CLASSIFYING CLUSTERING SCHEMES

GUNNAR CARLSSON AND FACUNDO MÉMOLI

ABSTRACT. Many clustering schemes are defined by optimizing an objective function defined on the partitions of the underlying set of a finite metric space. In this paper, we construct a framework for studying what happens when we instead impose various structural conditions on the clustering schemes, under the general heading of *functoriality*.

Functoriality refers to the idea that one should be able to compare the results of clustering algorithms as one varies the data set, for example by adding points or by applying functions to it. We show that within this framework, one can prove a theorems analogous to one of J. Kleinberg [Kle02], in which for example one obtains an existence and uniqueness theorem instead of a non-existence result.

We obtain a full classification of all clustering schemes satisfying a condition we refer to as excisiveness. The classification can be changed by varying the notion of maps of finite metric spaces. The conditions occur naturally when one considers clustering as the statistical version of the geometric notion of connected components. By varying the degree of functoriality that one requires from the schemes it is possible to construct richer families of clustering schemes that exhibit sensitivity to density.

1. INTRODUCTION

Clustering techniques play a very central role in various parts of data analysis. This type of methods can give important clues to the structure of data sets, and therefore suggest results and hypotheses in the underlying science. There are many interesting methods of clustering available, which have been applied to good effect in dealing with many datasets of interest, and are regarded as important methods in exploratory data analysis.

Most methods begin with a data set equipped with a notion of distance or metric. Starting with this input, a method might select a partition of the data set as one which optimizes a choice of an objective function. Other methods such as single, average, and complete linkage define clusterings, or rather nested families of clusterings, using a *linkage function* defined between clusters.

Desirable properties of clustering algorithms come from practitioners who have intuitive notions of what is a good clustering: "they know it when they see it". This is of course not satisfactory and a theoretical understanding needs to be developed. However, one thing this intuition reflects is the fact that density needs to be incorporated in the clustering procedures. Single linkage clustering, a procedure that enjoys several nice theoretical properties, is notorious for its insensitivity to density, which is manifested in the so called *chaining effect* [LW67].

Other methods such as average linkage, complete linkage [JD88, Chapter 3] and k-means share the property that they exhibit some sort of sensitivity to density, but are unstable in a sense which has been made theoretically precise [CM10a] and are therefore not well

supported by theory. We believe that this disconnect between theory and practice should not exist, and in particular, in [CM10b] we have constructed a theoretically sound framework that incorporates density via the use of 2-dimensional persistence ideas.

Given that there is relatively little theory surrounding clustering methods, the subject is regarded by many as consisting of a collection of ad hoc methods for which it is difficult to generate a coherent picture [GvLW09]. One exception to this rule is a paper of J. Kleinberg [Kle02], in which he proves a theorem which demonstrates the impossibility of constructing a clustering algorithm with three seemingly plausible properties. Kleinberg regards a clustering scheme as a map $\mathfrak{C}$ that assigns each finite metric space (X, d_X) one of its partitions. His "axioms" are

- *Scale invariance*: $\mathfrak{C}(X, d_X) = \mathfrak{C}(X, \lambda \cdot d_X)$ for all $\lambda > 0$;
- *Surjectivity*: for all partitions P of X there exists a metric d_X on X with $\mathfrak{C}(X, d_X) = P$; and
- *Consistency*: upon reducing distances between points in the same cluster of X (as produced by $\mathfrak{C}$), and increasing distances between points in different clusters, the result of applying the method to the new metric does not change.

Theorem 1.1 ([Kle02])**.** *There exists no clustering algorithm that simultaneously satisfies scale invariance, surjectivity and consistency.*

Kleinberg also points out that his results shed light on the trade-offs one has to make in choosing clustering algorithms. In this paper, we produce a variation on this theme, which we believe also has implications for how one thinks about and applies clustering algorithms.

In our earlier paper [CM10a], in the context of hierarchical methods, we have established a variant on Kleinberg's theorem, in which instead of impossibility, one actually proves an *existence and uniqueness* result, again assuming three plausible properties, slightly different from Kleinberg's. In the present paper, we make one more step towards understanding the theoretical foundations of clustering methods by expanding on the theme of [CM10a] and developing a context in which one can obtain much richer classification results, which include versions of a number of familiar methods, including analogues of the clique clustering methods familiar in network and graph problems. We now describe our context and results.

1.1. **Overview of our methods and results.** We regard clustering as the "statistical version" of the geometric notion of constructing the connected components of a topological space. Recall (see, e.g. [Mun75]) that the path components of a topological space X are the equivalence classes of points in the space under the equivalence relation $\sim_{path}$, where, for $x, x' \in X$, we have $x \sim_{path} y$ if and only if there is a continuous map $\varphi : [0, 1] \to X$ so that $\varphi(0) = x$ and $\varphi(1) = x'$. In other words, two points in X are in the same path component if they are connected by a continuous path in X. We let $\pi_0(X)$ denote the set of connected components of the space X. A crucial observation about π_0 is that it is a **functor**, i.e. that not only does one construct a set (of components) associated to every topological space, but additionally, to every continuous map $f : X \to Y$ one associates a map of sets $\pi_0(X) \to \pi_0(Y)$. We take the point of view that this **functoriality** is a key property to export from topological spaces to the finite metric spaces which form the input to clustering algorithms. Functoriality has been a very useful framework for the discussion of a variety of problems within mathematics over the last few decades, see [ML98].

Remark 1.1. *The reader may not be familiar with the notion of functoriality, as presented in this paper. The fundamental idea behind functoriality is that it is crucial to study the*

maps or transformations between mathematical objects as well as the objects themselves. Interpreted at this level of generality, one can readily see the importance of this notion via a number of examples.

(1) **Fourier analysis:** *The key observation of Fourier analysis in its various forms is the the study of the symmetry groups of a space allows one to construct very useful and conceptually important bases for spaces of functions on the space. So, the standard Fourier basis for complex valued functions on the circle is constructed by selecting functions which satisfy certain transformation laws under the rotational symmetries. This is of course also true for, e.g., the spherical harmonic basis for the functions on $\mathbb{R}^3$. Functoriality can be regarded as a transformation law.*

(2) **Normal forms for matrices:** *The analysis of the conjugation action of groups of invertible matrices on themselves gives rise to normal forms of matrices, such as Jordan normal form. This means that the conjugation symmetries on spaces of matrices allows one to understand their organization*

(3) **Algebraic topology:** *In algebraic topology, one associates to topological spaces algebraic objects called* homology groups. *It is critical for applications of these groups, and for computing them, that they obey a transformation law for any continuous map of spaces. This transformation law turns out to be very powerful, in that it is a key factor in making a small axiomatic description of the homology groups. Just as transformation laws for symmetries determine the Fourier basis, so the functoriality determines the homology groups.*

(4) **Galois theory:** *In Galois theory, groups of symmetries allow one to describe the sets of solutions of algebraic equations over fields, such as the rational numbers. The observation that equations were associated to symmetry groups revolutionized number theory, and to this day the study of various kinds of transformation laws in modular forms continues to have a powerful impact on the field.*

The point of the present paper is to illustrate that functoriality allows a very useful framework for classifying large families of clustering schemes. This framework will include clustering schemes which are sensitive to effects by density within the domain data set. Moreover, the functoriality of the construction of path connected components is a key feature in all aspects of topology, and the functoriality of clustering constructions should play a similar role in the computational topology of finite data sets.

Then, we require that a standard[1] clustering algorithm $\mathfrak{C}$ be a rule which assigns to every finite metric space X a set of clusters $\mathfrak{C}(X)$, subject to additional the requirement that every map $f : X \to Y$ induce a map of sets $\mathfrak{C}(f) : \mathfrak{C}(X) \to \mathfrak{C}(Y)$.

The question is what the replacement for continuous maps (i.e. *morphisms* of topological spaces) should be, and we find that there are many choices. We consider three nested classes of maps, namely isometries (called *iso*), distance non-increasing maps (called *gen*, for general), and distance non-increasing maps which are injections on the sets of points (called *inj*), with the hierarchy being given by $iso \subset inj \subset gen$. We find that functoriality with respect to *gen* is quite restrictive, in that it singles out single linkage clustering as the only functor which satisfies the *gen* functoriality requirement. On the other hand, functoriality with respect to *iso* appears not to be restrictive enough, in that it permits the specification of an arbitrary clustering on every isometry class of finite metric spaces.

---

[1]As opposed to hierarchical.

One of the features of our classification is the proof that for the *gen* and *inj* cases, a certain implicitly defined property of clustering schemes called **excisiveness**, which amounts to requiring that the clustering method is idempotent on each of the blocks of the partition it generates, is *equivalent* to the existence of a certain kind of *explicit generative model* for the scheme.

Mathematical results which prove the existence of a generative model given that some externally defined properties are satisfied are very powerful in mathematics. For example, the result that a system satisfying the axioms for a vector space over a field $\mathbb{F}$ admits a basis is very useful, as is the fact that a graph with no odd order cycles is bipartite.

The equivalence between excisiveness and the existence of a generative model for the clustering scheme has the interpretation that excisive schemes are parametrized by sets of "test metric spaces", which are those connected into a single cluster by the scheme. The method associated to such a family is then defined using a criterion which asserts that two points x and x' in a metric space X belong to the same cluster if there is a sequence of points $\{x_0, x_1, \ldots, x_n\}$ with $x_0 = x$, $x_n = x'$, and so that for each pair x_i, x_{i+1} there is a morphism in *gen* or *inj* from one of the test metric spaces into X with x_i and x_{i+1} in its image.

A key feature of the choice of morphisms in *inj* is that (since they are injective) this permits one to take account of *density* in a certain sense. If one considers functoriality with respect to *gen*, one finds that the morphisms might collapse a large dense collection of points into a single point, which makes it impossible to take density into account in building the clusters. Our classification of functors based on *inj* includes analogues of "clique clustering" [PDFV05] and the DBSCAN algorithm [EpKSX96], and such methods clearly have the effect of taking a version of density into account.

Whereas our classification of *gen* functorial schemes yields that single linkage is the unique possible scheme, which turns out to be excisive, we find large classes of non-excisive clustering schemes in *inj*. Our construction of such families relies on a study of the metric invariants that are well behaved under *inj* morphisms. Finally, it turns out that all schemes in *inj* that admit a *finite* generative model arise as the composition of single linkage clustering with an application that changes the underlying metric of the input metric space. This result has clear practical consequences which we discuss.

We believe that functoriality, incorporating as it does the idea that clustering should not operate only on finite metric spaces as isolated objects, but rather on the metric space together with other metric spaces which are related to it via morphisms, is a very desirable property. We can point to its success within topology as an extremely powerful tool which pins down many constructions as the unique solution to existence problems, and which is the key to most applications of topological methods. It is also true that functoriality is necessary in order to extend the methods of clustering to higher order connectivity information with computational topology of finite data sets [Car09].

In this paper we demonstrate that the use of restricted notions of functoriality allows us to study clustering schemes which take proxies for density into account. This is important because it is clear that some of the value of alternate schemes such as average and complete linkage (which are not functorial) is due precisely to their ability to do this. Therefore, some of the methods we study will not suffer from the problems inherent in single linkage clustering, such as chaining [JS71, Section 7.4] [CM10b].

1.2. **Organization of the paper.** In Section 2 we discuss basic background concepts that we use in our presentation. Section 3 introduces the concept of a category and gives many

examples and constructions we use in the paper. The concept of a functor and functoriality are discussed in Section 4. We present our main characterization results for standard clustering methods in Section 6. We study hierarchical methods in Section 7. Concluding remarks are given in Section 8.

2. Preliminaries

For a finite set X we denote by $\mathcal{P}(X)$ the set of all partitions of X. For each $k \in \mathbb{N}$, $\mathcal{M}_k$ denotes the collection of all finite metric spaces with cardinality k and $\mathcal{M} = \bigcup_{k \in \mathbb{N}} \mathcal{M}_k$ denotes the collection of all finite metric spaces. For $(X, d_X) \in \mathcal{M}$ let $\mathrm{sep}(X, d_X) := \min_{x \neq x'} d_X(x, x')$ be the *separation* of X, $\mathrm{diam}(X, d_X) := \max_{x, x' \in X} d_X(x, x')$ be the *diameter* of X, and for $\lambda > 0$ we denote by $\lambda \cdot X$ the metric space $(X, \lambda \cdot d_X)$. For $\delta > 0$ and $k \geqslant 2$ we let $\Delta_k(\delta)$ denote the metric space with k points whose inter-point distances equal δ.

Definition 2.1. *On $(X, d_X) \in \mathcal{M}$, for each $\delta \geqslant 0$, we define the equivalence relation $\sim_\delta$, where $x \sim_\delta x'$ if and only if there is a sequence $x_0, x_1, \ldots, x_k \in X$ so that $x_0 = x, x_k = x'$, and $d_X(x_i, x_{i+1}) \leqslant \delta$ for all i. We denote by $[x]_\delta$ the equivalence class of $x \in X$ under $\sim_\delta$.*

The following lemma follows directly from the definition of $\sim_\delta$.

Lemma 2.1. *Let $(X, d_X) \in \mathcal{M}$ and $\delta \geqslant 0$. Then, $[x]_\delta \neq [x']_\delta$ for some $x, x' \in X$ if and only if*

$$d_X(a, a') > \delta, \text{ for all } a \in [x]_\delta \text{ and } a' \in [x']_\delta.$$

Let X be any finite set and $W_X : X \times X \to \mathbb{R}_+$ be a symmetric function s.t. $W_X(x, x) = 0$ for all $x \in X$. Then, the *maximal sub-dominant ultrametric* relative to W_X is given by

$$(2\text{--}1) \qquad \mathcal{U}(W_X)(x, x') = \min \left\{ \max_i W_X(x_i, x_{i+1}), \; \{x_i\}_{i=0}^k \text{ s.t. } x_0 = x \text{ and } x_k = x' \right\}.$$

This construction and its correctness are standard [NS08, §6.4.3].

We define objects which will encode the notion of "multiscale" or "multiresolution" partitions.

