Stability for UMAP

J.F. Jardine*

Department of Mathematics
University of Western Ontario
London, Ontario, Canada

jardine@uwo.ca

November 26, 2020

Abstract

This paper displays the Healy-McInnes UMAP construction V(X,N) as an iterated ushout of Vietoirs-Rips objects $V(X, D_x)$, which are associated to extended pseudo metric spaces (ep-metric spaces) defined by a system N of neighbourhoods of the elements of a finite set X. An inclusion of finite sets $X \subset Y$ defines a map of UMAP systems $V(X, N) \to V(Y, N')$ in the presence of a compatible system of neighbourhoods N' for Y. There is also an induced map of ep-metric spaces $(X, D) \to (Y, D')$, where D and D' are colimits (global averages) of the metrics defined by the respective neighbourhood systems. We prove a stablity result for the restriction of this ep-metric space map to global components. This stability result translates, via excision for path components, to a stability result for global components of the UMAP systems.

The main result of [1] says that if X is a finite extended pseudo-metric space (ep-metric space), then the canonical map

$$\eta : V(X)_s \to S(X)_s$$

is a weak equivalence for all distance parameters s. Here, $V(X)$ is the Vietoris-Rips system and $X \mapsto S(X)$ is the singular functor.

In this paper, we use this result to model the UMAP construction, and we prove a stability result for the resulting hierarchies of clusters.

For the general program, we start with sets N_x (disjoint from x) for each $x \in X$, and distances (or weights) $d_x(x, y) \geq 0$ for all $y \in N_x$. These distances canonically extend to an ep-metric space structure (U_x, D_x) on the set

$$U_x = \{x\} \sqcup N_x,$$

*Supported by NSERC.

and then to an ep-metric space structure (X, D_x) on all of X, for which $D_x(y, z) = \infty$ unless both y and z are in U_x. The metric space structures (X, D_x) can be glued together along ep-metric space morphisms $(X, \infty) \to (X, D_x)$ to produce an ep-metric space

$$(X, D) = \vee_{x \in X}\ (X, D_x).$$

Similarly, the Vietoris-Rips systems $V(X, D_x)$ can be glued together along the maps $X \to V(X, D_x)$ to produce a system

$$V(X, N) = \vee_{x \in X}\ V(X, D_x).$$

The notation $V(X, N)$ reflects the fact that this system of spaces depends on the family $N = \{N_x,\ x \in X\}$ of neighbourhoods, which includes choices of weights d_x within each neighbourhood N_x.

The object $V(X, N)$, for suitable choices of neighbourhoods and weights, gives the various models for the UMAP system.

The original UMAP system $S(X, N)$ of Healy and McInnes [2], is constructed from Spivak's singular functor [3], [1], with

$$S(X, N) := \vee_{x \in X}\ S(X, D_x).$$

There is a sectionwise weak equivalence $V(X, N) \to S(X, N)$ by the main result of [1], and we use the Vietoris-Rips construction $V(X, N)$ since it is more familiar and easier to manipulate.

The choice of neighbourhood sets N_x can be arbitrary, but in [2] it is the set of k-nearest neighbours. The collections of distances $d_x(x, y)$ are also arbitrary, but are defined in [2], variously, as the original distance $d_x(x, y) = d(x, y)$ or the probability $d_x(x, y) = \frac{1}{r_x} d(x, y)$, or $d_x(x, y) = \frac{1}{r_x}(d(x, y) - s_x)$. Here, $r_x = \max_{y \in N_x} d(x, y)$ and $s_x = \min_{y \in N_x} d(x, y)$.

All corresponding constructions $V(X, N)$ are variants of the UMAP construction, and they are easily compared. The shrpest results on the general structure of $V(X, N)$ require the weights $d_x(x, y) > 0$ for $y \in N_x$, and this is assumed for most of the paper.

