A walk through a TDA pipeline. Case study of cells organization

Maria Jose Jimenez
Associate profesor
Departamento de Matematica Aplicada I
Universidad de Sevilla

majiro@us.es

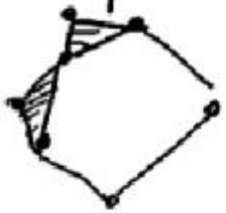
TOPOLOGICAL DATA ANALYSIS (TDA)

Study of - shape - of data

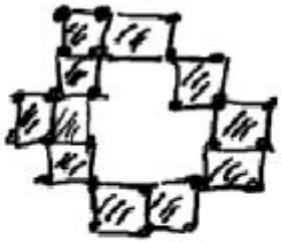
What type of data?

point clouds

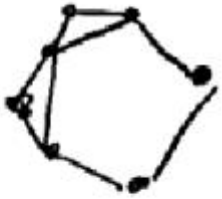
simplicial complexes



cubical complexes

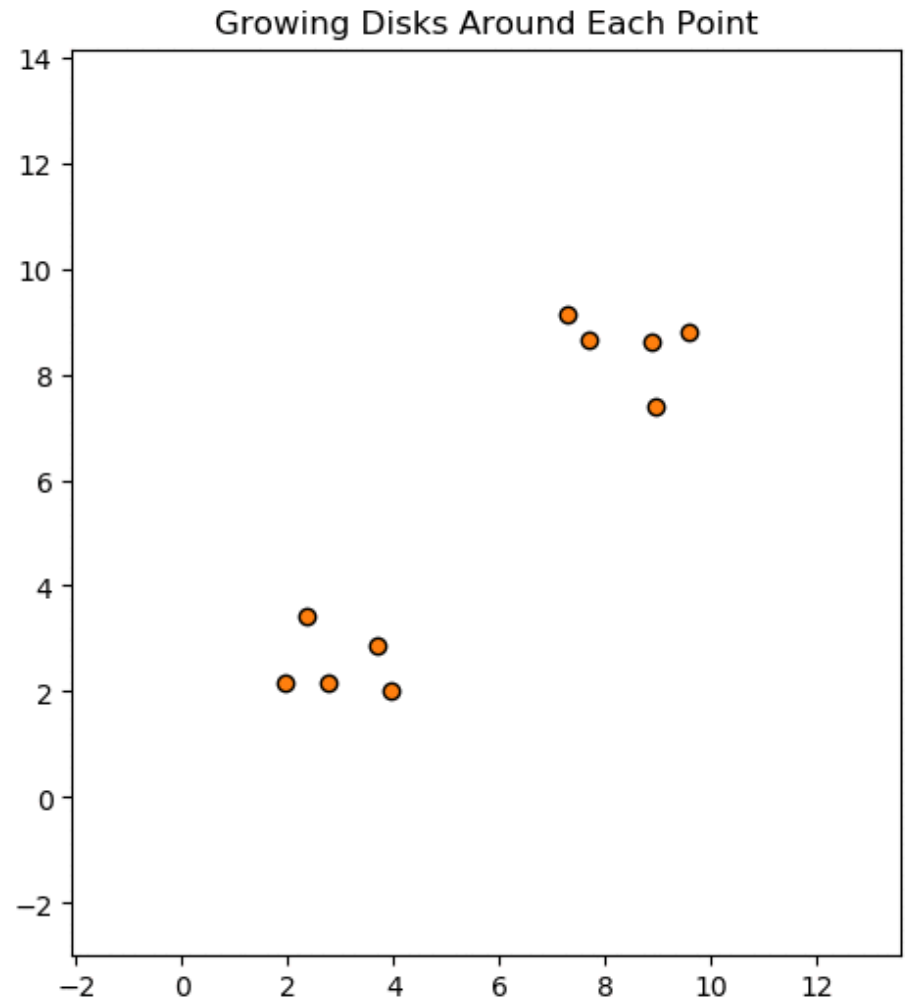


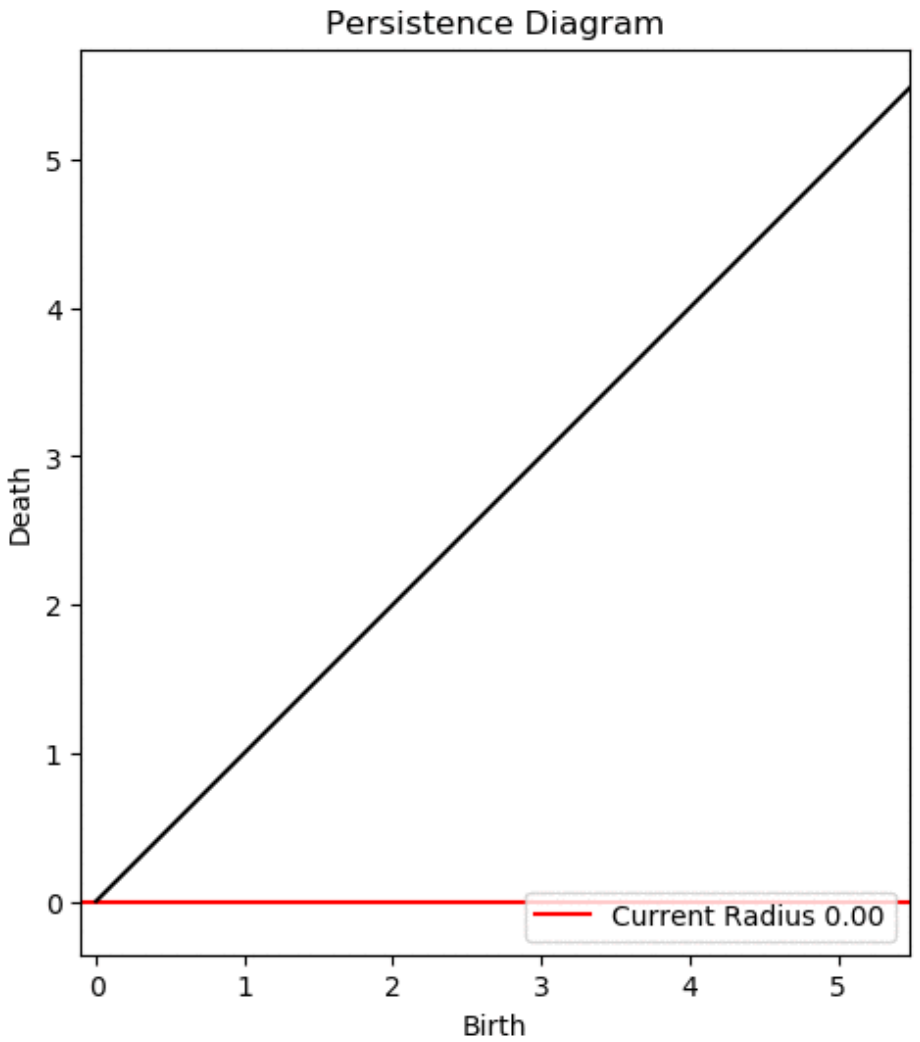
graphs



MOTIVATION

SHAPE OF DATA



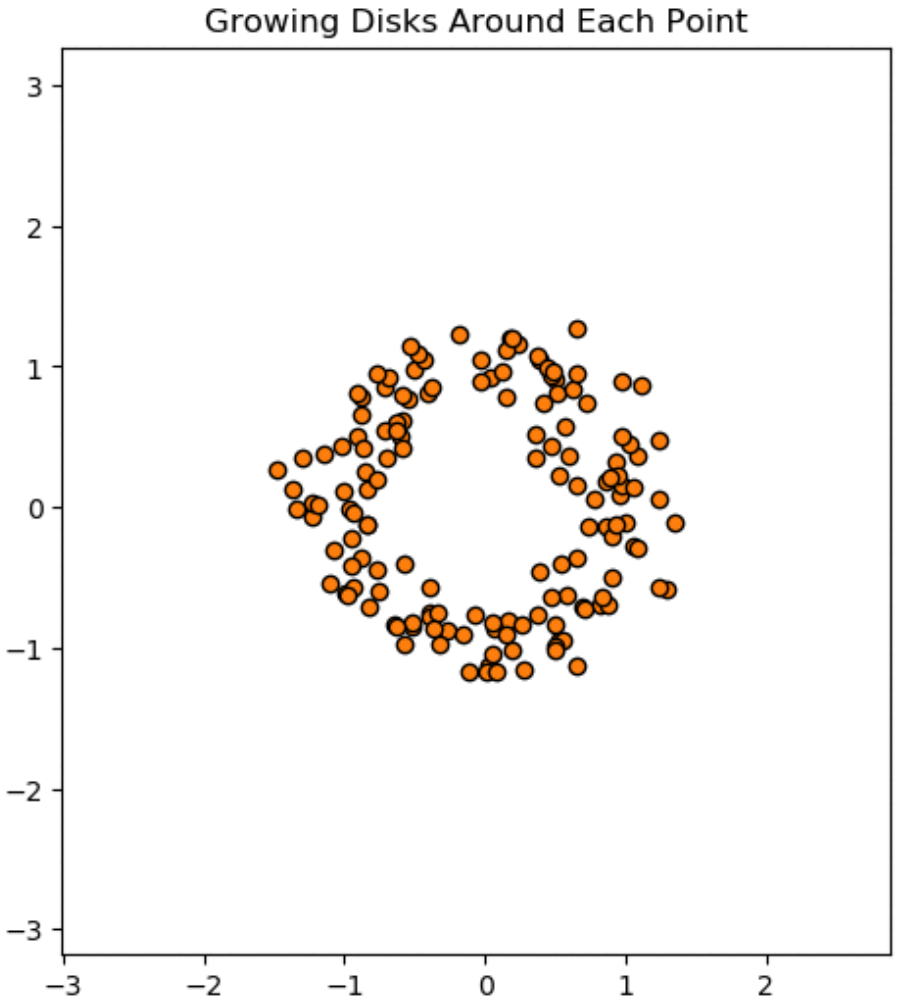


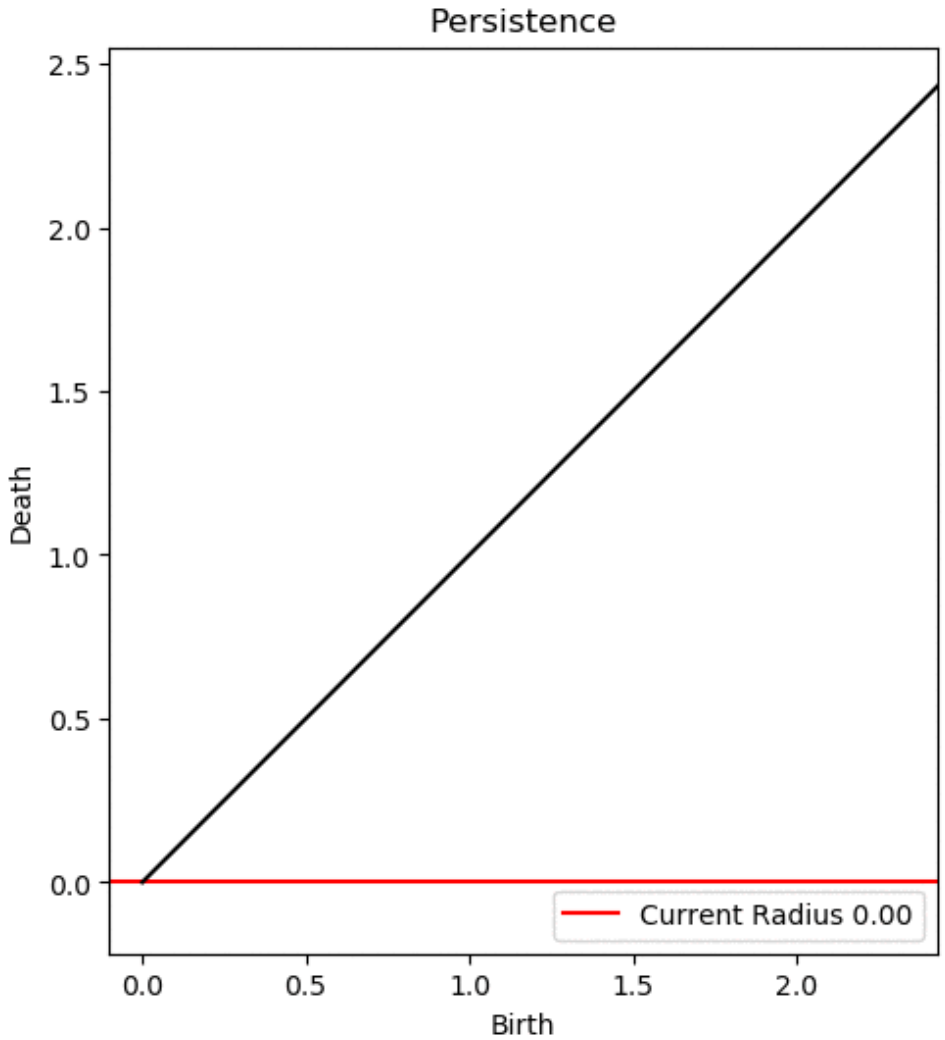
0-dimensional persistence diagram.

Credit: Animation by Gary Koplik
https://towardsdatascience.com/persistent-homology-with-examples-1974d4b9c3d0

MOTIVATION

SHAPE OF DATA





1-dimensional persistence diagram.

Credit: Animation by Gary Koplik
https://towardsdatascience.com/persistent-homology-with-examples-1974d4b9c3d0

MOTIVATION

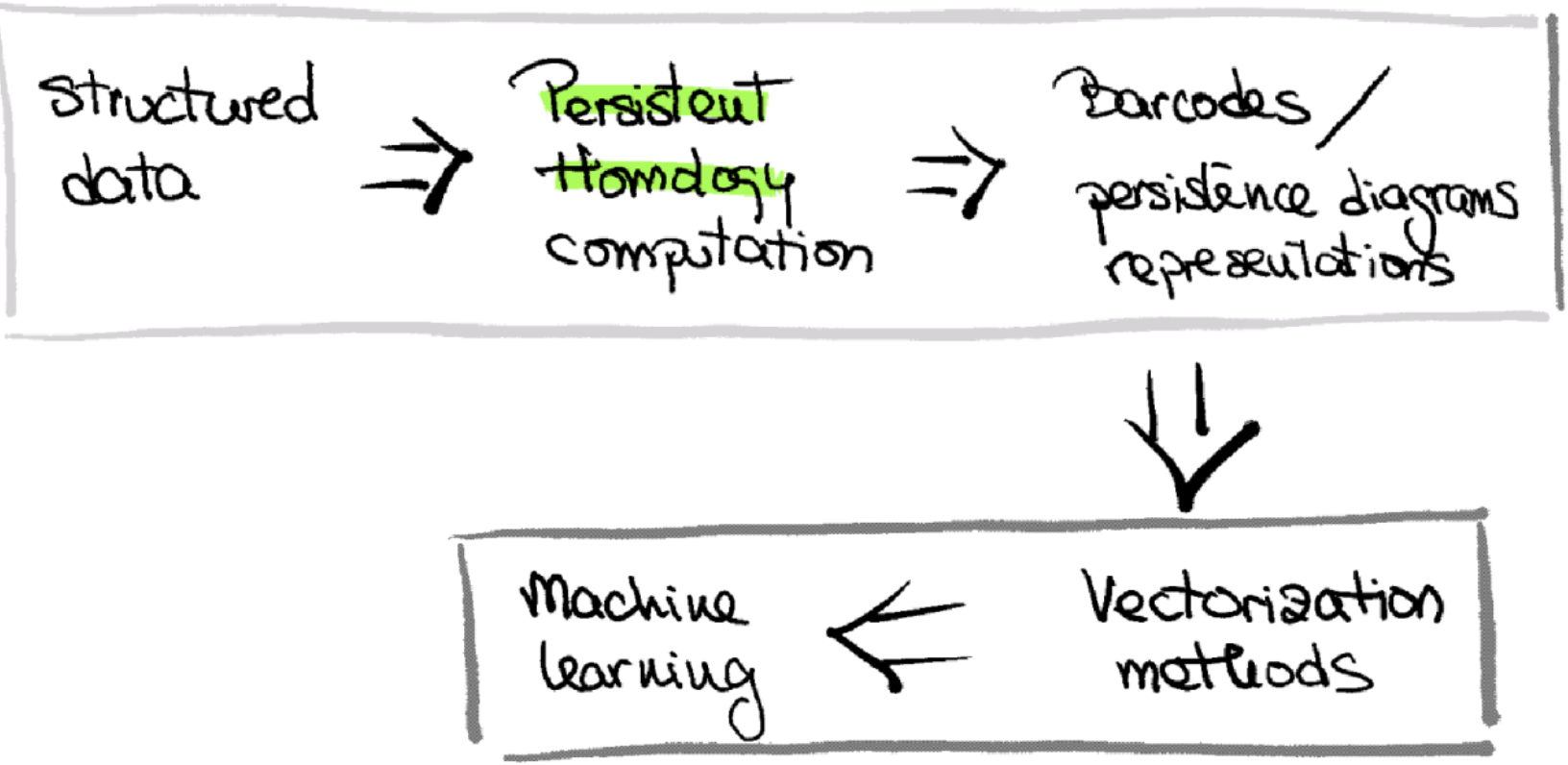
SHAPE OF DATA

APLICATIONS TO:

- Material science
- 3D shape analysis
- Time series analysis
- Diagnosis in medicine
- Genomics
- Chemistry
- Sensor networks
- transportation

...

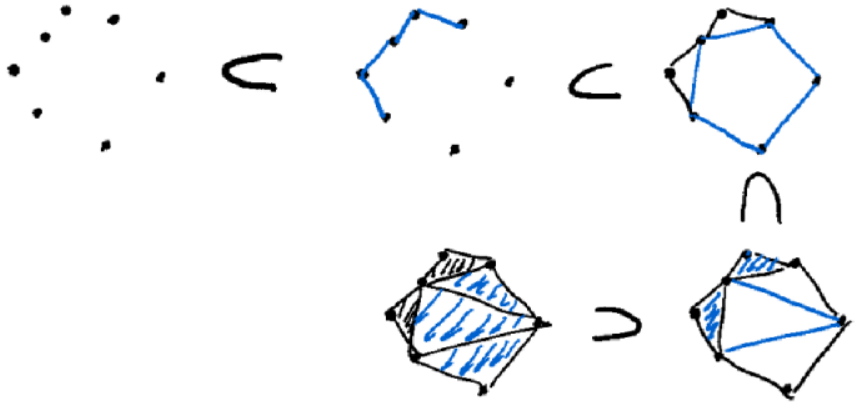
TDA PIPELINE



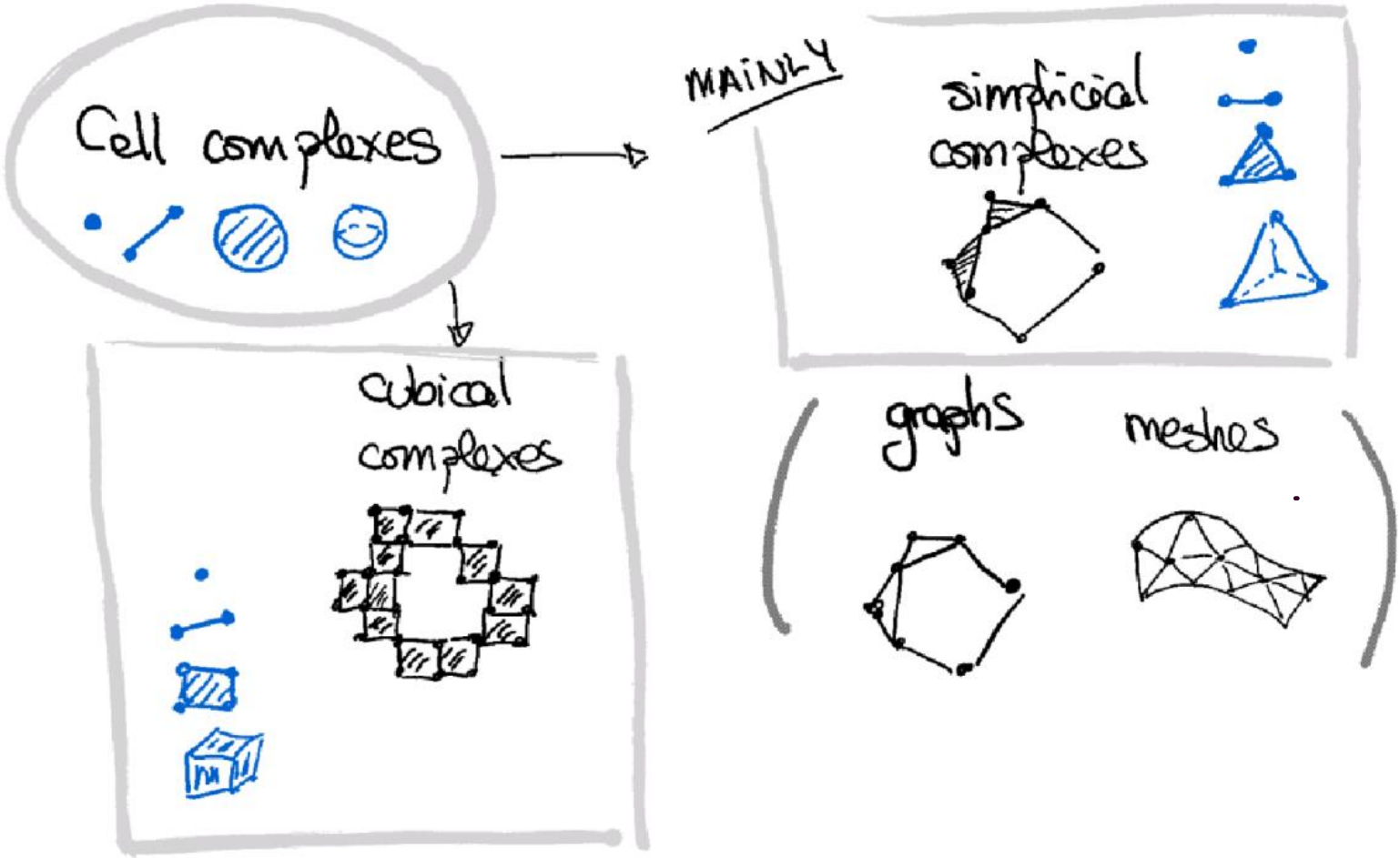
STRUCTURED DATA

Input data X

Increasing family of cell complexes $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$



