Mind Change Optimal Learning Of Bayes Net Structure

Oliver Schulte

School of Computing Science
Simon Fraser University
Vancouver, Canada

oschulte@cs.sfu.ca `

with Wei Luo (SFU, wluoa@cs.sfu.ca) and
Russ Greiner (U of Alberta, greiner@cs.ualberta.ca)

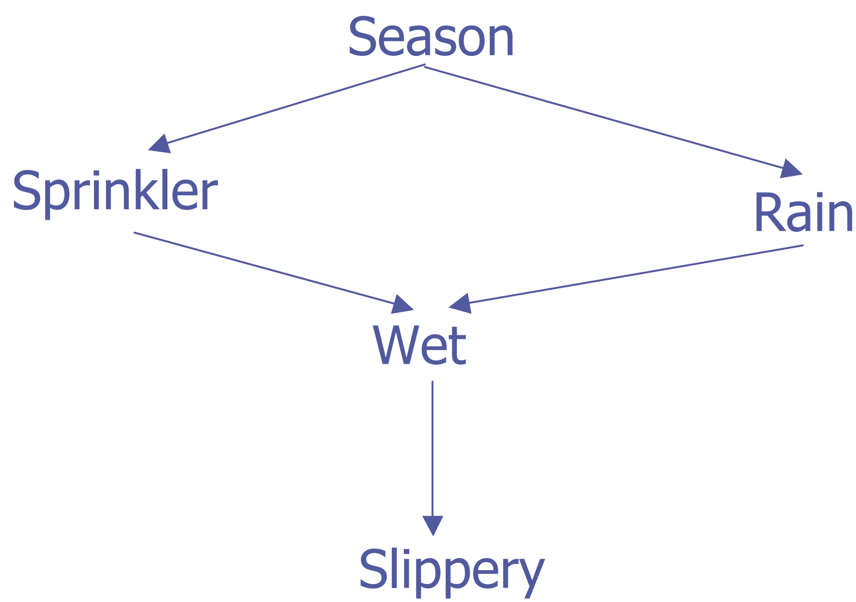
Outline

1. Brief Intro to Bayes Nets (BNs).
2. Language Learning Model for BN Structure Learning.
3. Mind Change Complexity of BN Learning.
4. Mind Change, Convergence Time Optimality.
5. NP-hardness of Optimal Learner.

