Probabilistic Reasoning

Inference

and

Relational Bayesian Networks

Outline

- Brief introduction to inference
  - Variable Elimination
  - Markov Chain Monte Carlo
- Application: Citation Matching
  - Relational Bayesian Networks
  - MCMC in action

A Leading Question

- Bayesian networks help us represent things compactly… but how can that translate into better inference algorithms?
- How would you do inference using a Bayes net?

BN inference: a naïve approach

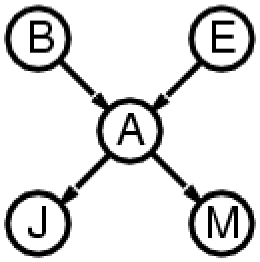
- When we had a full joint, we did inference by summing up entries.
- Well, a BN represents a full joint…
- But don't fill in the whole table! Remember:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}\, \mathbf{P}(X_i \mid Parents(X_i))$$

- To look up any joint entry, we just set the values of all the variables and take the product of their probabilities.

BN inference: a naïve approach

$$P(J \wedge M \wedge A \wedge B \wedge E)$$
$$= P(J \mid A) P(M \mid A) P(A \mid B, E) P(B) P(E)$$



So,

$$P(B \mid m, j) = \frac{\sum_{e,a} P(j \mid a) P(m \mid a) P(a \mid B, e) P(B) P(e)}{\sum_{e,a,b} P(j \mid a) P(m \mid a) P(a \mid b, e) P(b) P(e)}$$

$$= \frac{P(B) \sum_{e} P(e) \sum_{a} P(j \mid a) P(m \mid a) P(a \mid B, e)}{\sum_{b} P(b) \sum_{e} P(e) \sum_{a} P(j \mid a) P(m \mid a) P(a \mid B, e)}$$

…but this involves a lot of repeated computation.

Solution: use dynamic programming

Variable Elimination

- Essentially enumeration with dynamic programming (and clever summation order.)
- If the BN graph is a tree, this is O(# of CPT entries)
  - Which could still be bad if # of parents is not bounded
- Else… can be exponential in "induced clique size."

Bayesian Network Inference

Many options:

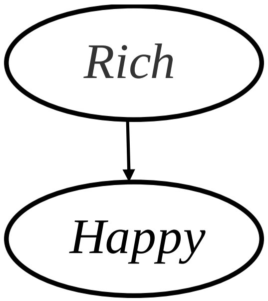
- Exact (intractable in general)
  - ***Variable Elimination***
  - Clustering/Join Tree algorithms
- Approximate (also intractable in general, but faster in practice)
  - Loopy Belief Propagation
  - Variational Methods
  - Sampling (Rejection Sampling, Likelihood Weighing, ***Markov Chain Monte Carlo***)

Sampling Methods

- Idea: instead of summing over all assignments, sum over samples from the assignment space.
- Sampling without evidence obviously trivial, but how do we get evidence in there?
  - Sample full network, ***reject*** samples that do not conform to evidence
  - Sample only non-evidence, ***reweigh*** each sample using evidence
  - Use the magic of ***Markov Chains***

