Data and homotopy types

J.F. Jardine

August 17, 2019

Abstract

This paper presents explicit assumptions for the existence of interleaving homotopy equivalences of both Vietoris-Rips and Lesnick complexes associated to an inclusion of data sets. Consequences of these assumptions are investigated on the space level, and for corresponding hierarchies of clusters and their sub-posets of branch points. Hierarchy posets and branch point posets admit a calculus of least upper bounds, which is used to show that the map of branch points associated to the inclusion of data sets is a controlled homotopy equivalence.

Introduction

This paper is a discussion of homotopy theoretic phenomena that arise in connection with inclusions $X \subset Y \subset \mathbb{R}^n$ of data sets in topological data analysis. The manuscript is not in final form, and comments are welcome.

Suppose that $r > 0$, and say that X is r-dense in Y if for every point y in Y there is an x in X such that the distance $d(y, x) < r$ in the ambient metric space.

The first result in this direction (Corollary 2 below) implies that, under the r-density assumption, the induced map $V_s(X) \to V_s(Y)$ of Vietoris-Rips complexes is a homotopy equivalence if $2r < t - s$, where t is the smallest number such that $V_t(X) \neq V_s(X)$.

This result can be extended to the assertion that the inclusion $L_{s,k}(X) \to L_{s,k}(Y)$ of Lesnick complexes (with fixed density parameter k) is a homotopy equivalence if every point y in the "configuration space" Y^{k+1}_{dis} of $k+1$ distinct elements of Y has an element $x \in X^{k+1}_{dis}$ such that $d(y, x) < r$ in $\mathbb{R}^{n(k+1)}$. This result appears as Corollary 4 in this paper.

Corollary 2 and Corollary 4 follow from Theorem 1 and Theorem 3, respectively. Both theorems are "interleaving" homotopy type results (see also [1]) that follow from the respective r-density assumptions, by methods that amount to manipulation of barycentric subdivisions. Theorem 1 is a special case of Theorem 3 (it is the case $k = 0$), but Theorem 1 was found initially, and its proof has a certain clarity in isolation.

For a fixed k, the sets $\pi_0 L_{s,k}(X)$ of path components, as s varies, define a tree $\Gamma_k(X)$ with elements $(s,[x])$, $[x] \in \pi_0 L_{s,k}(X)$. The inclusion $X \subset Y$ defines a poset morphism $\Gamma_k(X) \to \Gamma_k(Y)$.

The tree $\Gamma_k(X)$ is the object studied by the HDBSCAN clustering algorithm, while the individual sets of clusters $\pi_0 L_{s,k}(X)$ are the objects of interest for the DBSCAN algorithm.

The poset $\Gamma_k(X)$ has a subobject $\mathrm{Br}_k(X)$ whose elements are the branch points of $\Gamma_k(X)$, suitably defined — see Section 2. The branch points of $\Gamma_k(X)$ are in one to one correspondence with the stable components for $\Gamma_k(X)$ that are defined in [3], in the sense that every such stable component starts at a unique branch point. Thus, we can (and do) replace the stable component discussion of [3] with the branch point poset $\mathrm{Br}_k(X)$, and make particular use of its ordering.

The tree $\Gamma_k(X)$ has least upper bounds, and these restrict to least upper bounds for the subobject $\mathrm{Br}_k(X)$ of branch points. This notion of least upper bounds is an extension of and a potential replacement for the distance function that is introduced by Carlsson and Mémoli [2] in their description of the ultrametric structure on a data set X that arises from the single linkage cluster hierarchy. The Carlsson-Mémoli theory does not apply in general to the tree $\Gamma_k(X)$, because the vertex sets of the Lesnick complexes $L_{k,s}(X)$ vary with changes of the parameter s.

The calculus of least upper bounds and its relation with branch points is described in Lemmas 7–12 below.

The branch point tree $\mathrm{Br}_k(X)$ can be thought of as a highly compressed version of the hierarchy $\Gamma_k(X)$ that is produced by the HDBSCAN algorithm.

