Root Cause Identification for Collective Anomalies in Time Series given an Acyclic Summary Causal Graph with Loops

Charles K. Assaad
EasyVista

Imad Ez-zejjari
EasyVista

Lei Zan
EasyVista
Univ Grenoble Alpes,
CNRS, Grenoble INP, LIG

Abstract

This paper presents an approach for identifying the root causes of collective anomalies given observational time series and an acyclic summary causal graph which depicts an abstraction of causal relations present in a dynamic system at its normal regime. The paper first shows how the problem of root cause identification can be divided into many independent subproblems by grouping related anomalies using *d-separation*. Further, it shows how, under this setting, some root causes can be found directly from the graph and from the time of appearance of anomalies. Finally, it shows, how the rest of the root causes can be found by comparing direct effects in the normal and in the anomalous regime. To this end, an adjustment set for identifying direct effects is introduced. Extensive experiments conducted on both simulated and real-world datasets demonstrate the effectiveness of the proposed method.

1 INTRODUCTION

The need for high availability of information systems requires efficient monitoring tools and a new generation of AIOps software to automate the identification of actionable root causes of anomalies in an IT monitoring system that can be used to eliminate the anomalies. In recent years, many approaches have been developed to identify the root causes of anomalies in multivariate time series. The most common

direction that is explored considers discovering the causal graph (Pearl, 2000; Spirtes et al., 2000) that represents the anomalous regime of the dynamic system (Wang et al., 2018; Meng et al., 2020) using observational time series. Causal discovery methods are known to rely on strong assumptions which imply that they necessitate validation by an expert especially when there is no guarantee that these assumptions are satisfied. In addition, causal discovery methods usually need large sample sizes (Malinsky and Danks, 2018; Glymour et al., 2019; Assaad et al., 2022a). However, in many domains, the size of anomalous data depends on the sampling rate of the system. Thus systems with low sampling rates will collect a small size of anomalous data compared to systems with high sampling rates. In consequence, sometimes, it is difficult to have enough data for causal discovery methods. And even if a sufficient amount of data is collected, the process of validation of the causal graph by an expert is time consuming and can delay the elimination of anomalies.

To tackle this issue, we follow a different approach for root cause analysis which consists of discovering and reasoning about the summary causal graphs (Assaad et al., 2022a) which depicts an abstraction of causal relations present in a dynamic system at its normal regime. Usually, the size of data collected in the normal regime is significantly greater than the size of data collected in the anomalous regime since anomalies are supposed to be rare. In addition, system experts can have sufficient time to validate the graph long before the appearance of anomalies. Note that in this work, we do not adress the problem of causal discovery of the summary causal graph of the normal regime and we assume that the graph is already learned and validated by a system expert.

This paper presents a new method for root cause identification, which we call EasyRCA, which consists of using a summary causal graph of the normal regime in order to divide the problem of root cause

identification into many independent subproblems by grouping related anomalies using *d-separation*. Then for each group, EasyRCA finds the root causes either directly from the graph and the time of appearance of anomalies or by comparing direct effects in the normal and anomalous regime. To this end, an adjustment set for identifying direct effects is introduced.

The remainder of the paper is organized as follows: Section 2 introduces some terminology and formalizes the problem. Section 3 describes related work. Section 4 presents our method EasyRCA which is evaluated on simulated and real datasets in Section 5. Finally, Section 6 concludes the paper.

2 Problem setup

In this section, we first introduce some terminology, tools, and assumptions which are standard for the major part. Then, we formalize the problem we are going to solve.

Suppose that a dynamic system can be represented by a structural causal model (SCM) (Pearl, 2000) in which each point in a time series is given by a function (so-called causal mechanism) of its parents and an unobserved noise:

$$Y_t := f_t^y(Parents(Y_t), \xi_t^y) \qquad (1)$$

where the noise variables are jointly statistically independent so that there are no hidden confounding , i.e., causal sufficiency is satisfied (Spirtes et al., 2000). The qualitative causal relations induced by such SCM can be represented by a causal graph in which, under the causal Markov condition (Spirtes et al., 2000) each vertex is independent of all other vertices given its parents, except for its descendants. In dynamic systems, these causal graphs are referred to as full-time causal graphs. An example of such a graph is presented in Figure 1a. The main difficulty in working with this type of graph is that it is infinite and so in practice, inferring it is unfeasible. However, it is very likely that causal relations between two time series will hold throughout time as such relations are generally associated with underlying physical processes. Thus we can assume consistency throughout time.

Assumption 1 (Consistency throughout time, (Assaad et al., 2022a)). *A full time causal graph is said to be* consistent throughout time *if all the causal relationships remain constant in direction throughout time.*

Under this assumption, the full time causal graph can be contracted to give a finite graph which is called a window causal graph. It is a representation of the causal relations through a time window, the size of

which depends on the maximum lag between a cause and an effect in the full time causal graph. An example of a window causal graph is given in Figure1b. This said, it is usually difficult for an expert to validate, analyze let alone provide a window causal graph because it is difficult to determine the temporal lag between a cause and an effect. Thus, experts usually rely on the so-called summary causal graph which is a compact version of the window causal graph that represents the causal relations between time series without giving any information about the temporal lags of these relations. In this work, we assume that the summary causal graph is acyclic but loops are allowed to represent temporal dependencies within the same time series. An example of such graph is given in Figure 1c and formally it is defined as follows:

Definition 1 (Acyclic summary causal graph with loops). *Consider* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *is a* summary causal graph. *The set of vertices in that graph consists of the set of time series. The arcs* $\mathcal{E}$ *of the graph are defined as follows:* $\forall X, Y \in \mathcal{V}$, X *causes* Y *if and only if there exists some time lag* γ *such that* $X_{t-\gamma}$ *causes* Y_t *such that* $\gamma \geq 0$ *for* $X \neq Y$ *and* $\gamma > 0$ *for* $X = Y$. *If* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *has no directed cycles other than the edges going from one vertex to itself, then* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *is said to be an* acyclic summary causal graph with loops *(ASCGL).*

In this work, we suppose that the ASCGL is given either using experts knowledge or learned directly from observational time series (Peters et al., 2013; Assaad et al., 2021, 2022b) or by first discovering a window causal graph[1] (Runge et al., 2019; Runge, 2020) and then deduce the ASCGL from it. The correctness of such learned graphs usually rests on untestable assumptions and depends on the quality of the data so it is important to validate it by an expert or to simplify the problem of causal discovery by providing some background knowledge.

Now, we turn our focus on anomalies. In this work, we assume that anomalies are collective.

