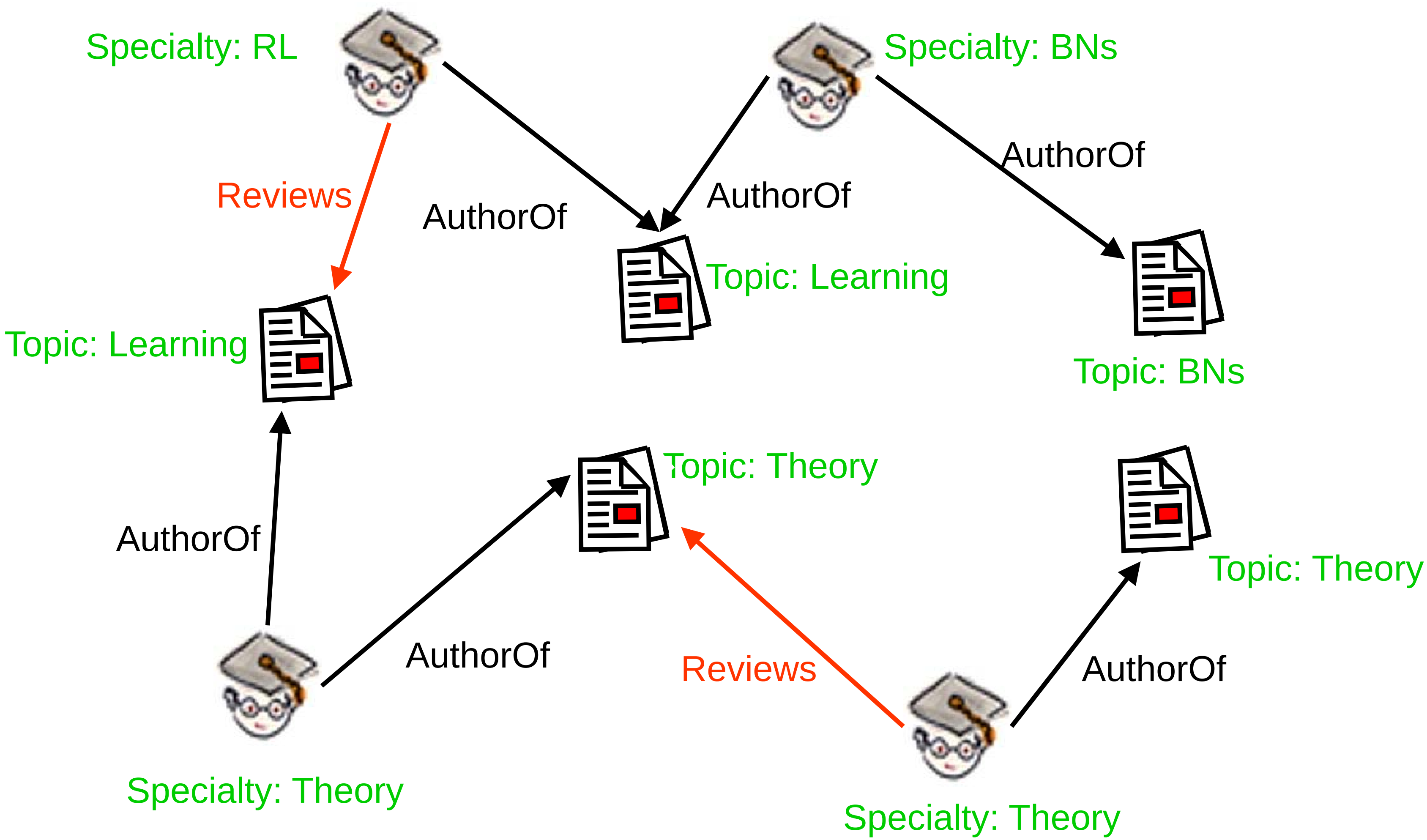
