Learning Dynamic Bayesian Networks: Algorithms and Issues

Alex Black, Kevin Korb, Ann Nicholson

Clayton School of Information Technology, Monash University

Introduction

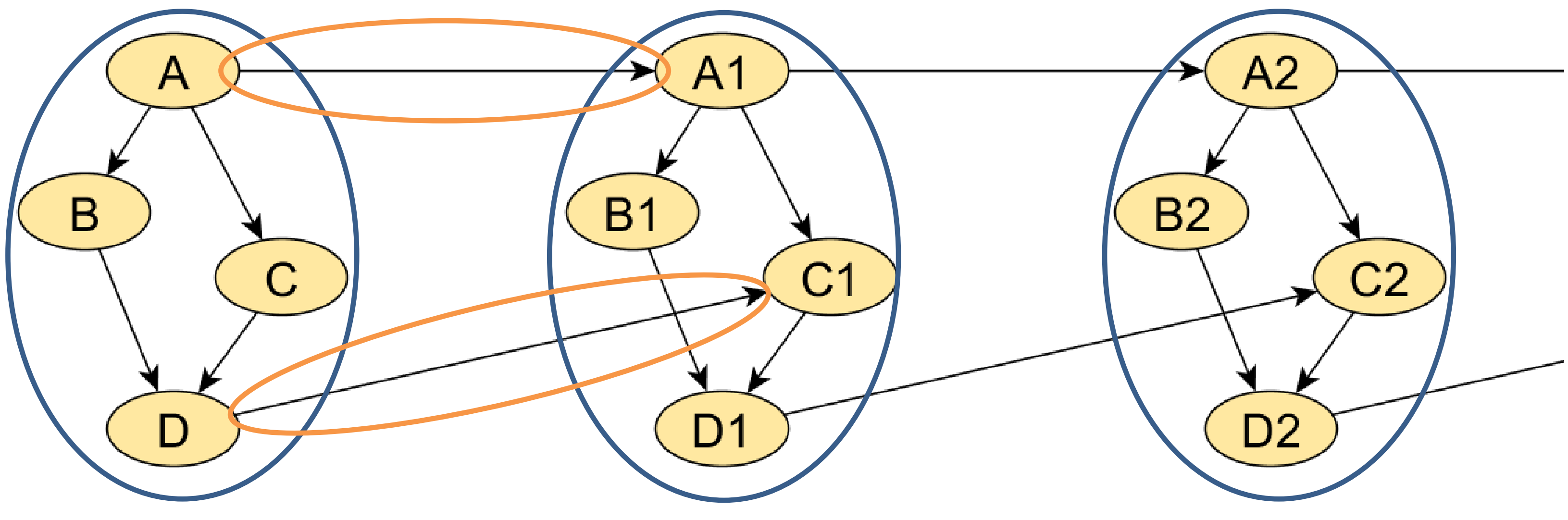
- Part 1: Background
  - Dynamic Bayesian Networks
  - DBN Learning
- Part 2: Our Research
  - CaMML DBN Learning
  - Experimental Evaluation

Dynamic Bayesian Networks

- Extension of BNs to time domain
- Why DBNs?
  - Temporal aspect to data or underlying process
  - Prediction over time

