A Study on Performance Comparison of Structural Learning Algorithms about Synthetic Pattern of Bayesian Network

Jae-seong Yoo
Baek-woon Yu
Rae-hyun Chang

June 14, 2017

Outline

Outline

Bayesian Network

A BN defines a unique joint probability distribution over X given by

$$P_B(X_1, \cdots, X_n) = \prod_{i=1}^{n} P_B(X_i | \prod_{X_i}).$$

- A BN encodes the independence assumptions over the component random variables of X.
- An edge (j, i) in E represents a direct dependency of X_i from X_j.
- The set of all Bayesian networks with n variables is denotes by B_n.

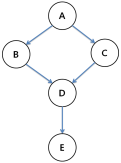
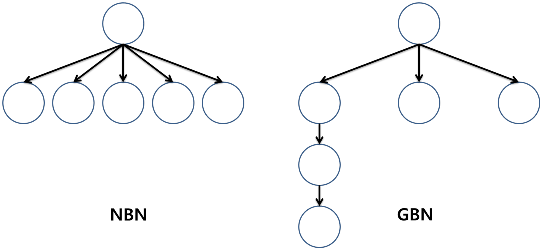


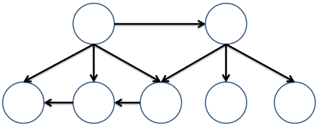
Figure: $P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|D)$

Various Types of Bayesian Network

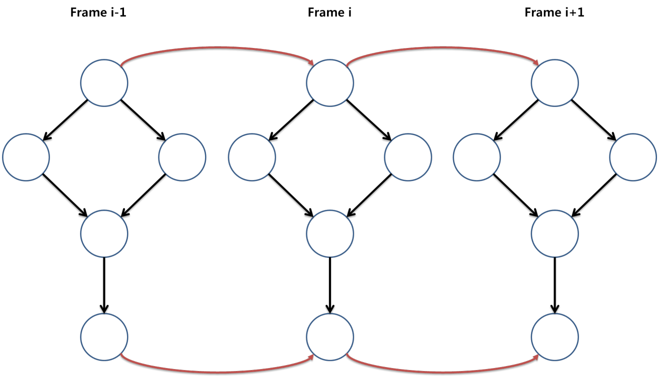
- Naive Bayesian Network (NBN)
- Generalized Bayesian Network (GBN)
- Tree-augmented Bayesian Network (TAN)
- Dynamic Bayesian Network (DBN)
- And so on $\cdots$

Various Types of Bayesian Network





Various Types of Bayesian Network



DBN

Bayesian Network Structure Learning

Learning a Bayesian network is as follows:

Given a data $T = \{y_1, \cdots, y_n\}$ and a scoring function ϕ, the problem of learning a Bayesian network is to find a Bayesian network $B \in B_n$ that maximizes the value $\phi(B, T)$.

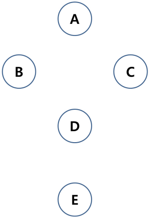


Figure: A model before learning structure

Bayesian Network Structure Learning

The number of DAGs may be computed by the recurrence relation as shown below.*

$$a_n = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} 2^{k(n-k)} a_{n-k}$$

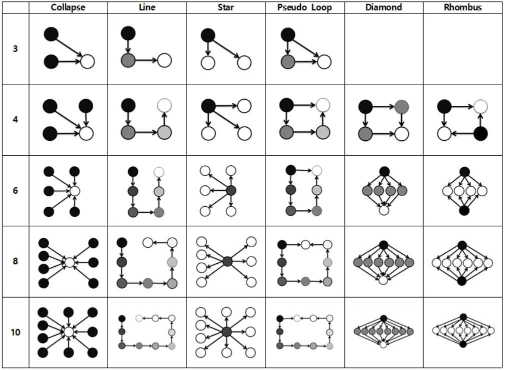
In other words, the number of DAGs on n labeled vertices, for n = 0, 1, 2, 3, $\cdots$ is 1, 1, 3, 25, 543, 29281, 3781503, $\cdots$.

These **exponentially increasing numbers makes difficulties** to check performance of each algorithms for structure learning.

* Robinson, R. W. (1973),

Counting labeled acyclic digraphs, *New Directions in the Theory of Graphs, Academic Press*, pp. 239–273.