The inclusion $\mathrm{Br}_k(X) \subset \Gamma_k(X)$ is a homotopy equivalence of posets, where the homotopy inverse is defined by taking the maximal branch point $(s_0,[x_0]) \leq (s,[x])$ below $(s,[x])$ for each object of $\Gamma_k(X)$. The existence of the maximal branch point below an object $(s,[x])$ is a consequence of Lemma 12.

The poset map $\Gamma_k(X) \to \Gamma_k(Y)$ defines a poset map $i_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(Y)$, via the homotopy equivalences for the data sets X and Y of the last paragraph.

The configuration space r-density assumption for Theorem 3 imples that there is a poset morphism $\theta_* : \mathrm{Br}_k(Y) \to \mathrm{Br}_k(X)$ that is induced by a morphism $(s,[x]) \to (s+2r,[\theta(x)])$, and that there are homotopies of poset maps $i_* \cdot \theta_* \simeq s_*$ and $\theta_* \cdot i_* \simeq s_*$, where s_* is defined by the shift operator $(s,[x]) \mapsto (s+2r,[x])$ on Y and X, respectively. These homotopies are "bounded" (or controlled) by the number $2r$.

In good circumstances, a branch point $(s,[x])$ is the maximal branch point below the shift $(s+2r,[x])$, and the inequalities defining the homotopies of the last paragraph become equalities in that case.

In summary, the poset morphism $\mathrm{Br}_k(X) \to \mathrm{Br}_k(Y)$ that is induced by an inclusion of data sets $X \subset Y$ has a homotopy theoretic character, and is measurably close to a homotopy equivalence if every sufficiently large group of distinct points of Y is close to a corresponding group of distinct points for the smaller data set X. Such a statement amounts to a stability result for hierarchies of branch points, albeit not in traditional terms.

Contents

1 Homotopy types

Suppose given (finite) data sets $X \subset Y \subset \mathbb{R}^n$. Suppose that $r > 0$.

Say that X is r-**dense** in Y, if for all $y \in Y$ there is an $x \in X$ such that $d(y, x) < r$.

Suppose that $s \geq 0$. Recall that $V_s(X)$ is the simplicial complex with simplices $(x_0, \dots, x_n)$ with $x_i \in X$ and $d(x_i, x_j) \leq s$.

Then we have

$$X = V_0(X) \subset V_s(X) \subset V_t(X) \subset \cdots \subset V_R(X) = \Delta^X$$

for $0 \leq s < t \leq R$, and for R sufficiently large, where $\Delta^X := \Delta^N$ and $N = |X| - 1$.

The inclusion $i : X \subset Y$ induces a map of systems of simplicial complexes $i : V_s(X) \subset V_s(Y)$.

The data sets X and Y are finite, so there is a finite string of parameter values

$$0 = s_0 < s_1 < \cdots < s_r,$$

consisting of the distances between elements of Y. This includes the list of distances between elements of X. I say that the s_i are the **phase-change** numbers.

Theorem 1. *Suppose that $X \subset Y \subset \mathbb{R}^n$, and that X is r-dense in Y.Then there is a homotopy commutative diagram*

$$\begin{array}{ccc} V_s(X) & \xrightarrow{j} & V_{s+2r}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ V_s(Y) & \xrightarrow[j]{} & V_{s+2r}(Y) \end{array}$$

in which the upper triangle commutes.

Proof. Define a function $\theta : Y \to X$ by specifying $\theta(x) = x$ for $x \in X$. For $y \in Y - X$ find $x \in X$ such that $d(x, y) < r$ and set $\theta(y) = x$.

If $(y_0, \dots, y_n)$ is a simplex of $V_s(Y)$,

$$d(\theta(y_i), \theta(y_j)) \leq d(\theta(y_i), y_i) + d(y_i, y_j) + d(y_j, \theta(y_j)) \leq r + s + r.$$

It follows that θ induces a simplicial complex map $\theta : V_s(Y) \to V_{s+2r}(X)$, such that the upper triangle commutes.

