New Local Move Operators for Learning the Structure of Bayesian Networks

Jimmy Vandel, Brigitte Mangin & Simon de Givry

INRA, Toulouse, France

Université Paul Sabatier TOULOUSE III

Unité BIA Toulouse
Biométrie et Intelligence Artificielle

INRA
Institut National de la Recherche Agronomique

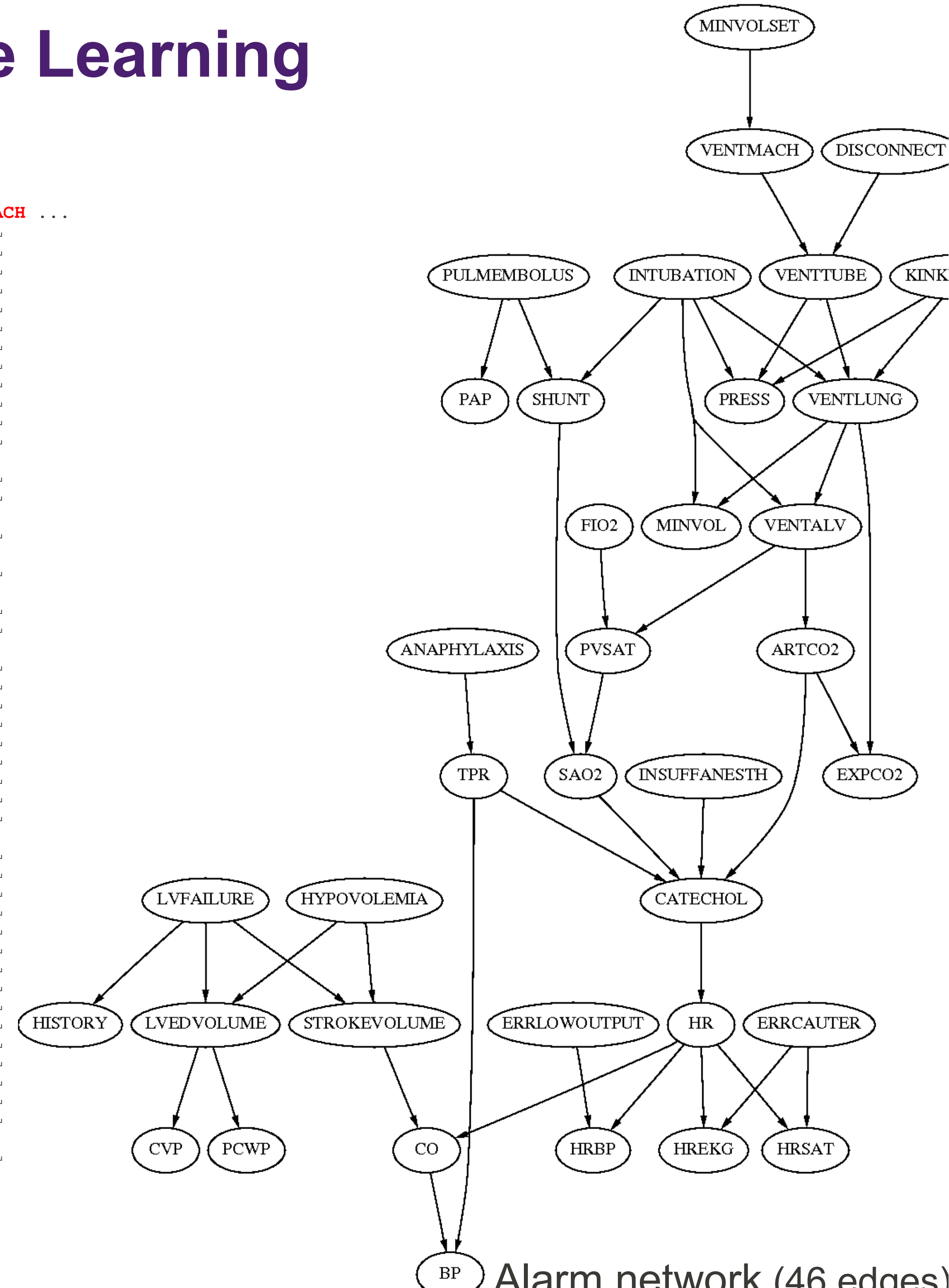
Structure Learning

SAO2	FIO2	PRESS	EXPCO2	MINVOL	MIVOLS	HYPOV	LVFAI	ANAPH	INSUF	VENTMACH	...
NORMAL	LOW	HIGH	ZERO	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
NORMAL	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	HIGH	ZERO	NORMAL	FALSE	FALSE	FALSE	TRUE	NORMAL	
LOW	LOW	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
NORMAL	NORMAL	HIGH	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	HIGH	
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	TRUE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
HIGH	NORMAL	HIGH	LOW	HIGH	LOW	FALSE	FALSE	FALSE	FALSE	LOW	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
HIGH	NORMAL	LOW	LOW	HIGH	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH	
LOW	NORMAL	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	HIGH	HIGH	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH	
HIGH	NORMAL	HIGH	LOW	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	HIGH	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	ZERO	
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	LOW	NORMAL	LOW	FALSE	FALSE	FALSE	FALSE	LOW	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	NORMAL	LOW	NORMAL	FALSE	TRUE	FALSE	TRUE	NORMAL	
LOW	NORMAL	HIGH	NORMAL	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	LOW	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
NORMAL	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	ZERO	LOW	LOW	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
HIGH	NORMAL	HIGH	LOW	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	LOW	ZERO	LOW	NORMAL	FALSE	FALSE	TRUE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL	
LOW	NORMAL	HIGH	ZERO	ZERO	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH	
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL	
...											

Alarm dataset (37 variables)

Structure Learning

SAO2	FIO2	PRESS	EXPCO2	MINVOL	MIVOLS	HYPOV	LVFAI	ANAPH	INSUF	VENTMACH ...
NORMAL	LOW	HIGH	ZERO	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
NORMAL	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	HIGH	ZERO	NORMAL	FALSE	FALSE	FALSE	TRUE	NORMAL
LOW	LOW	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
NORMAL	NORMAL	HIGH	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	HIGH
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	TRUE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
HIGH	NORMAL	HIGH	LOW	HIGH	LOW	FALSE	FALSE	FALSE	FALSE	LOW
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
HIGH	NORMAL	LOW	LOW	HIGH	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH
LOW	NORMAL	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	HIGH	HIGH	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH
HIGH	NORMAL	HIGH	LOW	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	HIGH	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	ZERO
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	LOW	NORMAL	LOW	FALSE	FALSE	FALSE	FALSE	LOW
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	NORMAL	LOW	NORMAL	FALSE	TRUE	FALSE	TRUE	NORMAL
LOW	NORMAL	HIGH	NORMAL	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	LOW	NORMAL	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
NORMAL	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	ZERO	LOW	LOW	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
HIGH	NORMAL	HIGH	LOW	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	LOW	ZERO	LOW	NORMAL	FALSE	FALSE	TRUE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	NORMAL
LOW	NORMAL	HIGH	ZERO	ZERO	HIGH	FALSE	FALSE	FALSE	FALSE	HIGH
LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	TRUE	FALSE	FALSE	FALSE	NORMAL
...										

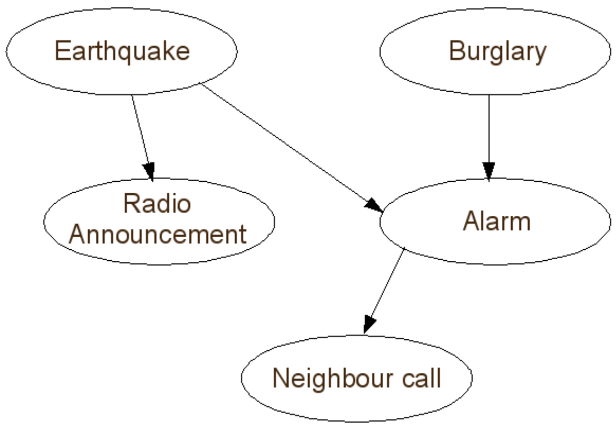


Alarm network (46 edges)

Bayesian Network

$\{ Earthquake \perp Burglary, \ldots \}$

- Directed Acyclic Graph (DAG)
- Conditional probability distribution of X_i, given its parents Pa_i in G: $P(X_i / Pa_i^j) = \theta_i^j$



Burglary network

Graphical representation of a joint probability distribution:

