The Long (and Fun) Road That Led To DSMGA-II

Tian-Li Yu

National Taiwan University
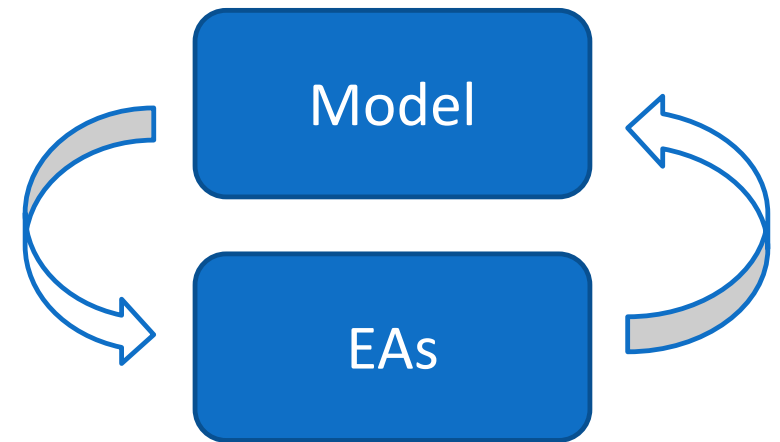
tianliyu@ntu.edu.tw

Outline

- Dependency detection

- From fully separable to overlapping structures

- Linkage trees and optimal mixing

- DSMGA-II

EDAs and Model Based EAs

- Model based EAs

  - Not necessarily estimating the distribution

- A metric to detect dependencies

- A way to regularize the model complexity

- A way to utilize the model (effective and efficient mixing)

What is DSMGA (Yu *et al.*, 2003) ?

| Pairwise Dependency Detection	→	Thresholding	→	Clustering	→	Recombination

[AFE]
[EDBCG]

- To me, DSMGA is more a decomposition of MBEAs than MBEA itself.
- So we can study each part independently.

Dependency-Detection Metrics

- ## Multivariate vs. Pairwise

    - Multivariate: eCGA, BOA, …

    - Pairwise: DSMGA, current GOMEA, DSMGA-II

    - Pairwise prevents from overfitting, especially when the population is small.

    - If you need 30 samples for good statistics, detecting dependencies among 5 bits in the multivariate way requires 1000 chromosomes.
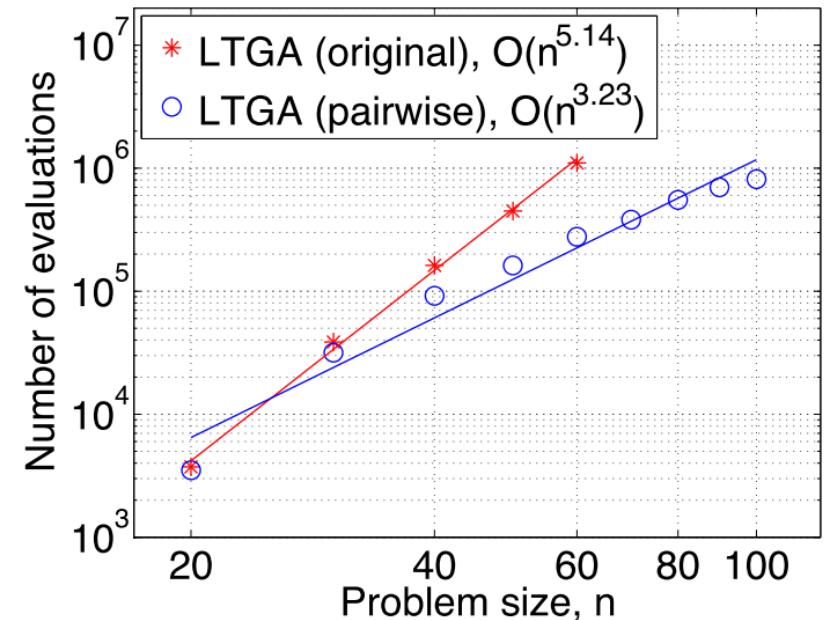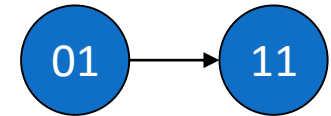

