Hybrid causal search in latent variable models

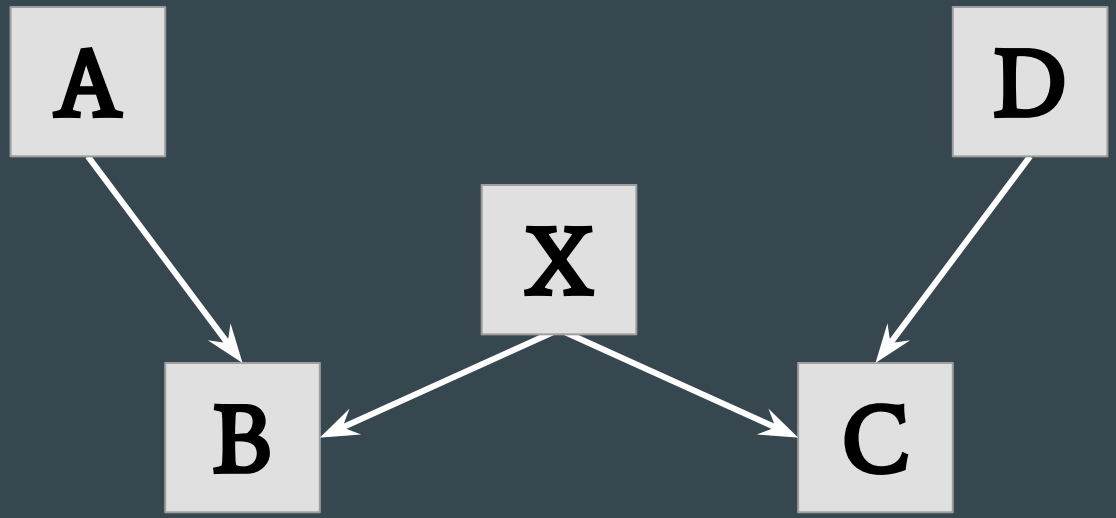
Juan Miguel Ogarrio
jogarrio@andrew.cmu.edu

Peter Spirtes
ps7z@andrew.cmu.edu

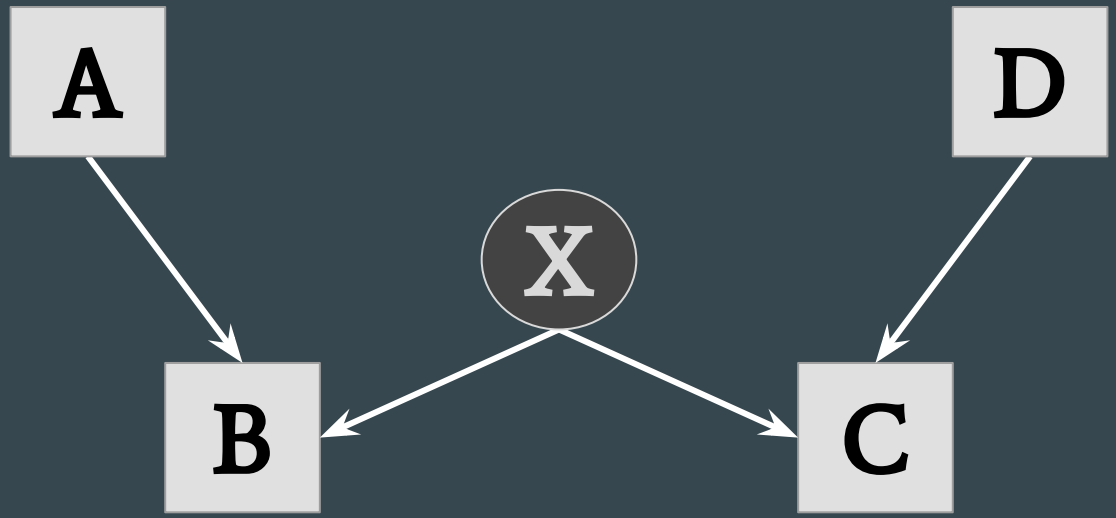
Joseph Ramsey
jdramsey@andrew.cmu.edu

Department of Philosophy
Carnegie Mellon University

DAG



DAG - with latents



Causal search limitations

- Few or no latents
- Strong model assumptions (e.g. linearity, n-factor models)
- Inaccurate with small samples

Outline

- Preliminaries
  - Assumptions
  - Patterns and PAGs
  - GES, FCI
- Greedy Fast Causal Inference
- Simulations

Assumptions

- Partition into observed and latent variables
  - Can only operate on observed variables
- i.i.d. sample, no selection bias (can be relaxed)

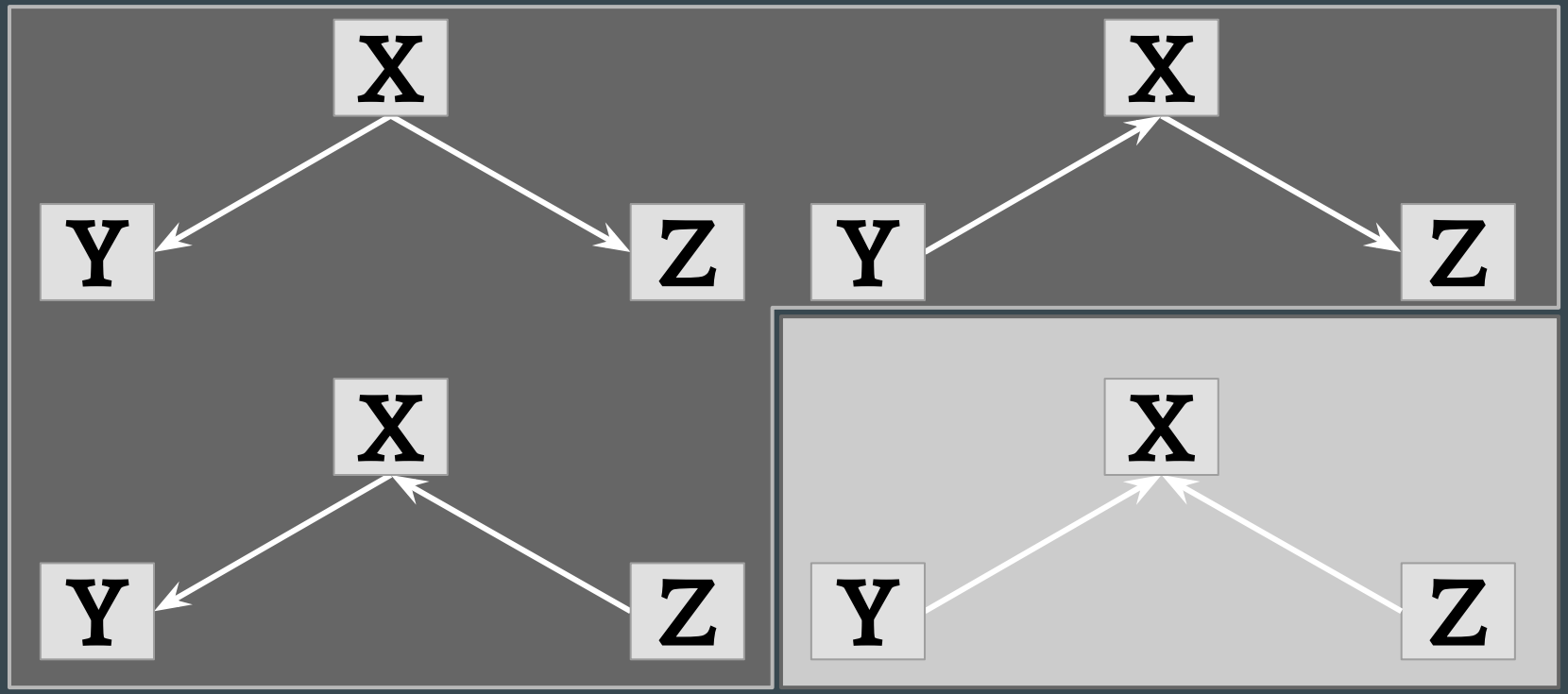
Assumptions

- Local Causal Markov assumption
  - d-separation → conditional independence
- Causal Faithfulness assumption
  - d-separation ← conditional independence

