Weakening Faithfulness: Some Heuristic Causal Discovery Algorithms

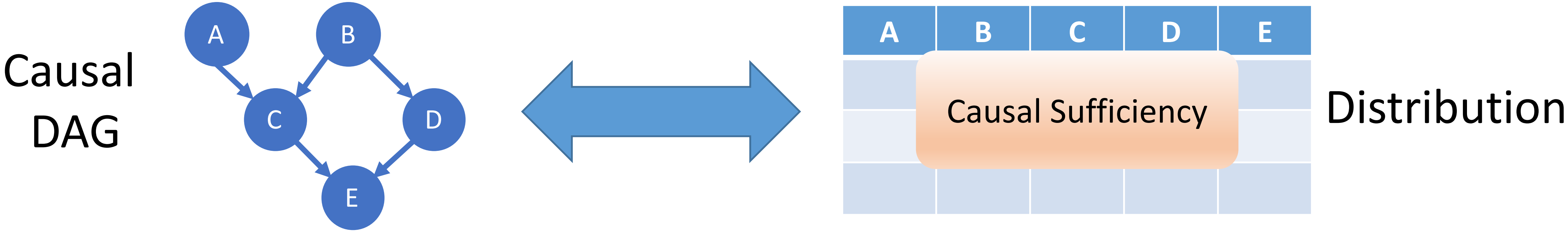
Zhalama[1] Jiji Zhang[2] · Wolfgang Mayer[1]

[1] University of South Australia

[2] Lingnan University

Causal DAG

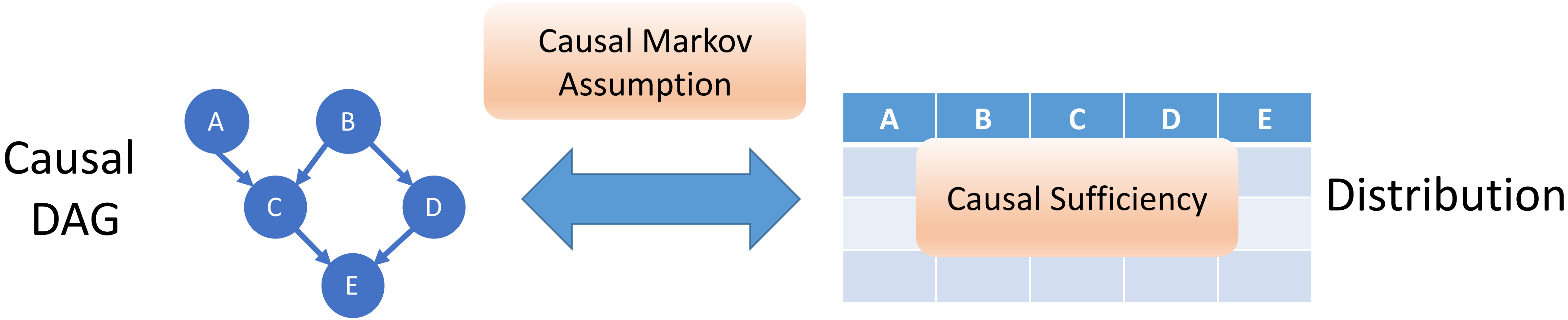
- Causal DAG $G = \langle V, E \rangle$
  Each edge $X \rightarrow Y$ represents a direct causal relation that X is a direct cause of Y relative to V
- Assumption: V is causally sufficient



Causal Inference Assumptions

- Causal Markov Condition: Every conditional independence statement entailed by the causal DAG over V is satisfied by the joint probability distribution of V.