Markov Chain Monte Carlo

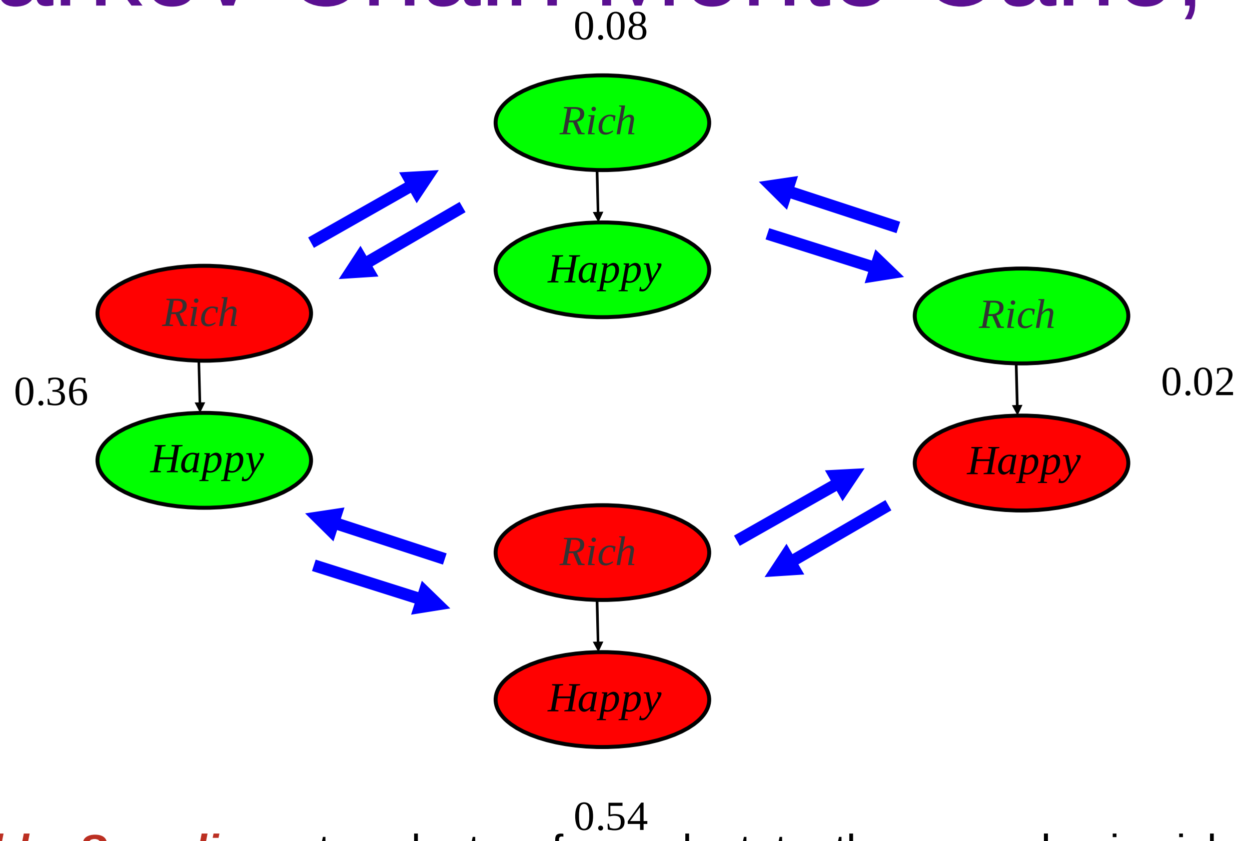
- Markov chain moves through the space of assignments, collects samples
- Transition probabilities set up so fraction of time spent in each state corresponds to its probability
- Example:



$P(Rich)=0.1$

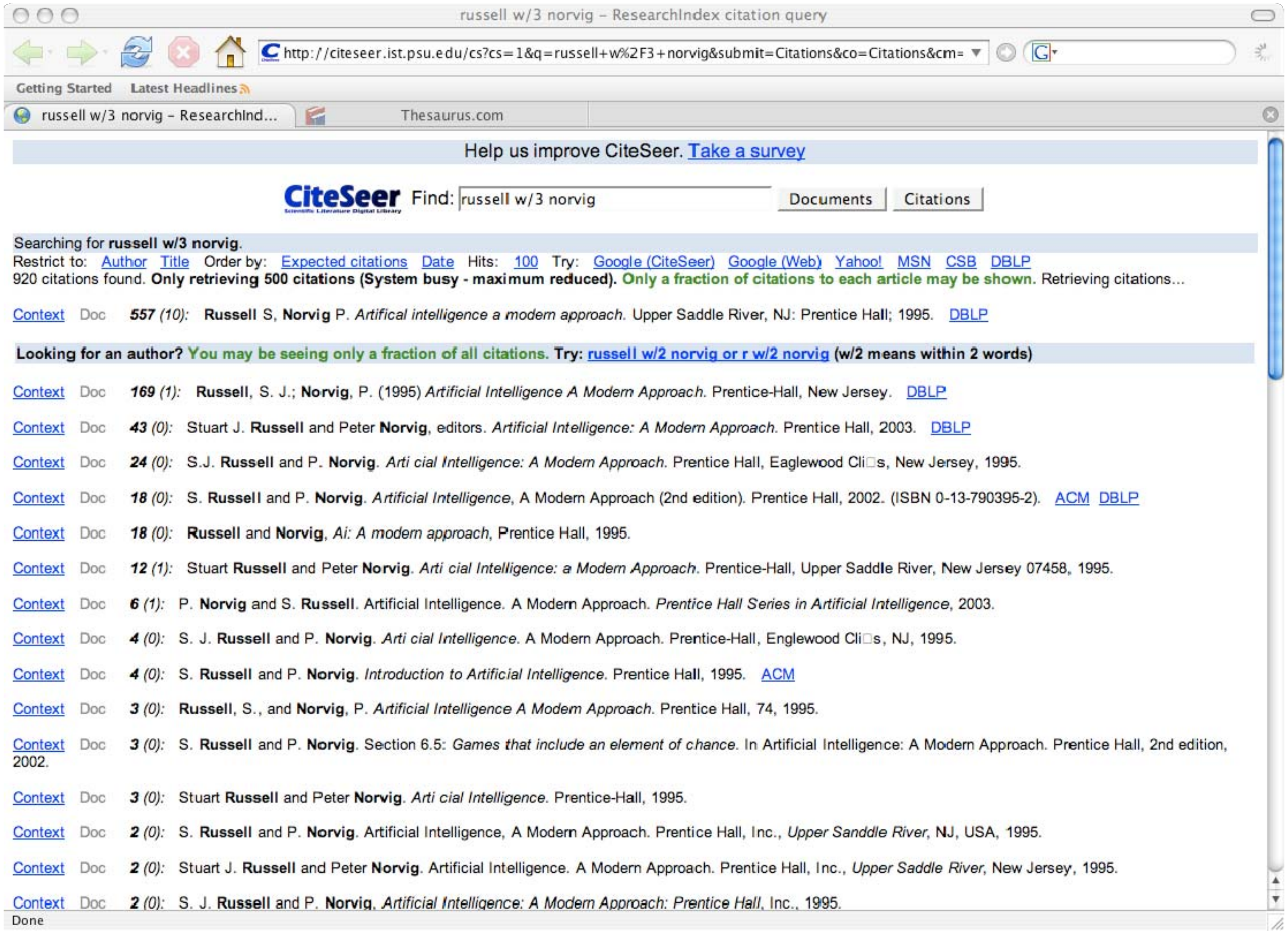
$P(Happy|Rich) = 0.8$
$P(Happy|\neg Rich) = 0.4$

