Marginal consistency of constraint-based causal learning

Anna Roumpelaki, Giorgos Borboudakis, Sofia Triantafillou and Ioannis Tsamardinos

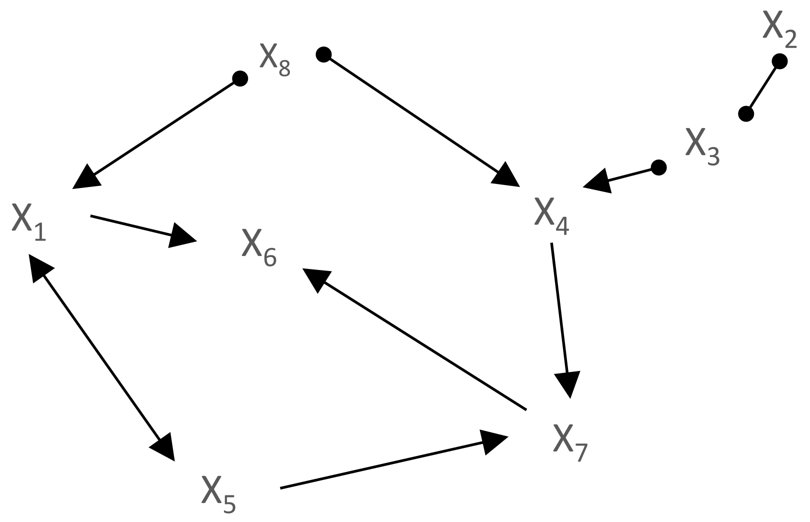
University of Crete, Greece.

Constraint-based Causal Learning

Measure $X_1, \dots, X_8$

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI

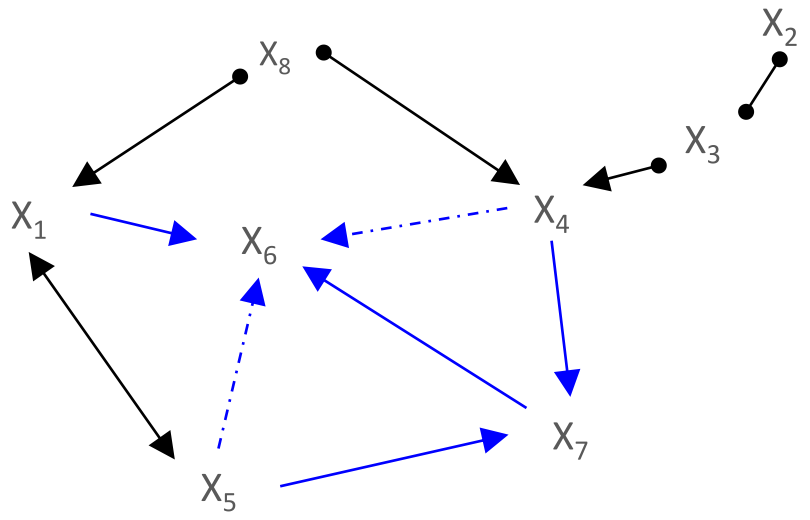


Constraint-based Causal Learning

Measure $X_1, \ldots, X_8$

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI

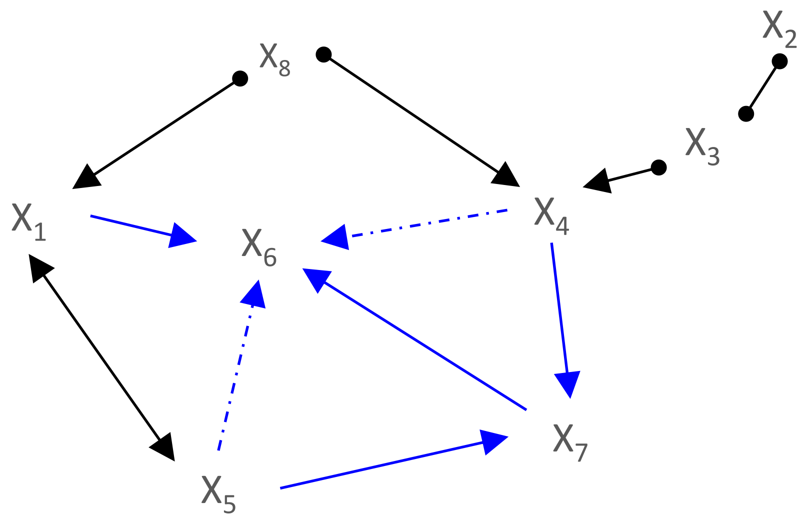


Constraint-based Causal Learning

Measure $X_1, \dots, X_8$

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI



How important is the choice of variables in learning causal relationships?

Would you have learnt the same relationships if you had chosen a subset of the variables?

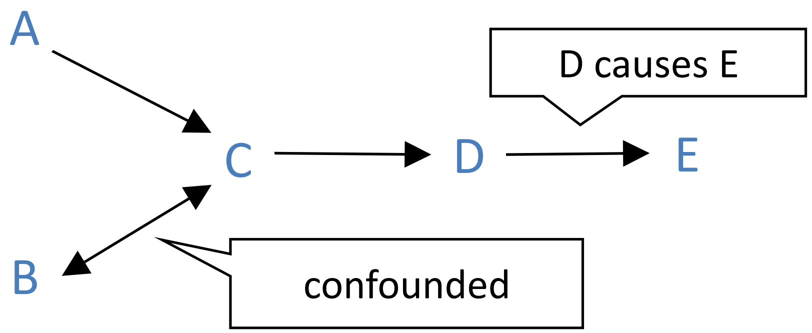
Maximal Ancestral Graphs

- Maximal Ancestral Graphs
- Capture the conditional independencies of the joint probability P over a set of variables under Causal Markov and Faithfulness conditions

- Directed edges denote causal ancestry.
- Bi-directed edges denote confounding.
- No directed cycles.
- No almost directed cycles $A \rightarrow B \rightarrow \cdots \rightarrow C \leftrightarrow B$.

- **Are closed under marginalization.**
  - $G = (V, E)$ is a MAG, $G[_L = (V \setminus L, E')$ is also a MAG.

- Can also handle selection (not here).

