

XAI Causal Anomaly Explainer: Algorithmic Analysis and Initial Implementation

Carlos Sanchez Gutierrez (A01412419)
Algoritmos Avanzados / IA Explicable

Abstract—This report analyzes an explainable AI (XAI) method for diagnosing anomalous events in multivariate telemetry by returning an explicit explanation object (minimal causal subgraph + ranking + fidelity). We also present the first runnable implementation and robustness experiments required to set up Midterm 2.

I. PART I — ALGORITHMIC ANALYSIS

A. Q1. Problem Definition, Motivation, and Assumptions

Problem. Given a multivariate time series $X_{1:n}(t)$ and a causal DAG G over variables, detect an anomalous time t and output an *explanation object* E describing *why* the event happened. **Inputs:** telemetry matrix, training window, DAG (known or estimated), detection threshold. **Output (explanation object):** a set of nodes V_E , edges $E_E \subseteq G$, a ranked list of suspicious nodes, and a scalar fidelity score. **Computation model:** local linear causal models per node; residual-based scoring; combinatorial selection of a small explanatory subgraph.

Assumptions. (i) Causal graph is correct enough to support path-based attribution. (ii) Local mechanisms are approximately linear under normal operation. (iii) Anomalies manifest as large residuals at t . **Relaxation impact.** If (i) is relaxed, paths may misattribute responsibility (lower faithfulness). If (ii) is relaxed, linear residuals become a weak proxy; nonlinear models (NN/Bayesian) would improve. If (iii) is relaxed, detection and explanation must incorporate temporal or distributional shifts, not only point outliers.

B. Q2. Core Algorithm and Correctness

Idea. Fit a local predictor for each node from its parents in G . At event time t , compute residual magnitudes as suspicion scores. Construct candidate root-to-node paths for top-scoring nodes and select a small subgraph that maximizes coverage of suspicious nodes while penalizing explanation size.

Invariant/axiom (completeness-style). Explanation should cover a high fraction of the suspicious set (coverage) while remaining small (parsimony). A fidelity score $F(E)$ combines coverage and size penalty.

Correctness sketch. Under correct G and stable mechanisms, a true causal upstream disturbance increases residuals along descendant paths; selecting minimal paths to top residual nodes recovers (approximately) the causal chain. Faithfulness is evaluated by fidelity and robustness stability under perturbations.

C. Q3. Complexity, Guarantees, and Limits

Training local regressors: $\sum_i O(T \cdot \deg(i)^2)$ for linear regression with T training points (implementation uses scikit-learn; practical cost dominated by matrix ops). Residual scoring at one time: $O(n + m)$ evaluations. Path extraction for k top nodes via shortest paths: $O(k(m + n))$ worst-case with BFS on DAG. Greedy set cover approximation for explanation nodes: $O(k \cdot n)$ with small constants.

Limits. If G is wrong, explanation may be plausible but not faithful. If anomalies are subtle or distributed, residual thresholding can miss them. Explanation selection resembles set cover, which is NP-hard; greedy gives standard approximation behavior but no exact optimality guarantee.

D. Q4. Robustness and Scalability

Robustness. We stress-test with additive noise and missingness. Regularization is applied by smoothing detection scores with EMA before selecting suspicious nodes, then re-running the explanation. Stability is measured by Jaccard similarity of explanation node sets between baseline and perturbed runs.

Scalability. The method is parallelizable across nodes (local models) and across robustness scenarios. It is GPU-optional (linear algebra can accelerate) but CPU is sufficient for medium n . Sampling cost is low (single-pass residuals) compared to perturbation-heavy methods like SHAP.

E. Q5. Limitations and Algorithmic Next Steps

(1) Replace linear local models with Bayesian or neural predictors for nonlinear mechanisms. (2) Jointly learn/validate the DAG and propagate uncertainty into explanations. (3) Formalize robustness with bounds on explanation stability under bounded noise/missingness and design stronger regularizers than EMA.

II. PART II — PROJECT REPORT (M1–M2)

A. Project Overview

Goal: build an XAI-first diagnosis tool for anomalous telemetry that outputs an explicit explanation object (subgraph + ranking + fidelity), not only an anomaly flag.

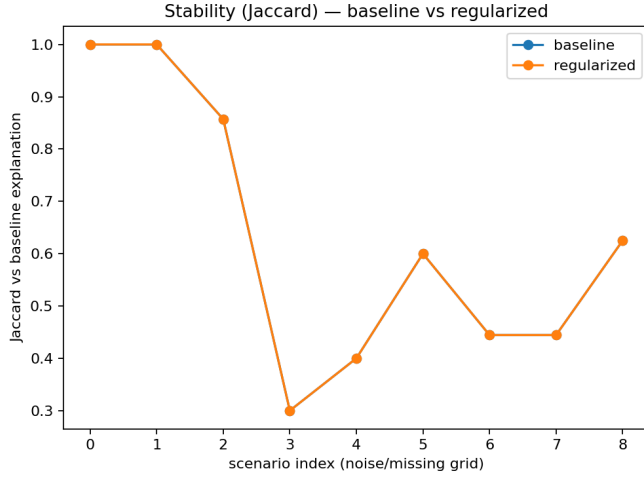


Fig. 1. Stability of explanations under noise and missingness. Jaccard similarity vs the baseline explanation, comparing baseline vs EMA-regularized detection.

B. M1 — XAI Feature (Implemented)

Implemented: runnable repository that (i) simulates causal telemetry on a DAG, (ii) detects an event via z-score thresholding, (iii) generates an explanation object by selecting a small causal subgraph covering top residual nodes. Evidence is provided in Appendix A (JSON explanation object).

C. M2 — Robustness / Regularization (Implemented)

Implemented robustness experiment grid: noise $\sigma \in \{0, 0.5, 1.0\}$ and missing rate $r \in \{0, 0.05, 0.15\}$. Regularization: EMA smoothing of detection score series prior to suspicious-node selection. Evidence is Appendix B (robustness table) and Figure 1.

REFERENCES

APPENDIX A

APPENDIX A: M1 EVIDENCE (EXPLANATION OBJECT)

Paste here the JSON block `explanation_baseline` from `outputs/.../results.json`. Use verbatim in a small excerpt if needed.

APPENDIX B

APPENDIX B: M2 EVIDENCE (ROBUSTNESS TABLE)

Paste here the robustness rows (noise_sigma, missing_rate, jaccard_baseline, jaccard_regularized) from `results.json`.