Root Cause Identification for Collective Anomalies in Time Series given a Summary Causal Graph

Simon Ferreira, Charles Assaad

ENS de Lyon

19^{th} April 2023

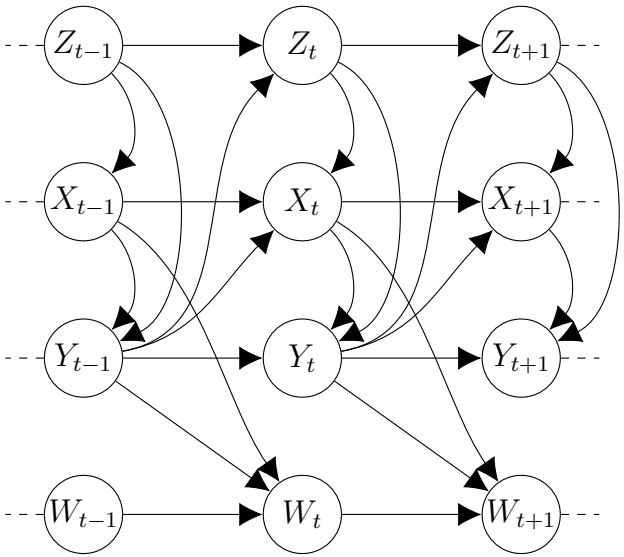
Outline

SCM

$$
\begin{aligned}
Z_t &= a_{zz} Z_{t-1} + a_{yz} Y_{t-1} + \xi_t^z \\
X_t &= a_{xx} X_{t-1} + a_{zx} Z_t + a_{yx} Y_{t-1} + \xi_t^x \\
Y_t &= a_{yy} Y_{t-1} + a_{zy} Z_t + a_{xy} X_t + \xi_t^y \\
W_t &= a_{ww} W_{t-1} + a_{xw} X_{t-1} + a_{yw} Y_{t-1} + \xi_t^w
\end{aligned}
\qquad \left(\begin{array}{c} \text{Consistent} \\ \text{throughout time} \end{array} \right)
$$

Causal Graphs

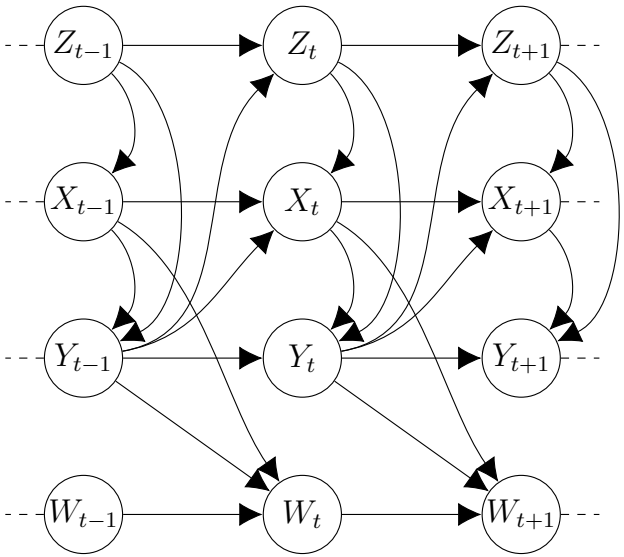
Representations of time series causal graphs[1] :



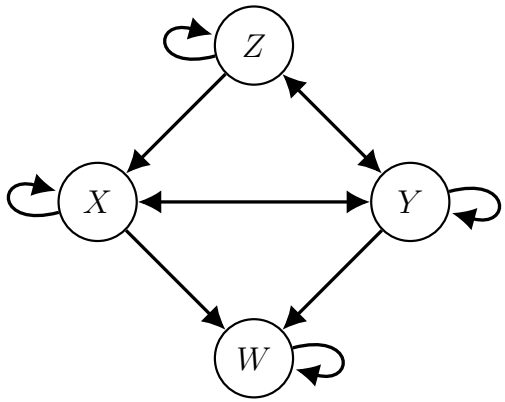
(a) (Acyclic) Full Time Causal Graph

Causal Graphs

Representations of time series causal graphs[1] :



