MCBIF: MEASURING TOPOLOGICAL AUTOCORRELATION IN MULTISCALE CLUSTERINGS VIA 2-PARAMETER PERSISTENT HOMOLOGY

Anonymous authors
Paper under double-blind review

ABSTRACT

Datasets in applications often possess an intrinsic multiscale structure with meaningful descriptions at different levels of coarseness. Such datasets are naturally described as multi-resolution clusterings, i.e., not necessarily hierarchical sequences of partitions across scales. To analyse and compare such sequences, we use tools from topological data analysis and define the Multiscale Clustering Bifiltration (MCbiF), a 2-parameter filtration of abstract simplicial complexes that encodes cluster intersection patterns across scales. The MCbiF can be interpreted as a higher-order extension of Sankey diagrams and reduces to a dendrogram for hierarchical sequences. We show that the multiparameter persistent homology (MPH) of the MCbiF yields a finitely presented and block decomposable module, and its stable Hilbert functions characterise the topological autocorrelation of the sequence of partitions, providing topological feature maps for downstream machine learning tasks. In particular, at dimension zero, the MPH captures the degree of nestedness of the sequence, whereas at dimension one, it captures higher-order inconsistencies between clusters across scales. We demonstrate in our experiments that the MCbiF feature maps outperform information-based baseline features on both regression and classification tasks on non-hierarchical sequences of partitions. We also show an application of MCbiF to real-world data to measure non-hierarchies in wild mice social grouping patterns across time.

1 INTRODUCTION

In many areas of application, datasets possess an intrinsic multiscale structure, whereby meaningful descriptions exist at different levels of coarseness (or scales). Think, for instance, of the multi-resolution structure in commuter mobility patterns (Alessandretti et al., 2020; Schindler et al., 2023), communities in social networks (Beguerisse-Díaz et al., 2017) and thematic groups of documents (Blei et al., 2003; Altuncu et al., 2019; Grootendorst, 2022); the subgroupings in single-cell data (Hoekzema et al., 2022) or phylogenetic trees (Chan et al., 2013); and the functional substructures in proteins (Delvenne et al., 2010; Delmotte et al., 2011). In such cases, it is desirable to go beyond a single clustering and find a multi-resolution description in terms of a not necessarily hierarchical sequence of partitions across scales, described by a scale parameter t. One way to produce multiscale clusterings is by exploiting a diffusion process that reveals robust partitions of increasing coarseness, yet not necessarily hierarchical, as it explores the data geometry with an increasing time horizon t (Coifman et al., 2005; Azran & Ghahramani, 2006; Lambiotte et al., 2014). Similarly, multiscale descriptions also emerge from hierarchical clustering, where t corresponds to the depth of the dendrogram (Carlsson & Mémoli, 2010; Rosvall & Bergstrom, 2011), or temporal clustering, where t corresponds to physical time (Bovet et al., 2022; Djurdjevac Conrad et al., 2025).

A natural problem is then *how to analyse and compare non-hierarchical multi-resolution sequences of partitions that are organised by the scale t.* Here we address this question from the perspective of topological data analysis (Carlsson & Zomorodian, 2009; Carlsson et al., 2009; Botnan & Lesnick, 2023) by introducing the *Multiscale Clustering Bifiltration* (MCbiF), a 2-parameter filtration of abstract simplicial complexes that encodes the patterns of cluster intersections across all scales.

Problem definition. A *partition* π of a finite set $X = \{x_1, x_2, ..., x_N\}$ is a collection of mutually exclusive subsets $C_i \subseteq X$ (called *clusters*) that cover X, i.e., $\pi = \{C_1, \ldots, C_c\}$ such that $X = \bigcup_{i=1}^{c} C_i$, and $C_i \bigcap C_j = \emptyset$, $\forall i \neq j$. The cardinality $|\pi| = c$ is the number of clusters in π. For notational convenience, we use π_i to denote the i-th cluster C_i of π. Let Π_X denote the *space of partitions* of X. We write $\pi \leq \pi'$ if every cluster in π is contained in a cluster of π'. This *refinement* relation constitutes a partial order and leads to the *partition lattice* $(\Pi_X, \leq)$ with lower bound $\hat{0} := \{\{x_1\}, \ldots, \{x_N\}\}$ and upper bound $\hat{1} := \{X\}$ (Birkhoff, 1967).

Here, we consider a *sequence of partitions* defined as a piecewise-constant function $\theta : [t_1, \infty) \to \Pi_X$, $t \mapsto \theta(t)$ such that $\theta(t) \in \Pi_X$ and the *scale index* t has M *change points* $t_1 < t_2 < ... < t_M$. We write $x \sim_t y$ if $x, y \in \theta(t)_i$, and this defines an equivalence relation whose equivalence classes $[x]_t$ are the clusters $\theta(t)_i$ of $\theta(t)$ (Brualdi, 2010). The sequence θ is called *hierarchical* if $\theta(s) \leq \theta(t)$, $\forall s \leq t$. Moreover, the sequence θ is called *coarsening* if $|\theta(s)| \geq |\theta(t)|$, $\forall s \leq t$. [1] Conversely, θ is called *fine-graining* if $|\theta(s)| \leq |\theta(t)|$, $\forall s \leq t$.

Our goal is to characterise and analyse arbitrary sequences of partitions θ, including non-hierarchical, in an integrated manner, taking account of memory effects across scales t.

Remark 1. *We are not concerned with the task of computing θ from dataset X, for which several methods exist. Rather, we take the multiscale clustering θ as a given and aim to analyse its structure.*

Remark 2. *This problem is distinct from consensus clustering, which aims to produce a summary partition by combining a set of partitions obtained, e.g., from different optimisations or clustering algorithms (Strehl & Ghosh, 2002; Vega-Pons & Ruiz-Shulcloper, 2011).*

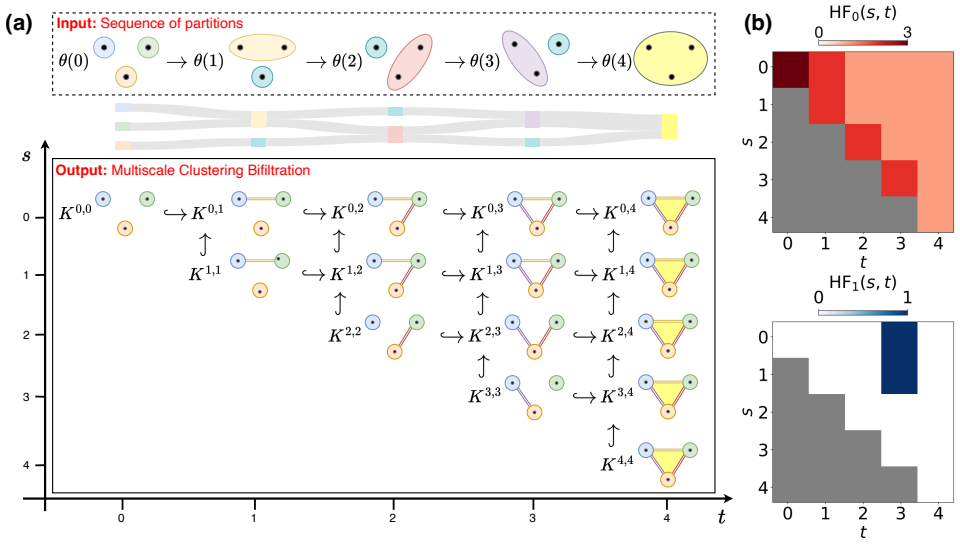

Figure 1: (a) The MCbiF $K^{s,t}$ is a bifiltration of abstract simplicial complexes that encodes the structure of the non-hierarchical sequence of partitions θ in an integrated manner. (b) The stable MCbiF Hilbert functions $\mathrm{HF}_k(s,t)$ measure the topological autocorrelation of θ, capturing nestedness at dimension $k = 0$ and higher-order cluster inconsistencies at dimension $k = 1$. These topological invariants can serve as feature maps describing θ for downstream machine learning tasks.

Contributions. To address this problem, we define the MCbiF, a bifiltration $(K^{s,t})_{t_1 \leq s \leq t}$ that aggregates the cluster intersection patterns from $\theta(s)$ to $\theta(t)$ for varying starting scale s and lag $t - s$. Using the machinery of multiparameter persistent homology (MPH) (Carlsson & Zomorodian, 2009; Carlsson et al., 2009; Botnan & Lesnick, 2023), we prove that the MCbiF leads to a block decomposable persistence module with stable Hilbert functions $\mathrm{HF}_k(s,t)$, and we show that these serve as measures of the *topological autocorrelation* of the sequence of partitions θ as induced by the scale t. In particular, $\mathrm{HF}_k(s,t)$ quantifies the non-hierarchy in θ in two complementary ways: at dimension $k = 0$ it measures the degree of nestedness of partitions, and at dimension $k = 1$ it quantifies

[1] Coarsening is equivalent to non-decreasing mean cluster size (see Remark 20 in the appendix).

the higher-order inconsistencies of cluster assignments across scales. In contrast, baseline methods such as ultrametrics (Carlsson & Mémoli, 2010) or information-based measures (Meilă, 2003) are restricted to pairwise cluster comparisons, hence cannot detect higher-order cluster inconsistencies. Furthermore, using its equivalent nerve-based construction, we show that MCbiF can be interpreted as a higher-order extension of a Sankey diagram. In the hierarchical case, the sequence of partitions reduces to a dendrogram, and the MCbiF can be obtained from the Vietoris-Rips filtration defined on the corresponding ultrametric space (Carlsson & Mémoli, 2010).

The Hilbert functions of the MCbiF provide interpretable feature maps useful in downstream machine learning tasks. In our experiments, the MCbiF feature maps outperform information-based baseline features (Meilă, 2007) on both regression and classification tasks on non-hierarchical sequences of partitions. We also show an application of MCbiF to real-world data to measure non-hierarchies in wild mice social grouping patterns across time (Bovet et al., 2022).

2 Related Work

Information-based comparison of clusterings. Information-based measures can be used to compare a pair of partitions. Assuming a uniform distribution on X, one can derive probability distributions for partitions interpreted as random variables and thus measure the information gain and loss between two partitions using the conditional entropy (CE) or the variation of information (VI), which is a metric on Π_X (Meilă, 2003; 2007). See Appendix C for detailed formulas. Extending information-based measures to more than two partitions is non-trivial. In consensus clustering, the average VI is used as a *consensus index* (CI) (Vinh & Epps, 2009; Vinh et al., 2010) for multiple partitions. However, the CI is independent of the ordering in the sequence and so cannot capture memory effects in sequences of partitions. Another limitation is that information-based measures rely only on the joint probability between pairs of random variables, hence higher-order inconsistencies between cluster assignments are not captured (see Section 5).