Markov Chain Monte Carlo, cont



- ***Gibbs Sampling***: at each step, for each state, the new value is picked according to the posterior given the current Markov blanket.
- Problem with MCMC: getting it to move around the whole space can be tricky

Citation Matching



A simplified example

Consider the following five citations:

1. *Killer Robots*, Jane Smith
2. *Killer Robots: A Modern Approach*, J. M. Smith
3. *Killer Robots*, J. Kowal
4. Jne M. Simth, *Killer Robotos*
5. *Hamlet: Shakespeare's meatiest play*, Jane Mary Smith

How would *you* match up the papers and the authors?

A Simplified example, cont.

1. *Killer Robots*, Jane Smith
2. *Killer Robtos: A Modern Approach*, J. M. Smith
3. *Killer Robots*, J. Kowal
4. Jne M. Simth, *Killer Robotos*
5. *Hamlet: Shakespeare’s meatiest play*, Jane Mary Smith

How would *you* match up the papers and the authors?

Are you sure you’re right?

Can you explain what general “rules” you used when making your decisions?

Can you express them using a BN?

Relational-Probabilistic Methods

- In the citeseer domain, we need to reason about:
  - Uncertainty
  - The set of objects in the world and their relationships
- In logic, first-order representations:
  - Can encode things more compactly than propositional ones. (E.g., rules of chess.)
  - Are applicable in new situations, where the set of objects is different.
- We want to do the same thing with probabilistic representations

Relational-Probabilistic Methods

- A general approach: defining network fragments that quantify over objects and putting them together once the set of objects is known.

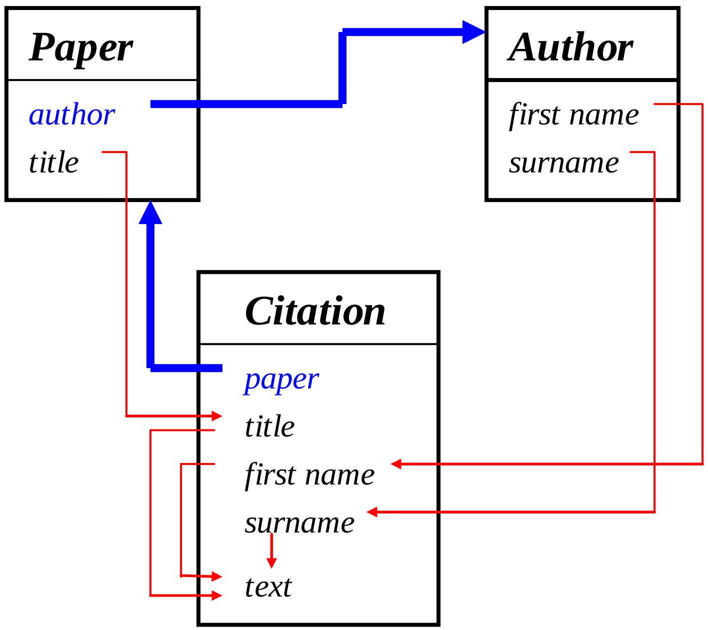
  In the burglar alarm example, we might quantify over people, their alarms, and their neighbours.

- One specific approach: ***Relational Probabilistic Models*** (RPMs.)

  (Like semantic nets; object-oriented.)