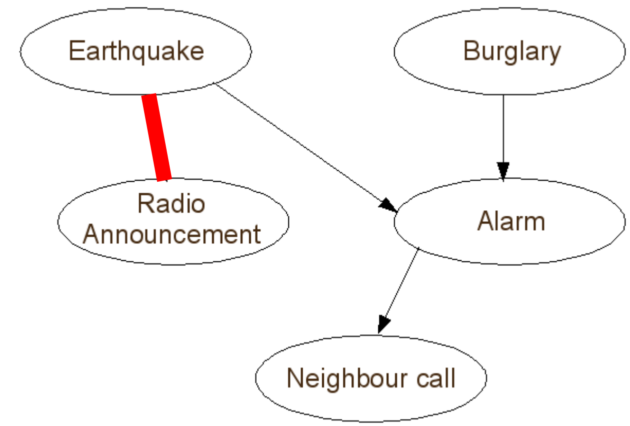
$$P(X) = \prod_{i=1}^{p} P(X_i / Pa_i)$$

Probability Distribution for the Alarm Node given the events of "Earthquakes" and "Burglaries"			
E	B	P(A | E,B)	P(!A | E,B)
E	B	0.90	0.10
E	! B	0.20	0.80
! E	B	0.90	0.10
! E	! B	0.01	0.99

Bayesian Network

$\{ Earthquake \perp Burglary, \ldots \}$

Partial DAG (PDAG)



Burglary network

- Directed Acyclic Graph (DAG)
- Conditional probability distribution of X_i, given its parents Pa_i in G: $P(X_i / Pa_i^j) = \theta_i^j$

Probability Distribution for the Alarm Node given the events of "Earthquakes" and "Burglaries"			
E	B	P(A | E,B)	P(!A | E,B)
E	B	0.90	0.10
E	! B	0.20	0.80
! E	B	0.90	0.10
! E	! B	0.01	0.99

Graphical representation of a joint probability distribution:

$$P(X) = \prod_{i=1}^{p} P(X_i / Pa_i)$$

Structure Learning strategy

- Find the graph $\hat{G} = argmax_{G_i} P(G_i / D)$ with dataset D

$$P(G_i / D) = \frac{P(D / G_i) P(G_i)}{P(D)}$$

$$\propto P(D / G_i) P(G_i)$$

  - $P(D / G_i)$: marginal likelihood of G_i
  - $P(G_i)$: prior probability of G_i
    $\rightarrow$ assumed to be uniform

- Maximize a scoring function easy to evaluate and which avoids over-fitting
  - Decomposable and penalized scores $f(G) = \sum_{i=1}^{p} f_{X_i}(G) = \sum_{i=1}^{p} f_{X_i}(Pa_i)$
    - BDeu score *(**D.Heckerman** Machine learning 1995)*
    - *BIC score (**G.Schwartz** Annals of statistics 1978)*
  - Local score change from G to G' after operation OP_i modifying Pa_i

$$\Delta_G OP_i = f(G') - f(G) = f_{X_i}(G') - f_{X_i}(G)$$ *(assuming G' is a DAG)*

Local search components

1. Search space

- Directed Acyclic Graph
- Partial DAG (PDAG)

- variable orders

Local search components

1. Search space

- Directed Acyclic Graph
- Partial DAG (PDAG)

- variable orders

2. Initial structure

- empty structure
- random structure
- informed structure (MWST, expert...)

Local search components

1. Search space
   - Directed Acyclic Graph
   - Partial DAG (PDAG)

   - variable orders

2. Initial structure
   - empty structure
   - random structure
   - informed structure (MWST, expert...)

3. Neighborhood operators
   - addition of an edge
   - deletion of an edge
   - reversal of an edge
   - k look-ahead
   - optimal reinsertion

Local search components

1. Search space
   - Directed Acyclic Graph
   - Partial DAG (PDAG)
   - variable orders

2. Initial structure
   - empty structure
   - random structure
   - informed structure (MWST, expert...)

3. Neighborhood operators
   - addition of an edge
   - deletion of an edge
   - reversal of an edge
   - k look-ahead
   - optimal reinsertion

4. Meta-heuristics
   - greedy search (**GS**) also called hill climbing (with restarts)
   - tabu search
   - simulated annealing
   - MCMC
   - genetic algorithms
   - ...

Local search components

1. Search space
   - Directed Acyclic Graph
   - Partial DAG (PDAG)
   - variable orders

2. Initial structure
   - empty structure
   - random structure
   - informed structure (MWST, expert...)

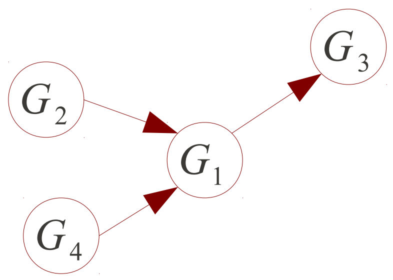
3. Neighborhood operators
   - addition of an edge
   - deletion of an edge
   - reversal of an edge
   - k look-ahead
   - optimal reinsertion

4. Meta-heuristics
   - greedy search (**GS**) also called hill climbing (with restarts)

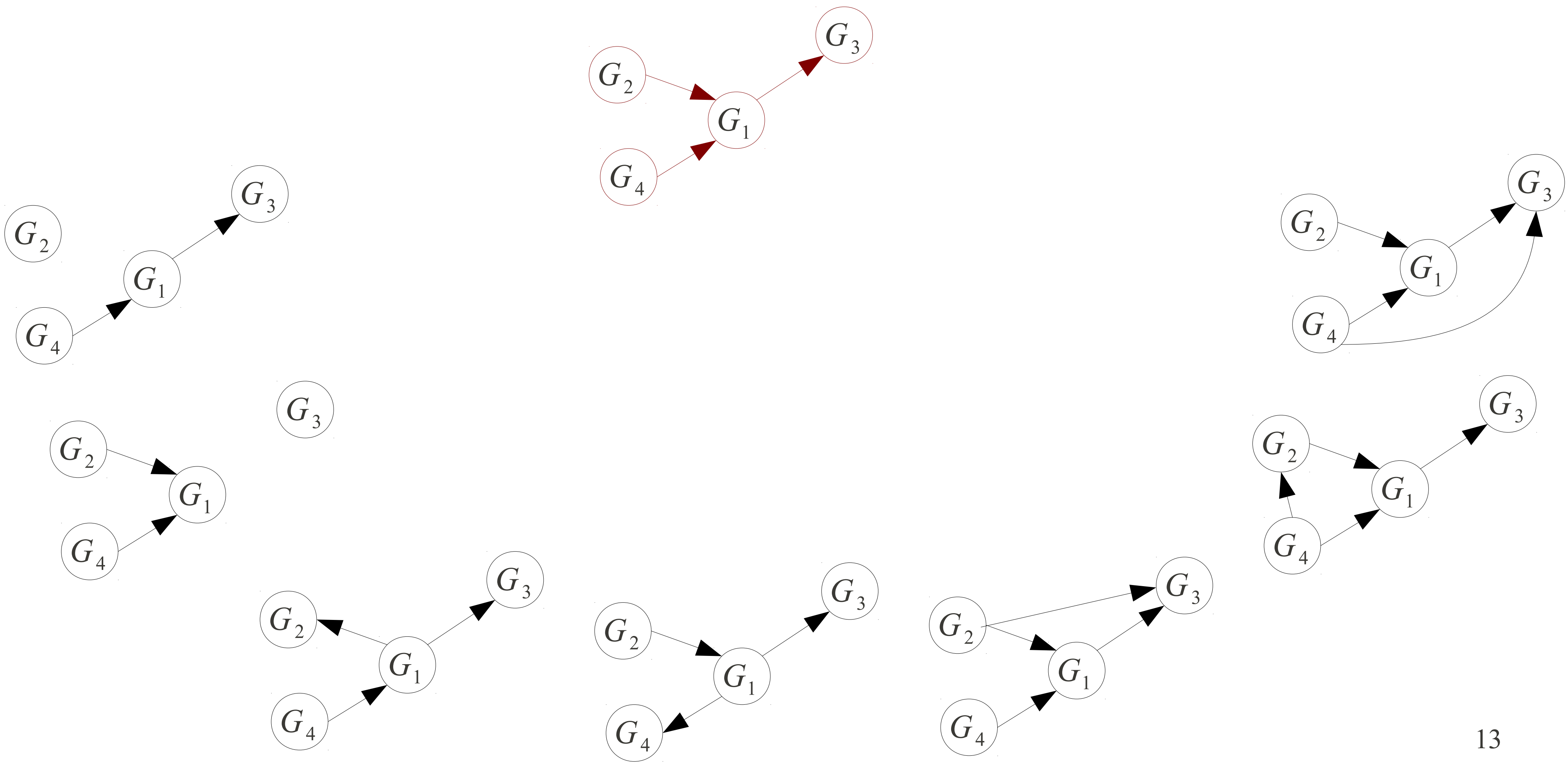
PDAG empty structure addition & deletion **GES**
(*Greedy Equivalence Search*, Chickering 2002)

DAG empty structure restricted 2 look-ahead **LAGD**
(*k Look-Ahead in l Good Directions*, Holland 2008)

