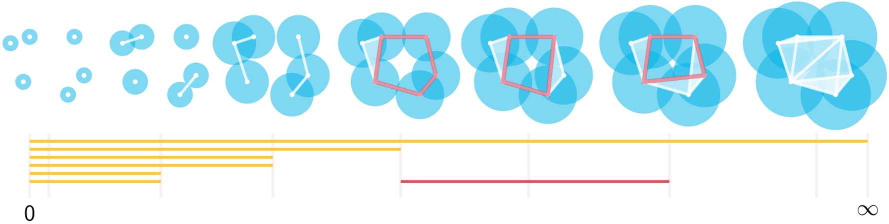
Topological structures and biological settings

Laboratoire Jean Alexandre Dieudonné
Université Côte d'Azur

11 March 2025



Jānis Lazovskis
University of Latvia

Slides at: `jlazovskis.com/talks`

Personal overview

- 2014 - 2019: University of Illinois at Chicago
- 2019 - 2020: University of Aberdeen
- 2020 - 2025: Riga Technical University
- 2023 - 2024: Printful
- 2025 - : University of Latvia

My work is...

- motivated by the deep structural tools that algebraic topology provides and which computational topology can realize
- often with teams and in interdisciplinary projects
- very frequently with a biological theme

Presentation overview

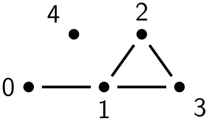
- Relevant mathematical structures
  - Graphs, simplicial complexes, invariants
- Neuro-topology: neural circuit structure and function through a topological lens
  - BlueBrain Project, digital reconstructions, simulations and classification
- Eco-topology: Ecological inferences for hypervolumes
  - Filtrations, the Vietoris–Rips complex, Persistent homology
  - Bounding distances between persistence diagrams
- Tracking dynamics: Topological changes at the foundations
  - Factoring the boundary matrix
  - Taking apart the machinery for computing persistence

Graphs

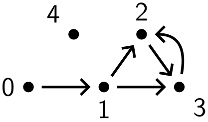
The basic mathematical concept throughout this talk will be a finite set.

$\{0, 1, 2, 3, 4\}$

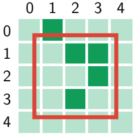
finite set V



undirected graph G



directed graph G'



	0	1	2	3	4
0		🟩			
1			🟩	🟩	
2				🟩	
3			🟩		
4					

adjacency matrix A

- The pair $G = (V, E)$ is an undirected graph, with $E = \{\{0, 1\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$.
- The pair $G' = (V, E')$ is a directed graph, with $E' = \{(0, 1), (1, 2), (1, 3), (2, 3), (3, 2)\}$.

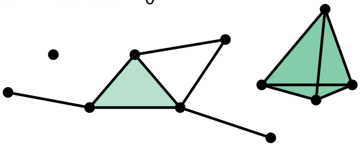
We are interested in particular subgraphs of G', with $V' = \{1, 2, 3\}$:

- V' and $(1, 2), (1, 3), (2, 3)$ form a *directed 3-clique*, as do $(1, 2), (1, 3), (3, 2)$
- V' and $(1, 2), (1, 3), (2, 3), (3, 2)$ form the *closed neighbourhood* of vertex 3 (or 2)

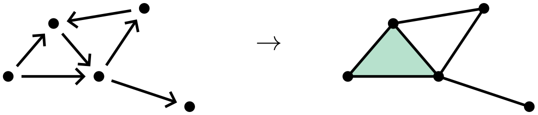
Simplicial complexes

Let K be a simplicial complex on the finite set K_0.

$K =$



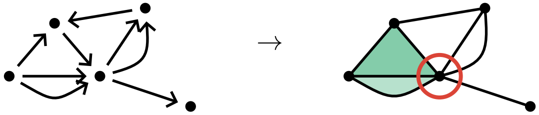
- A *0-simplex* is a set with a single element.
- An *n-simplex* is a set with $n+1$ elements, for which every subset of size $n-1$ is an $(n-1)$-simplex. The set of n-simplices is denoted K_n.
- A *simplicial complex* K is a set of simplices for which $\sigma \cap \tau \in K$, whenever $\sigma, \tau \in K$.



The simplicial complex associated to a directed graph is its *directed flag* (or *clique*) *complex*.