Bayes Nets: Overview

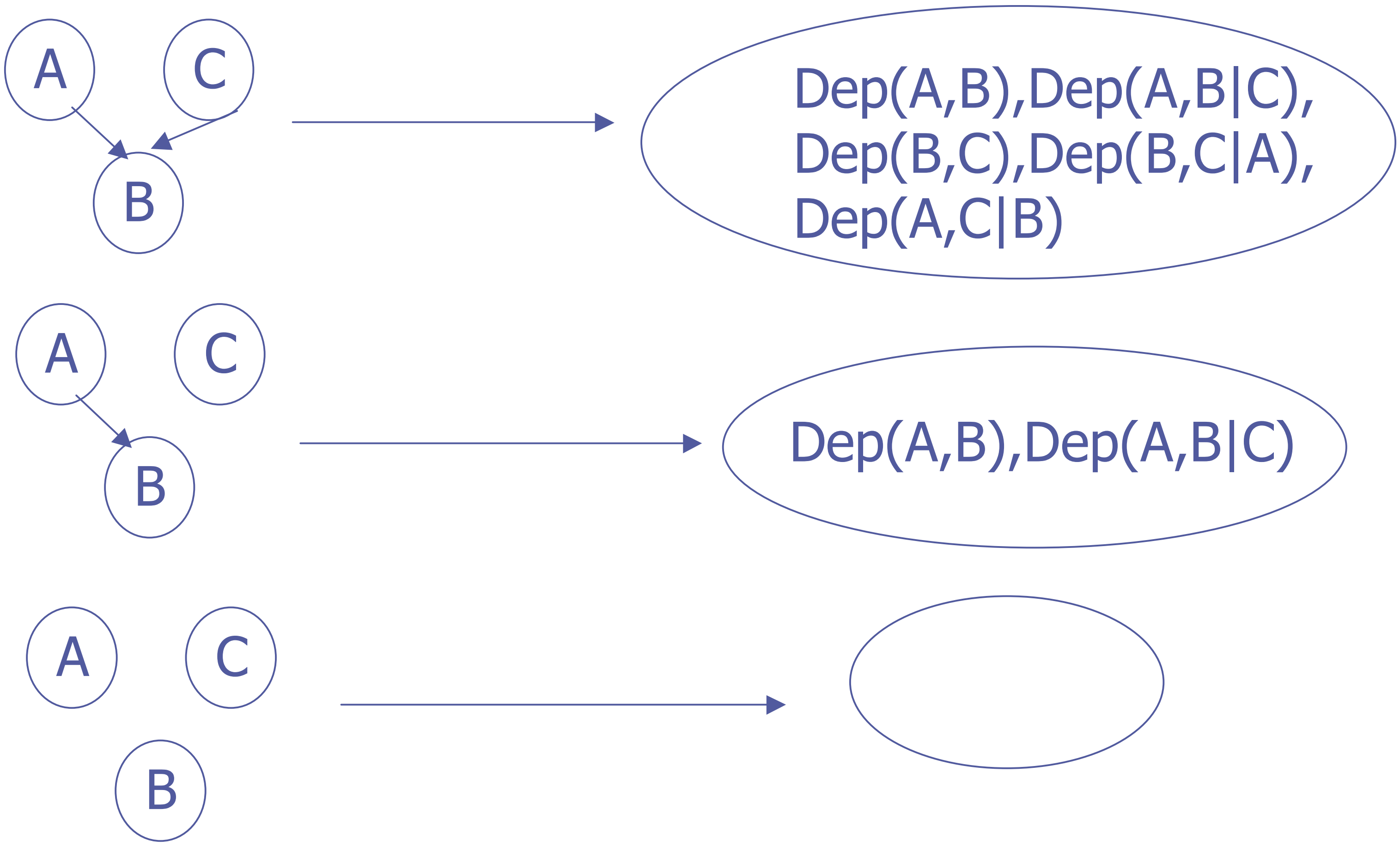
- Very widely used graphical formalism for probabilistic reasoning and KR in AI and machine learning.
- **Bayes Net Structure** = Directed Acyclic Graph.
- **Nodes** = Variables of Interest.
- **Arcs** = direct "influence", "association".
- Structure represents probabilistic <u>conditional dependencies</u> (correlations).

Example of Bayes Net Structure



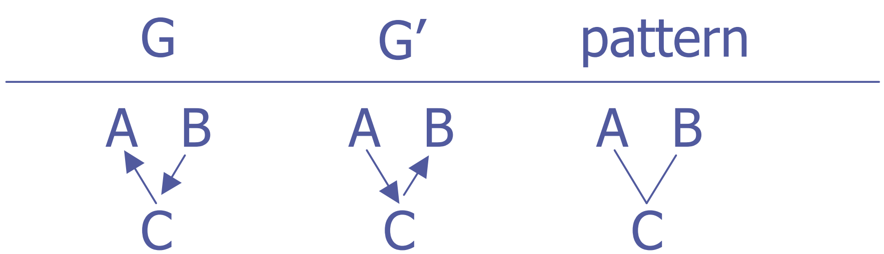
1. Season depends on Slippery.
2. Sprinkler depends on Rain.
3. Sprinkler does not depend on Rain given Season.
4. Sprinkler depends on Rain given Season, Wet.

Graphs entail Dependencies



Pattern = DAG Equivalence Class

- Write Dep(G) for the dependencies defined by DAG G.
- Natural Equivalence relation: $G \approx G' \Leftrightarrow Dep(G) = Dep(G')$.
- A partially directed graph, called a **pattern**, represents the equivalence class for a given DAG G.



