Establishing Markov equivalence in cyclic directed graphs

Tom Claassen & Joris M. Mooij
Radboud University Nijmegen, University of Amsterdam

39th Conference on UAI
Pittsburgh, 2 August, 2023

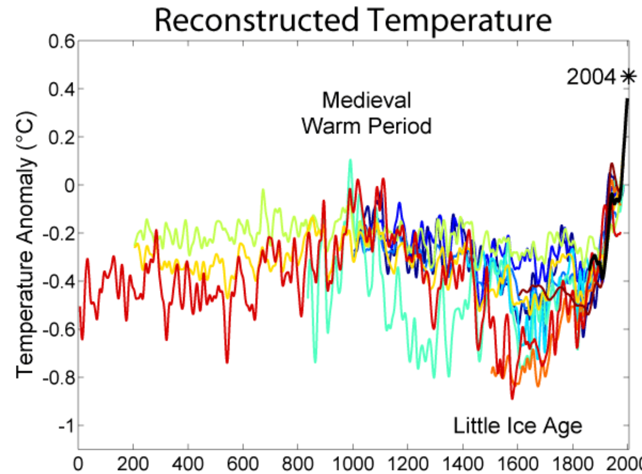
Radboud University Nijmegen

Outline

1. **Causal discovery and cyclic models**
2. The Cyclic Equivalence Theorem (CET)
3. An ancestral perspective on the CET
4. Establishing Markov equivalence and beyond