An RPM-style Model

- There are different ***classes*** of objects
- Classes have attributes:
  - ***Basic attributes*** take on discrete values (e.g., an author's surname)
  - ***Complex attributes*** map to objects (e.g. a paper's author)
  - Attribute values governed by probability models
- We instantiate objects of various classes (e.g. *Citation1*)
- ***Instance*** attributes -> BN nodes
- Probabilistic dependencies -> BN edges



An RPM-style Model

Priors needed for:

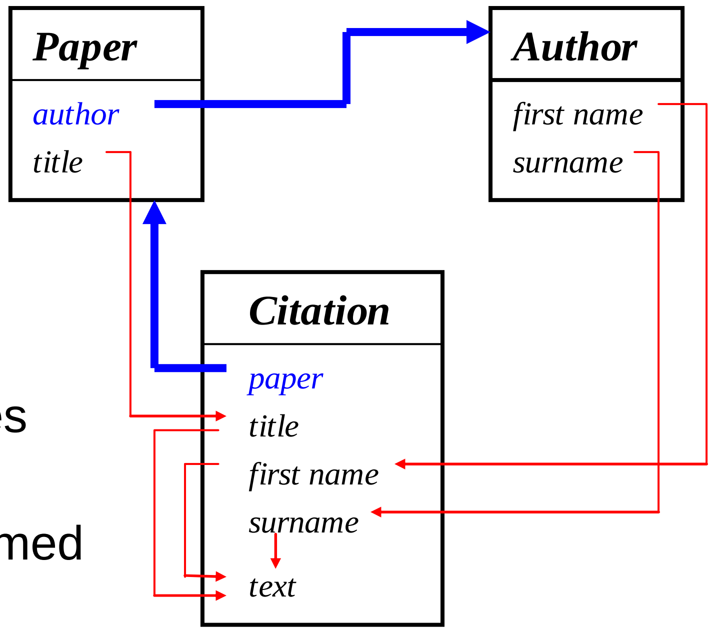
- The # of authors
- The # of papers
- Author names
- Paper titles

Conditional distributions needed for:

- Misspelt titles/names given real ones
- Citation text given citation fields

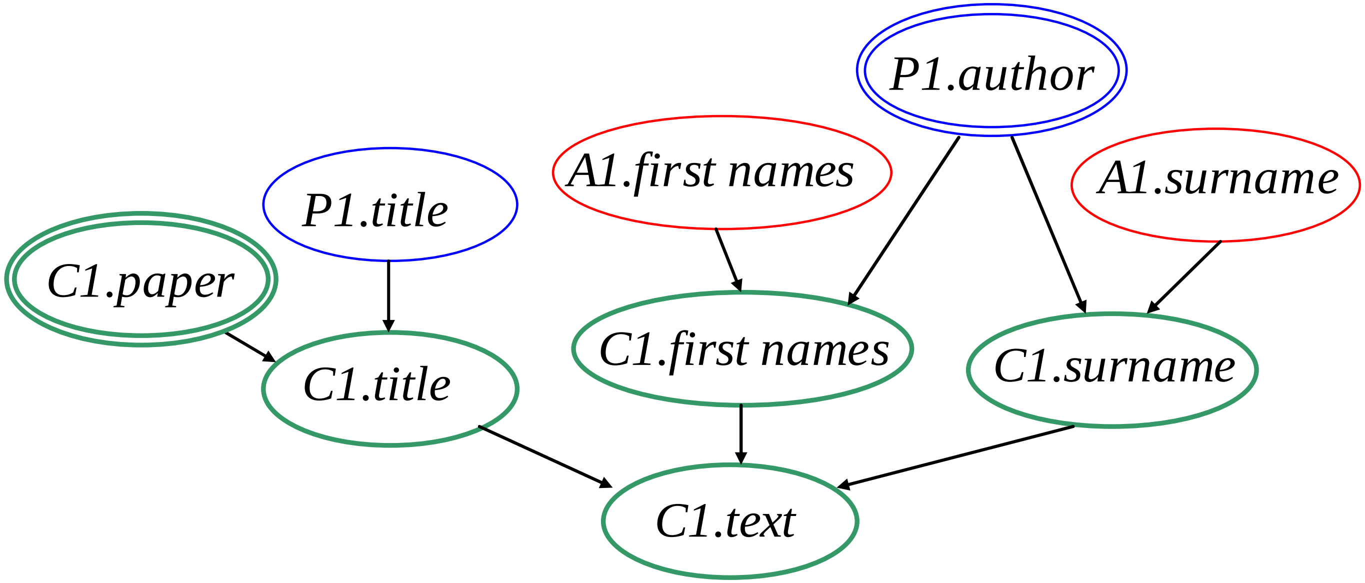
Distributions over authors/papers assumed uniform *a priori*.

Only ***text*** is observed.