Invariants

If G has a reciprocal edge or multiple edges, the associated topological space is a *cellular complex* K, which is defined by gluing copies of the *n-cell* D^n to K_{n-1} to form K_n.



From every graph or complex, we can compute values that represent some underlying feature.

Euler characteristic : $\chi(K) = \sum_n (-1)^n |K_n|$

transitive clustering coefficient : $\mathrm{tcc}(K, v) = \frac{\text{number of 3-cliques containing } v}{\text{number of theoretically possible 3-cliques containing } v}$

adjacency spectral gap : difference between the two largest moduli of eigenvalues of A

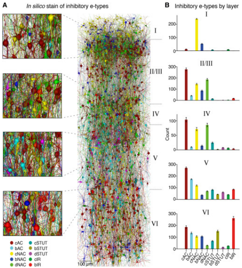
(*Bauer*) *Laplacian spectral gap* : Laplacian for directed graphs without strong connectivity

Main idea: **describe properties of closed neighbourhoods with numerical values.**

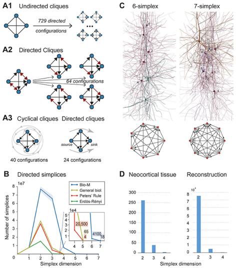
The BlueBrain Project

The Blue Brain Project (2005 - 2024) presented a complementary way to model the brain

- Encode electro-morphological neuron types of adult mouse in algorithms
- Reconstruct parts of neocortex and somatosensory cortex in a computer
- Stimulate and observe activity at neuron and synapse level



digital reconstruction



mathematical interpretation

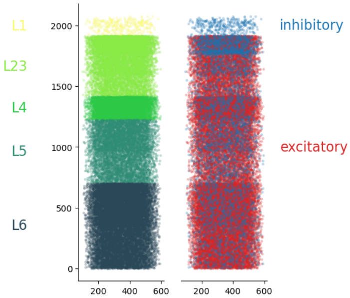
Blue
Brain
Project

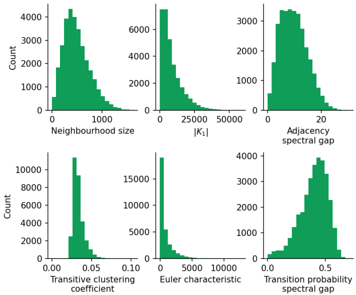
Open Brain
Institute

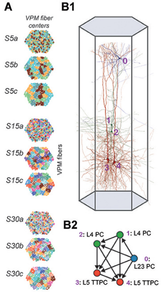
work continued by OBI

BBv5 connectome

- *Structure:* 31k neurons, 7.8m connections, 55 EM types
- *Topology:* Compared with random models, high dimensional cliques are overrepresented
- *Robustness:* Reciprocal connections preferentially appear in high dimensional cliques
- *Activity:* The number of active *n*-cliques can be used to classify simulations

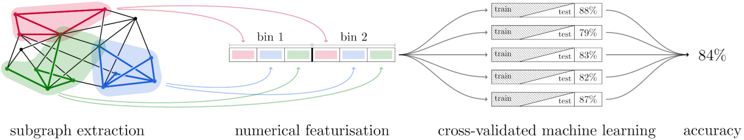






Classifying simulations by neighbourhoods

- Compute parameter p_1 for every neighbourhood, select highest / lowest n
- For each selected neighbourhood, compute parameter p_2 of m active subgraphs
- Classify feature vectors of length $n \cdot m \cdot 8 \cdot 557$ with SVM

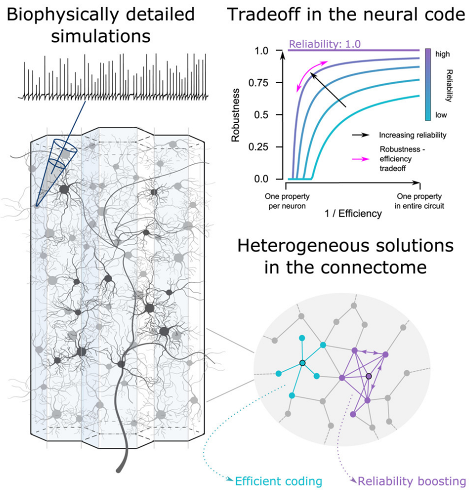


Validate: random selection, smaller regions, random regions, shuffled activity, NEST simulator

Main result (Conceição, Govc, Lazovskis, Levi, Riihimäki, Smith. 2022)

Simulations on neural circuits can be classified with high accuracy considering only the structure of and activity on a small number of neighbourhoods.