Definition 2 (Collective anomaly, Chandola et al. (2009)). *In time series, a collective anomaly is a sequence of data instances that is anomalous with respect to the entire time series.*

Point anomalies are disregarded because we are interested in finding actionable root causes that can eliminate anomalies. If an anomaly appears for one time instant and then disappears this means it was eliminated on its own and do not require any action to re-

---

[1]In our framework, if one decides using a causal discovery algorithm that learns a window causal graph, then one needs to carefully incorporate to these algorithms the constraint that the summary causal graphs should be acyclic.

(a) Full time causal graph   (b) Window causal graph   (c) Summary causal graph
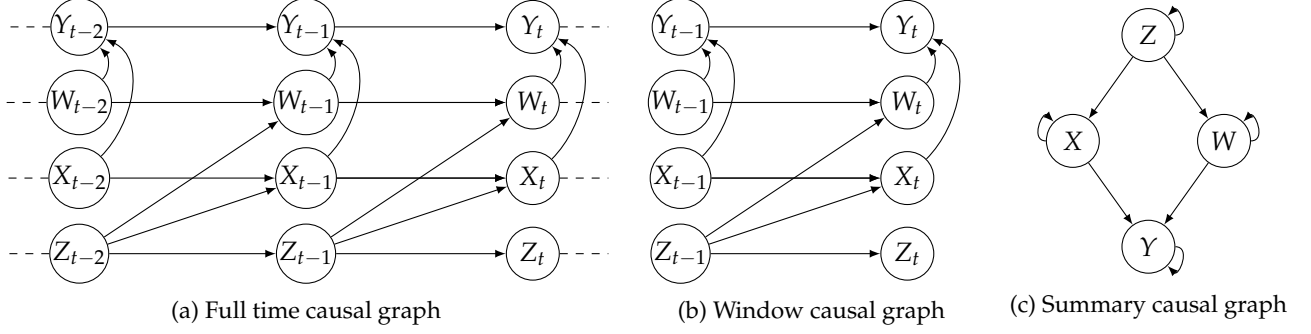
Figure 1: Different causal graphs that one can infer from three time series: full time causal graph (1a), window causal graph (1b) and summary causal graph (1c). Note that the first one gives more information but cannot be inferred in practice, the second one is a schematic viewpoint of the full behavior, whereas the last one is an abstraction and can be deduced from the window causal graph.

solve it. However, we also consider that those collective anomalies have a limited size since information systems are expected to be highly available (without anomalies). In addition, we assume that each anomaly can be eliminated by removing the intervention that caused it directly or via a causal path. This is given by the following assumption:

Assumption 2. *All anomalies are propagated from an external intervention through the structural causal model.*

Given the definition of ASCGL and collective anomalies, we define root causes[2] as follows:

Definition 3 (Root causes)**.** *Given an ASCGL and a set of anomalous vertices $\mathcal{A}$, the set of root causes $\mathcal{C}$ of $\mathcal{A}$ is a set of vertices that were affected by an external intervention which led to marginal distribution change in $\mathcal{A}$.*

In the literature, there exists two sorts of interventions (Eberhardt and Scheines, 2007) and both are crucial to root cause analysis. The first is known as parametric intervention and it is defined as follows:

Definition 4 (Parametric intervention)**.** *Consider an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. An intervention on a vertex $Y \in \mathcal{V}$ is parametric if the causal mechanism before intervention is different than after the intervention but $Parents(Y, \mathcal{G})$ remains unchanged.*

The second type of intervention is known as structural intervention[3] and it is defined as follows:

Definition 5 (Structural intervention)**.** *Consider an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. An intervention on a vertex $Y \in \mathcal{V}$*

is structural if $\exists X \in \mathcal{V}$ such that $X \in Parents(Y, \mathcal{G})$ before the intervention and $X \notin Parents(Y, \mathcal{G})$ after the intervention.

Structural interventions can be regarded as a special case of parametric interventions. But we distinguish between them because an expert might want to differentiate between interventions that provoke a disruption in the system from the ones that do not. As it will be shown in Section 4, to find these two types of interventions we will estimate the direct effects in the normal and anomalous regime. In consequence, we assume that causal mechanisms are fixed throughout time within the same regime (e.g., in Equation 1, $f^y(.)$ is fixed for all t) and we assume the minimality condition[4] (Spirtes et al., 2000) which implies that adjacent vertices in the ASCGL are statistically dependent in the normal regime. Finally, to simplify the problem we assume linear SCMs.

Now that we have introduced the needed tools and assumptions, the problem we are trying to solve is formalized as follows:

Problem. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set of anomalous vertices $\mathcal{A} \subset \mathcal{V}$, the distribution of the time series in the normal regime $\mathcal{N}$ and in the anomalous regime $\bar{\mathcal{N}}$, and the maximal lag between a cause and an effect γ_{max}, we want to find the smallest set of root causes $\mathcal{C}$ of $\mathcal{A}$.*

3 RELATED WORKS

Recently, there has been an increase in the popularity of automating the process of root cause analysis. Among the most popular unsupervised methods that deal with time series is CloudRanger (Wang et al.,

---

[2]The "root causes" are relative to the set of observed time series.

[3]Our definition of structural intervention is less restrictive than the classical definition which states that the intervention alone completely determines the probability distribution of the variable that underwent the intervention, i.e., this variable becomes independent of all of its parents.

[4]The minimality condition is usually assumed by causal discovery methods either directly or by assuming a stronger assumption called faithfulness which implies the minimality condition (Glymour et al., 2019).

2018) which is decomposed into two steps. First, it discovers the summary causal graph between anomalous time series using the PC algorithm (Spirtes et al., 2000) which was introduced for non-temporal data. Then, identifies root causes through random walk based on a transition matrix computed using the correlation between time series. As one might expect, the main limitation of this method is that it uses a non-temporal algorithm that does not take into account temporal lags that might exist between two time series. In addition, correlation does not necessarily represent the causal effect of one variable on another. To fix these issues, Meng et al. (2020) proposed a similar method, called MicroCause, where the causal discovery is done using the PCMCI[5] (Runge et al., 2019; Runge, 2020) algorithm, an extension of the PC algorithm for time series which infers a window causal graph. Then, MicroCause deduces from the inferred window causal graph a summary causal graph. Furthermore, to compute the transition matrix for the random walk, MicroCause estimates the partial correlation between each causally related time series given their parents in the graph. Conditioning on the parents is a sufficient condition to eliminate all spurious correlations when there are no hidden common causes. Note that PC and PCMCI use conditional independencies to infer the causal graph and such methods are known to be correct when the faithfulness condition is satisfied. Closer to our proposal, Budhathoki et al. (2021) introduced a formal method, that we will denote as WhyMDC, to detect the root cause of a change in a marginal distribution from non temporal data. WhyMDC considers that a directed acyclic causal graph is given and as far as we know, it is the first method to identify root causes by searching for changes in causal mechanisms.