Dendrograms and ultrametrics. A hierarchical sequence θ with $\theta(t_1 = 0) = \hat{0}$ and $\theta(t_M) = \hat{1}$ is also called a *dendrogram* and can be represented by an acyclic rooted merge tree (Jain et al., 1999). One can define an *ultrametric* D_θ from the first-merge times, which corresponds to the depth in the dendrogram. Carlsson & Mémoli (2010) showed that there is a one-to-one correspondence between dendrograms and ultrametrics, which can be used to efficiently compare two dendrograms via the Gromov-Hausdorff distance between the ultrametric spaces (Mémoli et al., 2023). When θ is non-hierarchical, however, the first-merge times no longer define the sequence uniquely because clusters that have merged can split off again. In this case, θ cannot be represented by a tree and D_θ does not fulfil the standard triangle inequality in general. Hence, ultrametrics cannot be used to analyse and compare non-hierarchical sequences of partitions (see Section 5).

3 Background

3.1 Sankey Diagrams

Non-hierarchical sequences of partitions θ are visualised by M-layered flow graphs $S(\theta) = (V = V_1 \uplus ... \uplus V_M, E = E_1 \uplus ... \uplus E_{M-1})$ called *Sankey diagrams* (Sankey, 1898; Zarate et al., 2018), where each level $m = 1, \dots, M$ corresponds to a partition and vertices V_m represent its clusters while the undirected edges E_m between levels indicate the overlap between clusters:

$$V_m := \{(m, i) \mid 1 \leq i \leq |\theta(t_m)|\},\ E_m = \{[(m, i), (m + 1, j)] \mid \theta(t_m)_i \cap \theta(t_{m+1})_j \neq \emptyset\}\,. \quad (1)$$

If θ is hierarchical, the Sankey diagram S reduces to a dendrogram. Sankey diagrams are studied in computer graphics as they allow for the visualisation of complex relational data. The nodes in each layer V_m are vertically ordered according to a ranking $\tau_m : V_m \to \{1, \dots, |V_m|\}$, and a crossing between two edges $[u, v], [u', v'] \in E_m$ occurs if $\tau_m(u) > \tau_i(u')$ and $\tau_m(v) < \tau_m(v')$ or vice versa. Ideally, a layered layout $\tau(S) := (\tau_1, \dots, \tau_M)$ for the nodes in the Sankey diagram should minimise the crossing number (Warfield, 1977) defined as:

$$\kappa(\tau(S)) := \sum_{m=1}^{M-1} \sum_{[u,v],[u',v'] \in E_m} \left[\tau_m(u) > \tau_m(u') \wedge \tau_m(v) < \tau_m(v')\right], \quad (2)$$

where we use the Iverson bracket notation. Minimising $\kappa(\tau(S))$ is NP-complete (Garey & Johnson, 2006) and finding efficient optimisation algorithms is an active research area (Zarate et al., 2018; Li et al., 2025). We denote the minimum crossing number of $S(\theta)$ by $\kappa(\theta) := \min_\tau \kappa(\tau(S))$.

3.2 Multiparameter Persistent Homology

Multiparameter persistent homology (MPH) is an extension of standard persistent homology to $n > 1$ parameters, first introduced by Carlsson & Zomorodian (2009). We present here basic definitions, see Carlsson & Zomorodian (2009); Carlsson et al. (2009); Botnan & Lesnick (2023) for details.

Multiparameter filtration. Let us define the parameter space $(P, \leq)$ as the product of $n \geq 1$ partially ordered sets $P = P_1 \times \cdots \times P_n$, i.e., $\boldsymbol{a} \leq \boldsymbol{b}$ for $\boldsymbol{a}, \boldsymbol{b} \in P$ if and only if $\boldsymbol{a}_i \leq \boldsymbol{b}_i$ in P_i for $i = 1, \ldots, n$. Let K be a *simplicial complex* defined for the set X, such that $K \subseteq 2^X$ and K is closed under the operation of building subsets. The elements of $\sigma \in K$ are called simplices, and $\dim(K)$ is defined as the largest dimension of any simplex in K. A collection of subcomplexes $(K^{\boldsymbol{a}})_{\boldsymbol{a}\in\mathbb{R}^n}$ with $K = \bigcup_{\boldsymbol{a}\in\mathbb{R}^n} K^{\boldsymbol{a}}$ and inclusion maps $\{i_{\boldsymbol{a},\boldsymbol{b}} : K^{\boldsymbol{a}} \to K^{\boldsymbol{b}}\}_{\boldsymbol{a}\leq\boldsymbol{b}}$ that yield a commutative diagram is called a *multiparameter filtration* (or *bifiltration* for $n = 2$). We denote by $\mathrm{birth}(\sigma) \subseteq P$ the set of parameters, called *multigrades* (or *bigrades* for $n = 2$), at which simplex $\sigma \in K$ first appears in the filtration. For example, the *sublevel filtration* $K^{\boldsymbol{a}} = \{\sigma \in K \mid f(\sigma) \leq \boldsymbol{a}\}$ for a filtration function $f : K \to P$ maps each simplex σ to a unique multigrade $f(\sigma)$, i.e., $|\mathrm{birth}(\sigma)| = 1$. A filtration is called *one-critical* if it is isomorphic to a sublevel filtration, and *multi-critical* otherwise.

Multiparameter persistent homology. Let H_k for $k \in \{0, \ldots, \dim(K)\}$ denote the k-dimensional *homology functor* with coefficients in a field (Hatcher, 2002). Then H_k applied to the multiparameter filtration leads to a *multiparameter persistence module*, i.e., a collection of vector spaces $(H_k(K^{\boldsymbol{a}}))_{\boldsymbol{a}\in\mathbb{R}^n}$, which are the homology groups whose elements are the generators of k-dimensional non-bounding cycles, and linear maps $\{\imath_{\boldsymbol{a},\boldsymbol{b}} := H_k(i_{\boldsymbol{a},\boldsymbol{b}}) : H_k(K^{\boldsymbol{a}}) \to H_k(K^{\boldsymbol{b}})\}_{\boldsymbol{a}\leq\boldsymbol{b}}$ that yield a commutative diagram called *multiparameter persistent homology* (MPH). For dimension $k = 0$, H_k captures the number of disconnected components and for $k = 1$, the number of holes. Note that we recover standard persistent homology (PH) for $n = 1$ (Edelsbrunner et al., 2002).

Hilbert function. While barcodes are complete invariants of 1-parameter PH ($n = 1$), the more complicated algebraic structure of MPH ($n \geq 2$) does not allow for such simple invariants in general; hence various non-complete invariants of the MPH are used in practice. We focus on the k-dimensional *Hilbert function* (Botnan & Lesnick, 2023) defined as

$$\mathrm{HF}_k : P \to \mathbb{N}_0, \mathbf{a} \mapsto \mathrm{rank}[H_k(i_{\boldsymbol{a},\boldsymbol{a}})] = \dim[H_k(K^{\mathbf{a}})], \tag{3}$$

which maps each filtration index $\mathbf{a}$ to the k-dimensional Betti number of the corresponding complex $K^{\mathbf{a}}$. The Hilbert distance is then defined as the L_2 norm on the space of Hilbert functions and can be used to compare persistence modules.

4 The Multiscale Clustering Bifiltration (MCbiF)

The central object of our paper is a novel bifiltration of abstract simplicial complexes that encodes cluster intersection patterns in the sequence of partitions θ across the scale t.

Definition 3. *Let $\theta : [t_1, \infty) \to \Pi_X$ be a sequence of partitions. We define the Multiscale Clustering Bifiltration (MCbiF) $\mathcal{M} := (K^{s,t})_{t_1\leq s\leq t}$ as a bifiltration of abstract simplicial complexes*

$$K^{s,t} := \bigcup_{t_1\leq s\leq r\leq t} \bigcup_{C\in\theta(r)} \Delta C, \quad t_1 \leq s \leq t. \tag{4}$$

Here we interpret each cluster C as a $(|C| - 1)$-dimensional solid simplex $\Delta C := 2^C$ following Schindler & Barahona (2025), and echoing the concept of clustering as coarse-graining (Lambiotte et al., 2014; Rosvall & Bergstrom, 2011; 2008) and the interpretation of clusters as equivalence classes (Brualdi, 2010). The MCbiF aggregates all clusters from partition $\theta(s)$ to $\theta(t)$ through the union operators, and thus depends not only on the lag $|t - s|$ but also on the starting scale s to measure topological autocorrelation, see Fig. 1.

Proposition 4. *$\mathcal{M}$ is a multi-critical bifiltration uniquely defined by its values on the finite grid $P = \{(s,t) \in [t_1,\dots,t_M]\times[t_1,\dots,t_M] \mid s \le t\}$ with partial order $(s,t) \le (s',t')$ if $s \ge s', t \le t'$.*

The proof is straightforward and can be found in Appendix A. The MCbiF leads to a triangular commutative diagram where the arrows indicate inclusion maps between abstract simplicial complexes, see Fig. 1. The sequence of partitions $\theta(t)$ is encoded by the complexes $K^{t,t}$ on the diagonal of the diagram. Moving along horizontal arrows corresponds to fixing a starting scale s and θ going forward, capturing the coarsening of partitions. Moving along vertical arrows corresponds to fixing an end scale t and aggregating θ going backwards, capturing the fine-graining of partitions.

Remark 5. *By fixing $s := t_1$, which corresponds to the top row in the diagram, the MCbiF recovers the 1-parameter Multiscale Clustering Filtration (MCF) defined by Schindler & Barahona (2025). The MCF, however, can only encode coarsening sequences of partitions. For example, a large cluster $C \in \theta(s')$ prevents MCF from detecting cluster assignment conflicts between points $x, y \in C$ for $t \ge s'$, see Section 4.1. MCbiF mitigates this lack of sensitivity in MCF by varying both the starting scale s and the lag $t - s$ to encode the full topological information contained in θ.*

4.1 Measuring Topological autocorrelation with MCbiF

Applying MPH (Carlsson & Zomorodian, 2009) to the bifiltration $\mathcal{M}$ at dimensions $k \le \dim K$, for $K = K^{t_M,t_M}$, leads to a triangular diagram of simplicial complexes $H_k(K^{s,t})$ called persistence module (see Section 3.2). We show in Proposition 21 that the MCbiF persistence module is *pointwise finite-dimensional*, *finitely presentable* and *block-decomposable* (see Botnan & Lesnick (2023) for definitions), which are strong algebraic properties that guarantee algebraic stability (Bjerkevik, 2021).The proof relies on the equivalent nerve-based construction of the MCbiF (see Proposition 23), and the exactness of the persistence module from which block-decomposability follows (Cochoy & Oudot, 2020). See Appendix A for details and full proof.