Many important research questions are rooted in causality

benefits of exercise and healthy nutrition

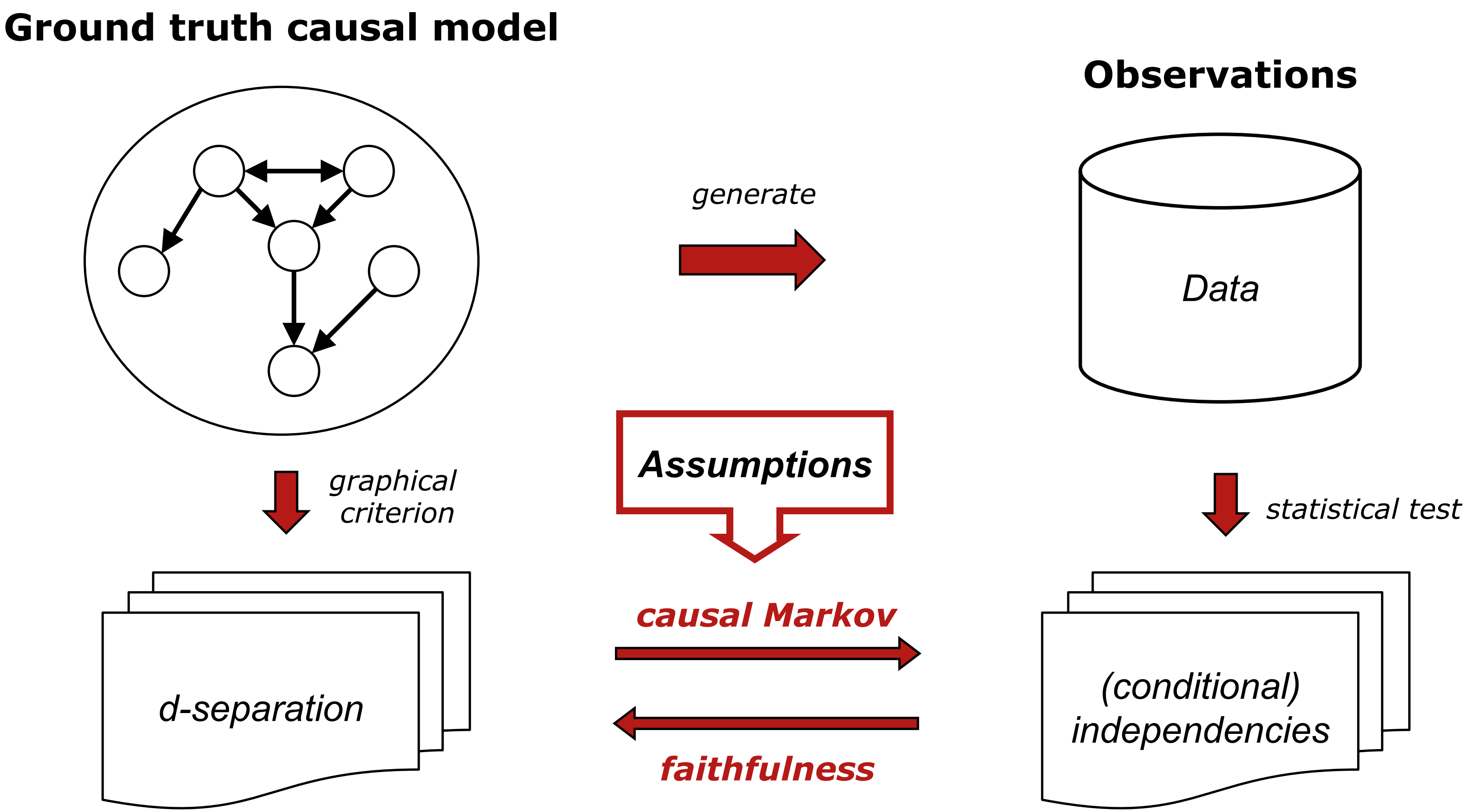


human activity and climate change

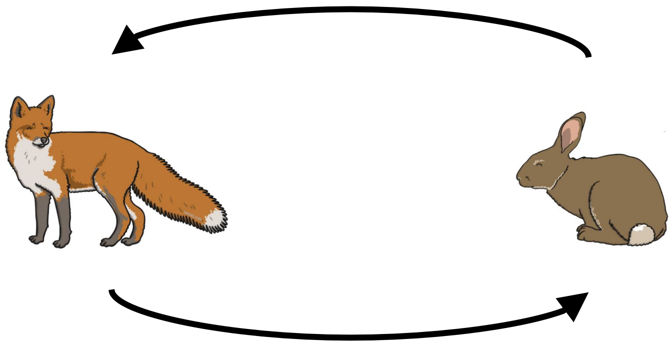
racial and gender bias in AI

Covid vaccine efficacy

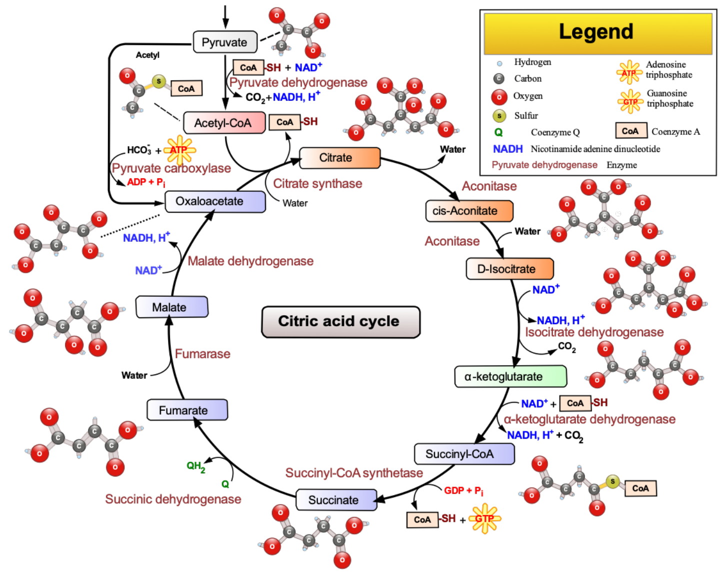
Basic constraint-based causal discovery



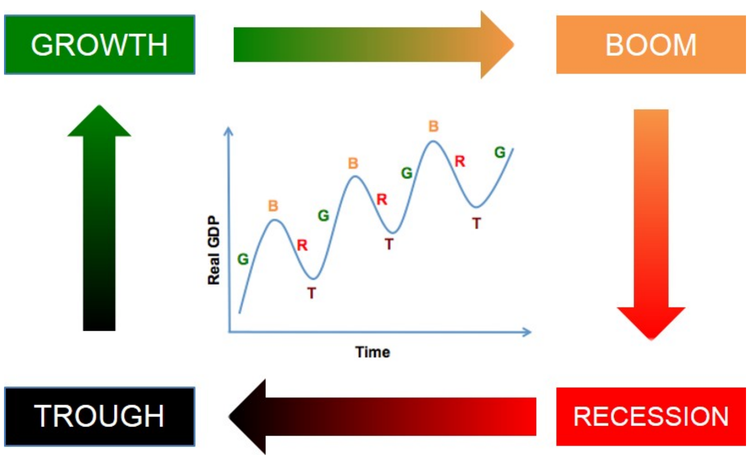
Meanwhile, in the real world …



