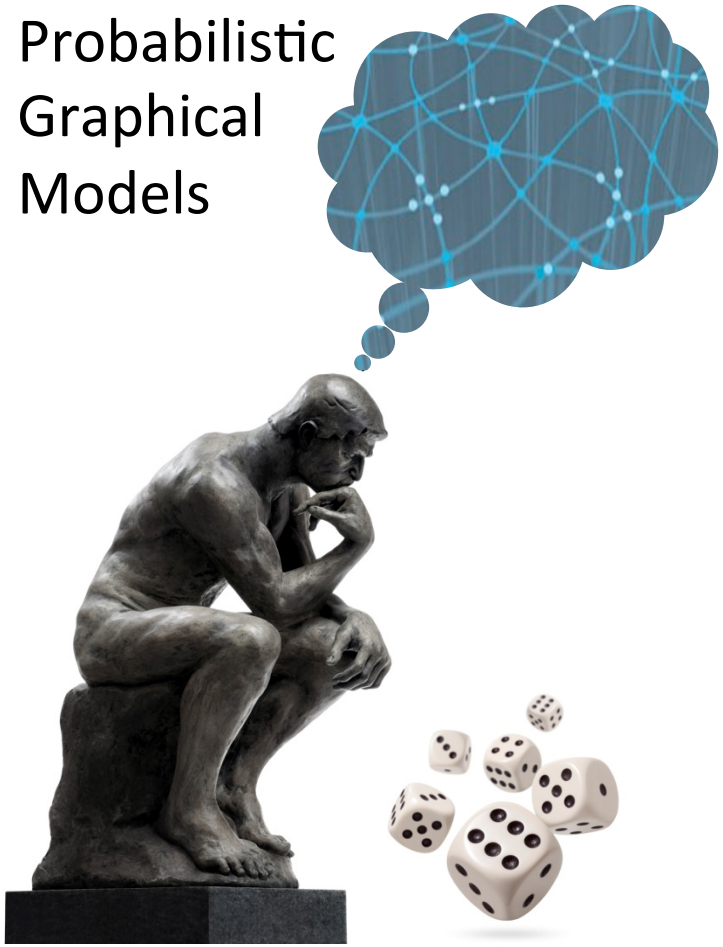
Probabilistic Graphical Models



Learning

BN Structure

General Graphs: Search

Optimization Problem

Input:

