Parametric Discrete Morse Theory

A Master Semester Project

Written by Luca Massimo Nyckees
Supervised by Celia Hacker & Stefania Ebli

Laboratory for Topology and Neuroscience
Directed by Prof. Kathryn Hess Bellwald
Spring semester of 2021

Abstract

Discrete Morse theory has recently found various interesting applications across different related domains of applied mathematics, such as homology computation [2] and topological data analysis [1]. It was developed as a discretized version of the classical smooth Morse theory, by Robin Forman [3]. Discrete Morse theory opened the gate to an approach called *parametric Morse theory* [5]. Parametric Morse theory introduces the notions of *birth* and *death* of critical cells along a sequence of discrete Morse functions $\{f_t : K \to \mathbb{R}\}_{t=1}^{n}$ on a simplicial complex K. This approach produces so-called *persistence parametric diagrams* of critical cells along a family of discrete Morse functions, which can be seen as an analog of persistence diagrams for critical cells. We consider parametric Morse theory as a toolbox to analyze a sequence of vertex real-valued functions on a simplicial complex K. We also provide an implementation of this method in Python.

Contents

Acknowledgments

For their support throughout this whole learning experience, their advice and their ideas, I would like to thank Celia Hacker and Stefania Ebli, who supervised this project and guided me until the end. For the general advice and the help with the code, I am very thankful to Kevin Knudson, Harish Kannan, and Areejit Samal. And finally, for making all of this possible, my sincerest thanks go to Kathryn Hess Bellwald, the director of the *Laboratory for Topology and Neuroscience* at EPFL.

1 Introduction

Discrete Morse theory is a discrete version of smooth Morse theory developed by Robin Forman [3], as a tool to study the geometry and topology of discrete structures in algebraic topology. It has various applications in the domain of applied mathematics, from configuration spaces to computing homology groups. The aspect we are interested in is the application of discrete Morse theory to topological data analysis. In particular, we consider *parametric Morse theory*, a theoretical setting in which one can define the notions of *birth* and *death* of critical cells across a sequence of discrete Morse functions $\{f_t : K \to \mathbb{R}\}_{t=1}^{n}$ defined on a simplicial complex K. This opens the way to interesting applications in the analysis of network dynamics. For example, it was used to detect cavities in a human brain via voxelized data in [5].

Parametric Morse theory studies the time-related behavior of critical cells, which are thus the fundamental object of our study. To better understand this object, we dedicate a part of this report to the presentation of smooth, discrete and parametric Morse theories. Moreover, we give an explanation to relate the notions of life coordinates (birth and death) in the smooth and discrete cases.

For completeness, we recall the basic objects of algebraic topology that are used in discrete Morse theory. We introduce simplicial complexes and CW-complexes in Section 3 and Section 4 respectively. Section 2 and Section 5 respectively deal with smooth and discrete Morse theories. In Section 2, we provide an intuition behind the basic ideas of smooth Morse theory, to be able to make the analogy with the concepts of birth and death in the discrete case.

The goal of this project is to describe a method to analyze the topological features of a sequence of node labelings on a fixed graph G, taking inspiration from parametric Morse theory. More precisely, we consider a sequence of graph node labelings where each labeling in time is determined by a mapping that associates to each node a specific real-valued continuous function. The output of the pipeline are what we call *parametric persistence diagrams* of critical cells that appear along the sequence. Parametric persistence diagrams can be thought of as persistence diagrams for critical cells. We further investigate the robustness of the whole process by controling the final impact of a possible perturbation of the initial node functions.

The general pipeline is implemented in Python. The source code, as well as an illustrative notebook, are available at the following GitHub repository.

GitHub Repository

The setting of parametric Morse theory is described in Section 6. Section 7 is dedicated to the description and implementation of Algorithm 1 in [4]. This is an algorithm designed to extend a node function $g : K^{(0)} \to \mathbb{R}$ on a simplicial complex K to a discrete Morse function $f : K \to \mathbb{R}$. Furthermore, it attains close to the theoretical minimal number of critical cells. In Section 8, we describe in detail the pipeline mentioned above. Results are presented in Section 9. In this section, we run the pipeline and investigate the associated stability analysis. In particular, we run our pipeline on various synthetic datasets.

2 Smooth Morse Theory

In this section, rather than trying to provide a rigorous introduction to the basics of smooth Morse theory, we simply describe a motivating example. Of course, discrete Morse theory has its own applications, but it is always interesting to understand the links between the smooth case and the discrete case. This not only give an intuition on the matter, but also helps understand why we deal with certain objects in discrete Morse theory.

In the following, M will denote a generic smooth manifold.

Notation. In this section, the notation $A \cong B$ indicates the existence of a homeomorphism between the two topological spaces A and B.

Definition 2.1 (Critical point)**.** *Let $f : M \to \mathbb{R}$ be a smooth map. A point $x \in M$ is said to be* critical *for f provided that we have $df_x = 0$.*

At critical points of a map $f : M \to \mathbb{R}$, we are able to define the notion of *Hessian* as follows.

Definition 2.2 (Hessian at critical points)**.** *Let $f : M \to \mathbb{R}$ be a map, and let $x \in M$ be critical for f. We define the* Hessian *of f at x, denoted $(df_x)^2$, by setting*

$$(df_x)^2(X, Y) := X \cdot (\tilde{Y} \cdot f)(x)$$

for all $X, Y \in T_xM$, where T_xM is the tangent space of M at x and $\tilde{Y}$ is a local extension of Y.

We say of a critical point x of a map $f : M \to \mathbb{R}$ that is is *non-degenerate* if the Hessian of f at x is non-singular.

Definition 2.3 (Morse function)**.** *A smooth map $f : M \to \mathbb{R}$ is a* Morse function *if all its critical points are non-degenerate.*

Example 2.4 (Intuition with a 2-torus)**.** *We choose any embedding of the 2-torus $\mathbf{T}$ in the Euclidean space $\mathbb{R}^3$, and consider a height function on $\mathbf{T}$, that projects any point of M to the vertical axis, as shown in Figure 1. We denote the height function by $p : \mathbf{T} \to \mathbb{R}$.*

The critical values are those for which the gradient of the height function vanishes and whose Hessian is non-singular. In this example, they are characterized by saddle points, local maxima and local minima. The non-degeneracy assumption for Morse functions ensures that we do not have

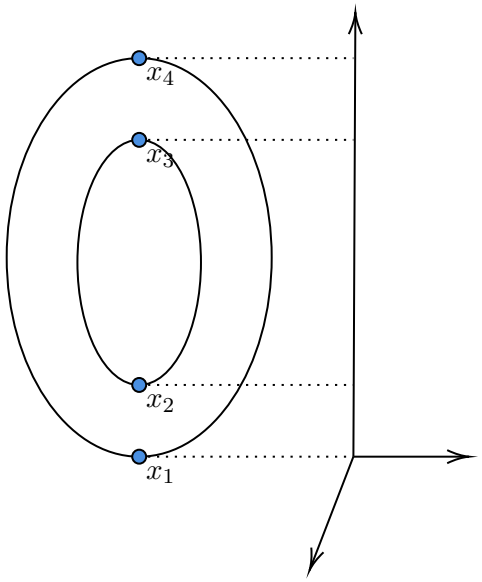


Figure 1: Critical points of a projection defined on the 2-torus.

non-isolated critical points. For example, if we replaced the torus by a vertical cylinder, then the points on top of the cylinder have vanishing gradient, but their Hessian is singular.

The idea behind Morse functions is that they capture enough information about the manifold M. This means we avoid maps $M \to \mathbb{R}$ for which critical points are situated at affine subvarieties of M, like plateaux. The important feature about critical points is that they encode the height levels at which the topology of the Morse function's level sets changes. More precisely, we consider a Morse function $f : M \to \mathbb{R}$ and define subspaces of M of the form $M_a := f^{-1}(]-\infty, a])$. If we progressively increase $a \in \mathbb{R}$, we find that the values at which we observe changes in the topology of the set M_a are precisely the heights of the critical points of f, i.e. the elements of the set $\{f(x)|x \text{ is critical}\}$.

For example, with notations of Figure 1, for all $a, b \in [p(x_1), p(x_2)[$, one has $M_a \cong M_b$. In this interval, the level sets are all circles. When increasing the height to a value of $p(x_2)$, the level set becomes a wedge of two circles, and going further gives the union of two disjoint circles. This provides us with a filtration of the underlying topological space $\mathbf{T}$, where each step indicates

a topological change :

$$\emptyset \subset M_{p(x_1)} \subset M_{p(x_2)} \subset M_{p(x_3)} \subset M_{p(x_4)} = \mathbf{T}.$$

Given a Morse function $f : M \to \mathbb{R}$, we introduce the notions of *stable* and *unstable* manifolds of a critical point of f. These concepts play a crucial role in making a link between the smooth case and the discrete case, when considering the notions of birth and death of critical cells (see Section 6).

First, we recall some notions of differential geometry. For a Morse function $f : M \to \mathbb{R}$, and a critical point $x \in M$, one can approximate the f-values of points y in an appropriate neighborhood of x by the expression

$$f(y) \sim f(x) + \frac{(d^2 f)_x(y - x, y - x)}{2}.$$

This follows from Taylor's theorem, recalling that the gradient of f vanishes at x since it is a critical point. A *Morse chart of* x is a neighborhood of x such that this approximation is actually an equality. Such charts always exist, by the so-called «Morse lemma».

Definition 2.5 (Pseudo-gradient)**.** *Let $f : M \to \mathbb{R}$ be a function. A* pseudo-gradient for f *is a vector field X satisfying*

(a) $(df_x)(f) \leq 0$ *for all* $x \in M$*,*

(b) $(df_x)(f) = 0$ *if and only if* $x \in M$ *is critical,*

(c) On any Morse chart around a critical point x, the vector fields X and $-\mathrm{grad}(f)$ *coincide.*

Definition 2.6 ((Un)stable Manifolds)**.** *Let $f : M \to \mathbb{R}$ be a Morse function, and let $x \in M$ be a critical point of f. Consider a pseudo gradient field X for f, and let ϕ^t denote the vector flows of X. The* stable *and* unstable *manifolds of x are defined, respectively, as*

$$W^s(x) = \{y \in M | \lim_{t \to \infty} \phi^t(y) = x\}, \text{ and}$$

$$W^u(x) = \{y \in M | \lim_{t \to -\infty} \phi^t(y) = x\}.$$

Example 2.7 ((Un)stable manifolds)**.** *We look at a manifold N embedded into $\mathbb{R}^3$, with associated height function $p : N \to \mathbb{R}$, as shown in Figure 2. The height function p is Morse, and has four critical points, denoted*

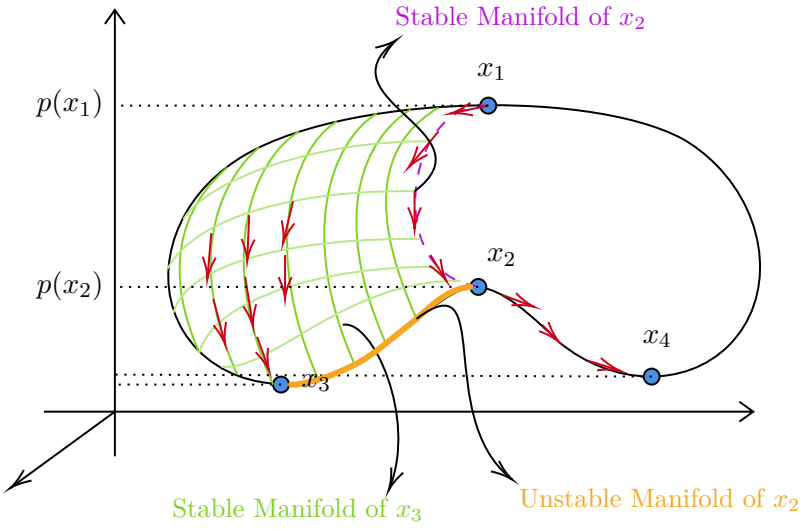


Figure 2: Stable and unstable manifolds of critical points with respect to a height function defined of a manifold N embedded in $\mathbb{R}^3$. The red arrows represent a vector flow $\phi : N \times \mathbb{R} \to \mathbb{R}$ of the vector field $\text{grad}(p)$.

$x_1, ..., x_4$. *We consider a vector flow* ϕ^t, *indicated by red arrows in the figure. Roughly speaking, the stable manifold of a critical point* x_i *consists in the set of points of* N *that are the starts of some paths in the vector flow* ϕ^t *ending up at* x_i. *Conversely, the unstable manifold of a critical point* x_i *is the set of points that are the endpoints of some paths in the vector flow starting at* x_i.

For example, we have $W^u(x_1) = N$, *as any point of the manifold is the endpoint of a path starting at* x_1. *Moreover, the stable manifold of* x_2 *is a circle passing through the points* x_1 *and* x_2. *Indeed, it consists of the only points of* N *that are the start of a path ending in* x_2. *See Figure 2 for an example.*

3 Simplicial Complexes

A simplical complex is a set of simplices (generalized triangles, e.g. tetrahedrons) that are glued together in a way that respects a specific hierarchy. More precisely, we have the following definitions. First, we recall the definition of *geometric simplex* to get a «Euclidean» intuition on simplices. That

said, we shall only consider *abstract simplices* later on.