(a) (Acyclic) Full Time Causal Graph



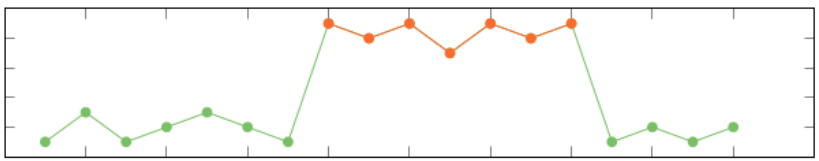
(b) (Cyclic) Summary Causal Graph with loops

Anomalies

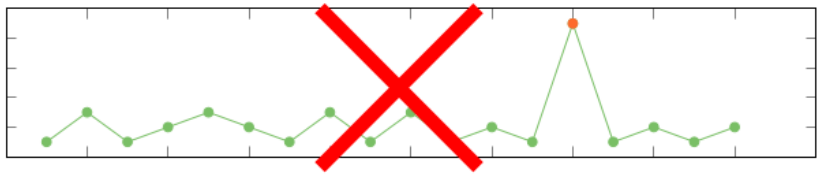
- A root cause is a vertex that underwent either :
  - an external *parametric* intervention,
  - or an external *structural* intervention.

Anomalies

- A root cause is a vertex that underwent either :
  - an external *parametric* intervention,
  - or an external *structural* intervention.
- Anomalies are collective[3] :



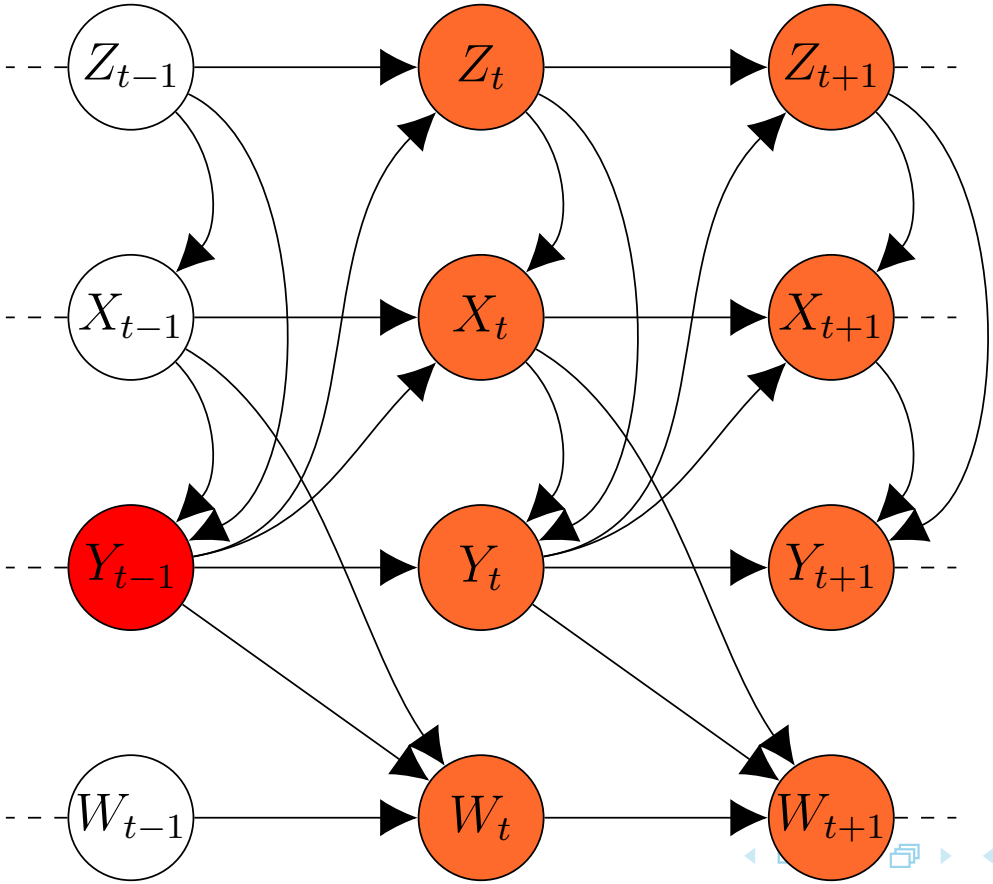
Time



Time

Anomalies

- Anomalies are propagated from external interventions through the SCM :