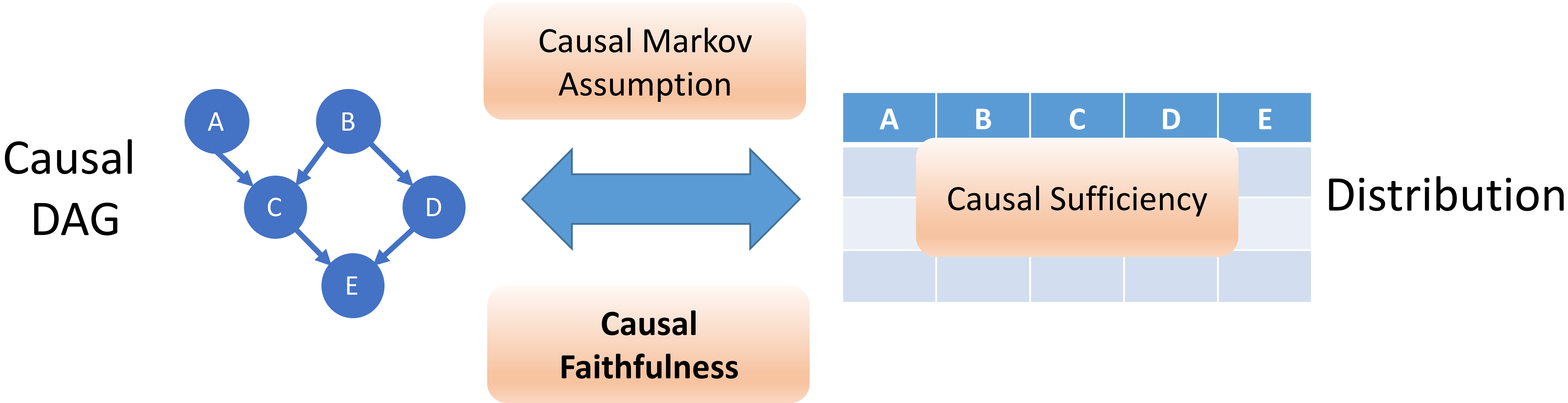
    i.e., X and Y are (causally) d-separated by $\mathrm{Z} \Longrightarrow X \perp Y \mid Z$.



Causal Inference Assumptions

- Faithfulness assumption: Every conditional independence statement satisfied by the joint distribution of V is entailed by the causal DAG over V.

  i.e., $X \perp Y \mid Z \implies X$ and Y are (causally) d-separated by Z.



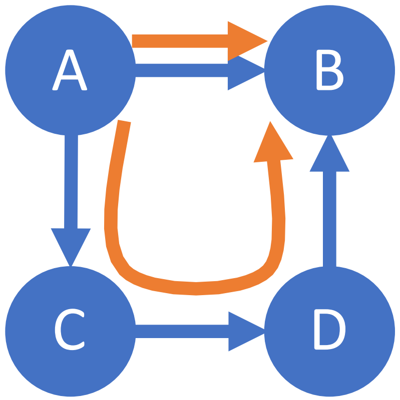
Causal Faithfulness Assumption

- More dubious than Causal Markov assumption.
- Even if Faithfulness is not exactly violated, the distribution may be sufficiently close to being unfaithful to make trouble with finite data.

- Can we **relax the Faithfulness assumption** and **adjust the causal discovery method** to make it more robust against unfaithfulness?
  - Adjacency unfaithfulness
  - Orientation unfaithfulness

Adjacency Faithfulness Violation

- Adjacency-Faithfulness: For every $X, Y \in V$, if X and Y are adjacent in the true causal DAG, then they are not independent conditional on any subset of $V \backslash \{X, Y\}$.