Dynamic Bayesian Networks

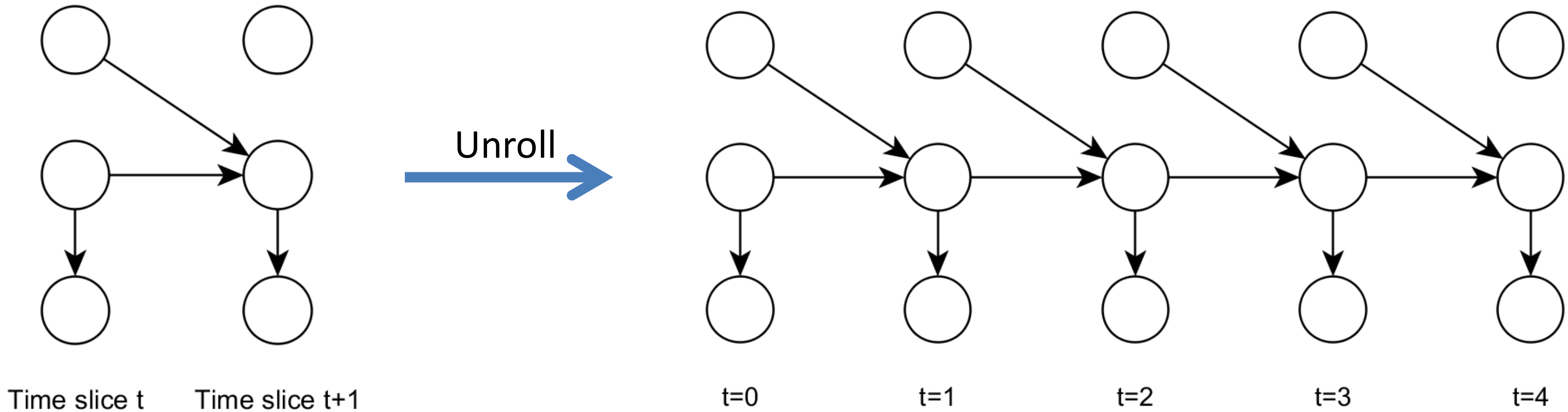
- Structure: same for all time steps
- Arcs span at most one time step
- Discrete data



Time = 0 Time = 1 Time = 2

2 Time Slice DBN

- Simplifies learning and representation
  - Learn t=0 and t=T+1 arcs/parameters
- For prediction
  - Unroll DBN



Learning Dynamic Bayesian Networks

Two learning approaches:

- Constraint Based
  - MIT, PC
  - Uses conditional independence tests

- Metric based - ‘Search and Score’
  - BIC, BDe, CaMML
  - Sample model space, score whole network

Learning DBNs

- BIC – Bayesian Information Criterion
- BDe
- MIT – Mutual Information Test
- CaMML – Causal Minimum Message Length

Static BN learners can be used

Implementations

- Research software
  - Variety of languages, input and output formats
- Issues: Maintenance? Documentation? GUI? Limitations?

Using Static BN Learners for DBNs

- Create new data set: Duplicate variables, offset by 1
  - Use tier prior constraints if available. (PC/Tetrad, CaMML)

Time	X	Y	Z
0	true	yes	high
1	false	yes	low
2	false	no	medium
3	false	no	low
4	true	yes	medium
...	...	...	...
N-2	false	yes	medium
N-1	true	yes	high

→

Time	X0	Y0	Z0	X1	Y1	Z1
0	true	yes	high	true	yes	high
1	false	yes	low	false	yes	low
2	false	no	medium	false	no	medium
3	false	no	low	false	no	low
4	true	yes	medium	true	yes	medium
...	...	...	...	...	...	...
N-2	false	yes	medium	false	yes	medium

- Why have separate DBN learners?
  - Computational speed
  - Improved results

PART 2: RESEARCH

Our Research: CaMML DBN Learning

CaMML – Causal Minimum Message Length

- Capable BN learner
- Extended for learning DBNs

Minimum Message Length:

- Information theoretic approach to statistical and inductive inference.

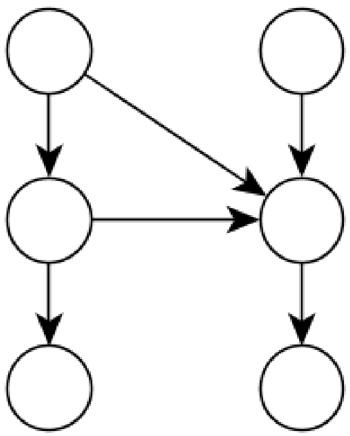
DBN Learning Algorithms

Q: How well do they work in practice?

