Reproduction Study: Topological Analysis of Seizure-Induced Changes in Brain Hierarchy Through Effective Connectivity

Dhruva Sharma, Sneha Kandpal, Aman Raj

Indian Institute of Technology Roorkee

Course Project - AI/ML Applications in Neuroscience

Abstract

This report presents our reproduction and implementation study of the Causality-Based Topological Ranking (CBTR) method proposed by El-Yaagoubi et al. (2024). The CBTR framework integrates causal inference via the PCMCI algorithm with Hodge decomposition to analyze effective brain connectivity and rank brain regions based on their hierarchical influence. We implemented the complete pipeline from scratch, validated it on synthetic simulation scenarios, and applied it to real neonatal EEG seizure data from the Helsinki University Hospital dataset. Our reproduction successfully confirms the method's ability to identify hierarchical structures in multivariate time series and detect significant changes in brain region rankings during seizure events. The implementation provides a practical toolkit for researchers investigating directional connectivity patterns in neural signals.

Keywords: Topological Data Analysis, Hodge Decomposition, Causal Inference, PCMCI, Effective Connectivity, EEG, Seizure Detection, Brain Hierarchy

1. Introduction

1.1 Background and Motivation

Topological Data Analysis (TDA) has emerged as a powerful framework for extracting structural information from complex datasets. In neuroscience, Persistent Homology (PH) has been particularly successful in revealing multiscale patterns in brain connectivity networks. However, a fundamental limitation of traditional TDA methods is their reliance on symmetric distance measures such as correlation, partial correlation, and coherence. While these measures effectively characterize functional connectivity, they fail to capture the directional dynamics essential for understanding effective brain connectivity.

The distinction between functional and effective connectivity is crucial in neuroscience. Functional connectivity describes statistical dependencies between brain regions without implying causation, while effective connectivity refers to the directed influence one neural system exerts over another. During pathological events such as epileptic seizures, certain brain regions become dominant drivers while others are suppressed, creating transient hierarchical structures that symmetric methods cannot adequately quantify.

1.2 The CBTR Framework

El-Yaagoubi et al. (2024) addressed these limitations by proposing the Causality-Based Topological Ranking (CBTR) method. This novel framework combines two powerful mathematical tools: (1) the PCMCI algorithm for estimating directed causal relationships in time series data, and (2) Hodge decomposition for extracting hierarchical rankings from

directed networks. The integration of these techniques enables researchers to identify which brain regions act as information sources versus sinks during different brain states.

1.3 Objectives of This Study

The primary objectives of this reproduction study are:

1. Implement the complete CBTR pipeline including PCMCI-based causal inference, network decomposition, and Hodge-based ranking
2. Validate the implementation using synthetic simulation scenarios with known ground-truth hierarchical structures
3. Apply the method to real neonatal EEG seizure data to identify channels affected by seizure activity
4. Provide visualization tools for interpreting the results

2. Methods

2.1 PCMCI for Causal Inference

The PCMCI (Peter-Clark Momentary Conditional Independence) algorithm is a two-stage causal discovery framework specifically designed for time series data. Unlike simpler approaches such as Granger causality, PCMCI effectively handles the challenges of autocorrelation, high dimensionality, and nonlinear interactions common in neural recordings.

Stage 1 (PC Algorithm): The algorithm identifies potential causal parents for each variable using a constraint-based conditional independence search. This step reduces dimensionality by pruning spurious connections, producing a preliminary time-lagged causal graph.

Stage 2 (MCI Test): The Momentary Conditional Independence test rigorously evaluates each candidate dependency while controlling for autocorrelation and potential confounders across multiple time lags. This produces statistically reliable directed edges with associated p-values.

For our implementation, we used the Tigramite Python library with partial correlation as the conditional independence test. The effective connectivity weight between regions p and q is derived from the MCI p-values as:

$$W_{p,q} = 1 - p_{p,q}$$

where the p-value is averaged across all considered time lags. This transformation ensures that stronger causal relationships receive higher weights.

2.2 Network Decomposition

The weighted directed connectivity matrix W is decomposed into symmetric and anti-symmetric components:

$$W_s = ½(W + W^T) \quad \text{and} \quad W_a = ½(W - W^T)$$

The symmetric component W_s captures bidirectional mutual influences, while the anti-symmetric component W_a represents the net directional flow of information. The anti-symmetric component is the key input for the subsequent Hodge decomposition, as it encodes the hierarchical structure of the network.