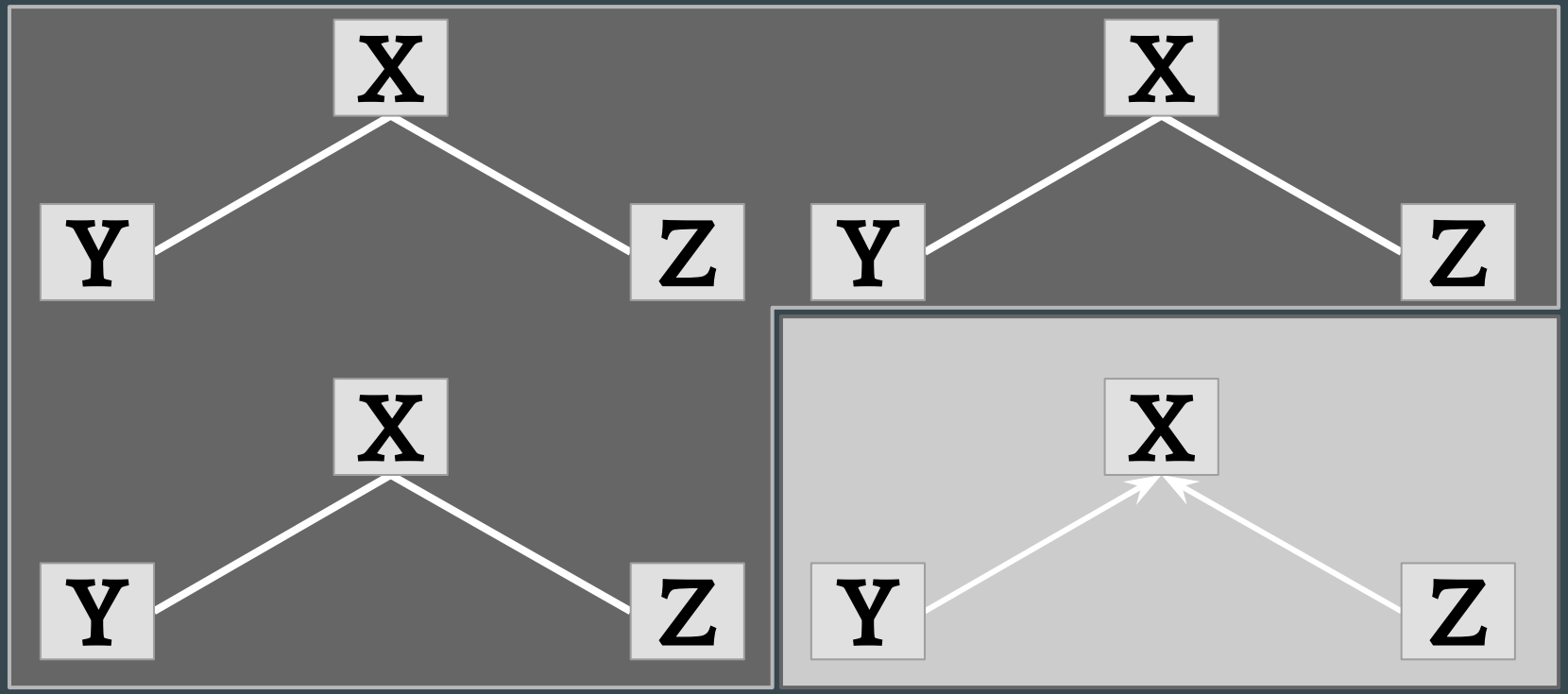
Underdetermination

- True model contains no latents
  - Pattern - Markov Equivalence class
- True model may contain latents
  - Partial Ancestral Graph (PAG) - Observational ME class