The algebraic stability of MCbiF means that we can summarise the topology of MCbiF using the Hilbert functions $\mathrm{HF}_k(s,t)$ (equation 3) as simple interpretable invariants, since finite presentation property implies stability of $\mathrm{HF}_k(s,t)$ with respect to small changes in the module (Oudot & Scoccola, 2024, Corollary 8.2.).

Remark 6. *Values of $\mathrm{HF}_k(s,t)$ further away from the diagonal ($s = t$) are more robust to permutations of partitions in θ, see Proposition 22 in the appendix. In particular, $\mathrm{HF}_k(t_1,t_M)$ only depends on $\mathrm{im}(\theta)$, the set of distinct partitions, and could be used as an overall measure of consistency in θ.*

We now show how $\mathrm{HF}_k(s,t)$ can be used to detect cluster-assignment conflicts. We focus on $k = 0, 1$, for which MPH is implemented in `RIVET` (Lesnick & Wright, 2015).

Nestedness of Partitions. Hierarchy is usually understood as a refinement of partitions captured by the partition lattice. We can quantify the degree of nestedness through $\mathrm{HF}_0(s,t)$ and 0-conflicts.

Definition 7. *We say that $x_1, x_2, x_3, x_4 \in X$ lead to a 0-conflict in θ for $s \le t$ if the following two conditions hold: a) Opposite splitting: $\exists r_1, r_2 \in [s,t]$: $[x_1]_{r_1} = [x_2]_{r_1}$, $[x_3]_{r_1} \ne [x_4]_{r_1}$ and $[x_3]_{r_2} = [x_4]_{r_2}$, $[x_1]_{r_2} \ne [x_2]_{r_2}$. b) No global transitivity: $\nexists r \in [s,t]$: $[x_1]_r = [x_2]_r = [x_3]_r = [x_4]_r$.*

Proposition 8. *(i) $\mathrm{HF}_0(s,t) \le \min_{s\le r\le t} |\theta(r)|$, $\forall t_1 \le s \le t$. (ii) The following two conditions are equivalent for $s \le t$: **C1**: $\exists r \in [s,t]$ such that $\mathrm{HF}_0(s,t) = |\theta(r)|$. **C2**: $\exists r \in [s,t]$ such that $\theta(r') \le \theta(r)$, $\forall s \le r' \le t$. (iii) $\neg$ **C1** (or $\neg$ **C2**) iff there is a 0-conflict for $s < t$.*

See Appendix A for a full proof. Note that **C2** means that the subposet $\theta([s,t])$ has an upper bound.

To obtain a single measure of nestedness in θ we define the *average persistent 0-conflict* as:

$$0 \le \bar{c}_0(\theta) := 1 - \frac{2}{|t_M - t_1|^2} \int_{t_1}^{t_M} \int_{s}^{t_M} \frac{\mathrm{HF}_0(s,t)}{\mathrm{HF}_0(t,t)} \mathrm{d}s \, \mathrm{d}t \le 1. \tag{5}$$

Higher values of $\bar{c}_0(\theta)$ indicate a lower degree of nestedness. The next corollary follows immediately from Proposition 8.

Corollary 9. *(i) $\mathrm{HF}_0(t,t) = |\theta(t)|$ for all $t \ge t_1$. (ii) If θ is hierarchical on the interval $[s,t]$ then $\mathrm{HF}_0(s,t) = |\theta(t)|$. (iii) $\bar{c}_0(\theta) = 0$ iff θ is strictly hierarchical.*

Higher-order Inconsistencies between Clusters. Another way of measuring non-hierarchy is to track higher-order inconsistencies of cluster assignments across scales captured by the higher-dimensional homology groups (Schindler & Barahona, 2025). We can measure this with $\mathrm{HF}_1(s,t)$.

Proposition 10. *(i) If* $\mathrm{HF}_1(s,t) \geq 1$ *then* $\neg$ ***C1****. (ii) In particular, every homology generator* $[z] \in H_1(K^{s,t})$ *induces a 0-conflict for* $s \leq t$ *and we call this a 1-conflict. (iii) If* θ *is hierarchical on the interval* $[s,t]$ *then* $\mathrm{HF}_1(s,t) = 0$.

See Appendix A for a proof. Note that not every 0-conflict is a 1-conflict, Example 25 in the appendix. However, two 0-conflicts can lead to a 1-conflict, see Fig. 1.

Remark 11. *1-conflicts arise from higher-order cluster inconsistencies across scales. We can interpret a non-bounding cycle* $[z] \in H_1(K^{s,t})$ *as an opportunity to unify the points involved in* $[z]$ *into a single cluster. Hence, when* $[z]$ *dies in the MPH at* (s,t')*,* $t < t'$*, we say that* $\theta(t')$ *is conflict-resolving partition, see Schindler & Barahona (2025).*

To obtain a single measure of 1-conflicts in θ, we define the *average persistent 1-conflict*:

$$0 \leq \bar{c}_1(\theta) := \frac{2}{|t_M - t_1|^2} \int_{t_1}^{t_M} \int_{s}^{t_M} \mathrm{HF}_1(s,t) \mathrm{d}s \, \mathrm{d}t. \quad (6)$$

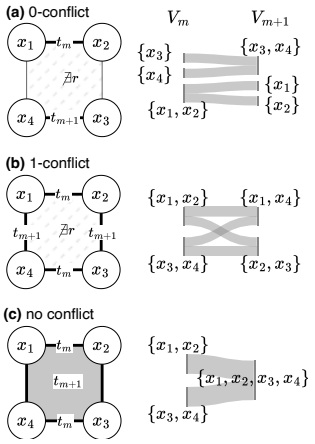

Figure 2: Relationship between conflicts and Sankey diagrams.

4.2 MCbiF as a Higher-Order Sankey Diagram

With V_m as in equation 1, we define the disjoint union $A(\ell, m) := V_\ell \uplus \ldots \uplus V_m$, $1 \leq \ell \leq m$, which assigns an index to each cluster in $\theta(t)$ for $t \in [t_\ell, t_m]$. We use this to define the nerve-based MCbiF.

Definition 12. *For* $s \in [t_\ell, t_{\ell+1}),\ \ell = 1, \ldots, M-1$*, and* $t \in [t_m, t_{m+1}),\ m = \ell, \ldots, M-1$ *or* $t \geq t_m$ *for* $m = M$ *we define the nerve-based MCbiF* $\tilde{K}^{s,t} := \{\sigma \subseteq A(\ell, m) : \bigcap_{(n,i)\in\sigma} \theta(t_n)_i \neq \emptyset\}$.

The nerve-based MCbiF $\tilde{\mathcal{M}} = (\tilde{K}^{s,t})_{t_1 \leq s \leq t}$ is a 1-critical bifiltration with simplices representing clusters and their intersections, whereas the simplices in $\mathcal{M}$ (equation 4) represent points in X and their equivalence relations. Despite these different perspectives, $\tilde{\mathcal{M}}$ and $\mathcal{M}$ lead to the same MPH and can be considered as equivalent, see Proposition 23 in the appendix, which follows from an extension of results by Schindler & Barahona (2025). However, their dimensionality can differ:

Proposition 13. *(i)* $\dim K^{s,t} = \max_{s \leq r \leq t} \max_{c \in \theta(r)} |C| - 1,\ \forall t_1 \leq s \leq t$.
(ii) $\dim \tilde{K}^{t_m, t_{m+n}} = n,\ \forall 1 \leq m \leq M, 0 \leq n \leq M - m$.

See Appendix A for the proof. The nerve-based MCbiF is computationally advantageous when $M < \max_{t_1 \leq t} \max_{C \in \theta(t)} |C| - 1$, making it the preferred choice in many applications.

Remark 14. *We can interpret* $\tilde{\mathcal{M}}$ *as a higher-order extension of the Sankey diagram* $S(\theta)$ *(equation 1) that not only records pairwise intersections between clusters in consecutive partitions of* θ *but also higher-order intersections between clusters in subsequences of* θ*. In particular, the graph* $S(\theta)$ *is a strict 1-dimensional subcomplex of* $\tilde{K}^{t_1, t_M}$ *because* $\tilde{K}^{t_m} = V_m$ *and* $\tilde{K}^{t_m, t_{m+1}} = E_m$*. Moreover, we can retrieve* $S(\theta)$ *from the zigzag filtration:*

$$\cdots \hookleftarrow \tilde{K}^{t_m, t_m} \hookrightarrow \tilde{K}^{t_m, t_{m+1}} \hookleftarrow \tilde{K}^{t_{m+1}, t_{m+1}} \hookrightarrow \cdots$$

The 0- and 1-conflicts that can arise in a single layer E_m of the Sankey diagram can be fully characterised. A 0-conflict arises when there exist nodes $u \in V_m$ and $v \in V_{m+1}$ that are both incident to at least two edges in E_m, and a 1-conflict arises when there is an even-cycle in E_m, see Fig. 2. Importantly, a 1-conflict in $K^{t_m, t_{m+1}}$ leads to a crossing in E_m that cannot be undone. This implies that the superdiagonal of HF_1 provides a lower bound for the minimal crossing number $\kappa(\theta)$ (equation 2).

Corollary 15. $\sum_{m=1}^{M-1} \mathrm{HF}_1(t_m, t_{m+1}) \leq \kappa(\theta)$.

Note that 1-conflicts arising through multiple partitions not necessarily lead to crossings, see Fig. 1. However, we hypothesise that the full HF_0 and HF_1 feature maps give insights into more complicated crossings that arise in the Sankey layout due to interactions between different layers.

5 MATHEMATICAL LINKS OF MCBIF TO OTHER METHODS

MCbiF has links and provides generalisations to the related methods introduced in Section 2.

Ultrametrics. Given θ, define the matrix of first-merge times conditional on starting scale s:

$$D_{\theta,s}(x_i, x_j) := \min\{t \geq s \mid \exists\, C \in \theta(t) : x_i, x_j \in C\}.$$

For $s = 0$, we recover the standard matrix of first-merge times $D_\theta := D_{\theta,0}$. If θ is hierarchical with $\theta(t_1 = 0) = \hat{0}$ and $\theta(t_M) = \hat{1}$ then D_θ is an ultrametric, i.e., it fulfils the *strong triangle inequality* $D_\theta(x, z) \leq \max\left(D_\theta(x, y), D_\theta(x, z)\right) \forall x, y, z \in X$, and Proposition 8 implies that the number of branches in the dendrogram at level t is equal to $\mathrm{HF}_0(s, t)$ for $s \leq t$. Hence, $\mathrm{HF}_0(s, t)$ contains the same information as the ultrametric in the hierarchical case, see also Schindler & Barahona (2025). If θ is non-hierarchical, 0-conflicts can lead to violations of the (strong) triangle inequality.