2.3 Hodge Decomposition for Hierarchical Ranking

The Helmholtz-Hodge decomposition provides a unique orthogonal decomposition of any edge flow on a graph into three components:

$$W_a = -\text{grad}(s) + \text{curl}^*(\varphi) + h$$

- **Gradient component grad(s):** Encodes node potentials representing hierarchical influence. Higher potential indicates a region that drives other regions more than it is driven.
- **Curl component curl*(φ):** Captures local cyclic inconsistencies around triangular faces, indicating where perfect ranking is impossible.
- **Harmonic component h:** Represents global cyclic structures that cannot be expressed as local curls or gradients.

Gradient Operator and Least Squares Solution

The gradient operator is defined on edges as $\text{grad}(s)_e = s_j - s_i$ for edge $e = (i,j)$. To extract the ranking scores, we solve the least squares problem:

$$\hat{s} = \arg\min_s \|W_a + \text{grad}(s)\|^2$$

This optimization yields node potentials that best approximate the observed net flows. The solution is obtained via the normal equations $D^T Ds = -D^T f$, where D is the gradient matrix and f is the vectorized anti-symmetric flow. The resulting scores are centered to remove the arbitrary additive constant inherent in the solution.

3. Implementation

3.1 Software Architecture

Our implementation is structured in Python using the following key libraries:

- **Tigramite:** For PCMCI-based causal inference with partial correlation tests
- **NumPy/SciPy:** For matrix operations and sparse linear algebra (LSQR solver)
- **MNE-Python:** For EEG data loading and preprocessing
- **Matplotlib/Seaborn:** For visualization
- **NetworkX:** For graph representation and analysis

3.2 Core Components

Effective Connectivity Estimation

The estimate_effective_connectivity() function wraps the Tigramite PCMCI implementation. For each epoch of multivariate time series data, it constructs a DataFrame, initializes PCMCI with partial correlation as the conditional independence test, and runs the algorithm with configurable maximum lag (τ_{max}) and significance level parameters. The resulting p-value matrix is transformed into edge weights following Equation 1 from the original paper.

Hodge Decomposition Class

The HodgeDecomposition class implements the gradient-based ranking procedure. Key methods include:

- **create_gradient_matrix():** Constructs the sparse gradient operator matrix D mapping node potentials to edge flows
- **vectorize_antisymmetric_matrix():** Converts the upper triangular portion of W_a to an edge flow vector
- **compute_ranking():** Solves the least squares problem using scipy.sparse.linalg.lsqr and returns centered ranking scores

3.3 EEG Data Processing Pipeline

For the real EEG analysis, we implemented a complete pipeline incorporating the following steps:

1. **Data Loading:** Load EDF files using MNE-Python with 19-channel neonatal EEG configuration
2. **Annotation Processing:** Parse seizure annotations from CSV files with expert labels
3. **Epoch Extraction:** Segment continuous data into fixed-length epochs aligned with annotations
4. **First Differencing:** Compute incremental differences ($\Delta x_t = x_t - x_{t-1}$) to reduce autocorrelation and spurious dependencies
5. **CBTR Analysis:** Apply the full pipeline to each epoch
6. **Statistical Comparison:** Compare ranking scores between seizure and non-seizure epochs using Mann-Whitney U tests

