Persistent entropy: a scale-invariant topological statistic for analyzing cell arrangements

N. Atienza, L. M. Escudero, M. J. Jimenez, M. Soriano-Trigueros

May 20, 2019

Abstract

In this work, we develop a method for detecting differences in the topological distribution of cells forming epithelial tissues. In particular, we extract topological information from their images using persistent homology and a summary statistic called persistent entropy. This method is scale invariant, robust to noise and sensitive to global topological features of the tissue. We have found significant differences between chick neuroepithelium and epithelium of Drosophila wing discs in both, larva and prepupal stages. Besides, we have tested our method, with good results, with images of mathematical tesselations that model biological tissues.

remark: this paper has been submitted to the journal pattern recognition letter.

1 Introduction

Topology is the branch of mathematics that deals with properties of space that keep invariant under continuous transformations. These properties are extremely important in networks.

Nowadays, computational topology and geometry are playing an increasing role in quantitative biology and biomedical engineering. In particular, the data analysis tool persistent homology [1, 2] has been successfully applied in fields such as tumor segmentation [3], analysis of biological networks [4] or diagnostic of chronic obstructive pulmonary disease [5].

Epithelia are packed tissues formed by tightly assembled cells. Their apical surfaces are similar to convex polygons forming a natural tessellation. Recently, many studies have been focused on their natural organization, in part justified because changes in this organization may indicate an early onset of disease [6, 7]. Therefore, being able to quantify differences in their topological distribution is an interesting problem.

The study of epithelial organization has been mainly focused on the distributions of the number of neighbors of the cells. In [8], the authors compared the polygon distributions of natural and mathematical tesselations of the plane (called centroidal Voronoi tessellations or CVTs); in [9], tissues are compared using the graphlet degree distribution agreement distance (GDD), which measures differences in terms of small subgraphs (graphlets) contained in the network representing the cells and their neighboring relations. Despite their achievements, both measurements rely on local properties of the network. Therefore, a value sensitive to global topological properties may offer new insights. In particular, the fact that such a value was a real number is advantageous for two reasons: first, due to the small number of samples available for this kind of study, univariate statistical tests are the best option; second, a number could be easily combined with other values in the Epigraph tool [9] to improve it.

Two examples of topological statistics which are numbers and are able to measure global features of the network are persistent entropy [10, 11] and the sum of the bars in the persistence barcode (see barcode definition in section 2), which appears in [5] together with the upper star filtration under the name of upwards complexity.

In this paper, we show how persistent entropy is useful for reflecting different topological distributions in epithelial tissues. A first approach to use Topological Data Analysis for that aim was presented in

[12]. There, a weighted network was constructed to model the cell arrangements and persistent homology was computed to describe it. However, alpha complexes are known as a better model to represent a map of convex regions as it is the case of a segmented image of a biological tissue. For that reason, we started an experimental work in [13] using such a model and computing persistent entropy. In this paper, we improve the method, focusing in the ability of describing different topological arrangements of cells independently of the scale of the image. We also further analyse statistically the experimental results to reach well-founded conclusions.

2 Material and methods

In this section, we recall the tools from Topological Data Analysis (TDA) used and then, how we adapt them to the posed problem.

2.1 TDA concepts

Simplicial complex and filtration An *n-simplex* is the convex hull of a set of n linearly independent points $\tau = \{p_1, \dots, p_n\}$. Each k-simplex contained in τ with $k < n$ is called a *face.* A *simplicial complex* $\mathcal{K}$ is formed by a set of simplices satisfying:

1. Every face of a simplex in $\mathcal{K}$ is also in $\mathcal{K}$.

2. The intersection of two simplices in $\mathcal{K}$ is a face of both simplices.

The *dimension* of a simplicial complex is the maximum of the dimensions of its simplices.

A *filtration* over a simplicial complex $\mathcal{K}$ is a finite increasing sequence of simplicial complexes

$$\mathcal{K}_1 \subset \mathcal{K}_2 \ldots \subset \mathcal{K}_n = \mathcal{K}$$

It is commonly defined using a monotonic function $f : \mathcal{K} \to \mathbb{R}$ by which we mean that for any two simplices $\delta, \tau \in \mathcal{K}$, if δ is a face of τ, then $f(\delta) \leq f(\tau)$. That way, if $a_1 \leq \ldots \leq a_n$ are the function values of all the simplices in $\mathcal{K}$, then the subcomplexes $\mathcal{K}_i = f^{-1}(-\infty, a_i]$, for $i = 1 \ldots n$ define a filtration over $\mathcal{K}$.

Voronoi diagram and α-complex A *Voronoi diagram* is a partitioning of the plane depending on a set of vertices $\{v_1, \dots, v_n\}$: for each vertex v_i we define the function $d_i(x) = d(v_i, x)$ and a region given by

$$V_i = \{x \mid d_i(x) \leq d_j(x), \quad \forall j = 1, \cdots n\}.$$

That is, each region is formed by points for which that vertex is the closest one (see the first image in Figure 1). An *α-complex* is a filtration [14, p. 68] that can be defined starting from a Voronoi diagram. Consider B_r^i as the ball of center v_i and radius r. For each r, consider the region $U_r^i = B_r^i \cap V_i$ and define the simplicial complex $\mathcal{K}_r$ with simplices

$$\tau = [v_0 \dots v_k] \in \mathcal{K}_r \Leftrightarrow \bigcap_{i=0\dots k} U_r^i \neq \emptyset.$$

In other words, a simplex is in $\mathcal{K}_r$ if and only if the intersection of balls with radius r and centers its vertices together with their Voronoi regions is not empty. If the points are in general position in the plane, no i-simplex will arise with i greater than 2 and the final simplicial complex is known as *Dealunay triangulation* [14, p. 63], see Figure 1.