There exist other root cause identification methods which are beyond the scope of this paper. For example, Budhathoki et al. (2022) proposed a root cause analysis framework to detect the root cause of a point anomaly using non-temporal structural causal models and Zhang et al. (2022) proposed a supervised learning approach to find root causes.

4 ROOT CAUSE IDENTIFICATION USING ASCGLs

4.1 Grouping related anomalies

We first give an extension of the concept of d-seperation to ASCGL and then show how it can be

---

[5]There exists two versions of PCMCI, one that allows for instantaneous relations (Runge, 2020) and one that does not (Runge et al., 2019). In the experimentation section, we use the version that allows instantaneous relations.

---

used to divide the root cause identification problem into many independent subproblems.

A path is said to be *blocked* by a set of vertices $\mathcal{Z} \in \mathcal{V}$ if it contains a chain $X \rightarrow W \rightarrow Y$ or a fork $X \leftarrow W \rightarrow Y$ and $W \in \mathcal{Z}$, or it contains a collider $X \rightarrow W \leftarrow Y$ such that no descendant of W is in $\mathcal{Z}$. A path is said to be *active* if it is not blocked. Using blocked paths, the notion of d-separation is defined as follows:

Definition 6 (d-separation, Pearl (2000))**.** *Given a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ if every path between a vertex in $\mathcal{X}$ and a vertex in $\mathcal{Y}$ is blocked by $\mathcal{Z}$. We write d-separated as $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{Y} \mid \mathcal{Z}$.*

Note that d-separation was introduced for directed acyclic graph (DAG), so it is directly applicable for full time graphs and window causal graphs but not for summary causal graphs. However, it turned out that the extention to ASCGL is simple. If there are no loops, d-separation in an ASCGL is equivalent to the one in Definition 6. For example, in Figure 1c, if we omit the loops then it is obvious that $X \perp\!\!\!\perp_{\mathcal{G}} W \mid Z$. If there are loops, at first glance, Definition 6 might seem to fail. However, The following proposition shows how Definition 6 can still be used for ASCGL.

Proposition 1. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$, $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{Y} \mid \mathcal{Z}$ if $\mathcal{Z} = Parents(\mathcal{X}, \mathcal{G}) \cup Parents(\mathcal{Y}, \mathcal{G})$ and $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}'} \mathcal{Y} \mid \mathcal{Z}'$ such that $\mathcal{G}'$ is a DAG identical to $\mathcal{G}$ but loops are omitted and $\mathcal{Z}' = Parents(\mathcal{X}, \mathcal{G}') \cup Parents(\mathcal{Y}, \mathcal{G}')$.*

Proof. Consider a DAG $\mathcal{G}'$ without loops. Suppose $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{V}'$, $\mathcal{Z}^x = Parents(\mathcal{X}, \mathcal{G}')$ and $\mathcal{Z}^y = Parents(\mathcal{Y}, \mathcal{G}')$ such that $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}'} \mathcal{Y} \mid \mathcal{Z}^x \cup \mathcal{Z}^y$. Now consider an identical graph $\mathcal{G}$ but with loops. $\forall X_{t-\gamma_{xy}}, Y_t$ such that $\gamma_{xy} \in \mathbb{N}$, the set $\mathcal{Z}^{x,t} = \mathcal{Z}^x_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{Z}^x_{t-\gamma_{xy}-\gamma_{max}}$ contains all parents of $X_{t-\gamma}$ in $\mathcal{Z}^x$ and none of its decedants, thus $\mathcal{Z}^{x,t}$ blocks all active paths going into $X_{t-\gamma_{xy}}$ that does not pass by the past $X_{t-\gamma_{xy}}$. It follows that $\mathcal{Z}^{x,t} \cup X_{t-\gamma_{xy}-1}, \cdots, X_{t-\gamma_{xy}-\gamma_{max}}$ blocks all active paths going into $X_{t-\gamma_{xy}}$. Similarly, $\mathcal{Z}^y_t \cup \cdots \cup \mathcal{Z}^y_{t-\gamma_{max}} \cup Y_{t-1}, \cdots, Y_{t-\gamma_{max}}$ blocks all active paths going into Y_t. Therefore, $\mathcal{X} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{Y} \mid \mathcal{Z}^x \cup \mathcal{Z}^y$ in $\mathcal{G}$. $\qquad\square$

Note that we focused on parents and excluded ancestors to avoid separation sets of infinite size. For example, in Figure 1c, we can explain why $X \not\perp\!\!\!\perp_{\mathcal{G}} W \mid Z$ by looking at the compatible full-time causal graph in Figure 1a where $X_t \not\perp\!\!\!\perp_{\mathcal{G}} W_t \mid Z_{t-1}, X_{t-1}, W_{t-1}$.

Assumption 2 and Definition 3 imply that root causes of anomalous vertices $\mathcal{A}$ is a subset of $\mathcal{A}$. So in the following, we consider that $\mathcal{A} = \mathcal{C} \cup \bar{\mathcal{C}}$. Such that $\mathcal{C}$ represents root causes and $\bar{\mathcal{C}}$ represents non root causes.

Proposition 2. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and anomalous vertices $\mathcal{A} \subseteq \mathcal{V}$ such that $\mathcal{A} = \mathcal{C} \cup \bar{\mathcal{C}}$, $\forall \mathcal{S} \subseteq \mathcal{V} \backslash \mathcal{A}, \mathcal{C} \not\perp\!\!\!\perp_{\mathcal{G}} \bar{\mathcal{C}} \mid \mathcal{S}$.*

Proof. Consider anomalous vertices $\mathcal{A} = \mathcal{C} \cup \bar{\mathcal{C}}$, such that $\mathcal{C}$ is the set of root causes of $\bar{\mathcal{C}}$. If $\exists \mathcal{S} \subseteq \mathcal{V} \backslash \mathcal{A}$ such that $\mathcal{C} \perp\!\!\!\perp_{\mathcal{G}} \bar{\mathcal{C}} \mid \mathcal{S}$ then by definition of d-separation all paths between $\mathcal{C}$ and $\bar{\mathcal{C}}$ are blocked given $\mathcal{S}$ which means all directed paths from $\mathcal{C}$ to $\bar{\mathcal{C}}$ are blocked given $\mathcal{S}$. In this case, $\forall X \in \mathcal{C}, \forall Y \in \bar{\mathcal{C}}$, there exists no directed path π from X to Y such that each vertex on π belongs to $\mathcal{A}$. It follows that $\bar{\mathcal{C}}$ is not propagated from $\mathcal{C}$ which contradicts Assumption 2. Hence it must be the case that $\forall \mathcal{S} \subseteq \mathcal{V} \backslash \mathcal{A}, \mathcal{C} \not\perp\!\!\!\perp_{\mathcal{G}} \bar{\mathcal{C}} \mid \mathcal{S}$. □