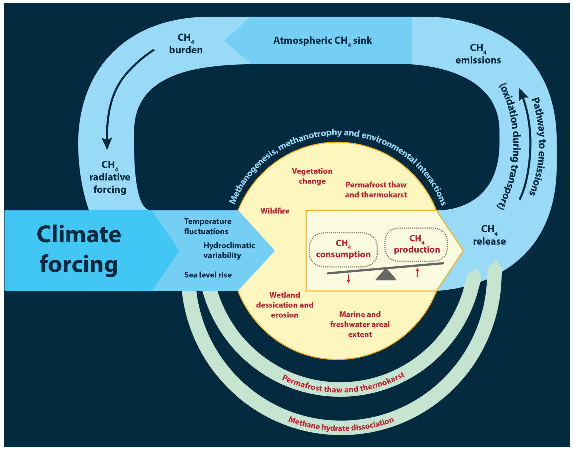
predator-prey interaction



cell microbiology processes

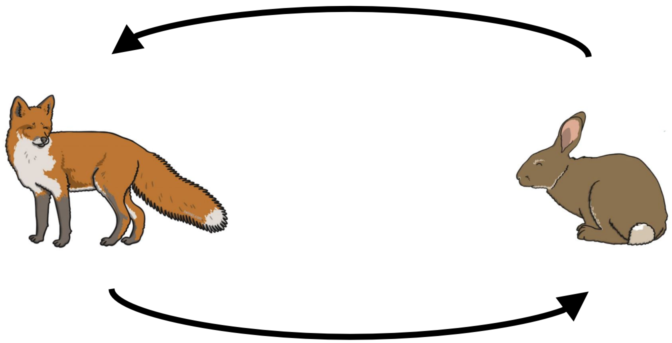


economic growth cycles



methane climate feedback

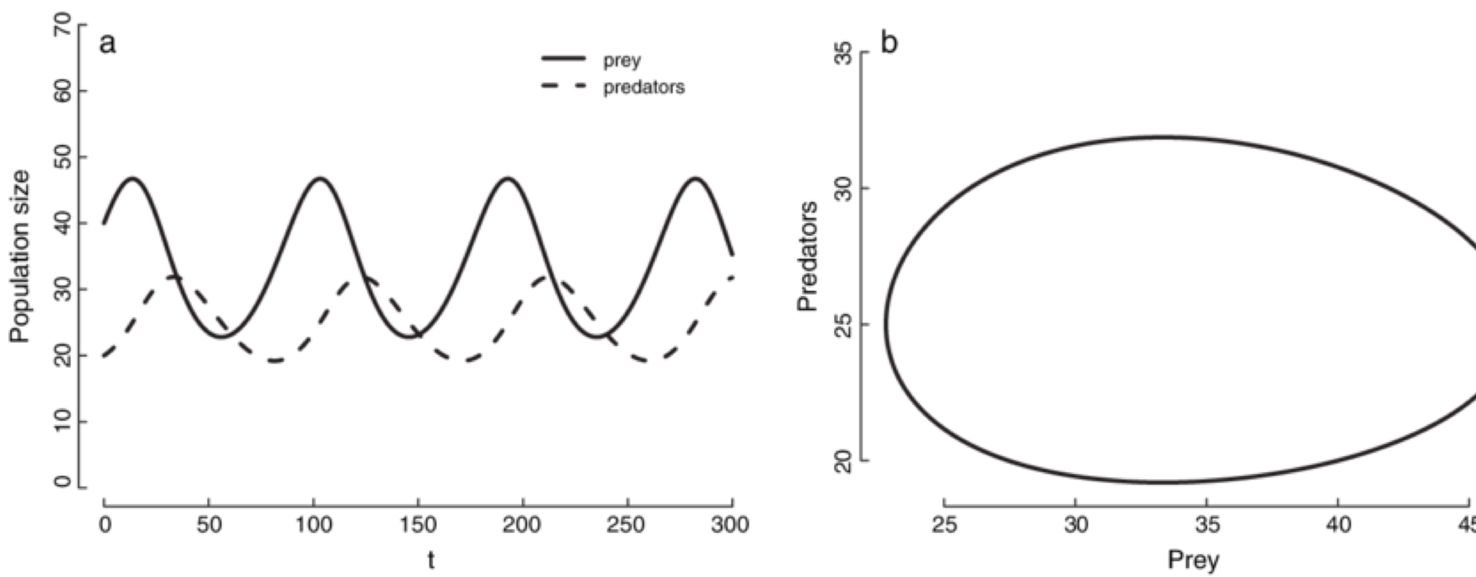
The rabbit and the fox



$$\text{rabbit:} \quad \frac{dx}{dt} = \alpha x - \beta x y$$

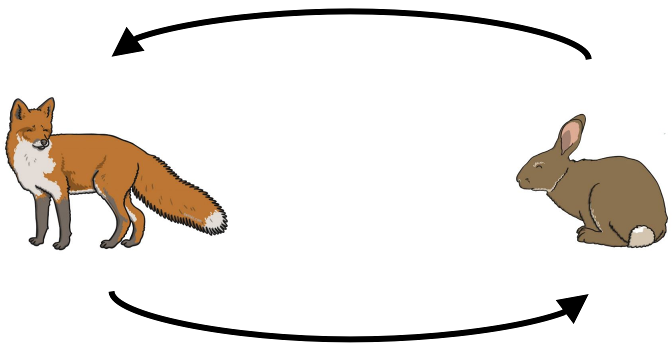
$$\text{fox:} \quad \frac{dy}{dt} = \delta x y - \gamma y$$

