Probabilities of Causation and Root Cause Analysis with Quasi-Markovian Models

Eduardo Rocha Laurentino[13], Fabio Gagliardi Cozman[2], Denis Deratani Mauá[2], Daniel Angelo Esteves Lawand[2], Davi Goncalves Bezerra Coelho[3], and Lucas Martins Marques[2]

[1]Instituto de Ciência e Tecnologia Itaú (ICTi)
[2]Escola Politécnica, Universidade de São Paulo (EP-USP)
[3]Instituto de Matemática e Estatística, Universidade de São Paulo (IME-USP)

Abstract. Probabilities of causation provide principled ways to assess causal relationships but face computational challenges due to partial identifiability and latent confounding. This paper introduces both algorithmic simplifications, significantly reducing the computational complexity of calculating tighter bounds for these probabilities, and a novel methodological framework for Root Cause Analysis that systematically employs these causal metrics to rank entire causal paths.

Keywords: Root cause analysis · Partial identifiability · Causal inference.

1 Introduction

Causal reasoning underpins intelligent decision-making and effective problem-solving across diverse domains. Several frameworks for causal inference have emerged, with Pearl's Structural Causal Model (SCM) being particularly influential due to its rigorous algorithmic foundation and applicability [8]. Pearl's approach has seen successful adoption in various contexts, including medicine [9], manufacturing [6], and reliability engineering.

In particular, *probabilities of causation*, such as Probability of Necessity (PN), Probability of Sufficiency (PS) and Probability of Necessity and Sufficiency (PNS), have demonstrated value in multiple fields, such as medical diagnostics, explainable artificial intelligence (XAI), climate change analysis and representation learning. Recent works highlight their effectiveness: Watson et al. (2022) developed local explanations explicitly utilizing necessity and sufficiency; Cai et al. (2022) optimized GNN explanations by maximizing PNS; and Yang et al. (2023) employed a PNS-based risk metric for robust representation learning. These examples illustrate the practical relevance and theoretical robustness of these causal probabilities.

However, the calculation of probabilities of causation poses significant computational challenges. In particular, probabilities of causation often suffer from partial identifiability issues. Latent confounding and non-monotonic relationships prevent exact identification, forcing reliance on computationally intensive

optimization methods that yield intervals rather than precise values. These computations involve multilinear programming subject to linear constraints, representing a substantial barrier for practical applications.

Addressing this complexity, we first provide theoretical advancements that significantly simplify these multilinear programs. By extending recent techniques developed for interventional inference to the realm of counterfactual inference, we offer techniques that reduce the cost of computating bounds for probabilities of causation. These results, detailed in Section 3, streamline the calculation and enhance the feasibility of applying these metrics in real-world settings.

Our methodological contributions center around the use of probabilities of causation within the practical domain of Root Cause Analysis (RCA). Indeed, many practical circumstances call for a search for causes: for instance, we may search for the cause of an illness in a patient, or the cause of a service malfunction in a banking transaction. In general, an abnormal observation of a target quantity might admit several causes. For instance, a service outage may be due to a memory leak that was caused by a buggy security upgrade. The memory leak fails to be a *root* cause as it is in turn caused by the buggy upgrade.

Traditional RCA methods often struggle under conditions of limited observability, where latent factors confound observable metrics. Probabilities of causation, with their intrinsic ability to capture counterfactual scenarios, provide a rigorous metric aligned precisely with the fundamental queries of RCA: assessing whether an observed failure would persist without a candidate cause or would necessarily occur when the cause is present. We propose a novel RCA methodology leveraging probabilities of causation not merely as isolated indicators but as core metrics for systematically comparing candidate root causes.

The methodological proposal involves employing these probabilities to score and rank entire causal paths within a DAG, rather than isolated variables alone. This approach, detailed in Section 4.2, allows identification not only of the direct root causes but also of the complete causal narrative leading to an incident. Our empirical results, conducted using the microservices domain, demonstrate the effectiveness of this causal-path scoring technique, providing clear, actionable insights into complex incident chains.

Thus, this paper makes two complementary contributions: it advances theoretical computation of probabilities of causation, and it introduces a practical RCA methodology that exploits these probabilities to find causal paths under realistic conditions of limited observability. After reviewing essential background material in Section 2, we elaborate on computational improvements in Section 3, then we present our methodological proposal in Section 4.2 and, finally, we discuss its experimental validation in Section 4.3. We summarize our findings and suggest future research in Section 5.

2   Causal Inference and Structural Causal Models

A *Structural Causal Model* (SCM) is a tuple $(\mathcal{G}, \mathbf{V}, \mathbf{U}, \mathcal{F}, \Pr(\mathbf{U}))$ whose elements are as follows. $\mathcal{G}$ is a directed graph whose node set is $\mathbf{V} \cup \mathbf{U}$, where $\mathbf{V}$ are inner
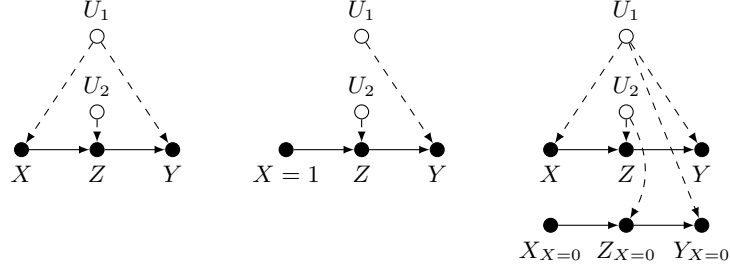
Fig. 1. Left: a simple SCM, capturing Pearl's Smoking/Tar/Cancer example [8]. Middle: the intervened graph, where the edge from U_1 to X is removed, while X is fixed at 1. Right: a twin network where the counterfactual "world" is generated by $X = 0$.

nodes called *endogenous* and $\mathbf{U}$ are root nodes called *exogenous*. $\mathcal{F}$ is a set of functions $f : \mathrm{val}(\mathrm{Pa}(X)) \to \mathrm{val}(X)$, each one of them called a *mechanism*, one mechanism for each endogenous variable X in $\mathcal{G}$, with $\mathrm{Pa}(X)$ denoting the parents of X in the directed graph. Finally, $\mathrm{Pr}(\mathbf{U})$ is a joint probability distribution over the exogenous random variables $\mathbf{U}$.