Definition 7 (Linked anomalous graph). *Given an AS-CGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a set of anomalous vertices $\mathcal{A} \subset \mathcal{V}$. $\mathcal{L} = \{\mathcal{L}^1, \cdots, \mathcal{L}^m\}$ is a set of linked anomalous graphs if $\forall i \in \{1, \cdots, m\}$ $\mathcal{L}^i = (\mathcal{A}^i, \mathcal{E}^i)$ is a subgraph of $\mathcal{G}$ such that $\mathcal{A}^i \subset \mathcal{A}$ and there exists a set of vertices $\mathcal{S} \subset \mathcal{V} \backslash \mathcal{A}$ such that $\mathcal{A}^i \perp\!\!\!\perp_{\mathcal{G}} \mathcal{A} \backslash \mathcal{A}^i \mid \mathcal{S}$.*

For example, consider the ASCGL in Figure 2: B, C, D, W, X, Y, Z are anomalous vertices and A is a normal vertex. Since $B, C, D \perp\!\!\!\perp_{\mathcal{G}} W, X, Y, Z \mid A$ and B, C, D and W, X, Y, Z are respectively d-connected given A, then B, C, D and W, X, Y, Z form two linked anomalous graphs.

Proposition 3. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if the set of linked anomalous graphs is $\mathcal{L} = \{\mathcal{L}^1, \cdots, \mathcal{L}^m\}$, then $\forall i, j \in \{1, \cdots, m\}, \mathcal{L}^i \cap \mathcal{L}^j = \emptyset$.*

Proof. Consider two different linked anomalous graphs $\mathcal{L}^1 = (\mathcal{A}^1, \mathcal{E}^1)$ and $\mathcal{L}^2 = (\mathcal{A}^2, \mathcal{E}^2)$ such that $\mathcal{A}^1 \perp\!\!\!\perp_{\mathcal{G}} \mathcal{A} \backslash \mathcal{A}^1 \mid \mathcal{S}^1$ and $\mathcal{A}^2 \perp\!\!\!\perp_{\mathcal{G}} \mathcal{A} \backslash \mathcal{A}^2 \mid \mathcal{S}^2$ such that $\mathcal{S}^1, \mathcal{S}^2 \subseteq \mathcal{V} \backslash \mathcal{A}$. It follows that if $\mathcal{L}^1 \cap \mathcal{L}^2 \neq \emptyset$ then $\exists X \in \mathcal{V}$ such that $X \in \mathcal{A}^1$ and $X \in \mathcal{A}^2$. In consequence, $\nexists \mathcal{S} \subseteq \mathcal{V} \backslash \mathcal{A}$ such that $\mathcal{A}^1 \not\perp\!\!\!\perp_{\mathcal{G}} \mathcal{A}^2 \mid \mathcal{S}$. Which means according to Definition 7, $\mathcal{A}^1$ and $\mathcal{A}^2$ belong to the same linked anomalous graph. □

Proposition 4. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if the set of linked anomalous graphs is $\mathcal{L} = \{\mathcal{L}_1, \cdots, \mathcal{L}_m\}$, then $\forall i, j \in \{1, \cdots, m\}, \mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ such that $\mathcal{C}_i$ is the set of root causes of $\mathcal{L}_i$ and $\mathcal{C}_j$ is the set of root causes of $\mathcal{L}_j$.*

Proof. This follows from Proposition 3. □

Propositions 3 and 4 suggest that linked anomalous graphs are modular with respect to each other, which implies that the set of root causes of each linked anomalous graph can be identified independently of the rest of the anomalies in the graph.

Next, we will show how to detect a subset of the root causes uniquely by looking at the graph and the time of the first appearance of anomalies on each vertex.
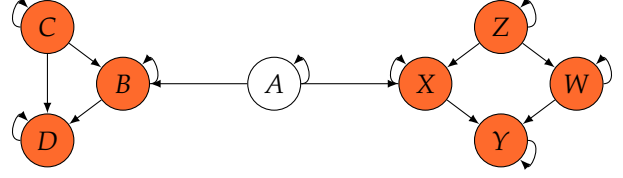

Figure 2: An ASCGL with two linked anomalous graphs. White vertices represents normal vertices and orange vertices represents anomalous vertices.

4.2 Identifying root causes from the graph

Definition 8 (Sub-root vertex). *A sub-root vertex is root vertex in a linked anomalous graph.*

Definition 9 (Time defying vertex). *Consider a linked anomalous graph $\mathcal{L}^i = (\mathcal{A}^i, \mathcal{E}^i)$. Y is a time defying vertex if and only if $\forall X \in Parents(Y, \mathcal{L}^i)$ the appearance time of the anomaly on Y precedes the appearance time of the anomaly on X.*

Proposition 5. *Given a linked anomalous graph $\mathcal{L}^i = (\mathcal{A}^i, \mathcal{E}^i)$, its set of sub-root vertices $\mathcal{R}^i$, and its set of time defying vertices $\mathcal{T}^i$, then $\mathcal{R}^i \cup \mathcal{T}^i \subseteq \mathcal{C}^i$ such that $\mathcal{C}^i$ is the true set of root causes in $\mathcal{L}^i$.*

Proof. Consider a linked anomalous graph $\mathcal{L}^i = (\mathcal{A}^i, \mathcal{E}^i)$. 1) By Definition 8, a sub-root vertex in $\mathcal{L}^i$ does not have any anomalous parent, it follows from Assumption 2 and Propositions 3,4 that the anomaly in this vertex cannot be propagated from other vertices, which implies that it was itself directly affected by an external intervention which means it is a root cause. 2) Consider two anomalous vertices $X, Y \in \mathcal{A}^i$ such that $Parents(Y, \mathcal{L}^i) = \{X\}$. According to Assumption 2, the anomaly in X which appeared at time t was propagated to Y according to the lag $\gamma \geq 0$ between X and Y of the SCM. It follows that if the appearance time of the anomaly on Y is t' such that $t' < t \leq t + \gamma$ then the anomaly on Y was not propagated from X, which implies that Y was itself directly affected by an external intervention which means it is a root cause. Using induction, suppose this is true for $Parents(Y, \mathcal{L}^i) = \{X^1, \cdots, X^p\}$. If $Parents(Y, \mathcal{L}^i) = \{X^1, \cdots, X^{p+1}\}$ such that t' precedes the appearance time of the anomalies on $\{X^1, \cdots, X^p\}$ then if t' does not precede the appearance time on X^{p+1}, then the anomaly on Y could have been propagated from X^{p+1}, otherwise, we conclude that Y was directly affected by an external intervention which means it is a root cause. □

Proposition 5 states that if a vertex is a sub-root or a time-defying vertex in a linked anomalous graph then it belongs to the set of root causes. However, this does not mean that every element in the set of root causes

is necessarily a sub-root or a time-defying vertex. Therefore, there might be a vertex $X \in \mathcal{A}^i \backslash \{\mathcal{R}^i \cup \mathcal{T}^i\}$, such that $X \in \mathcal{C}^i$.