Reliability in neural circuits



- Does there exist a random directed graph model with predetermined clique counts?
- How can firing rate be decoupled from every other parameter in active subgraphs?

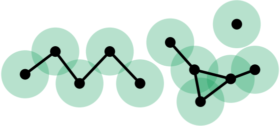
The level of tradeoff between efficiency and robustness depends on the *reliability* of the sytstem

- The brain finds the sweet spot by having *heterogeneous connectivity*
- This is quantified by neighbourhoods with high dimensional simplices that are very robust, but very inefficient

Distances and filtrations

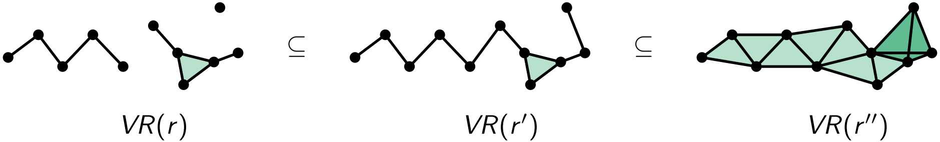
The *Vietoris–Rips complex* of a finite set is determined by pairwise distances among elements.

$X =$



$$VR(r)_n = \left\{ \sigma \subseteq X \ : \ \begin{array}{l} |\sigma| = n+1 \\ d(x,y) \leqslant r \ \forall \ x, y, \in \sigma \end{array} \right\}$$

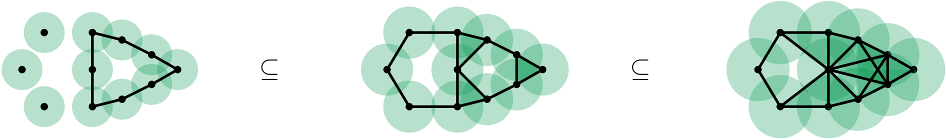
Increasing the threshold $r < r' < r''$ induces a *filtration*, or a sequence of nested spaces.



The *simplex-wise* filtration of a simplicial complex has one simplex added at each step.

Persistent homology

At each step r of the filtration $\mathcal{F}$, compute the homology $H_*(\mathcal{F}_r)$ of the space $\mathcal{F}_r$.



$\subseteq$ $\subseteq$

$H_0(\mathcal{F}_r) = \mathbf{Z}^4, \quad H_1(\mathcal{F}_r) = \mathbf{Z}$

$H_0(\mathcal{F}_{r'}) = \mathbf{Z}, \quad H_1(\mathcal{F}_{r'}) = \mathbf{Z}^2$

$H_0(\mathcal{F}_{r''}) = \mathbf{Z}, \quad H_1(\mathcal{F}_{r''}) = \mathbf{Z}$

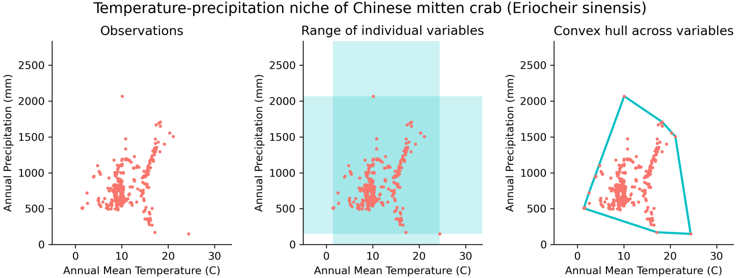
Compose these assignments as $\mathbf{R} \xrightarrow{\mathcal{F}} \mathrm{Top} \xrightarrow{H_*} \mathrm{Vect}$ to get the *persistent homology* of $\mathcal{F}$.

$$\begin{tikzcd}
\mathbf{Z} \arrow[r] & 0 \arrow[r] & 0 \\
\mathbf{Z} \arrow[r] & 0 \arrow[r] & 0 \\
\mathbf{Z} \arrow[r] & 0 \arrow[r] & 0 \\
\mathbf{Z} \arrow[r] & \mathbf{Z} \arrow[r] & \mathbf{Z} \\
r & r' & r''
\end{tikzcd}
\qquad
\begin{tikzcd}
0 \arrow[r] & \mathbf{Z} \arrow[r] & \mathbf{Z} \\
\mathbf{Z} \arrow[r] & \mathbf{Z} \arrow[r] & 0 \\
r & r' & r''
\end{tikzcd}$$

The *birth* and *death* times of each interval are determined by unique module decomposition.