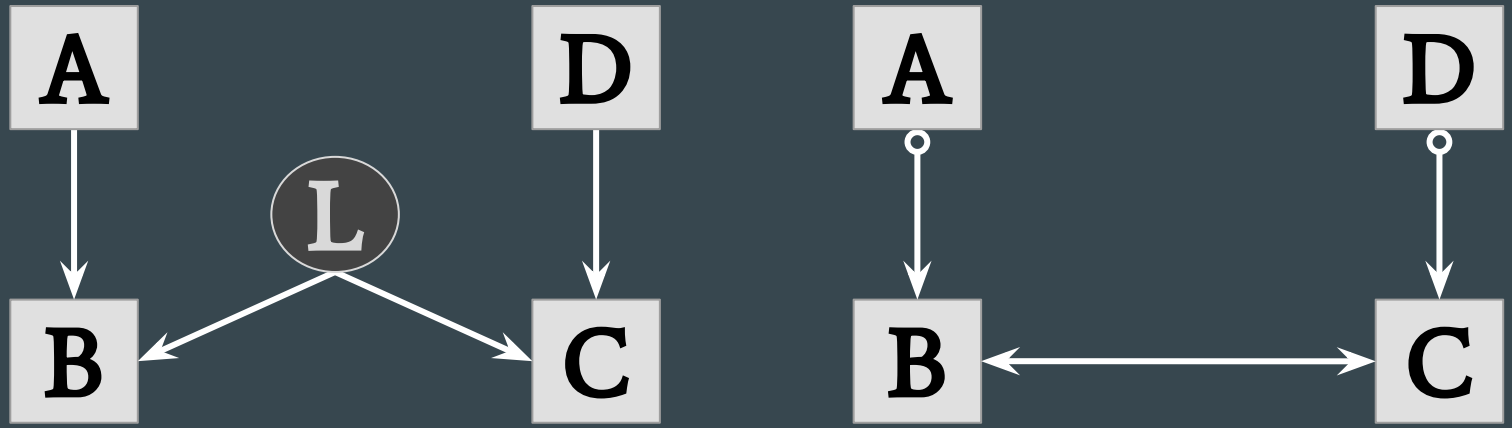
Patterns - examples



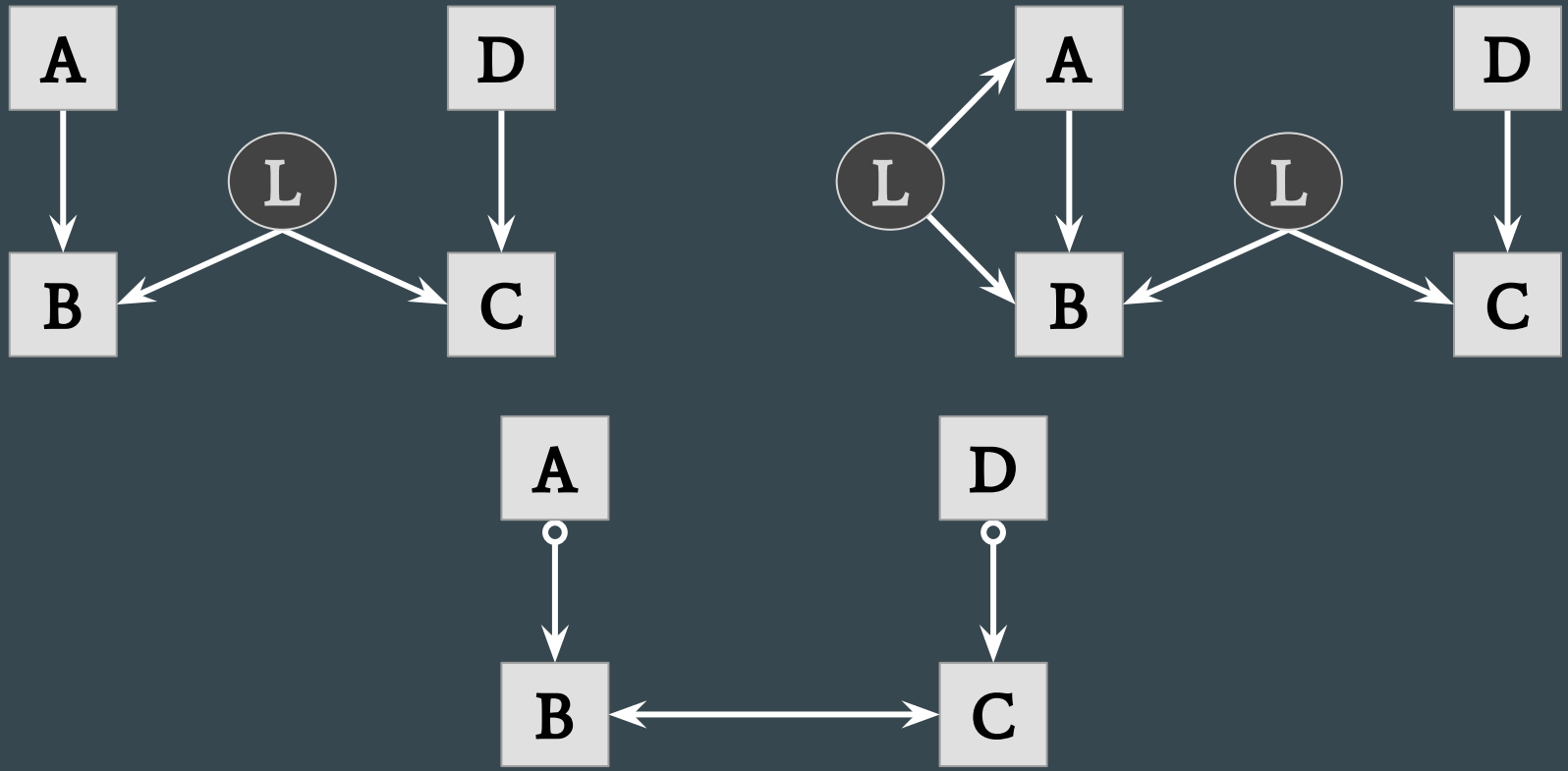
Patterns - examples



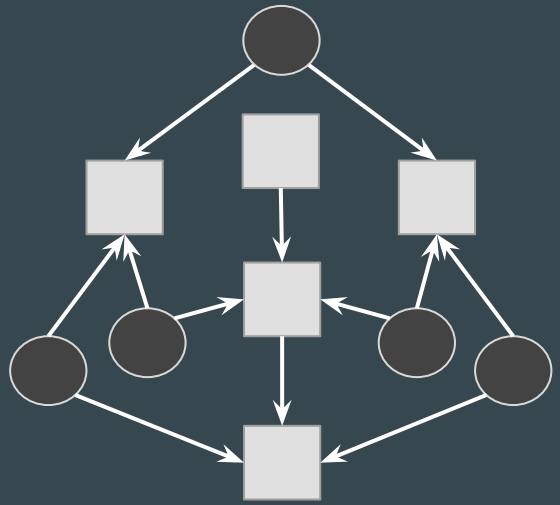
PAG example

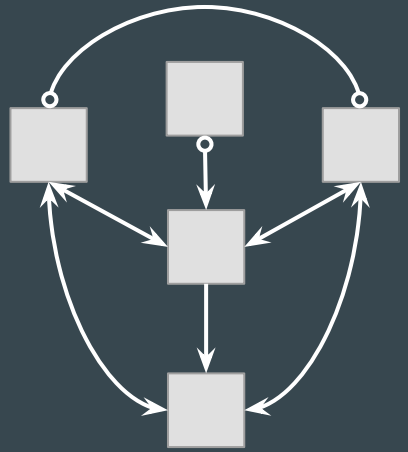


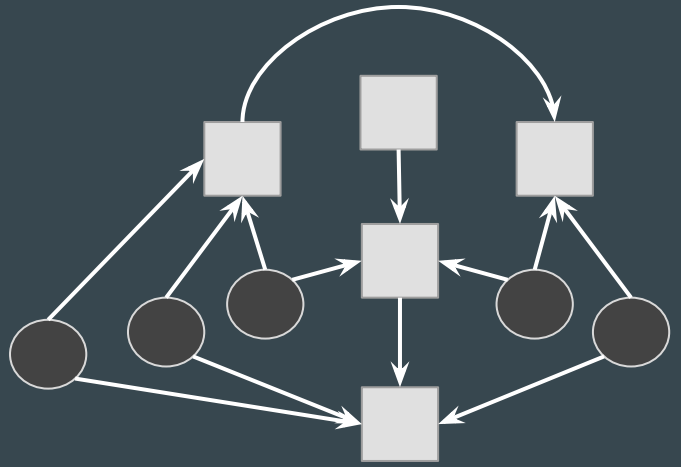
O-equivalent DAGs



O-equivalent DAGs







State of the art algorithms: Score-based

GES (Greedy Equivalence Search)

- Score-based
  - Bayesian Information Criterion (BIC) Score
- Outputs Markov and minimal pattern
- Fast Greedy Search (FGS)
  - Optimized version

GES (Greedy Equivalence Search)

- Efficient traversal of search space
  - Two phases
    - Forward phase
      - Single directed edge additions
      - Output Markov

GES (Greedy Equivalence Search)

- Efficient traversal of search space
  - Two phases
    - Forward phase
    - Backward phase
      - Single directed edge removals
      - Markov input, Markov and minimal output

GES (Greedy Equivalence Search)

- Pros
  - Fast
  - Accurate
  - No latents → Markov and faithful pattern
- Cons
  - Latents → Markov and minimal pattern
  - BIC score limited to some distributions

State of the art algorithms: Constraint-based

FCI (Fast Causal Inference)

- Constraint-based algorithm
  - Can use any conditional independence test
- Outputs Markov and faithful PAG
  - Can account for latents and selection bias
- Many variants/modifications
  - RFCI, FCI+

Fast Causal Inference (FCI)

- Overview
  - ➢ Initial adjacency phase
    - → Pre-orientation phase
    - → Final adjacency phase
    - → Final orientation phase
    - → PAG

Fast Causal Inference (FCI)

- Pros
  - Can use any independence test
  - Accounts for latents
- Cons
  - Inaccurate in practice
  - Bad worst-case performance
  - Non-parametric independence tests are slow

Improvement: pre-process data to improve initial stages of FCI

GFCI (Greedy FCI)

- Hybrid, same assumptions as FCI
- Overview
  - Preprocess data using GES
  - Feed adjacency and unshielded triples to FCI
  - Proceed with FCI
  - Output PAG

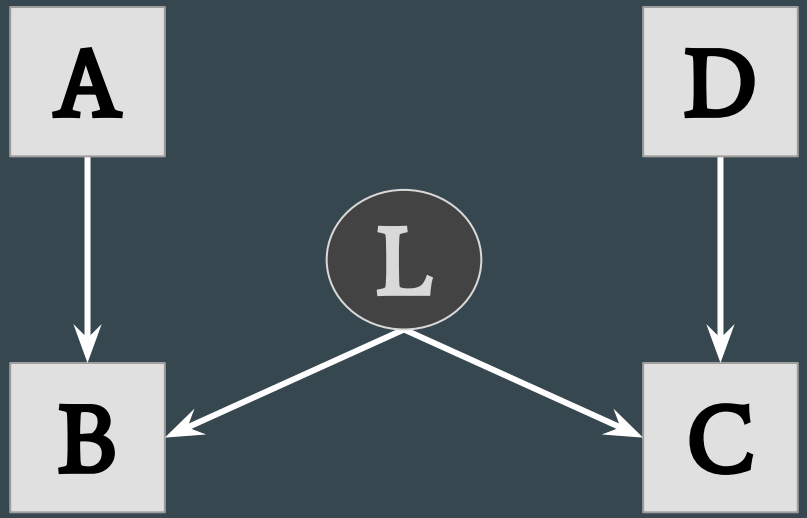
GFCI - GES preprocessing

- GES output:
  - Markov and minimal
    - No faithfulness assumption
  - Contains superset of adjacencies
  - Unshielded triples reflect triples in PAG

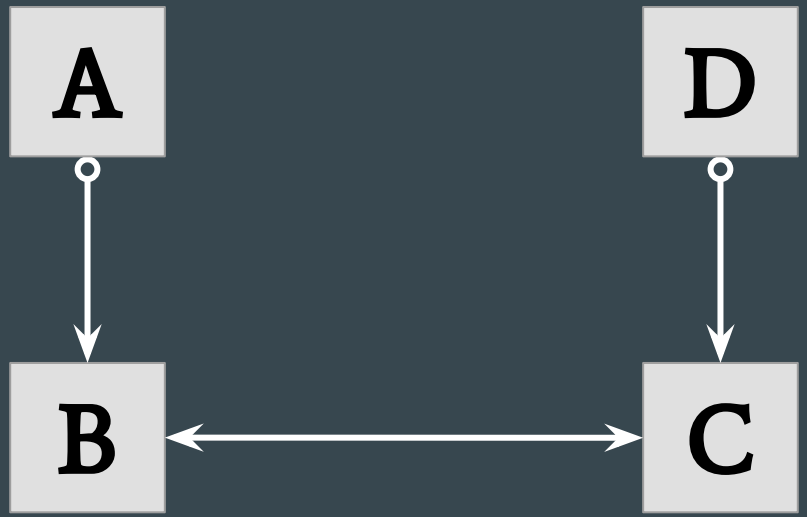
GFCI - FCI stage

- Remove additional adjacencies
- Copy unshielded triples
- Finalize orientations