We adopt a few assumptions about an SCM in this work: (i) the model is *quasi-Markovian*, meaning that each endogenous variable V_i has at most one exogenous parent; (ii) the support of the endogenous variables is finite; (iii) the exogenous variables are independent; and (iv) the graph $\mathcal{G}$ is acyclic. Assumptions (iii) and (iv) are quite common in causal inference, as they allow us to interpret the paths in the graph in terms of probabilistic (in)dependences. Assumption (i) is arguably the most restrictive, but still covers a sufficiently large class of models of interest. Relaxing Assumption (i) makes the subsequent optimization problems much harder, and often practically intractable.

Figure 1 (left) depicts a quasi-Markovian SCM with three endogenous variables and two exogenos variables.

A *confounded component* (for short, c-component) of a graph $\mathcal{G}$ is the set of endogenous nodes of a connected component of the undirected graph obtained by discarding edges whose endpoints are endogenous variables, then dropping direction of the remaining edges [12]. We say that an exogenous variable is associated with an endogenous variable if they appear in the same connected component of the previous undirected graph. For a quasi-Markovian SCM, every c-component is associated with exactly one exogenous node.

In Figure 1 we have two c-components: $\{X, Y\}$ (associated with U_1), and $\{Z\}$ (associated with U_2).

For any endogenous variable V in c-component $\mathbf{C}$, denote by $\mathbf{W}_V$ the endogenous variables that are topologically smaller than V in the union of variables in $\mathbf{C}$ and their endogeneous parents. Tian [12] showed that the probability distribution of the endogenous random variables in quasi-Markovian SCMs factorizes

as a product over c-components:

$$\Pr(\mathbf{V}) = \prod_{\mathbf{C}} \prod_{V \in \mathbf{C}} \Pr(V|\mathbf{W}_V). \tag{1}$$

For any c-component $\mathbf{C}$ associated with exogenous U, Expression (1) leads to the following system of linear equalities:

$$\prod_{V \in \mathbf{C}} \Pr(V|\mathbf{W}_V) = \sum_{u:f_V(\mathrm{Pa}(V))=V,V \in \mathbf{C}} \Pr(U = u), \tag{2}$$

where the summation is over the values of U that are consistent with the configurations of $\mathbf{C}$ and $\mathrm{Pa}(\mathbf{C}) \cap \mathbf{V}$.

Suppose the cardinalities of exogenous variables are not known. When endogenous variables are categorical, one can always produce categorical exogenous variables by a process called *canonicalization* [15], which, in essence, enumerates all possible mechanisms as latent variables. For quasi-Markovian graphs, canonicalization reduces every exogenous variable to a categorical random variable whose state space has cardinality $\prod_{V \in \mathbf{C}} |\mathrm{val}(V)|^{|\mathrm{val}(\mathrm{Pa}(V))|}$, where $\mathbf{C}$ is the corresponding c-component. Each value $u \in \mathrm{val}(U)$ specifies a mechanism $f_u : \mathrm{Pa}(V) \cap \mathbf{V} \to V$ for each V in the c-component. Thus, we assume without loss of generality in the rest that every exogenous variable is categorical.

A simple *intervention* (a.k.a. treatment, exposure, action) $\mathrm{do}(\mathbf{X} = \mathbf{x})$, for $\mathbf{X} \subseteq \mathbf{V}$, modifies an SCM by substituting each mechanism f_X, $X \in \mathbf{X}$, with the constant value x consistent with $\mathbf{x}$, thus obtaining a new SCM where the intervened variables have no parent. This intervened model induces a new (post-intervention) distribution over the variables, denoted by $\Pr(\mathbf{V}|\mathrm{do}(\mathbf{X} = \mathbf{x}))$. Given a variable $Y \in \mathbf{V}$, a *potential outcome* $Y_{\mathbf{X}=\mathbf{x}}(\mathbf{u})$ is the value that *unit* $Y(\mathbf{u})$ would obtain had $\mathbf{X}$ been intervened upon to be $\mathbf{x}$.[1] The probability distribution of $Y_{\mathbf{X}=\mathbf{x}}$ is then the probability induced by the distribution of the exogenous random variables. This is the same marginal distribution one would obtain had we generated data from the intervened SCM $\mathrm{do}(\mathbf{X} = \mathbf{x})$. The same definitions extend to any subset of random variables $\mathbf{Y} \subseteq \mathbf{V}$, giving us sets of counterfactual random variables $\mathbf{Y}_{\mathbf{X}=\mathbf{x}}$ with joint marginal distribution $\Pr(\mathbf{Y}_{\mathbf{X}=\mathbf{x}})$.

Figure 1 (middle) depicts the intervened version of the SCM in Figure 1 (left), where X is set to 1.

