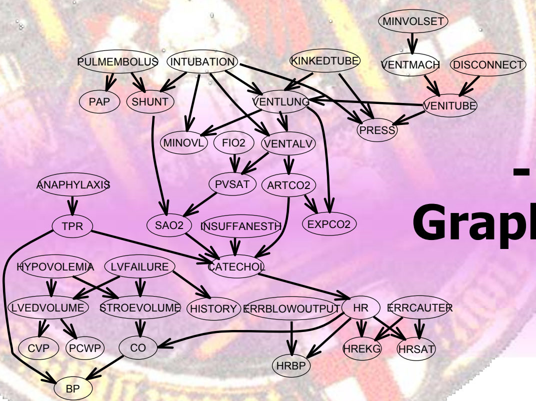
Based on Cussens and Kersting's ICML 2004 tutorial, De Raedt and Kersting's ECML/PKDD 2005 tutorial, and Friedman and Koller's NIPS 1999 tutorial

Advanced I WS 06/07



- Relational - Graphical Models

Wolfram Burgard, Luc De Raedt, Kristian Kersting, Bernhard Nebel

Albert-Ludwigs University Freiburg, Germany

Advanced I WS 06/07

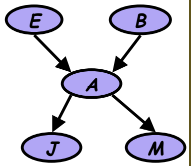
Outline

- Introduction
- Reminder: Probability theory
- Basics of Bayesian Networks
- Modeling Bayesian networks
- Inference (VE, Junction tree)
- [Excourse: Markov Networks]
- Learning Bayesian networks
- Relational Models

Advanced I WS 06/07

Bayesian Networks

1. Finite, acyclic graph
2. Nodes: (discrete) random variables
3. Edges: direct influences
4. Associated with each node: a table representing a conditional probability distribution (CPD), quantifying the effect the parents have on the node



$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} \boxed{P(X_i \mid \mathrm{pa}(X_i))}$$

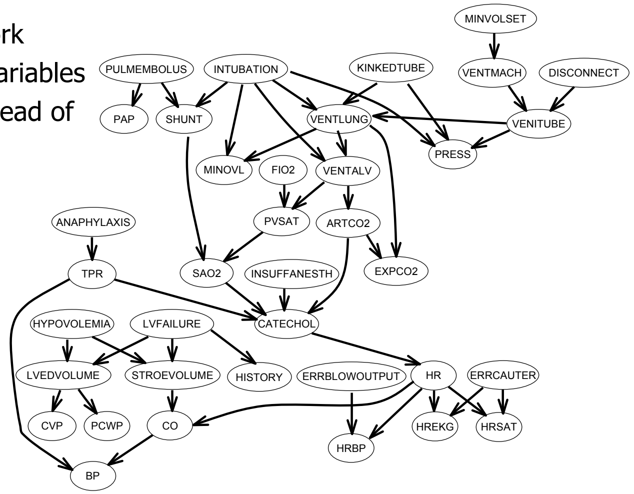
Advanced I WS 06/07

Bayesian Networks

The "ICU alarm" network

- 37 binary random variables
- 509 parameters instead of

$2^{37} = 10^{12}$



Bayesian Networks

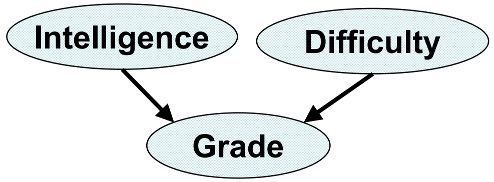
- Effective (and to some extend efficient) inference algorithms
  - Variable elimination
  - Junction Trees
  - MPE
- Effective (and to some extend efficient) learning approaches
  - Expectation Maximization
  - Gradient Ascent

Probability: Dealing with noisy data, missing data and hidden variables

Learning: Knowledge Acquisition Bottleneck, Data cheap

Bayesian Networks: Problem

- Bayesian nets use propositional representation
- Real world has objects, related to each other

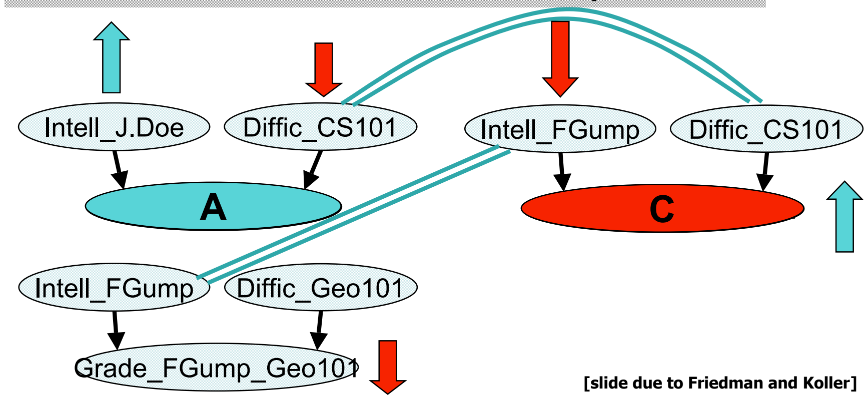


[slide due to Friedman and Koller]

Bayesian Networks: Problem

- Bayesian nets use propositional representation
- Real world has objects, related to each other

These “instances” are not independent!

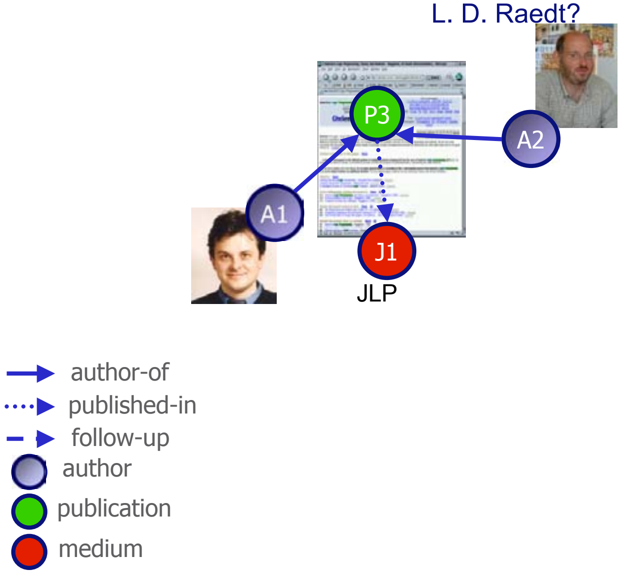


[slide due to Friedman and Koller]

How to Craft and Publish Papers

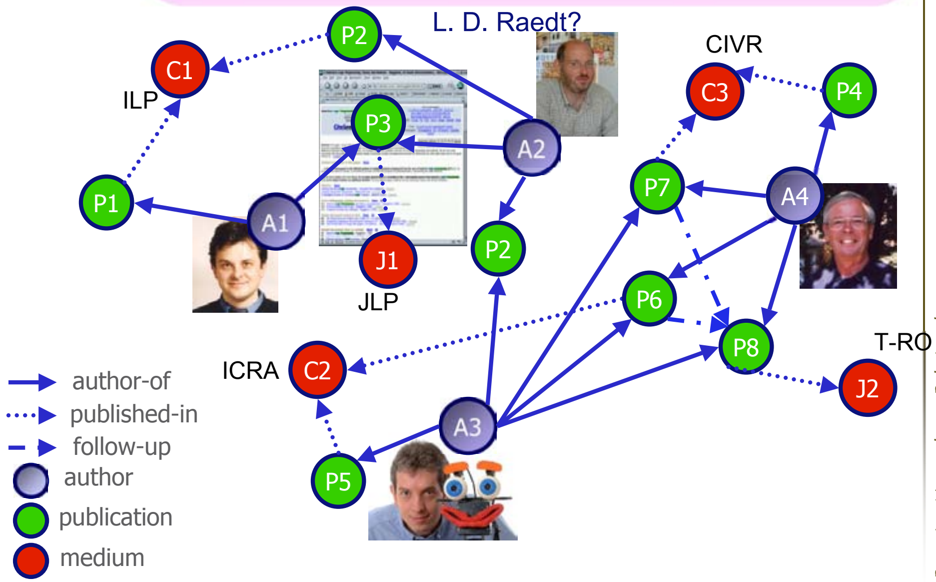
- Are there similar papers?
- Which papers are relevant?
- Keywords Extraction
- Does anybody know L. D. Raedt?