The ecological niche and hypervolume

Environmental parameters (temperature, rain, sunlight, ...) are key to every species' survival.



Given a particular species:

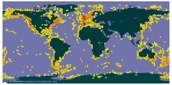
- its *niche* is bounded by the range of environmental parameters in which it can exist
- its *realized niche* is the subset in which the species is observed
- its *hypervolume* is an approximation of the realized niche as a subset of Euclidean space

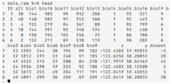
Constructing a biologically accurate hypervolume

- The size of a climatic niche is hypothesized to drive species *diversification rates*
- The similarity of species' environmental or functional trait hypervolumes measures niche divergence or packing, which may influence *species coexistence* and richness patterns
- Niche similarity also helps compare individuals within a species, assessing *climate change* impacts and niche shifts during *invasions*

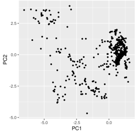
Inherent problem: **observations are heavily skewed to humans (with database access)**

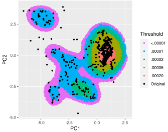
GBIF Global Biodiversity Information Facility





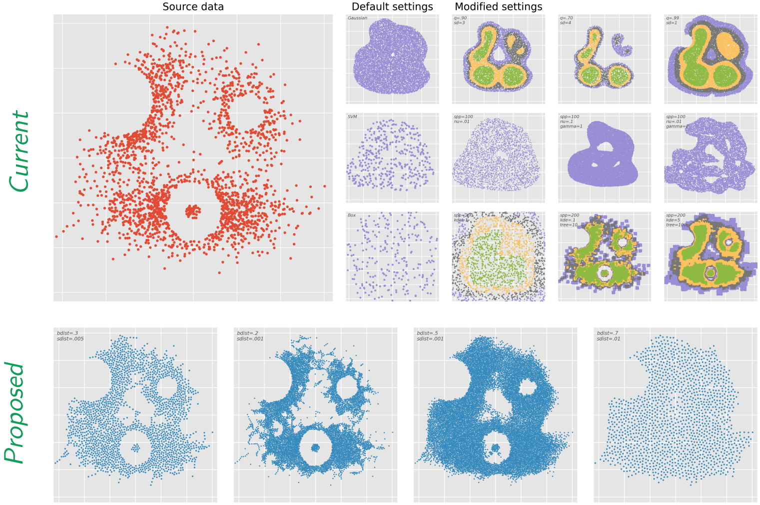
```
> data_raw %>% head
  ID alt bio1 bio10 bio11 bio12 bio13 bio14 bio15 bio16 bio17 b
1  1  30  144   185    99   932   206     1    93   521     9
2  2  48  140   187    91   753   168     1    92   423     9
3  3   4  155   219    84   361    80     1    85   197     6
4  4  10  144   191    92   740   174     1    95   430     7
5  5   0  150   194   102   356    73     1    86   186     5
6  6   2  152   211    87   518   119     1    90   287     6
  bio3 bio4 bio5 bio6 bio7 bio8 bio9         x        y dcoast
1   52 3393  244   50  194   99  182 -122.4583 37.95833     -3
2   57 3812  273   33  240   91  183 -122.6250 38.20833     20
3   47 5333  315   27  288   84  218 -121.7917 38.04167     42
4   54 3936  270   37  233   92  188 -122.4583 38.12500      5
5   52 3613  253   49  204  111  191 -122.0417 37.45833      3
6   49 4891  299   30  269   87  209 -122.0417 38.20833     28
> █
```





public database → species observations → environmental parameters → project to subspace → compute hypervolume

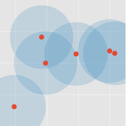
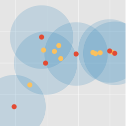
Current and proposed hypervolume methods



