Exploiting Time-Varying Relationships in Statistical Relational Models

Umang Sharan   Jennifer Neville

Department of Computer Science
Purdue University, West Lafayette

Joint 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis, 2007

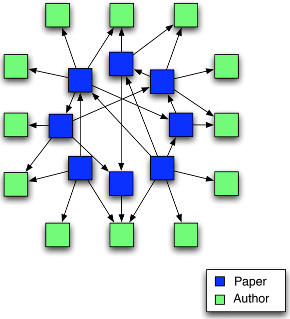
Outline

Outline

1. Introduction
   - The Basic Problem That We Studied
   - Related Work
2. Our Results/Contributions
   - Algorithm
   - Experiments and Results
3. Conclusions
   - Summary
   - Future Work

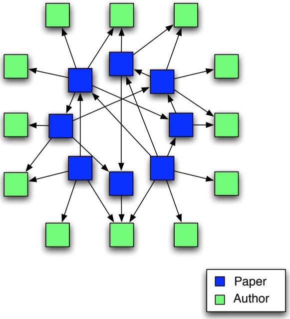
Example Task

- **Data**: Citation data with papers, authors, references.
- **Task**: Predict paper topic given coauthor and reference information.
- **Current Models**: Relational learning has achieved significant performance gains by exploiting homophily.



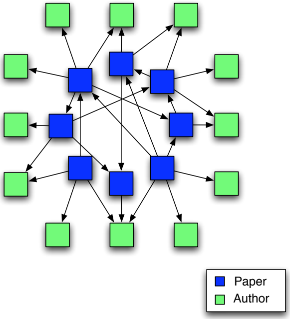
Example Task

- **Data**: Citation data with papers, authors, references.
- **Task**: Predict paper topic given coauthor and reference information.
- **Current Models**: Relational learning has achieved significant performance gains by exploiting homophily.

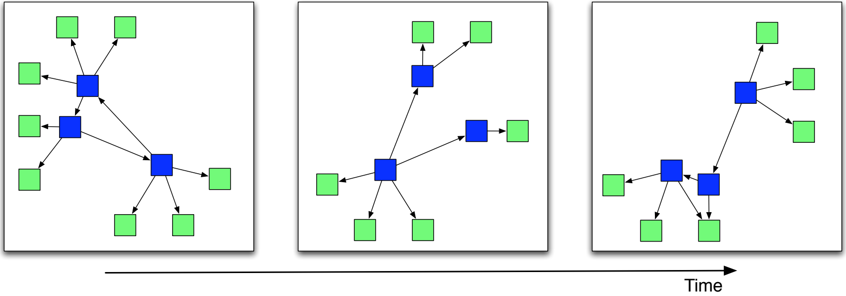


Example Task

- **Data**: Citation data with papers, authors, references.
- **Task**: Predict paper topic given coauthor and reference information.
- **Current Models**: Relational learning has achieved significant performance gains by exploiting homophily.

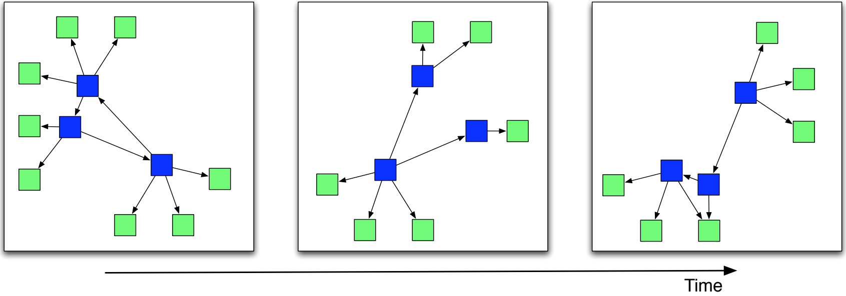


Why model time?



