# Graph persistence

Mattia G. Bergomi[1], Massimo Ferri[2], Lorenzo Zuffi[2]

[1] Champalimaud Center for the Unknown, Lisbon, Portugal
[2] ARCES and Dept. of Mathematics, Univ. of Bologna, Italy
mattia.bergomi@neuro.fchampalimaud.org, massimo.ferri@unibo.it,
lorenzo.zuffi@studio.unibo.it,

**Abstract.** We set up a framework for an abstract definition of persistence functions. We then study some constructions of simplicial complexes from graphs and define persistence functions on them, out of classical persistent homology. Finally, we give two examples of persistence functions on weighted graphs, not based on algebraic topology.

**Keywords:** Persistence function, coherent covering, clique, independent set, neighborhood, enclaveless set, block, edge-block, hierarchical recurrent networks.

## 1 Introduction

Is persistence just persistent homology? Indeed, from the very beginning it has been so: A suitable choice of topological spaces and of continuous filtering functions made it possible to convey a more natural concept of shape in pattern recognition;"size functions" (which would eventually become "persistent 0-Betti numbers") were the tool for the first applications [16,15,14]. Not much later, persistent homology was conceived for deducing the true topology of sampled objects [35,13]. A noticeable exception is the construction of persistence diagram for the analysis of faulty programs [36]. As will be seen in the next subsection, persistence has also been used in graph theory, but always through the construction of simplicial complexes and their homology.

The aim of this paper is to show that the persistence paradigm — comparing pairs of sublevel sets of a suitable filtering function — can be implemented in other mathematical contexts. It actually applies to any concrete category in which structures, typical of that category, can be counted and respect certain requirements (Def. 1). We are aware of the recent categorical generalizations of persistence [3,26,11], but they seem to be rather far from the agile tool for applications we want to make available to the scientific community. In particular, we think that emergent research fields in which weighted graphs are essential tools — e.g. complex social networks and neural nets [12,25] — deserve a dedicated persistence theory, not necessarily mediated by either complexes or homology.

The article has three technical sections. The core of Sect. 2 is the definition of *persistence function* (Def. 1) and of *coherent covering* (Def. 2) in a general concrete category. The section also provides the definition of *natural pseudodistance*, a measure of dissimilarity (Def. 7), the definition of *persistence diagram* (Def. 5) in this wider context, and the fact that the bottleneck distance is a lower bound for the natural pseudodistance (Thm. 1). The persistence diagrams of faulty programs are also discussed as a particular case. Section 3 analyzes some constructions of simplicial complexes from a graph in the framework of the previous section: Simplicial complexes of cliques, of independent sets, of neighborhoods and of enclaveless sets. Section 4 gives two examples of persistence functions, on a weighted graph, which don't rely (at least not directly) on simplicial complexes and homology. They can then be considered the first two examples of graph persistence which don't come from algebraic topology.

### 1.1 State of art

Graphs and persistence are bound together since the early days, long before the term "persistence" was even coined [42]. Graphs were a tool for managing the discretization of filtered spaces in applied contexts.

As far as we know, a true use of persistence in the study of graphs *per se* started with [22], where the clique and neighborhood complexes were built on a time-varying network for application in statistical mechanics (see also [30]). A research aiming at the physical application of persistence to polymer models of hypergraphs is developed in [1].

Complex networks have been studied with persistent homology also in [23,31]. In both cases, the main example is a network of collaborating people; simplices are formed on the basis of relationship measures between members.

Brain connections have been studied through complexes associated to graphs by various authors with different viewpoints and techniques, with exciting results: [32,38,34].

## 2 General preliminaries

Apart from fixing terminology in Subsection 2.1 and establishing the main category of work in Subsection 2.3, the present section is devoted to presenting, in the long Subsection 2.2, a framework for defining functions, with the convenient features of persistent homology, in general concrete categories.

### 2.1 Simplicial complexes and graphs

First, we recall that a *simplicial complex* $K$ (an *abstract* simplicial complex in the terminology of many authors) is a set of simplices, where a *simplex* is a finite set of elements (*vertices*) of a given set $V(K)$, such that

(i) any set of exactly one vertex is a simplex of $K$
(ii) any nonempty subset of a simplex of $K$ is a simplex of $K$ [40, Sect. 3.1].

A simplex consisting of $n+1$ vertices is said to have *dimension* $n$ and to be an $n-$simplex. The dimension of $K$ is the maximum dimension of its simplices. In the remainder all simplicial complexes will be finite.

We refer to [40] for terminology and notions of simplicial and algebraic topology; another very nice reference is [20]. For the sake of clarity, the simplex having vertices $v_0, \ldots, v_n$ will be denoted as $\langle v_0, \ldots, v_n \rangle$.

For the purposes of this article, graphs will be defined as simplicial complexes of dimension 1. So, in graph-theoretical terms they are finite simple graphs. The category **Graph** will have graphs as objects and simplicial maps as morphisms. For both graph-theoretical notions and terminology we refer to [2].

Since it is always possible to build a geometric *realization* of a finite-dimensional simplicial complex $K$, i.e. to associate a topological space $|K|$ to $K$ (its *space* or *body*) and to embed it in a suitable Euclidean space [40, Sect. 3.2, Thm. 9], we shall do it in both text and figures throughout the article.

### 2.2 Persistence functions

We set $\Delta^+ = \{(u,v) \in \mathbb{R} \,|\, u < v\}$, $\Delta = \{(u,v) \in \mathbb{R} \,|\, u = v\}$ and $\overline{\Delta}^+ = \Delta^+ \cup \Delta$.

We recall that a *concrete* category [28, Sect. I.7] is a pair $(\mathbf{C},\mathcal{U})$ where $\mathbf{C}$ is a category and $\mathcal{U}$ is a faithful functor $\mathcal{U} : \mathbf{C} \to \mathbf{Set}$. We require that in $\mathbf{C}$ every subobject (in the sense of [28, Sect.V.7]) of an object $X$ corresponds to a unique object $Y$ of $\mathbf{C}$; we will then call $Y$ itself a *subobject* of $X$. Of all the monics of $\hom(Y,X)$, the one which $\mathcal{U}$ applies to a set inclusion will be called an *embedding* of $Y$ into $X$. In the remainder of this subsection $(\mathbf{C},\mathcal{U})$ will be such a pair.

We recall that $\boldsymbol{R}$ is the poset category of the real numbers. Adapting from [26], we define a *filtration* in $\mathbf{C}$ to be a functor $\mathcal{F} : \boldsymbol{R} \to \mathbf{C}$ such that $\mathcal{F}$ maps each morphism of $\hom(\boldsymbol{R})$ to an embedding of subobjects.

**Proposition 1.** *Let $(X,f)$ be a pair such that $X \in \mathrm{Obj}(\mathbf{C})$ and $f : \mathcal{U}(X) \to \mathbb{R}$ is a function such that for any $t \in \mathbb{R}$ there is exactly one subobject $Y_t$ of $X$ with $\mathcal{U}(Y_t) = f^{-1}\big((-\infty, t]\big)$. Then $\mathcal{F}_{(X,f)}$ defined by $\mathcal{F}_{(X,f)}(t) = Y_t$ is a filtration on $\mathbf{C}$ and $\mathcal{S}_{(X,f)} = \mathcal{U} \circ \mathcal{F}$ is a filtration in $\mathcal{U}(\mathbf{C})$.* □

The function $f$ is called a *filtering* function for $X$.

*Remark 1.* The filtrations of Prop. 1 are generalizations of what is called *sublevel set filtration* in [26].

Let now an integer valued function $\Lambda$ be defined on all embeddings $\iota$ of subobjects in $\mathbf{C}$. In particular, given any pair $(X,f)$ as in Prop. 1, $\lambda_{(X,f)} : \Delta^+ \to \mathbb{Z}$ is defined by $\lambda_{(X,f)}(u,v) = \Lambda(\iota)$ where $\iota$ is the embedding of subobjects sent by $\mathcal{U}$ to the inclusion map of $f^{-1}\big((-\infty,u]\big)$ into $f^{-1}\big((-\infty,v]\big)$.

**Definition 1.** *$\lambda_{(X,f)} : \Delta^+ \to \mathbb{Z}$ is said to be a* persistence function *if*

1. *$\lambda_{(X,f)}(u,v)$ is nondecreasing in $u$ and nonincreasing in $v$;*

2. *for all* $u_1, u_2, v_1, v_2 \in \mathbb{R}$ *such that* $u_1 \leq u_2 < v_1 \leq v_2$ *the following inequality holds:* $\lambda_{(X,f)}(u_2, v_1) - \lambda_{(X,f)}(u_1, v_1) \geq \lambda_{(X,f)}(u_2, v_2) - \lambda_{(X,f)}(u_1, v_2)$
3. *given an analogous pair* $(Y, g)$*, if a* $\mathbf{C}$*-isomorphism* $\psi : X \to Y$ *exists such that* $\sup_{p \in \mathcal{U}(X)} |f(p) - g\big((\mathcal{U}(\psi))(p)\big)| \leq h$ $(h > 0)$*, then for all* $(u, v) \in \Delta^+$ *the inequality* $\lambda_{(X,f)}(u - h, v + h) \leq \lambda_{(Y,g)}(u, v)$ *holds.*

*Remark 2.* The persistent Betti number functions (also called Rank Invariants), at all homology degrees, are the most relevant example of persistence functions, when $\mathbf{C}$ is the category of topological spaces, $\mathcal{U}$ is the forgetful functor and the filtering functions are continuous. The same holds in the category of simplicial complexes, where the filtering function respects the condition that its value on each simplex $\sigma$ is greater than or equal to its value on each face of $\sigma$. See, e.g. [16,35,13].

*Remark 3.* Conditions 1 and 2 of Def. 1 correspond to Prop. 1 and Lemma 1 of [18], where discontinuities of "size functions" (what would later be called 0-th persistent Betti number functions) were studied. Condition 3, which appears here as part of a definition, is also present as a proposition in [17, Thm. 3.2], [10, Prop.10].