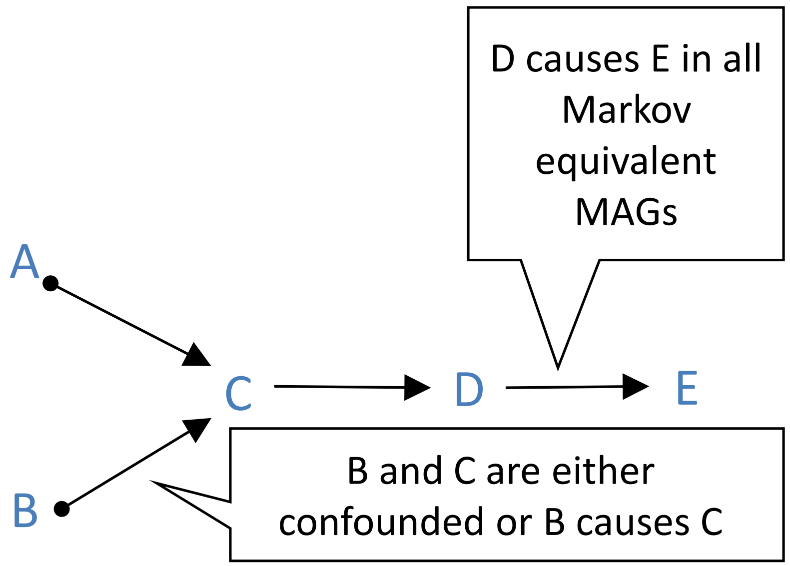


Partially Oriented Ancestral Graphs

- Summarize pairwise features of a Markov equivalence class of MAGs.
- Circles denote ambiguous endpoints.
- Can be learnt from data using FCI.

- FCI:
  - Sensitive to error propagation.
  - Order-dependent (if you change the order of variables in the data set you get a different result).

- Extensions of FCI [4]:
  - Order independent (iFCI) [1]:
    - output does not depend on the order of the variables.
  - Conservative FCI (cFCI) [3]:
    - Makes additional tests of independence for more robust orientations.
    - Forgoes orientations that are not supported by all tests.
  - Majority rule (mFCI) [3]:
    - conservative FCI that orients ambiguous triplets based on majority of test results for each triplet.

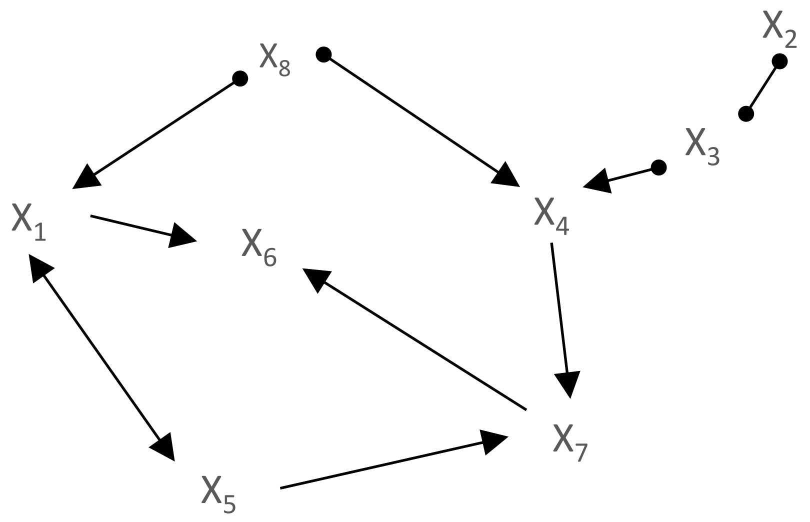


What happens when you marginalize out variables?

Measure $X_1, \dots, X_8$

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI

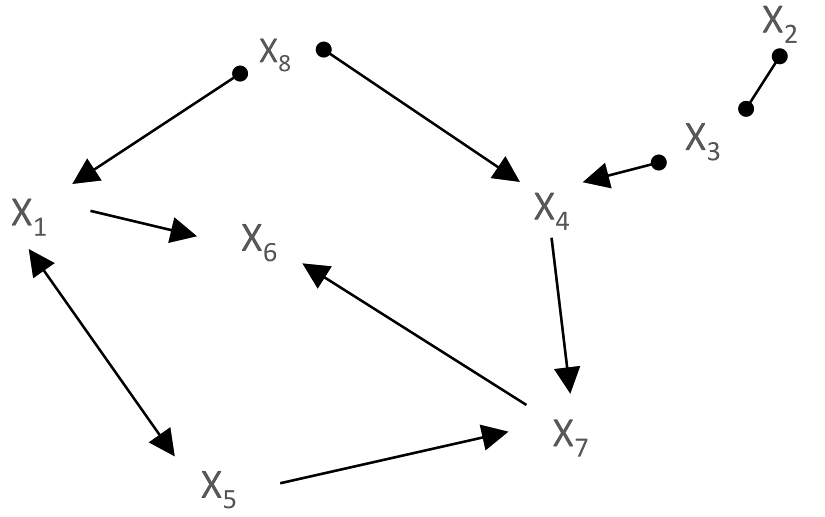


What happens when you marginalize out variables?

Measure $X_1, \ldots, X_8$

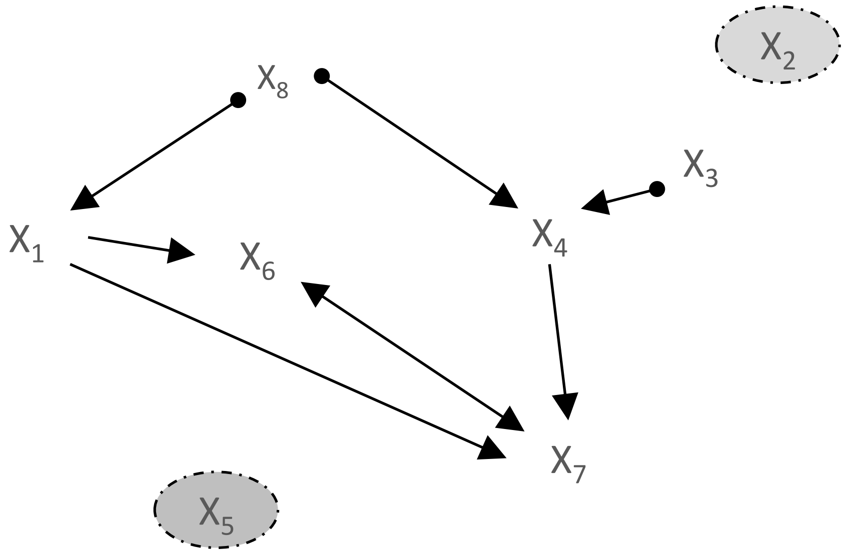
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI



marginal dataset without variables X_2, X_5

X_1	X_3	X_4	X_6	X_7	X_8

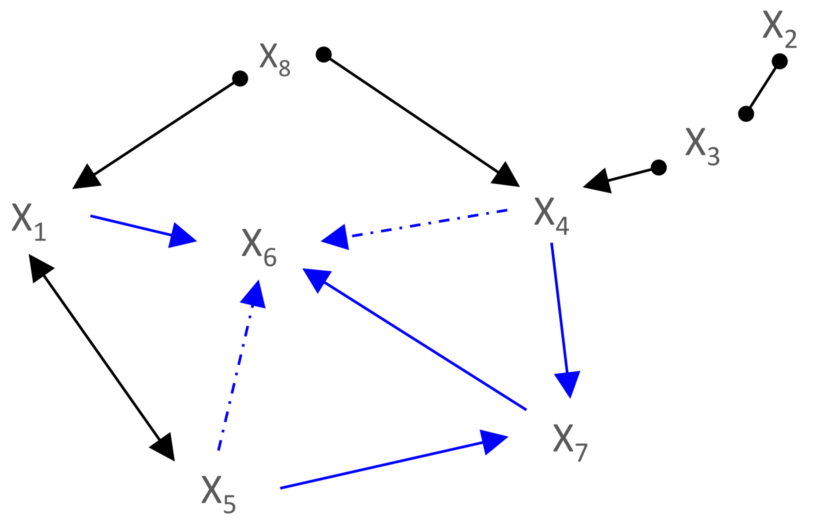


What happens when you marginalize out variables?

Measure $X_1, \ldots, X_8$

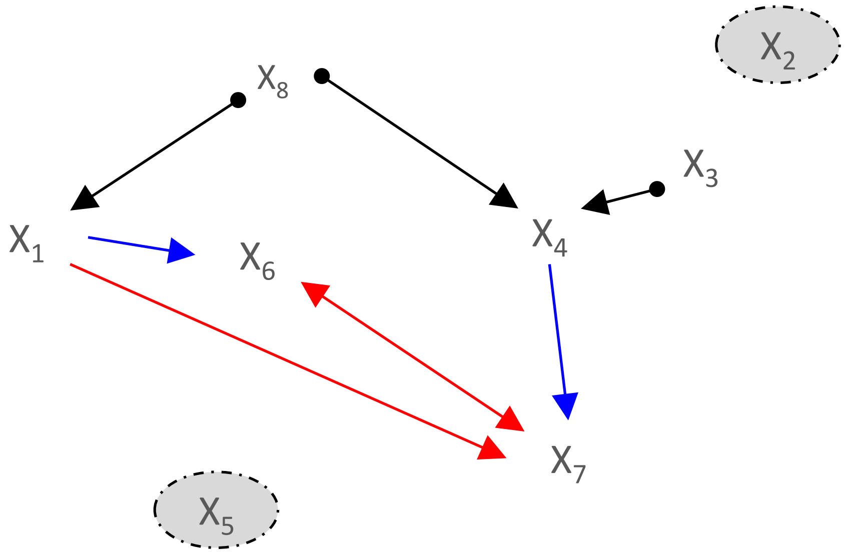
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8

FCI

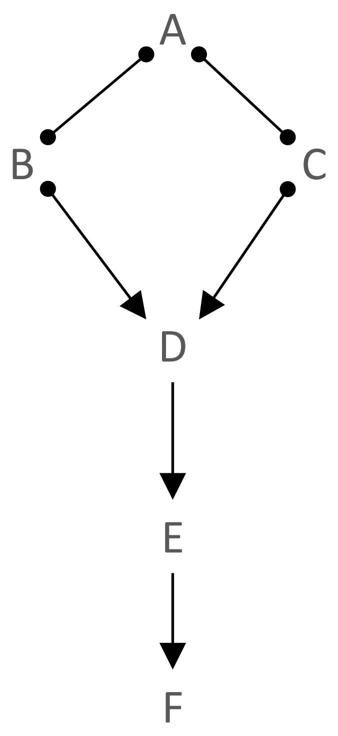


marginal dataset without variables X_2, X_5

X_1	X_3	X_4	X_6	X_7	X_8



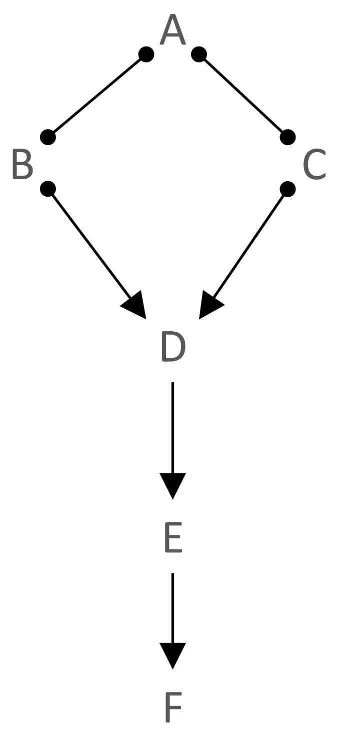
Finding $NAnc_P$



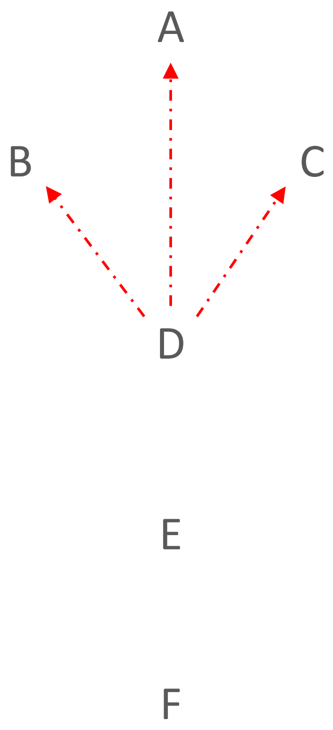
PAG P

- $(X, Y) \in NAnc_P$: X is **not an ancestor** of Y in all MAGs represented by P.
- X has **no potentially directed** path to Y.