Definition 2.2. *A persistent set is a pair (X, θ_X), where X is a finite set, and θ_X is a function from the non-negative real line $[0, +\infty)$ to $\mathcal{P}(X)$ so that the following properties hold.*
  (1) *If $r \leqslant s$, then $\theta_X(r)$ refines $\theta_X(s)$.*
  (2) *For any r, there is a number $\epsilon > 0$ so that $\theta_X(r') = \theta_X(r)$ for all $r' \in [r, r + \epsilon]$.*
If in addition there exists $t > 0$ s.t. $\theta_X(t)$ consists of the single block partition for all $r \geqslant t$, then we say that (X, θ_X) is a dendrogram.[2]

Example 2.1. *Let (X, d_X) be a finite metric space. Then we can associate to (X, d_X) the persistent set whose underlying set is X, and where for each $r \geqslant 0$, blocks of the partition $\theta_X(r)$ consist of the equivalence classes under the equivalence relation $\sim_r$.*

Example 2.2. *Here we consider the family of Agglomerative Hierarchical clustering techniques, [JD88]. We (re)define these by the recursive procedure described next. Let $X = \{x_1, \ldots, x_n\}$ and let $\mathcal{L}$ denote a family of* linkage functions, *i.e. functions which one uses*

---

[2]In the paper we will be using the word dendrogram to refer both to the object defined here and to the standard graphical representation of them.

for defining the distance between two clusters. Fix $l \in \mathcal{L}$. For each $R > 0$ consider the equivalence relation $\sim_{l,R}$ on blocks of a partition $\Pi \in \mathcal{P}(X)$, given by $\mathcal{B} \sim_{l,R} \mathcal{B}'$ if and only if there is a sequence of blocks $\mathcal{B} = \mathcal{B}_1, \ldots, \mathcal{B}_s = \mathcal{B}'$ in Π with $l(\mathcal{B}_k, \mathcal{B}_{k+1}) \leqslant R$ for $k = 1, \ldots, s-1$. Consider the sequences $r_1, r_2, \ldots \in [0, \infty)$ and $\Theta_1, \Theta_2, \ldots \in \mathcal{P}(X)$ given by $\Theta_1 := \{x_1, \ldots, x_n\}$ and for $i \geqslant 1$, $\Theta_{i+1} = \Theta_i / \sim_{l,r_i}$ where $r_i := \min\{l(\mathcal{B}, \mathcal{B}'), \mathcal{B}, \mathcal{B}' \in \Theta_i, \mathcal{B} \neq \mathcal{B}'\}$. Finally, we define $\theta^l : [0, \infty) \to \mathcal{P}(X)$ by $r \mapsto \theta^l(r) := \Theta_{i(r)}$ where $i(r) := \max\{i | r_i \leqslant r\}$.

Standard choices for l are single linkage*:*

$$l(\mathcal{B}, \mathcal{B}') = \min_{x \in \mathcal{B}} \min_{x' \in \mathcal{B}'} d_X(x, x');$$

complete linkage

$$l(\mathcal{B}, \mathcal{B}') = \max_{x \in \mathcal{B}} \max_{x' \in \mathcal{B}'} d_X(x, x');$$

and average linkage*:*

$$l(\mathcal{B}, \mathcal{B}') = \frac{\sum_{x \in \mathcal{B}} \sum_{x' \in \mathcal{B}'} d_X(x, x')}{|\mathcal{B}| \cdot |\mathcal{B}'|}.$$

It is easily verified that the notion discussed in Example 2.1 is equivalent to θ^l when l is the single linkage function. Note that, unlike the usual definition of agglomerative hierarchical clustering, at each step of the inductive definition we allow for more than two clusters to be merged.

3. Categories

3.1. **Definitions and Examples.** In this section, we will give a brief description of the theory of categories and functors, which will be the framework in which we state the constraints that we require of our clustering algorithms. An excellent reference for these ideas is [ML98].

Categories are useful mathematical constructs that encode the nature of certain objects of interest together with a set of admissible maps between them. This formalism is extremely useful for studying classes of mathematical objects which share a common structure, such as sets, groups, vector spaces, or topological spaces. The definition is as follows.

Definition 3.1. *A **category** $\underline{C}$ consists of:*
- *A collection of **objects** $\mathrm{ob}(\underline{C})$ (e.g. sets, groups, vector spaces, etc.)*
- *For each pair of objects $X, Y \in \mathrm{ob}(\underline{C})$, a set
  $\mathrm{Mor}_{\underline{C}}(X, Y)$, the **morphisms** from X to Y (e.g. maps of sets from X to Y, homomorphisms of groups from X to Y, linear transformations from X to Y, etc. respectively)*
- *Composition operations:
  $\circ : \mathrm{Mor}_{\underline{C}}(X, Y) \times \mathrm{Mor}_{\underline{C}}(Y, Z) \to \mathrm{Mor}_{\underline{C}}(X, Z)$, corresponding to **composition** of set maps, group homomorphisms, linear transformations, etc.*
- *For each object $X \in \underline{C}$, a distinguished element $id_X \in \mathrm{Mor}_{\underline{C}}(X, X)$, called the **identity** morphism.*

The composition is assumed to be associative in the obvious sense, and for any $f \in \mathrm{Mor}_{\underline{C}}(X, Y)$, it is assumed that $id_Y \circ f = f$ and $f \circ id_X = f$.

Example 3.1 (The category of sets)**.** *Denote by $\underline{Sets}$ the category whose objects are all sets, and the morphisms between two sets A and B are all the functions from A to B.*

A morphism $f : X \to Y$ which has a two sided inverse $g : Y \to X$, so that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$, is called an **isomorphism**. Two objects which are isomorphic are intuitively thought of as "structurally indistinguishable" in the sense that they are identical except for naming or choice of coordinates. For example, in the category of sets, the sets $\{1, 2, 3\}$ and $\{A, B, C\}$ are isomorphic, since they are identical except for the choice made in labelling the elements.

Example 3.2. *Here we consider four very simple categories, that we are going to refer to as $\underline{0}, \underline{1}, \underline{2}$ and $\underline{3}$ respectively.*

- *The category $\underline{0}$ has $ob(\underline{0}) = \varnothing$ and all the conditions in the definition above are trivially satisfied.*
- *Consider the category $\underline{1}$ with exactly one object A and one morphism: $\mathrm{Mor}_{\underline{1}}(A, A) = f$. It follows that f must be the identity morphism id_A. This is represented graphically as follows:*

$$A \;\; \overset{\curvearrowright}{id_A}$$

- *The category $\underline{2}$ has exactly two objects A and B and three morphisms: the identities from A to A and from B to B and exactly one morphism in $\mathrm{Mor}_{\underline{2}}(A, B)$:*

$$id_A \overset{\curvearrowright}{} A \overset{f}{\longrightarrow} B \overset{\curvearrowright}{id_B}$$

- *Finally, the category $\underline{3}$ has exactly three objects A, B and C and six morphisms: the identities from A to A, from B to B and C to C, and three more morphisms, $\mathrm{Mor}_{\underline{3}}(A, B) = f$, $\mathrm{Mor}_{\underline{3}}(B, C) = g$ and $\mathrm{Mor}_{\underline{3}}(A, C) = h$:*

$$id_A \overset{\curvearrowright}{} A \overset{f}{\longrightarrow} B \overset{\curvearrowright}{id_B}$$
$$\overset{h}{\searrow} \qquad \overset{g}{\swarrow}$$
$$C$$
$$\overset{\curvearrowright}{id_C}$$

Now, note that in order to satisfy the composition rule one must have $h = g \circ f$.

Example 3.3. *A more surprising example of a category is given by the following construction. A monoid $(M, *, e)$ consists of a set M equipped with a binary operation $* : M \times M \to M$ s.t. $(A * B) * C = A * (B * C)$ for all $A, B, C \in M$ and an element $e \in M$ s.t. $e * A = A * e = A$ for all $A \in M$. A monoid can be made into a category $\underline{M}$ with a single object O and $\mathrm{Mor}_{\underline{M}}(O, O) = M$ and id_O being represented by e. The composition of morphisms is induced by the operation $*$.*

Example 3.4. *Let $\mathbb{E}$ denote the extended positive real line $[0, +\infty]$. We construct the category $\underline{\mathbb{E}}$ with $ob(\underline{\mathbb{E}}) = \mathbb{E}$ and $\mathrm{Mor}_{\underline{\mathbb{E}}}(a, b) = \{(a, b)\}$ if $a \geqslant b$, and $\mathrm{Mor}_{\underline{\mathbb{E}}}(a, b) = \varnothing$. The composition is given by $(a, b) \circ (b, c) = (a, c)$.*

Now we discuss the core constructions for this paper.

Definition 3.2 ($\underline{\mathcal{C}}$, a category of outputs of standard clustering schemes). *Let Y be a finite set, $P_Y \in \mathcal{P}(Y)$, and $f : X \to Y$ be a set map. We define $f^*(P_Y)$ to be the partition of X whose blocks are the sets $f^{-1}(B)$ where B ranges over the blocks of P_Y. We construct the category $\underline{\mathcal{C}}$ of outputs of standard clustering algorithms with $ob(\underline{\mathcal{C}})$ equal to all possible*

pairs (X, P_X) where X is a finite set and P_X is a partition of X: $P_X \in \mathcal{P}(X)$. For objects (X, P_X) and (Y, P_Y) one sets $\text{Mor}_{\underline{C}}\big((X, P_X), (Y, P_Y)\big)$ to be the set of all maps $f : X \to Y$ with the property that P_X is a refinement of $f^*(P_Y)$.

Example 3.5. *Let X be any finite set, $Y = \{a, b\}$ a set with two elements, and P_X a partition of X. Assume first that $P_Y = \{\{a\}, \{b\}\}$ and let $f : X \to Y$ be any map. Then, in order for f to be a morphism in $\text{Mor}_{\underline{C}}\big((X, P_X), (Y, P_Y)\big)$ it is necessary that x and x' be in different blocks of P_X whenever $f(x) \neq f(x')$. Assume now that $P_Y = \{a, b\}$ and $g : Y \to X$. Then, the condition that $g \in \text{Mor}_{\underline{C}}\big((Y, P_Y), (X, P_X)\big)$ requires that $g(a)$ and $g(b)$ be in the same block of P_X.*

We will also construct a category of *persistent sets*, which will constitute the output of hierarchical clustering functors.

Definition 3.3 ($\underline{\mathcal{P}}$, **a category of outputs of hierarchical clustering schemes**)**.** *Let $(X, \theta_X), (Y, \theta_Y)$ be persistent sets. A map of sets $f : X \to Y$ is said to be* persistence preserving *if for each $r \in \mathbb{R}$, we have that $\theta_X(r)$ is a refinement of $f^*(\theta_Y(r))$. We define a category $\underline{\mathcal{P}}$ whose objects are persistent sets, and where $\text{Mor}_{\underline{\mathcal{P}}}((X, \theta_X), (Y, \theta_Y))$ consists of the set maps from X to Y which are persistence preserving.*

It is easily verified that the composite of persistence preserving maps is persistence preserving, and that any identity map is persistence preserving. A simple example is shown in Figure 1.
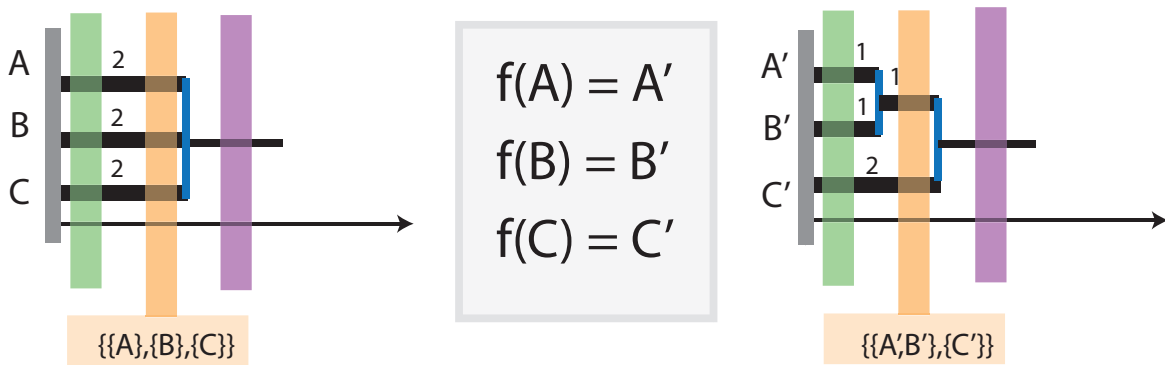

FIGURE 1. Two persistent sets (X, θ_X) and (Y, θ_Y) represented by their dendrograms. On the left one defined in the set $X = \{A, B, C\}$ and on the right one defined on the set $Y = \{A', B', C'\}$. Consider the given set map $f : X \to Y$. Then we see that f is persistence preserving since for each $r \geq 0$, the partition $\theta_X(r)$ is a refinement of $f^*(\theta_Y(r))$. Indeed, there are three interesting ranges of values of r. Pick for example r like in the orange shaded area: $r \in [1, 2)$. Then $\theta_Y(r) = \{\{A', B'\}, \{C'\}\}$ and hence $f^*(\theta_Y(r)) = \{f^{-1}(\{A', B'\}), \{f^{-1}(C')\}\} = \{\{A, B\}, \{C\}\}$ which is indeed refined by $\theta_X(r) = \{\{A\}, \{B\}, \{C\}\}$. One proceeds similarly for the other two cases.

3.2. **Three categories of finite metric spaces.** We will describe three categories $\underline{\mathcal{M}}^{iso}$, $\underline{\mathcal{M}}^{inj}$, and $\underline{\mathcal{M}}^{gen}$, whose collections of objects will all consist of the collection of finite metric

spaces $\mathcal{M}$. For (X, d_X) and (Y, d_Y) in $\mathcal{M}$, a map $f : X \to Y$ is said to be **distance non increasing** if for all $x, x' \in X$, we have $d_Y(f(x), f(x')) \leqslant d_X(x, x')$. It is easy to check that composition of distance non-increasing maps are also distance non-increasing, and it is also clear that id_X is always distance non-increasing. We therefore have the category $\underline{\mathcal{M}}^{gen}$, whose objects are finite metric spaces, and so that for any objects X and Y, $\mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, Y)$ is the set of distance non-increasing maps from X to Y, cf. [Isb64] for another use of this class of maps. It is clear that compositions of injective maps are injective, and that all identity maps are injective, so we have the new category $\underline{\mathcal{M}}^{inj}$, in which $\mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, Y)$ consists of the **injective distance non-increasing maps**. Finally, if (X, d_X) and (Y, d_Y) are finite metric spaces, $f : X \to Y$ is an **isometry** if f is bijective and $d_Y(f(x), f(x')) = d_X(x, x')$ for all x and x'. It is clear that as above, one can form a category $\underline{\mathcal{M}}^{iso}$ whose objects are finite metric spaces and whose morphisms are the isometries. Furthermore, one has inclusions

$$(3\text{--}2) \qquad\qquad \underline{\mathcal{M}}^{iso} \subseteq \underline{\mathcal{M}}^{inj} \subseteq \underline{\mathcal{M}}^{gen}$$

of subcategories (defined as in [ML98]). Note that although the inclusions are bijections on object sets, they are proper inclusions on morphism sets, i.e. in general they are not surjective.