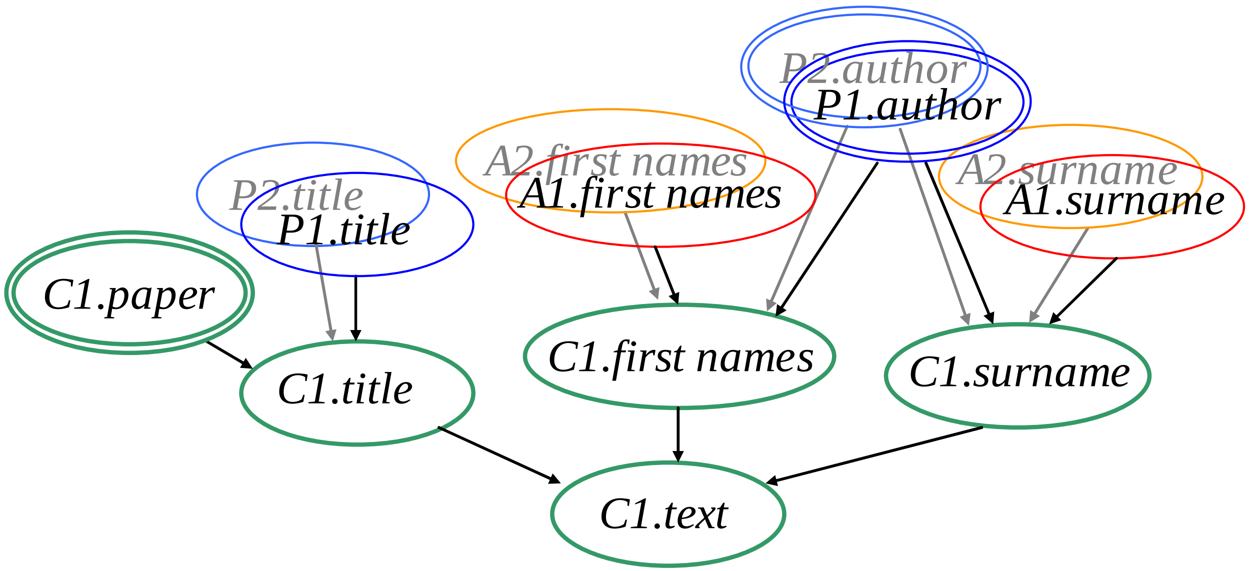
Building the BN

Given instances citation C1, paper P1, author A1, we get:



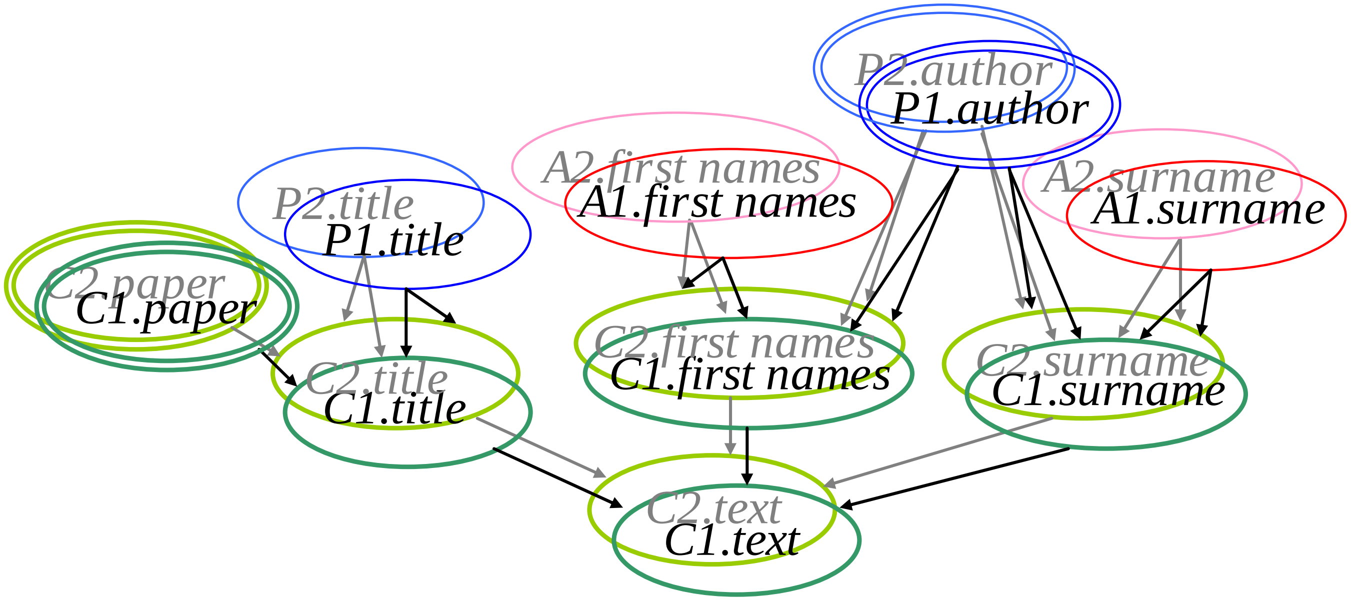
Building the BN

- Of course, the idea is that each citation could correspond to many different papers…



