UNIVERSITY OF CALIFORNIA
SANTA CRUZ

JOINT PROBABILISTIC INFERENCE OF CAUSAL STRUCTURE

Dhanya Sridhar

Lise Getoor

U.C. Santa Cruz

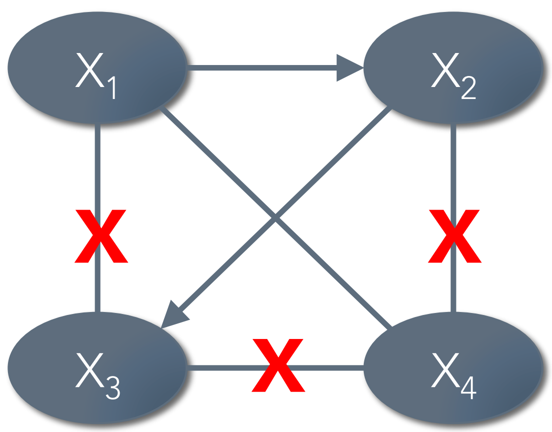
KDD Workshop on Causal Discovery

August 14th, 2016

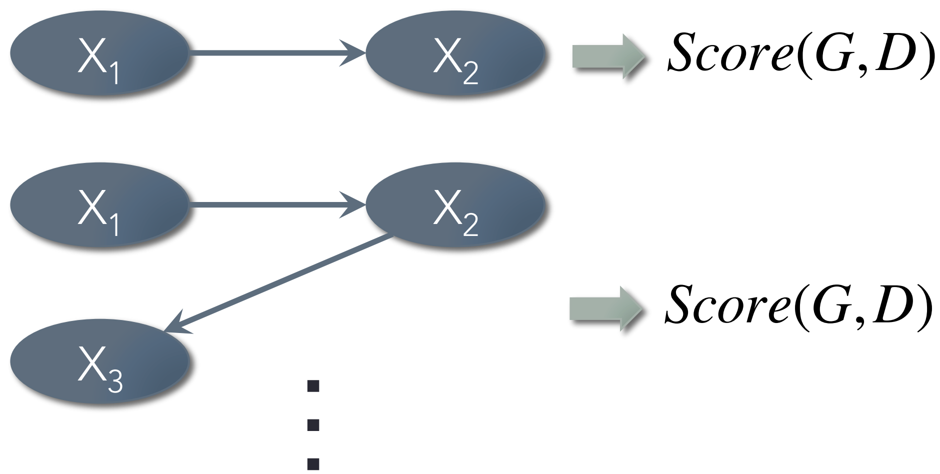
Outline

- Motivation
- Problem Formulation
- Our Approach
- Preliminary Results

Traditional to Hybrid Approaches