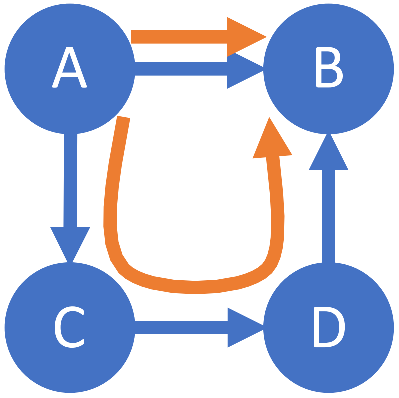
True Graph

$A \perp D \mid \{C\}$
$C \perp B \mid \{A, D\}$

The distribution satisfies one extra independence
$A \perp B \mid \emptyset$

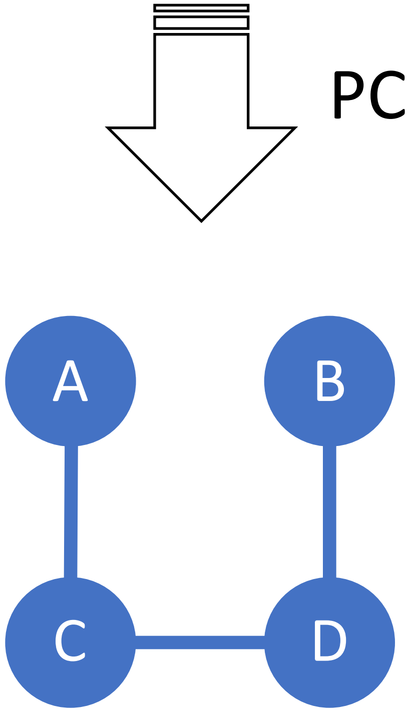
PC under Adjacency Faithfulness Failure

1. **Adjacency step**: for every pair of variables X and Y, search for a set of variables given which X and Y are conditionally independent, and infer them to be adjacent if and only if no such set is found.
   - Justified by adjacency faithfulness assumption
2. **Orientation step**: for every unshielded triple $(X;\ Y;\ Z)$, infer that it is a collider if and only if the set found in step 1 that renders X and Z conditionally independent does not include Y
   - Justified by orientation faithfulness assumption



True Graph

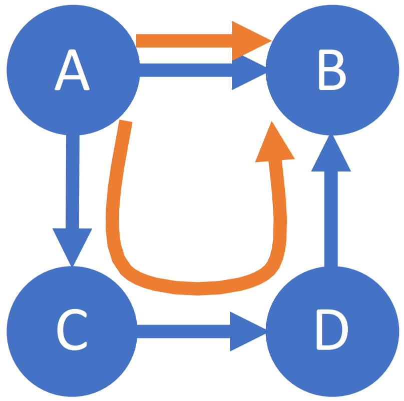
P: $A \perp B \mid \emptyset$



GES

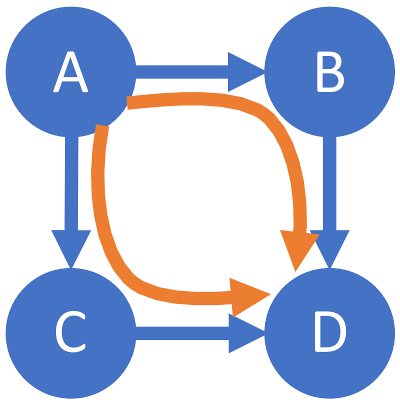
- Searches for a pattern that maximizes a score over the space of patterns
- Proceeds from one pattern to a neighbor by adding or removing edges, one at a time
- Forward phase:
  - Greedily add edges until the score cannot improve further
- Backward phase:
  - Remove edges until the score cannot improve further

- GES seems to be robust against Adjacency unfaithfulness



Orientation Faithfulness Violation

- Orientation-Faithfulness: For every unshielded triple (X, Y, Z)
  - If $X \rightarrow Y \leftarrow Z$ is a collider, then X and Z are not conditionally independent given any subset of $V \backslash \{X, Z\}$ that includes Y.
  - Otherwise, X and Z are not conditionally independent given any subset of $V \backslash \{X, Z\}$ that excludes Y.