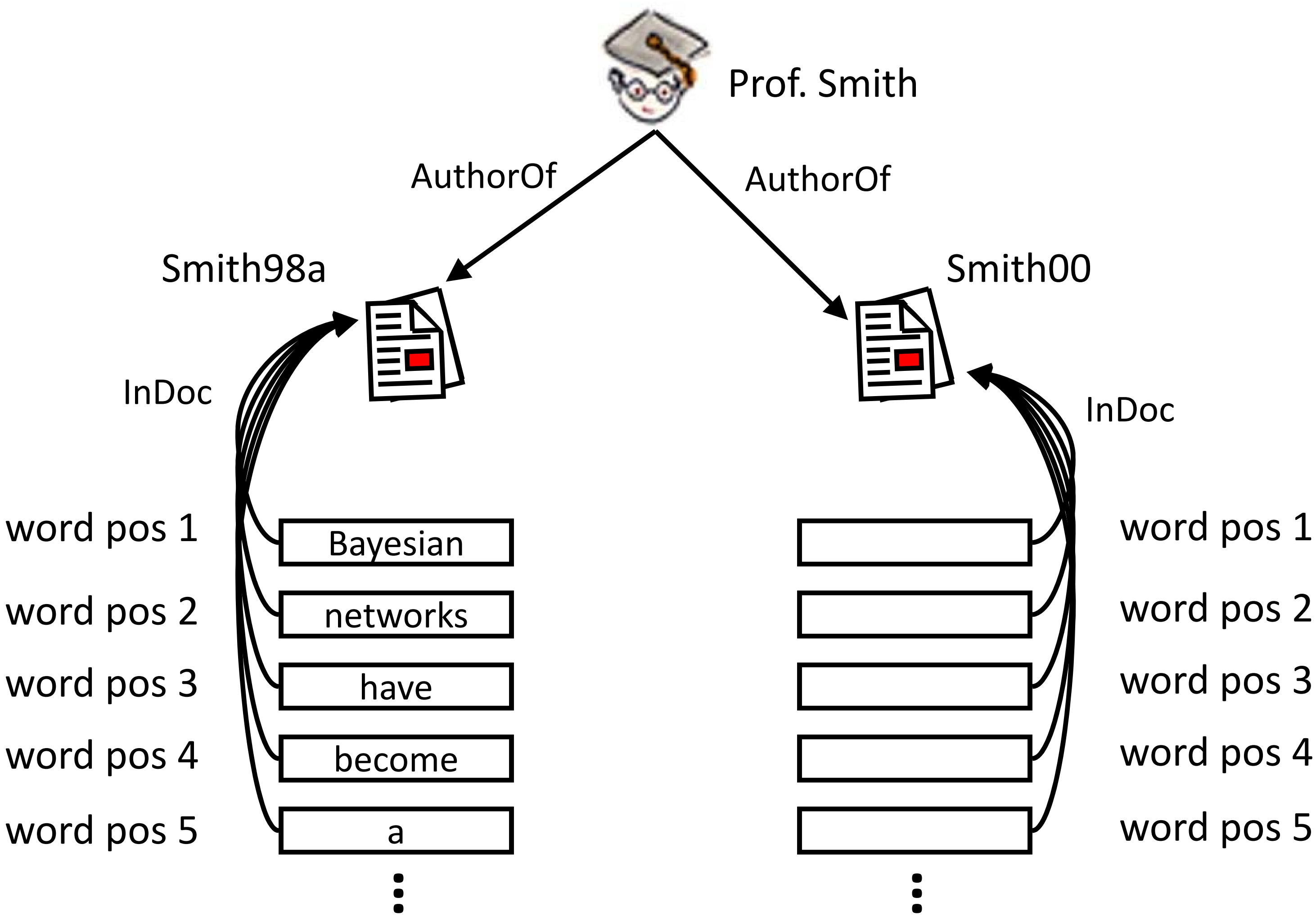
Relational Probability Models