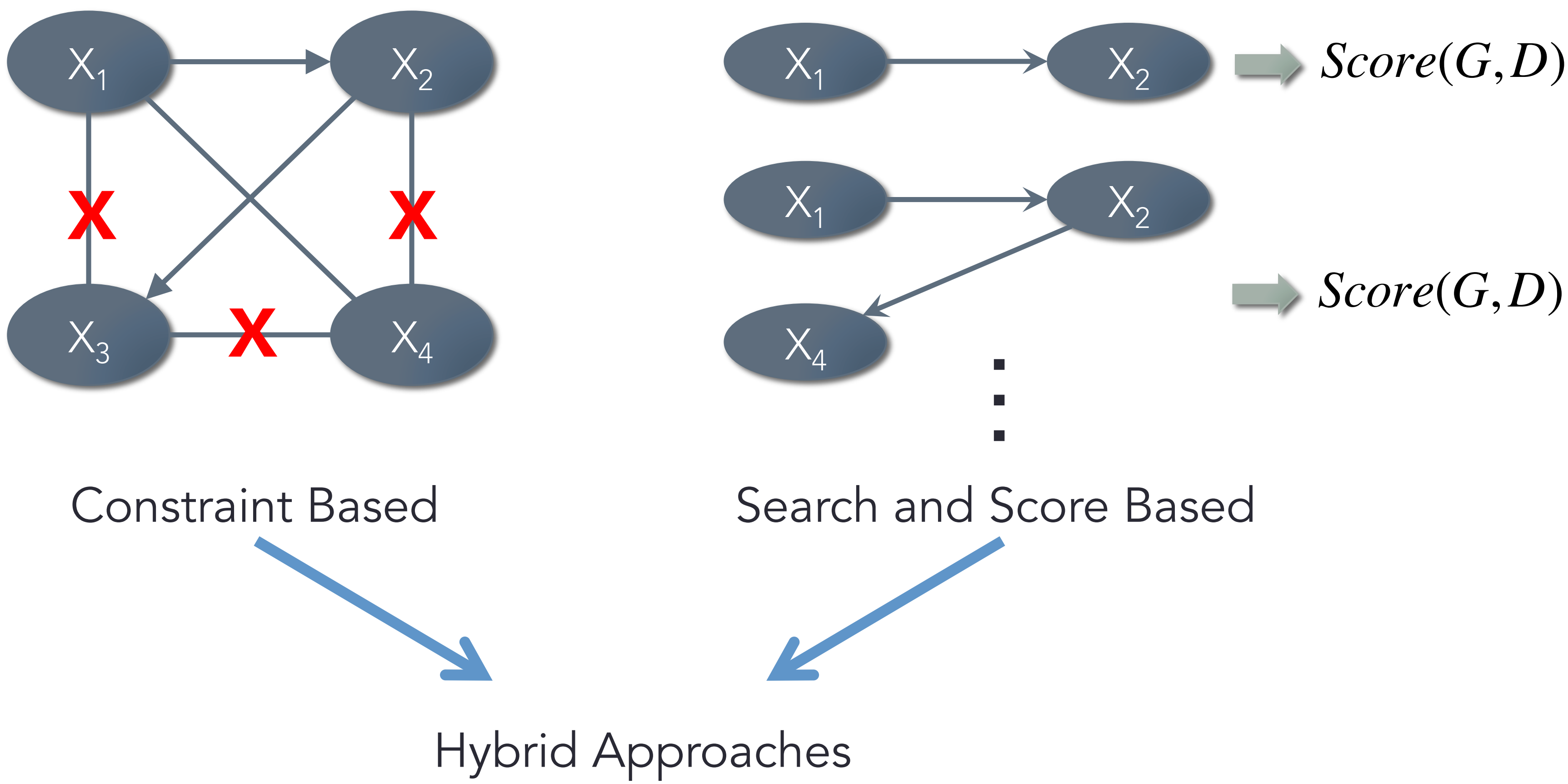


Constraint Based

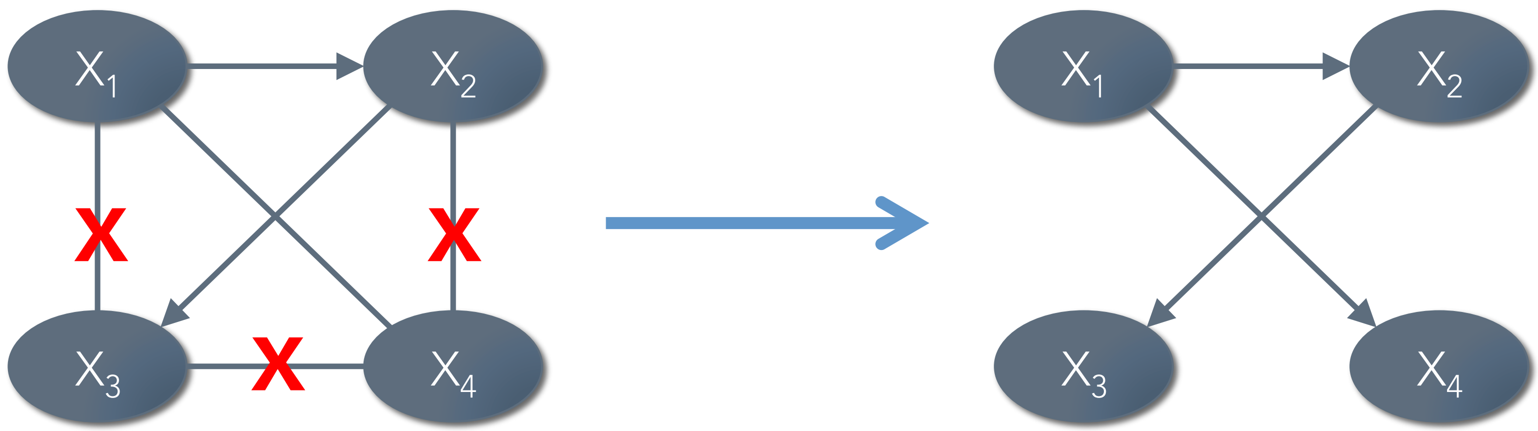


Search and Score Based

Traditional to Hybrid Approaches



Traditional to Hybrid Approaches



Constraint Based

$\Rightarrow$ $Score(G,D)$

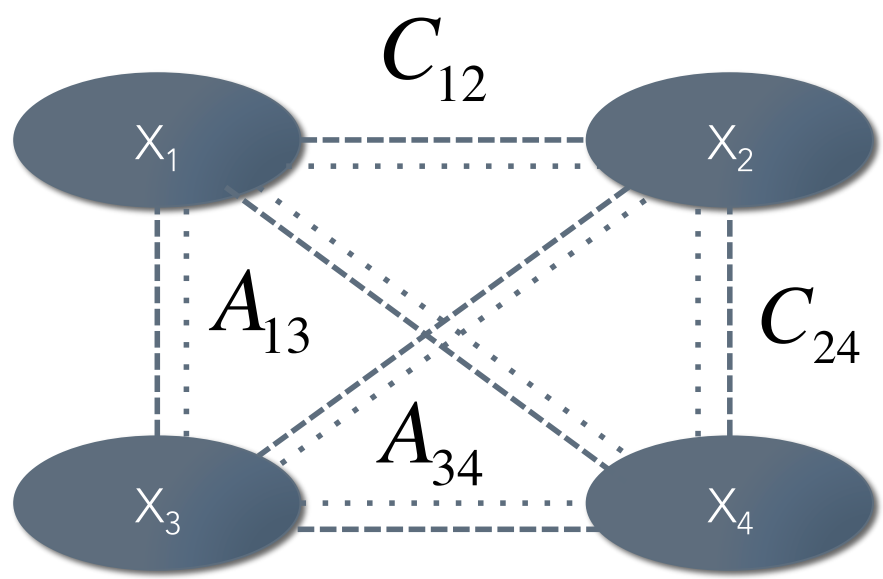
Hybrid Approaches:

- PC-based DAG Search – Dash and Drudzel, UAI 99