Constraint-Based BN Learning as Language Learning

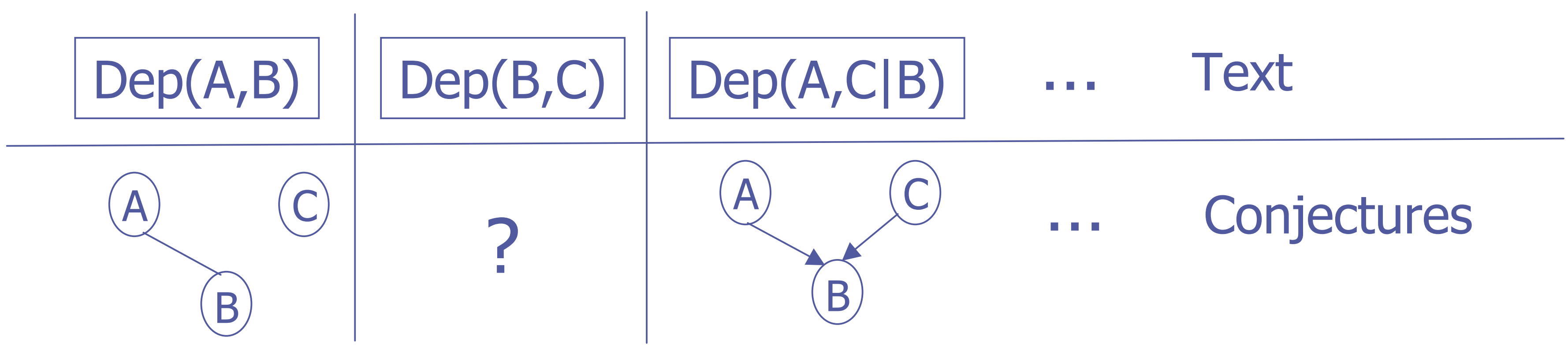
Constraint-Based Approach: Learn BN from *(in)dependency information*. Spirtes, Glymour, Shines (2000); Pearl and Verma (2000); Margaritis and Thrun (1999); Cheng and Greiner (2001).

Bayes Net	Gold paradigm
conditional dependence: Dep(X,Y|**Z**) **Z**= set of variables	string
dependency relation	language
pattern	index

A **BN learner** maps a sequence of dependencies (repetitions allowed) to a pattern or to ?.

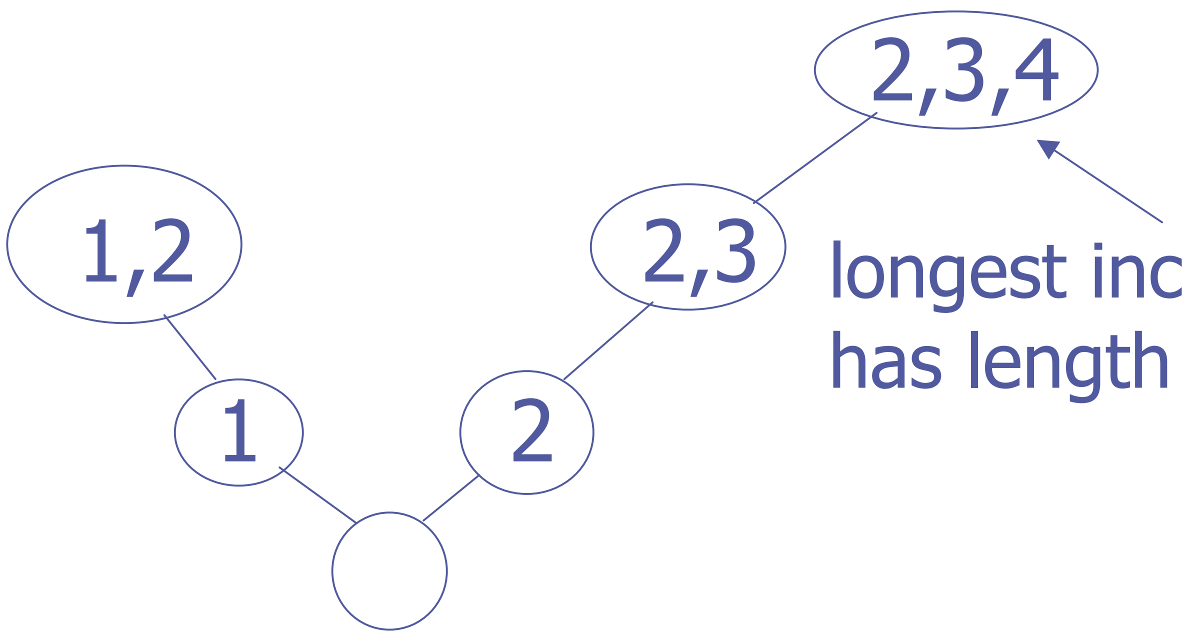
Identification with Bounded Mind Changes