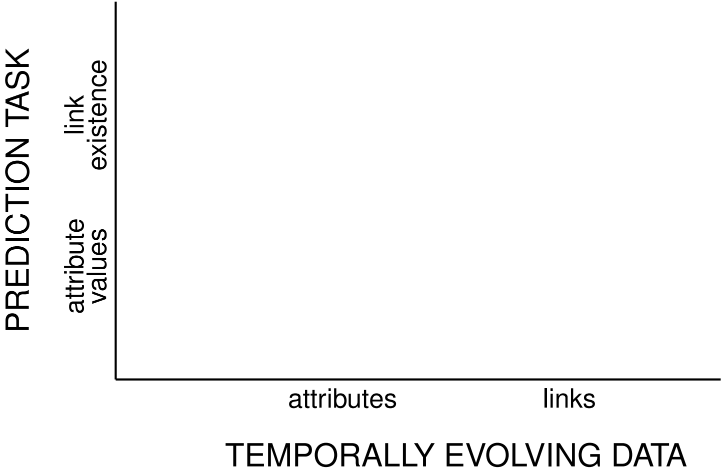
- Many relational domains have temporal dynamics (e.g fraud detection, web analysis, bioinformatics, etc).
- Temporal aspects contain information that is important to model (e.g. locality, recurrence).

Why model time?

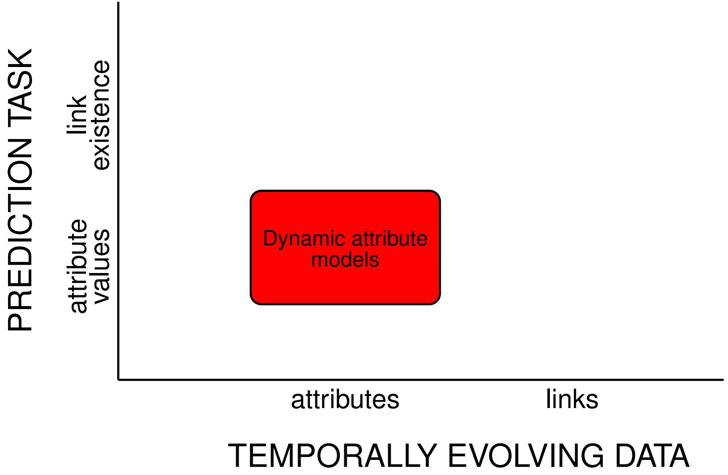


- Many relational domains have temporal dynamics (e.g fraud detection, web analysis, bioinformatics, etc).
- Temporal aspects contain information that is important to model (e.g. locality, recurrence).

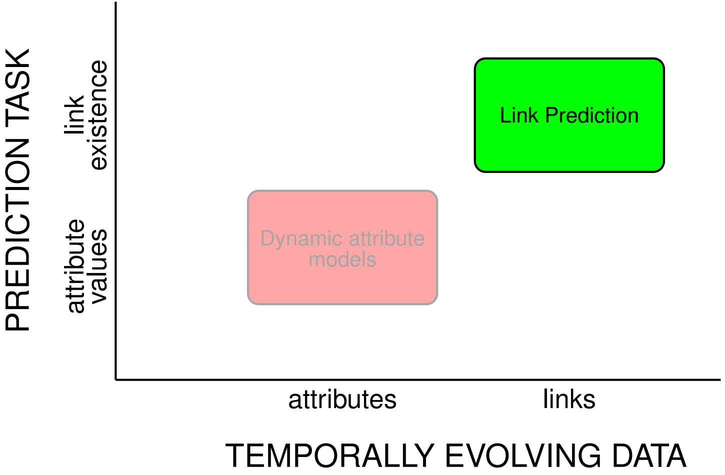
Temporal Relational Models



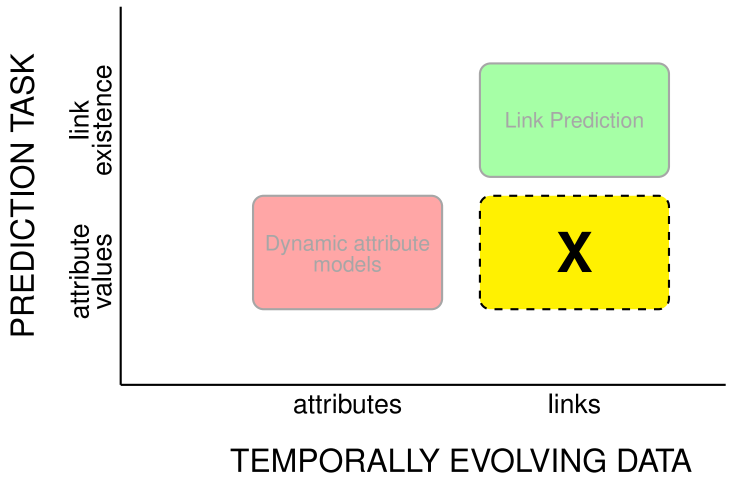
Temporal Relational Models



Temporal Relational Models



Temporal Relational Models



Outline

Time Varying RBC

- Model domains where link occurrence varies over time.
- Our algorithm takes a two-step approach:
    - Graph Summarization to capture link dynamics.
    - Weighted Relational Bayes Classifier to incorporate link strength into prediction.