Suppose given an ep-metric space map $i : (X, d_X) \to (Y, d_Y)$, where the underlying function is an injection, and X and Y are finite. The assumption that i is an ep-metric space morphism means that i compresses distance in the sense that $d_Y(i(x), i(y)) \leq d_X(x, y)$ for all $x, y \in X$.

We can assume for now that X and Y are metric spaces, and are therefore *globally connected* in the sense that $d(x, y) < \infty$ for all x, y. In that case, since X is finite, we define the *compression factor* $m(i)$ by

$$m(i) = \max_{x \neq y}\ \frac{d_X(x, y)}{d_Y(i(x), i(y))}$$

This makes sense because none of the distances in the ratio are either 0 or ∞.

If we further assume that for every $y \in Y$ there is an $x \in X$ such that $d_Y(y, x) \leq r$, then the same argument as for the ordinary Rips stability theorem produces a homotopy interleaving

$$\begin{array}{ccc} V(X)_s & \xrightarrow{\sigma} & V(X)_{m(i)(s+2r)} \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ V(Y)_s & \xrightarrow[\sigma]{} & V(Y)_{m(i)(s+2r)} \end{array}$$

This statement appears as Proposition 8 in this paper.

Suppose now that $i : X \subset Y$ is an inclusion of finite sets, and we have made choice of neighbourhoods $N_x,\ x \in X$ and $N_y,\ y \in Y$. Suppose that

1) the inclusion i induces inclusions $i : N_x \subset N_{i(x)}$, and

2) the weights are chosen such that $d_x(x, x') > 0$ for $x' \neq x \in N_x$, $d_y(y, y') > 0$ for $y' \neq y \in N_y$, and $d_{i(x)}(i(x), i(y)) \leq d_x(x, y)$ for all $y \in N_x$.

The assumptions imply that the inclusion i induces an ep-metric space map $i : (X, D) \to (X, D')$, and the global connected components of both ep-metric spaces are metric spaces. If E is a global connected component of (X, D), then there is a global connected component F of (Y, D') such that i restricts of a ep-metric space morphism $i : (E, D) \to (F, D')$ of metric spaces.

Subject to the assumptions of the last paragraph, it follows from Proposition 8 that, if for every $y \in F$ there is an $x \in E$ such that $d_Y(y, i(x)) \leq r$, then there is a homotopy interleaving

$$\begin{array}{ccc} V(E,D)_s & \xrightarrow{\sigma} & V(E,D)_{m(i)(s+2r)} \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ V(F,D')_s & \xrightarrow[\sigma]{} & V(F,D')_{m(i)(s+2r)} \end{array} \qquad (1)$$

This is a componentwise stability result for the ep-metric space morphism $i : (X, D) \to (Y, D')$, which appears as Theorem 9 in this paper. The input for this result involves compatible choices of neighbourhoods and weights within those neighbourhoods, rather than distance.

The canonical ep-metric space maps $(X, D_x) \to (X, D)$ induce a map of systems

$$\phi : V(X, N) = \vee_{x \in X}\ V(X, D_x) \to V(X, D)$$

which is natural with respect to inclusions $i : X \subset Y$ satisfying the conditions above.

The ep-metric space (X, D) is a disjoint union of its global connected components E, and the system $V(X, D)$ is a disjoint union of the systems $V(E, D)$. This splitting defines a disjoint union structure

$$V(X, N) = \bigsqcup_E\ V(X, N)(E),$$

where $V(X,N)_E$ is the pullback of the system $V(E)$ under the map ϕ. The induced map

$$\phi_* : \pi_0 V(X,N)(E) \to \pi_0 V(E)$$

is an isomorphism of systems of sets, by path component excision (Lemma 2).