Proposition 16. *D_θ is not an ultrametric if $\exists x_1, x_2, x_3 \in X$ that lead to a 0-conflict.*

See Appendix A for a proof. This proposition shows that $\bar{c}_0(\theta)$ can be interpreted as a measure of how much the ultrametric property of D_θ is violated. Recall that, although not a metric, $D_{\theta,s}$ is a *dissimilarity measure* that can be used to define a filtration (Chazal et al., 2014). Extending results from Schindler & Barahona (2025), it can be shown that a 2-parameter Vietoris-Rips filtration from $D_{\theta,s}$ has the same zero-dimensional MPH as MCbiF but has a trivial higher-dimensional MPH, see Proposition 24. This suggests that the pairwise merge times $D_{\theta,s}$ can capture the nestedness of partitions but not the inconsistencies that arise through higher-order interactions between clusters.

Conditional Entropy. CE is only defined for pairs of partitions ($\theta(s)$,$\theta(t)$) and can be computed as the expected information of the *conditional probability* of $\theta(t)$ given $\theta(s)$, denoted by $P_{t|s}$, see Appendix C. For the special case of $M = 2$, $\mathrm{HF}_0(t_1, t_2)$ follows directly from the spectral properties of the matrix $P_{t_2|t_1} P_{t_2|t_1}^T$, which can be interpreted as an undirected graph.

Proposition 17. $\mathrm{HF}_0(t_1, t_2) = \dim(\ker L)$ *for graph Laplacian* $L := \mathrm{diag}(P_{t_2|t_1}\mathbf{1}) - P_{t_2|t_1} P_{t_2|t_1}^T$.

The proof follows from Proposition 23, which shows that $P_{t_2|t_1} P_{t_2|t_1}^T$ and $\tilde{K}^{t_1,t_2}$ have the same 1-simplices and $\dim(\ker L)$ is equal to the number of connected graph components (Chung, 1997). Note that $P_{t|s}$ only encodes the pairwise relationship between clusters, and does not capture higher-order inconsistencies between cluster assignments. In particular, CE cannot detect 1-conflicts arising across more than two scales, see Example 26 in appendix.

6 EXPERIMENTS

6.1 REGRESSION TASK: MINIMAL CROSSING NUMBER

In our first experiment, we predict the minimal crossing number $\kappa(\theta)$ (equation 2). This task is of relevance in computer graphics and related to the NP-complete minimisation of the crossing number in Sankey diagram layouts (Zarate et al., 2018; Li et al., 2025).

Definition 18. *Let Π_N^M denote the space of coarsening sequences $\theta : [0, \infty) \to \Pi_X$ with $|X| = N$ so that $\theta(0) = \hat{0}$, $\theta(M-1) = \hat{1}$ and the scale t has M change points $t_m = m - 1,\ \forall 1 \leq m \leq M$.*

For $M = 20$, we define two datasets by sampling 20,000 random samples $\theta \in \Pi_N^M$ for $N = 5$ and $N = 10$, respectively. For each θ, we compute the information-based pairwise conditional entropy matrix CE (Meilă, 2003) (see equation 10) and our MCbiF Hilbert functions (HF_0 and HF_1) as feature maps. We define the minimal crossing number $y = \kappa(\theta)$ as our prediction target, which we computed with the `OmicsSankey` algorithm (Li et al., 2025). See Appendix D.1 for details. Predicting y is harder for $N = 10$ because the increased complexity of Π_M^N allows for more complicated crossings.

We first compute the Pearson correlation, r, between y, CI (given by the average CE, see equation 11), $\bar{c}_0$ and $\bar{c}_1$. The correlation between CI and y is low ($r = 0.20$ for $N = 5$ and $r = 0.06$ for $N = 10$), higher for $\bar{c}_0$, and highest for $\bar{c}_1$ ($r = 0.47$ for $N = 5, 10$) (see Fig. 6 in appendix). This is

Table 1: Regression task. Test R^2 score of LR, CNN and MLP models trained on different features for $N = 5$ and $N = 10$. See Appendix D.1 for train R^2 scores.

N	Method	Raw θ	HF_0	HF_1	HF_0 & HF_1	CE
5	LR	0.001	0.147	0.486	0.539	0.392
	CNN	-0.006	0.155	0.504	**0.544**	0.492
	MLP	-0.002	0.150	0.491	0.541	0.409
10	LR	-0.012	0.214	0.448	**0.516**	0.457
	CNN	0.000	0.211	0.448	0.507	0.454
	MLP	0.000	0.212	0.450	0.514	0.458

consistent with our theoretical results that link the crossing number with HF_1, see Section 4.2. For the regression task, we split each dataset into train (64%), validation (16%) and test (20%). For each feature map (or their combinations), we train three different models: linear regression (LR), multi-layer perceptron (MLP), and convolutional neural network (CNN).We use the mean-squared error (MSE) as our loss function and the validation set for hyperparameter tuning. See Appendix D.1 for details. We evaluate the model performances on the *unseen test data* using the coefficient of determination (R^2) and observe that MCbiF features outperform information-based features (Table 1). In particular, the combined HF_0 and HF_1 features lead to a significantly better model performance than CE ($p < 0.0001$, t-test). Moreover, HF_0 and HF_1 yield $R^2 = 0.544$ for $N = 5$ and $R^2 = 0.516$ for $N = 10$ whereas CE only achieves $R^2 = 0.492$ and $R^2 = 0.458$, respectively. The strong performance of LR demonstrates the interpretability of the MCbiF features, important for explainable AI (XAI) (Adadi & Berrada, 2018).

6.2 Classification Task: Order-Preservation

In our second experiment, we classify whether a sequence θ is order-preserving or not. This task is of relevance in utility theory and the study of preference relations (Roberts, 2009).

Definition 19. *A partition $\theta(t_m)$ equipped with a total order $<_m$ is called an ordered partition and induces a total preorder $\lesssim_m$ on X (Stanley, 2011), i.e., $x \lesssim_m y$ if $[x]_t <_m [y]_t$. We call θ order-preserving if there exist total orders $(<_1, \ldots, <_M)$ such that the total preorders $(\lesssim_1, \ldots, \lesssim_M)$ are compatible, i.e., $\forall \ell, m$ we have $x \lesssim_\ell y$ iff $x \lesssim_m y$, $\forall x, y \in X$.*

We generate a balanced dataset of 3,700 partitions $\theta \in \Pi_{500}^{30}$ with 50% of the sequences order-preserving ($y = 0$) and 50% not order-preserving ($y = 1$), see Appendix D.2 for details. We choose $N = 500$ and $M = 30$ to demonstrate the scalability of our method. For each θ we compute CE and HF_k, $k = 0, 1$, using the computationally advantageous nerve-based MCbiF.

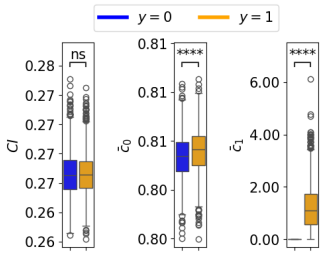

Figure 3: Difference between order-preserving ($y = 1$) and not order-preserving ($y = 0$) sequences (**** indicates $p < 0.0001$, Mann- U test).

Table 2: Classification task. Test accuracy of logistic regression trained on different features.

Raw θ	HF_0	HF_1	CE
0.53	0.56	**0.97**	0.50

Whereas there is no significant difference between the two classes for CI, we observe a statistically significant increase of $\bar{c}_0$ and $\bar{c}_1$ for order-preserving sequences (Fig. 3). For the classification task, we split our data into train (80%) and test (60%). We train logistic regression on each feature map separately, and evaluate the accuracy on the test split, see Appendix D.2. We observe that HF_1 predicts y with high accuracy (0.95) in contrast to CE that cannot improve on a random classifier (Table 2). Our results demonstrate the high sensitivity of MCbiF to order-preservation in θ.

6.3 Application to Real-world Temporal Data

In our final experiment, we apply MCbiF to temporal sequences of partitions θ_τ computed from real-world contact data of free-ranging house mice (Bovet et al., 2022). Each partition $\theta_\tau(t)$ describes mice social groupings at week $t \in [1, \ldots, 9]$ and the scale t corresponds here to the nine weeks in the study period (28 February-1 May 2017), throughout which $N = 281$ individual mice were present. Each sequence captures the fine-graining of social groups over the transition from winter to

spring, and larger values of the temporal resolution $\tau > 0$ lead to finer temporal community structure (Fig. 8). We use MCbiF to compare the temporal sequences θ_{τ_i} for nine different resolutions τ_i, $i = 1, \ldots, 9$, as provided inBovet et al. (2022). See Appendix D.3 for details.

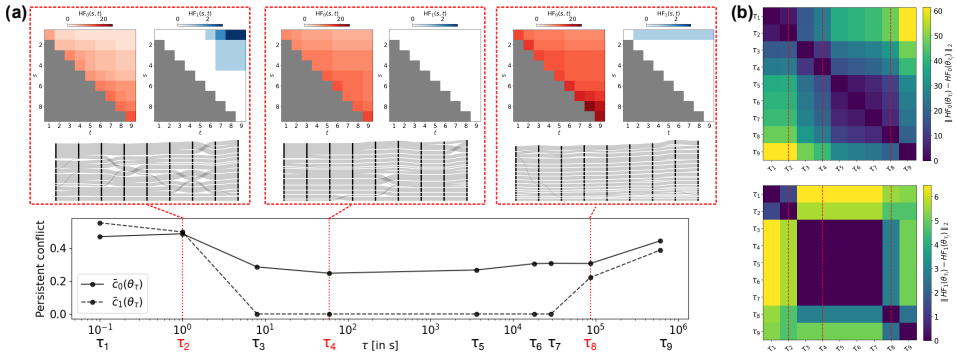

Figure 4: (a) Persistent conflict in temporal sequences $\theta_{\tau_i}(t)$ for different levels of temporal resolution τ_i from mice social data. Vertical red lines indicate robust τ_i identified by Bovet et al. (2022), for which we display Sankey diagrams and MCbiF features. (b) The robust τ_i correspond to stable temporal regimes captured by the block structure in the MCbiF Hilbert distances.