- Min-max Hill Climbing – Tsamardinos et al., JMLR 06

Joint Inference for Structure Discovery

Joint Inference of Variables:

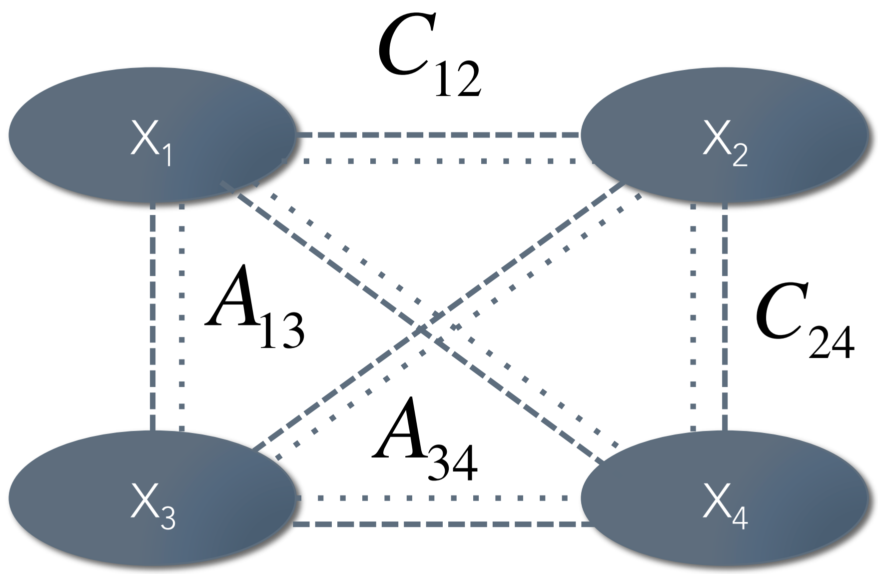
- Causal Edge C_{ij}
- Adjacency Edges A_{ij}



Joint Inference for Structure Discovery

Joint Inference of Variables:

Causal Edge C_{ij}
Adjacency Edges A_{ij}



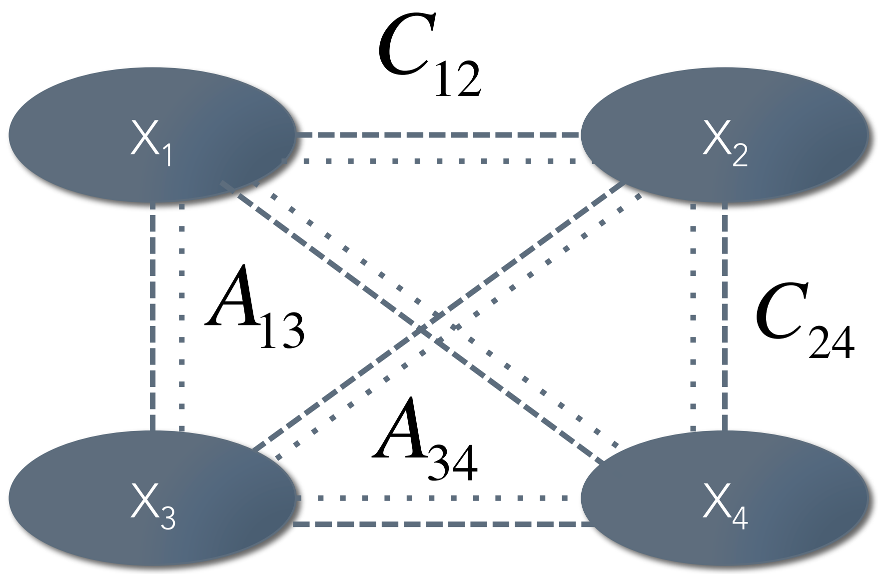
Joint Inference Approaches:

- Linear Programming Relaxations, Jaakkola et al., AISTATS 10

Joint Inference for Structure Discovery

Joint Inference of Variables:

Causal Edge C_{ij}
Adjacency Edges A_{ij}



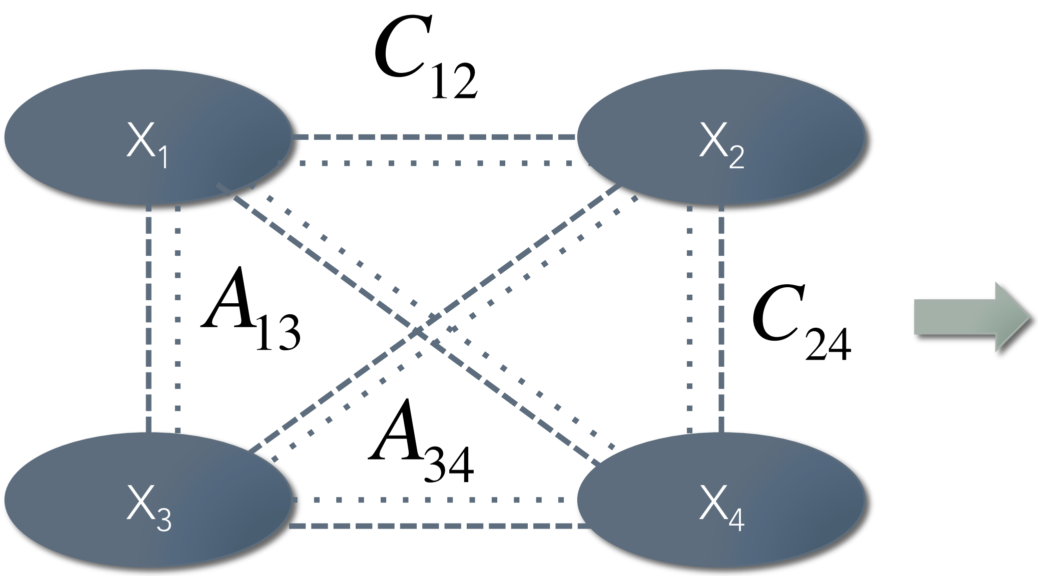
Joint Inference Approaches:

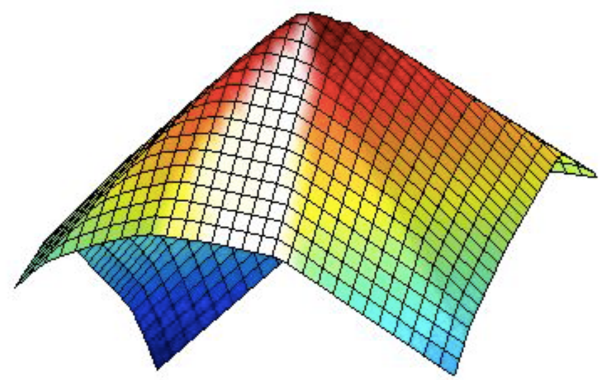
- Linear Programming Relaxations, Jaakkola et al., AISTATS 10
- MAX-SAT, Hyttinen et al., UAI 13

Outline

- Motivation
- **Problem Formulation**
- Our Approach
- Preliminary Results

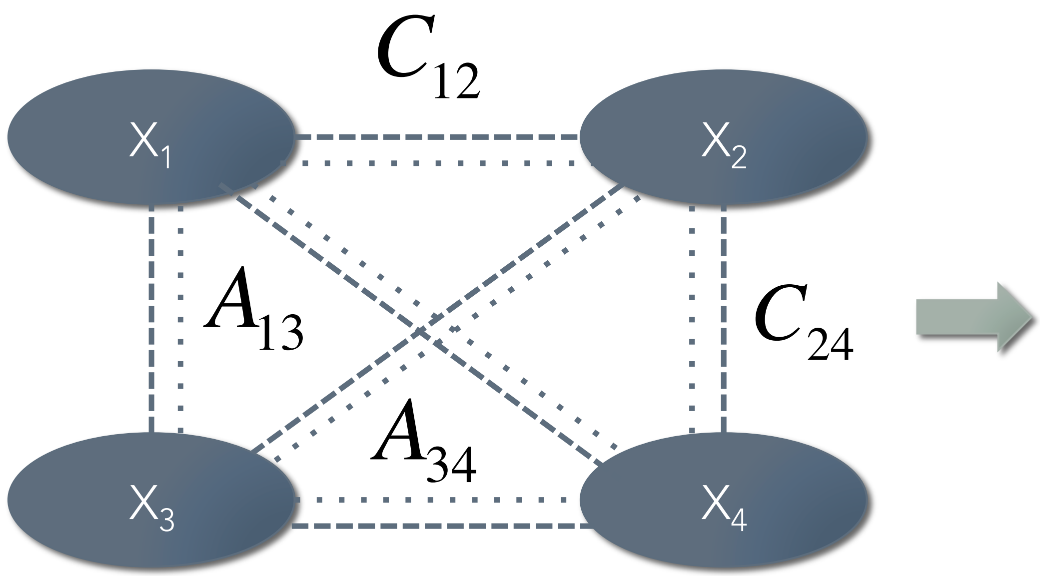
Probabilistic Joint Model of Causal Structure

