Functorial Metric Clustering with Overlaps: Possibilities and Impossibilities

2016 SIAM Conference on Imaging Science

Dan P. Guralnik (Penn)

joint work with

Jared Culbertson (AFRL)

Peter Stiller (Texas A&M)

May 23, 2016

Supported by Air Force Office of Scientific Research under the LRIR 12RY02COR, LRIR 15RYCOR153, and MURI FA9550-10-1-0567 grants.

Approaches to Distance-Based Clustering

Vague Objective: Given a space[1] X with distance/weight assignment $d : X \times X \to \mathbb{R}_{\geq 0}$, denoted $(x, y) \mapsto d_{xy}$, classify the points of X in a manner consistent with treating d as a measure of dissimilarity among them.

[1] All our spaces X, Y, etc. are finite.

Approaches to Distance-Based Clustering

The Meaning of the word ‘Classify’: construct a map

$$\left\{ \textbf{metric spaces} \right\} \xrightarrow{\mathbf{F}} \left\{ \textbf{classifying objects} \right\}$$

Traditional classifiers:

- Partitioning the data into disjoint “blocks”:
  - Kleinberg inconsistency: axiomatic approaches seem to fail [Kle03]
  - Consistency through axioms about quality, e.g. CQMs [BDA09]
- The data set is covered by overlapping “clusters”:
  - Some general desiderata laid out by Jardine-Sibson in [JS71]
  - Projections to *tree metrics* à-la Bunemann [Bun71, MS99]
  - More generally, *split decompositions* [BD92] produce *networks* [DHM01]

Approaches to Distance-Based Clustering

The Meaning of the word 'Classify': construct a ~~map~~ functor

$$\left\{ \begin{array}{c} \text{category of} \\ \textbf{metric spaces} \text{ and} \\ \text{non-expansive maps} \end{array} \right\} \xrightarrow{\mathbf{F}} \left\{ \begin{array}{c} \text{category of} \\ \textbf{classifying objects} \\ \text{and refining maps} \end{array} \right\}$$

Instead[1], use a *category structure* to encode consistency constraints:

- Admissible transformations *between data sets* translate consistently, via **F**, into refinement relations between classifiers.
- More classifiers $\Leftrightarrow$ Fewer obstructions to consistency

  $\rightsquigarrow$ *e.g. replacing partitions with dendrograms removes Kleinberg's obstruction*

- Fewer morphisms between spaces $\Leftrightarrow$ Fewer Constraints

  $\rightsquigarrow$ *e.g. more functors* $\mathbf{Met}^{inj} \to \mathbf{Dendro}$ *than* $\mathbf{Met} \to \mathbf{Dendro}$.

[1] Carlsson-Mémoli [CM08]

Formally Modeling the Data (1): Weights

Space of Weights. For any finite set $X \neq \varnothing$, define:

$$\mathbf{W}_X := \left\{ w : X \times X \to \mathbb{R}_{\geq 0} \;\middle|\; w_{xy} \equiv w_{yx}\,,\; w_{xx} \equiv 0 \;\right\}$$

Ordering the Weights. For any X and $u, v \in \mathbf{W}_X$:

$$u \leq v \quad \overset{def.}{\Leftrightarrow} \quad u_{xy} \leq v_{xy} \text{ for all } x, y \in X$$

$\rightsquigarrow$ *(X, v) is "better resolved" than (X, u)*

Pull-back. For $f : X \to Y$ define the pull-back on weights:

$$f^* : \mathbf{W}_Y \to \mathbf{W}_X\,, \quad (f^* w)_{xx'} := w_{f(x)f(x')}$$

$\rightsquigarrow$ *Provides a notion of consistency across spaces.*

Formally Modeling the Data (2): Weight Categories

The Category W of Weights:

- **Objects** are pairs (X, u), with $u \in \mathbf{W}_X$;
- **Morphisms** are non-expansive maps:

$$f\colon (X,u) \to (Y,v) \text{ in } \mathbf{W} \;\Leftrightarrow\; f\colon X \to Y \text{ in } \mathbf{Set} \text{ has } f^*v \leq u$$

A **Weight Category** is a full sub-category[2] $\mathbf{C}$ of $\mathbf{W}$ satisfying $f^*(\mathbf{C}_Y) \subseteq \mathbf{C}_X$ for all set maps[3] $f : X \to Y$, where

$$\mathbf{C}_X := \{w \in \mathbf{W}_X \mid (X, w) \in \mathbf{C}\}$$

For example: the categories **Met** and **Ult** of metric and ultra-metric spaces, respectively[4].