- Learner Ψ changes its mind on text T at stage k+1 $\Leftrightarrow$ $\Psi(T[k]) \neq \Psi(T[k+1])$ or $\Psi(T[k]) \neq ?$ and $\Psi(T[k+1]) = ?$.
- Learner Ψ identifies language collection $\mathcal{L}$ with k mind changes $\Leftrightarrow$ Ψ identifies $\mathcal{L}$ and changes its mind at most k times on any text for a language in $\mathcal{L}$.
- $\mathcal{L}$ is identifiable with k mind changes $\Leftrightarrow$ there is a learner Ψ that identifies $\mathcal{L}$ with k mind changes.



Inclusion Depth and Mind Change Bounds

Proposition (Luo and Schulte 2006)
Suppose that $\mathcal{L}$ has finite thickness. Then the best mind change bound for $\mathcal{L}$ is given by the length of the longest **inclusion chain** $L_1 \subset L_2 \subset \ldots L_k$ formed by languages in $\mathcal{L}$.



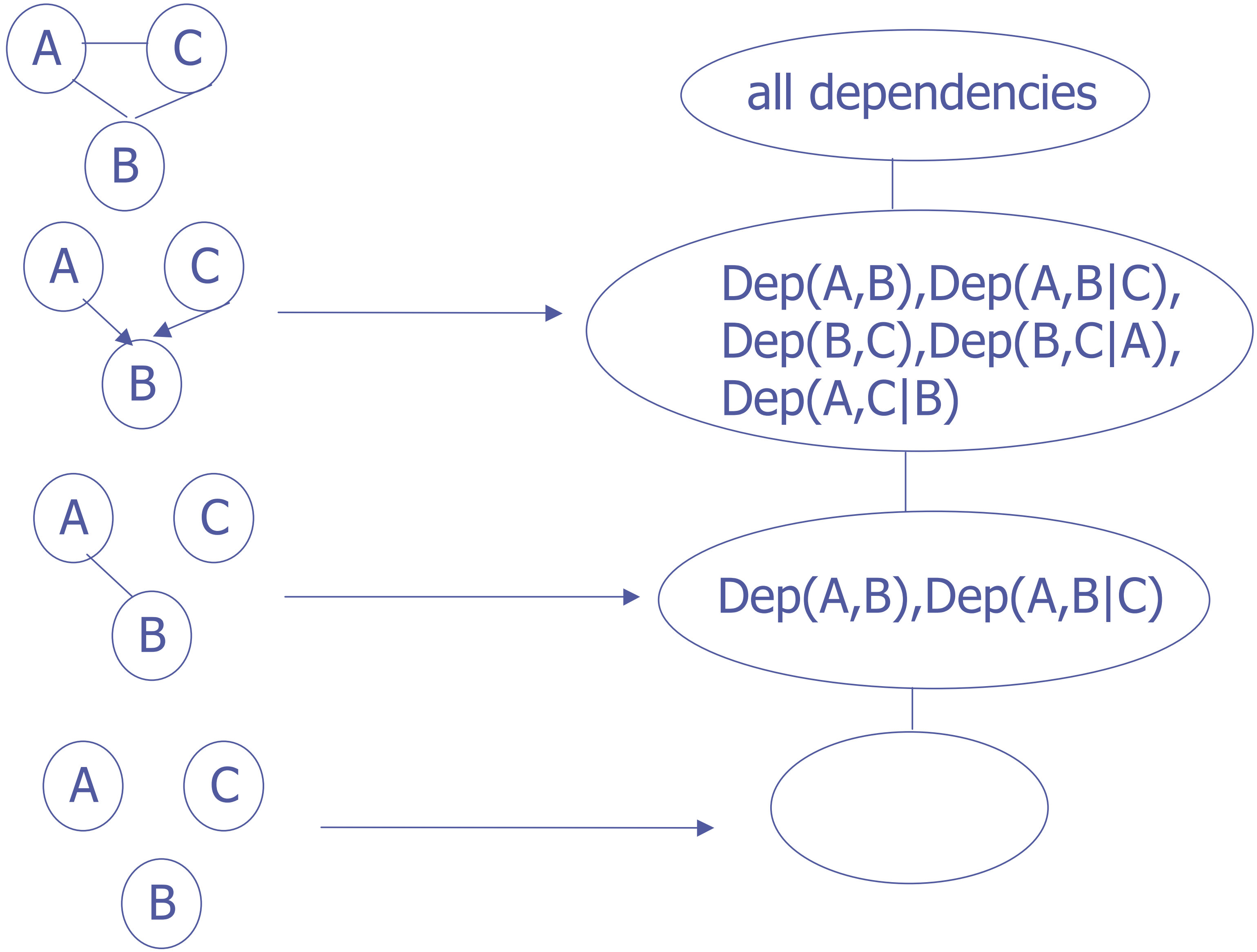
longest inclusion chain has length 4

Mind Change Complexity of BN Learning

Let $\mathcal{L}_V$ be the collection of dependency relations definable by Bayes nets with variables V.