The following definition is meant to generalize the type of coverings that classically generate persistence functions, e.g. connected components, path-connected components (giving rise to 0-Betti numbers in Čech and singular homology respectively). This generalization will be used to define a persistence function from the blocks of a graph in Section 4.1.

**Definition 2.** *A* coherent covering $\mathcal{V}$ *on* $(\mathbf{C}, \mathcal{U})$ *is the assignment to each* $X \in \mathrm{Obj}(\mathbf{C})$ *of a set* $\mathcal{V}(X)$ *of subsets of* $\mathcal{U}(X)$*, such that*

1. $\mathcal{V}(X)$ *is a finite covering of* $\mathcal{U}(X)$*;*
2. *if* $\mathcal{U}(X_1) \subseteq \mathcal{U}(X_2)$*, then each element of* $\mathcal{V}(X_1)$ *is contained in exactly one element of* $\mathcal{V}(X_2)$*;*
3. *if* $\psi : X \to Y$ *is a* $\mathbf{C}$*-isomorphism, then* $\mathcal{V}(Y) = \Big(\mathcal{U}(\psi)\Big)\big(\mathcal{V}(X)\big)$*.*

Because of the many symbols referring to various sets, we suggest that the reader keeps in mind the example of persistent 0-Betti numbers, where the $X$s are sublevel sets and the $Z$s are their path-connected components.

**Proposition 2.** *Let a coherent covering* $\mathcal{V}$ *be given on* $(\mathbf{C}, \mathcal{U})$*; for all objects* $X$ *of* $\mathbf{C}$*, for all filtering functions* $f : X \to \mathbb{R}$*, the function* $\lambda_{(X,f)} : \Delta^+ \to \mathbb{Z}$*, where* $\lambda_{(X,f)}(u, v)$ *is defined as the number of elements of* $\mathcal{V}\Big(f^{-1}\big((-\infty, v]\big)\Big)$ *containing at least one element of* $\mathcal{V}\Big(f^{-1}\big((-\infty, u]\big)\Big)$*, is a persistence function.*

*Proof.* By condition 1 of Def. 2, $\lambda_{(X,f)}(u, v)$ is a nonnegative integer. For the remainder of this proof, for any real number $w$ we set $X_w = \mathcal{S}_{(X,f)}(w) = f^{-1}\big((-\infty, w]\big)$. We now prove this easy claim:

**(*)** Let $X' \subseteq X'' \subseteq X''' \subseteq X$. For each element $Z' \in \mathcal{V}(X')$ (i.e. a subset of $X'$ belonging to its covering) there are exactly one $Z'' \in \mathcal{V}(X'')$ and one $Z''' \in \mathcal{V}(X''')$ such that $Z' \subseteq Z'' \subseteq Z'''$.

In fact, the existence and unicity of $Z''$ are condition 2 of Def. 2. The existence and unicity of $Z'''$ comes from the same condition and from the transitivity of inclusion.

We shall now prove all claimed inequalities by showing that it is possible to define suitable injective maps.

1. Let $u_1 < u_2 < v$; if $\overline{Z}, \hat{Z} \in \mathcal{V}(X_v)$, with $\overline{Z} \neq \hat{Z}$, contain elements of $\mathcal{V}(X_{u_1})$ (necessarily different by condition 2 of Def. 2), then by (*) they also contain elements of $\mathcal{V}(X_{u_2})$, which are different by condition 2 of Def. 2; then $\lambda_{(X,f)}(u_1, v) \leq \lambda_{(X,f)}(u_2, v)$.
   Let $u < v_1 < v_2$; if $Z'', Z^{**} \in \mathcal{V}(X_{v_2})$, with $Z'' \neq Z^{**}$, contain elements $Z, \tilde{Z}$ of $\mathcal{V}(X_u)$ (necessarily different by condition 2 of Def. 2), then by (*) there exist $Z', Z^* \in \mathcal{V}(X_{v_1})$ (necessarily different by condition 2 of Def. 2) such that $Z \subseteq Z'$, $\tilde{Z} \subseteq Z^*$; therefore $\lambda_{(X,f)}(u, v_1) \geq \lambda_{(X,f)}(u, v_2)$.
2. Let now $u_1 \leq u_2 < v_1 \leq v_2$. For $i = 1, 2$, the difference $\lambda_{(X,f)}(u_2, v_i) - \lambda_{(X,f)}(u_1, v_i)$ is the number of elements of $\mathcal{V}(X_{v_i})$ which contain at least an element of $\mathcal{V}(X_{u_2})$ but no elements of $\mathcal{V}(X_{u_1})$.
   Let $Z'' \in \mathcal{V}(X_{v_2})$ contain an element $Z^{**} \in \mathcal{V}(X_{u_2})$ but no elements of $\mathcal{V}(X_{u_1})$. Then by (*) there exists $Z' \in \mathcal{V}(X_{v_1})$ such that $Z^{**} \subseteq Z' \subseteq Z''$. No element of $\mathcal{V}(X_{u_1})$ can be contained in $Z'$, otherwise it would also be contained in $Z''$. If $\overline{Z}'' \in \mathcal{V}(X_{v_2})$ is in the same situation as $Z''$ but different from it, then the corresponding $\overline{Z}' \in \mathcal{V}(X_{v_1})$ is different from $Z'$ by condition 2 of Def. 2. Therefore $\lambda_{(X,f)}(u_2, v_1) - \lambda_{(X,f)}(u_1, v_1) \geq \lambda_{(X,f)}(u_2, v_2) - \lambda_{(X,f)}(u_1, v_2)$
3. Given an analogous pair $(Y, g)$, let a $C$-isomorphism $\psi : X \to Y$ exist such that $\sup_{p \in \mathcal{U}(X)} |f(p) - g\big((\mathcal{U}(\psi))(p)\big)| \leq h$ $(h > 0)$. For any $u > h, v > u$, we have $X_{u-h} \subseteq \psi^{-1}(Y_u) \subseteq \psi^{-1}(Y_v) \subseteq X_{v+h}$. Then, by applying (*) twice, if $Z'''' \in \mathcal{V}(X_{v+h})$ contains an element $Z' \in \mathcal{V}(X_{u-h})$, then there exist uniquely determined $Z'' \in \mathcal{V}\big(\psi^{-1}(Y_u)\big)$, $Z''' \in \mathcal{V}_X\big(\psi^{-1}(Y_v)\big)$ such that $Z' \subseteq Z'' \subseteq Z''' \subseteq Z''''$. By condition 3 of Def. 2 we have $\psi(Z''') \in \mathcal{V}(Y_v)$ and $\psi(Z'') \in \mathcal{V}(Y_u)$, and also $\psi(Z'') \subseteq \psi(Z''')$. If $\overline{Z}'''' \in \mathcal{V}(X_{v+h})$ is in the same situation as $Z''''$ but different from it, then the corresponding $\overline{Z}''$ and $\overline{Z}'''$, and their images under $\psi$, are different from $Z''$ and $Z'''$ and their images under $\psi$ respectively, by condition 2 of Def. 2. Therefore $\lambda_{(X,f)}(u-h, v+h) \leq \lambda_{(Y,g)}(u, v)$. $\square$

For the remainder of this section $\lambda_{(X,f)} : \Delta^+ \to \mathbb{Z}$ will be a persistence function. The following simple propositions have the same proofs as the quoted propositions of [18].

**Proposition 3.** [18, Cor. 1] *The following statements hold:*

1. *If $\overline{u}$ is a discontinuity point for $\lambda_{(X,f)}(\cdot, \overline{v})$ and $\overline{u} < v < \overline{v}$ then $\overline{u}$ is a discontinuity point also for $\lambda_{(X,f)}(\cdot, v)$*

2. *If $\overline{v}$ is a discontinuity point for $\lambda_{(X,f)}(\overline{u},\cdot)$ and $\overline{u} < u < \overline{v}$ then $\overline{v}$ is a discontinuity point also for $\lambda_{(X,f)}(u,\cdot)$.* □

**Proposition 4.** [18, Lemma 2] *Any open path-connected neighborhood of a discontinuity point of $\lambda_{(X,f)}$ contains at least one discontinuity point either in the first or in the second variable.* □

**Proposition 5.** [18, Prop. 6] *For every point $(\overline{u},\overline{v}) \in \Delta^+$ an $\varepsilon > 0$ exists such that the open set $W_\varepsilon = \{(u,v) \in \mathbb{R}^2 \mid 0 < |u-\overline{u}| < \varepsilon, 0 < |v-\overline{v}| < \varepsilon\ \}$ does not contain any discontinuity point of $\lambda_{(X,f)}$.* □

**Definition 3.** [18, Def. 4] [10, Def. 4] *For every point $p(u,v) \in \Delta^+$ we define the* multiplicity *of $p$ for $\lambda_{(X,f)}$ to be the number $\mu(p)$ equal to the minimum, over all positive real $\varepsilon$ with $u+\varepsilon < v-\varepsilon$, of*

$$\lambda_{(X,f)}(u+\varepsilon,v-\varepsilon)-\lambda_{(X,f)}(u-\varepsilon,v-\varepsilon)-\lambda_{(X,f)}(u+\varepsilon,v+\varepsilon)+\lambda_{(X,f)}(u-\varepsilon,v+\varepsilon)$$

*We shall call any $p \in \Delta^+$ with positive multiplicity a* proper cornerpoint*.*

**Definition 4.** [18, Def. 5] [10, Def. 5] *For every vertical line $r$, with equation $u = k$, let us identify $r$ with the pair $(k,\infty)$ and define its multiplicity for $\lambda_{(X,f)}$ to be the number $\mu(r)$ equal to the minimum, over all positive real $\varepsilon$ with $k+\varepsilon < 1/\varepsilon$, of*

$$\lambda_{(X,f)}(k+\varepsilon,1/\varepsilon)-\lambda_{(X,f)}(k-\varepsilon,1/\varepsilon)$$

*We shall call any $r$ with positive multiplicity a* cornerpoint at infinity*.*

**Proposition 6.** [18, Lemma 3]

1. *If $\overline{u}$ is a discontinuity point for $\lambda_{(X,f)}(\cdot,\overline{v})$ with $\overline{u} < \overline{v}$, then either there is a cornerpoint of $\lambda_{(X,f)}$ on the closed half-line $\{(\overline{u},v) \in \mathbb{R}^2 \mid \overline{v} \le v\}$, or the line $u = \overline{u}$ is a cornerpoint at infinity, or both cases occur.*
2. *If $\overline{v}$ is a discontinuity point for $\lambda_{(X,f)}(\overline{u},\cdot)$ with $\overline{u} < \overline{v}$, then there is a cornerpoint of $\lambda_{(X,f)}$ on the closed half-line $\{(u,\overline{v}) \in \mathbb{R}^2 \mid u \le \overline{u}\}$.* □

*Remark 4.* Propositions 3 to 6, depending only on conditions 1 and 2 of Def. 1, grant the appearance with overlapping triangles typical of persistent Betti number functions.