A *counterfactual inference* is more intricate. Given some fixed (factual) observation $\mathbf{E} = \mathbf{e}$, the notation $\Pr(\mathbf{Y}_{\mathbf{X}=\mathbf{x}}|\mathbf{E} = \mathbf{e})$ denotes the distribution of $\mathbf{Y}_{\mathbf{X}=\mathbf{x}}$ induced by $\Pr(\mathbf{U}|\mathbf{E} = \mathbf{e})$ [7]. That is, the observation is taken before the intervention is realized. For example, we might ask for $\Pr(\mathbf{Y}_{\mathbf{X}=\mathbf{x}_0} = \mathbf{y}_0|\mathbf{X} = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}_1)$, the probability of $\mathbf{Y}$ taking on some values $\mathbf{y}_0$ had we intervened upon $\mathbf{X}$ to be $\mathbf{x}_0$ when our actual observation was that $\mathbf{Y} = \mathbf{y}_1$ and $\mathbf{X} = \mathbf{x}_1$.

Twin networks are often used to compute counterfactual quantities [8]. A twin network is produced by replicating the endogenous variables and leaving

---

[1] Note: by construction, once the values of the exogenous random variables are fixed at some $\mathbf{U} = \mathbf{u}$, the endogenous variables become deterministic.

the new variables with the same exogenous parents (the exogenous variables are not replicated). For instance, Figure 1 (right) shows the twin network related to Figure 1 (left) so as to compute say $\Pr(Y_{X=0} = 0|Y = 1)$.

3 Computing the Probabilities of Causation

Suppose we have, as usual in causal modeling, a given SCM and an *input distribution* $\hat{\Pr}$ over the endogenous variables $\mathbf{V}$. The latter distribution $\hat{\Pr}$ might come from a large dataset of observed endogenous variables. The assumption here is that endogenous variables are observed and exogenous variables are not.

In looking for causes of a particular observation in this SCM, we can use three quantities based on counterfactual reasoning [7]. The first is the *probability of necessity* (PN) of a cause X for an effect Y, defined as:

$$PN = \Pr(Y_{X=0} = 0|X = 1, Y = 1).$$

That is, PN tries to capture the degree to which $X = 1$ is necessary for $Y = 1$ to happen. The *probability of sufficiency* (PS) is then:

$$PS = \Pr(Y_{X=1} = 1|X = 0, Y = 0).$$

That is, PS tries to capture the degree to which $X = 1$ is sufficient for $Y = 1$ to occur. Finally, the *probability of necessity and sufficiency* (PNS) is:

$$PNS = \Pr(Y_{X=1} = 1, Y_{X=0} = 0).$$

These three quantities can be computed by replicating the variables in $\mathbf{V}$, often referred to as "factual" variables, so that each intervention produces a fresh set of "counterfactual" variables. That is, to compute the PN, we have $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{V}_{X=0}$, and the latter is associated with the same mechanisms associated with $\mathbf{V}$ except for the intervened X.

In the context of root cause analysis, Oliveira et al. [6] proposed interventional quantities as proxies for the previous definitions. Inspired by their proposals, we define, for a tentative cause X and an effect of interest Y:

- The *weak probability of necessity* (w-PN) is $\Pr(Y = 0|\mathrm{do}(X = 0))$.
- The *weak probability of sufficiency* (w-PS) is $\Pr(Y = 1|\mathrm{do}(X = 1))$.

Usually, the cardinality of exogenous variables is not known; as noted in Section 2, the canonicalization procedure can then be employed to build the exogenous variables.

There are interventional probabilities that can be identified using previous results (that is, a single real number is obtained from graph and input distribution). For instance, for the graph in Figure 1 (left), the w-PS can be precisely computed using the front-door criterion [8]. However, in general, interventional and counterfactual probabilities cannot be precisely identified, and we must resort to computing tight lower and upper bounds for them. Hence we must focus, for any Q, either on its *lower value* $\underline{Q} = \inf Q$, or its *upper value* $\overline{Q} = \sup Q$.

Previous work has shown that the computation of lower and upper values for interventional quantities, given a canonicalized SCM an an input distribution, leads to minimization/maximization of a multilinear objective function subject to linear constraints [14]. For a quasi-Markovian SCM, the degree of the multilinear objective function is, in principle, equal to the number of exogenous variables. Hence the feasible region is closed and the upper probabilities are obtained by maximization while the lower probabilities are obtained by minimization [15].

In a recent breakthrough, Shridharam and Iyengar [11] have shown that interventional quantifies require a multilinear objective function whose degree is *only* equal to the number of distinct intervened c-components. Hence, tight bounds for interventions in a single c-component require only linear programs! Such results exploit properties of the input distribution to remove exogenous variables from the calculations.

Alas, these previous results by Shridharan and Iyengar do *not* apply to counterfactual quantities such as PN or PS. This can be already seen in connection with the simple network in Figure 1 (left). Suppose we wish to compute PN; to do so, we resort to the twin network in Figure 1 (right). However, in that network we cannot remove U_2; there is no corresponding simplification to be made to the ensuing multilinear objective function.

Is it possible to remove exogenous variables while computing probabilities of causation PN, PS, PNS? We now present novel results regarding this question, in essence by combining and extending previous efforts.

We start by noting that the PN and the PS can be written as

$$\text{PN} = \frac{\Pr(Y_{X=0} = 0, X = 1, Y = 1)}{\hat{\Pr}(X = 1, Y = 1)}, \quad \text{PS} = \frac{\Pr(Y_{X=1} = 1, X = 0, Y = 0)}{\hat{\Pr}(X = 0, Y = 0)}, \quad (3)$$

where numerators are computed using the twin network, and where denominators depend only on the factual variables whose distribution is the input $\hat{\Pr}$. The latter fact let us avoid fractional objective functions while computing PN or PS.