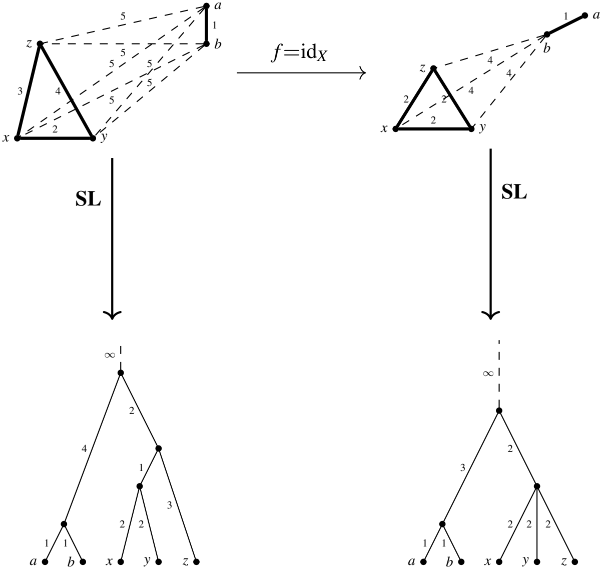
---

[2] Possibly fewer objects, same morphisms.

[3] Weakened versions may be considered, with fewer maps.

[4] Semi-metrics are allowed.

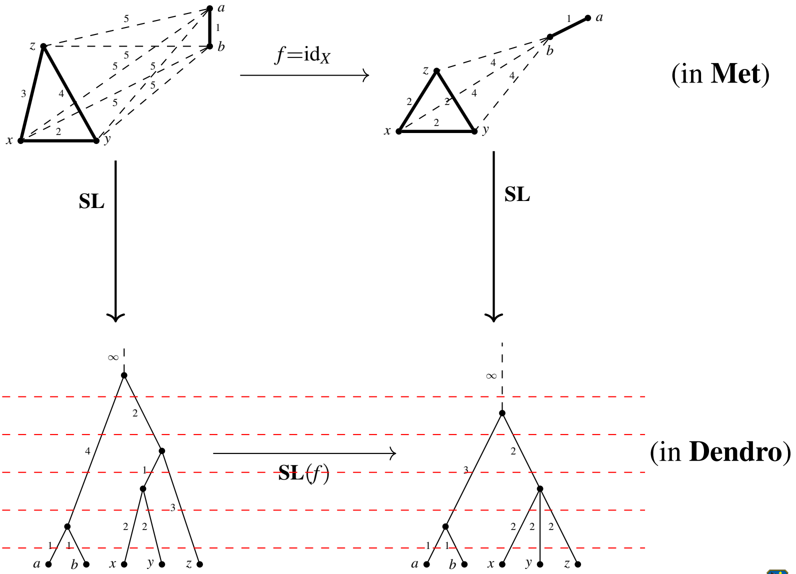
Example: Hierarchical Single Linkage [CM08]