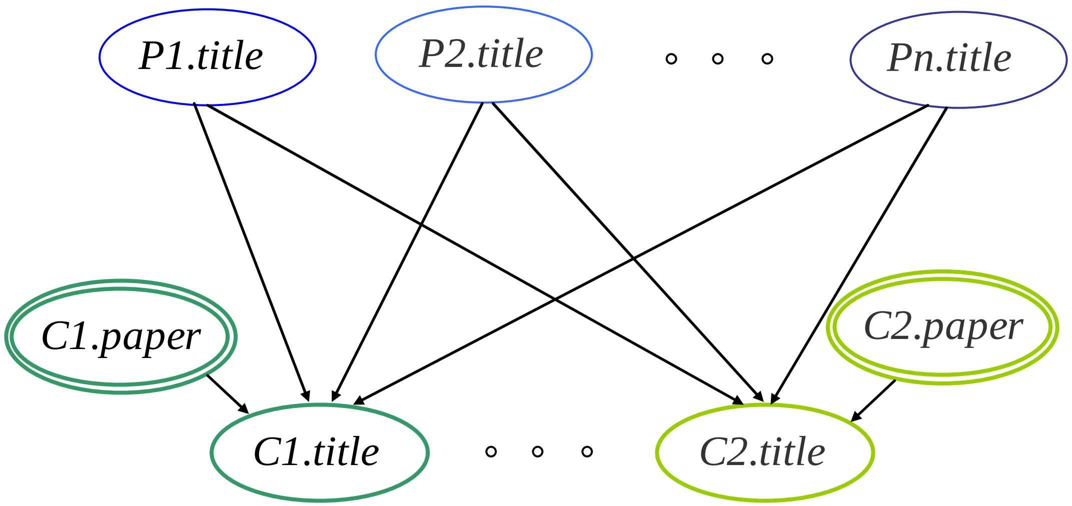
Constructing the BN

- And that we are reasoning about multiple citations…



BN structure, close up

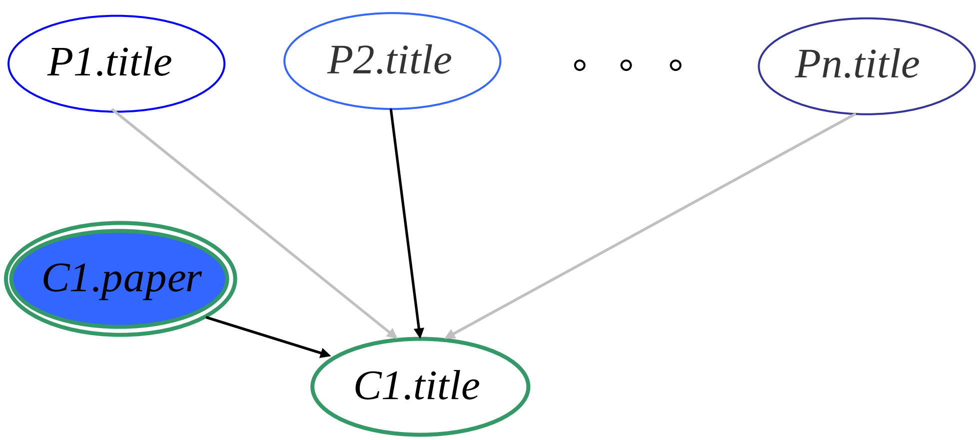
- Let us zoom in:



- A very highly-connected model: inference seems a bit hopeless

Context-specific Independence

- Note: distribution at *C1.title* is essentially a multiplexer.



- For any particular *paper* value, network structure is simplified

Identity Uncertainty

- But wait, we are given only citations. Where are all the papers and authors coming from?
- In broad terms:
  - We create one for each citation and each author in a citation
  - We then use an equivalence relation to group the co-referring ones together

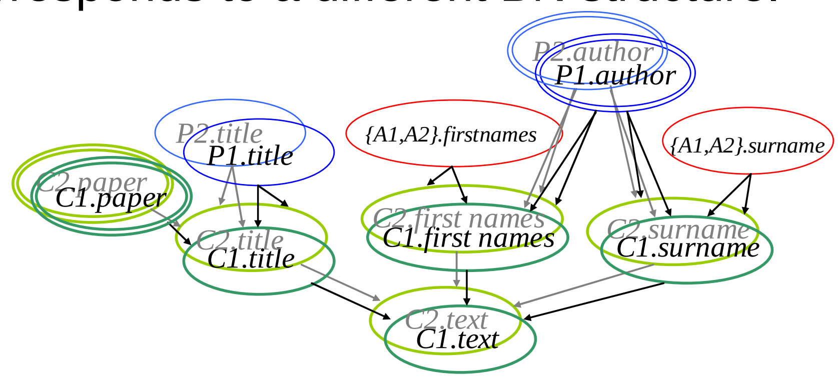
    Example: {{P1, P3, P4}{P2}{P5}{A1, A2, A3, A4}{A5}}
  - Each set corresponds to a paper or an author

    ({A1, A2} can be thought of as “The author known as ‘Jane Smith’ and ‘J.M.Smith’”
- To do inference properly, we should sum over all possible equivalence relations