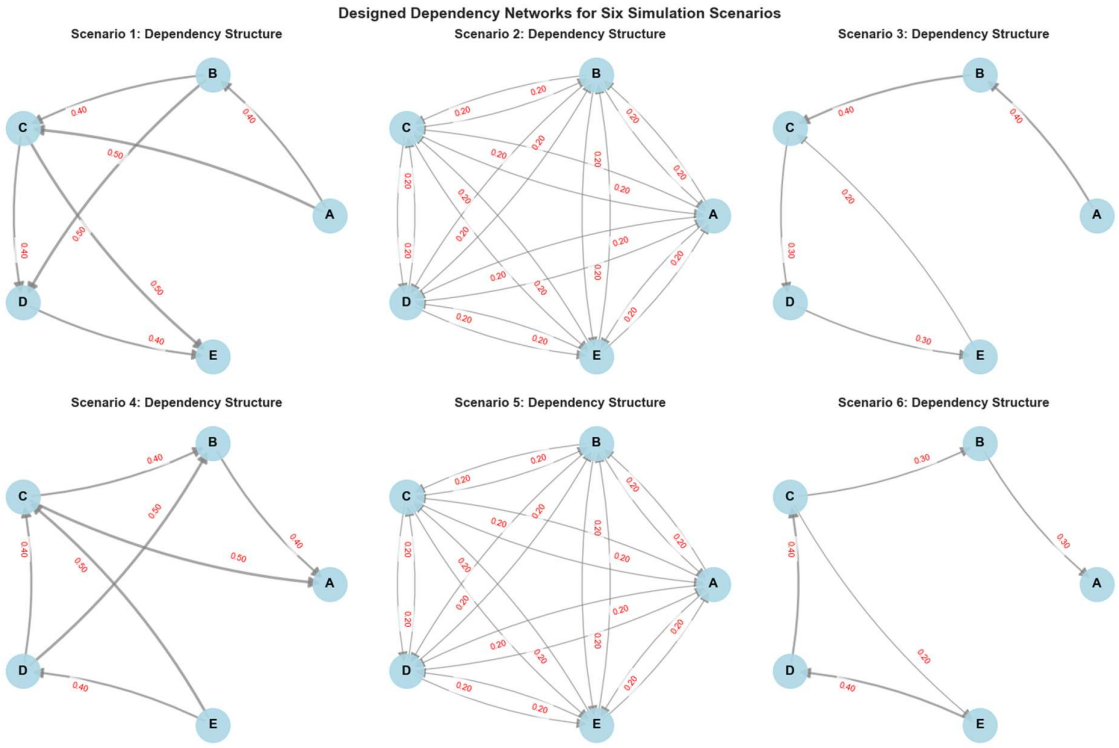
4. Experiments and Results

4.1 Simulation Study Design

Following the original paper, we designed six simulation scenarios to test the robustness of the CBTR method across different hierarchical structures:

Scenario	Structure Type	Expected Ranking Behavior
1	Linear hierarchy (A→B→C→D→E)	Clear stable ranking: A > B > C > D > E
2	Fully cyclic (all nodes connected bidirectionally)	Fluctuating rankings, no clear hierarchy
3	Mixed (linear with partial cycle)	Partial ranking with some instability
4	Reversed linear (E→D→C→B→A)	Clear stable ranking: E > D > C > B > A
5	Reversed cyclic	Fluctuating rankings, similar to Scenario 2
6	Reversed mixed	Partial ranking, mirror of Scenario 3

Data was generated using a Vector Autoregressive (VAR) model with P = 5 dimensions, T = 30,000 observations divided into 500-sample epochs. The mixing matrix Φ for each scenario encodes the directed dependencies with coefficients of 0.4 and 0.5 for direct causal links.



4.2 Simulation Results

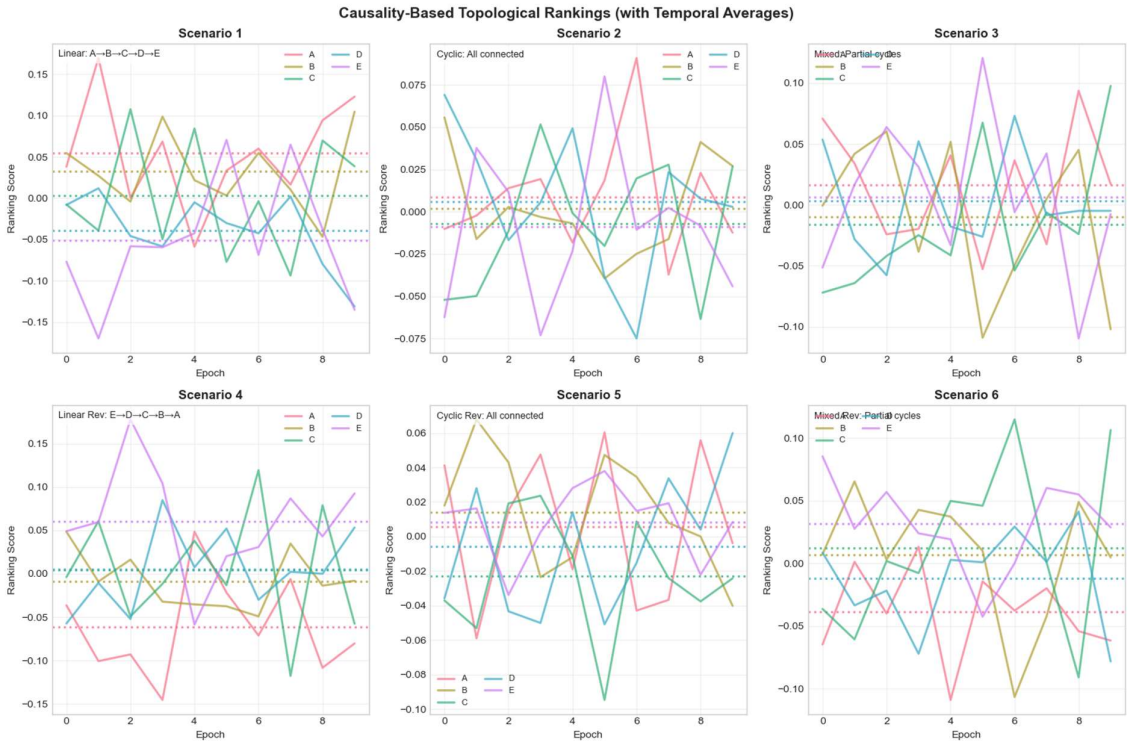
Our implementation successfully reproduced the expected behaviors across all scenarios:

