distributions over the parameters and corresponding sampling methods
  - RNB: Beta (conjunctive prior)
  - RLR: Normal (Metropolis-Hastings sampler)

Parameters	Predictions: Fixed Point	Predictions: Stochastic
Fixed Point	*Relational EM*	–
Stochastic	***Relational Stochastic EM***	***Relational Data Augmentation***

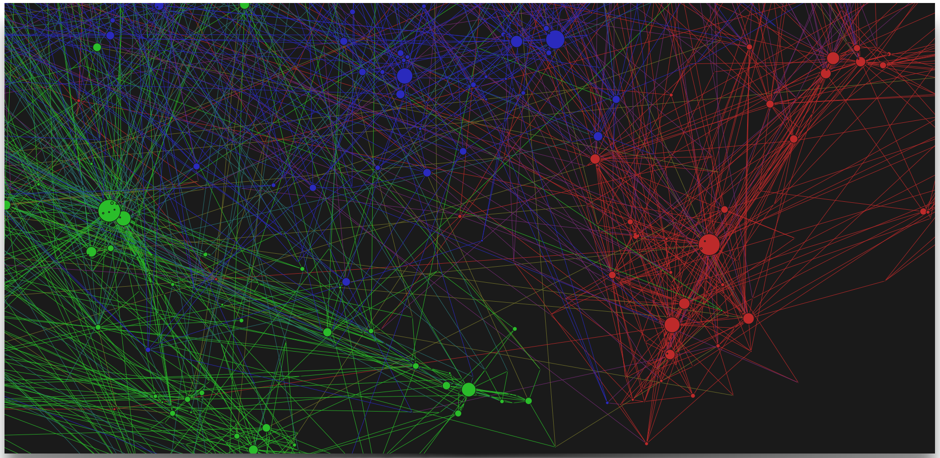
```
Alternate Between:
    Gibbs sample of labels
```
$$\tilde{\mathbf{Y}}_U^t \sim P_{\mathscr{Y}}^t(\mathbf{Y}_U | \mathbf{Y}_L, \mathbf{X}, \mathbf{E}, \tilde{\Theta}_{\mathscr{Y}}^{t-1})$$
```
    Sample Parameters
```
$$\tilde{\Theta}_{\mathscr{Y}}^t \sim P^t(\Theta_{\mathscr{Y}} | \tilde{\mathbf{Y}}_U^t, \mathbf{Y}_L, \mathbf{X}, \mathbf{E})$$

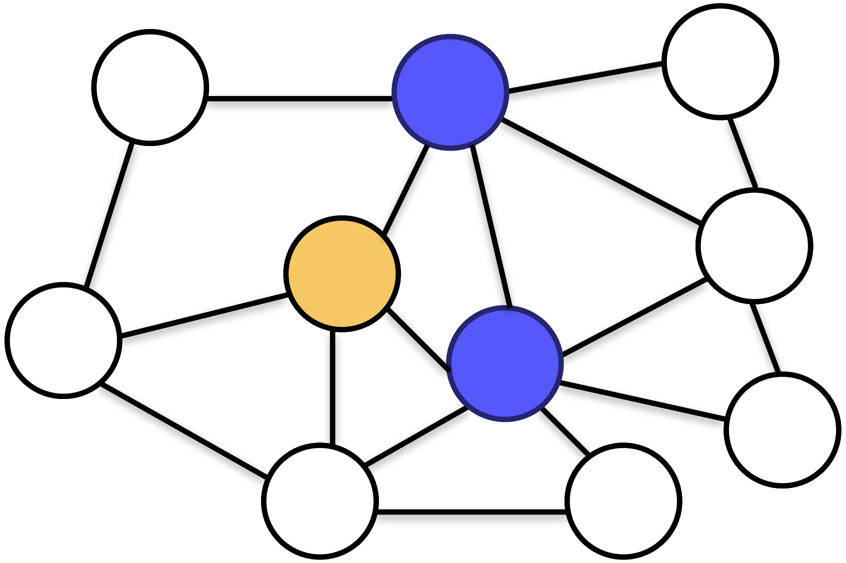
`Final Parameters:` $$\hat{\Theta}_{\mathscr{Y}} = \frac{1}{T}\sum_t \tilde{\Theta}_{\mathscr{Y}}^t$$