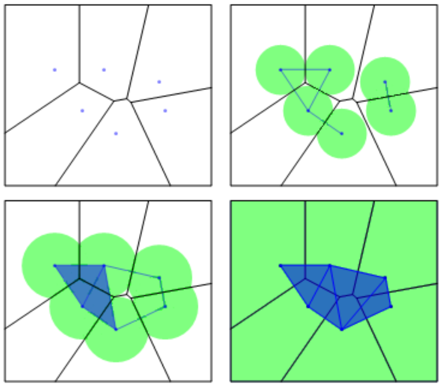


Figure 1: An example of alpha complex filtration. The first image coincides with the Voronoi diagram and the last one with the Dealunay triangulation

Persistent homology and barcodes In Algebraic Topology, the concept of *homology class* allows to define holes rigorously and compute them using linear algebra. Intuitively, an n-dimensional hole in a simplicial complex is a cavity when $n = 2$, a cycle when $n = 1$ and a connected component when $n = 0$. *Persistent homology* is the main tool in TDA defined for tracking the persistence (birth, lifetime and death) of holes along a filtration. For example, in the case of an α-complex constructed from a point cloud on the plane, 0-dimensional and 1-dimensional persistent homology describe, respectively, the evolution of connected components and cycles as the radius of the balls increase.

The persistence of each n-dimensional hole is represented using an interval of the form $[a, b]$, where a is the birth time, if K_a is the simplicial complex (in the filtration) where the hole first appeared and b is the death time if K_b is the first simplicial complex (in the filtration) where the hole disappeared. If the hole remains until the final simplicial complex, we write $b = \infty$. This representation is called the n-dimensional *persistence barcode* and each interval a *bar*. We give an example in figure 2. A formal definition of homology and persistent homology together with fast algorithms for computing it and proofs of robustness to noise can be found in [14].

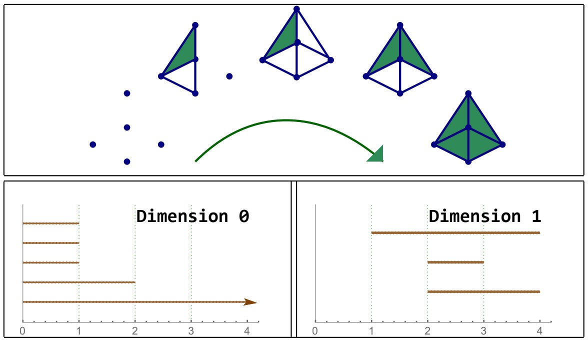
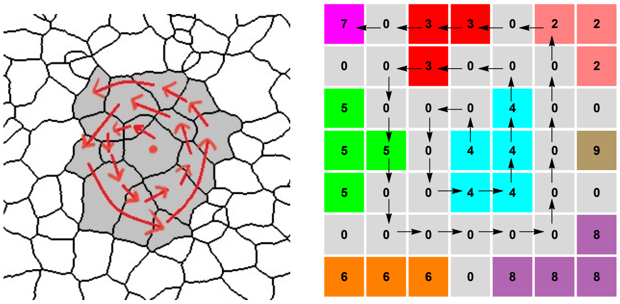


Figure 2: Top: example of a filtration $\mathcal{K}$. Bottom: barcodes representing connected components and cycles.

Persistent entropy Barcodes are very intuitive but their statistical analysis is rather complex [15]. Landscapes are an equivalent representation which are more comfortable for it, but statistics on them lack of specifics tests so far and rely on the law of large numbers [16]. Therefore, in order to perform a useful



7	0	3	3	0	2	2
0	0	3	0	0	0	2
5	0	0	0	4	0	0
5	5	0	4	4	0	9
5	0	0	4	4	0	0
0	0	0	0	0	0	8
6	6	6	0	8	8	8

Figure 3: Left: Intuition behind the algorithm. We pick cells following a spiral until the desired number is reached. Right: Toy example. Taking as input $n = 5$, the output is the set of labels $\mathcal{C} = \{4, 3, 5, 2, 7\}$.

statistical analysis of persistent homology for small samples, we need a real number that encapsulates the information contained in the barcode. *Persistent entropy* [10, 11] is an stable topological statistic [17] and can be seen as an adaptation of Shannon entropy (Shannon index in ecology) to the persistent homology context. Given a barcode without infinite bars $B = \{[a_i, b_i]\}_{i=1...n}$ consider the length of the bars $\ell_i = b_i - a_i$ and their sum $L(B) = \ell_1 + \ldots + \ell_n$. Then, its *persistent entropy* is:

$$PE(B) = \sum_{i=1}^{n} -\frac{\ell_i}{L(B)} \log\left(\frac{\ell_i}{L(B)}\right)$$

Note that when calculating persistent entropy the lengths of the bars are normalized with respect to their total sum.

2.2 Normalization of number of cells and selection method

The common workflow in persistent homology consists in taking the data, defining a filtration over it and computing its persistent homology and the associated barcode. For this study, we will take the centroids of the cells in each image as input data, use the Voronoi diagram to approximate the boundaries of the cells and compute its α-complex. After that, we will analyze it statistically using persistent entropy. Nevertheless, in order to make sure that persistent entropy measures what we want, we need to adapt it to our problem.