The componentwise stability result displayed in the interleaving (1) therefore specializes to interleavings in clusters

$$\begin{array}{ccc} \pi_0 V(X,N)(E)_s & \xrightarrow{\sigma} & \pi_0 V(X,N)(E)_{m(i)(s+2r)} \\ {\scriptstyle i}\downarrow & \nearrow^{\theta} & \downarrow{\scriptstyle i} \\ \pi_0 V(Y,N')(F)_s & \xrightarrow[\sigma]{} & \pi_0 V(Y,N')(F)_{m(i)(s+2r)} \end{array} \qquad (2)$$

This is a stability result for UMAP, which appears as Theorem 10 below. Theorem 10 is the main result of this paper.

Contents

1 General constructions

Suppose that we have a set X with a finite list of ep-metric space structures (X, d_i), $i = 1, \dots, k$. We can also endow X with a discrete ep-metric space structure, so that $d_\infty(x,y) = \infty$ for all $x, y \in X$. Suppose that X has a total ordering.

There are canonical ep-metric space morphisms $(X, d_\infty) \to (X, d_i)$, all of which are the identity on X. Write (X, D) for the colimit in $ep - \mathbf{Met}$, giving a diagram

$$\begin{array}{ccc} (X, d_\infty) & \longrightarrow & (X, d_i) \\ \downarrow & & \downarrow{\scriptstyle \tau_i} \\ (X, d_j) & \xrightarrow[\tau_j]{} & (X, D) \end{array}$$

The maps $\tau_i : (X, d_i) \to (X, D)$ are the canonical maps into the colimit. Recall [1] that the colimit (X, D) is formed by taking the colimit of the underlying functions, and endowing it with a metric, in this case D. The colimit of functions, which are identity functions on X, is X again, so that the notation (X, D) makes sense.

We also write

$$(X, D) = \vee_i \ (X, d_i)$$

to reflect the fact that we are gluing together the ep-metric spaces (X, d_i) along the underlying set X.

Formally,

$$D(x, y) = \inf_P \ (\sum \ d_{i_j}(x_j, x_{j+1})),$$

indexed over all polygonal paths

$$x = x_0, x_1, \dots, x_n = y$$

and choices of metrics d_{i_j} in the list d_i, $1 \leq i \leq k$. The pair x, y forms a polygonal path, so that

$$D(x, y) \leq d_i(x, y)$$

for all i. In this sense, the ep-metric D optimizes the metrics d_i.

There may not be a polygonal path P and metrics d_{i_j} such that all $d_{i_j}(x_i, x_{i+1})$ are finite. In that case, we have $D(x, y) = \infty$.

If X is a finite set, then the collection of polygonal paths from x to y in X is finite, and so

$$D(x, y) = \sum \ d_{i_j}(x_j, x_{j+1})$$

for some choice of polygonal path P and metrics d_{i_j}. In that case, $d_{i_j}(x_j, x_{j+1})$ must be minimal among all $d_k(x_j, x_{j+1})$.

The maps $(X, d_\infty) \to (X, d_i)$ induce maps $X \to V(X, d_i)$ into Vietoris-Rips systems, and we form the iterated pushout

$$V(X, D) := \vee_i \ V(X, d_i) \tag{3}$$

in the category of systems. This means that the object (3) is the colimit of all maps

$$X \to V(X, d_i), \tag{4}$$

over the discrete system X. The maps (4) are sectionwise monomorphisms, so the object $V(X, D)$ is a type of homotopy colimit.

Remark 1. In practice and in general, although one tends to be notationally lazy, it is better to replace the Vietoris-Rips system $s \mapsto V_s(X)$ with the homotopy equivalent system $s \mapsto BP_s(X)$, where $P_s(X)$ is the poset of non-degenerate simplices of $V_s(X)$, and $BP_s(X)$ is the nerve of $P_s(X)$. The poset $P_s(X)$ can be described explicitly as the collection of subsets σ of X such that $d(x, y) \leq s$ for all $x, y \in \sigma$. The structure of the poset $P_s(X)$ does not depend on an ordering of the set X.