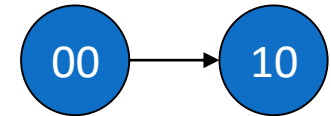Pelikan *et al.*, 2011

Dependency Detection Metrics

- Nonlinearity
  - LINC, LIMD

  $$|f_{00} + f_{11} - f_{10} - f_{01}|$$

- Simultaneity

- Differential mutual complement

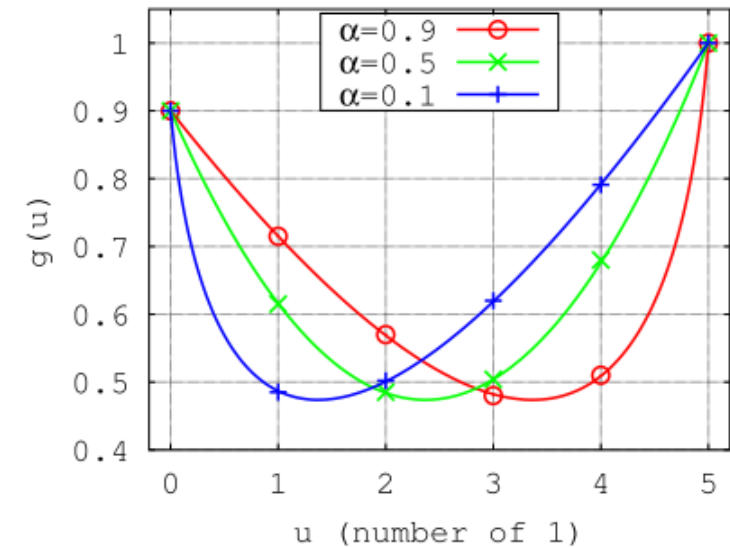 $$p_{00} \cdot p_{11} \pm p_{01} \cdot p_{10}$$

- Entropy-based (mutual information)

  $$\sum_{xy} p_{xy} \cdot \log \frac{p_{xy}}{p_x p_y}$$

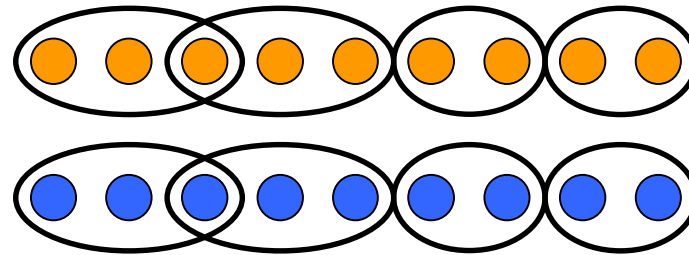 - ecGA, BMDA, BOA, GOMEA, DSMGA-I/II

How Do Metrics Differ?

- Our work (Fan and Yu, 2011) indicated that all metrics are similar qualitatively.

- With optimum thresholds, all metrics performs equally on fully separable problems.

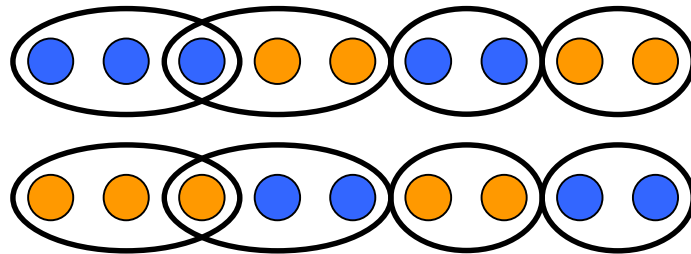- Most MBEAs adopt MDL for thresholding, but not optimum.