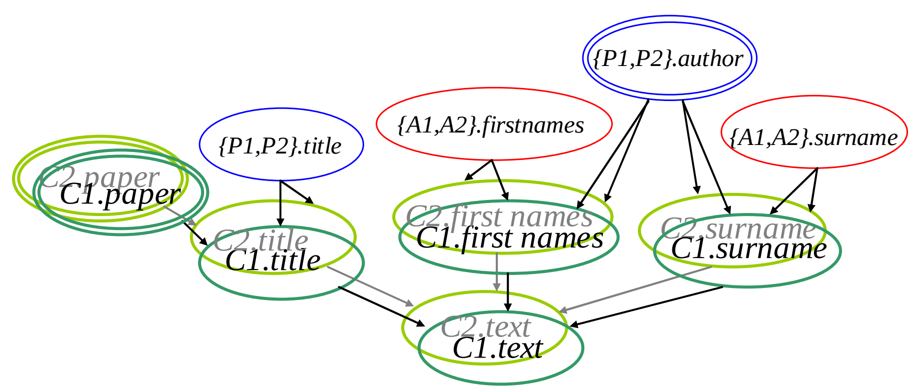
Identity Uncertainty and Inference

- To do inference properly, we should sum over all possible equivalence relations. But there are intractably many…
- Also, note that each relation corresponds to a different BN structure. E.g.,

{{P1}{P2}{A1,A2}}



{{P1,P2}{A1,A2}}

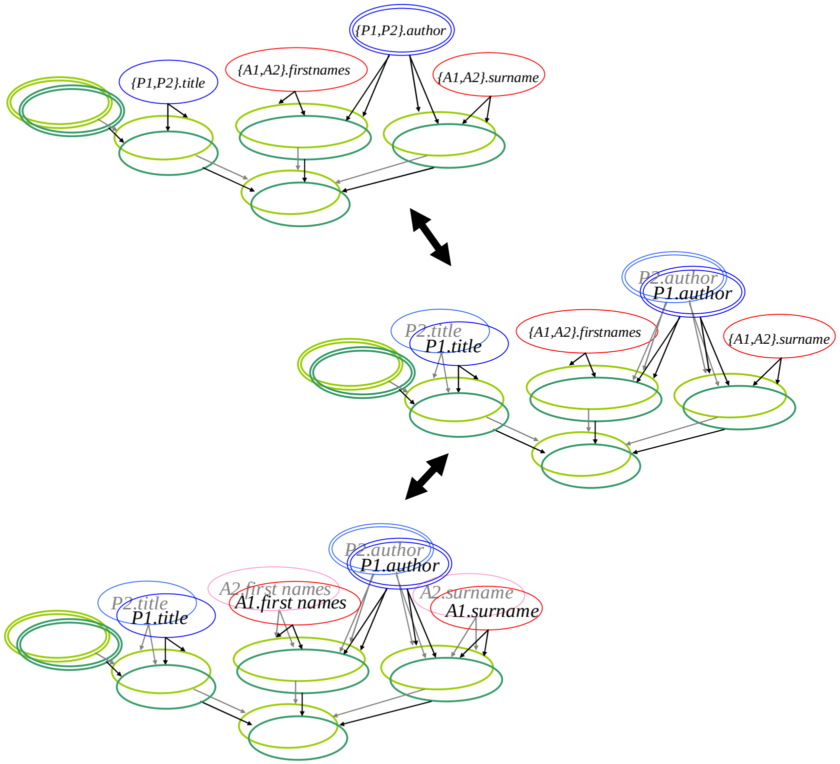


MCMC, again

- We could sample from the space of equivalence relations--using, perhaps, MCMC
- Also, recall that each MCMC state is a fully specified assignment to all the variables--so we can use the simplified network structures!