The ep-metric space maps $(X, d_i) \to (X, D)$ induce commutative diagrams

$$\begin{array}{ccc} BP(X, d_i) & \longrightarrow & BP(X, D) \\ {\scriptstyle \gamma} \downarrow {\scriptstyle \simeq} & & \downarrow {\scriptstyle \gamma_*} \\ V(X, d_i) & \longrightarrow & V(X, D) \end{array}$$

of maps of systems, where the map γ is a sectionwise weak equivalence defined by subdivision, and the induced map γ_* is a sectionwise weak equivalence arising from the displayed comparison of homotopy colimits.

From this perspective, we can write

$$V(X, D) = BP(X, D) = \vee_i\ BP(X, d_i) = \vee_i\ V(X, d_i)$$

as sectionwise homotopy types.

Lemma 2 (Excision)**.** *Suppose that X is a finite set, with a finite collection of ep-metric structures d_i.*

Then the canononical map

$$\phi : \vee_i\ V(X, d_i) \to V(X, D)$$

induces bijections

$$\phi_* : \pi_0(\vee_i\ V(X, d_i))_s \xrightarrow{\cong} \pi_0 V(X, D)_s$$

for all s.

Proof. The map ϕ is the identity on vertices, so that ϕ_* is surjective.

Suppose that $D(x, y) \leq s$ in (X, D). There is a polygonal path

$$P : x = x_0, x_1, \dots, x_n = y$$

and metrics d_{i_j} such that

$$D(x, y) = \sum_j\ d_{i_j}(x_j, x_{j+1}) \leq s,$$

since X is finite. This means that $d_{i_j}(x_j, x_{j+1}) \leq s$ for all j, and so there are 1-simplices (x_j, x_{j+1}) in $V(X, d_{i_j})_s$ which together describe a path from x to y in $\vee_X\ V_s(X, d_i)$.

It follows that, if x, y are in the same path component of $V(X, D)_s$, then x, y are in the same path component of $\vee_i V(X, d_i)_s$. □

2 UMAP

The UMAP algorithm of [2] starts with a finite metric space X. We assume that X has a total ordering.

For each point $x \in X$ one finds the list

$$N_x := \{x_1, \dots, x_k\}$$

of distinct k-nearest neighbours with $x_i \neq x$, with maximum distance $r_x = \max_i\ d(x, x_i)$.

The set N_x is the set of *neighbours* of x.

In much of what follows, the choices of the sets N_x can be quite arbitrary. In all applications, one assigns distances $d_x(x,y)$ for all neighbours $y \in N_x$, and then one extends functorially to an ep-metric D_x on X. This is done for all $x \in X$.

Examples: Possibilities for $d_x(x,y)$ include $\frac{1}{r_x}d(x,y)$, $\frac{1}{r_x}(d(x,) - \eta_x)$ where η_x is the distance from x to a nearest neighbour. We can also use the ambient metric $d_x(x,y) = d(x,y)$ from X.

Remark 3. Explicitly, given $x \in X$ we find a set (and a listing) $N_x = \{x_1, \dots, x_k\}$ of k-nearest neighbours, by finding an element x_1 (in the total order) such that $d(x,x_1)$ is minimal (x_1 is a nearest neighbour). Then $x_2 \in X - \{x, x_1\}$ is chosen such that $d(x,x_2)$ is minimal and x_2 is the first element in the total order that has this property, and so on.

The algorithm is set up such that the sublist $\{x_i, x_{i+1}, \dots, x_k\}$ of elements having $d(x,x_j) = r$ has $x_i < x_{i+1} < \cdots < x_k$ in the total order.

Assumptions: Suppose that X is a finite set. Suppose given a system of neighbourhoods N_x for $x \in X$, and define distances $d_x(x,y) > 0$ for each $y \in N_x$.