The next move is to use the results of Shpitser and Pearl [10] that allow us to simplify the twin networks. In fact, Shpitser and Pearl studied general settings in which many different counterfactual "worlds" must be considered, moving away from the two-world setting of twin networks. Their *counterfactual graphs* have as many copies of the endogenous variables as there are different interventions. More importantly, a counterfactual graph removes variables that are not necessary, in some cases by merging variables that are actually identical. The next example illustrates the properties of counterfactual graphs in an interventional setting.

Example 1. Consider the quasi-Markovian SCM in Figure 2 (left). A twin network might be built and, based on it, a multilinear objective function for minimization of PN for cause X and effect Y would have degree 6. The counterfactual graph is shown in Figure 2 (middle), taking intervention $X = 1$. Note that R and W (and associated exogenous variables) are removed as they cannot affect the computation of interest (they are not ascendants of relevant variables). Also, the duplicates of S and T are not included in the counterfactual graph, because
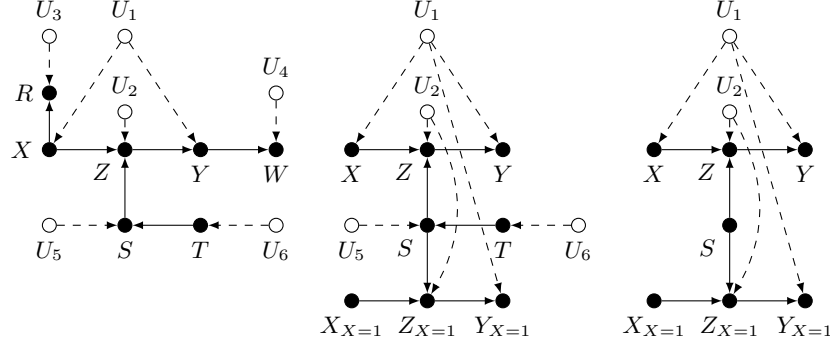
Fig. 2. Left: a quasi-Markovian SCM. Middle: the counterfactual graph for cause X and effect Y. Right: the reduced counterfactual graph (applying Theorem 1).

they *must* be identical to factual S and T as they are associated with identical mechanisms and no intervened variables. □

A counterfactual graph can lead to a substantial simplification: in this example, we moved from a multilinear program with degree 6 to one with degree 4. Can we remove additional exogenous variables? Under some conditions, yes:

Theorem 1. *Suppose we have a set of endogenous variables* $\mathbf{Z}$ *in a counterfactual graph such that no variable in* $\mathbf{Z}$ *is a descendant of the counterfactual target* $Y_{X=x}$ *(where x can be zero or one, depending on whether PN or PS is to be computed), the factual intervened* X *or the factual target* Y *; moreover, suppose* $\mathbf{Z}$ *d-separates, given the factual intervened* X *and the factual target* Y *, all its ascendants from* $Y_{X=x}$. *Then the computation of lower and upper versions of PN and of PS are not affected if all ascendants of* $\mathbf{Z}$ *(including exogenous variables) are removed from the counterfactual graph, and the marginal distribution of* $\mathbf{Z}$ *is then be set at* $\hat{\Pr}(\mathbf{Z})$ *(from the input distribution* $\hat{\Pr}$*).*

Proof. Let $\mathbf{W}$ denote the ascendants of $\mathbf{Z}$ not in $\mathbf{Z}$. Since $\mathbf{Z}$ d-separates $Y_{X=x}$ and $\mathbf{W}$ given X and Y, we can compute $\Pr(Y_{X=x} = y | X = 1 - x, Y = 1 - x)$ as

$$\sum_{\mathbf{Z}} \Pr(Y_{X=x} = y | X = 1 - x, Y = 1 - x, \mathbf{Z}) \sum_{\mathbf{W}} \Pr(\mathbf{Z}, \mathbf{W} | X = 1 - x, Y = 1 - x).$$

Now the conditional probability distribution over $\mathbf{Z}, \mathbf{W}$ is invariant to the intervention (as it refers only to pre-treatment variables). Hence, it can be computed from the input distribution (i.e., it is identifiable) $\Pr(\mathbf{Z} | X = 1 - x, Y = 1 - x)$. This is the same estimand one obtains if we start from a modified model where $\mathbf{W}$ is removed and $\mathbf{Z}$ is replaced by a single node whose marginal is $\hat{\Pr}(\mathbf{Z})$. □

This theorem resembles simplifications by Shridharan and Iyengar [11]; however, they do not deal with counterfactual variables. And their results can be adapted to prove:

Corollary 1. *Suppose a counterfactual graph is reduced using Theorem 1. Tight bounds for PN or PS can be computed by minimizing/maximizing a multilinear objective function whose degree is equal to the number of exogenous variables in the reduced graph, subject to linear constraints.*

Example 2. Consider again Example 1. Using Theorem 1, we can reduce the counterfactual graph in Figure 2 (middle) to the graph in Figure 2 (right); U_5 and U_6 have been removed. Bounds on PN or on PS can be obtained minimizing/maximizing appropriate multilinear objective functions of degree 2 (instead of degree 4 using the counterfactual graph), as shown in Example 3. □

Now consider the computation of PNS. This is somewhat different from PN or PS, as now we have two interventions (thus producing two counterfactual "worlds"). However, the two counterfactual variables in $\Pr(Y_{X=1} = 1, Y_{X=0} = 0)$ are not descendants of the factual endogenous variables, hence the latter variables are not present in the counterfactual graph. Moreover, the simplifications to the counterfactual graph that are described by Theorem 1 can be adapted to the computation of PNS as follows:

Theorem 2. *Suppose we have a set of endogenous variables $\mathbf{Z}$ in a counterfactual graph such that no variable in $\mathbf{Z}$ is a descendant of counterfactual variables $Y_{X=1}$ and $Y_{X=0}$; moreover, suppose $\mathbf{Z}$ d-separates all its ascendants from $Y_{X=1}$ and $Y_{X=0}$. Then the computation of lower and upper versions of PNS are not affected if all ascendants of $\mathbf{Z}$ (including exogenous variables) are removed from the counterfactual graph, and the marginal distribution of $\mathbf{Z}$ is then be set at $\hat{\Pr}(\mathbf{Z})$ (from the input distribution $\hat{\Pr}$).*

A parallel to Corollary 1 can then be derived (that is, tight bounds are obtained minimizing/maximizing a multilinear expression whose degree is the number of exogenous variables in a reduced graph).

Example 3. Consider again Example 1 and suppose we wish to compute <u>PN</u>. As noted previously, we have to process the graph in Figure 2 (right) to obtain a bilinear objective function (that is, a multilinear expression of degree 2) subject to linear constraints. Consider first the linear constraints, produced out of Expression (2). Writing these linear constraints down requires specifying an encoding for the mechanisms (as done by Balke and Pearl in their original discussion of interventional bounds [1]). Start with the c-component associated with U_1 to impose constraints on $\Pr(U_1)$. Note that U_1 has 8 values; denote $\Pr(U_1 = j)$ by $q_{1,j}$ with $j \in \{0, \ldots, 7\}$. Suppose we take $X = 0$ when $j \in \{0, 1, 2, 3\}$ and $X = 1$ otherwise; and we take $Y = 0$ when $j = 0$ or $j = 4$, $Y = Z$ when $j = 1$ or $j = 5$, $Y = 1 - Z$ when $j = 2$ or $j = 6$, and $Y = 1$ when $j = 3$ or $j = 7$. We then have 8 constraints, one per configuration of X, Y, Z, using the template

$$\sum_j q_{1,j} [\![f_X(j) = x]\!] [\![f_Y(z, j) = y]\!] = g_1(xyz),$$

where $[\![\theta]\!]$ is the indicator function of equality θ (that is, 1 is θ holds, 0 otherwise), and $g_1(xyz) := \hat{\Pr}(Y = y | X = x, Z = z)\hat{\Pr}(X = x)$. Now consider the c-component associated with U_2 to impose constraints on $\Pr(U_2)$. Note that U_2

has 16 values, and adopt $q_{2,k} := \Pr(U_2 = k)$. Write k as a binary number with bits $b_0 b_1 b_2 b_3$, and take the mechanism $f_Z(s, x, k)$ as follows: if $sx = 00$, then the mechanism f_Z yields the first bit b_0 of k; if $sx = 01$, then f_Z yields b_1; if $sx = 10$, then f_Z yields b_2; and if $sx = 11$, then f_Z yields b_3. We then have 8 constraints, one per configuration of S, X, Z, using the template

$$\sum_k q_{2,k} [\![f_Z(s, x, k) = z]\!] = g_2(sxz),$$

where $g_2(sxz) := \hat{\Pr}(Z = z | S = s, X = x)$. Then the value of PN, to be minimized, is the bilinear expression: $q_{1,1} h_1 + q_{1,2} h_2$, where

$$
\begin{aligned}
h_1 &:= (1 - \gamma)(q_{2,4} + q_{2,5} + q_{2,6} + q_{2,7}) + \gamma(q_{2,1} + q_{2,5} + q_{2,9} + q_{2,13}), \\
h_2 &:= (1 - \gamma)(q_{2,8} + q_{2,9} + q_{2,10} + q_{2,11}) + \gamma(q_{2,2} + q_{2,6} + q_{2,10} + q_{2,14}),
\end{aligned}
$$

with $\gamma = \Pr(S = 1)$. A similar expression is obtained for PNS. Even though the PNS requires two interventions, it does not require conditioning — for this reason, the reduced counterfactual graph to be processed is almost identical to the graph in Figure 2 (right): the only difference is that we have $X_{X=0}$ instead of X, $Z_{X=0}$ instead of Z, and $Y_{X=0}$ instead of Y. The value of PNS, to be minimized, is the following expression (again subject to the same linear constraints described before): $(q_{1,1} + q_{1,5}) h_1 + (q_{1,2} + q_{1,6}) h_2$.

To illustrate, take the following randomly generated vector, containing values of the input distribution $\hat{\Pr}(X = x, Y = y, Z = z, S = s)$ ordered as if $xyzs$ were a binary number: $[0.30375, 0.0075, 0.03375, 0.3, 0.12, 0.0225, 0.03, 0.2025, 0.0675,$ $0.01, 0.0075, 0.4, 0.01, 0.01125, 0.0025, 0.10125]$. We can use the Gurobi optimization solver,[2] to easily obtain the PNS interval $[0.35, 0.49]$, in less than 10ms, and the PN interval $[0.175, 0.245]$, also in less than 10ms. $\square$

4   Probabilities of Causation in Root Cause Analysis

In the previous section we introduced new techniques to simplify the computation of probabilities of causation. We now investigate how to use those probabilities in a practical challenge, namely, in the search for root causes.

4.1   Root Cause Analysis: A Very Brief Overview

An anomaly occurs when a system exhibits a set of symptoms signaling a potential failure, triggering the need to identify its underlying causes to rectify or mitigate the issue. Finding root causes is, however, a complex task. If a module fails, and its behavior depends on a sequence of previous actions, it may difficult to find the specific root source of the failure.