Finding $NAnc_P$

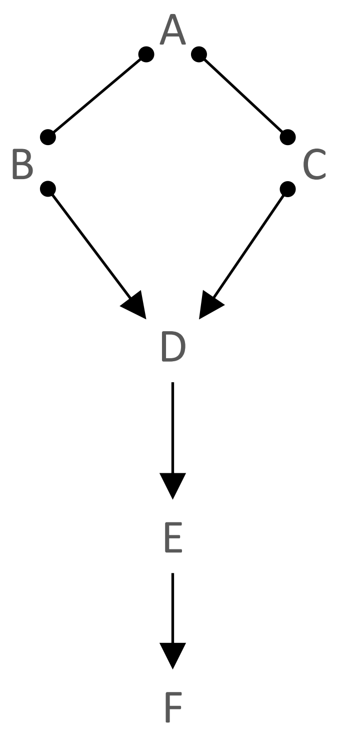


PAG P

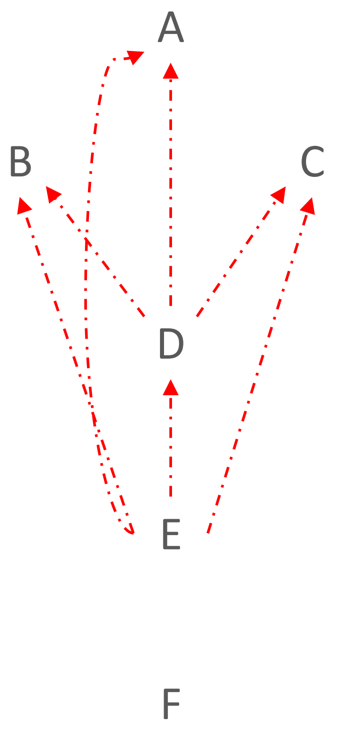


- $(X, Y) \in NAnc_P$: X is **not an ancestor** of Y in all MAGs represented by P.
- X has **no potentially directed** path to Y.
- $(D, A), (D, B), (D, C)$

Finding $NAnc_P$



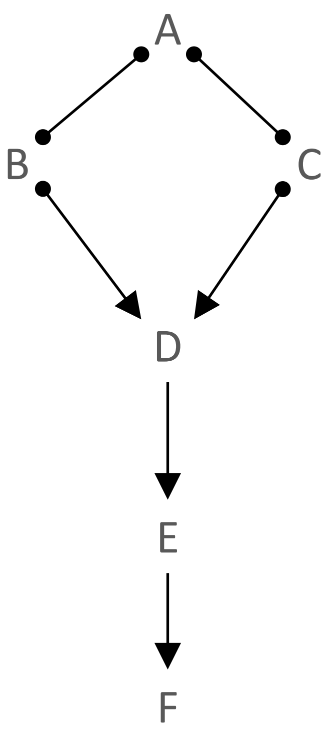
PAG P



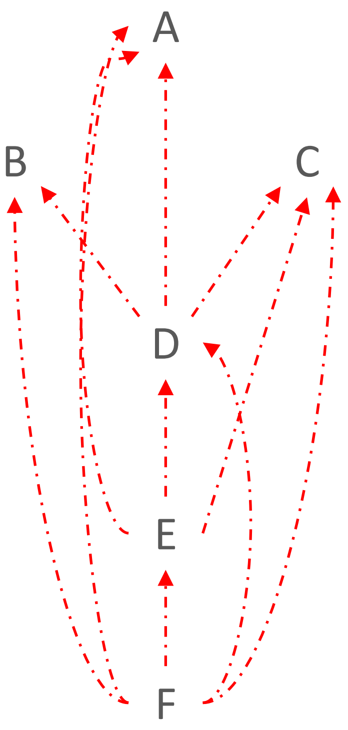
- $(X, Y) \in NAnc_P$: X is **not an ancestor** of Y in all MAGs represented by P
- X has **no potentially directed** path to Y.

- $(D, A), (D, B), (D, C)$
- $(E, A), (E, B), (E, C), (E, D)$

Finding $NAnc_P$



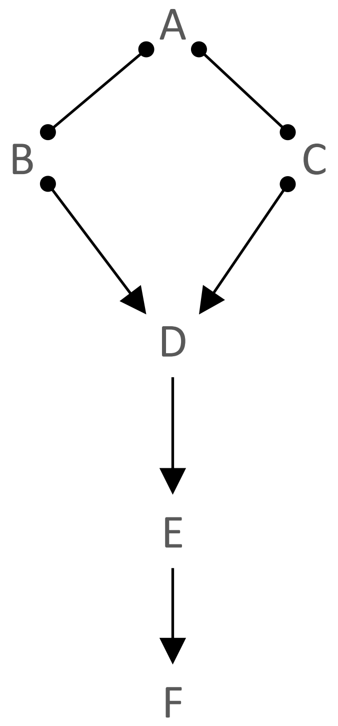
PAG P



- $(X, Y) \in NAnc_P$: X is **not an ancestor** of Y in all MAGs represented by P
- X has **no potentially directed** path to Y.