GFCI example - DAG



GFCI example - true PAG



GFCI example - GES phase

A

D

B

C

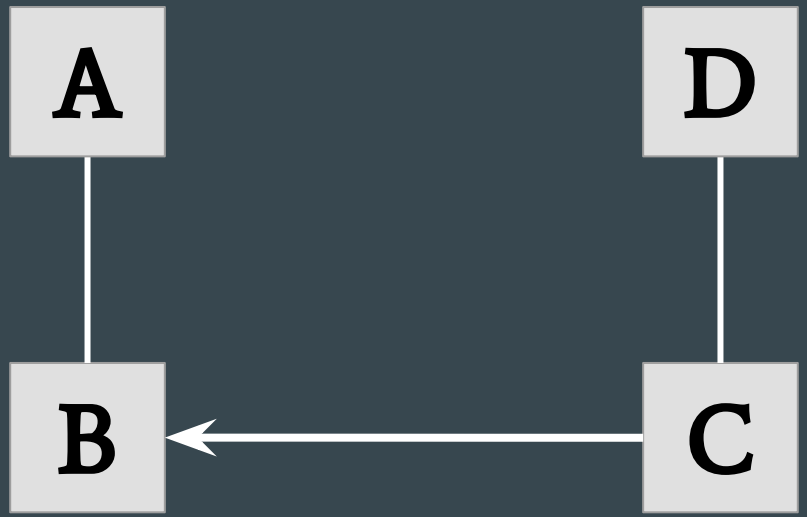
GFCI example - GES phase

GFCI example - GES phase

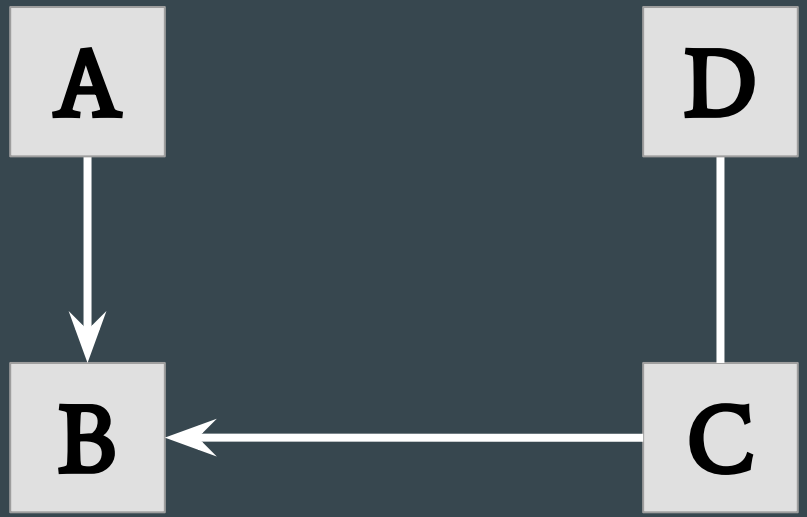
GFCI example - GES phase

GFCI example - GES phase

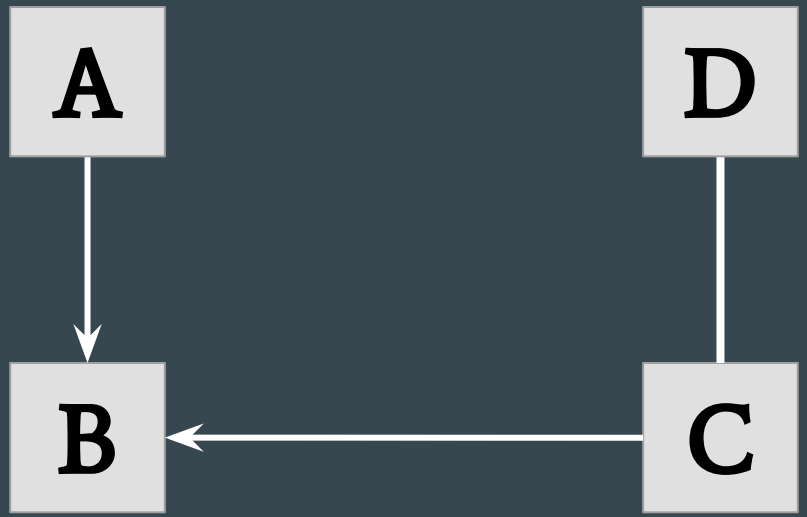
GFCI example - GES phase



GFCI example - GES phase

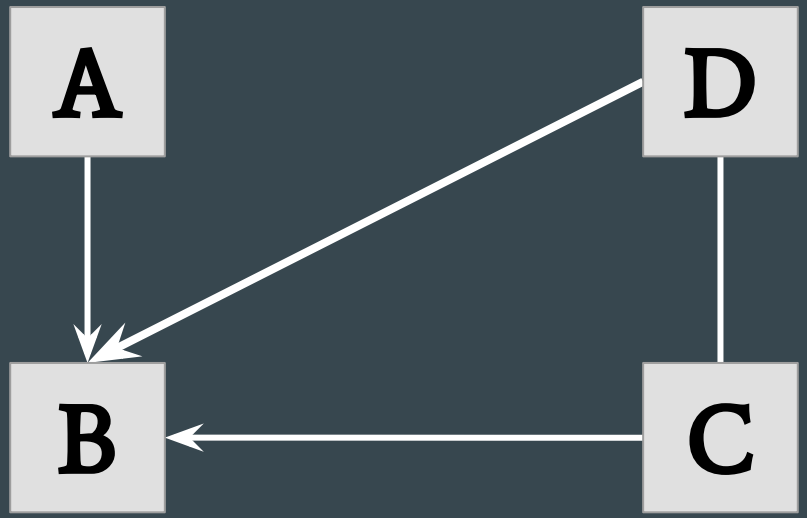


GFCI example - GES phase

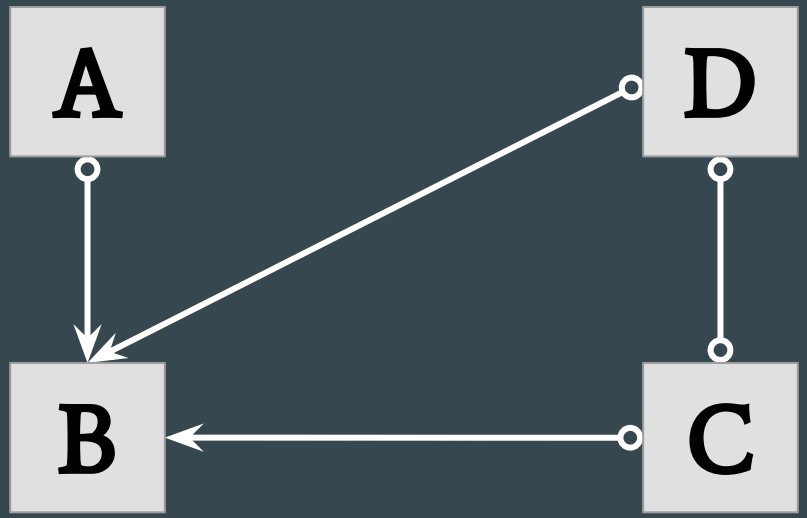


$B \perp D \mid C$

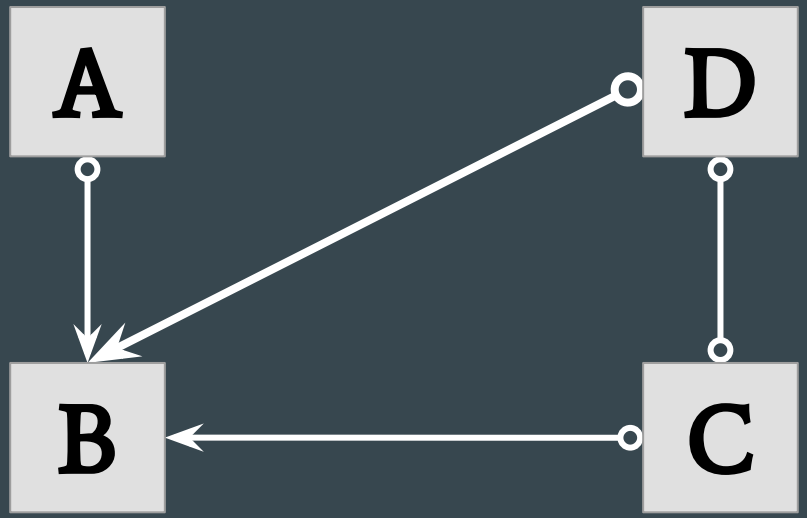
GFCI example - GES phase



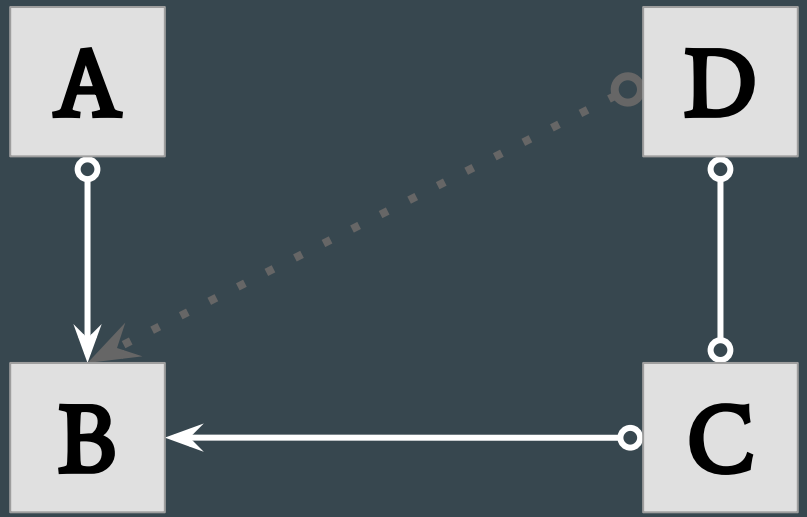
GFCI example - FCI phase



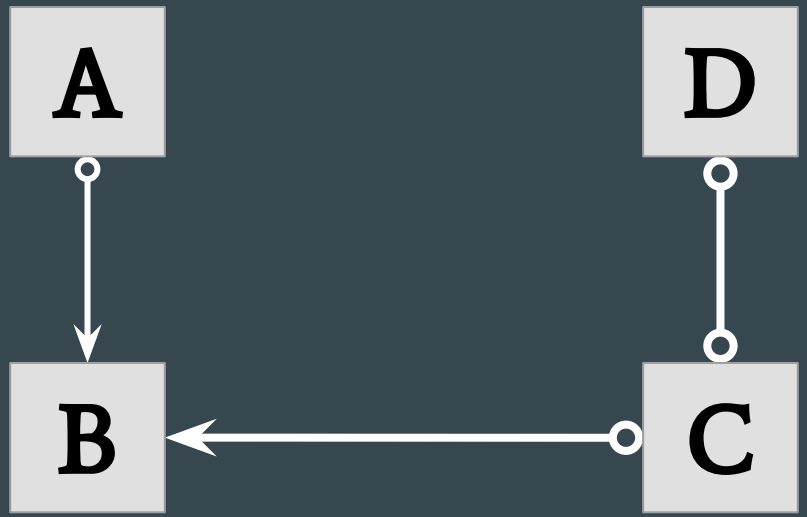
GFCI example - FCI phase



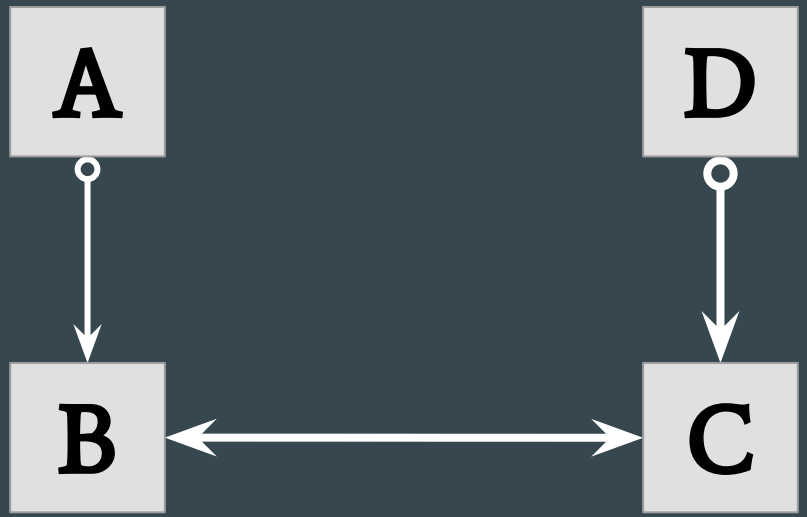
GFCI example - FCI phase



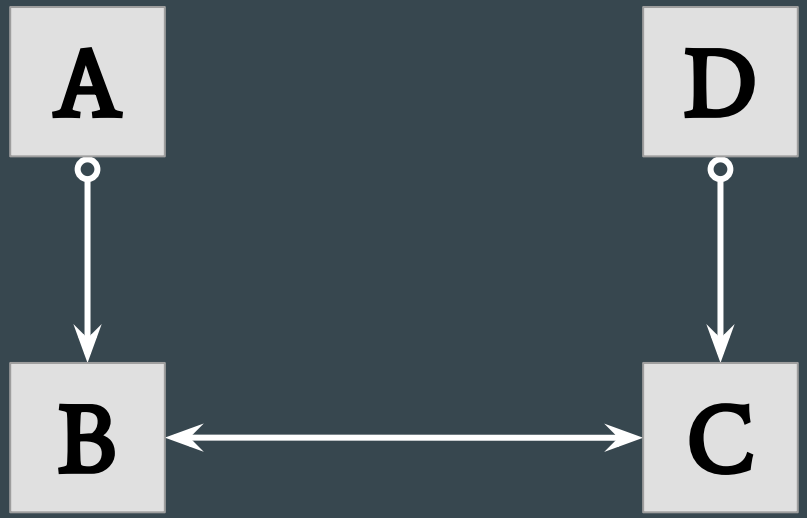
GFCI example - FCI phase



GFCI example - FCI phase



GFCI example - FCI phase



Simulations

Experiments

- Compare accuracy/speed
- FCI, RFCI, FCI+, GFCI
- Linear Gaussian models

Experiments - parameters

- Graphs
  - DAG sizes: 100, 1000 nodes
  - DAG connectivity: 100, 200 edges / 1000, 2000 edges
  - Latents: 5, 20 latents / 50, 200 latents
- Linear Gaussian models
- Fisher's Z-test with $\alpha : \{0.01, 0.05, 0.1\}$
- Around 100 trials per parametrization

Experiments - parameters

- Algorithms
  - FCI
  - RFCI
  - FCI+
  - uRFCI (unbounded RFCI)
  - GFCI (penalty = 4)

Experiments - results

- Accuracy
  - GFCI substantially better
  - All algorithms struggle finding ↔ adjacencies
- Speed
  - GFCI not as fast, but no practical difference
    - Can be sped up
    - Scaling?