Definition 3.1 (Geometric Simplex)**.** *Let $k \in \mathbf{N}$. A* k-dimensional simplex *(or* k-simplex*)* α *is the convex hull built on a set of affinely independent points $V = \{v_0, ..., v_{k+1}\} \subset \mathbb{R}^k$. Symbolically, one has*

$$\alpha = \{\sum_{i=0}^{k} a_i v_i | \sum_{i=0}^{k} a_i = 1\}.$$

It follows that simplices are, apart from vertices and edges, higher dimensional triangles (e.g. tetrahedrons). For example, a 2-simplex is a triangle, a 3-simplex is a tetrahedron and a 4-simplex is a 4-cell (homeomorphic to the disk D^4). *Regular* k-simplices are k-simplices where, with notations as in the definition above, the points $v_1, ..., v_k$ are the standard unitary vectors of $\mathbb{R}^k$ and $v_0 = \mathbf{0}$. Simplices are most often studied up to homotopy, *i.e.* continuous deformation. Thus, one can always consider the case of regular simplices and imagine nice regular triangles and tetrahedrons. We define the *face* of a k-simplex to be a convex hull built on a subset (that is not the entire simplex) of the points that constitute the k-simplex. Faces are simplices themselves. For two simplices α and β, we denote by $\alpha < \beta$ (or equivalently $\beta > \alpha$) the fact that α is a face of β.

Definition 3.2 (Abstract Simplex)**.** *An* abtract simplex σ *is a finite set of vertices of the form $\{v_i\}_{i=0}^n$, which we denote by $\sigma = [v_0, ..., v_n]$. An abtract n-simplex is an abtract simplex built on n vertices.*

Definition 3.3 (Abstract Simplicial Complex)**.** *An* abstract simplicial complex X *is a set of abstract simplices such that both conditions (1) and (2) expressed below are satisfied.*

(1) $\beta < \alpha \in X \implies \beta \in X$

(2) $\alpha, \beta \in X$ *and* $\alpha \cap \beta \neq \emptyset \implies \alpha \cap \beta < \alpha$ *and* $\alpha \cap \beta < \beta$

When considering an abstract simplicial complex, we shall simply use the term *simplicial complex*. Simplicial complexes can be obtained from graphs by completing some cliques, in the sense that we add higher-dimensional interactions between vertices. Simplicial complexes are the structure we choose for running the experiments of this project, since they are the most natural object on which to define discrete Morse theory.

4 CW-Complexes

Although we do not use CW-complexes in our computations, this class of objects plays an important role in understanding the basics of discrete Morse theory, especially in the formulation of minimal decomposition results (e.g. Theorem 5.9). In particular, they lead to very elegant analogies with the case of smooth Morse theory, and thus meaningful interpretations.

CW-complexes are topological spaces that are built inductively, by attaching cells progressively to some original space. We have the following definitions.

Definition 4.1 (Cellular attachment)**.** *Let X be a topological space and let $n \in \mathbb{N}$. Let e^n be an n-cell (i.e. homeomorphic to the n-disk) and consider a continuous map $f : S^{n-1} \to X$. The* glued space $X \cup_f e^n$ *is the topological space that makes the diagram below commute.*

$$\begin{array}{ccc} S^{n-1} & \xrightarrow{f} & X \\ {\scriptstyle i}\big\downarrow & & \big\downarrow \\ D^n & \longrightarrow & X \cup_f e^n \end{array}$$

The glued space $X \cup_f e^n$ is said to be obtained by attaching *an n-cell to X via the gluing application $f : S^{n-1} \to X$. The map $i : S^{n-1} \to D^n$ is the natural embedding application of the boundary of the disk D^n into itself.*

Definition 4.2 (CW-complex)**.** *A CW-complex is a space X that can be obtained by induction on a chain of the form $X_0 \subset X_1 \subset \cdots \subset X_d = X$, where the space X_i is obtained from attaching a cell to the space X_{i-1}.*

5 Discrete Morse Theory

In this section, we introduce the notion of *discrete Morse theory*, as defined by Robin Forman (see [3]), and have a look at some of its related fundamental results.

Notation. For the entirety of sections 5 and 6, we consider a generic simplicial complex K. We shall also denote its underlying set of simplices as K, for practical purposes. For a simplex $\alpha \in K$, we denote by $\mathcal{B}_\alpha^F$ and $\mathcal{B}_\alpha^C$ the sets of its codimension-1 faces and cofaces, respectively. Finally, we shall use the symbol $\simeq$ to denote a homotopy equivalence between two spaces.

Definition 5.1. *Let $f : K \to \mathbb{R}$ be a real-valued function. If conditions (1) and (2) below are satisfied, we say f is a* discrete Morse function *on X.*

(1) $\sum_{\beta \in \mathcal{B}_\alpha^F} \mathbb{1}_{\{f(\beta) \geq f(\alpha)\}} \leq 1$, $\forall \alpha \in K$

(2) $\sum_{\beta \in \mathcal{B}_\alpha^C} \mathbb{1}_{\{f(\beta) \leq f(\alpha)\}} \leq 1$, $\forall \alpha \in K$

The idea behind such a function is that, with at most one break of the rule, it assigns numbers to simplices in an increasing way, when considering the ordering given by the binary relation $\alpha \prec \beta \iff \dim(\alpha) < \dim(\beta)$.

Definition 5.2. *If the sums in Definition 5.1 are both zero for the same given simplex $\alpha \in K$, then we say that $\alpha \in K$ is a* critical simplex*. We will denote by $\mathrm{Cr}_i^f(X)$ the set of i-dimensional critical cells of X with respect to the discrete Morse function $f : K \to \mathbb{R}$.*

We now put those concepts in practise, by considering the case of the sphere. It is worth the time to make a link with smooth Morse theory. For a smooth manifold M with homeomorphic simplicial complex Y, we have the following correspondence, provided that the simplicial decomposition attains the minimum number of critical cells (this is explained later on): critical 0-cells, critical 1-cells and maximum-dimensional critical cells of Y correspond to local minima, saddle points and local maxima, respectively.

Example 5.3 (Discrete Morse function on the sphere S^2)**.** *Let us consider the simplicial complex X given by a triangulation of the sphere S^2 in Figure 3, where edges and vertices are identified as shown by the coloring. We build a discrete Morse function on X as described by the values on the vertices and edges. This way, we obtain exactly one critical vertex and one critical 2-cell, namely the vertex with value 9 and the 2-cell with value 21.*

We see this agrees with the note above : those critical cells indicate the existence of exactly one local maximum and one local minimum of S^2 when considering the smooth Morse theory approach.

We come to the main result of discrete Morse theory. It resonates with the observation that the sphere S^2 admits a CW-complex decomposition made of one 0-cell and one 2-cell (we build the sphere by contracting the boundary of a disk D^2 to a point, *i.e.* we attach a 2-cell to a 0-cell).

Theorem 5.4. *Let $f : K \to \mathbb{R}$ be a discrete Morse function on a simplicial complex $X = (V, K)$. Then X admits a homotopy equivalent CW-complex decomposition consisting of exactly $|\mathrm{Cr}_i^f(X)|$ i-cells.*

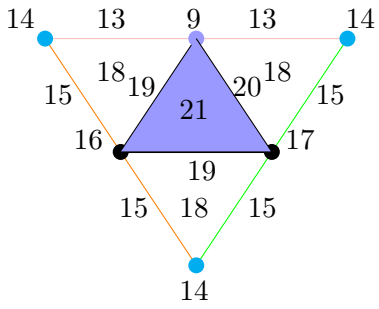


Figure 3: A discrete Morse function on S^2. The two critical cells (a vertex and a triangle) are highlighted in violet (taking Morse values 9 and 21).

Before giving a proof of Theorem 5.4, we make a few observations.

Definition 5.5. *Let $\alpha < \beta$ be simplices in X. If β is the only coface of α, then we say α is a* free face *of β.*

Remark 5.6. *Whenever we have an inclusion of simplicial complexes $X_a \subset X_b$ such that $X_b \setminus X_a = \{\alpha, \beta\}$, where α is a free face of β, we can find a deformation retraction from X_b to X_a. This process of deformation retraction by canceling exactly two simplices is called a* simplicial collapse. *For example, one can collapse a triangle to a point in three steps of simplicial collapsing. We say that the triangle is* collapsible. ♠

Notation. Given our generic simplicial complex K, we define, for any real value $x \in \mathbb{R}$, a subcomplex of K by setting $K(x) := \cup_{\alpha<\beta} \cup_{\beta \in f^{-1}(x)} \beta$.

Proposition 5.7. *Let $x < y$ be two values in $\mathbb{R}$. If $K(y) \setminus K(x)$ contains no critical cell, then we can deformation retract $K(y)$ to $K(x)$.*

Proof. If no critical cells are contained in $K(y) \setminus K(x)$, it means we only have pairs of regular cells. By the definition of a discrete Morse function, given a pair of regular cells $\alpha < \beta$, α is necessarily a free face of β. Hence, we can obtain a series of simplicial collapses, which proves the claim. ■

Proposition 5.8. *Let $x < y$ be two values in $\mathbb{R}$. If the set $K(y) \setminus K(x)$ contains exactly one critical cell $\alpha^{(d)}$ for some $d \in \mathbb{N}$, then we can find a gluing map $g : S^{d-1} \to D^d$ so that we have a glued space $K(x) \cup_g D^d \simeq K(y)$.*

Proof. One obtains $K(y)$ by adding $\alpha^{(d)}$ to $K(x)$ in a way that coincides with attaching a d-cell to $K(x)$. Indeed, $\alpha^{(d)}$ is a d-cell and it is glued to

$K(x)$ by being patched to its boundary. This follows from the fact that α is critical : its boundary already belongs to $K(x)$. For example, if $d = 2$ we would patch a 2-cell to an empty triangle, obtaining a cell e^2. Hence, what we really do each time is adding a critical cell of dimension d to a subcomplex which is covering (or patching) a $d-1$-dimensional hole of $K(x)$. ■

Proof of Theorem 5.4. Let us consider a filtration

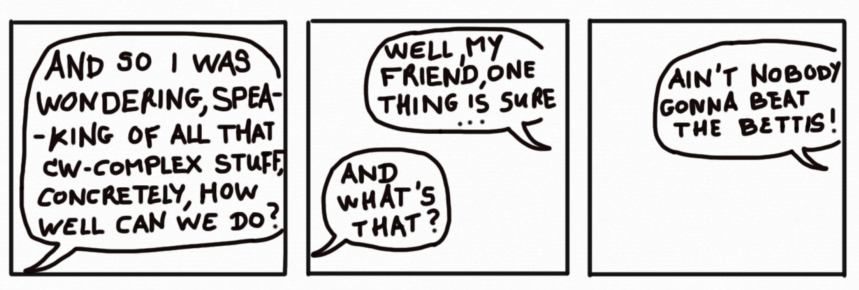
$$K(x_0) \subset K(x_1) \subset \cdots \subset K(x_m) = K,$$

where $\{x_0, ..., x_m\}$ is precise enough so that each complement set of the form $K(x_i) \setminus K(x_{i-1})$, $i \in \{1, ..., m\}$, contains either no critical cell or exactly one critical cell. Then it follows from the two propositions above that K is obtained, up to homotopy equivalence, by consecutively attaching the critical cells of the Morse function f. Indeed, if $\{x_{j_0}, ..., x_{j_k}\} \subseteq \{x_0, ..., x_m\}$ represents the steps of the filtration in which we add the critical cells, then we know that K has a CW-complex decomposition given by the filtration

$$K(x_{j_0}) \subset K(x_{j_1}) \subset \cdots \subset K(x_{j_k}).$$

Hence the result. ■

Now, one may wonder why this result is important. What it tells us is that finding a *good* Morse function on a simplicial complex X, *i.e.* with a minimal number of critical cells, ensures that there exists a *good CW*-complex decomposition of X, *i.e.* with a minimal number of cells in each dimension. This minimality refers to the proof of the result below, that mathematically answers the question *how well can we do?*



Theorem 5.9 (Morse Inequalities)**.** *Let m_d denote the number of critical d-simplices for some discrete Morse function $f : K \to \mathbb{R}$ on K, let $b_d = \dim(H_d(K))$, and let c_d denote the number of d-cells appearing in some CW-complex decomposition of X. One obtains*

(1) $b_d \leq m_d \ \forall d \in \{0, ..., \dim(K)\}$,

(2) $b_d \leq c_d \ \forall d \in \{0, ..., \dim(K)\}$.

