

# ProRCA: A Causal Python Package for Actionable Root Cause Analysis in Real-world Business Scenarios

Ahmed Dawoud\*

Shravan Talupula†

## Abstract

Root Cause Analysis (RCA) is becoming ever more critical as modern systems grow in complexity, volume of data, and interdependencies. While traditional RCA methods frequently rely on correlation-based or rule-based techniques, these approaches can prove inadequate in highly dynamic, multi-layered environments. In this paper, we present a pathway-tracing package built on the *DoWhy* causal inference library. Our method integrates *conditional anomaly scoring*, *noise-based attribution*, and *depth-first path exploration* to reveal multi-hop causal chains. By systematically tracing entire causal pathways from an observed anomaly back to the initial triggers, our approach provides a comprehensive, end-to-end RCA solution. Experimental evaluations with synthetic anomaly injections demonstrate the package’s ability to accurately isolate triggers and rank root causes by their overall significance.

## 1. Introduction

effect back to its initial triggers.

Modern operational landscapes, spanning domains such as retail, healthcare, finance, and software systems, are increasingly characterized by complex interdependencies and massive data streams. In such settings, anomalies rarely arise from a single isolated factor; rather, they emerge as the cumulative effect of multi-hop causal chains. Existing RCA methods typically focus on detecting outliers or isolating single nodes based on correlation or localized attribution. However, these approaches do not provide a *complete* explanation of why a failure occurred. In other words they do not systematically trace *all* possible causal pathways from an observed

The primary motivation for our work is to address this limitation by developing a package that systematically reconstructs the full causal pathway from an observed anomaly back to its root cause. By leveraging the strengths of the *DoWhy* causal inference library, our method extends existing techniques to not only identify individual anomalous nodes but also trace entire multi-hop causal chains. This end-to-end approach enables practitioners to intervene precisely at the earliest disruption points, thereby reducing the risk of recurring failures and improving overall system reliability.

---

\*Head of Decision Science, ProfitOps; [adawoud@profitops.ai](mailto:adawoud@profitops.ai)

†Founder and CEO, ProfitOps; [stalupula@profitops.ai](mailto:stalupula@profitops.ai)

## 2. Literature Review

Root Cause Analysis (RCA) is crucial for maintaining the reliability and performance of complex systems, spanning diverse domains [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. However, existing RCA methodologies exhibit significant variations in their approach to causality, ranging from purely correlational techniques to those grounded in formal causal inference. This review categorizes and critically assesses these approaches, highlighting their strengths, limitations, and, crucially, whether they address the problem of tracing **all** root cause pathways. We identify three primary methodological groupings.

### 2.1 Correlation-Based Approaches

Methods in this category primarily rely on statistical correlations between metrics, often combined with graph structures representing system topology or communication patterns, but they do **not** explicitly model causal mechanisms or employ formal causal inference techniques (e.g., do-calculus, interventions). For instance, **MicroRCA** [12] and **MicroHECL** [1] fall squarely within this category by leveraging service call graphs and correlating performance symptoms with resource utilization. While these approaches are practical and scalable, they are susceptible to spurious correlations and may miss root causes that do not manifest in obvious resource spikes. Similarly, **ExplainIt!** [9] uses attributed graphs and correlations, making it weakly causal at best, as it explores a causality graph but the core methodology remains correlational. In another example, **Chen et al. (2023)** [8] explores the use of Large Language Models (LLMs) for root cause analysis; however,

this approach is weakly causal because it primarily depends on correlations and prompt engineering, thus being vulnerable to spurious relationships and potentially missing subtle causal pathways. Likewise, **Wang et al. (2023) - REASON** [2] introduces interdependent causal networks and uses a hierarchical GNN. Despite the term "causal networks," the core methodology is grounded in learning correlations and propagating information through the network, and although the GNN learns representations, it does not explicitly model interventions or counterfactuals in a formally causal manner.

These methods are often efficient and applicable to large-scale systems. However, their reliance on correlation makes them vulnerable to confounding and limits their ability to provide robust, causally sound explanations. They do **not** aim to trace all root cause pathways.

### 2.2 Hybrid Approaches: Causal Discovery with Limitations

Methods in this group incorporate elements of causal discovery but have significant limitations that prevent them from being considered fully causal. For example, **PORCA** [13] uses "magnified structural causal models" and score-based causal discovery, which is a step toward genuine causal inference. Nevertheless, its effectiveness depends heavily on the novel score function and scheduling process, and the extent to which these truly capture causal relationships requires further scrutiny. Meanwhile, **Wang et al. (2023) - Incremental** [3] employs causal discovery with trigger points; although this is indeed a causal approach, its scalability is constrained by the computational cost of causal discovery, and it does not aim for complete pathway tracing. Ad-

ditionally, **Okati et al. (2024)** [5] explores a traversal-based approach in combination with a causal graph model and Shapley values, making it weakly causal.