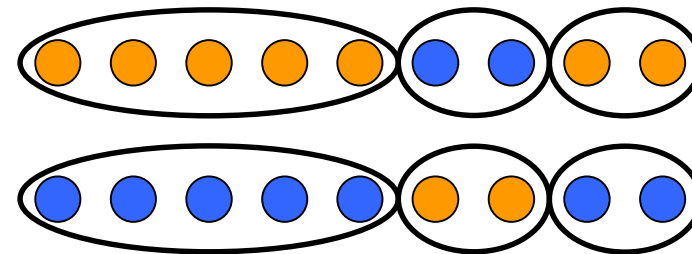Bézier Trap

Mixing with Overlapping BBs
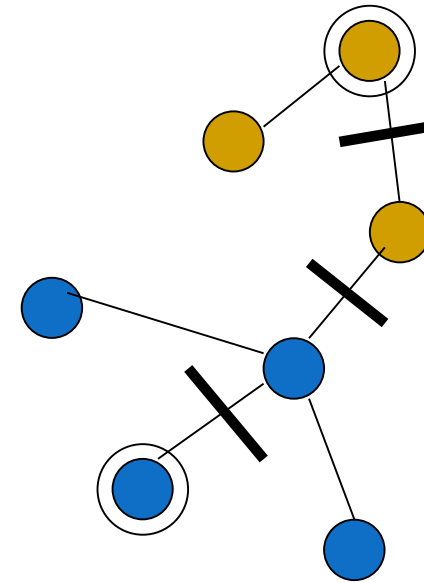
- To group or not to group?

Break one module

One big module

For problems with m BBs, you can at most break $O(m^{1/2})$ of them (Yu et al., 2004).
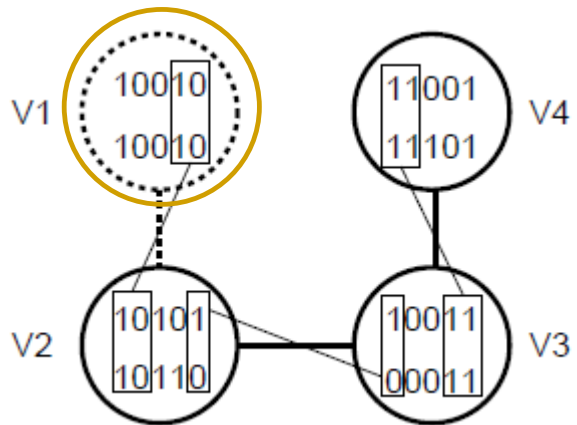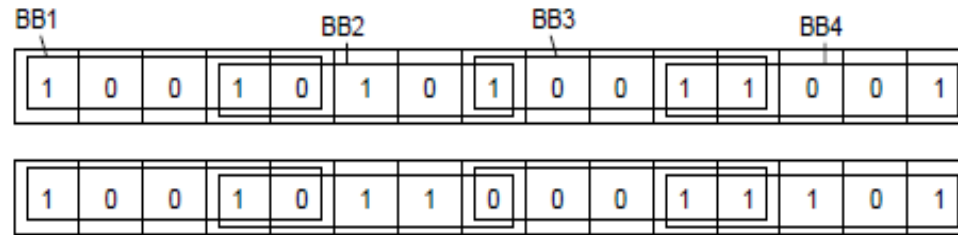
minCut (Yu *et al.,* 2005)

(1) Every BB needs to be recombined with non-zero probability.

(2) The recombination needs to be least disruptive.

- Build a BB graph.

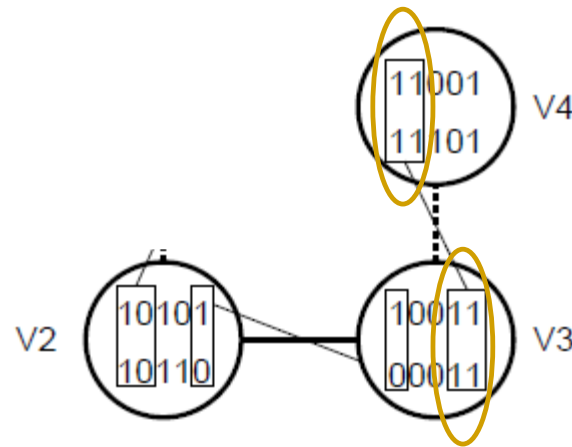- Randomly choose two nodes and find the mincut that separates them.
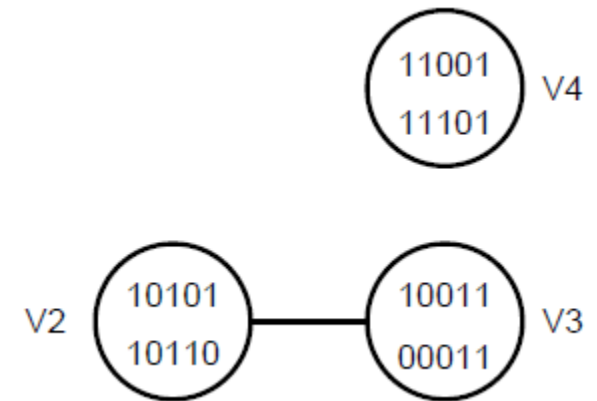
minCut+ (Tsuji *et al.*, 2006)