STRUCTURED DATA



PERSISTENT HOMOLOGY

Homology

p-dimensional homology H_p : counts the number of p-dimensional holes

vector space

0-dim. homology H_0 : rank 7

1-dim. homology H_1 : rank 0

0-dim. homology H_0: rank 1

1-dim. homology H_1: rank 3

PERSISTENT HOMOLOGY

$$K_0 \subset K_1 \subset K_2 \subset K_3 \subset K_4$$

Filtration of simplicial complexes

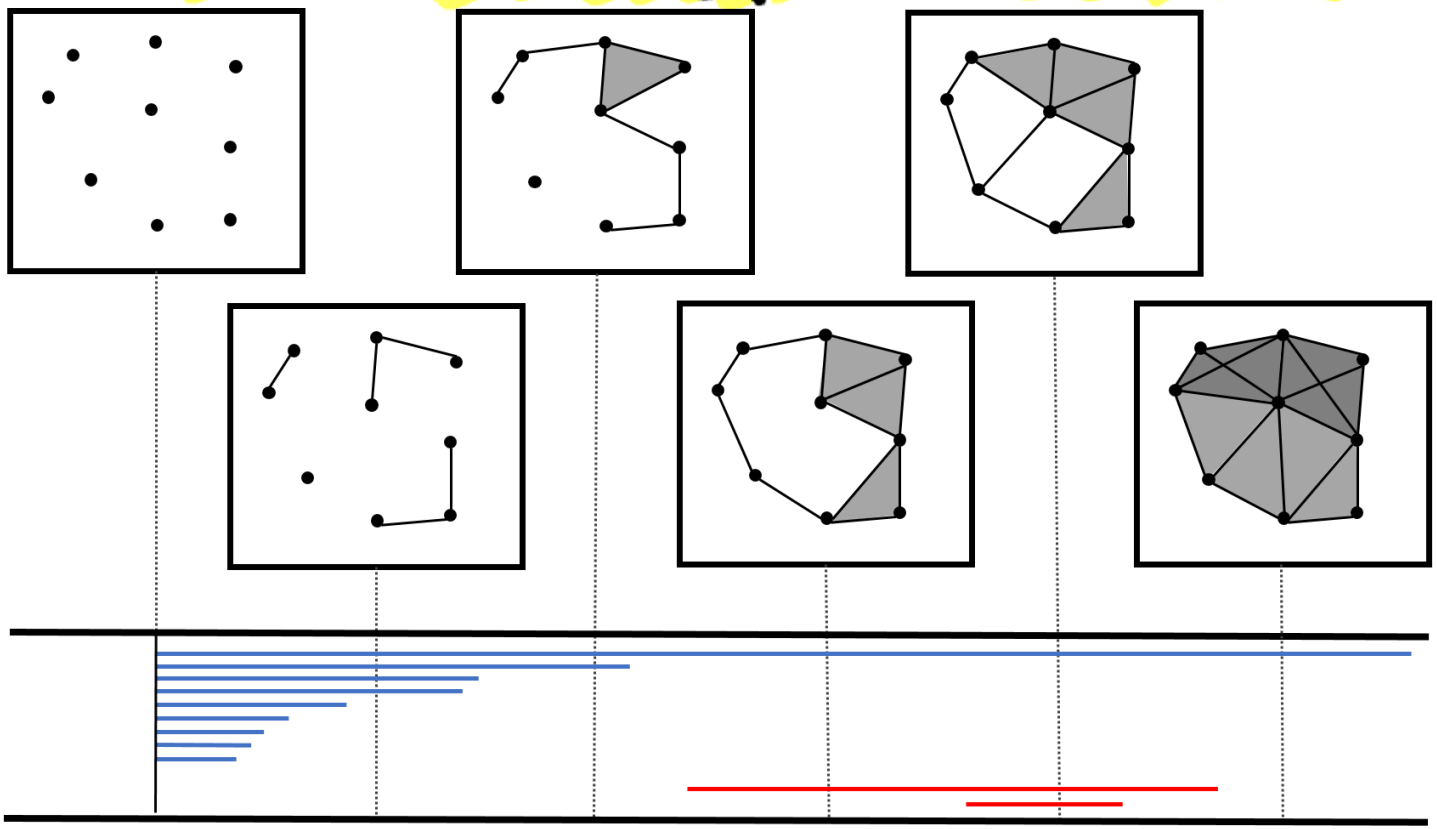
Homology $H_p(\cdot)$

$$H_p(K_0) \to H_p(K_1) \to H_p(K_2) \to H_p(K_3) \to H_p(K_4)$$

$p=0$: "Track of 0-holes along the sequence"

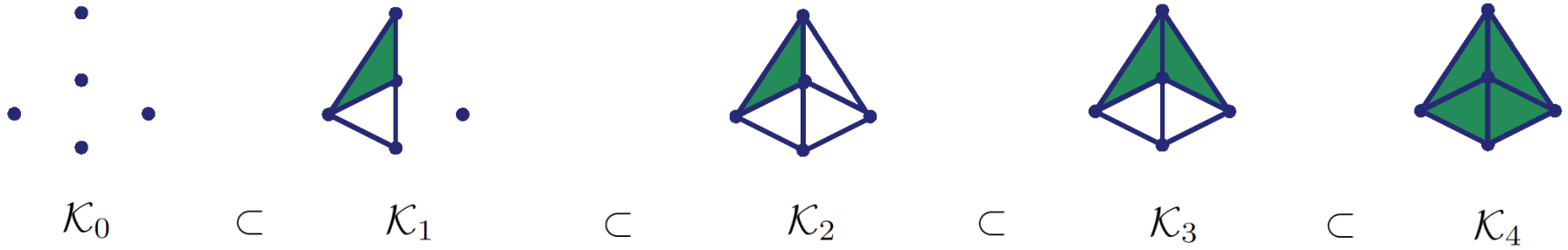
$p=1$: "Track of 1-holes along the sequence"

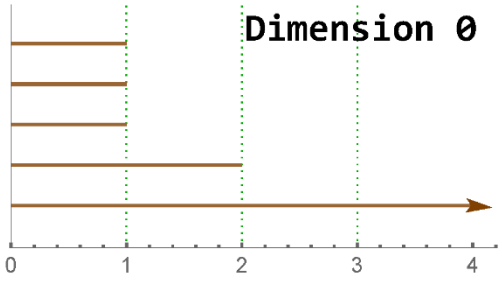
PERSISTENT HOMOLOGY REPRESENTATION

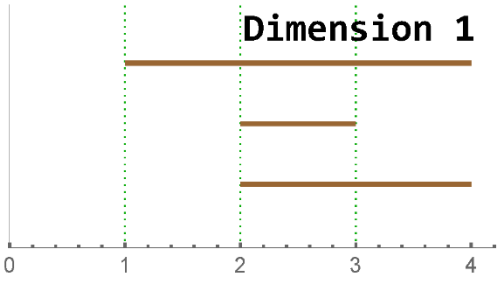


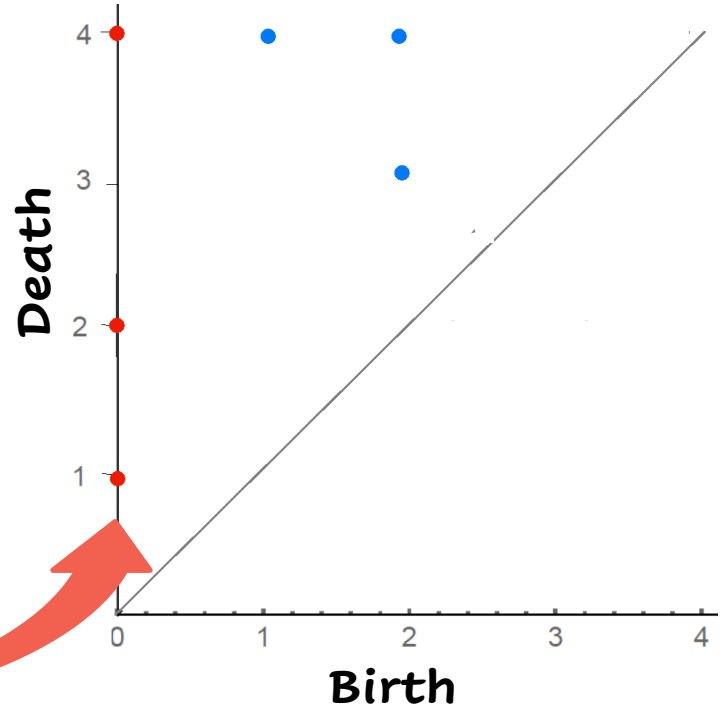
From: Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

Persistent Homology Representation

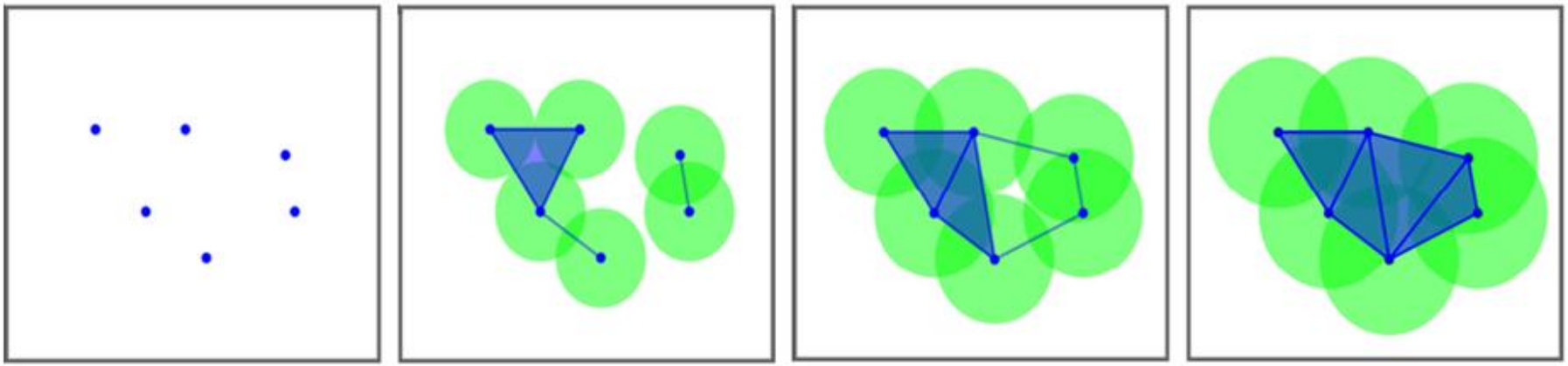






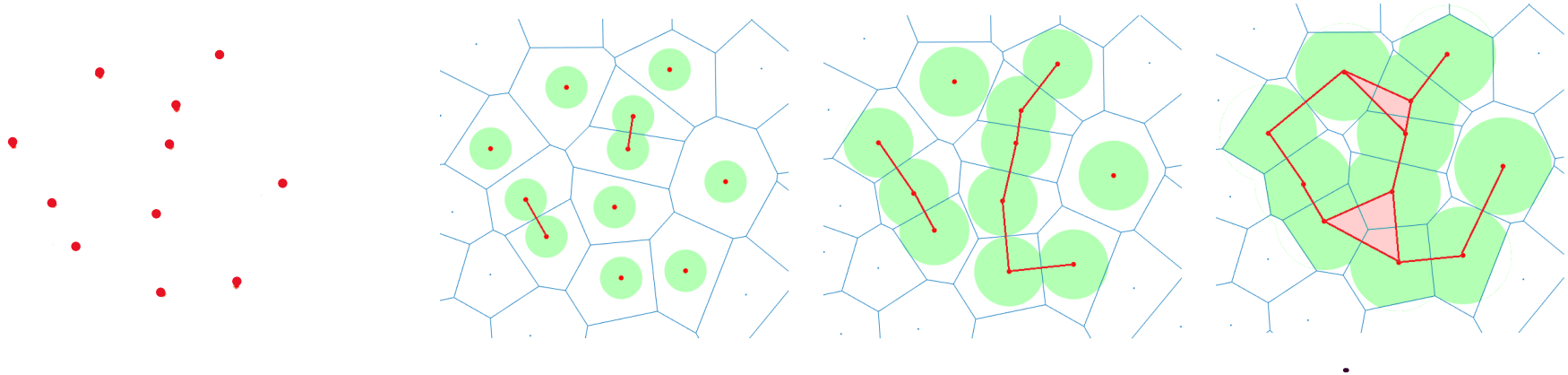


Vietoris Rips Filtration



$$K_0 \subset K_1 \subset K_2 \subset K_3$$

ALPHA COMPLEX Filtration



$$K_0 \subset K_1 \subset K_2 \subset K_3$$

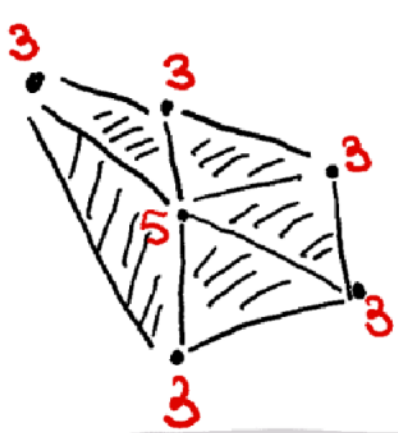
FILTER FUNCTION ON VERTICES

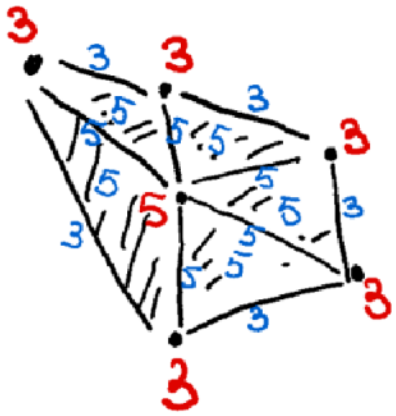
$f: K_0 \to \mathbb{R}_{\geq 0}$

$v \longmapsto f(v)$

Induces $f: K \to \mathbb{R}_{\geq 0}$

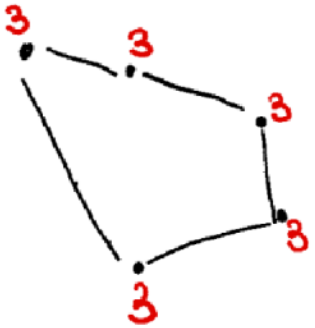
$\sigma \longmapsto \max\{f(v)\}_{v \in \sigma}$



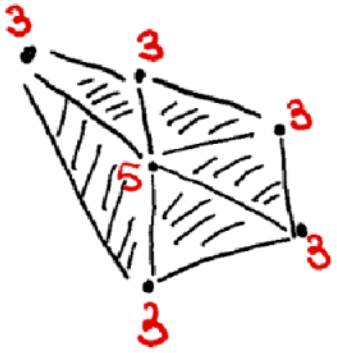


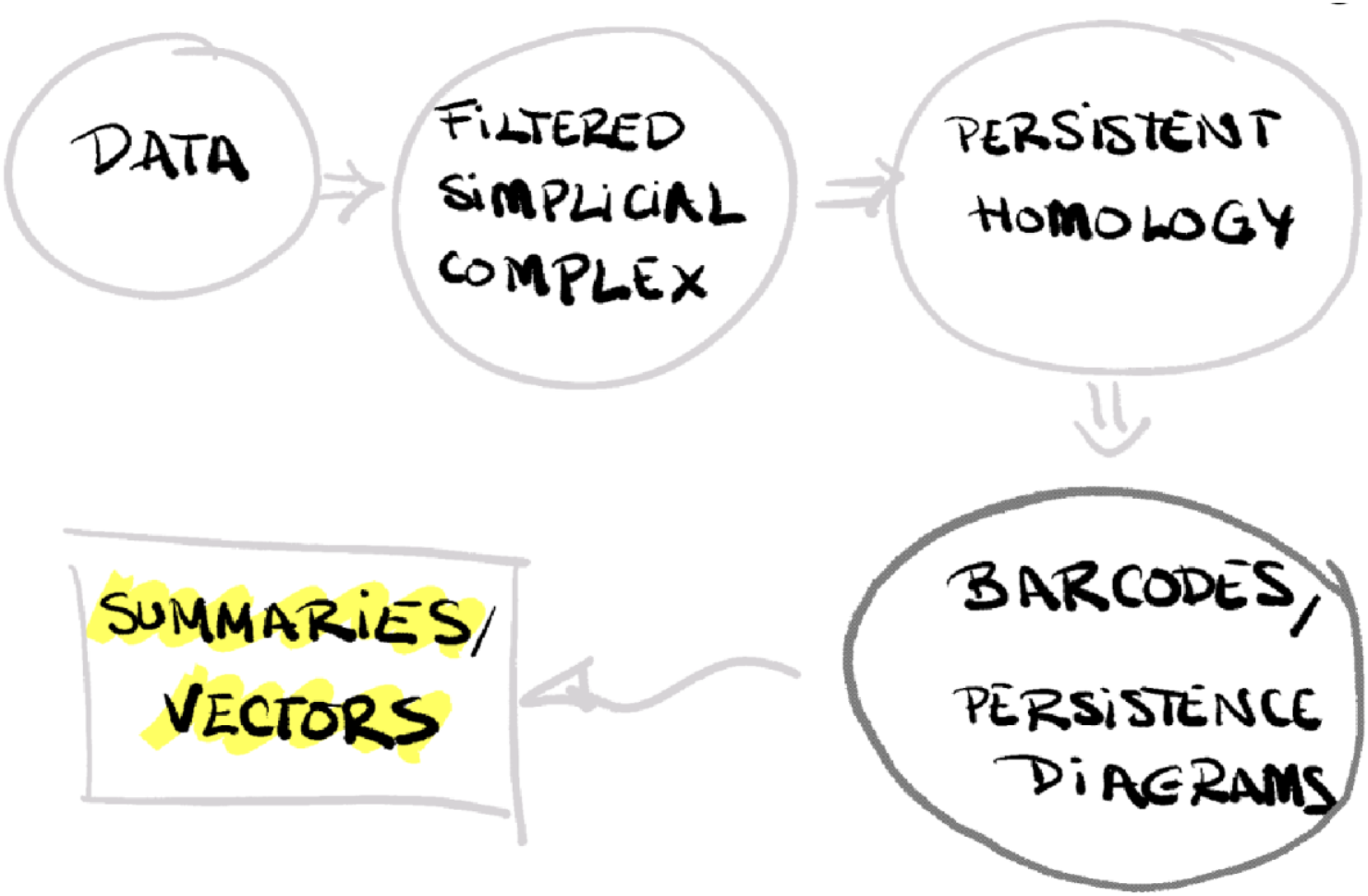
$\sigma < \tau$

$f(\sigma) \leq f(\tau)$



$\subset$




DATA
FILTERED SIMPLICIAL COMPLEX
PERSISTENT HOMOLOGY
BARCODES / PERSISTENCE DIAGRAMS
SUMMARIES / VECTORS

VECTORIZATION METHODS

Statistical vectorizations

$l_i = y_i - x_i$

$m_i = \frac{y_i + x_i}{2}$

$P.E. = -\sum \frac{l_i}{L} \log \frac{l_i}{L}$

$$\left[\text{mean}\begin{matrix}x_i\\y_i\\l_i\\m_i\end{matrix}, \text{std.}\begin{matrix}x_i\\y_i\\l_i\\m_i\end{matrix}, \text{median}\begin{matrix}x_i\\y_i\\l_i\\m_i\end{matrix}, \text{percentiles}\begin{matrix}x_i\\y_i\\l_i\\m_i\end{matrix}, \right.$$

range, IQR, Persistent Entropy, ...]

Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

Algebraic vectorizations

$[P_1(x,y), P_2(x,y) \ldots]$

- **Algebraic functions**
  = Adcock-Carlsson coordinates
  Ex. $\sum_i x_i(y_i - x_i)$

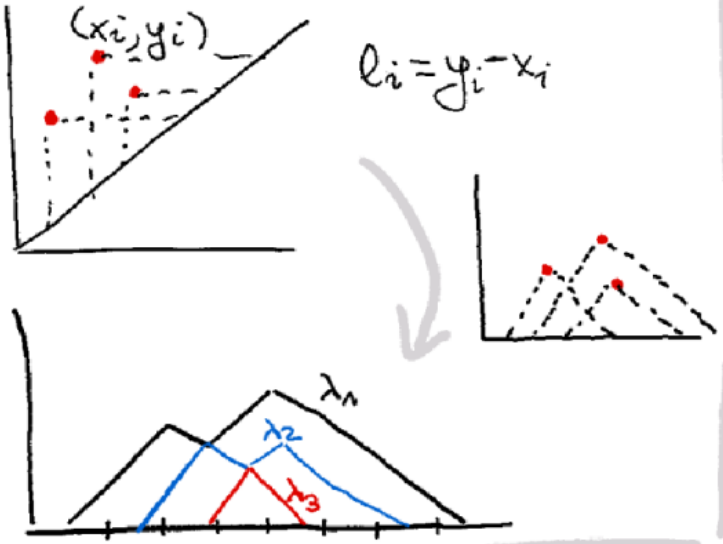
- **Tropical coordinates**
  $(min, max, +, -)$ on x_i, y_i
  Ex. $\max(y_i - x_i)$

- **Complex polynomials**
  (x_i, y_i) = roots of complex polynomial
  highest coefficients

Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

Curve vectorizations

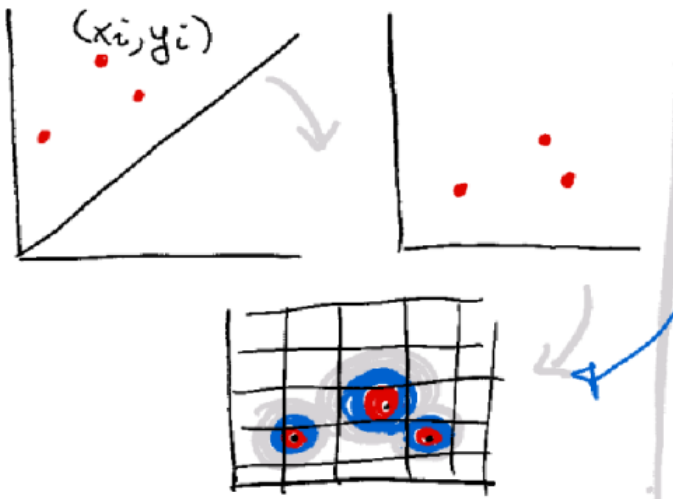


- **Betti curve**
  ~ number of "bars alive"
- **Lifespan curve**
  ~ Betti curve + weights
- **Persistence landscape**
  $\lambda_1, \lambda_2, \lambda_3 \ldots$
- **Persistence silhouette**
  ~ weighted sum of landscapes

Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

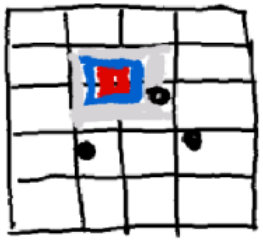
Functional vectorizations



- Persistence images

Weighted gaussian blurring + sampling

- Template functions



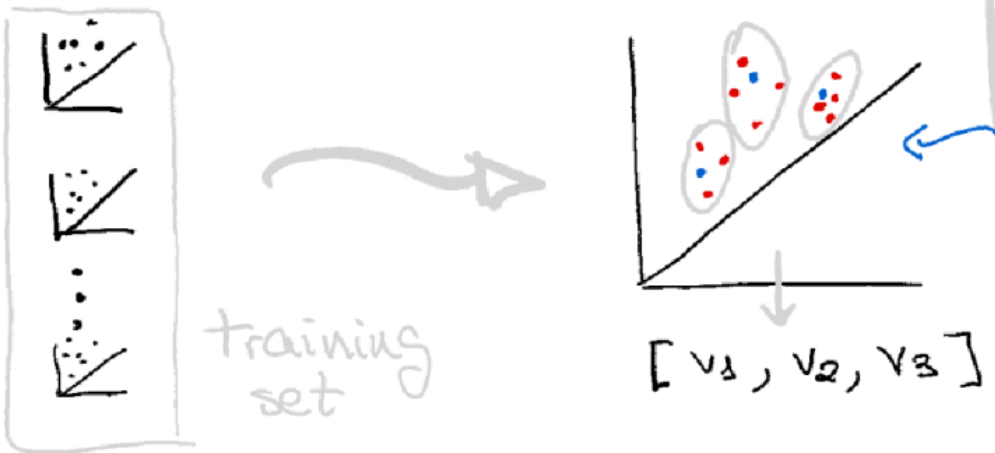
$(f_1, \dots f_n)$

$\sum_{(x_i,y_i)} f_j(x_i y_i)$

Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

Ensemble vectorizations



- Adaptive Template Systems

  Template functions on ellipses $\{E_j\}$

- ATOL

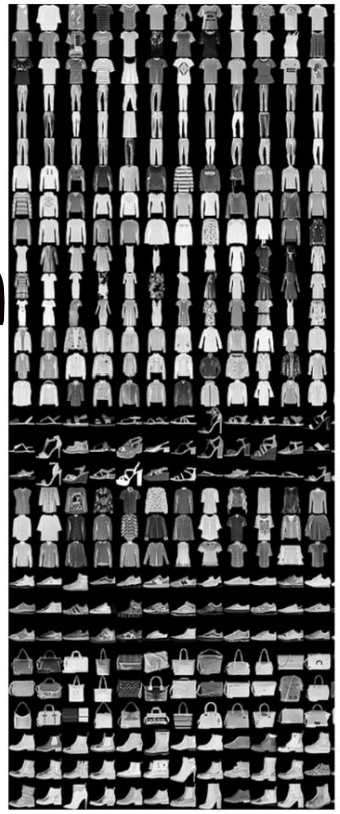
  b clusters
  ↳ centres $z_1, \dots z_b$
  contrast function against $z_1, \dots, z_b$

Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023)
A survey of vectorization methods in T.D.A
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

Boundary detection + Swepts

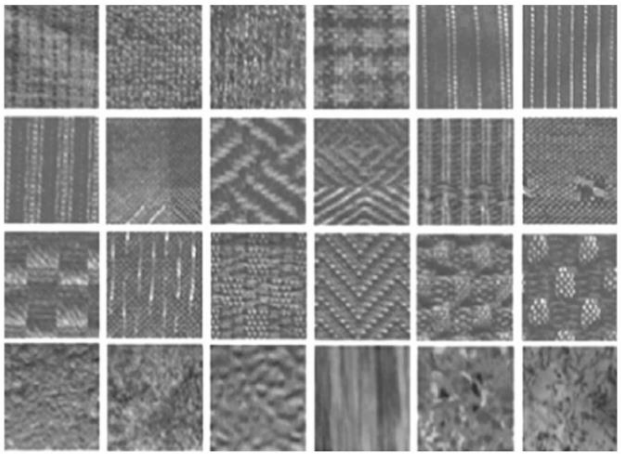
Fashion MNIST



SHREC14

HKS

Outex

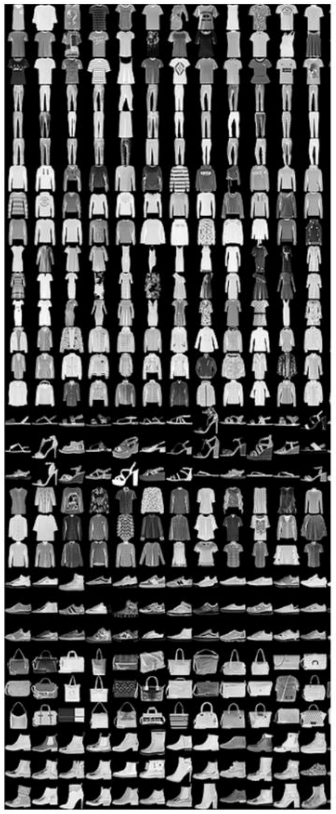


gray levels

From: Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023) A survey of vectorization methods in T.D.A IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

VECTORIZATION METHODS

Fashion MNIST

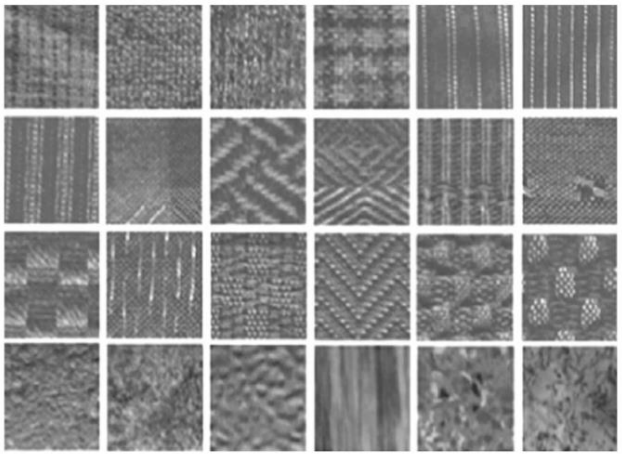


75%

SHREC14

95%

Outex



93%

From: Ali, D., Asaad, A., Jimenez M.J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M., (2023) A survey of vectorization methods in T.D.A IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

TDA for the organization of regions

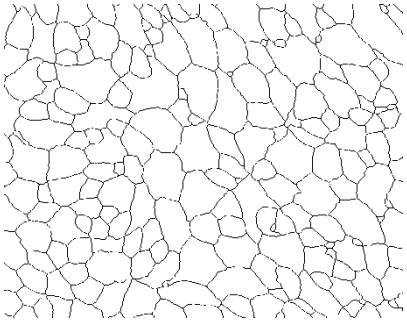
Can TDA be applied to characterize an image partitioned into regions?

* Neigbouring relations
* Relative sizes
* "Regularity" of shapes

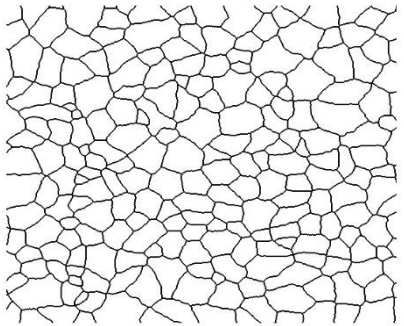
TDA for the organization of regions

Initial Motivation

Chicken

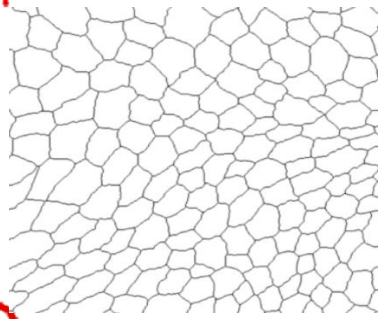


cEE

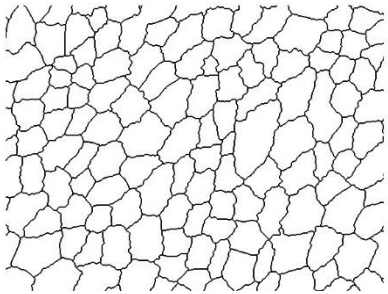


cNT

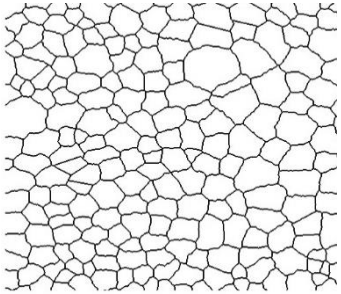
Drosophila



dNP



dWL



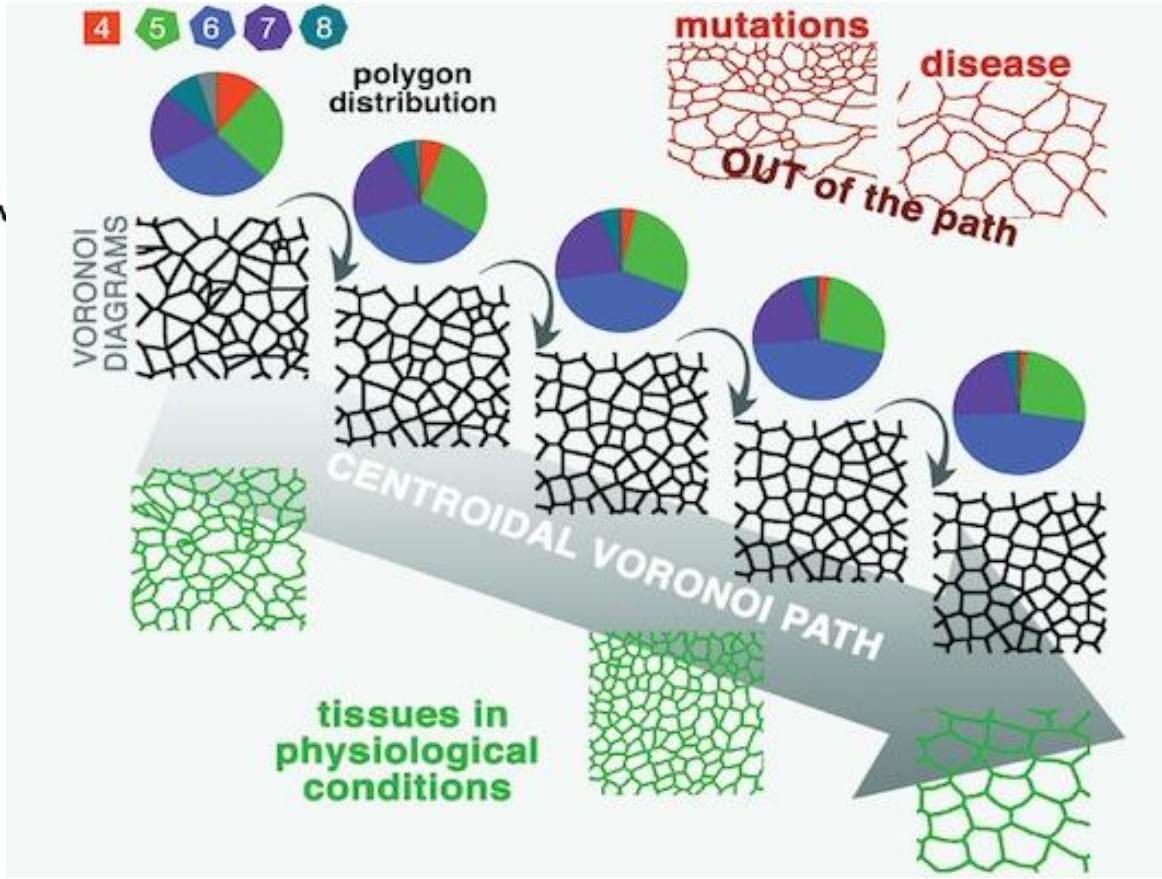
dWP

PREVIOUS WORKS

Luis M. Escudero's team

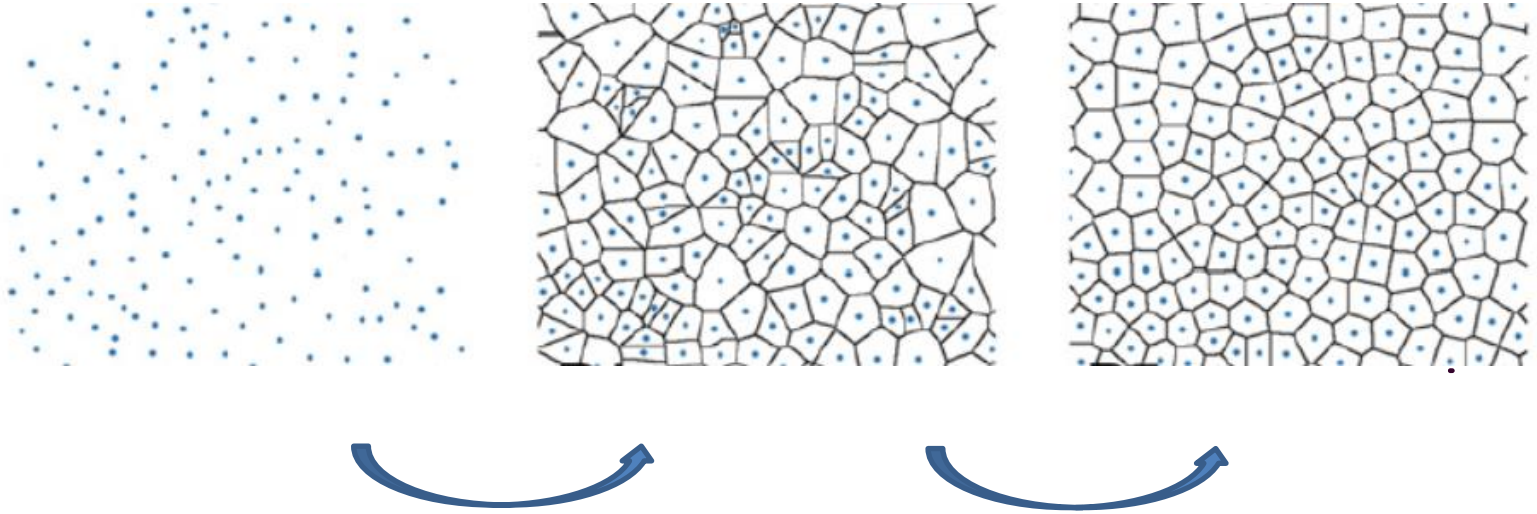
INSTITUTO DE BIOMEDICINA

UNIVERSIDAD DE SEVILLA



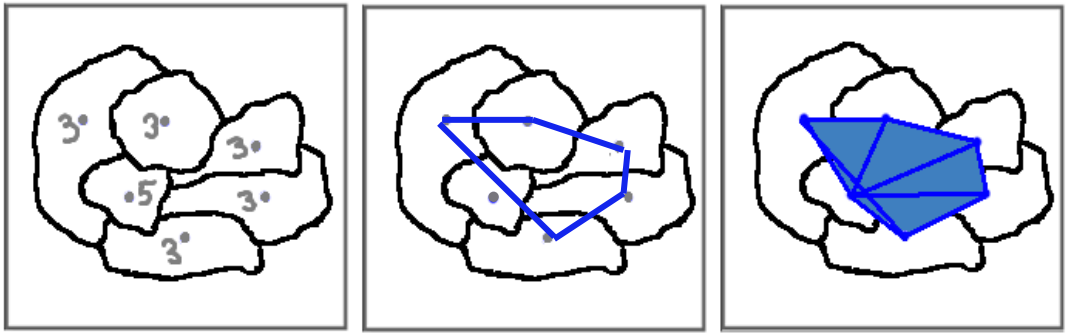
Sanchez-Gutierrez, D.; Tozluoglu, M.; et al., L.M.E. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO J.* **2016**, *35*, 77–88. doi:10.15252/embj.201592374.