From now on, we shall also assume that every function $\lambda_{(X,f)}$ has a finite number of cornerpoints (proper and at infinity).

We now generalize the notion of persistence diagram [8,7] to the wider context of functions $\lambda_{(X,f)}$, of the type described in this section.

**Definition 5.** *The* persistence diagram *of $\lambda_{(X,f)}$ is the multiset of its cornerpoints (proper and at infinity), each repeated as many times as its multiplicity, together with all points of the diagonal $\Delta$, each counted with infinite ($\aleph_0$) multiplicity. For the sake of simplicity it will just be denoted as $D(f)$.*

**Definition 6.** *Given the persistence diagrams* $D(f), D(g)$ *of* $\lambda_{(X,f)}, \lambda_{(Y,g)}$ *respectively, let* $\Gamma$ *be the set of all bijections between the multisets* $D(f)$ *and* $D(g)$*. We define the* bottleneck *(formerly* matching*)* distance *as the real number*

$$d(D(f), D(g)) = \inf_{\gamma \in \Gamma} \sup_{p \in D(f)} \|p - \gamma(p)\|_\infty$$

*Remark 5.* As in the "classical" persistence theory, this distance checks the maximum displacement between corresponding points for a given matching either between cornerpoints of the two diagrams or between cornerpoints and their own projections on the diagonal $\Delta$, and takes the minimum among these maxima. Minima and maxima are actually attained because of the requested finiteness.

Always in the concrete category $(\mathbf{C}, \mathcal{U})$ let us now consider two pairs $(X, f)$, $(Y, g)$ with $X, Y \in \mathrm{Obj}(\mathbf{C})$ and $f : \mathcal{U}(X) \to \mathbb{R}$, $g : \mathcal{U}(Y) \to \mathbb{R}$. Set also $H = \mathcal{U}(\overline{H})$, where $\overline{H}$ is the (possibly empty) set of $\mathbf{C}$-isomorphisms between $X$ and $Y$. we can now generalize some definitions given in [19,10,26].

**Definition 7.** *The* natural pseudodistance *of* $(X, f)$ *and* $(Y, g)$ *is*

$$\delta\big((X,f),(Y,g)\big) = \begin{cases} \infty & \text{if} \quad H = \emptyset \\ \inf_{\phi \in H} \sup_{p \in \mathcal{U}(X)} \big|f(p) - g\big(\phi(p)\big)\big| & \text{otherwise} \end{cases}$$

The following theorem depends on the previous propositions and on condition 3 of Def. 1. The proof is the same of the corresponding Thm. 29 of [10].

**Theorem 1 (Stability).** *For pairs* $(X, f), (Y, g)$ *as above,*

$$d\big(D(f), D(g)\big) \leq \delta\big((X,f),(Y,g)\big) \qquad \square$$

#### 2.2.1 Persistence for the analysis of faulty programs

This subsection is intended to embed the construction of [36] into the present theory. For the notions of *Hierarchical Recurrent Network* (*HRN*), of *desired data output*, of *data flow*, of *magnitude of a data flow error* and for the algorithm of *generation of P-intervals* we refer to that nice paper.

Let $H$ be a HRN consisting of $\bar{h}$ subprograms $H^1, \ldots, H^{\bar{h}}$ and $\delta = (\delta^1, \ldots, \delta^{\bar{h}})$ the $\bar{h}$-tuple of fixed desired data outputs. For the sake of intuition, we report in Fig. 1 the toy example discussed in [36, Section 2].

We build a category $\mathbf{C}(H, \delta)$ as follows: Objects are *data flow h-tuples* $D = (D^1_{m(1)}, \ldots, D^h_{m(h)})$, with $1 \leq h \leq \bar{h}$, where, as in [36], for each data flow $D^n_{m(n)}$ the positive integer $m(n)$ is the number of times that the data flow circulates through the cycle $H^n$. Note that, by the structure itself of HRNs, we cannot have intermediate vanishing indices $m(n)$; still we admit *partial* data flow $h$-tuples (with $h < \bar{h}$) which represent programs which do not reach the end of the HRN.

There is exactly one morphism from $D' = (D^1_{m'(1)}, \ldots, D^{h'}_{m'(h')})$ to $D = (D^1_{m(1)}, \ldots, D^h_{m(h)})$ if and only if $h' \leq h$ and for $1 \leq n \leq h'$ we have $m'(n) \leq$

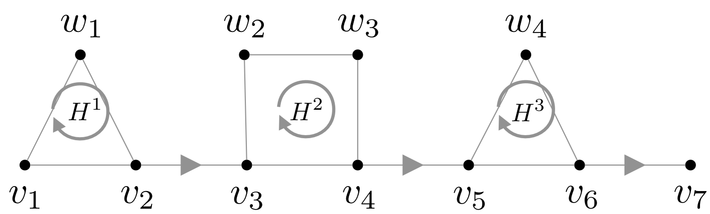


**Fig. 1.** A simple example of hierarchical recurrent network. Each subprogram is represented as a cycle as $v_1 \to w_1 \to v_2$.

$m(n)$. A subobject of $D = (D^1_{m(1)}, \ldots, D^h_{m(h)})$ is a data flow $h'$-tuple $D' = (D^1_{m'(1)}, \ldots, D^{h'}_{m'(h')})$ such that $h' \leq h$ and for $1 \leq i \leq h'$ and $m'(i) = m(i)$.

A functor $\mathcal{U}$, making $\big(\mathbf{C}(H, \delta), \mathcal{U}\big)$ a concrete category with the required features, can be defined by $\mathcal{U}\big((D^1_{m(1)}, \ldots, D^h_{m(h)})\big) = \{\big(1, m(1)\big), \ldots, \big(h, m(h)\big)\}$.

For each object $D = (D^1_{m(1)}, \ldots, D^h_{m(h)})$ we define a function $f : \mathcal{U}(D) \to \mathbb{R}$ simply by $f\Big(\big(n, m(n)\big)\Big) = n$ for all $1 \leq n \leq h$. Then, by Prop. 1, $(D, f)$ induces a filtration in $\mathbf{C}(H, \delta)$ and a filtration in $\mathcal{U}\big(\mathbf{C}(H, \delta)\big)$.

Given a program represented by a data flow $\bar{h}$-tuple $\bar{D} = (D^1_{m(1)}, \ldots, D^{\bar{h}}_{m(\bar{h})})$, the sets $S^1$ of the indices of faulty subprograms and $S^2$ of able subprograms are defined (see [36]). For $(u, v) \in \mathbb{R}$, $1 \leq u < v$, let $h, k$ be the greatest integers such that $h \leq u$, $k \leq v$. Then we have an embedding of $D_h = f^{-1}\big((-\infty, u]\big) = (D^1_{m(1)}, \ldots, D^h_{m(h)})$ into $D_k = f^{-1}\big((-\infty, v]\big) = (D^1_{m(1)}, \ldots, D^k_{m(k)})$. Now, following Algorithm 1 of [36] (Generation of P-intervals), we define $\lambda_{(D,f)}(u, v)$ as the sum of magnitudes of the data flow errors created by the faulty subprograms whose indices (in $S^1$) are less than or equal to $h$, which have not been resolved by the able subprograms whose indices (in $S^2$) are less than or equal to $k$; i.e. the dimension gap created by $D_h$ and not resolved by $D_k$. For $u < 1, v > u$ we set $\lambda_{(D,f)}(u, v) = 0$.

**Proposition 7.** *The function $\lambda_{(D,f)} : \Delta^+ \to \mathbb{Z}$ is a persistence function.*

*Proof.* we can limit our arguments to integer pairs in $\Delta^+$. We just need to check that $\lambda_{(D,f)}$ satisfies the conditions of Def. 1.

1. If $h < h' < k$, then the data flow errors occurred in $D_h$ are still not resolved by $D_k$; so $\lambda_{(D,f)}(h, k) \leq \lambda_{(D,f)}(h', k)$.
   If $h < k < k'$, then $D_k$ might lack some able subprograms present in $D_{k'}$ and miss to resolve some errors created by $D_h$; so $\lambda_{(D,f)}(h, k) \geq \lambda_{(D,f)}(h, k')$.
2. Let $h \leq h'$ and $k \leq k'$. First consider $\lambda_{(D,f)}\,(h', k)$: This expression represents the sum of the magnitudes of errors arising in $D_{h'}$ that are yet not fixed by able subprograms in $D_k$. Now $\lambda_{(D,f)}\,(h', k) - \lambda_{(D,f)}\,(h, k)$ is the magnitude sum of errors created in $D_{h'}$, but not in $D_h$, and not solved in $D_k$. It is now sufficient to recall that $k \leq k'$ to deduce that

$$\lambda_{(D,f)}\,(h', k) - \lambda_{(D,f)}\,(h, k) \geq \lambda_{(D,f)}\,(h', k') - \lambda_{(D,f)}\,(h, k')\,.$$

Indeed, there could be able subprograms $\left\{D_{i_j}\right\}_{k<i_j<k'}$ that contribute to lower the sum magnitude of the errors.

