In search of topological summaries for multispecies spatial patterns

Maria Jose Jimenez MAJIRO@US.ES
Universidad de Sevilla, Spain

Editors: Michael Bleher, Freya Jensen, Levin Maier, Diaaeldin Taha, and Anna Wienhard

Abstract

We present here, informally, ideas on ongoing work for analyzing chromatic (labeled) point clouds, with the goal of understanding spatial interactions between cell types in biological settings. We describe our starting framework and state some open questions that we are currently addressing. Two specific contexts attract our interest: (1) tumor microenvironment and (2) cell differentiation process.

1. Introduction

The field of topological data analysis (TDA) is concerned with the interpretation of datasets as geometric objects, which are analyzed using tools from algebraic topology (see (Carlsson and Zomorodian, 2009; Edelsbrunner and Harer, 2010) as foundational references). Given a point cloud dataset $X \subset \mathbf{R}^d$, a filtration of topological spaces is typically constructed (under the form of a nested sequence of simplicial complexes) to approximate the underlying space from which the data are presumed to be sampled. Persistent homology is usually encoded under the form of either a barcode or a persistence diagram and this information is typically vectorized (Ali et al., 2023) so that it can serve as input for machine learning methods for classification tasks, for example. Among a huge variety of applications of TDA, a significant quota of them is developed in the biomedical setting. The survey Amézquita et al. (2020) showcases TDA applications to various biological fields, like structural biology, molecular biology, and evolution, introducing different TDA methods, depending on the application. The set of TDA tools used for biological applications is also extensive: multiparameter persistent homology (Carlsson and Zomorodian, 2007) for the study of tumor microenvironment (Vipond et al., 2021) or spatial transcriptomic data (Benjamin et al., 2024); an adaptation of the mapper algorithm (Singh et al., 2007) to describe cell differentiation structures (Imoto and Hiraoka, 2023); or zigzag persistence (Carlsson and de Silva, 2010) to understand the dynamics of macrophages on agent-based model simulations of a tumour (Yang et al., 2025).

In simple terms, spatial biology is the study of how molecules, cells, and tissues are organized and interact with each other within their natural 2D or 3D spatial environment. More specifically, our two spatial biology problems of interest are:

1. Characterizing cell-type interactions in cancer tissues: recent techniques such as multiplex immunohistochemistry allow researchers to study the spatial distribution of

cell types within cancerous tissues. In fact, multiplex images generate chromatic or labeled datasets, where data points are categorized by distinct labels and interactions exist among points of different categories. Hence, the description of interactions between cancer cells and immune cells help to understand the disease progression and therapeutic responses. Some references in the field include (Bull et al., 2020; Bull and Byrne, 2023; Bull et al., 2024a,b; Stolz et al., 2024; Cheung et al., 2024).

2. Understanding stem cell differentiation through spatial signaling: studying stem cell differentiation involves understanding how spatial and temporal dynamics of signaling pathways, like BMP, guide cells toward specific fates. Many papers, such as (Teague et al., 2024; Camacho-Aguilar et al., 2024; Etoc et al., 2016; Heemskerk et al., 2019), model the signals guiding cell state, in some cases, in a constrained spatial environment that leads to micropatterns, but no interaction between cells in an unrestricted environment has been considered so far.

In these scenarios, it is crucial not only to understand the overall spatial structure of the labelled point cloud, but also the relationships among different labeled subsets of data, such as cell types in biological systems. This shift in focus reveals the necessity for developing new TDA methods that can handle the complexity of labeled data, as traditional approaches are insufficient for capturing the full richness of these interrelationships. Recent advances have extended TDA to handle chromatic point clouds—datasets in which each point is assigned a label or "color". In the case of bicolored (or bilabeled) data, Stolz et al. (2024) proposes Dowker persistence to provide information about the topological features that are co-living in both colors at the same time. Other related work is Wagner et al. (2024), where they combine standard persistent homology with image persistent homology to define a way of characterizing shapes and interactions between two pointsets in any dimension. A richer information can be obtained by using the chromatic alpha filtration proposed by di Montesano et al. (2024a). They propose a geometric construction that generalizes the alpha filtration for the chromatic case and that allows to deal with several colors at once. This geometric construction allows to encode both proximity relations between points of the same color as well as those between points of different colors, keeping compatibility with the alpha complexes of each color set separately. A set of 6 persistence diagrams, including the one of the image persistent homology, can be derived from the inclusion of labeled datasets in a common embedded space.