Brian Milch
MIT 9.66
November 27, 2007

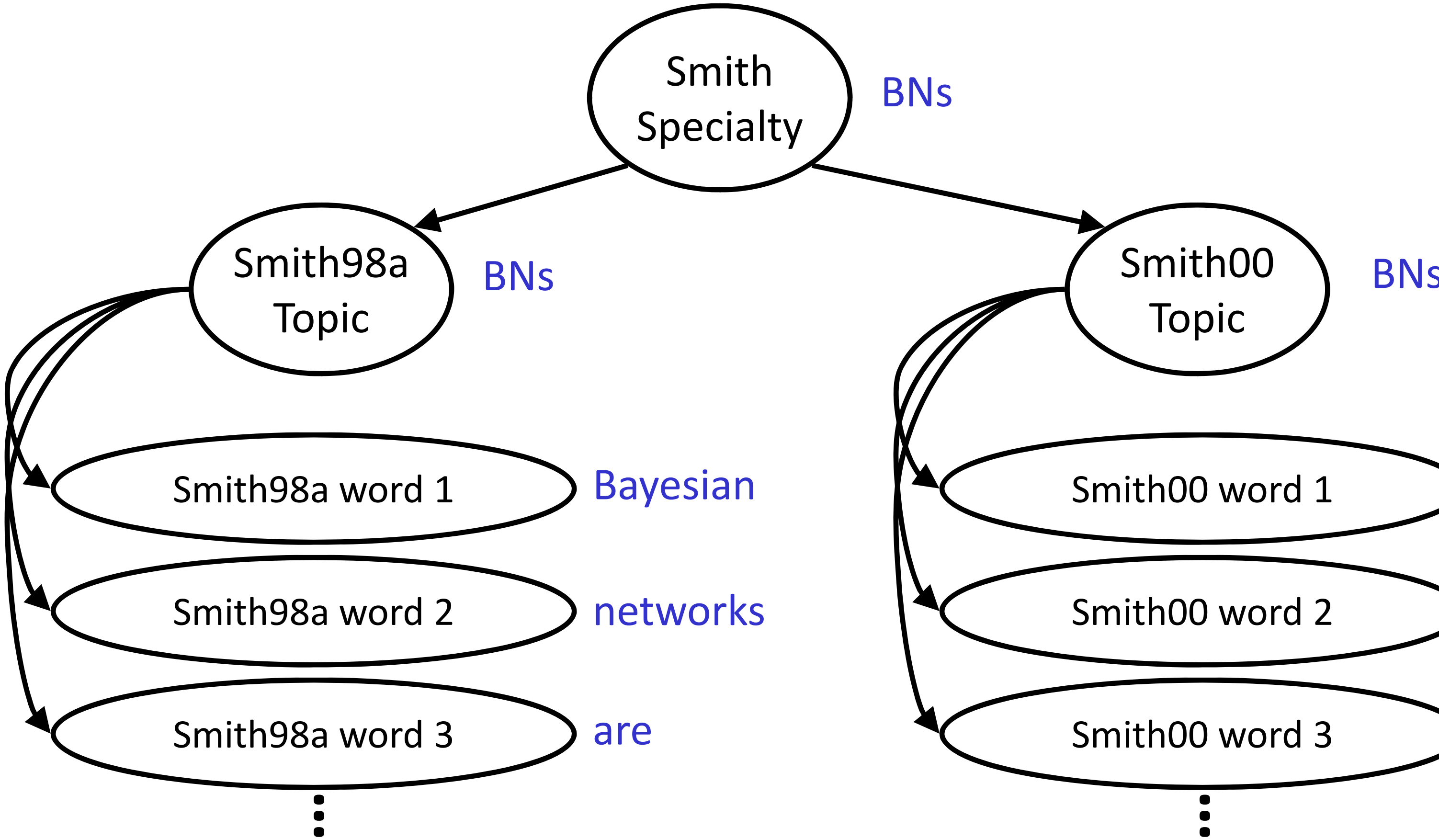
Objects, Attributes, Relations



Specific Scenario



Graphical Model for This Scenario



- Dependency models are repeated at each node
- Graphical model is specific to Smith98a and Smith00

Abstract Knowledge

- Humans have abstract knowledge that can be applied to any individuals
  - Within a scenario
  - Across scenarios
- How can such knowledge be:
  - Represented?
  - Learned?
  - Used in reasoning?

