The Human Brain as a Combinatorial Complex

Valentina Sánchez   V.SANCHEZMELCHOR@TILBURGUNIVERSITY.EDU
Tilburg University, The Netherlands

Çiçek Güven   C.GUVEN@TILBURGUNIVERSITY.EDU
Tilburg University, The Netherlands

Koen Haak   K.HAAK@TILBURGUNIVERSITY.EDU
Tilburg University, The Netherlands

Theodore Papamarkou   THEODORE@POLYSHAPE.COM
PolyShape, Greece

Gonzalo Nápoles   G.NAPOLES@TILBURGUNIVERSITY.EDU
Tilburg University, The Netherlands

Marie Šafář Postma   M.SAFARPOSTMA@TILBURGUNIVERSITY.EDU
Tilburg University, The Netherlands

Abstract

We propose a framework for constructing combinatorial complexes (CCs) from fMRI time series data that captures both pairwise and higher-order neural interactions through information-theoretic measures, bridging topological deep learning and network neuroscience. Current graph-based representations of brain networks systematically miss the higher-order dependencies that characterize neural complexity, where information processing often involves synergistic interactions that cannot be decomposed into pairwise relationships. Unlike topological lifting approaches that map relational structures into higher-order domains, our method directly constructs CCs from statistical dependencies in the data. Our CCs generalize graphs by incorporating higher-order cells that represent collective dependencies among brain regions, naturally accommodating the multi-scale, hierarchical nature of neural processing. The framework constructs data-driven combinatorial complexes using O-information and S-information measures computed from fMRI signals, preserving both pairwise connections and higher-order cells (e.g., triplets, quadruplets) based on synergistic dependencies. Using NetSim simulations as a controlled proof-of-concept dataset, we demonstrate our CC construction pipeline and show how both pairwise and higher-order dependencies in neural time series can be quantified and represented within a unified structure. This work provides a framework for brain network representation that preserves fundamental higher-order structure invisible to traditional graph methods, and enables the application of topological deep learning (TDL) architectures to neural data.

Keywords: Topological Deep Learning, Combinatorial Complex, Higher-Order Networks, Information Theory, Brain Networks

1. Introduction

The human brain constitutes one of nature's most remarkable examples of organized complexity, exhibiting rich patterns of connectivity and information processing across multiple spatial and temporal scales. As an organized complex system composed of billions of interconnected neurons, the brain demonstrates the defining characteristics of complexity: nontrivial interactions between components that generate emergent properties not reducible to individual parts, and hierarchical organization spanning from molecular to systems-level dynamics (Sporns, 2022).

Current approaches to brain network analysis predominantly rely on graph-based representations, where brain regions are modeled as nodes and their pairwise statistical relationships form edges. This framework has become the standard in network neuroscience, enabling fundamental insights into functional connectivity, network topology, and brain organization (Bassett and Sporns, 2017; Sporns, 2010). The success of this approach has been so profound that neuroscience increasingly needs network science to understand the brain's complex organizational principles (Barabási et al., 2023). However, graph-based representations face a fundamental limitation: they are inherently restricted to pairwise relationships and cannot capture higher-order interactions—collective dependencies among three or more brain regions that emerge from their joint activity patterns (Giusti et al., 2016; Battiston et al., 2021; Boccaletti et al., 2023).

This limitation is particularly problematic for understanding brain function, where information processing often involves synergistic interactions that cannot be decomposed into pairwise relationships. Recent neuroscientific evidence suggests that higher-order interactions may represent important organizing principles of brain networks (Luppi et al., 2022; Santoro et al., 2024), motivating our research question: How can we capture higher-order interactions in neural data?

Recent efforts in topological data analysis (TDA) have contributed important tools for studying neural data, including fMRI, by characterizing global shape and connectivity. Most approaches emphasize topological invariants—such as Betti numbers and persistent homology—derived from geometric or distance-based filtrations (Santoro et al., 2024). In contrast, we adopt a data-driven information-theoretic approach without requiring geometric embeddings or distance-based filtrations. We extend graph-theoretical methods beyond pairwise connectivity by constructing combinatorial complexes using multivariate information measures that directly exploit the statistical dependencies in fMRI signals.

From an information-theoretic perspective, these higher-order dependencies can be formally characterized using measures that distinguish between redundant and synergistic information (Varley, 2025). Redundant information refers to information that is present in multiple source variables simultaneously, while synergistic information is information that is only accessible when multiple source variables are observed together and cannot be obtained from any individual source. Recent developments in multivariate information theory provide the mathematical foundation for identifying and quantifying these higher-order effects in neural data (Rosas et al., 2024; Mediano et al., 2022).