`Final Inference:` $$\hat{\mathbf{Y}}_U^t = \frac{1}{T}\sum_t \tilde{\mathbf{Y}}_U^t$$

Outline

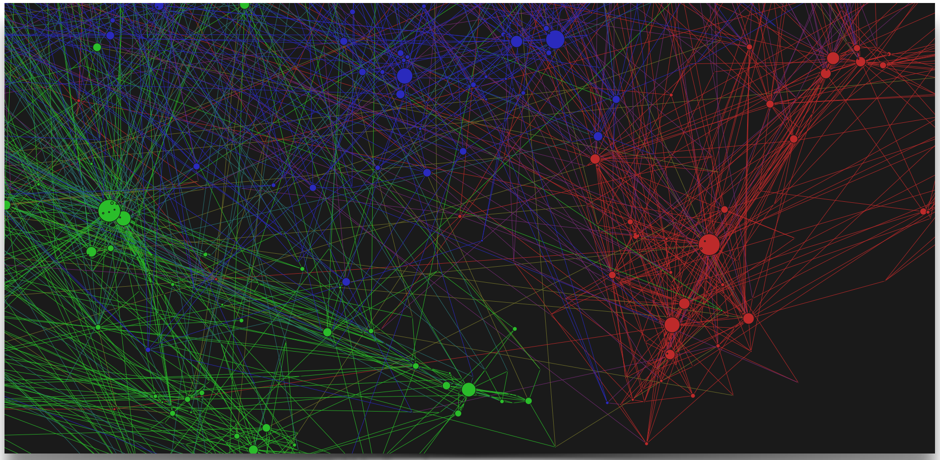
- Introduction
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- **Relational Stochastic EM and Relational Data Augmentation**
- Experiments
- Conclusions

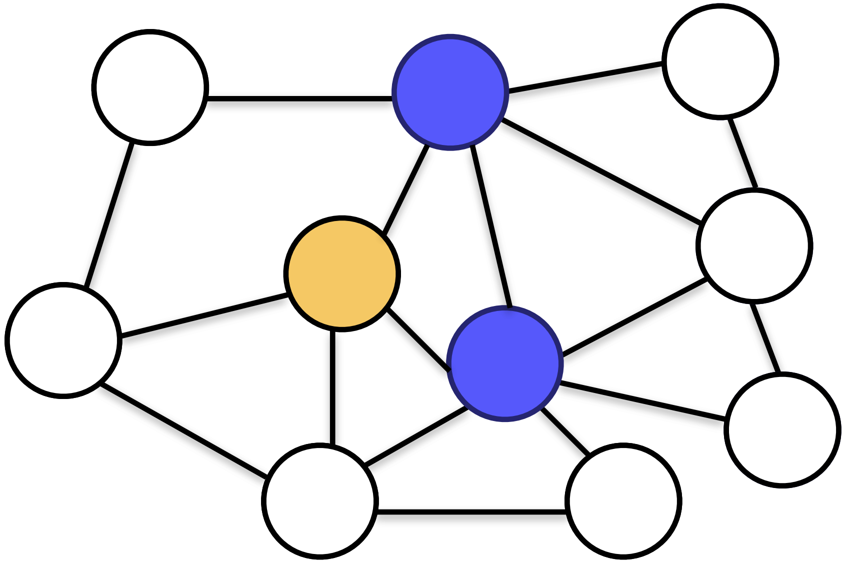




Outline

- Introduction
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- **Experiments**
- Conclusions