Root Cause Analysis (RCA) range from manual, diagram-based inspections to advanced automated systems leveraging statistical or causal inference techniques [13]. These approaches are broadly employed in domains like epidemiology, medical diagnostics and, increasingly, system observability, particularly for

---

[2] Gurobi: https://gurobi.com/

microservices and distributed infrastructures [3, 5]. The complexity and opacity of modern systems, due to hidden components or limited monitoring, pose significant challenges for traditional RCA methods that rely on correlations between monitored signals [2, 13]. Unobserved factors (latent confounders) frequently obscure the true causal relationships, complicating root cause identification.

Existing causal RCA methods typically involve constructing causal graphs either from domain expertise or automated causal discovery algorithms, subsequently assigning scores to each node reflecting their contribution to the observed anomaly [3, 4]. Several scoring metrics have been proposed, including Shapley-style attributions derived from soft interventions, as implemented in libraries such as DoWhy. These methods yield a ranked list of candidate root causes based on their causal relevance to the incident [2]. Even though there are approaches that employ causal tools, the literature is very sparse about using probabilities of causation for RCA, particularly under scenarios of limited information.

4.2   A Proposal: Causal Narratives via Probabilities of Causation

Probabilities of causation are effective in establishing to what degree a particular variable can be deemed a cause of an event. Yet, assessing whether a particular variable is a root cause of an incident requires considering distinctions between degrees of causation between variables. We now propose a technique that uses probability of causation to identify what we call *causal narratives*, roughly understood as a causal chain of events that connect a single root cause to an incident observed in sample data containing examples of both normal and abnormal behavior. The root cause is thus the head of such causal chain.

Algorithm 1 operates on probability scores, using a pruned-DFS (Depth-First Search) to rank the causal chains and assessing whether they can recover the injected root-cause narratives from the resulting probability values. A few details about the algorithm:

Search Metric. Let $G = (V, E)$ be a directed acyclic graph (DAG) whose edges encode causal relations between observed variables and let $Y \in V$ be a distinguished *incident node*. For every $X \in V \setminus \{Y\}$ we compute a local *probability-of-causation*-based metric $\mathrm{PC}(X, Y)$ of choice. For instance, $\mathrm{PC}(X, Y)$ may be chosen to refer to the Probability of Necessity (PN).

Search Strategy. We perform a depth-first traversal that walks *upstream* from the incident node. Whenever the absolute change $\Delta_{k+1} = |\mathrm{PC}_{k+1} - \mathrm{PC}_k|$ between two consecutive ancestors exceeds a fixed multiple $\alpha > 1$ of the current median $\mathcal{M}_k = \mathrm{median}\{\Delta_1, \ldots, \Delta_k\}$, we *prune* the branch and declare the present node a plausible root cause. This rule removes spurious upstream nodes whose causal strength drops abruptly while preserving coherent high-PC chains.

Path Significance. For every surviving path $\pi = \langle X_r, \ldots, X_1, Y \rangle$ we define a linear *path-significance score* define as $S(\pi) = w \cdot \mathrm{PC}(X_r, Y) + \sum_{j=1}^{r-1} \mathrm{PC}(X_j, Y)$, where $w \geq 1$ is an adjustable weight that emphasizes the putative root cause X_r. Paths are ranked by $S(\pi)$; the top entry yields both the most likely root cause and its dominant causal chain.

Algorithm 1 Root-Cause Path Ranking

Require: DAG G, PC scores PC$[\cdot]$, target Y, drop factor α, root weight w
1: **function** RANK(G, PC, Y)
2:     out $\leftarrow [\,]$
3:     **procedure** DFS(v, P, Δ)
4:         **if** $parents(v) = \emptyset$ **then**
5:             REC(P) **return**
6:         **end if**
7:         **for** $p \in parents(v)$ **do**
8:             $\delta \leftarrow |\text{PC}[p] - \text{PC}[v]|$
9:             $\beta \leftarrow \alpha \cdot \text{median}(\Delta)$
10:             **if** $|\Delta| \geq 1 \wedge \delta > \beta$ **then**
11:                 REC($P \cup \{p\}$); **continue**
12:             **end if**
13:             DFS(p, $P \cup \{p\}$, $\Delta \cup \{\delta\}$)
14:         **end for**
15:     **end procedure**
16:     **function** REC(P)
17:         $s \leftarrow w \cdot \text{PC}[P_0] + \sum\limits_{x \in P \setminus \{Y\}} \text{PC}[x]$
18:         push (P^{rev}, s) into out
19:     **end function**
20:     DFS(Y, $[Y]$, $[\,]$)
21:     **return** out sorted by $s \downarrow$
22: **end function**

HeavyTraffic

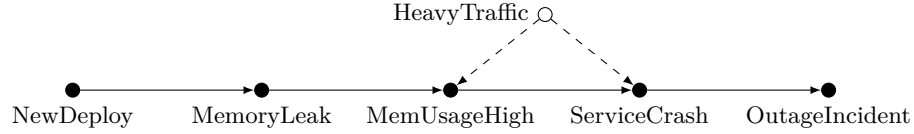NewDeploy   MemoryLeak   MemUsageHigh   ServiceCrash   OutageIncident

Fig. 3. Model 1: Small-scale observability causal model. Dashed arrows indicate latent confounding by a surge in traffic.