2. Spatial relations from inclusion maps

One way of studying spatial relations in data $X \subset \mathbb{R}^d$ is to study the induced inclusion maps between filtrations and the corresponding maps on persistent homology. Then those maps on persistent homology can be studied in terms of invariants such as their kernel, cokernel, image, domain, and codomain (Cohen-Steiner et al., 2009).

More specifically, let $X = X_0 \sqcup \cdots \sqcup X_s \subset \mathbb{R}^d$ be a chromatic set, also denoted by (X, μ), where $\mu = \{X_0, \ldots, X_s\}$ is a coloring of X. If we can construct a filtration in such a way that it is functorial with respect to the input dataset, then we can study the induced inclusion maps among these filtered subcomplexes. For that aim, the well-known Vietoris–Rips,

$\mathrm{VR}_\bullet(X)$, and Čech, $\check{C}_\bullet(X)$, filtrations are suitable; but these constructions are filtrations of the complete simplicial complex on the vertex set X in which the number of simplices scales exponentially with the number n of data points, discarding them for practical applications with large datasets.

On the other hand, if X is in general position, then the alpha filtration, $A_\bullet(X)$, is a filtration of the Delaunay triangulation $\mathrm{Del}(X)$ associated with X, which is sparse in the sense that the number of simplices in $\mathrm{Del}(X)$ is bounded by a polynomial in n of degree $\lfloor d/2 \rfloor$, with d the dimension of the ambient space, so the computation is feasible when d is low. Moreover, $A_r(X)$ is homotopy equivalent to $\check{C}_r(X)$ for any nonnegative r, so the alpha filtration captures the same topological information as the Čech filtration. Unfortunately, alpha filtrations and the underlying Delaunay triangulations are not functorial with respect to inclusions of point clouds; that is, $\mathrm{Del}(X_0)$ is not necessarily a subcomplex of $\mathrm{Del}(X_0 \sqcup X_1)$.

This deficiency is addressed by the chromatic Delaunay triangulation $\mathrm{Del}(X, \mu)$ introduced by Biswas et al. (2022) (extending the previous concept of coupled alpha complex defined in Reani and Bobrowski (2021)), and the corresponding chromatic alpha filtration $A_\bullet(X, \mu)$, introduced in di Montesano et al. (2024a). Roughly speaking, the chromatic Delaunay triangulation $\mathrm{Del}(X, \mu)$ is constructed by lifting points of color i to another "copy" of the initial d-dimensional space, for each i, and taking the Delaunay triangulation of the lifted points in $\mathbb{R}^{d+s}$. The chromatic alpha filtration defined in di Montesano et al. (2024a) has the good properties of the functoriality of the Čech filtration (with respect to inclusion of points) and the sparsity of the alpha filtration, while preserving the same spatial information.

However, the algorithm for computing the chromatic alpha filtration value for a simplex is computationally expensive. In order to simplify this computation, we introduced in Natarajan et al. (2024) the chromatic Delaunay–Čech filtration $\mathrm{Del}\,\check{C}_\bullet(X, \mu)$ and the chromatic Delaunay–Rips filtration $\mathrm{Del}\,\mathrm{VR}_\bullet(X, \mu)$. These are filtrations of the chromatic Delaunay triangulation where the filtration value of each simplex is given by its filtration value in the Čech and Vietoris–Rips filtrations, respectively. Then, using Morse theory techniques, we proved that there are simplicial collapses

$$\check{C}_r(X) \searrow \mathrm{Del}\,\check{C}_r(X, \mu) \searrow A_r(X, \mu), \qquad \text{and also} \qquad \mathrm{Del}\,\check{C}_r(X, \mu) \searrow \mathrm{Del}\,\check{C}_r(X).$$

The former generalizes an analogous result for the non-chromatic setting in Bauer and Edelsbrunner (2016).

These results have practical implications for computations with large datasets where the chromatic alpha filtration values can become a computational bottleneck. The implementations of these filtrations are available in the Python package `chalc`: https://github.com/abhinavnatarajan/Chalc.