These methods represent a move toward causal reasoning, but they often rely on simplifying assumptions (e.g., about the form of causal relationships) or have limitations in scalability or completeness. They typically aim to **identify** or **localize** root causes, but not to trace all pathways.

### 2.3 Genuinely Causal Approaches

The final group includes methods that are firmly grounded in formal causal inference frameworks, such as Structural Causal Models (SCMs) and do-calculus. **Budhathoki et al. (2021)** [6] explicitly use SCMs and Shapley values to quantify the "Intrinsic Causal Contribution" (ICC) of changes in causal mechanisms, offering a theoretically sound approach but requiring the causal graph (DAG) to be known, which is a strong assumption in practice. **Blöbaum et al. (2022)** [7] (DoWhy-GCM) presents a software package for causal inference, supporting various causal queries based on SCMs; however, it does not prescribe a specific method for root cause *pathway* tracing. Additionally, **Strobl and Lasko (2022, 2024)** [4, 10] introduce the heteroscedastic noise model (HNM) and Generalized Root Causal Inference (GRCI), which is genuinely causal but focuses on identifying patient-specific root causes in a medical context and does not address the *all pathways* problem. Finally, **Han et al. (2023)** [11] proposes RootCLAM, a framework that uses a causal graph autoencoder.

These methods offer the strongest theoretical guarantees regarding causality. Crucially, none

of these methods explicitly address the problem of tracing all possible root cause pathways.

### 2.4 Genuine Causal Pathway Tracking Approaches

The PyRCA package is the only package we found that tracks complete causal pathways. It incorporates a spectrum of methods, some firmly grounded in causal inference theory, while others rely more on statistical correlations. For example, techniques like epsilon-Diagnosis and random walk are primarily driven by changes in metric values and graph-based heuristics, respectively. These approaches flag anomalies based on statistically significant deviations or network connectivity but do not explicitly model interventions or counterfactuals. As a result, their causality claims are limited; they tend to capture correlations rather than true causation.

In contrast, methods such as the Bayesian inference approach and the hypothesis-testing (CIRCA) algorithm explicitly leverage the known causal structure to assess how interventions propagate through the system. These methods build and utilize a structural causal model that mirrors the true data-generating process, evaluating whether observed changes in metrics can be directly attributed to a disruption at a specific node. When the underlying assumptions hold—such as the accuracy of the DAG, the absence of hidden confounders, and appropriate model specifications—these methods provide a more rigorous, truly causal analysis of root causes.

However, according to the package’s GitHub repository, all these causal methods from PyRCA have proven to be inaccurate—except for hypothesis-testing (HT), which performed

perfectly on the simulated dataset. Yet, HT is limited to continuous variables and uses only linear regression. In contrast, DoWhy employs nine different models for each node and selects the best one, allowing for robust fitting of causal mechanisms and accommodating varying levels of complexity in modeling relationships across different data types. Theoretically, DoWhy offers a more robust approach; therefore, extending it should enable us to achieve better results.

### 3. Overall Gap and Positioning of the Proposed Work

The Literature review revealed a significant gap in the existing literature. While many methods exist for RCA, most rely on correlations or make simplifying assumptions that limit their causal validity. Even the genuinely causal approaches do not focus on the exhaustive tracing of all possible root cause pathways, which is the core contribution of the proposed work. By addressing this gap, the proposed technique offers a novel and potentially more robust approach to RCA, particularly in complex systems where multiple interacting factors may contribute to a failure.

We argue that complete pathway tracing is essential for robust RCA. Unlike methods that pinpoint a single failure point, our approach seeks to reveal the full **sequence** of causal events, providing a comprehensive understanding of the failure’s etiology. This holistic view uncovers hidden interactions and dependencies, including feedback loops, that might be missed by localized analyses.

Furthermore, a complete pathway reveals multiple potential intervention points, enabling earlier and potentially less disruptive mitigation strategies. By considering all pathways,

our analysis is less susceptible to noise and spurious correlations, leading to a more reliable and, crucially, more **explainable** diagnosis. In essence, while identifying a root cause offers a valuable starting point, tracing **all pathways** provides the complete, robust, and actionable understanding necessary for effective failure prevention in complex systems, clearly distinguishing our approach.

## 4. Methodology

Our proposed package for root cause pathway analysis leverages core functionalities from the DoWhy Package—such as Structural Causal Model (SCM) construction, noise attribution, and conditional anomaly scoring—and extends them with novel contributions designed to extract, trace, and rank multi-hop causal pathways. In the section, we first briefly describe the inherited components and then focus on detailing our original contributions.