Remark 3.1. *The category $\underline{\mathcal{M}}^{gen}$ is special in that for any pair of finite metric spaces X and Y, $\mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, Y) \neq \varnothing$. Indeed, pick $y_0 \in Y$ and define $\phi : X \to Y$ by $x \mapsto y_0$ for all $x \in X$. Clearly, $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, Y)$. This is not the case for $\underline{\mathcal{M}}^{inj}$ since in order for $\mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, Y) \neq \varnothing$ to hold it is necessary (but not sufficient in general) that $|Y| \geqslant |X|$.*

4. FUNCTORS AND FUNCTORIALITY.

Next we introduce the key concept in our discussion, that of a *functor*. We give the formal definition first, and several examples will appear as different constructions that we use in the paper.

Definition 4.1 (**Functor**). *Let $\underline{C}$ and $\underline{D}$ be categories. Then a* functor *from $\underline{C}$ to $\underline{D}$ consists of:*

- *A map of sets $F : \mathrm{ob}(\underline{C}) \to \mathrm{ob}(\underline{D})$.*
- *For every pair of objects $X, Y \in \underline{C}$ a map of sets $\Phi(X, Y) : \mathrm{Mor}_{\underline{C}}(X, Y) \to \mathrm{Mor}_{\underline{D}}(FX, FY)$ so that*
  - *(1) $\Phi(X, X)(id_X) = id_{F(X)}$ for all $X \in \mathrm{ob}(\underline{C})$, and*
  - *(2) $\Phi(X, Z)(g \circ f) = \Phi(Y, Z)(g) \circ \Phi(X, Y)(f)$ for all $f \in \mathrm{Mor}_{\underline{C}}(X, Y)$ and $g \in \mathrm{Mor}_{\underline{C}}(Y, Z)$.*

*Given a category $\underline{C}$, an **endofunctor** on $\underline{C}$ is any functor $F : \underline{C} \to \underline{C}$.*

Remark 4.1. *In the interest of clarity, we will always refer to the pair (F, Φ) with a single letter F. See diagram (4–4) below for an example.*

Example 4.1. (**Forgetful functors**) *When one has two categories $\underline{C}$ and $\underline{D}$, where the objects in $\underline{C}$ are objects in $\underline{D}$ equipped with some additional structure and the morphisms in $\underline{C}$ are simply the morphisms in $\underline{D}$ which preserve that structure, then we obtain the "forgetful functor" from $\underline{C}$ to $\underline{D}$, which carries the object in $\underline{C}$ to the same object in $\underline{C}$, but regarded without the additional structure. For example, a group can be regarded as a set with the additional structure of multiplication and inverse maps, and the group homomorphisms are simply the set maps which respect that structure. Accordingly, we have the functor from the category of groups to the category of sets which "forgets the multiplication and inverse".*

Similarly, we have the forgetful functor from $\underline{\mathcal{C}}$ to the category of sets, which forgets the presence of P_X in the output set (X, P_X).

Example 4.2. *The inclusions $\underline{\mathcal{M}}^{iso} \subseteq \underline{\mathcal{M}}^{inj} \subseteq \underline{\mathcal{M}}^{gen}$ are both functors.*

Example 4.3 (Scaling functor)**.** *For any $\lambda > 0$ we define an endofunctor $\sigma_\lambda : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{M}}^{gen}$ on objects by $\sigma_\lambda(X, d_X) = (X, \lambda \cdot d_X)$ and on morphisms by $\sigma_\lambda(f) = f$. One easily verifies that if f satisfies the conditions for being a morphism in $\underline{\mathcal{M}}^{gen}$ from (X, d_X) to (Y, d_Y), then it readily satisfies the conditions of being a morphism from $(X, \lambda \cdot d_X)$ to $(Y, \lambda \cdot d_Y)$. Clearly, σ_λ can also be regarded as an endofunctor in $\underline{\mathcal{M}}^{iso}$ and $\underline{\mathcal{M}}^{inj}$.*

Similarly, we define a functor $s_\lambda : \underline{\mathcal{P}} \to \underline{\mathcal{P}}$ by setting $s_\lambda(X, \theta_X) = (X, \theta_X^\lambda)$, where $\theta_X^\lambda(r) = \theta_X(\frac{r}{\lambda})$.

4.1. **Clustering algorithms as functors.** The notion of categories, functors and functoriality provide useful framework for studying algorithms. One first defines a class of input objects $\underline{\mathcal{I}}$ and a class of output objects $\underline{\mathcal{O}}$. Moreover, one associates to each of these classes a class of natural maps, the morphisms, between objects. For the problem of HC for example, the input class is the set of finite metric spaces and the output class is that of dendrograms. An algorithm is to be regarded as a functor between a category of input objects and a category of output objects.

An algorithm will therefore be a procedure that assigns to each $I \in \underline{\mathcal{I}}$ an output $O_I \in \underline{\mathcal{O}}$ with the further property that it respects relations between objects in the following sense. Assume $I, I' \in \underline{\mathcal{I}}$ such that there is natural map $f : I \to I'$. Then, the algorithm has to have the property that the relation between O_I and $O_{I'}$ has to be represented by a natural map for output objects.

Example 4.4. *Assume that $\underline{\mathcal{I}}$ is such that $\mathrm{Mor}_{\underline{\mathcal{I}}}(X, Y) = \varnothing$ for all $X, Y \in \underline{\mathcal{I}}$ with $X \neq Y$. In this case, since there are no morphisms between input objects any functor $\mathfrak{A} : \underline{\mathcal{I}} \to \underline{\mathcal{O}}$ can be specified arbitrarily on each $X \in \underline{\mathcal{O}}$.*

We view any given clustering scheme as a procedure which takes as input a finite metric space (X, d_X), and delivers as output either an object in $\underline{\mathcal{C}}$ or $\underline{\mathcal{P}}$:

- **Standard clustering**: a pair (X, P_X) where P_X is a partition of X. Such a pair is an object in the category $\underline{\mathcal{C}}$.
- **Hierarchical clustering**: a pair (X, θ_X) where θ_X is a persistent set over X. Such a pair is an object in the category $\underline{\mathcal{P}}$.

The concept of **functoriality** refers to the additional condition that the clustering procedure should map a pair of input objects into a pair of output objects in a manner which is consistent with respect to the morphisms attached to the input and output spaces. When this happens, we say that the clustering scheme is **functorial**. This notion of consistency is made precise in Definition 4.1 and described by diagram (4–4). Let $\underline{\mathcal{M}}$ stand for any of $\underline{\mathcal{M}}^{gen}$, $\underline{\mathcal{M}}^{inj}$ or $\underline{\mathcal{M}}^{iso}$.

According to Definition 4.1, in order to view a standard clustering scheme as a functor $\mathfrak{C} : \underline{\mathcal{M}} \to \underline{\mathcal{C}}$ we need to specify:

(1) how it maps objects of $\underline{\mathcal{M}}$ (finite metric spaces) into objects of $\underline{\mathcal{C}}$, and
(2) how a morphism $f : (X, d_X) \to (Y, d_Y)$ between two objects (X, d_X) and (Y, d_Y) in the input category $\underline{\mathcal{M}}$ induces a map in the output category $\underline{\mathcal{C}}$, see diagram (4–4).

(4–3)
$$\begin{array}{ccc} (X, d_X) & \xrightarrow{\ f\ } & (Y, d_Y) \\ \downarrow{\mathfrak{C}} & & \downarrow{\mathfrak{C}} \\ (X, P_X) & \xrightarrow{\mathfrak{C}(f)} & (Y, P_Y) \end{array}$$

Similarly, in order to view a hierarchical clustering scheme as a functor $\mathfrak{H} : \underline{\mathcal{M}} \to \underline{\mathcal{P}}$ we need to specify:

(1) how it maps objects of $\underline{\mathcal{M}}$ (finite metric spaces) into objects of $\underline{\mathcal{P}}$, and
(2) how a morphism $f : (X, d_X) \to (Y, d_Y)$ between two objects (X, d_X) and (Y, d_Y) in the input category $\underline{\mathcal{M}}$ induces a map in the output category $\underline{\mathcal{P}}$, see diagram (4–4).

(4–4)
$$\begin{array}{ccc} (X, d_X) & \xrightarrow{\ f\ } & (Y, d_Y) \\ \downarrow{\mathfrak{H}} & & \downarrow{\mathfrak{H}} \\ (X, \theta_X) & \xrightarrow{\mathfrak{H}(f)} & (Y, \theta_Y) \end{array}$$

Precise constructions will be discussed in Sections 6 and 7, where we study different clustering algorithms, both standard or flat, and hierarchical, using the idea of functoriality.

We have 3 possible "input" categories ordered by inclusion (3–2). The idea is that studying functoriality over a larger category will be more stringent/demanding than requiring functoriality over a smaller one. We will consider different clustering algorithms and study whether they are functorial over our choice of the input category. The least demanding one, $\underline{\mathcal{M}}^{iso}$ basically enforces that clustering schemes are not dependent on the way points are labeled. We will prove uniqueness results for functoriality over the most stringent category $\underline{\mathcal{M}}^{gen}$, and finally we study how relaxing the conditions imposed by the morphisms in $\underline{\mathcal{M}}^{gen}$, namely, by restricting ourselves to the smaller but intermediate category $\underline{\mathcal{M}}^{mon}$, we permit more functorial clustering algorithms.

5. Invariants of metric spaces

Now we consider maps from a metric space into the ordered extended positive real line $\underline{\mathbb{E}}$ that behave well under morphisms of the categories $\underline{\mathcal{M}}^{iso}$, $\underline{\mathcal{M}}^{inj}$, and $\underline{\mathcal{M}}^{gen}$, respectively. Such maps will be used later on for defining certain clustering functors, and the richness of such maps, or lack thereof, signals the degree to which these clustering functors exhibit different interesting behaviors.

Definition 5.1. *Let $\underline{\mathcal{M}}$ be any of $\underline{\mathcal{M}}^{gen}$, $\underline{\mathcal{M}}^{inj}$ or $\underline{\mathcal{M}}^{iso}$, then an **invariant** is any functor $\mathfrak{I} : \underline{\mathcal{M}} \to \underline{\mathbb{E}}$.*

Recall that this means that $\mathfrak{I}(X) \geqslant \mathfrak{I}(Y)$ whenever $\mathrm{Mor}_{\underline{\mathcal{M}}}(X, Y) \neq \varnothing$. Diagrammatically, this is the behavior we seek:

(5–5)
$$\begin{array}{ccc} (X, d_X) & \xrightarrow{\ \phi\ } & (Y, d_Y) \\ \downarrow{\mathfrak{I}} & & \downarrow{\mathfrak{I}} \\ \mathfrak{I}(X) & \xrightarrow{\ \geqslant\ } & \mathfrak{I}(Y) \end{array}$$

5.1. **The case of $\underline{\mathcal{M}}^{iso}$.** The collection of invariants in $\underline{\mathcal{M}}^{iso}$ is the largest possible among $\underline{\mathcal{M}}^{iso}, \underline{\mathcal{M}}^{inj}$ and $\underline{\mathcal{M}}^{gen}$, and this could be seen as a consequence of the fact that $\underline{\mathcal{M}}^{iso}$ is the category with the fewest morphisms cf. (3–2). In this case all $\underline{\mathcal{M}}^{iso}$-functorial invariants are given in the following manner: $\mathfrak{I}(X) = \Psi(d_X)$ where $\Psi : \mathbb{R}_+^{|X|} \times \mathbb{R}_+^{|X|} \to \mathbb{E}$ is such that $\Psi(A) = \Psi(\pi' \cdot A \cdot \pi)$ for any permutation matrix π of size $|X| \times |X|$ and $A \in \mathbb{R}_+^{|X|} \times \mathbb{R}_+^{|X|}$.

5.2. **The case of $\underline{\mathcal{M}}^{inj}$.**

Example 5.1. *A first example of a $\underline{\mathcal{M}}^{inj}$-invariant is any non-increasing function $\zeta : \mathbb{N} \to \mathbb{E}$ of the cardinality $|X|$ of a finite metric space X: clearly, since $\mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, Y) \neq \varnothing$ requires that $|X| \leqslant |Y|$, then $\zeta(|X|) \geqslant \zeta(|Y|)$. Hence, the construction is functorial.*

Another simple example is $\mathfrak{I}_{\mathrm{sep}} : \underline{\mathcal{M}}^{inj} \to \underline{\mathbb{E}}$, which assigns to any finite metric space X its separation $\mathrm{sep}(X)$. This example belongs to a larger class.

Example 5.2. *For each $k \geqslant 2$ let $\mathfrak{I}_k : \underline{\mathcal{M}}^{inj} \to \mathbb{E}$ be the functor given by*
$$\mathfrak{I}_k^-(X) := \inf \left\{ \varepsilon \geqslant 0 \,\middle|\, \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(\Delta_k(\varepsilon), X) \neq \varnothing \right\}$$
for $X \in \mathrm{ob}(\underline{\mathcal{M}}^{inj})$ with $|X| \geqslant k$. If $|X| < k$ we put $\mathfrak{I}_k(X) = +\infty$.

Notice that $\mathfrak{I}_{sep} = \mathfrak{I}_2^-$.
That this definition is indeed $\underline{\mathcal{M}}^{inj}$-functorial can be seen easily.

Proof. Assume that $X, Y \in \mathrm{ob}(\underline{\mathcal{M}}^{inj})$ and $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, Y)$, then, we need to check that $\mathfrak{I}_k^-(X) \geqslant \mathfrak{I}_k^-(Y)$. If $|X|, |Y| < k$, or $|X| \geqslant k$ and $|Y| < k$, there's nothing to prove. Let's assume that $|X|, |Y| \geqslant k$. Pick $\varepsilon > \mathfrak{I}_k^-(X)$ and $f \in \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(\Delta_k(\varepsilon), X)$. Then, $\phi \circ f \in \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(\Delta_k(\varepsilon), Y)$, and hence $\varepsilon \geqslant \mathfrak{I}_k^-(Y)$. The conclusion follows since $\varepsilon > \mathfrak{I}_k^-(X)$ was arbitrary. $\qquad\square$

Another family of $\underline{\mathcal{M}}^{inj}$-functorial invariants with a similar structure is the following.

Example 5.3. *For each $k \in \mathbb{N}$ define*
$$\mathfrak{I}_k^+(X) := \sup \left\{ \varepsilon \geqslant 0 \,\middle|\, \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, \Delta_k(\varepsilon)) \neq \varnothing \right\}$$
if $|X| \leqslant k$, and $\mathfrak{I}_k^+(X) = 0$ otherwise.

Finally, these two examples can be generalized in the following manner.