Algorithm	Software	Notes
CaMML – DBN Learner	CaMML	
CaMML – BN Learner	CaMML	BN learner with tier prior
MIT	GlobalMIT	No intraslice arcs
BDe	Banjo	No parameters
BIC (2-Step)	Bayes Net Toolbox	Learn interslice + interslice arcs separately. Not ideal. No complete BIC implementation available.
PC	Tetrad	BN learner with tier prior

Experimental Design

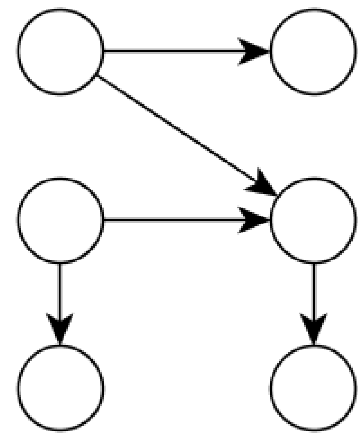
Goal: Evaluate performance of DBN learning algorithms



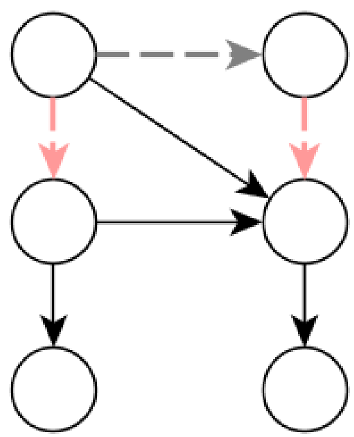
Step 1: Start with existing DBN

Time	X	Y	Z
0	true	yes	high
1	false	yes	low
2	false	no	medium
3	false	no	low
4	true	yes	medium
...	...	...	...
N	false	yes	medium

Step 2: Generate Data



Step 3: Re-learn DBN using each algorithm



Step 4: Compare

Our Experiments:
7 DBNs with
6 to 28 variables

Evaluating BNs/DBNs

- Edit Distance
  - Score 1 for every arc error
- Kullback-Leibler (KL) Divergence
- Causal Kullback-Leibler (CKL) Divergence

