Constraint-based network reconstruction with consistent separating set

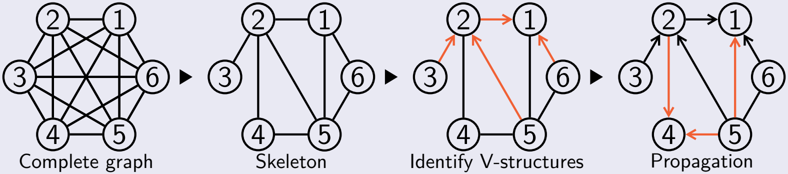
Honghao LI

Isambert Team, Institut Curie

September 27, 2019

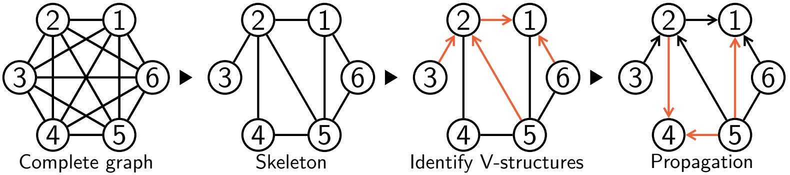
Motivation

Constraint-based method in general



1. (Conditional) independence test gives graph skeleton
2. Signature of causality (V-structure) allows for the orientation of edges.
3. Additional assumptions allows for the propagation of orientations.
4. In the oriented graph, each (un)directed edge represents a direct (non-)causal relation between variables.

Motivation



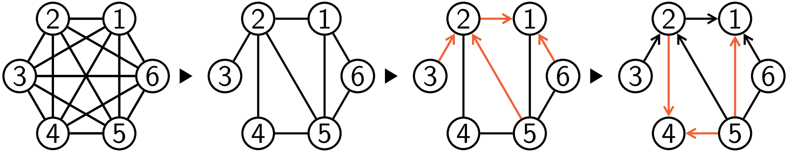
Interpretability: separating set

Conditional independence relations:
$(1 \perp\!\!\!\perp 4 \mid 2, 5)$, $(1 \perp\!\!\!\perp 3 \mid 2, 5)$, $(3 \perp\!\!\!\perp 4 \mid 2, 5)$, $(4 \perp\!\!\!\perp 6 \mid 5)$.

In the oriented graph, these results explain

1. the missing of direct relation between two variables;
2. which variable(s) contribute to the passing of information between the two variables.

Motivation



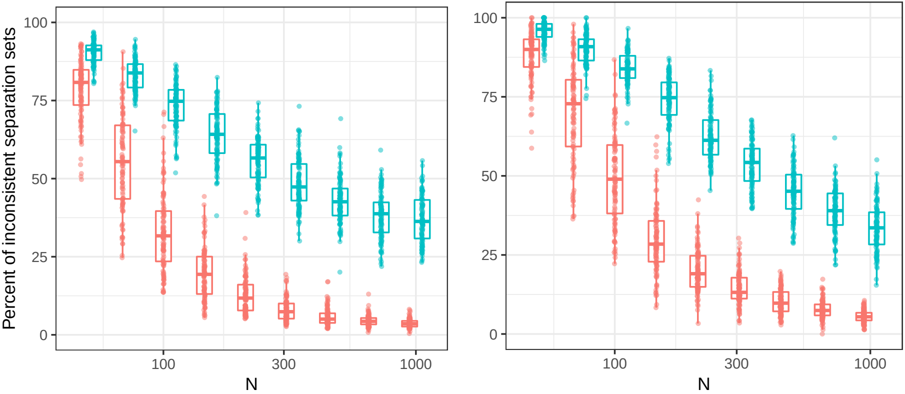
Complete graph Skeleton Identify V-structures Propagation

Separating set: inconsistency

type I: $(2 \perp\!\!\!\perp 6 \mid 3)$ There is no path between 2 and 6 that goes through 3, inconsistent with respect to the skeleton;
type II: $(3 \perp\!\!\!\perp 6 \mid 1)$ The vertex 1 is a descendant of vertex 6 and 3, inconsistent with respect to the oriented graph.

In practice, these results, even if correct in terms of dependence relation, are **not interpretable**

Motivation

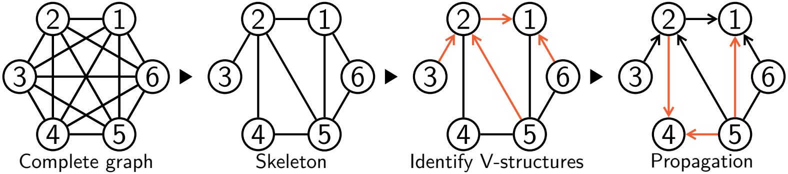


Inconsistency of separating set of the original PC-stable algorithm

Red: percent of inconsistent edges in skeleton;

Blue: percent of inconsistent edges in oriented graph.