Bovet et al. (2022) identified that the temporal resolutions $\tau_2 = 1\,\mathrm{s}$, $\tau_4 = 60\,\mathrm{s}$ and $\tau_8 = 24\,\mathrm{h}$ lead to robust sequences of partitions. Using the Hilbert distance, i.e., the L_2-norm on the zero- and one-dimensional MCbiF Hilbert functions, we find these temporal resolutions to be representative for three distinct temporal regimes, which are characterised by different degrees of non-hierarchy as measured by $\bar{c}_0$ and $\bar{c}_1$ (Fig. 4). In particular, high $\bar{c}_0$ indicates that mice tend to split off groups over time, and high $\bar{c}_1$ indicates that mice meet in overlapping subgroups but never jointly in one nest box. Note that θ_{τ_2} has strong non-hierarchies because the large-scale mice social clusters get disrupted in the transition to spring. In contrast, θ_{τ_8} is more hierarchical as it captures the underlying stable social groups revealed by the higher temporal resolution. Finally, the Hilbert distances also capture an increased time reversibility in the sequence θ_{τ_8} due to the increased stability of social groupings over time, see Fig. 9 in the appendix.

7 Conclusion

We introduce the MCbiF, a novel bifiltration that encodes the cluster intersection patterns of multiscale, non-hierarchical sequences of partitions, θ. Its stable Hilbert functions quantify the topological autocorrelation of θ by measuring nestedness of partitions and higher-order cluster inconsistencies. Our measures of persistent conflict $\bar{c}_0(\theta)$ and $\bar{c}_1(\theta)$ are global, history-dependent and sensitive to the ordering of the partitions in θ. The MCbiF extends the 1-parameter MCF defined by Schindler & Barahona (2025) to a 2-parameter filtration, leading to richer algebraic invariants that describe the full topological information in θ. We demonstrate with numerical experiments that the MCbiF Hilbert functions provide topological feature maps for downstream machine learning tasks, which outperform information-based features on regression and classification tasks on non-hierarchical sequences of partitions. Moreover, the grounding of MCbiF features in algebraic topology enhances interpretability, a crucial attribute for XAI and applications to real-world data.

Limitations and future work Our analysis of the MCbiF MPH is restricted to dimensions zero and one due to limitations of the `RIVET` software Lesnick & Wright (2015) used in our numerical experiments. However, analysing topological autocorrelation for higher dimensions would allow us to capture more complex higher-order cluster inconsistencies. We also focused on the Hilbert functions of MCbiF as our topological invariants because of their computational efficiency and analytical simplicity, which facilitates our theoretical analysis. In future work, we plan to use algebraically richer feature maps by exploiting the block decomposition of the MCbiF persistence module, which leads to barcodes Bjerkevik (2021), or by using multiparameter persistence landscapes (Vipond, 2020). Finally, we plan to analyse minimal cycle representatives of the MPH (Li et al., 2021) to localise 0 and 1-conflicts in the sequence of partitions, which is of interest in many applications.

Reproducibility Statement

Detailed proofs of all theoretical results can be found in Appendix A and extensive documentation of our experiments in Appendix D. The dataset studied in Section 6.3 is publicly available at: `https://dataverse.harvard.edu/file.xhtml?fileId=5657692`. We will release code upon publication.

References

Amina Adadi and Mohammed Berrada. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2870052.

Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587 (7834):402–407, November 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2909-1.

M. Tarik Altuncu, Erik Mayer, Sophia N. Yaliraki, and Mauricio Barahona. From free text to clusters of content in health records: An unsupervised graph partitioning approach. *Applied Network Science*, 4(1):23, December 2019. ISSN 2364-8228. doi: 10.1007/s41109-018-0109-9.

A. Azran and Z. Ghahramani. Spectral Methods for Automatic Multiscale Data Clustering. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pp. 190–197, June 2006. doi: 10.1109/CVPR.2006.289.

Mariano Beguerisse-Díaz, Amy K. McLennan, Guillermo Garduño-Hernández, Mauricio Barahona, and Stanley J. Ulijaszek. The 'who' and 'what' of #diabetes on Twitter. *DIGITAL HEALTH*, 3, January 2017. ISSN 2055-2076, 2055-2076. doi: 10.1177/2055207616688841.

Garrett Birkhoff. *Lattice Theory*. Number 25 in Colloquium Publications / American Mathematical Society. American Mathematical Society, Providence, RI, 3. ed. edition, 1967. ISBN 978-0-8218-1025-5.

Håvard Bakke Bjerkevik. On the Stability of Interval Decomposable Persistence Modules. *Discrete Comput. Geom.*, 66(1):92–121, July 2021. ISSN 0179-5376. doi: 10.1007/s00454-021-00298-0.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, March 2003. ISSN 1532-4435.

Magnus Bakke Botnan and Michael Lesnick. An introduction to multiparameter persistence. In Aslak Bakke Buan, Henning Krause, and Øyvind Solberg (eds.), *EMS Series of Congress Reports*, volume 19, pp. 77–150. EMS Press, 1 edition, November 2023. ISBN 978-3-98547-054-9 978-3-98547-554-4. doi: 10.4171/ecr/19/4.

Alexandre Bovet, Jean-Charles Delvenne, and Renaud Lambiotte. Flow stability for dynamic community detection. *Science Advances*, 8(19):eabj3063, May 2022. doi: 10.1126/sciadv.abj3063.

Richard A. Brualdi. *Introductory Combinatorics*. Pearson/Prentice Hall, Upper Saddle River, N.J, 5th ed edition, 2010. ISBN 978-0-13-602040-0.

Gunnar Carlsson and Facundo Mémoli. Characterization, Stability and Convergence of Hierarchical Clustering Methods. *Journal of Machine Learning Research*, 11(47):1425–1470, 2010. ISSN 1533-7928.

Gunnar Carlsson and Afra Zomorodian. The Theory of Multidimensional Persistence. *Discrete & Computational Geometry*, 42(1):71–93, July 2009. ISSN 1432-0444. doi: 10.1007/s00454-009-9176-0.

Gunnar Carlsson, Gurjeet Singh, and Afra Zomorodian. Computing Multidimensional Persistence. In Yingfei Dong, Ding-Zhu Du, and Oscar Ibarra (eds.), *Algorithms and Computation*, pp. 730–739, Berlin, Heidelberg, 2009. Springer. ISBN 978-3-642-10631-6. doi: 10.1007/978-3-642-10631-6_74.

Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, November 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1313480110.

Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, December 2014. ISSN 0046-5755, 1572-9168. doi: 10.1007/s10711-013-9937-z.

Fan R. K. Chung. *Spectral Graph Theory*. Number no. 92 in Regional Conference Series in Mathematics. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I, 1997. ISBN 978-0-8218-0315-8.

Jérémy Cochoy and Steve Oudot. Decomposition of Exact pfd Persistence Bimodules. *Discrete & Computational Geometry*, 63(2):255–293, March 2020. ISSN 1432-0444. doi: 10.1007/s00454-019-00165-z.

R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, May 2005. doi: 10.1073/pnas.0500334102.

A. Delmotte, E. W. Tate, S. N. Yaliraki, and M. Barahona. Protein multi-scale organization through graph partitioning and robustness analysis: Application to the myosin–myosin light chain interaction. *Physical Biology*, 8(5):055010, August 2011. ISSN 1478-3975. doi: 10.1088/1478-3975/8/5/055010.

Jean-Charles Delvenne, Sophia N. Yaliraki, and Mauricio Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, 107(29):12755–12760, July 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0903215107.

Nataša Djurdjevac Conrad, Elisa Tonello, Johannes Zonker, and Heike Siebert. Detection of dynamic communities in temporal networks with sparse data. *Applied Network Science*, 10(1):1, January 2025. ISSN 2364-8228. doi: 10.1007/s41109-024-00687-3.

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological Persistence and Simplification. *Discrete & Computational Geometry*, 28(4):511–533, November 2002. ISSN 1432-0444. doi: 10.1007/s00454-002-2885-2.

M. R. Garey and D. S. Johnson. Crossing Number is NP-Complete. *SIAM Journal on Algebraic Discrete Methods*, July 2006. doi: 10.1137/0604033.

Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. (arXiv:2203.05794), March 2022. doi: 10.48550/arXiv.2203.05794.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge ; New York, 2002. ISBN 978-0-521-79160-1 978-0-521-79540-1.

Renee S. Hoekzema, Lewis Marsh, Otto Sumray, Thomas M. Carroll, Xin Lu, Helen M. Byrne, and Heather A. Harrington. Multiscale Methods for Signal Selection in Single-Cell Data. *Entropy*, 24 (8):1116, August 2022. ISSN 1099-4300. doi: 10.3390/e24081116.

A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31 (3):264–323, September 1999. ISSN 0360-0300. doi: 10.1145/331499.331504.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017.

Renaud Lambiotte, Jean-Charles Delvenne, and Mauricio Barahona. Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. *IEEE Transactions on Network Science and Engineering*, 1(2):76–90, July 2014. ISSN 2327-4697. doi: 10.1109/TNSE.2015.2391998.

Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*, pp. 255–258. MIT Press, Cambridge, MA, USA, October 1998. ISBN 978-0-262-51102-5.

Michael Lesnick and Matthew Wright. Interactive Visualization of 2-D Persistence Modules, December 2015.

Lu Li, Connor Thompson, Gregory Henselman-Petrusek, Chad Giusti, and Lori Ziegelmeier. Minimal Cycle Representatives in Persistent Homology Using Linear Programming: An Empirical Study With User's Guide. *Frontiers in Artificial Intelligence*, 4, 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.681117.

Shiying Li, Bowen Tan, Si Ouyang, Zhao Ling, Miaozhe Huo, Tongfei Shen, Jingwan Wang, and Xikang Feng. OmicsSankey: Crossing Reduction of Sankey Diagram on Omics Data, June 2025.

Marina Meilă. Comparing Clusterings by the Variation of Information. In Bernhard Schölkopf and Manfred K. Warmuth (eds.), *Learning Theory and Kernel Machines*, Lecture Notes in Computer Science, pp. 173–187, Berlin, Heidelberg, 2003. Springer. ISBN 978-3-540-45167-9. doi: 10.1007/978-3-540-45167-9_14.

Marina Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, May 2007. ISSN 0047-259X. doi: 10.1016/j.jmva.2006.11.013.

Facundo Mémoli, Zane Smith, and Zhengchao Wan. The Gromov-Hausdorff distance between ultrametric spaces: Its structure and computation. *Journal of Computational Geometry*, pp. 78–143 Pages, September 2023. doi: 10.20382/JOCG.V14I1A4.

Steve Oudot and Luis Scoccola. On the Stability of Multigraded Betti Numbers and Hilbert Functions. *SIAM Journal on Applied Algebra and Geometry*, 8(1):54–88, March 2024. doi: 10.1137/22M1489150.

Fred S. Roberts. *Measurement Theory: With Applications to Decisionmaking, Utility and the Social Sciences*. Number 7 in Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge, digitally printed version edition, 2009. ISBN 978-0-521-10243-8 978-0-521-30227-2.

Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, January 2008. doi: 10.1073/pnas.0706851105.

Martin Rosvall and Carl T. Bergstrom. Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems. *PLOS ONE*, 6(4):e18209, April 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0018209.

Matthew Sankey. Introductory note on the thermal efficiency of steam-engines. In *Minutes of Proceedings of the Institution of Civil Engineers*, volume 134, pp. 278–283, 1898.

Juni Schindler and Mauricio Barahona. Analysing Multiscale Clusterings with Persistent Homology, April 2025.

Juni Schindler, Jonathan Clarke, and Mauricio Barahona. Multiscale mobility patterns and the restriction of human movement. *Royal Society Open Science*, 10(10):230405, October 2023. doi: 10.1098/rsos.230405.

Richard P. Stanley. *Enumerative Combinatorics. Volume 1*. Number 49 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, NY, 2nd ed edition, 2011. ISBN 978-1-107-01542-5 978-1-107-60262-5.

Alexander Strehl and Joydeep Ghosh. Cluster Ensembles — A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3(Dec):583–617, 2002. ISSN ISSN 1533-7928.

Sandro Vega-Pons and José Ruiz-Shulcloper. A SURVEY OF CLUSTERING ENSEMBLE ALGORITHMS. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03): 337–372, May 2011. ISSN 0218-0014, 1793-6381. doi: 10.1142/S0218001411008683.

Nguyen Xuan Vinh and Julien Epps. A Novel Approach for Automatic Number of Clusters Detection in Microarray Data Based on Consensus Clustering. In *2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering*, pp. 84–91, June 2009. doi: 10.1109/BIBE.2009.19.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *The Journal of Machine Learning Research*, 11:2837–2854, December 2010. ISSN 1532-4435.

Oliver Vipond. Multiparameter Persistence Landscapes. *Journal of Machine Learning Research*, 21 (61):1–38, 2020. ISSN 1533-7928.

John N. Warfield. Crossing Theory and Hierarchy Mapping. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(7):505–523, July 1977. ISSN 2168-2909. doi: 10.1109/TSMC.1977.4309760.

David Cheng Zarate, Pierre Le Bodic, Tim Dwyer, Graeme Gange, and Peter Stuckey. Optimal Sankey Diagrams Via Integer Programming. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 135–139, April 2018. doi: 10.1109/PacificVis.2018.00025.

A Proofs of Theoretical Results

Remark 20. *It is a simple fact that θ is coarsening iff $\frac{1}{|\theta(s)|}\sum_i^{|\theta(s)|}|\theta(s)_i| \leq \frac{1}{|\theta(t)|}\sum_j^{|\theta(t)|}|\theta(t)_j|$ for $s \leq t$. The proof follows directly from the fact that $\sum_i^{|\theta(s)|}|\theta(s)_i| = \sum_i^{|\theta(t)|}|\theta(t)_i| = N$.*

We provide a proof for the multi-criticality of the MCbiF filtration stated in Proposition 4.

Proof of Proposition 4. The MCbiF is indeed a bifiltration because $K^{s,t} \subseteq K^{s',t'}$ if $s \geq s'$ and $t \leq t'$. It is uniquely defined by its values on the finite grid $[t_1, \ldots, t_M] \times [t_1, \ldots, t_M]$ because θ has change points $t_1 < \cdots < t_M$. It is a multi-critical bifiltration because for $x \in X$ we have $[x] \in K^{s,t}$ for all $s, t \in [t_1, \infty)^{\mathrm{op}} \times [t_1, \infty)$. In particular, $x \in K^{t_1,t_1}$ and $x \in K^{t_1+\delta,t_1+\delta}$ for $\delta > 0$ but (t_1, t_1) and $(t_1 + \delta, t_1 + \delta)$ are incomparable in the poset $[t_1, \infty)^{\mathrm{op}} \times [t_1, \infty)$. □

Next, we provide the proof for Proposition 21, which shows that the MCbiF persistence module is pointwise finite-dimensional, finitely presented and block-decomposable (see Botnan & Lesnick (2023) for definitions).

Proposition 21. *For any $k \leq \dim K$, the MCbif persistence module $H_k(K^{s,t})$ is pointwise finite-dimensional, finitely presented and block-decomposable.*

Proof of Proposition 21. We prove the proposition for the nerve-based MCbiF, which leads to the same persistence module. As MCbiF is pointwise finite-dimensional and defined on a finite grid, we can use Theorem 9.6 by Cochoy & Oudot (2020) that implies block-decomposability of the persistence module if it is *exact*. Hence, it suffices to show that for all $t \leq t' \leq t'' \leq t'''$ the diagram

$$\begin{array}{ccc} H_k(\tilde{K}^{t,t''}) & \longrightarrow & H_k(\tilde{K}^{t,t'''}) \\ \uparrow & & \uparrow \\ H_k(\tilde{K}^{t',t''}) & \to & H_k(\tilde{K}^{t',t'''}) \end{array}$$

induces an exact sequence:

$$H_k(\tilde{K}^{t',t''}) \to H_k(\tilde{K}^{t,t''}) \oplus H_k(\tilde{K}^{t',t'''}) \to H_k(\tilde{K}^{t,t'''}) \tag{7}$$

By construction of the MCbiF, $\tilde{K}^{t,t'''} = \tilde{K}^{t,t''} \cup \tilde{K}^{t',t'''}$. Furthermore, $\tilde{K}^{t,t''} = \tilde{K}^{t,t'} \cup \tilde{K}^{t',t''}$ and $\tilde{K}^{t',t'''} = \tilde{K}^{t',t''} \cup \tilde{K}^{t'',t'''}$. Without loss of generality, $t = t_k$, $t' = t_\ell$, $t'' = t_m$, $t''' = t_n$ for change

points $t_k < t_\ell < t_m < t_n$ of θ such that $A(k,\ell) \cap A(m,n) = \emptyset$. Hence, $\tilde{K}^{t,t'} \cap \tilde{K}^{t'',t'''} = \emptyset$ and $\tilde{K}^{t',t''} = \tilde{K}^{t,t''} \cap \tilde{K}^{t',t'''}$. This means that equation 7 is a Mayer-Vietoris sequence for all $k \geq 0$, implying exactness (Hatcher, 2002) and proving the block decomposability (Cochoy & Oudot, 2020). As the MCbiF is defined uniquely by its values on a finite grid (Proposition 4), its persistence module decomposes into finitely many blocks and is therefore finitely presented. □

It follows from the construction of MCbiF that the Hilbert functions are invariant to certain permutations of partitions in θ.

Proposition 22. $\mathrm{HF}_k(s,t)$ *is invariant to permutations of the sequence between* s *and* t.

Proof. The proof follows directly from the construction of MCbiF, see equation 4. □

We now provide a proof for Proposition 8 on properties of the zero-dimensional Hilbert function of the MCbiF.

Proof of Proposition 8. Statement (i) follows directly from Proposition 12 in Schindler & Barahona (2025). To prove (ii), consider first that **C2** is true and $\theta(r)$ is an upper bound for the partitions $\theta(r')$, $s \leq r' \leq t$. Then all simplices in $K^{s,t}$ are nested within $K^{r,r}$, which has $|\theta(r)|$ disconnected components, i.e., $\mathrm{HF}_0 = |\theta(r)|$ implying **C1**. We prove the other direction by contradiction. Assume that **C1** is true but **C2** is false. Then there exists $\theta(r')$ for $s \leq r' \leq t$ such that either $\theta(r)$ and $\theta(r')$ are not comparable in the partition lattice, otherwise $\theta(r) < \theta(r')$ contradicting **C1**. This implies that there are points $x, y \in X$ such that $x, y \in C'$ for a cluster $C' \in \theta(r')$ but there exists no cluster $C \in \theta(r)$ such that $x, y \in C$. This implies that $K^{r,r} \cup \Delta C' \subseteq K^{s,t}$ has strictly fewer connected components than $K^{r,r}$, i.e., $\mathrm{HF}_0(s,t) < |\theta(r)|$ contradicting **C1**.

To prove (iii), we first show that $\neg$ **C2** implies that there is a 0-conflict. If $\neg$ **C2**, $\exists r_1, r_2 \in [s,t]$ such that $\theta(r_1) \not\leq \theta(r_2)$ and $\theta(r_2) \not\leq \theta(r_1)$, otherwise there would be a total order on $\theta([s,t])$ with an upper bound. Hence, $\exists C \in \theta(r_1)$ such that C is split in $\theta(r_2)$ and $\exists C' \in \theta(r_2)$ such that C' is split in $\theta(r_1)$. This means there are $x_1, x_2 \in C$ and $x_3, x_4 \in C'$ such that $[x_1]_{r_1} = [x_2]_{r_1}$, $[x_3]_{r_1} \neq [x_4]_{r_1}$ and $[x_3]_{r_2} = [x_4]_{r_2}$, $[x_1]_{r_2} \neq [x_2]_{r_2}$. Without loss of generality, $\nexists C'' \in \theta(r)$ for any $r \in [s,t]$ such that $C \cup C' \subseteq C''$, otherwise $\theta([s,t])$ would have an upper bound. Hence, $\nexists r \in [s,t]$: $[x_1]_r = [x_2]_r = [x_3]_r = [x_4]_r$. This shows there is a 0-conflict. To show the other direction, assume there is a 0-conflict. Then the opposite splitting implies $\theta(r_1) \not\leq \theta(r_2)$ and $\theta(r_2) \not\leq \theta(r_1)$ and the lack of global transitivity implies $\nexists r \in [s,t]$ such that $\theta(r_1) \leq \theta(r)$ and $\theta(r_2) \leq \theta(r)$. This shows **C2**, □

We can now prove Proposition 10 on 1-conflicts.

Proof of Proposition 10. (i) We show that **C1** implies $\mathrm{HF}_0(s,t) = 0$. As **C1** and **C2** are equivalent, it suffices to show the Proposition for **C2**. Let us assume that $\theta(r)$ is an upper bound for the partitions $\theta(r')$, $s \leq r' \leq t$. Then all simplices in $K^{s,t}$ are nested within $K^{r,r}$, which is a disjoint union of solid simplices that all have trivial higher-dimensional homology. Hence, $\mathrm{HF}_1(s,t) = \mathrm{HF}_1(r,r) = 0$. (ii) If $\mathrm{HF}_1(s,t) \geq 1$ there exists a 1-dimensional chain $z = [x_1, x_2] + \cdots + [x_{n-1}, x_n] + [x_n, x_1]$ that is non-bounding, i.e., $[z] \neq 0$ in $H_1(K^{s,t})$. If there was a cluster $C \in \theta(r)$ for some $s \leq r \leq t$ such that $x_1, \ldots, x_n \in C$ then there would be a two-dimensional chain of which z is the boundary, contradicting $[z] \neq 0$. (iii) If θ is hierarchical, then **C2** is true and the statement follows from (i). □

Next, we provide the proof about the equivalence between MCbiF and nerve-based MCbiF.