Questions/Comments

jogarrio@andrew.cmu.edu

ps7z@andrew.cmu.edu

jdramsey@andrew.cmu.edu

Code available as part of the Tetrad program

http://www.phil.cmu.edu/tetrad/

https://github.com/cmu-phil/tetrad

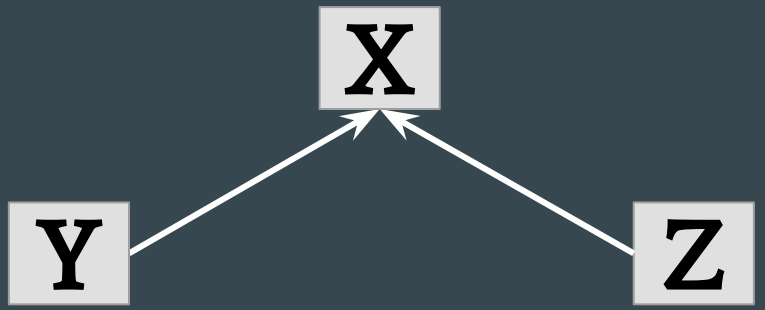
Additional slides

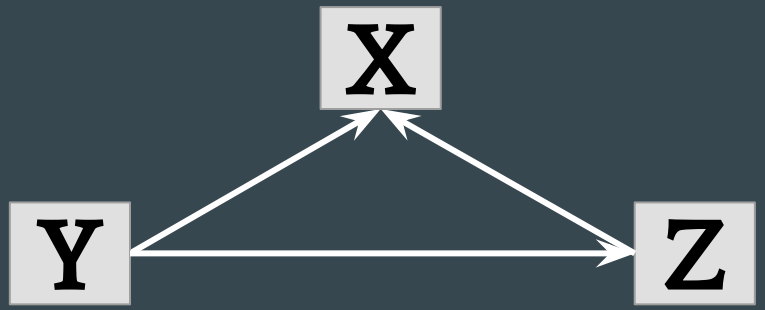
Patterns

- Same nodes as DAG
- Directed edges
    - Shared by every DAG
- Undirected edges
    - Distinct for at least two DAGs

Patterns - characterization

- DAG's pattern/ME class determined by:
    - adjacencies
    - unshielded colliders



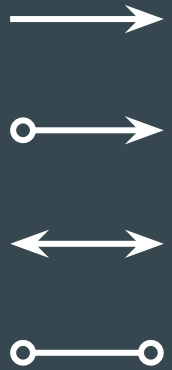


PAG

- Nodes same as O nodes in DAG
- Edges
  - Nodes cannot be d-separated
- Edge marks
  - Arrow : non-ancestor in every DAG
  - Tail : ancestor in every DAG
  - Circle : ancestor in some, non-ancestor in some

PAG

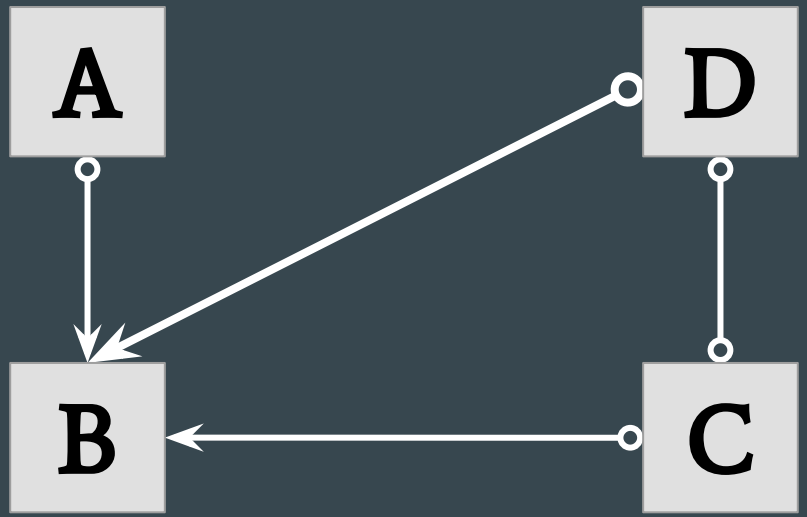
- Edge types
  - Directed
  - Semi-directed
  - Bi-directed
  - Undirected