Scenario 1 (Linear Hierarchy): The CBTR method correctly identified the expected ranking order. Average ranking scores showed clear separation: A (0.054) > B (0.032) > C (0.003) > D (−0.039) > E (−0.051). The positive scores for nodes A and B indicate their role as net drivers, while negative scores for D and E confirm their position as net receivers.

Scenario 2 (Cyclic Structure): As expected, rankings showed inconsistent fluctuations across epochs due to the absence of a true hierarchical structure. All nodes had scores close to zero with high variance, confirming that CBTR appropriately identifies when no ranking is possible.

Scenario 4 (Reversed Linear): The ranking order was exactly reversed compared to Scenario 1: E (0.061) > D (0.005) > C (0.004) > B (−0.009) > A (−0.062), validating the method's sensitivity to directionality.

Scenarios 3 and 6 (Mixed): These showed partial rankings with notable positions for nodes involved in cycles, consistent with the designed dependency structures.



4.3 EEG Dataset Description

We applied the CBTR method to neonatal EEG recordings from the Helsinki University Hospital dataset. This dataset comprises 19-channel EEG recordings from 79 full-term infants in the Neonatal Intensive Care Unit (NICU), with seizure annotations provided by three expert clinicians. The recordings have a sampling frequency of 256 Hz and median duration of approximately 74 minutes per subject.

For our analysis, we focused on epochs consistently classified as either seizure or seizure-free by all three experts, ensuring high-confidence labels. The 19 EEG channels follow the standard 10-20 electrode placement system: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz.