Proposition 23. *The bifiltrations* $\mathcal{M}$ *and* $\tilde{\mathcal{M}}$ *lead to the same persistence module.*

Proof. The proof follows from Proposition 30 in Schindler & Barahona (2025), which extends directly to the 2-parameter case. □

Next, we prove Proposition 13 about the dimension of the nerve-based MCbiF.

Proof of Proposition 13. Statement (i) follows directly from the definition in equation 4. We show statement (ii) by induction. From the definition of the nerve-based MCbiF, it follows directly that $\dim N^{t_m,t_m} = 1$ because the indices in $A(m,m)$ correspond to mutually exclusive clusters. Let us assume that $\dim N^{t_m,t_{m+n}} = n$, then there exist $C_1,\dots,C_n \in \theta([t_m,t_{m+n}])$ such that $C_1 \cap \dots \cap C_n \neq \emptyset$. As the clusters in partition $\theta(t_{m+n+1})$ cover the set X there exist a cluster $C \in \theta(t_{m+n+1})$ such that $C \cap C_1 \cap \dots \cap C_n \neq \emptyset$. Hence, $\dim N^{t_m,t_{m+n}} \geq n+1$. If $\dim N^{t_m,t_{m+n}} > n+1$ there would exist a second cluster $C' \in \theta(t_{m+n+1})$ with $C' \cap C \cap C_1 \cap \dots \cap C_n \neq \emptyset$ but $C' \cap C \neq \emptyset$ contradicts that clusters of $\theta(t_{m+n+1})$ are mutually exclusive. Hence, $\dim N^{t_m,t_{m+n}} = n+1$, proving statement (ii) by induction. □

We continue by proving that 0-conflicts can induce violations of the strong triangle inequality as stated in Proposition 16.

Proof of Proposition 16. If $\exists x_1,x_2,x_3 \in X$ that induce a 0-conflict for $s \leq t$ then $\exists r_1,r_2 \in [s,t]$ such that $[x_1]_{r_1} = [x_2]_{r_1}$ and $[x_2]_{r_2} = [x_3]_{r_2}$. Without loss of generality, $r_1 < r_2$ and suppose there is $r_3 > r_2$ such that $[x_1]_{r_3} = [x_3]_{r_3}$. Then $D_\theta(x_1,x_3) > \max(D_\theta(x_1,x_2), D_\theta(x_2,x_3))$.

□

Finally, we show that the zero-dimensional MPH of MCbiF corresponds to the zero-dimensional MPH of a Vietoris-Rips bifiltration constructed from $D_{\theta,s}$.

Proposition 24. *Let us define the Vietoris-Rips bifiltration $\mathcal{L} = (L^{s,t})_{t_1 \leq s \leq t}$ based on $D_{\theta,s}$ as*

$$L^{s,t} = \{\sigma \subset X \mid \forall x,y \in \sigma : \ D_{\theta,s}(x,y) \leq t\}. \tag{8}$$

Then the zero-dimensional MPH of $\mathcal{L}$ and of the MCbiF $\mathcal{M}$ are equivalent. However, the higher-dimensional MPH of $\mathcal{L}$ is always trivial.

Proof. The proof follows from a simple extension of Proposition 32 in Schindler & Barahona (2025) to the 2-parameter case. □

B Toy Examples

The first example illustrates the difference between 0- and 1-conflicts.

Example 25. *Let $x_1,x_2,x_3,x_4 \in X$ such that $x_1 \sim_t x_2 \sim_{t'} x_3 \sim_{t''} x_4$ for $t_1 \leq t < t' < t''$. If $x_i \neq x_j$ for $i,j = 1,2,3$ then we know that the partitions $\theta(t)$ and $\theta(t')$ are not nested, which we can measure with $\mathrm{HF}_0(t,t') < min(|\theta(t)|,|\theta(t')|)$ (0-conflict). If additionally $x_1 = x_4$ we get a cycle $z = [x_1,x_2] + [x_2,x_3] + [x_3,x_1]$ in the chain complex associated to $K^{t,t''}$ that indicates a tight relationship between x_1,x_2,x_3, see Fig. 1. A coarse-graining perspective on clustering suggests that there is an opportunity to unify the three points in a single cluster. If the cycle is non-bounding, i.e., $[z] \neq 0$ in $H_1(K^{t,t''})$, we have $\mathrm{HF}_1(t,t'') \geq 1$, which indicates that there is no cluster that unifies the three points, i.e., $\nexists r$, $t \leq r \leq t''$, such that $x_1 \sim_r x_2 \sim_r x_3$. We call this a 1-conflict, which arises through the higher-order interactions between partitions across scales. If instead $x_1 \neq x_4$ then there is no 1-conflict but only a 0-conflict.*

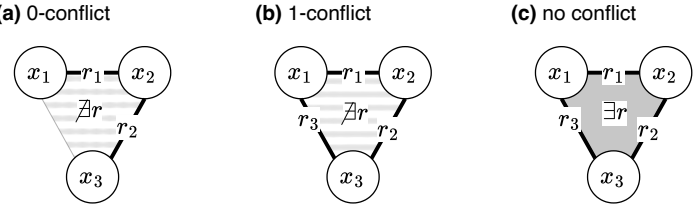

Figure 5: Illustration of (a) 0-conflict that violates the triangle inequality, (b) a 1-conflict and (c) no conflict due to global transitivity.

The next example demonstrates that information-based measures cannot detect 1-conflicts in general.

Example 26. *Let us consider a set $X = \{x_1, x_2, x_3, x_4\}$ and two different sequences of partitions $\theta(t)$ and $\eta(t)$ such that:*

$$\theta(1) = \eta(1) = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$$
$$\theta(2) = \eta(2) = \{\{x_1\}, \{x_2, x_3\}, \{x_4\}\}$$
$$\theta(3) = \{\{x_1, x_3\}, \{x_2\}, \{x_4\}\} \neq \theta(3) = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}\}$$

The two sequences θ and η are only different at $t = 3$. In particular, $t = 3$ induces a 1-conflict inθ but η has only 0-conflicts. Note that θ corresponds to the toy example in Fig. 1 with one additional isolated point. We get the following results:

The one-dimensional Hilbert functions can detect that only θ leads to a 1-conflict because

$$\mathrm{HF}_1(\theta(1), \theta(3)) = 1 > \mathrm{HF}_1(\eta(1), \eta(3)) = 0$$

.

In contrast, conditional entropy cannot distinguish between the two sequences which lead to the same pairwise conditional entropies (see equation 10):

$$\mathrm{H}(\theta(2)|\theta(1)) = \mathrm{H}(\theta(3)|\theta(1)) = \mathrm{H}(\theta(3)|\theta(2)) = \frac{1}{2}\log 2$$

$$\mathrm{H}(\eta(2)|\eta(1)) = \mathrm{H}(\eta(3)|\eta(1)) = \mathrm{H}(\eta(3)|\eta(2)) = \frac{1}{2}\log 2$$

C Details on Information-based Baseline Methods

Information-based measures can be used to compare arbitrary pairs of partitions in the sequence θ (Meilă, 2007). Assuming a uniform distribution on X, the conditional probability distribution of $\theta(t)$ given $\theta(s)$ is given by

$$P_{t|s}[i|j] = \frac{|\theta(t)_i \cap \theta(s)_j|}{|\theta(s)_j|}. \tag{9}$$

Similarly, we can define the joint probability $P_{s,t}[i, j]$. The conditional entropy (CE) $\mathrm{H}(t|s)$ is then given by the expected Shannon information:

$$\mathrm{H}(t|s) = -\sum_{i=1}^{|\theta(t)|}\sum_{j=1}^{|\theta(s)|} P_{s,t}[i, j]\log(P_{t|s}[i|j]) \tag{10}$$

It measures how much information about $\theta(t)$ we gain by knowing $\theta(s)$. If $\theta(s) \leq \theta(t)$ there is no information gain and $\mathrm{H}(t|s) = 0$. We denote the conditional entropy matrix $\mathrm{CE}_{s,t} = H(t|s)$. Furthermore, we can compute the variation of information (VI) $\mathrm{VI}(s, t) = \mathrm{H}(s|t) + \mathrm{H}(t|s)$, which is a metric. Both CE and VI are bounded by $\log N$.

Extending information-based measures for the analysis and comparison of more than two partitions is non-trivial. However, the pairwise comparisons can be summarised with the *consensus index* (CI) (Vinh et al., 2010) which can be computed as the average VI:

$$\mathrm{CI}(\theta) := \frac{\sum_{i=1,i<j}^{M} \mathrm{VI}(t_i, t_j)}{M(M-1)/2} \tag{11}$$

D Details on Experiments

D.1 Regression task

Figure 6 shows the correlation between the crossing number y and information- and MCbiF-based summary statistics. In addition to the results described in the main text, we also observe that the correlation between CI and $\bar{c}_0$ ($r = -0.32$ for $N = 5$, $r = -0.48$ for $N = 10$) is stronger than with $\bar{c}_1$ ($r = -0.12$ for $N = 5$, $r = -0.34$ for $N = 10$). This can be explained by the fact that CI and $\bar{c}_0$ can both be computed from pairwise interactions of clusters in contrast to $\bar{c}_1$, see Section 5.

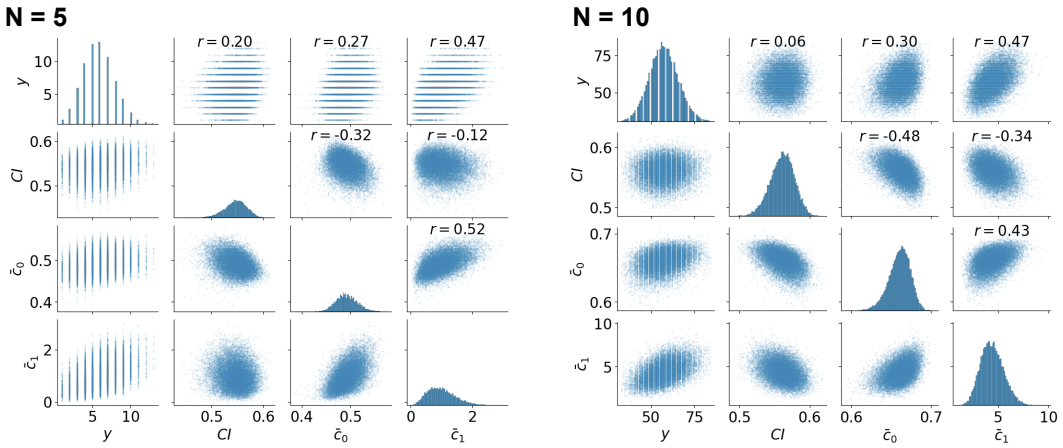

Figure 6: Pearson correlation (r) between crossing number y, information-based consenus indices $\overline{\text{CE}}$ and $\overline{\text{VI}}$ and MCbiF-based conflict measures $\bar{c}_0$ and $\bar{c}_1$ for $N = 5$ and $N = 10$.

