# Learning Bayesian Network Structure by Self-Generating Prior Information: The Two-Step Clustering-Based Strategy

Yikun Zhang[1]

[1] *School of Mathematics, Sun Yat-Sen University, China*

Joint work with Yang Liu[2] and Jiming Liu[2].

[2] *Department of Computer Science, Hong Kong Baptist University*

W3PHIAI-2018

# Bayesian Networks

- Probabilistic graphical models
- Annotated directed acyclic graphes
- Address uncertainty and causal relationships

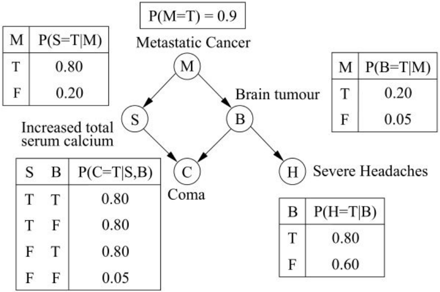



Figure: A simple Bayesian network model for the metastatic cancer problem: structure and CPTs (Twardy et al., 2006)

# Bayesian Networks and Health Intelligence

- Meningitis epidemic outbreaks modeling (Beresniak et al., 2011)
- Survival prediction and treatment selection in lung cancer (Sesen et al., 2013)
- ...

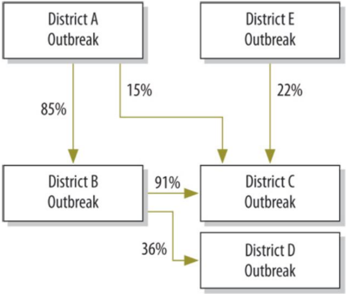



Figure: Simple Bayesian network model for the occurrence of a meningitis epidemic (Beresniak et al., 2011)

# Bayesian Network Structure Learning

Identify a network that uncovers **conditional independence relations** (or **cause-effect relationships**) among the variables given the data set

Existing structure learning algorithms:

- Constraint-based algorithms: Markov Property
- Score-based algorithms: statistically motivated score
- Hybrid algorithms

# Current Problems

For constraint-based methods,

- Sensitive to the failures in (conditional) independence tests

For score-based methods,

- NP-hard
- Heuristic searching algorithms may get stuck in a local maximum

# Current Problems

For constraint-based methods,

- Sensitive to the failures in (conditional) independence tests

For score-based methods,

- NP-hard
- Heuristic searching algorithms may get stuck in a local maximum

Time-consuming and less accurate on large-scale data sets!!!

# Possible Solution

One feasible approach to address these problems:
**Prior information**

# Possible Solution

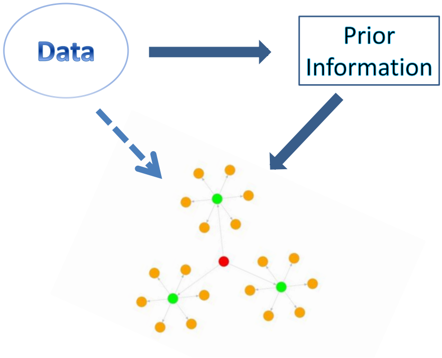
One feasible approach to address these problems:
**Prior information**

However, under real-world scenarios,

1. Expert knowledge may be scarce
2. A structure, an ordering of nodes, or distribution knowledge of nodes and arcs cannot be specified

# Our Expectation

1. Generate prior information (existence of arcs) from data
2. Improve time efficiency, accuracies, or both



# Inspirations

1. Cluster-tree decomposition in the “Sparse Candidate” algorithm (Friedman et al., 1999)
2. Dividing the super-structure (a pre-assumed skeleton for the network) into clusters (Kojima et al., 2010)

# Inspirations

1. Cluster-tree decomposition in the “Sparse Candidate” algorithm (Friedman et al., 1999)
2. Dividing the super-structure (a pre-assumed skeleton for the network) into clusters (Kojima et al., 2010)

Why not group similar variables into clusters and learn the within-cluster and between-cluster arcs in two steps?

