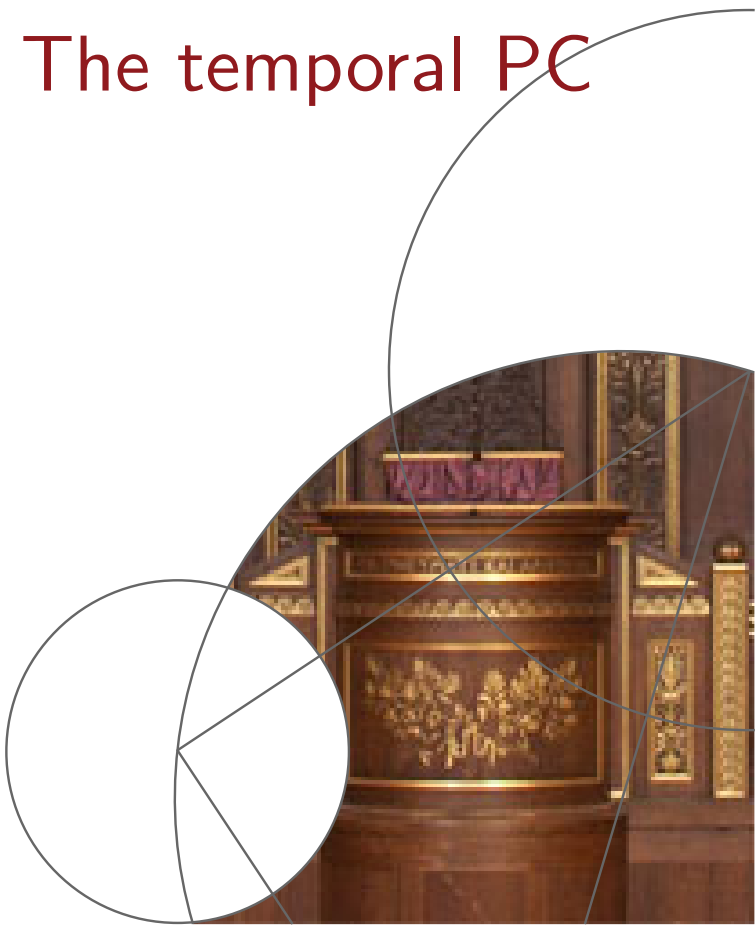
UNIVERSITY OF COPENHAGEN

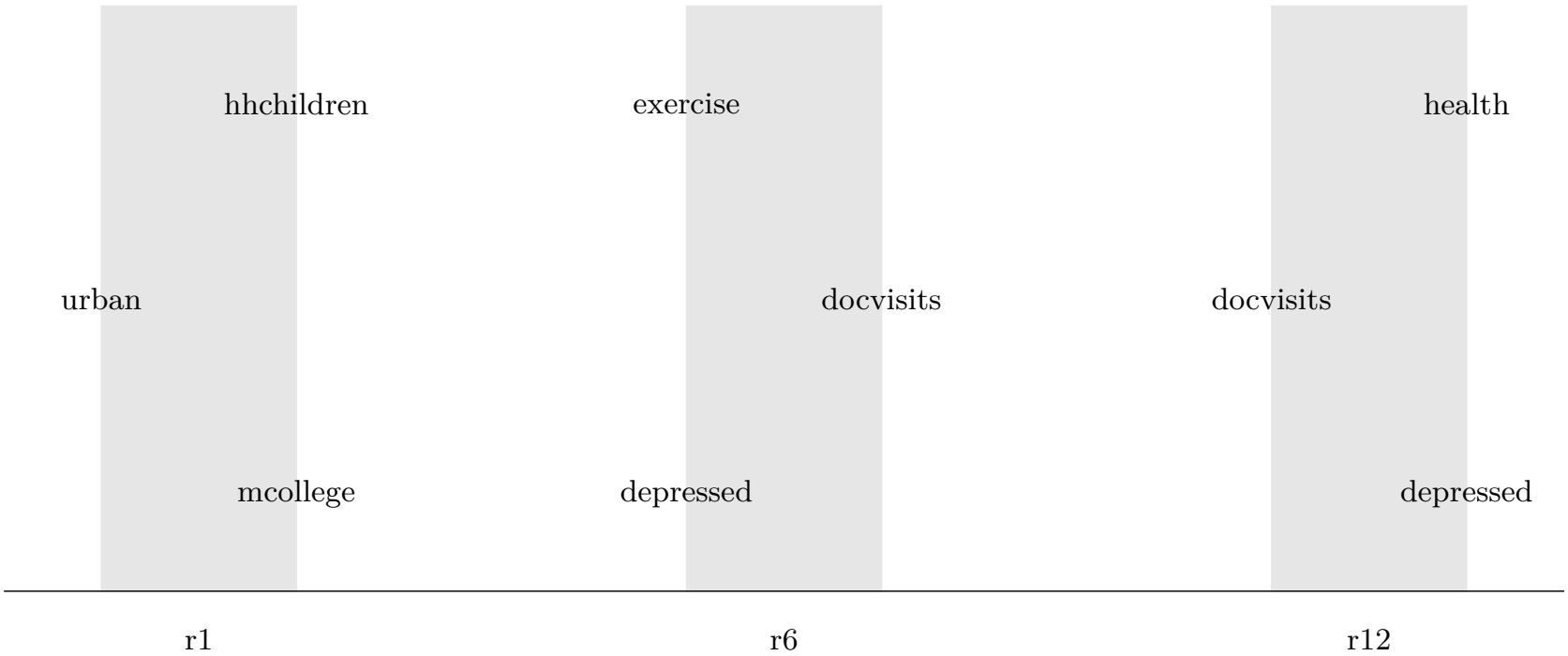
Introduction to causal discovery: The temporal PC algorithm

Anne Helby Petersen



Tiered/temporal background information

Setup: All variables can be assigned to a tier (e.g. period), and we know the ordering of these tiers.