Definition 5.2. *Let $\Omega \subset \mathcal{M}$ be a collection of finite metric spaces and $\iota : \underline{\mathcal{M}}^{iso} \to \underline{\mathbb{E}}$ any $\underline{\mathcal{M}}^{iso}$-functorial invariant. Then, one defines*
$$\mathfrak{I}_\Omega^-(X) := \inf \left\{ \iota(\omega) \,\middle|\, \exists\, \omega \in \Omega \text{ s.t. } \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(\omega, X) \neq \varnothing \right\}$$
and
$$\mathfrak{I}_\Omega^+(X) := \sup \left\{ \iota(\omega) \,\middle|\, \exists\, \omega \in \Omega \text{ s.t. } \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, \omega) \neq \varnothing \right\}.$$

In these definitions we use the convention that inf over an empty set equals $+\infty$ whereas sup over an empty set equals 0. The proof that these these two constructions are $\underline{\mathcal{M}}^{inj}$-functorial follows the same line of argument as for the case of $\mathfrak{I}_k^-$ above. That these definitions generalize the ones of $\mathfrak{I}_k^\pm$ above can be seen by letting $\Omega = \{\Delta_k(\delta),\, \delta > 0\}$ and ι equal to the separation invariant.

We do not know a complete classification of all the $\underline{\mathcal{M}}^{inj}$-functorial invariants but it seems that progress could be made using our techniques. Notice that the family of $\underline{\mathcal{M}}^{inj}$-functorial invariants is rich in the sense that it is closed under addition and multiplication.

5.3. **The case of $\underline{\mathcal{M}}^{gen}$.** This case is particularly easy to study: since the collection of all morphisms is quite large (it is a super-set of those of $\underline{\mathcal{M}}^{iso}$ and $\underline{\mathcal{M}}^{inj}$), one would expect that there are relatively few $\underline{\mathcal{M}}^{gen}$-functorial invariants. This is indeed the case and one has that the only $\underline{\mathcal{M}}^{gen}$-functorial invariants that are possible are the constant ones: as we saw in Remark 3.1, one property of the category $\underline{\mathcal{M}}^{gen}$ is that for any $X, Y \in \mathrm{ob}(\underline{\mathcal{M}}^{gen})$ there exist morphisms $\phi_1 \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, Y)$ and $\phi_2 \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(Y, X)$. Functoriality of $\mathfrak{I}$ would then require that $\mathfrak{I}(X) \geqslant \mathfrak{I}(Y) \geqslant \mathfrak{I}(X)$, and hence, since X, Y were arbitrary, $\mathfrak{I}$ must be constant.

6. STANDARD CLUSTERING

Definition 6.1. *For each $\delta > 0$ we define the **Vietoris-Rips clustering functor***

$$\mathfrak{R}_\delta : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{C}}$$

as follows. For a finite metric space (X, d_X), we set $\mathfrak{R}_\delta(X, d_X)$ to be $(X, P_X(\delta))$, where $P_X(\delta)$ is the partition of X associated to the equivalence relation $\sim_\delta$ defined in Example 2.1. We define how $\mathfrak{R}_\delta$ acts on maps $f : (X, d_X) \to (Y, d_Y)$: $\mathfrak{R}_\delta(f)$ is simply the set map f regarded as a morphism from $(X, P_X(\delta))$ to $(Y, P_Y(\delta))$ in $\underline{\mathcal{C}}$.

We now check that $\mathfrak{R}_\delta$ is in fact $\underline{\mathcal{M}}^{gen}$-functorial. Fix finite metric spaces (X, d_X) and (Y, d_Y) and a morphism $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, Y)$. It is enough to prove that whenever $x, x' \in X$ are s.t. $x \sim_\delta x'$, then $\phi(x) \sim_\delta \phi(x')$ in Y as well. Let $x = x_0, x_1, \ldots, x_n = x'$ be points in X with $d_X(x_i, x_{i+1}) \leqslant \delta$ for all $i = 0, 1, \ldots, n-1$. Then, since ϕ is distance non-increasing, $d_Y(\phi(x_i), \phi(x_{i+1})) \leqslant \delta$ for all $i = 0, 1, \ldots, n-1$ as well, and by Definition 2.1, then $\phi(x) \sim_\delta \phi(x')$.[3]

By restricting $\mathfrak{R}_\delta$ to the subcategories $\underline{\mathcal{M}}^{iso}$ and $\underline{\mathcal{M}}^{inj}$, we obtain functors $\mathfrak{R}_\delta^{iso} : \underline{\mathcal{M}}^{iso} \to \underline{\mathcal{C}}$ and $\mathfrak{R}_\delta^{inj} : \underline{\mathcal{M}}^{inj} \to \underline{\mathcal{C}}$. We will denote all these functors by $\mathfrak{R}_\delta$ when there is no ambiguity.

Remark 6.1 (The Vietoris-Rips functor is surjective.). *Among the desirable conditions singled out by Kleinberg [Kle02], one has that of surjectivity (which he referred to as "richness"). Given a finite set X and $P_X \in \mathcal{P}(X)$, surjectivity calls for the existence of a metric d_X on X such that $\mathfrak{R}_\delta(X, d_X) = (X, P_X)$. Indeed this is the case: pick any $\alpha > 1$ and let $d_X(x, x') = \delta$ whenever x and x', $x \neq x'$, are in the same block of P_X, and $d_X(x, x') = \alpha\delta$ when x and x' are in different blocks of P_X. With this definition it is easy to verify that d_X is an ultrametric on X and hence a metric. That $\mathfrak{R}_\delta(X, d_X) = (X, P_X)$ follows directly from Lemma 2.1 and the definition of the Vietoris-Rips functor.*

For $\underline{\mathcal{M}}$ being any one of our choices $\underline{\mathcal{M}}^{iso}$, $\underline{\mathcal{M}}^{inj}$ or $\underline{\mathcal{M}}^{gen}$, a clustering functor in this context will be denoted by $\mathfrak{C} : \underline{\mathcal{M}} \to \underline{\mathcal{C}}$. **Excisiveness** of a clustering functor refers to the property that once a finite metric space has been partitioned by the clustering procedure, it should not be further split by subsequent applications of the same algorithm.

Definition 6.2 (Excisive clustering functors). *We say that a clustering functor $\mathfrak{C}$ is **excisive** if for all $(X, d_X) \in \mathrm{ob}(\underline{\mathcal{M}})$, if we write $\mathfrak{C}(X, d_X) = (X, \{X_\alpha\}_{\alpha \in A})$, then*

$$\mathfrak{C}\left(X_\alpha, d_{X|X_\alpha \times X_\alpha}\right) = (X_\alpha, \{X_\alpha\}) \text{ for all } \alpha \in A.$$

---

[3]The Vietoris-Rips functor is actually just **single linkage clustering** as it is well known, see [CM08, CM10a]. The name comes from certain simplicial constructions used by Vietoris in the 1930s to study Čech cohomology theories of metric spaces. Much later, Rips used similar constructions in the context of geometric group theory, see [Hau95].
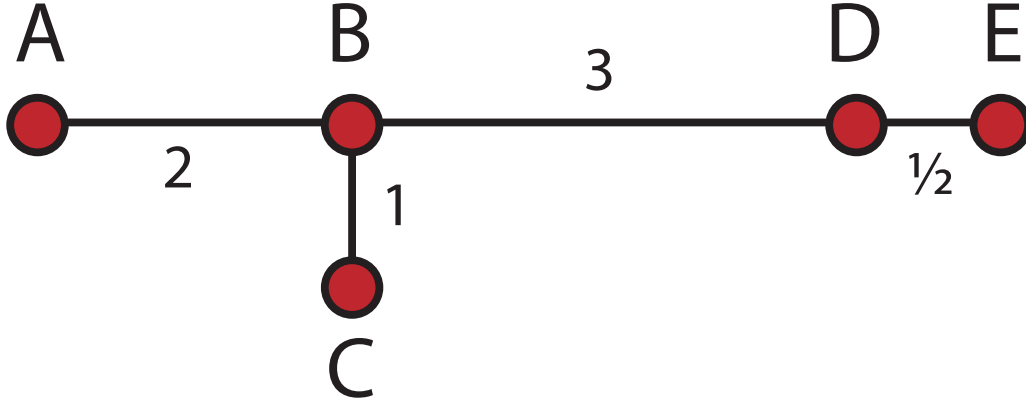
FIGURE 2. Metric space used to prove that the functor $\widehat{\mathfrak{R}} : \underline{\mathcal{M}}^{inj} \to \underline{\mathcal{C}}$ is not excisive. The metric is given by the graph distance on the graph.

Remark 6.2 (The Vietoris-Rips functor is excisive). *Fix $\delta > 0$ and consider the functor $\mathfrak{R}_\delta$ defined in Example 7.1. We claim that $\mathfrak{R}_\delta$ is excisive. Indeed, write $\mathfrak{R}_\delta(X, d) = (X, \{X_\alpha, \alpha \in A\})$. Then, X_α, $\alpha \in A$ are the different blocks of the partition of X into equivalence classes of $\sim_\delta$ ($x \sim_\delta x'$ if and only if $x, x' \in X_\alpha$ for some $\alpha \in A$). Fix $\alpha \in A$ and let $x, x' \in X_\alpha$. Let $x_0, x_1, \ldots, x_m \in X_\alpha$ be s.t. $x_0 = x$, $x_m = x'$ and $d_X(x_i, x_{i+1}) \leqslant \delta$ for $i \in \{0, \ldots, m-1\}$. But then it follows that $x_i \in X_\alpha$ for all $i \in \{0, \ldots, m\}$ as well and hence $\mathfrak{R}_\delta\left(X_\alpha, d_{X|_{X_\alpha \times X_\alpha}}\right) = (X_\alpha, \{X_\alpha\})$.*

There exist large families of non-excisive clustering functors in $\underline{\mathcal{M}}^{inj}$.

Example 6.1 (A family of non-excisive functors in $\underline{\mathcal{M}}^{inj}$). *For each invariant $\mathfrak{I} : \underline{\mathcal{M}}^{inj} \to \underline{\mathbb{E}}$ (recall their construction in §5) and non-increasing function $\eta : \mathbb{E} \to \mathbb{E}$, consider the clustering functor $\widehat{\mathfrak{R}} : \underline{\mathcal{M}}^{inj} \to \underline{\mathcal{C}}$ defined as follows: For a finite metric space (X, d_X), we define $\widehat{\mathfrak{R}}(X, d_X)$ to be $(X, \widehat{P}_X)$, where $\widehat{P}_X$ is the partition of X associated to the equivalence relation $\sim_{\eta(\mathfrak{I}(X))}$ on X. That $\widehat{\mathfrak{R}}$ is a functor follows from the fact that whenever $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(X, Y)$ and $x \sim_{\eta(\mathfrak{I}(X))} x'$, then $\phi(x) \sim_{\eta(\mathfrak{I}(Y))} \phi(x')$.*

Indeed, let $x, x' \in X$ be s.t. $x \sim_{\eta(\mathfrak{I}(X))} x'$ and let $x = x_0, \ldots, x_n = x'$ in X s.t. $d_X(x_i, x_{i+1}) \leqslant \eta(\mathfrak{I}(X))$ for $i = 0, \ldots, n-1$. Then, for $i = 0, \ldots, n-1$, $d_Y(\phi(x_i), \phi(x_{i+1})) \leqslant d_X(x_i, x_{i+1}) \leqslant \eta(\mathfrak{I}(X))$. Since η is non-increasing and $\mathfrak{I}(X) \geqslant \mathfrak{I}(Y)$, it follows that for $i = 0, 1, \ldots, n-1$, $d_Y(\phi(x_i), \phi(x_{i+1})) \leqslant \eta(\mathfrak{I}(Y))$, and hence $\phi(x) \sim_{\eta(\mathfrak{I}(Y))} \phi(x')$.

*Now, the functors $\widehat{\mathfrak{R}}$ are **not excisive** in general. An explicit example is the following, which applies to $\mathfrak{I} = \mathfrak{I}_{sep}$ and $\eta(z) = z^{-1}$. Consider the metric space (X, d_X) depicted in figure on the right, where the metric is given by the graph metric on the underlying graph (the edge weights are shown in the figure). Note that $sep(X, d_X) = 1/2$ and thus $\eta(\mathfrak{I}(X)) = 2$. We then find that $\widehat{\mathfrak{R}}(X, d_X) = (X, \{\{A, B, C\}, \{D, E\}\})$. But, $sep\left(\{A, B, C\}, \left(\begin{smallmatrix} 0 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 1 & 0 \end{smallmatrix}\right)\right) = 1$ and hence $\eta(\mathfrak{I}(\{A, B, C\})) = 1$. Therefore,*

$$\widehat{\mathfrak{R}}\left(\{A, B, C\}, \left(\begin{smallmatrix} 0 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 1 & 0 \end{smallmatrix}\right)\right) = (\{A, B, C\}, \{A, \{B, C\}\}),$$

and we see that $\{A, B, C\}$ gets further partitioned by $\widehat{\mathfrak{R}}$.

It is interesting to point out that the same construction of a non-excisive functor in $\underline{\mathcal{M}}^{gen}$ would not work, since as we saw in §5, all the invariants $\mathfrak{I} : \underline{\mathcal{M}}^{gen} \to \underline{\mathbb{E}}$ are constant. We will see in §6.5 that $\underline{\mathcal{M}}^{gen}$ permits the existence of only one (well behaved) clustering functor which turns out to be the Vietoris-Rips functor, which is excisive as we saw in Remark 6.2.

6.1. **The case of $\underline{\mathcal{M}}^{iso}$.** It is easy to describe all $\underline{\mathcal{M}}^{iso}$-functorial clustering schemes. Let $\mathcal{I}$ denote the collection of all isometry classes of finite metric spaces. Clearly, one must specify the clustering functor on each class of $\mathcal{I}$ in a manner which is invariant to self-symmetries of (an element of) the class. The precise statement is as follows. For each $\zeta \in \mathcal{I}$ let (X_ζ, d_{X_ζ}) denote an element of the class ζ, G_ζ the isometry group of (X_ζ, d_{X_ζ}), and Ξ_ζ the set of all fixed points of the action of G_ζ on $\mathcal{P}(X_\zeta)$. For completeness we state as a theorem the following obvious fact:

Theorem 6.1 (**Classification of $\underline{\mathcal{M}}^{iso}$-functorial clustering schemes**). *Any $\underline{\mathcal{M}}^{iso}$-functorial clustering scheme determines a choice of $p_\zeta \in \Xi_\zeta$ for each $\zeta \in \mathcal{I}$, and conversely, a choice of p_ζ for each $\zeta \in \mathcal{I}$ determines an $\underline{\mathcal{M}}^{iso}$-functorial scheme.*

6.2. **Representable Clustering Functors.** In what follows, $\underline{\mathcal{M}}$ is either of $\underline{\mathcal{M}}^{inj}$ or $\underline{\mathcal{M}}^{gen}$. For each $\delta > 0$ the Vietoris-Rips functor $\mathfrak{R}_\delta : \underline{\mathcal{M}} \to \underline{\mathcal{C}}$ can be described in an alternative way which relies on the ability to construct maps from the two point metric space $\Delta_2(\delta)$ into a given input finite metric space (X, d_X). A first trivial observation is that the condition that $x, x' \in X$ satisfy $d_X(x, x') \leqslant \delta$ is equivalent to requiring the existence of a map $f \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\Delta_2(\delta), X)$ with $\{x, x'\} \subset \mathrm{Im}(f)$. Using this observation, we can reformulate the condition that $x \sim_\delta x'$ by the requirement that there exist $z_0, z_1, \ldots, z_k \in X$ with $z_0 = x$, $z_k = x'$, and $f_1, f_2, \ldots, f_k \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\Delta_2(\delta), X)$ with $\{x_{i-1}, x_i\} \subset \mathrm{Im}(f_i)$ $\forall i = 1, 2, \ldots, k$. Informally, this points to the interpretation that $\{\Delta_2(\delta)\}$ is the "parameter" in a "generative model" for $\mathfrak{R}_\delta$.