How to Craft and Publish Papers



Advanced I WS 06/07

How to Craft and Publish Papers

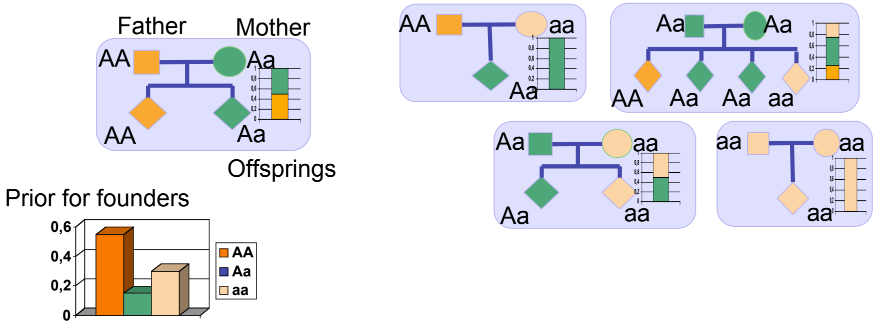


Bayesian Networks - Relational

Advanced I WS 06/07

Blood Type / Genetics/ Breeding

- 2 Alleles: A and a
- Probability of Genotypes AA, Aa, aa ?

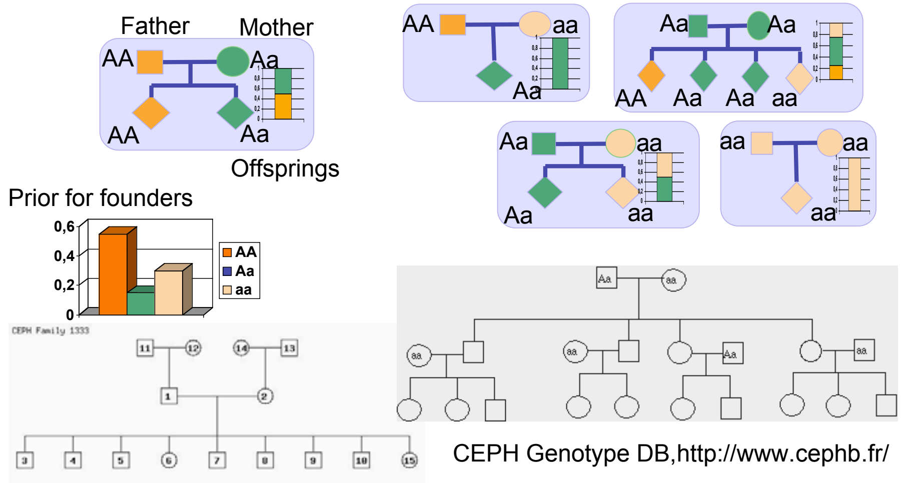


Bayesian Networks - Relational

Advanced I WS 06/07

Blood Type / Genetics/ Breeding

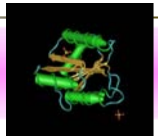
- 2 Alleles: A and a
- Probability of Genotypes AA, Aa, aa ?

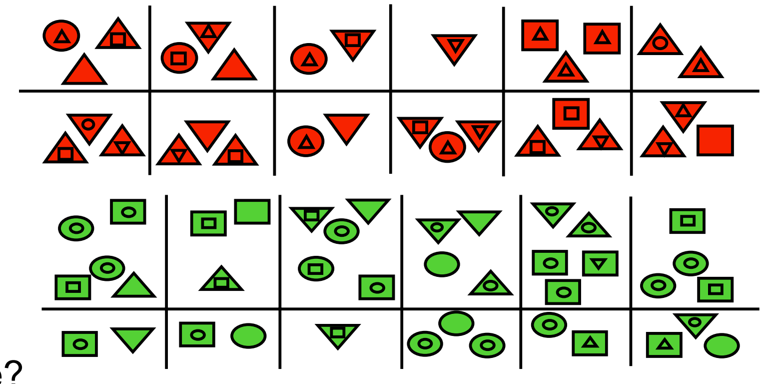


CEPH Genotype DB,http://www.cephb.fr/

Bayesian Networks - Relational

Bongard´s Problems



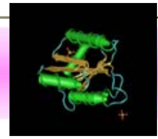


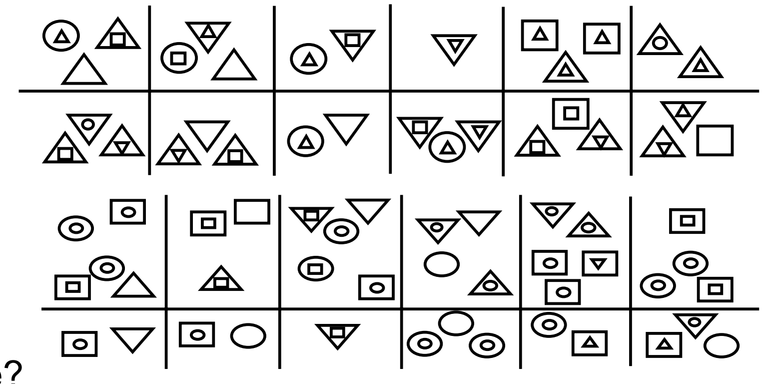
Noise?

Some objects are opaque?

(e.g. `in` relation is not always observed)

Bongard´s Problems





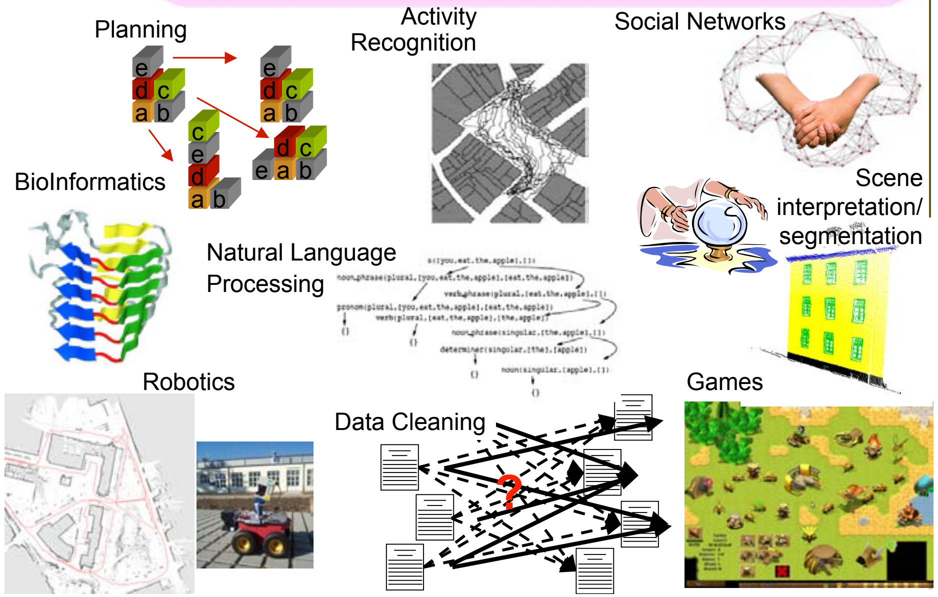
Noise?

Some objects are opaque?

(e.g. `in` relation is not always observed)

Clustering?

... Other Application Areas



Why do we need relational models?

- Rich Probabilistic Models
- Comprehensibility
- Generalization (similar situations/individuals)
- Knowledge sharing
- Parameter Reduction / Compression
- Learning
  - Reuse of experience (training one RV might improve prediction at other RV)
  - More robust
  - Speed-up

When to apply relational models ?

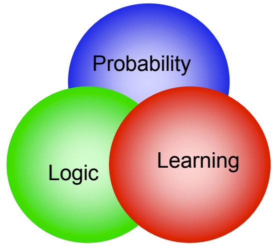
- When it is impossible to elegantly represent your problem in attribute value form
  - variable number of 'objects' in examples
  - relations among objects are important

A1	A2	A3	A4	A5	A6
true	true	?	true	false	false
?	true	?	?	false	false
...	...	...	...	...	...
true	false	?	false	true	?

attribute value form

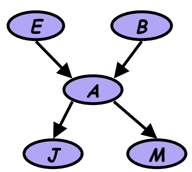
Statistical Relational Learning

... deals with machine **learning** and data mining in **relational** domains where observations may be **missing, partially observed**, and/or **noisy** ...



and is one of the key open questions in AI.