Now, starting from a filtration of a simplicial complex K,

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K,$$

applying the p-th homology functor, we get a sequence of vector spaces:

$$H_p(K_0) \longrightarrow \cdots \longrightarrow H_p(K_{i-1}) \longrightarrow H_p(K_i) \longrightarrow \cdots \longrightarrow H_p(K_n).$$

If L is a subcomplex of K, compatible with the filtration values, then

$$\begin{array}{ccccccccccc} H_p(K_0) & \longrightarrow & \cdots & \longrightarrow & H_p(K_i) & \longrightarrow & H_p(K_{i+1}) & \longrightarrow & \cdots & \longrightarrow & H_p(K_n) \\ \uparrow & & & & \uparrow & & \uparrow & & & & \uparrow \\ H_p(L_0) & \longrightarrow & \cdots & \longrightarrow & H_p(L_i) & \longrightarrow & H_p(L_{i+1}) & \longrightarrow & \cdots & \longrightarrow & H_p(L_n) \end{array}$$

where the vertical maps $\kappa_i : H_p(L_i) \to H_p(K_i)$ are induced by the inclusions $L_i \subseteq K_i$. These maps have kernels, images, and cokernels that are also persistence modules. Cohen-Steiner et al. (2009) provides explicit algorithms for computing persistent homology of them, and the `chalc` package also includes their implementation. In di Montesano et al. (2024a), using the chromatic alpha filtration and considering the subcomplex L generated by a subset of colors, they propose to study these persistence diagrams together with those of the domain (associated to the L), the codomain (associated to the complex K), and the relative homology $H_\bullet(K, L)$, forming what they call the *6-pack* of persistence diagrams.

In this way, some features of the shape of the point cloud are reflected in these diagrams. For example, in the two colors setting for 2D point sets, the existence of a loop of one color that is "filled" by points of the other color can be reflected as a prominent point in the kernel diagram (dimension 1); the existence of a loop formed by two arcs, one of each color, is represented by a long-lived feature in the cokernel (dimension 1).

3. Directions

Although the theory described in previous section provides a framework for real applications, there is no a straightforward path to follow. There are many choices that need to be carefully taken: what subcomplex shall we consider for the inclusion? What combinations of colors should generate each subcomplex? The choice should be "standard" in the sense that it is not dependent on the differences in density of the point sets. In di Montesano et al. (2024b), the authors propose the use of the so called *k-chromatic sub-filtration*, consisting of all simplices with points of at most k distinct colors. This way, the number of combinations are reduced and the process is somehow standarized. However, it is also desirable that the relevant points in the persistence diagrams have an intuitive interpretation, which also depends on the choice of the subcomplex. For example, as shown in Fig.1, the kernel diagram for the 2-chromatic subfiltration of the set of points on the left has an infinite feature at degree 1, which is not directly interpretable in the set of colored points. So, is there a more suitable subcomplex to provide intuitive information of the spatial relations between colors?

When restricting to the bicolored case, we do not have those artifacts using the 1-chromatic inclusion. Then, a natural question arises regarding what relations exist between Dowker persistence and the information that can be obtained from the 6-pack (or part of it). Can the information encoded in Dowker persistence always be deciphered inside the 6-pack? Are there any algebraic relations between them, or at least with some of the diagrams? Are there any specific organizational patterns for which Dowker persistence is advantageous?

Also, for the sake of interpretability, an important goal for us is to define numerical descriptors from those diagrams whose combination can claim robust statements about the spatial

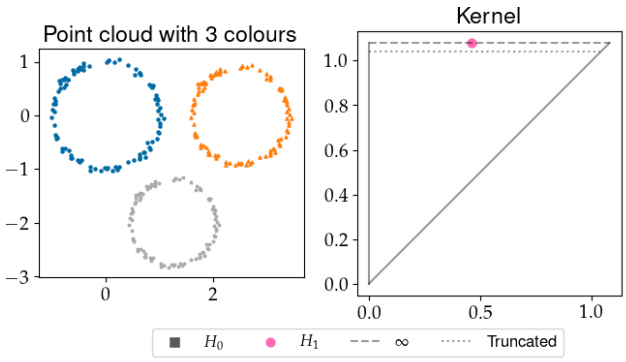

Figure 1: A trichromatic point set and the corresponding kernel persistence diagram using the 2-chromatic inclusion.

distribution of the labeled point clouds. In this sense, a descriptor was defined in di Montesano et al. (2024b), called the *MST-ratio*, which relates the lengths of bars of 0-homology classes of the disjoint union of the sets of points with the one of the union of points. This is something that can always be done without computing the 6-pack and involves only the domain and codomain. However, we think that the most interesting information about the interactions between the different sets of points lies in the kernel, image, and cokernel.