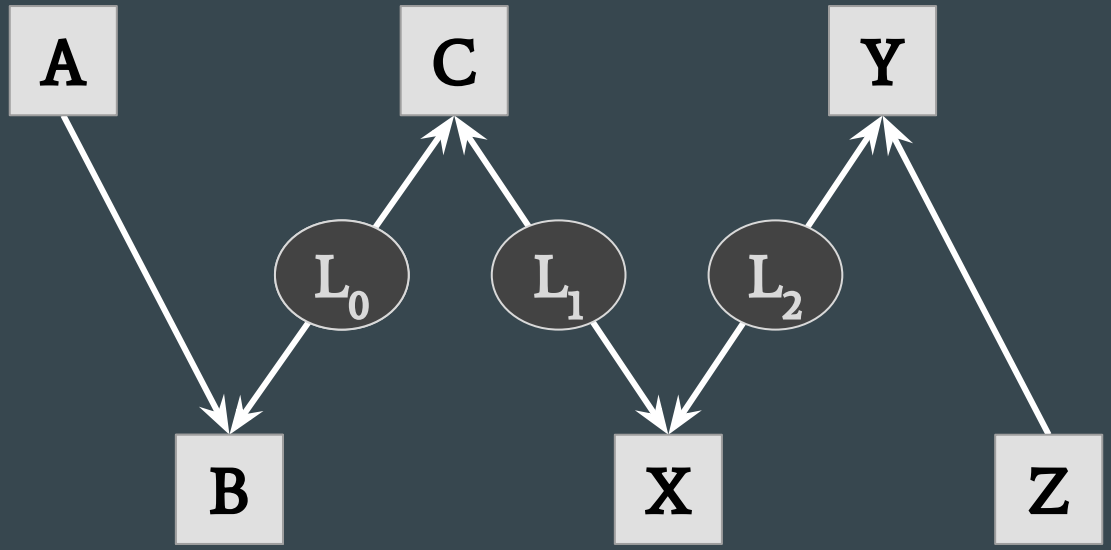
GFCI - Triangle conjecture

- Optimization
  - Additional adjacencies found inside of triangles

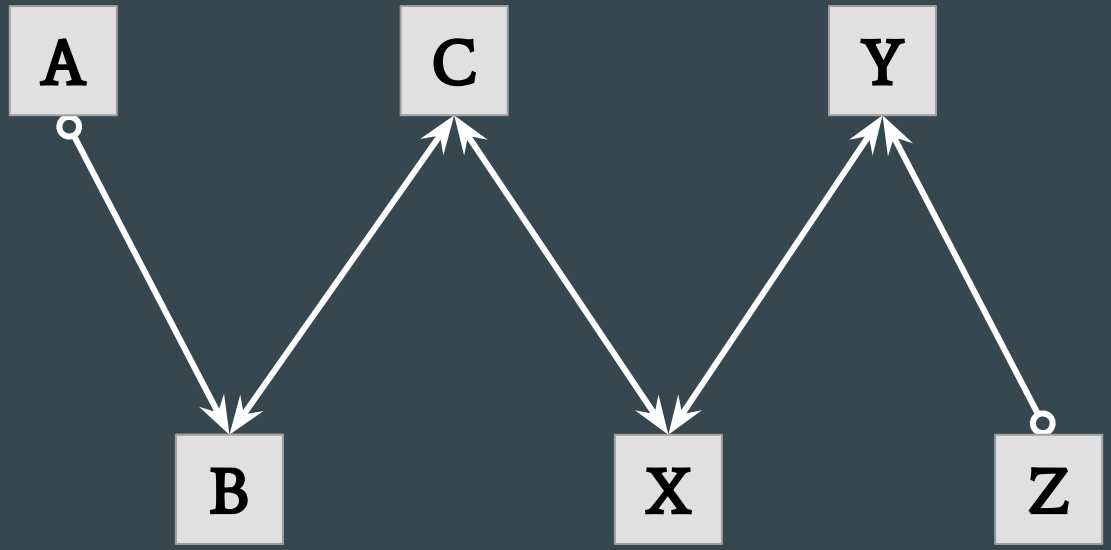
Triangle conjecture



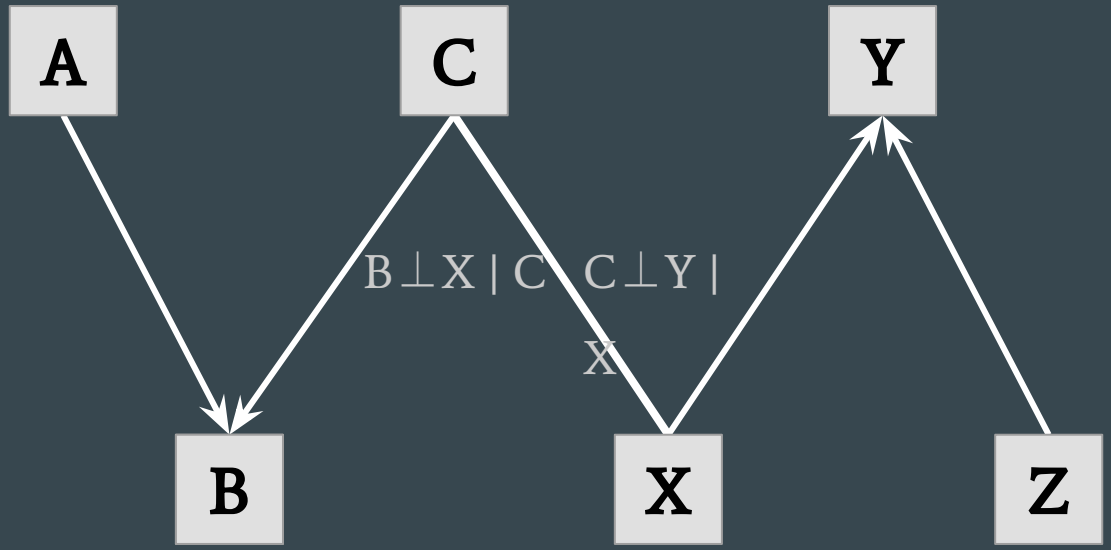
GFCI example - DAG



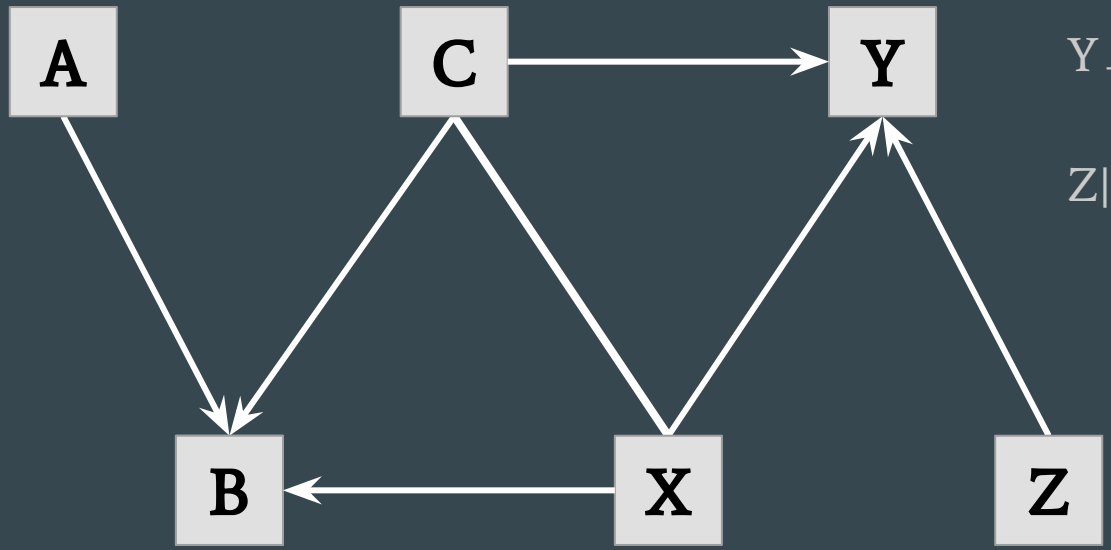
GFCI example - true PAG



GFCI example - intermediate step



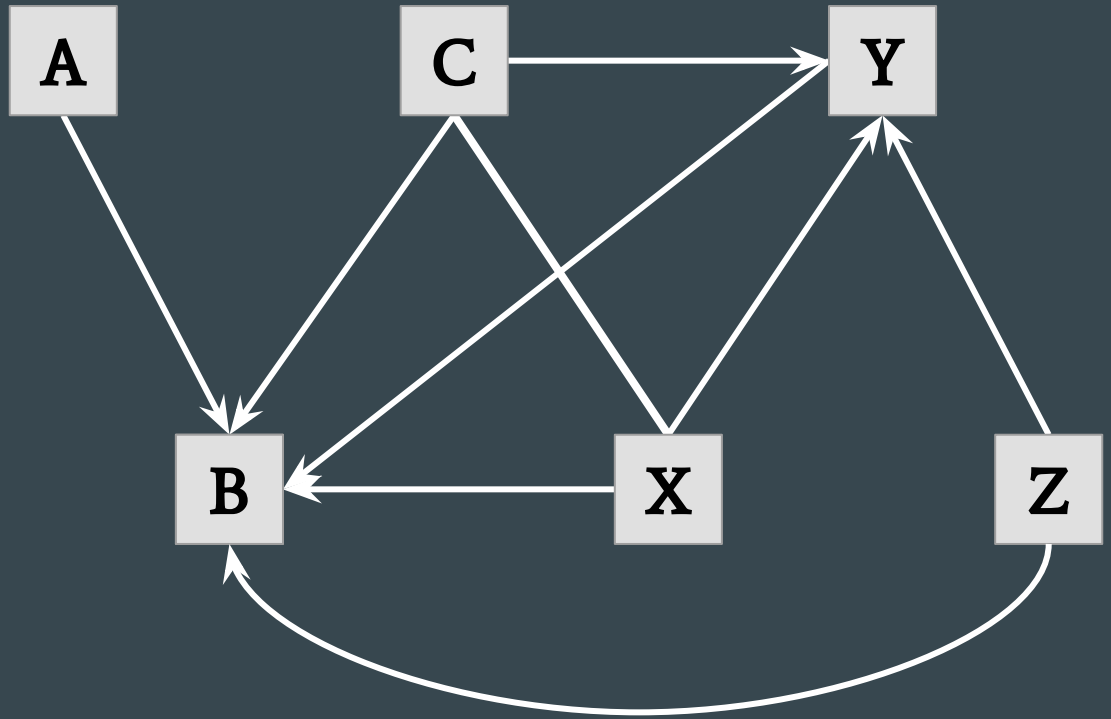
GFCI example - intermediate step



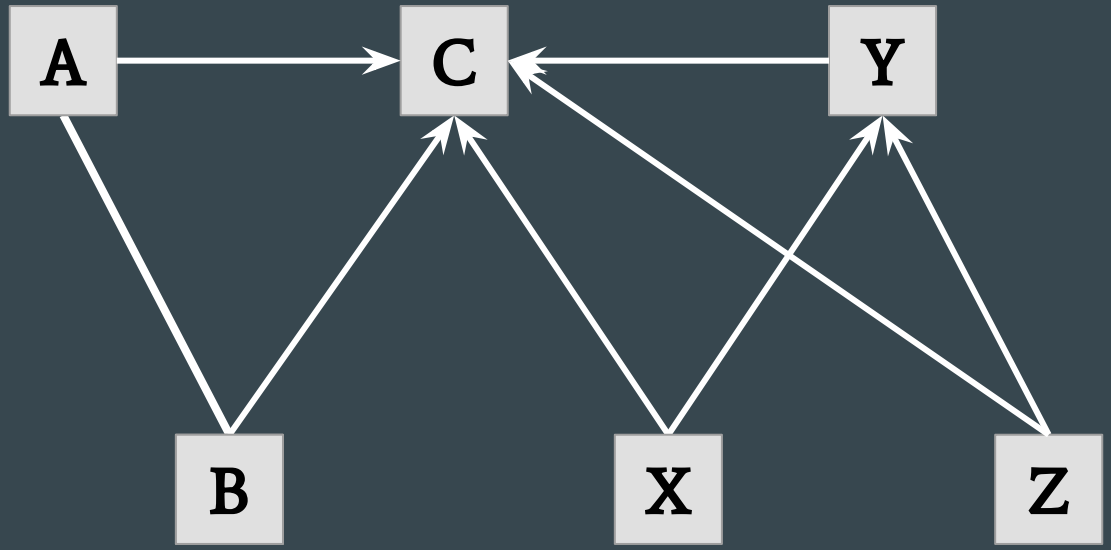
$Y \perp B \mid \{C, X\}$

$Z|B \mid \{C,X,Y\}$

GFCI example - additional edges



GFCI example - other parameters



Experiments - parameters

- Graphs
  - DAG sizes: 100, 1000 nodes
  - DAG connectivity: 100, 200 edges / 1000, 2000 edges
  - Latents: 5, 20 latents / 50, 200 latents