Proof. By using Theorem 5.4, it easy to see that proving condition (2) is sufficient. We state and prove a family of inequalities, from which condition (2) is quickly deduced. In fact, we show that for any $d \in \{0, ..., \dim(K)+1\}$, one has $\sum_{i=0}^{d-i} c_i \geq \sum_{i=0}^{d-i} b_i$. This follows from the calculation provided below, where $\delta_i : C_i(X) \to C_{i-1}(X)$ are the boundary maps used to define the homology theory of CW-complexes ($C_i(X)$ is the $\mathbb{Z}$-module generated by the i-cells appearing in the CW-complex decomposition of X). One has

$$\begin{aligned}
\sum_{i=0}^{d}(-1)^{d-i}b_i &= \sum_{i=0}^{d}(-1)^{d-i}\dim(H_i(X)) \\
&= \sum_{i=0}^{d}(-1)^{d-i}[\dim(\ker(\delta_i)) - \dim(\mathrm{im}(\delta_{i+1}))] \\
&= \sum_{i=0}^{d}(-1)^{d-i}\dim(\ker(\delta_i)) + \sum_{i=0}^{d}(-1)^{d-i+1}\dim(\mathrm{im}(\delta_{i+1})) \\
&= \sum_{i=0}^{d}(-1)^{d-i}\dim(\ker(\delta_i)) + \sum_{i=1}^{d+1}(-1)^{d-i}\dim(\mathrm{im}(\delta_i)) \\
&= \sum_{i=0}^{d}(-1)^{d-i}\dim(\ker(\delta_i)) + \sum_{i=0}^{d}(-1)^{d-i}\dim(\mathrm{im}(\delta_i)) \\
&\quad - (-1)^{d}\dim(\mathrm{im}(\delta_0)) - \dim(\mathrm{im}(\delta_{d+1})) \\
&= \sum_{i=0}^{d}(-1)^{d-i}\dim(C_i(X)) - \dim(\mathrm{im}(\delta_{d+1})) \\
&= \sum_{i=0}^{d}(-1)^{d-i}c_i - \dim(\mathrm{im}(\delta_{d+1})) \\
&\leq \sum_{i=0}^{d}(-1)^{d-i}c_i,
\end{aligned}$$

where we use that $\dim(\mathrm{im}(\delta_0)) = 0$. ■

Manually constructing a discrete Morse function can easily become complicated when considering non-trivial simplicial complexes. For this reason, a recurring question in discrete Morse theory is the way we should build algorithmically discrete Morse functions. There are algorithms designed to

execute this task in a rather efficient way (e.g. in [6] or [4]). Before presenting one of these algorithm, we will now present a nice tool allowing to better understand the geometry of discrete Morse functions, namely the notion of *gradient vector field*. We use the notation $\gamma^{(p)}$ to indicate the fact that a simplex γ is a p-simplex.

Definition 5.10 (Discrete vector field)**.** *We define a* discrete vector field V *on* K *to be a set of pairs* $\{\alpha^{(p)} < \beta^{(p+1)}\} \in K \times K$ *such that* $|A_\gamma| \leq 1$ *for each* $\gamma \in K$*, where we define the set* $A_\gamma := \{(\alpha, \beta) \in V | \gamma \in \{\alpha, \beta\}\}$.

Definition 5.11 (V-path)**.** *Let* V *be a discrete vector field on* K*. A* V-path *is a finite sequence of simplices of the form* $\{\alpha_1^{(p)}, \beta_1^{(p+1)}, ..., \alpha_n^{(p)}, \beta_n^{(p+1)}, \alpha_{n+1}^{(p)}\}$ *such that* $(\alpha_i, \beta_i) \in V$ *satisfy the relations* $\alpha_i < \beta_i > \alpha_{i+1}$ *and* $\alpha_i \neq \alpha_{i+1}$ *for all* $i \in \{1, ..., n\}$*. Such a* V*-path is a* V-cycle *provided that* $\alpha_1 = \alpha_{n+1}$.

We consider again a discrete Morse function $f : K \to \mathbb{R}$. One observes that non-critical (or *regular*) simplices come in pairs: suppose $\alpha \in K$ is a codimension-1 face of $\beta \in K$ with $f(\alpha) \geq f(\beta)$. Then, by symmetry, $\beta \in K$ is a codimension-1 coface of $\alpha \in K$ with $f(\beta) \leq f(\alpha)$ and, by definition of discrete Morse function, this inversion they share is their only inversion (*i.e.* they belong to no other such pair). Now, in this case, we adopt the convention of drawing an arrow from α to β, so as to get an ordered pair $(\alpha, \beta) \in K \times K$, which we call a *regular pair*. A *gradient vector field of* f is a discrete vector field that is built with respect to f, meaning that its pairs are regular pairs in the sense described above.

Definition 5.12 (Gradient vector field)**.** *Let* $f : K \to \mathbb{R}$ *be a discrete Morse function on* K*. A* gradient vector field of f *is a discrete vector field* $V = \{(\alpha, \beta)\}$ *where each pair* (α, β) *is a regular pair.*

A connected finite sequence of arrows (*i.e.* regular pairs) is a *gradient path*.

Definition 5.13 (Gradient path)**.** *Let* V *be a discrete vector field on* K*, and* $f : K \to \mathbb{R}$ *a discrete Morse function on* K*. A* V*-path is a* gradient path of f *if* V *is a gradient vector field of* f.

Now, we observe that gradient paths of a discrete Morse function $f : K \to \mathbb{R}$ are in particular sequences along which f takes decreasing values. With notations as in the definition above, one has the inequalities below.

$$f(\alpha_1) \geq f(\beta_1) > f(\alpha_2) \geq f(\beta_2) > f(\alpha_3) \geq \cdots$$

Indeed, let us fix $i \in \{1, ..., n\}$. Then $f(\alpha_i) \geq f(\beta_i)$ follows from the fact that (α_i, β_i) is a regular pair. *Ab absurdo*, suppose that $f(\beta_i) \leq f(\alpha_{i+1})$.

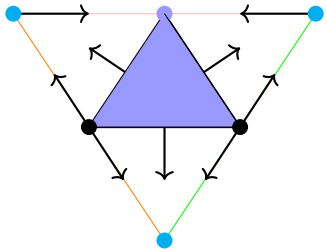


Figure 4: Discrete vector field on a triangulation of the sphere S^2. In purple are the two critical cells. The identifications of the triangulation between the various simplices are indicated with same coloring.

Then, since α_{i+1} is a face of β_i, we have that (α_{i+1}, β_i) is a regular pair. But regular pairs are pairwise disjoint and thus $\alpha_i = \alpha_{i+1}$, a contradiction. What is elegant about this is the fact that the converse is true as well (see Proposition 5.14).

Proposition 5.14. *Let V be a gradient vector field of some discrete Morse function f. Then a finite sequence $\{\alpha_1^{(p)}, \beta_1^{(p+1)}, ..., \alpha_n^{(p)}, \beta_n^{(p+1)}, \alpha_{n+1}^{(p)}\}$ is a gradient path iff we have*

(1) $\alpha_i < \beta_i$ and $\beta_i > \alpha_{i+1}$ for all $i \in \{1, ..., n\}$

(2) $f(\alpha_1) \geq f(\beta_1) > f(\alpha_2) \geq f(\beta_2) > f(\alpha_3) \geq \cdots$

As a consequence, gradient paths do not form cycles. It turns out that the converse statement is also true, *i.e.* discrete vector fields with no cycles admit some potential. We can say a discrete vector field V is derived from a potential f if V is the gradient vector field of the discrete Morse function f. More precisely, we formulate the result below.

Theorem 5.15. *For any discrete vector field V with no cycles, there exists a discrete Morse function $f : K \to \mathbb{R}$ for which V is a gradient path.*

For a proof of Theorem 5.15, see [3].

Example 5.16 (Discrete vector field on the sphere S^2)**.** *We go back to the sphere S^2, triangulated as a tetrahedron. One can build a discrete vector field as the set of arrows shown in Figure 4. This is the gradient vector field of the Morse function exhibited in Example 5.3.*

Remark 5.17 (General flow behavior)**.** *Generally speaking, an interesting phenomenon can be seen : following the arrows along a* maximal *path leads*

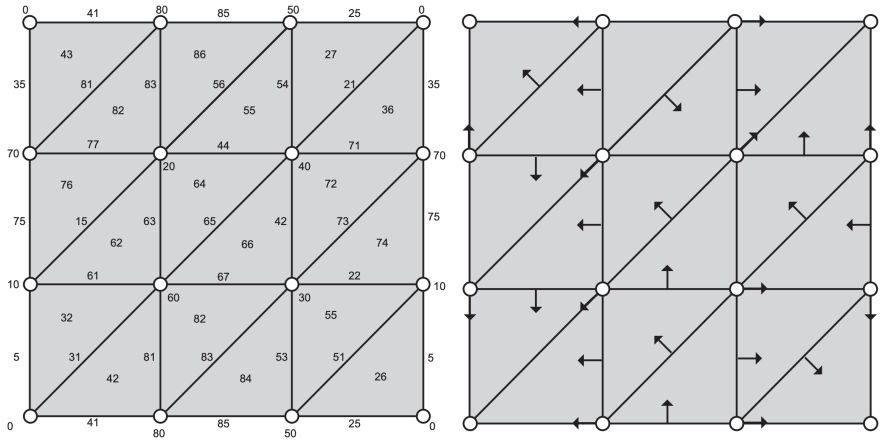


Figure 5: Building a discrete Morse function on a triangulated torus. On the left, we have a discrete Morse function and on the right, the associated gradient vector field. This figure is borrowed from [6].

to a critical cell. One can play around with the gradient vector field build on the 2-dimensional torus in Figure 5. But we also note that not all critical cells can be accessed from following a path. For example, the cell with value 21, critical for our Morse function built on S^2, is not the endpoint of any path.

♠

6 Parametric Morse Theory

In this section, we lay the foundations of *parametric Morse theory*, a theory developed by H. King, N. Mramor and K. Knudson in the paper [5]. In this paper they consider a method aiming to understand the evolution of critical cells in a parameter-varying network, e.g. a time parameter. More precisely, they look at a given sequence $\{f : K \to \mathbb{Z}\}_{f \in \mathcal{F}}$ of discrete Morse functions each defined on the same fixed simplicial complex K, and try to track the apparition and the persistence of critical cells. Once again, this has been done in the case of smooth Morse theory, from which one can build an intuition; moreover, the links between the two theories, especially in regard of parametric Morse theory, is a domain with many open questions. We refer to Section 7 of [5] for insights into related future work.

Remark 6.1. *Here, we consider the case where the simplicial complex decomposition X of the space we consider does not change trough time. The general case, in which the simplicial complex can change through time, is treated in Section 4 of [5]. This is done by introducing the notions of* cellulation refining *and* discrete vector field refining*.* ♠

Definition 6.2 (Finite Morse parametrization)**.** *A* finite Morse parametrization *of X is a sequence of the form $\{(f_i, V_i)\}_{i=1}^n$ where $\{f_i : K \to \mathbb{R}\}_{i=1}^n$ is a family of discrete Morse functions on K and for $i \in \{1, ..., n\}$, V_i is the gradient vector field of f_i.*

For any k, we now define the notions of *strong* and *forward connections* between two k-cells of X, each critical for some Morse functions f_i and f_j, respectively. For the remainder of the section, we consider a finite Morse parametrization $\{(f_i, V_i)\}_{i=1}^n$ of X. We denote by X^l the set of l-cells of X.

Definition 6.3 (Forward and strong connections)**.** *Let α_i and α_j be k-cells of X, critical for f_i and f_j respectively, for some $i \neq j \in \{1, ..., n\}$. We say there is a* forward connection *from α_i to α_j, provided the existence of a k-cell γ with a V_i-path $\{\alpha_i, ..., \gamma\} \subseteq X^{k-1} \cup X^k$ and a V_j-path $\{\gamma, ..., \alpha_j\} \subseteq X^k \cup X^{k+1}$. If there is a forward connection from α_i to α_j and vice-versa, we say the that there is a* strong connection *between the two cells. Two critical cells sharing a strong connection are said to be* strongly connected*.*

In the paper [5], we have the following definition.

Definition 6.4 (Birth and death)**.** *Let $\alpha \in K$ be a cell that is critical for V_i. We say α* is born *(resp.* dies*) at index i provided that no strong connection exists between α and any cell that is critical for V_{i-1} (resp. V_{i+1}). If there is cell $\beta \in K$ that is critical for V_{i+1}, such that the only strong connection of α and β is between themselves, then we say α* moves *(or* mutes*) to β.*

For Definitions 6.5 and 6.6, we consider G to be the graph whose nodes are the critical cells appearing along a finite Morse parametrization, and whose directed edges are exactly the strong connections, in the direction of increasing time value.

Definition 6.5 (Transcendence path)**.** *We say of a maximal path in G containing a critical cell α_i associated to some gradient vector field V_i that is it a* transcendence path of α_i*.*

Definition 6.6 (Parametric path)**.** *Let $\alpha_i \in K$ be a critical cell associated to gradient a vector field V_i. Let $\{P\}_{k=1}^m$ be the family of all transcendence*

paths of α_i. If all paths P_k start at a critical cell that is associated to the same fixed time slice, then we define the parametric path of α_i *to be (one of) the longest of the paths P_k. Else, we define the* parametric path of α_i *to be (one of) the oldest of the paths P_k, in the sense that it starts at a critical cell that is associated to the earliest time slice.*

Having defined a transcendence path for a given finite Morse parametrization, we prefer to consider the following equivalent definition for clarity purposes. This is also the approach we use in the code implementation. In this project, we only consider cases where critical cells belong to exactly one transcendence path.

Definition 6.7 (Birth and death coordinates)**.** *Let $\alpha \in K$ be a cell that is critical for V_i. Let P be the parametric path associated to α. Suppose P starts at a cell that is critical for gradient vector field V_j and ends at a cell that is critical for gradient vector field V_k. Then we define the birth and death coordinates of α to be given by $(j, k) \in \mathbb{R} \times \mathbb{R}$.*

Now, just as one can plot the bifurcation diagram of a family of smooth Morse functions $f : \mathbb{R} \times I \to I$, where I is the unit interval, *i.e.* we plot the evolution of the critical points in the plane $\mathbb{R} \times I$, one can consider what we call the *birth-death diagram* of a finite Morse parametrization of X. Thus, instead of continuous curves in the plane, we look at point sequences in the plane that trace the behavior of critical cells as the time parameter changes. Birth-death diagrams encode the connectivity of critical cells across the «time-line».