### 4.1 Inherited Components from DoWhy

The baseline components used in our package are as follows:

- **SCM Construction:** We build a directed acyclic graph (DAG)  $G = (V, E)$ , where  $V$  is the set of variables (nodes) and  $E$  is the set of causal edges. Each node  $v \in V$  is assigned a causal mechanism  $M(v)$  that maps parent inputs to its output. These steps—including automatic node inference, mechanism assignment, and model fitting—are implemented using DoWhy’s standard routines.
- **Noise Attribution:** For each node

$v$ , a noise contribution vector  $C(v) \in \mathbb{R}^{n_v}$  is computed using DoWhy’s `attribute_anomalies` function with a `MedianCDFQuantileScorer`. This vector quantifies the deviation of observed values from their expected distribution.

- **Conditional Anomaly Scoring:** For nodes with parents, conditional anomaly scores  $S(v)$  are derived by comparing the observed value  $x_v$  against the prediction of the mechanism  $M(v)$  given the parent values  $\mathbf{x}_{P(v)}$ . These scores are computed using DoWhy’s `anomaly_conditional_anomaly_scores` function.

## 4.2 Our Contributions

Our work extends the DoWhy package with the following novel components:

### 4.2.1 Combined Node Score

We introduce a *combined node score* that fuses the structural anomaly score and the noise contribution into a single metric for each node  $v$ . Specifically, let:

- (a) The average structural anomaly score over the anomaly subset  $S_A(v)$  be

$$\bar{S}(v) = \frac{1}{|S_A(v)|} \sum_{s \in S_A(v)} s. \quad (1)$$

- (b) The maximum absolute noise contribution be

$$\tilde{C}(v) = \max_{i=1, \dots, n_v} \{|C(v)_i|\}. \quad (2)$$

Our combined score is then defined as a weighted sum:

$$\text{CombinedScore}(v) = \alpha \bar{S}(v) + (1 - \alpha) \tilde{C}(v), \quad (3)$$

where  $\alpha$  is a weight parameter (set to  $\alpha = 0.7$  in our experiments). Equation (3) represents our primary novel contribution, effectively integrating both model-driven and empirical anomaly information.

### 4.2.2 Root-Cause Path Discovery

We propose a depth-first search (DFS) algorithm to trace multi-hop causal pathways from a target node (e.g., `PROFIT_MARGIN`) back to potential root causes. Let the target node be denoted by  $v_0$  and consider a path:

$$P = \{v_0, v_1, \dots, v_k\},$$

where each  $v_{i+1}$  is a parent of  $v_i$ . Our algorithm extends a path based on the following threshold criteria:

- (i) For each intermediate node  $v_i$  ( $i = 1, \dots, k - 1$ ), require:

$$\text{CombinedScore}(v_i) \geq \beta\theta, \quad (4)$$

where  $\theta$  is a predefined threshold (e.g.,  $\theta = 0.8$ ) and  $\beta$  is set to 0.7.

- (ii) For the terminal (root) node  $v_k$ , require:

$$\text{CombinedScore}(v_k) \geq \theta. \quad (5)$$

Thus, the set of accepted paths is defined as:

$$\mathcal{P} = \left\{ P = \{v_0, v_1, \dots, v_k\} \mid \begin{array}{l} \forall i = 1, \dots, k - 1 : \text{CombinedScore}(v_i) \geq \beta\theta, \\ \text{and } \text{CombinedScore}(v_k) \geq \theta \end{array} \right\}. \quad (6)$$

This DFS-based path discovery is our original extension, enabling the exploration of multi-hop causal chains based on a composite importance measure.

#### 4.2.3 Path Significance Evaluation

To rank the candidate causal pathways, we introduce a novel metric that evaluates each path based on a combination of the root node’s noise anomaly and the causal consistency along the path.

**(a) Noise Component:** Let the average noise contribution at the terminal node  $v_k$  be

$$\mu_{v_k} = \frac{1}{n_{v_k}} \sum_{i=1}^{n_{v_k}} C(v_k)_i. \quad (7)$$

**(b) Causal Consistency Component:** For each adjacent pair  $(v_i, v_{i+1})$  along the path, compute the absolute Pearson correlation coefficient between their noise vectors:

$$\rho_{i,i+1} = \left| \text{corr}(C(v_i), C(v_{i+1})) \right|. \quad (8)$$

Then, the overall consistency of the path is given by:

$$\text{Consistency}(P) = \frac{1}{k} \sum_{i=0}^{k-1} \rho_{i,i+1}. \quad (9)$$

**(c) Overall Path Significance:** We compute the significance of the path as a weighted sum of the noise and consistency components:

$$\text{Significance}(P) = \gamma \mu_{v_k} + (1-\gamma) \text{Consistency}(P), \quad (10)$$

with  $\gamma$  set to 0.7. Equations (7)–(10) together constitute our novel metric for ranking multi-hop causal pathways.