For the simplex $(y_0, \ldots, y_n)$ of $V_s(Y)$, the string of elements

$$(y_0, \ldots, y_n, \theta(y_0), \ldots, \theta(y_n))$$

defines a simplex of $V_{s+2r}(Y)$, since

$$d(y_i, \theta(y_j) \leq d(y_i, y_j) + d(y_j, \theta(y_j) \leq s + r.$$

Set

$$\gamma(y_0, \ldots, y_n) = (y_0, \ldots, y_n, \theta(y_0), \ldots, \theta(y_n)).$$

This assignment defines a morphism $\gamma : NV_s(Y) \to NV_{s+2r}(Y)$ of posets of non-degenerate simplices, and there are homotopies (natural transformations)

$$j \to \gamma \leftarrow i \cdot \theta$$

which are defined by face inclusions. □

Corollary 2. *Suppose that $X \subset Y \subset \mathbb{R}^n$, and that X is r-dense in Y. Suppose that $2r < s_{i+1} - s_i$. Then the map*

$$i : V_{s_i}(X) \to V_{s_i}(Y)$$

is a weak homotopy equivalence.

Proof. In the homotopy commutative diagram

$$\begin{array}{ccc} V_{s_i}(X) & \xrightarrow{j} & V_{s_i+2r}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ V_{s_i}(Y) & \xrightarrow[j]{} & V_{s_i+2r}(Y) \end{array}$$

the horizontal morphisms j are isomorphisms (identities), and so $V_{s_i}(X)$ is a deformation retract of $V_{s_i}(Y)$. □

Corollary 2 has consequences for both persistent homology and clustering.

Suppose that k is a non-negative integer. The Lesnick subcomplex $L_{s,k}(X)$ is the full subcomplex of $V_s(X)$ on those vertices x for which there are at least k distinct vertices $x_i \neq x$ such that $d(x, x_i) \leq s$ [5], [3], [4].

A simplex $\sigma = (y_0, \ldots, y_n)$ of $V_s(X)$ is in $L_{s,k}(X)$ if and only if each vertex y_i has at least k distinct neighbours in $V_s(X)$ — this is the meaning of the assertion that $L_{s,k}(X)$ is a full subcomplex of $V_s(X)$.

There is an array of subcomplexes

$$\begin{array}{ccc} V_s(X) & \longrightarrow & V_t(X) \\ \uparrow & & \uparrow \\ L_{s,k}(X) & \longrightarrow & L_{t,k}(X) \\ \uparrow & & \uparrow \\ L_{s,k+1}(X) & \longrightarrow & L_{t,k+1}(X) \end{array}$$

We have the following observations:

1) $L_{s,0}(X) = V_s(X)$.

2) $L_{s,k}(X)$ could be empty for small s and large k. In general, for $s \leq t$, $L_{s,k}(X)$ and $L_{t,k}(X)$ may not have the same vertices.

3) Every inclusion $i : X \subset Y \subset \mathbb{R}^n$ induces maps $i : L_{s,k}(X) \to L_{s,k}(Y)$ which are natural in s and k.

I say that s is a **spatial parameter** and that k is a **density parameter** (also a **valence**, or degree). Lesnick says [4] that $\{L_{s,k}(X)\}$ is the **degree Rips filtration** of the system $\{V_s(X)\}$.

Write X^{k+1}_{dis} for the set of $k+1$ distinct points of X, and think of it as a subobject of $(\mathbb{R}^n)^{k+1}$.

Theorem 3. *Suppose that $X \subset Y \subset \mathbb{R}^n$ and that X^{k+1}_{dis} is r-dense in Y^{k+1}_{dis} and that $L_{s,k}(Y) \neq \emptyset$. Then there is a homotopy commutative diagram*

$$\begin{array}{ccc} L_{s,k}(X) & \xrightarrow{j} & L_{s+2r,k}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ L_{s,k}(Y) & \xrightarrow[j]{} & L_{s+2r,k}(Y) \end{array}$$

in which the upper triangle commutes in the usual sense.