BNs = Probabilistic Propositional Logic




```
E.
B.
A  :- E, B.
J  :- A.
M  :- A.
```

+ CPDs +

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}\boxed{P(X_i|\,\mathrm{pa}(X_i))}$$

Logic Programming

```
father(rex,fred).      mother(ann,fred).
father(brian,doro).    mother(utta, doro).
father(fred,henry).    mother(doro,henry).
pc(rex,a).  mc(rex,a).
pc(ann,a).  mc(ann,b).
...
```

The maternal information `mc/2` depends on the maternal and paternal `pc/2` information of the mother `mother/2`:

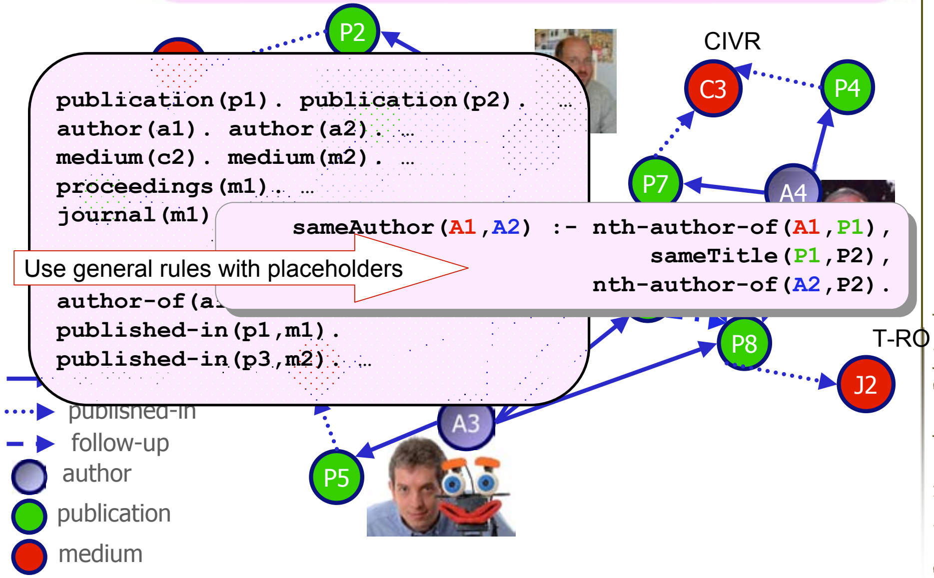
```
mchrom(fred,a). mchrom(fred,b),...
```

or better

```
mc(P,a) :- mother(M,P), pc(M,a), mc(M,a).
mc(P,a) :- mother(M,P), pc(M,a), mc(M,b).
mc(P,b) :- mother(M,P), pc(M,a), mc(M,b).
...
```

Placeholder
Could be `rex`, `fred`, `doro`, ...

How to Craft and Publish Papers




```
publication(p1). publication(p2). …
author(a1). author(a2). …
medium(c2). medium(m2). …
proceedings(m1). …
journal(m1)
author-of(a…
published-in(p1,m1).
published-in(p3,m2). …
```

Use general rules with placeholders

```
sameAuthor(A1,A2) :- nth-author-of(A1,P1),
                     sameTitle(P1,P2),
                     nth-author-of(A2,P2).
```

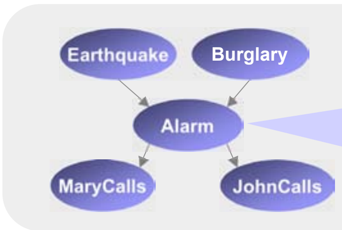

Outline Relational Models

- Relational Models
  - Probabilistic Relational Models
  - Baysian Logic Programs
  - Relational Markov networks
  - Markov Logic

Probabilistic Relational Models (PRMs)

[Koller,Pfeffer,Getoor]

- Database theory
- Entity-Relationship Models
  - Attributes = RVs

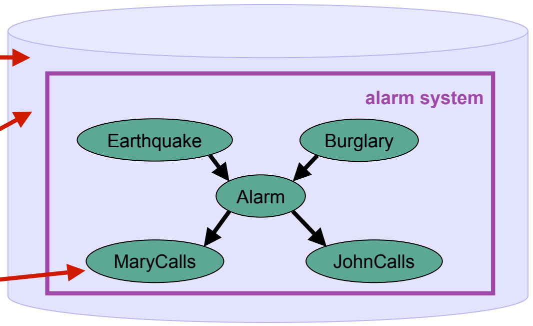


E	B	P(A | B,E)	
e	b	0.9	0.1
e	$\bar{b}$	0.2	0.8
$\bar{e}$	b	0.9	0.1
$\bar{e}$	$\bar{b}$	0.01	0.99

Database

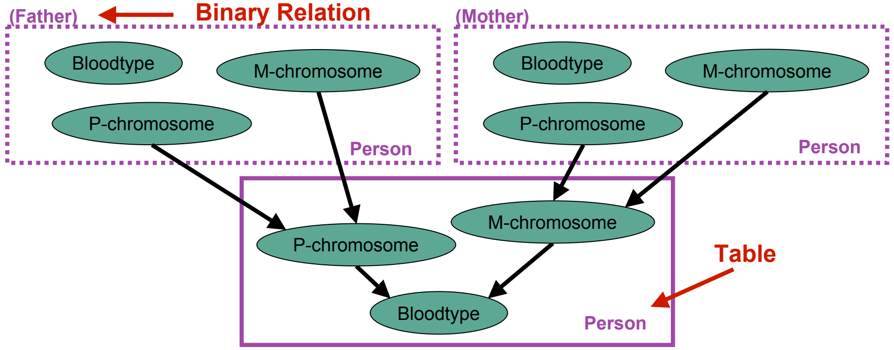
Table

Attribute



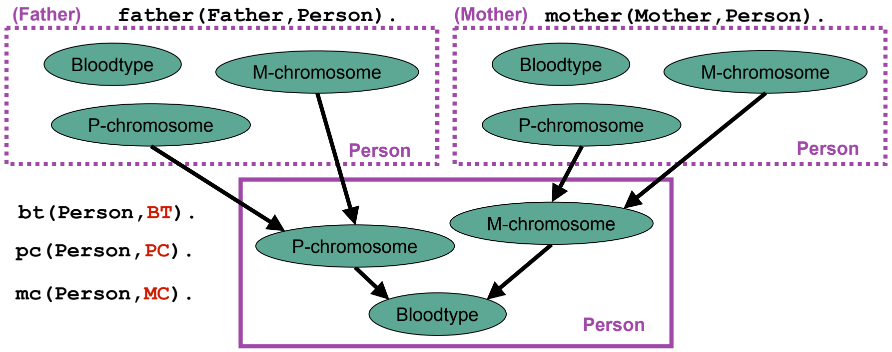
Probabilistic Relational Models (PRMs)

[Koller,Pfeffer,Getoor]



Probabilistic Relational Models (PRMs)

[Koller,Pfeffer,Getoor]



Dependencies (CPDs associated with):

```
bt(Person,BT) :- pc(Person,PC), mc(Person,MC).
pc(Person,PC) :- pc_father(Father,PCf), mc_father(Father,MCf).
```

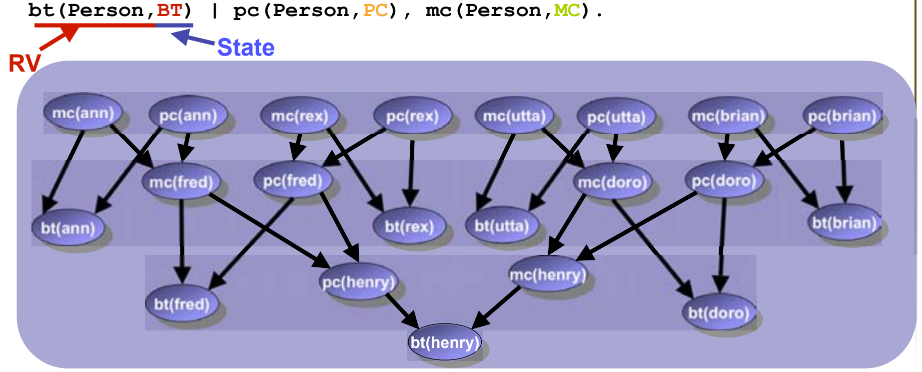
View :

```
pc_father(Person,PCf) | father(Father,Person),pc(Father,PC).
...
```


Probabilistic Relational Models (PRMs)