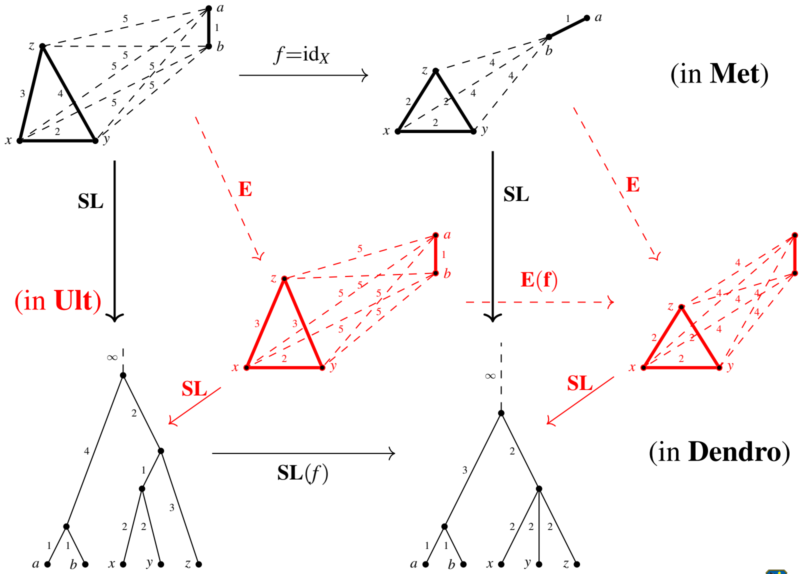
(in **Met**)

(in **Dendro**)

Example: Hierarchical Single Linkage [CM08]



Example: Hierarchical Single Linkage [CM08]

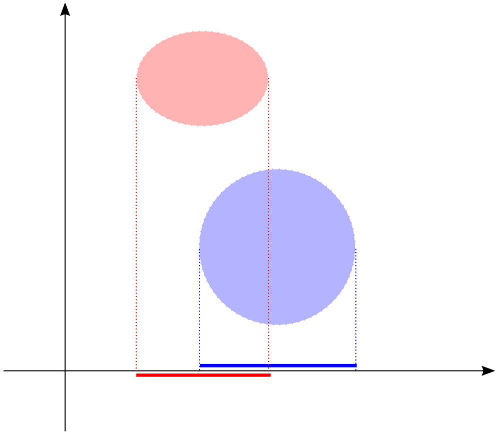


Overview of Results

- Carlsson-Mémoli categorical framework extends to an equivalence of categories between **Met** and a generalization, **Sieve**, of **Dendro**.
- Moreover, extends beyond real coefficients.
- Hierarchical Single Linkage is but one instance of an abundant family of clustering functors (with overlaps), characterized uniquely by their images in **Sieve**.
- No (non-trivial) functorial clustering from **Met** to any class of metrics containing the class of tree metrics.
- No functorial clustering onto cut-metrics.
- New class(es) of classifier objects, arising from study of injective envelopes.
- Is there a way to "categorify" the Bunemann and Dress split-based projections?
- How about K-means, spectral methods and friends?

Why Clustering with Overlaps?

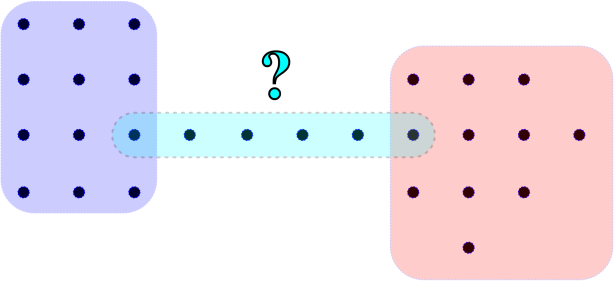
Data is almost always, at best, a projection of objective reality.



Even well-resolved classes will, generally, appear as overlapping.

Why Clustering with Overlaps?

Forced ambiguity is information, too.

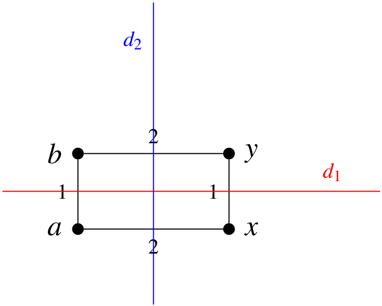


- The inability to decide on membership in one of two categories may motivate the introduction of a new semantic category[5].
- This problem is especially acute in the absence of a natural ambient geometry.