4.3 Identifying root causes from data

To find root causes that are neither sub-roots nor time-defying vertices, we search for changes in the causal mechanisms. In some cases, we can find these changes, for each $X \to Y$ in a given linked anomalous graph, by estimating the total effect[6] (Pearl, 2000) of X on Y in the normal regime, defined as

$$TE^N_{X_{t-\gamma_{xy}} \to Y_t} = E_N[Y_t | do(X_{t-\gamma_{xy}} = x)] \qquad (2)$$
$$- E_N[Y_t | do(X_{t-\gamma_{xy}} = x')]$$

and the total effect of X on Y in the anomalous regime, defined as

$$TE^{\bar{N}}_{X_{t-\gamma_{xy}} \to Y_t} = E_{\bar{N}}[Y_t | do(X_{t-\gamma_{xy}} = x)] \qquad (3)$$
$$- E_{\bar{N}}[Y_t | do(X_{t-\gamma_{xy}} = x')],$$

where E_N and $E_{\bar{N}}$ are respectivly the expectations in the normal regime and in the anomalous regime and the temporal lag γ_{xy} is between $\bar{\gamma}_{xy}$ and γ_{max} such that $\bar{\gamma}_{xy}$ can be found by the substracting the time of appearance of anomalies on Y from the time of appearance of anomalies on X. If there is no directed path from X to Y (other than $X \to Y$) or if directed paths between X and Y exist but we know that all vertices (other than X and Y) on these paths cannot be root causes, then if $TE^N_{X_{t-\gamma_{xy}} \to Y_t} \neq TE^{\bar{N}}_{X_{t-\gamma_{xy}} \to Y_t}$ we can conclude that there is a change in the causal mechanism of Y provoked by an external intervention on Y.

In order to estimate $TE^N_{X_{t-\gamma_{xy}} \to Y_t}$ and $TE^{\bar{N}}_{X_{t-\gamma_{xy}} \to Y_t}$ from observational data we need to eliminate the *do* from the total effect expression. This can be achieved, when there is no hidden common causes, by the back-door criterion (Pearl, 2000) which searches for an adjustment set of vertices, called back-door set, that eliminates all spurious correlations between X and Y. However, the back-door criterion cannot directly be applied to ASCGL because of loops. In the following we present an adjustment set for identifying total effects[7] from ASCGLs:

Definition 10 (An adjustment set for total effects in an ASCGL). *Consider an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a maximal lag γ_{max}, two vertices X and Y such that $X \to Y$ in $\mathcal{G}$, and the temporal lag γ_{xy} between*

---

[6]The *do* operator represents an external intervention.
[7]A similar result was presented in (Eichler and Didelez, 2007) for summary causal graphs assuming there is no instantaneous relations and allowing for sets of infinite size.

X and Y. The adjustment set for identifying the total effect relative to $(X_{t-\gamma_{xy}}, Y_t)$ is $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}} \cup \mathcal{X}$ such that: 1. $\mathcal{B} = Parents(X, \mathcal{G}) \backslash \{X\}$; 2. $\mathcal{X} = \{X_{t-\gamma_{xy}-1}, \cdots, X_{t-\gamma_{xy}-\gamma_{max}}\}$ if there exists a loop on X in $\mathcal{G}$, otherwise $\mathcal{X} = \{\varnothing\}$.

Proposition 6. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a maximal lag γ_{max}, if $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}} \cup \mathcal{X}$ satisfies Definition 10 in $\mathcal{G}$ relative to $(X_{t-\gamma_{xy}}, Y_t)$ then $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}} \cup \mathcal{X}$ blocks all activated paths between $X_{t-\gamma_{xy}}$ and Y_t going into $X_{t-\gamma_{xy}}$ in every window causal graph associated with $\mathcal{G}$.*

Proof Sketch. If X and Y have no loops in $\mathcal{G}$, then $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}}$ is sufficient to block all paths between $X_{t-\gamma_{xy}}$ and Y_t going into $X_{t-\gamma_{xy}}$ since all possible parents of $X_{t-\gamma_{xy}}$ are in $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}}$. Given that $\mathcal{G}$ is acyclic, $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}}$ cannot block any directed path nor create any new activated path between $X_{t-\gamma_{xy}}$ and Y_t as there cannot be any descendant of $X_{t-\gamma_{xy}}$ in $\mathcal{B}_{t-\gamma_{xy}} \cup \cdots \cup \mathcal{B}_{t-\gamma_{xy}-\gamma_{max}}$. If X and Y have loops then adjusting on $\mathcal{B}$ up to γ_{max} cannot block all back-door paths because there will always be an activated path passing by $\mathcal{B}_{t-\gamma_{xy}-\gamma_{max}-i}$ such that $i > 0$. The only way to block this path is to add the past of $X_{t-\gamma_{xy}}$ to the adjustment set. $\square$

When there exist directed path between X and Y the total effect would no longer be reliable to detect external interventions. For example, in Figure 2, if there's an external intervention on B then $TE^N_{C_{t-\gamma_{cd}} \to D_t} \neq TE^{\bar{N}}_{C_{t-\gamma_{cd}} \to D_t}$ due to the change in the causal mechanism of the mediator B of C and D. To avoid such cases, for each $X \to Y$ in a given linked anomalous graph, we need to estimate the direct effect (Pearl, 2000) of X on Y in the normal regime, defined as

$$DE^N_{X_{t-\gamma_{xy}} \to Y_t} = E_N[Y_t | do(X_{t-\gamma_{xy}} = x, \mathcal{W} = w)] \quad (4)$$
$$- E_N[Y_t | do(X_{t-\gamma_{xy}} = x', \mathcal{W} = w)]$$

and the direct effect of X on Y in the anomalous regime, defined as

$$DE^{\bar{N}}_{X_{t-\gamma_{xy}} \to Y_t} = E_{\bar{N}}[Y_t | do(X_{t-\gamma_{xy}} = x, \mathcal{W} = w)] \quad (5)$$
$$- E_{\bar{N}}[Y_t | do(X_{t-\gamma_{xy}} = x', \mathcal{W} = w)],$$

where $\mathcal{W} = \mathcal{V}_{t-\gamma_{max}} \cup \cdots \cup \mathcal{V}_{t-\gamma_{xy}} \backslash \{X_{t-\gamma_{xy}}\} \cup \cdots \cup \mathcal{V}_t \backslash \{Y_t\}$.

