What About Higher-Order Cellular Complexity? An Inquiry with single-cell Topological Simplicial Analysis

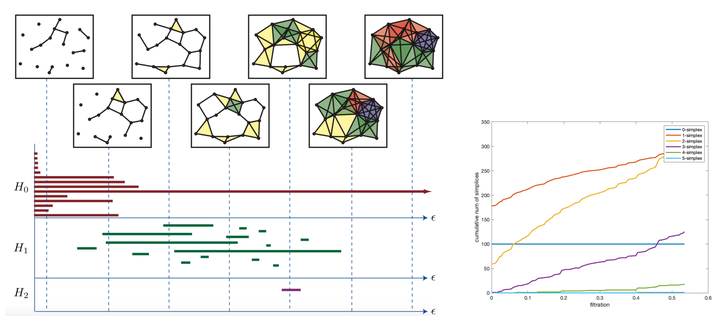
Baihan Lin[1,2,3], Raul Rabadan[1], Nikolaus Kriegeskorte[2]

[1]Department of Systems Biology, Columbia University Medical Center, New York, NY 10027, USA
[2]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
[3]Department of Applied Mathematics, University of Washington, Seattle, WA 98105, USA

Abstract

The lack of a formal link between cell-cell cohabitation and its emergent dynamics into cliques during development has hampered our understanding of how cell populations proliferate, differentiate, and compete, i.e. the cell ecology. With the advancement of single-cell RNA-sequencing (RNA-seq), we have now come closer to describing such a link by taking cell-specific transcriptional programs into account, constructing graphs of a network that reflect the similarity of gene expression, and analyzing these graphs using algebraic topology. We proposed single-cell topological simplicial analysis (scTSA). Applying this approach to single-cell gene expression profiles from local networks of cells in different developmental stages with different outcomes revealed a previously unseen topology of cellular ecology. These networks contain an abundance of cliques of single-cell profiles bound into cavities that guide the emergence of more complicated habitation forms. We visualize these ecological patterns with topological simplicial architectures of these networks, compared with the null models. Benchmarked on single-cell RNA-seq of zebrafish embryogenesis over 25 cell types and 12 time steps, our approach highlights the gastrulation as the most critical stage, consistent with consensus in developmental biology. As a nonlinear, model-independent, and unsupervised framework, our approach can also be applied to tracing multi-scale cell lineage, identifying critical stages, or creating pseudo-time series.

Simplicial Filtration



Unlike traditional TDA, we computed simplices into high dimension (towards 7) during the entire filtration process (as above).

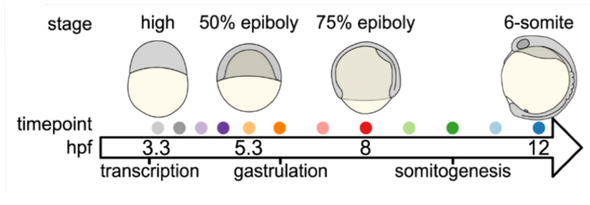
Lazy Witness Complex

As the number of cells collected in each time points vary differently, direct simplicial computation can be incomparable. For these larger datasets, if we include every data point as a vertex, the filtrated simplicial complexes can quickly contain too many simplices for efficient computation. Instead, we extracted lazy witness complexes by sampling n data points [1] with sequential maxmin procedure [2] setting nearest neighbor inclusion of 2 (as in "lazy").

Control Models for scTSA

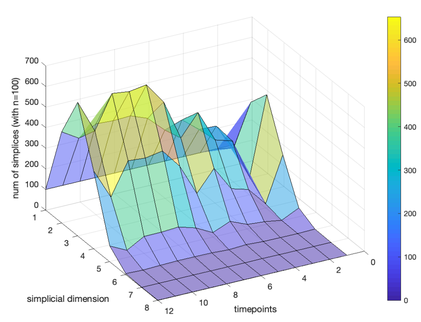
Usually for binary connectivity data (like brain connectome), Erdős-Rényi random graph [3] can be used as control. However, in similarity-based data, the connectivity is entirely dependent on the filtration factor. To avoid this caveat, we take a different approach by shuffling the pairwise distances of the data points, such that the MDS embedding can form different connectivity profiles but maintain the same distance distribution.

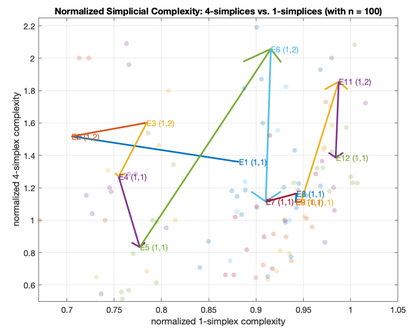
Data Preprocessing

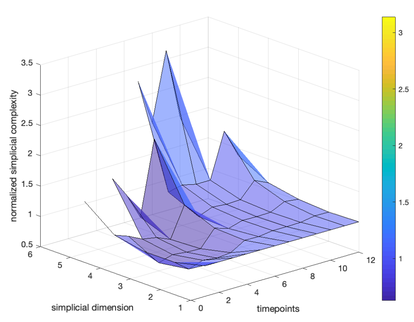


We applied scTSA to zebrafish single-cell data with 38,731 cells, 25 cell types, over 12 time steps [4]. The data has dimension of 103 corresponding to the expression levels of 103 genes selected by scTDA [5].