Persistent entropy is very sensitive to the number n of bars appearing in the barcode. A usual technique in information theory is to divide the result by $\log(n)$ in order to obtain a value between 0 and 1. Unfortunately, apart from loosing the stability properties [17], the variance of $PE(B)/\log(n)$ still depends on n and therefore the number of bars affects to this summary statistic. If we want to be sure we are finding differences in the topological distribution and not in the number of cells appearing in the image, we must set a common number of cells in all the processed images. This way, the 0-dimensional persistence barcodes will have the same number of bars (one per cell) and although there may be differences in the number of cycles (1-dimensional persistent homology) this would be good, since it would represent differences in the topological distribution.

Now, the question is how to select the cells that will take part in the analysis. In order to avoid changes in the topological arrangement of the selected cells with respect to the original image, we follow Algorithm 1 to take the cells following a spiral (Figure 3, left). Note that this process is expected to work well due to the convexity of the cells. Our input is a labeled image M, where each region corresponding to a cell is labeled with a natural number and the boundaries are tagged with 0, as shown in Figure 3, right.

Algorithm 1 Spiral selection of regions

```
procedure SPIRAL($M, n$)                                   ▷ $M$ is an image and $n$ a number
    $\mathcal{C} := \{\,\}$
    $(x, y) := center\,(M)$                                ▷ central coordinates of $M$
    if $M(x, y) \neq 0$ then
        $\mathcal{C} := \{M(x, y)\}$
    end if
    $i := 0$
    while $\#\mathcal{C} < n$ do                            ▷ # is the number of elements
        $i := i + 1$
        for $j \in (1, \ldots, i)$ do                       ▷ repeat $i$ times
            if $\#\mathcal{C} < n$ then
                $x := x + (-1)^i$
                if $M(x, y) \neq 0$ and $M(x, y) \notin \mathcal{C}$ then
                    $\mathcal{C} := \mathcal{C} \cup \{M(x, y)\}$
                end if
            end if
        end for
        for $j \in (1, \ldots, i)$ do                       ▷ repeat $i$ times
            if $\#\mathcal{C} < n$ then
                $y := y + (-1)^i$
                if $M(x, y) \neq 0$ and $M(x, y) \notin \mathcal{C}$ then
                    $\mathcal{C} := \mathcal{C} \cup \{M(x, y)\}$
                end if
            end if
        end for
    end while
    return $\mathcal{C}$                                    ▷ return the first $n$ labels around the center
end procedure
```

2.3 Scale invariance of the method

In addition, we expect this method to be scale invariant. Otherwise, the result may vary from images of one tissue to other even if they have the same topological distribution. Persistent entropy, by definition, is invariant with respect to the scale of the barcode with finite bars. Nevertheless, in barcodes with infinity bars this invariance may not be satisfied depending on the treatment we make of these bars. In our case, as α-complex always have one connected component and none cycles for a radius big enough, there is only one 0-dimensional infinity bar which is born at moment 0, so that it gives no information about the topology of the cells and it can be removed. Actually, if we do not remove it but limit its endpoint to a fixed value, we eliminate the scale invariance of the method since the same pattern re-scaled will have different barcodes as shown in Figure 4. Therefore, we remove the infinity bar of all barcodes in this experiment. This is the main improvement with respect to the communication [13] and the differences in the results are analyzed in Section 3.3.

2.4 Proposed method

We summarize here the steps of our method (see Figure 6 for a sketch):

1. Set a fixed number of cells and apply Algorithm 1.

2. Consider the point cloud given by the centroids of the selected cells and construct the α-complex.

3. Compute 0 and 1-dimensional persistence barcodes.

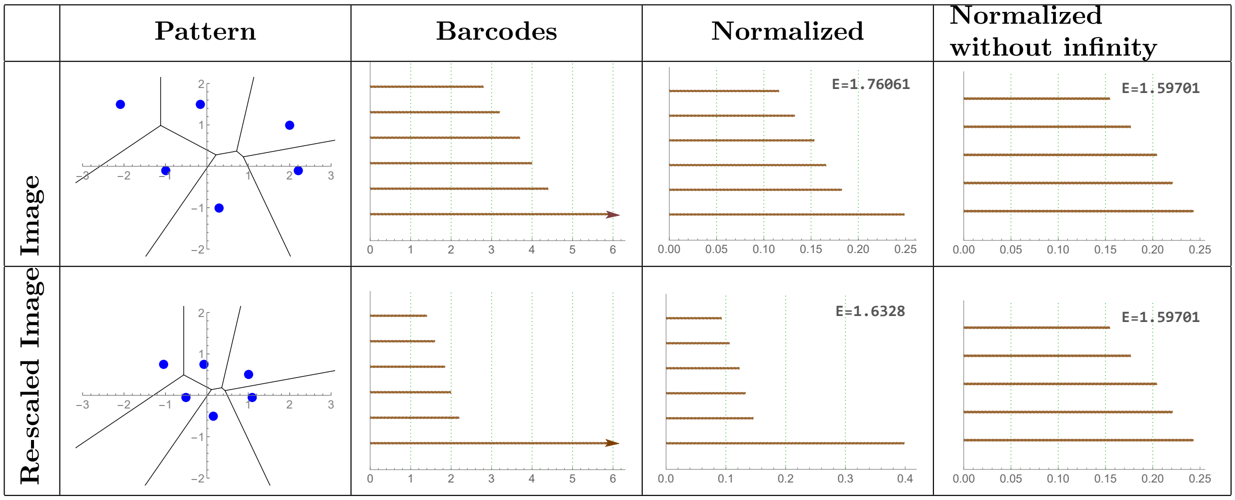


Figure 4: The Voronoi diagrams in the first column are computed from images with the same pattern differing with a scalar factor of two. In the barcodes of the second column the infinite bar have been limited to 6. In the third column, the normalization of the bars produces different barcodes. The last column shows how normalizing bars after removing the infinity bar provides the same barcodes and, hence, equal entropy.