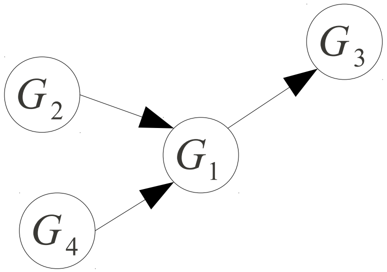
Greedy Search (GS) algorithm

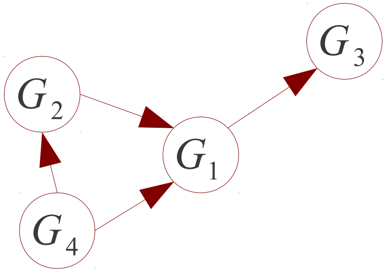


Greedy Search (GS) algorithm

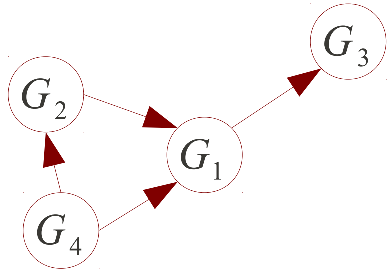


Greedy Search (GS) algorithm

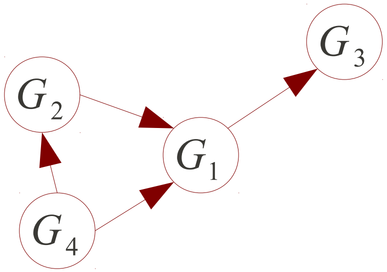




Greedy Search (GS) algorithm



Greedy Search (GS) algorithm



- **Repeat this process until a local maximum is reached**

- **Property 1** (Gamez, Mateo, Puerta, 2011)

Assuming a dataset of n iid fully observed samples of some distribution P and a locally consistent scoring function (BDeu),

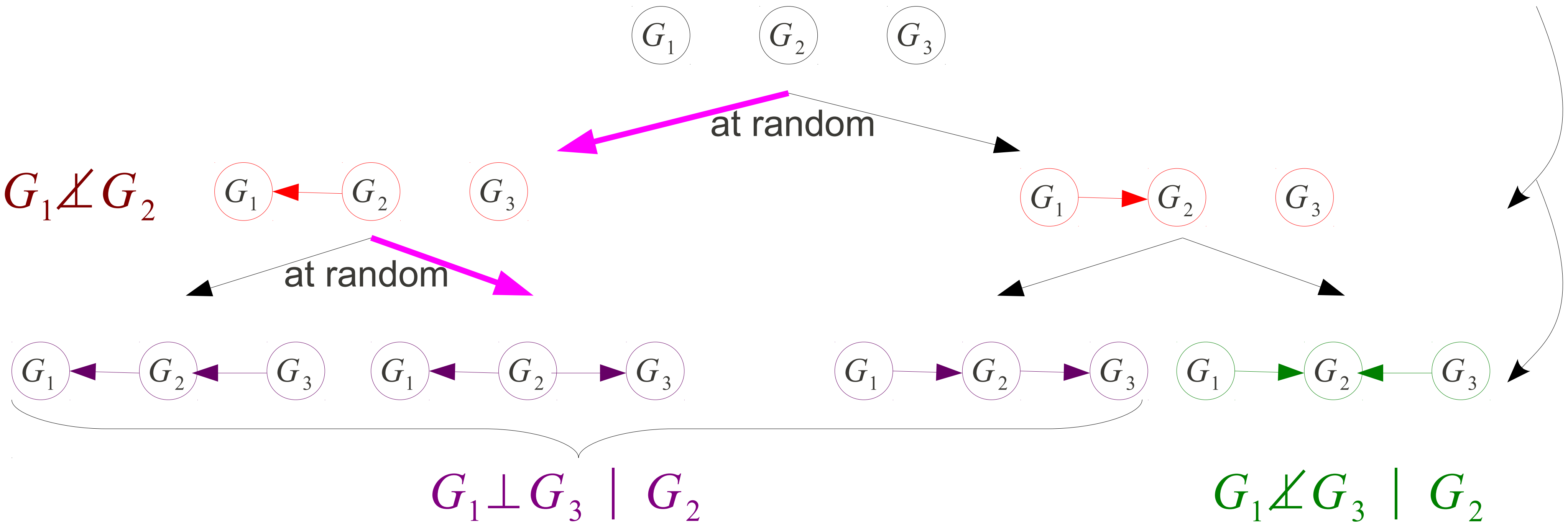
GS returns a minimal independence map of P as the sample size n grows large

- **Restart with another initial random network**

Stochastic Greedy Search (SGS) algorithm

SGS = GS + random orientation for Marvov-equivalent structures

- Markov-equivalent structures in Bayesian networks



Stochastic Greedy Search (SGS) algorithm

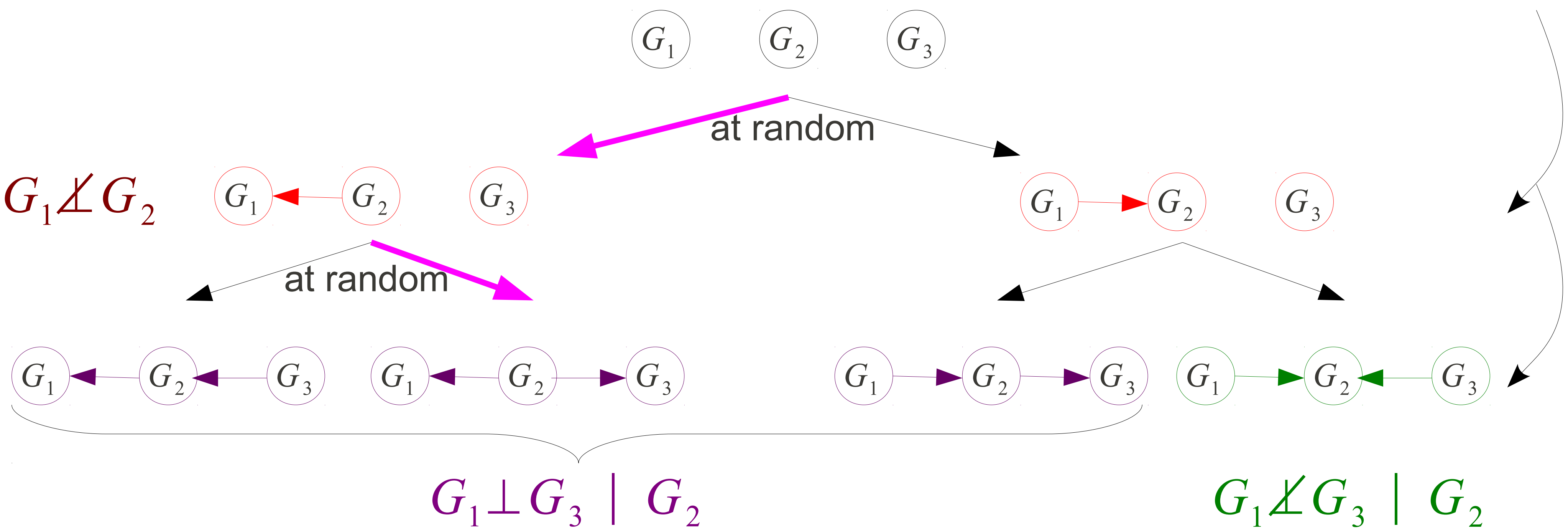
SGS = GS + random orientation for Marvov-equivalent structures $\simeq GES$

- **Property 2** (Chickering 2002)

Assuming a dataset of n iid fully observed samples of some faithful distribution P and a locally consistent scoring function (BDeu),

SGS returns a perfect independence map of P as both the sample size n and the number of restarts r grows large

- Markov-equivalent structures in Bayesian networks



Swap Operator

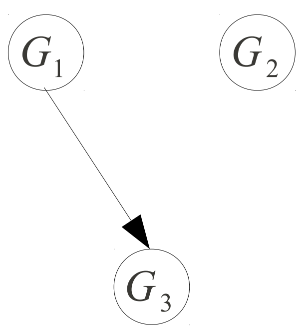
- addition
- deletion
- reversal (deletion + addition on the same pair)
- **swap (deletion + addition on the same target node)**

Swap Operator

- addition
- deletion
- reversal (deletion + addition on the same pair)
- **swap (deletion + addition on the same target node)**

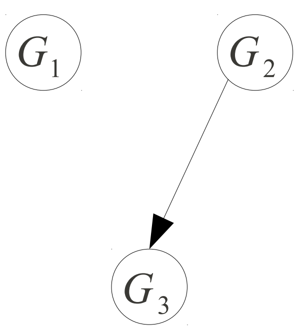
Example:

Current situation



$$\Delta_{\emptyset}\, Add\,(\overrightarrow{G_2, G_3}) > \Delta_{\emptyset}\, Add\,(\overrightarrow{G_1, G_3}) > 0$$

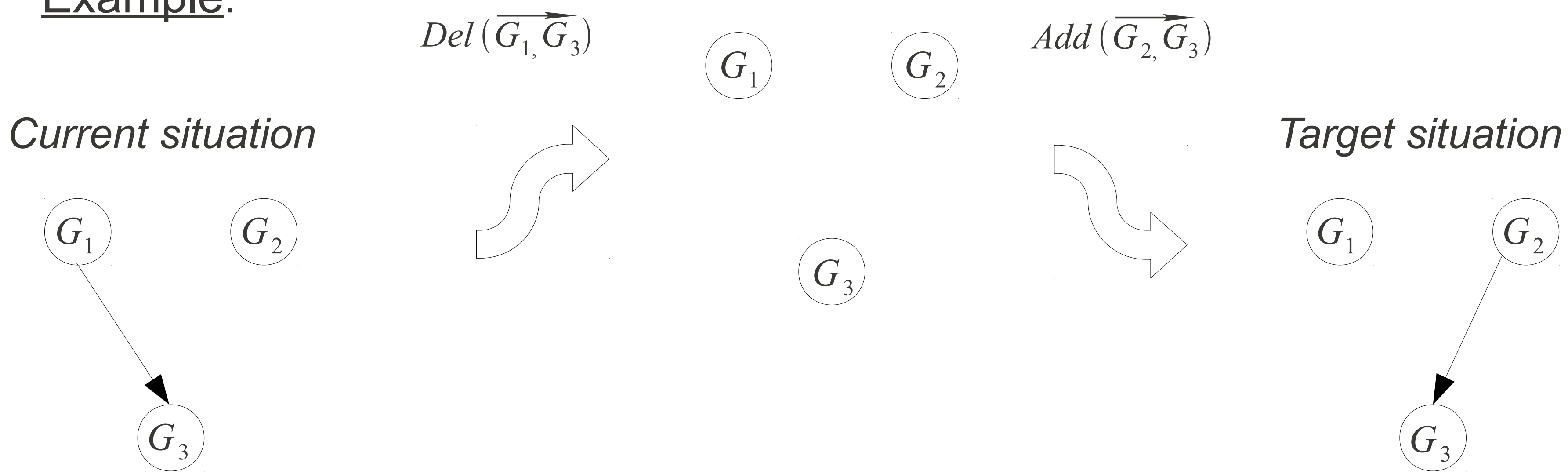
Target situation



Swap Operator

- addition
- deletion
- reversal (deletion + addition on the same pair)
- **swap (deletion + addition on the same target node)**

Example:

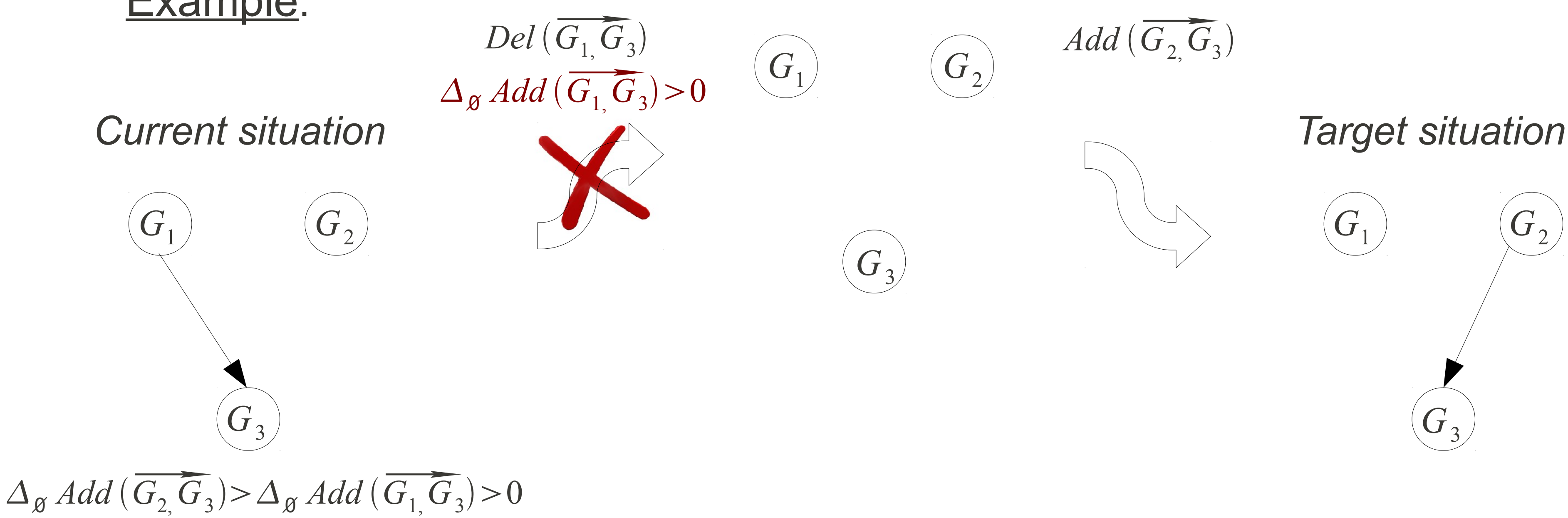


$$\Delta_{\emptyset} Add(\overrightarrow{G_2, G_3}) > \Delta_{\emptyset} Add(\overrightarrow{G_1, G_3}) > 0$$

Swap Operator

- addition
- deletion
- reversal (deletion + addition on the same pair)
- **swap (deletion + addition on the same target node)**

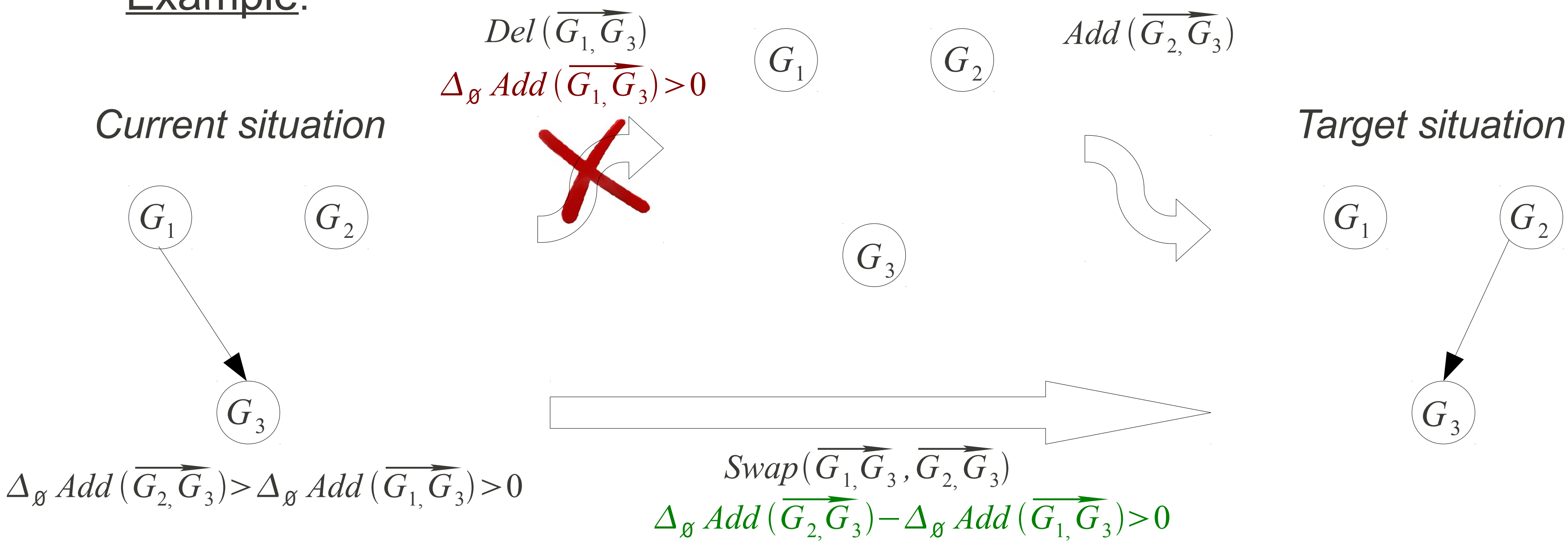
Example:



Swap Operator

- addition
- deletion
- reversal (deletion + addition on the same pair)
- **swap (deletion + addition on the same target node)**

Example:



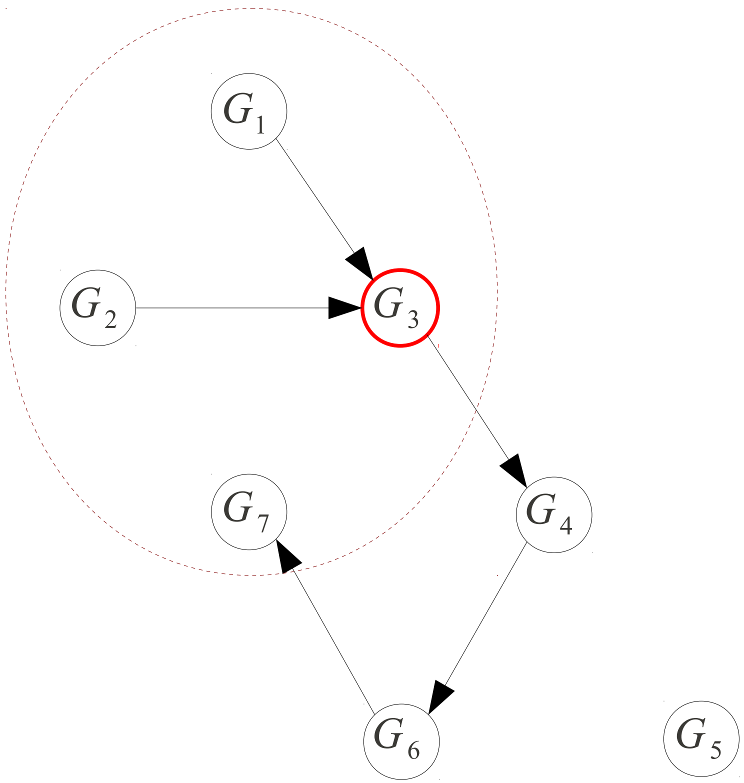
→ escape from some local maxima

Swap★ Operator

$$Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3})\ ?$$

Current situation

$\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) > 0$



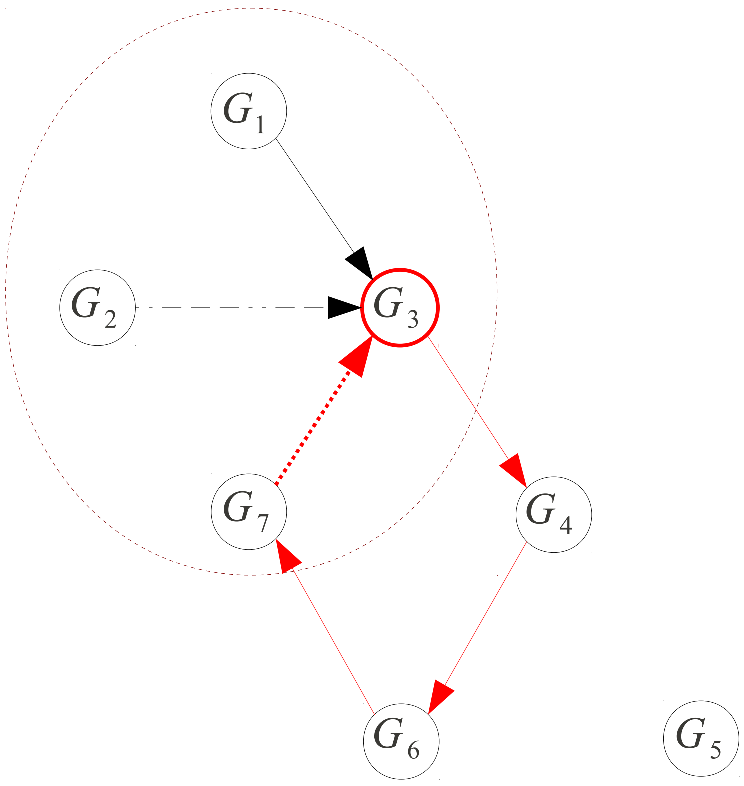
Swap★ Operator

$$Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

Current situation

$$\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) > 0$$

$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$



Swap* Operator

$$Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

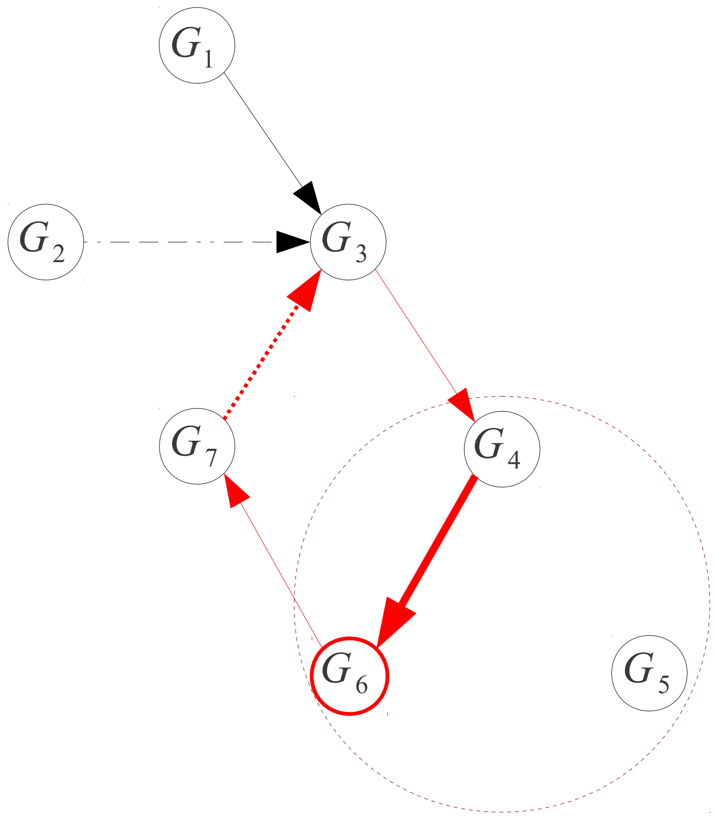
Current situation

$\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) > 0$

$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$

While there exist a directed cycle and ! STOP

Select the edge of the cycle maximizing $\Delta_G Del(e)$

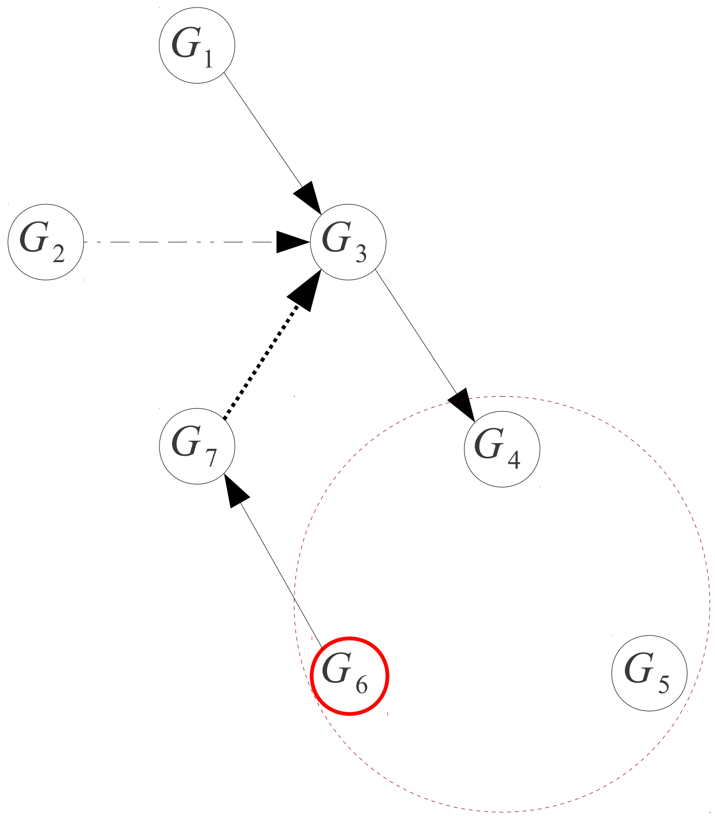


Swap★ Operator

$$Swap(\overrightarrow{G_2, G_3}, \overrightarrow{G_7, G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

Current situation

$\Delta_G Swap(\overrightarrow{G_2, G_3}, \overrightarrow{G_7, G_3}) > 0$



$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$

While there exist a directed cycle and ! STOP

- Select the edge of the cycle maximizing $\Delta_G Del(e)$
- Try to delete it

Swap★ Operator

$$Swap(\overrightarrow{G_{2,}G_3}, \overrightarrow{G_{7,}G_3}) \longrightarrow Cycle\{G_{3,} G_{4,} G_{6,} G_7\}$$