# Outline of the Algorithm (Step 1)

***Two-Step Clustering-Based*** **Bayesian Network Structure Learning Strategy**

- Data set $\mathcal{D} = \{X_1, X_2, ..., X_N\}$ with $N$ variables
- The number of clusters: $K$ (Parameter)

**Step 1:**

1: Compute the dissimilarity matrix.
2: Carry out clustering analysis via *average linkage agglomerative clustering method* and cut the dendrogram into $K$ groups (clusters).
3: Learn Bayesian network structures within each cluster using a traditional algorithm $A$[1].

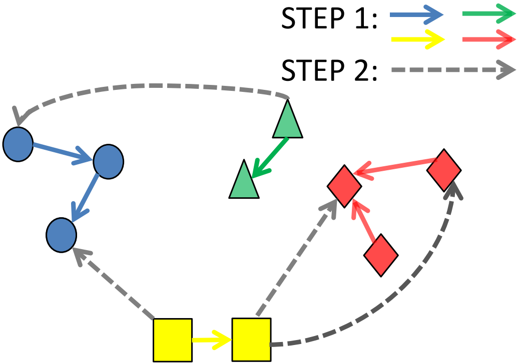
[1] This could be any traditional structure learning algorithm, like the Grow-Shrink algorithm (Margaritis, 2003).

# Outline of the Algorithm (Step 2)

**Step 2:**

1: Apply the algorithm $A$ again on all variables with the retained arcs to combine clusters.

**Output:** Bayesian network structure learned from the data set $\mathcal{D}$.



# Dissimilarity Metric

Data sets with only discrete variables:

- Negative *mutual information*

Data sets with only continuous variables:

- *1-(Pearson's) correlation*

Hybrid data sets: See our paper for details.

# Accuracy and Time Efficiency Evaluation

Accuracy metric (Metz, 1978):

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Negative}{\sum Total\ Population}$$

Time efficiency metric:

Average elapsed times of repeated experiments

# Traditional Structure Learning Algorithms

Constraint-based:

- Grow-Shrink (GS) algorithm (Margaritis, 2003)
- Incremental Association Markov Blanket (IAMB) algorithm (Tsamardinos et al., 2003a)
- Interleaved Incremental Association (Inter-IAMB) algorithm (Yaramakala and Margaritis, 2005)

Score-based:

- Hill-Climbing (HC) algorithm
- Tabu greedy search (TABU) algorithm

Hybrid:

- Max-Min Parents and Children (MMPC) algorithm (Tsamardinos et al., 2003b)

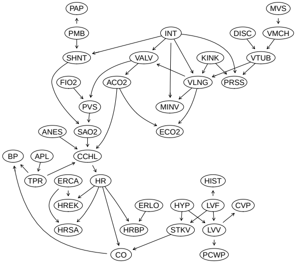
# Accuracy Comparisons (I)

| Methods | “asia” | “insurance” | “alarm” | “hepar2” |
|---|---|---|---|---|
| GS | **0.9096** (0.8918) | **0.9309** (0.9263) | **0.9662** (0.9602) | **0.9763** (0.9753) |
| IAMB | **0.9084** (0.8896) | **0.9287** (0.9218) | **0.9715** (0.9686) | **0.9747** (0.9741) |
| Inter-IAMB | **0.9082** (0.8936) | **0.9281** (0.9208) | **0.9716** (0.9689) | **0.9748** (0.9742) |
| MMPC | **0.8557** (0.8546) | **0.9259** (**0.9259**) | **0.9649** (0.9646) | **0.9732** (0.9728) |
| HC | **0.9766** (**0.9766**) | **0.9328** (0.9293) | **0.9768** (0.9724) | **0.9824** (0.9822) |
| TABU | **0.9664** (0.9657) | **0.9422** (0.9312) | **0.9788** (0.9744) | **0.9814** (0.9810) |

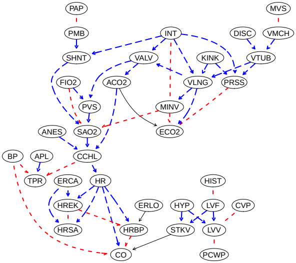
Table: TSCB Strategy vs Embedded Traditional Algorithms (inside round brackets).