Theorem The longest inclusion chain in $\mathcal{L}_V$ is of length $\binom{|V|}{2}$ =
the <u>number of edges in a complete graph</u>.

Maximal Length Inclusion Chain

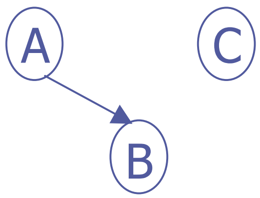
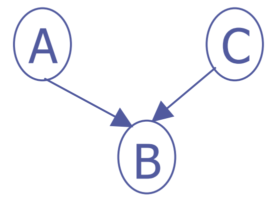


Mind Change Optimal Learning

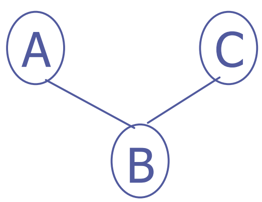
- Learner Ψ is **MC-optimal** for language collection $\mathcal{L} \Leftrightarrow$ if given any data sequence σ, the learner Ψ identifies $\mathcal{L}$ with the best possible mind change bound for the language collection {L: L is in $\mathcal{L}$ and consistent with σ}.
- **Proposition** A BN learner identifying $\mathcal{L}$ is MC-optimal $\Leftrightarrow$ for all dependency sequences σ, if there is no unique edge-minimal pattern consistent with σ, then $\Psi(\sigma) = ?$.

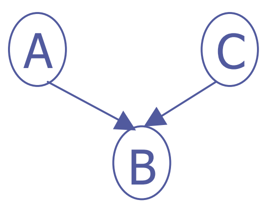
Proof follows from general characterization of MC-optimality in Luo and Schulte (2005,2006).