Goal



The goal of our team is

- to compare the performance of the structure learning algorithm based on the above synthetic model*

and

- to confirm whether each algorithm learns the pattern of the synthetic model*, well.

Outline

Available constraint-based learning algorithms

Grow-Shrink (GS) based on the Grow-Shrink Markov Blanket, the first (and simplest) Markov blanket detection algorithm used in a structure learning algorithm.

Incremental Association (IAMB) based on the Markov blanket detection algorithm of the same name, which is based on a two-phase selection scheme (a forward selection followed by an attempt to remove false positives).

Available Score-based Learning Algorithms

Hill-Climbing (HC) a hill climbing greedy search on the space of the directed graphs. The optimized implementation uses score caching, score decomposability and score equivalence to reduce the number of duplicated tests.

Tabu Search (TABU) a modified hill climbing able to escape local optima by selecting a network that minimally decreases the score function.

Available Hybrid Learning Algorithms

Max-Min Hill-Climbing (MHHC) a hybrid algorithm which combines the Max-Min Parents and Children algorithm (to restrict the search space) and the Hill-Climbing algorithm (to find the optimal network structure in the restricted space).

Restricted Maximization (RSMAX2) a more general implementation of the Max-Min Hill-Climbing, which can use any combination of constraint-based and score-based algorithms.

Outline

The Number of Graphical Errors in the Learnt Structure

In terms of the number of graphical errors in the learnt structure.

		Target Network	Learnt Network	Direction
C	(Correct Arcs)	exist	exist	correct
M	(Missing Arcs)	exist	not exist	
WO	(Wrongly Oriented Arcs)	exist	exist	wrong
WC	(Wrongly Corrected Arcs)	not exist	exist	

Network Scores

In all four cases, the higher the value of the metric, the better the network.

BDe $BDe(B,T) = P(B,T) = P(B) \times \prod_{i=1}^{n} \prod_{j=1}^{q_i} (\frac{\Gamma(N'_{ij})}{\Gamma(N_{ij}+N'_{ij})} \times \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk}+N'_{ijk})}{\Gamma(N'_{ijk})})$

$$\phi(B|T) = LL(B|T) - f(N)|B|,$$

$$LL(B|T) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log(\frac{N_{ijk}}{N_{ij}}).$$

Log-Likelihood(LL) If $f(N) = 0$, we have the **LL** score.

AIC If f(N) $= 1$, we have the **AIC** scoring function:

BIC If $f(N) = \frac{1}{2}\log(N)$, we have the **BIC** score.

Outline

Asia DataSet

Description Small synthetic data set from Lauritzen and Spiegelhalter (1988) about lung diseases (tuberculosis, lung cancer or bronchitis) and visits to Asia.

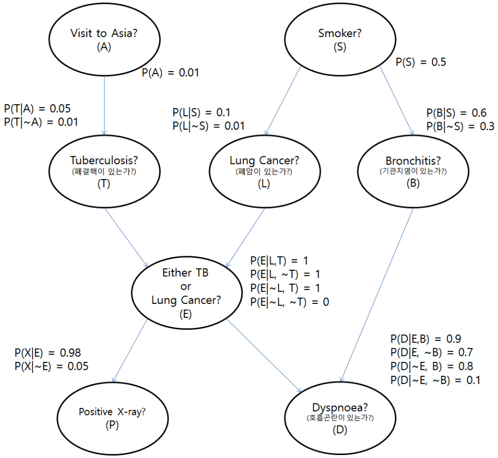
Number of nodes 8

Number of arcs 8

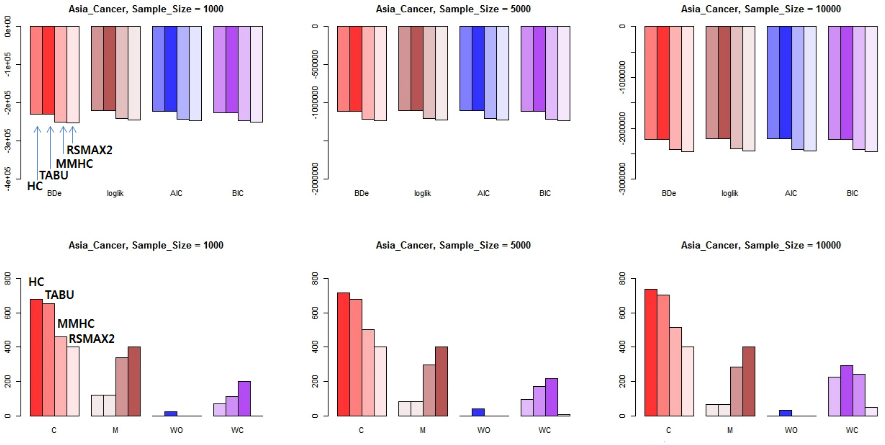
Number of parameters 18

Source Lauritzen S, Spiegelhalter D (1988).
"Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion)".
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 50(2), 157-224.