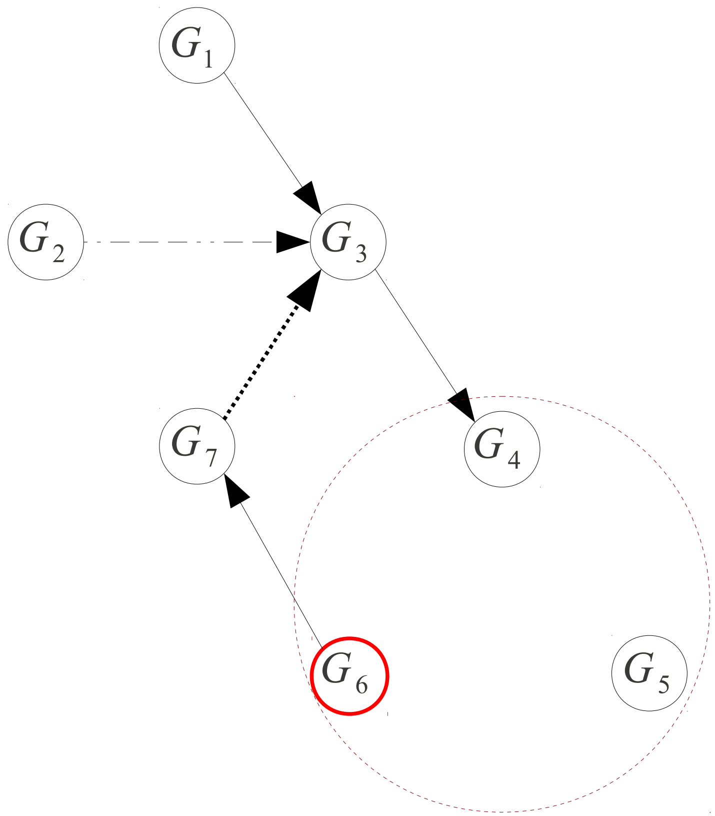
Current situation

$\Delta_G Swap(\overrightarrow{G_{2,}G_3}, \overrightarrow{G_{7,}G_3}) > 0$

$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$

While there exist a directed cycle and ! STOP

- Select the edge of the cycle maximizing $\Delta_G Del(e)$
- Try to delete it
- If $\Delta_G Swap(\overrightarrow{G_{2,}G_3}, \overrightarrow{G_{7,}G_3}) + \Delta_G Del(\overrightarrow{G_{4,}G_6}) \leq 0$
- Else
  - Record $Del(\overrightarrow{G_{4,}G_6})$

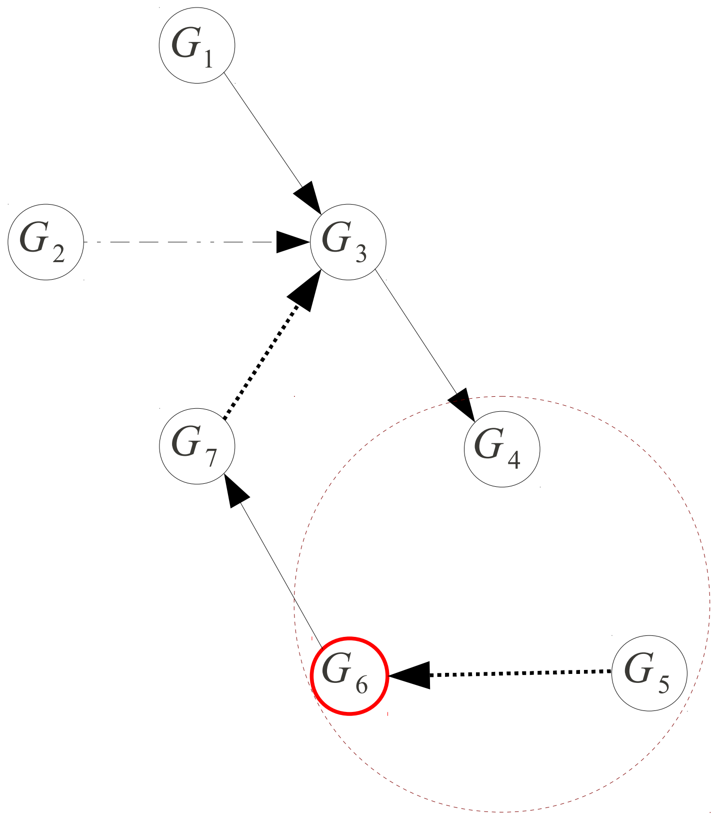


Swap* Operator

$$Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

Current situation

$\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) > 0$



$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$

While there exist a directed cycle and ! STOP

- Select the edge of the cycle maximizing $\Delta_G Del(e)$
- Try to delete it
- If $\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) + \Delta_G Del(\overrightarrow{G_4 G_6}) \leq 0$
  - Try to swap this edge
- Else
  - Record $Del(\overrightarrow{G_4 G_6})$

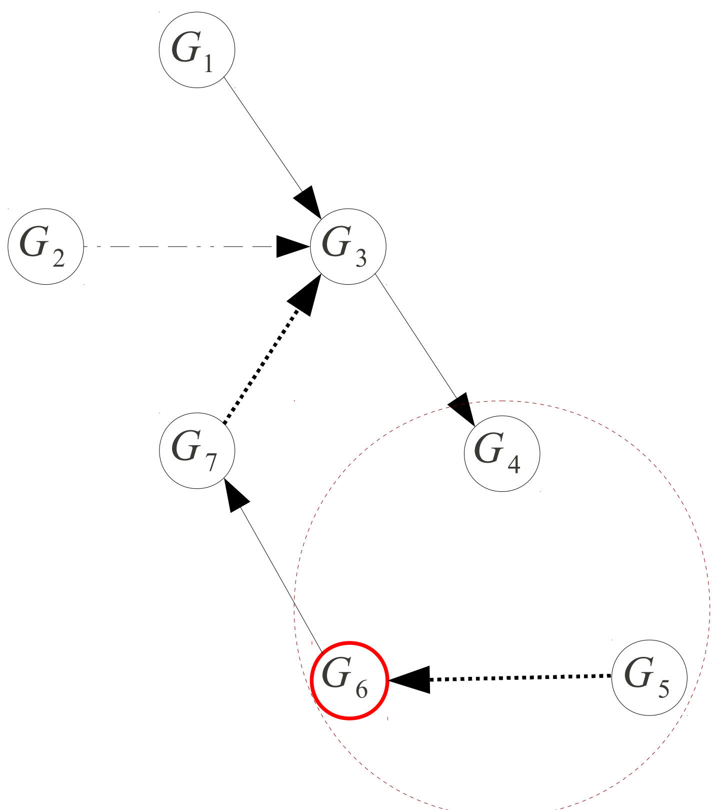
Swap★ Operator

$$Swap(\overrightarrow{G_2G_3}, \overrightarrow{G_7G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

Current situation

$\Delta_G Swap(\overrightarrow{G_2G_3}, \overrightarrow{G_7G_3}) > 0$

$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$



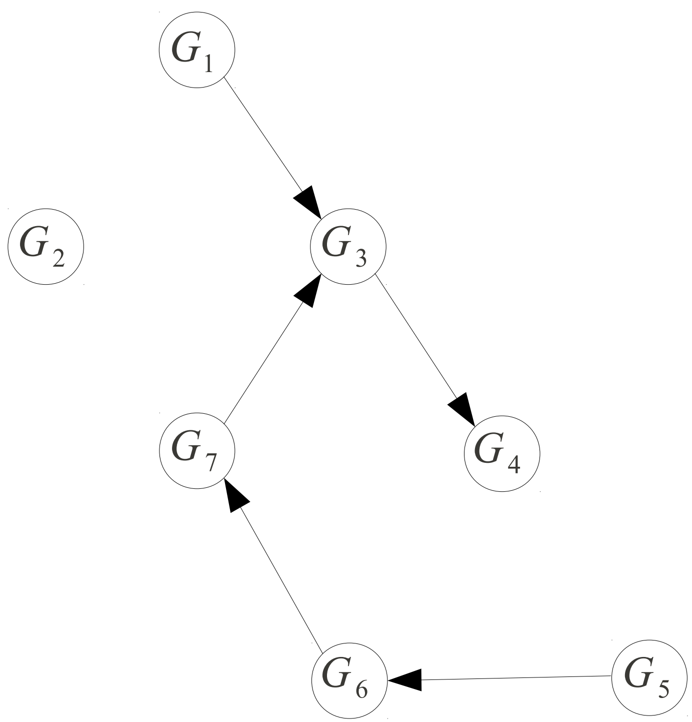
While there exist a directed cycle and ! STOP
- Select the edge of the cycle maximizing $\Delta_G Del(e)$
- Try to delete it
- If $\Delta_G Swap(\overrightarrow{G_2G_3}, \overrightarrow{G_7G_3}) + \Delta_G Del(\overrightarrow{G_4G_6}) \leq 0$
  - Try to swap this edge
  - If $\Delta_G Swap(\overrightarrow{G_2G_3}, \overrightarrow{G_7G_3}) + \Delta_G Swap(\overrightarrow{G_4G_6}, \overrightarrow{G_5G_6}) \leq 0$
    - STOP
  - Else
    - Record $Swap(\overrightarrow{G_4G_6}, \overrightarrow{G_5G_6})$
- Else
  - Record $Del(\overrightarrow{G_4G_6})$