Assuming linearity, the *do* from the direct effect expression can be eliminated (i.e., direct effect can be identified) using any adjustment set given by the single-door criterion (Pearl, 2000) which is not applicable in ASCGLs. In the following, we present an adjustment set for direct effects in ASCGLs:

Definition 11 (An adjustment set for direct effects in an ASCGL). *Consider an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a maximal lag γ_{max}, two vertices X and Y such that $X \rightarrow Y$ in $\mathcal{G}$, and the temporal lag γ_{xy} between X and Y. An adjustment set for identifying the direct effect relative to $(X_{t-\gamma_{xy}}, Y_t)$ is $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}$ such that: 1. $\mathcal{B} = Parents(Y, \mathcal{G}) \backslash \{X, Y\}$; 2. $\mathcal{X} = \{X_t, \cdots, X_{t-\gamma_{max}}\} \backslash \{X_{t-\gamma_{xy}}\}$ and $\mathcal{Y} = \{Y_{t-1}, \cdots, Y_{t-\gamma_{max}}\}$ if there exists a loop on Y in $\mathcal{G}$, otherwise $\mathcal{Y} = \{\varnothing\}$.*

Proposition 7. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a maximal lag γ_{max}, if $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}$ satisfies Definition 11 in $\mathcal{G}$ relative to $(X_{t-\gamma_{xy}}, Y_t)$ then $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}$ blocks all activated paths between $X_{t-\gamma_{xy}}$ and Y_t in every window causal graph associated with $\mathcal{G}$ except the direct path $X_{t-\gamma_{xy}} \rightarrow Y_t$.*

Proof Sketch. If X and Y have no loops in $\mathcal{G}$, then $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}}$ is sufficient to block all paths from $X_{t-\gamma_{xy}}$ to Y_t except $X_{t-\gamma_{xy}} \rightarrow Y_t$ since and all possible parents of Y_t are in $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}}$. Given that $\mathcal{G}$ is acyclic $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}}$ cannot create any new activated path that is not blocked by $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}}$ since no vertex in $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}}$ is a descendant of Y_t. If X and Y have loops then adjusting on $\mathcal{B}$ up to γ_{max} cannot block all paths because there can be an activated path passing by the future of $X_{t-\gamma_{xy}}$ or by the past of Y_t. The only way to block these types of paths is to add $\mathcal{X} \cup \mathcal{Y}$ to the adjustment set. $\qquad \square$

4.4 An algorithm for root cause identification

Here, we describe our main method called EasyRCA[8], in which the pseudocode is provided in Algorithm 1. The algorithm starts by finding linked anomalous graphs (line 1). Then for each linked anomalous graph, it searches for the sub-roots and time-defying vertices (line 3). Finally, it searches for the rest of the root causes by comparing direct effects in the normal regime with direct effects in the anomalous regime (lines 4-16). The conditions in lines 12 and 14 need the minimality condition because if X and Y are statistically independent given the adjustment set in the normal regime then an intervention on Y might not imply any change to the statistical distribution thus one cannot conclude on the presence of interventions. The for-loop in line 2 can be parallelized since as showed in Proposition 3 and 4, linked anomalous graphs are modular.

Theorem 1. *Given an ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a set anomalous vertices $\mathcal{A} \subseteq \mathcal{V}$, the distribution of the time series in the normal regime $\mathcal{N}$ and in the anomalous regime $\bar{\mathcal{N}}$, and*

[8]Code available at https://github.com/ckassaad/EasyRCA

the maximal lag between a cause and an effect γ_{max}, under Assumption 2 and the minimality condition, EasyRCA is capable of identifying the set of root cause $\mathcal{C}$ of $\mathcal{A}$.

Proof Sketch. It follows from Propositions 3, 4, 5, 7. $\qquad \square$

Note that we can also distinguish between parametric and structural interventions. Given that $DE^N_{X_{t-\gamma_{xy}} \rightarrow Y_t} \neq DE^{\bar{N}}_{X_{t-\gamma_{xy}} \rightarrow Y_t}$, if $DE^{\bar{N}}_{X_{t-\gamma_{xy}} \rightarrow Y_t} = 0$, we conclude that the intervention on Y is structural, otherwise, we conclude that it is parametric.

Algorithm 1 EasyRCA

Require: ASCGL $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, distribution of the time series in the normal regime $\mathcal{N}$ and in the anomalous regime $\bar{\mathcal{N}}$, maximal lag γ_{max}, Anomalies $\mathcal{A}$
1: $\mathcal{L}^1, \cdots, \mathcal{L}^m$ = list of linked anomalous graphs as in Definition 7
2: **for** $i \in \{1, \cdots, m\}$ **do**
3: $\quad$ Identify sub-root vertices $\mathcal{R}^i$ and time defying vertices $\mathcal{T}^i$ using Definition 8 and 9
4: $\quad$ Let $\mathcal{D}^i = []$
5: $\quad$ Let $\mathcal{A}^i$ be the set of vertices in $\mathcal{L}^i$
6: $\quad$ **for** Y in $\mathcal{A}^i \backslash \{\mathcal{R}^i \cup \mathcal{T}^i\}$ **do**
7: $\qquad$ **for** X in $Parents(Y, \mathcal{G})$ **do**
8: $\qquad\quad$ $\bar{\gamma}_{xy}$: anomaly lag between X and Y
9: $\qquad\quad$ **for** γ_{xy} in $\{\bar{\gamma}_{xy}, \cdots, \gamma_{max}\}$ **do**
10: $\qquad\qquad$ Identify $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}$ using Definition 11
11: $\qquad\qquad$ Estimate $DE^N_{X_{t-\gamma_{xy}} \rightarrow Y_t}$
12: $\qquad\qquad$ **if** $DE^N_{X_{t-\gamma_{xy}} \rightarrow Y_t} \neq 0$ **then**
13: $\qquad\qquad\quad$ Estimate $DE^{\bar{N}}_{X_{t-\gamma_{xy}} \rightarrow Y_t}$
14: $\qquad\qquad\quad$ **if** $DE^N_{X_{t-\gamma_{xy}} \rightarrow Y_t} \neq DE^{\bar{N}}_{X_{t-\gamma_{xy}} \rightarrow Y_t}$ **then**
15: $\qquad\qquad\qquad$ $\mathcal{D}^i = [\mathcal{D}^i, Y]$
16: $\qquad\qquad\qquad$ Break
17: **Return** $\mathcal{R}, \mathcal{T}, \mathcal{D}$

---

5 EXPERIMENTS

We propose first an extensive analysis on simulated data, generated from random causal graphs; then we perform an analysis on a real word dataset.