Lotka-Volterra equations



periodic solution over time

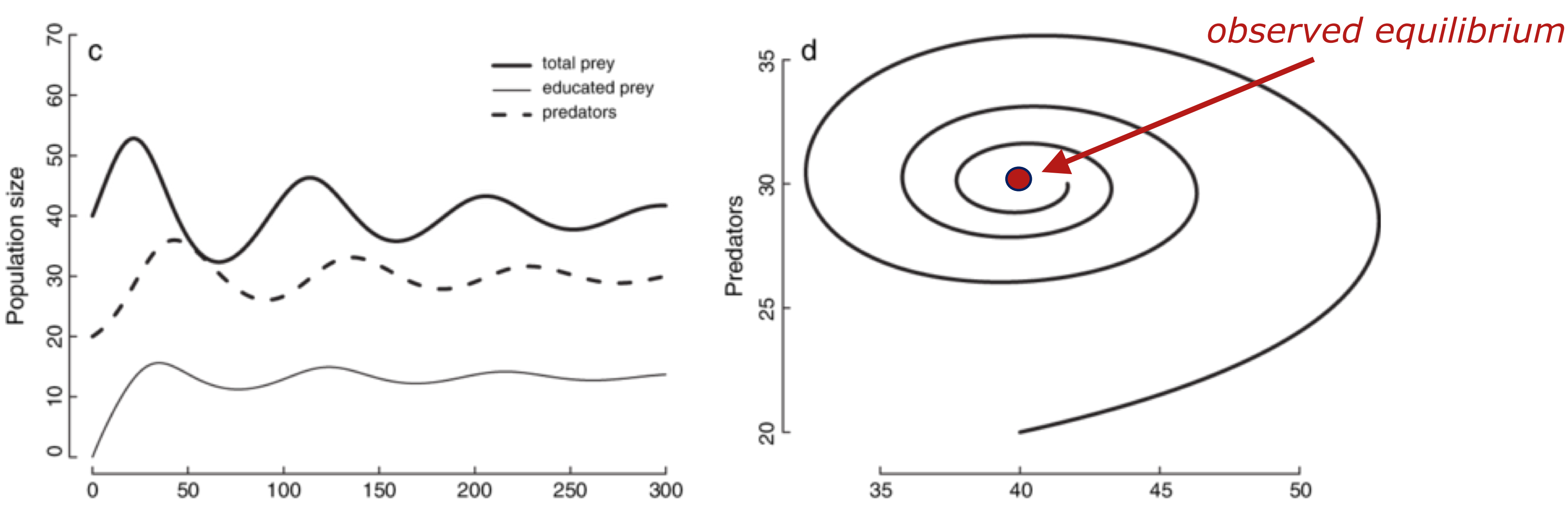
Static equilibrium solutions



rabbit : $$\frac{dx}{dt} = \alpha x - \beta xy$$

fox: $$\frac{dy}{dt} = \delta xy - \gamma y$$

Lotka-Volterra equations



with `damping': equilibration towards unique, static solution

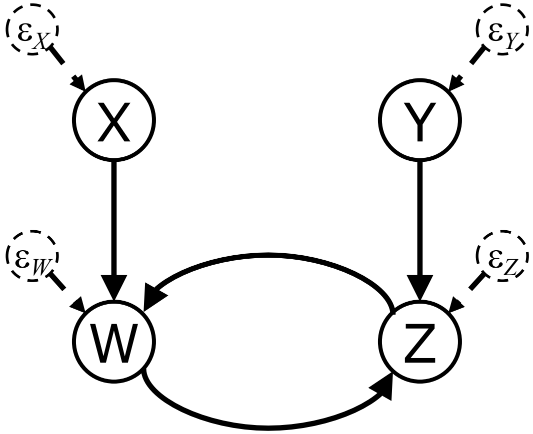
Outline

1. Causal discovery and cyclic models
2. **The Cyclic Equivalence Theorem (CET)**
3. An ancestral perspective on the CET
4. Establishing Markov equivalence and beyond

Linear/discrete cyclic causal models

$$
\begin{aligned}
X &= \epsilon_X \quad (\sim \mathcal{N}(0,1)) \\
Y &= \epsilon_Y \quad (\sim \mathcal{N}(0,1)) \\
Z &= \alpha_{ZW} W + \alpha_{ZY} Y + \epsilon_Z \quad (\sim \mathcal{N}(0,1)) \\
W &= \alpha_{WZ} Z + \alpha_{WX} X + \epsilon_W \quad (\sim \mathcal{N}(0,1))
\end{aligned}
$$

linear cyclic Gaussian SCM



matching causal graph G

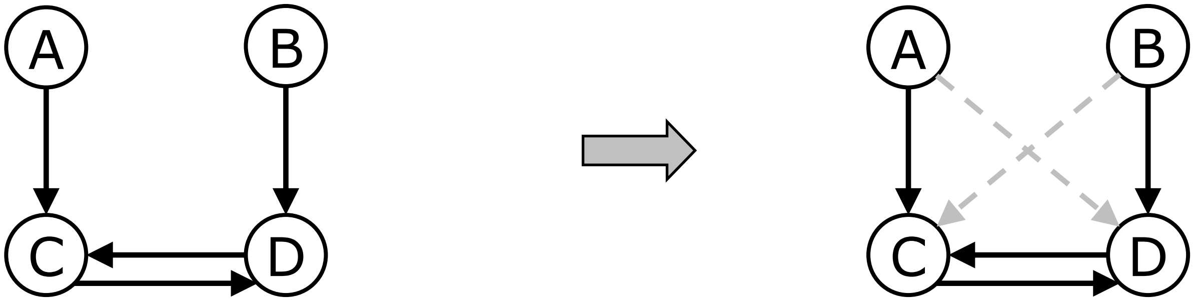
Key implication

- directed global Markov property still holds *[Spirtes,1994; Bongers et al.,2021]*
- standard *d*-separation still applies …
- … but leads to a few extra quirks relative to acyclic models *[Richardson,1996*]

Goal

- discovery aims for Markov equivalence class (cyclic PAG)