KL Divergence & CKL

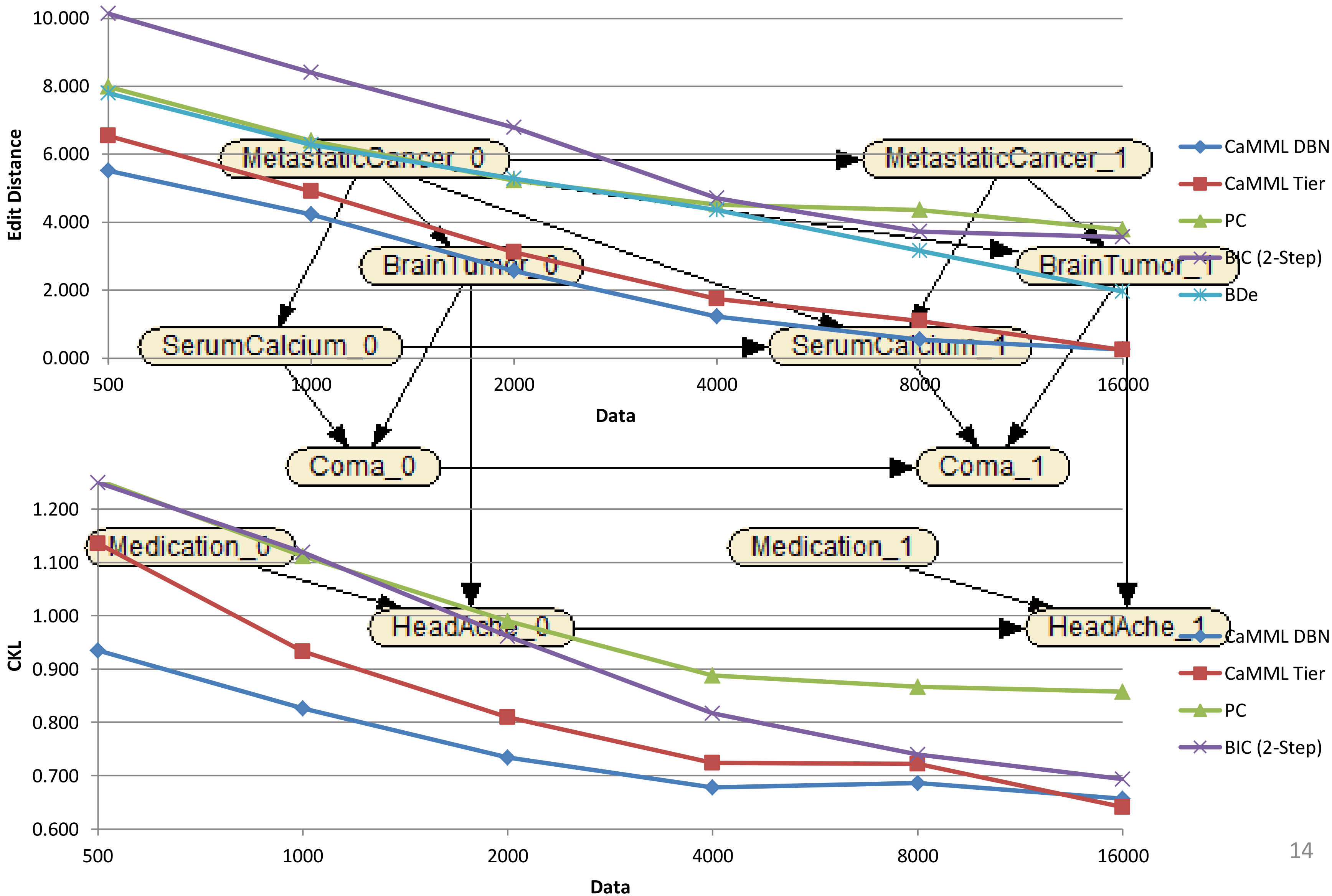
- Kullback-Leiber Divergence

$$D_{KL}(P||Q) = \sum_i P(x_i) \log_{\mathrm{e}} \left(\frac{