PREVIOUS WORKS

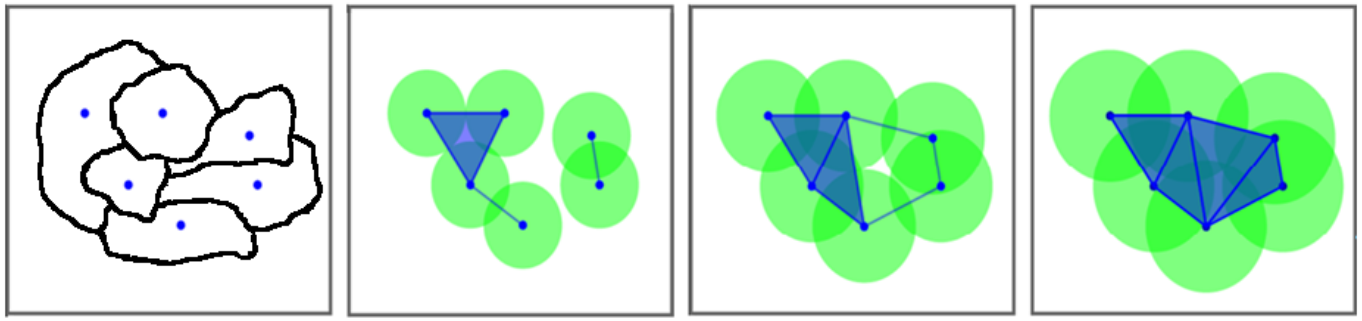


Sanchez-Gutierrez, D.; Tozluoglu, M.; et al., L.M.E. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO J.* **2016**, *35*, 77–88. doi:10.15252/embj.201592374.

OUR APPROACH



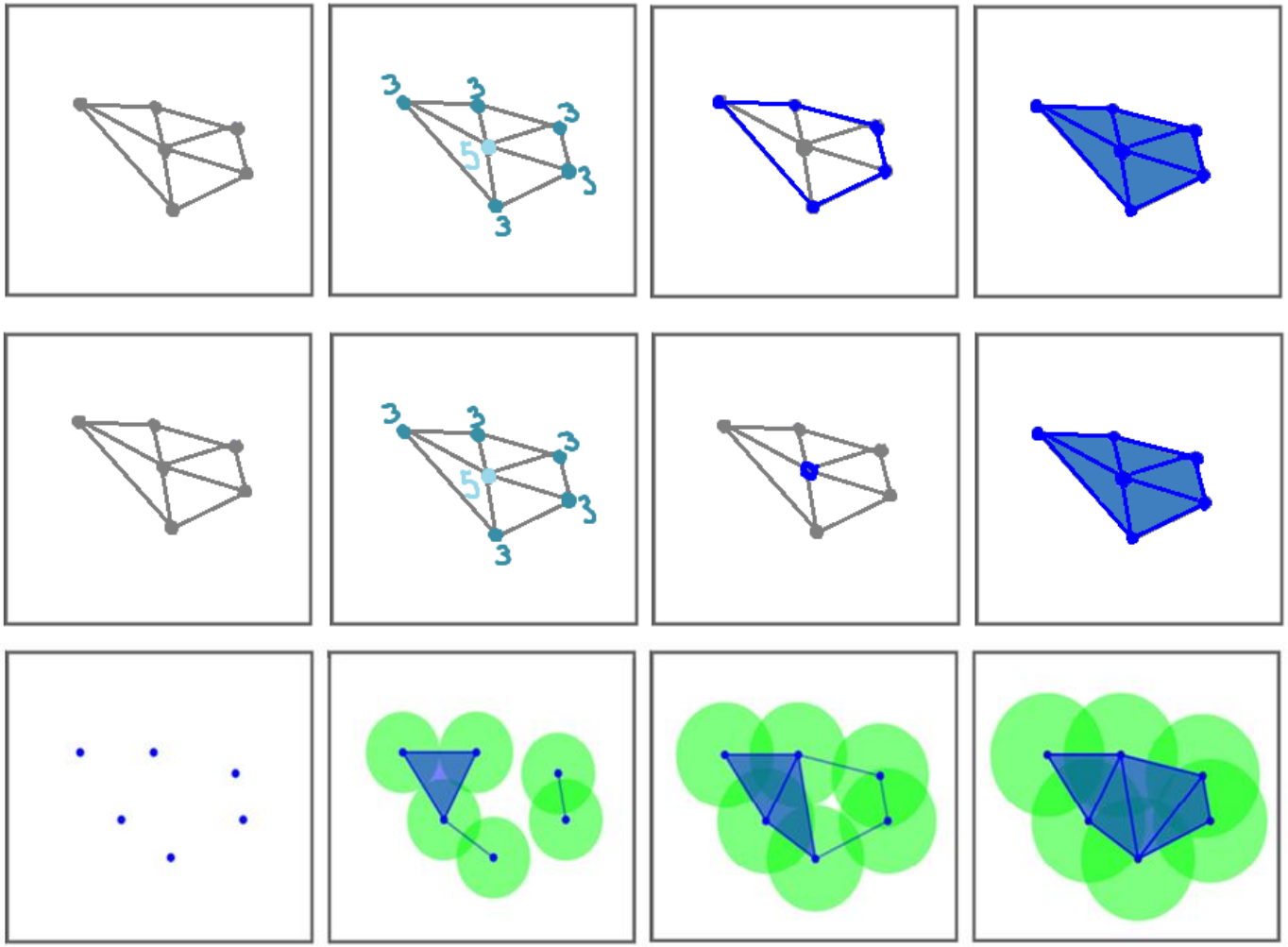
Contact Graph
+
N. of neighbours



VR filtration on centroids

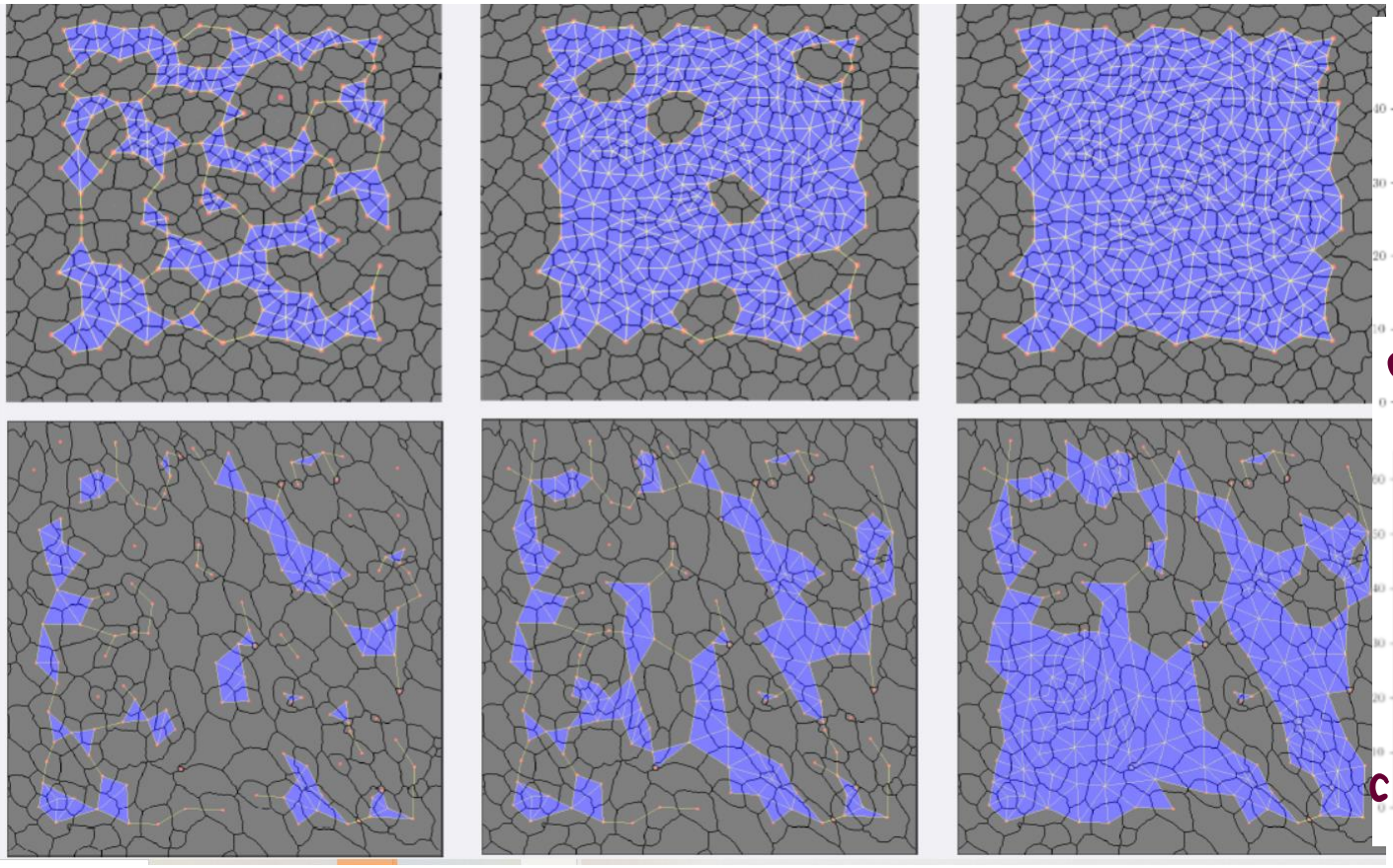
Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

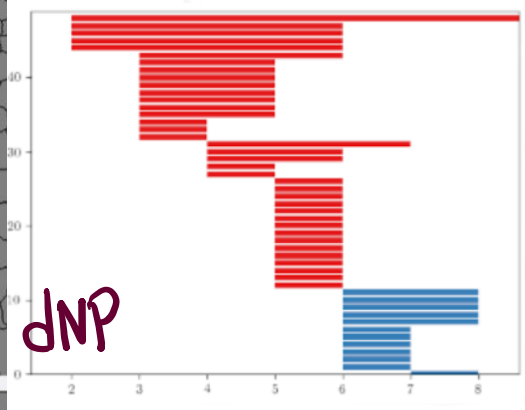
OUR APPROACH

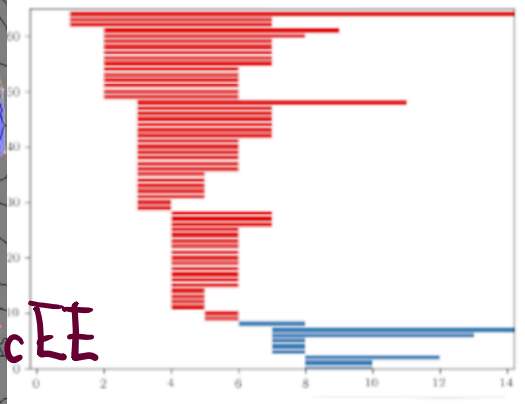


Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M. *Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue*. Mathematics 2021, 9, 1723.

OUR APPROACH



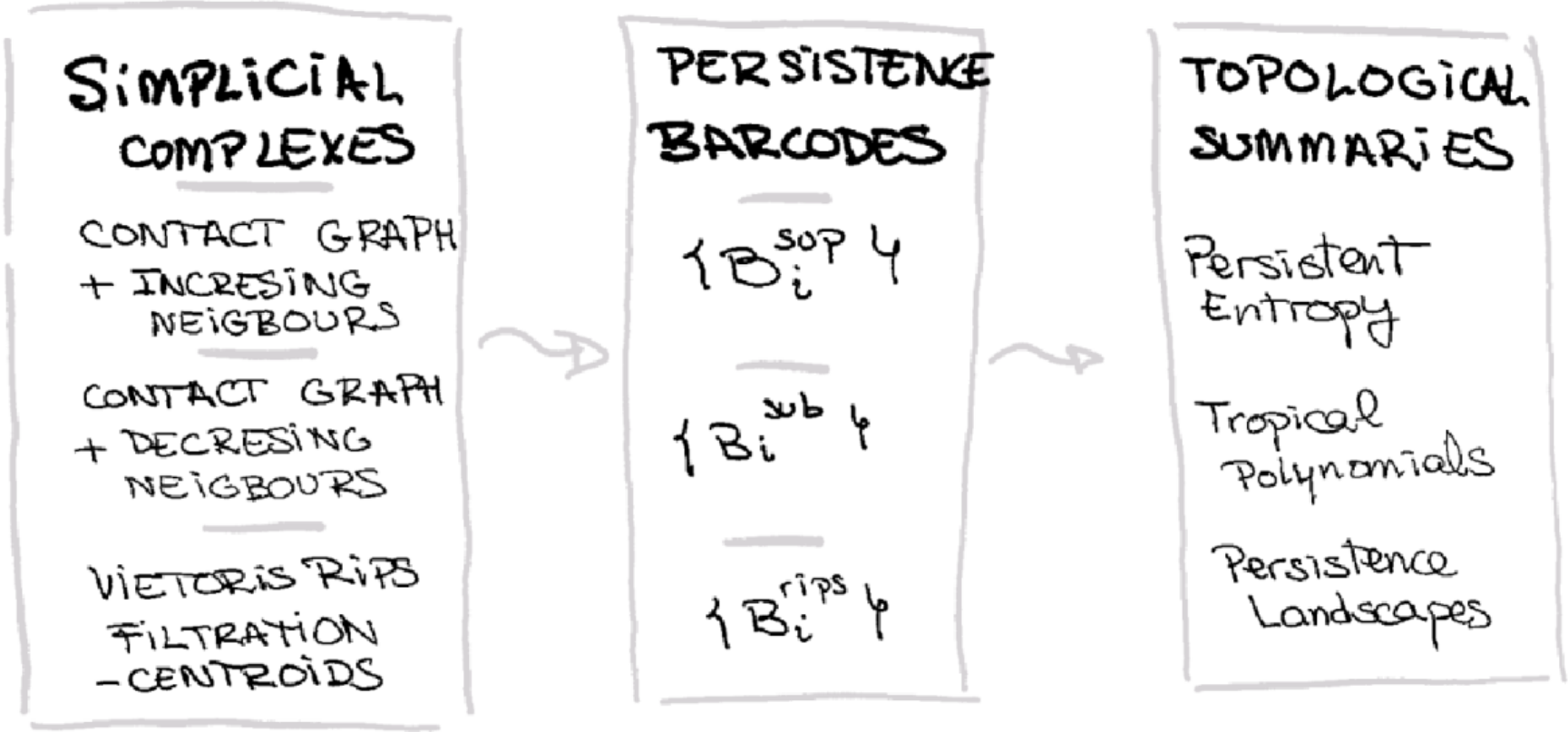




Increasing number of neighbors in the contact graph

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH



Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - PERSISTENT ENTROPY

The *persistent entropy* of a barcode $B = \{[b_i, d_i)\}_{i=1...n}$

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

where $\ell_i = d_i - b_i$ and $L(B) = \ell_1 + \ldots + \ell_n$.

Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M.J., Krim, H.: An entropy-based persistence barcod. Pattern Recognition **48** (2) 391–401 (2015)

. Rucco, M., Gonzalez-Diaz, R., Jimenez, M.J., Atienza, N., Cristalli, C., Concettoni, E., Ferrante, A. and Merelli, E.: A new topological entropy-based approach for measuring similarities among piecewise linear functions. Signal Processing, **134**, 130-–138 (2017)

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - TROPICAL POLYNOMIALS

Max-plus semiring $(\mathbb{R} \cup \{-\infty\}, \boxplus, \odot)$, addition and multiplication being defined as:

$$a \boxplus b := \max(a, b) \qquad a \odot b := a + b.$$

Max-plus polynomials can be defined on the Barcodes space on the lenghts of the bars:

$$\max(a_1 + a_1^1 \ell_1 + \ldots + a_q^1 \ell_q, \ldots, a_r + a_1^r \ell_1 + \ldots + a_q^r \ell_q).$$

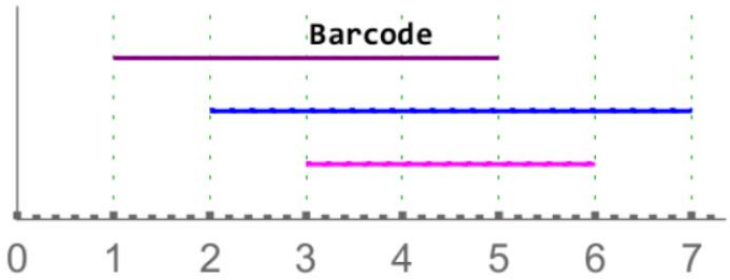
ℓ_i = i-th maximum length of intervals in the barcode

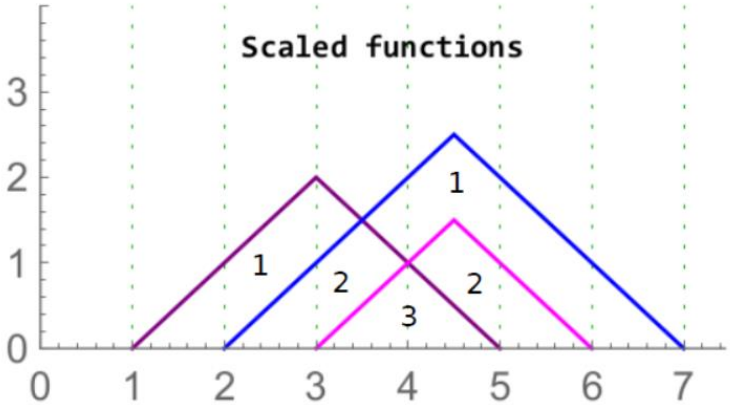
Kališnik, S. Tropical Coordinates on the Space of Persistence Barcodes. *Foundations of Computational Mathematics* **2018**, pp. 101–129. doi:10.1007/s10208-018-9379-y.

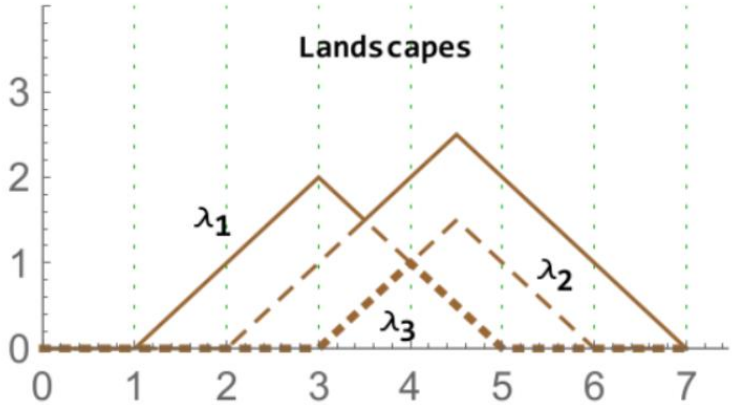
Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - PERSISTENCE LANDSCAPES







$\|\lambda_i\|_1$

Bubenik, P. Statistical Topological Data Analysis Using Persistence Landscapes. *J. Mach. Learn. Res.* **2015**, *16*, 77–102.
Bubenik, P. The Persistence Landscape and Some of Its Properties. *Topological Data Analysis. Abel Symposia* **2020**, *15*, 97–117. doi:10.1007/978-3-030-43408-3_4.

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - STATISTICAL ANALYSIS

TOPOLOGICAL SUMMARIES

- Persistent Entropy
- Tropical Polynomials
- Persistence Landscapes

× 3 FILTRATIONS = 57 summaries / image

Kruskall-Wallis test

+

Dunn test for pairs

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - STATISTICAL ANALYSIS

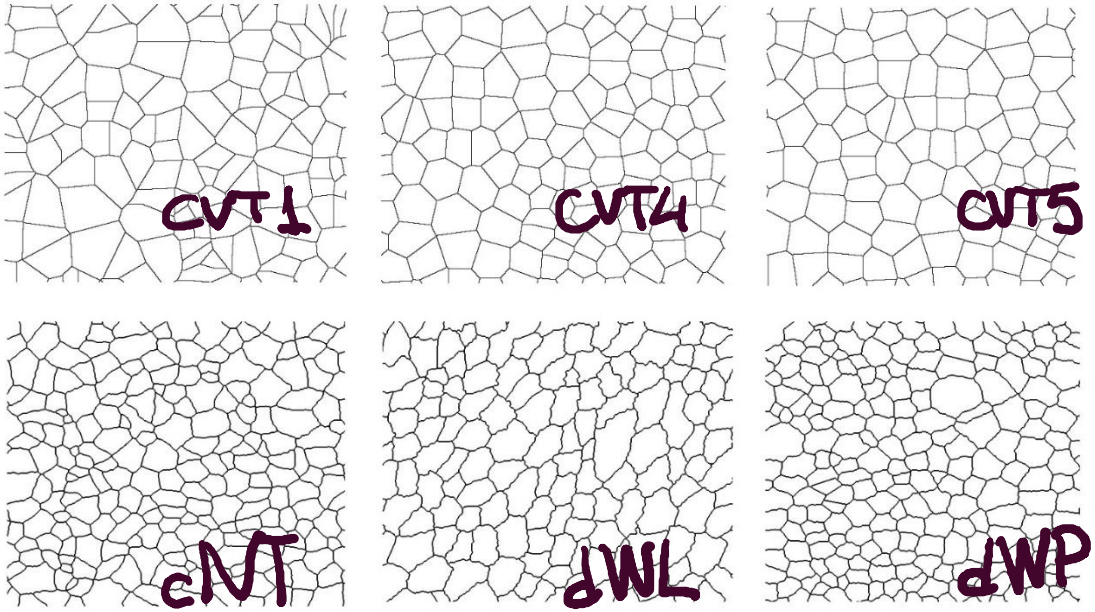
187 cells	$\lvert\lambda_1^{sub}(9)\rvert$	$\lvert\lambda_0^{sup}(5)\rvert$	$\ell_1^{sup}(2)+\ell_1^{sup}(3)$	$\ell_0^{rips}(9)$	PE_0^{rips}
cEE vs cNT	✓	×	×	×	×
cEE vs dNP	✓	×	✓	×	✓
cNT vs dNP	✓	×	✓	×	×
cEE vs dWL	✓	×	✓	✓	✓
cNT vs dWL	×	×	✓	✓	✓
dNP vs dWL	×	×	×	✓	×
cEE vs dWP	✓	✓	✓	✓	✓
cNT vs dWP	×	✓	✓	✓	✓
dNP vs dWP	×	×	×	✓	✓
dWL vs dWP	×	✓	×	×	×

✓ = p-value smaller than 0.01 in Dunn Test

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M.
Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue. Mathematics 2021, 9, 1723.