Proof. Suppose that $y \in L_{s,k}(X)_0 - L_{s,k}(X)_0$. Then there are k points $y_1, \dots, y_k$ of Y, distinct from y such that $d(y, y_i) < s$. There is a $(k+1)$-tuple $(x_0, x_1, \dots, x_k)$ such that

$$d((x_0, \dots, x_k), (y, y_1, \dots, y_k)) < r,$$

by assumption. Then $d(y, x_0) < r$, $d(y_i, x_i) < r$, and so $d(x_0, x_i) < s + 2r$, so that $x_0 \in L_{s+2r,k}(X)$. Set $\theta(y) = x_0$, and observe that $d(y, \theta(y)) < r$.

If $(y_0, \dots, y_p)$ is a simplex of $L_{s,k}(Y)$ then $(\theta(y_0), \dots, \theta(y_p)$ is a simplex of $L_{s+2r,k}(X)$, as is the string

$$(y_0, \dots, y_p, \theta(y_0), \dots, \theta(y_p)).$$

Finish according to the method of proof for Theorem 1. □

Corollary 4. *Suppose that $X \subset Y \subset \mathbb{R}^n$ and that X^{k+1}_{dis} is r-dense in Y^{k+1}_{dis}. Suppose that $2r < s_{i+1} - s_i$. Then the inclusion $i : L_{s_i,k}(X) \to L_{s_i,k}(Y)$ is a weak homotopy equivalence.*

Lemma 5. *Suppose that X^{k+1}_{dis} is r-dense in Y^{k+1}_{dis}, and that $Y^{k+1}_{dis} \neq \emptyset$. Then X^k_{dis} is r-dense in Y^k_{dis}.*

Proof. Suppose that $\{y_0, \dots, y_{k-1}\}$ is a set of k distinct points of Y. Then there is a $y_k \in Y$ which is distinct from the y_i, so that $(y_0, y_1, \dots, y_k)$ is a $(k+1)$-tuple of distinct points of Y. There is a $(k+1)$-tuple $(x_0, \dots, x_k)$ of distinct points of X such that

$$d((y_0, \dots, y_{k-1}, y_k), (x_0, \dots, x_{k-1}, x_k)) < r.$$

It follows that

$$d((y_0, \dots, y_{k-1}), (x_0, \dots, x_{k-1})) < r.$$

□

Corollary 6. *Suppose that $X \subset Y \subset \mathbb{R}^n$ and that X^{k+1}_{dis} is r-dense in Y^{k+1}_{dis}. Suppose that $2r < s_{i+1} - s_i$. Then the inclusion $i : L_{s_i,p}(X) \to L_{s_i,p}(Y)$ is a weak homotopy equivalence for $0 \leq p \leq k$.*

2 Branch points and upper bounds

Fix the density k and suppose that $L_{s,k}(X) \neq \emptyset$ for s sufficiently large. Apply the path component functor to the $L_{s,k}(X)$, to get a diagram of functions

$$\dots \to \pi_0 L_{s,k}(X) \to \pi_0 L_{t,k} \to \dots$$

There is a graph $\Gamma_k(X) := \Gamma(\pi_0 L_{*,k}(X))$ with vertices $(s, [x])$ with $[x] \in \pi_0 L_{s,k}(X)$, and edges $(s, [x]) \to (t, [x])$ with $s \leq t$. This graph underlies a contractible poset, and is therefore a tree (or hierarchy).

To reflect the poset structure of $\Gamma_k(X)$, I write the morphisms of $\Gamma_k(X)$ as relations $(s, [x]) \leq (t, [y])$. The existence of such a relation means that $[x] = [y] \in \pi_0 L_{t,k}(X)$, or that the image of $[x] \in \pi_0 L_{s,k}(X)$ under the induced function $\pi_0 L_{s,k}(X) \to \pi_0 L_{t,k}(X)$ is $[y]$.

Remarks: 1) Partitions of X given by the set $\pi_0 V_s(X)$ are standard **clusters**. The tree $\Gamma_0(X) = \Gamma(V_*(X))$ defines a **hierarchical clustering** (similar to, but not the same as single linkage clustering).

2) The set $\pi_0 L_{s,k}(X)$ gives a partitioning of the set of elements of X having at least k neighbours of distance $\leq s$, which is the subject of the **DBSCAN** algorithm. The tree $\Gamma_k(X) = \Gamma(\pi_0 L_{*,k}(X))$ is the basis of the **HDBSCAN** algorithm.