This immediately suggests considering more general clustering functors constructed in the following manner. Let Ω be any fixed collection of finite metric spaces. Define a clustering functor $\mathfrak{C}^\Omega : \underline{\mathcal{M}} \to \underline{\mathcal{C}}$ as follows: let $(X, d) \in \mathrm{ob}(\underline{\mathcal{M}})$ and write $\mathfrak{C}^\Omega(X, d) = (X, \{X_\alpha\}_{\alpha \in A})$. One declares that points x and x' belong to the same block X_α if and only if there exist

- a sequence of points $z_0, \ldots, z_k \in X$ with $z_0 = x$ and $z_k = x'$,
- a sequence of metric spaces $\omega_1, \ldots, \omega_k \in \Omega$ and
- for each $i = 1, \ldots, k$, pairs of points $(\alpha_i, \beta_i) \in \omega_i$ and morphisms $f_i \in \mathrm{Mor}_{\underline{\mathcal{M}}}(w_i, X)$ s.t. $f_i(\alpha_i) = z_{i-1}$ and $f_i(\beta_i) = z_i$.

Also, we declare that $\mathfrak{C}^\Omega(f) = f$ on morphisms f. Notice that above one can assume that $z_0, z_1, \ldots, z_k$ all belong to X_α.

Remark 6.3. *Let Ω be a collection of finite metric spaces. Consider the clustering functor $\mathfrak{C}^\Omega : \underline{\mathcal{M}} \to \underline{\mathcal{C}}$ constructed as above. Then, clearly, $\mathfrak{C}^\Omega(\omega, d_\omega) = (\omega, \{\omega\})$ for all $\omega \in \Omega$. As a consequence, we have the following: If X is any finite metric space and $\omega \in \Omega$, then all $f \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega, X)$ are s.t. $\mathrm{Im}(f)$ is fully contained in a single block of the partition of X by $\mathfrak{C}^\Omega$.*

Remark 6.4. *Let $\Omega \supset \Omega' \neq \varnothing$, then $\mathfrak{C}^{\Omega'}$ produces partitions that are not coarser than those produced by $\mathfrak{C}^\Omega$.*

Definition 6.3. *We say that a clustering functor $\mathfrak{C}$ is **representable** whenever there exists a collection of finite metric spaces Ω such that $\mathfrak{C} = \mathfrak{C}^\Omega$. In this case, we say that $\mathfrak{C}$ is*

represented by Ω. We say that $\mathfrak{C}$ is **finitely representable** whenever $\mathfrak{C} = \mathfrak{C}^\Omega$ for some finite *collection of finite metric spaces* Ω.

As we saw above, the Vietoris-Rips functor $\mathfrak{R}_\delta$ is (finitely) represented by $\{\Delta_2(\delta)\}$.

6.3. **Representability and Excisiveness.** We now prove that representability and excisiveness are equivalent properties. Notice that excisiveness is an axiomatic statement whereas representability asserts existence of generative model for the clustering functor.

Theorem 6.2. *Let $\underline{\mathcal{M}}$ be either of $\underline{\mathcal{M}}^{inj}$ or $\underline{\mathcal{M}}^{gen}$. Then any clustering functor on $\underline{\mathcal{M}}$ is excisive if and only if it is representable.*

Proof. Assume that $\mathfrak{C}$ is an excisive clustering functor. Let

$$(6\text{–}6) \qquad \Omega := \big\{(B, d_{|B \times B}),\, B \in P,\, \text{where } (X,P) = \mathfrak{C}(X,d) \text{ and } (X,d) \in \mathcal{M}\big\}.$$

That is, Ω contains all the possible blocks (endowed with restricted metrics) of all partitions obtained by applying $\mathfrak{C}$ to all finite metric spaces. Notice that by definition of Ω, and since $\mathfrak{C}$ is excisive,

$$(6\text{–}7) \qquad\qquad\qquad \mathfrak{C}(\omega, d_\omega) = (\omega, \{\omega\}) \text{ for all } \omega \in \Omega.$$

We will now prove that $\mathfrak{C}$ is represented by Ω, that is $\mathfrak{C} = \mathfrak{C}^\Omega$. Fix any finite metric space (X, d_X) and write $(X,P) = \mathfrak{C}(X, d_X)$ and $(X, P') = \mathfrak{C}^\Omega(X, d_X)$. We know that Ω contains all blocks $B \in P$, and since the inclusion $B \hookrightarrow X$ is a morphism in $\underline{\mathcal{M}}^{inj}$, by Remark 6.3 one sees that $\mathfrak{C}^\Omega(B, d_{X_{|B \times B}}) = (B, \{B\})$. Now, by functoriality of $\mathfrak{C}^\Omega$, any two points x, x' in B must belong to the same block of P'. Since B was any block of P it follows that P is a refinement of P'.

This is depicted by the diagram below:

$$(6\text{–}8)$$
$$
\begin{array}{ccc}
(B, d_{X_{|B \times B}}) & \xrightarrow{\;\hookrightarrow\;} & (X, d_X) \\
{\scriptstyle \mathfrak{C}^\Omega}\big\downarrow & & \big\downarrow{\scriptstyle \mathfrak{C}^\Omega} \\
(B, \{B\}) & \xrightarrow{\;\mathfrak{C}^\Omega(\hookrightarrow)\;} & (X, P')
\end{array}
$$

Conversely, assume $x, x' \in B' \in P'$ and let $\omega_1, \dots, \omega_k \in \Omega$, $(\alpha_i, \beta_i) \in \omega_i$ for $i = 1, 2, \dots, k$, $z_0, z_1, \dots, z_k \in X$ and $f_1, f_2, \dots, f_k$ s.t. $f_i \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_i, X)$, $f_i(\alpha_i) = z_{i-1}$ and $f_i(\beta_i) = z_i$ for $i = 1, 2, \dots, k$. Consider for each $i \in \{1, 2, \dots, k\}$ the diagram on the right, where we have used (6–7). Then, since for each $i \in \{1, \dots, k\}$ α_i, β_i are in the same (unique) block of ω_i, z_{i-1} and z_i have to be in the same block of P.

$$
\begin{array}{ccc}
(\omega_i, d_{\omega_i}) & \xrightarrow{\;f_i\;} & (X, d_X) \\
{\scriptstyle \mathfrak{C}}\big\downarrow & & \big\downarrow{\scriptstyle \mathfrak{C}} \\
(\omega_i, \{\omega_i\}) & \xrightarrow{\;\mathfrak{C}(f_i)\;} & (X, P)
\end{array}
\;.
$$

Since this has to happen for all i, we conclude that $x = z_0$ and $z_k = x'$ ought to be in the same block of P. Thus P' is a refinement of P. We have proved that P refines and is refined by P' and therefore these two partitions must be equal. Since X was any finite metric space, we conclude that $\mathfrak{C}$ is represented by Ω and we are done.

Assume now that $\mathfrak{C}$ is representable, i.e. $\mathfrak{C} = \mathfrak{C}^\Omega$ for some collection of finite metric spaces Ω. We will prove that $\mathfrak{C}$ is excisive. Fix a finite metric space (X, d_X) and write

$\mathfrak{C}^\Omega(X, d_X) = (X, P)$. Pick any block B of P and any two points $x, x' \in B$. Let $\omega_1, \ldots, \omega_k \in \Omega$, $f_1 \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_1, X), \ldots, f_k \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_k, X)$; $(\alpha_1, \beta_1) \in \omega_1, \ldots, (\alpha_k, \beta_k) \in \omega_k$; and $z_0, z_1, \ldots, z_k$ be s.t. $z_0 = x$, $z_k = x'$, $f_i(\alpha_i) = z_{i-1}$, $f_i(\beta_i) = z_i$ for all $i = 1, 2, \ldots, k$. One can assume that all $z_i \in B$ and moreover, by Remark 6.3, $\mathrm{Im}(f_i) \subseteq B$. Hence, we can regard each f_i as a morphism in $\mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_i, B)$ where B is endowed with the restricted metric.

By Remark 6.3, $\mathfrak{C}(\omega, d_\omega) = (\omega, \{\omega\})$ for all $\omega \in \Omega$. Hence, by functoriality we have that $x, x' \in B$ are in the same block of P_B, where $(B, P_B) = \mathfrak{C}^\Omega(B, d_{X|_{B \times B}})$. Since $x, x' \in B$ were arbitrary we conclude that P_B consists of only one block. The arbitrariness in the choices of $B \in P$ and $X \in \mathrm{ob}(\underline{\mathcal{M}})$ finishes the proof that $\mathfrak{C} = \mathfrak{C}^\Omega$ is excisive. $\square$

6.4. A factorization theorem.

We now prove that all finitely representable clustering functors in $\underline{\mathcal{M}}^{gen}$ and $\underline{\mathcal{M}}^{inj}$ arise as the composition of the Vietoris-Rips functor with a functor that changes the metric.

For a given collection Ω of finite metric spaces let

$$(6\text{–}9) \qquad \qquad \mathfrak{T}^\Omega : \underline{\mathcal{M}} \to \underline{\mathcal{M}}$$

be the endofunctor that assigns to each finite metric space (X, d_X) the metric space (X, d_X^Ω) with the same underlying set and metric given by $d_X^\Omega = \mathcal{U}(W_X^\Omega)$,[4] where $W_X^\Omega : X \times X \to \mathbb{R}_+$ is given by

$$(6\text{–}10) \qquad (x, x') \mapsto \inf \left\{ \lambda > 0 \mid \exists w \in \Omega \text{ and } \phi \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\lambda \cdot \omega, X) \text{ with } \{x, x'\} \subset \mathrm{Im}(\phi) \right\},$$

for $x \neq x'$, and by 0 on $\mathrm{diag}(X \times X)$. Above we assume that the inf over the empty set equals $+\infty$. Note that $W_X^\Omega(x, x') < \infty$ for all $x, x' \in X$ as long as $|\omega| \leqslant |X|$ for some $\omega \in \Omega$. Also, $W_X^\Omega(x, x') = \infty$ for all $x \neq x'$ when $|X| < \inf\{|\omega|, \omega \in \Omega\}$.

Theorem 6.3. *Let $\underline{\mathcal{M}}$ be either $\underline{\mathcal{M}}^{gen}$ or $\underline{\mathcal{M}}^{inj}$ and $\mathfrak{C}$ be any $\underline{\mathcal{M}}$-functorial finitely representable clustering functor represented by some $\Omega \subset \mathcal{M}$. Then, $\mathfrak{C} = \mathfrak{R}_1 \circ \mathfrak{T}^\Omega$.*

Proof. Fix any finite metric space (X, d_X) and write $\mathfrak{C}(X, d_X) = (X, P)$ and $\mathfrak{R}_1 \circ \mathfrak{T}^\Omega(X, d_X) = (X, P')$. Assume that $x, x' \in B \in P$ and let $x = z_0, z_1, \ldots, z_k = x' \in X$, $\omega_1, \ldots, \omega_k \in \Omega$, $f_i \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_i, X)$ and $(\alpha_i, \beta_i) \in \omega_i$ s.t. $f_i(\alpha_i) = z_{i-1}$, $f_i(\beta_i) = z_i$ for $i = 1, 2, \ldots, k$. Then, from (6–10) for all $i = 1, 2, \ldots, k$, $W_X^\Omega(z_{i-1}, z_i) \leqslant 1$ and in particular $\max_i W_X^\Omega(z_{i-1}, z_i) \leqslant 1$. It follows from (2–1) that $d_X^\Omega(x, x') \leqslant 1$ and consequently x and x' are in the same block of P'. Thus we see that P refines P'.

Assume now that $x, x' \in B' \in P'$. By definition of the Vietoris-Rips functor, $x \sim_1 x'$ in (X, d_X^Ω), which implies the existence of $x = x_0, x_1, \ldots, x_k = x'$ in X with $d_X^\Omega(x_i, x_{i+1}) \leqslant 1$ for all $i = 1, \ldots, k - 1$. Now, by (2–1), for each $i = 0, 1, \ldots, k$ there exist $y_0^{(i)}, y_1^{(i)}, \ldots, y_{k_i}^{(i)}$ in X with $y_0^{(i)} = x_i$, $y_{k_i}^{(i)} = x_{i+1}$ and $W_X^\Omega\left(y_j^{(i)}, y_{j+1}^{(i)}\right) \leqslant 1$ for $j = 0, 1, \ldots, k_i - 1$. By concatenating all sequences $\left\{y_j^{(i)}\right\}_{j=1}^{k_i}$ for $i = 0, 1, \ldots, k$, we obtain a sequence $z_0, z_1, \ldots, z_N \in X$ with $z_0 = x$, $z_N = x'$ and $W_X^\Omega(z_i, z_{i+1}) \leqslant 1$. By definition of W_X^Ω and because Ω is finite, one sees that for every $i = 1, 2, \ldots, N$ there exists $\lambda_i \in (0, 1]$, $\omega_i \in \Omega$ and $\phi_i \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\lambda_i \cdot \omega_i, X)$ with $\{z_{i-1}, z_i\} \subset \mathrm{Im}(\phi_i)$. But then, obviously, $\phi_i \in \mathrm{Mor}_{\underline{\mathcal{M}}}(\omega_i, X)$ as well, and since $\mathfrak{C}$ is represented by Ω one sees that x and x' must be in the same block of P. This proves that P' refines P, and finishes the proof. $\square$

6.5. The case of $\underline{\mathcal{M}}^{gen}$.

---

[4] Recall (2–1), the definition of $\mathcal{U}$.