Moreover, we can pose the question whether all the diagrams are really necessary to characterize the spatial distribution. There are algebraic relations between the triples (kernel, domain, image), (image, codomain, cokernel), and (cokernel, relative homology, kernel), given by short exact sequences, that lead us to think that the kernel, image, and cokernel are the most powerful diagrams to characterize the spatial interactions of a labeled point cloud if we restrict to applications in 2D, like the ones we are concerned here.

In Wagner et al. (2024), they devise a new representation, the *mixup barcode*, in which they combine the information of the bars in the domain with the bars in the image. This is an important starting point for developing a whole set of *mingling numbers* to be defined out of the kernel, image, and cokernel to describe the spatial distribution of the labeled point sets, including, at least, the cases of two colors as well as three colors.

Our ultimate goal is to define a set of topological summaries that can characterize, at least, the organizational patterns that are usually of interest for biologists, like the ones listed in (Bull et al., 2024b): "unstructured", "aggregation", "exclusion", "architecture". However, when considering three or more colors, designing interpretable topological descriptors from the 6-pack of persistence diagrams (or part of it) for describing the shape of such labeled datasets becomes a tougher challenge.

Acknowledgements

These are ideas and directions we are heading towards in our recently started research, funded by the Spanish government, MCIU, under grant PID2024-155867NB-I00, in col-

laboration with Joshua A. Bull, Helen M. Byrne, Elena Camacho-Aguilar, Heather A. Harrington, Abhinav Natarajan, Miguel Navarro-Castro and Bernadette J. Stolz. The author also thanks the IMUS–María de Maeztu grant CEX2024-001517- M funded by MICIU/AEI/10.13039/501100011033.

References

Dashti Ali, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. ISSN 0162-8828.

Erik J Amézquita, Michelle Y Quigley, Tim Ophelders, Elizabeth Munch, and Daniel H Chitwood. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Developmental Dynamics*, 249:816–833, 2020. ISSN 1058-8388.

Ulrich Bauer and Herbert Edelsbrunner. The morse theory of Čech and delaunay complexes. *Transactions of the American Mathematical Society*, 369, 2016. ISSN 0002-9947. doi: 10.1090/tran/6991.

Katherine Benjamin, Aneesha Bhandari, Jessica D Kepple, Rui Qi, Zhouchun Shang, Yanan Xing, Yanru An, Nannan Zhang, Yong Hou, and Tanya L Crockford. Multiscale topology classifies cells in subcellular spatial transcriptomics. *Nature*, pages 1–7, 2024. ISSN 0028-0836.

Ranita Biswas, Sebastiano Cultrera di Montesano, Ondřej Draganov, Herbert Edelsbrunner, and Morteza Saghafian. On the size of chromatic delaunay mosaics. *arXiv preprint arXiv:2212.03121*, 2022.

Joshua A. Bull and Helen M. Byrne. Quantification of spatial and phenotypic heterogeneity in an agent-based model of tumour-macrophage interactions. *PLoS Computational Biology*, 19, 2023. ISSN 15537358. doi: 10.1371/journal.pcbi.1010994.

Joshua A. Bull, Philip S. Macklin, Tom Quaiser, Franziska Braun, Sarah L. Waters, Chris W. Pugh, and Helen M. Byrne. Combining multiple spatial statistics enhances the description of immune cell localisation within tumours. *Scientific Reports*, 10, 2020. ISSN 20452322. doi: 10.1038/s41598-020-75180-9.

Joshua A Bull, Eoghan J Mulholland, Simon J Leedham, and Helen M Byrne. Extended correlation functions for spatial analysis of multiplex imaging data. *Biological Imaging*, 4:e2, 2024a. ISSN 2633-903X.

Joshua A Bull, Eoghan J Mulholland, Joshua W Moore, Jesús J Bosque, Bernadette J Stolz, Joseph Boen, Holly R Eggington, Hayley L Belnoue-Davis, Helen Jones, and Chandler D Gatenbee. Integrating diverse statistical methods to analyse stage-discriminatory cell interactions in colorectal neoplasia. *bioRxiv*, pages 2024–2026, 2024b.

Elena Camacho-Aguilar, Sumin T Yoon, Miguel A Ortiz-Salazar, Siqi Du, M Cecilia Guerra, and Aryeh Warmflash. Combinatorial interpretation of bmp and wnt controls the decision

between primitive streak and extraembryonic fates. *Cell Systems*, 15:445–461, 2024. ISSN 2405-4712.