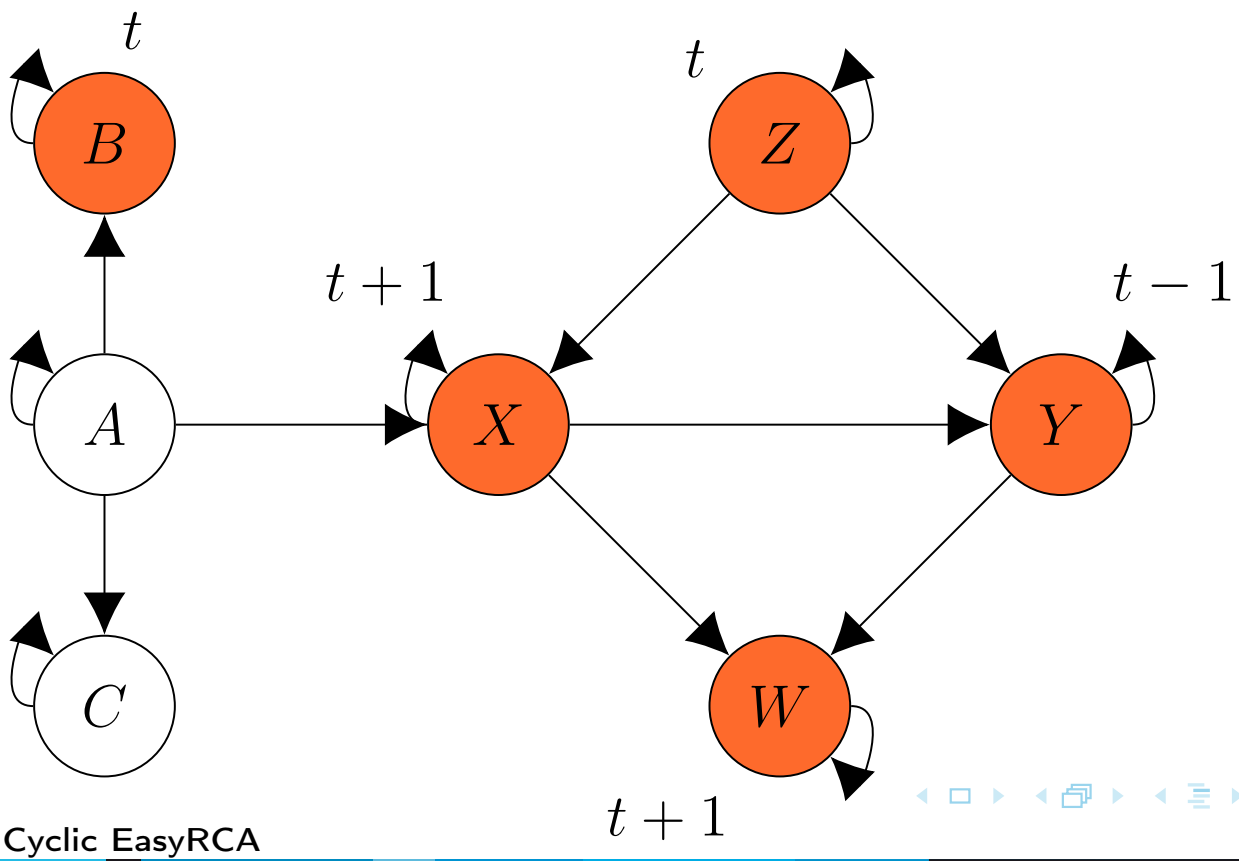
Problem

Find root causes of anomalies given :

- Summary causal graph of the normal regime.
- Maximal lag between a cause and an effect.
- Observational time series.
- Time of appearance of anomalies.