A *branch point* in the tree $\Gamma_k(X)$ is a vertex $(t, [x])$ such that either of following two conditions hold:

1) there is an $s_0 < t$ such that for all $s_0 \leq s < t$ there are distinct vertices $(s, [x_0])$ and $(s, [x_1])$ with $(s, [x_0]) \leq (t, [x])$ and $(s, [x_1]) \leq (t, [x])$, or

2) there is no relation $(s, [y]) \leq (t, [x])$ with $s < t$.

The second condition means that a representing vertex x is not a vertex of $L_{s,k}(X)$ for $s < t$. Write $\mathrm{Br}_k(X)$ for the set of branch points $(s, [x])$ in $\Gamma_k(X)$.

Every branch point $(s, [x])$ of $\Gamma_k(X)$ has $s = s_i$, where s_i is a phase change number for X.

The branch point poset $\mathrm{Br}_k(X)$ is a tree, because the element $(s_r, [x])$ corresponding to the highest phase change number s_r is maximal.

The set of branch points $\mathrm{Br}_k(X)$ inherits a partial ordering from the poset $\Gamma_k(X)$, and the inclusion $\mathrm{Br}_k(X) \subset \Gamma_k(X)$ of the set of branch points defines a monomorphism of trees.

Suppose that $(s, [x])$ and $(t, [y])$ are vertices of the graph $\Gamma_k(X)$. There is a vertex $(v, [w])$ such that $(s, [x]) \leq (v, [w])$ and $(t, [y]) \leq (v, [w])$. The two relations mean, among other things, that $[x] = [z] = [y]$ in $\pi_0 L_{v,k}(X)$.

It follows that there is a unique smallest vertex $(u, [z])$ which is an upper bound for both $(s, [x])$ and $(t, [y])$. The number u is the smallest parameter v such that $[x] = [y]$ in $\pi_0 L_{u,k}(X)$, and so $[z] = [x] = [y]$.

I write

$$(s, [x]) \cup (t, [y]) = (u, [z]).$$

The vertex $(u, [z])$ is the **least upper bound** (or join) of $(s, [x])$ and $(t, [y])$.

Every finite collection of points $(s_1, [x_1]), \dots, (s_p, [x_p])$ has a least upper bound

$$(s_1, [x_1]) \cup \dots \cup (s_p, [x_p])$$

in $\Gamma_k(X)$.

Lemma 7. *The least upper bound of branch points $(u, [z])$ of $(s, [x])$ and $(t, [y])$ is a branch point.*

Proof. For numbers $s, t < v < u$, $(v, [x])$ and $(v, [y])$ are distinct, because $(u, [z])$ is a least upper bound.

Otherwise, $s = u$ or $t = u$, in which case $(u, [z]) = (s, [x])$ or $(u, [z]) = (t, [y])$, respectively. $\square$

It follows from Lemma 7 that any two branch points $(s, [x])$ and $(t, [y])$ have a least upper bound in $\mathrm{Br}_k(X)$, and that the poset inclusion $\alpha : \mathrm{Br}_k(X) \to \Gamma_k(X)$ preserves least upper bounds.

Remark: The poset $\Gamma_k(X)$ also has greatest lower bounds (or meets). The greatest lower bound

$$(t_1, [y_1]) \cap \dots \cap (t_r, [y_r])$$

is the least upper bound of all $(s, [x])$ such that $(s, [x]) \leq (t_j, [y_j])$ for all j.

We have the following triviality:

Lemma 8. *Suppose that $(s_1, [x_1]), (s_2, [x_2])$ and $(s_3, [x_3])$ are vertices of $\Gamma_k(X)$. Then*

$$(s_1, [x_1]) \cup (s_3, [x_3]) \leq ((s_1, [x_1]) \cup (s_2, [x_2])) \cup ((s_2, [x_2]) \cup (s_3, [x_3])).$$

Proof. We have the identity

$$((s_1,[x_1])\cup(s_2,[x_2]))\cup((s_2,[x_2])\cup(s_3,[x_3])) = (s_1,[x_1])\cup(s_2,[x_2])\cup(s_3,[x_3]),$$

and then

$$(s_1,[x_1])\cup(s_3,[x_3]) \leq (s_1,[x_1])\cup(s_2,[x_2])\cup(s_3,[x_3]).$$

□

Carlsson and Mémoli [2] define an ultrametric d on $X = V_0(X)$, for which they say that $d(x,y) = s$, where s is the minimum parameter value such that $[x] = [y] \in \pi_0 V_s(X)$.