4. Remove the only infinity bar in the 0-dimensional barcode and compute persistent entropies PE_0 and PE_1.

5. Perform a statistical study to analyze the results.

3 Computational experiments

3.1 Data and experiments

The images for our experiments come from three types of epithelial tissues and three artificial tesselations (called centroidal Voronoi tessellations or CVTs). Tissues are chick neuroepithelium (cNT), wing disc in the larva (dWL) and prepupal stages (dWP) from Drosophila, see Figure 5, second row. The first is relatively distinct to the other two so we expect to find differences in this topological distribution. The other two tissues are taken from two proliferative stages separated by only 24 hours of developments, and topological differences (if exist) are expected to be very difficult to prove. Further information about the way images were obtained and segmented can be found in [18].

CVTs are created inductively. In the first step, images are given by the Voronoi diagram of random set of points generated using a poisson distribution. The following images are created taking the centroid of each Voronoi region and computing a new Voronoi diagram. Iterating, we obtain a CVT path of images. We call the set images obtained in the Nth step as CVT N, see Figure 5, first row, for some examples.

In [8] cNT, dWL and dWP were compared with CVTs using the proportion of polygons of each image. Multivariate statistical test was not able to find differences between CVT4-dWL and CVT5-dWP. Significant differences between cNT and CVT1, CVT2 were found with a narrow margin. It was conjectured that the distribution of polygons in each tissue could be compared with the existing in the CVT paths, with dWL and dWP being equivalent to CVT4 and CVT5 respectively and cNT lying between CVT1 and CVT2.

We will assign persistent entropy values to each image (PE_0 and PE_1, corresponding respectively to the zero and one dimensional persistence barcodes) and then apply univariate non parametric statistical tests to find significant differences in their distributions. We use the techniques explained in the previous

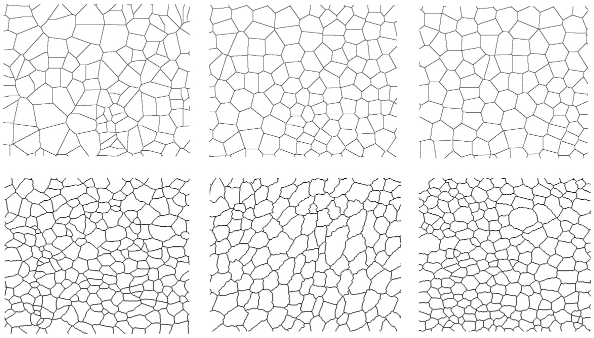


Figure 5: Top row: first, fourth and fifth image of a CVT path. Bottom row: images of tissues coming from cNT, dWL and dWP.

Table 1: The number of cells in each picture.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
cNT	666	661	565	573	669	532	419	592	743	527	594	473	704	747	469	834
dWP	748	805	566	414	454	654	751	713	503	430	387	516	413	455	271	249
dWL	426	555	491	522	510	936	890	789	977	913	604	835	785	747	622	

section, as displayed in Figure 6. Table 1 shows the number of valid cells (completely contained) in each image. Taking into account the normalization in the number of cells explained in 2.2 and that we do not want to discard any image, we consider 245 cells from each one. Functions used can be found in `https://github.com/Cimagroup` and require [19] and [20].

3.2 Results

Comparing CVTs We calculate PE_0 and PE_1 of the barcodes obtained from the first 245 cells in each image of CVT1, CVT4 and CVT5. Results are shown in Figure 7. First, we perform a Kruskal-Wallis test to see if there are significant differences between the three groups and then a Dunn test for pairwise comparison. We could distinguish all of them as shown in Table 2.

Comparing tissues We repeat the experiment with the first 245 cells in each cNT, dWL and dWP image. Results are displayed in Figure 8. Again, we perform a Kruskal-Wallis test to see if there are significant differences between the three groups and then a Dunn test to find pairwise differences. Significant differences were found between cNT and dWP, dWL but not between these last two. See Table 3.

We have also compared persistent entropy values between cNT - CVT1, dWL - CVT4 and dWP - CVT5 using the MannWhitney U test, finding significant differences between them (see Table 4).

Table 2: Kruskal-Wallis Test and Dunn Test for comparing the persistent entropies of CVT images.