We propose that recent advances in Topological Deep Learning (TDL) offer a promising solution for representing the now quantifiable higher-order interactions in neural time series data. With information-theoretic measures enabling systematic quantification of synergistic

dependencies, we can construct combinatorial complexes (CCs)—mathematical structures that generalize graphs to include higher-order relationships (Hajij et al., 2022; Papillon et al., 2023). CCs are particularly well-suited for this application because their flexible structure allows direct integration of continuous information-theoretic measures into discrete topological representations, representing both pairwise and higher-order interactions without the closure or inclusion constraints required by other topological structures.

Our data-driven approach generates higher-order cells directly from multivariate dependencies in neural time series. The resulting combinatorial complexes can provide structured representations for combinatorial complex neural networks (CCNNs), which could perform message passing across multiple topological ranks. This would preserve both local pairwise structure and global higher-order patterns, potentially enabling improved learning tasks such as brain state classification or cognitive task decoding.

We present this work as a first-generation approach that establishes foundational principles while acknowledging current challenges that require future methodological development. This framework bridges topological deep learning and network neuroscience not via topological lifting from lower-order relational scaffolds (e.g., graphs), but through data-driven construction of combinatorial complexes from multivariate statistical dependencies.

2. From Brain Graphs to Combinatorial Complexes

2.1. Mathematical Framework

Traditional brain network analysis constructs an undirected graph $G = (V, E)$, where nodes represent brain regions and edges encode pairwise statistical dependencies. These dependencies are typically inferred from region-wise time series using statistical measures such as Pearson correlation, partial correlation, mutual information, or coherence (Smith et al., 2011). This representation captures at most $\binom{N}{2}$ pairwise relationships, typically stored in an $N \times N$ adjacency matrix.

Combinatorial complexes (CCs) generalize graph-based representations by formally incorporating higher-order cells. In our framework, CCs store both pairwise interactions (edges) and higher-order relationships (e.g., triplets) in a unified, hierarchical structure. The full CC consists of a structured collection of subsets of varying sizes, with up to $\sum_{k=0}^{n} \binom{N}{k+1}$ potential cells. This makes the CC a discrete, higher-dimensional object that grows combinatorially with the number of nodes, posing significant computational challenges common across higher-order network modeling approaches.

To identify and filter higher-order dependencies, we turn to information-theoretic measures. While several options exist—including mutual information, Partial Information Decomposition (PID), O-information, and S-information (Varley, 2025)—we focus on the latter two due to their computational tractability and suitability for capturing synergy and redundancy in multivariate systems. CCs relax the closure and embedding constraints imposed by other topological structures (e.g., simplicial complexes), enabling integration with information-theoretic methods in our pipeline.

A **combinatorial complex** (Hajij et al., 2022) is a triple $(S, \mathcal{X}, \mathrm{rk})$ where S is a finite set, $\mathcal{X} \subseteq \mathcal{P}(S) \setminus \{\emptyset\}$, and $\mathrm{rk} : \mathcal{X} \to \mathbb{Z}_{\geq 0}$ satisfies: (i) $\{s\} \in \mathcal{X}$ for all $s \in S$; and (ii) if $x \subseteq y$ with $x, y \in \mathcal{X}$, then $\mathrm{rk}(x) \leq \mathrm{rk}(y)$. The rank function induces a hierarchy: rank-0 cells are

individual nodes, rank-1 cells are edges, and rank-k cells correspond to $k + 1$-tuples (e.g., triplets, quadruplets).

2.2. Combinatorial Complex Construction

To construct higher-order cells, we quantify multivariate statistical dependencies using two information-theoretic measures: **S-information** (Σ) and **O-information** (Ω) (Rosas et al., 2024; Varley, 2025). These measures quantify complementary aspects of multivariate statistical dependence across subsets.

S-information. Given random variables $X_1, \dots, X_n$, S-information quantifies the total statistical interdependence within the system:

$$\Sigma(X) = TC(X) + DTC(X) \tag{1}$$

where TC and DTC are total and dual total correlations. Higher Σ values indicate stronger multivariate dependencies, including potential higher-order synergistic interactions.