parents

BB1			BB2				BB3			BB4				
1	0	0	1	0	1	0	1	0	0	1	1	0	0	1
1	0	0	1	0	1	1	0	0	0	1	1	1	0	1

Remove unnecessary nodes
(alleles in BBs are identical)

Remove unnecessary edges
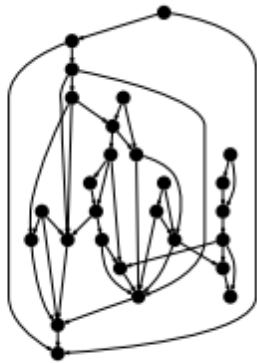(alleles in overlaps are identical)

resulted graph

The Threshold and Bayesian Ways...

- For problems with overlapping structures, using thresholds is dangerous.
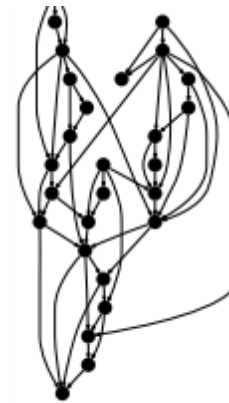
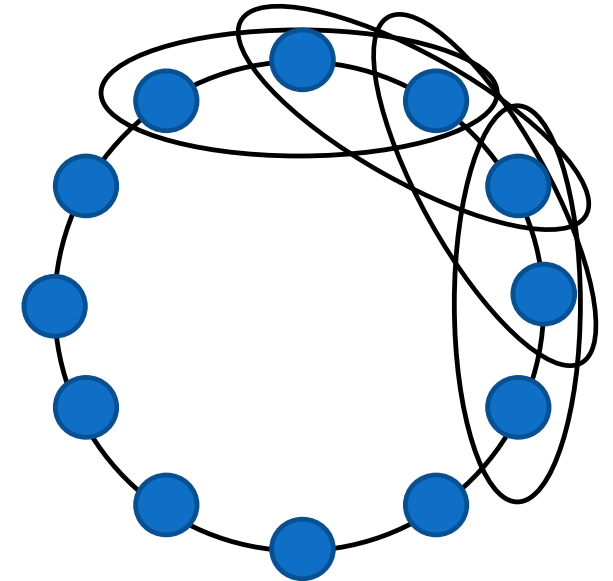- Bayesian way is not easily comprehensive for humans.
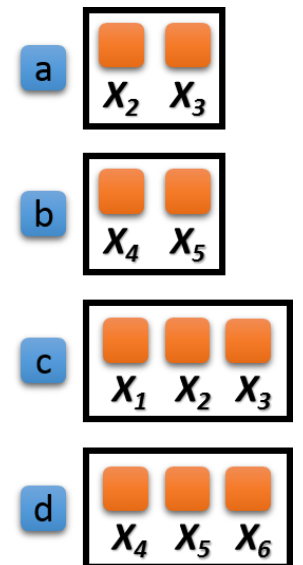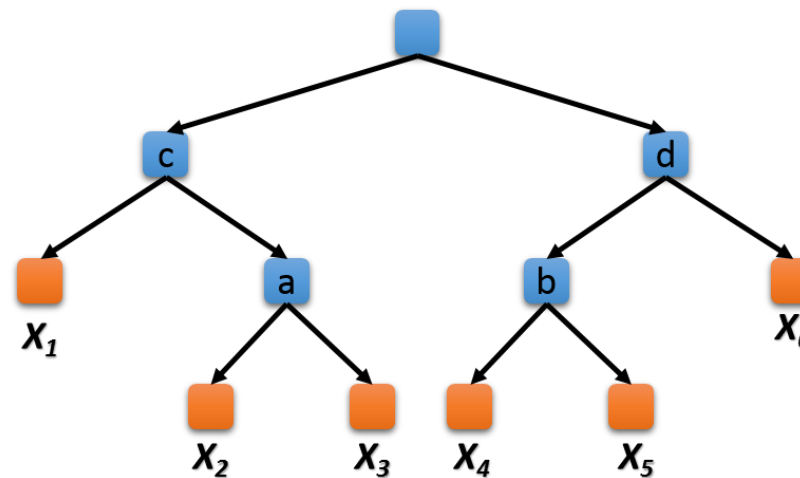