[Koller,Pfeffer,Getoor]

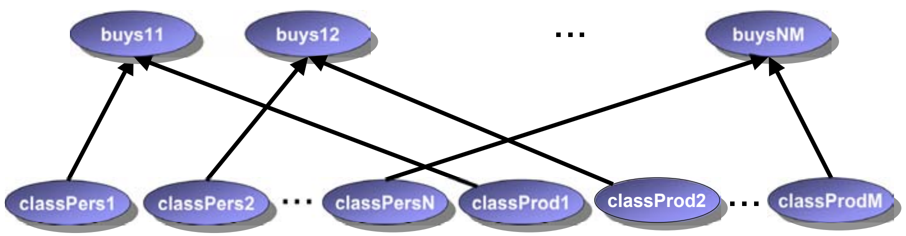
```
father(rex,fred).      mother(ann,fred).
father(brian,doro).    mother(utta, doro).
father(fred,henry).    mother(doro,henry).

pc_father(Person,PCf) | father(Father,Person),pc(Father,PC).
...
mc(Person,MC) | pc_mother(Person,PCm), pc_mother(Person,MCm).
pc(Person,PC) | pc_father(Person,PCf), mc_father(Person,MCf).
bt(Person,BT) | pc(Person,PC), mc(Person,MC).
```



PRM Application: Collaborative Filtering

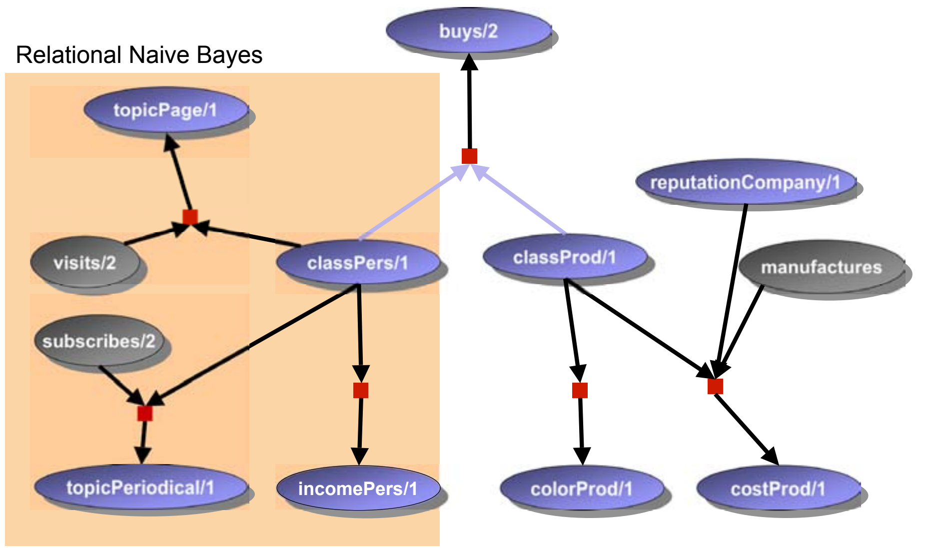
[Getoor, Sahami]

- User preference relationships for products / information.
- Traditionally: single dyactic relationship between the objects.



PRM Application: Collaborative Filtering

[Getoor, Sahami; simplified representation]



Probabilistic Relational Models (PRMs)

[Koller,Pfeffer,Getoor]

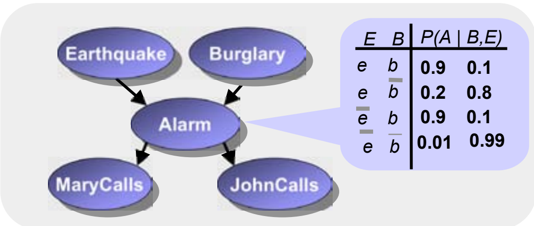
- Database View
- Unique Probability Distribution over finite Herbrand interpretations
  - No self-dependency
- Discrete and continuous RV
- BN used to do inference
- Graphical Representation

Outline Relational Models

- Relational Models
  - Probabilistic Relational Models
  - Baysian Logic Programs
  - Relational Markov networks
  - Markov Logic

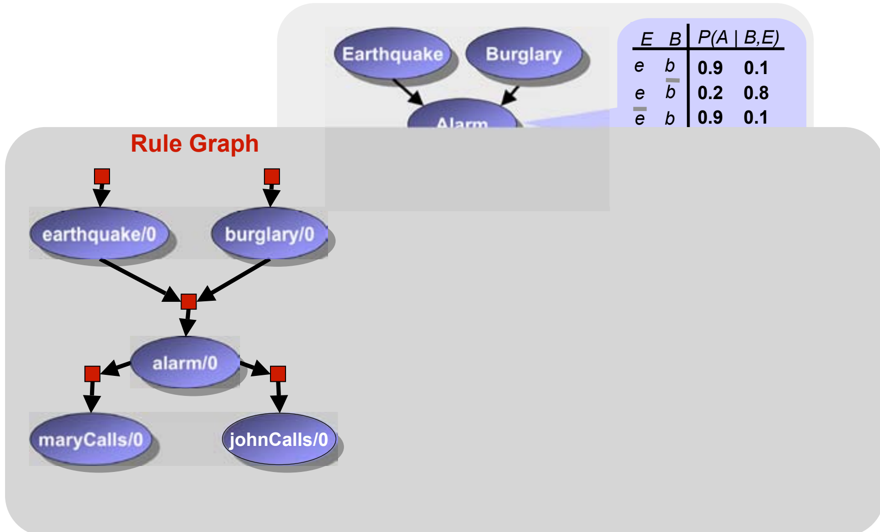
Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]



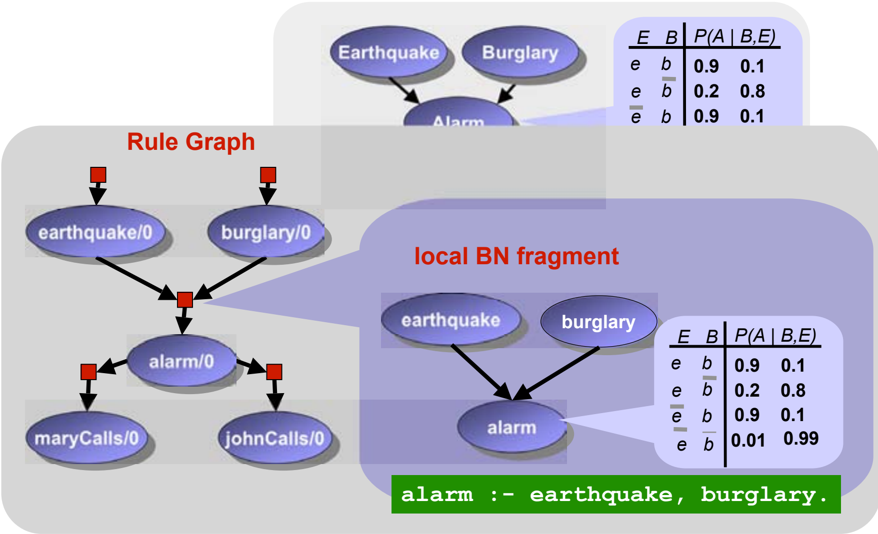
Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]



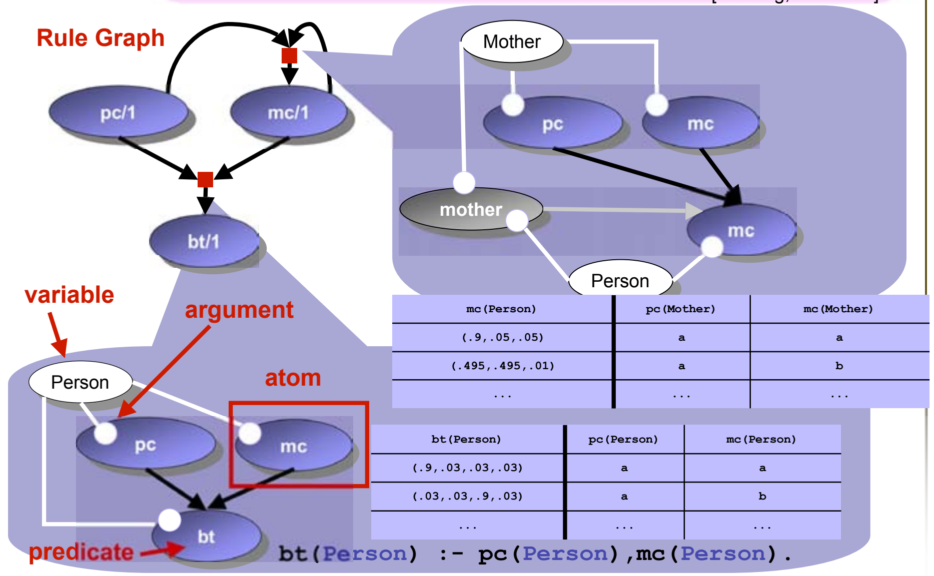
Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]



Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]

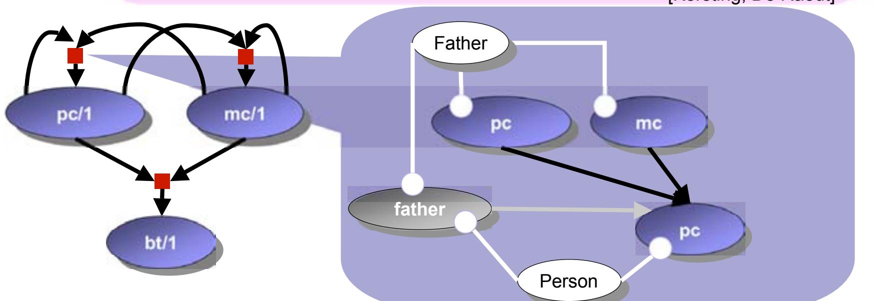


mc(Person)	pc(Mother)	mc(Mother)
(.9,.05,.05)	a	a
(.495,.495,.01)	a	b
...	...	...

bt(Person)	pc(Person)	mc(Person)
(.9,.03,.03,.03)	a	a
(.03,.03,.9,.03)	a	b
...	...	...

Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]



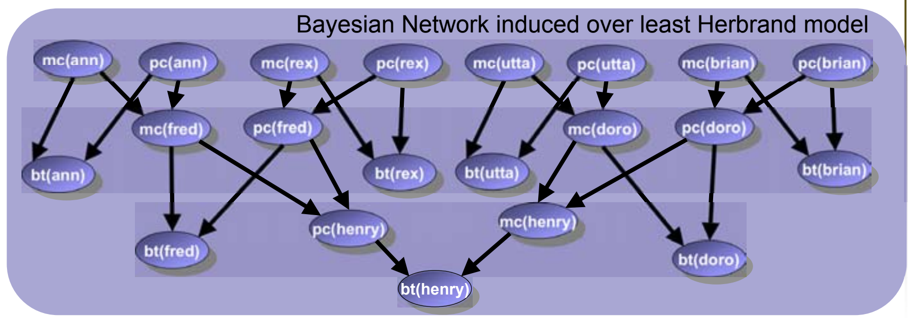
pc(Person)	pc(Father)	mc(Father)
(.9,.05,.05)	a	a
(.495,.495,.01)	a	b
...	...	...

```
mc(Person) | mother(Mother,Person), pc(Mother),mc(Mother).
pc(Person) | father(Father,Person), pc(Father),mc(Father).
bt(Person) | pc(Person),mc(Person).
```


Bayesian Logic Programs (BLPs)

[Kersting, De Raedt]

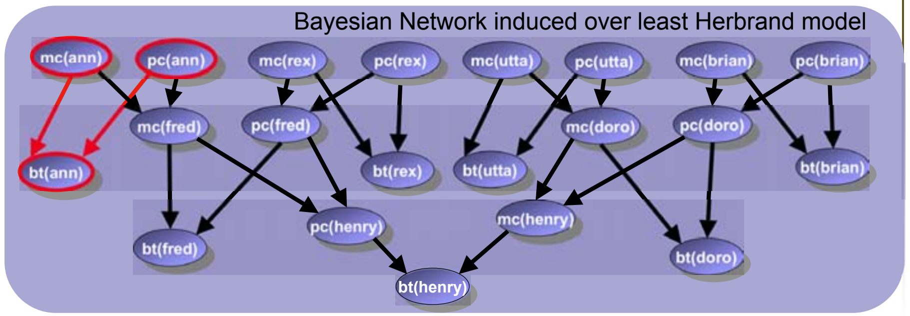
```
father(rex,fred).      mother(ann,fred).
father(brian,doro).    mother(utta, doro).
father(fred,henry).    mother(doro,henry).

mc(Person) | mother(Mother,Person), pc(Mother),mc(Mother).
pc(Person) | father(Father,Person), pc(Father),mc(Father).
bt(Person) | pc(Person),mc(Person).
```



Answering Queries

P(`bt(ann)`) **?**

Bayesian Network induced over least Herbrand model



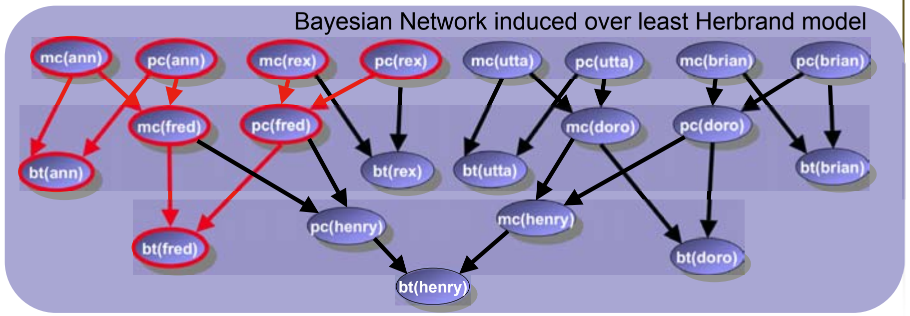
Answering Queries

Bayes' rule

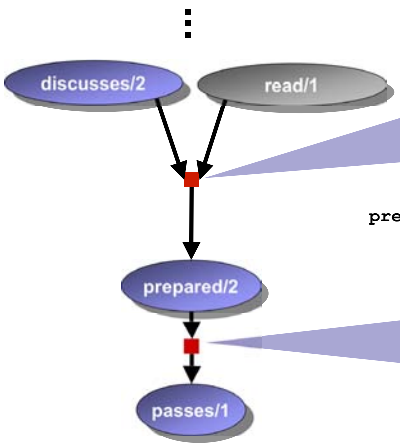
$$\mathbf{P}(\texttt{bt(ann)} \mid \texttt{bt(fred)}) = \frac{\mathbf{P}(\texttt{bt(ann)},\texttt{bt(fred)})}{\mathbf{P}(\texttt{bt(fred)})}$$

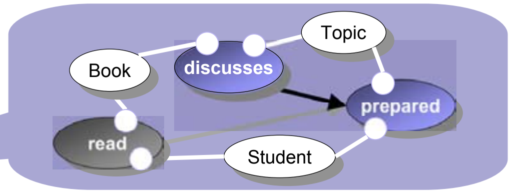
P(`bt(ann), bt(fred)`) **?**

Bayesian Network induced over least Herbrand model

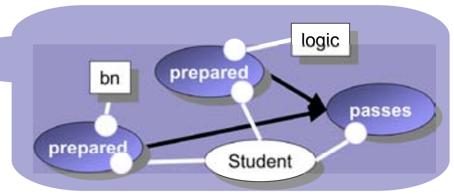


Combining Partial Knowledge



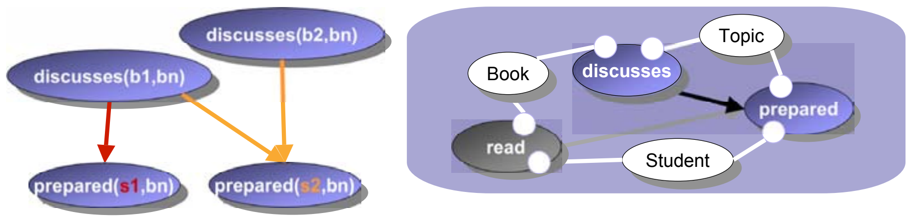



```
prepared(Student,Topic) | read(Student,Book),
                          discusses(Book,Topic).
```



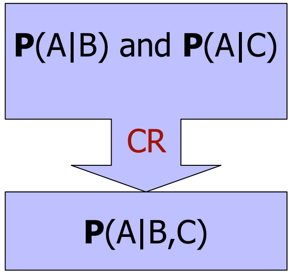

```
passes(Student) | prepared(Student,bn),
                  prepared(Student,logic).
```