O-information. For random variables $X_1, \dots, X_n$, the O-information measures the system's overall redundancy–synergy bias:

$$\Omega(X) = TC(X) - DTC(X) \tag{2}$$

Negative Ω indicates net synergy; positive values indicate redundancy. In this proof-of-concept, we compute both measures and apply dual thresholding criteria: Σ for overall statistical strength and Ω to select synergy-dominated structures.

Combinatorial Complex Structure Across Ranks

Rank 0 (Nodes)   *Rank 1 (Edges)*   *Rank 2 (Triplets as Hyperedges)*

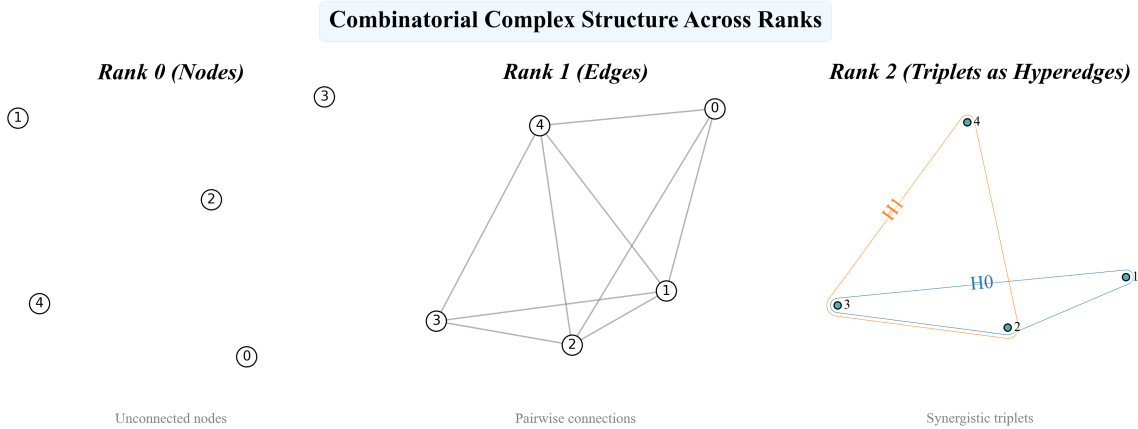Unconnected nodes   Pairwise connections   Synergistic triplets

Figure 1: Toy combinatorial complex from NetSim neural time series. **Rank-0** cells are brain regions (nodes); **Rank-1** edges capture pairwise interactions (mutual information ≥ 0.02); **Rank-2** hyperedges denote synergistic triplets (S-information ≥ 0.45, O-information $\lesssim 0$). Top triplets: (2,3,4) with $S = 0.51$, $\Omega = 0.06$, and (1,2,3) with $S = 0.48$, $\Omega = 0.04$. Construction pipeline and implementation details are provided in Appendix A.

Let S denote a set of variables (e.g., neural time series). For each candidate subset $x \subseteq S$, we compute $\Sigma(x)$ and $\Omega(x)$. Subsets with $\Sigma(x) > \tau$ and $\Omega(x) \lesssim 0$ are retained as rank-k cells in the combinatorial complex, where $k = |x| - 1$. This dual criterion targets

synergistic higher-order interactions—collective dependencies that emerge only from joint activity and cannot be decomposed into pairwise relationships. Pairwise dependencies are included as rank-1 cells based on a separate statistical measure (e.g., Pearson correlation).

This yields a combinatorial complex whose structure reflects higher-order informational dependencies inherent in the data (Figure 1). In contrast to topological filtrations, which rely on geometric constraints, our approach selects cells based on statistical dependencies, uncovering structure intrinsic to the data rather than imposed by geometric priors.

3. Limitations & Future directions

While combinatorial complexes offer a novel way to represent higher-order dependencies in brain data, several limitations remain. The combinatorial growth of candidate cells poses significant computational challenges, which could be mitigated through techniques such as locality-sensitive hashing (Indyk and Motwani, 1998), Kalman filtering (Kalman, 1960), and similarity-based preselection. Our current reliance on fixed thresholding may overlook weak but meaningful interactions; future work should explore adaptive criteria, statistical testing, and robust estimators beyond the Gaussian assumption (Varley, 2025). Finally, results may depend strongly on how brain regions are defined (e.g., ICA vs. anatomical parcellation (Smith et al., 2011)), underscoring the need for standardized benchmarks in higher-order network modeling (Bechler-Speicher et al., 2024). This framework generalizes to any multivariate time series data and could be integrated with TopoBench (Telyatnikov et al., 2025) for topological deep learning applications.

