MCbiF: Measuring Topological Autocorrelation in Multiscale Clusterings via 2-Parameter Persistent Homology

ICLR 2026, 23–27 April 2026 in Rio de Janeiro, Brazil

Juni Schindler[1, 2] and Mauricio Barahona [1]

1 Department of Mathematics, Imperial College London, UK
2 Department of Mathematical Modeling and Machine Learning, University of Zurich, Switzerland

Email: juni.schindler@uzh.ch

Problem Definition & Related Work

Motivation: Multiscale clusterings appear in unsupervised learning across various domains including clustering of single cell data, temporal clustering, diffusion-based clustering.

Definition: A *sequence of partitions* is a piecewise-constant function $\theta : [t_1, \infty) \to \Pi_X, t \mapsto \theta(t) \in \Pi_X$ with continuous scale parameter t and M change points $t_1 < t_2 < \ldots < t_M$. We call θ a multiscale clustering.

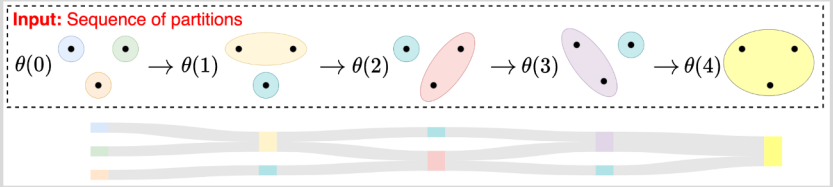


Figure 1. A (non-hierarchical) sequence of partitions θ can be illustrated by its Sankey diagram.

Note: The scale t is often related to the average cluster size.

Goal: Characterise and analyse sequences of partitions θ, ***including non-hierarchical ones***, in an integrated manner, taking full account of memory effects across the scale t.

Idea: We use tools from TDA (multiparameter persistent homology [1]) to capture the structure of θ as a whole.

Related work:

- Ultrametrics [2] encode structure for dendrograms
  ➢ *only applicable to hierarchical sequences*
- Measures for pairwise comparisons of partitions, e.g., Conditional entropy (CE) & Variation of Information (VI) [3], Adjusted Rand Index (ARI) [4], Maximum Overlap Distance (MOD) [5]
  ➢ *do not capture higher-order cluster inconsistencies*

[1] Carlsson, G. & Zomorodian, A. The Theory of Multidimensional Persistence. *Discrete Comput Geom* 42, 71–93 (2009).
[2] Carlsson, G. & Mémoli, F. Characterization, Stability and Convergence of Hierarchical Clustering Methods. *JMLR* 11, 1425–1470 (2010).
[3] Meilă, M. Comparing Clusterings by the Variation of Information. in *Learning Theory and Kernel Machines* (eds Schölkopf, B. & Warmuth, M. K.) 173–187 (2003).
[4] Hubert, L. & Arabie, P. Comparing partitions. *Journal of Classification* 2, 193–218 (1985).
[5] Peixoto, T. P. Revealing Consensus and Dissensus between Network Partitions. *Phys. Rev. X* 11, 021003 (2021).

Method: Multiscale Clustering Bilfitration (MCbiF)