Suppose given $[x]$ and $[y]$ in $\pi_0 L_{s,k}(X)$ (equivalently, points $(s,[x])$ and $(s,[y])$ in $\Gamma_k(X)$). Write $d([x],[y]) = u - s$, where $(s,[x])\cup(s,[y]) = (u,[w])$.

Lemma 9. *Given $[x],[y]$ and $[z]$ in $\pi_0 L_{s,k}(X)$, we have a relation*

$$d([x],[z]) \leq \max\{d([x],[y]), d([y],[z])\}.$$

Proof. Suppose that $(s,[x])\cup(s,[y])\cup(s,[z]) = (v,[w])$. Then

$$v - s = \max\{d([x],[y]), d([y],[z])\}$$

and

$$d([x],[z]) \leq v - s$$

by Lemma 8. □

Corollary 10. *The function*

$$d : \pi_0 L_{s,k}(X) \times \pi_0 L_{s,k}(X) \to \mathbb{R}_{\geq 0}$$

of Lemma 9 gives the set $\pi_0 L_{s,k}(X)$ the structure of an ultrametric space.

Remark: One could define a "distance" function d on the full set of points of $\Gamma_k(X)$ by setting

$$d((s,[x]),(t,[y])) = \max\{u - s, u - t\},$$

where $(s,[x])\cup(t,[y]) = (u,[z])$.

The ultrametric property of Lemma 9 fails for the points $(s,[x]),(t,[x])$ and $(u,[x])$ where $s < t < u$, since it is not the case that $u - s \leq \max\{t - s, u - t\}$

Lemma 11. *Every vertex $(s,[x])$ of $\Gamma_k(X)$ has a unique largest branch point $(s_0,[x_0])$ such that $(s_0,[x_0]) \leq (s,[x])$.*

Proof. The least upper bound of the finite list of the branch points $(t,[y])$ such that $(t,[y]) \leq (s,[x])$ is a branch point, by Lemma 7. □

In the situation described by Lemma 11, I say that $(s_0,[x_0])$ is the **maximal branch point below** $(s,[x])$.

Lemma 12. *Suppose that* $(s_0,[x_0])$ *and* $(t_0,[y_0])$ *are maximal branch points below the points* $(s,[x])$ *and* $(t,[y])$*, respectively.*

Then $(s_0,[x_0])\cup(t_0,[y_0])$ *is the maximal branch point below* $(s,[x])\cup(t,[y])$*.*

Proof. Suppose that $s \leq t$.

We have

$$(s_0,[x_0])\cup(t_0,[y_0]) \leq (s,[x])\cup(t,[y]).$$

and $(s_0,[x_0])\cup(t_0,[y_0])$ is a branch point by Lemma 7. Write

$$(v,[z]) = (s_0,[x_0])\cup(t_0,[y_0]).$$

Suppose that $v \leq t$. Then $(t_0,[y_0]) \leq (t,[y])$ and $(t_0,[y_0]) \leq (v,[z])$, so that $(v,[z]) \leq (t,[y])$ since $v \leq t$. Also, $(s_0,[x_0]) \leq (s,[x])$ and $(s_0,[x_0]) \leq (v,[z]) \leq (t,[y])$ so that $(s,[x]) \leq (t,[y])$. Then $(s_0,[x_0]) \leq (t_0,[y_0])$ by maximality, and it follows that

$$(s_0,[x_0])\cup(t_0,[y_0]) = (t_0,[y_0])$$

is the maximal branch point of

$$(s,[x])\cup(t,[y]) = (t,[y])$$

Suppose that $v > t$. Then $(s,[x]) = (s,[x_0]) \leq (v,[z])$ and $(t,[y]) = (t,[y_0]) \leq (v,[z])$ because $s \leq t < v$, so that

$$(s,[x])\cup(t,[y]) \leq (s_0,[x_0])\cup(t_0,[y_0]),$$

Thus, $(s_0,[x_0])\cup(t_0,[y_0]) = (s,[x])\cup(t,[y])$ is a branch point, by Lemma 7. □