References

D. L. Barabási, G. Bianconi, E. Bullmore, M. Burgess, D. Chialvo, J. R. Cohen, C. Constantinidis, G. Deco, L. Douw, J. Duch, K. J. Friston, D. Garlaschelli, G. Gong, P. Hagmann, C. J. Honey, A. E. Kahn, R. Kana, F. M. Krienen, R. Lambiotte, V. Latora, D. Marinazzo, N. T. Markov, D. Meunier, S. Muldoon, M. Perc, J. M. Shine, O. Sporns, C. J. Stam, M. Turalska, P. Van Mieghem, B. T. T. Yeo, and A. Zalesky. Neuroscience needs network science. *The Journal of Neuroscience*, 43(34):5989–5995, 2023.

Danielle S. Bassett and Olaf Sporns. Network neuroscience. *Nature Neuroscience*, 20(3): 353–364, 2017.

Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, Megan M. Murray, Tiago P. Peixoto, Francesco Vaccarino, and Giovanni Petri. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10): 1093–1098, 2021.

Maya Bechler-Speicher, Ben Finkelshtein, Fabrizio Frasca, Luis Müller, Jan Tönshoff, Antoine Siraudin, Viktor Zaverkin, Michael M Bronstein, and Mathias Niepert. Position: Graph learning will lose relevance due to poor benchmarks. *arXiv preprint arXiv:2402.02287*, 2024.

Stefano Boccaletti, Pietro De Lellis, Charo I. del Genio, Karin Alfaro-Bittner, Regino Criado, Sarika Jalan, and Miguel Romance. The structure and dynamics of networks with higher order interactions. *Physics Reports*, 1018:1–64, 2023.

Chad Giusti, Robert Ghrist, and Danielle S Bassett. Two's company, three (or more) is a simplex. *Journal of Computational Neuroscience*, 41(1):1–14, 2016. doi: 10.1007/s10827-016-0608-6.

Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K. Dey, Soham Mukherjee, Shreyas N. Samaga, Neal Livesay, Robin Walters, Paul Rosen, and Michael T. Schaub. Topological deep learning: Going beyond graph data. *arXiv preprint arXiv:2206.00606*, 2022. arXiv:2206.00606 [cs.LG].

Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC)*, pages 604–613. ACM, 1998. doi: 10.1145/276698.276876.

Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, 82(Series D):35–45, 1960. doi: 10.1115/1.3662552.

Joseph T. Lizier. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11, 2014.

Andrea I. Luppi, Pedro A. M. Mediano, Fernando E. Rosas, Nicola Holland, Tim D. Fryer, John T. O'Brien, James B. Rowe, David K. Menon, Daniel Bor, and Emmanuel A. Stamatakis. A synergistic core for human brain evolution and cognition. *Nature Neuroscience*, 25(6):771–782, 2022.

Pedro A. M. Mediano, Fernando E. Rosas, Andrea I. Luppi, Henrik J. Jensen, Anil K. Seth, Adam B. Barrett, Robin L. Carhart-Harris, and Daniel Bor. Greater than the parts: a review of the information decomposition approach to causal emergence. *Philosophical Transactions of the Royal Society A*, 380(2227):20210246, 2022.

Mathilde Papillon, Sophia Sanborn, Mustafa Hajij, and Nina Miolane. Architectures of topological deep learning: A survey on topological neural networks. *arXiv preprint arXiv:2304.10031*, 2023. arXiv:2304.10031 [cs.LG].

Fernando E. Rosas, Pedro A. M. Mediano, Andrea I. Luppi, Thomas F. Varley, Joseph T. Lizier, Sebastiano Stramaglia, Henrik J. Jensen, and Daniele Marinazzo. Information decomposition and the informational architecture of the brain. *Trends in Cognitive Sciences*, 28(4):352–368, 2024.

Andrea Santoro, Federico Battiston, Maxime Lucas, Giovanni Petri, and Enrico Amico. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *Nature Communications*, 15(1):9899, 2024.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *NeuroImage*, 54(2):875–891, 2011.

Olaf Sporns. *Networks of the Brain.* MIT Press, Cambridge, MA, 2010.

Olaf Sporns. The complex brain: connectivity, dynamics, information. *Trends in Cognitive Sciences*, 26(12):1066–1067, 2022.