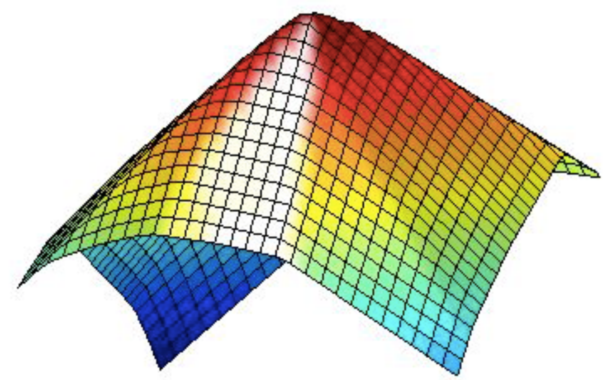




Extending joint approaches:
probabilistic model over causal structures

Probabilistic Joint Model of Causal Structure

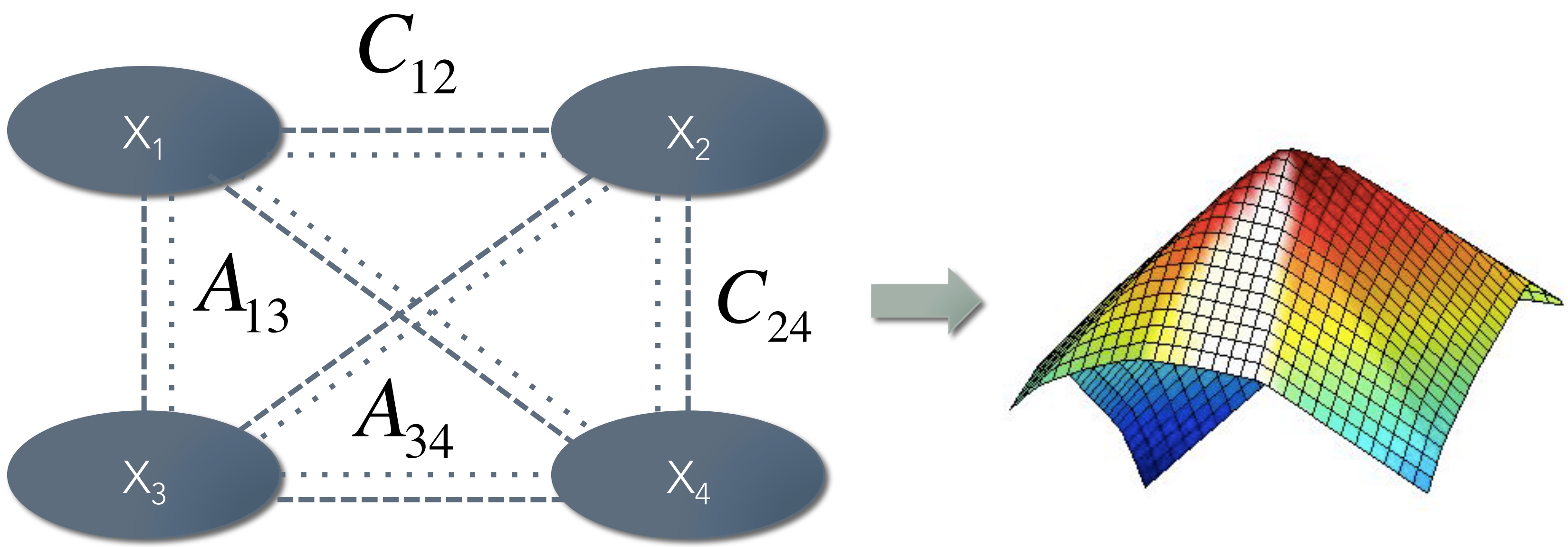




Independence Tests

$$\arg\max_{\mathbf{C},\mathbf{A}} P(C_{ij}, A_{ij} | I_{ij}) \quad \forall i, j$$

Probabilistic Joint Model of Causal Structure



Combining logical and structural constraints and probabilistic reasoning

Outline

- Motivation
- Problem Formulation
- **Our Approach**
- Preliminary Results

Probabilistic Soft Logic (PSL)

- Logic-like syntax with probabilistic, soft constraints
- Describes an undirected graphical model

5.0: *Causes(A, B)* ^ *Causes(B, C)* ^ *Linked(A,C)* → *Causes(A, C)*

↑ Weighted rules

Bach et. al (2015). "Hinge-loss Markov Random Fields and Probabilistic Soft Logic." arXiv.
Open source software: https://psl.umiacs.umd.edu

Probabilistic Soft Logic (PSL)

- Logic-like syntax with probabilistic, soft constraints
- Describes an undirected graphical model

5.0: *Causes(A, B)* ^ *Causes(B, C)* ^ *Linked(A,C)* → *Causes(A, C)*

Weighted rules

Predicates are continuous random variables!