Key cyclic terminology *[Richardson,1996/97]*

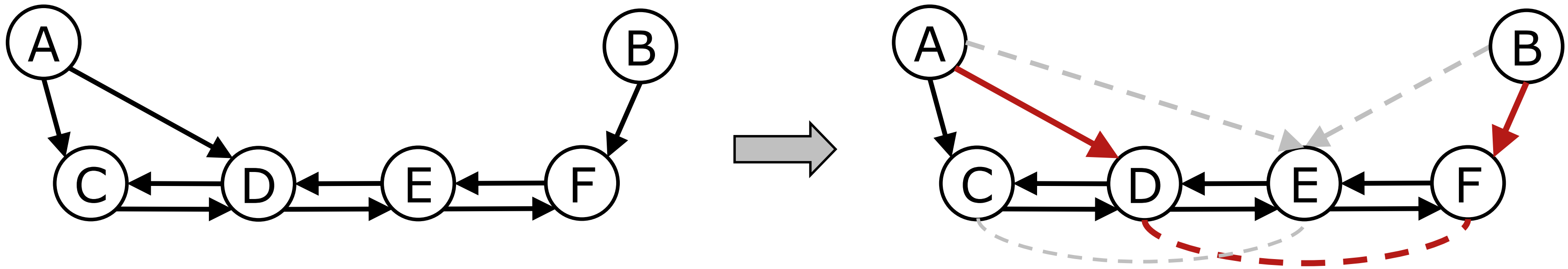


cyclic graph with two virtual edges

- *A* and *D* are *virtually adjacent* if they are not adjacent, but have a common child *C* in a cycle with *A* and/or *D*,
- an *itinerary* is a path over real and/or virtual edges (e.g. <*A*,*C*,*B*>)
- an itinerary is *uncovered* if no two nodes on the path are (virtually) adjacent, other than the neighbours along the path

- an itinerary <*A*,*C*,*D*> is a *conductor* if *C* is ancestor of *A* or *D*, otherwise it is a *nonconductor*,
- a nonconductor <*A*,*C*,*B*> is *perfect* if *C* is a descendant of a common child of *A* and *B*, otherwise it is *imperfect*

Key cyclic terminology *[Richardson,1996/97]*



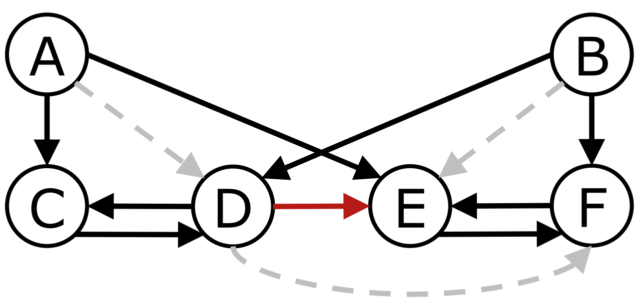
cyclic graph where <A,D,F> and <D,F,B> are
m.e. conductors w.r.t. uncovered itinerary <A,D,F,B>

- two triples <*A*,*B*,*C*> and <*X*,*Y*,*Z*> are mutually exclusive (m.e.) *conductors w.r.t. an uncovered itinerary* <*A*,*B*,*C*,..,*X*,*Y*,*Z*> if each consecutive triple along the itinerary is a conductor, all nodes are ancestor of each other but not of A or Z, and no two nodes are (virtually) adjacent, except along the itinerary itself

Cyclic Equivalence Theorem *[Richardson,1997]*

Two directed graphs G_1 and G_2 are *d*-separation equivalent iff they have:

i. the same (virtual) adjacencies,
ii. the same unshielded conductors,
iii. the same perfect nonconductors, (= '*v*-structures')
iv. the same m.e. conductors w.r.t. some uncovered itinerary,
v. if <*A*,*X*,*B*> and <*A*,*Y*,*B*> are unshielded imperfect nonconductors in G_1 and G_2, then *X* is ancestor of *Y* in G_1 iff *X* is ancestor of *Y* in G_2,
vi. if <*A*,*B*,*C*> and <*X*,*Y*,*Z*> are m.e. conductors and <*A*,*M*,*Z*> is an unshielded imperfect nonconductor in G_1 and G_2, then *M* is a descendant of *B* in G_1 iff *M* is a descendant of *B* in G_2.



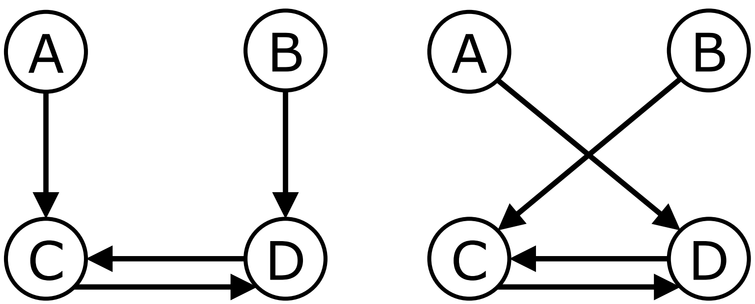
Example CET rule v : invariant edge D->E between two cycles C-D and E-F.