Swap★ Operator

$$Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) \longrightarrow Cycle\{G_3, G_4, G_6, G_7\}$$

Current situation

$\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) > 0$



$$\Delta_G OP_i = f(G') - f(G) \simeq f_{X_i}(G') - f_{X_i}(G)$$

While there exist a directed cycle and ! STOP

- Select the edge of the cycle maximizing $\Delta_G Del(e)$
- Try to delete it
- If $\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) + \Delta_G Del(\overrightarrow{G_4 G_6}) \leq 0$
  - Try to swap this edge
  - If $\Delta_G Swap(\overrightarrow{G_2 G_3}, \overrightarrow{G_7 G_3}) + \Delta_G Swap(\overrightarrow{G_4 G_6}, \overrightarrow{G_5 G_6}) \leq 0$
    - STOP
  - Else
    - Record $Swap(\overrightarrow{G_4 G_6}, \overrightarrow{G_5 G_6})$
- Else
  - Record $Del(\overrightarrow{G_4 G_6})$

If ! STOP

- Validate all recorded moves

SGS algorithms

- **SGS¹**: Addition + Deletion + Reversal
- **SGS²**: Addition + Deletion + Reversal + Swap
- **SGS³**: Addition★ + Deletion + Reversal★ + Swap★

- One parameter: number of restarts **r**
  (stochastic edge orientations for score-equivalent neighbors)

Experimentation

- 4 benchmark networks:

	Alarm	*Insurance*	*Hailfinder*	*Pigs*
Nodes	37	27	56	441
Edges	46	52	66	592
In-degree	4	3	4	2

- Data generated from known conditional probabilities distributions:
  100 datasets with n=500 and n=5,000 sample sizes

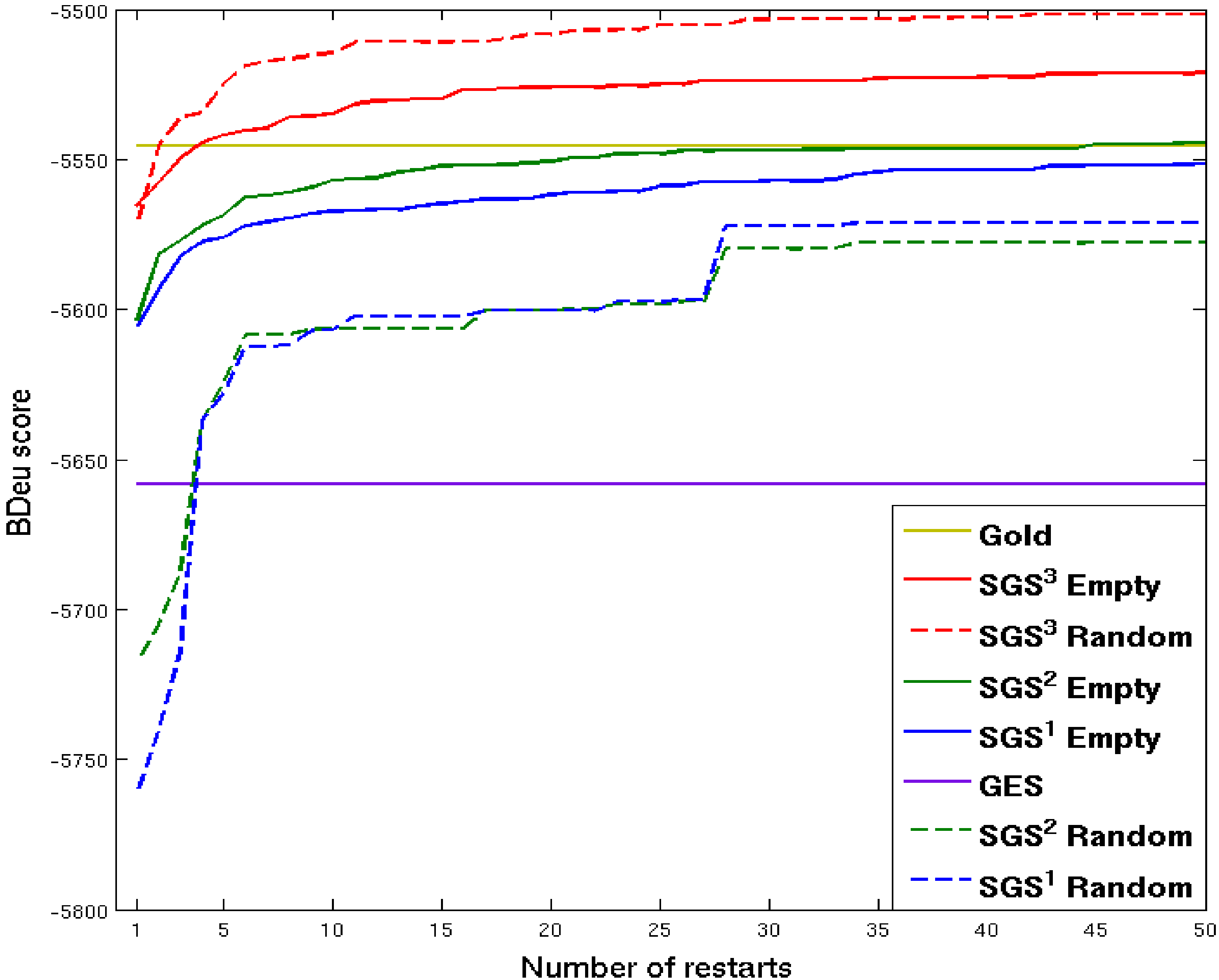
- **SGS** compared to: **LAGD** (2 look-ahead in 5 good directions)
  GES

- Limit on the maximum number of parents : 5

- Pre-filtering candidate parents under condition for Pigs network with SGS

$$\Delta_{\emptyset} Add(\overrightarrow{Candidate, Target}) > 0$$

Impact of the number of restarts *r*



Alarm network (mean BDeu score on **30** datasets with n=**500** samples)

BDeu comparison between SGS³, LAGD & GES

- *4* benchmark networks, n=**500** and **5,000** samples
- Best of r=**10** runs for **SGS³** and **LAGD** (random variable order in input dataset)
- **All methods** initialized with an **empty network**

Wilcoxon test 5%	*Alarm*		*Insurance*		*Hailfinder*		*Pigs*	
	500	5,000	500	5,000	500	5,000	500	5,000
SGS³ vs **GES**	+	+	+	+	+	+	+	-
SGS³ vs **LAGD**	+	+	+	+	~	+	n/a	n/a
LAGD vs **GES**	+	~	+	+	+	+	n/a	n/a

SHD comparison between SGS³, LAGD & GES

Structural Hamming Distance (SHD) = False Positive + False Negative

(edge orientations not taken into account)

	Alarm		*Insurance*		*Hailfinder*		*Pigs*	
	500	5,000	500	5,000	500	5,000	500	5,000
SGS³	**11***	8	**24***	**10***	41	**29***	32	41
LAGD	15	10	**24***	16	47	39	n/a	n/a
GES	**11***	**6***	25	15	**39***	33	**9***	**0***

** : best result*

Conclusion & Perspectives

We

- Propose a new algorithm Stochastic Greedy Search (SGS)
- Propose a new local operator SWAP and its iterative version for breaking cycles
- Improve BDeu scores of learned networks with these operators
- Analyse the impact of initial structures depending on the set of operators

TODO list:

- try other meta-heuristics
- improve SHD results (post-processing rule: SGS^3 was 6/8 better than GES)
- reduce the number of restarts **r** required

Results (4/4)

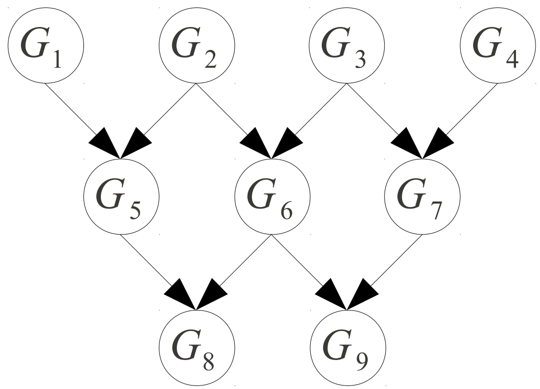
- Comparison of **Hamming distances** for **SGS³**, **LAGD** and **GES**