# Accuracy Comparisons (II)

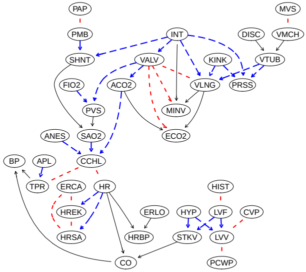
Network Configurations of the “alarm” Data Set.




(a) Actual Network




(b) Learned by the GS Algorithm




(c) Learned by the TSCB Strategy

Red dashed line: *False Positive*; Blue dashed line: *False Negative*.

# Time Comparisons

| Mean Elapsed Times / s | "asia" | "insurance" | "alarm" | "hepar2" |
|---|---|---|---|---|
| Clustering | 0.00230 | 0.00788 | 0.01076 | 0.04432 |
| Within clusters | 0.00464 | 0.01670 | 0.05012 | 0.04744 |
| Between clusters | 0.00962 | 0.16420 | 0.24640 | 1.46168 |
| TSCB | 0.01656 | **0.18878** | **0.30728** | **1.55344** |
| Traditional | **0.01010** | 0.19362 | 0.35900 | 1.65584 |

Table: Mean Elapsed Times Comparison.

Sampling repeated times: $50$; Recording repeated times: $10$; Experiments repeated times: $50 \times 10 = \mathbf{500}$.

# Contributions

- An automatic way to generate prior information from data
- A wide range of Bayesian network structure learning algorithms can be improved
  1. Accuracy
  2. Time efficiency

# Future work

1. Small clusters ($\leq 3$ variables) $\sim$ "network motifs" (Milo et al., 2002) ?

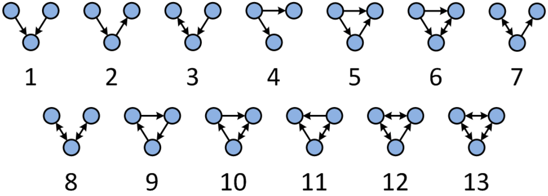



Figure: All 13 types of three-node connected subgraphs (Milo et al., 2002)

2. Existence of latent variables $\Rightarrow$ TSCB strategy ?
3. . . .

# References I


Charles R. Twardy, Ann E. Nicholson, Kevin B. Korb, John McNeil (2006)
Epidemiological Data Mining of Cardiovascular Bayesian Networks
*e-Journal of Health Informatics*, 1(1): e3

A. Beresniak, E. Bertherat, W. Perea, G. Soga, R. Souley, D. Dupont, and S. Hugonnet (2012)
A Bayesian Network Approach to the Study of Historical Epidemiological Databases: Modelling Meningitis Outbreaks in the Niger
*Bulletin of the World Health Organization*, 90:412-417A.

M. Berkan Sesen, Ann E. Nicholson, Rene Banares-Alcantara, Timor Kadir, Michael Brady (2013)
Bayesian Networks for Clinical Decision Support in Lung Cancer Care
*PLOS ONE*, 8(12): e82349.

C. Metz (1978)
Basic Principles of ROC Analysis
*Seminars in Nuclear Medicine*, 8(4).

Dimitris Margaritis (2003)
Learning Bayesian Network Model Structure from Data
*Ph.D. Dissertation*, CMU-CS-03-153, Pittsburgh, USA.

# References II

Ioannis Tsamardinos, Constantin Aliferis, Alexander Statnikov, and Er Statnikov (2003a)
Algorithms for Large Scale Markov Blanket Discovery
In *Proceedings of The Sixteenth International FLAIRS Conference, St*, 376-380. AAAI Press.

Sandeep Yaramakala and Dimitris Margaritis (2003)
Speculative Markov Blanket Discovery for Optimal Feature Selection
In *Proceedings of the Fifth IEEE International Conference on Data Mining*, ICDM '05, 809-812. TX, USA: IEEE Computer Society

Ioannis Tsamardinos, Constantin Aliferis, and Alexander Statnikov (2003b)
Time and Sample Efficient Discovery of Markov Blankets and Direct Causal Relations
In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, 673-678. New York, USA: ACM.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon (2002)
Network Motifs: Simple Building Blocks of Complex Networks
*Science*, 298(5594): 824-827.

# Thank you!

## Suggestions, Comments, Questions?

Contact: yikunzhang@foxmail.com