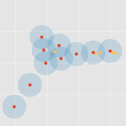
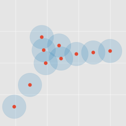
- ▶ (Gaussian) KDE infers new data nearby each true observation
- ▶ Topology infers new data between collections of true observations
- ▶ Both methods infer more samples and uniformize with respect to density

A complementary approach

Barycentric subdivision: Add the average of $k+1$ elements ($k=1,2$) within a distance $\delta > 0$:

$X =$  , $B(X,\delta) =$  $= X \cup \bigcup_k \left\{ \frac{\sum \sigma_i}{k+1} : \sigma = \{\sigma_0, \ldots, \sigma_k\} \in VR_\delta(X)_k \right\}$

Sparsification: Remove all other elements within a distance $\varepsilon > 0$ of each element:

$X =$  , $S(X,\varepsilon) =$  $= \{x \in X \ : \ d(x,x') > \varepsilon \ \forall \ x' \in X, x' \neq x\}$

Topological guarantees

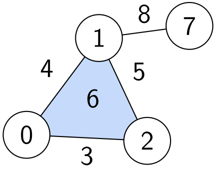
Main result (Lazovskis, Levi, Morimoto. 202x)

Let $X \subseteq \mathbf{R}^N$ be a finite set and D_X its persistence diagram of the Vietoris–Rips filtration in degree 0. Then $d_W(D_X, D_{B(X,\delta)}) \leqslant \frac{\delta}{4}$ and $d_W(D_X, D_{S(X,\varepsilon)}) \leqslant \frac{\varepsilon}{2}$ for every $\delta, \varepsilon > 0$.

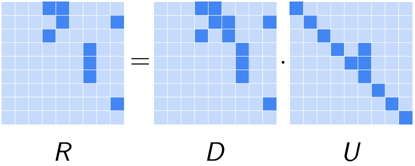
$$\begin{tikzcd}
S(X,\varepsilon) \arrow[r, "\mathrm{erule}_{S(X,\varepsilon)}", "\simeq"'] \arrow[d, hook, "\iota_S"'] & D_{S(X,\varepsilon)} \arrow[d, hook, dashed, "\varphi_S"] \\
X \arrow[r, "\mathrm{erule}_X", "\simeq"'] \arrow[d, hook, "\iota_B"'] & D_X \arrow[d, hook, dashed, "\varphi_B"] \\
B(X,\delta) \arrow[r, "\mathrm{erule}_{B(X,\delta)}", "\simeq"'] & D_{B(X,\delta)}
\end{tikzcd}$$

- Constructed relation between persistent homology computation and hardware limitations
- Respects original data and maintains uniform density in hypervolume
- Next: Align to grid, take complement, use duality for top-dimensional holes. Track over time.

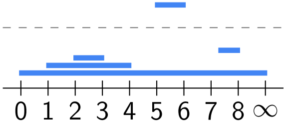
Persistence diagrams



filtered simplicial complex



bounday matrix factorization



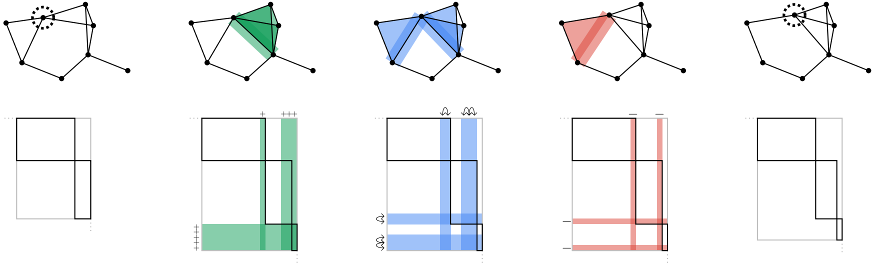
barcode from pivots of R

The standard barcode algorithm performs Gaussian elimination on D to produce the barcode.

```
Algorithm 1: SBA - Standard Barcode Algorithm
  Input: Boundary matrix $D$, identity matrix $I$
  Output: Reduced matrix $R$, operations matrix $U$
$R \leftarrow D$
$U \leftarrow I$
for $j = 1, \dots, n$
    while there exists $j' < j$ for which $\mathrm{piv}(R[j']) = \mathrm{piv}(R[j]) \neq 0$
        add $R[j']$ to $R[j]$
        add $U[j']$ to $U[j]$
return $R$, $V$
```

Dynamic data

How does the boundary matrix of the Vietoris–Rips filtration change in dynamic settings?