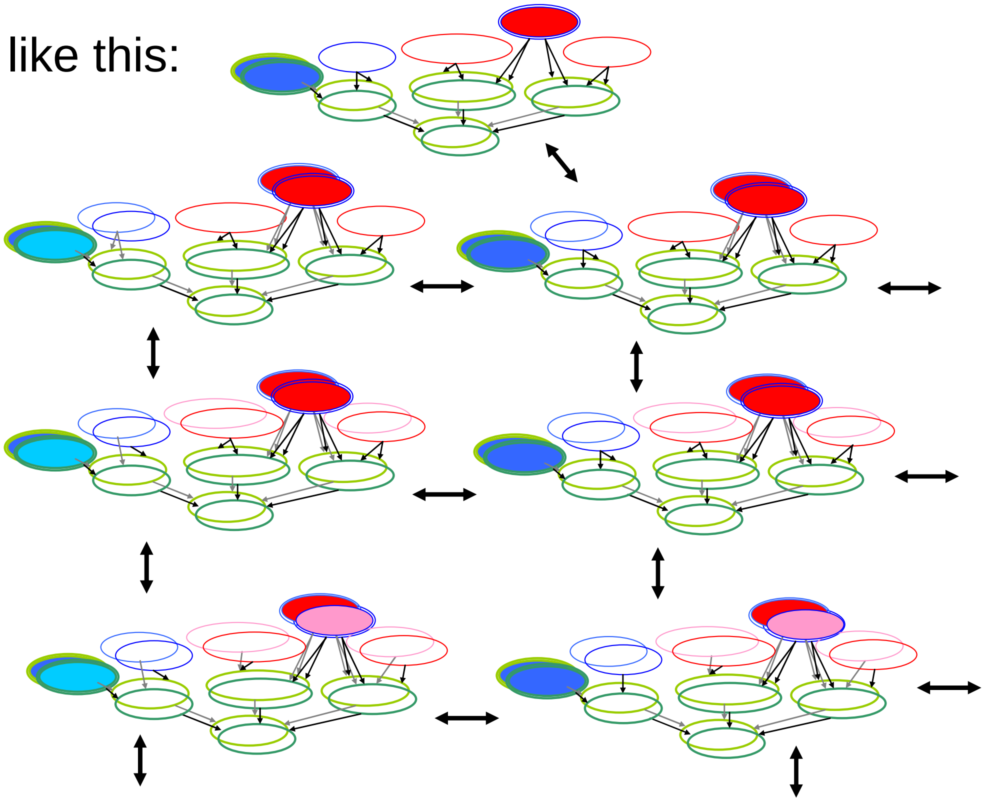
MCMC in action

The state space will look a little like this:

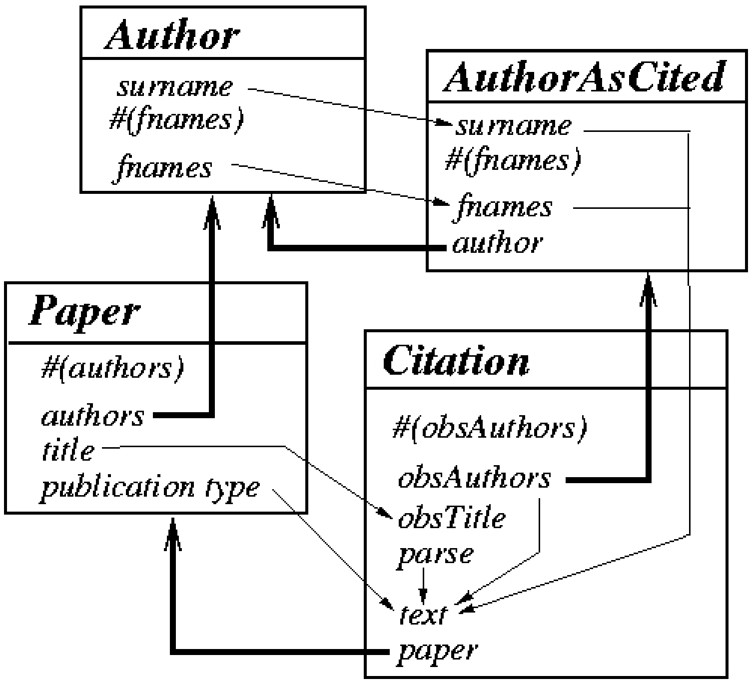


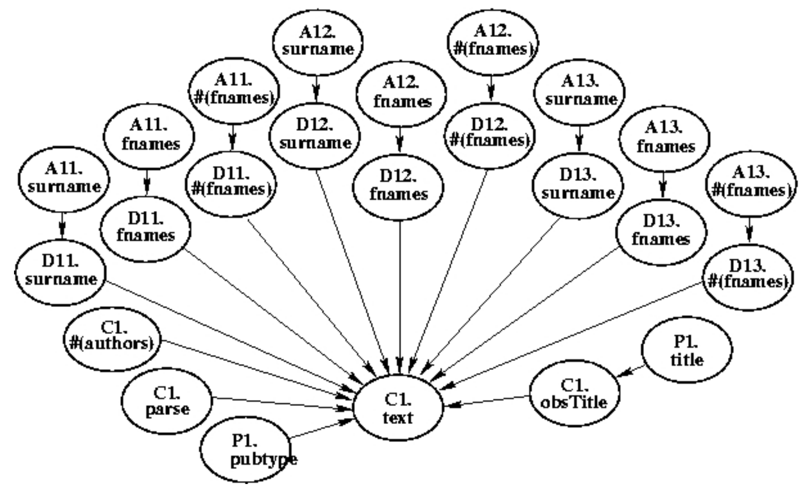
MCMC in action

Actually, more like this:



The Actual Model





Generative Models

- The models we have looked at today are all examples of ***generative models***.

  They define a full distribution over everything in the domain, so possible worlds can be generated from them.

  Generative models can answer any query.
- An alternative is ***discriminative models***.

  There, the query and evidence are decided upon in advance and only the conditional distribution P(Q|E) is learnt.
- Which type is “better”? Current opinion is divided.