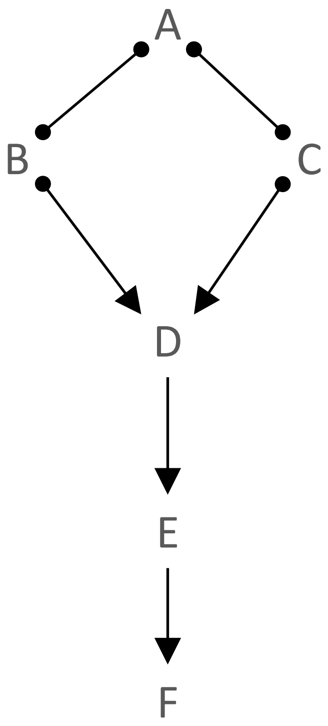
- $(D, A), (D, B), (D, C)$
- $(E, A), (E, B), (E, C), (E, D)$
- $(F, A), (F, B), (F, C), (F, D), (F, E)$

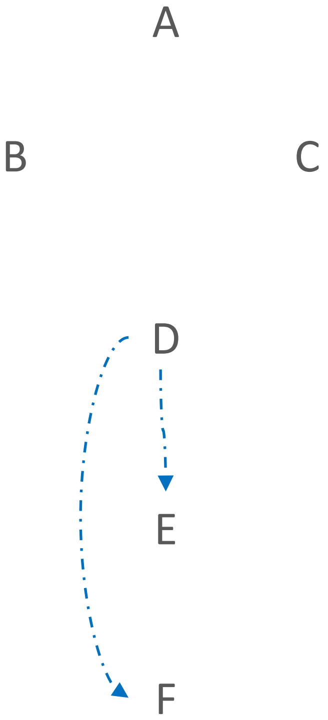
Finding Anc_P



- $(X, Y) \in NAnc_P$: X is **an ancestor** of Y in all MAGs represented by P.
- Take the **transitive closure** of directed edges in the PAG.

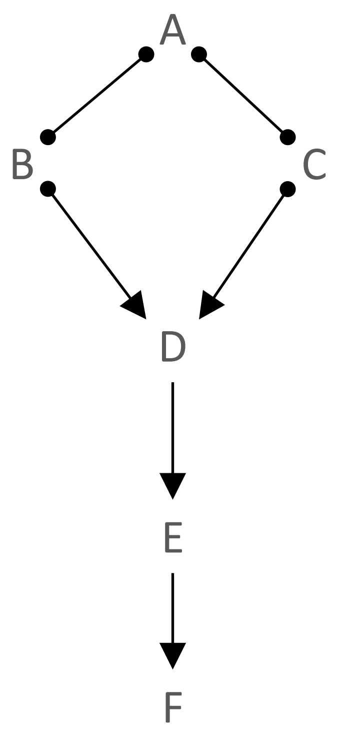
Finding Anc_P

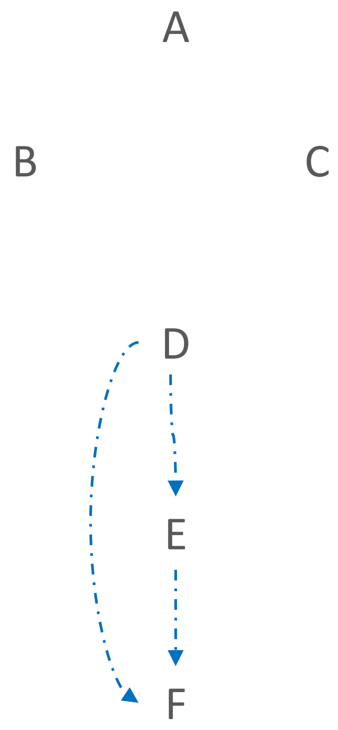




- $(X, Y) \in NAnc_P$: X is **an ancestor** of Y in all MAGs represented by P.
- Take the **transitive closure** of directed edges in the PAG.
- $(D, E), (D, F)$
- (E, F)

Finding Anc_P



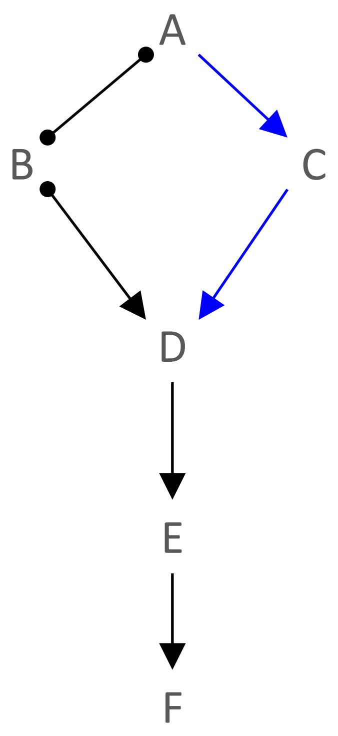


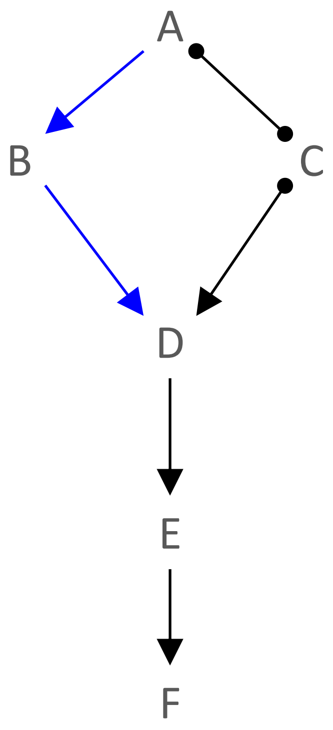
- $(X, Y) \in NAnc_P$: X is **an ancestor** of Y in all MAGs represented by P.

- Take the **transitive closure** of directed edges in the PAG.

- $(D, E), (D, F)$
- (E, F)

Finding Anc_P





- $(X, Y) \in NAnc_P$: X is **an ancestor** of Y in all MAGs represented by P.

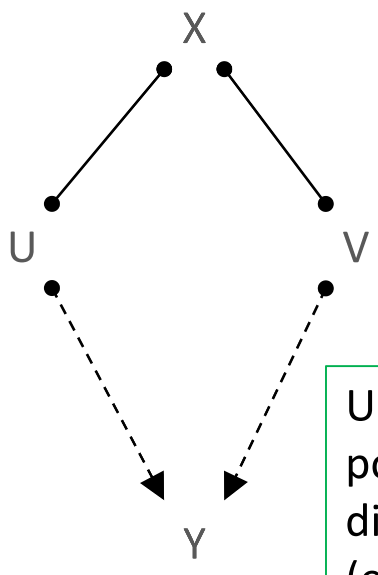
- Take the **transitive closure** of directed edges in the PAG.