3. By construction, each isomorphism class in $\mathbf{C}(H,\delta)$ contains just one object. Moreover for each object only one filtering function of the described type is possible, so this point is trivially satisfied. $\square$

*Remark 6.* The function just described produces the persistence diagrams of [36] but gives no more information. In particular, property 3 of Def. 1 and the consequent "stability" are disappointing and practically meaningless. It would be interesting to find other more interesting and useful formalizations with a different concrete category or a different filtering function. Also note that other algorithms could be designed (as duly remarked in [36]) which could give different persistence functions.

### 2.3 Weighted graphs

We now put an important constraint on the vertices of the graphs we consider. When filtering a graph we are interested only in subgraphs induced by edge sets.

**Definition 8 (Weighted graphs).** *The category* **wGraph** *of* weighted graphs *is defined as the category whose objects are pairs $(G,f)$ where $G$ is a graph with no isolated vertices and the* weight *$f : E(G) \to \mathbb{R}$ is a map on its edges. A morphism $\varphi : (G,f) \to (G',f')$ is a simplicial map $\varphi : G \to G'$ such that, for each $e \in E(G)$, $f(e) \geq f'(\varphi(e))$*

The functor $\mathcal{U}$ making (**wGraph**,$\mathcal{U}$) a concrete category will be the forgetful functor sending a graph $G$ either to its vertex set or to its edge set. Def. 7 applies also here, so we have a natural pseudodistance between weighted graphs. We then have a straightforward characterization:

**Proposition 8.** *$\delta\left((G,f),(G',f')\right) = 0$ iff $(G,f)$ and $(G',f')$ are isomorphic in* **wGraph**. $\square$

## 3 Topological graph persistence

The "leit-motiv" of this section is to study simplicial complexes built from weighted graphs, to study then the filtration of the complexes with the methods of persistent homology. Our goal is actually to spot particular classes of sets in a graph such that the conditions (i) and (ii) of the definition of simplicial complex hold. A thorough treatise on these configurations can be found in [24].

### 3.1 Complex of cliques

A *clique* in a graph is a nonempty set of vertices whose induced subgraph is complete.

**Proposition 9.** *Given a graph $G$, the set $Cl(G)$ of its cliques is a simplicial complex.* $\square$

*Remark 7.* Not every simplicial complex $K$ is $Cl(G)$ for some $G$. In fact, let $K$ be the complex formed by an $h$-simplex $s$ $(h > 1)$ and by all of its faces. Then $K$ and $K - \{s\}$ have the same 1-skeleton $K^1$; $Cl(K^1)$ is isomorphic to $K$; $K - \{s\}$ is not the complex of cliques of any graph. The next easy proposition amends this gap.

**Proposition 10.** *For any simplicial complex $K$, let $K'$ be its barycentric subdivision. Then for the graph $G = (K')^1$ we have that $Cl(G)$ is isomorphic to $K'$ and $|Cl(G)|$ is homeomorphic to $|K|$.* $\square$

If $K = Cl(G)$, then its suspension $\Sigma(K)$ is the complex of cliques of

$$CSusp(G) = \big(V(G) \cup \{x, y\}, E(G) \cup \{\langle x, v\rangle, \langle v, y\rangle \mid v \in V(G)\}\big)$$

where $x, y \notin V(G)$. In particular, any sphere of dimension $h \geq 1$ can be triangulated by the clique complex of a suitable nonempty graph, e.g. starting from a 4-cycle for $\mathbb{S}^1$ and applying $CSusp$ the necessary number of times.

**Corollary 1.** *For any finite sequence $\sigma$ of nonnegative integers, there exists a graph $G$ such that $\sigma$ is the sequence of Betti numbers of $Cl(G)$.* $\square$

The next proposition makes it possible to associate a filtration of simplicial complexes (hence also of their spaces) to a filtration of graphs.

**Proposition 11.** *If $G$ is a subgraph of $H$, $Cl(G)$ is a subcomplex of $Cl(H)$.*

*Proof.* Every clique of $G$ is also a clique of $H$. $\square$

Let now $(G, f)$ be a weighted graph. We define a filtering function $f_{Cl}$ : $Cl(G) \to \mathbb{R}$ as follows:

- for every 0-simplex $\sigma = \langle v\rangle$, $f_{Cl}(\sigma)$ is the minimum value of $f$ on the edges incident on $v$;
- for every $k$-simplex $\sigma$ $(k \geq 1)$, i.e. for every $(k+1)$-clique, $f_{Cl}(\sigma)$ is the maximum value of $f$ on the edges of the induced complete subgraph.

**Proposition 12.** $\big(Cl(G), f_{Cl}\big)$ *is a filtered complex.*

*Proof.* By construction, the value of every simplex is $\geq$ the value of each of its faces. $\square$

### 3.2 Complex of independent sets

An *independent* (or *stable*) nonempty set in a graph is a set of vertices such that the induced subgraph does not contain any edge. Recall that, given a graph $G = (V, E)$, its *complement* is the graph $G^c = (V, E')$ where for all $u, v \in V,\ u \neq v$, $\langle u, v\rangle \in E'$ if and only if $\langle u, v\rangle \notin E$; i.e. it has the same vertex set as $G$ and its edge set is complementary to $E$ with respect to the complete graph with the same vertices. Then a set of vertices is independent in $G$ if and only if it is a clique in $G^c$ and conversely.

**Proposition 13.** *Given a graph $G$, the set $I(G)$ of its independent sets is a simplicial complex.* □

*Remark 8.* Of course a remark analogous to Rem. 7 applies, and we have the following proposition as a remedy.

**Proposition 14.** *For any simplicial complex $K$, let $K'$ be its barycentric subdivision. Then for the graph $G = (K')^1$ we have that $I(G^c)$ is isomorphic to $K'$ and $|I(G^c)|$ is homeomorphic to $|K|$.* □

If $K = I(G)$, then its suspension $\Sigma(K)$ is the complex of independent sets of

$$ISusp(G) = \big(V(G) \cup \{x, y\}, E(G) \cup \{\langle x, y\rangle\}\big)$$

where $x, y \notin V(G)$. In particular, a sphere of any dimension can be triangulated by a suitable $I(G)$.

**Corollary 2.** *For any finite sequence $\sigma$ of nonnegative integers, there exists a graph $G$ such that $\sigma$ is the sequence of Betti numbers of $I(G)$.* □

The sort of duality between cliques and independent sets implies that the monotonically increasing correspondence of Prop. 11 becomes decreasing here.

**Proposition 15.** *If $G$ is a subgraph of $H$, $I(H)$ is a subcomplex of $I(G)$.* □

### 3.3 Complex of neighborhoods

In a graph $G = (V, E)$, given $v \in V$ its *neighborhood* in $G$ is the set $N_G(v) = \{v\} \cup \{u \in V \mid \langle v, u\rangle \in E\}$.

**Proposition 16.** [27,29] *Given a graph $G$, the set $Nb(G)$ of all nonempty subsets of neighborhoods of vertices of $G$ is a simplicial complex.* □

Not all simplicial complexes can be obtained as complex of neighborhoods of a graph: the boundary of a triangle for one. Here, the barycentric subdivision trick of Prop. 10 does not work. Still, we think this construction to be most interesting, because it diverges very much from the topology of $G$ as a simplicial complex, so the topology of $Nb(G)$ can give information on combinatorial aspects of $G$ detached from its homeomorphism type. See, e.g., Fig. 2, where three obviously homeomorphic cycles give rise to complexes which are even non-homotopic. Moreover, as stressed in [29,22], it seems to be a precious tool in applications to complex network analysis.

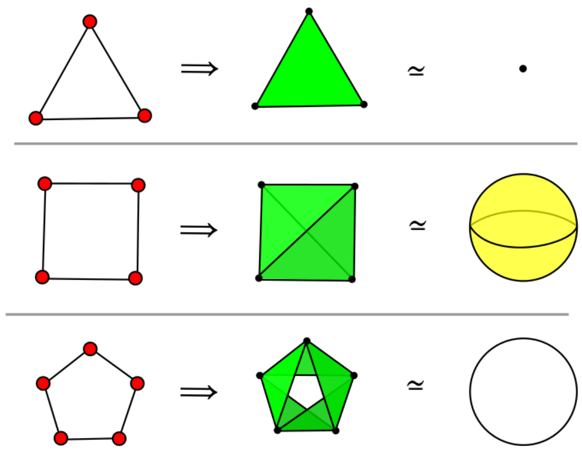

**Fig. 2.** Nonhomotopic complexes of neighborhoods of three cycles.

**Proposition 17.** *If $G$ is a subgraph of $H$, $Nb(G)$ is a subcomplex of $Nb(H)$.*

*Proof.* The neighborhood of every vertex $v$ of $G$ is a subset of the neighborhood of $v$ in $H$. $\square$

Let now $(G, f)$ be a weighted graph. We define a filtering function $f_{Nb}$ : $Nb(G) \to \mathbb{R}$ as follows:

- for every 0-simplex $\sigma = \langle v \rangle$, $f_{Nb}(\sigma)$ is the minimum value of $f$ on the edges incident on $v$;
- let now $\sigma$ be a $k$-simplex ($k \geq 1$) and let $S = \big(\sigma, E(S)\big)$ be the subgraph of $G$ induced by $\sigma$; then there exist at least one vertex in $\sigma$, but possibly $h$ (i.e. $v_1, \ldots, v_h \in \sigma$), such that $\sigma = N_S(v_1) = \ldots = N_S(v_h)$; we then set

$$f_{Nb}(\sigma) = \min_{i=1,\ldots,h} \max\{f(\langle v_i, u \rangle) \,|\, u \in N_S(v_i)\}$$

**Proposition 18.** $\big(Nb(G), f_{Nb}\big)$ *is a filtered complex.*

*Proof.* By construction, the value of every simplex is greater than or equal to the value of each of its faces. $\square$

### 3.4 Complex of enclaveless sets

Separation is as important as closeness in a network, and this is well represented by the duality clique/independent set. Somehow, there is a concept that merges these two aspects: In a graph $G = (V, E)$ a set $X \subseteq V$ is said to be *dominating* if every vertex of $G$ either belongs to $X$ or is adjacent to at least one of its vertices. Unfortunately, the inheritance property (ii) of simplicial complexes does

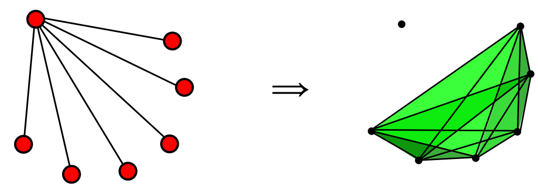

**Fig. 3.** The smallest graph containing an enclaveless set of cardinality 6 and the corresponding complex.

not hold for dominating sets; on the contrary, every superset of a dominating set is dominating. So we turn to their complementary sets. A set $Y \subseteq V$ is said to be *enclaveless* if for no $v \in Y$ we have $N(v) \subseteq Y$.