True Graph

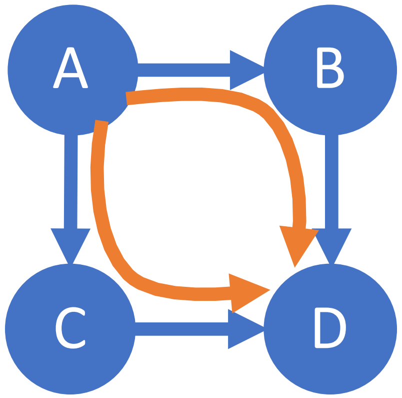
$A \perp D \mid \{B, C\}$

$B \perp C \mid \{A\}$

The distribution satisfies one extra independence

$A \perp D \mid \emptyset$

GES under Orientation Faithfulness Violation



True Graph

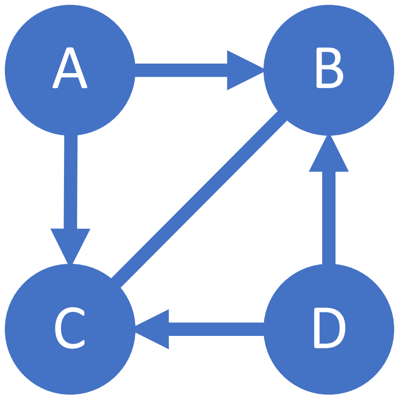
$A \perp D | \{B, C\}$

$B \perp C \mid \{A\}$

The distribution satisfies one extra independence

$A \perp D | \emptyset$

GES



α −Conservative Orientation

- Given a skeleton and a unshielded triple therein, consider all subsets of the variables adjacent to X or of the variables that are adjacent to Z that render (X, Z) consitionally independent

$$r = \frac{number\ of\ sets\ that\ include\ Y}{number\ of\ sets}$$

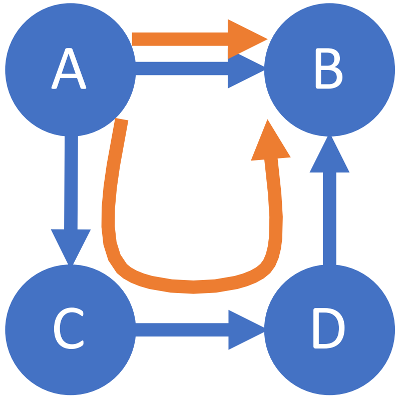
  - If $r \leq \alpha$, the triple is marked as a collider.
  - If $r \geq 1 - \alpha$, the triple is marked as a non-collider.
  - Otherwise it is ambiguous
- CPC(Ramsey et al, 2006) : $\alpha = 0$: too cautious
- Majority rule orientation(Colombo and Maathuis, 2014) : $\alpha = 0.5$: not conservative enough
- We use $\alpha = 0.4$

Proposed Hybrid Methods

- PC+GES
  - Run PC first, use the output pattern as a starting point for GES
  - Mitigate PC's vulnerability to adjacency faithfulness violations
- GES+c
  - Run GES first, then apply the α-conservative orientation rules and Meek's orientation rules(Meek, 1996)
  - Mitigate GES's vulnerability to orientation faithfulness violations
- PC+GES+c
  - Run PC+GES first, then apply the α-conservative orientation rules and Meek's orientation rules(Meek, 1996)
  - Mitigate both vulnerabilities

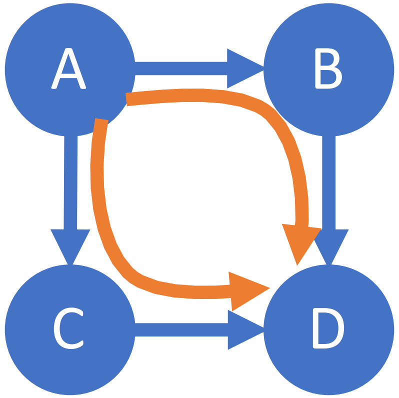
Simulations – Examples of exact Faithfulness violations

Adjacency unfaithfulness



	PC	PC-stable	PC+GES	GES	MMHC
True adj. rate	0.75	0.75	0.95	0.93	0.76
False adj. rate	0.01	0.01	0.02	0.06	0.02