- Training data
- Scoring function
- Set of possible structures

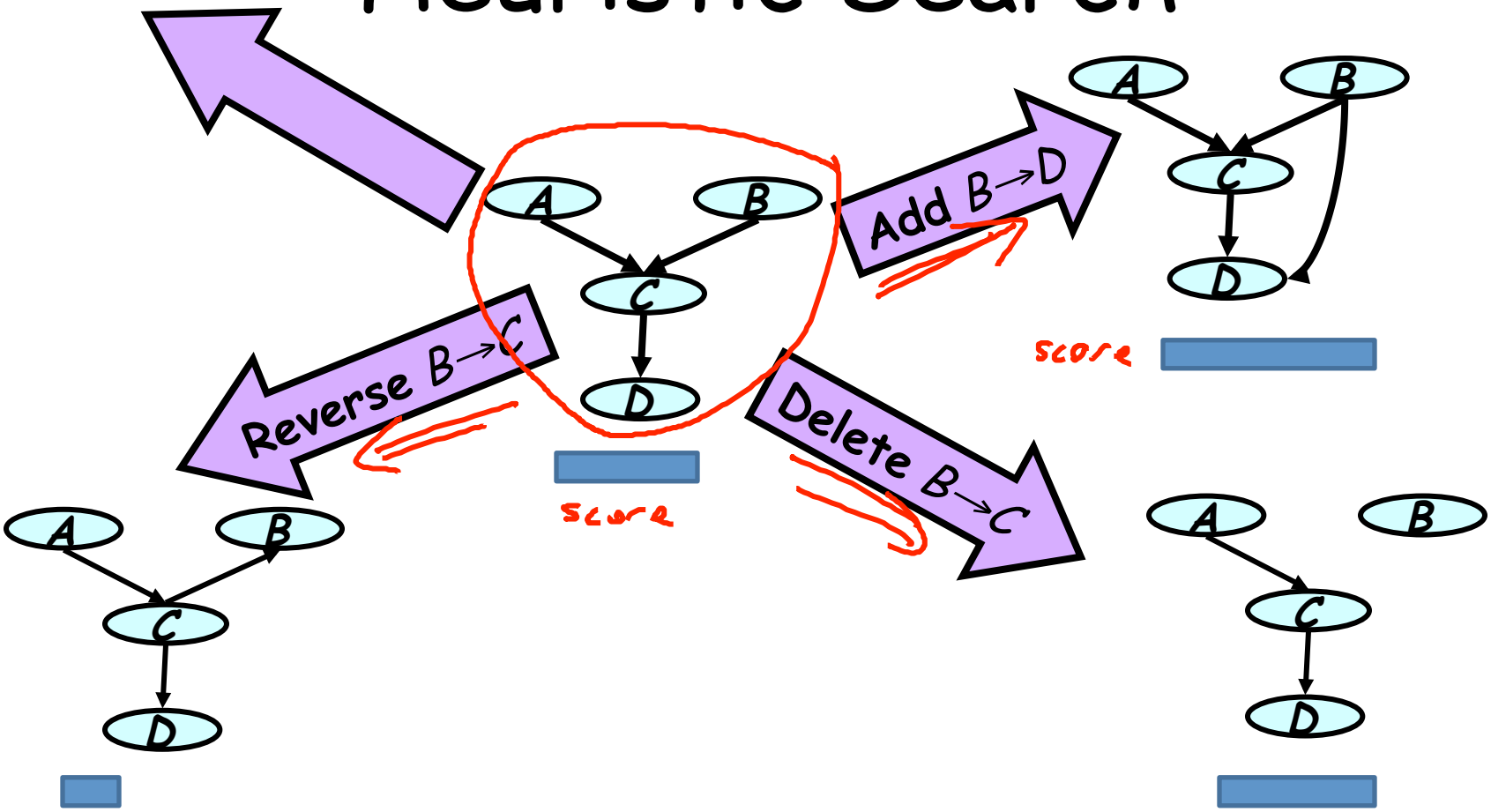
Output: A network that maximizes the score

Beyond Trees

- Problem is not obvious for general networks
  - Example: Allowing two parents, greedy algorithm is no longer guaranteed to find the optimal network

- Theorem:
  - Finding maximal scoring network structure with at most *k* parents for each variable is NP-hard for *k>1*

Heuristic Search



Heuristic Search

- Search operators:
  - local steps: edge addition, deletion, reversal
  - global steps

- Search techniques:
  - Greedy hill-climbing
  - Best first search
  - Simulated Annealing
  - ...

Search: Greedy Hill Climbing

- Start with a given network
  - empty network
  - best tree
  - a random network
  - prior knowledge
- At each iteration
  - Consider score for all possible changes
  - Apply change that most improves the score
- Stop when no modification improves score

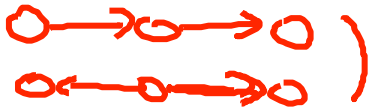
greedy

local max.