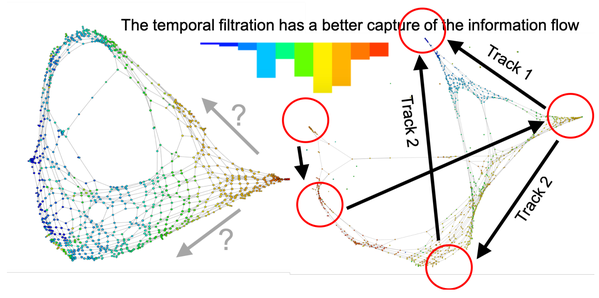
Simplicial Dynamics Across Developmental Stages

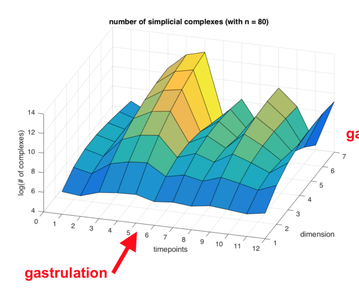


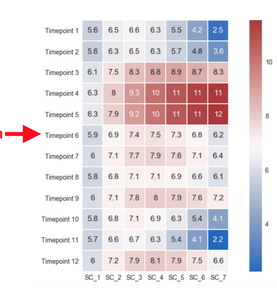




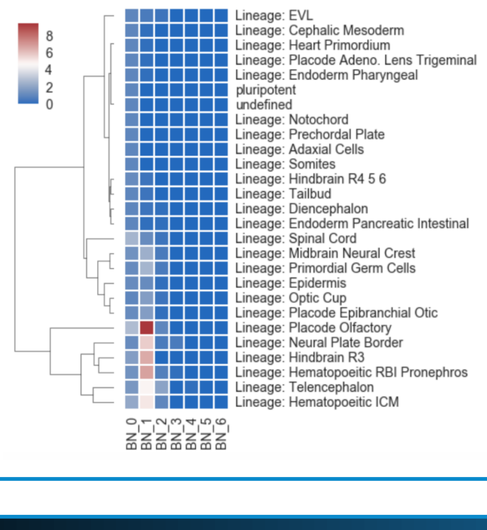
Temporal TDA (tTDA) Reveals Developmental Trajectories

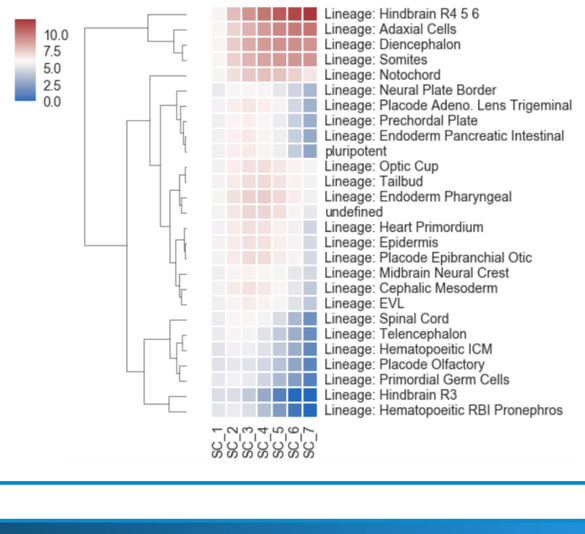






Lineage Tracing with Filtrated Simplicial Architecture





Possible Questions to Explore

- Can we determine developmental stages without physiological features?
- Can we generate pseudo-time series based on scRNA-Seq?
- Does the vast presence of high-dimensional cliques suggest that the interaction between these cells is organized into fundamental building blocks of increasing complexity?

References

[1] Vin De Silva and Gunnar E Carlsson. Topological estimation using witness complexes. *SPBG*, 4:157–166, 2004.

[2] Henry Adams and Gunnar Carlsson. On the nonlinear statistics of range image patches. *SIAM Journal on Imaging Sciences*, 2(1):110–117, 2009.

[3] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[4] Jeffrey A Farrell, Yiqun Wang, Samantha J Riesenfeld, Karthik Shekhar, Aviv Regev, and Alexander F Schier. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392):eaar3131, 2018.

[5] Abbas H Rizvi, Pablo G Camara, Elena K Kandror, Thomas J Roberts, Ira Schieren, Tom Maniatis, and Raul Rabadan. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. *Nature biotechnology*, 35(6):551, 2017.