The poset inclusion $\alpha : \mathrm{Br}_k(X) \to \Gamma_k(X)$ has an inverse

$$max : \Gamma_k(X) \to \mathrm{Br}_k(X),$$

up to homotopy.

In effect, Lemma 11 implies that every vertex $(s,[x])$ of $\Gamma_k(X)$ has a unique maximal branch point $(s_0,[x_0])$ such that $(s_0,[x_0]) \leq (s,[x])$. Set

$$max(s,[x]) = (s_0,[x_0]).$$

The maximality condition implies that max preserves the ordering. The composite $max \cdot \alpha$ is the identity on $\mathrm{Br}_k(X)$, and the relations $(s_0,[x_0]) \leq (s,x)$ define a homotopy $max \cdot \alpha \leq 1$.

Return to the inclusion $i : X \subset Y \subset \mathbb{R}^n$ of finite data sets. Suppose that X^{k+1}_{dis} is r-dense in Y^{k+1}_{dis} and that $L_{s,k}(Y)$ is non-empty, as in the statement of Theorem 3.

Write $i_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(Y)$ for the composite

$$\mathrm{Br}_k(X) \xrightarrow{\alpha} \Gamma_k(X) \xrightarrow{i_*} \Gamma_k(Y) \xrightarrow{max} \mathrm{Br}_k(Y)$$

This map takes a branch point $(s, [x])$ to the maximal branch point below $(s, [i(x)])$. The map i_* preserves least upper bounds by Lemma 7.

Poset morphisms $\theta_* : \mathrm{Br}_k(Y) \to \mathrm{Br}_k(X)$ and $s_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(X)$ are similarly defined, by the poset morphism $\theta : \Gamma_k(Y) \to \Gamma_k(X)$ with $(t, [y]) \mapsto (t + 2r, [\theta(y)])$, and the shift morphism $s : \Gamma_k(X) \to \Gamma_k(X)$ with $(s, [x]) \mapsto (s + 2r, [x])$.

The construction of the poset map $i_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(Y)$ is not functorial in maps of the form $X \to Y$, but it is functorial up to coherent homotopy.

Similarly, the map $i_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(Y)$ only preserves least upper bounds up to homotopy. Suppose that $(s, [x])$ and $(t, [y])$ are branch points of X, and let $(s_0, [x_0]) \leq (s, [i(x)])$ and $(t_0, [y_0]) \leq (t, [i(y)])$ be maximal branch points below the images of $(s, [x])$ and $(t, [y])$ in $\Gamma_k(Y)$. Then

$$(s_0, [x_0]) \cup (t_0, [y_0]) \leq (s, [i(x)]) \cup (t, [i(y)]),$$

so that

$$i_*(s, [x]) \cup i_*(t, [y]) \leq i_*((s, [x]) \cup (t, [y])).$$

Similar inequalities hold for least upper bounds with respect to the other maps that one encounters, namely $\theta_* : \mathrm{Br}_k(Y) \to \mathrm{Br}_k(X)$ and the shift map $s_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(X)$.

1) Consider the poset maps

$$\mathrm{Br}_k(X) \xrightarrow{i_*} \mathrm{Br}_k(Y) \xrightarrow{\theta_*} \mathrm{Br}_k(X).$$

If $(s, [x])$ is a branch point for X, choose maximal branch points $(s_0, [x_0]) \leq (s, [i(x)]$ for Y, $(s_1, [x_1]) \leq (s_0 + 2r, [\theta(x_0)])$ and $(v, [y]) \leq (s + 2r, [x])$ below the respective objects.