- $(D, E), (D, F)$
- (E, F)

Some relations are not so obvious!

A causes D in all MAGs repr. by P.

Finding Anc_P

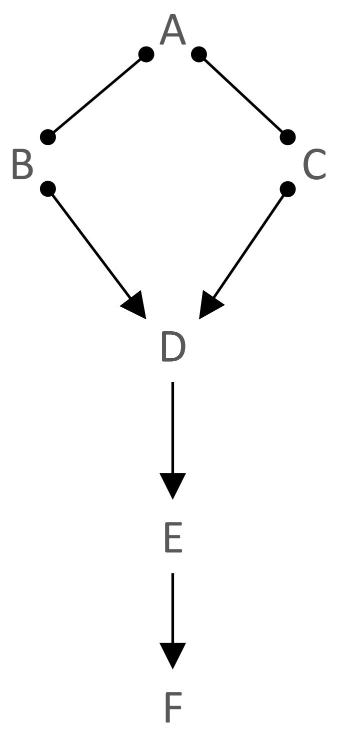


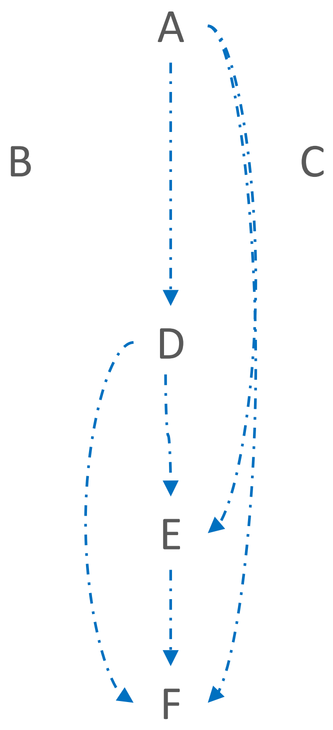
Uncovered potentially directed path (every triplet is an unshielded non-collider).

Theorem:
If (X,Y) is not in the transitive closure of PAG P, X is an ancestor of Y if and only if ∃ U,V, U≠V such that:

1. There are uncovered potentially directed paths from X to Y via U and V in P, and
2. <U,X,V> is an unshielded definite non-collider in P.

Finding Anc_P



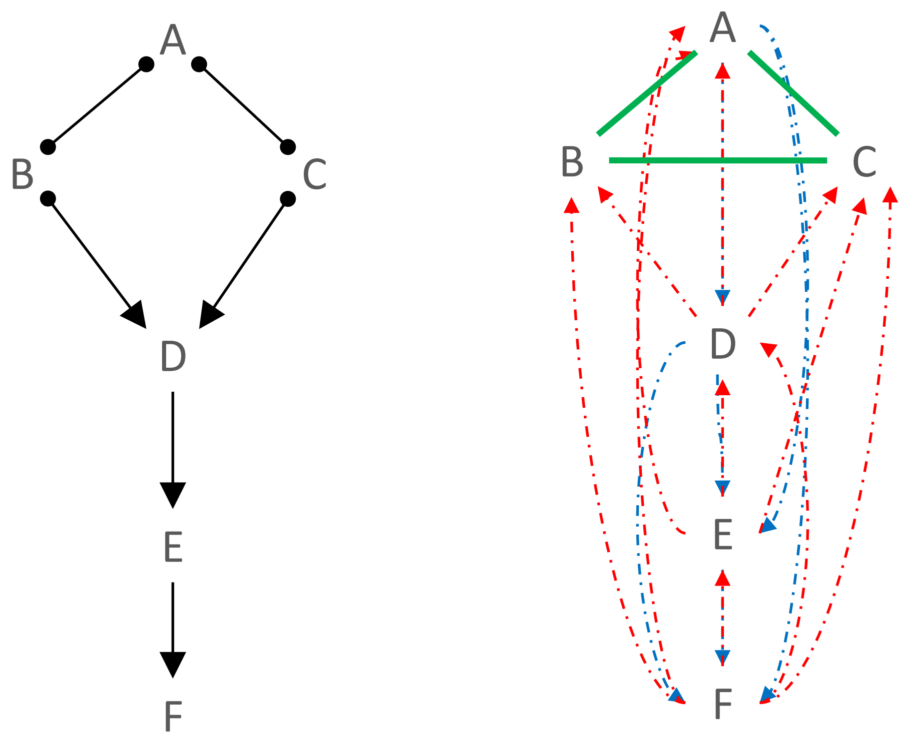


- $(X, Y) \in NAnc_P$: X is **an ancestor** of Y in all MAGs represented by P.

- Take the **transitive closure** of directed edges in the PAG.

- $(D, E), (D, F)$
- (E, F)
- $(A, D), (A, E), (A, F)$

Ambiguous relationships



- If $(X, Y) \notin Anc_P \cup NAnc_P$: Ambiguous relationship.
- If you marginalize out variables, (non) ancestral relationships can become ambiguous.

Marginal Consistency

Under perfect statistical knowledge:

- Ancestral relations in the marginal are ancestral in the original PAG (over $V \setminus$ L)
  - $Anc(P[_L) \subset Anc(P)_{V \setminus L}$ (d=0, e=0)
- Non-ancestral relations in the marginal non-ancestral in the original PAG (over $V \setminus$ L)
  - $NAnc(P[_L) \subset NAnc(P)_{V \setminus L}$ (c=0, f=0)

		marginal P[L Anc	N Anc
original P	Anc	$\lvert An_{P[_L} \rvert$	0
	N Anc	0	$\lvert NAn_{P[_L} \rvert$
	Amb	0	0
		$\lvert An_{P[_L} \rvert$	$\lvert NAn_{P[_L} \rvert$

In reality:

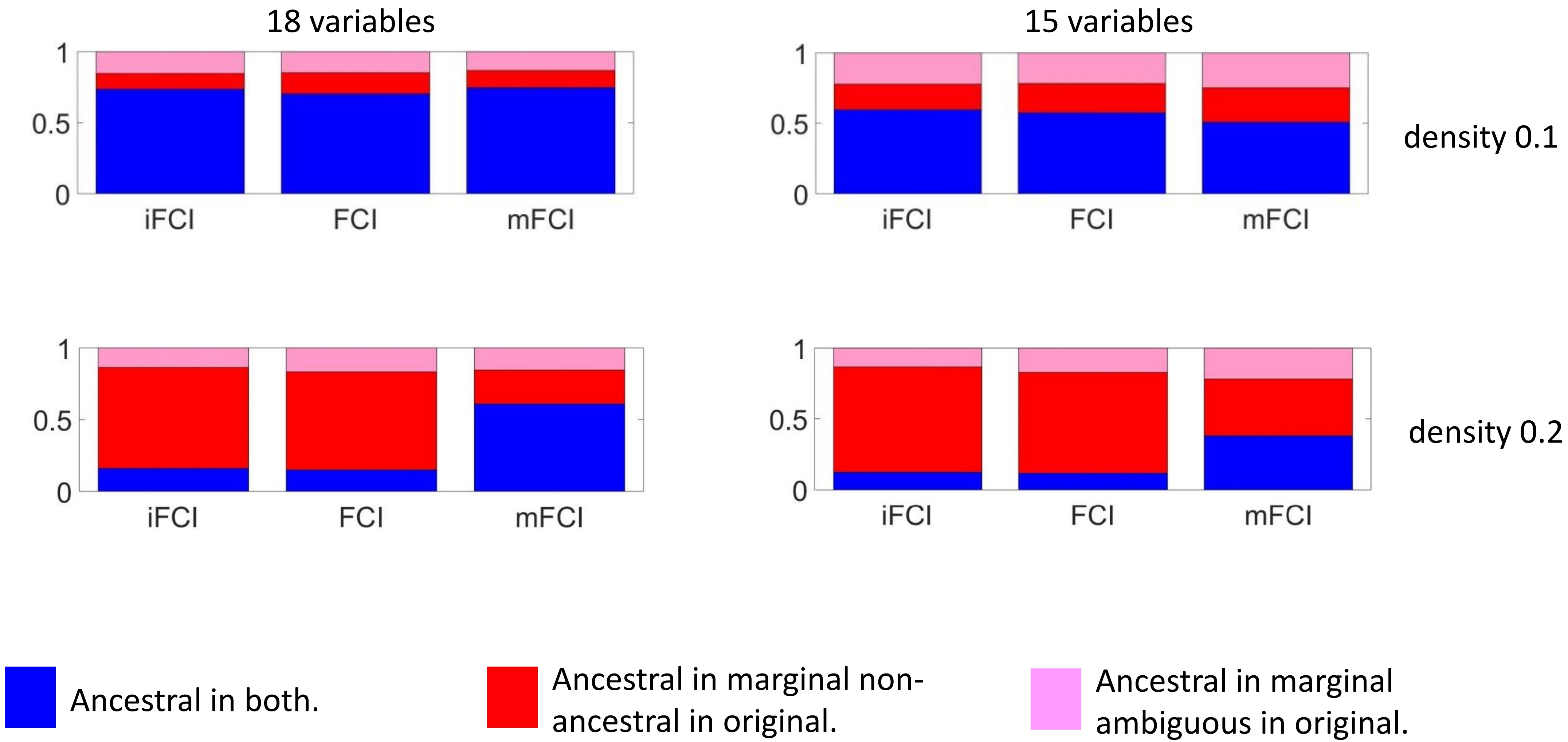
- Ancestral relations in the marginal can be:
  - non-ancestral in the original PAG (d).
  - ambiguous in the original PAG (e)
- Non-ancestral relations in the marginal can be:
  - Ancestral in the original PAG (c).
  - ambiguous in the original PAG (f)

		marginal P[L Anc	NAnc
original P	Anc	p	c
	NAnc	d	n
	Amb	e	f
		$\lvert An_{P[_L} \rvert$	$\lvert NAn_{P[_L} \rvert$

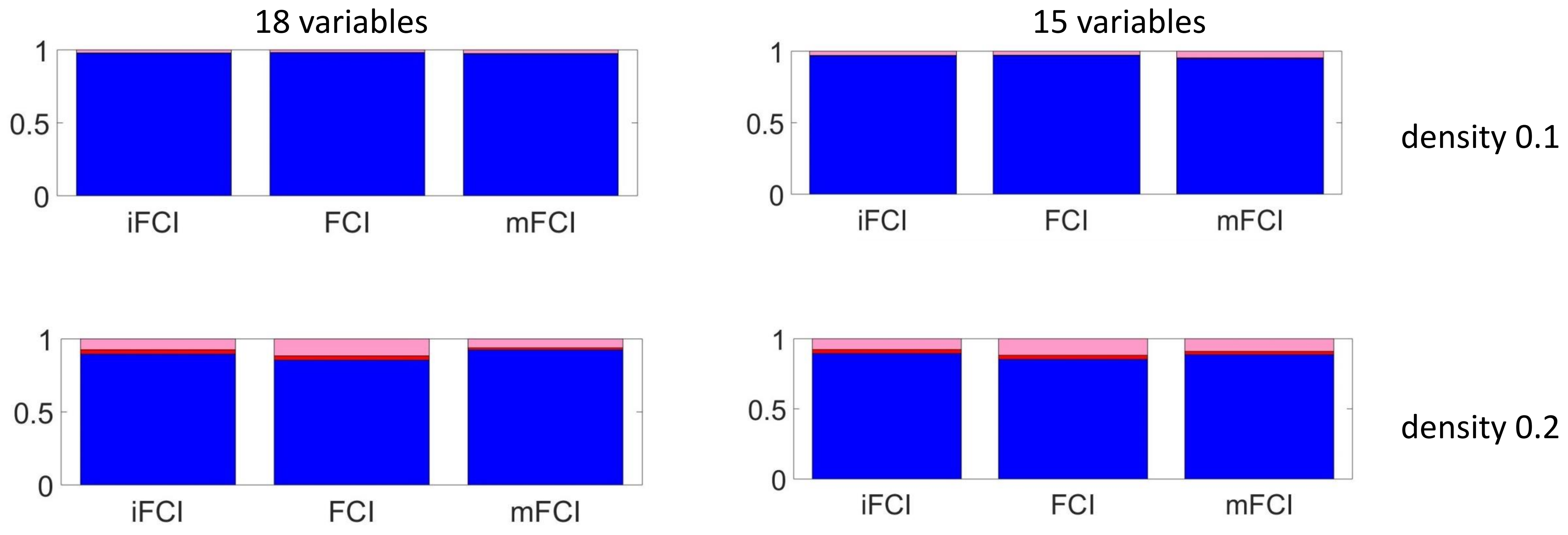
Experiments

- 50 DAGs of 20 variables.
- Graph density: 0.1 and 0.2.
- Continuous data sets with 1000 samples.
- 100 random marginals of 18 and 15 variables.
- Algorithms used (pcalg) [2]:
  - FCI
  - iFCI
  - mFCI

Results (Ancestral relationships)



Results (Non Ancestral Relationships)



Non-Ancestral in both.