(a) Gen:1   (b) Gen:3   (c) Gen:5   (d) Gen:7

In a cyclic topology, all genes depend on each other.

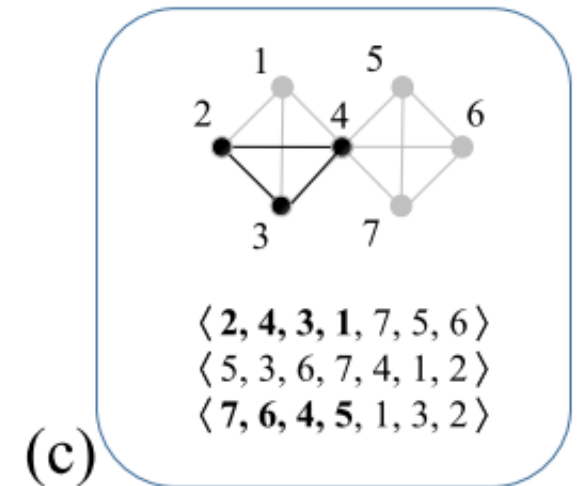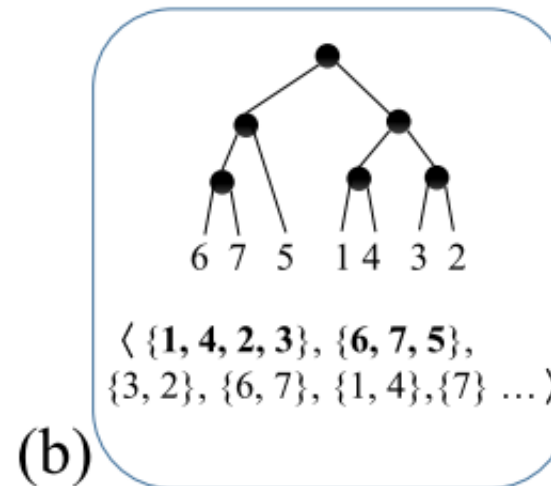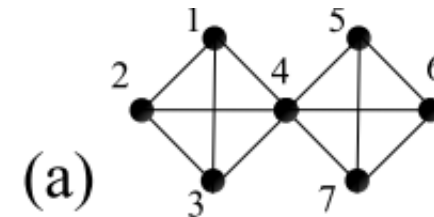# The Linkage Tree Way and OM (Thierens and Bosman *et al.* 2010~)

- Does not need any threshold when identifying BBs

- OM : noise-free decision making

- Population sizing
  - Supply: $2^k \log m$
  - Decision making: $2^k m^{0.5} \log m$
  - Model building: $2^{2k} m \log m$

Incremental Linkage Set (ILS)

{2}, {2,4}, {2,4,3}, {2,4,3,1}, {2,4,3,1,7}, …

- Starting from one gene, incrementally add the next most dependent gene one-by-one.

- More expressive

(a)

(b)

⟨ {1, 4, 2, 3}, {6, 7, 5},
{3, 2}, {6, 7}, {1, 4},{7} …⟩

(c)

⟨ 2, 4, 3, 1, 7, 5, 6⟩
⟨ 5, 3, 6, 7, 4, 1, 2⟩
⟨ 7, 6, 4, 5, 1, 3, 2⟩

Restricted Mixing (RM)

- All genes in the mask are altered

    - Same effect as restrictedly choosing the donor

- RM does not explicitly require a donor. Instead, it adopts **supply check**:

    - Masks stop growing when the corresponding patterns are not in the population.