Orientation unfaithfulness



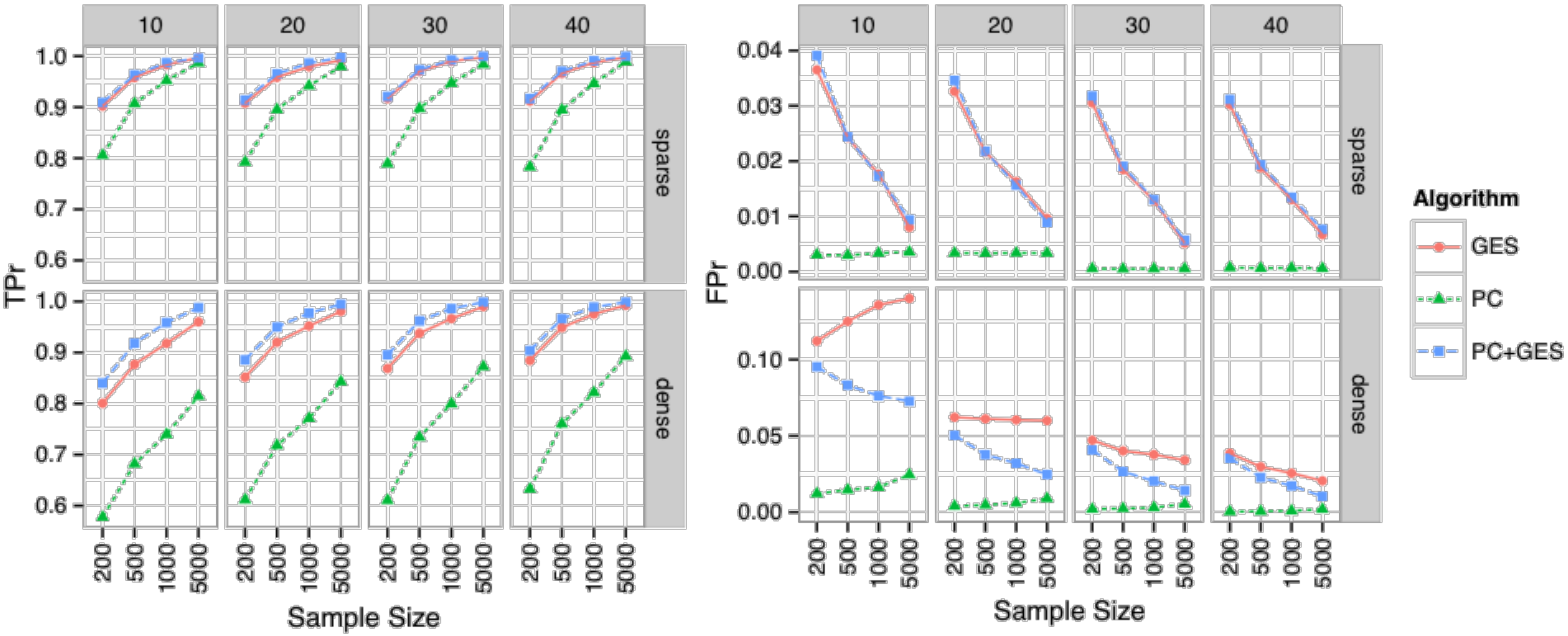
GES	GES+c	PC	CPC	MMHC
0.35	0.96	0.49	0.99	0.56