Example of Mind Change Optimal Learner

Dep(A,B)	Dep(B,C)	Dep(A,C|B)	...	Text
	?		...	Conjectures

Alternatives:





Convergence Time

- Convergence Time = number of observed dependencies - important to minimize
- *Def* (Gold) Learner Ψ is **uniformly faster** than learner $\Phi \Leftrightarrow$
  1. Ψ converges at least as fast as Φ on every text T, and
  2. Ψ converges strictly faster on some text T.

Define

$$\Psi_{\text{fast}}(\sigma) = \begin{cases} G & \text{if } G \text{ is the unique edge-minimal pattern consistent with } \sigma \\ ? & \text{otherwise} \end{cases}$$

Proposition The learner Ψ_{fast} is uniformly faster than **any other MC-optimal BN learner.**

Complexity Analysis

A list of dependencies is compactly represented by a **dependency oracle**.

Dep(A,B)	Dep(B,C)	Dep(A,C|B)	dependency list

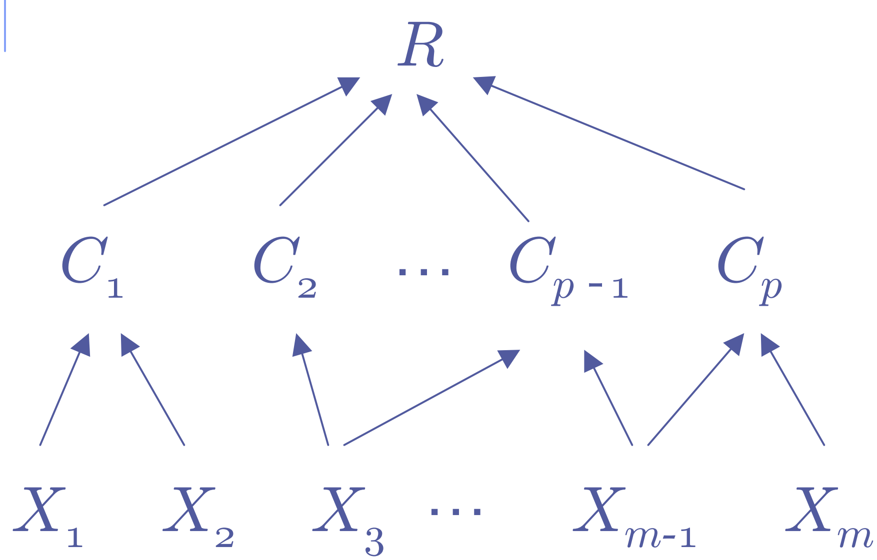
Q:"Dep(B,C)" → $\boxed{\text{Oracle}}$ → yes

Q:"Dep(A,C)" → $\boxed{\text{Oracle}}$ → ?

Unique *k* O-cover Given a dependency oracle O, and a bound *k*, is there a DAG G covering the dependencies in O with ≤ *k* edges s.t. all other DAGs G' covering the dependencies in O have more edges than G?

NP-hardness result

Theorem *Unique X3set-Cover* reduces to *Unique k O-Cover*. So if P = RP, then UMOC is NP-hard.
Basic Idea: Construct a dependency oracle that forces a tree.



- Universe: $X_1,\ldots,X_m$.
- Sets: $C_1,\ldots,C_p$.
- All elements must be dependent on R.

Conclusion

- Constraint-based approach to BN learning analyzed as language learning problem.
- Mind Change Complexity = $\binom{n}{2}$, where n is the number of variables.
- Number of edges: new intuitive notion of *simplicity* for a BN, based on learning theory.
- Unique fastest mind-change optimal method is NP-hard.

Future Work

- Heuristic Implementation of MC-optimal Learner (GES search).
- Leads to a new BN learning algorithm with good performance.

References

- W. Luo and O. Schulte. *Mind change efficient learning.* In COLT 2005, pages 398-412.
- W. Luo and O. Schulte. *Mind change efficient learning*. Information and Computation 204:989-1011, 2006.

THE END