Asia DataSet



Asia DataSet

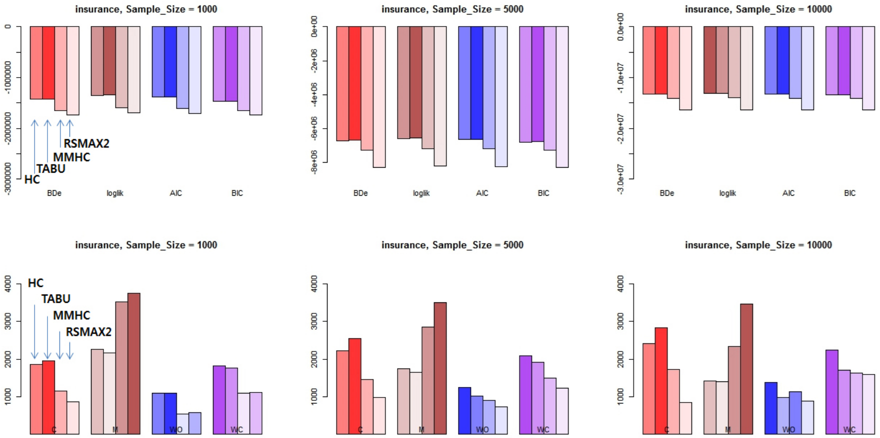


Insurance DataSet

Description	Insurance is a network for evaluating car insurance risks.
Number of nodes	27
Number of arcs	52
Number of parameters	984
Source	Binder J, Koller D, Russell S, Kanazawa K (1997). "Adaptive Probabilistic Networks with Hidden Variables". Machine Learning, 29(2-3), 213-244.

Insurance DataSet

Insurance DataSet



Alarm DataSet

Description The ALARM ("A Logical Alarm Reduction Mechanism") is a Bayesian network designed to provide an alarm message system for patient monitoring.

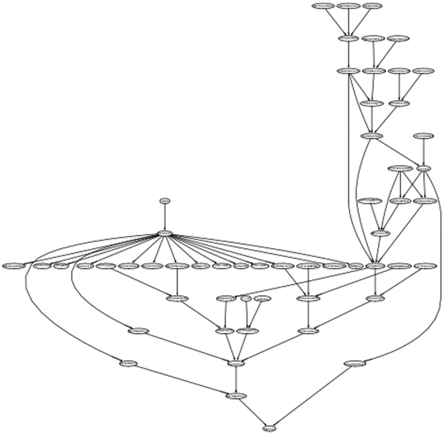
Number of nodes 37

Number of arcs 46

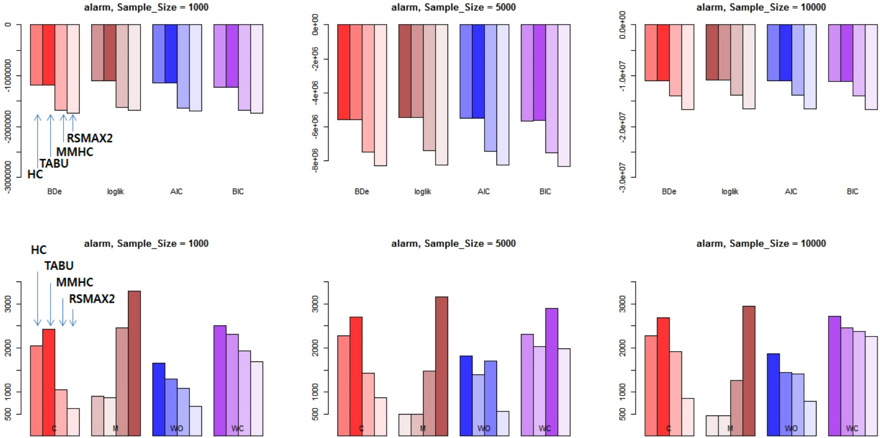
Number of parameters 509

Source Beinlich I, Suermondt HJ, Chavez RM, Cooper GF (1989).
"The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks."
In "Proceedings of the 2nd European Conference on Artificial Intelligence in Medicine", pp. 247-256. Springer-Verlag.