5.1 Experimental Setup

In practice, to test if $DE^N_{X_{t-\gamma_{xy}} \rightarrow Y_t} \neq DE^{\bar{N}}_{X_{t-\gamma_{xy}} \rightarrow Y_t}$, we fit 11 multiple linear regressions:

$$Y_t = \hat{a}_x X_{t-\gamma_{xy}} + \sum_{B_{t-\gamma_{by}} \in \mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}} \hat{a}_b B_{t-\gamma_{by}} + \epsilon^y_t,$$

such that $\mathcal{B}_t \cup \cdots \cup \mathcal{B}_{t-\gamma_{max}} \cup \mathcal{X} \cup \mathcal{Y}$ is identified using Definition 11. One of them fitted on the anomalous data and 10 fitted on different chunks of the normal data. Then using the Grubbs-test (Grubbs, 1950) we check if the coefficient $\hat{a}_x$ of the anomalous data is significantly different than the 10 others $\hat{a}_x$ from the normal data. To test if $TE^{\bar{N}}_{X_{t-\gamma_{xy}} \to Y_t} = 0$, we use a t-test on the coefficient $\hat{a}_x$ of the anomalous data.

Baselines: We compare EasyRCA with three other methods[9]: CloudRanger, MicroCause and a naive adaptation of WhyMDC to time series for which we provide a window causal graph. Since CloudRanger and MicroCause try to solve a harder problem compared to EasyRCA by inferring the summary causal graph from anomalous data (while EasyRCA considers that the summary causal graph is given), we also consider another version of EasyRCA, denoted as EasyRCA*, where we suppose that the summary causal graph is not given. In the first step of EasyRCA*, we infer the window causal graph from normal data using PCMCI (Runge, 2020), the same causal discovery algorithm used by MicroCause, and then we deduce the summary causal graph from it. Note that the summary causal graph obtained from the window graph that is inferred by PCMCI can be cyclic even if the true summary causal graph is acyclic (due to estimation errors). In such cases, we consider that EasyRCA* does not identify any root cause.

Hyper-parameters: For EasyRCA, EasyRCA* and MicroCause, we set the maximal lag γ_{max} to 3 and for all methods (even though the true γ_{max} is smaller in our simulation study), we set the significance threshold to 0.01. For EasyRCA*, CloudRanger and MicroCause we use a Fisher-z-test (Kalisch and Bühlmann, 2007), which is commonly used for causal discovery when linearity and gaussianity are satisfied. Furthermore, for CloudRanger and MicroCause we set the walk length to 1000 and the backward step threshold to 0.1. Lastly, other hyper-parameters in WhyMDC were set to the default values in the DoWhy package.

Evaluation: To assess the quality of identifying root causes, we use the F1-score. Since by construction EasyRCA identifies sub-root and time defying vertices as root causes we do not evaluate their detection. This ensures a fair comparison with other methods.

5.2 Simulated Data

For simulated data, we start by randomly generating 30 different ASCGLs such that each graph contains 6 vertices, has a maximal degree between 4 and 5, and

---

[9]We implemented CloudRanger and MicroCause and we adapted WhyMDC based on the DoWhy package.

has one root vertex. We consider that all lags in the window causal graph associated with any of the generated ASCGL are equal to 1. So the generative process (the SCM) is the following:

$$Y_t = \sum_{X_{t-1} \in Parents(Y_t, \mathcal{G}_w)} a X_{t-1} + 0.1\xi^y_t$$

where $a \sim U\{0.1, 1\}$, $\xi^y_t \sim \mathcal{N}(0, 1)$, Y_t denotes the value of the vertex at time t, $Parents(Y_t)$ denotes the direct parents of Y_t in the window causal graph.

For each ASCGL, we choose two root causes (two vertices that will undergo an intervention): the first is the root of the ASCGL and the second is a randomly chosen vertex among the non-root vertices. We propagate the effect of each intervention according to the generating process toward all the descendants of the vertex which underwent an intervention. We set the starting time of each intervention according to the generative process. For example, if the root vertex X has an intervention at time t and a vertex Y randomly selected to undergo an intervention is a child of X then the starting time of the intervention on Y is $t + 1$. In general, the starting time of the intervention does not need to respect the generative process, but we chose to respect it to avoid any time-defying vertices which would give an advantage to our method.

In our experiments, we consider the two types of interventions separately and we vary the anomaly size between 100 and 2000. In the case of structural interventions, values of the root vertex and values of a randomly chosen non-root vertex in the anomalous interval are replaced by data drawn form the distribution $Exp(2)$. In the case of parametric interventions, values of the root are set similarly to structural interventions. Then values of the non-root vertex are re-generated with new coefficients from $U(0.1, 1)$.

Results: In Figure 3, we report the performance of each method at detecting structural interventions with respect to the anomaly size. As one can see, EasyRCA and EasyRCA* clearly outperform other methods, and their performances increases (and their variance decreases) significantly between the anomaly of size 100 and the anomaly of size 2000 reaching an F1-score of 1 for both EasyRCA and EasyRCA*. The small difference in the performance of EasyRCA and EasyRCA* shows that our method is robust with respect to small errors in the ASCGL. WhyMDC and MicroCause have similar results and outperform CloudRanger. In Figure 4, we report the performance of each method at detecting parametric interventions with respect to the anomaly size. As before, EasyRCA and EasyRCA* outperforms other methods but now the difference between EasyRCA and EasyRCA* is more visible and all other meth-

ods suffer. However, it is worth noting, that unlike CloudRanger, WhyMDC and MicroCause were able to detect the root vertex of each graph as a root cause.
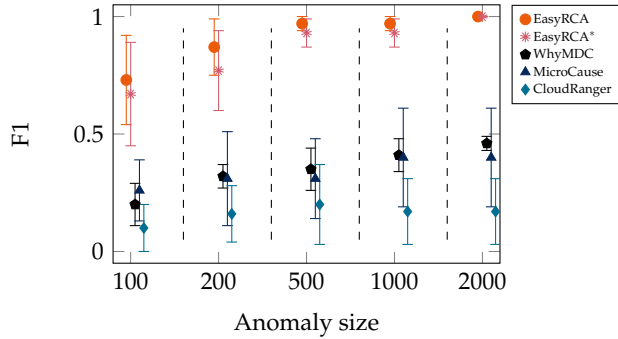

Figure 3: Mean and variance of F1-scores with respect to structural interventions over 30 graphs containing one linked anomalous graph with one sub-root vertex and one structural intervention.
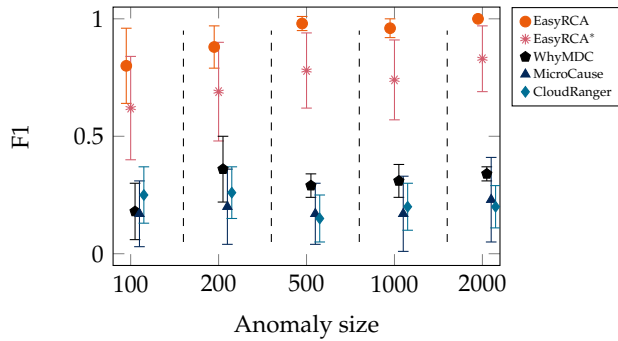

Figure 4: Mean and variance of F1-scores with respect to parametric interventions over 30 graphs containing one linked anomalous graph with one sub-root vertex and one parametric intervention.