### 4.3 Summary of the Methodology

ProRCA’s methodology integrates core functionalities of the DoWhy package with novel contributions for complete causal pathway analysis. First, inheriting from DoWhy, we construct a Structural Causal Model (SCM), building a directed acyclic graph (DAG)  $G = (V, E)$  and assigning causal mechanisms  $M(v)$  to each node  $v$ . We then compute, for each node, a noise contribution vector  $C(v)$  and conditional anomaly scores  $S(v)$ , also leveraging DoWhy’s established procedures.

Building upon this foundation, ProRCA introduces a combined node score, fusing structural anomaly and noise contribution:  $\text{CombinedScore}(v) = 0.7 \bar{S}(v) + 0.3 \tilde{C}(v)$ . This single metric integrates both model-based and empirical anomaly information. To trace root cause pathways, we perform a depth-first search (DFS) starting from the target node  $v_0$ , identifying paths  $P = \{v_0, v_1, \dots, v_k\}$  that meet specific threshold criteria for both intermediate and terminal nodes (Equations (4) and (5)). This yields a set of accepted paths,  $\mathcal{P}$  (Equation (6)).

Finally, we evaluate the significance of each discovered path  $P$  by combining the terminal node’s average noise contribution ( $\mu_{v_k}$ ) with a measure of causal consistency along the path ( $\text{Consistency}(P)$ ):  $\text{Significance}(P) = 0.7 \mu_{v_k} + 0.3 \text{Consistency}(P)$ . This novel metric prioritizes paths with both significant root cause noise and strong causal coherence. In essence, ProRCA extends DoWhy with a combined scoring system, a DFS-based pathway discovery algorithm, and a path significance evaluation, providing a comprehensive package for multi-hop root cause analysis.

## 5. Complete ProRCA Workflow

This section details the process of anomaly injection, the implementation pipeline, and the evaluation metrics. The complete code, including data generation scripts and analysis notebooks, is publicly available in the *notebook* folder of our GitHub repository.<sup>1</sup>

### 5.1 Synthetic Data Generation

The synthetic dataset is produced by a Python-based simulation that approximates real-world retail transactions in an intricate, non-linear manner.

```
# Example usage
start_date = '2023-01-01'
end_date = '2023-12-30'
df = generate_fashion_data_with_brand(
    start_date, end_date)
```

Listing 1: Example Data Generation Code 2023

At its core, the code assigns product attributes—such as category, brand, and hierarchical merchandise structure—to each generated record, reflecting a diverse range of items (e.g., `Apparel.Men.Shirts.Casual`, `Footwear.Women.Boots.Winter`). Transaction-level features exhibit pronounced non-linearity through lognormal draws for `PRICEEACH`, Poisson sampling for `QUANTITYORDERED`, and multiplicative modifiers for costs and discounts. Randomization of sales channels, geographic territories, and promotion codes further amplifies the complexity, while weekly seasonality (weekends vs. weekdays) influences base demand.

Key financial metrics, including `SALES`, `DISCOUNT`, `NET_SALES`, `FULFILLMENT_COST`, `COST_OF_GOODS_SOLD`, `MARKETING_COST`, and

`RETURN_COST`, are computed by non-linear functions that factor in brand/category parameters, loyalty tiers, and location-specific multipliers.

This layered approach ensures that transactional outcomes—such as `PROFIT` and `PROFIT_MARGIN`—arise from realistic, interdependent processes. Specifically, the generation script recasts each row’s profit by combining net sales, shipping revenue, and cost components in a manner that imitates real-world operations, introducing the potential for complex interactions and non-trivial outliers.

### 5.2 Controlled Anomaly Injection

To rigorously assess ProRCA’s root cause detection, the code injects anomalies at specified dates and scopes—either delimited by `SALES_CHANNEL` or product hierarchy—using non-linear transformations. By recalculating every downstream metric within each affected transaction, the injection script ensures robust, interlinked disruptions that realistically resemble operational mishaps. Such non-linear perturbations impose additional challenges for detection, necessitating causal inference techniques capable of disentangling multi-hop dependencies.

```
anomaly_schedule = {
    '2023-01-10': ('Discount', 0.6, 'Apparel'),
    '2023-06-10': ('COGs', -0.8, 'Footwear'),
    '2023-09-10': ('Fulfill', -2, 'Beauty'),
    '2023-12-10': ('Return', 10, 'Accessories')
}

df_anomalous = inject_anomalies_by_date(df,
    anomaly_schedule)
```

Listing 2: Anomaly Injection Schedule

Figure 1 shows four types of anomalies are introduced on different dates and product lines,