Example 6.8 (Birth-Death diagram)**.** *As an example, we consider a cellulation of the circle S^1, as it is rather simple space on which to generate various discrete vector fields. We take the same vector fields as in Figure* 4 *of [5] (see Figure 6 and 7). In Figure 8, the evolution of the time parameter is illustrated clockwise.*

Definition 6.9 (Parametric persistence diagram)**.** *We define the* parametric persistence diagram *of a finite Morse parametrization $\{(f_i, V_i)\}_{i=1}^n$ to be the subset of the space $\{(x, y) \in \mathbb{R}_+ \times \mathbb{R}_+ | y \geq x\}$ formed by the set of birth-death coordinates of all critical cells in the parametrization.*

Remark 6.10. *In a way, parametric persistence diagrams are analogous to persistence diagrams. The difference is that we do not consider the usual topological features but rather critical cells.* ♠

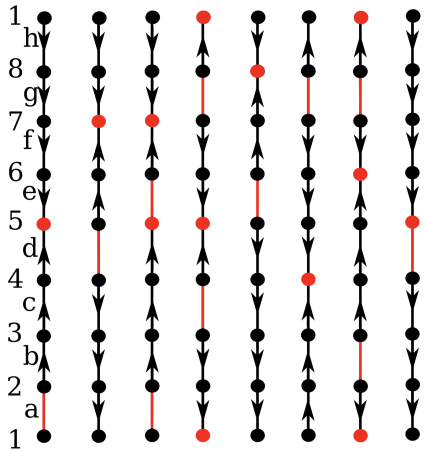


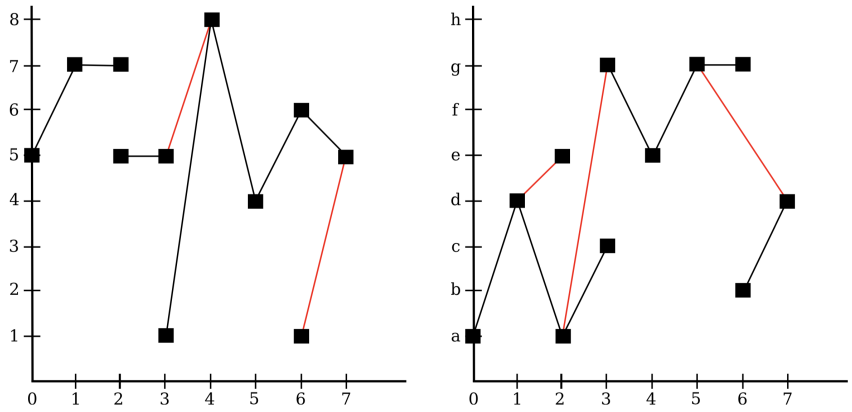
Figure 6: A Morse parametrization on S^1 for Example 6.8. The critical cells are indicated with red coloring. This figure is borrowed from [5].

Remark 6.11. *A parametric persistence diagram encodes less information than the corresponding birth-death diagram. Indeed, in a birth-death diagram, one can easily retrieve the life coordinates of critical cells, and can further observe the behavior of critical cells through connectivity curves.* ♠

Remark 6.12. *The passage from the setting of gradient vector fields to the drawing of birth-death diagrams relies, of course, on finding all possible forward and strong connections between critical cells. To this end, an algorithm was developed in [5], to which we refer for more complex examples. This algorithm is implemented in the code we provide on GitHub.* ♠

6.1 An Analogy

Let M be a smooth manifold. The notions of birth and death in smooth Morse theory are notions that naturally appear by considering a family



(A) 0-cells (B) 1-cells

Figure 7: The birth-death diagram for Example 6.8. The strong connections are indicated with black coloring. This figure is borrowed from [5].

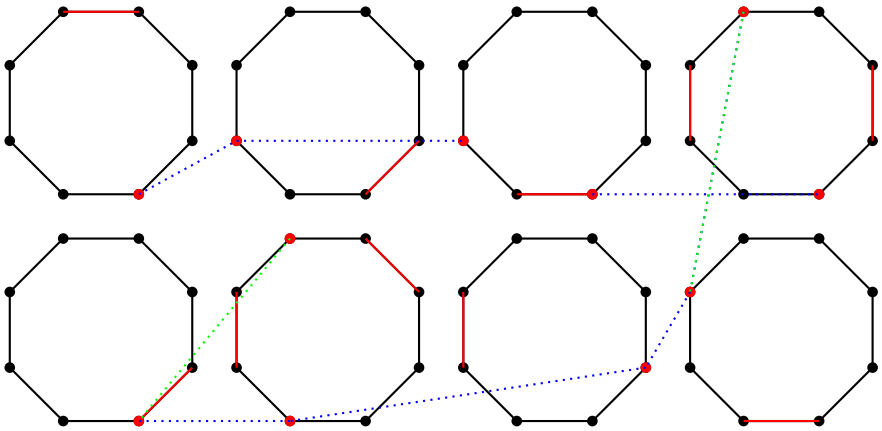


Figure 8: Continuation of Example 6.8 - this is a compact visualization for the case of 0-cells : blue edges represent strong connections, green edges represent simple forward connections. The diagram is read clockwise. Critical cells are emphasized in red color. One can imagine that we bent the natural timeline into a U-shaped curve.

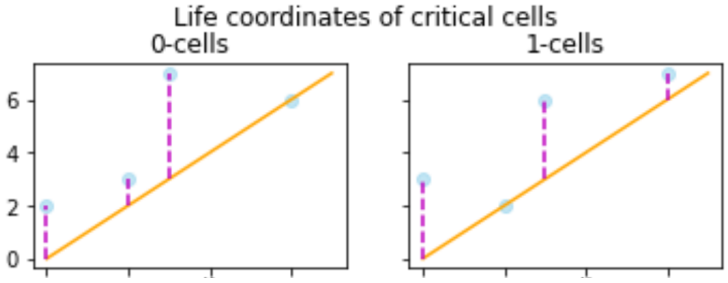


Figure 9: The parametric persistence diagram for Example 6.8.

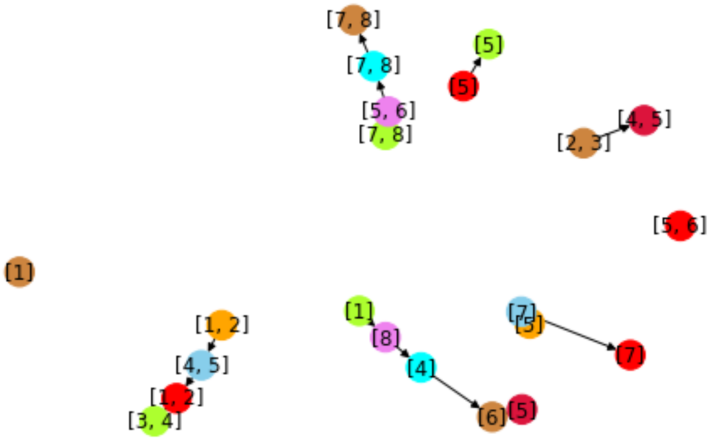


Figure 10: A visualization of connectivity information for Example 6.8. Each critical cell is colored according to its associated time-slice. There is a time slice where cell $[2, 3]$ mutes into cell $[4, 5]$. More precisely, this is a representation of the transcendence paths of the finite Morse parametrization.

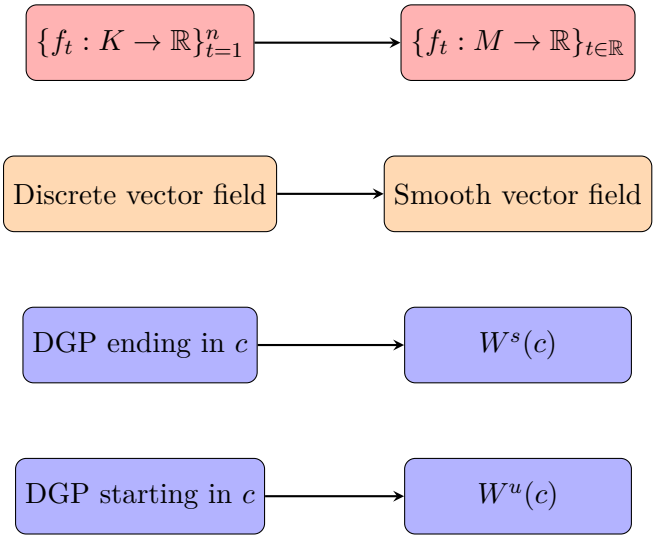


Figure 11: An analogy for the notions of birth and death of critical cells in the smooth case versus the discrete case. Here, «DGP» stands for «discrete gradient path».

of smooth Morse functions $\{f_t : M \to \mathbb{R}\}_{t\in\mathbb{R}}$ such that the function $f : M \times \mathbb{R} \to \mathbb{R}$:

$$f : \begin{cases} M \times \mathbb{R} \to \mathbb{R} \\ (x,t) \mapsto f_t(x) \end{cases}$$

is smooth. We suggest an analogy as described in Figure 11.

Idea. We observe the following. We define the notions of birth and death of critical cells in a finite Morse parametrization with the gradient path existence condition «There is a k-cell γ with a V_i-path $\{\alpha_i, ..., \gamma\} \subseteq X^{k-1} \cup X^k$ and a V_j-path $\{\gamma, ..., \alpha_j\} \subseteq X^k \cup X^{k+1}$.» This translates to the smooth Morse theory setting as a condition of the form «There is a point $x \in M$ such that $x \in W^u(c_i) \cup W^s(c_j)$.», where $c_i \in M$ (resp. $c_j \in M$) is the critical point corresponding to critical cell α_i (resp. α_j).

The smooth setting in which the study of time-related behavior of critical cells is the *Cerf theory*, to which we refer to for deeper insights into the subtleties of the smooth case.

7 On Building Discrete Morse Functions

As mentioned in the previous section, a great deal in discrete Morse theory is how to build or generate discrete Morse functions. In most real-life applications, we are given a set of point data, and thus we usually start working with simplicial complexes with vertex values. An important step, to apply the theory we have seen, is thus to extend a vertex-valued function to a discrete Morse function on the whole complex. Various algorithms based on the notions of discrete vector field and *Hasse diagram* exist, and we refer in particular to [6] for a detailed example of algorithm. Here we shall present and work with an implementation described in [**Saucan et al.2019**]).

7.1 Algorithm

In [6], the authors propose an implementation allowing one to extend any injective function defined on a 0-skeleton $h : K_0 \to \mathbb{R}$ to a discrete Morse function $\tilde{h} : K \to \mathbb{R}$ in a way that preserves the global behavior of h. Moreover, the algorithm involves a way of deleting pairs of critical cells, to a certain degree. More precisely, one can choose a *persistence level* $p \in \mathbb{R} \cup \{\infty\}$ so that pairs of critical cells that are connected by exactly one gradient path (with respect to a temporary discrete Morse function based on h) and whose difference of h-values is lower than p, are canceled. This process aims not only at canceling a maximum of critical cells (with $p = \infty$) so as to simplify homology computations, for example, but also at deleting topological features that are noise-related (with $p \in \mathbb{R}$) since those usually do not persist through time for too long.

Although the method in [6] is designed with a very performant way of minimizing the number of critical cells, we choose to run our experiments with another, more recent algorithm that is more practical and easier to implement. More precisely, we follow **Algorithm 1** of [4], whose pseudo-code is reported below. Moreover, we provide an implementation of the code written in Python, making use of the *Gudhi* library. For a given simplicial complex, the output function achieves close to the theoretical minimum of critical cells. In the code hereunder, the variable `Flag[]` encodes the numbers

$$n_\alpha := \sum_{\beta \in \mathcal{B}_\alpha^F} \mathbb{1}_{\{f(\beta) \geq f(\alpha)\}}$$

for each $\alpha \in K$, where we look at a simplicial complex K. Moreover, the input g denotes a function $K^{(0)} \to \mathbb{R}$ defined on the vertices of the complex.