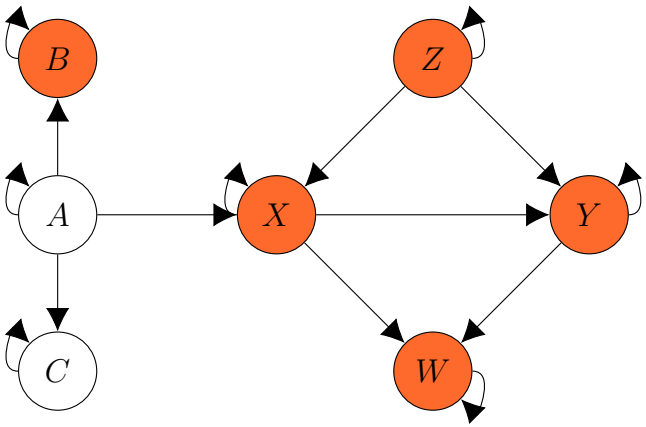
Running Exemple

The EasyRCA algorithm[2] is described here in the case of an **acyclic** summary causal graph.



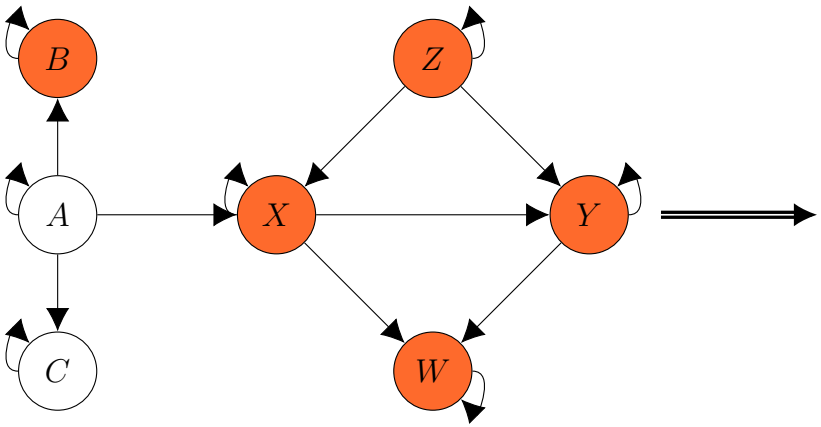
Step 1 : Linked Anomalous Graphs

Compute the connexe anomalous components called linked anomalous graphs.

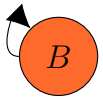


Step 1 : Linked Anomalous Graphs

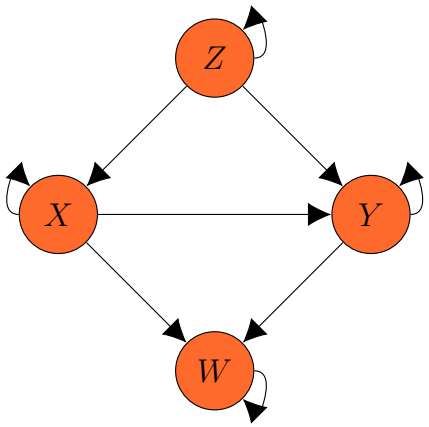
Compute the connexe anomalous components called linked anomalous graphs.



(a) Linked Anomalous Graph 1



(b) Linked Anomalous Graph 2

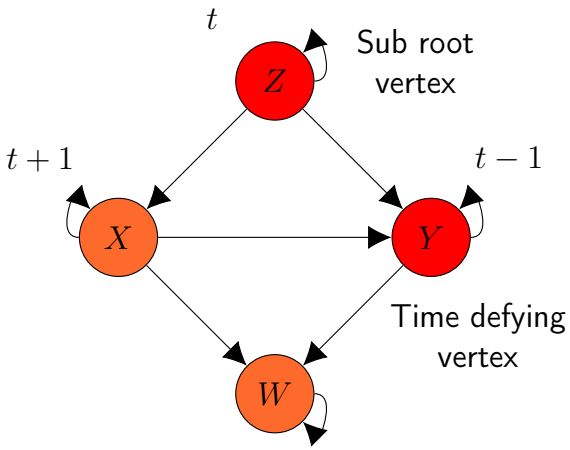


Step 2 : Finding External Intervention from Graph

- *Sub root vertex* : A sub root vertex is root vertex in a linked anomalous graph.
- *Time defying vertex* : A time defying vertex is a vertex on which the anomaly has appeared strictly earlier than on its parents. (Temporal Priority)