**Proposition 19.** [21] *Given a graph $G = (V, E)$, a nonempty set $Y \subseteq V$ is enclaveless if and only if $V - Y$ is dominating.* □

**Proposition 20.** *Given a graph $G$, the set $El(G)$ of all its enclaveless sets is a simplicial complex.* □

Not all simplicial complexes can be obtained as complex of enclaveless sets of a graph: e.g. the one formed by a single n-simplex and its faces cannot; the smallest $G$ such that El(G) contains an n-simplex must also contain an extra point. See Fig. 3 for $n = 5$. Still, spheres of any dimension and suspensions can be obtained.

If $K = El(G)$, then its suspension $\Sigma(K)$ is the complex of enclaveless sets of

$$ElSusp(G) = \big(V(G) \cup \{x, y\}, E(G) \cup \{\langle x, y\rangle\}\big)$$

where $x, y \notin V(G)$. It turns out to be the same construction as $ISusp$ for the complex of independent sets. Spheres can be built also in another way:

**Proposition 21.** *For $n \geq 2$ the space of $El(K_n)$ is homeomorphic to $\mathbb{S}^{n-2}$*

*Proof.* Minimal dominant sets in a complete graph $K_n$ are all singletons; so the maximal enclaveless sets are all sets of $n-1$ vertices; they together form the boundary of an $(n-1)$-simplex, while the $(n-1)$-simplex itself is not present in $El(K_n)$. □

**Proposition 22.** *If $G$ is a subgraph of $H$, $El(G)$ is a subcomplex of $El(H)$.*

*Proof.* Every dominating set of $H$ is also a dominating set of $G$. □

Let now $\big(G = (V, E), f\big)$ be a weighted graph. We define a filtering function $f_{El} : El(G) \to \mathbb{R}$ as follows:

- for every 0-simplex $\sigma = \langle v\rangle$, $f_{El}(\sigma)$ is the minimum value of $f$ on the edges incident on $v$;

– let now $\sigma$ be a $k$-simplex ($k \geq 1$); then there exist at least one set of edges in $E$, but possibly $h$ (i.e. $E_1, \ldots, E_h \subseteq E$) such that the respective induced subgraphs of $G$ all have $\sigma$ as vertex set; we then set

$$f_{El}(\sigma) = \min_{i=1,\ldots,h} \max\{f(e) \mid e \in E_i\}$$

**Proposition 23.** *$\big(El(G), f_{El}\big)$ is a filtered complex.*

*Proof.* By construction, the value of every simplex is greater than or equal to the value of each of its faces. □

### 3.5 Other complexes from graphs

There are several other classes of sets in a graph $G = (V, E)$ which respect the definition of simplicial complex [24]. We have done a preliminary study on the following ones.

– the nonempty sets $\sigma \subseteq V$ such that the subgraph induced by $\sigma$ is acyclic;
– the nonempty sets $\sigma \subseteq E$ such that the subgraph induced by $\sigma$ is acyclic;
– (with $G$ connected) the nonempty sets $\sigma \subseteq E$ such that the subgraph induced by $\sigma$ is acyclic and the subgraph induced by $E - \sigma$ is connected.
– for fixed positive $j < k$, the sets of $k$-cycles which share at least $j$ vertices;
– for fixed positive $l$, the sets of maximal cliques sharing at least $l$ vertices.

We have decided to postpone the study of these complexes: The first three because, contrarily to the preceding examples, simplices of the same dimensions are not isomorphic as induced subgraphs, so much structure is forgotten. As for the last two, the dependence on $j, k$ and $l$ respectively suggests that this type of complex might be of use in very specific applications.

### 3.6 Persistence on complexes from graphs

**Proposition 24.** *The functions persistent $r$-Betti numbers of $\big(Cl(G), f_{Cl}\big)$, of $\big(Nb(G), f_{Nb}\big)$ and of $\big(El(G), F_{El}\big)$ are persistence functions.* □

So, out of the same weighted graph $(G, f)$ we can get different persistence functions and different persistence diagrams which give us information on both the graph and its filtration.

**Proposition 25.** *Let $(G, f)$, $(G', f')$ be weighted graphs and $D(f)$, $D(f')$ be the persistence diagrams of the persistent $r$-Betti numbers of $\big(K(G), f_K\big)$, $\big(K(G', f'_K\big)$ respectively, for $r$ fixed, $K = Cl, Nb, El$. Then*

$$d\big(D(f), D(f')\big) \leq \delta\big((G, f), (G', f')\big)$$

*Proof.* Let the graph $G = (V, E)$ be isomorphic to $G'$; then $K(G)$ is a complex isomorphic to $K(G')$ (and the polyhedra $|K(G)|$, $|K(G')|$ are homeomorphic). To each isomorphism from $G$ to $G'$ there corresponds an isomorphism from $K(G)$ to $K(G')$ (and a homeomorphism from $|K(G)|$ to $|K(G')|$), but not conversely, in general. So

$$\delta\Big(\big(K(G), f_K\big), \big(K(G'), f'_K)\big)\Big) = \min_{\varphi \in \tilde{H}} \max_{\sigma \in K(G)} |f_K(\sigma) - f'_K\big(\varphi(\sigma)\big)| \leq$$
$$\leq \min_{\psi \in \overline{H}} \max_{e \in E} |f(e) - f'\big(\psi(e)\big)| = \delta\big((G, f), (G', f')\big)$$

where $\tilde{H}$ is the set of all simplicial isomorphisms from $K(G)$ to $K(G')$ and $\overline{H}$ is the set of all graph isomorphisms from $K(G)$ to $K(G')$. On the other side,

$$d\big(D(f), D(f')\big) \leq \delta\Big(\big(K(G), f_K\big), \big(K(G'), f'_K)\big)\Big)$$

is the classical stability result for filtered complexes or topological spaces (see, e.g., [8,7,10,26]). □

*Remark 9.* Unfortunately, the inequality between the natural pseudodistances of filtered complexes and filtered graphs may be strict, so we cannot get an optimality result yet, although the elegant construction of [26, Prop. 5.8] applies to $K = Cl$ through Prop. 1.

As the reader may have noted, $I(G)$ is excluded from this subsection, because of its monotonically decreasing behaviour with respect to inclusion (Prop. 15). However, there is an interesting "extended" diagram that comes exactly from this phenomenon.

### 3.7 An extended persistence of Ramsey type

The celebrated Ramsey principle [33], in its most common graph-theoretical version [2, sect. 12.3] makes it clear that one should also consider independent sets if one is interested in cliques and conversely. An equivalent viewpoint is: When interested in the information conveyed by the cliques of $G$, it is natural to also take into account the cliques of $G^c$. In the persistence field, the concept of *extended persistence* [9] explores the lower half-plane $\Delta^- = \{(u, v) \in \mathbb{R} \,|\, u > v\}$ by relative homology. We somehow merge these two philosophies in the following setting.

Given a weighted graph $\big(G = (V, E), f\big)$, build the pair $(\overline{G}, \overline{f})$, where $\overline{G} = (V, \overline{E})$ is the complete graph on the vertex set $V$, and

$$\begin{aligned} \overline{f} : \overline{E} \to & \quad \mathbb{R} \cup \{+\infty\} \\ e \mapsto & \begin{cases} f(e) & \text{if } e \in E \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

Now, with a slight abuse we have that $\big(Cl(\overline{G}, (-\overline{f})\big)$ is a filtered simplicial complex. For any weighted graph $(H, h)$ let the function $\beta^r_{(H,h)}$ be the persistent $r$-Betti number function of $\big(Cl(H), h_{Cl}\big)$.