6.5.1. *A Uniqueness theorem.* In $\underline{\mathcal{M}}^{gen}$ clustering functors are very restricted, as reflected by the following theorem.

Theorem 6.4. *Assume that $\mathfrak{C} : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{C}}$ is a clustering functor for which there exists $\delta_{\mathfrak{C}} > 0$ with the property that*
- *$\mathfrak{C}(\Delta_2(\delta))$ is in one piece for all $\delta \in [0, \delta_{\mathfrak{C}}]$, and*
- *$\mathfrak{C}(\Delta_2(\delta))$ is in two pieces for all $\delta > \delta_{\mathfrak{C}}$.*

Then, $\mathfrak{C}$ is the Vietoris-Rips functor with parameter $\delta_{\mathfrak{C}}$. i.e. $\mathfrak{C} = \mathfrak{R}_{\delta_{\mathfrak{C}}}$.

Recall that the Vietoris-Rips functor is excisive, cf. Remark 6.2.

Remark 6.5 (About uniqueness in $\underline{\mathcal{M}}^{gen}$). *For each $\delta > 0$ consider the clustering functor $\overset{\circ}{\mathfrak{R}}_{\delta}$ that assigns points x, x' in a metric space (X, d_X) to the same block if and only if there exist $x_0, x_1, \ldots, x_k \in X$ with $x_0 = x$, $x_k = x'$ and $d_X(x_i, x_{i+1}) < \delta$ for $i = 0, 1, \ldots, k-1$. Note the* strict *inequality, cf. with the definition of the Vietoris-Rips functor in Example 7.1.*

If one changes the conditions in Theorem 6.4 in a way such that
- *$\mathfrak{C}(\Delta_2(\delta))$ is in one piece for all $\delta \in [0, \delta_{\mathfrak{C}})$, and*
- *$\mathfrak{C}(\Delta_2(\delta))$ is in two pieces for all $\delta \geqslant \delta_{\mathfrak{C}}$,*

then, $\mathfrak{C} = \overset{\circ}{\mathfrak{R}}_{\delta_{\mathfrak{C}}}$.

Proof of Theorem 6.4. Fix a finite metric space (X, d_X) and write $(X, P) = \mathfrak{C}(X, d_X)$ and $(X, P') = \mathfrak{R}_{\delta_{\mathfrak{C}}}(X, d_X)$.

(1) Let $x, x' \in B$ for some block $B \in P'$. We will prove that then x, x' are contained in the same block of P. Let $x_0, x_1, \ldots, x_k \in X$ be s.t. $x_0 = x$, $x_k = x'$ and $d_X(x_i, x_{i+1}) \leqslant \delta_{\mathfrak{C}}$ for all $i \in \{1, \ldots, k-1\}$. We now prove that each pair $\{x_i, x_{i+1}\}$ belongs to the same block of P', what will yield the claim. Indeed, since for each i, $d_X(x_i, x_{i+1}) \leqslant \delta_{\mathfrak{C}}$, then for each i there exists $f_i \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(\Delta_2(\delta_{\mathfrak{C}}), X)$ with $\mathrm{Im}(f_i) = \{x_i, x_{i+1}\}$. Thus, by functoriality and the definition of $\delta_{\mathfrak{C}}$, x_i and x_{i+1} must fall in the same block of P'. This concludes the first half of the proof.

(2) Assume that $x_0, x_0' \in X$ are in different blocks of P'. We will prove that then x_0 and x_0' must belong to different blocks of P. Let $\delta := \min\{d_X(x, x'), x \in X \backslash [x_0']_{\delta_{\mathfrak{C}}}, x' \in [x_0']_{\delta_{\mathfrak{C}}}\}$. Notice that by Lemma 2.1, $\delta > \delta_{\mathfrak{C}}$. Write $\Delta_2(\delta) = \left(\{p, q\}, \left(\begin{smallmatrix} 0 & \delta \\ \delta & 0 \end{smallmatrix}\right)\right)$ and define the map $\phi : X \to \Delta_2(\delta)$ given by $\phi(x) = p$ for all $x \in X \backslash [x_0']_{\delta_{\mathfrak{C}}}$, and $\phi(x') = q$ for all $x' \in [x_0']_{\delta_{\mathfrak{C}}}$. We claim that $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, \Delta_2(\delta))$. Indeed, if $x, x' \in X \backslash [x_0']_{\delta_{\mathfrak{C}}}$ or $x, x' \in [x_0']_{\delta_{\mathfrak{C}}}$, $\phi(x) = \phi(x')$ and there's nothing to check. Assume then that $x \in X \backslash [x_0']_{\delta_{\mathfrak{C}}}$ and $x' \in [x_0']_{\delta_{\mathfrak{C}}}$, and note that by definition of δ, $d_X(x, x') \geqslant \delta$. On the other hand, $p = \phi(x)$ and $q = \phi(x')$ are at distance δ from each other. Thus ϕ is distance non-increasing. Now, we apply $\mathfrak{C}$ to X and $\Delta_2(\delta)$ and note that in this case, by our assumption on $\mathfrak{C}$, since $\delta > \delta_{\mathfrak{C}}$, then $\mathfrak{C}(\Delta_2(\delta))$ is in two pieces. By functoriality we conclude that x_0 and x_0' cannot lie in the same block of P. $\qquad\square$

6.6. **Scale invariance in $\underline{\mathcal{M}}^{gen}$ and $\underline{\mathcal{M}}^{inj}$.** It is interesting to consider the effect of imposing Kleinberg's scale invariance axiom on $\underline{\mathcal{M}}^{gen}$-functorial and $\underline{\mathcal{M}}^{inj}$-functorial clustering schemes. It turns out that in $\underline{\mathcal{M}}^{gen}$ there are only two possible clustering schemes enjoying scale invariance, which turn out to be the trivial ones:

Theorem 6.5. *Let $\mathfrak{C} : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{C}}$ be a clustering functor s.t. $\mathfrak{C} \circ \sigma_{\lambda} = \mathfrak{C}$ for all $\lambda > 0$. Then, either*

- $\mathfrak{C}$ *assigns to each finite metric space X the partition of X into singletons, or*
- $\mathfrak{C}$ *assigns to each finite metric the partition with only one block.*

Proof. Let $\mathfrak{C} : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{C}}$ be a scale invariant clustering functor. Consider first the metric space $\Delta_2(1)$ consisting of two points at distance 1. There are two possiblities: either **(1)** $\mathfrak{C}(\Delta_2(1))$ is in one piece, or **(2)** $\mathfrak{C}(\Delta_2(1))$ in two pieces.

Assume **(1)**. By scale invariance, the partition of $\Delta_2(\delta)$ given by $\mathfrak{C}(\Delta_2(\delta))$ is in one piece for all $\delta > 0$. Fix any finite metric space (X, d_X) and write $\mathfrak{C}(X, d_X) = (X, P)$. Pick any $x, x' \in X$, let $\delta = d_X(x, x')$, and consider the morphism $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(\Delta_2(\delta), X)$ s.t. $\phi(p) = x$ and $\phi(q) = x'$, where p and q are the two elements of $\Delta_2(\delta)$. By scale invariance and functoriality, x and x' must then be in the same block of P. Since x and x' were arbitrary, P must be in one piece as well. It follows from the arbitrariness of X that $\mathfrak{C}$ then assigns the single block partition to any finite metric space.

Now assume **(2)**. By scale invariance $\mathfrak{C}(\Delta_2(\delta))$ is in two pieces for all $\delta > 0$. Pick any finite metric space (X, d_X) and write $\mathfrak{C}(X, d_X) = (X, P)$. Pick any $x, x' \in X$, and a partition of X into subsets A and B such that $x \in A$ and $x' \in B$. Let $\delta = \mathrm{sep}(X)$, and consider the map $\psi : X \to \Delta_2(\delta)$ given by $\psi(a) = p$ for all $a \in A$ and $\psi(b) = q$ for all $b \in B$. It is easy to verify that $\psi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{gen}}(X, \Delta_2(\delta))$. By scale invariance and functoriality, x and x' cannot be in the same block of P. Since x and x' were arbitrary, no two points belong to the same block of P, and hence $\mathfrak{C}$ assigns X the partition into singletons. It follows from the arbitrariness of X that $\mathfrak{C}$ assigns to every finite metric space its partition into singletons. This finishes the proof. $\qquad\qquad\square$

By refining the proof of the previous theorem, we find that the behavior of any $\underline{\mathcal{M}}^{inj}$-functorial clustering functor is also severely restricted in each $\mathcal{M}_k$, $k \in \mathbb{N}$.[5]

Theorem 6.6. *Let $\mathfrak{C} : \underline{\mathcal{M}}^{inj} \to \underline{\mathcal{C}}$ be a clustering functor s.t. $\mathfrak{C} \circ \sigma_\lambda = \mathfrak{C}$ for all $\lambda > 0$. Let*

$$K(\mathfrak{C}) := \big\{ k \geqslant 2;\ \mathfrak{C}(\Delta_k(1))\ \text{is in one piece} \big\}.$$

- *If $K(\mathfrak{C}) = \varnothing$, then $\mathfrak{C}$ assigns to each finite metric space X its partition into singletons.*
- *Otherwise, let $k_{\mathfrak{C}} = \min K(\mathfrak{C})$. Then,*
  - *For each $k \geqslant k_{\mathfrak{C}}$, $\mathfrak{C}$ assigns to each finite metric space $X \in \mathcal{M}_k$ the partition of X with only one block.*
  - *For each $2 \leqslant k < k_{\mathfrak{C}}$, $\mathfrak{C}$ assigns to each finite metric space $X \in \mathcal{M}_k$ the partition of X into k singletons.*

Proof. Assume first that $K(\mathfrak{C}) = \varnothing$, pick any finite metric space X and let $k = |X|$. Pick any two distinct points $x, x' \in X$. Choose any injective map $\phi : X \to \Delta_k(\mathrm{sep}(X))$ such that $\phi(x)$ and $\phi(x')$ fall in different blocks of the partition of $\Delta_k(\mathrm{sep}(X))$ produced by $\mathfrak{C}$. Then, functoriality guarantees that x and x' must belong to different blocks of the partition of X generated by $\mathfrak{C}$. Since $x, x' \in X$ were arbitrary we conclude that $\mathfrak{C}$ partitions X into singletons.

Now assume that $K(\mathfrak{C}) \neq \varnothing$. Pick $k \geqslant k_{\mathfrak{C}}$ and $X \in \mathcal{M}_k$. Consider any morphism $\phi \in \mathrm{Mor}_{\underline{\mathcal{M}}^{inj}}(\Delta_k(\mathrm{diam}(X)), X)$ arising from arbitrarily choosing an injection from the underlying set of $\Delta_k(1)$ to X. Then, by functoriality, this forces $\mathfrak{C}(X, d_X)$ to produce a partition of X into exactly one block.

---

[5]Recall that $\mathcal{M}_k$ denotes the collection of all k-point metric spaces.

Finally, for $2 \leqslant k < k_{\mathfrak{C}}$, $\mathfrak{C}$ partitions $\Delta_k(1)$ into more than piece. An argument similar to the one given for the case $K(\mathfrak{C}) = \varnothing$ proves that then $\mathfrak{C}$ partitions any $X \in \mathcal{M}_k$ into (k) singletons.

$\square$

6.7. Clustering functors on $\underline{\mathcal{M}}^{inj}$ and density.

In the case of $\underline{\mathcal{M}}^{inj}$, there is a richer class of allowed clustering schemes that includes functors other than the Vietoris-Rips one. We already saw in Example 6.1 that in $\underline{\mathcal{M}}^{inj}$ there are non-excisive clustering schemes that by construction are different from the Vietoris-Rips functor. A class of excisive $\underline{\mathcal{M}}^{inj}$-functorial clustering schemes which has practical interest is the following. Consider for each $n \in \mathbb{N}$ and
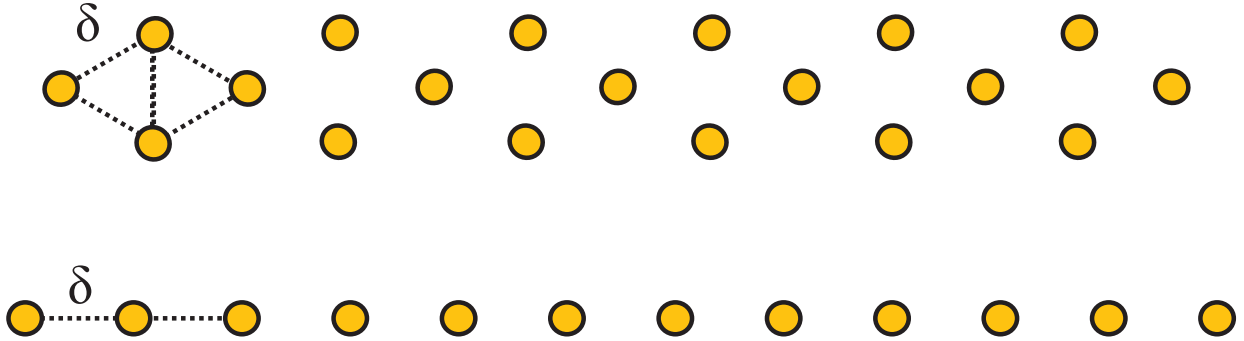

FIGURE 3. Fix $\delta > 0$ and consider the two clustering functors $\mathfrak{C}^{\Delta_2(\delta)}$ and $\mathfrak{C}^{\Delta_3(\delta)}$ on $\underline{\mathcal{M}}^{inj}$. Then $\mathfrak{C}^{\Delta_2(\delta)}$ applied to both metric spaces in the Figure will yield a single cluster. Meanwhile, $\mathfrak{C}^{\Delta_3(\delta)}$ will yield a single cluster for the metric space on right, but it will not connect the points of the metric space on the left. This admits the interpretation that in order for $\mathfrak{C}^{\Delta_3(\delta)}$ to produce a cluster, a certain degree of agglomeration of the points is necessary. This reflects the fact that this clustering scheme is more sensitive to density than its two-point counterpart $\mathfrak{C}^{\Delta_2(\delta)} = \mathfrak{R}_\delta$. Similar considerations apply to the general case of $\mathfrak{C}^{\Delta_k(\delta)}$, any $k \in \mathbb{N}$

$\delta \in \mathbb{R}_+$ the clustering functor $\mathfrak{C}^{\Delta_n(\delta)}$, and notice that this clustering method implements the notion that in order to deem a two points x and x' in a finite metric space X to be in the same cluster one requires that they are "tightly packed together" in the sense that there exists a sequence of points $x_0, x_1, \ldots, x_k$ connecting x to x' such that each consecutive pair $\{x_i, x_{i+1}\}$ is contained in a subset S_i of X of cardinality n whose inter-point distances do not exceed δ. This notion is closely related to proposals in network clustering [PDFV05] and the DBSCAN algorithm [EpKSX96], and for $n > 2$ provides a way to tackle the so called "chaining effect" exhibited by single linkage clustering (i.e. $\mathfrak{C}^{\Delta_2(\delta)}$), see Figure 3. Note that Theorem 6.3 implies that for a given finite metric space (X, d_X), $\mathfrak{C}^{\Delta_3(\delta)}(X, d_X)$ can be obtained by first applying a change of metric to X and then applying the standard Vietoris-Rips functor to the resulting metric space. The change of metric specified by the theorem leads to finding all 3-cliques in X whose diameter is not greater than δ.