Hamming distance = False Positive + False Negative

	Alarm		*Insurance*		*Hailfinder*		*Pigs*	
	500	5 000	500	5 000	500	5 000	500	5 000
SGS³	**11***	8	**24***	**10***	41	**29***	32	41
LAGD	15	10	**24***	16	47	39	n/a	n/a
GES	**11***	**6***	25	15	**39***	33	**9***	**0***

* best result

Pigs network



Results (4/4)

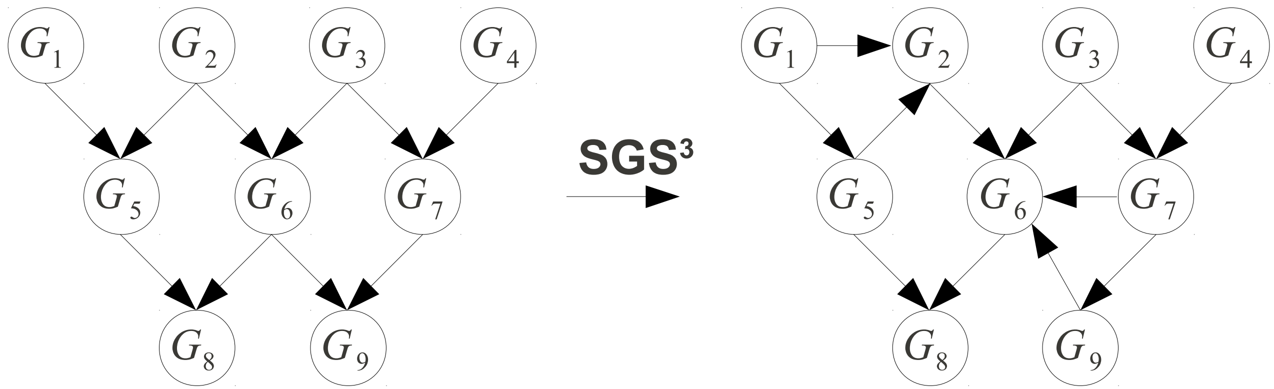
- Comparison of **Hamming distances** for **SGS³**, **LAGD** and **GES**

Hamming distance = False Positive + False Negative

	Alarm		*Insurance*		*Hailfinder*		*Pigs*	
	500	5 000	500	5 000	500	5 000	500	5 000
SGS³	**11***	8	**24***	**10***	41	**29***	32	41
LAGD	15	10	**24***	16	47	39	n/a	n/a
GES	**11***	**6***	25	15	**39***	33	**9***	**0***

* best result

Pigs network



Results (4/4)

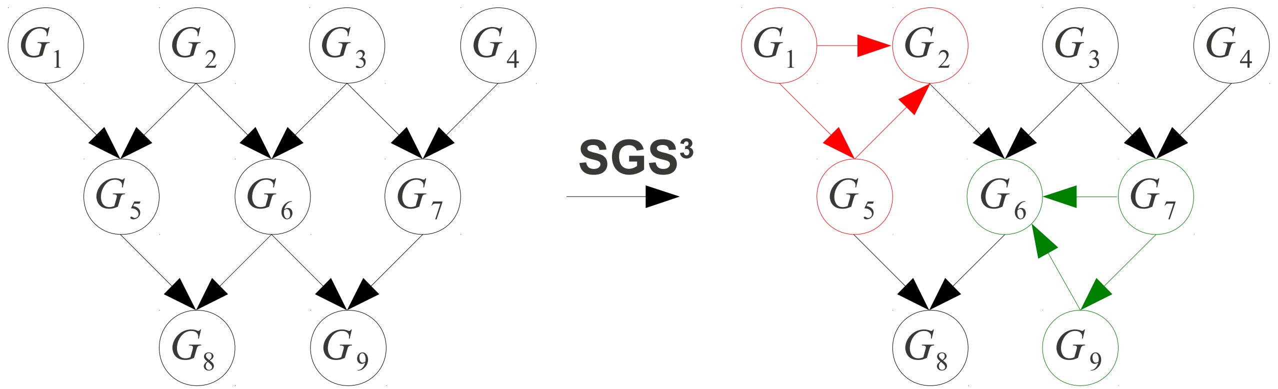
- Comparison of **Hamming distances** for **SGS³**, **LAGD** and **GES**

Hamming distance = False Positive + False Negative

	Alarm		*Insurance*		*Hailfinder*		*Pigs*	
	500	5 000	500	5 000	500	5 000	500	5 000
SGS³	**11***	8	**24***	**10***	41	**29***	32	41
LAGD	15	10	**24***	16	47	39	n/a	n/a
GES	**11***	**6***	25	15	**39***	33	**9***	**0***

* best result

Pigs network



Results (4/4)

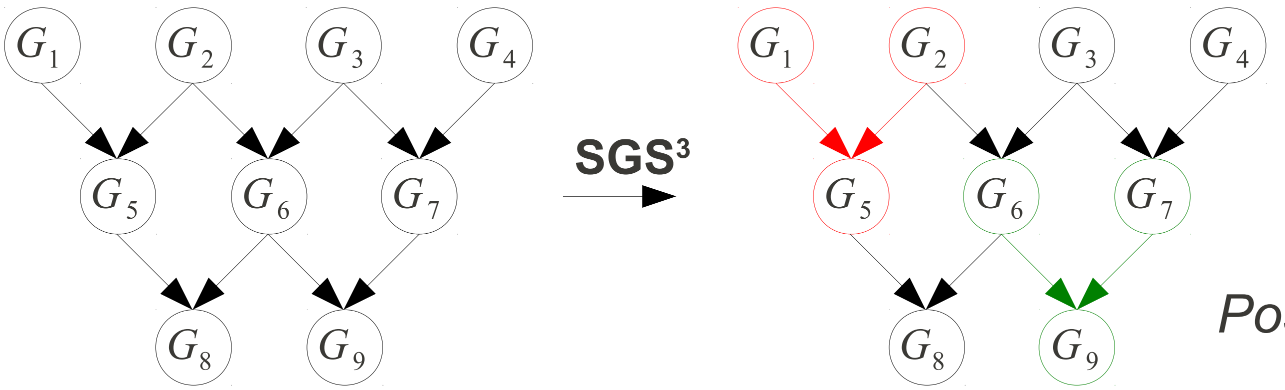
- Comparison of **Hamming distances** for **SGS³**, **LAGD** and **GES**

Hamming distance = False Positive + False Negative

	Alarm		*Insurance*		*Hailfinder*		*Pigs*	
	500	5 000	500	5 000	500	5 000	500	5 000
SGS³	**11***	8	**24***	**10***	41	**29***	32	41
LAGD	15	10	**24***	16	47	39	n/a	n/a
GES	**11***	**6***	25	15	**39***	33	**9***	**0***

* best result

Pigs network

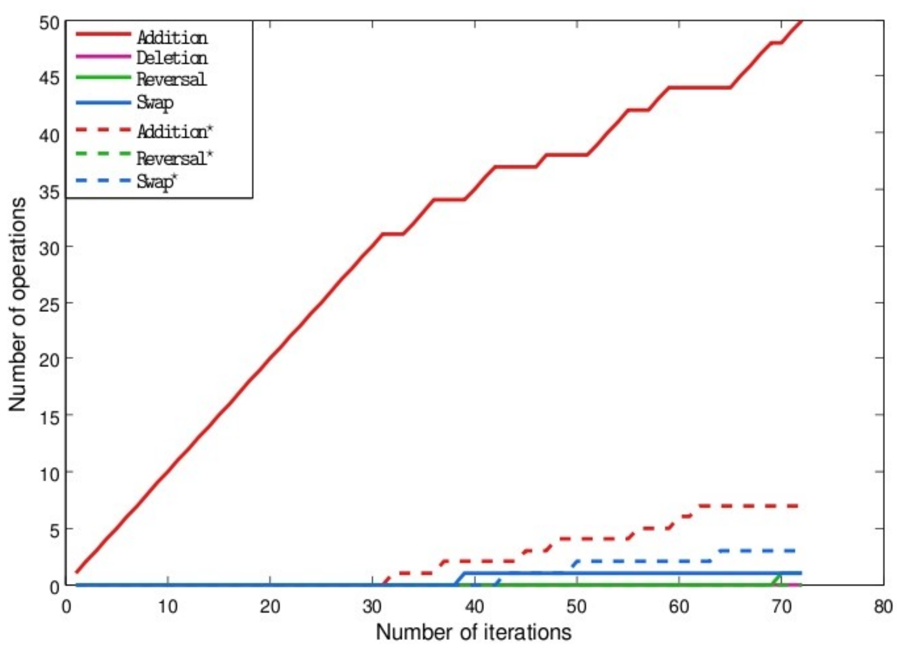


Post-processing

Results (2/4)

- Number of applied operators by type during the search
- *Alarm* network
- **1** run of **SGS³** (r=1) with 500 samples
- **SGS³** Initialized with **empty** and **random network** (2 parents max)

empty network



random network