P(x_i)}{Q(x_i)} \right)$$

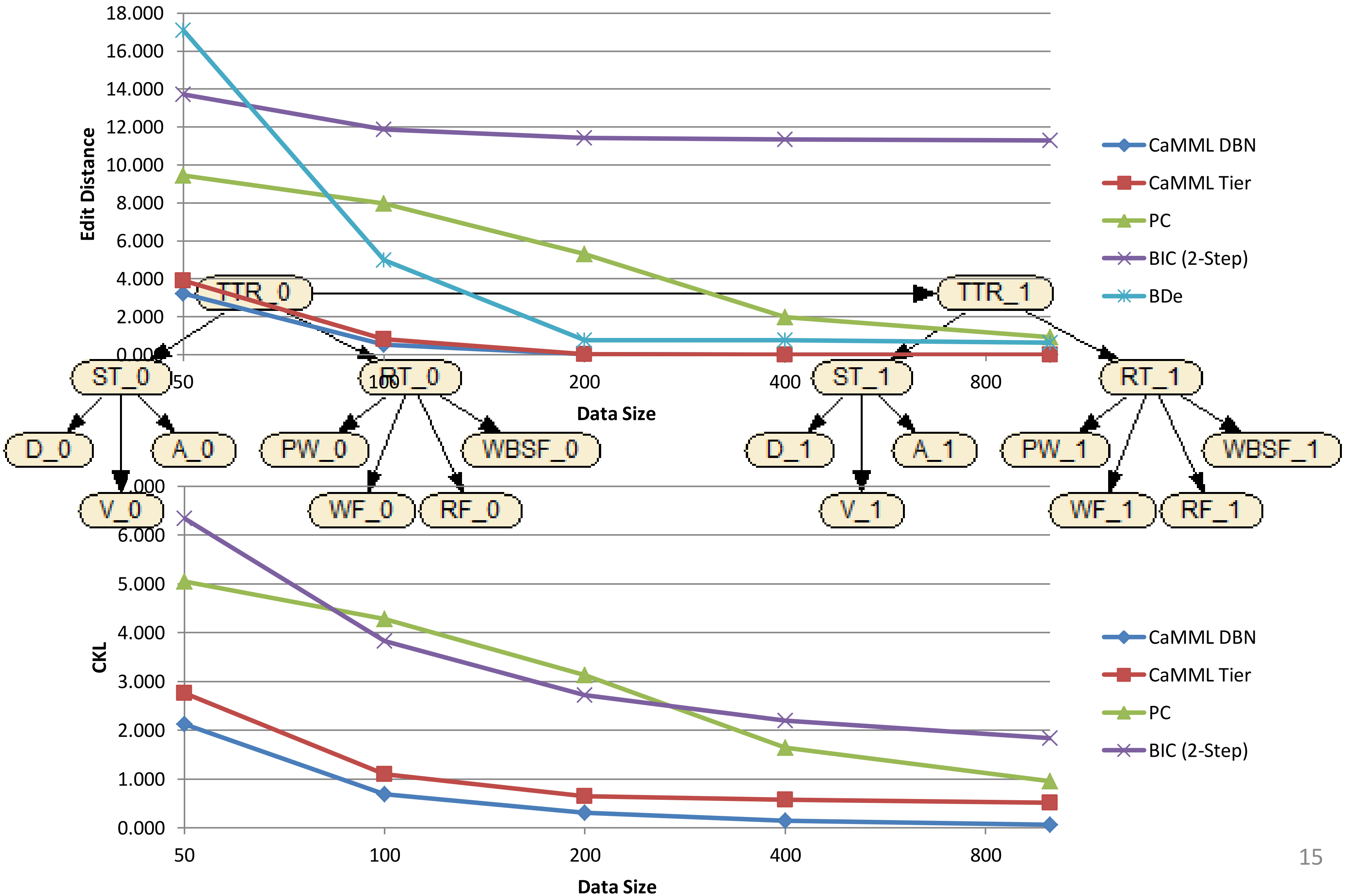
  - Difference between probability distributions
  - But: Ignores network structure