- More efficient

    - Every different mask results in different chromosome.

Back Mixing (BM)

- Successful RM → good pattern
- Use that pattern to OM all chromosomes in the population
- Losing diversity quickly
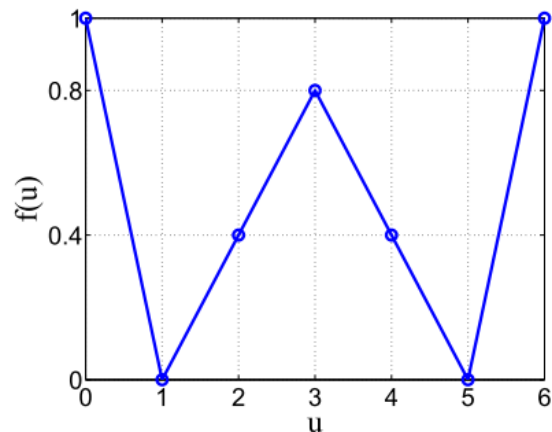- Seems risky, but actually safe
- Refining DSM
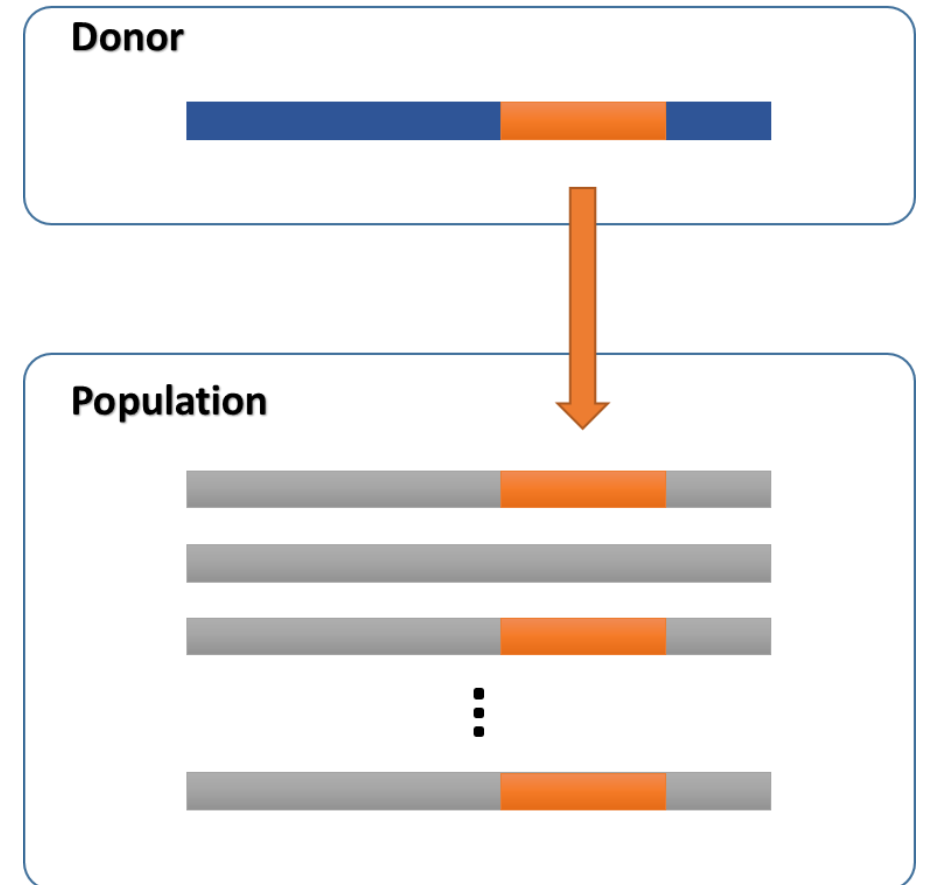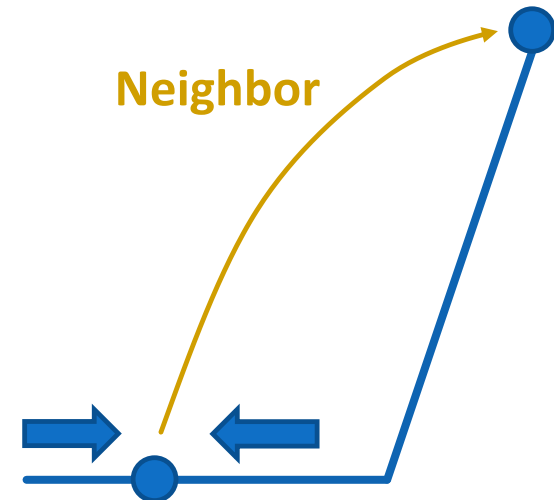
DSMGA-II Performances (Hsu and Yu, 2015)
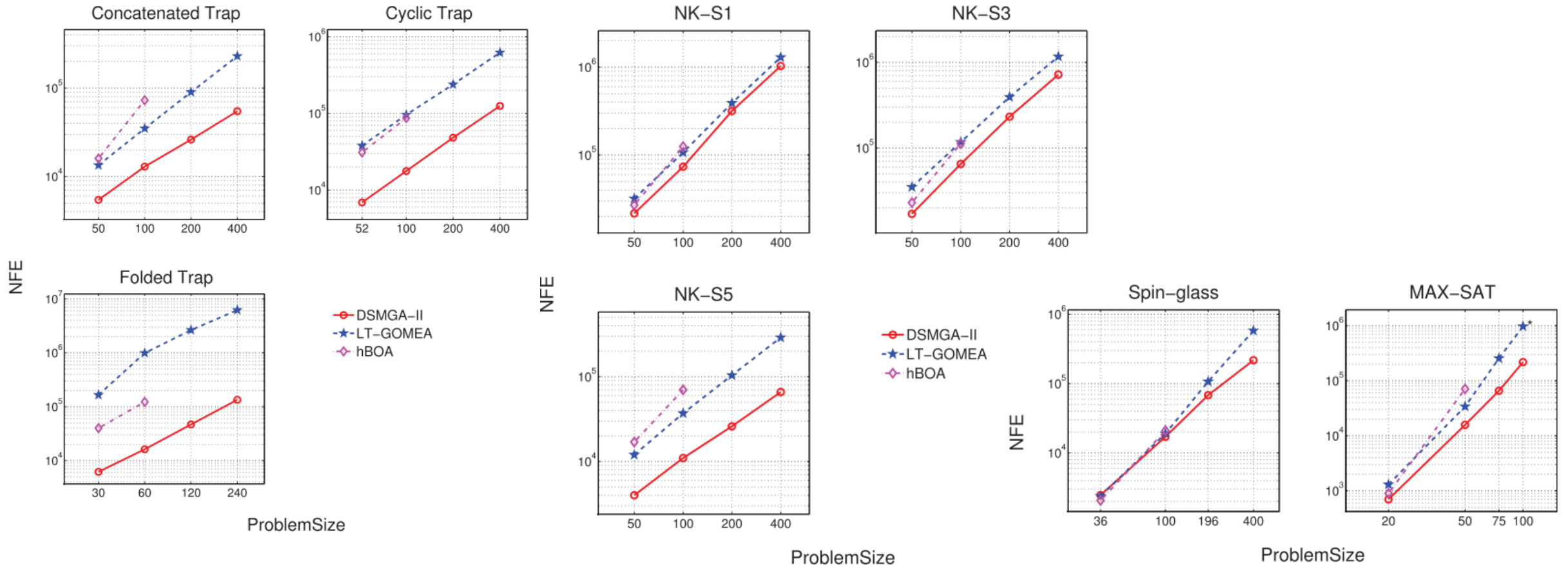
Problems with Plateau

- DSMGA-II does not actually explore plateau.

- Instead, back mixing shrinks the plateau into one point, and then neighborhood is established from that point to the optimum.

Folded Trap

Neighbor

Conclusion

- Pairwise detection is more robust than multivariate.

- Using threshold is dangerous for problems with overlapping structures.

- More expressive linkage models may result in more function evaluations without proper ways to reduce trials.

- Losing diversity is not always bad for EAs.

- OM seems nothing more than a hill climber at the first glance, but with adaptive neighborhood, it is powerful.

- Finally, we need more theoretical supports.