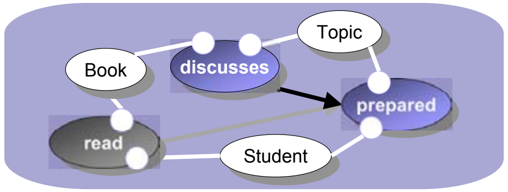

Combining Partial Knowledge



- variable # of parents for `prepared/2` due to `read/2`
  - whether a student prepared a topic depends on the books she read
- CPD only for one book-topic pair

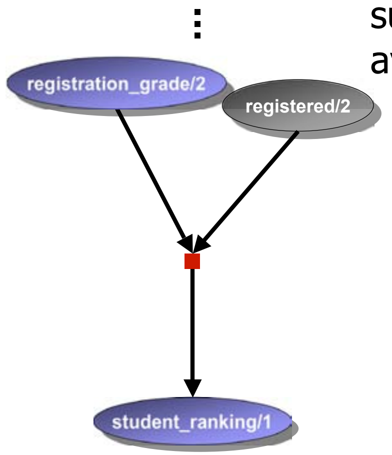
Combining Rules





- Any algorithm which
  - has an empty output if and only if the input is empty
  - combines a set of CPDs into a single (combined) CPD
- E.g. noisy-or, regression, ...

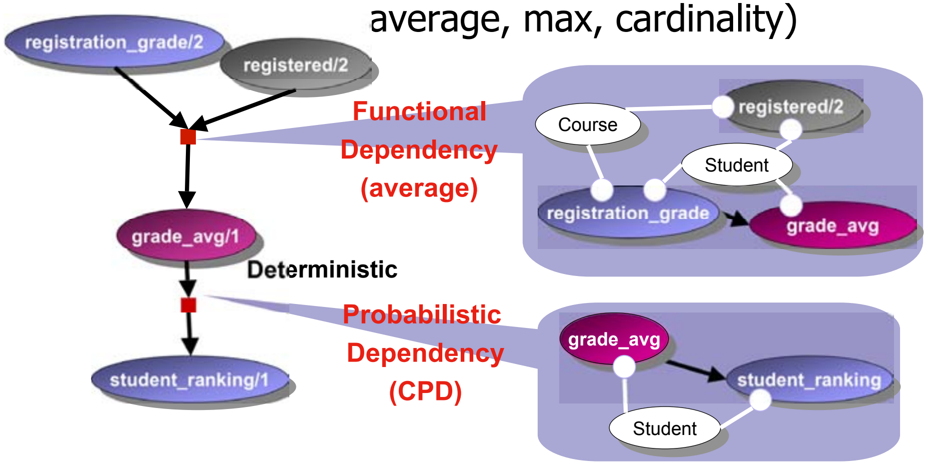
Aggregates



Map multisets of values to summary values (e.g., sum, average, max, cardinality)

Aggregates

Map multisets of values to summary values (e.g., sum, average, max, cardinality)



Experiments

KDD Cup 2001 localization task

predict the localization based on local features and interactions

862 training genes
381 test genes

>1000 interactions

16 classes

WebKB

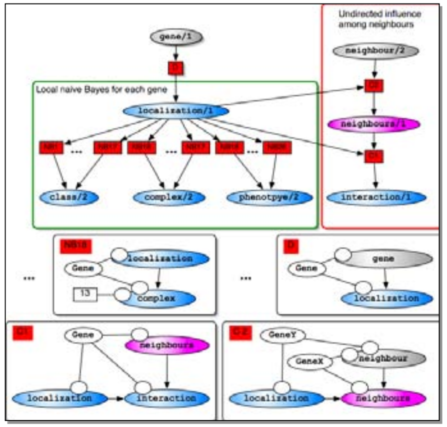
predict the type of web pages

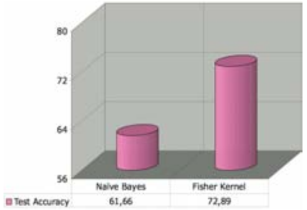
877 web pages from 4 CS department

1516 links

6 classes

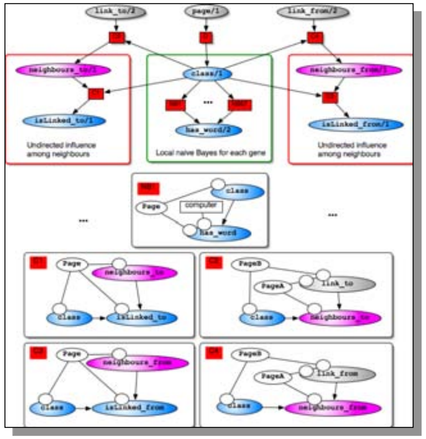
KDD Cup: Protein Localization





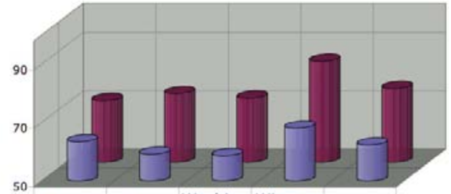
RFK (**72.89%**) better then Hayashi et al.'s KDD Cup 2001 winning nearest- neighbour approach (**72.18%**)

WebKB: Web Page Classification



Collective NB ~ PRMs [Getoor et al. 02]
RFK outperforms PRMs

PRM with structural uncertainty over the links , best acc. (68%) on Washington



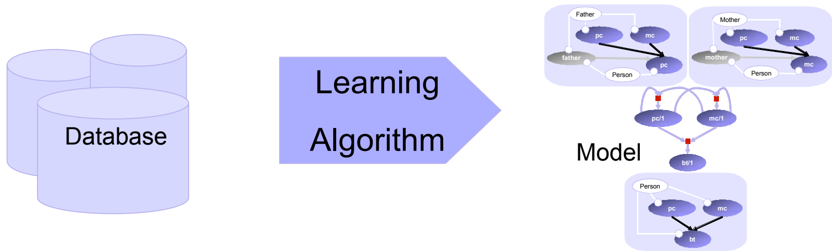
	Cornell	Texas	Washington	Wisconsin	Mean
Naïve Bayes	63,44	59,2	58,65	68,07	62,34
Fisher Kernels	71,08	73,52	71,93	84,59	75,28

Leave-one-university-out cross-validation

Bayesian Logic Programs (BLPs)

- Unique probability distribution over Herbrand interpretations
  - Finite branching factor, finite proofs, no self-dependency
- Highlight
  - Separation of qualitative and quantitative parts
  - Functors
- Graphical Representation
- Discrete and continuous RV

Learning Tasks



- Parameter Estimation
  - Numerical Optimization Problem
- Model Selection
  - Combinatorical Search

What is the data about?

RVs + States = (partial) Herbrand interpretation

Probabilistic learning from interpretations

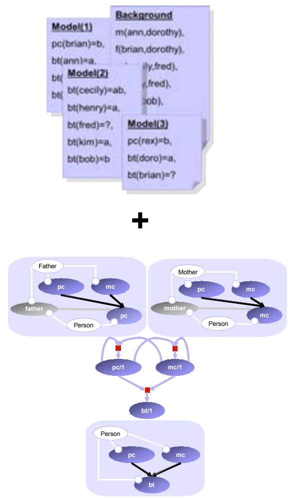
Family(1)
pc(brian)=b,
bt(ann)=a,
bt(brian)=?,
bt(dorothy)=a

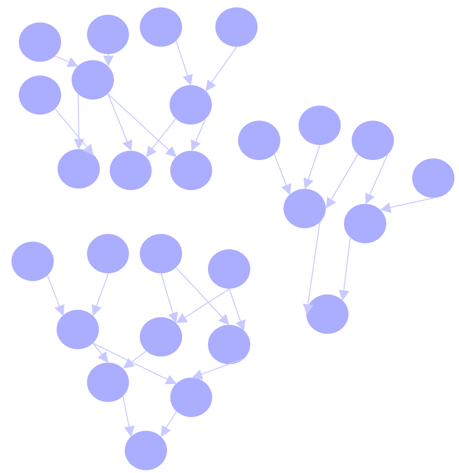
Background
m(ann,dorothy),
f(brian,dorothy),
m(cecily,fred),
f(henry,fred),
f(fred,bob),
m(kim,bob),
...

Family(3)
pc(rex)=b,
bt(doro)=a,
bt(brian)=?