OUR APPROACH

TOPOLOGICAL SUMMARIES - STATISTICAL ANALYSIS



257 cells	$\lvert\lambda_0^{\text{sub}}(0.15N)\rvert$	PE_0^{rips}	$Poly_0^{\text{rips}}(1, 0.05N)$	$\ell_0^{\text{rips}}(0.10N)$
cNT vs CVT_1	✓	✓	✓	✓
dWL vs CVT_4	×	✓	✓	×
dWP vs CVT_5	×	✓	×	✓

✓ = p-value smaller than 0.01 in Mann-Whitney U test

Atienza, N.; Jimenez, M.-J.; Soriano-Trigueros, M. *Stable Topological Summaries for Analyzing the Organization of Cells in a Packed Tissue*. Mathematics 2021, 9, 1723.

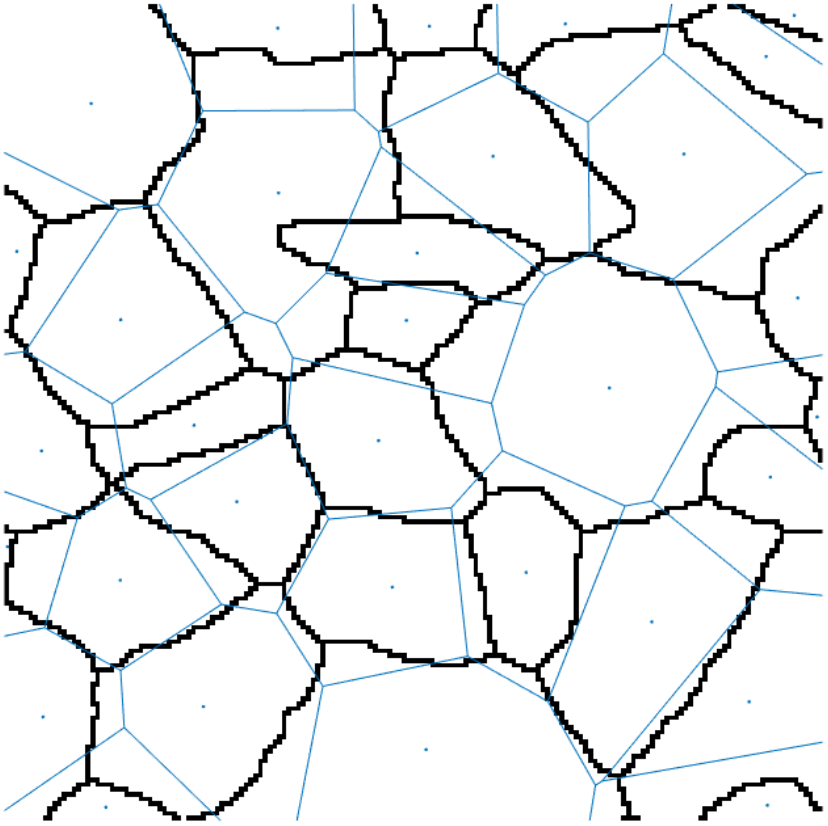
CAN WE IMPROVE THE MODEL?

Can we build a filtration on a simplicial complex that gathers information of

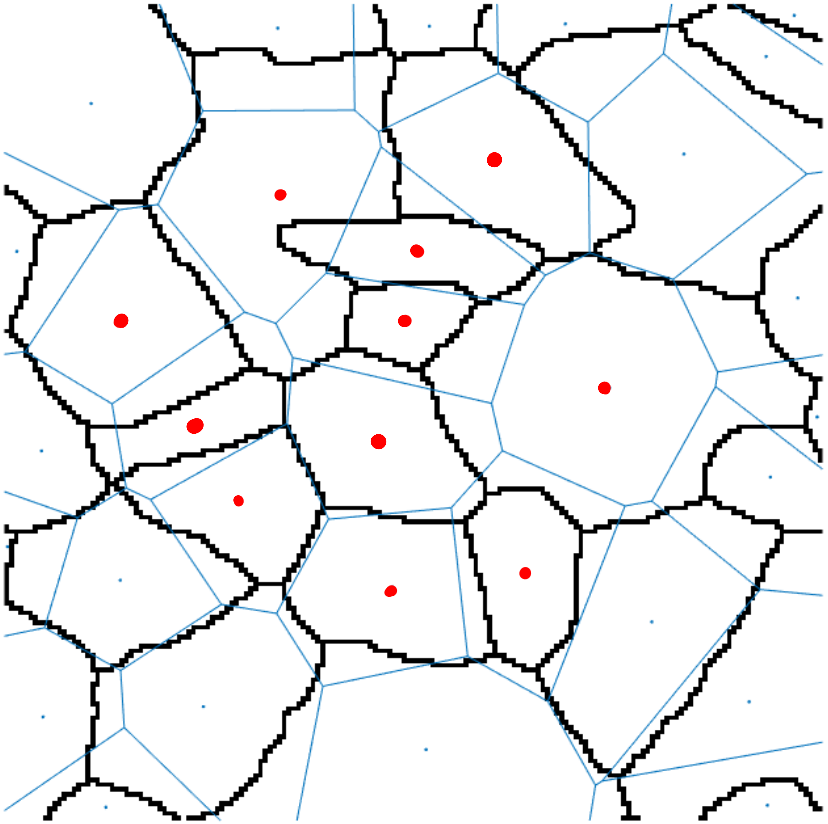
- neighbouring relations
- spatial distribution of centroids
- More on the actual shape of the regions

CAN WE IMPROVE THE MODEL?

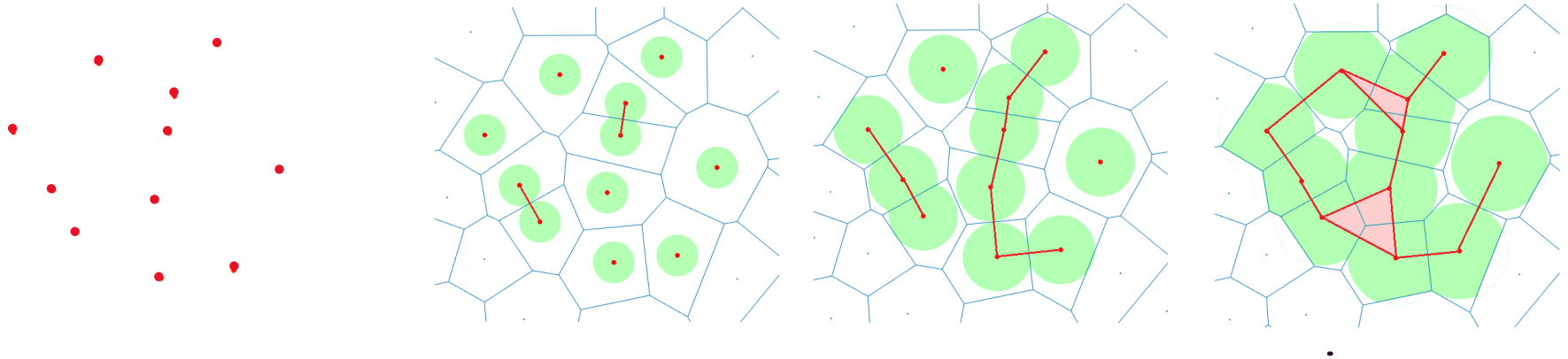
In the biological setting, it is natural to approximate regions (cells) by their Voronoi region.



CAN WE IMPROVE THE MODEL?



ALPHA COMPLEX Filtration

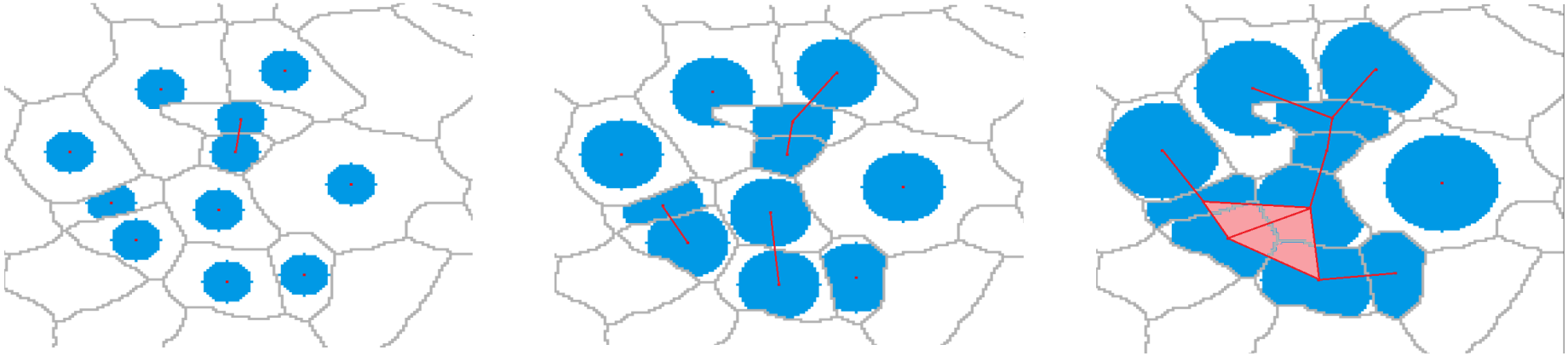


$K_0 \subset K_1 \subset K_2 \subset K_3$

$$\{v_0 \dots v_k\} \in \mathcal{K}_\alpha \Leftrightarrow \bigcap_{i=0\dots k} U_\alpha^i \neq \emptyset.$$

$$U_\alpha^i = B_\alpha^i \cap V_i$$

INSPIRED ON THE ALPHA COMPLEX



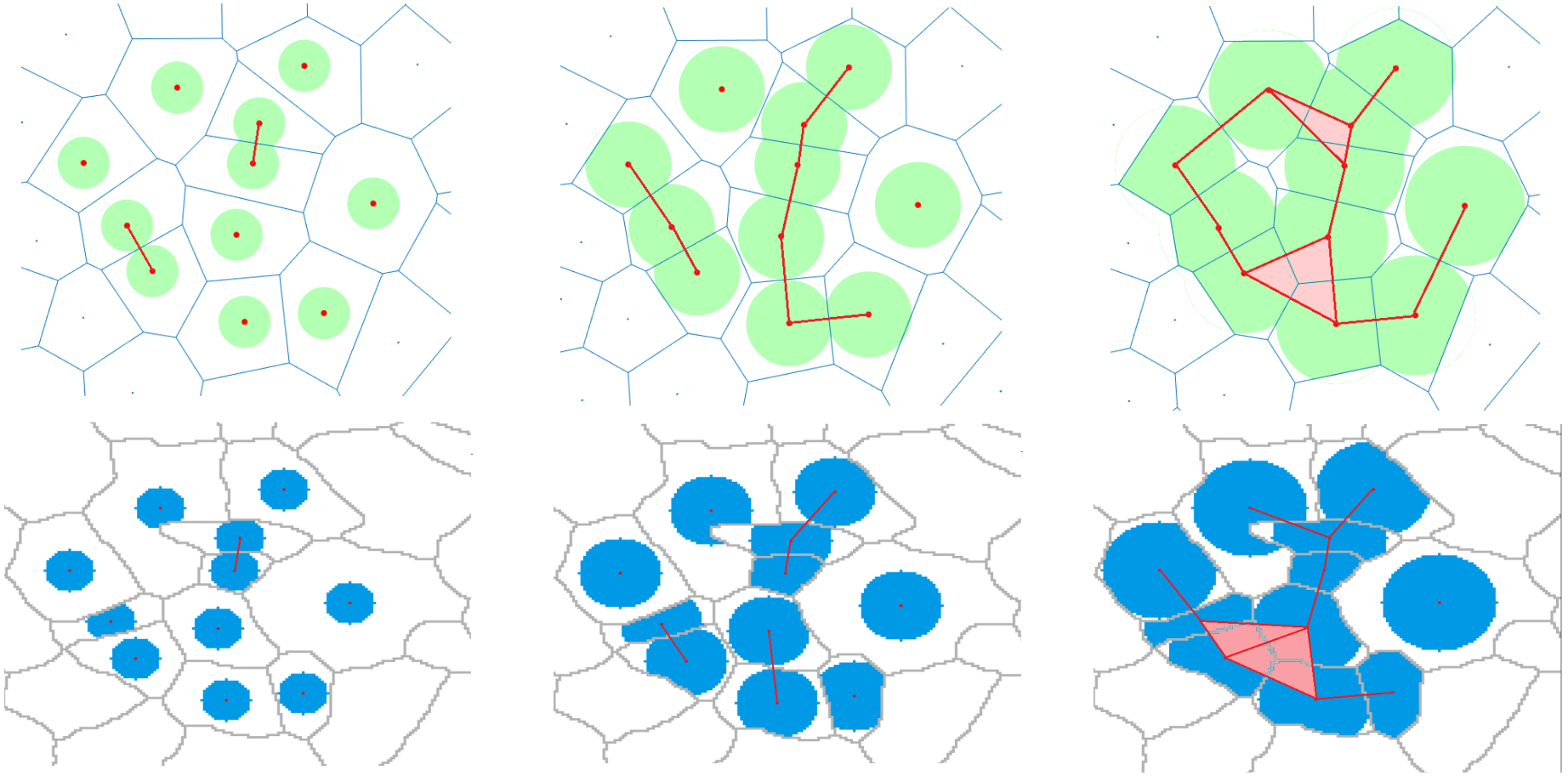
$$\{c_0 \dots c_k\} \in \mathcal{K}_\alpha^{\mathcal{R}} \Leftrightarrow \bigcap_{i=0\dots k} U_\alpha^{\mathcal{R}}(c_i) \neq \emptyset.$$

Simple segmentation complex

$$U_\alpha^{\mathcal{R}}(c_i) = B_\alpha(c_i) \cap R_i$$

Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: *DGMM 2022*, pp. 123-135.

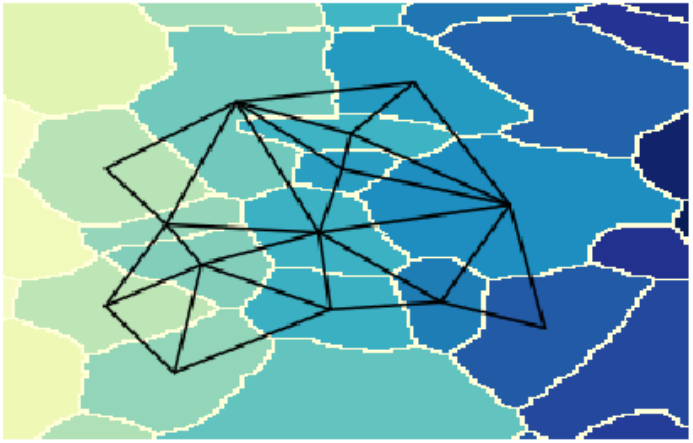
Inspired on the Alpha complex



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: *DGMM 2022*, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

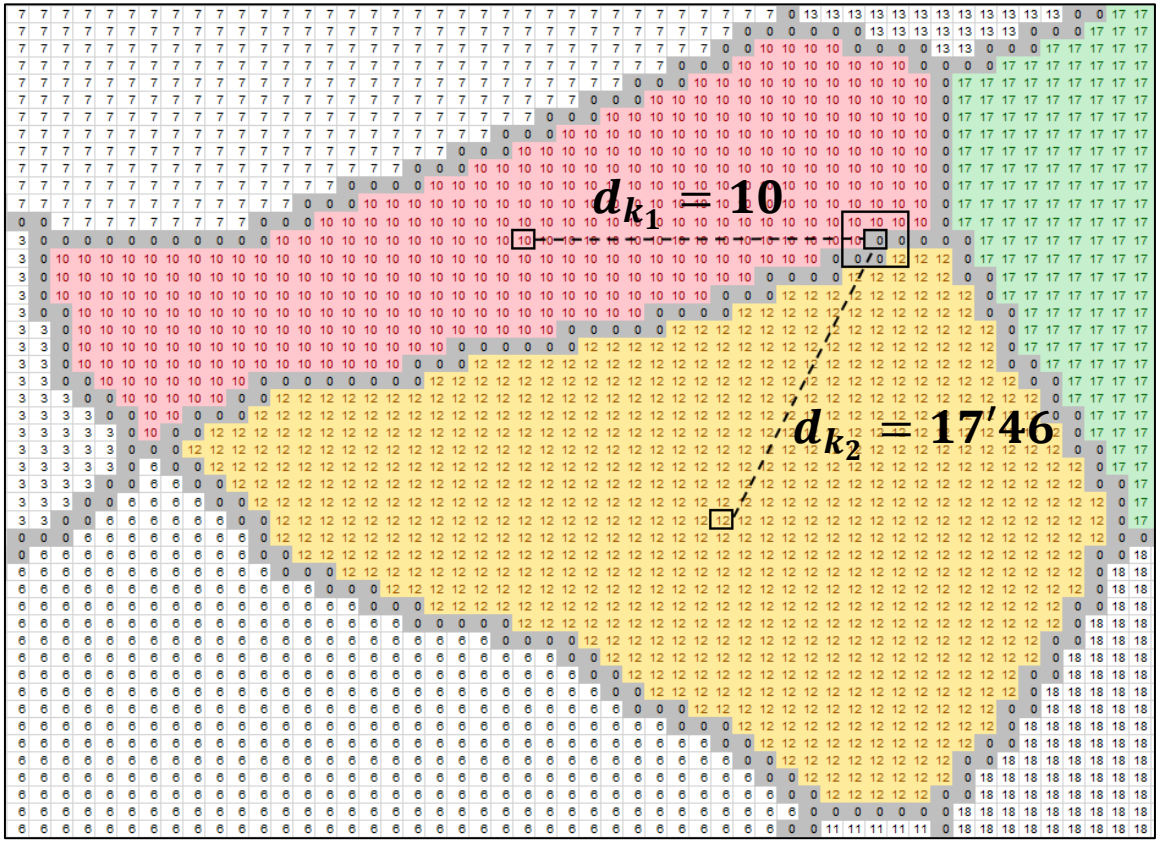
1. Construct the contact graph



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: *DGMM 2022*, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