Outline

- Logic: first-order versus propositional
- Relational probability models (RPMs): first-order logic meets probability
- Relational uncertainty in RPMs
- Thursday: models with unknown objects

Kinds of possible worlds

- **Atomic**: each possible world is an atom or token with no internal structure. E.g., Heads or Tails
- **Propositional**: each possible world defined by values assigned to variables. E.g., propositional logic, graphical models
- **First-order**: each possible world defined by objects and relations

[Slide credit: Stuart Russell]

Specialties and Topics

Propositional

Spec_Smith_BNs $\rightarrow$ Topic_Smith98a_BNs
Spec_Smith_Theory $\rightarrow$ Topic_Smith98a_Theory
Spec_Smith_Learning $\rightarrow$ Topic_Smith98a_Learning

Spec_Smith_BNs $\rightarrow$ Topic_Smith00_BNs
Spec_Smith_Theory $\rightarrow$ Topic_Smith00_Theory
Spec_Smith_Learning $\rightarrow$ Topic_Smith00_Learning

First-Order

$$\forall r \, \forall t \, \forall p \; [(\mathrm{Spec}(r, t) \wedge \mathrm{AuthorOf}(r, p)) \rightarrow \mathrm{Topic}(p, t)]$$

AuthorOf(Smith, Smith00)
AuthorOf(Smith, Smith98a)

Expressiveness matters

- Expressive language => concise models
  => fast learning, sometimes fast reasoning

E.g., rules of chess:
  1 page in first-order logic,
  ~100000 pages in propositional logic,
  ~100000000000000000000000000000000000000 pages as atomic-state model
  (Note: chess is a teeny problem)

[Slide credit: Stuart Russell]

Brief history of expressiveness

	atomic	*propositional*	*first-order*
probabilistic	histogram 17th–18th centuries	Probabilistic logic [Nilsson 1986], Graphical models late 20th century	First-order probabilistic languages (FOPLs) 20th-21st centuries
deterministic		Boolean logic 5th century B.C. - 19th century	First-order logic 19th - early 20th century

First-Order Logic Syntax

- Constants: Brian, 2, AIMA2e, MIT,...
- Predicates: AuthorOf, >,...
- Functions: PublicationYear,$\sqrt{}$,...
- Variables: x,y,a,b,...
- Connectives: $\wedge \vee \neg \rightarrow \leftrightarrow$
- Equality: =
- Quantifiers: $\forall \exists$

[Slide credit: Stuart Russell]

Terms

- A term refers (according to a given possible world) to an object in that world
- Term =
  – function($term_1$,...,$term_n$) or
  – constant symbol or
  – variable
- E.g., PublicationYear(AIMA2e)
- Arbitrary nesting $\Rightarrow$ infinitely many terms

[Slide credit: Stuart Russell]

Atomic sentences

- Atomic sentence =
  - $\text{predicate}(\text{term}_1,\ldots,\text{term}_n)$ or
  - term1=term2
- E.g.,
  - AuthorOf(Norvig,AIMA2e)
  - NthAuthor(AIMA2e,2) = Norvig
- Can be combined using connectives, e.g.,
  (Peter=Norvig) $\Rightarrow$(NthAuthor(AIMA2e,2) = Peter)

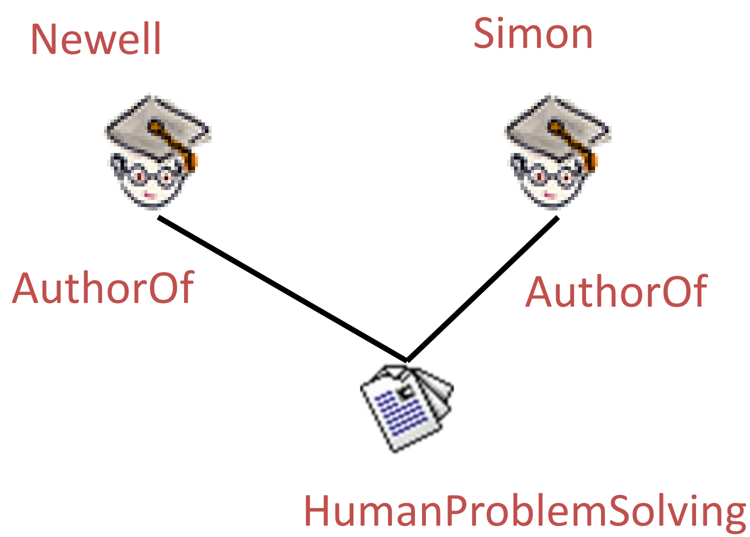
[Slide credit: Stuart Russell]