4.3   Experiments and Evaluation: RCA for Microservices

We describe experiments that assess the performance of our proposed metrics in the context of RCA for microservices characterized by limited information. Microservice outages typically stem from cascading failures, where an initial fault propagates through dependent components. Effective RCA must therefore accurately trace the causal pathways that lead to such incidents.

We investigate metrics based on PN, PS, PNS, w-PN and w-PS as defined in Section 3. The estimation of probabilities of causation typically yields interval estimates rather than precise point values. To practically use these interval estimates, we apply heuristic strategies to reduce each interval to a single scalar value, such as Minimum (to consider the lower bound as the score, thus prioritizing the estimates of the metrics by how far from zero, of no importance, they are), Maximum (to use the upper bound of the interval as the score, thus prioritizing the estimates of the metrics by their maximum possible importance to the observed effect), Mean and Midpoint.

We designed experiments based on synthetic causal models and datasets to evaluate how effectively these metrics identify root causes under scenarios involving latent confounders and partial observability. Each synthetic model represents a simplified but realistic microservice system, where nodes correspond to binary representations of service-level metrics (such as high latency or error occurrences, determined by specified thresholds). Given space restrictions, we fo-
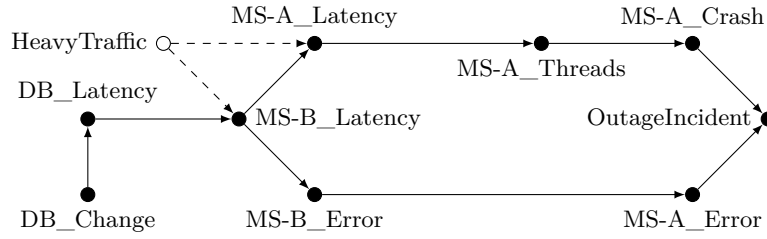
Fig. 4. Model 2: Medium-scale microservice observability causal model. Dashed arrows indicate latent heavy traffic confounding latency between Microservice B and in Microservice A. Target node is the OutageIncident, indicating unavailability.
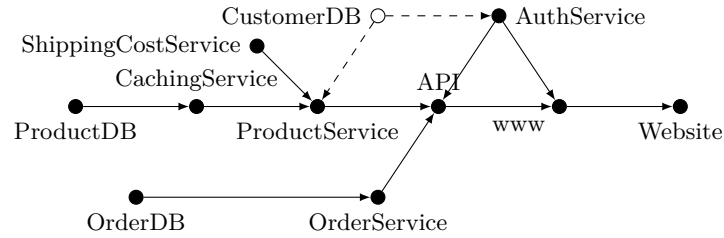

Fig. 5. Model 3: Finding the Root Cause of large latencies in a microservice architecture. The latency of customers databases is not monitored, so it is a latent confounder. Website is the target node for which we want to run a RCA.

cused on models that can illustrate the mechanics of our techniques, and not on models engineering to display computational gains produced by the techniques in the previous sections.

To evaluate our proposed metrics rigorously, we designed three synthetic causal models of varying complexity to simulate typical fault scenarios encountered in microservices architectures, as illustrated in Figures 3-5. Each model's mechanisms and corresponding synthetic datasets explicitly encode known causal narratives, allowing us to objectively assess how well each causation-based metric recovers these true underlying causal relationships.

Our results with experiments in Models 1-3 are summarized in Tables 1-2. For each model-narrative pair with ground truth, we present the identified root cause when using each of the proposed metrics within the Algorithm 1. We can see that metrics derived from *necessity* (PN and w-PN) consistently identifies the correct root cause pathways across various scenarios. In contrast, *sufficiency*-based metrics (PS and w-PS) consistently underperformed, often leading to incorrect conclusions. PNS exhibited inconsistent performance across causal narratives.

To interpret our experimental findings, it is essential to clarify the specific nature of the RCA we are interested in. The umbrella term *Root Cause Analysis* encompasses diverse techniques, ranging from analyses of isolated incidents, anomalies, or outliers occurring at specific points in time, to more generalized evaluations over extended periods, such as structural changes or system degra-

Table 1. Causal Path-Ranking results with *strong* probabilities ($\alpha = 2$; $w = 2$)

Model	Narrative	Ground Truth	PN	PS	PNS
M1	N1	MemLeak	MemLeak	MemLeak	MemLeak
	N2	HeavyTraffic	ServiceCrash	MemUsageHigh	ServiceCrash
	N3	ServiceCrash	ServiceCrash	ServiceCrash	ServiceCrash
	N4	MemUsageHigh	MemUsageHigh	MemUsageHigh	MemUsageHigh
M2	N1	DB_Latency	DB_Latency	DB_Latency	DB_Latency
	N2	MS-A_Threads	MS-A_Threads	MS-A_Threads	MS-A_Threads
	N3	MS-B_Error	MS-B_Error	MS-B_Error	MS-B_Error
	N4	MS-A_Latency	MS-A_Latency	MS-A_Threads	MS-A_Latency
M3	N1	CachingService	CachingService	ProductService	CachingService
	N2	OrderDB	OrderDB	API	OrderService

Table 2. Causal Path Ranking results with *weak* measurements ($\alpha = 2$; $w = 2$)

Model	Narrative	Ground Truth	weak PS	weak PN
M1	N1	MemLeak	NewDeploy	MemLeak
	N2	HeavyTraffic	NewDeploy	NewDeploy
	N3	ServiceCrash	NewDeploy	ServiceCrash
	N4	MemUsageHigh	NewDeploy	MemUsageHigh
M2	N1	DB_Latency	DB_Change	DB_Change
	N2	MS-A_Threads	MS-A_Error	MS-A_Threads
	N3	MS-B_Error	MS-A_Crash	MS-B_Error
	N4	MS-A_Latency	MS-A_Threads	MS-B_Latency
M3	N1	CachingService	ProdutctService	CachingService
	N2	OrderDB	API	OrderService

dation. Even though counterfactual inference emphasizes individual-level *what-if* questions, our proposed method diverges slightly by aggregating binary observations over specified intervals, aiming to determine the general root cause behind persistent inconsistencies observed in a target variable.