One defines an ep-metric D_x first on the set

$$U_x = \{x\} \sqcup N_x$$

and then one extends to all of X with the decomposition

$$X = U_x \sqcup \Big(\bigsqcup_{y \in X - U_x} \{y\} \Big). \tag{5}$$

The ep-metric space structure on the set U_x is given by the wedge

$$(U_x, D_x) = \vee_{y \in N_x} \ (\{x,y\}, d_x)$$

over x of the 2-element metric spaces $(\{x,y\}, d_x)$, in the category of ep-metric spaces. The metric D_x on U_x has the property that $D_x(x,y) = d_x(x,y)$ for $y \in N_x$. The triangle inequality forces

$$D_x(y,z) \leq d_x(y,x) + d_x(x,z)$$

for $y \neq z$ in N_x. At the same time, the sum $d_x(y,x) + d_x(x,z)$ is the length of the shortest polygonal path (y,x,z) between y and z in U_x, and so it follows that

$$D_x(y,z) = d_x(y,x) + d_x(x,z)$$

for $y \neq z$ in N_x.

Use the decomposition (5) to extend D_x to an ep-metric on all of X. This forces $D_x(y,z) = \infty$ unless y and z are both in U_x.

Define systems of simplicial sets $V(X,N)$ and $S(X,N)$ by setting

$$V(X,N) = \vee_{x\in X} V(X,D_x)$$

and

$$S(X,N) = \vee_{x\in X} S(X,D_x),$$

respectively. Here, $V(X,N)$ is the iterated pushout of the cofibrations $X \to V(X,D_x)$, where the set X is identified with a constant, discrete system. Similarly, $S(X,N)$ is the iterated pushout of the cofibrations $X \to S(X,D_x)$.

The maps $\eta : V(X,D_x) \to S(X,D_x)$ are sectionwise weak equivalences by [1], and therefore induce a sectionwise weak equivalence

$$\eta : V(X,N) \to S(X,N) \tag{6}$$

by comparison of iterated pushouts (or homotopy colimits).

Spivak's realization construction Re preserves colimits, and there is a natural isomorphism $\mathrm{Re}(V(X,D_x)) \cong (X,D_x)$ (see [1]), so that the realization

$$\mathrm{Re}(V(X,D)) \cong \vee_X \ (X,D_x) = (X,D)$$

is the iterated pushout of the maps $X \to (X,D_x)$ in the ep-metric space category, as in the first section.

Remark 4. The set X is finite. The distances d_x have the property that $d_x(x,y) > 0$ for all $y \in N_x$, $x \in X$, and one can show that $D(u,v) = 0$ in (X,D) forces $u = v$.

In effect, $D(u,v)$ is a sum

$$D(u,v) = \sum D_{z_i}(x_i,x_{i+1})$$

which is defined by a particular polygonal path $P : u = x_0,\dots,x_n = v$ (since there are only finitely many such paths). Then $D(u,v) = 0$ forces all

$$D_{z_i}(x_i,x_{i+1}) = d_{z_i}(x_i,z_i) + d_{z_i}(z_i,x_{i+1})$$

to be 0, so that $x_i = z_i = x_{i+1}$ for all i, and $u = v$.

Remark 5. It is time for a homotopy theory interlude.

Suppose that each map

$$V = \{0,1,\dots,n\} \subset \Delta^n = X_i,\ i \geq 0,$$

is the inclusion of the set of vertices V of the standard n-simplex Δ^n, and let $Y_k = X_0 \cup \dots \cup X_k$ be an iterated pushout of n-simplices over the common vertex set V.

