THE REPRESENTATION THEORY OF FILTERED HIERARCHICAL CLUSTERING

Ulrich Bauer (TUM)

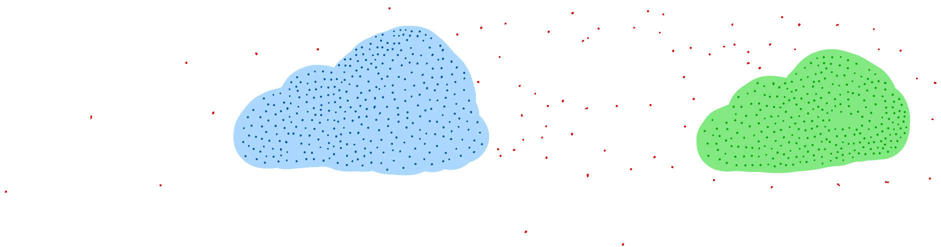
WORKSHOP GEOMETRY AND TOPOLOGY IN DATA ANALYSIS

ICIAM 2019, VALENCIA

Jul 16, 2019

joint work with:

Magnus Botnan / Steffen Oppermann / Johan Steen



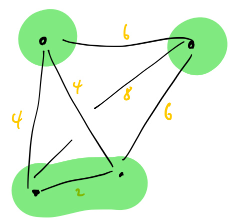
CLUSTERING FUNCTIONS

X : finite set

Clustering function φ :

maps a metric $d : X \times X \to \mathbb{R}$ (distance matrix)

to a partition of X.

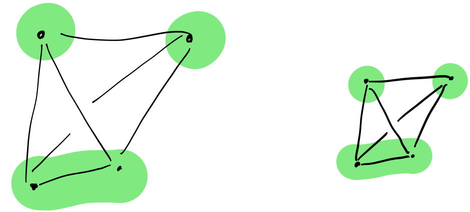


KLEINBERG'S AXIOMS

Desirable properties

- scale invariance:

$$\varphi(d) = \varphi(\lambda \cdot d).$$

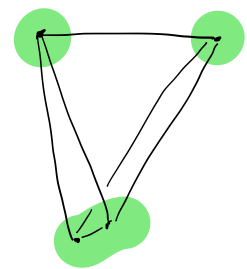


- richness:

every partition is obtained from some d.

- consistency:

decreasing d within clusters /
increasing d across clusters
does not change the result.



KLEINBERG'S IMPOSSIBILITY THEOREM

Thm [Kleinberg 2002] No clustering function satisfies

- scale-invariant,
- richness, and
- consistency.

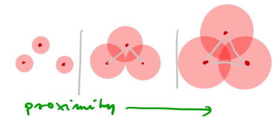
Motivates the use of a scale parameter

⇝ hierarchical clustering

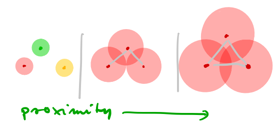
CLUSTERING FROM CONNECTED COMPONENTS

proximity graph

- filter edges by proximity



π_0 (connected components)



single-linkage clustering

H_0 (homology in deg. 0 with coeffs in K)

$K^3 \to K \to K$

proximity

persistent homology

$H_0 = F \circ \pi_0$

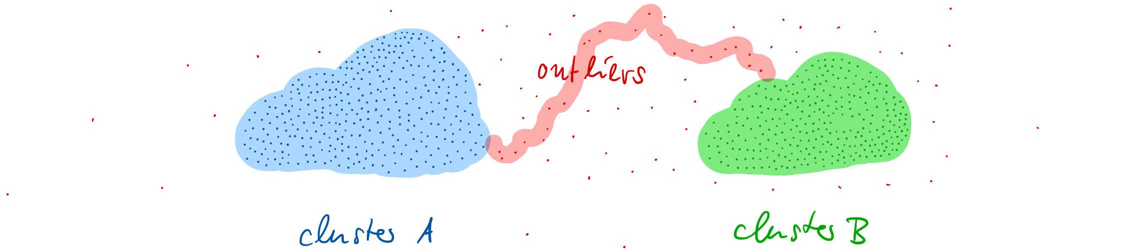
HIERARCHICAL CLUSTERING: EXISTENCE & UNIQUENESS

Thm [Carlsson, Mémoli 2010] Single-linkage clustering is the unique hierarchical clustering method satisfying [... certain axioms similar to Kleinberg's].

But ...

CHAINING EFFECT

Single-linkage clustering is sensitive to outliers

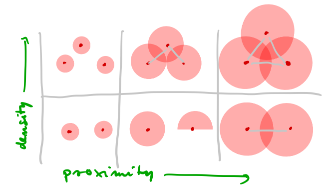


→ not used much in practice!