Semantics: Truth in a world

- Each possible world contains ≥1 objects (domain elements), and maps…
  - Constant symbols → objects
  - Predicate symbols → relations (sets of tuples of objects satisfying the predicate)
  - Function symbols → functional relations
- An atomic sentence predicate($term_1$,...,$term_n$) is true iff the objects referred to by $term_1$,...,$term_n$ are in the relation referred to by predicate

[Slide credit: Stuart Russell]

Example



AuthorOf(Newell,HumanProblemSolving) is true in this world

[Slide credit: Stuart Russell]

Outline

- Logic: first-order versus propositional
- Relational probability models (RPMs): first-order logic meets probability
- Relational uncertainty in RPMs
- Thursday: models with unknown objects

Relational Probability Models

Abstract probabilistic model for attributes

+

Relational skeleton: objects & relations

⇩

Graphical model

Representation

- Have to represent
  - Set of variables
  - Dependencies
  - Conditional probability distributions (CPDs)
- Many proposed languages
- We'll use Bayesian logic (BLOG)
  [Milch *et al.* 2005]

All depend on relational skeleton

Typed First-Order Logic

- Objects divided into types

  Boolean, Researcher, Paper, WordPos, Word, Topic

- Express attributes and relations with functions (predicates are just Boolean functions)

  FirstAuthor(paper) $\rightarrow$ Researcher (non-random)
  Specialty(researcher) $\rightarrow$ Topic (random)
  Topic(paper) $\rightarrow$ Topic (random)
  Doc(wordpos) $\rightarrow$ Paper (non-random)
  WordAt(wordpos) $\rightarrow$ Word (random)

Set of Random Variables

- For random functions, have random variable for each tuple of argument objects

  Researcher: Smith, Jones   Paper: Smith98a, Smith00, Jones00
  WordPos: Smith98a_1, …, Smith98a_3212, Smith00_1, etc.

Specialty: Specialty(Smith) Specialty(Jones)

Topic: Topic(Smith98a) Topic(Smith00) Topic(Jones00)

WordAt: WordAt(Smith98a_1) … WordAt(Smith98a_3212)

WordAt(Smith00_1) … WordAt(Smith00_2774)

WordAt(Jones00_1) … WordAt(Jones00_4893)

Dependency Statements

```
                              BNs    RL    Theory
Specialty(r) ~ TabularCPD[[0.5, 0.3, 0.2]];
```

```
                           BNs     RL     Theory
Topic(p) ~ TabularCPD[[0.90, 0.01, 0.09],      | BNs
                      [0.02, 0.85, 0.13],      | RL
                      [0.10, 0.10, 0.80]]      | Theory
               (Specialty(FirstAuthor(p)));
```

Logical term identifying parent node

```
                            the           Bayesian  reinforcement
WordAt(wp) ~ TabularCPD[[0.03,..., 0.02, 0.001,...],     | BNs
                        [0.03,..., 0.001, 0.02,...],     | RL
                        [0.03,..., 0.003, 0.003,...]]    | Theory
               (Topic(Doc(wp)));
```

Variable Numbers of Parents

- What if we allow multiple authors?
  - Let skeleton specify predicate AuthorOf(***r***, ***p***)
- Topic(***p***) now depends on specialties of multiple authors

$\boxed{\text{Number of parents depends on skeleton}}$

Aggregation

- Aggregate distributions

multiset defined by formula

```
Topic(p) ~ TopicAggCPD({Specialty(r) for Researcher r :
                                          AuthorOf(r, p)});
```

mixture of distributions conditioned on individual elements of multiset [Taskar *et al*., IJCAI 2001]