Non-Ancestral in marginal, ancestral in original.

Non-Ancestral in marginal ambiguous in original.

Conclusions

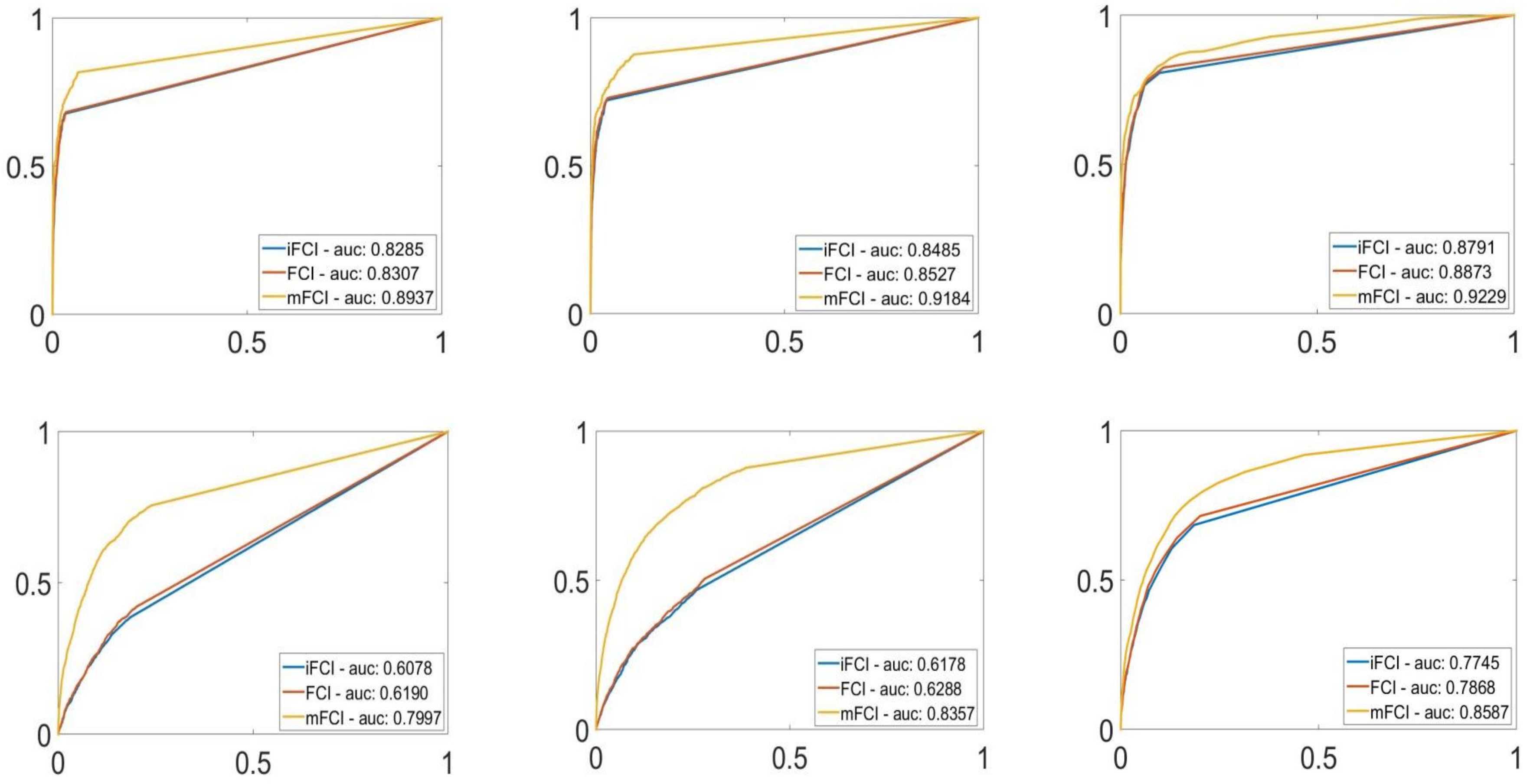
- Constraint-based methods are sensitive to marginalization.
- Consistency of causal predictions drops for denser networks/smaller marginals.
- Non-causal predictions are very consistent.
- Majority rule FCI outperforms other FCI variants.

Ranking based on marginal consistency

- Can you use marginal consistency to find more robust predictions?
- Are predictions that are frequent in the marginals more robust?
- Compare with bootstrapping.

Results

18 variables | 15 variables | bootstrapping



density 0.1

density 0.2

Conclusion/Future work

- Ranking by marginal consistency can help identify robust causal predictions.
- Bootstrapping is more successful in ranking causal predictions (by a small margin).

- Ranking by marginals can become much faster by caching tests of independence.
- Try it for much larger data sets (number of variables).
- Combine with bootstrapping.
- Try to identify a marginal that maximizes marginal consistency.
- Try to identify a graph that is consistent for more marginals.

References

1. Diego Colombo and Marloes H Maathuis. Order-independent constraint-based causal structural learning. JMLR 2014.

2. Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H Maathuis, and Peter Bühlmann. Causal inference using graphical models with the R package pcalg. JSS 2012.

3. Joseph Ramsey, Jiji Zhang, and Peter Spirtes. Adjacency-faithfulness and conservative causal inference UAI 2006.

4. Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction,and Search. MIT Press, 2000.