Step 2 : Finding External Intervention from Graph

- *Sub root vertex* : A sub root vertex is root vertex in a linked anomalous graph.
- *Time defying vertex* : A time defying vertex is a vertex on which the anomaly has appeared strictly earlier than on its parents. (Temporal Priority)



$$\left.\begin{array}{l}\text{Sub root vertices}\\ +\\ \text{Time defying vertices}\end{array}\right\} \subseteq \text{Root causes}$$

Step 3 : Finding External Intervention from Data

X is a root cause if :

the direct effect[4] in the normal regime

is different from

the direct effect in the anomalous regime

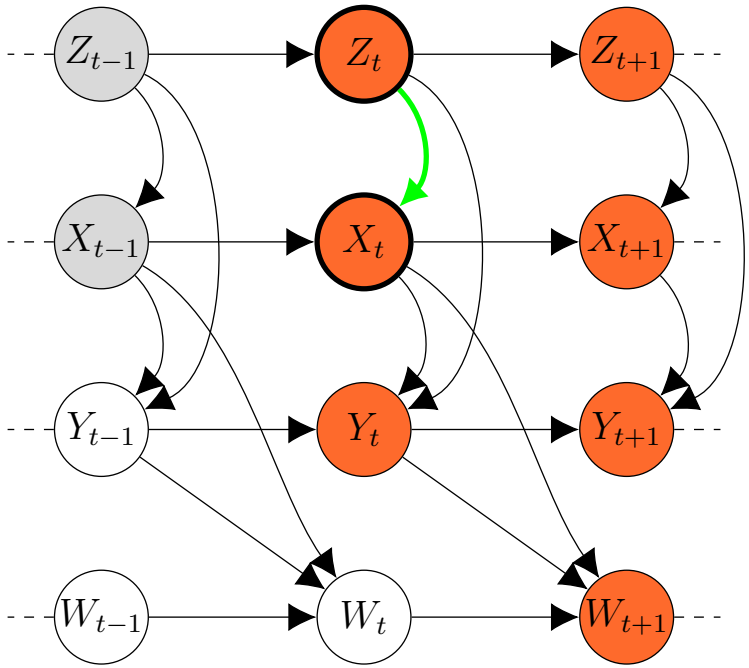
$$E_N[X_t|do(Z_{t-\gamma_{zx}} = z)] - E_N[X_t|do(Z_{t-\gamma_{zx}} = z')]$$

$$\neq$$

$$E_{\bar{N}}[X_t|do(Z_{t-\gamma_{zx}} = z)] - E_{\bar{N}}[X_t|do(Z_{t-\gamma_{zx}} = z')]$$

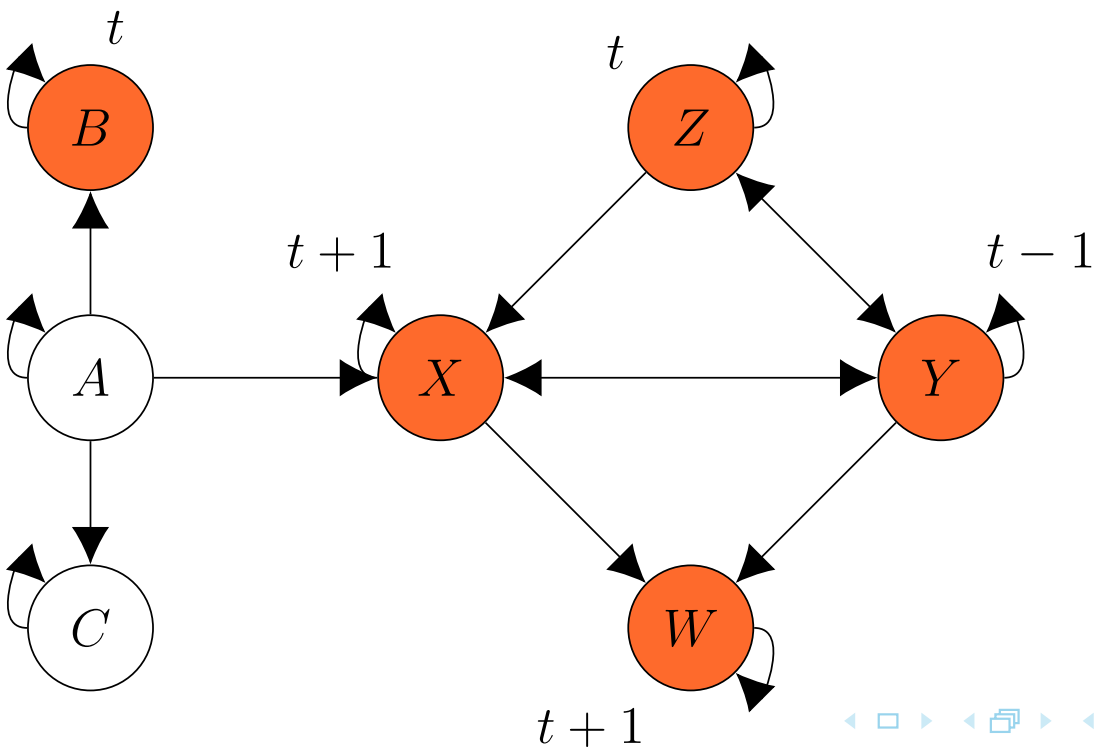
Step 3 : Finding External Intervention from Data