Time Varying RBC

- Model domains where link occurrence varies over time.
- Our algorithm takes a two-step approach:
  - Graph Summarization to capture link dynamics.
  - Weighted Relational Bayes Classifier to incorporate link strength into prediction.

Time Varying RBC

- Model domains where link occurrence varies over time.
- Our algorithm takes a two-step approach:
  - Graph Summarization to capture link dynamics.
  - Weighted Relational Bayes Classifier to incorporate link strength into prediction.

Time Varying RBC

- Model domains where link occurrence varies over time.
- Our algorithm takes a two-step approach:
  - Graph Summarization to capture link dynamics.
  - Weighted Relational Bayes Classifier to incorporate link strength into prediction.

Graph Summarization

- Represent time-varying data as a temporal sequence of graphs $(G_1, G_2, \ldots, G_t)$.
- Summarize graph G_t^S as a weighted sum of graphs at each time step:

$$G_t^S = \begin{cases} (1-\theta)G_{t-1}^S + \theta G_t & \text{if } t > 1 \\ \theta G_t & \text{if } t = 1 \end{cases}$$

- θ is a weighing parameter defining the strength of a link (i.e. relationship) through time.

Graph Summarization

- Represent time-varying data as a temporal sequence of graphs $(G_1, G_2, \ldots, G_t)$.
- Summarize graph G_t^S as a weighted sum of graphs at each time step:

$$G_t^S = \begin{cases} (1-\theta)G_{t-1}^S + \theta G_t & \text{if } t > 1 \\ \theta G_t & \text{if } t = 1 \end{cases}$$

- θ is a weighing parameter defining the strength of a link (i.e. relationship) through time.

Graph Summarization

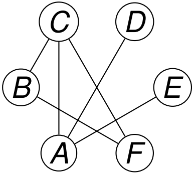
- Represent time-varying data as a temporal sequence of graphs $(G_1, G_2, \ldots, G_t)$.
- Summarize graph G_t^S as a weighted sum of graphs at each time step:

$$G_t^S = \begin{cases} (1-\theta)G_{t-1}^S + \theta G_t & \text{if } t > 1 \\ \theta G_t & \text{if } t = 1 \end{cases}$$

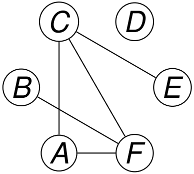
- θ is a weighing parameter defining the strength of a link (i.e. relationship) through time.

Graph Summarization - an example

Graph Summarization - an example

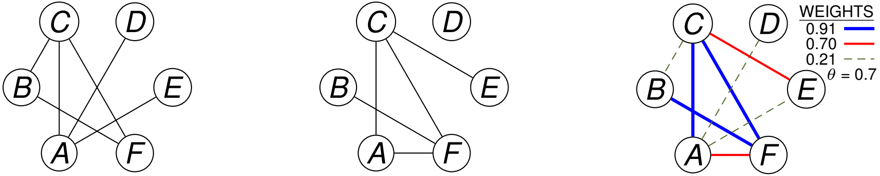


G_{1991}



G_{1992}

Graph Summarization - an example



G_{1991} G_{1992} G^S_{1992}

Relational Bayes Classifier

Baseline RBC (Neville et al. '03)

For a class label C, attributes $\mathbf{X}$, and related items R, the RBC calculates the probability of C for an item i of type $G(i)$ as follows:

$$P(C^i|\mathbf{X},R) \propto \prod_{X_m \in \mathbf{X}^{\mathbf{G(i)}}} P(X_m^i|C) \cdot \prod_{j \in R} \prod_{X_k \in \mathbf{X}^{\mathbf{G(j)}}} P(X_k^j|C) \cdot P(C)$$

TV-RBC

We incorporate the weights from the summary graph (each edge (i,j) has weight w_{ij}) as:

$$P(C_t^i|\mathbf{X},R) \propto \prod_{X_m \in \mathbf{X}^{\mathbf{G(i)}}} P(X_m^i|C) \cdot \prod_{j \in R} \prod_{X_k \in \mathbf{X}^{\mathbf{G(j)}}} w_{ij}^t \cdot P(X_k^j|C) \cdot P(C)$$