Algorithm 1 Extending vertex labels to a discrete Morse function

procedure ExtendToMorse(K, g)
    $d \leftarrow \dim(K)$
    for $p = 0, ..., d$ **do**
        for $\alpha^{(p)} \in K$ **do**
            $n_\alpha \leftarrow 0$
    for $\alpha^{(0)} \in K$ **do**
        $f(\alpha) \leftarrow g(\alpha)$
    for $p = 1, ..., d$ **do**
        for $\alpha^{(p)} \in K$ **do**
            Let `Faces` be a list of all $(p-1)$-dimensional faces of α
            Sort `Faces` by decreasing order of f-values
            $\gamma_0 \leftarrow$ `Faces[0]`
            $\gamma_1 \leftarrow$ `Faces[1]`
            if `Flag`$[\gamma_0] = 0$ and $f(\gamma_0) > f(\gamma_1)$ **then**
                $f(\alpha) = (f(\gamma_0) + f(\gamma_1))/2$
                `Flag`$[\gamma_0] \leftarrow 1$
            else
                $\epsilon \leftarrow \text{UniformRandom}(0, 0.5)$
                $f(\alpha) \leftarrow f(\gamma_0) + \epsilon$
    return f

In the code, the function `getskel(K,p)` returns the set of p-cells of K. We choose to encode the label function f and the `Flag` variable as dictionaries `dict`, for practical implementation purposes. For example, if we have $f([1,2,3]) = 2.1$ for some triangle $[1,2,3] \in K$, then the label dictionary `f` will contain an entry of the form `"[1,2,3]":2.1`.

```
// Extending vertex labels to a discrete Morse function
import gudhi as gd

def build_morse_function(K,d,g):
    """
    K : simplicial complex (gudhi.SimplexTree)
    d : dimension of K
    g : labels on nodes given by a dictionary
    """
    Flag = {}
    f = {}

    for p in range(d):
        for simplex in list(K.get_skeleton(p)):
            Flag[str(simplex[0])]=0

    for simplex in list(K.get_skeleton(0)):
        f[str(simplex[0])] = g[str(simplex[0])]

    for p in range(1,d+1):

        for simplex in getskel(K,p):
            Faces = [face[0] for face in
                list(K.get_boundaries(simplex[0]))]
            Faces = [str(face) for face in Faces]
            Faces = sorted(Faces, key=lambda x: f[x], reverse=True)
            gamma_0 = Faces[0]
            gamma_1 = Faces[1]

            if (Flag[Faces[0]]==0 and f[gamma_0]>f[gamma_1]):
                f[str(simplex[0])]=(f[gamma_0]+f[gamma_1]) / 2
                Flag[gamma_0]=1

            else:
                epsilon = np.random.uniform(low=0.0, high=0.5)
                f[str(simplex[0])] = f[gamma_0] + epsilon

    return [f, Flag]
```

Simplex	function f_1	function f_2	`Flag`	`IsCritical`
[0]	3.1	3.1	1	`False`
[1]	2.1	2.1	1	`False`
[2]	1.1	1.1	0	`True`
[3]	3.2	3.2	1	`False`
[4]	3.3	3.3	1	`False`
[5]	3.4	3.4	1	`False`
[6]	4.1	4.1	1	`False`
[7]	0.1	0.1	0	`True`
[8]	4.3	4.3	1	`False`
[0,1]	2.60	2.60	0	`False`
[1,2]	1.60	1.60	0	`False`
[2,3]	2.15	2.15	0	`False`
[3,4]	3.25	3.45	0	`False`
[4,2]	3.45	2.2	1	`False`
[2,5]	2.25	2.25	0	`False`
[5,7]	3.48	3.52	0	`True`
[1,7]	2.35	2.37	0	`True`
[7,0]	3.15	3.57	1	`False`
[7,6]	2.10	2.10	0	`False`
[7,8]	2.20	2.20	0	`False`
[0,1,7]	2.875	3.086	0	`False`
[2,3,4]	3.350	2.825	0	`False`

Figure 12: Extending a vertex label to a discrete Morse function

Example. We reproduce the experiment of [4] (*cf.* Figure 1 of the paper), and compare the results. We note that we obtain the same final `Flag` values, as well as the same critical cells. However, note that we add a uniformly distributed random noise term in line 18 of the pseudo-code. Hence, the method contains a stochastic step and our output function thus slightly changes. We denote by f_1 and f_2 the discrete Morse functions reported in the experiment of [4] and with our code, respectively. The table in Figure 7.1 contains all the information on the input (K, g); indeed, K is formed of all simplices reported in the first column, and g is equal to the restriction of the functions f_1 and f_2 to the vertices.

7.2 The Triangle

We consider a simple example, by defining the simplicial complex K to be a triangle and setting $G = K^{(0)}$. The point of this subsection is that the injectivity of the initial node function has a major impact on the output number of critical cells. Indeed, we show with the examples below that an injective setting is best to produce a minimal number of critical cells. We run the extension algorithm (**Algorithm 1**) for non-injective and injective settings in Figures 13 and 14, respectively. In the non-injective setting, the algorithm will produce five critical cells (see Figure 13), while in the injective setting it will produce the minimal number of critical cells (see Figure 14).

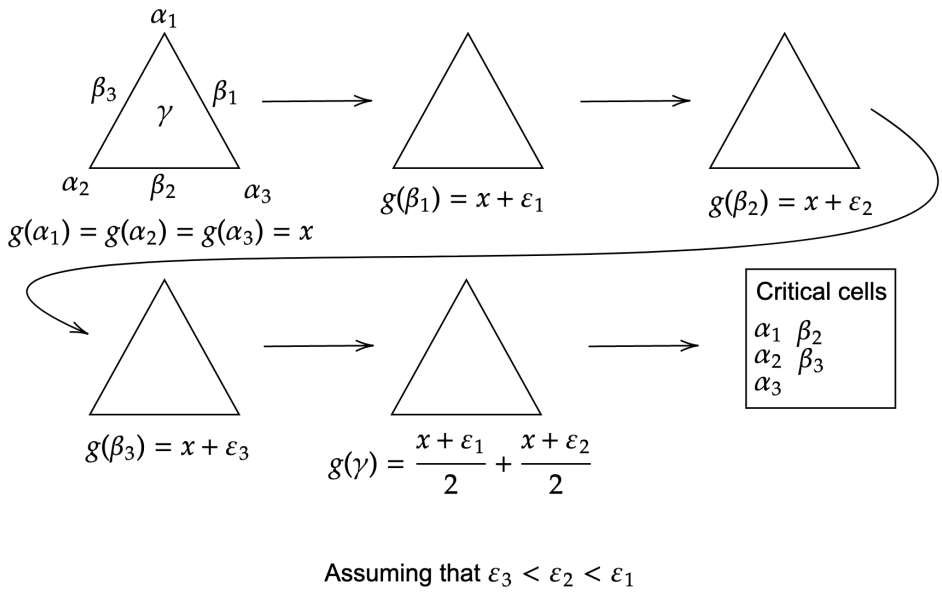


Figure 13: Extension algorithm with non-injective node function on the triangle.

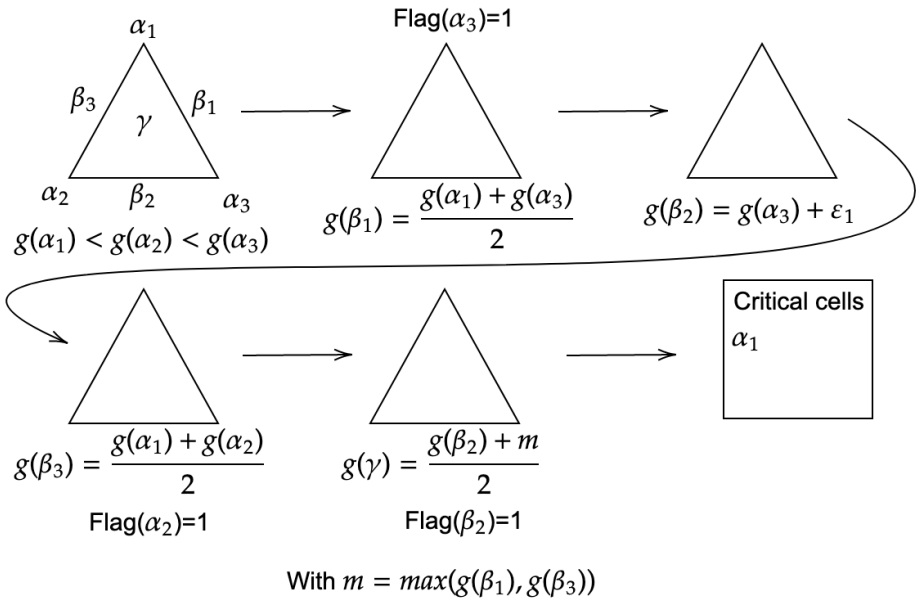


Figure 14: Extension algorithm with injective node function on the triangle.

8 An Application

8.1 Morse features of finite node-function sequences

In this subsection, we report the precise process we go through to analyze the discrete Morse features associated to a sequence of graphs with vertex values obtained from an initial input graph G.

Clique completion. We form a simplicial complex K obtained from the input graph G by clique completion.

Assigning vertex values. We form a sequence of pairs $\{(K, g_t)\}_{t=1}^{n}$ where each $g_t : K \to \mathbb{R}$ is a node labeling function, defined based on a parameter-dependent sampling process. More precisely, we choose an initial mapping

$$\mathbf{h} : \begin{cases} G \to \mathcal{C}^0(\mathbb{R}) \\ i \to h_i \end{cases},$$

where $\mathcal{C}^0(\mathbb{R})$ denotes the space of continuous functions $\mathbb{R} \to \mathbb{R}$. Then, we choose a set of parameters $(x_0, \delta, n) \in \mathbb{R} \times \mathbb{R} \times \mathbb{N}$ and define, for a given time

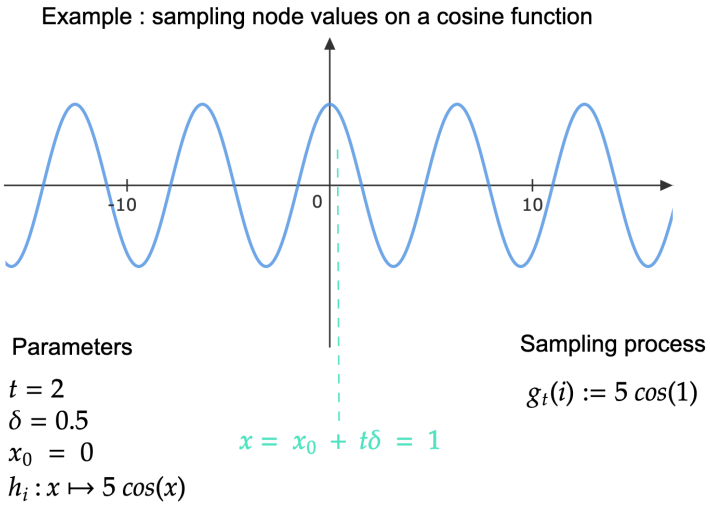


Figure 15: Sampling process illustrated for a node i and associated node function $h_i : x \mapsto 5\cos(x)$, with parameters $(x_0, \delta) = (0, 0.5)$ and time $t = 2$.

slice $t \in \{1, ..., n\}$, the function

$$g_t : \begin{cases} G \to \mathbb{R} \\ i \to g_t(i) := h_i(x_0 + t\delta) \end{cases} .$$

We illustrate this process in Figure 15.

Generating Morse functions. Now, we can obtain a family of discrete Morse functions $\{(K, f_t)\}_{t\in[T]}$ by extending each $g_t : K^{(0)} \to \mathbb{R}$ to a map $f_t : K \to \mathbb{R}$ via the algorithm described in the previous section. Finally, we are able to run the full parametric Morse theory analysis on $\{(K, f_t)\}_{t\in[T]}$, *i.e.* we report, at each time slice $t \in [T]$, the current critical cells and the gradient vector field of f_t. Moreover, we can plot the corresponding parametric persistence diagrams.

Plotting parametric persistence diagrams. Finally, we are able to plot the parametric persistence diagrams of the time-series $\{(K, f_t)\}_{t\in[T]}$. Those can be thought of, in some sense, as persistence diagrams for *all* critical cells

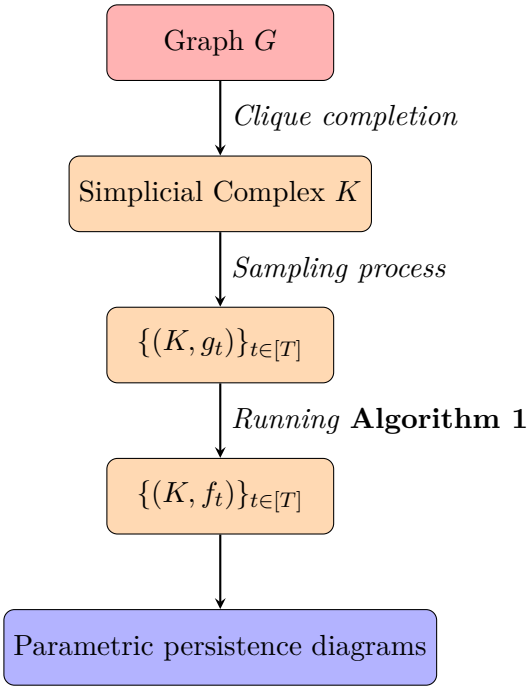


Figure 16: Main Pipeline. This diagram illustrates the process of obtaining parametric persistence diagrams starting from an input graph G.

of the complex K (*i.e.* with respect to all discrete Morse functions $(f_t)_t$).

Pipeline illustration. Figure 8.1 illustrates the process followed by the pipeline, starting from a graph G together with node functions as input and obtaining parametric persistence diagrams as output, through the steps described above.

8.2 Towards Robustness

When following such a method, one can naturally think of the pipeline as a mapping

$$F : \mathcal{G} \to 2^{\mathbb{R}\times\mathbb{R}},$$

where $\mathcal{G}$ is the set of all pairs of the form $(G, \mathbf{h})$ where G is a graph and $\mathbf{h} : G \to \mathcal{C}^0(\mathbb{R})$ is a mapping as described earlier. The mapping F sends the pair $(G, \mathbf{h})$ to its associated parametric persistence diagram. Following

this way of thinking, one may wonder about the robustness of the method, or analogously about the *stability* of F. More precisely, given a perturbation in the early steps of the method, we look at how much the final output is affected. There are many ways of perturbing the method in the early steps. For example, one could try to answer this question by running the same experiment on a graph G two times, where the stochastic map $\{(K, g_t)\}_{t\in[T]} \to \{(K, f_t)\}_{t\in[T]}$ is implemented with a significant amount of noise. This means we get two time-series $\{(K, f_t)\}_{t\in[T]}$ and $\{(K, \tilde{f}_t)\}_{t\in[T]}$. Instead, we choose another way of considering perturbations, which we describe below.