Lev Telyatnikov, Guillermo Bernardez, Marco Montagna, Mustafa Hajij, Martin Carrasco, Pavlo Vasylenko, Mathilde Papillon, Ghada Zamzmi, Michael T. Schaub, Jonas Verhellen, Pavel Snopov, Bertran Miquel-Oliver, Manel Gil-Sorribes, Alexis Molina, Victor Guallar, Theodore Long, Julian Suk, Patryk Rygiel, Alexander Nikitin, Giordan Escalona, Michael Banf, Dominik Filipiak, Max Schattauer, Liliya Imasheva, Alvaro Martinez, Halley Fritze, Marissa Masden, Valentina Sánchez, Manuel Lecha, Andrea Cavallo, Claudio Battiloro, Matt Piekenbrock, Mauricio Tec, George Dasoulas, Nina Miolane, Simone Scardapane, and Theodore Papamarkou. TopoBench: A Framework for Benchmarking Topological Deep Learning. *arXiv preprint arXiv:2406.06642*, 2025. URL https://arxiv.org/abs/2406.06642.

Thomas F. Varley. Information theory for complex systems scientists: What, why, & how? *arXiv preprint arXiv:2304.12482*, 2025. arXiv:2304.12482 [cs.IT].

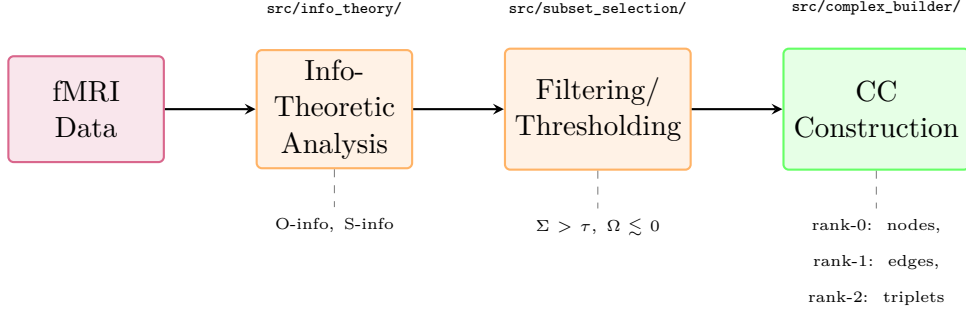## Appendix A. Construction Pipeline And Implementation Details

Figure 2: Pipeline for constructing combinatorial complexes (CCs) from fMRI data.

This appendix describes the implementation used to construct combinatorial complexes (CCs) from NetSim (Smith et al., 2011) synthetic data and to generate the results shown in Figure 1, with the overall workflow summarised in Figure 2. Our approach focuses on information-theoretic analysis of higher-order dependencies using established tools, with future extensions aimed at applying CC assembly to larger real-fMRI datasets and integrating the framework with topological deep learning methods.

We use NetSim synthetic BOLD time series that simulate realistic fMRI dynamics via a dynamic causal modelling (DCM) neural process coupled to a nonlinear balloon–Windkessel haemodynamic forward model. Each dataset consists of an $N \times T$ matrix $\mathbf{X}$ representing N brain regions over T time points, with neural activity convolved through region-specific haemodynamic response functions, realistic inter-regional HRF variability, and additive Gaussian noise at the BOLD level. While NetSim provides a graph-level ground truth, we use only the generated time series, making it a small, well-controlled, and interpretable yet physiologically informed testbed for proof-of-concept higher-order network construction.

A.1. Information-theoretic analysis

The core contribution of our pipeline (illustrated in Figure 2) lies in integrating higher-order information-theoretic analysis with combinatorial complex (CC) construction. We use the Java Information Dynamics Toolkit (JIDT) (Lizier, 2014), interfaced via Python using JPype.

Higher-order measures. For each candidate triplet (i, j, k), we compute S-information and O-information as described in the main text. These quantify the strength of statistical interdependence and the net redundancy–synergy bias within the triplet.

The JIDT library provides robust estimators for multivariate information measures, handling the computational complexity of higher-order statistics. Our wrapper functions in `src/info_theory/jidt_interface/` standardize the input/output format and provide error handling for the JPype bridge.