Relational Bayes Classifier

Baseline RBC (Neville et al. '03)

For a class label C, attributes $\mathbf{X}$, and related items R, the RBC calculates the probability of C for an item i of type $G(i)$ as follows:

$$P(C^i|\mathbf{X},R) \propto \prod_{X_m \in \mathbf{X^{G(i)}}} P(X_m^i|C) \cdot \prod_{j \in R} \prod_{X_k \in \mathbf{X^{G(j)}}} P(X_k^j|C) \cdot P(C)$$

TV-RBC

We incorporate the weights from the summary graph (each edge (i,j) has weight w_{ij}) as:

$$P(C_t^i|\mathbf{X},R) \propto \prod_{X_m \in \mathbf{X^{G(i)}}} P(X_m^i|C) \cdot \prod_{j \in R} \prod_{X_k \in \mathbf{X^{G(j)}}} w_{ij}^t \cdot P(X_k^j|C) \cdot P(C)$$

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

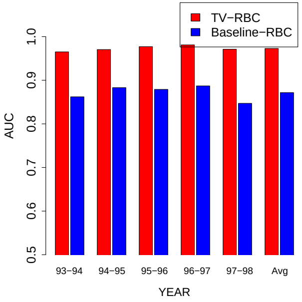
Experimental Setup

- **Dataset**: 4,330 CORA machine learning papers published between 1981 and 1998.
- **Task**: Predict paper topic given topics of references and coauthor papers.
- **Models**:
  - *TV-RBC*: learn weighted RBC on G_t^S, apply on G_{t+1}^S.
  - *Baseline RBC*: learn on snapshot at t, apply on snapshot at $t+1$.
- **Experiments**:
  - *Fixed* θ: Measure AUC for θ=0.7.
  - *Varying* θ: Measure AUC for varying θ.

Results: Fixed θ

Variation of AUC across different years for θ = 0.7

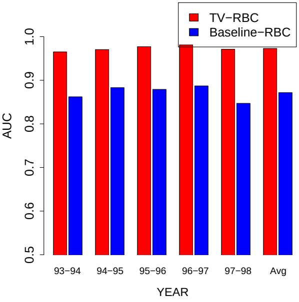
Results: Fixed θ



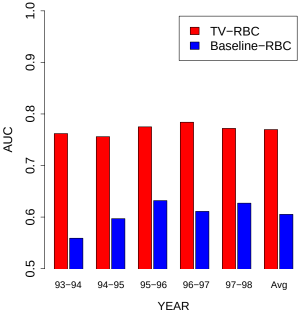
Refs + Coauthors

Variation of AUC across different years for θ = 0.7

Results: Fixed θ



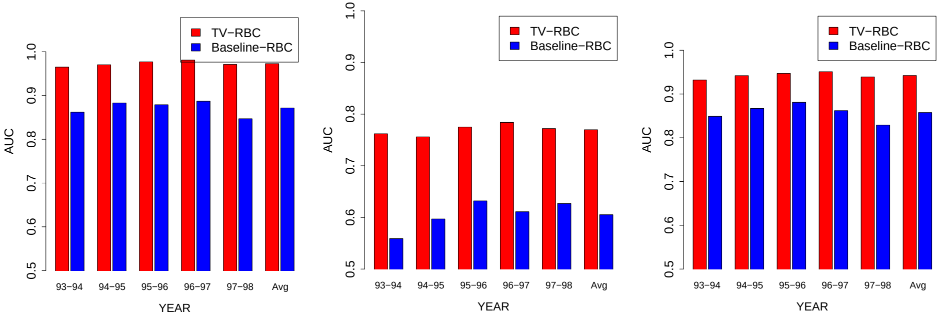
Refs + Coauthors



Coauthors only.

Variation of AUC across different years for θ = 0.7

Results: Fixed θ



Refs + Coauthors

Coauthors only.

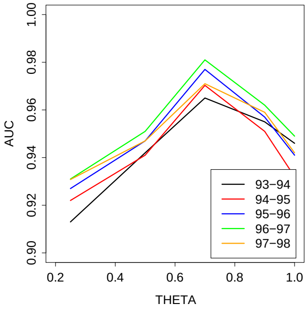
References only.