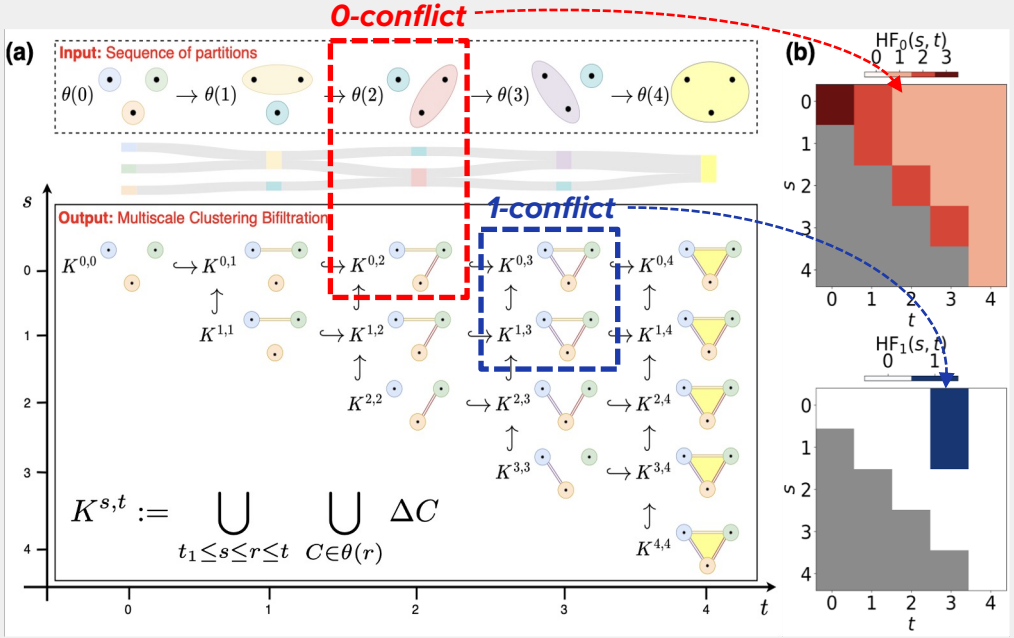


Figure 2.
a) MCbiF encodes θ as a bifiltration of abstract simplicial complexes.

b) Hilbert functions (HF) summarise the topological autocorrelation of the bifiltration.

Consider *Hilbert functions* (HF):

$$\mathrm{HF}_k : [t_1, \infty)^{\mathrm{op}} \times [t_1, \infty) \to \mathbb{N}_0, (s,t) \mapsto \mathrm{HF}_k(s,t) := \dim[H_k(K^{s,t})]$$

Definition: Average 0-conflict:

$$0 \leqslant \bar{c}_0(\theta) := 1 - \frac{2}{|T - t_1|^2} \int_{t_1}^{T} \int_{s}^{T} \frac{\mathrm{HF}_0(s,t)}{\min_{r \in [s,t]} \mathrm{HF}_0(r,r)} \mathrm{d}s \; \mathrm{d}t \leqslant 1$$

Corollary 9: *0-conflict ⇒ θ non-hierarchical*

Definition: Average 1-conflict:

$$0 \leqslant \bar{c}_1(\theta) := \frac{2}{|T - t_1|^2} \int_{t_1}^{T} \int_{s}^{T} \mathrm{HF}_1(s,t) \mathrm{d}s \; \mathrm{d}t$$

Corollary 14:

- *1-conflict ⇔ θ has higher-order cluster inconsistencies.*
- *Nested clusters ⇒ no 1-conflict.*

Experimental Validation

Regression task: Minimal crossing number of Sankey layout

Synthetic data: Coarse-graining sequences of partitions of length M=20 defined on sets with N=5, 10 elements.

➢ 20,000 such sequences generated computationally

Target: Minimal crossing number of the Sankey layout.

Experimental setup:

- Compute MCbiF and baseline features
- Train ML models on regression task with different features and apply representation learning.
- Evaluate performance with R2 score on test split.

N	Method	Raw label encoding	Sankey graph	HF_0	HF_1	HF_0 & HF_1	CE	ARI	MOD
5	LR	0.078	-	0.147	0.486	0.539	0.392	0.166	0.413
	CNN	0.267	-	0.155	0.504	**0.544**	0.492	0.422	0.354
	MLP	0.104	-	0.150	0.491	0.541	0.409	0.214	0.351
	GCN	-	0.416	-	-	-	-	-	-
10	LR	0.038	-	0.214	0.448	**0.516**	0.457	0.246	0.345
	CNN	0.072	-	0.211	0.448	0.507	0.454	0.294	0.312
	MLP	0.036	-	0.212	0.450	0.514	0.458	0.256	0.246
	GCN	-	0.229	-	-	-	-	-	-

Results: *MCbiF HFs have highest predictive power for the global property of the minimal crossing number.*

Classification task: Non-order-preserving sequences of partitions

Synthetic data: Coarse-graining sequences of partitions of length M=30 defined on a set of N=500 elements, where 50% of the sequences are non-order preserving (y=1) and the other 50% are order preserving (y=0).

➢ 3,700 such sequences generated computationally

Target: Binary label y

Experimental setup:

- Compute MCbiF and baseline features.
- Train logistic regression on binary classification task with different features or raw label encoding.
- Evaluate performance of test split with accuracy.