Bach et. al (2015), arXiv
Open source software: https://psl.umiacs.umd.edu

Probabilistic Soft Logic (PSL)

- Logic-like syntax with probabilistic, soft constraints
- Describes an undirected graphical model

Relaxations of Logical Operators

5.0: *Causes(A, B)* ^ *Causes(B, C)* ^ *Linked(A,C)* → *Causes(A, C)*

Weighted rules

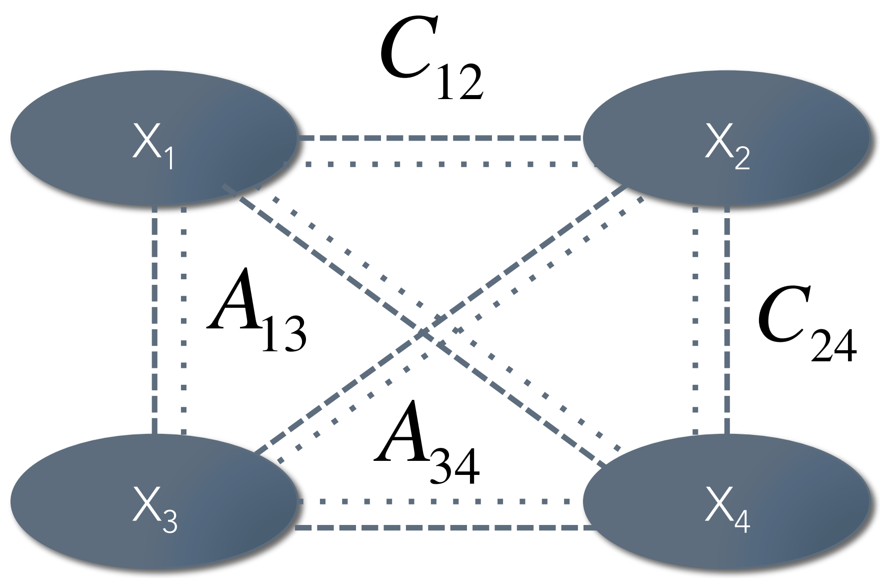
Predicates are continuous random variables!

Bach et. al (2015), arXiv
Open source software: https://psl.umiacs.umd.edu

Probabilistic Soft Logic (PSL)

- Rules instantiated with values from real network

5.0: *Causes(A, B)* ^ *Causes(B, C)* ^ *Linked(A,C)* → *Causes(A, C)*



Probabilistic Soft Logic (PSL)

- Rules instantiated with variables from real network

C_{24}

5.0: $Causes(X_1, X_2) \wedge Causes(X_2, X_4) \wedge Linked(X_1, X_4) \rightarrow Causes(X_1, X_4)$

C_{12} C_{14} A_{14}

Soft Logic Relaxation

5.0: *Causes*(X_1, X_2) ^ *Causes*(X_2, X_4) ^ *Linked*(X_1,X_4) → *Causes*(X_1, X_4)

Convex relaxation of implication
and distance to rule satisfaction

$$\max\{\ell(C_{12}, C_{24}, A_{14}, C_{14}), 0\}$$

Linear Function

Bach et al. (2015), arXiv

Hinge-loss Markov Random Fields

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(w, \mathbf{X})} \exp\left[-\sum_{j=1}^{m} w_j \left[\max\left\{\ell_j(\mathbf{Y}, \mathbf{X}), 0\right\}\right]^{\{1,2\}}\right]$$

Conditional random field

Bach et al. (2015), arXiv

Hinge-loss Markov random fields

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(w, \mathbf{X})} \exp \left[-\sum_{j=1}^{m} w_j \left[\max \{ \ell_j(\mathbf{Y}, \mathbf{X}), 0 \} \right]^{\{1,2\}} \right]$$

Conditional random field

Feature functions are <u>hinge-loss functions</u>

Bach et al. (2015), arXiv

Hinge-loss Markov random fields

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(w, \mathbf{X})} \exp\left[-\sum_{j=1}^{m} w_j \left[\max\left\{\ell_j(\mathbf{Y}, \mathbf{X}), 0\right\}\right]^{\{1,2\}}\right]$$