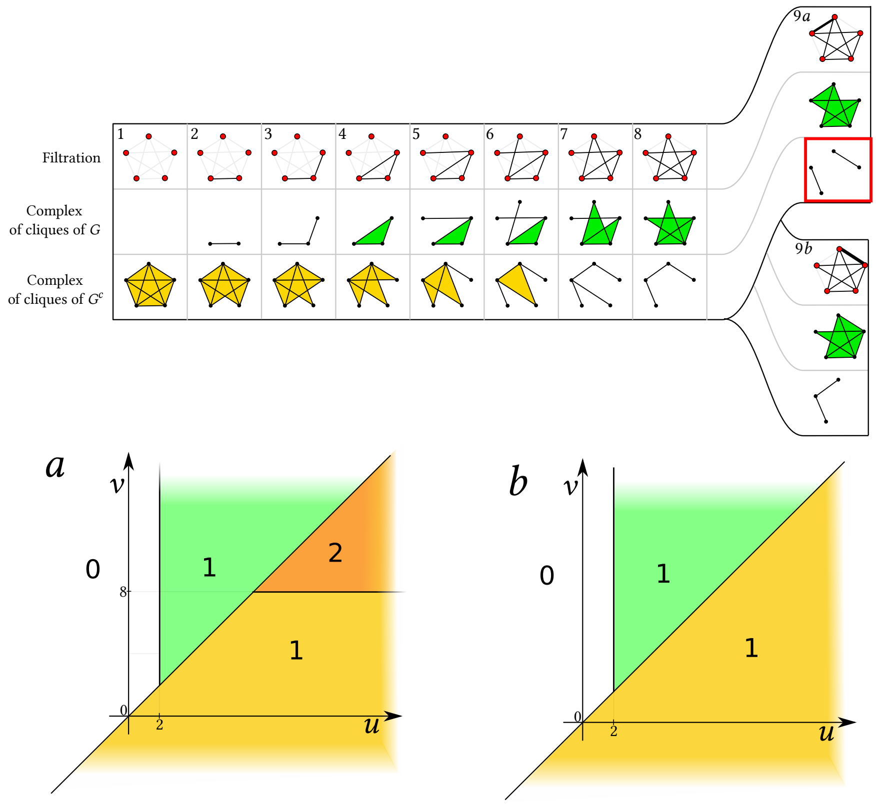


**Fig. 4.** Two filtrations and their extended persistent 0-Betti number functions.

**Definition 9.** *The* extended persistent $r$-Betti number function *of* $(G, f)$ *is* $\overline{\beta}^r_{(G,f)} : \mathbb{R}^2 \to \mathbb{Z}$, *such that*

$$(u, v) \mapsto \begin{cases} \beta^r_{(G,f)}(u, v) & \text{if} \quad u \leq v \\ \beta^r_{\left(\overline{G}, -\overline{f}\right)}(-u, -v) & \text{if} \quad u > v \end{cases}$$

These functions are actually carrying more information than the ones relative just to $\left(Cl(G), f_{Cl}\right)$, as Fig. 4 shows on two filtrations of a same graph, which differ just in the last step. The two filtrations are indistinguishable using just the persistent Betti functions on $\left(Cl(G), f_{Cl}\right)$: the persistent 0-Betti number functions are equal, and for $n \geq 1$ the persistent $n$-Betti number functions are trivially zero. On the contrary, as shown in the figure, the extended persistent 0-Betti number functions are different.

*Remark 10.* Of course, the same extension can be defined also for the other constructions, i.e. by considering neighborhoods of $G$ and of $G^c$, enclaveless sets of $G$ and of $G^c$, etc. For the moment we restrict our attention to this case, because of the role of cliques and independent sets both in theory and applications.

## 4 Graph-theoretical persistence

Now we want to take advantage of the general setting established in Subsection 2.2 for defining two persistence functions on weighted graphs, without passing through the construction of a simplicial complex and the computation of homology.

### 4.1 Blocks

We recall that in a (loopless) graph $G$ a *cut vertex* (or *separating vertex*) is a vertex $v \in V(G)$ whose deletion (along with incident edges) makes the number of connected components of $G$ increase. A *block* is a connected graph which does not contain any cut vertex. A block of a graph $G$ is a maximal subgraph $H$ such that $H$ is a block [2]. We shall now work in the concrete category $(\mathbf{Graph}, \mathcal{U})$, where $\mathcal{U}$ is the forgetful functor mapping each graph to its edge set.

**Proposition 26.** *The assignment $\mathcal{B}$, which maps each graph $G$ to the partition $\mathcal{B}(G)$ of its edge set $E(G)$ into the edge sets of its blocks, is a coherent covering.*

*Proof.* Let $G$ be a graph.

1. Each edge of $G$ belongs to at least one (actually exactly one) block. The set of blocks of a finite graph is finite.
2. If $H$ is a subgraph of $G$, each block of $H$ is contained in exactly one block of $G$.
3. Blocks correspond through graph isomorphisms. □

**Definition 10.** *Given a weighted graph $(G, f)$, we call* persistent block number *the function $bl_{(G,f)} : \Delta^+ \to \mathbb{Z}$ which maps the pair $(u, v)$ to the number of blocks of $\mathcal{U}^{-1}\big(f^{-1}(-\infty, v]\big)$ containing at least one block of $\mathcal{U}^{-1}\big(f^{-1}(-\infty, u]\big)$.*

**Corollary 3.** *$bl_{(G,f)}$ is a persistence function.*

*Proof.* By Prop. 2. □

An example of persistent block number function can be seen in Fig. 5. We can then associate to $bl_{(G,f)}$, via Def. 5, a *persistent block diagram* $D_{bl}(f)$ with all classical features granted by the propositions of Section 2.2.

**Corollary 4.** *Given weighted graphs $(G, f)$, $(G', f')$ and the respective persistent block diagrams $D_{bl}(f)$ and $D_{bl}(f')$, we have*

$$d\big(D_{bl}(f), D_{bl}(f')\big) \leq \delta\big((G, f), (G', f')\big)$$

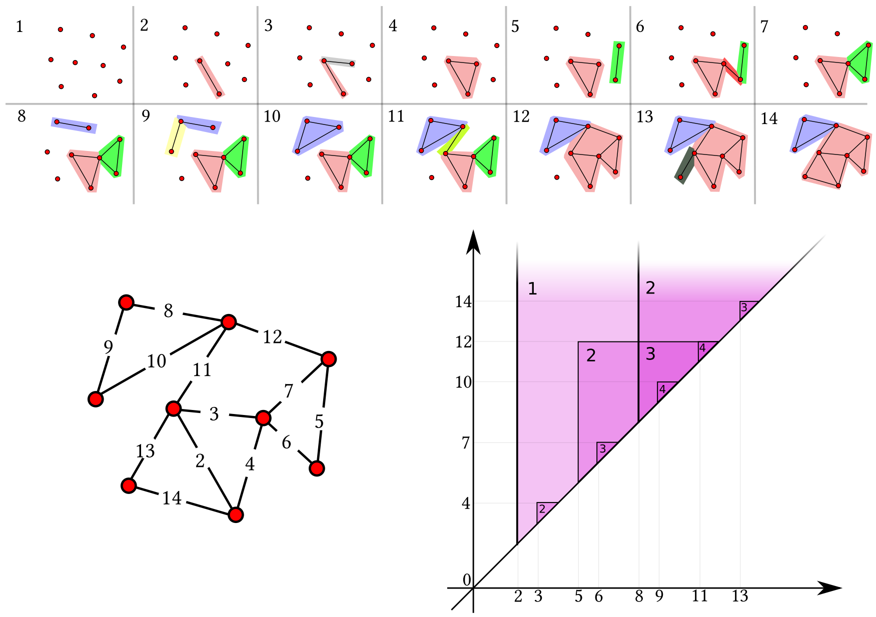


**Fig. 5.** Example of a filtration on a graph. The persistence block number function is based on the subgraphs induced by edge sets, so isolated vertices are not considered as blocks.

*Proof.* By Thm. 1. □

The next construction follows the idea underlying Lemma 30 of [10]. If a weighted graph $(G, f)$ has several blocks (so at least one cut vertex), then all blocks arising in the filtration are contained in one of the blocks of $G$ itself. So the construction can be performed for each of the blocks of $G$.

**Lemma 1.** *Let $bl_{(G,f)}$, $bl_{(G',f')}$ be the persistent block number functions of two weighted graphs $(G, f)$, $(G', f')$ which have no cut vertices; let $D_{bl}(f)$, $D_{bl}(f')$ be the respective persistent block diagrams. Then there exist weighted graphs $(M, h)$, $(M, h')$ such that*

1. $D_{bl}(f) = D_{bl}(h)$, $D_{bl}(f') = D_{bl}(h')$
2. $d\big(D_{bl}(h), D_{bl}(h')\big) = \delta\big((M, h), (M, h')\big) = \max_{e \in E(L)} |h(e) - h'(e)|$

*Proof.* There is at least one bijection $\gamma$ between the multisets $D_{bl}(f)$ and $D_{bl}(f')$ which realizes the distance $d = d\big(D_{bl}(f), D_{bl}(f')\big)$. There are $\overline{p}$, $\overline{p}'$, cornerpoints at infinity of $D_{bl}(f)$, $D_{bl}(f')$ respectively, and points $p_1, \ldots, p_m$, $p'_1, \ldots, p'_m$, $q_1, \ldots, q_r$, $q'_1, \ldots, q'_s$ where the $p_i$ and $q_i$ are proper cornerpoints of $D_{bl}(f)$, the $p'_i$ and $q'_i$ are proper cornerpoints of $D_{bl}(f')$ such that ($\pi$ being the orthogonal projection on the diagonal $\Delta$)

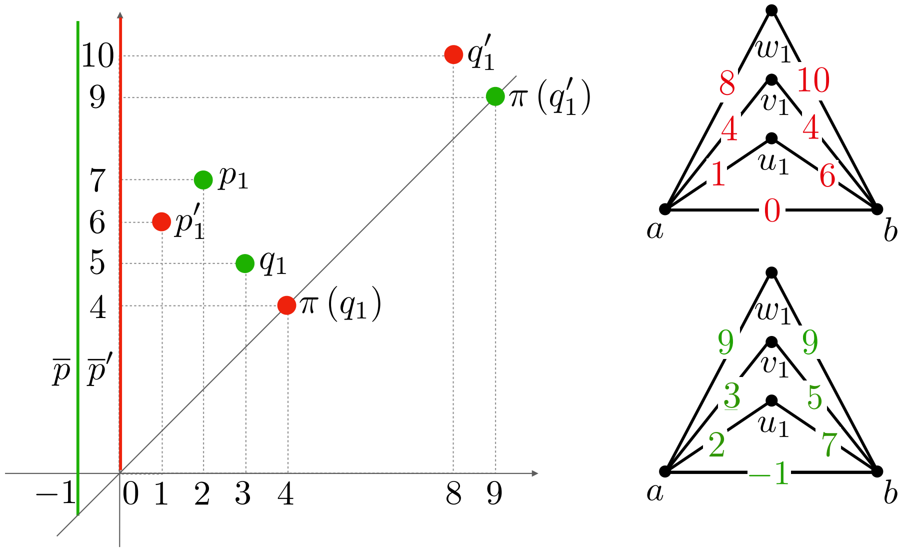


**Fig. 6.** Example of the construction of Lemma 1. The diagram $D(f) = D(h)$ and the function $h$ are in green; $D(f') = D(h')$ and $h'$ are in red.