Variation of AUC across different years for θ = 0.7

Results: Varying θ

Variation of AUC with θ.

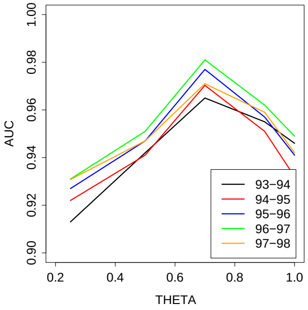
Results: Varying θ



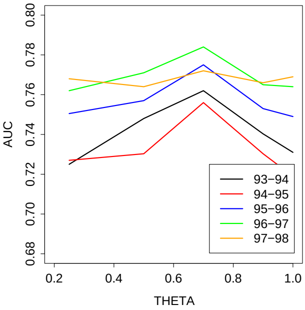
Refs + Coauthors

Variation of AUC with θ.

Results: Varying θ



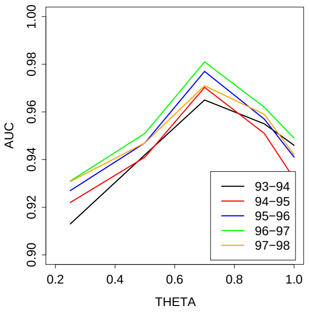
Refs + Coauthors



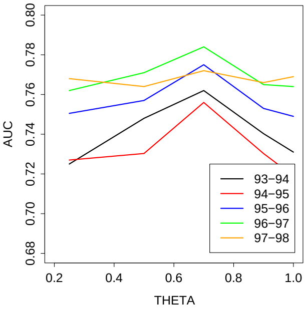
Coauthors only.

Variation of AUC with θ.

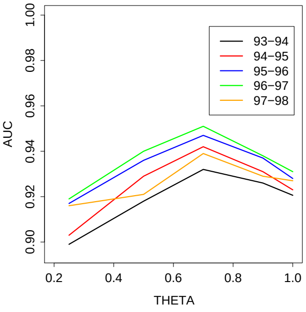
Results: Varying θ



Refs + Coauthors



Coauthors only.



References only.

Variation of AUC with θ.

Outline

Conclusions

- New approach to modeling relational data with time-varying link structure.
- Time-Varying RBC - uses summarized link graph and weighted relational bayes classifier for prediction.
- Our approach to summarization can be used with other SRL models as well (provided attribute values can be weighted).

Conclusions

- New approach to modeling relational data with time-varying link structure.
- Time-Varying RBC - uses summarized link graph and weighted relational bayes classifier for prediction.
- Our approach to summarization can be used with other SRL models as well (provided attribute values can be weighted).

Conclusions

- New approach to modeling relational data with time-varying link structure.
- Time-Varying RBC - uses summarized link graph and weighted relational bayes classifier for prediction.
- Our approach to summarization can be used with other SRL models as well (provided attribute values can be weighted).

Future Work

- Evaluate the algorithm on other real-world datasets.
- Compare the performance of TV-RBC with other modified SRL models (e.g. RPTs).
- Develop a temporal cross-validation approach to set the value of θ automatically during learning.
- Extend the approach to model temporally-varying attributes.

Future Work

- Evaluate the algorithm on other real-world datasets.
- Compare the performance of TV-RBC with other modified SRL models (e.g. RPTs).
- Develop a temporal cross-validation approach to set the value of θ automatically during learning.
- Extend the approach to model temporally-varying attributes.

PURDUE
UNIVERSITY

Future Work

- Evaluate the algorithm on other real-world datasets.
- Compare the performance of TV-RBC with other modified SRL models (e.g. RPTs).
- Develop a temporal cross-validation approach to set the value of θ automatically during learning.
- Extend the approach to model temporally-varying attributes.

Future Work

- Evaluate the algorithm on other real-world datasets.
- Compare the performance of TV-RBC with other modified SRL models (e.g. RPTs).
- Develop a temporal cross-validation approach to set the value of θ automatically during learning.
- Extend the approach to model temporally-varying attributes.

Thank You

- Umang Sharan,
  Department of Computer Science,
  Purdue University
  e-mail: usharan@cs.purdue.edu

- Prof. Jennifer Neville,
  Department of Computer Science,
  Purdue University
  e-mail: neville@cs.purdue.edu