There is a pushout diagram

$$\begin{array}{ccc} V & \longrightarrow & X_1 \\ \downarrow & & \downarrow \\ X_0 & \longrightarrow & X_0 \cup_V X_1 \end{array}$$

in which both X_0 and X_1 are contractible. It follows that $X_0 \cup_V X_1$ has the homotopy type of the suspension $X_1/V \simeq \Sigma V$ for a suitable choice of base point of the discrete set V — choose 0. Then $V = \{0,1\} \vee \{0,2\} \vee \cdots \vee \{0,n\}$ is a wedge of n copies of S^0, and ΣV is a wedge of n copies of $\Sigma S^0 = S^1$. Thus,

$$Y_1 = X_0 \cup_V X_1 \simeq S^1 \vee \cdots \vee S^1 \ \ (n \text{ summands}).$$

More generally, consider the space $Y_k = X_0 \cup_V X_1 \cup_V \cdots \cup_V X_k$. Collapsing the contractible space X_0 to a point gives

$$Y_k \simeq (X_1/V) \vee (X_2/V) \vee \cdots \vee (X_k/V).$$

Each X_i/V is an n-fold wedge of circles, by the above, so that Y_k is a $k \cdot n$-fold wedge of circles.

Suppose that the finite set X has $M+1$ elements. Then $V(X,D)_\infty$ is an iterated pushout of the maps $X \subset V(X,D_x)_\infty$. Each $V(X,D_x)_\infty$ is a copy of the M-simplex Δ^M, and each map $X \subset V(X,D_x)_\infty$ is a copy of the inclusion of vertices $\mathbf{M} \subset \Delta^M$.

It follows that $V(X,D)_\infty$ is a large wedge of circles. Explicitly, there is a weak equivalence

$$V(X,D)_\infty \simeq \vee_{M^2}\ S^1.$$

This space is path connected.

The system of path component sets

$$s \mapsto \pi_0 V(X,d)_s$$

therefore describes a hierarchy, as in the standard algorithms of topological data analyis.

3 Stability

Suppose that (X,d) is an ep-metric space, and that $x \in X$. The *global connected component* of x is the collection of $y \in X$ such that $d(x,y) < \infty$. Say that X is *globally connected* if $d(x,y) < \infty$ for all $x,y \in X$.

Global connectedness has the following general properties:

1) Every ep-metric space (X,d) is a disjoint union of its set $\pi_\infty(X,d)$ of global components.

2) An ep-metric space morphism $f : (X, d_X) \to (Y, d_Y)$ preserves global connected components: if $d_X(x, y) < \infty$ then

$$d_Y(f(x), f(y)) \leq d_X(x, y) < \infty,$$

so that $f(y)$ is in the connected component of $f(x)$. We therefore have an induced function $f_* : \pi_\infty(X, d_X) \to \pi_\infty(Y, d_Y)$.

3) Every metric space is globally connected.

Example 6. Suppose that X is a finite set with a system of neighbourhoods N_x and associated distances d_x for all $x \in X$ as in the list of Assumptions above, with the resulting ep-metric space (X, D).

The ep-metric space (X, D) has the property that $D(x, y) = 0$ forces $x = y$, by Remark 4. It follows that the global connected components of the ep-metric space (X, D) are metric spaces.

Say that a pair of elements (x, y) of X is a *neighbourhood pair* if $x \in N_y$ or $y \in N_x$. The argument of Remark 4 shows that elements u and v of (X, D) are in the same global connected component if and only if there is a polygonal path

$$P : u = x_0, x_1, \dots, x_n = v$$

such that each pair (x_i, x_{i+1}) is a neighbourhood pair.

Suppose that (X, d_X) and (Y, d_Y) are finite metric spaces, and that there is a monomorphism $i : X \subset Y$ that defines a map of ep-metric spaces, so that $d_Y(x, y) \leq d_X(x, y)$ for all $x, y \in X$. Set

$$m(i) = \max_{x \neq y \in X} \{ \frac{d_X(x, y)}{d_Y(i(x), i(y))} \}.$$

Then $1 \leq m(i) < \infty$ since X is finite.

The number $m(i)$ is the *compression factor* for the monomorphism i.

Example 7. Suppose that $i : X \subset Y$ is an inclusion of finite sets, and choose systems of neighbourhoods N_x, $x \in X$ and N'_y, $y \in Y$, with distances d_x and d'_y. Suppose that $N_x \subset N'_{i(x)}$ for all $x \in X$, and that

$$d'_{i(x)}(i(x), i(z)) \leq d_x(x, z) \tag{7}$$

for all $z \in N_x$, and for all $x \in X$.