Cyclic partial ancestral graphs (CPAGs)

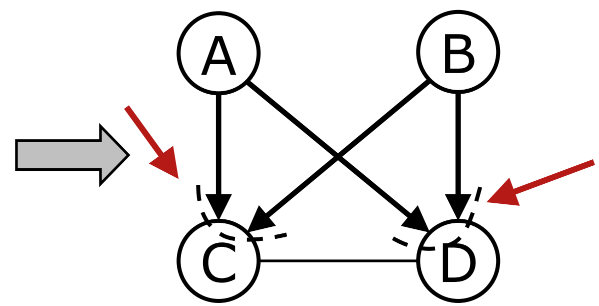
- compact graphical representation to uniquely identify Markov equivalence class $\{G\}$ for cyclic directed graph G, similar to standard (acyclic) PAGs

CPAG

- edge between each pair of (virtually) adjacent nodes,
- arrowhead/tail marks to indicate invariant (non)ancestors, circle marks for noncommitted edge marks,
- dashed-underlined $A \rightarrow \underline{B} \leftarrow C$ iff <*A*,*B*,*C*> is an *imperfect nonconductor*.



Markov equivalent cyclic graphs



CPAG with dashed underlined triples

Outline

1. Causal discovery and cyclic models
2. The Cyclic Equivalence Theorem (CET)
3. **An ancestral perspective on the CET**
4. Establishing Markov equivalence and beyond

Reflecting on the CET - CPAG

- the famous *Cyclic Causal Discovery* (CCD) algorithm *[Richardson,1996]* was an efficient CET-based implementation to reconstruct a CPAG from data,
- an analogous version based on *d*-separation could be used to establish Markov equivalence between cyclic graphs,
- yet, despite their central role in the CET, the dreaded 'm.e. conductors on an uncovered itinerary' never need to be recorded explicitly in the CPAG …

Motivating question

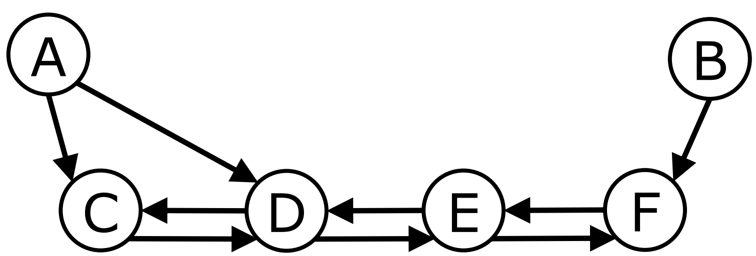
- does this mean we can simplify the CET? (spoiler: yes!)

Key idea (based on the familiar DAG-MAG-PAG trilogy from acyclic graphs):

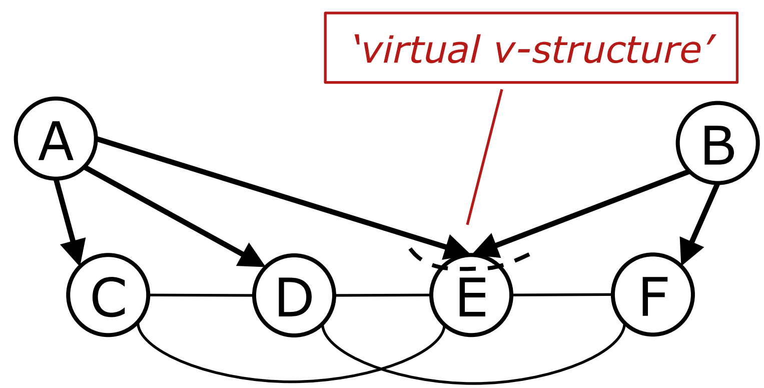
- introduce the CMAG as intermediate ancestral representation,
- rephrase CET in terms of invariant elements in Markov equivalent CMAGs

The **cyclic maximal ancestral graph (CMAG)** *M* for directed graph *G* has:

- an edge between every pair of (virtually) adjacent nodes in *G*,
- a tail mark X —∗ Y iff X is an ancestor of Y in *G*,
- arrowhead X ←∗ Y iff X is *not* an ancestor of Y,
- dashed-underlined $A \rightarrow \underline{E} \leftarrow B$ for *v*-structures in *M* with a virtual edge in *G.*



directed cyclic graph



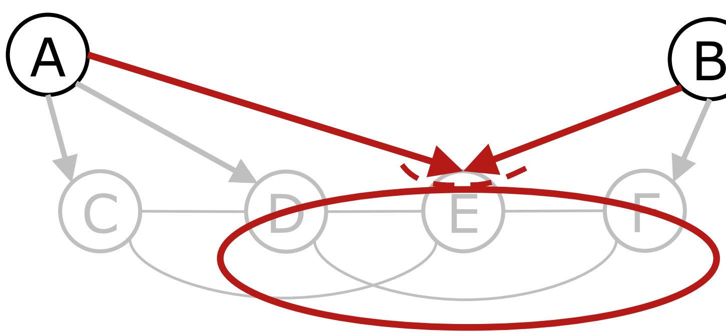
cyclic MAG

What about the m.e. conductors?

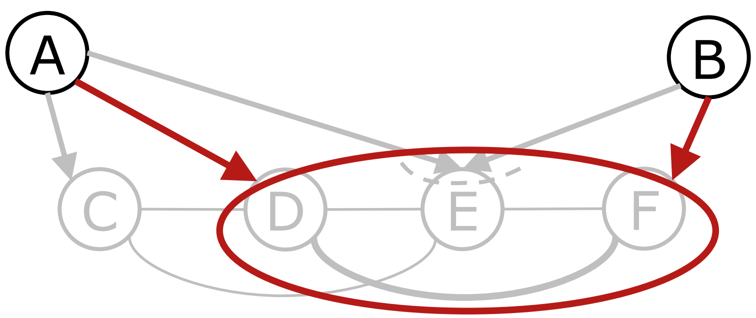
- In a CMAG *M*, quadruple <X,Z,Z′,Y> is a *u*-structure, iff there is an uncovered path X → Z – .. – Z′ ← Y in *M*.
- (Lemma 1) every *u*-structure in *M* matches a pair of m.e. conductors w.r.t. an uncovered itinerary in *G* and v.v.,
- comparing with virtual *v*-structures, both can be seen as arcs into a cycle

Merge into one!