[5] Interesting empirical results by Gama–Segarra–Ribeiro [GSR15].

Why Clustering with Overlaps?

Manage incompatible clusterings at different scales.



(X, d) decomposes as a sum $d = d_1 + d_2$ of
positive multiples of split semi-metrics.

This idea yields the theory of *split decompositions* [BD92]:

- Instead of **Ult**, the range of a clustering map should be all cut-metric spaces. **Can this be achieved with functors?**

Desired Axioms for Overlapping Clustering

Contraction as Coarsening / Loss of Information [JS71]
Jardine–Sibson [JS71]: a clustering of (X, d) is a symmetric, reflexive relation $\mathscr{R}$:

- Clusters are of the form

$$[x]_{\mathscr{R}} = \left\{ y \in X \,\middle|\, x\mathscr{R}y \right\}.$$

Viewing such $\mathscr{R}$ as defining covers of X, it makes sense to define:

Definition (set pairwise linked by a cover)

$F \subseteq X$ is *pairwise linked* by a cover $\mathcal{U}$ of X, if any $\{x, y\} \subseteq F$ is contained in some element of $\mathcal{U}$.

Desired Axioms for Overlapping Clustering

Transitivity of $\mathscr{R}$ is too strong. It is equivalent to restricting attention to partition-based classifiers.

Definition (set linked by a cover)

$F \subseteq X$ is *linked* by a cover $\mathcal{U}$ of X if there exists an element of $\mathcal{U}$ containing F.

Weakening: when should F be linked by $\mathcal{U}$?

Think of the prototypical relation $d_{xy} \leq \delta$:

- Not transitive, but
- Often forms large cliques.

One wants to work with maximal such cliques.

The Category $\mathbf{Cov}_{\triangleright}$ of Non-Nested Flag Covers

Objects of the form $(X, \mathcal{U})$ where $\mathcal{U}$ is a set cover of $X \neq \varnothing$.

Additionally objects $(X, \mathcal{U})$ must satisfy, for any $A, B \subseteq X$:

(non - nesting) If $A, B \in \mathcal{U}$ and $A \subseteq B$ then $A = B$;

(flag condition) If A is pairwise linked by $\mathcal{U}$ then it is linked by $\mathcal{U}$.

Morphisms $f\colon (X, \mathcal{U}) \to (Y, \mathcal{V})$ are maps $f\colon X \to Y$ such that $f(U)$ is linked by $\mathcal{V}$ for every $U \in \mathcal{U}$. $\rightsquigarrow$ *"clusters map into clusters"*

Denote the set of non-nested flag covers of X by $\mathbf{Cov}_{\triangleright}(X)$.

Observe that both $\bot := \{\{x\}\}_{x \in X}$ and $\top := \{X\}$ lie in $\mathbf{Cov}_{\triangleright}(X)$.

Flat Clustering with Overlaps: The Extremes (1)

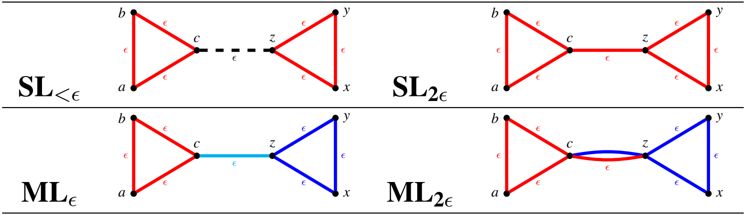
The δ-Threshold Graph[5] $R_\delta(X,d)$ is the simple graph (X,E) with

$$xy \in E \;\Leftrightarrow\; d_{xy} \leq \delta.$$

Two clustering functors derived from $R_\delta(X,d)$ bound the behavior of any $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Cov}_{\triangleright}$:

Single Linkage, $\mathbf{SL}_\delta$, Covers X with the components of $R_\delta(X,d)$

Maximal Linkage, $\mathbf{ML}_\delta$, Covers X with maximal cliques of $R_\delta(X,d)$



[5] This is the 1-skeleton of the Vietoris-Rips complex at resolution δ.