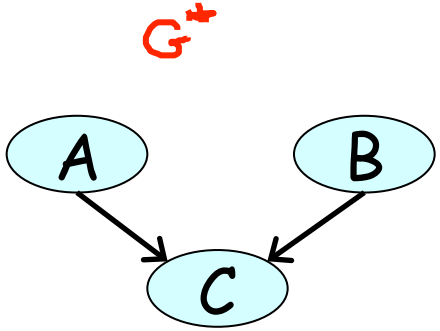
Greedy Hill Climbing Pitfalls

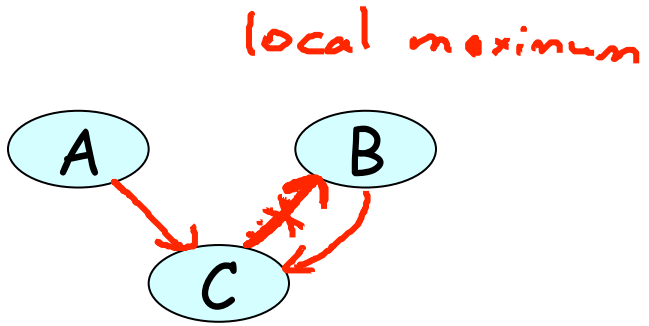
- Greedy hill-climbing can get stuck in:
  - Local maxima

  - Plateaux
    - Typically because equivalent networks are often neighbors in the search space