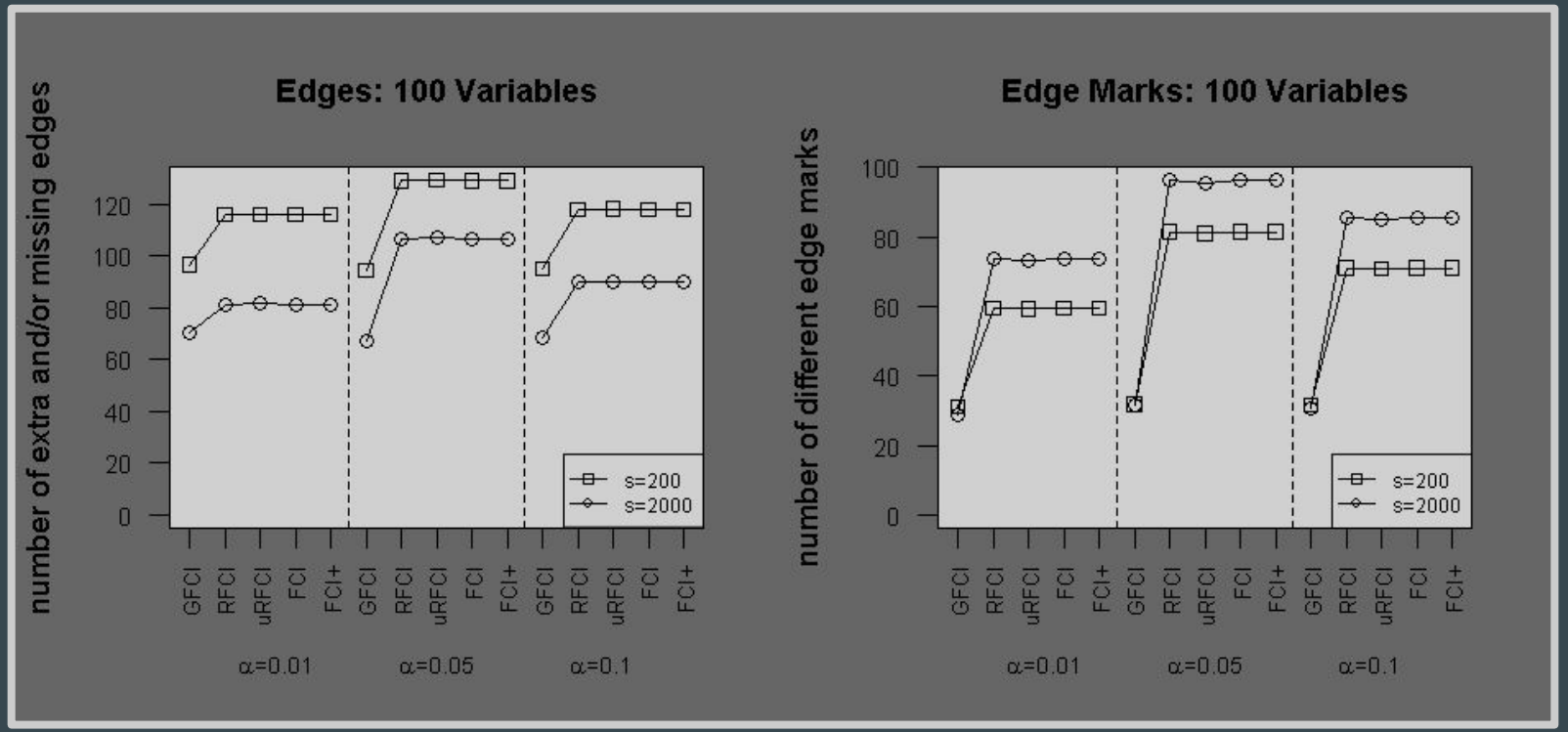
Experiments - parameters

- Models
  - Gaussian variables
    - Mean = 0
    - Variance : [1, 3]
  - Edges
    - Coefficients: ±[0.2, 1.5]

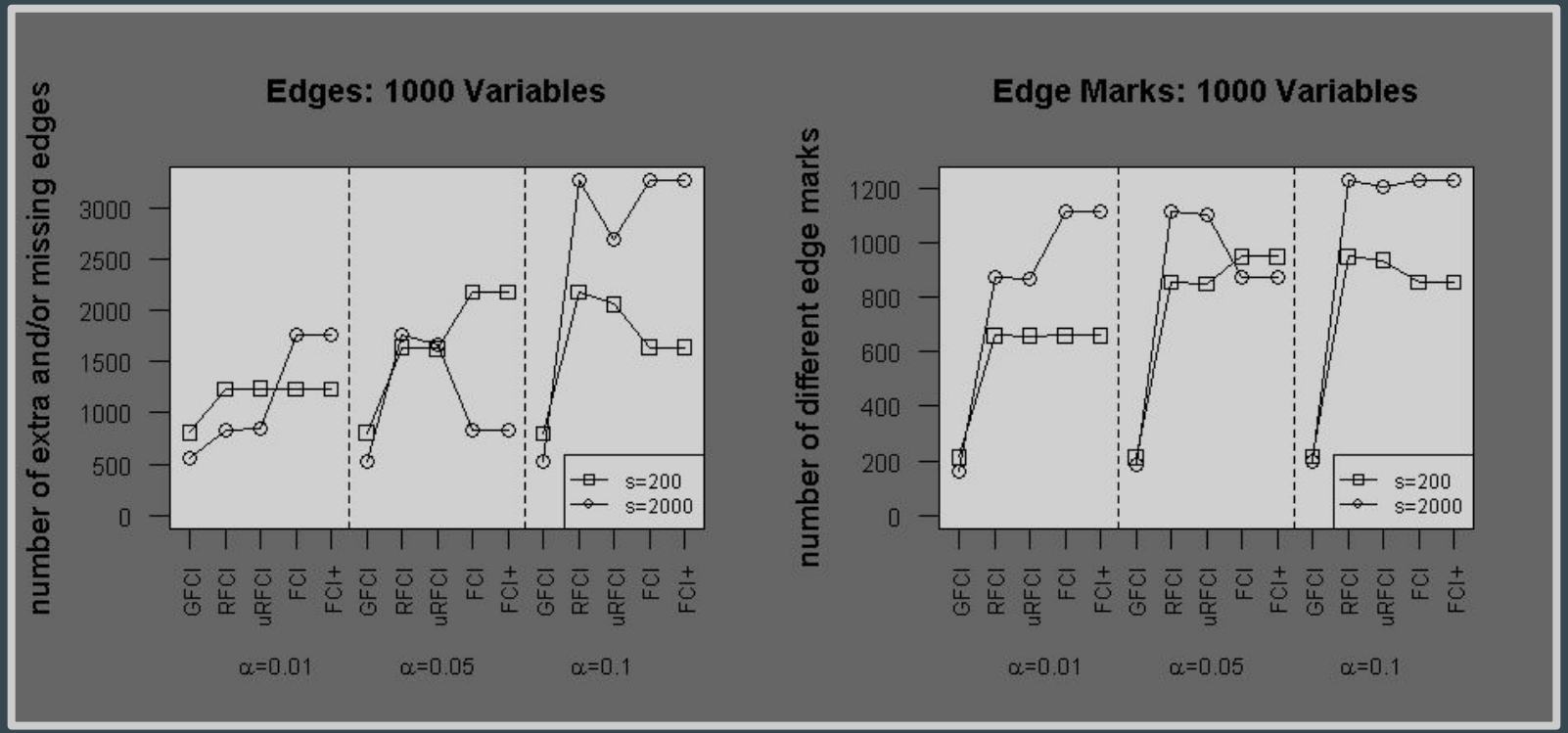
Experiments - parameters

- Independence test
  - Fisher's Z-test
    - $\alpha : \{0.01, 0.05, 0.1\}$
- Random samples
  - Size : 200, 2000
- Around 100 trials

Experiments - accuracy overview



Experiments - result overview



Experiments - time

100 variables, time in msec.					
	Min.	1st Qu.	Med.	3rd Qu.	Max.
GFCI	18	90	134	266	1002
RFCI	0	3	7	14	52
uRFCI	136	332	392	493	1568
FCI+	14	49	91	145	381

Experiments - time

1000 variables, time in sec.					
	Min.	1st Qu.	Med.	3rd Qu.	Max.
GFCI	2.330	4.700	6.390	12.760	29.010
RFCI	0.150	0.798	3.540	6.942	62.900
uRFCI	1.300	2.610	6.840	16.390	18530.000
FCI+	1.770	4.500	8.795	13.570	83.140