- Causal Kullback-Leibler Divergence
  - Both structure and probability

Results – Metastatic Cancer DBN



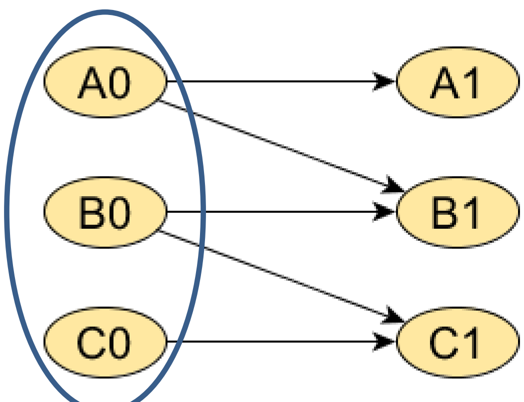
Results – Threat DBN



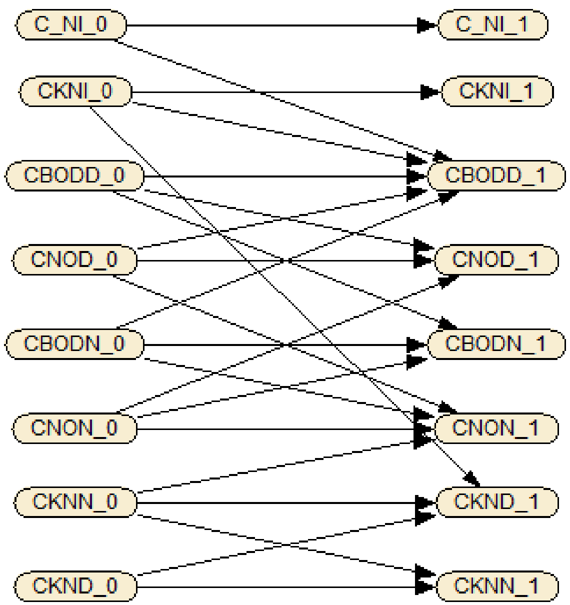
DBNs: No Arcs Within Time Slice

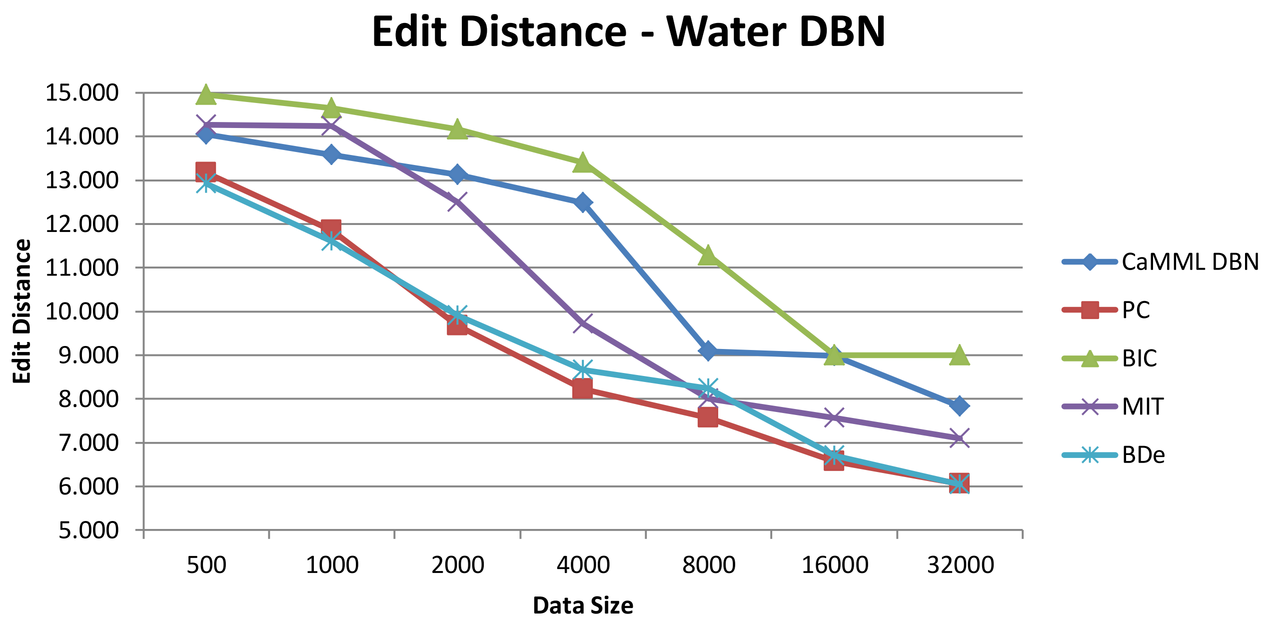
- Some learners: no arcs within time slice
  - MIT, some BIC and BDe software
  - Common in bioinformatics

- Assumption holds: Good results, fast
- Assumption violated: Poor results



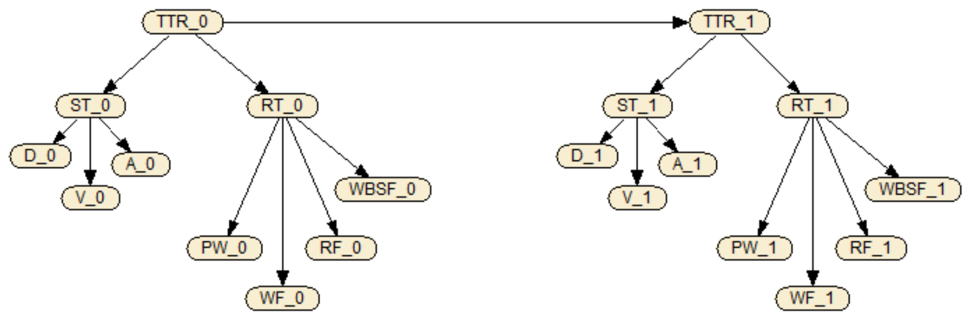
Learning DBNs: No Arcs Within Time Slices