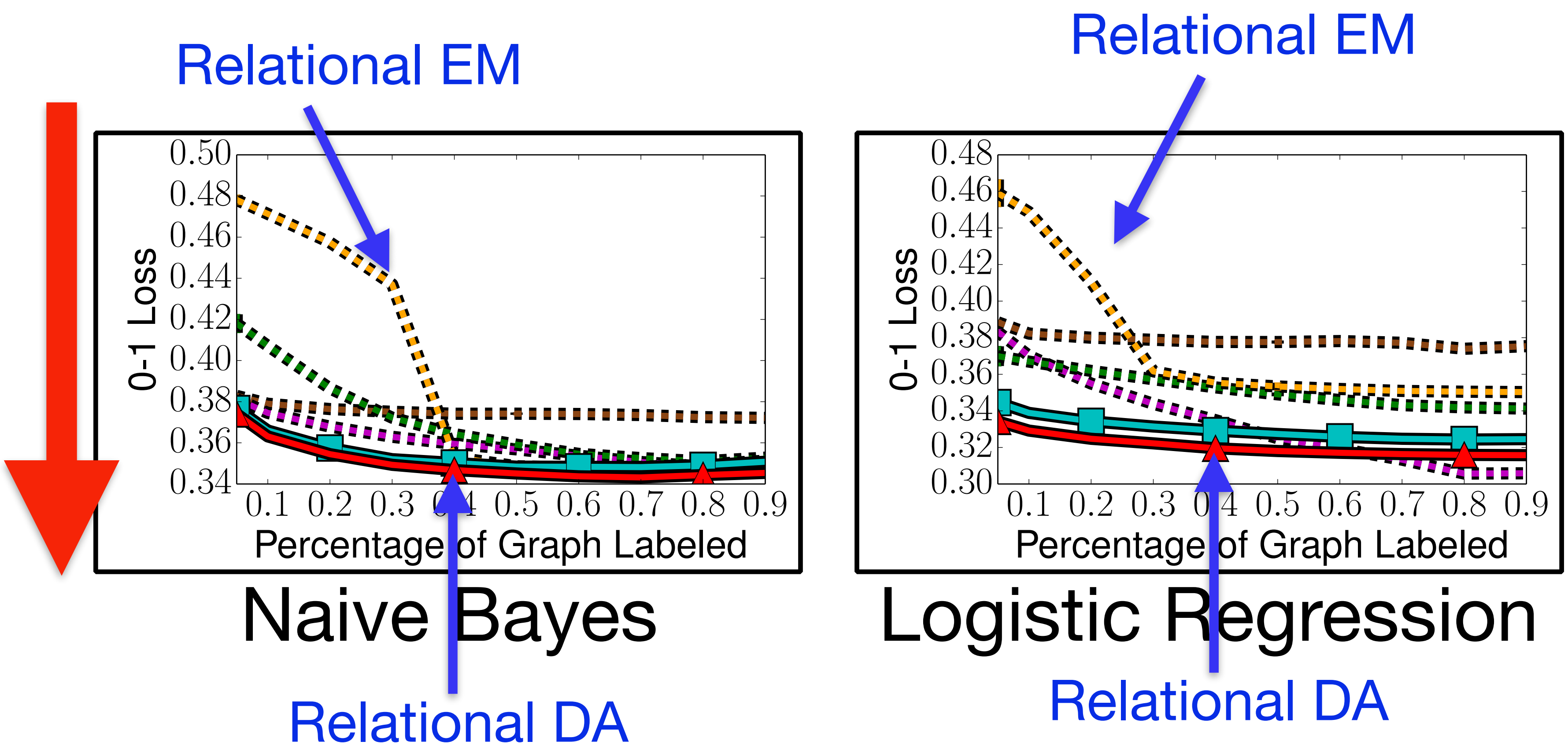


Experiments - Setup

- Compare on four real world networks
- Competing Models
  - Independent
  - Relational (No CI and CI)
  - Relational EM
  - **Relational SEM**
  - **Relational DA**
- Conditional Models
  - Relational Naive Bayes
  - Relational Logistic Regression
- Error measures
  - Zero-One Loss (0-1 Loss)
  - Mean absolute error (MAE)