Then the inclusions $i : N_x \subset N'_{i(x)}$ and the relations (7) define system morphisms $V(X, D_x) \to V(Y, D'_{i(x)})$ and $V(X, D) \to V(Y, D')$, as well as ep-metric space morphisms $(X, D) \to (Y, D')$.

The ep-metric space map $(X, D) \to (Y, D')$ is the realization of the system morphism $V(X, D) \to V(Y, D')$.

The ep-metric space morphism $(X, D) \to (Y, D')$ preserves global connected components, and the global connected components of (X, D) and (Y, D') are finite metric spaces.

For the following, recall that if (X,d) is an ep-metric space, then $P_s(X,d)$ is the poset of subsets σ of X such that $d(x,y) \leq s$ for all $x,y \in \sigma$. Recall further that the nerve $BP_s(X,d)$ is the barycentric subdivision of the Vietoris-Rips complex $V(X,d)_s$, so that the systems $BP(X,d)$ and $V(X,d)$ are naturally sectionwise homotopy equivalent.

Proposition 8. *Suppose that $i : X \subset Y$ is an inclusion of finite sets. Suppose that X and Y have metric space structures such that i defines a morphism $i : (X,d_X) \to (Y,d_Y)$ of ep-metric spaces. Suppose that for esvery $y \in Y$ there is an $x \in X$ such that $d_Y(y,i(x)) < r$ in Y.*

Then there are diagrams of poset morphisms

$$\begin{array}{ccc} P_s(X,d_X) & \xrightarrow{\sigma} & P_{m(i)\cdot(s+2r)}(X,d_X) \\ {\scriptstyle i}\downarrow & \nearrow{\scriptstyle \theta} & \downarrow{\scriptstyle i} \\ P_s(Y,d_Y) & \xrightarrow[\sigma]{} & P_{m(i)\cdot(s+2r)}(Y,d_Y) \end{array}$$

for all $0 \leq s < \infty$, in which the upper triangle commutes and the lower triangle homotopy commutes rel $P_s(X,d_X)$.

Proof. Define a function $\theta : Y \to X$ by setting $\theta(x) = x$ for $x \in X$, and by choosing $\theta(y)$ such that $d_Y(y,i(\theta(y))) < r$ for y outside of X.

Then

$$\begin{aligned} d_Y(i(\theta(y_1)),i(\theta(y_2))) &\leq d_Y(i(\theta(y_1)),y_1) + d_Y(y_1,y_2) + d_Y(y_2,i(\theta(y_2))) \\ &< d_Y(y_1,y_2) + 2r, \end{aligned}$$

and it follows that

$$d_X(\theta(y_1),\theta(y_2)) \leq m(i)\cdot(d_Y(y_1,y_2)+2r).$$

If $\sigma = \{y_1,\dots,y_n\}$ is a subset of Y such that $d(y_j,y_k) \leq s$ for all j,k, then $\theta(\sigma) = \{\theta(y_1),\dots,\theta(y_n)\}$ has $d(\theta(y_j),\theta(y_k)) \leq m(i)\cdot(s+2r)$ for all j,k.

The subset $\sigma \cup i(\theta(\sigma))$ of Y has distance between any two elements bounded above by $m(i)\cdot(s+2r)$. The natural inclusions

$$\sigma \subset \sigma \cup i(\theta(\sigma)) \supset i(\theta(\sigma))$$

define the required homotopies. □

As in Example 7, suppose that $i : X \subset Y$ is an inclusion of finite sets, and choose systems of neighbourhoods N_x, $x \in X$ and N'_y, $y \in Y$, with distances d_x and d'_y. Suppose that $N_x \subset N'_{i(x)}$ for all $x \in X$, and that

$$0 \neq d'_{i(x)}(i(x),i(z)) \leq d_x(x,z)$$

for all $z \in N_x$, for all $x \in X$. Form the corresponding ep-metric space morphism $i : (X,D) \to (Y,D')$.