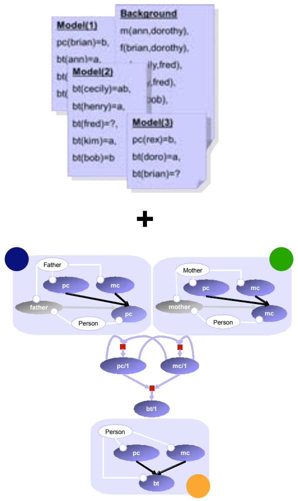
Family(2)
bt(cecily)=ab,
pc(henry)=a,
mc(fred)=?,
bt(kim)=a,
pc(bob)=b

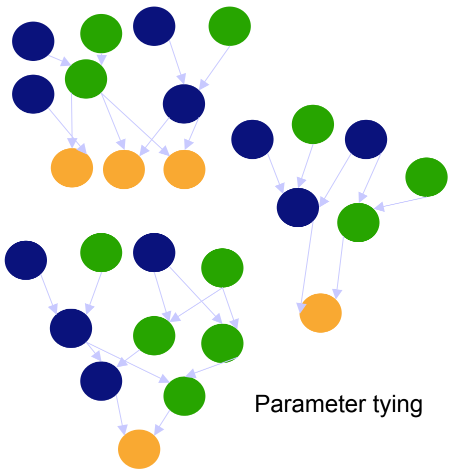
Parameter Estimation





Parameter Estimation



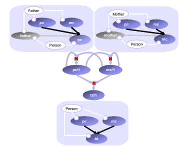


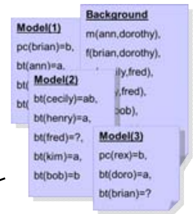
Parameter tying

Expectation Maximization

EM-algorithm:
iterate until convergence

Logic Program L





Initial Parameters θ0

Current Model
(M,θk)

Expectation

Inference

Expected counts of a clause

$$\sum_{\substack{\text{Ground Instance}\\ \text{GI}}} \sum_{\substack{\text{DataCase}\\ \text{DC}}} \mathbf{P(\,head(GI),\ body(GI)\mid DC\,)}$$

$$\frac{\displaystyle\sum_{\substack{\text{Ground Instance}\\ \text{GI}}} \sum_{\substack{\text{DataCase}\\ \text{DC}}} \mathbf{P(\,head(GI),\ body(GI)\mid DC\,)}}{\displaystyle\sum_{\substack{\text{Ground Instance}\\ \text{GI}}} \sum_{\substack{\text{DataCase}\\ \text{DC}}} \mathbf{P(\,body(GI)\mid DC\,)}}$$

Maximization
Update parameters
(ML, MAP)

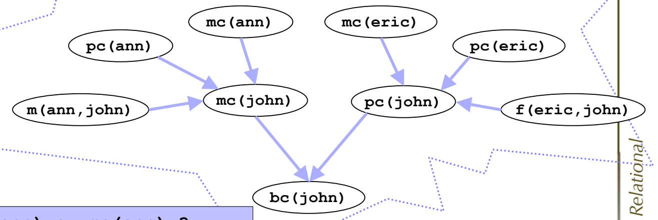
Model Selection

- Combination of ILP and BN learning
- Modify the general rules syntactically:
  - Add atoms: b(X,a)
  - Delete atoms
  - Unify placeholders: m(X,Y) -> m(X,X)
  - ...
- **Add, (reverse, and) delete bunches of edges simultaniously**

Example

Original program

```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```



Data cases

```
{m(ann,john)=true, pc(ann)=a, mc(ann)=?,
 f(eric,john)=true, pc(eric)=b, mc(eric)=a,
 mc(john)=ab, pc(john)=a, bt(john) = ? }

...
```

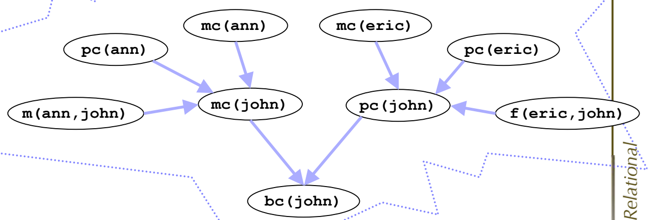

Example

Original program

```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```



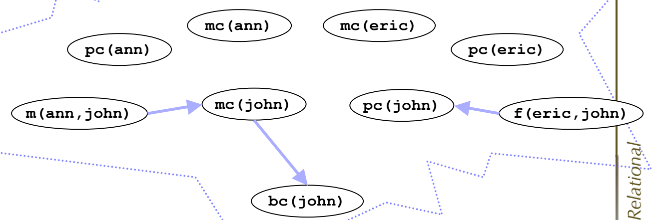
Example

Original program

```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```



Example

Original program

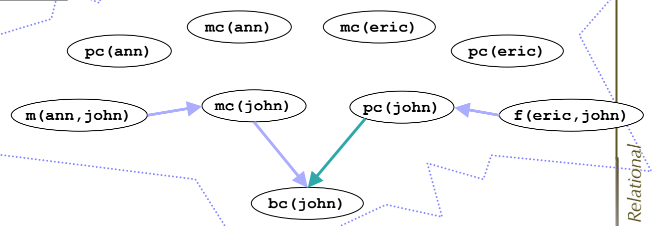
```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```

Refinement

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```



Advanced I WS 06/07

Example

Original program

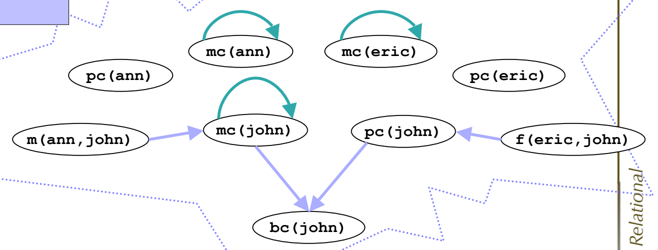
```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```

Refinement

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```



Refinement

```
mc(X) | m(M,X),mc(X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```


Advanced I WS 06/07

Example

Original program

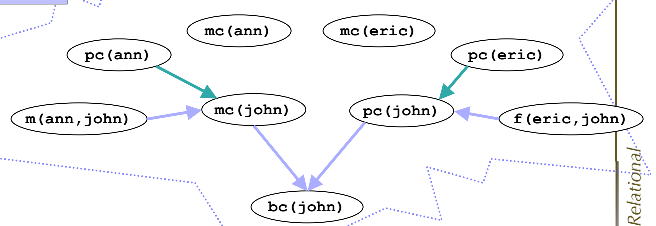
```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```

Refinement

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```



Refinement

```
mc(X) | m(M,X),pc(X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```


Advanced I WS 06/07

Example

Original program

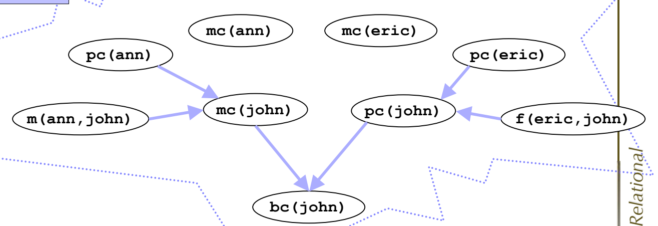
```
mc(X) | m(M,X), mc(M), pc(M).
pc(X) | f(F,X), mc(F), pc(F).
bt(X) | mc(X), pc(X).
```

Initial hypothesis

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X).
```

Refinement

```
mc(X) | m(M,X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```



Refinement

```
mc(X) | m(M,X),pc(X).
pc(X) | f(F,X).
bt(X) | mc(X), pc(X).
```

...