Use the single-door criterion to estimate remove the $do()$ operator.

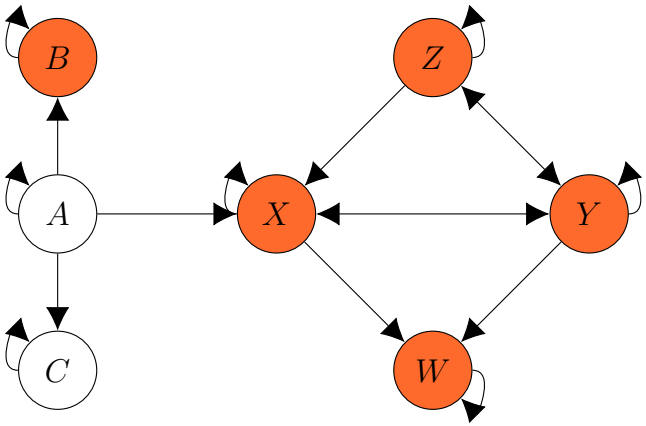


Running Exemple

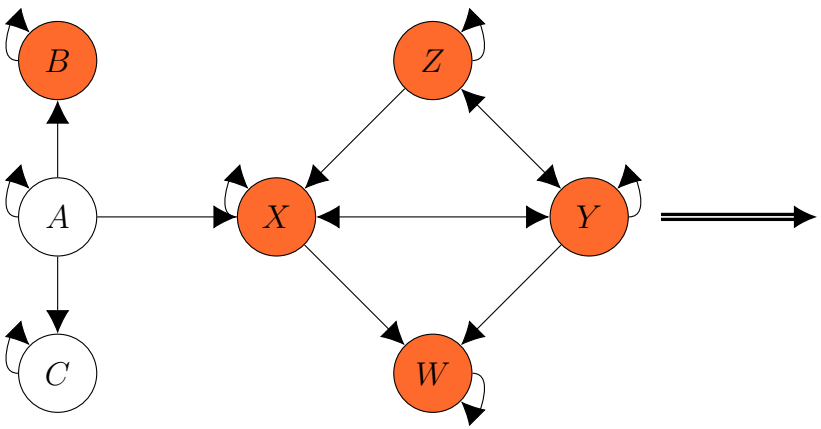
How can the EasyRCA algorithm be adapted in the case of a **cyclic** summary causal. Note that the full-time causal graph is still acyclic.