Flat Clustering with Overlaps: The Extremes (2)

The clustering parameter of a functor $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Cov}_{\triangleright}$ is:

$$\delta_{\mathbf{F}} := \inf\{\epsilon > 0 \,\big|\, \mathbf{F}(\bullet \overset{\epsilon}{\leftrightarrow} \bullet) = \bot\}$$

At this resolution one has:

Theorem (C.–G.–Hansen–S. [CGHS15])

Suppose $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Cov}_{\triangleright}$. *Then, for any* $(X,d) \in \mathbf{Met}$ *one has:*

$$\mathbf{ML}_{\delta_{\mathbf{F}}}(X,d) \succeq \mathbf{F}(X,d) \succeq \mathbf{SL}_{\delta_{\mathbf{F}}}(X,d)\,.$$

- When $\mathrm{Image}(\mathbf{F}) \subseteq \mathbf{Part}$, the category of partitions, this forces $\mathbf{F} = \mathbf{SL}_{\delta_{\mathbf{F}}}$, recovering the Carlsson–Mémoli characterization [CM10].
- Additional flat clustering functors are considered in [CGHS15].

Hierarchical Classifiers with Overlaps: Sieves

Objects in the category **Sieve** are pairs (X, ξ), called sieves, where ξ is a map $\xi : \mathbb{R}_{\geq 0} \to \mathbf{Cov}_{\triangleright}(X)$, $t \mapsto \xi_t$ satisfying:

- If $s < t$ then $\xi_s \succeq \xi_t$;
- Every $s \geq 0$ has $\epsilon > 0$ with $\xi_t = \xi_s$ for all $t \in [s, s+\epsilon)$;
- There exists $t \geq 0$ such that $\xi_t = \top$.

Morphisms $f\colon (X, \xi) \to (Y, \eta)$ in the category **Sieve** are set maps $f\colon X \to Y$ which are morphisms $(X, \xi_t) \to (Y, \eta_t)$ in $\mathbf{Cov}_{\triangleright}$ for every resolution $t \geq 0$.

A Sieving Functor is a functor $\mathbf{F} : \mathbf{W} \to \mathbf{Sieve}$ fibering over **Set**:

- $\mathbf{F}(X, w)$ has the form $(X, \xi_{\mathbf{F},w})$ for any X and weight $w \in \mathbf{W}_X$.
- $\mathbf{F}(f) = f$ for any morphism f in $\mathbf{W}$.

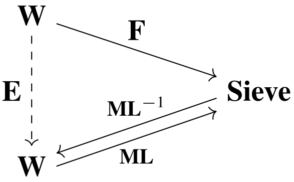
$\leadsto$ *Labels of data sets remain the same; clustering only changes the attached information.*

Maximal Linkage as an Equivalence

Theorem (C.–G.–S. [CGS15])

Let $\mathbf{ML}$ *be the sieving functor defined by* $\mathbf{ML}(X,w)_\delta := \mathbf{ML}_\delta(X,w)$. *Then* $\mathbf{ML} : \mathbf{W} \to \mathbf{Sieve}$ *is an equivalence of categories*[6].

In other words, every sieving functor on $\mathbf{W}$ factors as $\mathbf{ML} \circ \mathbf{E}$, with $\mathbf{E} : \mathbf{W} \to \mathbf{W}$ a functor taking the form $(\dagger)$ below:



$$(\dagger) \quad \begin{cases} \mathbf{E}(X,u) = (X, E_X u) & \text{for all } X \text{ and all } u \in \mathbf{W}_X \\ \mathbf{E}(f) = f & \text{whenever } f : (X,u) \to (Y,v) \text{ in } \mathbf{W} \end{cases}$$

[6] restricting on **Ult** to its well-known [CM08] equivalence with **Dendro**.