Why Edge Reversal

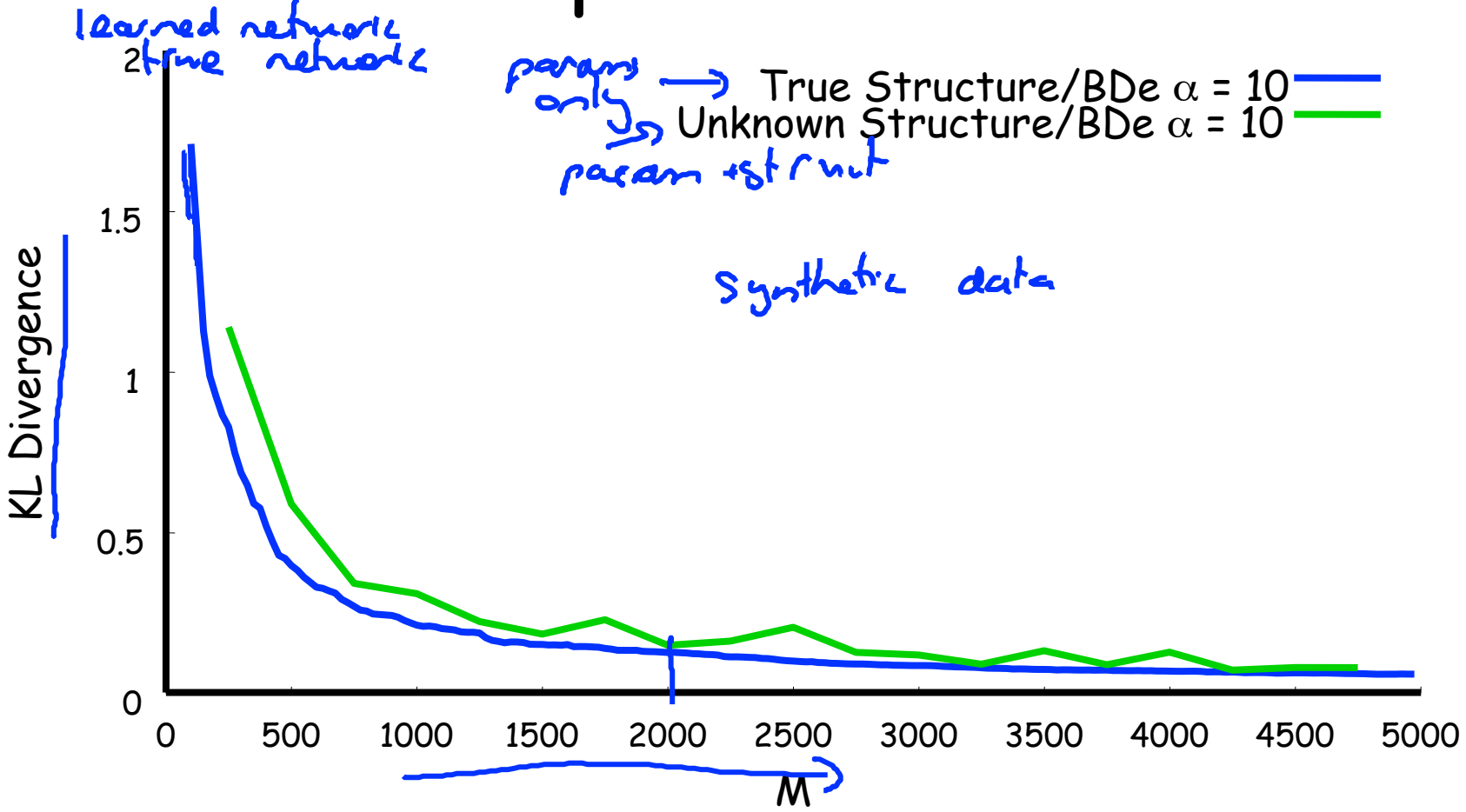




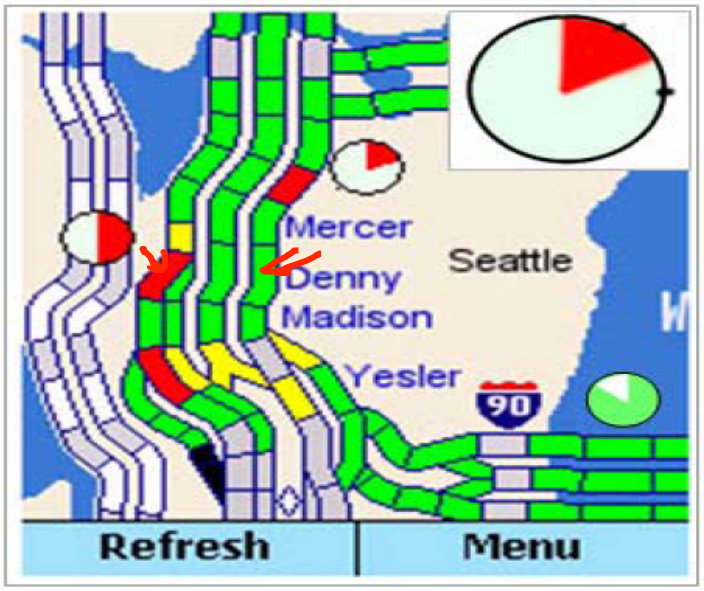
A Pretty Good, Simple Algorithm

- Greedy hill-climbing, augmented with:
- Random restarts:
  - When we get stuck, take some number of random steps and then start climbing again
- Tabu list:
  - Keep a list of K steps most recently taken
  - Search cannot reverse any of these steps