Conditional random field

Feature function for each instantiated rule

Bach et al. (2015), arXiv

Hinge-loss Markov random fields

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(w, \mathbf{X})} \exp\left[-\sum_{j=1}^{m} w_j \left[\max\{\ell_j(\mathbf{Y}, \mathbf{X}), 0\}\right]^{\{1,2\}}\right]$$

Conditional random field

5.0: *Causes*(X_1, X_2) ^ *Causes*(X_2, X_4) ^ *Linked*(X_1,X_4) → *Causes*(X_1, X_4)

Bach et al. (2015), arXiv

Hinge-loss Markov random fields

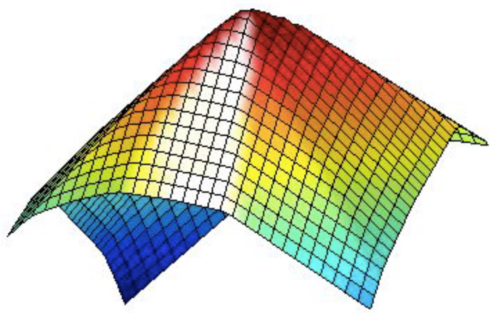
$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(w, \mathbf{X})} \exp \left[-\sum_{j=1}^{m} w_j \left[\max \{ \ell_j(\mathbf{Y}, \mathbf{X}), 0 \} \right]^{\{1,2\}} \right]$$

Conditional random field

MAP Inference Intuition: minimize distances to satisfaction!

Bach et al. (2015), arXiv

Fast Inference in Hinge-loss MRFs



Convex, continuous inference objective…

Convex optimization!

- Solved using efficient, message-passing algorithm called Alternating Direction Method of Multipliers
- Algorithms for weight learning and reasoning with latent variables

Bach et al. (2015), arXiv
Open source software: https://psl.umiacs.umd.edu

Encoding PC Algorithm with PSL

- PC Algorithm:
  - No latent variables and confounders
  - Constraint-based approach
- PC with PSL:
  - Use all independence tests
  - All rule weights set to 1.0

PSL Causal Structure Discovery

$$\textsc{Independent}(A, B, SepSet) \rightarrow \neg\textsc{Adj}(A, B)$$

Multiple independence tests with various separation sets

No early pruning!

PSL Causal Structure Discovery

$$\text{Adj}(A,B) \wedge \text{Adj}(B,C) \wedge \text{NotInSepSet}(B,A,C) \rightarrow \text{Causes}(A,B)$$
$$\text{Adj}(A,B) \wedge \text{Adj}(B,C) \wedge \text{NotInSepSet}(B,A,C) \rightarrow \text{Causes}(C,B)$$

Colliders in triples using d-separation

PSL Causal Structure Discovery

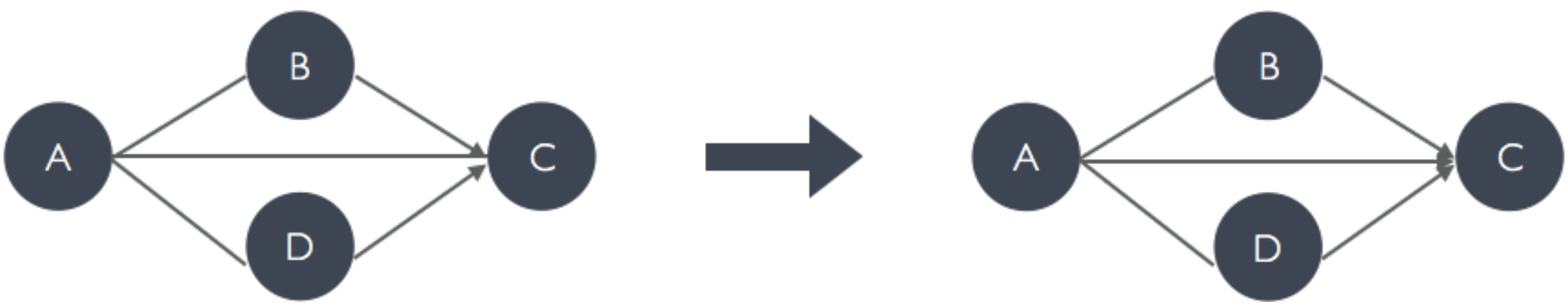
$$\textsc{Adj}(A, B) \wedge \textsc{Causes}(A, C) \wedge \textsc{Causes}(C, B) \rightarrow \textsc{Causes}(A, B)$$

PSL Causal Structure Discovery

$$\text{Adj}(A, B) \land \text{Causes}(A, C) \land \text{Causes}(C, B) \rightarrow \text{Causes}(A,B)$$