<sup>1</sup><https://github.com/Economist-Ahmed-Dawoud/ProRca>

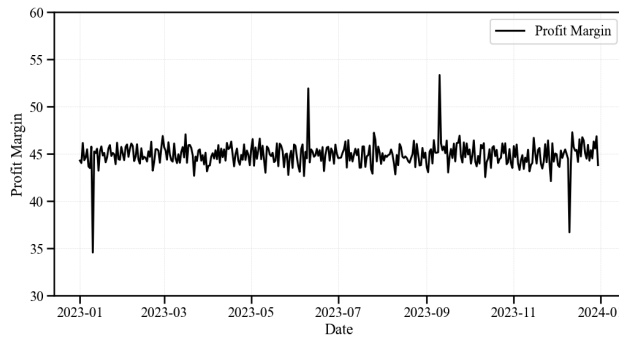
each with a unique root cause, severity, and non-linear effect on key variables:

**ExcessiveDiscount (2023-01-10, Apparel):** Simulates a pricing error by boosting the discount to an unusually high proportion of SALES. This anomaly drastically lowers NET\_SALES and propagates through downstream metrics, including RETURN\_COST and PROFIT\_MARGIN.

**COGSOverstatement (2023-06-10, Footwear):** Models a supplier issue by artificially increasing UNIT\_COST, thereby raising COST\_OF\_GOODS\_SOLD. The resulting effect lowers PROFIT and, in turn, reduces PROFIT\_MARGIN.

**FulfillmentSpike (2023-09-10, Beauty):** Represents a logistics problem that inflates FULFILLMENT\_COST in a non-linear manner. This spike ripples through PROFIT calculations, ultimately diminishing PROFIT\_MARGIN.

**ReturnSurge (2023-12-10, Accessories):** Emulates a quality control issue causing an abrupt jump in RETURN\_COST. This leads to marked reductions in PROFIT and PROFIT\_MARGIN, with potential downstream effects on other revenue- and cost-related features.



**Figure 1:** Profit Margin Analysis Over Time.

### 5.3 Anomaly Detection

In this step, the `AnomalyDetector` identifies dates where the PROFIT\_MARGIN metric deviates from expected patterns. This module flags anomalous dates; it does not attempt to explain why the anomalies occurred. The `AnomalyDetector` is implemented as a Python class, leveraging the 'InterQuartileRangeAD' detector from the 'adtk' library. This detector is a standard, non-parametric method for identifying outliers in time series data. It flags values that fall outside a dynamically calculated range based on the interquartile range (IQR) of the data.

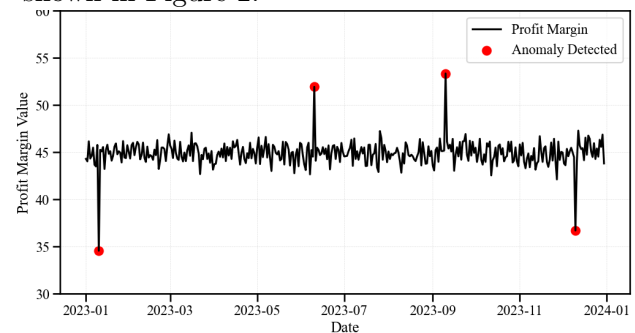
The code for the `AnomalyDetector` is as follows:

```
detector = AnomalyDetector(df_agg, date_col="ORDERDATE", value_col="PROFIT_MARGIN")

anomalies = detector.detect()
anomaly_dates = detector.get_anomaly_dates()
detector.visualize(figsize=(12,6))
```

Listing 3: Anomaly Detection Code

The `AnomalyDetector` class provides three key methods: `detect()`, `get_anomaly_dates()`, and `visualize()`. The `detect()` identifies anomalies in the specified `value_col`. The `get_anomaly_dates()` method extracts the dates where anomalies were detected. The `visualize()` method plots the time series data, highlighting the detected anomalies in red, as shown in Figure 2.





**Figure 2:** Detected Anomaly Dates in Red.

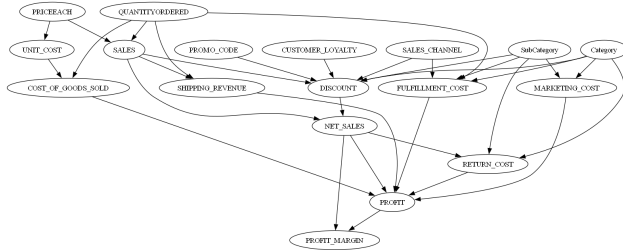
Table 1 exemplifies the output of the anomaly detection, indicating that June 10, July 15, and October 30 are flagged as suspicious dates.