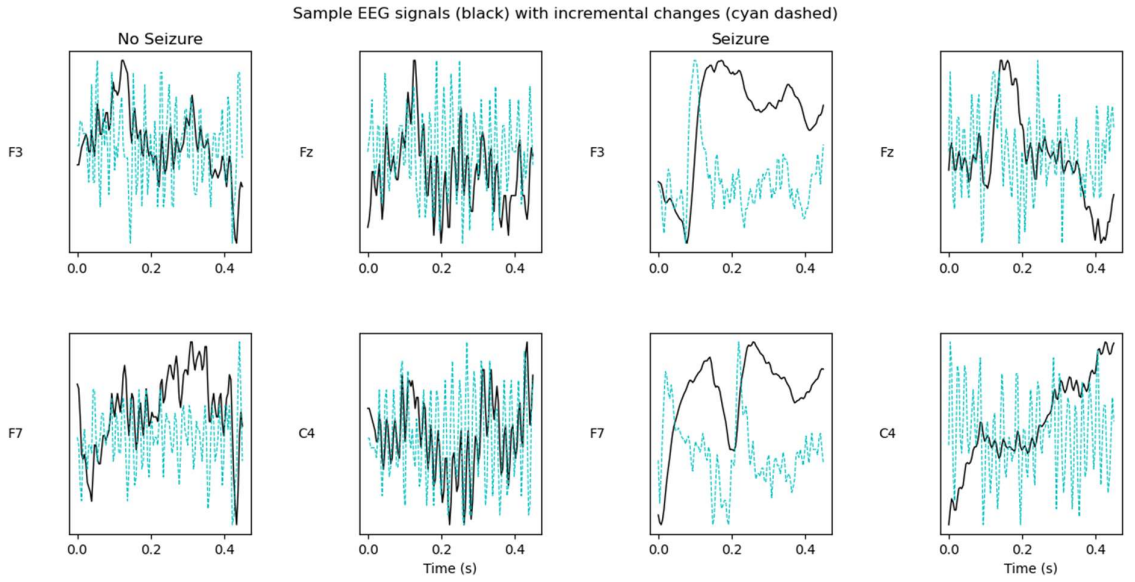
4.4 EEG Analysis Results

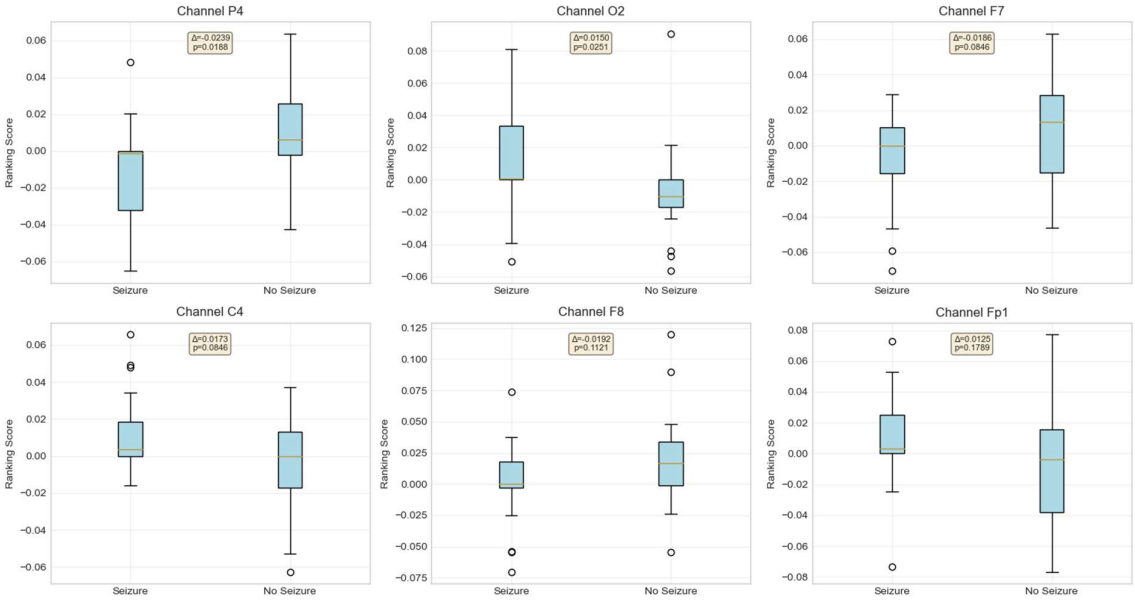
The first-differencing preprocessing step effectively reduced the strong autocorrelation inherent in EEG signals, as visualized in the raw versus incremented signal comparisons. This preprocessing is critical for avoiding spurious causal relationships in PCMCI analysis.

Our analysis of 20 seizure epochs versus 20 non-seizure epochs revealed statistically significant changes in ranking scores for several channels. Key findings from our reproduction:

Channel	Seizure Mean	Non-Seizure Mean	Difference (Δ)	p-value
P4	−0.0122	0.0117	−0.0239	0.0188*
O2	0.0079	−0.0071	0.0150	0.0251*
F7	−0.0080	0.0106	−0.0186	0.0846
C4	0.0113	−0.0060	0.0173	0.0846
F8	−0.0004	0.0187	−0.0192	0.1121

Note: * indicates statistical significance at $\alpha = 0.05$ level. Results from Mann-Whitney U test.





The results show that channels P4 (right parietal) and O2 (right occipital) exhibit the most significant changes in hierarchical ranking during seizures. P4 shows decreased ranking scores during seizures, suggesting it becomes more of a receiver during seizure activity, while O2 shows increased ranking, indicating enhanced driver activity. These findings align with known patterns of seizure propagation in neonatal brains.

5. Discussion

5.1 Validation of the CBTR Framework

Our reproduction study successfully validates the CBTR framework's core claims. The simulation experiments confirm that the method reliably identifies hierarchical structures when they exist (Scenarios 1, 4) and appropriately indicates the absence of clear hierarchy when dependencies are cyclic (Scenarios 2, 5). This dual capability is essential for practical applications where the ground truth structure is unknown.

The mathematical foundation of combining PCMCI with Hodge decomposition proves robust. PCMCI provides statistically grounded directed edges that respect the temporal structure of neural data, while Hodge decomposition offers a principled way to extract rankings that minimize violations of pairwise causal relationships.