Suppose that E is a global connected component of (X,D) and that F is a global connected component of (Y,D') such that $i(E) \subset F$. Consider the restriction of the ep-metric space morphism $i : (X,D) \subset (Y,D')$ to the ep-metric space morphism $i : (E,D) \to (F,D')$. Suppose that $m(i)$ is the compression factor for the map i of global components.

The objects (E,D) and (F,D') are metric spaces, by the choices of all weights d_x and d'_y — see Example 6.

The following result is a corollary of Proposition 8.

Theorem 9. *Suppose that the map $i : (E,D) \to (F,D')$ is the ep-metric space morphism between metric spaces that is described above. Suppose that for every $y \in F$ there is an $x \in E$ such that $D'(y,i(x)) < r$. Then there are diagrams*

$$\begin{array}{ccc} P_s(E,D) & \xrightarrow{\sigma} & P_{m(i)\cdot(s+2r)}(E,D) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ P_s(F,D') & \xrightarrow[\sigma]{} & P_{m(i)\cdot(s+2r)}(F,D') \end{array}$$

for all $0 \leq s < \infty$, in which the upper triangle commutes and the lower triangle homotopy commutes rel $P_s(E)$.

The canonical map

$$\phi : V(X,N) = \vee_x \ V(X,D_x) \to V(X,D),$$

is induced by the ep-metric space maps $(X,D_x) \to (X,D)$.

The ep-metric space (X,D) is a disjoint union of its global connected components E, and the system $V(X,D)$ is a disjoint union of the systems $V(E,D)$. Form the pullback diagram

$$\begin{array}{ccc} V(X,N)(E) & \longrightarrow & V(X,N) \\ {\scriptstyle \phi}\downarrow & & \downarrow{\scriptstyle \phi} \\ V(E,D) & \longrightarrow & V(X,D) \end{array}$$

The disjoint union

$$V(X,D) = \bigsqcup_{E \in \pi_\infty(X,D)} V(E,D)$$

pulls back to a disjoint union structure

$$V(X,N) = \bigsqcup_{E \in \pi_\infty(X,D)} V(X,N)(E)$$

on $V(X,N)$.

The excision isomorphisms

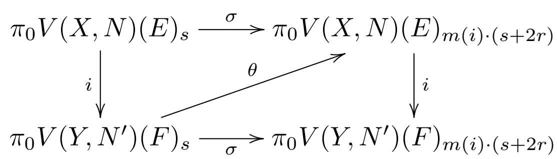
$$\phi_* : \pi_0 V(X,N)_s \xrightarrow{\cong} \pi_0 V(X,D)_s$$

of Lemma 2 restrict to isomorphisms

$$\phi_* : \pi_0 V(X,N)(E)_s \xrightarrow{\cong} \pi_0 V(E,D)_s. \tag{8}$$

We finish with a corollary of Theorem 9:

Theorem 10. *Suppose that the map $i : (E,D) \to (F,D')$ is the ep-metric space morphism between metric spaces that is described above. Suppose that for every $y \in F$ there is an $x \in E$ such that $D'(y, i(x)) < r$. Then there are commutative diagrams*



for all $0 \leq s < \infty$, where $m(i)$ is the compression factor for the map i.

Theorem 10 is a stability result for clustering in UMAP.

References

[1] J.F. Jardine. Metric spaces and homotopy types. Preprint, `http://uwo.ca/math/faculty/jardine/`, 2020.

[2] Leland McInnes and John Healy. UMAP: uniform manifold approximation and projection for dimension reduction. *CoRR*, abs/1802.03426, 2018.

[3] D.I. Spivak. Metric realization of fuzzy simplicial sets. Preprint, 2009.