Dataset	Vertices	Edges	Attributes
Facebook	5,906	73,374	2
IMDB	11,280	426,167	28
DVD	16,118	75,596	28
Music	56,891	272,544	26

Experiments - DVD



Relational EM's instability in sparsely labeled domains causes poor performance

Experiments - Facebook

Relational EM

Relational EM

Relational Data Augmentation can outperform Relational Stochastic EM

(In Paper): Cast collective inference as a nonlinear dynamical system to analyze the convergence of Relational SEM

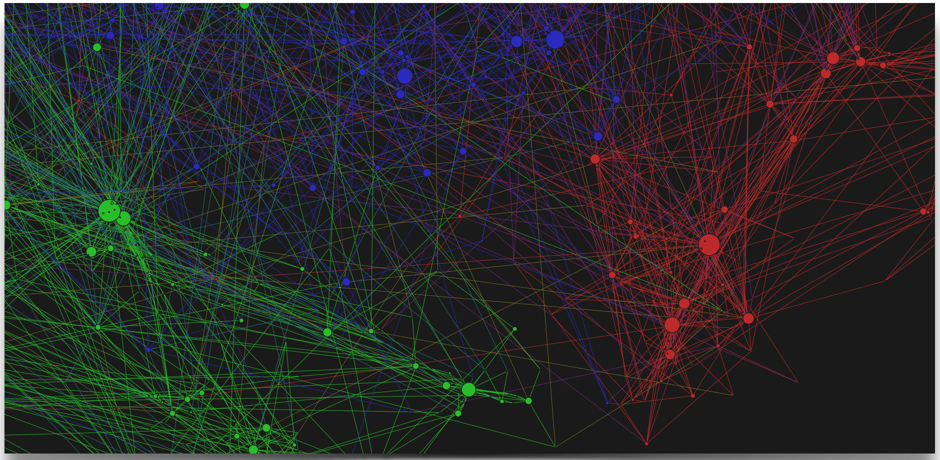
Relational DA

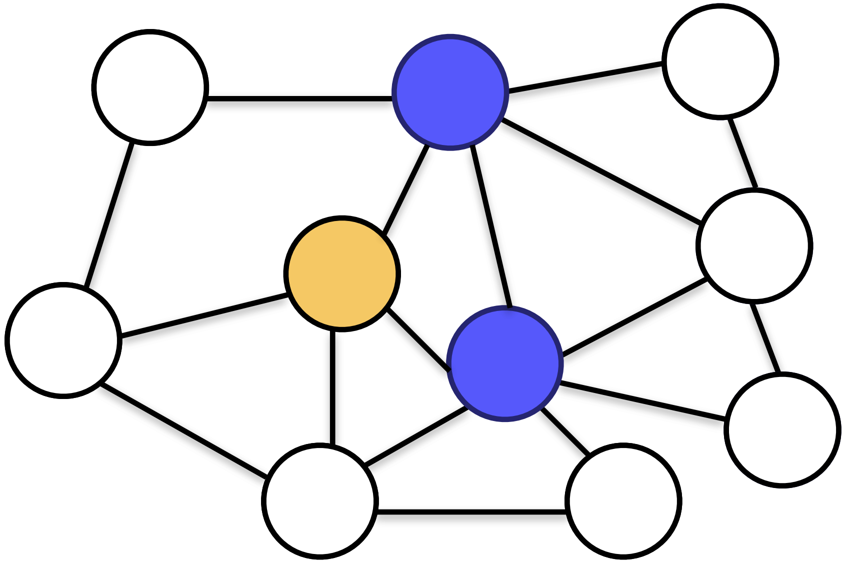
Relational DA

(Finding): Similar to Relational EM, Relational SEM can sometimes learn parameters that result in an unstable system