Gunnar Carlsson and Vin de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10, 2010. ISSN 16153375. doi: 10.1007/s10208-010-9066-0.

Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. In *Proceedings of the twenty-third annual symposium on Computational geometry*, pages 184–193, 2007.

Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete and Computational Geometry*, 42, 2009. ISSN 01795376. doi: 10.1007/s00454-009-9176-0.

Alison M Cheung, Wenchao Han, Snow Zhou, Dan Wang, Vishwesh Ramanathan, Owen Qu, Anne L Martel, and Martin J Yaffe. Spatial analysis of immune cells in breast cancer using k-nearest neighbor graphs and louvain-community clustering of immunofluorescent protein multiplexing images. In *17th International Workshop on Breast Imaging (IWBI 2024)*, volume 13174, pages 261–269. SPIE, 2024.

David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistent homology for kernels, images, and cokernels. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 2009. doi: 10.1137/1.9781611973068.110.

S Cultrera di Montesano, ONDREJ Draganov, HERBERT Edelsbrunner, and MORTEZA Saghafian. Chromatic alpha complexes. *Foundations of Data Science*, 2024a.

Sebastiano Cultrera di Montesano, Ondrej Draganov, Herbert Edelsbrunner, and Morteza Saghafian. Chromatic topological data analysis. *arXiv preprint arXiv:2406.04102*, 2024b.

Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

Fred Etoc, Jakob Metzger, Albert Ruzo, Christoph Kirst, Anna Yoney, M. Zeeshan Ozair, Ali H. Brivanlou, and Eric D. Siggia. A balance between secreted inhibitors and edge sensing controls gastruloid self-organization. *Developmental Cell*, 39, 2016. ISSN 18781551. doi: 10.1016/j.devcel.2016.09.016.

Idse Heemskerk, Kari Burt, Matthew Miller, Sapna Chhabra, M. Cecilia Guerra, Lizhong Liu, and Aryeh Warmflash. Rapid changes in morphogen concentration control self-organized patterning in human embryonic stem cells. *eLife*, 8, 2019. ISSN 2050084X. doi: 10.7554/eLife.40526.

Yusuke Imoto and Yasuaki Hiraoka. V-mapper: topological data analysis for high-dimensional data with velocity. *Nonlinear Theory and Its Applications, IEICE*, 14, 2023. doi: 10.1587/nolta.14.92.

Abhinav Natarajan, Thomas Chaplin, Adam Brown, and Maria-Jose Jimenez. Morse theory for chromatic delaunay triangulations. *arXiv preprint arXiv:2405.19303*, 2024.

Yohai Reani and Omer Bobrowski. A coupled alpha complex. *arXiv preprint arXiv:2105.08113*, 2021.

Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *4th Symposium on Point Based Graphics, PBG@Eurographics 2007*, 2007.

Bernadette J Stolz, Jagdeep Dhesi, Joshua A Bull, Heather A Harrington, Helen M Byrne, and Iris H R Yoon. Relational persistent homology for multispecies data with application to the tumor microenvironment. *Bulletin of Mathematical Biology*, 86:128, 2024. ISSN 0092-8240.

Seth Teague, Gillian Primavera, Bohan Chen, Zong Yuan Liu, Li Ang Yao, Emily Freeburne, Hina Khan, Kyoung Jo, Craig Johnson, and Idse Heemskerk. Time-integrated bmp signaling determines fate in a stem cell model for early human development. *Nature Communications*, 15, 2024. ISSN 20411723. doi: 10.1038/s41467-024-45719-9.

Oliver Vipond, Joshua A Bull, Philip S Macklin, Ulrike Tillmann, Christopher W Pugh, Helen M Byrne, and Heather A Harrington. Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. *Proceedings of the National Academy of Sciences*, 118:e2102166118, 2021. ISSN 0027-8424.

Hubert Wagner, Nickolas Arustamyan, Matthew Wheeler, and Peter Bubenik. Mixup barcodes: quantifying geometric-topological interactions between point clouds. *arXiv preprint arXiv:2402.15058*, 2024.

Jingjie Yang, Heidi Fang, Jagdeep Dhesi, Iris HR Yoon, Joshua A Bull, Helen M Byrne, Heather A Harrington, and Gillian Grindstaff. Topological classification of tumour-immune interactions and dynamics: J. yang et al. *Journal of Mathematical Biology*, 91(3):25, 2025.