5.3 Real Data

For real data, we consider a dataset[10] which consists of eight time series collected from an IT monitoring system with a one-minute sampling rate provided by EasyVista[11] such that each of these time series is considered anomalous and all collective anomalies are considered to have the same time of appearance and of size 100. The corresponding ASCGL is provided in Figure 5 where PMDB represents the extraction of some information about the messages received by the Storm ingestion system; MDB refers to an activity of a
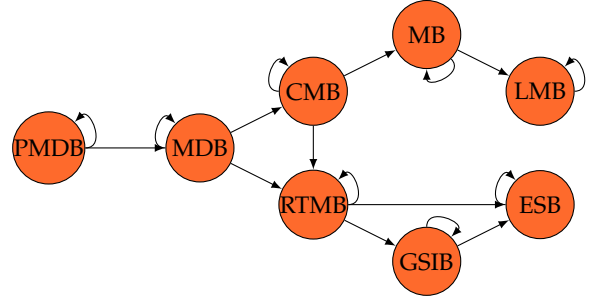
---

[10]The real IT monitoring data is available at https://easy vista2015-my.sharepoint.com/:f:/g/personal/aait-bachir_ easyvista_com/ElLiNpfCkO1JgglQcrBPP9IBxBXzaINrM5 f0ILz6wbgoEQ?e=OBTsUY

[11]https://www.easyvista.com/fr/produits/ev-observe

Figure 5: ASCGL of the normal regime of an IT monitoring system. All vertices are anomalous in the anomalous regime. According to EasyVista's system experts, PMDB and ESB are expected to be the root causes of these anomalies.

process that orient messages to other process with respect to different types of messages; CMB represents the activity of extraction of metrics from messages; MB represents the activity of insertion of data in a database; LMB reflects the updates the last values of metrics in Cassandra; RTMB represents the activity of searching to merge of data with information coming from the check message bolt; GSIB represents the activity of insertion of historical status in database. ESB represents the activity of writing data in Elasticsearch. According to EasyVista's system experts, PMDB and ESB are expected to be the root causes of these anomalies.

EasyRCA inferred 3 roots causes, PMDB as a root vertex, in addition to RTMB and ESB as structural interventions. EasyRCA* inferred 5 roots causes, PMDB, GSIB and MB as root vertices, in addition to RTMB and ESB as structural interventions. MicroCause inferred that PMDB and MB are the root causes of the anomalies. CloudRanger inferred that GSIB and MDB are the root causes. We did not apply WhyMDC in this real world application because the true window causal graph is unknown. In terms of the trade-off between false positives and false negatives, EasyRCA gives the best result.

6 Conclusion

We adressed the problem of identifying root causes of collective anomalies using observational time series and an ASCGL of the normal regime of a given system. We showed that the problem can be divided into many independent subproblems and that all root causes can be identified using the graph and the data. For future work, it would be interesting to extend this method for cyclic summary causal graphs, for nonlinear SCMs and to allow for hidden common causes.

Acknowledgements

References

Charles K. Assaad, Emilie Devijver, Eric Gaussier, and Ali Ait-Bachir. A mixed noise and constraint-based approach to causal inference in time series. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 453–468, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86486-6.

Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *J. Artif. Int. Res.*, 73, apr 2022a. doi: 10.1613/jair.1.13428.

Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Entropy-based discovery of summary causal graphs in time series. *Entropy*, 24(8), 2022b. ISSN 1099-4300. doi: 10.3390/e24081156.

Kailash Budhathoki, Dominik Janzing, Patrick Bloebaum, and Hoiyi Ng. Why did the distribution change? In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1666–1674. PMLR, 13–15 Apr 2021.

Kailash Budhathoki, Lenon Minorics, Patrick Bloebaum, and Dominik Janzing. Causal structure-based root cause analysis of outliers. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 2357–2369. PMLR, 17–23 Jul 2022.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41:15:1–15:58, 2009.

Frederick Eberhardt and Richard Scheines. Interventions and causal inference. *Philosophy of Science*, 74 (5):981–995, 2007. ISSN 00318248, 1539767X.

Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, UAI'07, page 109–116, Arlington, Virginia, USA, 2007. AUAI Press. ISBN 0974903930.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graph-ical models. *Frontiers in Genetics*, 10, 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00524.

Frank E. Grubbs. Sample Criteria for Testing Outlying Observations. *The Annals of Mathematical Statistics*, 21(1):27 – 58, 1950. doi: 10.1214/aoms/1177729885.

Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *J. Mach. Learn. Res.*, 8:613–636, may 2007. ISSN 1532-4435.

Daniel Malinsky and David Danks. Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1):e12470, 2018. doi: https://doi.org/10.1111/phc3.12470.

Yuan Meng, Shenglin Zhang, Yongqian Sun, Ruru Zhang, Zhilong Hu, Yiyin Zhang, Chenyang Jia, Zhaogang Wang, and Dan Pei. Localizing failure root causes in a microservice through causality inference. In *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*, pages 1–10, 2020. doi: 10.1109/IWQoS49365.2020.9213058.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-77362-8.

Jonas Peters, D. Janzing, and B. Schölkopf. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems 26*, pages 154–162, 2013.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 1388–1397. PMLR, 03–06 Aug 2020.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), 2019. doi: 10.1126/sciadv.aau4996.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.

Ping Wang, Jingmin Xu, Meng Ma, Weilan Lin, Disheng Pan, Yuan Wang, and Pengfei Chen. Cloudranger: Root cause identification for cloud native systems. In *2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 492–502, 2018. doi: 10.1109/CCGRID.2018.00076.

Chaoli Zhang, Zhiqiang Zhou, Yingying Zhang, Linxiao Yang, Kai He, Qingsong Wen, and Liang Sun. Netrca: An effective network fault cause localization algorithm, 2022.