A Curious Property of **SL**

In the case of **SL**, the induced endo-functor is special:

Definition (Projection)

Let $\mathbf{C}$ be a weight category. A projection is an endo-functor $\mathbf{E}$ of $\mathbf{C}$ satisfying:

1. $\mathbf{E}$ has the form $(\dagger)$;
2. $E_X w \leq w$ for all X and $w \in \mathbf{C}_X$;

   $\rightsquigarrow$ *"Points must aggregate at the resolution given by their weight"*
3. $E_X^2 = E_X$ for all X.

   $\rightsquigarrow$ *"All clustering decisions are final"*

HOW TO CONSTRUCT PROJECTIONS?

Uniqueness of Projections

Proposition (C.–G.–S. [CGS15])

Projections are uniquely determined by their images.

Proof. Suppose $\mathbf{E}, \mathbf{F}$ are projections having the same image. Fix X.

- By functoriality, if $u \leq v$ in $\mathbf{C}_X$ then $\mathrm{id}_X \colon (X, v) \to (X, u)$ is a morphism. If $\mathbf{E}$ is any projection in $\mathbf{C}$, then

$$\mathrm{id}_X = \mathbf{E}(\mathrm{id}_X) \colon (X, E_X v) \to (X, E_X u)$$

  is a morphism. Hence $E_X u \leq E_X v$.
- For any $u \in \mathbf{C}_X$ we have $E_X F_X u = F_X u$ and $F_X E_X u = E_X u$, since the images and fixed point sets of E_X and F_X coincide.
- Now, for any $u \in \mathbf{C}_X$ we calculate:

$$E_x u \leq u \;\Rightarrow\; E_X u = F_X E_X u \leq F_X u$$

  Symmetrically, we obtain $F_X u \leq E_X u$. ∎

Existence of Projections

A Clustering Domain is a weight category $\mathbf{C}$ which is sup-closed:

- for all X, if $S \subset \mathbf{C}_X$ is bounded above by an element of $\mathbf{W}$ then the pointwise supremum of S lies in $\mathbf{C}_X$.

Note:

1. $\mathbf{W}$ is a clustering domain;
2. max-closed and closed implies sup-closed.

Theorem (C.–G.–S. [CGS15])

Suppose $\mathbf{C}$ is a clustering domain and $\mathbf{D}$ is a full weight sub-category of $\mathbf{C}$. Then $\mathbf{D}$ is the image of a projection (in $\mathbf{C}$) if and only if $\mathbf{D}$ is a clustering domain. Moreover, the unique projection of $\mathbf{C}$ onto $\mathbf{D}$ is given by

$$E_X(w) = \sup\{u \in \mathbf{D} \,\big|\, u \leq w\}\,.$$

Find clustering domains with interesting / useful / tractable geometry / combinatorics.

Examples of Clustering (non/)Domains

Examples.

- **W** with the identity functor;
- **Met** with the induced path metric functor;
- q-metrics, $d_{xy}^q \leq d_{xz}^q + d_{yz}^q$, $q \in [1, \infty]$ with similar projection [SCMR];
- **Ult** with Single Linkage;
- Integer weights with the floor functor;
- Intersection of two clustering domains with $(?)$;
- λ-Infra-metrics $d_{xy} \leq \lambda \cdot \max\{d_{xz}, d_{yz}\}$, with $(?)$.

Non-examples.

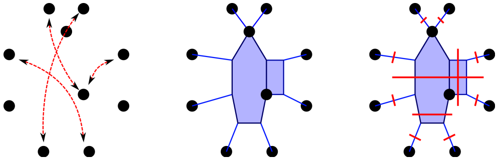
- A clustering domain containing the tree metrics contains **Met**.
- Tree metrics, Cut metrics do not possess projections.

New Class of Metrics to Care About: A-spaces[7]

Definition (A-spaces, C.–G.–S. [CGS15])

A metric space (X, d) is an A-space if every $x \in X$ has $y \in X$ with $d_{xy} = \mathtt{diam}(X, d)$.

Recall that the *injective envelope* $\epsilon(X, d)$ of (X, d) is, in a sense, a minimal "complete geodesic filling-in" of (X, d) [Isb64, Dre84].