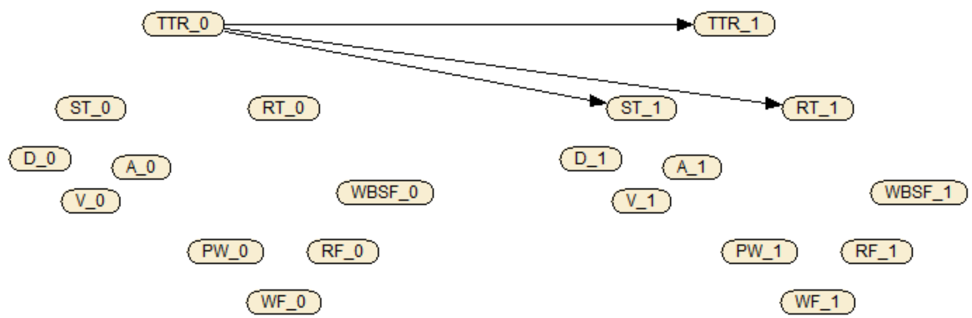


No Intraslice Arcs: Invalid Assumption

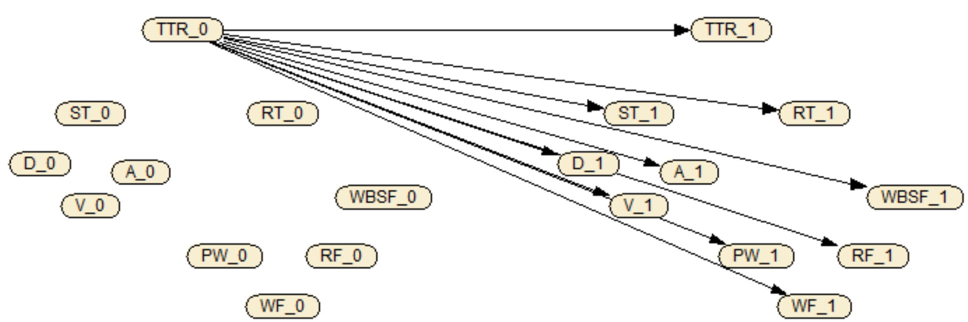
True Network



MIT, BDe
data = 100 obs.



BIC, MIT, BDe
data = 1000 obs.



Summary of DBN Learning Results

- CaMML DBN learner
  - Better performance than other algorithms
  - Conservative: Errors mostly due to missing arcs

- BDe: Can overfit with small data sets.

- No arcs within time slice:
  - Some BIC/BDe implementations may be much faster
  - Poor results if assumption does not hold

Final Thoughts

- CaMML

  Supports BN and DBN learning

  bayesian-intelligence.com/software/

- Paper on CaMML / DBN Learning – 2014
- Practical applications of CaMML DBN underway
- Learning DBNs? Time series data? Talk to us.

DBN Structure Learning Software

Name	Structure Learning	Parameter Learning	DBN Algorithms	GUI	URL
CaMML	Yes	Yes	CaMML	Yes	bayesian-intelligence.com
Bayes Net Toolbox*	Partial[1]/2-Step[1] Yes[2]	Yes	BIC[1,2], BDe[2]	No	code.google.com/p/bnt/
BNFinder	Partial[1]	?	BIC, BDe, MIT	No	bioputer.mimuw.edu.pl/software/bnf/
GlobalMIT*	Partial[1]	No	MIT	No	code.google.com/p/globalmit/
Tetrad	No/Manual[3]	Yes	(PC, others)	Yes	www.phil.cmu.edu/tetrad/
GeNIe	No/Manual[3,4]	Yes	(PC, K2, other)	Yes	genie.sis.pitt.edu
Banjo	Yes	No	BDe	No	cs.duke.edu/~amink/software/banjo/

Kevin Murphy's Bayesian Network Software List:
http://people.cs.ubc.ca/~murphyk/Software/bnsoft.html

* Requires Matlab. GlobalMIT: May be possible to use Octave (free) instead of Matlab
[1] Supports DBN learning with interslice arcs only (i.e. no arcs within time slices)
[2] With DBmcmc extension (bioss.ac.uk/~dirk/software/DBmcmc/) but binary/ternary data/attributes only
[3] No official support for learning DBNs. Can adapt BN algorithms using tier priors etc.
[4] DBN parameter learning, but no structure learning. Supports DBN inference, unrolling etc.

References

- Friedman, N., K. Murphy, and S. Russell (1998). *Learning the structure of dynamic probabilistic networks*. 14th UAI, pp. 139–147.
- Korb, K. and A. Nicholson (2011). *Bayesian Artificial Intelligence*, 2nd ed. CRC Press.
- O'Donnell, R., K. Korb, and Lloyd Allison. *Causal KL: Evaluating causal discovery*. Technical Report 2007/207, Clayton School of IT, Monash University, 2007.
- Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Springer.
- Vinh, N. et al (2011). GlobalMIT: learning globally optimal dynamic bayesian network with the mutual information test criterion. Bioinformatics, Vol. 27 no. 19, pp. 2765–2766