Mean Arrow Precision

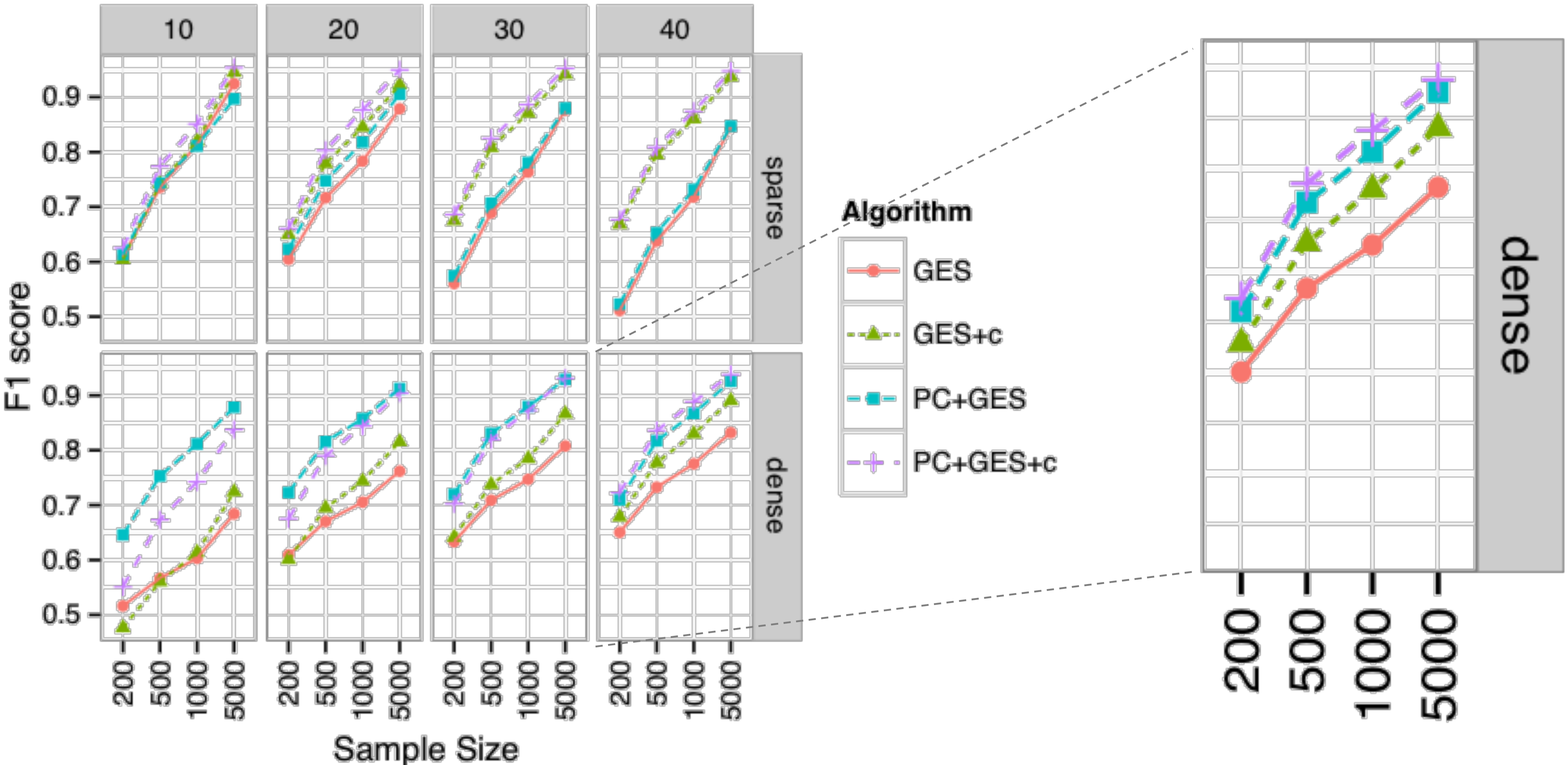
More comprehensive simulations(without exact unfaithfulness)

- Number of variables (dimension) $\in \{10, 20, 30, 40\}$
- Expected vertex degree (sparsity) $\in \{2, 4\}$
- Sample size $\in \{200, 500, 1000, 5000\}$
- For each setting, 100 random DAGs are generated, and on each DAG a linear Gaussian model is randomly built:
  - Edge coefficients are uniformly drawn from [-1, -0.1] ∩ [0.1, 1]
  - Variances of error terms are uniformly drawn from [0.5, 1]
- From each model, 50 datasets at each sample size are generated.

Adjacency on Random Graphs



Orientation on Random Graphs



Conclusion and Outlook

- PC and GES are vulnerable to violations of Faithfulness
- Heuristic hybrid algorithms shown to be able to mitigate some adjacency and orientation issues
  - even if faithfulness is not exactly violated

- Try to develop efficient methods for causal inference under weaker faithfulness assumptions (e.g. triangle faithfulness)