The temporal PC algorithm (TPC)

Tiered background information may be used in two ways in the PC algorithm:

1. We can **skip certain statistical tests**: If S is a set of variables d-separating X and Y, S cannot be later than both X and Y. Hence we do not have to test $X \perp\!\!\!\perp Y \mid S$.
2. We can **orient cross-tier edges**: If X is temporally prior to Y, we can rule out the orientation $Y \to X$.

These ideas were originally proposed by Spirtes, Glymour & Scheines (2000), and a specific TPC algorithm implementation was proposed by Petersen, Osler & Ekstrøm (2021).

Two (similar) R-implementations available: `causalDisco` (Petersen 2022) and `tpc` (Witte & Foriatta 2023)

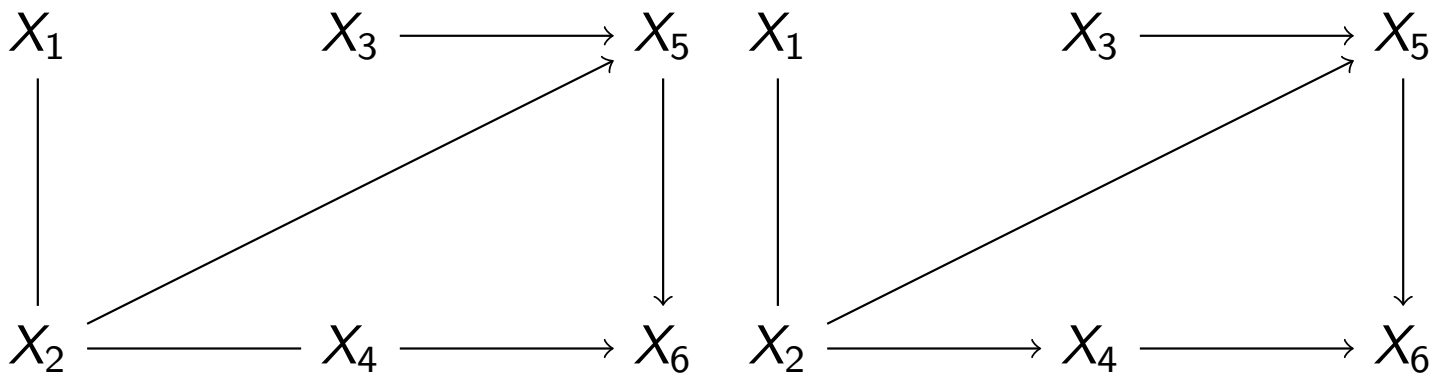
Temporal Peter-Clark (TPC) algorithm summary

Input: Information about conditional independencies **and tiered background information on the variables.**

1. Start with fully connected undirected graph
2. Repeat: For each pair of variables (A, B), look for separating sets S among variables adjacent to A or B **which are *not* later than both A and B in the temporal ordering** s.t. $A \perp\!\!\!\perp B \mid S$. If such an S exists: Remove edge between A and B.
3. **First, orient cross-tiers edges according to time.** Then, orient v-structures **whenever this does not induce orientations against the direction of time**. Finally, recursively apply Meek's three orientation rules.

Output: **Tiered** partially directed acyclic graph (TPDAG)

Tiered partially directed acyclic graphs



- Edges of a TPDAG may be interpreted as for a CPDAG: Oriented edges are direct causal relationships, unoriented edges represent ambivalence about the edge orientation among the possible DAGs it represents.
- A TPDAG is more informative than a CPDAG, and no longer represents a (full) DAG equivalence class.
- A TPDAG is a type of maximal partially directed acyclic graph (MPDAG) (Bang & Didelez 2023).
- For MPDAGs, results concerning e.g. adjustment exist (Perkovic, Kalisch & Maathuis 2017).

TPC properties

- If the tiered background knowledge is correct, TPC inherits its **correctness** directly from PC (Bang et al. 2024).
- The TPC algorithm is **complete**: No further causal information can be derived from the combination of conditional independence information and tiered background knowledge (Bang et al. 2024).
- With perfect conditional independence information, **only Meeks 1st rule** (avoiding introducing new v-structures directly) is invoked (Bang & Didelez 2023). Unclear what happens on finite data.
- Conjecture: TPC results in **less statistical error** – skipping unnecessary tests is a more efficient use of data.

References

Bang & Didelez (2023). Do we become wiser with time? On causal equivalence with tiered background knowledge. In *Proceedings of Uncertainty in Artificial Intelligence*.

Bang, Witte, Foraita & Didelez (2024). Efficient use of tiered background knowledge for causal discovery with cohort data. Soon to appear on *arXiv*.

Perkovic, Kalisch & Maathuis (2017). Interpreting and using CPDAGs with background knowledge. In *Proceedings of Uncertainty in Artificial Intelligence*.

Petersen (2022). causalDisco: Tools for Causal Discovery on Observational Data. R package available on CRAN.

Spirtes, Glymour & Scheines (2000). *Causation, Prediction, and Search*.

Witte & Foraita (2023). tpc: Tiered PC Algorithm. R package available on CRAN.