Furthermore, we observe a strong correlation between $\bar{c}_0$ and $\bar{c}_1$ ($r = 0.52$ for $N = 5$ and $r = 0.43$ for $N = 10$) because of the dependencies between 0- and 1-conflicts, see Section 4.1

Note that we can consider our information- and MCbiF-based feature maps as $M \times M$ greyscale images, where HF_0 and HF_1 are symmetric and CE is asymmetric. For our regression task, we train a simple CNN (LeCun & Bengio, 1998) with one convolution and max-pool layer and one fully connected layer and also a simple MLP with one or two hidden layers and dropout. For each feature map (or their combinations) separately, we perform hyperparameter optimisation for the number of filters (ranging from 2 to 6) and kernel size (chosen as 4, 8, 16, 32 or 64) in the CNN and the number of nodes (chosen as 4, 8, 16, 32, 64, 128 or 256), number of layers (1 or 2) and dropout rate (chosen as 0.00, 0.25 or 0.50) in the MLP. We use the Adam optimiser (Kingma & Ba, 2017) with learning rate chosen as 0.01, 0.005, 0.001, 0.0005 or 0.0001 for training.

We detail the hyperparameters for the best MCbiF- and CE-based models, which we determined through a full grid search of the hyperparameter space using the train and validation split of our data:

- A CNN with 4 filters, kernel size 3, and learning rate 0.001 is best for HF_0 & HF_1 at $N = 5$.
- A CNN with 8 filters, kernel size 2, and learning rate 0.005 is best for CE at $N = 5$ with the following hyperparameters:
- LR is best for HF_0 & HF_1 at $N = 10$.
- An MLP with a single layer of 256 nodes, no dropout and a learning rate of 0.001 is best for CE at $N = 10$ with the following hyperparameters:

In the following, we present test R^2 scores on MCbiF-, information-based and mixed features.

Table 3: Test R^2 scores of different models and combinations of feature maps for $N = 5$ and $N = 10$, $M = 20$.

N	Method	Raw θ	HF_0	HF_1	HF_0 & HF_1	CE	CE & HF_0	CE & HF_1	CE & HF_0 & HF_1
5	LR	0.001	0.147	0.486	0.539	0.392	0.507	0.532	0.538
	CNN	-0.006	0.155	0.504	**0.544**	0.492	0.494	0.539	**0.557**
	MLP	-0.002	0.150	0.491	0.541	0.409	0.470	0.528	0.536
10	LR	-0.012	0.214	0.448	**0.516**	0.457	0.491	0.503	0.512
	CNN		0.211	0.448	0.507	0.454	0.456	0.491	0.510
	MLP	0.000	0.212	0.450	0.514	0.458	0.465	0.496	0.512

We also present the train R^2 scores:

Table 4: Train R^2 scores of different models and combinations of feature maps for $N = 5$ and $N = 10$, $M = 20$.

N	Method	Raw θ	HF_0	HF_1	HF_0 & HF_1	CE	CE & HF_0	CE & HF_1	CE & HF_0 & HF_1
5	LR	0.005	0.163	0.493	0.550	0.409	0.516	0.539	0.557
	CNN	0.000	0.170	0.509	**0.562**	0.515	0.549	0.586	**0.568**
	MLP	0.006	0.160	0.499	0.547	0.439	0.527	0.565	0.570
10	LR	0.013	0.230	0.456	**0.522**	0.464	0.501	0.517	0.530
	CNN	0.009	0.220	0.456	0.519	0.476	0.500	0.508	0.514
	MLP	0.003	0.218	0.453	0.515	0.468	0.485	0.506	0.516

D.2 Classification Task

Details on Synthetic Data. We generate order-preserving ($y = 0$) sequences $\theta \in \Pi_N^M$ through the following scheme: Let us assume that we have a total order $X = \{x_1, \dots, x_N\}$ given by the point labels, i.e., $x_i < x_j$ if $i < j$. We construct each $\theta(t_m)$, $m = 0, \dots, M-1$, by cutting X into clusters of the form $C = \{x_i, x_{i+1}, \dots, x_{i+n}\}$. It is easy to verify that θ is indeed order-preserving. We adapt this scheme to generate sequences $\theta \in \Pi_N^M$ that are not order-preserving ($y = 0$): Again, we start by constructing each sequence $\theta(t_m)$ through cutting the ordered set X as before. Additionally, with probability $p = 0.1$, we swap the cluster assignments in $\theta(t_m)$ for two arbitrary points $x, y \in X$. If N and M are large enough, the so-generated sequence θ is almost surely not order-preserving. We chose $N = 500$ and $M = 30$ to demonstrate the scalability of the MCbiF method.

The number of clusters of all our generated sequences of partitions $\theta \in \Pi_N^M$ for both classes is decreasing linearly, see Fig. 7 (a). Moreover, the average number of swaps for sequences with $y = 1$ is 2.98 for our choice of $p = 0.1$, see Fig. 7 (b).

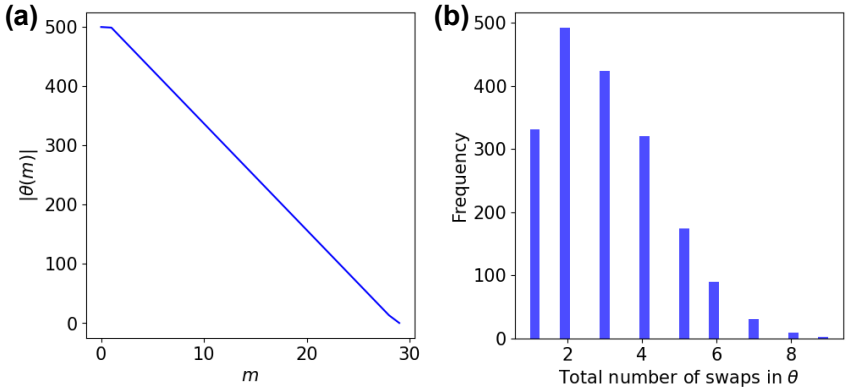

Figure 7: (a) Number of clusters $|\theta(m)|$ over $0 \leq m \leq M = 30$. (b) Histogram for the number of swaps of the sequences θ with $y = 1$.

D.3 Application to Temporal Data

Data Preprocessing. The temporal sequences of partitions computed by Bovet et al. (2022) are available at: `https://dataverse.harvard.edu/file.xhtml?fileId=5657692`. We restricted the partitions to the $N = 281$ mice that were present throughout the full study period to ensure well-defined sequences of partitions. Moreover, we only considered the first nine temporal resolution values τ_i, $i = 1, \dots, 9$ because θ_{τ_10} is an outlier. Note that the sequences tend to be fine-graining, see Fig. 8.

Time Reversibility. In the main text, we restricted our analysis to the so-called *forward* Flow Stability sequences of partitions. However, by reversing time direction Bovet et al. (2022) computed a second set of *backward* sequences. For each temporal resolution τ_i, we thus get a forward and

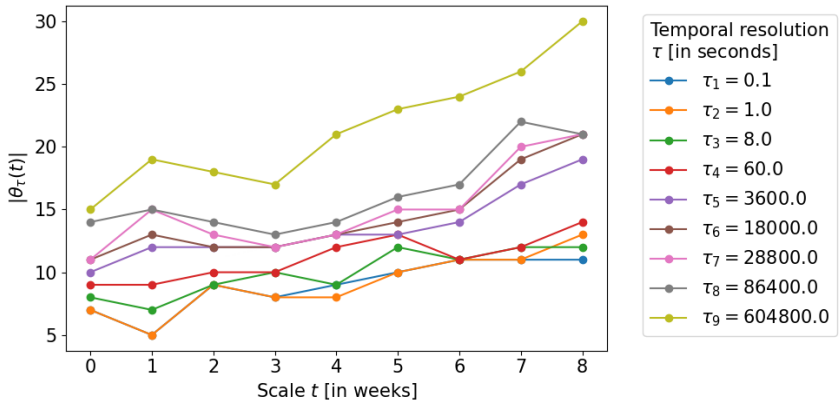

Figure 8: Number of clusters over weeks t.

backward sequence denoted by $\theta^f_{\tau_i}$ and $\theta^b_{\tau_i}$, respectively. Here we use the MCbiF to compare the forward and backward sequences of partitions for different τ_i and we compute the Hilbert distance $\| HF_k(\theta^f_{\tau_i}) - HF_k(\theta^b_{\tau_i}) \|_2$ for $k = 0, 1$, see Fig. 9.

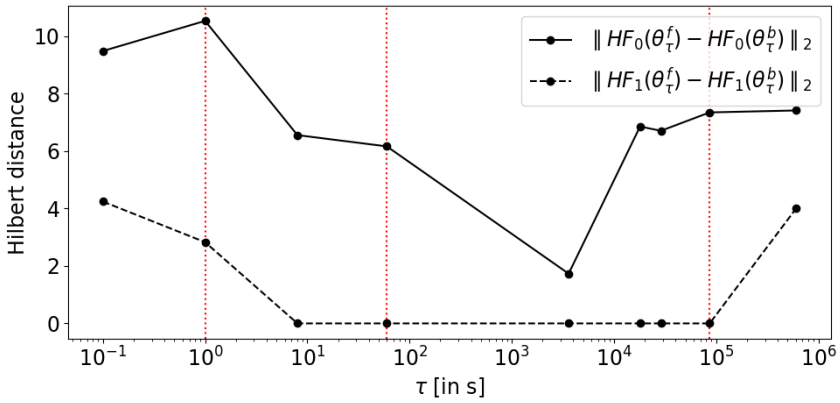

Figure 9: Hilbert distance between forward and backward Flow Stability sequences.

We observe that the Hilbert distance between forward and backward sequences is high for τ_2 because the large-scale group structure changes significantly over the study period, so that the temporal flows at low resolution τ_2 are not reversible. In contrast, the Hilbert distance between forward and backward sequences is low for τ_8 because the underlying social groups are more stable over the study period, leading to increased time reversibility at the high temporal resolution.