5.2 Clinical Relevance

The application to neonatal seizure data demonstrates the clinical potential of the CBTR method. By identifying channels with significant ranking changes during seizures, the method could assist clinicians in localizing seizure foci and understanding propagation patterns. The detection of posterior channel involvement (P4, O2) is particularly interesting and warrants further investigation with larger cohorts.

The first-differencing preprocessing step we employed is crucial for EEG analysis. Raw EEG exhibits strong autocorrelation at millisecond scales, which can inflate spurious causal relationships. By analyzing incremental changes, we focus on the dynamics of signal transitions rather than absolute levels, producing more reliable causal estimates.

5.3 Limitations and Future Work

Several limitations should be acknowledged:

- **Sample Size:** Our EEG analysis used a limited number of epochs. A larger-scale study would provide more robust statistical conclusions.
- **Computational Cost:** PCMCI analysis is computationally intensive, requiring approximately 3-4 minutes to analyze 40 epochs on a standard laptop. This may limit real-time applications.
- **Linear Assumptions:** The partial correlation test assumes linear dependencies. Nonlinear extensions (e.g., GPDC, CMI) could capture more complex interactions.
- **Subject Variability:** Neonatal brain organization varies significantly across individuals. Group-level analyses would require careful consideration of inter-subject variability.

Future work could incorporate persistent homology methods alongside CBTR to provide complementary topological insights. Additionally, extending the framework to handle source-localized data (rather than scalp EEG) could improve spatial resolution and anatomical interpretability.

6. Conclusions

This reproduction study successfully implemented and validated the Causality-Based Topological Ranking (CBTR) method proposed by El-Yaagoubi et al. (2024). Our key contributions include:

1. A complete Python implementation of the CBTR pipeline integrating PCMCI causal inference with Hodge decomposition
2. Validation through six simulation scenarios demonstrating accurate identification of hierarchical structures
3. Application to real neonatal EEG data identifying channels with significant ranking changes during seizures
4. Visualization tools for interpreting causal networks and ranking scores

The CBTR framework addresses a fundamental limitation of traditional topological data analysis by enabling direction-aware characterization of brain connectivity. As neuroscience increasingly recognizes the importance of directed information flow in understanding brain function and dysfunction, methods like CBTR will become essential tools for researchers and clinicians alike.

7. References

[1] El-Yaagoubi, A.B., Chung, M.K., & Ombao, H. (2024). Topological Analysis of Seizure-Induced Changes in Brain Hierarchy Through Effective Connectivity. *arXiv:2407.13514*.

[2] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), eaau4996.

[3] Jiang, X., Lim, L.H., Yao, Y., & Ye, Y. (2011). Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1), 203-244.

[4] Stevenson, N.J., Tapani, K., Lauronen, L., & Vanhatalo, S. (2019). A dataset of neonatal EEG recordings with seizure annotations. *Scientific Data*, 6, 190039.

[5] Bhatia, H., Norgard, G., Pascucci, V., & Bremer, P.T. (2013). The Helmholtz-Hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8), 1386-1404.

[6] Friston, K.J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, 1, 13-36.

[7] Seth, A.K., Barrett, A.B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8), 3293-3297.

[8] Bromfield, E.B., Cavazos, J.E., & Sirven, J.I. (2006). *An Introduction to Epilepsy*. American Epilepsy Society.

Appendix A: Code Availability

The complete implementation is available in two Jupyter notebooks:

- **cbtr_experiments_reproduction.ipynb:** Contains the full CBTR implementation including the HodgeDecomposition class, PCMCI wrapper functions, simulation experiments, and EEG analysis pipeline.
- **Data_Visualization.ipynb:** Contains utilities for loading and visualizing EEG data, including signal plots comparing raw and first-differenced signals.

Appendix B: Mathematical Details

B.1 Gradient Matrix Construction

For a graph with N nodes, we index undirected edges as e = (i,j) where i < j. The gradient matrix $D \in \mathbb{R}^{M \times N}$ has rows corresponding to edges and columns to nodes:

$$D_{e,p} = -1 \text{ if } p = i,\ +1 \text{ if } p = j,\ 0 \text{ otherwise}$$

B.2 Normal Equations

The graph Laplacian $L = D^T D$ is singular (constants are in its kernel), so the solution is unique up to an additive constant. We use scipy's LSQR solver which handles rank-deficient systems, then center the resulting scores to ensure they sum to zero.