Outline

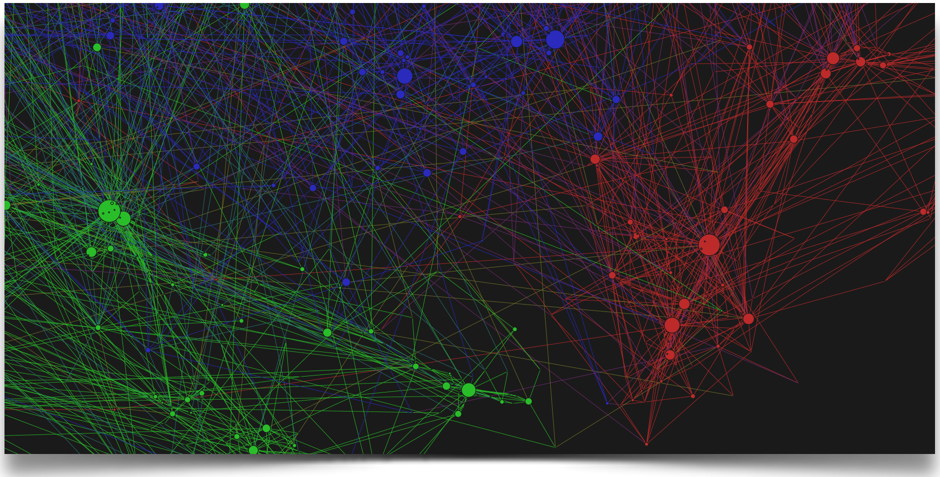
- Introduction
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- **Experiments**
- Conclusions

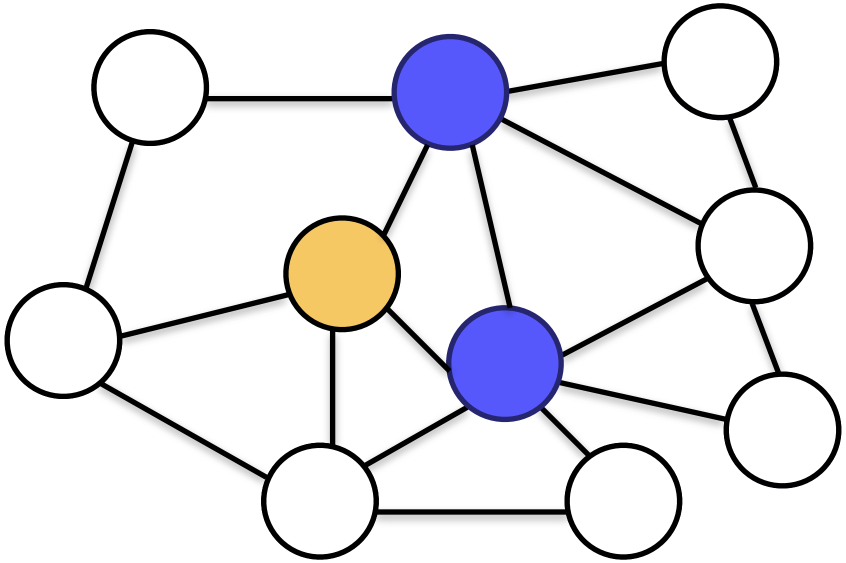




Outline

- Introduction
- Background: Semi-supervised Relational Machine Learning
- Analysis of Relational EM convergence in real world networks
- Relational Stochastic EM and Relational Data Augmentation
- Experiments
- **Conclusions**





Conclusions

- **Findings**
  - Fixed point approximation error led Relational EM to not converge
  - *Overpropagation* and *Overcorrection*
  - *(In Paper) Fixed point EM methods can result in unstable systems during inference*

- **Developed**
  - Relational Stochastic EM method has lower variance in parameter estimates
  - Relational Data Augmentation for computing a posterior predictive distribution for the unlabeled instances
    - Models the uncertainty over the parameter estimates, meaning it can outperform the Relational Stochastic EM approach
  - Both work well in conjunction with Composite Likelihood approximation
  - Both significantly outperformed a variety of competitors under a number of testing scenarios

Thank you!

jpfeiffer@purdue.edu
neville@cs.purdue.edu
paul.n.bennett@microsoft.com

Jennifer Neville

Paul Bennett