Raw label encoding	HF_0	HF_1	CE	ARI	MOD
0.53	0.56	**0.97**	0.50	0.49	0.46

Results: *In contrast to pairwise measures, MCbiF HFs (HF1) have high sensitivity to order-preservation in sequences of partitions, due to capturing higher-order cluster inconsistencies.*

Application to Real-World Temporal Data

Task: Analyse level of non-hierarchy in temporal interaction data

Data: Sequences of partitions describe mice social groupings for N=281 individuals over 9 weeks. Sequences were computed at 9 different temporal resolutions from coarse (small τ) to fine (large τ), and robust resolutions τ_2, τ_4 and τ_8 were previously identified [1].

Experimental setup:

- We compute MCbiF HFs for sequences of partitions for 9 different temporal resolutions.
- We compute the average 0- and 1-conflict.

Results:

- MCbiF measures show that τ_2 leads to a less hierarchical sequence of partitions than τ_4 and τ_8.
- τ_4 leads to the most hierarchical sequence of partitions with lowest 0-conflict and no 1-conflict.

➢ *Quantitative results consistent with qualitative findings in [1].*

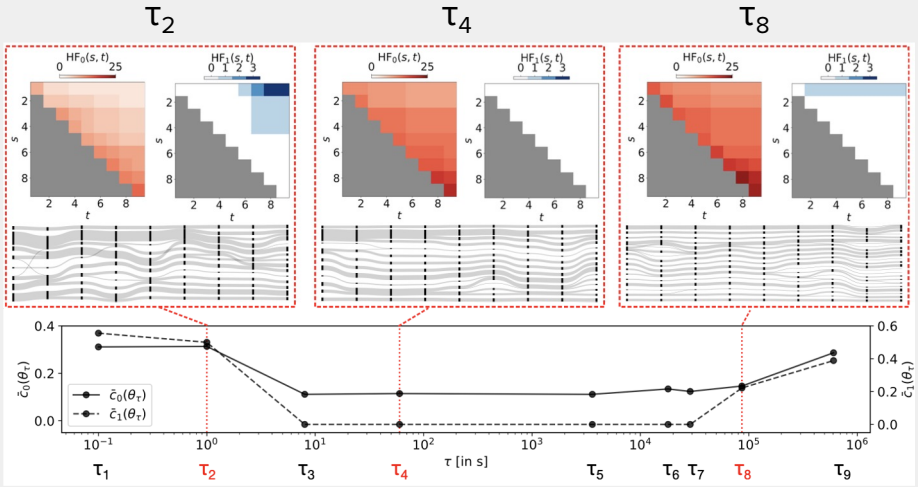


Figure 3. Analysing non-hierarchical sequences of temporal partitions compiled from social interactions of a mice population over 9 weeks.

[1] Bovet, A., Delvenne, J.-C. & Lambiotte, R. Flow stability for dynamic community detection. *Science Advances* 8, eabj3063 (2022).

Conclusion

MCbiF is a novel ***bifiltration*** that encodes the cluster intersection patterns in ***multiscale, non-hierarchical*** sequences of partitions θ.

Our theoretical analysis shows:

- MCbiF HFs are ***stable*** and quantify ***topological autocorrelation*** in θ across all scales
- At dimension 0 we capture violations of strict refinements in θ (***0-conflicts***).
- At dimension 1 we capture higher-order cluster inconsistencies in θ (***1-conflicts***).

Our numerical experiments show:

- ***MCbiF HFs outperform other methods*** on regression and classification tasks on non-hierarchical sequences of partitions.

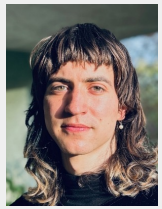
Paper: Schindler, J. & Barahona, M. MCbiF: Measuring Topological Autocorrelation in Multiscale Clusterings via 2-Parameter Persistent Homology.

OpenReview: https://openreview.net/forum?id=E7D6uybODJ

arXiv: https://arxiv.org/abs/2510.14710

GitHub: barahona-research-group/MCbiF

Funding: UKRI Engineering and Physical Sciences Research Council



Dr. Juni Schindler
(juni.schindler@uzh.ch)

Prof. Mauricio Barahona