$\gamma(\overline{p}) = \overline{p}'$
$\gamma(p_1) = p_1', \ldots, \gamma(p_m) = p_m'$
$\gamma(q_1) = \pi(q_1), \ldots, \gamma(q_r) = \pi(q_r)$
$\gamma\big(\pi(q_1')\big) = q_1', \ldots, \gamma\big(\pi(q_s')\big) = q_s'$

(some of the numbers $m,\ r,\ s$ might be null). The distance $d$ is then the maximum of the distance in the $L^\infty$ norm of corresponding points (considering $\infty - \infty = 0$). We now construct a new graph $M$ as follows.

$$V(M) = \{a, b,\ u_1, \ldots, u_m,\ v_1, \ldots, v_r,\ w_1, \ldots, w_s\}$$

$E(M)$ contains the edge $\langle a, b\rangle$, one edge joining $a$ to each of the $u_i, v_i, w_i$, one edge joining $b$ to each of the $u_i, v_i, w_i$. We define $h, h' : E(M) \to \mathbb{R}$ by

$$h(\langle a,b\rangle) = x_{\overline{p}}, \quad h'(\langle a,b\rangle) = x_{\overline{p}'}$$
$$h(\langle a,u_i\rangle) = x_{p_i}, \quad h(\langle u_i,b\rangle) = y_{p_i}, \qquad h'(\langle a,u_i\rangle) = x_{p_i'}, \quad h'(\langle u_i,b\rangle) = y_{p_i'}$$
$$h(\langle a,v_i\rangle) = x_{q_i}, \quad h(\langle v_i,b\rangle) = y_{q_i}, \qquad h'(\langle a,v_i\rangle) = \tfrac{x_{q_i}+y_{q_i}}{2}, \quad h'(\langle v_i,b\rangle) = \tfrac{x_{q_i}+y_{q_i}}{2}$$
$$h(\langle a,w_i\rangle) = \tfrac{x_{q_i'}+y_{q_i'}}{2}, \quad h(\langle w_i,b\rangle) = \tfrac{x_{q_i'}+y_{q_i'}}{2}, \qquad h'(\langle a,w_i\rangle) = x_{q_i'}, \quad h'(\langle w_i,b\rangle) = y_{q_i'}$$

for all indices $i$ in the relevant intervals. Then a straightforward check shows that $D_{bl}(f) = D_{bl}(h),\ D_{bl}(f') = D_{bl}(h')$ and that the identity of $M$ realizes the natural pseudodistance $\delta\big((M,h),(M,h')\big) = d$ $\square$

A toy example is given in Fig. 6. We now follow the logical line of Thm. 32 of [10] for proving the optimality of the bottleneck (or matching) distance among the lower bounds for the natural pseudodistance which can come from distances between persistent block diagrams.

**Theorem 2.** *If $\tilde{d}$ is a distance for persistent block diagrams such that*

$$\tilde{d}\big(D_{bl}(f), D_{bl}(f')\big) \leq \delta\big((G,f),(G',f')\big)$$

*for any persistent block diagrams $D_{bl}(f),\ D_{bl}(f')$ of weighted graphs $(G,f),\ (G',f')$, with $G,\ G'$ isomorphic, then*

$$\tilde{d}\big(D_{bl}(f), D_{bl}(f')\big) \leq d\big(D_{bl}(f), D_{bl}(f')\big)$$

*Proof.* We prove the inequality by contradiction: assume $(G,f),\ (G',f')$, with $G,\ G'$ isomorphic, exist such that

$$d\big(D_{bl}(f), D_{bl}(f')\big) < \tilde{d}\big(D_{bl}(f), D_{bl}(f')\big)$$

By Lemma 1, weighted graphs $(M,h),\ (M,h')$ exist such that $D_{bl}(h) = D_{bl}(f)$, $D_{bl}(h') = D_{bl}(f')$ and

$$d\big(D_{bl}(h), D_{bl}(h')\big) = \delta\big((M,h),(M,h')\big)$$

Of course, $\tilde{d}\big((D_{bl}(f), D_{bl}(f')\big) = \tilde{d}\big((D_{bl}(h), D_{bl}(h')\big)$. Therefore we would have

$$\delta\big((M,h),(M,h')\big) = d\big(D_{bl}(h), D_{bl}(h')\big) = d\big(D_{bl}(f), D_{bl}(f')\big) <$$

$$< \tilde{d}\big((D_{bl}(f), D_{bl}(f')\big) = \tilde{d}\big((D_{bl}(h), D_{bl}(h')\big) \leq \delta\big((M,h),(M,h')\big)$$

yielding a contradiction. □

### 4.2 Edge-blocks

We recall that in a graph $G$ a *cut edge* (or *bridge*) is an edge $e \in E(G)$ whose deletion makes the number of connected components of $G$ increase [2]. We define an *edge-block* as a connected graph which does not contain any cut edge. An edge-block of a graph $G$ is a maximal subgraph $H$ such that $H$ is an edge-block.

We shall now work in the concrete category $(\mathbf{Graph}, \mathcal{U})$, where $\mathcal{U}$ is, this time, the forgetful functor mapping each graph to its vertex set. The proofs of the next statements are totally analogous to those of Section 4.1, except for Lemma 2, where the construction needs a slight, intuitive modification exemplified in Fig. 8.

**Proposition 27.** *The assignment $\mathcal{E}$, which maps each graph $G$ to the partition $\mathcal{E}(G)$ of its vertex set $V(G)$ into the vertex sets of its edge-blocks, is a coherent covering.* □

**Definition 11.** *Given a weighted graph $(G,f)$, we call* persistent edge-block number *the function $ebl_{(G,f)} : \Delta^+ \to \mathbb{Z}$ which maps the pair $(u,v)$ to the number of edge-blocks of $\mathcal{U}^{-1}\big(f^{-1}(-\infty,v]\big)$ containing at least one edge-block of $\mathcal{U}^{-1}\big(f^{-1}(-\infty,u]\big)$.*

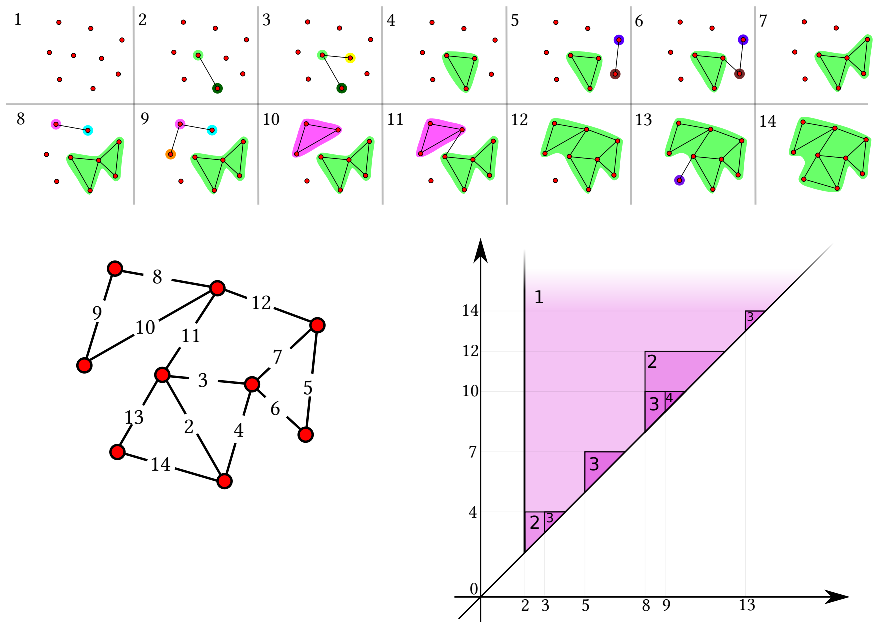


**Fig. 7.** Example of persistent edge-block number function. It is based on the subgraphs induced by edge sets too, so isolated vertices are not considered as edge-blocks.

**Corollary 5.** *$ebl_{(G,f)}$ is a persistence function.* □

An example of persistent edge-block number function can be seen in Fig. 7. We can associate to $ebl_{(G,f)}$, via Def. 5, a *persistent edge-block diagram* $D_{ebl}(f)$.

**Corollary 6.** *Given weighted graphs $(G, f)$, $(G', f')$ and the respective persistent edge-block diagrams $D_{ebl}(f)$ $D_{ebl}(f')$, we have*

$$d\big(D_{ebl}(f), D_{ebl}(f')\big) \leq \delta\big((G, f), (G', f')\big) \qquad \square$$

**Lemma 2.** *Let $ebl_{(G,f)}$, $ebl_{(G',f')}$ be the persistent edge-block number functions of two weighted graphs $(G, f)$, $(G', f')$ which have no cut edges; let $D_{ebl}(f)$, $D_{ebl}(f')$ be the respective persistent block diagrams. Then there exist weighted graphs $(M, h)$, $(M, h')$ such that*

1. $D_{ebl}(f) = D_{ebl}(h)$, $D_{ebl}(f') = D_{ebl}(h')$
2. $d\big(D_{ebl}(h), D_{ebl}(h')\big) = \delta\big((M, h), (M, h')\big) = \max_{e \in E(L)} |h(e) - h'(e)|$ □

Fig. 8 shows the construction needed for Lemma 2, in the case of the diagrams (this time to be considered as persistent edge-block diagrams) of Fig. 6, left.