PSL Causal Structure Discovery

$$\textsc{Adj}(A, B) \wedge \textsc{Adj}(A, C) \wedge \textsc{Causes}(C, B) \wedge \textsc{Adj}(A, D) \wedge \textsc{Causes}(D, B) \wedge \neg\, \textsc{Adj}(C, D) \rightarrow \textsc{Causes}(A,B)$$

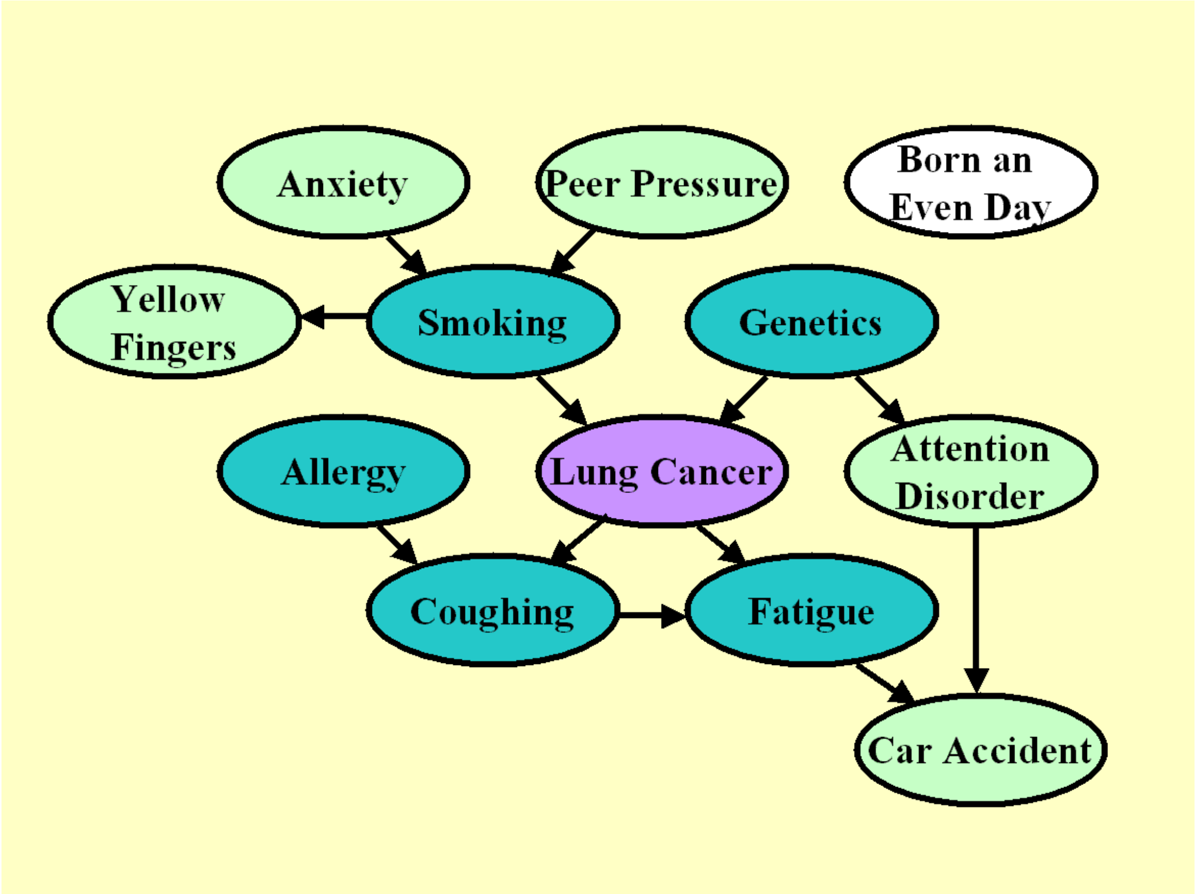


Outline

- Motivation
- Problem Formulation
- Our Approach
- **Preliminary Results**

Evaluation Dataset

Synthetic Causal DAG Dataset – 2000 examples



Causality Challenge: http://www.causality.inf.ethz.ch/data/LUCAS.html

Evaluation

- Experimental setup:
  - G^2 Independence Tests for both PC and PSL
  - Max separation set of size 3

- Evaluation details
  - Run PC and PC-PSL algorithms and compare to causal ground truth
  - For PSL, round with threshold selected by cross-validation on causal edges

Causal Edge Prediction Results

Average causal edge prediction accuracy and F1 score on 3-fold cross validation

	Accuracy	F1 Score
PC Algorithm	0.91 ± 0.06	0.53 ± 0.26
PC-PSL	0.94 ± 0.02	0.58 ± 0.19

Summary and Future Directions

- Joint inference of causal structure using probabilistic, soft constraints
- Incorporate prior and domain knowledge for causal edges from text-mining, ontological constraints, and variable selection methods
- Extensive, cross-validation experiments on multiple datasets