KWT	PE_0	PE_1
p-value	3.398e-10	9.554e-11

DT p-value adjusted	CVT 1 vs 4	CVT 4 vs 5	CVT 5 vs 1
PE_0	2.165e-05	3.9e-02	3.054e-10
PE_1	7.305e-05	9.122e-03	4.889e-11

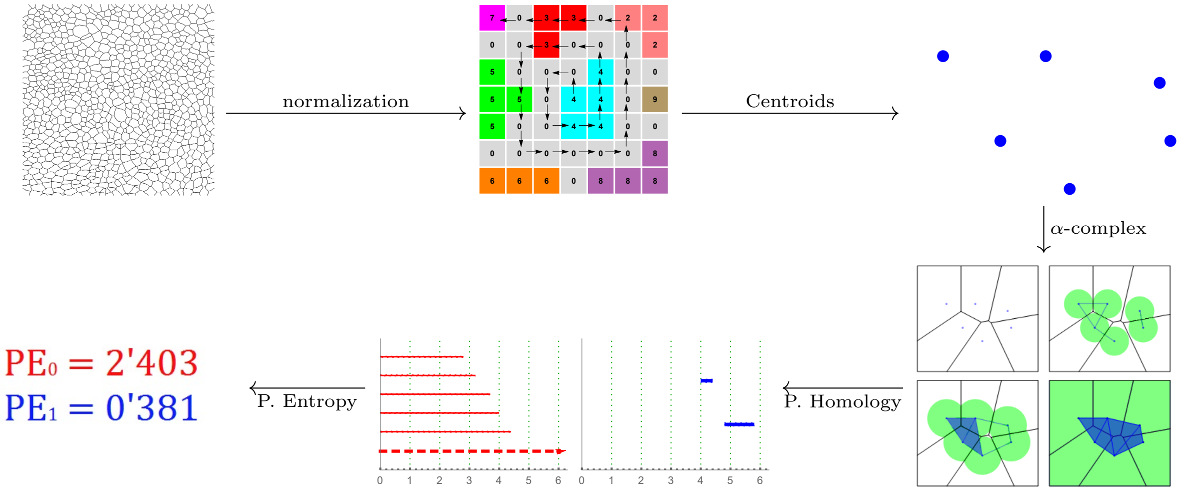


Figure 6: Steps followed to obtain the persistent entropy values. First, we extract the same number of cells from each image using the spiral algorithm. Then, calculate their centroids and α-complexes. Finally, we obtain their 0 or 1st dimensional barcode and calculate its persistent entropy.

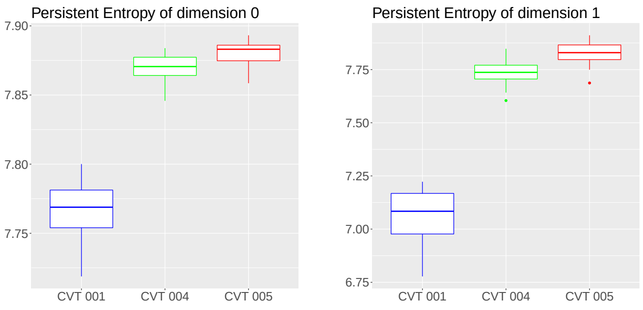


Figure 7: Boxplot representation of the results for the first, fourth and fifth images of a set of CVT paths.

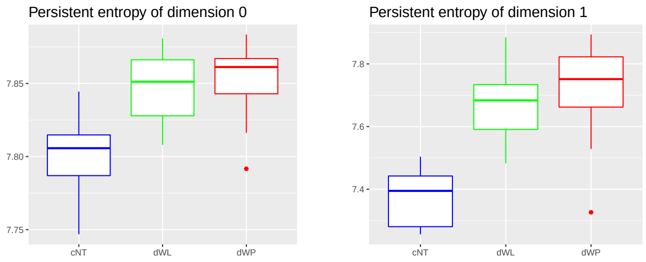


Figure 8: Boxplot representation of persistent entropy for the zero and first persistent homology of the three tissues.

Table 3: Kruskal-Wallis Test and Dunn Test for comparing the persistent entropies of tissues images.

KWT	PE_0	PE_1
p-value	1.846e-05	7.731e-07

DT p-value adjusted	dWL vs dWP	cNT vs dWL	dWP vs cNT
PE_0	0.653	3.561e-04	5.93 e-05
PE_1	0.3765	1.145e-04	1.816e-06

Table 4: Comparing cNT, dWL and dWP with CVT1, CVT4 and CVT5 respectively, using the U of Mann-Whitney test.

U-MW p-value adjusted	cNT vs CVT1	dWL vs CVT4	dWP vs CVT5
PE_0	0.001	0.002	2.454e-05
PE_1	2.737e-10	0.028	0.009

3.3 Comparison with previous results

A previous version of the method explained in this paper was discussed in the communication [13]. The main difference lies in how the infinity bar of the 0-dimensional persistence barcode is treated. As we have reasoned in 2.3, limiting the infinity bar to a fixed value may affect to the scale invariance of entropy. However, in [13] we kept the infinity bar, considering its length to be the maximum of all the values of the filtration function determining the α-complex. This affected the scale invariance of the method, and therefore the size of the image changed the value of persistent entropy. In that experiment, the p-values of Kruskal-Wallis Test for both persistent entropies of 0 and 1-dimensional barcodes, were also under 0.05. As it can be seen in Table 5, the p-value in the 0-dimensional case of dWL vs dWP is much smaller than the obtained when removing the infinity bar. This fact suggests that perhaps the differences detected in the previous method were due to change of scale between images dWL and dWP and not because of a different topological distribution, that is, the tissues may present similar organization but with cells of different sizes. Therefore, if we use a statistic sensitive to scale, like the sum of the lengths $L(B)$, we should find significant differences between dWL and dWP. This is exactly the case, see Table 6.