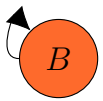
Step 1 : Linked Anomalous Graphs



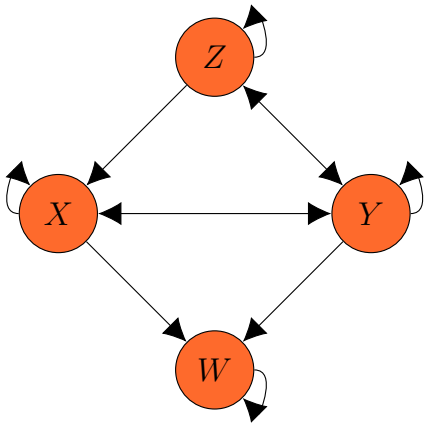
Step 1 : Linked Anomalous Graphs



(a) Linked Anomalous Graph 1

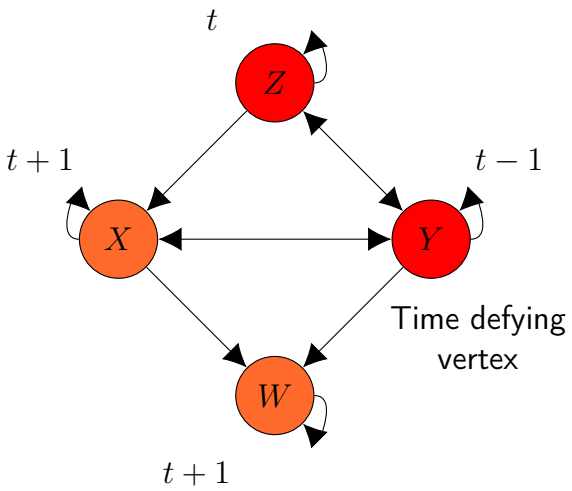


(b) Linked Anomalous Graph 2



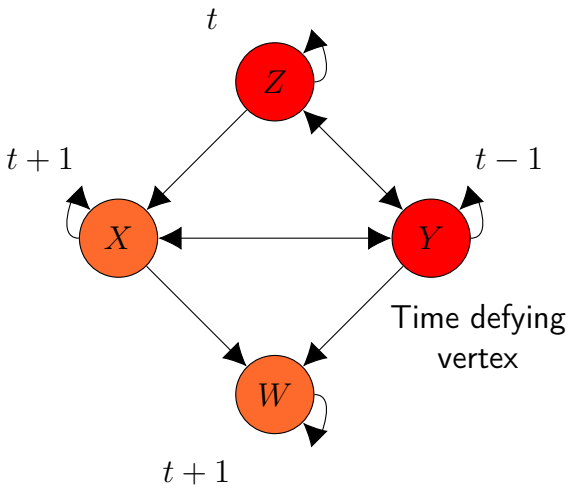
We will consider cyclic linked anomalous graphs.