Date	Suspicious?
2023-01-10	Yes
2023-06-10	Yes
2023-09-10	Yes
2023-12-10	Yes

**Table 1:** Initial Anomaly Detection

## 5.4 Building the DAG and Fitting the SCM

The `CausalRootCauseAnalyzer` operates on a Structural Causal Model (SCM) comprising a Directed Acyclic Graph (DAG) and corresponding causal mechanisms. In our controlled experiments, the DAG and functional forms are known, enabling direct construction of the SCM. Figure 3 shows the causal structure among the key variables in the data frame. These relationships reflect typical business logic in retail.



**Figure 3:** Directed Acyclic Graph (DAG)

To construct the SCM, the `ScmBuilder` class takes a list of directed edges (e.g., (PRICE\_EACH, SALES)) and assigns appropriate causal mechanisms to each node based on the data as illustrated in listing 2.

```

scm = ScmBuilder(edges=causal_edges)
scm.build(df_agg)
analyzer = CausalRootCauseAnalyzer(scm,
    min_score_threshold=0.3)
analysis_results = analyzer.analyze(
    df_agg, anomaly_dates)

```

Listing 4: SCM Building and Pathway Tracking

Once the SCM is built, `CausalRootCauseAnalyzer` is initialized for anomaly analysis. Here, `min_score_threshold` governs the sensitivity of the root-cause pathways, determining which nodes qualify as anomalous. The entire pipeline, from DAG specification to root cause analysis, is thus encapsulated within the `CausalRootCauseAnalyzer` class.

## 5.5 Results: Root Cause Pathways

The ProRCA package was evaluated on the detected anomaly dates from the step 5.2 (2023-01-10, 2023-06-10, 2023-9-10, and 2023-12-10). For each date, the system extracted a complete causal pathway from the outcome node (`PROFIT_MARGIN`) back to a candidate root cause. The package computes two key metrics for each node: a *Combined Score* (a weighted sum of the structural anomaly score and noise contribution) and the *Noise Contribution* (the maximum absolute deviation from expected behavior). In addition, the overall *Path Causal Significance* (PCS) is determined as a weighted combination of the terminal node's average noise and the causal consistency along the path.

The `min_score_threshold` parameter sets the minimum acceptable combined score for a node to be included in the DFS-based causal pathway. A higher threshold (e.g., 0.8) admits only nodes with very strong anomaly signals, while a lower threshold (e.g., 0.3) increases sen-

sitivity by including nodes with lower scores. In our experiments, both settings yielded robust complete pathways; however, the high threshold ensured that only the most significant nodes were retained. The reported results below are based on a conservative threshold of 0.8

### 5.5.1 Accuracy of Pathway Identification

Table 2 summarizes the causal pathways identified by ProRCA for the three anomaly dates. For each anomaly date (first column), the corresponding complete causal path (second column) is reported, along with the overall Path Causal Significance (PCS) in the third column.

Notably, the pathway identified for 2023-01-10 is `PROFIT_MARGIN`  $\rightarrow$  `NET_SALES`  $\rightarrow$  `DISCOUNT`, which directly corresponds to the originally injected *ExcessiveDiscount* anomaly. Similarly, the pathway for 2023-06-10, namely `PROFIT_MARGIN`  $\rightarrow$  `PROFIT`  $\rightarrow$  `COST_OF_GOODS_SOLD`  $\rightarrow$  `UNIT_COST`, aligns with the *COGSOVerstatement* anomaly that was introduced. The third pathway for 2023-9-10, `PROFIT_MARGIN`  $\rightarrow$  `PROFIT`  $\rightarrow$  `FULFILLMENT_COST`, reflects the impact of the injected *FulfillmentSpike* anomaly.

Finally, the third pathway for 2023-12-10, `PROFIT_MARGIN`  $\rightarrow$  `PROFIT`  $\rightarrow$  `RETURN_COST`, reflects the impact of the injected *ReturnSurge* anomaly. Thus, ProRCA successfully and accurately identified the correct multi-hop causal pathways, confirming its effectiveness in pinpointing the root causes corresponding to the pre-defined anomalies.

Anomaly Date	Identified Causal Path
2023-01-10	<code>PROFIT_MARGIN</code> $\rightarrow$ <code>NET_SALES</code> $\rightarrow$ <code>DISCOUNT</code>
2023-06-10	<code>PROFIT_MARGIN</code> $\rightarrow$ <code>PROFIT</code> $\rightarrow$ <code>COST_OF_GOODS_SOLD</code> $\rightarrow$ <code>UNIT_COST</code>
2023-9-10	<code>PROFIT_MARGIN</code> $\rightarrow$ <code>PROFIT</code> $\rightarrow$ <code>FULFILLMENT_COST</code>
2023-12-10	<code>PROFIT_MARGIN</code> $\rightarrow$ <code>PROFIT</code> $\rightarrow$ <code>RETURN_COST</code>