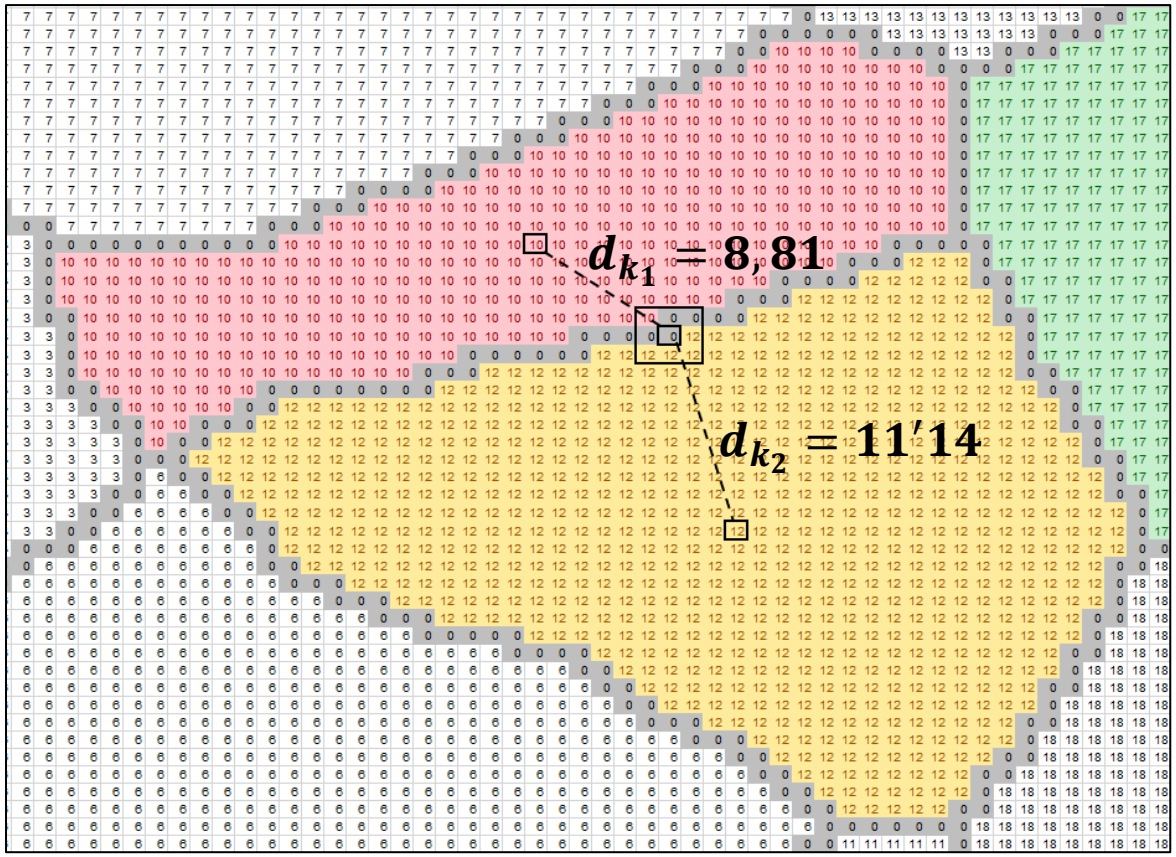
2. For each pair of neighbour regions, compute the edge value for the filtration



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

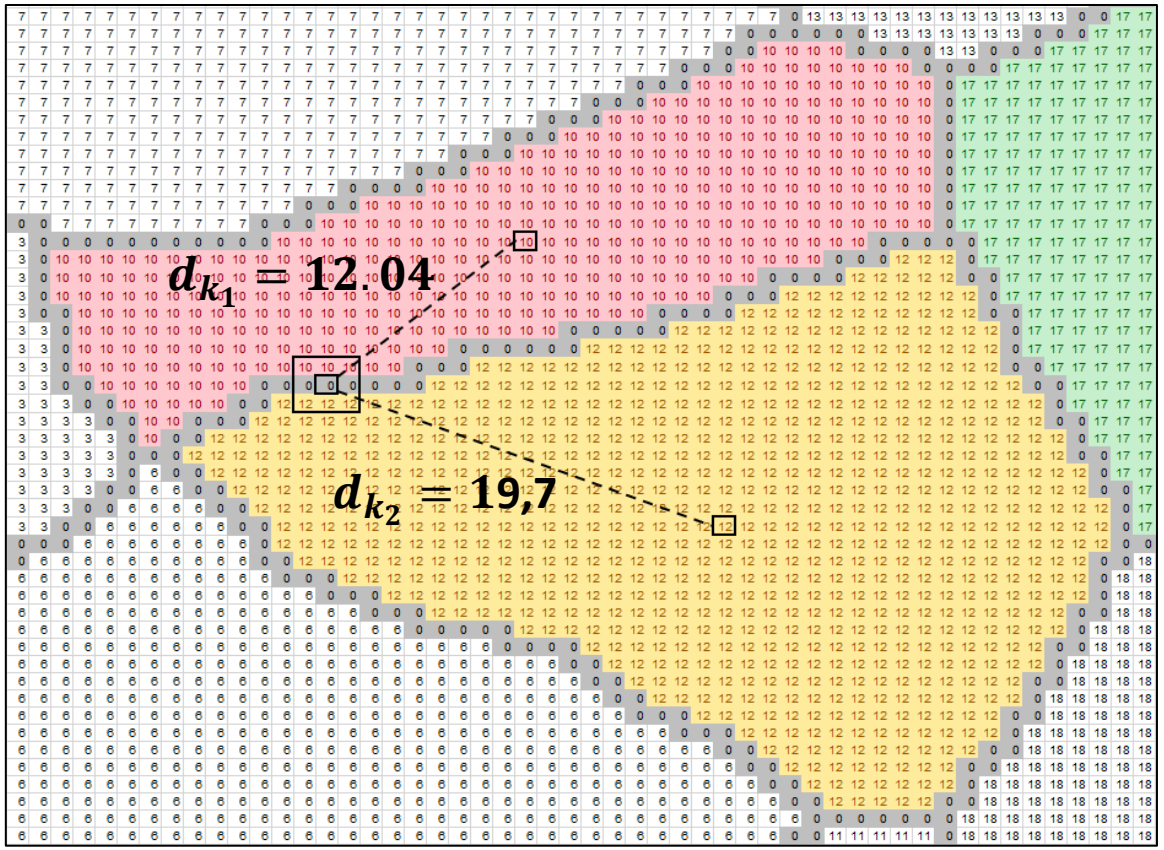
2. For each pair of neighbour regions, compute the edge value for the filtration



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

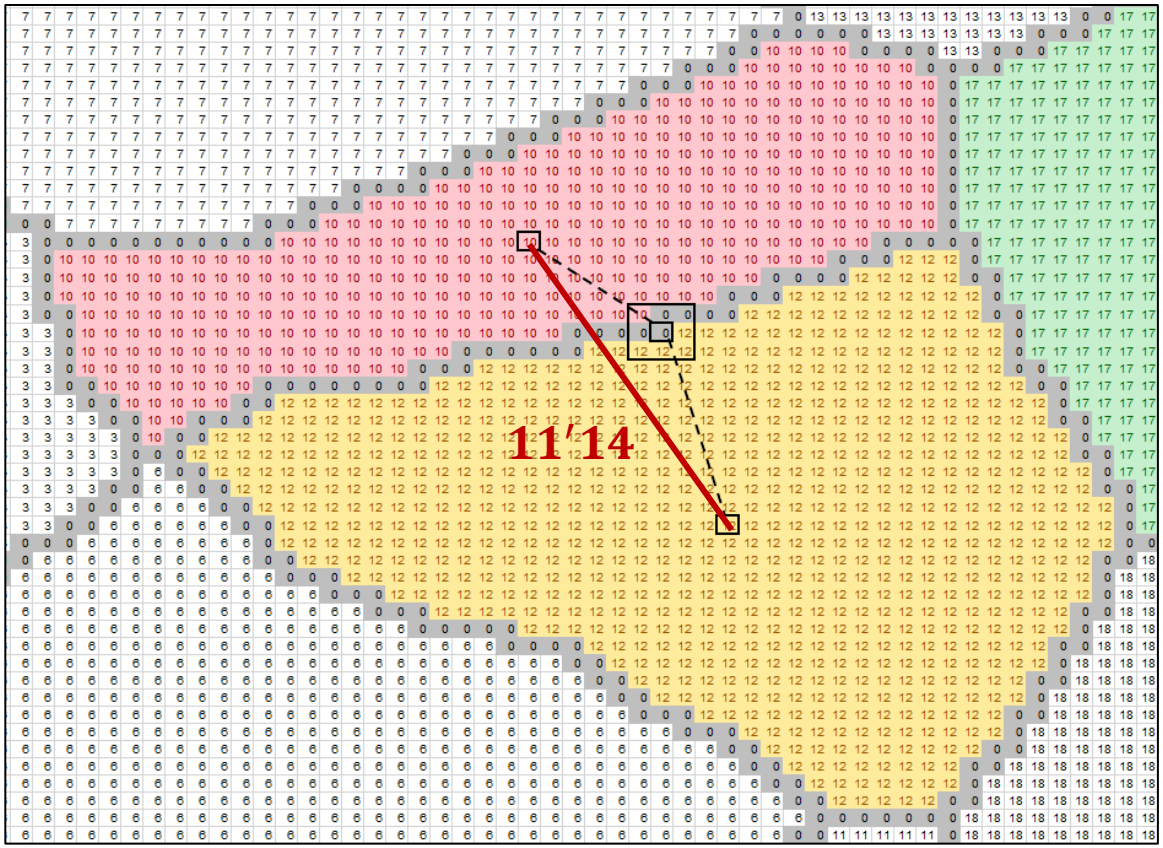
2. For each pair of neighbour regions, compute the edge value for the filtration



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

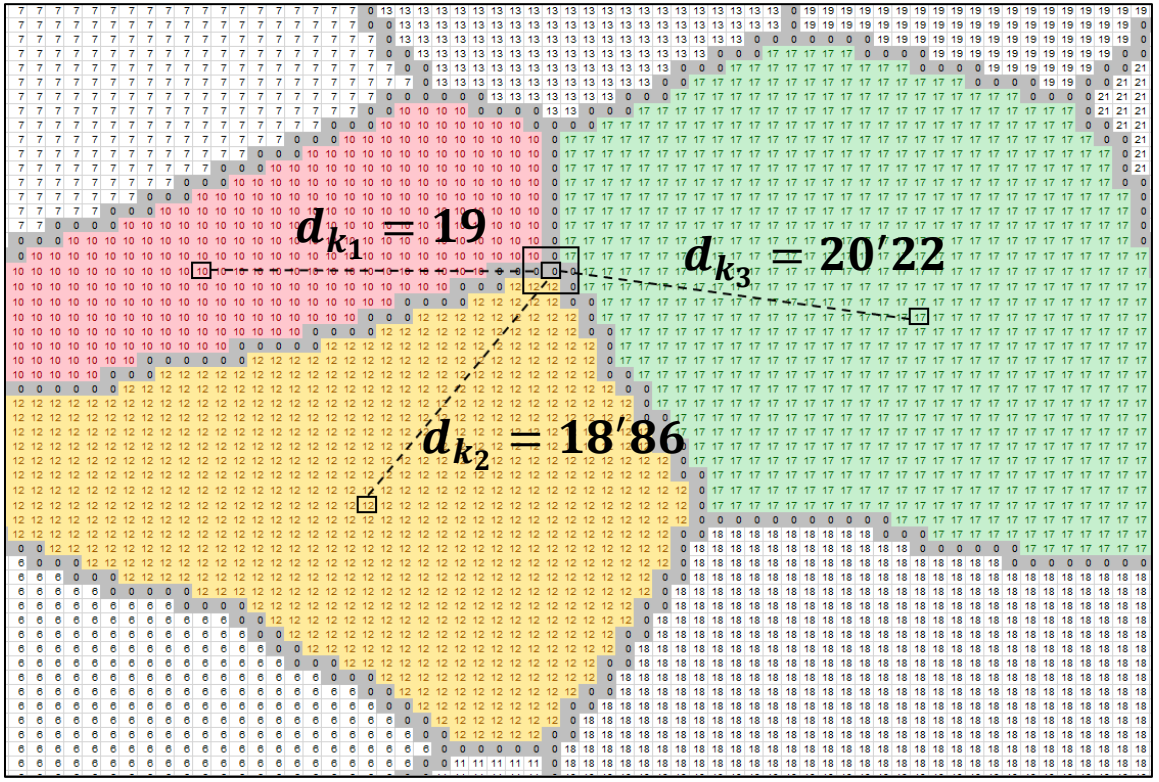
2. For each pair of neighbour regions, compute the edge value for the filtration



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

COMPUTING THE REGIONS COMPLEX

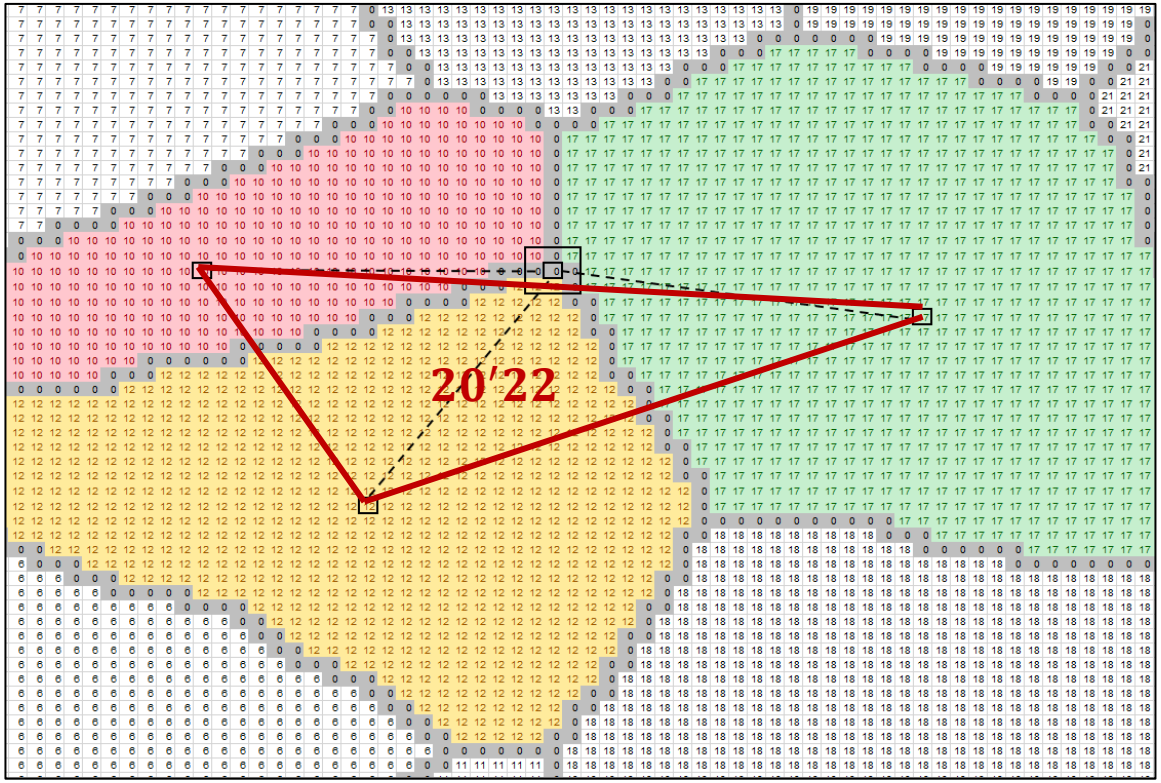
3. For each three (four) adjacent regions, compute the triangle (tetrahedra) value for the filtration



Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

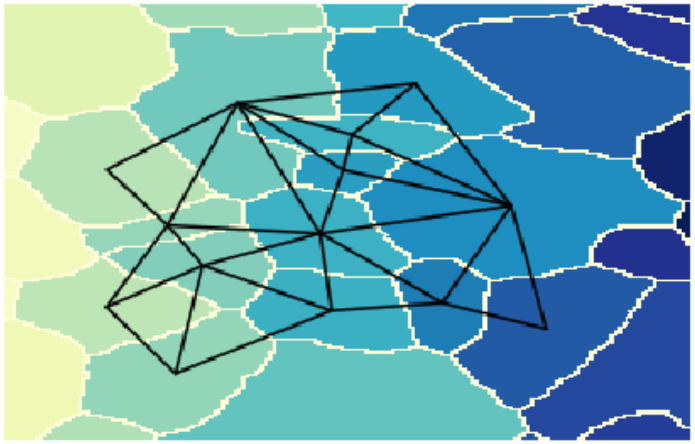
COMPUTING THE REGIONS COMPLEX

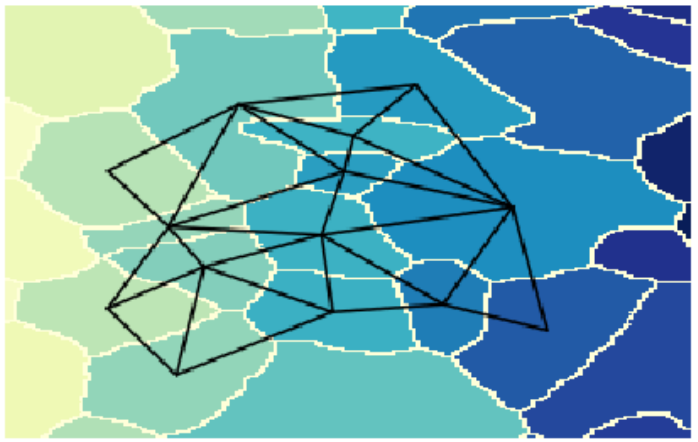
3. For each three (four) adjacent regions, compute the triangle (tetrahedra) value for the filtration

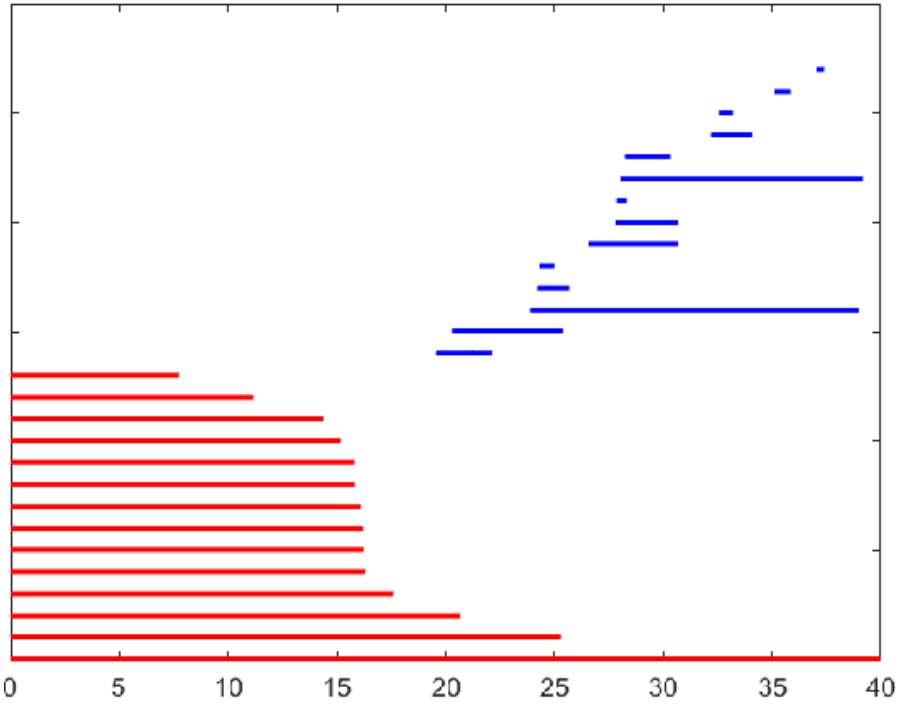


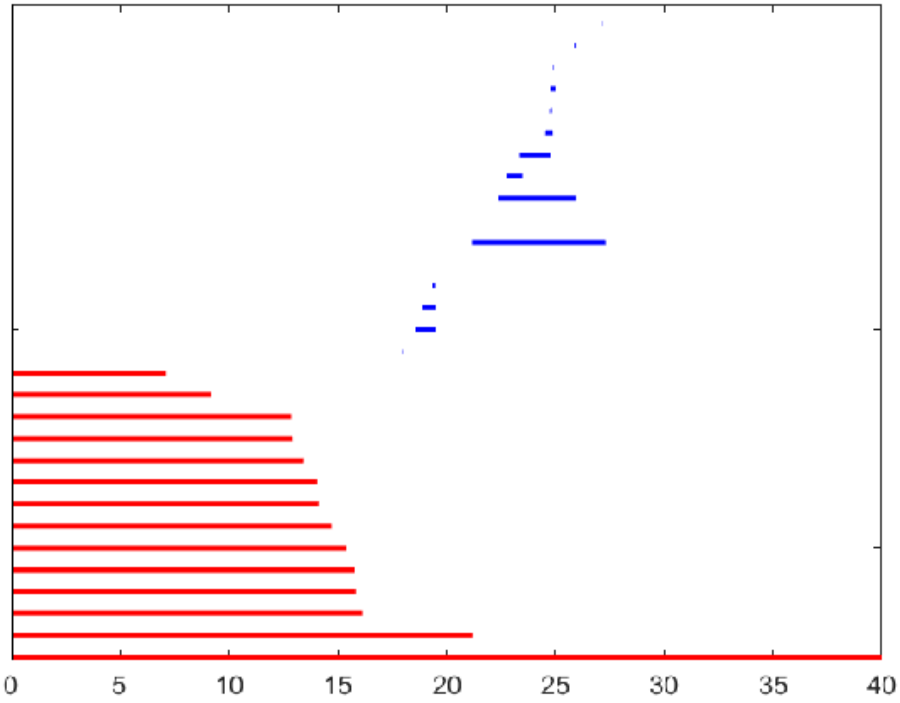
Jimenez, M. J., Medrano, B. *Topological Analysis of Simple Segmentation Maps*. In Proceedings of Discrete Geometry and Mathematical Morphology: DGMM 2022, pp. 123-135.