Step 2 : Finding External Intervention from Graph



Sub root vertices
$+$
Time defying vertices
$\Big\} \subseteq$ Root causes

Step 2 : Finding External Intervention from Graph

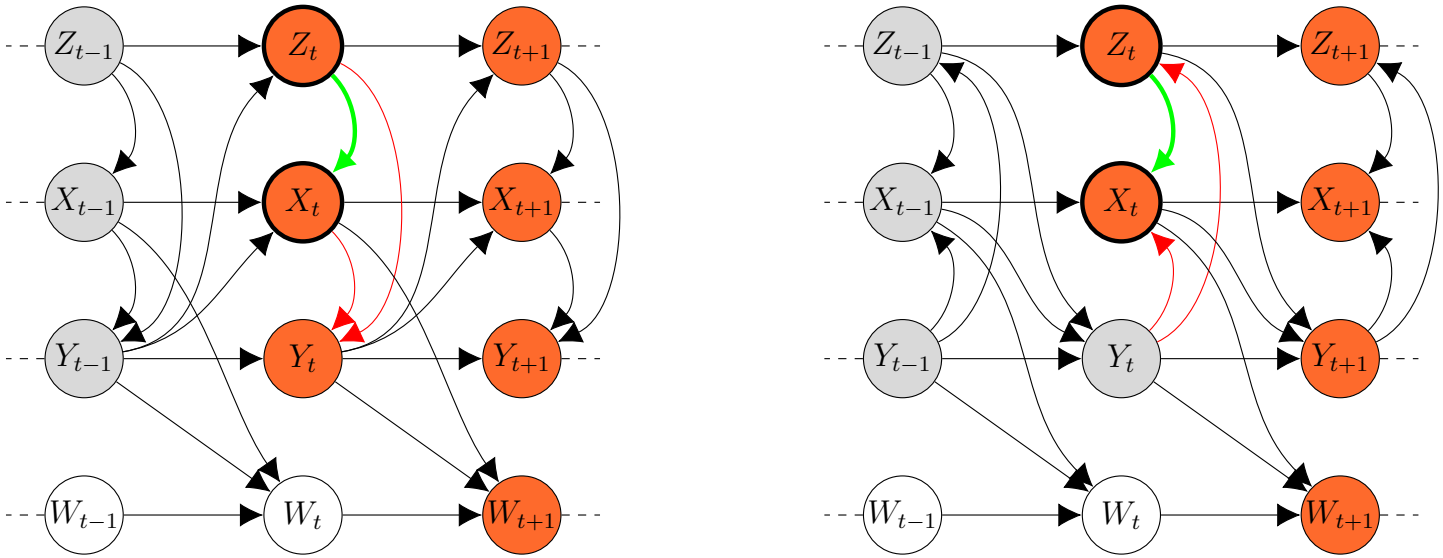


Sub root vertices
+
Time defying vertices
$\}$ $\subseteq$ Root causes

- *Time defying vertex* : No problem in the cyclic case.
- *Sub root vertex* : There will be no sub root vertex in a cycle. Therefore, there may be a linked anomalous graph without any sub root vertices.

Step 3 : Finding External Intervention from Data

This step is the crux of the problem. An adaptation of the single-door criterion to cyclic summary causal graph is needed.



In this example, Y_t can be either a collider or a common cause on a path from Z_t to X_t thus identification is not easy.

Futur Work

- Finish this work by adapting the single-door criterion to cyclic summary causal graph. This may be done by adapting the work presented yesterday by Anouar Meynaoui to the single-door criterion.
- Relax the hypothesis of acyclic full time causal graph.

Thank You

Simon Ferreira : simon.ferreira@ens-lyon.fr

[1] Charles K Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73 :767–819, 2022.

[2] Charles K. Assaad, Imad Ez-zejjari, and Lei Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. *AISTATS*, 2023.

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection : A survey. *ACM Comput. Surv.*, 41(3), jul 2009.

[4] Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK : CambridgeUniversityPress*, 19(2), 2000.

Results of EasyRCA[2]

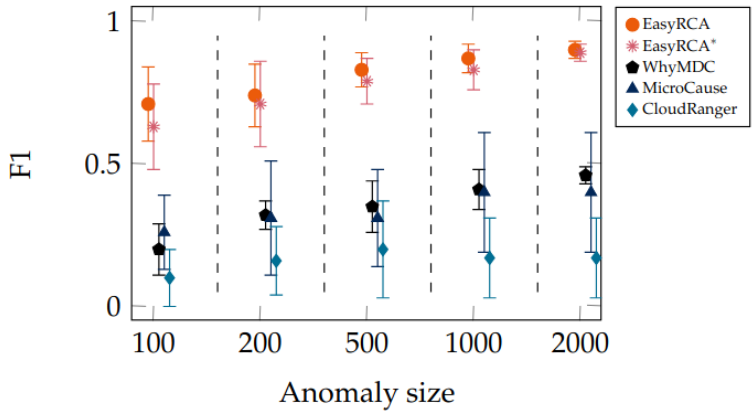


Figure 3: Mean and variance of F1-scores with respect to structural interventions over 30 graphs containing one linked anomalous graph with one sub-root vertex and one structural intervention.

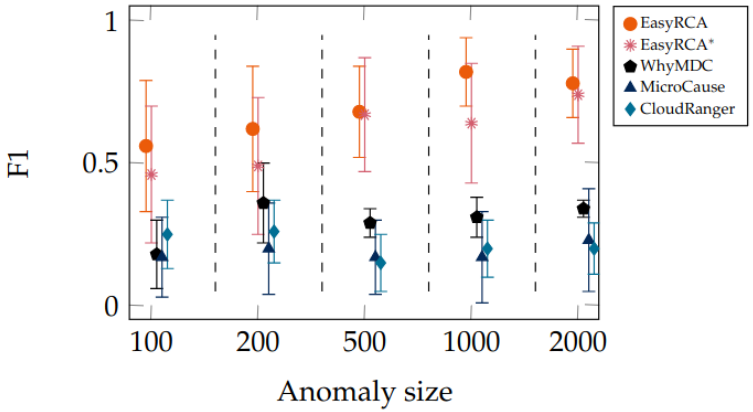


Figure 4: Mean and variance of F1-scores with respect to parametric interventions over 30 graphs containing one linked anomalous graph with one sub-root vertex and one parametric intervention.