Example: ICU-Alarm

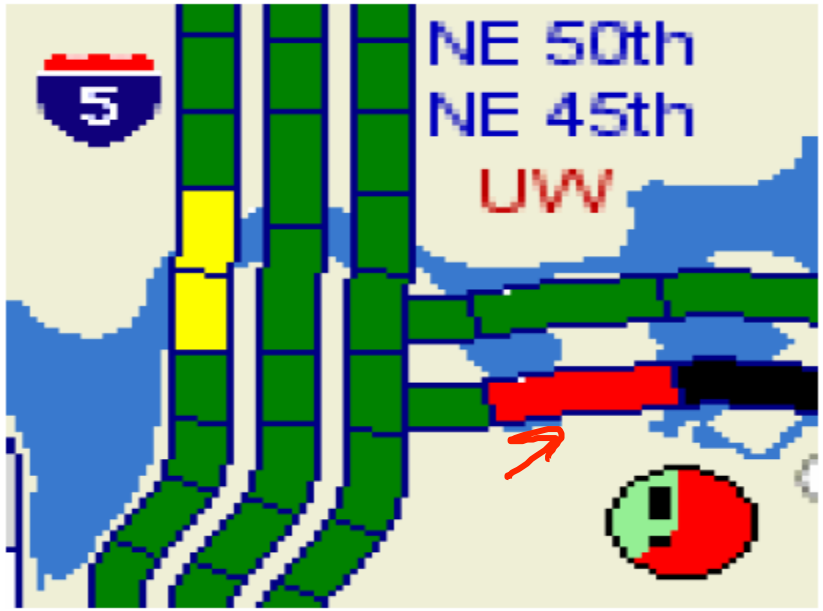


JamBayes



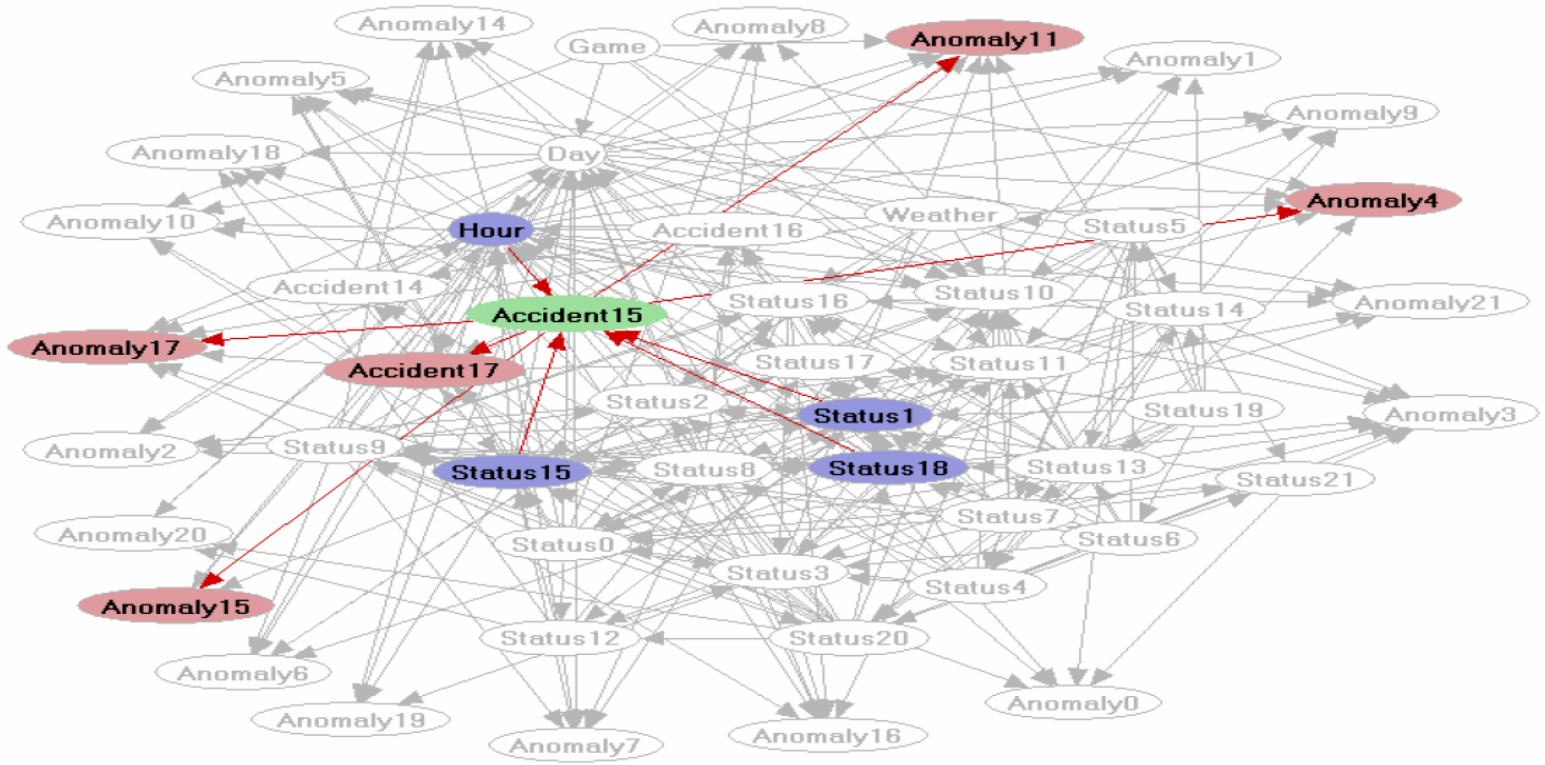
Horvitz, Apacible, Sarin, & Liao, UAI 2005