A triple <X,Z,Y> is a virtual collider triple iff it is a virtual *v*-structure, or it is part of a *u*-structure <X,Z,Z′,Y> or <X,Z′,Z,Y>.



virtual v-structure



CMAG with u-structure <A,D,F,B>

Ancestral Cyclic Equivalence Theorem

Two CMAGs M_1 and M_2 are *d*-separation equivalent iff they have:

i. the same skeleton,
ii. the same *v*-structures,
iii. the same virtual collider triples,
iv. if $<A,X,B>$ and $<A,Y,B>$ are virtual collider triples in M_1 and M_2, then X is ancestor of Y in M_1 iff X is ancestor of Y in M_2.

$\Longrightarrow$ Basis for efficient procedure to establish Markov equivalence

Outline

Establishing Markov equivalence

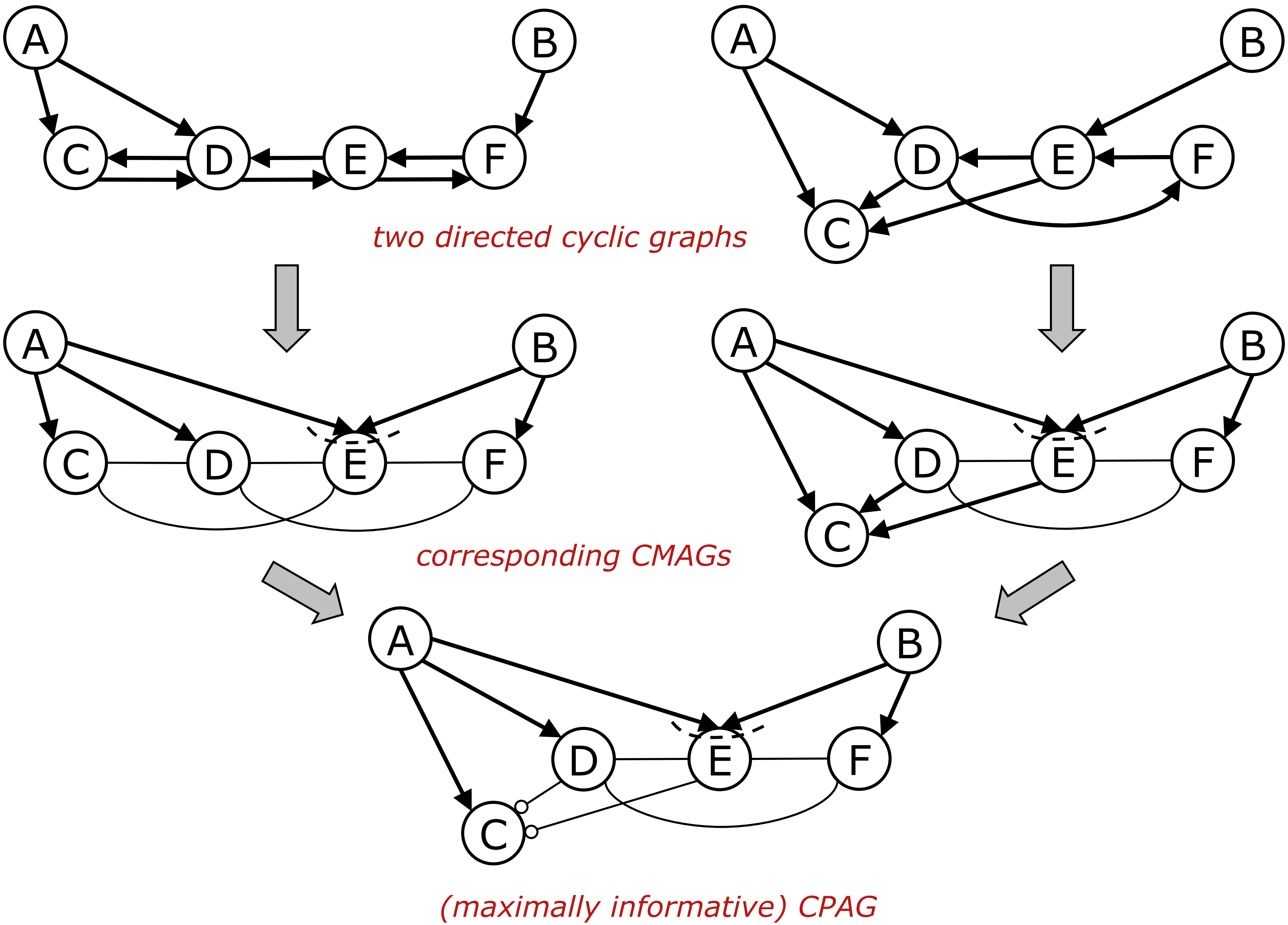
Ancestral CET suggests straightforward Cyclic-Graph-to-CPAG procedure:

- convert directed graph G into CMAG M
- copy skeleton and v-structures in M to CPAG P
- use CMAG M to identify and orient virtual collider triples in CPAG P
- orient remaining edges between cycles via CET-rule (iv)
- compare resulting CPAGs

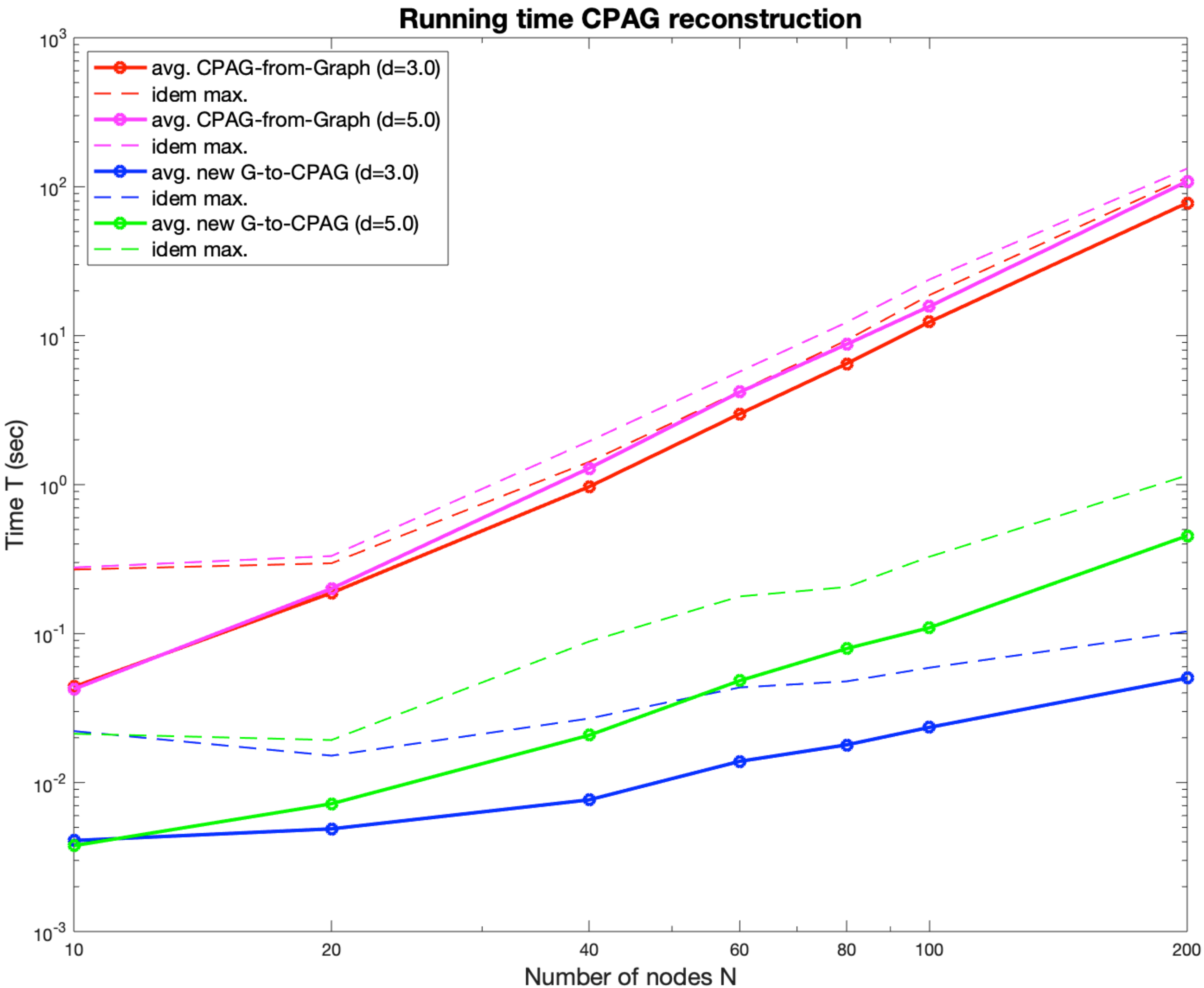
No need for d-separation tests!

- computational complexity scales as $O(N^2 * d^3)$
- worst case $O(N^5)$ compared to previous $O(N^7)$
- avg. scaling much better than worst case (for both versions)

Example: Cyclic-Graph-to-CPAG procedure



Experimental results – computational complexity



Conclusions & future work

New ancestral perspective on the CET proved very useful:

- significantly simplified CET characterization
- helpful intermediate CMAG representation
- fast, graphical procedure to establish Markov equivalence

Next steps

- recover *maximally informative* CPAG,
- merge *d*-separation (linear/discrete systems) and σ-separation (nonlinear systems) approaches via the CMAG,

Most promising

- formulation via 'virtual collider triples' suggests a natural extension of the CET to unobserved confounders, similar to the acyclic case, in the form of '*virtual triples with order*' *[Ali et al,2009; Claassen&Bucur,2022]*

Thank you!

(poster 504 @11:00)