3.4 Discussion about the number of cells

In 2.2 we reasoned that it was necessary a normalization in the number of cells. In order to have a broader view, we have repeated the experiment again up to 385 cells (removing the two dWP images with ID numbers 15 and 16, which have fewer valid cells), and plotted the p-values of each case. When examining real tissues, differences (if existing) are found even with few cells: depending on the tissue, it might be necessary from 5 to 20 cells. Differences between CVTs are also found with few cells, except for CVT4 vs CVT5. In this case, PE_1 performs much better than PE_0 and 150 cells are required to start finding differences. Finally, when comparing dWL with CVT4 and dWP with CVT5, PE_0 works better than PE_1. In the first case, requiring 150 cells to find differences while in the second, only 30. These plots

Table 5: Dunn Test results when the infinity bar is not removed

DT p-value adjusted	dWL vs dWP	cNT vs dWL	dWP vs cNT
PE_0	0.0267	0.01600541	7.574294e-06
PE_1	0.3272	5.792e-05	2.162e-06

Table 6: Dunn Test for the sum of the lengths

DT p-value adjusted	dWL vs dWP	cNT vs dWL	dWP vs cNT
PE_0	0.0098	0.2335	0.1504

can be found in Figures 9, 10, 11. Note that, in all the cases in which we have found significant differences, there is a minimum number of cells from which the existence of those differences are guaranteed. This is, in fact, what is expected from a good measure of topological distributions.

4 Conclusions

We have used persistent entropy to find differences between the topology of the epithelium cNT and the topologies of both, dWL and dWP. This method is robust to noise, scale invariant and sensitive to global properties of the cell arrangement, making it a great statistic for summarizing the topology of any distribution. These characteristics above are particularly important for the analysis of biological and biomedical samples which often present a limited availability. In particular, this technique opens the door to perform non-parametric tests for finding differences in the topological distribution of two biological structures.

The method has been designed to analyze 2D surfaces of epithelial tissues and we hope this technique may help to improve existing topological tools for tissues such as (19). However, recent works make very clear that the biological complexity of epithelial organization lies in their 3D architecture [21]. Our procedure can be generalized to analyze the connectivity of the cells in 3D tissues. Therefore, this approach has the potential to quantify cell packing information from physiological conditions that can be compared with mutants and disease samples.

References

[1] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discret. & Comput. Geom.*, 28(4):511–533, November 2002.

[2] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discret. & Comput. Geom.*, 33(2):249–274, November 2004.

[3] T. Qaiser, Y.-W. Tsang, and N. Rajpoot et al. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Méd Image Anal.*, 55:1–14, July 2019.

[4] E. Merelli, M. Rucco, P. Sloot, and L. Tesei. Topological characterization of complex systems: Using persistent entropy. *Entropy*, 17(12):6872–6892, October 2015.

[5] F. Belchi, M. Pirashvili, and J. Brodzki et al. Lung topology characteristics in patients with chronic obstructive pulmonary disease. *Sci. Rep.*, 8(1), March 2018.

[6] V. Emmanuele, A. Kubota, and M. Hirano et al. Fhl1 w122s causes loss of protein function and late-onset mild myopathy. *Hum. Mol. Genet.*, 24(3):714–726, September 2014.

[7] J.-Ah Park, J. H. Kim, and J.J. Fredberg et al. Unjamming and cell shape in the asthmatic airway epithelium. *Nat. Mater.*, 14(10):1040–1048, August 2015.

[8] D. Sanchez-Gutierrez, M. Tozluoglu, and L. M. Escudero et al. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO J.*, 35(1):77–88, November 2015.

[9] P. Vicente-Munuera, P. Gómez-Gálvez, and L.M. Escudero et al. Epigraph: an open-source platform to quantify epithelial organization. *bioRxiv:217521*, 2019.

[10] H. Chintakunta, T. Gentimis, R. Gonzalez-Diaz, M.J. Jimenez, and H. Krim. An entropy-based persistence barcode. *Pattern Recognit.*, 48(2):391–401, February 2015.

[11] M. Rucco, F. Castiglione, E. Merelli, and M. Pettini. Characterisation of the idiotypic immune network through persistent entropy. In *Proceedings of ECCS 2014*, pages 117–128. Springer International Publishing, 2016.

[12] M. J. Jimenez, M. Rucco, P. Vicente-Munuera, P. Gómez-Gálvez, and L. M. Escudero. Topological data analysis for self-organization of biological tissues. In *Lect. Notes in Comput. Sci.*, pages 229–242. Springer International Publishing, 2017.

[13] N. Atienza, L. M. Escudero, M. J. Jimenez, and M. Soriano-Trigueros. Characterising epithelial tissues using persistent entropy. In *Computational Topology in Image Context*, pages 179–190. Springer International Publishing, December 2018.

[14] H. Edelsbrunner and J.L. Harer. *Computational topology : an introduction.* American Mathematical Society, 2010.

[15] Y. Mileyko, S. Mukherjee, and J. Harer. Probability measures on the space of persistence diagrams. *Inverse Probl.*, 27(12):124007, nov 2011.

[16] P. Bubenik. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, January 2015.

[17] N. Atienza, R. González-Díaz, and M. Soriano-Trigueros. On the stability of persistent entropy and new summary functions for TDA. *arxiv*, abs/1803.08304, 2018.

[18] Luis M. Escudero, Luciano da F. Costa, and M. Madan Babu et al. Epithelial organisation revealed by a network of cellular contacts. *Nat. Commun.*, 2(1), sep 2011.