A.2. Combinatorial complex construction framework

Rank assignment. Brain regions are treated as rank-0 cells (nodes). Rank-1 cells (edges) are added based on pairwise mutual information (MI), using a fixed threshold of 0.02 to ensure sparsity while retaining meaningful connections. For rank-2 cells (triplets), we compute both S-information (Σ) and O-information (Ω) using Gaussian estimators from JIDT (Lizier, 2014). A triplet is retained as a rank-2 cell if both its Σ value exceeds 0.45 and $\Omega \lesssim 0$, reflecting strong multivariate dependencies with synergy-dominated character. This dual criterion ensures we capture higher-order interactions that are both informationally rich and synergistic rather than redundant. Although our current implementation focuses on ranks 0–2 for tractability, the framework generalizes naturally to higher-rank structures.

Thresholding strategy. The choice of fixed thresholds was guided by exploratory runs on the NetSim dataset. Σ is used to quantify the overall strength of statistical dependencies within a triplet, while Ω is computed in parallel to diagnose whether synergy or redundancy dominates. Together, they allow us to detect strongly coupled subsets and interpret the nature of their higher-order interactions. In this proof-of-concept, we implement dual selection criteria: $\Sigma > 0.45$ identifies subsets with strong statistical dependencies, while $\Omega \lesssim 0$ ensures we retain only synergy-dominated structures. We use "$\lesssim 0$" rather than "< 0" to accommodate weak synergy-dominant structures with small positive Ω values. This dual thresholding approach filters for both statistical strength and synergistic character. As emphasized by Varley (2025), fixed thresholds can miss weak but relevant interactions, motivating future integration of adaptive criteria, statistical tests, and more robust estimators.

Illustrative example and interpretation. Figure 1 shows the CC constructed from subject 1 in the `sim1.mat` NetSim dataset (50 subjects, 5 regions, 200 time points). We selected this example for its simplicity as a toy testbed. The top two triplets—(2,3,4) with $\Sigma = 0.51$, $\Omega = 0.06$, and (1,2,3) with $\Sigma = 0.49$, $\Omega = 0.04$—illustrate how our method captures structured higher-order dependencies that are invisible to graph-based representations. While Σ reflects the overall strength of multivariate dependencies, Ω helps reveal the balance between synergy and redundancy: the mildly positive values here suggest a mixed interaction pattern with a synergy-dominant core. In larger or more complex datasets, Ω will further aid in distinguishing interaction motifs. Even in this minimal setting, the resulting CC reveals triadic structures that reflect non-pairwise organization in the neural signals.

A.3. Computational considerations and software stack

Scalability. The computational cost of constructing combinatorial complexes grows combinatorially with rank. For a dataset with N regions, the number of candidate rank-k cells is $\binom{N}{k+1}$, and evaluating each requires computing information-theoretic quantities over $(k+1)$ variables. For example:

- Rank-1 (edges): $O(N^2)$

- Rank-2 (triplets): $O(N^3)$

- Rank-$(q-1)$: $O(N^q)$

In practice, our proof-of-concept restricts analysis to triplets for tractability. With typical NetSim dimensions ($N \approx 5$, $T \approx 200$), the pipeline executes efficiently, requiring evaluation of only $\binom{5}{3} = 10$ triplets. However, the combinatorial growth becomes prohibitive for realistic brain parcellations: standard atlases with $N = 100$ regions would require evaluating $\binom{100}{3} = 161,700$ triplets. Beyond $N \approx 50$, computational requirements exceed typical desktop capabilities without sophisticated optimization. The modular design in `src/info_theory/` supports future scalability improvements via candidate pre-filtering, parallelisation, and distributed computation, and could incorporate advanced strategies such as locality-sensitive hashing (Indyk and Motwani, 1998), Kalman filtering (Kalman, 1960), and similarity-based preselection to further reduce the search space.

Software dependencies.

The implementation builds on:

- **Core computation:** NumPy, SciPy for numerical operations

- **JIDT interface:** JPype1 for Java–Python bridging

- **Visualization:** Matplotlib for time series plots, HyperNetX for combinatorial complex visualisation (Figure 1)

A.4. Reproducibility and code availability

The public implementation and scripts to reproduce the figure are available at `TopoBrainX`.

Future extensions. The current implementation establishes a foundation for constructing combinatorial complexes directly from time series data. While this framework was designed with fMRI in mind, it generalizes to any multivariate time series or signal data. In this proof-of-concept, we focus on demonstrating the full pipeline rather than on exploratory analysis or interpretation. Planned developments include scaling to large datasets while mitigating combinatorial explosion, applying the method to real fMRI data for data analysis and interpretation, incorporating adaptive thresholding strategies, and integrating with TopoBench (Telyatnikov et al., 2025) for downstream topological deep learning applications.