Motivation



Goal

- Make sure all separating sets remain consistent with respect to the final graph;
- Retain the same level of performance (in terms of precision and recall) with respect to original algorithm;
- Resonable time complexity.

Method

Definition (Skeleton-consistent sets)

The set of consistent vertices with respect to (X, Y) and the skeleton of $\mathcal{G}$:

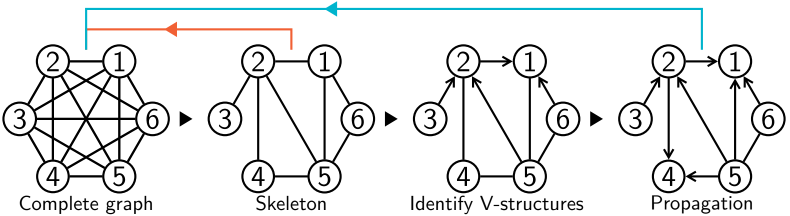
$$\text{Conskel}(X, Y \mid \mathcal{G}) = \{\, Z \in \boldsymbol{V} \setminus \{\, X, Y \} \mid \text{at least one path } \gamma_{XY}^{Z} \text{ exists} \,\}$$

Definition (Orientation-consistent sets)

The set of consistent vertices with respect to (X, Y) and $\mathcal{G}$:

$$\text{Consist}(X, Y \mid \mathcal{G}) =$$
$$\{\, Z \in \text{Conskel}(X, Y \mid \mathcal{G}) \mid Z \text{ is not a common descendant of } X \text{ and } Y. \,\}$$

Method

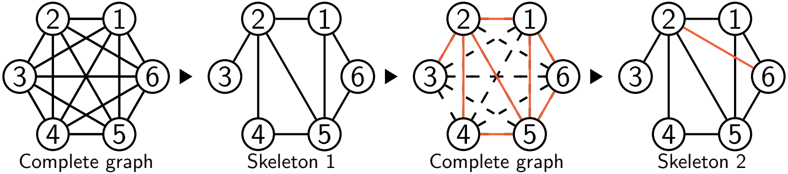


An iterative approach: two strategies

Using the previous definitions as constraint, we may

- either ensure that the skeleton is consistent, then orient the graph without breaking consistency;
- or ensure that the oriented graph is consistent.

Method

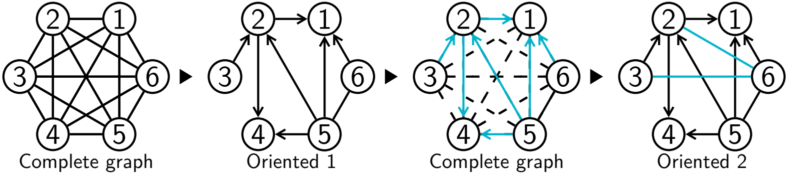


Skeleton-consistent approach

- In each iteration, the skeleton obtained is used as a constraint on finding separating set during the next iteration, until a consistent skeleton is found.
- Any inconsistency present in the current skeleton (e.g. $(2 \perp\!\!\!\perp 6 \mid 3)$ in Skeleton 1) will be corrected in the next interation (Skeleton 2).

The consistent skeleton obtained is then used to be oriented while keeping consistency.

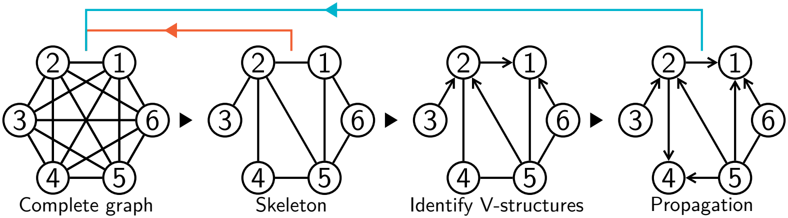
Method



Orientation-consistent approach

- In each iteration, the oriented graph is used as a constraint on finding separating set during the next iteration, until a consistent graph is found.
- Any inconsistency present in the current graph (e.g. $(2 \perp\!\!\!\perp 6 \mid 3)$ and $(3 \perp\!\!\!\perp 6 \mid 1)$ in Oriented 1) will be corrected in the next interation (Oriented 2).

Method



An iterative approach: two strategies

As a result, we make sure that in the final graph, all missing edges are interpretable with the corresponding separating set.

Honghao Li et al. (2019). "Constraint-based Causal Structure Learning with Consistent Separating Sets". In: *Advances in Neural Information Processing Systems 33*. Curran Associates, Inc.