Regions Complex vs Alpha Complex

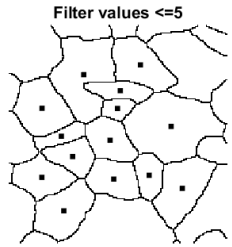


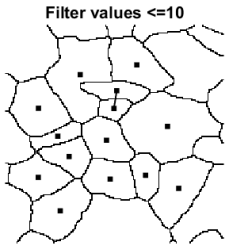


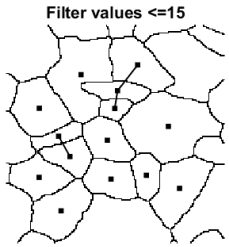


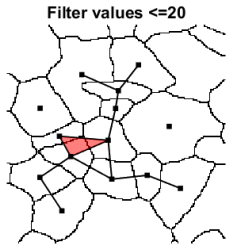


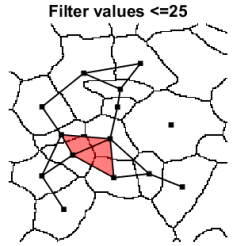
Regions Complex vs Alpha Complex

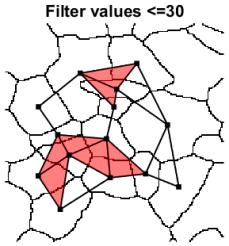


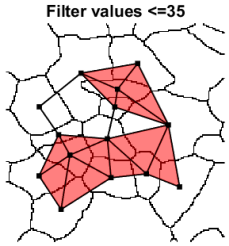


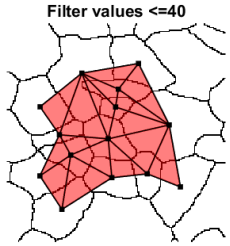


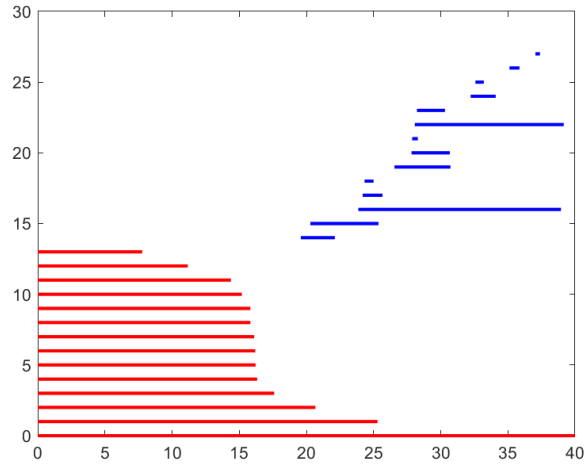


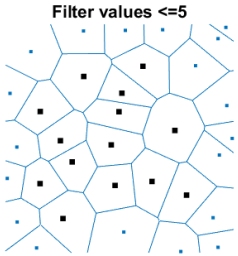


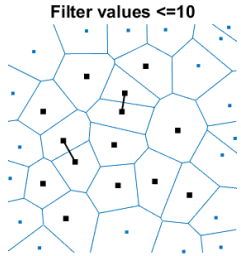


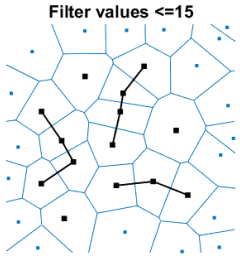


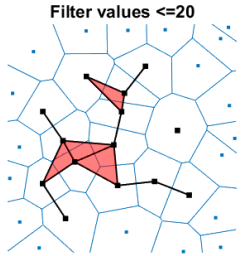


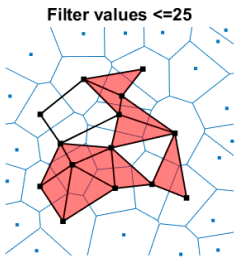


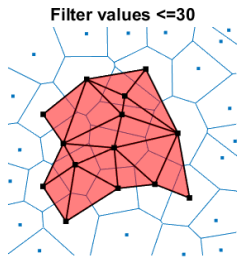


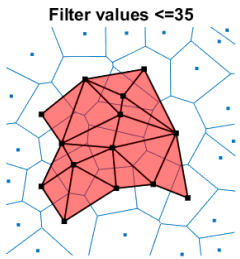


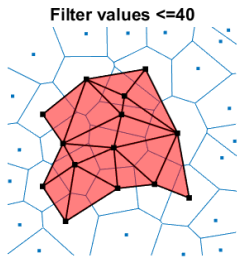


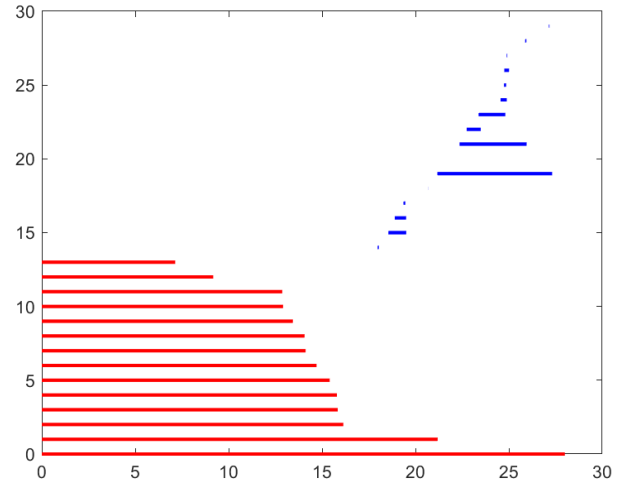




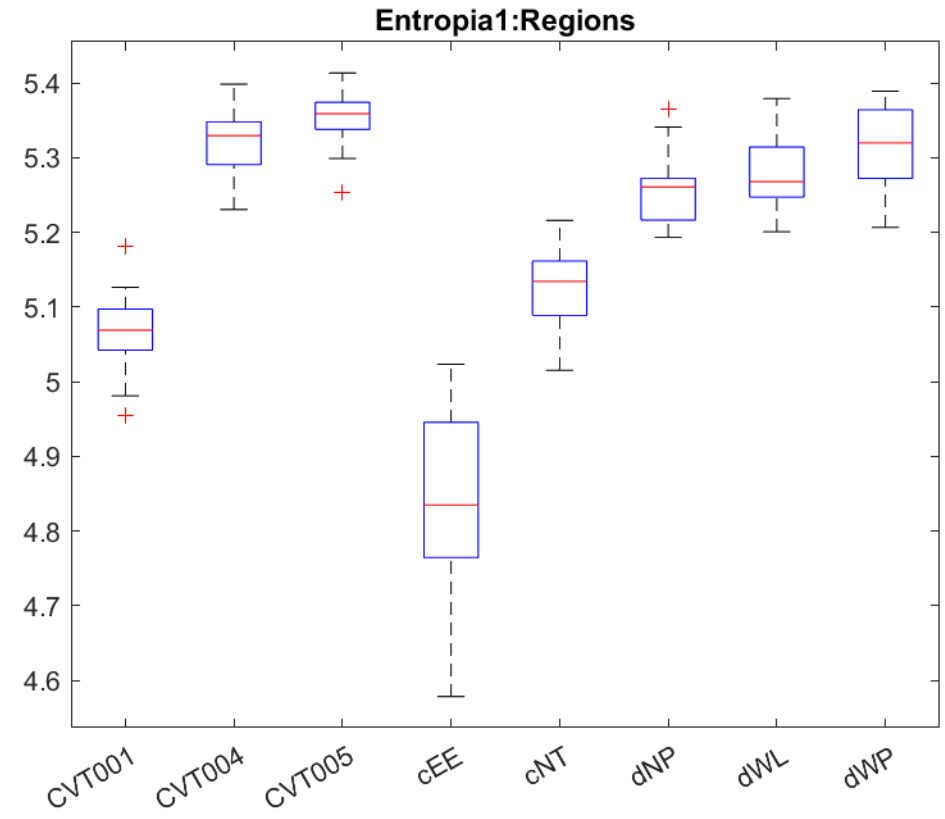


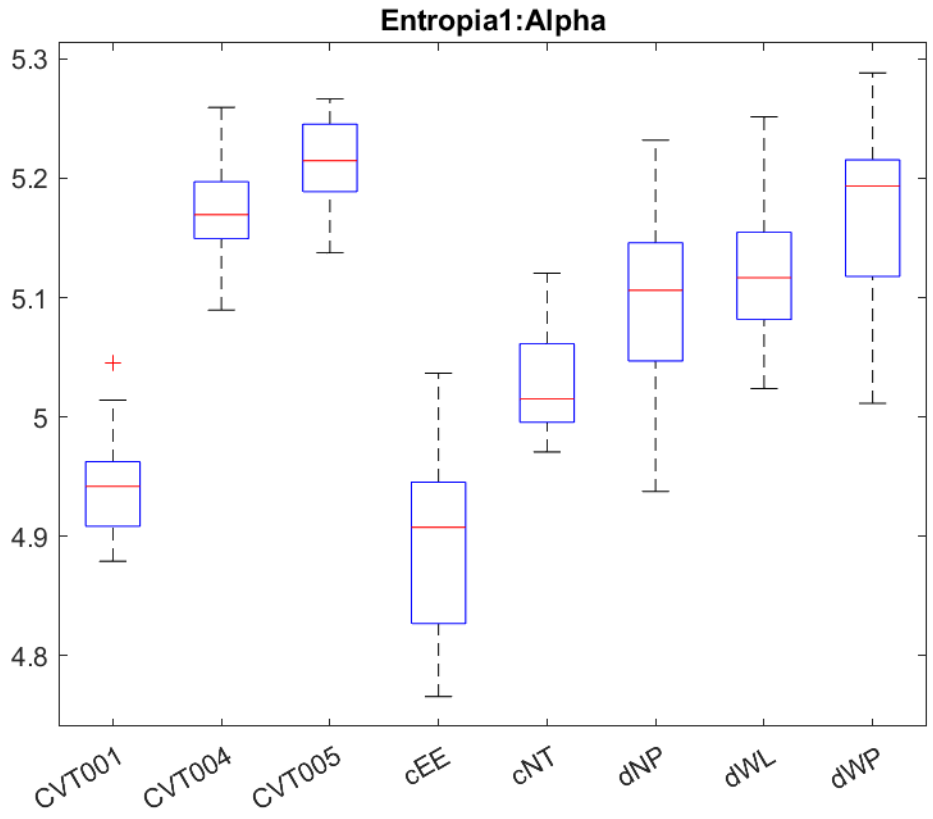




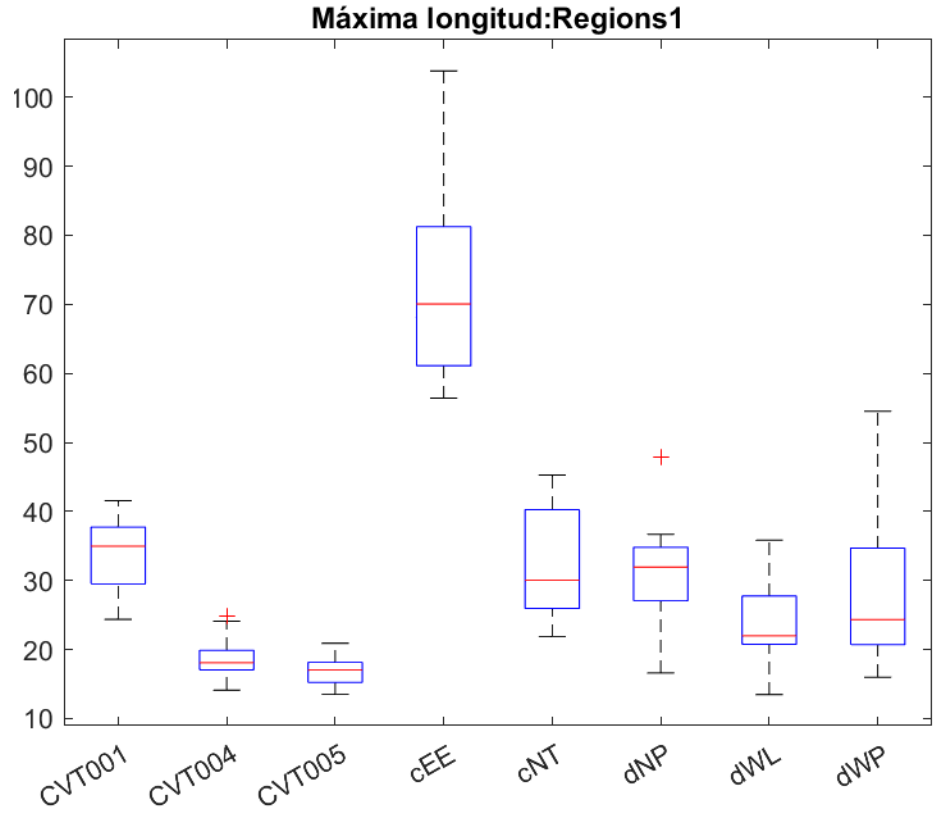


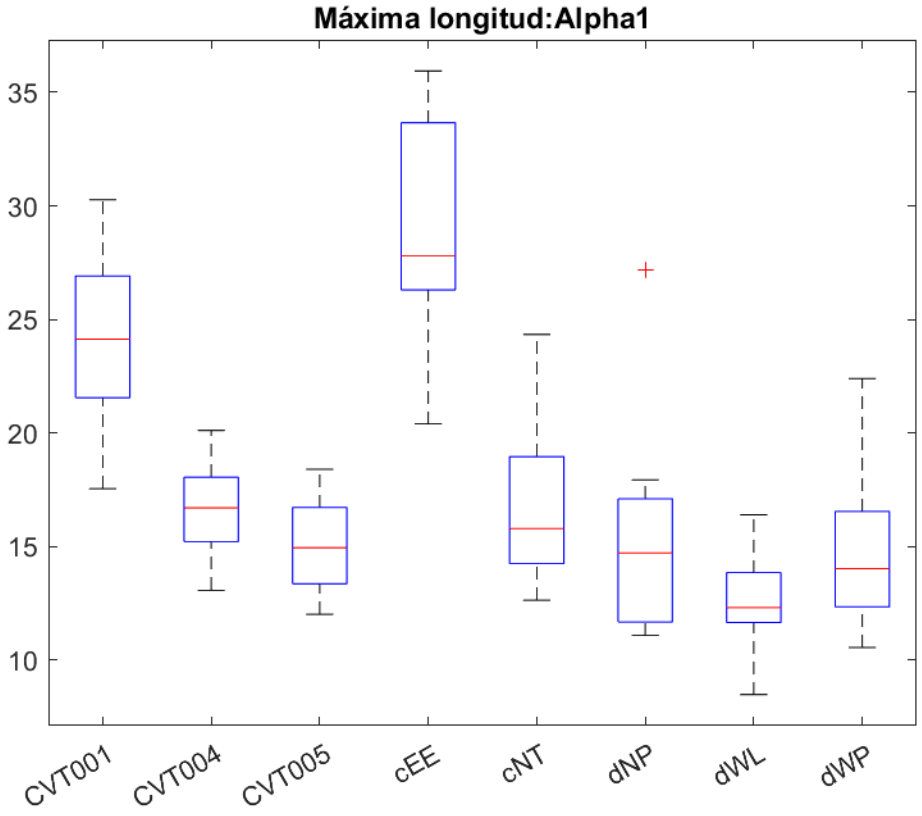
Regions Complex vs Alpha Complex





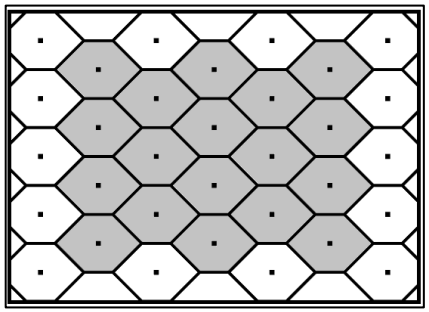
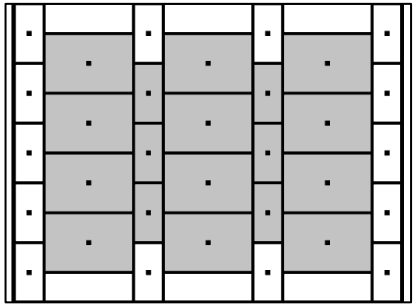
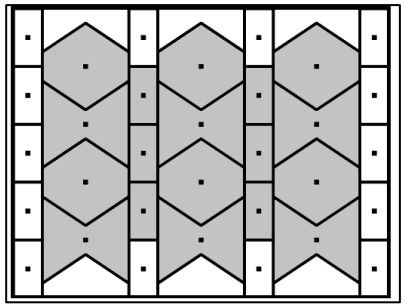
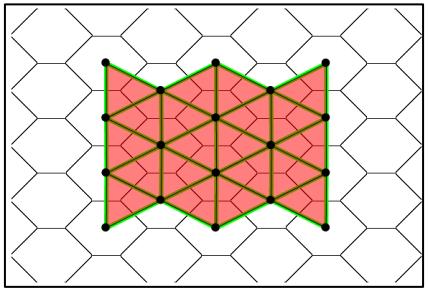
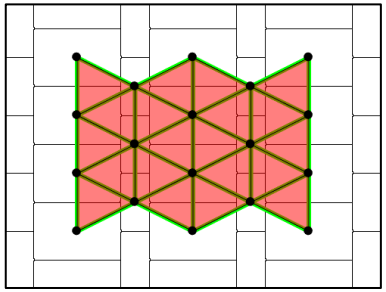
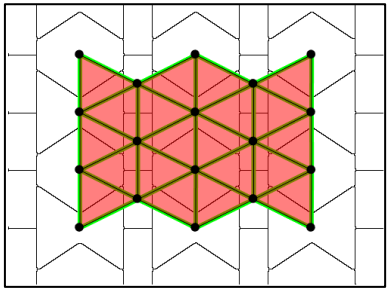
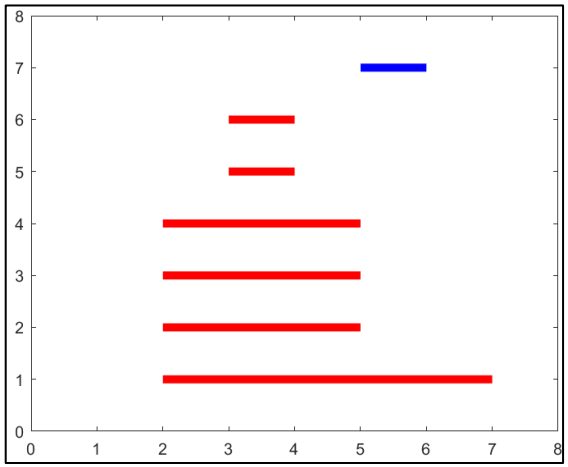
Regions Complex vs Alpha Complex





REGION COMPLEX CAPTURES 'MORE GEOMETRY'

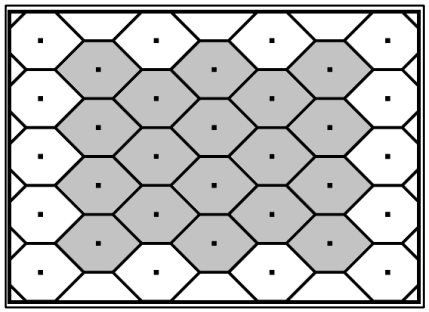
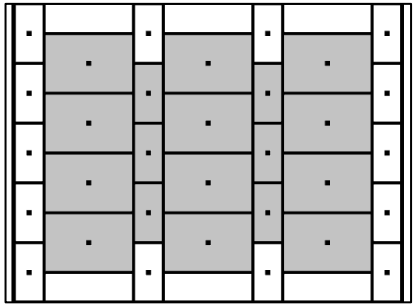
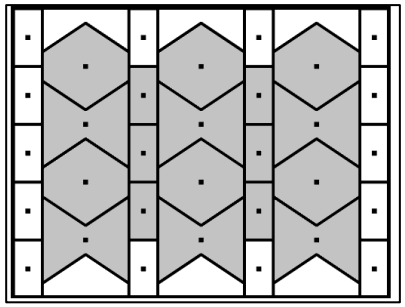
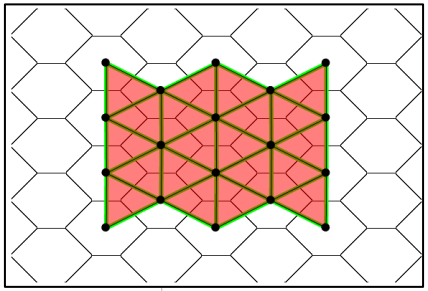
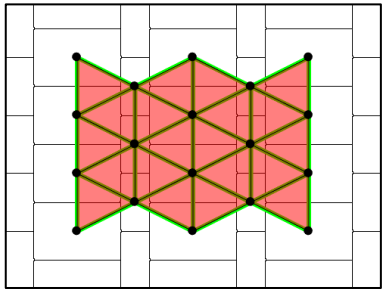
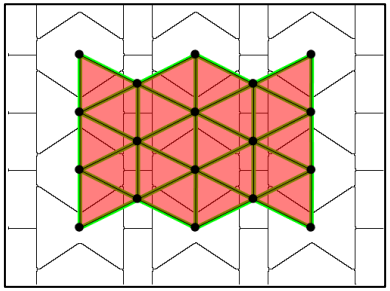
Same spacial distribution of centroids