Corollary 6.1. *On $\underline{\mathcal{M}}^{inj}$, for all $m \in \mathbb{N}$*

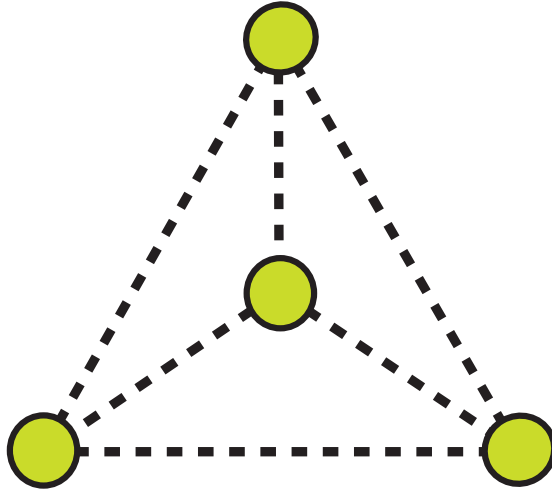$$\mathfrak{C}^{\Delta_m(\delta)} = \mathfrak{R}_\delta \circ \mathfrak{T}^m.$$

FIGURE 4. A metric space that encodes a notion of "density" different from that provided by $\Delta_3(\delta)$. .

Above, we have used the shorthand notation $\mathfrak{T}^m = \mathfrak{T}^\Omega$, for $\Omega = \{\Delta_m(\delta),\ \delta \geqslant 0\}$.

The formalism of representable clustering functors allows for far more generality than that provided by families $\{\Delta_n(\delta),\ n \in \mathbb{N},\ \delta > 0\}$. For example, one could encode the requirement that clusters be densely connected by considering a clustering functor represented by the metric space in Figure 4t, which consists of the vertices of an equilateral triangle of side δ, together with its center.

7. Hierarchical Clustering

Example 7.1 (A hierarchical version of the Vietoris-Rips functor). *We define a functor*

$$\mathfrak{R} : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{P}}$$

as follows. For a finite metric space (X, d_X), *we define* (X, d_X) *to be the persistent set* $(X, \theta_X^{\mathrm{VR}})$, *where* $\theta_X^{\mathrm{VR}}(r)$ *is the partition associated to the equivalence relation* $\sim_r$ *defined in Example 2.1. This is clearly an object in* $\underline{\mathcal{P}}$. *We also define how* $\mathfrak{R}$ *acts on maps* $f :$ $(X, d_X) \to (Y, d_Y)$: *The value of* $\mathfrak{R}(f)$ *is simply the set map* f *regarded as a morphism from* $(X, \theta_X^{\mathrm{VR}})$ *to* $(Y, \theta_Y^{\mathrm{VR}})$ *in* $\underline{\mathcal{P}}$. *That it is a morphism in* $\underline{\mathcal{P}}$ *is easy to check.*

Clearly, this functor implements the hierarchical version of single linkage clustering in the sense that for each $\delta \geqslant 0$, *if one writes* $\mathfrak{R}_\delta(X, d_X) = (X, P_X(\delta))$, *then* $P_X(\delta) = \theta_X^{\mathrm{VR}}(\delta)$.

7.1. **Functoriality over** $\underline{\mathcal{M}}^{iso}$**.** This is the smallest category we will deal with. The morphisms in $\underline{\mathcal{M}}^{iso}$ are simply the bijective maps between datasets which preserve the distance function. As such, functoriality of a clustering algorithms over $\underline{\mathcal{M}}^{iso}$ simply means that the output of the scheme doesn't depend on any artifacts in the dataset, such as the way the points are named or the way in which they are ordered.

Agglomerative hierarchical clustering, in standard form, as described for example in [JD88], begins with point cloud data and constructs a *binary* tree (or dendrogram) which describes

the merging of clusters as a threshold is increased. The lack of functoriality comes from the fact that when a single threshold value corresponds to more than one data point, one is forced to choose an ordering in order to decide which points to "agglomerate" first. This can easily be modified by relaxing the requirement that the tree be binary. This is what we did in Example 2.2 In this case, one can view these methods as functorial on $\underline{\mathcal{M}}^{iso}$, where the functor takes its values in arbitrary rooted trees. It is understood that in this case, the notion of morphism for the output ($\underline{\mathcal{P}}$) is simply isomorphism of rooted trees. In contrast, we see next that amongst these methods, when we impose that they be functorial over the larger (more demanding) category $\underline{\mathcal{M}}^{gen}$ then only one of them passes the test.[6]

7.2. **Complete and Average linkage are not functorial over $\underline{\mathcal{M}}^{inj}$.** It is interesting to point out why complete linkage and average linkage (agglomerative) clustering, as defined in Example 2.2 fail to be functorial on $\underline{\mathcal{M}}^{inj}$ (and therefore also on $\underline{\mathcal{M}}^{gen}$ as well) . A simple example explains this: consider the metric spaces $X = \{A, B, C\}$ with metric given by the edge lengths $\{4, 3, 5\}$ and $Y = (A', B', C')$ with metric given by the edge lengths $\{4, 3, 2\}$, as given in Figure 5. Obviously the map f from X to Y with $f(A) = A'$, $f(B) = B'$ and $f(C) = C'$ is a morphism in $\underline{\mathcal{M}}^{inj}$. Note that for example for $r = 3.5$ (shaded regions of the dendrograms in Figure 5) we have that the partition of X is $\Pi_X = \{\{A, C\}, B\}$ whereas the partition of Y is $\Pi_Y = \{\{A', B'\}, C'\}$ and thus $f^*(\Pi_Y) = \{\{A, B\}, \{C\}\}$. Therefore Π_X does not refine $f^*(\Pi_Y)$ as required by functoriality. The same construction yields a counter-example for average linkage.
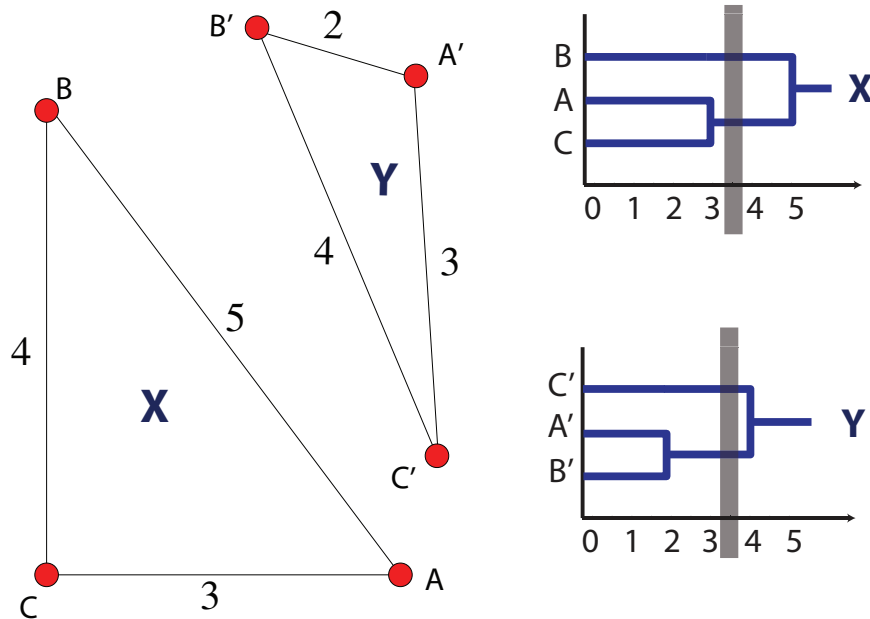

FIGURE 5. An example that shows why complete linkage fails to be functorial on $\underline{\mathcal{M}}^{inj}$ (and therefore also on $\underline{\mathcal{M}}^{gen}$).

---

[6]The result in Theorem 7.1 is actually more powerful in that it states that there is a unique functor from $\underline{\mathcal{M}}^{gen}$ to $\underline{\mathcal{P}}$ that satisfies certain natural conditions.

7.3. **Functoriality over $\underline{\mathcal{M}}^{gen}$: a uniqueness theorem.** We prove a theorem of the same flavor as the main theorem of [Kle02], except that we prove existence and uniqueness on $\underline{\mathcal{M}}^{gen}$ instead of impossibility in our context.

Theorem 7.1. *Let* $\mathfrak{H} : \underline{\mathcal{M}}^{gen} \to \underline{\mathcal{P}}$ *be a hierarchical clustering functor which satisfies the following conditions.*

(I): *Let* $\alpha : \underline{\mathcal{M}}^{gen} \to \underline{Sets}$ *and* $\beta : \underline{\mathcal{P}} \to \underline{Sets}$ *be the forgetful functors* $(X, d_X) \to X$ *and* $(X, \theta_X) \to X$, *which forget the metric and persistent set respectively, and only "remember" the underlying sets* X. *Then we assume that* $\beta \circ \mathfrak{H} = \alpha$. *This means that the underlying set of the persistent set associated to a metric space is just the underlying set of the metric space.*

(II): *For* $\delta \geqslant 0$ *let* $\Delta_2(\delta) = (\{p, q\}, \left(\begin{smallmatrix} 0 & \delta \\ \delta & 0 \end{smallmatrix}\right))$ *denote the two point metric space with underlying set* $\{p, q\}$, *and where* $dist(p, q) = \delta$. *Then* $\mathfrak{H}(\Delta_2(\delta))$ *is the persistent set* $(\{p, q\}, \theta_{\Delta_2(\delta)})$ *whose underlying set is* $\{p, q\}$ *and where* $\theta_{\Delta_2(\delta)}(t)$ *is the partition with one element blocks when* $t < \delta$ *and is the partition with a single two point block when* $t \geqslant \delta$.

(III): *Write* $\mathfrak{H}(X, d_X) = (X, \theta^{\mathfrak{H}})$, *then for any* $t < sep(X)$, *the partition* $\theta^{\mathfrak{H}}(t)$ *is the discrete partition with one element blocks.*

Then $\mathfrak{H}$ *is equal to the functor* $\mathfrak{R}$.

We should point out that another characterization of single linkage has been obtained in the book [JS71].

Proof of Theorem 7.1. Write $\mathfrak{H}(X, d_X) = (X, \theta_X^{\mathfrak{H}})$. For each $r \geqslant 0$ we will prove that **(a)** $\theta_X^{\mathrm{VR}}(r)$ is a refinement of $\theta_X^{\mathfrak{H}}(r)$ and **(b)** $\theta_X^{\mathfrak{H}}(r)$ is a refinement of $\theta_X^{\mathrm{VR}}(r)$. Then it will follow that $\theta_X^{\mathrm{VR}}(r) = \theta_X^{\mathfrak{H}}(r)$ for all $r \geqslant 0$, which shows that the objects are the same. Since this is a situation where, given any pair of objects, there is at most one morphism between them, this also determines the effect of the functor on morphisms.

Fix $r \geqslant 0$. In order to obtain (a) we need to prove that whenever $x, x' \in X$ lie in the same block of the partition $\theta_X^{\mathrm{VR}}(r)$, that is $x \sim_r x'$, then they both lie in the same block of $\theta_X^{\mathfrak{H}}(r)$.

skip It is enough to prove the following

Claim 7.1. *Whenever* $d_X(x, x') \leqslant r$ *then* x *and* x' *lie in the same block of* $\theta_X^{\mathfrak{H}}(r)$.

Indeed, if the claim is true, and $x \sim_r x'$ then one can find $x_0, x_1, \ldots, x_n$ with $x_0 = x$, $x_n = x'$ and $d_X(x_i, x_{i+1}) \leqslant r$ for $i = 0, 1, 2, \ldots, n - 1$. Then, invoking the claim for all pairs (x_i, x_{i+1}), $i = 0, \ldots, n - 1$ one would find that: $x = x_0$ and x_1 lie in the same block of $\theta_X^{\mathfrak{H}}(r)$, x_1 and x_2 lie in the same block of $\theta_X^{\mathfrak{H}}(r)$, ..., x_{n-1} and $x_n = x'$ lie in the same block of $\theta_X^{\mathfrak{H}}(r)$. Hence, x and x' lie in the same block of $\theta_X^{\mathfrak{H}}(r)$.

So, let's prove the claim. Assume $d_X(x, x') \leqslant r$, then the function given by $p \to x, q \to x'$ is a morphism $g : \Delta_2(r) \to (X, d_X)$ in $\underline{\mathcal{M}}^{gen}$. This means that we obtain a morphism

$$\mathfrak{H}(g) : \mathfrak{H}(\Delta_2(r)) \to \mathfrak{H}(X, d_X)$$

in $\underline{\mathcal{P}}$. But, by assumption (II), p and q lie in the same block of the partition $\theta_{\Delta_2(r)}(r)$. By functoriality it follows that $\mathfrak{H}(g)$ is persistence preserving and hence the elements $g(p) = x$ and $g(q) = x'$ lie in the same block of $\theta_X^{\mathfrak{H}}(r)$. This concludes the proof of (a).

skip For condition (b), assume that x and x' belong to the same block of the partition $\theta_X^{\mathfrak{H}}(r)$ for some $r \geqslant 0$. We will prove that necessarily $x \sim_r x'$. This of course will imply that x and x' belong to the same block of $\theta_X^{\mathrm{VR}}(r)$.

skip Consider the metric space (X_r, d_r) whose points are the equivalence classes of X under the equivalence relation $\sim_r$, and where the distance $d_r(\mathcal{B}, \mathcal{B}')$ between two equivalence classes $\mathcal{B}$ and $\mathcal{B}'$ is defined to be the maximal ultra-metric pointwisely less than or equal to

$$W_X(\mathcal{B}, \mathcal{B}') := \min_{x \in \mathcal{B}} \min_{x' \in \mathcal{B}'} d_X(x, x'),$$

that is $d_r = \mathcal{U}(W_X)$, see (2–1). It follows from the definition of $\sim_r$ and Lemma 2.1 that if two equivalence classes are distinct, then the distance between them is $> r$. Indeed, for otherwise there would be $x, x' \in X$ with $[x]_r \neq [x']_r$ and $d_r([x]_r, [x']_r) \leqslant r$, which would imply that $W_X([x]_r, [x']_r) \leqslant r$, which is a contradiction. Hence $\mathrm{sep}(X_r, d_r) > r$.