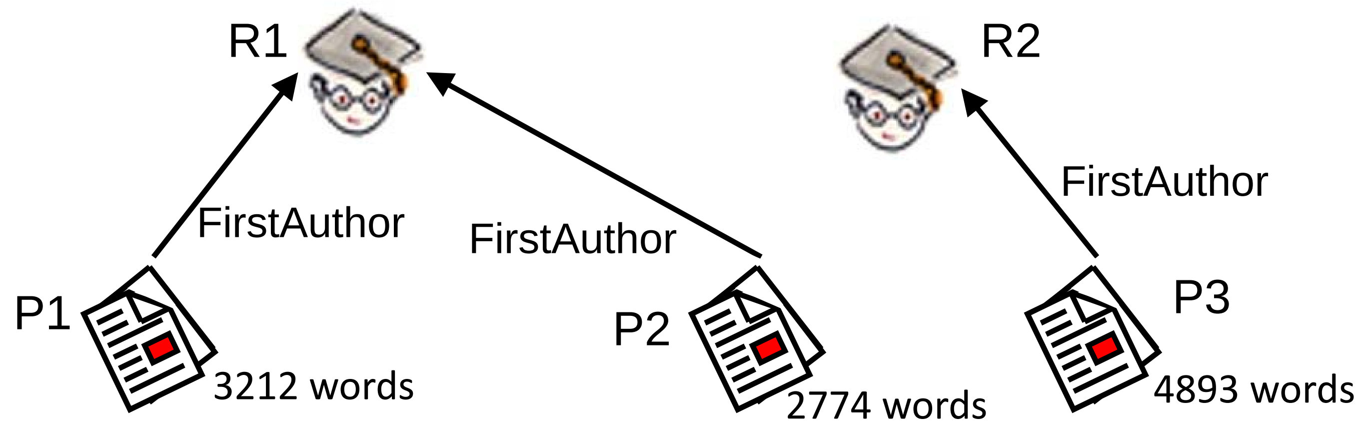
- Aggregate values

aggregation function

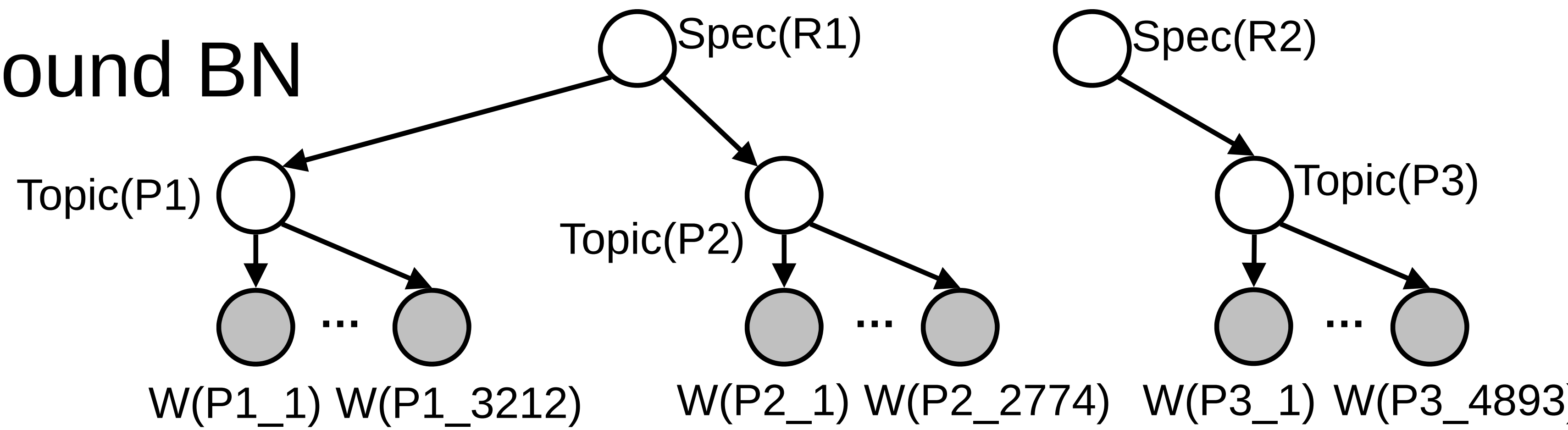
```
Topic(p) ~ TopicCPD(Mode({Specialty(r) for Researcher r :
                                          AuthorOf(r, p)}));
```