[19] B.T. Fasy, J. Kim, F. Lecci, C. Maria, and V. Rouvreau. *TDA: Statistical Tools for Topological Data Analysis*, 2017. The included GUDHI is authored by Maria, C. and Dionysus by Morozov, D. and PHAT by Bauer, U., Kerber, M., Reininghaus, J. R package version 1.6.

[20] D.H. Ogle, P. Wheeler, and A. Dinno. *FSA: Fisheries Stock Analysis*, 2018. R package version 0.8.22.

[21] Pedro Gómez-Gálvez, Pablo Vicente-Munuera, and Luis M. Escudero et al. Scutoids are a geometrical solution to three-dimensional packing of epithelia. *Nature Communications*, 9(1), July 2018.

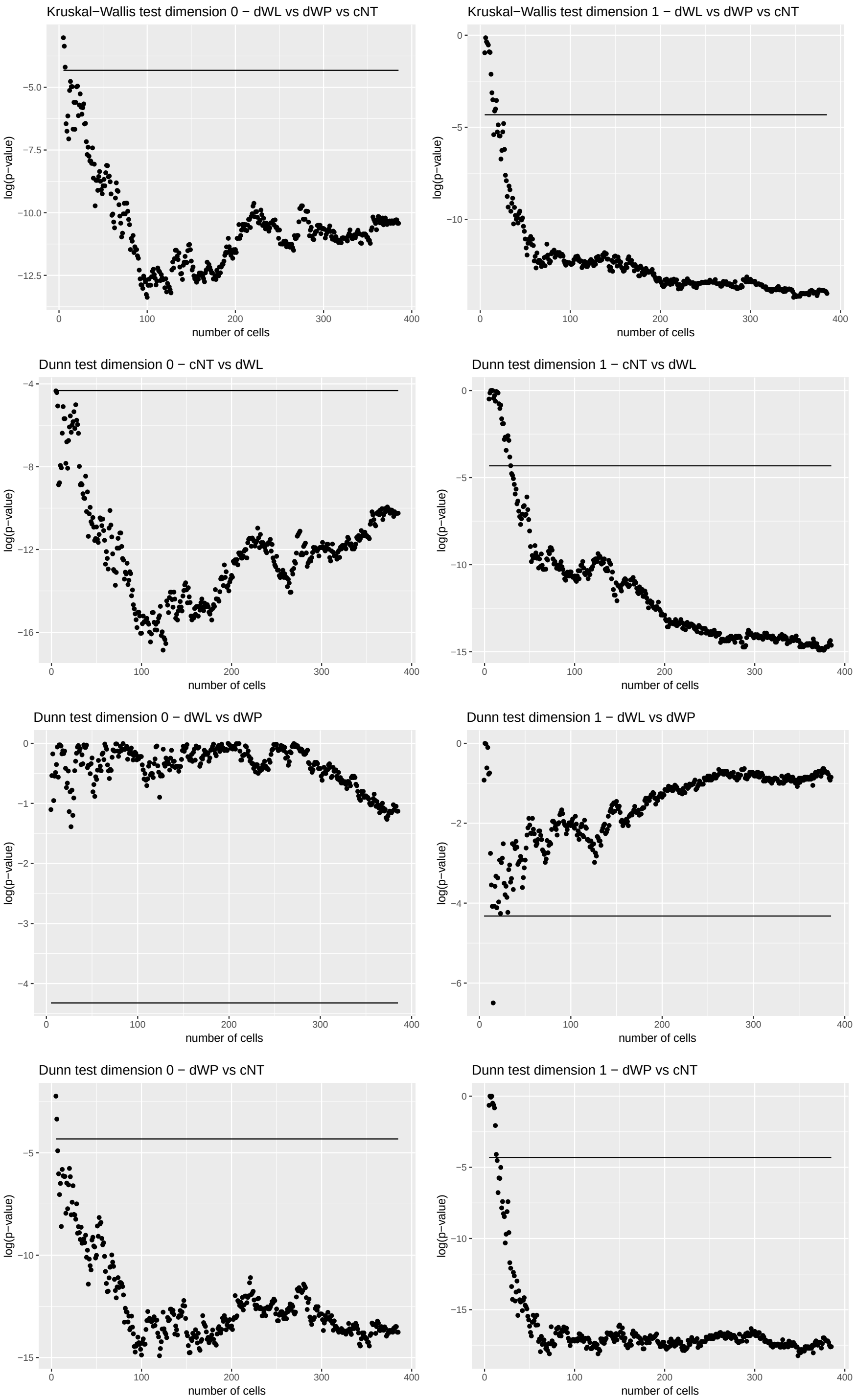


Figure 9: Statistical tests of cell tissues for different number of cells

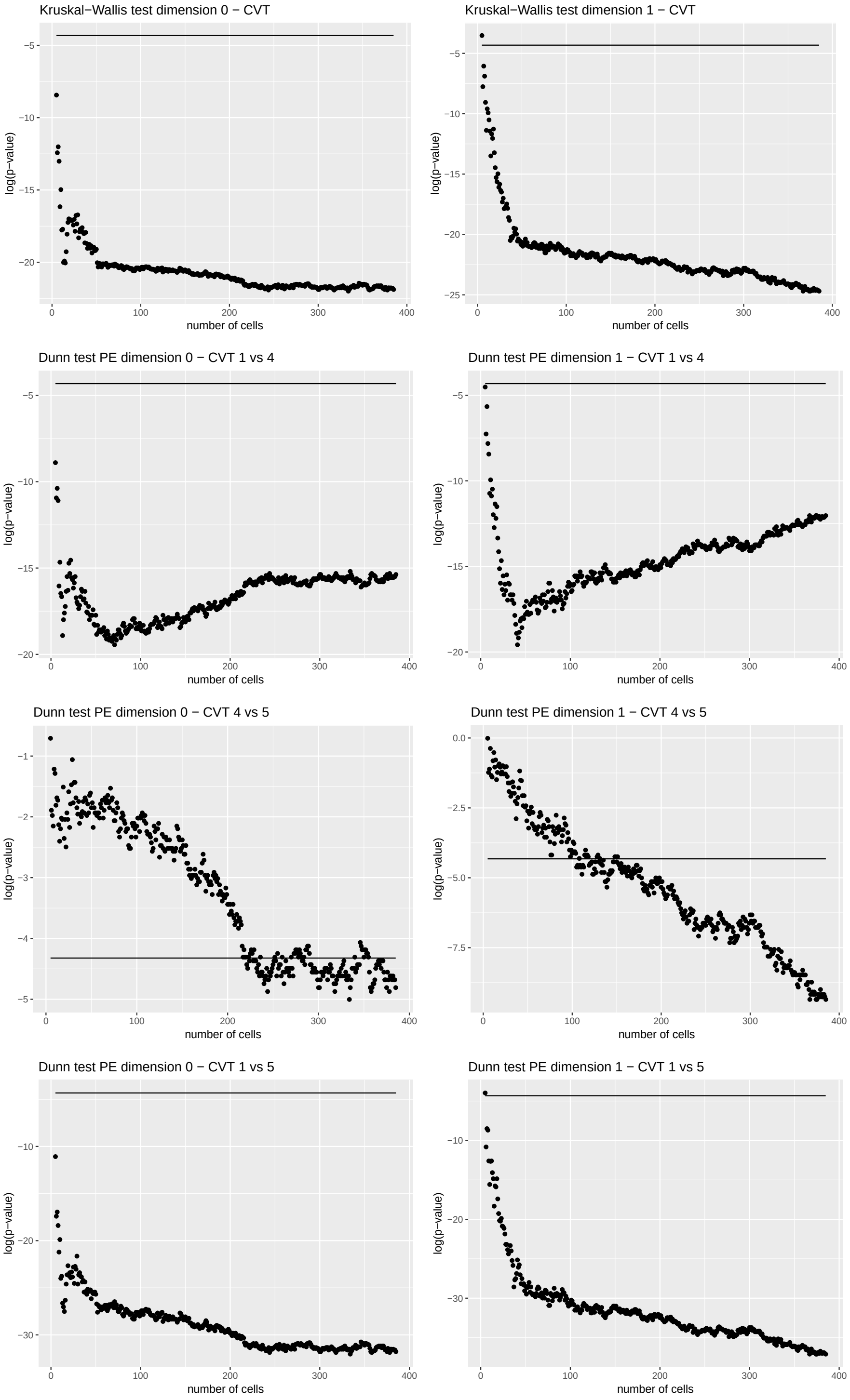


Figure 10: Statistical tests of CVT tesselation for different number of cells

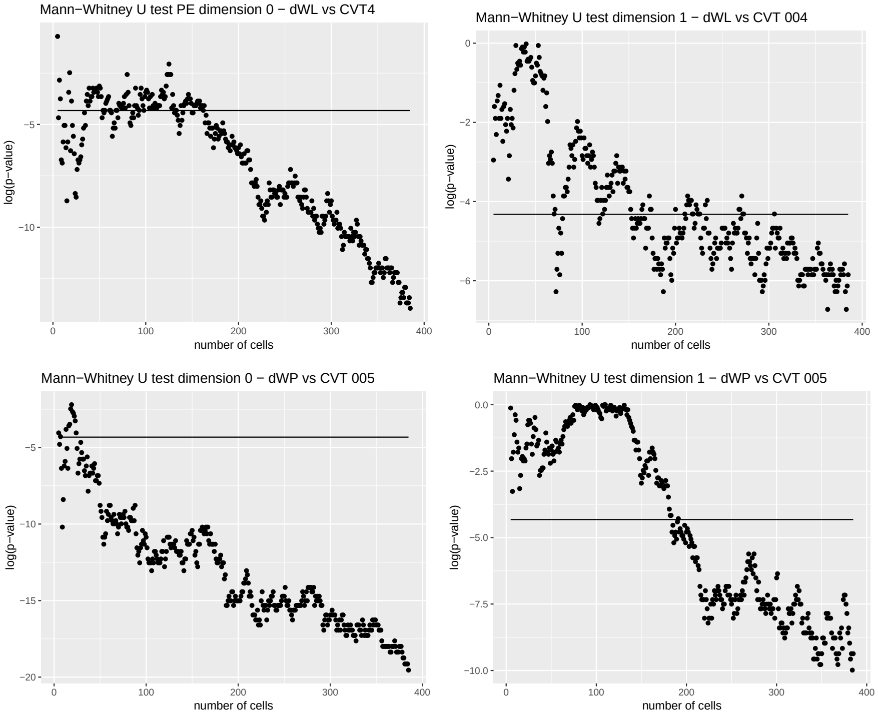


Figure 11: Top row: comparison of dWL and CVT4 for dimension 0 and 1. Bottom row: comparison of dWP and CVT5 for dimension 0 and 1.