Mantra: Starting with (X, d) (left), this canonical extension represents essential connectivity relations among points (center), inducing cuts and clusters (right).

[7] Not to be confused with *antipodal spaces* [HKM05, HKM04]

New Class of Metrics to Care About: A-spaces[7]

Our results about A-spaces:

- A-spaces form a clustering domain in **Met**, extending **Ult**, but not containing any tree metrics outside **Ult**;
- The projection to A-spaces is easy to compute:
  1. If edges of length $\mathtt{diam}(X, d)$ do not form a cover, collapse all top distances to 2nd-largest distances
  2. Repeat until done.
- (X, d) is an A-space if and only if $\epsilon(X, d)$ has a (unique) point at distance $\mathtt{diam}(X, d)/2$ from every point of X.

Thus, A-spaces are the class of metric spaces with a natural root, extending the analogous notion for dendrograms.

- Injective envelopes are extremely hard to compute [SY04].
- A-spaces are an example where this computation is redundant.
- What are the sieves corresponding to A-spaces? Are there "hereditary A-spaces" and what sieves do they generate?

[7] Not to be confused with *antipodal spaces* [HKM05, HKM04]

Hans-Jürgen Bandelt and Andreas W. M. Dress.
A canonical decomposition theory for metrics on a finite set.
Adv. Math., 92(1):47–105, 1992.

Shai Ben-David and Margareta Ackerman.
Measures of clustering quality: A working set of axioms for clustering.
In *Advances in neural information processing systems*, pages 121–128, 2009.

P. Bunemann.
The recovery of trees from measures of dissimilarity, pages 387–395.
Edinburgh University Press, Edinburgh, 1971.

Jared Culbertson, Dan P. Guralnik, Jakob Hansen, and Peter F. Stiller.
Consistency constraints for overlapping data clustering.
in preparation, 2015.

Jared Culbertson, Dan P. Guralnik, and Peter F. Stiller.
Functorial hierarchical clustering with overlaps.
in preparation, 2015.

Gunnar Carlsson and Facundo Mémoli.
Persistent clustering and a theorem of J. Kleinberg.
arXiv preprint arXiv:0808.2241, 2008.

Gunnar Carlsson and Facundo Mémoli.
Classifying clustering schemes.
arXiv preprint arXiv:1011.5270, 2010.

A. Dress, K. T. Huber, and V. Moulton.
Metric spaces in pure and applied mathematics.

In *Proceedings of the Conference on Quadratic Forms and Related Topics (Baton Rouge, LA, 2001)*, number Extra Vol., pages 121–139, 2001.

Andreas W. M. Dress.
Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces.
Adv. in Math., 53(3):321–402, 1984.

Fernando Gama, Santiago Segarra, and Alejandro Ribeiro.
Overlapping clustering of network data using cut metrics.
In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015.

Katharina T Huber, Jacobus H Koolen, and Vincent Moulton.
The tight span of an antipodal metric space: Part iiÑgeometrical properties.
Discrete & Computational Geometry, 31(4):567–586, 2004.

Katharina T Huber, Jacobus H Koolen, and Vincent Moulton.
The tight span of an antipodal metric spaceÑpart i: Combinatorial properties.
Discrete mathematics, 303(1):65–79, 2005.

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

Nicholas Jardine and Robin Sibson.
Mathematical taxonomy.
London etc.: John Wiley, 1971.

Jon Kleinberg.

An impossibility theorem for clustering.
Advances in neural information processing systems, pages 463–470, 2003.

Vincent Moulton and Mike Steel.
Retractions of finite distance functions onto tree metrics.
Discrete Applied Mathematics, 91(1):215–233, 1999.

Santiago Segarra, Gunnar Carlsson, Facundo Mémoli, and Alejandro Ribeiro.
Metric representations of network data.

Bernd Sturmfels and Josephine Yu.
Classification of six-point metrics.
Electron. J. Combin., 11(1):Research Paper 44, 16 pp. (electronic), 2004.