Alarm DataSet



Alarm DataSet



HailFinder DataSet

Description Hailfinder is a Bayesian network designed to forecast severe summer hail in northeastern Colorado.

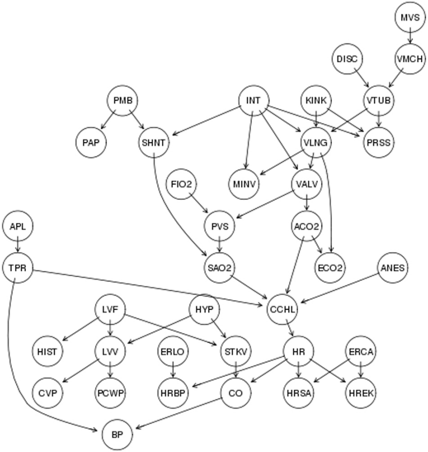
Number of nodes 56

Number of arcs 66

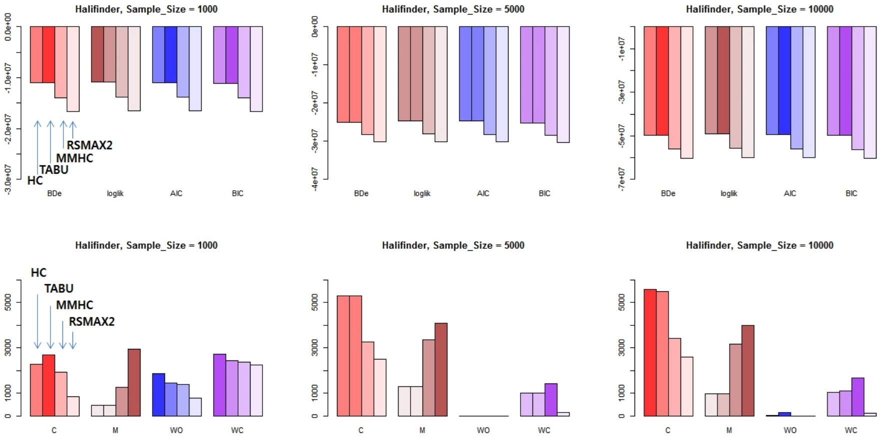
Number of parameters 2656

Source Abramson B, Brown J, Edwards W, Murphy A, Winkler RL (1996). "Hailfinder: A Bayesian system for forecasting severe weather". International Journal of Forecasting, 12(1), 57-71.

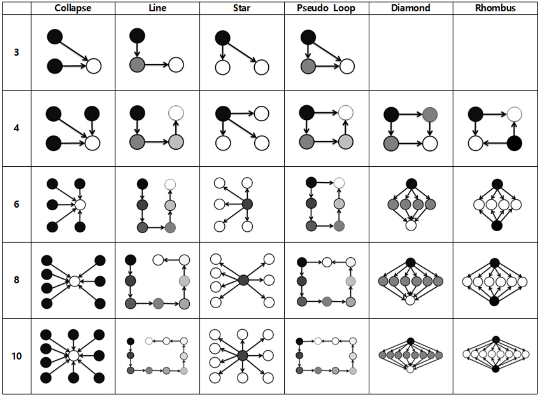
HailFinder DataSet



HailFinder DataSet



Varying topologies and number of nodes

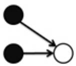
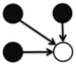
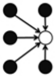


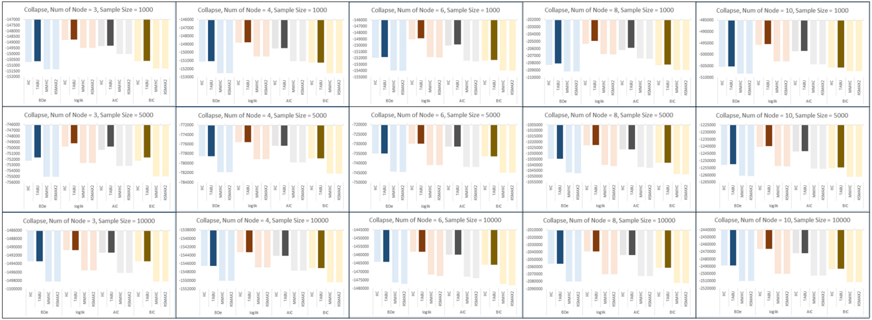
Eitel J. M. Lauría,
"An Information-Geometric Approach to Learning Bayesian Network Topologies from Data",
Innovations in Bayesian Networks Studies in Computational Intelligence Volume 156, 2008, pp 187-217