Predicting Surprises



Horvitz, Apacible, Sarin, & Liao, UAI 2005

Learned Model

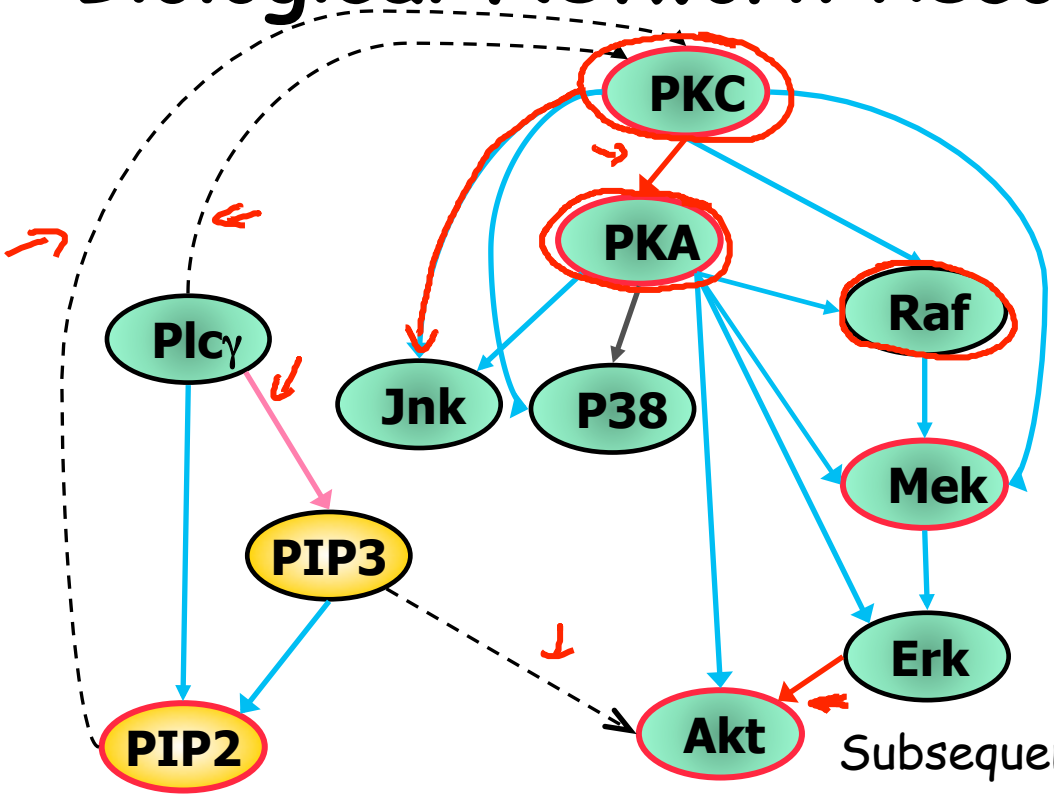


Horvitz, Apacible, Sarin, & Liao, UAI 2005

Influences in Learned Model

Horvitz, Apacible, Sarin, & Liao, UAI 2005

Biological Network Reconstruction



Phospho-Proteins
Phospho-Lipids
Perturbed in data

Known	———	15/17
Supported	———	2/17
Reversed	———	1
Missed	- - -	3

Subsequently validated in wetlab

From "Causal protein-signaling networks derived from multiparameter single-cell data"
Sachs et al., *Science* 308:523, 2005. **Reprinted with permission from AAAS.**

Daphne Koller

Summary

- Useful for building better predictive models:
  - when domain experts don't know the structure
  - for knowledge discovery
- Finding highest-scoring structure is NP-hard
- Typically solved using simple heuristic search
  - local steps: edge addition, deletion, reversal
  - hill-climbing with tabu lists and random restarts
- But there are better algorithms