**Table 2:** Summary of Identified Causal Pathways

### 5.5.2 Detailed Node-Level Statistics

The pathway extracted for **2023-01-10** consists of three nodes: `PROFIT_MARGIN`, `NET_SALES`, and `DISCOUNT`. All nodes exhibit high Combined Scores (0.8526, 0.8267, and 0.8479, respectively), indicating that each node strongly contributed to the anomaly signal. Notably, the Noise Contribution increases along the causal chain—from 0.0379 at `PROFIT_MARGIN` to 0.0998 at `NET_SALES` and reaching 0.1241 at `DISCOUNT`. This trend suggests that while the outcome node already exhibits anomalous behavior, the anomaly becomes more pronounced upstream, with `DISCOUNT` displaying the highest noise level. Such a result is consistent with the pre-injected *ExcessiveDiscount* anomaly, thereby confirming that ProRCA correctly identified the underlying causal pathway.

Node	Combined Score	Noise Contribution
<code>PROFIT_MARGIN</code>	0.8526	0.0379
<code>NET_SALES</code>	0.8267	0.0998
<code>DISCOUNT</code>	0.8479	0.1241

**Table 3:** Node Statistics for 2023-06-10.

The pathway extracted for **2023-06-10** comprises four nodes: `PROFIT_MARGIN`, `PROFIT`, `COST_OF_GOODS_SOLD`, and `UNIT_COST`.

All nodes exhibit high Combined Scores (ranging from 0.8381 to 0.9086), which confirms that each component strongly contributes to the anomaly signal. Notably, the Noise Contribution progressively increases along the causal chain—from 0.0382 at `PROFIT_MARGIN` and 0.0409 at `PROFIT` to 0.1247 at `COST_OF_GOODS_SOLD`, reaching a peak of 0.2496 at `UNIT_COST`. This trend indicates that the anomaly becomes more pronounced upstream, with `UNIT_COST` displaying the most significant deviation. Such a pattern is consistent with the pre-injected *COGSOverstatement* anomaly, thereby confirming that ProRCA has accurately identified the correct causal pathway.

Node	Combined Score	Noise Contribution
PROFIT_MARGIN	0.8655	0.0382
PROFIT	0.9086	0.0409
COGS	0.8711	0.1247
UNIT_COST	0.8381	0.2496

**Table 4:** Node Statistics for 2023-07-15.

**For 2023-09-10, the pathway comprises three nodes** , with the terminal node `FULFILLMENT_COST` displaying the highest Combined Score (0.8306) and Noise Contribution (0.2496). This indicates that the anomaly is primarily driven by an abnormal deviation in `FULFILLMENT_COST`, consistent with the injected *FulfillmentSpike* anomaly.

Node	Combined Score	Noise Contribution
PROFIT_MARGIN	0.8166	0.0384
PROFIT	0.8001	0.0326
FULFILLMENT_COST	0.8306	0.2496

**Table 5:** Node Statistics for 2023-09-10.

**For 2023-12-10, the pathway comprises three nodes** , with the terminal node `RETURN_COST` displaying the highest Combined Score (0.8654) and Noise Contribution (0.0831). This succinct trend indicates that the anomaly is primarily driven by an abnormal deviation in `RETURN_COST`, consistent with the injected *ReturnSurge* anomaly.

Node	Combined Score	Noise Contribution
PROFIT_MARGIN	0.8098	0.0382
PROFIT	0.8318	0.0390
RETURN_COST	0.8654	0.0831

**Table 6:** Node Statistics for 2023-10-30.

### 5.5.3 Discussion of min\_score\_threshold

When using a `min_score_threshold` of 0.3, the analyzer increased its sensitivity by admitting nodes with lower Combined Scores, thereby identifying a larger set of candidate causal pathways.

For example, on 2023-01-10, five potential pathways were detected, yet the most significant pathway ( $PCS = 0.0869$ ) correctly linked `PROFIT_MARGIN`  $\rightarrow$  `NET_SALES`  $\rightarrow$  `DISCOUNT`, in agreement with the pre-injected *ExcessiveDiscount* anomaly. Similarly, on 2023-06-10 the pathway with the highest PCS (0.1747) traversed from `PROFIT_MARGIN` through `PROFIT`

to `COST_OF_GOODS_SOLD` and `UNIT_COST`, consistent with the expected *COGSOverstatement* anomaly.

For 2023-09-10, among four candidate pathways, the dominant one ( $PCS = 0.17$ ) accurately implicated `FULFILLMENT_COST`, matching the injected *FulfillmentSpike* anomaly. For 2023-12-10, among four candidate pathways, the dominant one ( $PCS = 0.0578$ ) accurately implicated `RETURN_COST`, matching the injected *ReturnSurge* anomaly.