Perturbing the vertex values. Let us consider the node functions setting described in Section 8.1, *i.e.* we have, for an input graph G, a mapping

$$\mathbf{h} : \begin{cases} G \to \mathcal{C}^0(\mathbb{R}) \\ i \to h_i \end{cases} .$$

and we choose a set of parameters $(x_0, \delta, n) \in \mathbb{R} \times \mathbb{R} \times \mathbb{N}$ so as to define, for a fixed time slice $t \in \{1, ..., n\}$, the function

$$g_t : \begin{cases} G \to \mathbb{R} \\ i \to g_t(i) := h_i(x_0 + t\delta) \end{cases} .$$

Now, we choose a *perturbation term* $\varepsilon > 0$, and define, for a fixed time slice $t \in \{1, ..., n\}$, the *perturbed* node labeling function $\tilde{g}_t$ by setting

$$\tilde{g}_t : \begin{cases} G \to \mathbb{R} \\ i \to \tilde{g}_t(i) := h_i(x_0 + t\delta + \varepsilon) \end{cases} .$$

For the stability analysis, we choose to define a distance between the two sequences of node labeling functions $\mathbf{g} := \{g_t\}_{t=1}^n$ and $\tilde{\mathbf{g}} := \{\tilde{g}_t\}_{t=1}^n$ by setting

$$d(\mathbf{g}, \tilde{\mathbf{g}}) := \frac{1}{n} \sum_{t=1}^{n} ||g_t - \tilde{g}_t||_\infty,$$

where we set $||g_t - \tilde{g}_t||_\infty := \max_{i\in G} |g_t(i) - \tilde{g}_t(i)|$.

Let $K > 0$ be fixed and suppose that, for each $i \in G$, the associated continuous function $h_i \in \mathcal{C}^0(\mathbb{R})$ is L-Lipschitz. Then it follows that, for a perturbing

term ε, and sequences $\mathbf{g}$ and $\tilde{\mathbf{g}}$ as above, one has

$$\begin{aligned} d(\mathbf{g}, \tilde{\mathbf{g}}) &= \frac{1}{n} \sum_{t=1}^{n} ||g_t - \tilde{g}_t||_\infty && (1) \\ &= \frac{1}{n} \sum_{t=1}^{n} \max_{i \in G} |g_t(i) - \tilde{g}_t(i)| && (2) \\ &= \frac{1}{n} \sum_{t=1}^{n} \max_{i \in G} |h_i(x_0 + t\delta) - h_i(x_0 + t\delta + \varepsilon)| && (3) \\ &\leq \frac{1}{n} \sum_{t=1}^{n} L\varepsilon && (4) \\ &= L\varepsilon, && (5) \end{aligned}$$

and we are thus able to control the term $d(\mathbf{g}, \tilde{\mathbf{g}})$ For clarity purposes, we shall also denote the perturbed sequence $\tilde{\mathbf{g}}$ by $\mathbf{g}_\varepsilon$, given that the associated perturbation term is ε.

Evaluating the final impact. The effect of perturbing vertex values on the final output is then evaluated in terms of distances between the resulting parametric persistence diagrams. More precisely, we compute the Bottleneck and some p-Wasserstein distances (defined below) between parametric persistence diagrams obtained with and without perturbation.

Definition 8.1 (Bottleneck distance.)**.** *For two diagrams $D, \tilde{D} \in \mathbb{N} \times \mathbb{N}$, we define their Bottleneck distance as*

$$d_B(D, \tilde{D}) = \inf_{h: D \to \tilde{D}} \sup_{x \in D} ||x - h(x)||_\infty,$$

where the infimum is taken over all bijections $h : D \to \tilde{D}$. Note that a bijection is allowed here to send a point $(x, y) \in D$ to the diagonal of the $\tilde{D}$.

Definition 8.2 (p-Wasserstein distance.)**.** *For two diagrams $D, \tilde{D} \in \mathbb{R} \times \mathbb{R}$ of same cardinality, we define their p-Wasserstein distance as*

$$d_p(D, \tilde{D}) = \{\inf_{h: D \to \tilde{D}} \sup_{x \in D} ||x - h(x)||_\infty^p\}^{\frac{1}{p}}$$

where the infimum is taken over all bijections $h : D \to \tilde{D}$. Note that a bijection is allowed here to send a point $(x, y) \in D$ to the diagonal of the $\tilde{D}$.

Remark 8.3. *Note that the Wasserstein distances provide a way to generalize the Bottleneck distance. Indeed, the Bottleneck distance is equivalent to the 1-Wasserstein distance, i.e. $d_1(\cdot, \cdot) = d_B(\cdot, \cdot)$.* ♠

We may consider a distance between two discrete Morse parametrizations as a kind of mean quadratic error summed over all time slices. However, in the results section, we only consider a perturbation on the vertex values and not directly on the function $f : K \to \mathbb{R}$. The danger with the second method is that we may obtain non-Morse functions by perturbing a discrete Morse function.

Definition 8.4 (Parametric distance.)**.** *We define a distance d between two time-series $\{(K, f_t)\}_{t\in[T]}$ and $\{(K, \tilde{f}_t)\}_{t\in[T]}$ as above by setting*

$$d\left(\{(K, f_t)\}_{t\in[T]}, \{(K, \tilde{f}_t)\}_{t\in[T]}\right) = \frac{1}{|G|}\frac{1}{T}\sum_{i\in G}\sum_{t\in T}(f_t(i)) - \tilde{f}_t(i))^2.$$

The test of stability is illustrated in Figure 8.2 below. The set $\mathcal{D} = (D_k)_k$ (resp. $\tilde{\mathcal{D}} = (\tilde{D})_k$) contains parametric persistence diagrams obtained from the time-series $\{(K, f_t)\}_{t\in[T]}$ (resp. $\{(K, \tilde{f}_t)\}_{t\in[T]}$), where D_k (resp. $\tilde{D}_k$) consists of k-dimensional cells exclusively. Thus, we compute the distances between parametric persistence diagrams of k-critical cells, for some $k \in \{0, ..., \dim(K)\}$.

Remark 8.5. *Let D_k^ε denote the diagram $\tilde{D}_k$ obtained by running the pipeline with perturbation term ε. For the continuity analysis to make sense, we have to translate the birth and death coordinates of a perturbed diagram by the perturbation term used. More precisely, if we run the pipeline with perturbation term ε, and obtain to sets of diagrams $(D_k)_{k=1}^n$ and $(D_k^\varepsilon)_{k=1}^n$, then the births and deaths of a diagram D_k^ε will live in time-slices $t = 1 + \varepsilon, ..., n + \varepsilon$.* ♠

Complete Analysis. Our analysis further consists in computing and plotting a discrete curve

$$(\varepsilon, d(\mathbf{g}, \mathbf{g}_\varepsilon))_{\varepsilon\in E},$$

for some finite set $E \subset \mathbb{R}^+$, to observe how the node labeling functions react to perturbations. We compute and plot the curves

$$(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon\in E}$$

for $k = 1, ..., p$ where p is the maximum dimension of critical cells.

9 Results

In this section, we report the results of the stability analysis described in Section 8.2. We run the tests on two graphs. We consider the example of

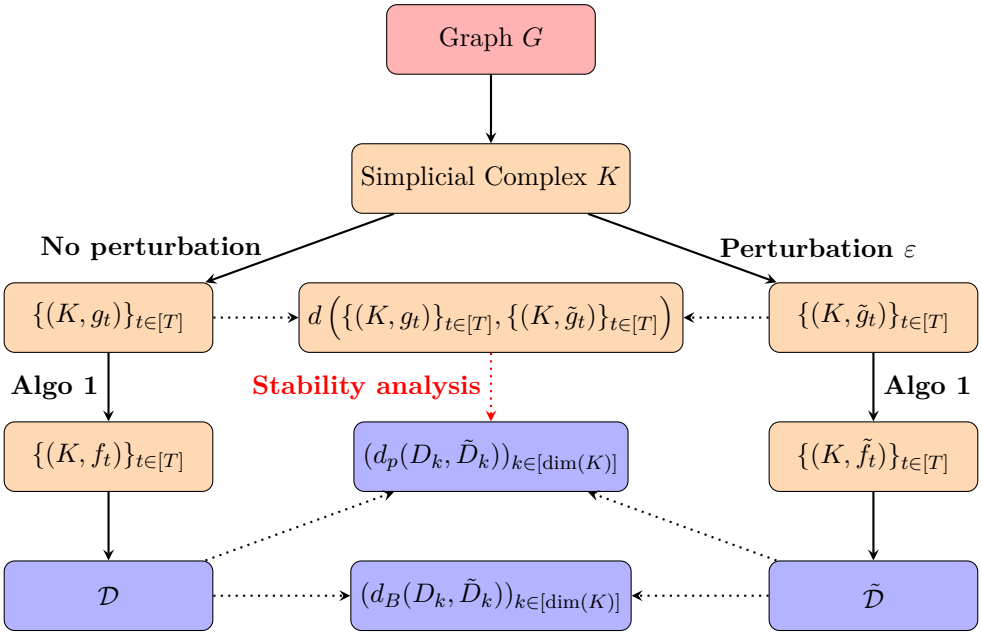


Figure 17: Testing stability process. This diagram illustrates the concept of the stability analysis : it consists of comparing the output of the pipeline when running it with and without a perturbation term.

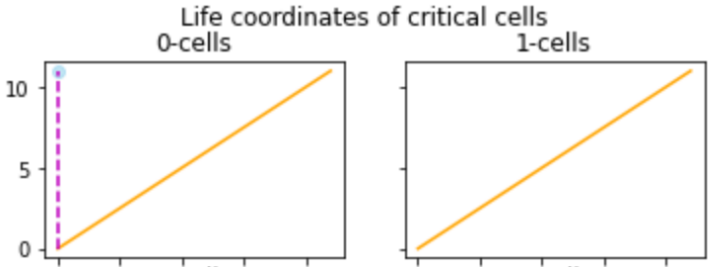


Figure 18: Parametric persistence diagram obtained with a triangle as input and cosine and sine wave functions on each vertex. Each parametric persistence diagram corresponds to a specific dimension of critical cells, as indicated.

the circle, and the sphere S^2 triangulated as a tetrahedron, with an initial mapping $\mathbf{h}$ that is injective.

Remark 9.1. *Generally speaking, we run experiments with a graph G and an* injective *mapping $\boldsymbol{h} : G \to \mathcal{C}^0(\mathbb{R})$. This enables the extension algorithm to produce close the theoretical number of critical cells (see Subsection 7.2).* ♠

9.1 The Triangle

We start with a very simple example, by considering our simplicial complex K to be a 2-simplex and set $G = K^{(0)}$. We choose an injective mapping $\mathbf{h}$ on the nodes, determined by $h_1 = \cos(2\cdot)$, $h_2 = \sin(3\cdot)$ and $h_3 = \cos(x)$. We obtain the parametric persistence diagram shown in Figure 18. We run the pipeline with no perturbation term, by considering a single time sequence. We choose the initial parameters so that we define

$$g_t : \begin{cases} G \to \mathbb{R} \\ i \to g_t(i) := h_i(5t) \end{cases} ,$$

where $t \in \{0, ..., 7\}$. We have a sequence of 8 node labeling functions $g_t : G \to \mathbb{R}$.

Stability Analysis. Once again, we investigate the stability analysis described at the end of Section 8. We plot the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for

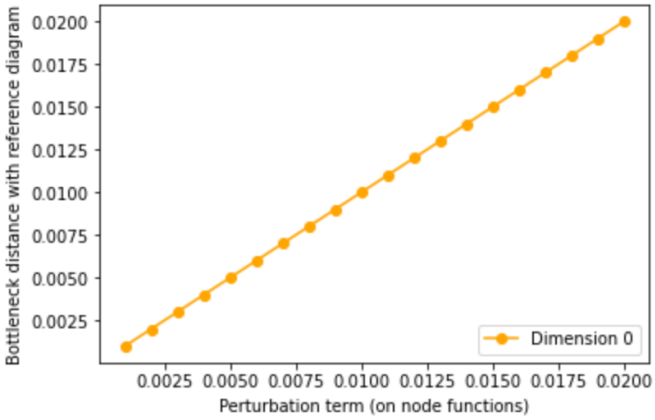


Figure 19: This figure represents the plotting of the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for the triangle with sine and cosine waves on each vertex.

each dimension $k \leq \dim(K)$, where d_B is the Bottleneck distance. The resulting plot is presented in Figure 21. The experiment produces the diagram shown in Figure 19. The procedure is deterministic and we obtain an identity curve. In fact, the perturbation of the initial node function does not affect at all the output parametric persistence diagram here. The only important factor in the experiment of the triangle is the injectivity of the initial node function. Apart from that, due to the symmetry of the triangle, the repartition of values on each node does not affect at all the resulting parametric persistence diagrams.