Outline Relational Models

- Relational Models
  - Probabilistic Relational Models
  - Baysian Logic Programs
  - Relational Markov networks
  - Markov Logic

Undirected Relational Models

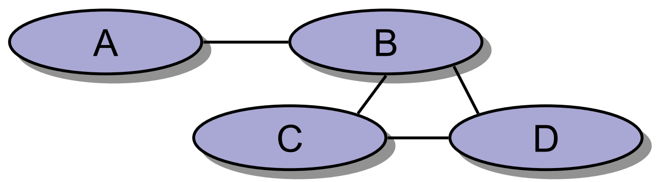
- So far, directed graphical models only
- Impose acyclicity constraint

- Undirected graphical models do not impose the acyclicity constraint

Undirected Relational Models

- Two approaches
  - Relational Markov Networks (RMNs)
    - (Taskar et al.)
  - Markov Logic Networks (MLNs)
    - (Anderson et al.)
- Idea
  - Semantics = Markov Networks
  - More natural for certain applications
- RMNs ~ undirected PRM
- MLNs ~ undirected BLP

Markov Networks



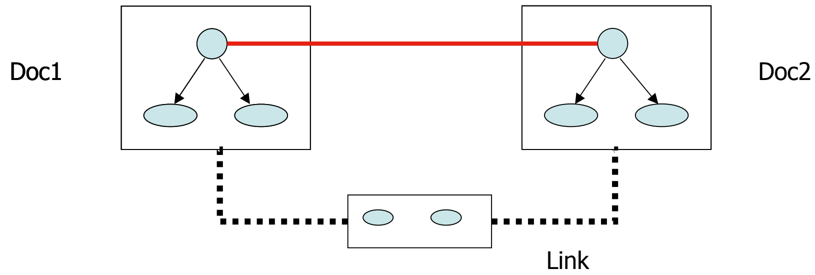
- To each clique c, a potential ϕ_c is associated
- Given the values $\mathbf{v}$ of all nodes in the Markov Network

$$P(\mathbf{v}) = \frac{1}{Z}\prod_{c \in C(G)} \phi_c(\mathbf{v}_c) \qquad Z = \sum_{\mathbf{v}} \prod_{c \in C(G)} \phi_c(\mathbf{v}'_c)$$

$$\log P(\mathbf{v}) = \sum_c \mathbf{w}_c \cdot \mathbf{f}_c(v_c) - \log Z = \mathbf{w} \cdot \mathbf{f}(v) - \log Z$$

Relational Markov Networks

```
SELECT doc1.Category,doc2.Category
FROM doc1,doc2,Link link
WHERE link.From=doc1.key and link.To=doc2.key
```



Markov Logic Networks

1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \big(Smokes(x) \Leftrightarrow Smokes(y)\big)$

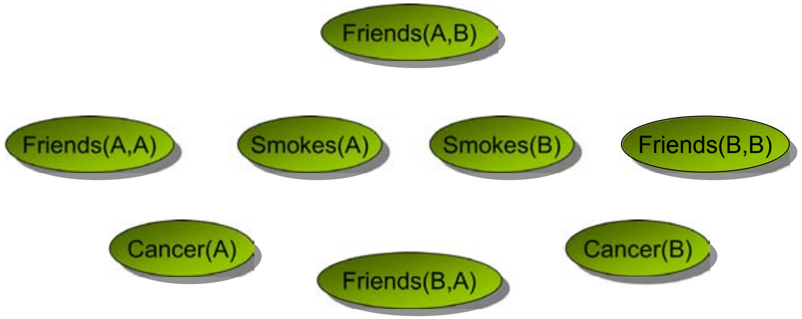
Suppose we have two constants: **Anna** (A) and **Bob** (B)

slides by Pedro Domingos

Markov Logic Networks

1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \big(Smokes(x) \Leftrightarrow Smokes(y)\big)$

Suppose we have two constants: **Anna** (A) and **Bob** (B)

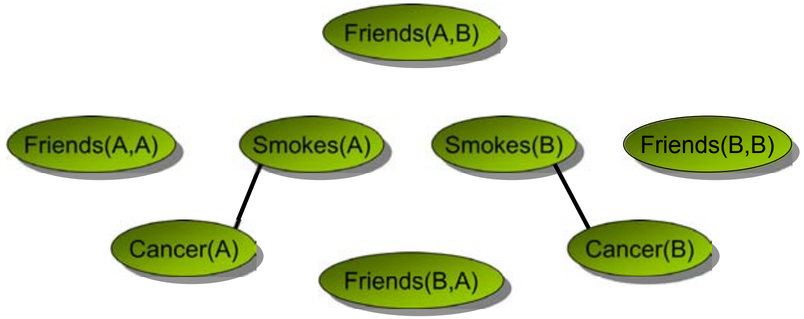


slides by Pedro Domingos

Markov Logic Networks

1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \big(Smokes(x) \Leftrightarrow Smokes(y)\big)$

Suppose we have two constants: **Anna** (A) and **Bob** (B)

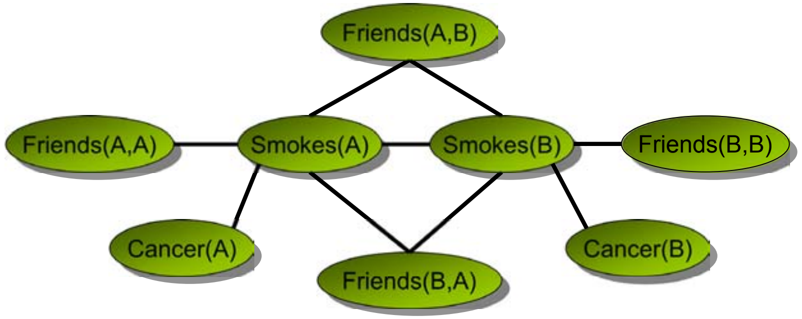


slides by Pedro Domingos

Markov Logic Networks

1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \big(Smokes(x) \Leftrightarrow Smokes(y)\big)$

Suppose we have two constants: **Anna** (A) and **Bob** (B)



slides by Pedro Domingos

Learning Undirected PRMs

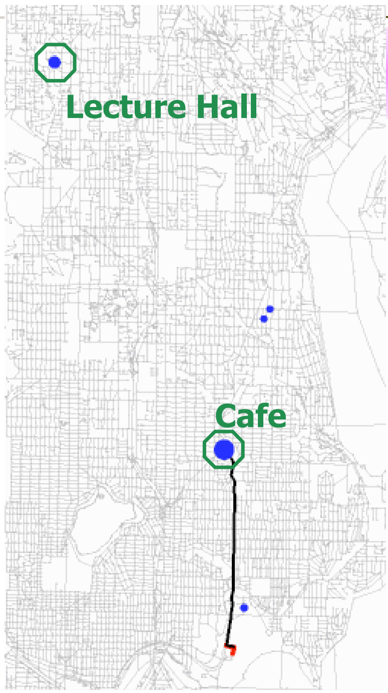
- Parameter estimation
  - discriminative (gradient, max-margin)
  - generative setting using pseudo-likelihood
- Structure learning
  - Similar to PRMs, BLPs

Applications

- Computer Vision
  - (Taskar et al.)
- Citation Analysis
  - (Taskar et al., Singla&Domingos)
- Activity Recognition
  - (Liao et al.)

Activity Recognition

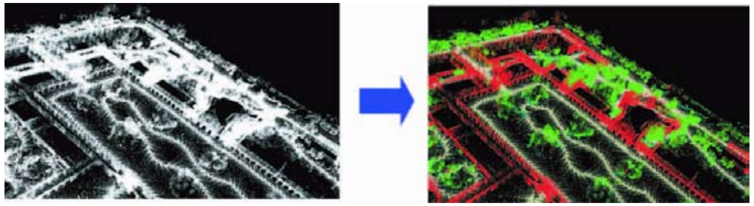
[Fox et al. IJCAI03]

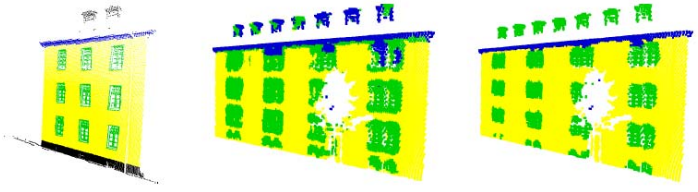


Will you go to the AdvancedAI lecture or will you visit some friends in a cafe?

3D Scan Data Segmentation
[Anguelov et al. CVPR05, Triebel et al. ICRA06]

How do you recognize the lecture hall?





(a) Training data (b) Bayes classifier (c) AMN + adaptive reduction

Outline Relational Models

- Relational Models
  - Probabilistic Relational Models
  - Baysian Logic Programs
  - Relational Markov networks
  - Markov Logic

Conclusions

- **SRL = Probability + Logic + Learning**
- Covers full AI spectrum: Logic, probability, learning, kernels, sequences, planning, reinforcement learning, …
- Considered to be a revolution in ML
- Logical variables/Placeholders: group random variables/states
- Unification: context-specific prob. information

Thanks

… for your attention

… and enjoy the other parts of the lecture !