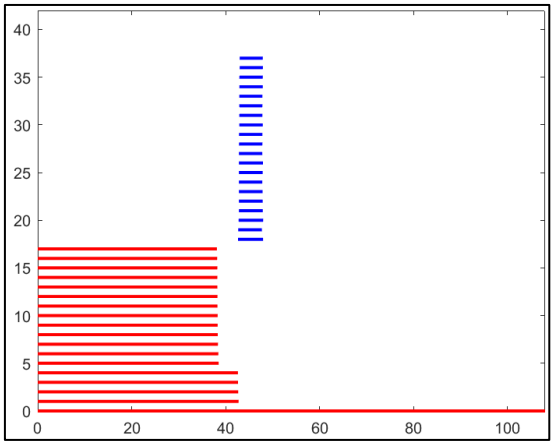








Same contact network

Region complex captures 'more geometry'

Same spacial distribution of centroids

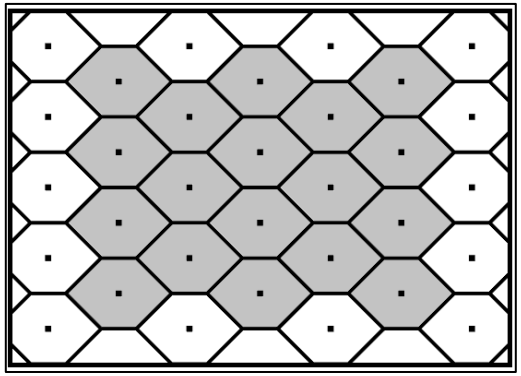


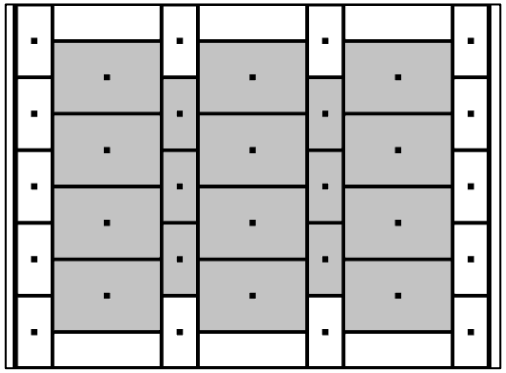
Same alpha complex!

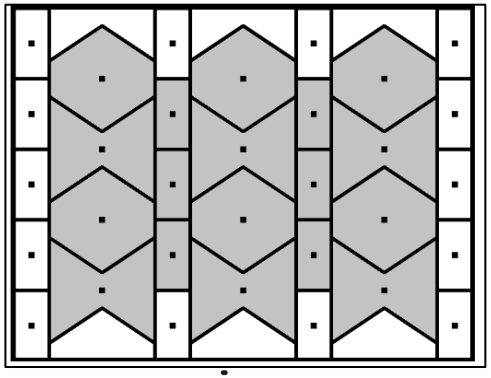
REGION COMPLEX CAPTURES 'MORE GEOMETRY'

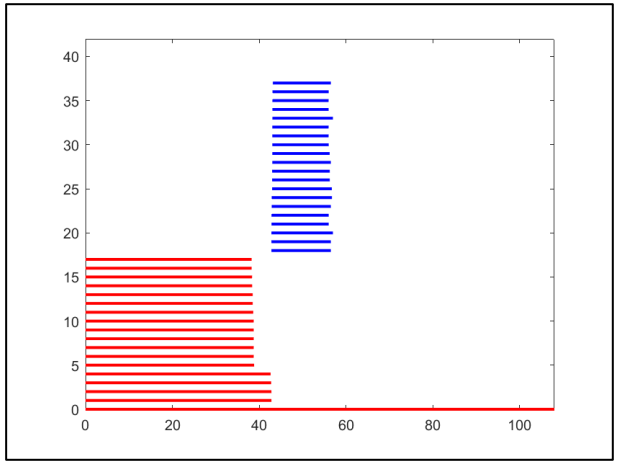
Same spacial distribution of centroids

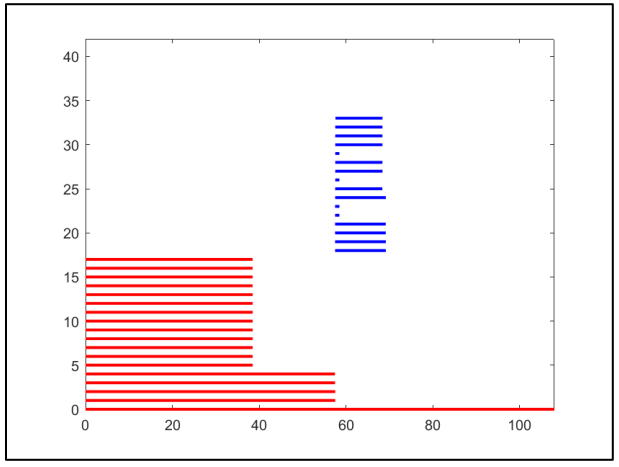
Same contact network

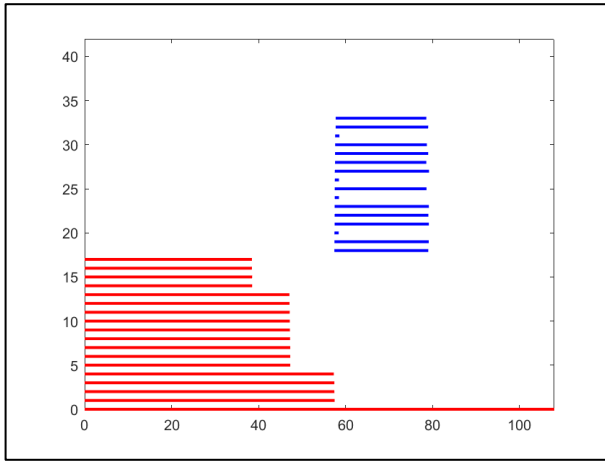










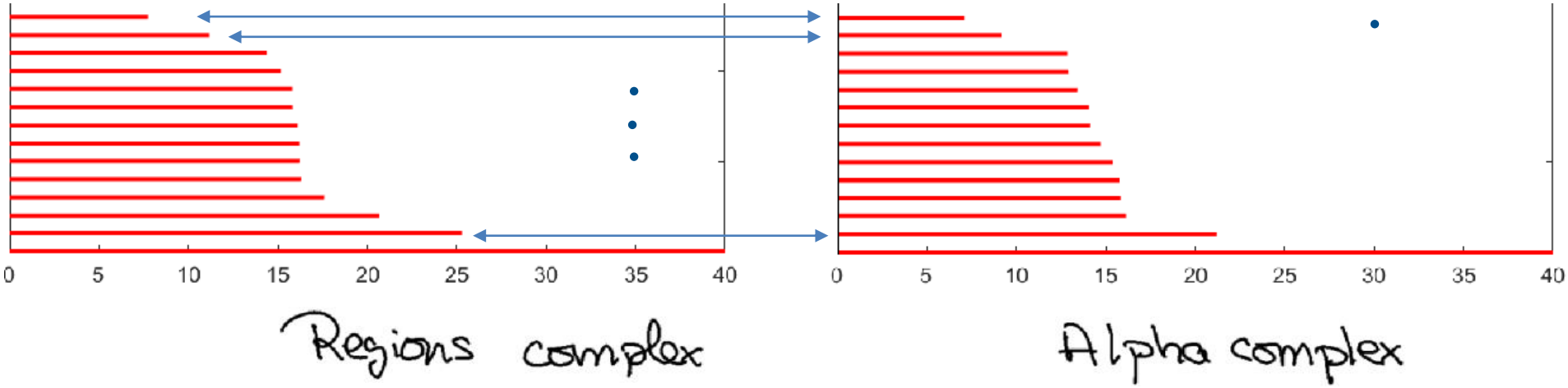


WORK IN PROGRESS

Idea

Value assigned to an edge in the Region complex $\geq$ Value assigned to an edge in the Alpha complex

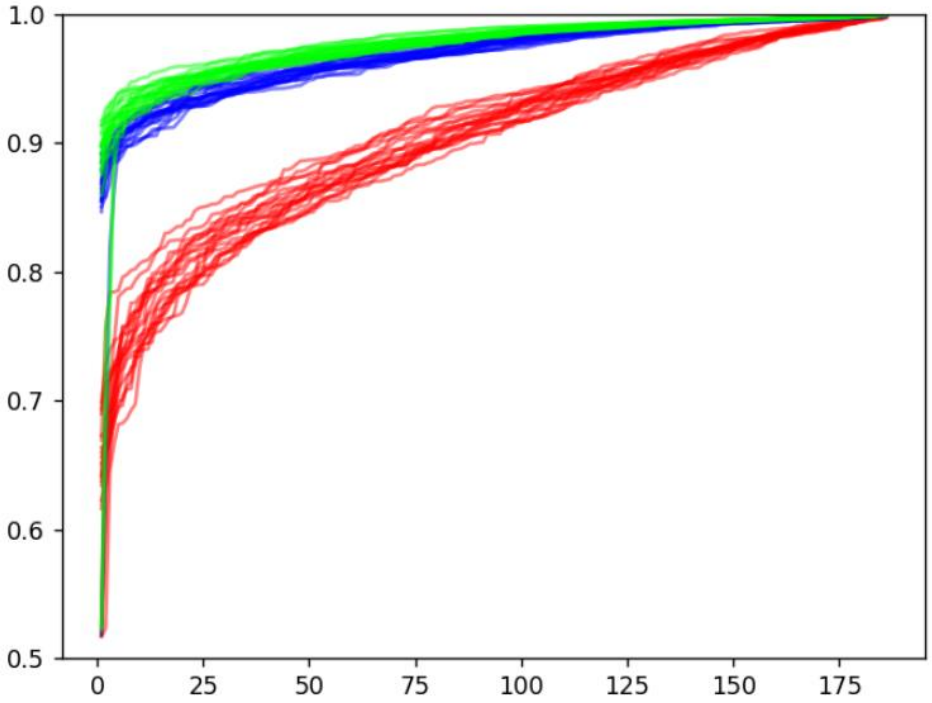
$$0 < \ell_i^{\alpha} / \ell_i^{R} \leq 1$$



Regions complex

Alpha complex

WORK IN PROGRESS

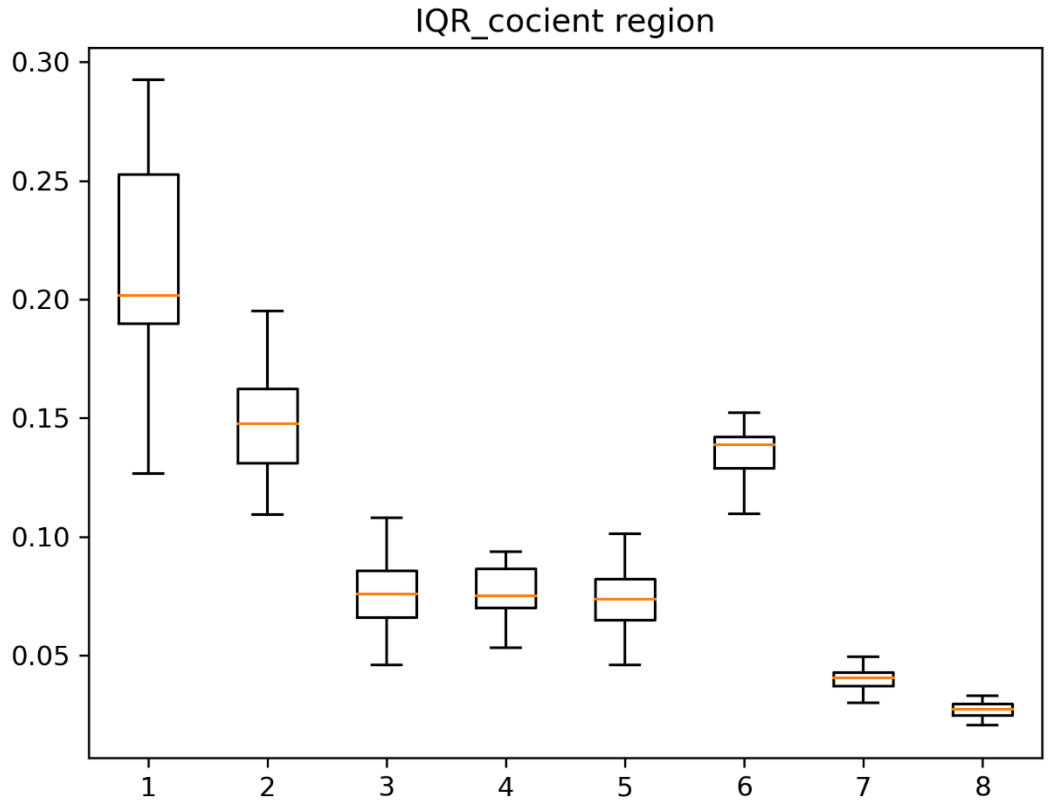


CVT1

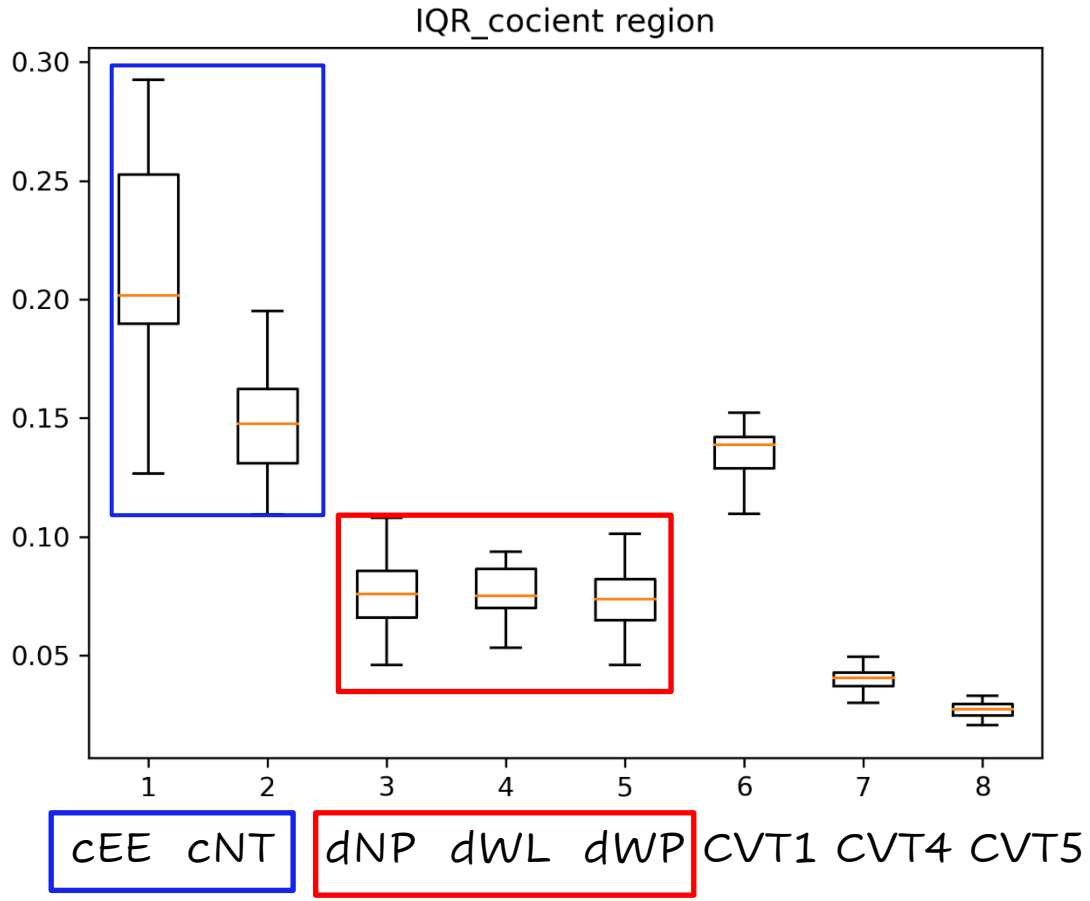
CVT4

CVT5

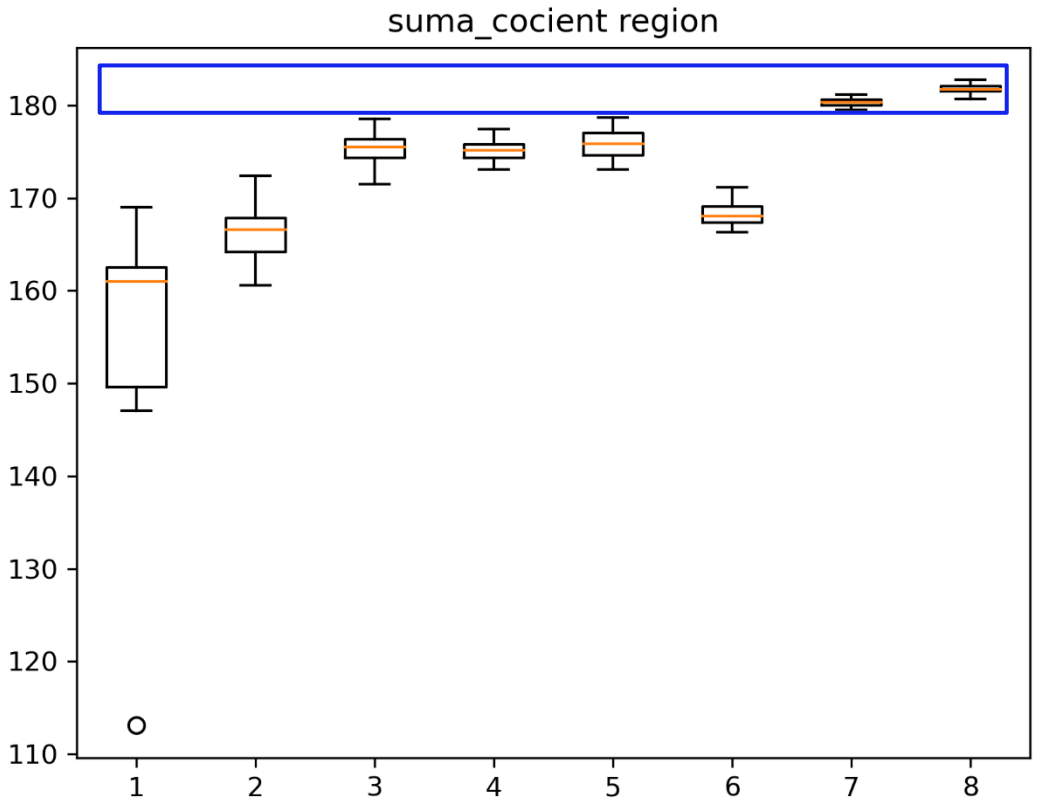
WORK IN PROGRESS



WORK IN PROGRESS



WORK IN PROGRESS



Acknowledgments: Nieves Atienza, Rocío González-Díaz, Belén Medrano, Eduardo Paluzo-Hidalgo, Manuel Soriano-Trigueros; Luis M. Escudero's Team.

Funding: Grant of Convocatoria de la Universidad de Sevilla para la recualificacion del sistema universitario español, 2021-23, funded by the European Union, NextGenerationEU. Spanish grant Ministerio de Ciencia e Innovacion TED2021-129438B-I00; Agencia Andaluza del Conocimiento, grant PAIDI-2020 P20-0114 1813012402; Junta de Andalucia under grant FQM-369; IMUS.