9.2 The Tetrahedron

In this subsection, we consider our simplicial complex K to be a tetrahedron. We set $G = K^{(0)}$, and choose an associated constant mapping

$$\mathbf{h} : \begin{cases} G \to \mathcal{C}^0(\mathbb{R}) \\ i \to h \end{cases},$$

determined by setting $h_1(\cdot) = \cos(2\cdot)$, $h_2(\cdot) = \sin(3\cdot)$, $h_3(\cdot) = \cos(\cdot)$ and $h_4(\cdot) = \sin^2(\cdot)$. We choose the initial parameters so that we define

$$g_t : \begin{cases} G \to \mathbb{R} \\ i \to g_t(i) := h(\frac{5}{2}t) \end{cases},$$

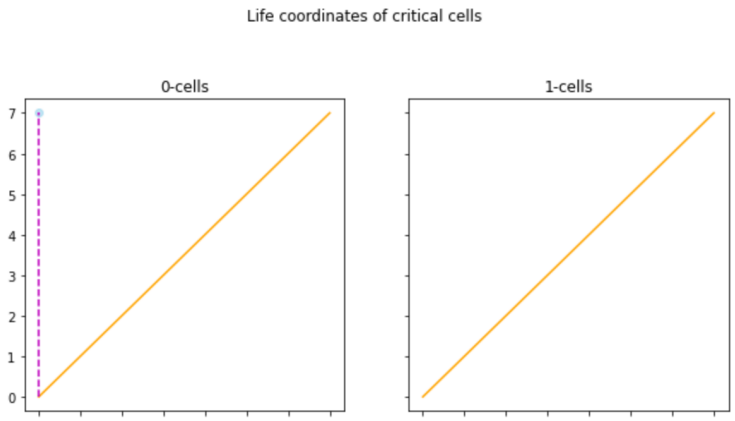


Figure 20: Parametric persistence diagrams obtained with a tetrahedron as input and **injective** sampling process. Each parametric persistence diagram corresponds to a specific dimension of critical cells, as indicated.

where $t \in \{0, ..., 11\}$. We have a sequence of 8 node labeling functions $g_t : G \to \mathbb{R}$. We run the experiment and obtain the parametric diagrams shown in Figure 20.

Interpretation. With a fixed noise term of $\varepsilon = 0.05$ in the stochastic step of the extension algorithm (**Algorithm 1**), the output diagram is always the same. This is due to the symmetry properties of the tetrahedron. Indeed, we are facing the same phenomenon as in the case of the triangle. More precisely, the initial node values do not affect at all the output diagrams.

Stability Analysis. We consider the robustness analysis described at the end of Section 8. Again, we plot the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for each dimension $k \leq \dim(K)$, where d_B is the Bottleneck distance. The corresponding plots are presented in Figure 21. As in the example of the triangle, the symmetry properties of the tetrahedron imply that we get an identity curve in the plot of $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$. Here, we set with $E = \{0.01m\}_{m=0}^{2}00$.

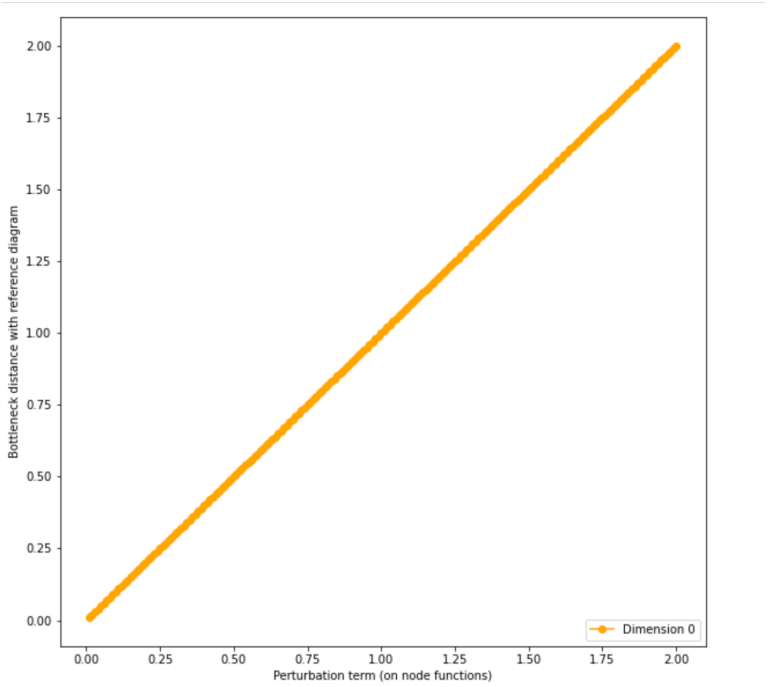


Figure 21: This figure represents the plotting of the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for the tetrahedron with injective sampling process.

9.3 The Cycle C_8

9.3.1 Different periodicity

For this experiment, we choose our first non-trivial example. To this end, we choose $K = C_8$, the cycle on eight vertices. We set an injective sampling process, but since we don't have the symmetry properties of the triangle or the tetrahedron, the output will be non-trivial. We set $G = K^{(0)}$, and choose an associated constant mapping

$$\mathbf{h} : \begin{cases} G \to \mathcal{C}^0(\mathbb{R}) \\ i \to h \end{cases} ,$$

determined by setting $h_1(\cdot) = \cos(2\cdot)$, $h_2(\cdot) = \sin(3\cdot)$, $h_3(\cdot) = \cos(\cdot)$, $h_4(\cdot) = \sin^2(\cdot)$, $h_5(\cdot) = \sin(\cdot)$, $h_6(\cdot) = \sin(2\cdot)$, $h_7(\cdot) = \sin(2\cdot) + \cos(2\cdot)$ and $h_8(\cdot) = \sin(2\cdot) + \cos(3\cdot)$. Note that not all functions have the same periodicity. We choose the initial parameters so that we define

$$g_t : \begin{cases} G \to \mathbb{R} \\ i \to g_t(i) := h(\frac{5}{2}t) \end{cases} ,$$

where $t \in \{0, ..., 11\}$. The node functions are illustrated in Figure 22.

Stability Analysis. For the stability analysis, we plot the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ with $E = \{0.01m\}_{m=0}^{2}00$. The result is reported in Figure 23. The periodicity of the labeling functions can be interpreted from the plot.

9.3.2 Same periodicity

We consider the exact same experiment as above, except from the fact that the mapping

$$\mathbf{h} : \begin{cases} G \to \mathcal{C}^0(\mathbb{R}) \\ i \to h \end{cases}$$

is determined by $h_i(\cdot) = i\cos(\cdot)$, so that we reproduce the injectivity of the sampling process, but all functions have the same periodicity. The node functions are illustrated in Figure 24.

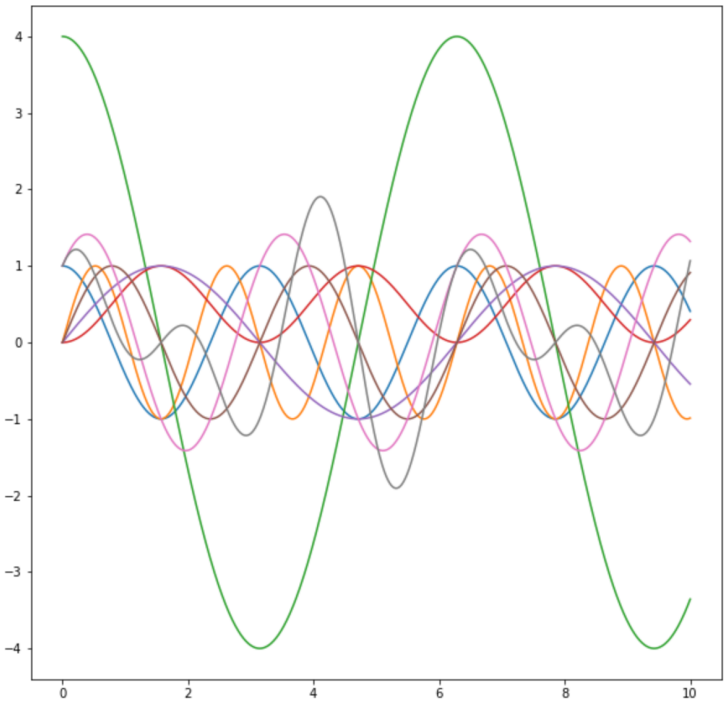


Figure 22: Plotting the node functions $h_1(\cdot) = \cos(2\cdot)$, $h_2(\cdot) = \sin(3\cdot)$, $h_3(\cdot) = \cos(\cdot)$, $h_4(\cdot) = \sin^2(\cdot)$, $h_5(\cdot) = \sin(\cdot)$, $h_6(\cdot) = \sin(2\cdot)$, $h_7(\cdot) = \sin(2\cdot) + \cos(2\cdot)$ and $h_8(\cdot) = \sin(2\cdot) + \cos(3\cdot)$.

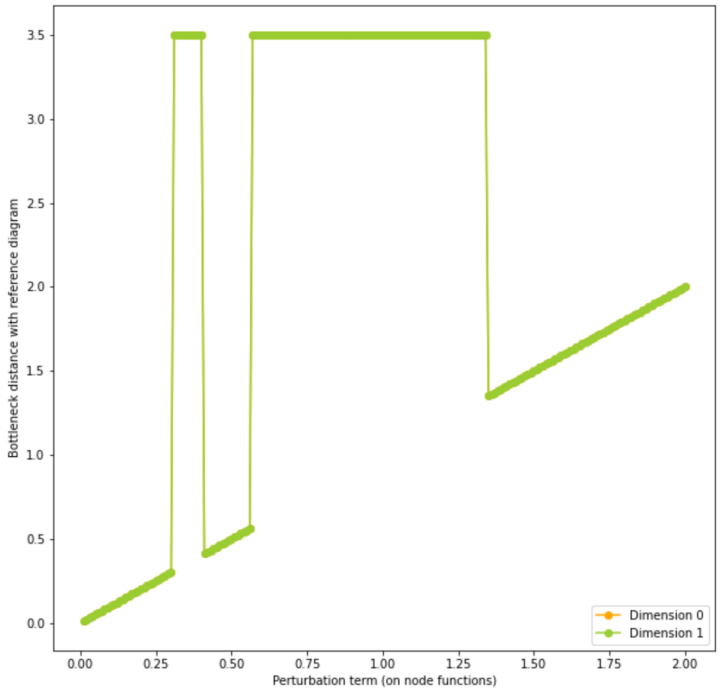


Figure 23: This figure represents the plotting of the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for the cycle C_8 with injective sampling process. The functions h_i do not all have the same periodicity. Here, the yellow and green curves coincide.

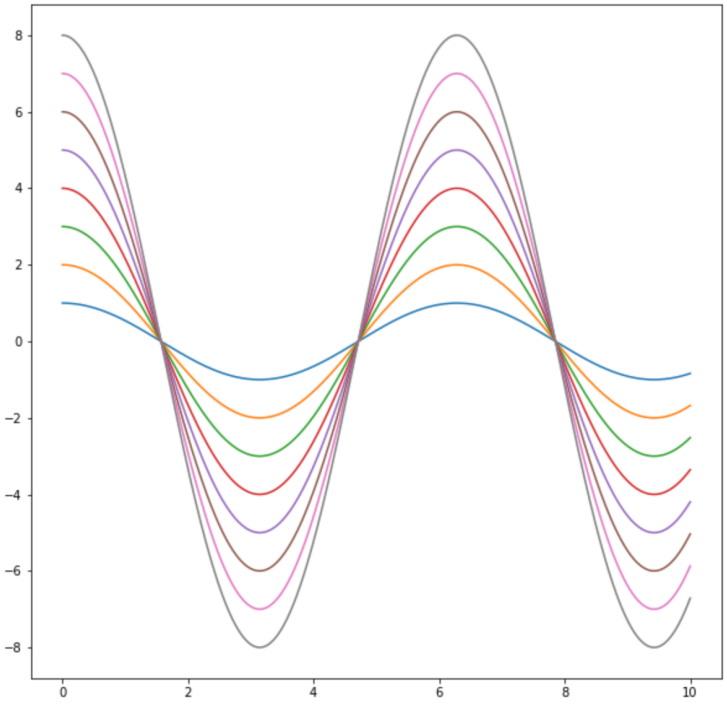


Figure 24: Plotting the node functions $h_i(\cdot) = i\cos(\cdot)$.

Stability analysis. Results are reported in Figure 25. The point of this example is the following. With a complex that does not have symmetry properties like the triangle or the tetrahedron, the main factor affecting the output is the sign of the first derivatives of the node functions. More precisely, consider a diagram D_1 obtained by running the pipeline once, and a diagram D_2 obtained by running the pipeline on the same example but with a perturbation term ε. Let $\{g_t\}_t$ and $\{\tilde{g}_t\}_t$ denote the node labelings of the first and the second experiment, respectively. Now, consider a node $i \in G$ and suppose the node functions are differentiable. A condition that can make a difference between D_1 and D_2 is, for some $t = 1, ..., T-1$,

$$g_t(i) < g_{t+1}(i) \text{ and } \tilde{g}_t(i) > \tilde{g}_{t+1}(i).$$

Indeed, this can drastically change the way the extension algorithm. In this experiment, the condition above never happens since the periodicity is the same for all functions h_i.

9.4 The Stochastic Block Model

One of the graph classes that we use for the computations is the stochastic block model, having for main characteristic that it naturally forms a partitioning of its vertices into disjoint clusters, also called «communities».

Definition 9.2. *The* stochastic block model $\text{Block}(n, k, \{C_a\}_{a=1}^k, \{p_{a,b}\}_{a,b=1}^k)$ *is a graph of n vertices built on k disjoint communities (or clusters), defined as follows. We consider all nodes, partitioned into k communities $\{C_a\}_{a=1}^k$, and draw an edge between $i \in C_a$ and $j \in C_b$ with probability $p_{a,b}$.*

We run the pipeline on a sotchastic block model G generated on a total of $n = 24$ nodes, for four clusters C_1 to C_4 of cardinality $\frac{n}{4} = 6$ with intra-cluster probabilities of 0.05 and inter-clusters probabilities of 0.8. The input graph is shown below, in Figure 27.