Prerequisite

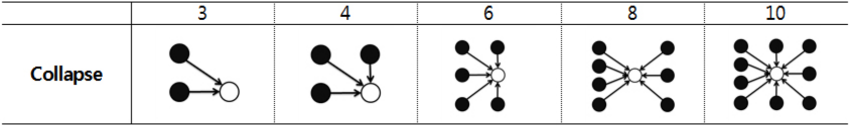
- Cardinality was limited to two.
- The probability value was set by U(0, 1) distribution.
- All experiments are repeated 100 times.
- Constraint-based Learning Algorithms often makes undirected arcs.
  So, this has been excluded from comparison.

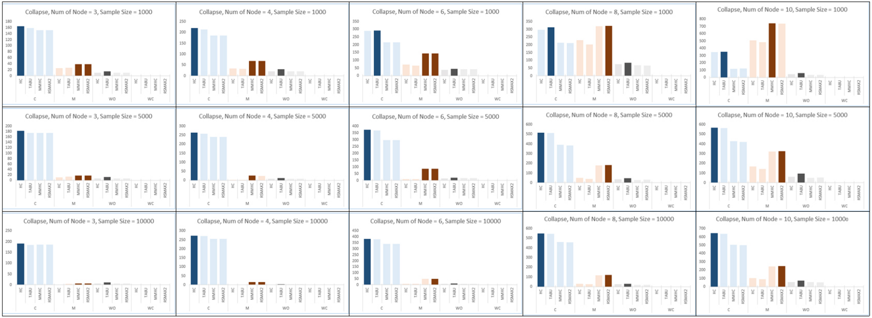
Collapse (Score)

	3	4	6	8	10
Collapse					

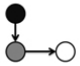
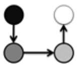
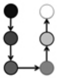
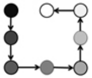
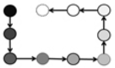


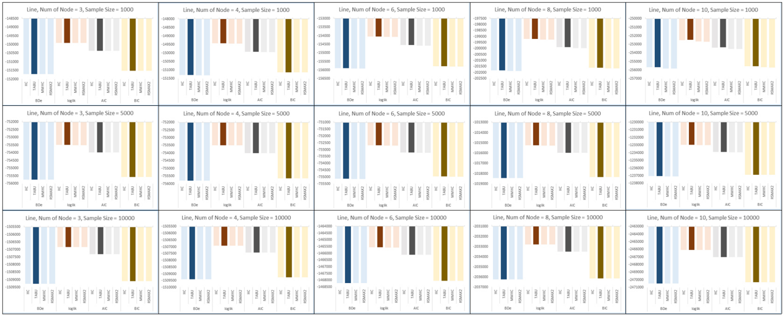
Collapse (Arcs)



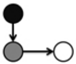
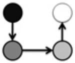
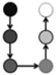
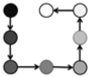
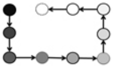


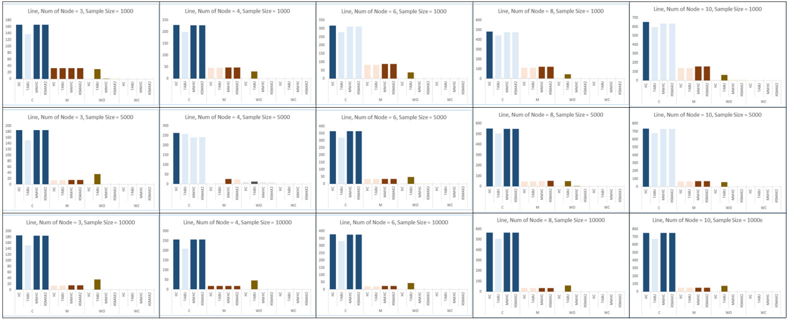
Line (Score)