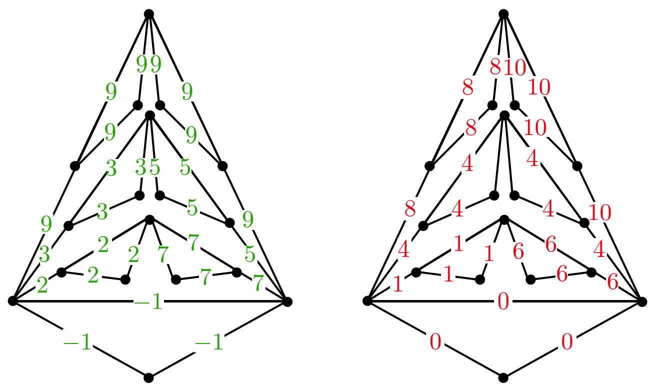


**Fig. 8.** Example of the construction necessary for Lemma 2.

**Theorem 3.** *If $\tilde{d}$ is a distance for persistent edge-block diagrams such that*

$$\tilde{d}\big(D_{ebl}(f), D_{ebl}(f')\big) \leq \delta\big((G, f), (G', f')\big)$$

*for any persistent edge-block diagrams $D_{ebl}(f)$, $D_{ebl}(f')$ of weighted graphs $(G, f)$, $(G', f')$, with $G$, $G'$ isomorphic, then*

$$\tilde{d}\big(D_{ebl}(f), D_{ebl}(f')\big) \leq d\big(D_{ebl}(f), D_{ebl}(f')\big) \qquad \square$$

## 5 Conclusions and future work

We have exposed a general framework for defining persistence functions, i.e. functions which can be described by persistence diagrams, well beyond the usual categories of topological spaces and simplicial complexes.

We have stressed the connection of these functions and diagrams with the natural pseudodistance, because we think that dissimilarity plays an essential role in all aspects of data analysis.

We have given some examples of this type of functions for weighted graphs in two ways: through construction of simplicial complexes and computation of their homology, and directly by using purely graph-theoretical concepts (blocks and edge-blocks).

We hope that this work paves the road to new applications of the persistence paradigm in various fields. We list a few possible developments by our team and hopefully by other researchers.

### 5.1 Topological graph persistence

After having defined the persistence functions of Sect. 3, it is important to study the information they convey in graph-theoretical terms.

We want to investigate further which complexes can be obtained from the considered constructions and which cannot.

Even in the cases when every sequence of Betti numbers can be produced by a certain construction (as is the case of Cor. 1), we need to know whether the natural pseudodistance between the resulting spaces coincides with the one between graphs, if we want to get an optimality (or universality) result like [26, Thm. 5.5].

More structures in a graph respect the inheritance property, necessary for building a simplicial complex. We intend to proceed in examining them within the proposed framework.

A possible connection with `Mapper` [37] deserves attention.

### 5.2 Graph-theoretical persistence

$k$-connected and $k$-edge-connected "components" are our next case of coherent covering of a graph to be studied for persistence.

Literally hundreds of graph-theoretical invariants can be explored for the possible production of persistence functions.

So far, we have only used $\mathbb{Z}$ as a range for persistence functions. In the comparison of two sublevel sets in a weighted graph, we could get something more interesting than an integer. We could have polynomials, e.g.: We are thinking of the chromatic polynomial, although the well-known recursive formula is likely to make this choice less dense of information. The connection with the study of hyperplane arrangements [41] makes this a worthy study anyway.

### 5.3 Beyond graphs

We believe that the persistence functions just defined can shed new light on hot research fields, e.g. social networks and neural nets.

So far we have just considered $\mathbb{R}$ as a parameter for filtrations, but there has been much progress in the study of filtering functions with $\mathbb{R}^k$ [19,6,5] or even $\mathbb{S}^1$ [4] as a range, and of spaces parametrized by a lattice [39]. The definition of persistence functions should be extended to these settings.

Persistence diagrams are but a shadow of much more general and powerful tools: persistence modules and further [3,26,11], on which the *interleaving distance* plays a central role. It is necessary to connect the ideas of the present paper to that research domain.

Our boldest hope is that researchers, working in areas which are far from topology and graph theory, find our setting useful for introducing persistence in their fields.

## Acknowledgments

We are indebted to Diego Alberici, Emanuele Mingione, Pierluigi Contucci (whose research originated the present one), Luca Moci, Fabrizio Caselli, Patrizio Frosini and Pietro Vertechi for many fruitful discussions. Article written within the activity of INdAM-GNSAGA.

## References


1. D. Alberici, P. Contucci, E. Mingione, and M. Molari. Aggregation models on hypergraphs. *Annals of Physics*, 376:412–424, 2017.
2. A. Bondy and U. Murty. *Graph Theory*. Graduate Texts in Mathematics. Springer London, 2011.
3. P. Bubenik and J. A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.
4. D. Burghelea and T. K. Dey. Topological persistence for circle-valued maps. *Discrete & Computational Geometry*, 50(1):69–98, 2013.
5. F. Cagliari, B. Di Fabio, and M. Ferri. One-dimensional reduction of multidimensional persistent homology. *Proceedings of the American Mathematical Society*, 138(8):3003–3017, 2010.
6. G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. In J. Erickson, editor, *Symposium on Computational Geometry*, pages 184–193. ACM, 2007.
7. F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *SCG '09: Proceedings of the 25th annual symposium on Computational geometry*, pages 237–246, New York, NY, USA, 2009. ACM.
8. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discr.Comput. Geom.*, 37(1):103–120, 2007.
9. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
10. M. d'Amico, P. Frosini, and C. Landi. Natural pseudo-distance and optimal matching between reduced size functions. *Acta Applicandae Mathematicae*, 109(2):527–554, 2010.
11. V. de Silva, E. Munch, and A. Stefanou. Theory of interleavings on $[0, \infty)$-actegories. *arXiv preprint arXiv:1706.04095*, 2017.
12. M. Dehmer, F. Emmert-Streib, S. Pickl, and A. Holzinger. *Big Data of Complex Networks*. CRC Press, 2016.
13. H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discr. Comput. Geom.*, 28(4):511–533, 2002.
14. M. Ferri, S. Gallina, E. Porcellini, and M. Serena. On-line character and writer recognition by size functions and fuzzy logic. *Proc. ACCV95*, pages 5–8, 1995.
15. M. Ferri, S. Lombardini, and C. Pallotti. Leukocyte classifications by size functions. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 223–229. IEEE, 1994.
16. P. Frosini. Measuring shapes by size functions. In *Intelligent Robots and Computer Vision X: Algorithms and Techniques, Boston, November 01, 1991*, pages 122–133. International Society for Optics and Photonics, 1992.
17. P. Frosini and C. Landi. Size functions and morphological transformations. *Acta Applicandae Mathematicae*, 49(1):85–104, 1997.
18. P. Frosini and C. Landi. Size functions and formal series. *Appl. Algebra Engrg. Comm. Comput.*, 12(4):327–349, 2001.
19. P. Frosini and M. Mulazzani. Size homotopy groups for computation of natural size distances. *Bull. of the Belg. Math. Soc.*, 6(3):455–464, 1999.
20. A. Hatcher. *Algebraic Topology*. Algebraic Topology. Cambridge University Press, 2002.

21. T. W. Haynes, S. Hedetniemi, and P. Slater. *Fundamentals of domination in graphs*. CRC Press, 1998.
22. D. Horak, S. Maletić, and M. Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, 2009.
23. W. Huang and A. Ribeiro. Persistent homology lower bounds on high-order network distances. *IEEE Transactions on Signal Processing*, 65(2):319–334, 2017.
24. J. Jonsson. *Simplicial complexes of graphs*, volume 3. Springer, 2008.
25. J. Lee, H. Kim, J. Lee, and S. Yoon. Transfer learning for deep learning on graph-structured data. In *AAAI*, pages 2154–2160, 2017.
26. M. Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, pages 1–38, 2015.
27. L. Lovász. Kneser's conjecture, chromatic number, and homotopy. *Journal of Combinatorial Theory, Series A*, 25(3):319–324, 1978.
28. S. Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.
29. S. Maletić, M. Rajković, and D. Vasiljević. Simplicial complexes of networks and their statistical properties. In *International Conference on Computational Science*, pages 568–575. Springer, 2008.
30. S. Maletić, Y. Zhao, and M. Rajković. Persistent topological features of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(5):053105, 2016.
31. S. Pal, T. J. Moore, R. Ramanathan, and A. Swami. Comparative topological signatures of growing collaboration networks. In *Workshop on Complex Networks CompleNet*, pages 201–209. Springer, 2017.
32. G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.
33. F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, s2-30(1):264–286, 1930.
34. M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11:48, 2017.
35. V. Robins. Towards computing homology from finite approximations. *Topology Proceedings*, 24(1):503–532, 1999.
36. N. A. Scoville and K. Yegnesh. A persistent homological analysis of network data flow malfunctions. *Journal of Complex Networks*, page cnx038, 2017.
37. G. Singh, F. Mémoli, and G. E. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100, 2007.
38. A. Sizemore, C. Giusti, R. F. Betzel, and D. S. Bassett. Closures and cavities in the human connectome. *arXiv preprint arXiv:1608.03520*, 2016.
39. P. Škraba and J. P. Costa. A lattice for persistence. *arXiv preprint arXiv:1307.4192*, 2013.
40. E. H. Spanier. *Algebraic topology*, volume 55. Springer Science & Business Media, 1994.
41. R. P. Stanley et al. An introduction to hyperplane arrangements. *Geometric combinatorics*, 13:389–496, 2004.
42. A. Verri, C. Uras, P. Frosini, and M. Ferri. On the use of size functions for shape analysis. *Biol. Cybern.*, 70:99–107, 1993.