Semantics: Ground BN

Skeleton



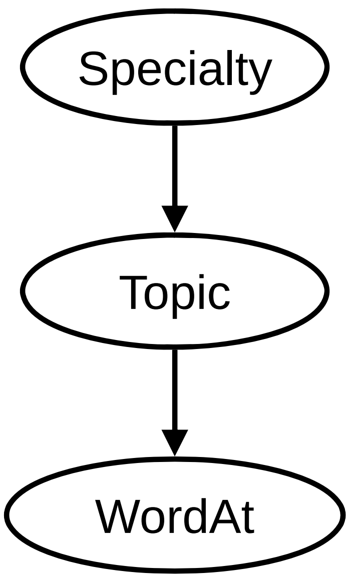
Ground BN



When Is Ground BN Acyclic?

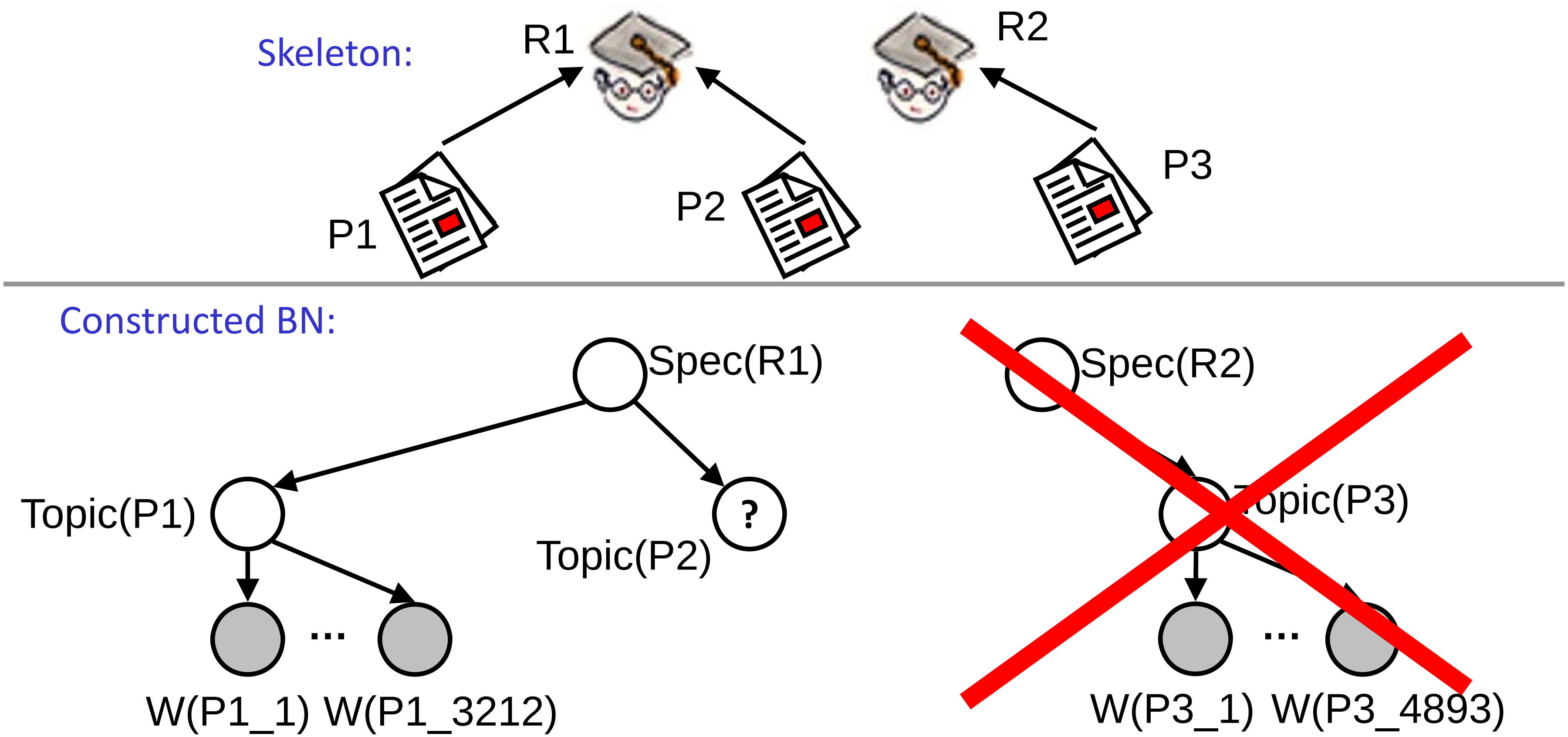
[Koller & Pfeffer, AAAI 1998]