	3	4	6	8	10
Line					

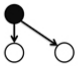
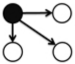
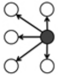
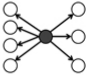


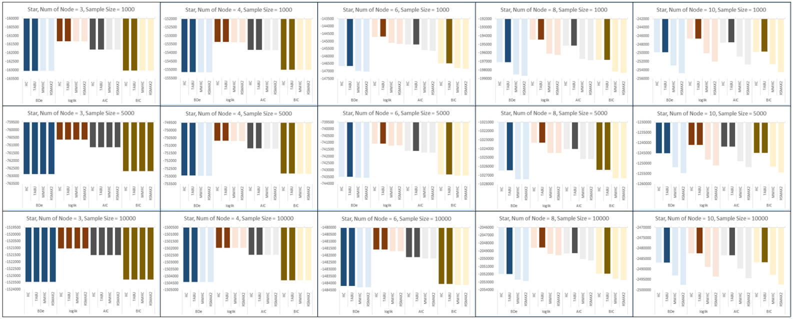
Line (Arcs)

	3	4	6	8	10
Line					

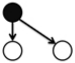
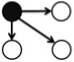
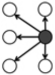
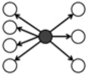


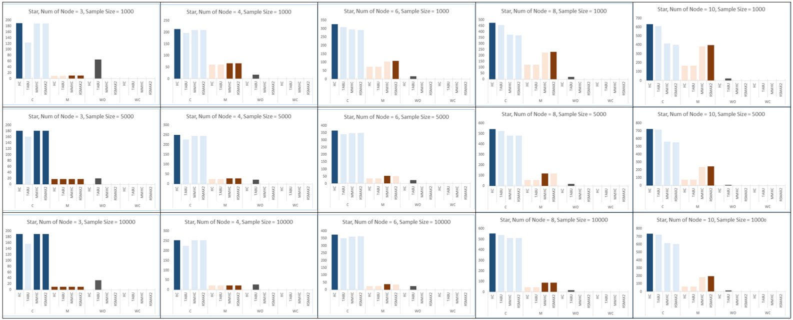
Star (Score)

	3	4	6	8	10
Star					

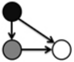
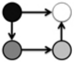
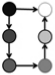
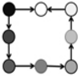
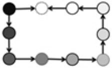


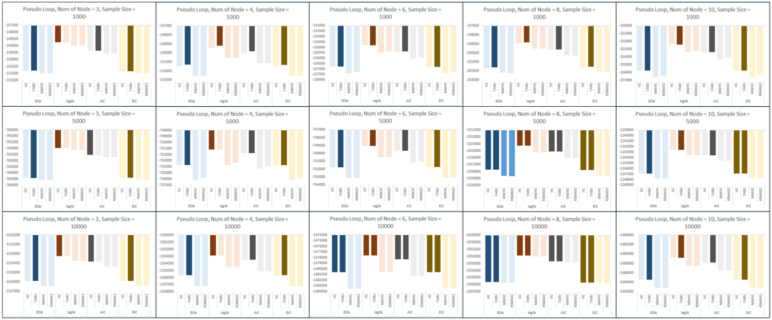
Star (Arcs)

	3	4	6	8	10
Star					

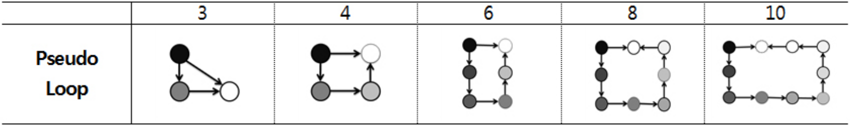


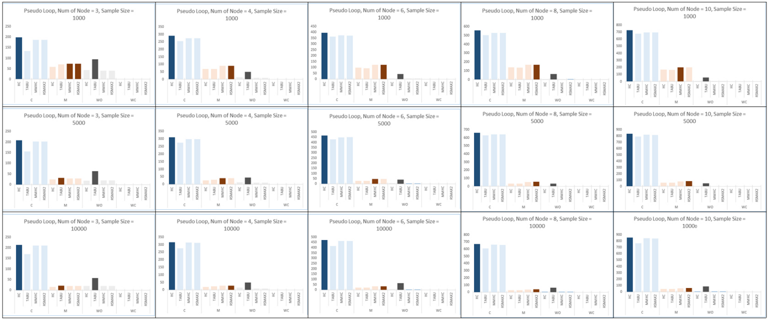
PseudoLoop (Score)