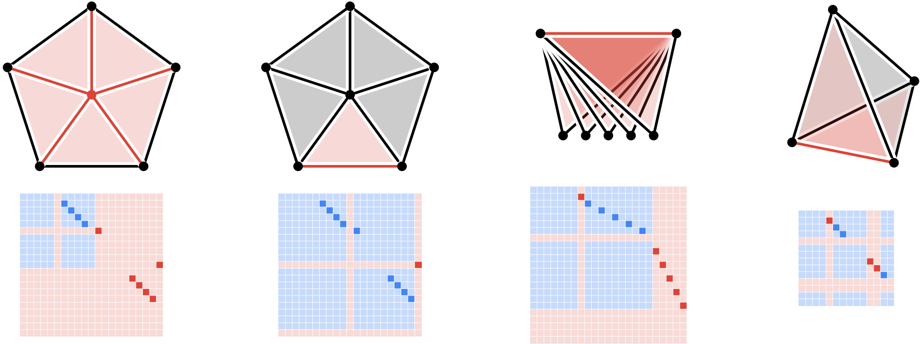


initial state *insert into filtration* *swap order* *remove from filtration* *final state*

- The number of simplices n is exponential in the number of points
- Insertion is reindexing, swapping is known to be $O(n)$

Removing a simplex

Removing a simplex $\sigma \in K$ means removing all cofaces of σ, described by the *star* $\mathrm{st}(\sigma)$ of σ.



The SiRUP algorithm:

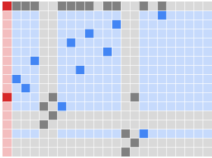
- For each simplex $\tau \in \mathrm{st}(\sigma)$, collect the off-diagonal entries in the column of τ in U
- Adjust the columns in R and U for each affected simplex to respect *class representatives*

Minimality

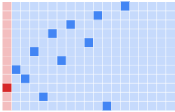
Main result (Giunti, Lazovskis. 2025)

Given the reduced boundary matrix of $\mathcal{F}$ as input, SiRUP will output the reduced boundary matrix of $\mathcal{F} \setminus L$ with $O(mn^2)$ operations, where $m = |L|$ and $n = |\mathcal{F}|$.

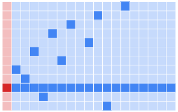
Combine repeated additions to involve each affected column exactly once in an operation.



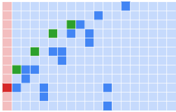
reduced boundary matrix



relevant submatrix



add back column of σ



apply standard barcode algorithm

Thank you

- F. Bauer. *Normalized graph laplacians for directed graphs.* Linear Algebra and its Applications, 2012.
- D. Cohen-Steiner, H. Edelsbrunner, D. Morozov. *Vines and vineyards by updating persistence in linear time.* Proceedings of the Twenty-Second Annual Sympo- sium on Computational Geometry, 2006.
- P. Conceição, D. Govc, J. Lazovskis, R. Levi, H. Riihimäki, J.P. Smith. *An application of neighbourhoods in digraphs to the classification of binary dynamics.* Network Neuroscience, 2022.
- W. Crawley-Boevey. *Locally finitely presented additive categories.* Communications in Algebra, 1994.
- G. Fagiolo. *Clustering in complex directed networks.* Physical Review E, 2007.
- B. Giunti, J.Lazovskis. *Pruning vineyards: Pruning vineyards: updating barcodes by removing simplices.* Manuscript in preparation, 2025.
- G.E. Hutchinson. *Population studies-animal ecology and demography - concluding remarks.* Cold Spring Harbor symposia on quantitative biology, 1957.
- J. Jordan et al. *Nest 2.18.0.* Zenodo, 2019.
- H. Markram et al. *Reconstruction and Simulation of Neocortical Microcircuitry.* Cell, 2015.
- D.E. Santander, C. Pokorny, A. Ecker, J. Lazovskis, M. Santoro, J.P. Smith, K. Hess, R. Levi, M.W. Reimann. *Heterogeneous and higher-order cortical connectivity undergirds efficient, robust, and reliable neural codes.* iScience, 2025.