- Look at symbol graph
  - Node for each random function
  - Read off edges from dependency statements
- *Theorem*: If symbol graph is acyclic, then ground BN is acyclic for every skeleton



Inference: Knowledge-Based Model Construction (KBMC)

- Construct relevant portion of ground BN



[Breese 1992; Ngo & Haddawy 1997]

Inference on Constructed Network

- Run standard BN inference algorithm
  - Exact: variable elimination/junction tree
  - Approx: Gibbs sampling, loopy belief propagation
- Exploit some repeated structure with lifted inference [Pfeffer *et al*., UAI 1999; Poole, IJCAI 2003; de Salvo Braz *et al*., IJCAI 2005]

References

- Wellman, M. P., Breese, J. S., and Goldman, R. P. (1992) "From knowledge bases to decision models". *Knowledge Engineering Review* 7:35-53.
- Breese, J.S. (1992) "Construction of belief and decision networks". *Computational Intelligence* 8(4):624-647.
- Ngo, L. and Haddawy, P. (1997) "Answering queries from context-sensitive probabilistic knowledge bases". *Theoretical Computer Sci.* 171(1-2):147-177.
- Koller, D. and Pfeffer, A. (1998) "Probabilistic frame-based systems". In *Proc. 15th AAAI National Conf. on AI*, pages 580-587.
- Friedman, N., Getoor, L., Koller, D.,and Pfeffer, A. (1999) "Learning probabilistic relational models". In *Proc. 16th Int'l Joint Conf. on AI*, pages 1300-1307.
- Pfeffer, A., Koller, D., Milch, B., and Takusagawa, K. T. (1999) "SPOOK: A System for Probabilistic Object-Oriented Knowledge". In *Proc. 15th Conf. on Uncertainty in AI*, pages 541-550.
- Taskar, B., Segal, E., and Koller, D. (2001) "Probabilistic classification and clustering in relational data". In *Proc. 17th Int'l Joint Conf. on AI*, pages 870-878.
- Getoor, L., Friedman, N., Koller, D., and Taskar, B. (2002) "Learning probabilistic models of link structure". *J. Machine Learning Res.* 3:679-707.
- Taskar, B., Abbeel, P., and Koller, D. (2002) "Discriminative probabilistic models for relational data". In *Proc. 18th Conf. on Uncertainty in AI*, pages 485-492.

References

- Poole, D. (2003) "First-order probabilistic inference". In *Proc. 18th Int'l Joint Conf. on AI*, pages 985-991.
- de Salvo Braz, R. and Amir, E. and Roth, D. (2005) "Lifted first-order probabilistic inference." In *Proc. 19th Int'l Joint Conf. on AI*, pages 1319-1325.
- Dzeroski, S. and Lavrac, N., eds. (2001) *Relational Data Mining*. Springer.
- Flach, P. and Lavrac, N. (2002) "Learning in Clausal Logic: A Perspective on Inductive Logic Programming". In *Computational Logic: Logic Programming and Beyond (Essays in Honour of Robert A. Kowalski),* Springer Lecture Notes in AI volume 2407, pages 437-471.
- Pasula, H. and Russell, S. (2001) "Approximate inference for first-order probabilistic languages". In *Proc. 17th Int'l Joint Conf. on AI*, pages 741-748.
- Milch, B., Marthi, B., Russell, S., Sontag, D., Ong, D. L., and Kolobov, A. (2005) "BLOG: Probabilistic Models with Unknown Objects". In *Proc. 19th Int'l Joint Conf. on AI*, pages 1352-1359.