	3	4	6	8	10
Pseudo Loop					

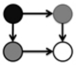
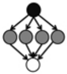
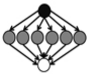
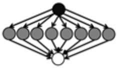


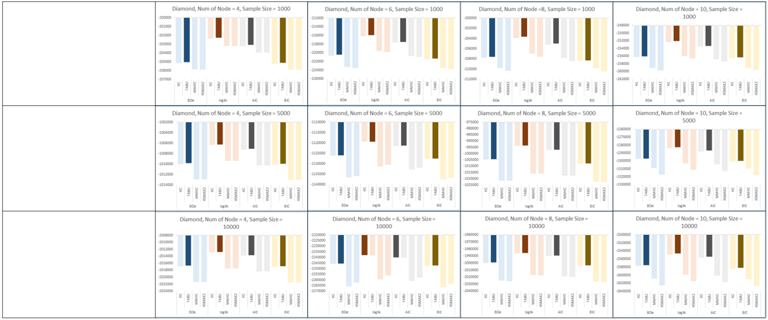
PseudoLoop (Arc)



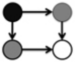
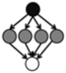
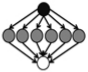
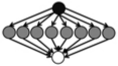


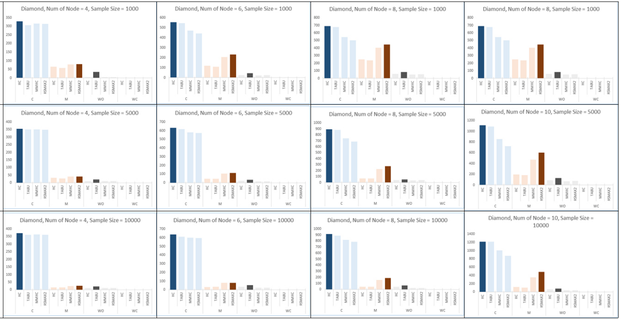
Diamond (Score)

	3	4	6	8	10
Diamond					

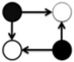
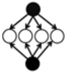
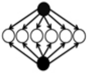
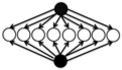


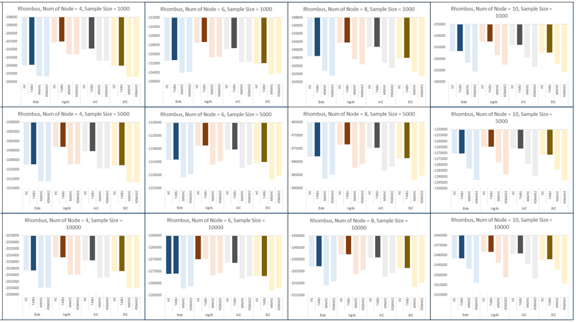
Diamond (Arc)

	3	4	6	8	10
Diamond					

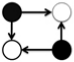
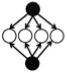
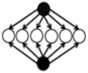
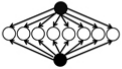


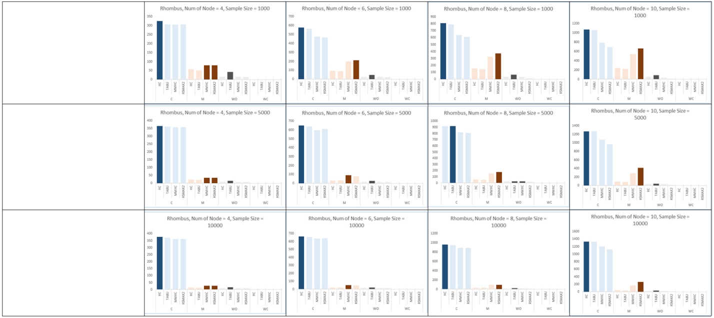
Rhombus (Score)

	3	4	6	8	10
Rhombus					



Rhombus (Arc)

	3	4	6	8	10
Rhombus					



Summary

Collapse: HC = TABU >>> MMHC = RSMAX2

Line: HC = MMHC = RSMAX2 > TABU

Star: HC > TABU = MMHC = RSMAX2

Pseudo Loop: HC >>> TABU = MMHC = RSMAX2

Diamond: HC = TABU >>> MMHC = RSMAX2

Rhombus: HC = TABU >>> MMHC = RSMAX2