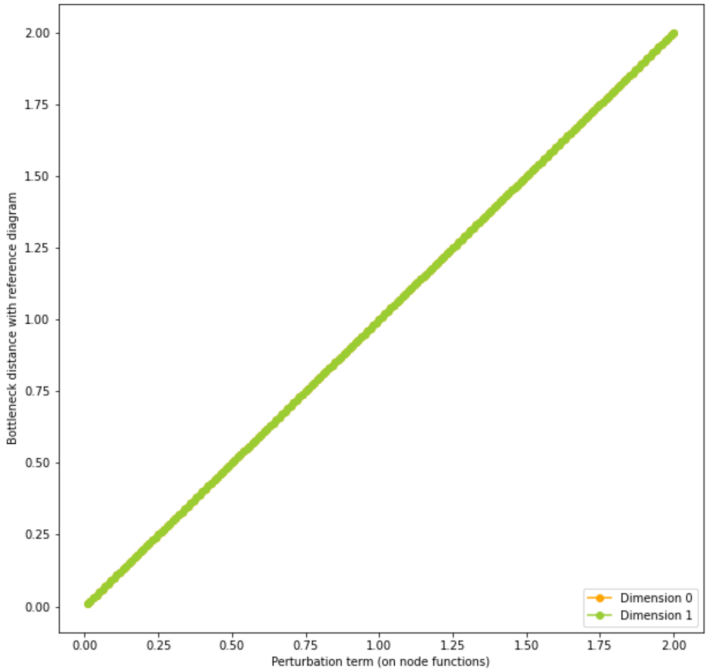


Figure 25: This figure represents the plotting of the curves $(\varepsilon, d_B(D_k, D_k^\varepsilon))_{\varepsilon \in E}$ for the cycle C_8 with injective sampling process. The functions h_i all have the same periodicity.

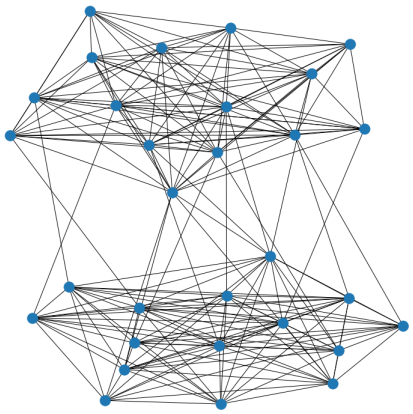


Figure 26: A stochastic block model generated on a total of $n = 30$ nodes, for two clusters C_1 and C_2 with intra-cluster probabilities $p_{1,1} = p_{2,2} = 0.7$ and inter-clusters probability $p_{1,2} = 0.05$. We set $|C_1| = |C_2| = \frac{n}{2} = 15$.

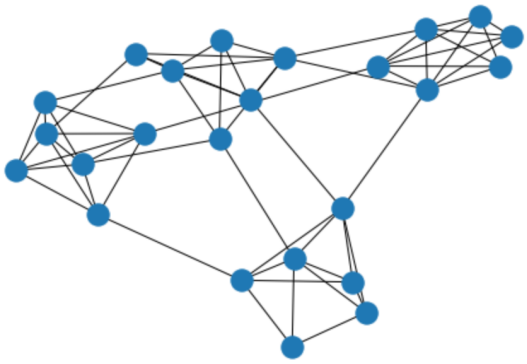


Figure 27: A stochastic block model generated on a total of $n = 24$ nodes, for four clusters C_1 to C_4 with intra-cluster probabilities of 0.05 and inter-clusters probabilities of 0.8. We set $|C_1| = |C_2| = |C_3| = |C_4| = \frac{n}{4} = 6$.

The final output of the pipeline are the parametric persistence diagrams (up to dimension $k = 3$) of the critical cells obtained with the finite discrete Morse parametrization built from G. Once again, we consider sine and cosine waves as functions assigned to each node. We run two experiments and report the results. In the first experiment, we consider a sampling process such that in each cluster C_k has a similar mapping $\mathbf{h} : C_k \to \mathbb{R}$ that is injective. In the second experiment, the sampling process is done in a way that each cluster C_k has a constant mapping, and this mapping is different for all clusters. Parametric persistence diagrams for the first and second experiments are reported in figures 28 and 29. We observe a similar distribution of life coordinates between the two sets of diagrams. In this case, it thus seems the shape of the graph has more importance than the actual node values for the final output.

10 Conclusion and Future Work

The content of this work aims, in particular, at developing ways to analyze the evolution of critical cells appearing along a sequence of real-valued vertex labelings $\{g : K^{(0)} \to \mathbb{R}\}_{t=1}^{n}$ extended to discrete Morse functions

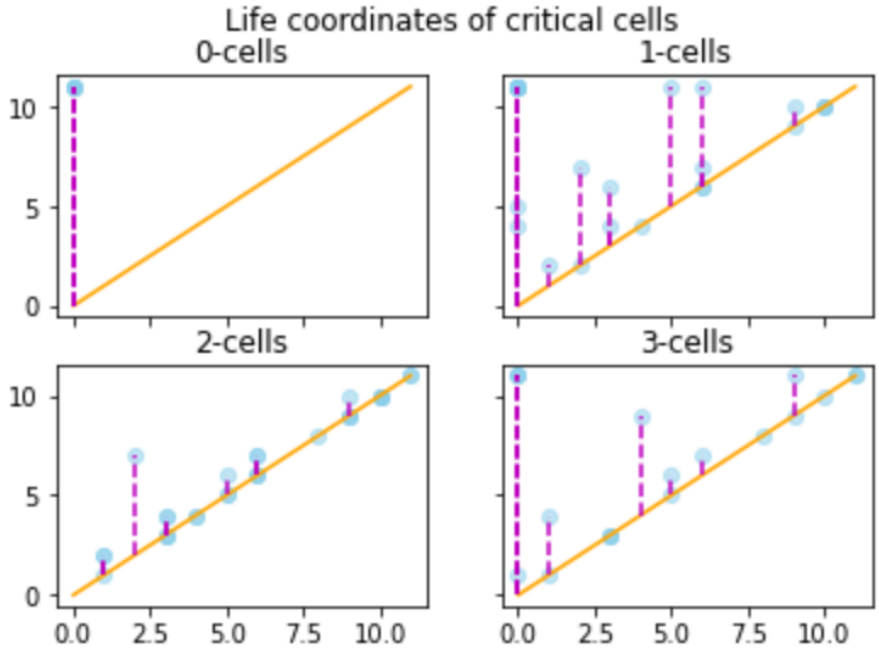


Figure 28: Parametric persistence diagrams for the stochastic block model G with four clusters. Sampling process is the same for each cluster and injective within each single cluster.

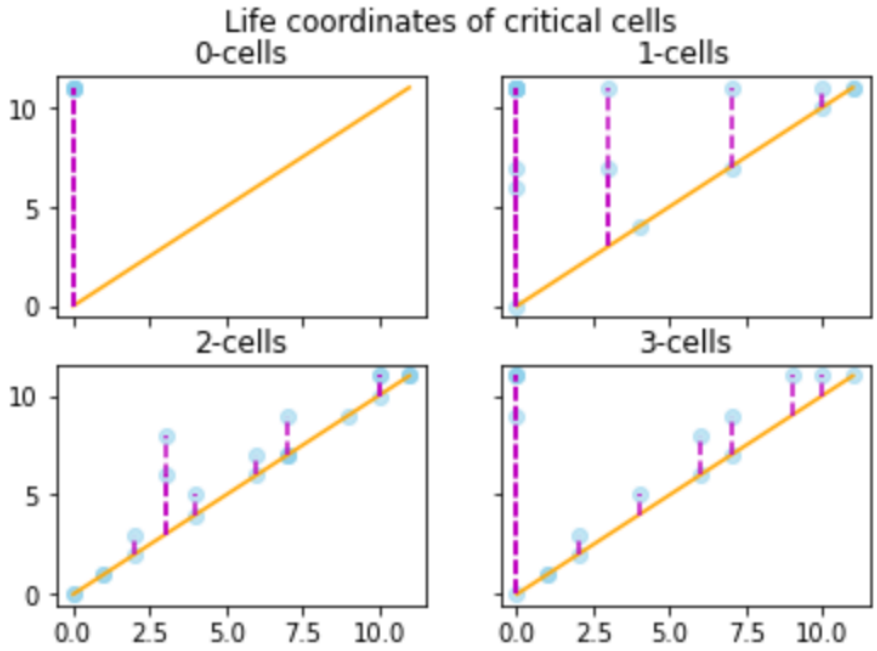


Figure 29: Parametric persistence diagrams for the stochastic block model G with four clusters. Sampling process is different for each cluster. Each single cluster is assigned a constant mapping.

$\{f : K \to \mathbb{R}\}_{t=1}^{n}$ on a simplicial complex K. We know that, when attaining the minimum number of critical cells for a given pair (K, f_t) where $f_t : K \to \mathbb{R}$ is a discrete Morse function, we have an elegant interpretation of what the term «critical» means. Indeed, we recall that 0-cells, 1-cells and maximum-dimensional cells of Y correspond to local minima, saddle points and local maxima respectively. Thus, parametric discrete Morse theory also provides tools to study the evolution (through time, perhaps) of fundamental geometric properties.

Extension algorithms. An interesting direction to look at in the future is the one of providing extension algorithms in higher dimensional labelings. More precisely, let $0 \leq d \leq \dim(K)$. One could provide a method to extend any real-valued labeling function $g : K^{(d)} \to \mathbb{R}$ to a discrete Morse function $f : K \to \mathbb{R}$. For $d = 2$, this enables the application of discrete Morse theory-based analyzes where the input data is a graph G that was assigned edge weights.

Spectral theory-based node labelings. An idea could be to consider initial node labelings according to a spectral theory principle, where a node i in time slice t would take the value of the i-component of the t-th eigenvector of the Laplacian of the graph given by $K^{(0)}$. The spectral theory setting might help understand the behavior of connected and almost disconnected components through time. If starting on a real-valued labeling sequence $\{g : K^{(d)} \to \mathbb{R}\}_{t=1}^{n}$ for some $1 \leq d \leq \dim(K)$, and using an extension algorithm as mentioned above, then we may consider higher-order Laplacians defined for simplicial complexes. We refer to [7] for more details. This would enable the understanding of connected and almost connected components through the study of not only 0-critical cells but also higher-dimensional critical cells.

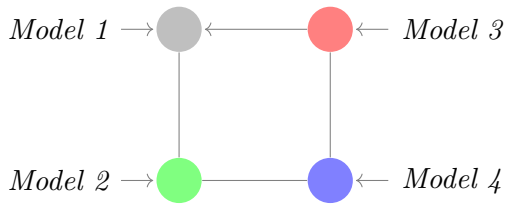
Smooth Morse theory. It seems like there are different questions to work on in the context of understanding the links between the smooth and the discrete case. An interesting direction could be to translate the concept of *Cerf theory* in the discrete case. This would be related to the notions of birth and death as studied in this project.

Birth-death diagrams. In this project, we consider parametric persistence diagrams rather than birth-death diagrams as we deal with Wasserstein and Bottleneck distances. However, birth-death diagrams encode more information than parametric persistence diagrams do, and thus it would be

interesting to define a notion of distance between birth-death diagrams.

Biological Modeling of Neural Networks. We now give some perspective to our approach by considering a possible application in the context of biological modeling of neural networks (for the purpose of studying neuron populations).

Example 10.1 (Neuron Populations)**.** *One could think of the input graph G as a neural network representing the generalized firing behavior of certain brain regions. More precisely, each node represents a specific neuron population (or a brain area) that share a biological feature, such as their firing model. By considering four of those populations, we are looking at the graph below, where* ***Model 1*** *could be, for example, the leaky integrate-and-fire model (LIF) and thus the function associated to the node of* ***Model 1*** *would be a LIF curve.*



Example 10.2 (The Stochastic Block Model for Neuron Firing Models)**.** *In the context of biological modeling of neural networks, one can extend the previous example by replacing every node (i.e. every neuron population) by a cluster of neurons. To this end, one can take as input graph G a stochastic block model where nodes within the same community share similar firing models.*

References

[1] Gunnar Carlsson. "Topology and Data". In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 255–308. DOI: 10.1090/S0273-0979-09-01249-X.

[2] Paweł Dłotko and Hubert Wagner. *Computing homology and persistent homology using iterated Morse decomposition*. 2012. arXiv: 1210.1429 [math.AT].

[3] Robin Forman. "A User's Guide to Discrete Morse Theory". In: *Séminaire Lotharingien de Combinatoire* 48 (2001). URL: https://www.emis.de/journals/SLC/wpapers/s48forman.pdf.

[4] Harish Kannan et al. "Persistent homology of unweighted complex networks via discrete Morse theory". In: *Scientific Reports* 9.1 (Sept. 2019). ISSN: 2045-2322. DOI: 10.1038/s41598-019-50202-3. URL: http://dx.doi.org/10.1038/s41598-019-50202-3.

[5] Henry King, Kevin Knudson, and Neza Mramor. *Birth and death in discrete Morse theory*. 2016. arXiv: 0808.0051 [math.AT].

[6] Henry King, Kevin Knudson, and Neza Mramor. "Generating Discrete Morse Functions from Point Data". In: *Experimental Mathematics* 14.4 (2005), pp. 435–444. URL: https://projecteuclid.org/journals/experimental-mathematics/volume-14/issue-4/Generating-Discrete-Morse-Functions-from-Point-Data/em/1136926974.full.

[7] Abubakr Muhammad and Magnus Egerstedt. "Control Using Higher Order Laplacians in Network Topologies". In: *Proc. of 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto*. 2006, pp. 1024–1038.