Write $\mathfrak{H}(X_r, d_r) = (X_r, \theta^{\mathfrak{H}}_{X_r})$. Since $\mathrm{sep}(X_r) > r$, hypothesis (III) now directly shows that the blocks of the partition $\theta^{\mathfrak{H}}_{X_r}(r)$ are exactly the equivalence classes of X under the equivalence relation $\sim_r$, that is $\theta^{\mathfrak{H}}_{X_r}(r) = \theta^{\mathrm{VR}}_X(r)$. Finally, consider the morphism

$$\pi_r : (X, d_X) \to (X_r, d_r)$$

in $\underline{\mathcal{M}}^{gen}$ given on elements $x \in X$ by $\pi_r(x) = [x]_r$, where $[x]_r$ denotes the equivalence class of x under $\sim_r$. By functoriality, $\mathfrak{H}(\pi_r) : (X, \theta^{\mathfrak{H}}_X) \to (X_r, \theta^{\mathfrak{H}}_{X_r})$ is persistence preserving, and therefore, $\theta^{\mathfrak{H}}(r)$ is a refinement of $\theta_{X_r}(r) = \theta^{\mathrm{VR}}_X(r)$. This is depicted as follows:

$$
\begin{array}{ccc}
(X, d_X) & \xrightarrow{\mathfrak{H}} & (X, \theta^{\mathrm{VR}}_X) \\
{\scriptstyle \pi_r} \downarrow & & \downarrow {\scriptstyle \mathfrak{H}(\pi_r)} \\
(X_r, d_r) & \xrightarrow{\mathfrak{H}} & (X_r, \theta^{\mathfrak{H}}_{X_r})
\end{array}
$$

This concludes the proof of (b). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

7.3.1. *Comments on Kleinberg's conditions.* We conclude this section by observing that analogues of the three (axiomatic) properties considered by Kleinberg in [Kle02] hold for $\mathfrak{R}$.

- Kleinberg's first condition was *scale-invariance*, which asserted that if the distances in the underlying point cloud data were multiplied by a constant positive multiple λ, then the resulting clustering decomposition should be identical. In our case, this is replaced by the condition (Recall notation of Example 4.3) that $\mathfrak{R} \circ \sigma_\lambda(X, d_X) = s_\lambda \circ \mathfrak{R}(X, d_X)$, which is trivially satisfied.
- Kleinberg's second condition, *richness*, asserts that any partition of a dataset can be obtained as the result of the given clustering scheme for some metric on the dataset. In our context, partitions are replaced by persistent sets. Assume that there exist $t \in \mathbb{R}_+$ s.t. $\theta_X(t)$ is the single block partition, i.e., impose that the persistent set is a dendrogram (cf. Definition 2.2). In this case, it is easy to check that any such persistent set can be obtained as $\mathcal{R}^{gen}$ evaluated for some (pseudo)metric on some dataset. Indeed,[7] let $(X, \theta_X) \in ob(\underline{\mathcal{P}})$. Let $\epsilon_1, \ldots, \epsilon_k$ be the (finitely many) transition/discontinuity points of θ_X. For $x, x' \in X$ define $d_X(x, x') = \min\{\epsilon_i\}$ s.t. x, x' belong to same block of $\theta_X(\epsilon_i)$.

  This is a pseudo metric on X. Indeed, pick points x, x' and x'' in X. Let ϵ_1 and ϵ_2 be minimal s.t. x, x' belong to the same block of $\theta_X(\epsilon_1)$ and x', x'' belong to the same block of $\theta_X(\epsilon_2)$. Let $\epsilon_{12} := \max(\epsilon_1, \epsilon_2)$. Since (X, θ_X) is a persistent set (Definition 2.2), $\theta_X(\epsilon_{12})$ must have a block $\mathcal{B}$ s.t. x, x' and x'' all lie in $\mathcal{B}$. Hence $d_X(x, x'') \leqslant \epsilon_{12} \leqslant \epsilon_1 + \epsilon_2 = d_X(x, x') + d_X(x', x'')$.

---

[7] We only prove triangle inequality.

- Finally, Kleinberg's third condition, *consistency*, could be viewed as a rudimentary example of functoriality. His morphisms are similar to the ones in $\underline{\mathcal{M}}^{gen}$.

7.4. **Functoriality over $\underline{\mathcal{M}}^{inj}$.** In this section, we illustrate how relaxing the functoriality permits more clustering algorithms. In other words, we will restrict ourselves to $\underline{\mathcal{M}}^{inj}$ which is smaller (less stringent) than $\underline{\mathcal{M}}^{gen}$ but larger (more stringent) than $\underline{\mathcal{M}}^{iso}$.

Fir a finite metric space (X, d_X) and $r \geqslant 0$. For $x \in X$, let $[x]_r$ be the equivalence class of x under the equivalence relation $\sim_r$, and define $c(x) = \big| [x]_r \big|$. For any positive integer m, we now define $X_m \subseteq X$ by $X_m = \{x \in X; \ c(x) \geqslant m\}$. We note that for any morphism $f : X \to Y$ in $\underline{\mathcal{M}}^{inj}$, we find that $f(X_m) \subseteq Y_m$. This property clearly does not hold for more general morphisms. For every r, we can now define a new equivalence relation $\sim_r^m$ on X, which refines $\sim_r$, by requiring that each equivalence class of $\sim_r$ which has cardinality $\geqslant m$ is an equivalence class of $\sim_r^m$, and that for any x for which $c(x) < m$, x defines a singleton equivalence class in $\sim_r^m$. We now obtain a new persistent set (X, θ_X^m), where $\theta_X^m(r)$ will denote the partition associated to the equivalence relation $\sim_r^m$. It is readily checked that $X \to (X, \theta_X^m)$ is functorial on $\underline{\mathcal{M}}^{inj}$.

Definition 7.1. *For each $m \in \mathbb{N}$ let $\mathfrak{H}^m : \underline{\mathcal{M}}^{inj} \to \mathcal{P}$ be the functor which to each finite metric space (X, d_X) assigns the persistent set (X, θ_X^m) where for each $r \geqslant 0$, $\theta_X^m(r)$ is the partition of X into equivalence classes of $\sim_r^m$.*

This scheme could be motivated by the intuition that one does not regard clusters of small cardinality as significant, and therefore makes points lying in small clusters into singletons, where one can then remove them as representing "outliers" or "noise".

Another construction arises from similar consideration to those that gave rise to the standard clustering algorithms $\mathfrak{C}^{\Delta_m(\delta)}$ in Section 6.7, cf. Corollary 6.1.

Definition 7.2. *For each $m \in \mathbb{N}$ define the functor $\mathfrak{R}^{\Delta_m} : \underline{\mathcal{M}}^{inj} \to \mathcal{P}$ given by*

$$\mathfrak{R}^{\Delta_m} = \mathfrak{R} \circ \mathfrak{T}^m,$$

where we have written $\mathfrak{T}^m = \mathfrak{T}^\Omega$ for $\Omega = \{\Delta_m(\delta), \ \delta \geqslant 0\}$, cf. (6–9).

That this definition produces a functor follows directly from the fact that $\mathfrak{R}$ and $\mathfrak{T}^m$ are functorial.

The resulting construction is in the same spirit as DBSCAN [EpKSX96].

Remark 7.1. *Notice first that $\mathfrak{R}^{\Delta_2} = \mathfrak{R}$. Also, for each $m \in \mathbb{N}$, the hierarchical clustering functor $\mathfrak{R}^{\Delta_m}$ is related to the standard clustering functor $\mathfrak{C}^{\Delta_m(\delta)}$ in the following manner. Fix $\delta \geqslant 0$ and write $\mathfrak{C}^{\Delta_m(\delta)}(X, d_X) = (X, P_X)$, then, $P_X = \theta_X^m(\delta)$.*

An example of dendrograms obtained with these methods is shown in Figure 6.

8. Discussion

The methods of this paper permit natural extensions to other situations, such as clustering of graphs and networks with, for example, the notion of clique clustering [PDFV05] fitting naturally into our context. It appears likely that from the point of view described here, it will in many cases be possible, given a collection of constraints on a clustering functor, to determine the *universal* one satisfying the constraints. One could therefore use sets of constraints as the definition of clustering functors.
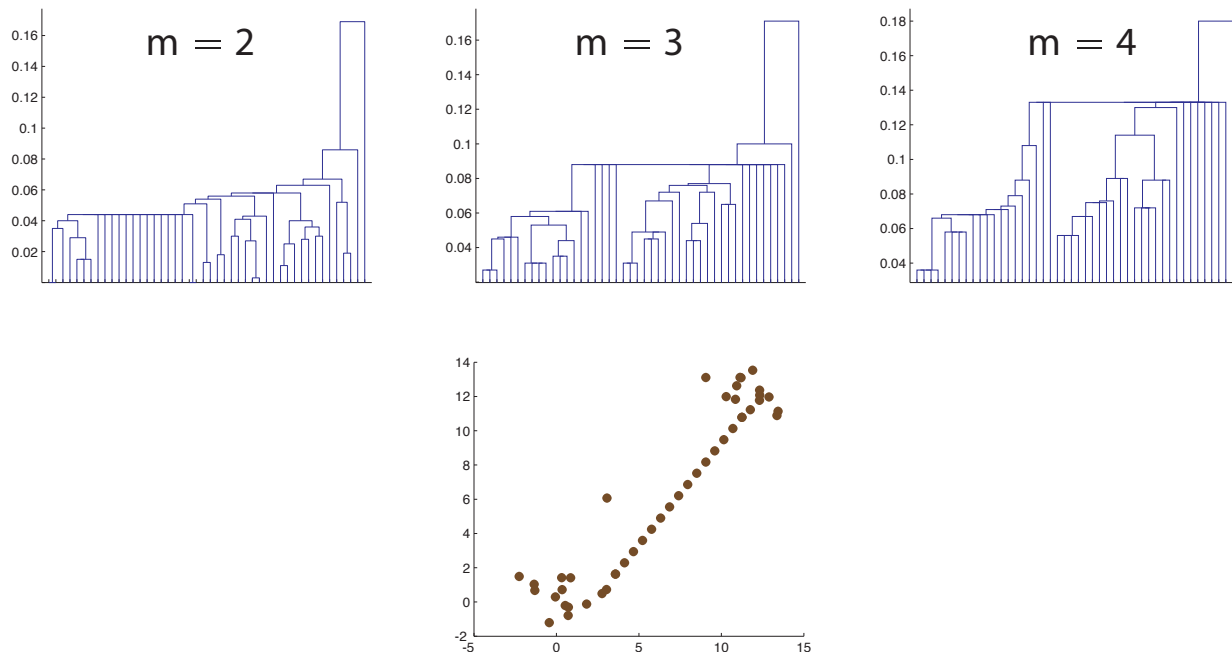
FIGURE 6. Dendrograms resulting from applying $\mathfrak{R}^{\Delta_m}$ to a randomly generated dataset consisting of 50 points in the plane. The dataset consists of two clusters of points which are joined by a thin chain of points with some additional outliers. On the top row, from left to right we show the dendrograms corresponding to $m = 2, 3$ and 4. Note how increasing m produces dendrograms that exhibit more clearly clustered subsets.

In addition to producing classifications of clustering functors, functoriality is a very desirable property when clustering is used as an ingredient in computational topology. Functorial clustering schemes produce diagrams of sets of varying shapes which can be used to construct simplicial complexes which in turn reflect the topology of connected components [Car09]. Functorial schemes also can be used to construct "zig-zag diagrams" [CdS10] which reflect the stability of clusters produced over a family of independent samples.

We believe that the conceptual framework presented here can be a useful tool in reasoning about clustering algorithms. We have also shown that clustering methods which have some degree of functoriality admit the possibility of certain kind of qualitative geometric analysis of datasets which can be quite valuable. The general idea that the morphisms between mathematical objects (together with the notion of functoriality) are critical in many situations is well-established in many areas of mathematics, and we would argue that it is valuable in this statistical situation as well.

REFERENCES

[Car09]   Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46:255–308, 2009.
[CdS10]   G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4), August 2010.
[CM08]   G. Carlsson and F. Mémoli. Persistent Clustering and a Theorem of J. Kleinberg. *ArXiv e-prints*, August 2008.

[CM10a] Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.

[CM10b] Gunnar Carlsson and Facundo Mémoli. Multiparameter clustering methods. In Claus Weihs Hermann Locarek-Junge, editor, *Classification as a Tool for Research. Proc. 11th Conference of the International Federation of Classification Societies (IFCS-09)*, pages 63–70, Heidelberg-Berlin, 2010. Springer-Verlag.

[EpKSX96] Martin Ester, Hans peter Kriegel, Jörg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

[GvLW09] I. Guyon, U. von Luxburg, and R. Williamson. Clustering: Science or art? Technical report, Paper presented at the NIPS 2009 Workshop Clustering: Science or Art? `http://www.kyb.tuebingen.mpg.de/bs/people/ule/publications/publication_downloads/IsClustering` 2009.

[Hau95] Jean-Claude Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. In *Prospects in topology (Princeton, NJ, 1994)*, volume 138 of *Ann. of Math. Stud.*, pages 175–188. Princeton Univ. Press, Princeton, NJ, 1995.

[Isb64] J. R. Isbell. Six theorems about injective metric spaces. *Comment. Math. Helv.*, 39:65–76, 1964.

[JD88] Anil K. Jain and Richard C. Dubes. *Algorithms for clustering data*. Prentice Hall Advanced Reference Series. Prentice Hall Inc., Englewood Cliffs, NJ, 1988.

[JS71] Nicholas Jardine and Robin Sibson. *Mathematical taxonomy*. John Wiley & Sons Ltd., London, 1971. Wiley Series in Probability and Mathematical Statistics.

[Kle02] Jon M. Kleinberg. An impossibility theorem for clustering. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *NIPS*, pages 446–453. MIT Press, 2002.

[LW67] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies 1. Hierarchical systems. *Computer Journal*, 9(4):373–380, February 1967.

[ML98] Saunders Mac Lane. *Categories for the working mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.

[Mun75] James R. Munkres. *Topology: a first course*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.

[NS08] Amiya Nayak and Ivan Stojmenovic. *Handbook of Applied Algorithms: Solving Scientific, Engineering, and Practical Problems*. Wiley-IEEE Press, 2008.

[PDFV05] Gergely Palla, Imre Derenyi, Illes Farkas, and Tamas Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, June 2005.

Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail address: {gunnar,memoli}@math.stanford.edu