Taking into account this qualitative assessment, we interpret the effectiveness of necessity metrics in RCA as inherently linked to their alignment with temporal causality, distinguishing between preventive and remedial analyzes. If the goal is to identify critical components that impact system incidents, sufficiency measures may initially appear more informative. However, to address underlying conditions triggering successive failures, necessity measures provide a more robust analytical foundation.

Accordingly, the consistent capability of necessity-based metrics to accurately recover the true causal paths underscores their adequacy for automated RCA in intelligent autonomous systems. From a practical perspective, the choice between PN and w-PN must consider computational and data requirements. While PN provides stronger discriminatory power, its inference involves substantially

higher computational costs compared to the intervention-based w-PN. Thus, the decision to deploy either metric within intelligent RCA systems must carefully weigh their respective performance benefits against these operational constraints.

5   Conclusion

To recap, we have presented contributions both on algorithmic and methodological aspects of causal reasoning.

On algorithms, we presented simplifications to counterfactual graphs that lead to multilinear programs of reduced degree, by exploiting the information in input distributions. We have also presented a few additional simplifications to the computation of probabilities of causation PN, PS and PNS, as these counterfactual quantities are particularly important in practice.

On methodological contributions for root cause analysis (RCA), we proposed a novel approach leveraging probabilities of causation as metrics for identifying and ranking causal paths within scenarios of limited observability, such as those in microservices observability contexts. By formulating RCA as the task of quantifying causal contributions of variables along potential causal paths, we developed and evaluated a ranking algorithm based on the concept of *path significance*, where significance scores were systematically derived from probabilities of causation (PN, weak-PN, PS, weak-PS, and PNS). We also analyzed realistic scenarios with cascading failures and latent confounders, so as to understand the behavior of probabilities of causation in practice. Necessity-based measures PN and w-PN emerged as winners in identifying root causes and causal paths leading to observed incidents.

Looking at potential avenues for future research, we suggest extending our methods to recently formalized definitions of probabilities of causation for discrete and continuous variables. Such advances would avoid the need for data binarization and potentially enhance RCA accuracy and applicability.

Acknowledgements

Bibliography

[1] Balke, A., Pearl, J.: Counterfactual probabilities: Computational methods, bounds and applications. In: Conf. on Uncertainty in Artificial Intelligence. pp. 46–54 (1994)

[2] Budhathoki, K., Minorics, L., Blöbaum, P., Janzing, D.: Causal structure-based root cause analysis of outliers. In: Int. Conf. on Machine Learning (ICML) (2022)

[3] Ikram, A., Chakraborty, S., Mitra, S., Saini, S.K., Bagchi, S., Kocaoglu, M.: Root cause analysis of failures in microservices through causal discovery. In: Advances in Neural Information Processing Systems (NeurIPS) (2022)

[4] Li, M., Li, Z., Yin, K., Nie, X., Zhang, W., Sui, K., Pei, D.: Causal inference-based root cause analysis for online service systems with intervention recognition. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). pp. 1575–1585 (2022)

[5] Mokhtari, A., Pang, H., Farakos, S.S., McKenna, C., Crowley, C., Cranford, V., Bowen, A., Phillips, S., Madad, A., Obenhuber, D., Doren, J.M.V.: Leveraging risk assessment for foodborne outbreak investigations: The quantitative risk assessment-epidemic curve prediction model. Risk Analysis **43**(2), 324–338 (2023)

[6] Oliveira, E.E., Miguéis, V.L., Borges, J.L.: Understanding overlap in automatic root cause analysis in manufacturing using causal inference. IEEE Access **10**, 191–201 (2022). https://doi.org/10.1109/ACCESS.2021.3139199

[7] Pearl, J.: Causality. Cambridge University Press (2009)

[8] Pearl, J., Glymour, M., Jewell, N.P.: Causal inference in statistics: a primer. Wiley, Chichester, West Sussex (2016)

[9] Richens, J.G., Lee, C.M., Johri, S.: Improving the accuracy of medical diagnosis with causal machine learning. Nature Comm. **11**, 3923 (2020)

[10] Shpitser, I., Pearl, J.: What counterfactuals can be tested. In: Conference on Uncertainty in Artificial Intelligence. p. 352359. AUAI Press (2007)

[11] Shridharan, M., Iyengar, G.: Causal bounds in quasi-Markovian graphs. In: Proceedings of the 40th International Conference on Machine Learning. vol. 202, pp. 31675–31692. PMLR (2023)

[12] Tian, J.: Studies in Causal Reasoning and Learning. Ph.D. thesis, UCLA (2002)

[13] Wang, T., Qi, G.: A comprehensive survey on root cause analysis in (micro) services: Methodologies, challenges, and trends. arXiv preprint arXiv:2408.00803 (2024)

[14] Zaffalon, M., Antonucci, A., Cabañas, R., Huber, D., Azzimonti, D.: Efficient computation of counterfactual bounds. International Journal of Approximate Reasoning pp. 1–24 (2024)

[15] Zhang, J., Tian, J., Bareinboim, E.: Partial counterfactual identification from observational and experimental data. In: Int. Conf. on Machine Learning. vol. 162, pp. 26548–26558. PMLR (2022)