Although lowering the threshold yielded additional, lower-significance paths, the primary causal pathways remained unchanged, thereby confirming that ProRCA reliably identifies the true root causes despite increased sensitivity.

## 6. Conclusion and Key Considerations

This paper introduced ProRCA, a python package built on top of the DoWhy causal inference library designed to trace multi-hop causal pathways in complex systems. Unlike conventional methods that rely on correlation or provide limited causal insights, ProRCA integrates structural anomaly scoring, noise-based attribution, and depth-first search to systematically reconstruct causal chains from observed effects to their root causes.

Experiments on a synthetic retail dataset with injected anomalies demonstrate ProRCA’s effectiveness in identifying true underlying causes. The package accurately isolated key drivers and exhibited robustness, maintaining stable causal pathways even under reduced sensitivity settings.

While this study validates ProRCA in a con-

trolled environment, several directions remain for future work. Extending the package to real-time streaming applications is crucial for operational deployment. Additionally, ProRCA’s accuracy is contingent on the correctness of the DAG, making causal discovery an important area for further research. We propose exploring large language models (LLMs) to enhance automated DAG construction.

Overall, ProRCA advances root cause analysis by offering a comprehensive and explainable approach to multi-hop causal reasoning. Its ability to generate actionable insights in complex, data-intensive environments makes it a valuable tool across industries.

## References

- [1] Li Liu, Dewei, Chuan He, Xin Peng, Fan Lin, Chenxi Zhang, Shengfang Gong, Ziang Li, Jiayu Ou, and Zheshun Wu. Microhecl: High-efficient root cause localization in large-scale microservice systems. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 338–347. IEEE, 2021.
- [2] Dongjie Wang, Zhengzhang Chen, Jingchao Ni, Tong Liang, Zheng Wang, Yanjie Fu, and Haifeng Chen. Interdependent causal networks for root cause localization. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5051–5060, 2023.
- [3] Dongjie Wang, Zhengzhang Chen, Yanjie Fu, Yanchi Liu, and Haifeng Chen. Incremental causal graph learning for online root

- cause analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2269–2278, 2023.
- [4] Eric V Strobl and Thomas A Lasko. Identifying patient-specific root causes with the heteroscedastic noise model. *arXiv preprint arXiv:2205.13085*, 2022. Accepted at the Journal of Biomedical Informatics.
- [5] Nastaran Okati, Sergio Hernan Garrido Mejia, William Roy Orchard, Patrick Blöbaum, and Dominik Janzing. Root cause analysis of outliers with missing structural knowledge. *arXiv preprint arXiv:2406.05014*, 2024.
- [6] Kailash Budhathoki, Dominik Janzing, Patrick Bloebaum, and Hoiyi Ng. Why did the distribution change? In *International Conference on Artificial Intelligence and Statistics*, pages 1666–1674. PMLR, 2021.
- [7] Patrick Blöbaum, Peter Götz, Kailash Budhathoki, Atalanti A Mastakouri, and Dominik Janzing. Dowhy-gcm: An extension of dowhy for causal inference in graphical causal models. *arXiv preprint arXiv:2206.06821*, 2022.
- [8] Yinfang Chen, Huaibing Xie, Minghua Ma, Yu Kang, Xin Gao, Liu Shi, Yunjie Cao, Xuedong Gao, Hao Fan, Ming Wen, et al. Automatic root cause analysis via large language models for cloud incidents. In *Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, pages 130–144, 2023.
- [9] Vimalkumar Jeyakumar, Omid Madani, Ali Parandeh, Ashutosh Kulshreshtha, Weifei Zeng, and Navindra Yadav. Explainit!– a declarative root-cause analysis engine for time series data (extended version). *arXiv preprint arXiv:1903.08132*, 2019.
- [10] Eric V Strobl and Thomas A Lasko. Identifying patient-specific root causes with the heteroscedastic noise model. *Journal of Biomedical Informatics*, 150:104585, 2024.
- [11] Xiao Han, Lu Zhang, Yongkai Wu, and Shuhan Yuan. On root cause localization and anomaly mitigation through causal inference. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 699–708, 2023.
- [12] Li Wu, Johan Tordsson, Erik Elmroth, and Odej Kao. Microrca: Root cause localization of performance issues in microservices. In *2020 IEEE/IFIP Network Operations and Management Symposium (NOMS)*, pages 1–9. IEEE, 2020.
- [13] Chang Gong, Di Yao, Chengyang Luo, Wenbin Li, Jingping Bi, Jin Wang, and Yongtao Xie. Porca: Root cause analysis with partially observed data. *ACM Transactions on Knowledge Discovery from Data*, 18(4):1–27, 2024.