Then $\theta_* i_*(s, [x]) = (s_1, [x_1])$, and there is a natural relation

$$\theta_* i_*(s, [x]) = (s_1, [x_1]) \leq (v, [y]) = s_*(s, [x]) \leq (s + 2r, [x]).$$

We therefore have a homotopy of poset maps

$$\theta_* i_* \leq s_* : \mathrm{Br}_k(X) \to \mathrm{Br}_k(X).$$

Note that $(s, [x]) \leq s_*(s, [x])$ since $(s, [x])$ is a branch point and $s_*(s, [x])$ is the maximal branch point below $(s + 2r, [x])$. This means that the shift morphism s_* is homotopic to the identity on $\mathrm{Br}_k(X)$.

The branch point $(s, [x])$ has a "close" shared upper bound $(s + 2r, [x])$ with the element $(s_0 + 2r, [\theta(x_0)])$, which is the image of the branch point $(s_0, [x_0])$ under the poset map $\theta_* : \Gamma_k(Y) \to \Gamma_k(X)$.

2) Similarly, if $(t, [y])$ is a branch point of Y, then

$$i_* \theta_*(t, [y]) \leq s_*(t, [y]) \geq (t, [y])$$

while $s_*(t, [y]) \leq (t + 2r, [y])$.

The element $(t + 2r, [y])$ is a close shared upper bound for $(t, [y])$ and an element of the form $(t_0, [i(y_0)])$, where $(t_0, [y_0])$ is a maximal branch point of X below $(s + 2r, [\theta(y)])$.

Remark: The subobject of $\mathrm{Br}_k(X)$ consisting of all branch points of the form $(s, [x])$ as s varies has an obvious notion of distance on it: the distance between points $(s, [x])$ and $(t, [x])$ is $|t - s|$. The closeness referred to in constructions 1) and 2) above can be expressed in terms of such a distance.

Suppose that $(s_1, [x_1])$ and $(s_2, [x_2])$ are branch points of X, let

$$(s, [x]) = (s_1, [x_1]) \cup (s_2, [x_2]),$$

and write

$$(t, [y]) = (s_1, [i(x_1)]) \cup (s_2, [i(x_2)]).$$

Then $(s, [i(x)])$ is an upper bound for $(s_1, [i(x_1)])$ and $(s_2, [i(x_2)])$, so that $(t, [y]) \leq (s, [i(x)])$.

The element $(t + 2r, [\theta(y)])$ is an upper bound for $(s_1, [x_1])$ and $(s_2, [x_2])$, so that

$$(s, [x]) \leq (t + 2r, [\theta(y)]) \leq (s + 2r, [x]). \tag{1}$$

It follows that $s - 2r \leq t \leq s$, which gives a constraint on the parameter t corresponding to the least upper bound $(t, [y])$ in $\Gamma_k(Y)$, in terms of the least upper bound $(s, [x])$ in $\Gamma_k(X)$, or in $\mathrm{Br}_k(X)$. The number $2r$ is a bound on the distances between the three points in (1).

If the bound $2r$ is sufficiently small, then $(s, [x])$ is the largest branch point below $(s + 2r, [x])$ and $s_*(s, [x]) = (s, [x])$ in that case. Similarly, in $\Gamma_k(Y)$, $s_*(t, [y]) = (t, [y])$ if $2r$ is sufficiently small.

Recall that if $(s, [x])$ is a branch point, then $s = s_i$ is one of the phase shift numbers. Then $(s_i, [x])$ is the maximal branch point below $(s_i + 2r, [x])$ if $2r < s_{i+1} - s_i$.

References

[1] Andrew J. Blumberg and Michael Lesnick. Universality of the homotopy interleaving distance. *CoRR*, abs/1705.01690, 2017.

[2] Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *J. Mach. Learn. Res.*, 11:1425–1470, 2010.

[3] J.F. Jardine. Stable components and layers. Preprint, 2019.

[4] M. Lesnick and M. Wright. RIVET: visualization and analysis of two-dimensional persistent homology. `http://rivet.online`, 2019.

[5] Leland McInnes and John Healy. Accelerated hierarchical density based clustering. In *2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017*, pages 33–42, 2017.