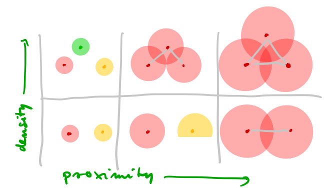
2-PARAMETER CLUSTERING

density-proximity graph

- filter points by density
- filter edges by proximity



π_0 (connected components)



H_0 (homology in deg. 0
with coeffs in K)

$$\begin{tikzcd} K^3 \arrow[r] & K \arrow[r] & K \\ K^2 \arrow[r] \arrow[u] & K^2 \arrow[r] \arrow[u] & K \arrow[u] \end{tikzcd}$$

density

proximity

DECOMPOSING DIAGRAMS OF VECTOR SPACES

$$
\begin{array}{ccc}
V & \xrightarrow{f} & W \\
\cong & & \\
\ker f & \longrightarrow & 0 \\
\oplus & & \\
\operatorname{im} f & \longrightarrow & \operatorname{im} f \\
\oplus & & \\
0 & \longrightarrow & \operatorname{coker} f
\end{array}
$$

TWO MAPS

$$
\begin{array}{cccccc}
 & U & \xrightarrow{g} & V & \xrightarrow{f} & W \\
\cong & & & & & \\
 & K^a & \longrightarrow & 0 & \longrightarrow & 0 \\
\oplus & & & & & \\
 & 0 & \longrightarrow & K^b & \longrightarrow & 0 \\
\oplus & & & & & \\
 & 0 & \longrightarrow & 0 & \longrightarrow & K^c \\
\oplus & & & & & \\
 & K^d & \longrightarrow & K^d & \longrightarrow & 0 \\
\oplus & & & & & \\
 & 0 & \longrightarrow & K^e & \longrightarrow & K^e \\
\oplus & & & & & \\
 & K^f & \longrightarrow & K^f & \longrightarrow & K^f
\end{array}
$$

SEQUENCES OF MAPS

$$V: \quad V_0 \to V_1 \to V_2 \to \cdots \to V_n$$

decomposes into summands of the form

$$\cdots \to 0 \to \underbrace{K \to \cdots \to K}_{i \qquad\qquad j} \to 0 \to \cdots$$

$V \cong$ "collection of intervals"

$\rightsquigarrow$ persistence barcode.

DROZD'S TRICHOTOMY

Given a finite indexing poset P. (k: algebr. closed)

What are the indecomposable diagrams with shape P?

3 cases (representation types):

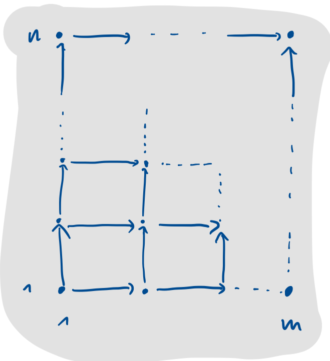
(a) A finite list. finite type

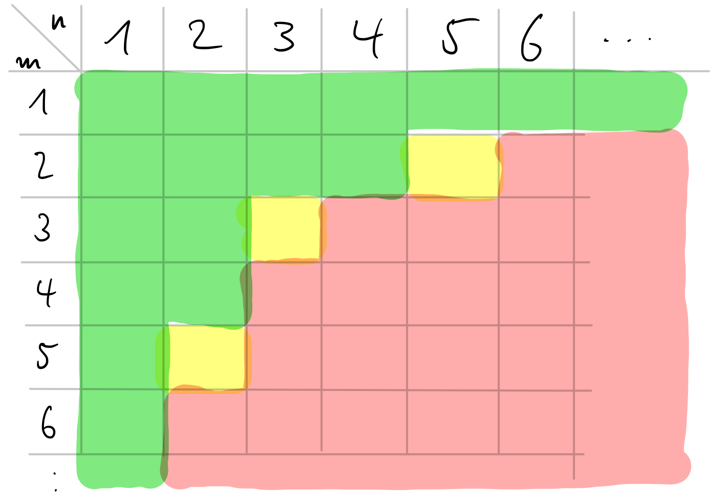
(b) A finite list (of 1-param. families). tame

(c) It's complicated. wild

(as complicated as modules over any finite-dim. algebra; including undecidable problems)

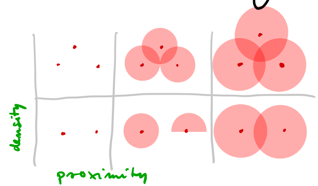
REPRESENTATION TYPES OF COMMUTATIVE GRIDS





GRID DIAGRAMS FROM CLUSTERING

Consider again 2-parameter clustering (proximity / density)



This yields diagrams of the form

$$\begin{tikzcd}
\vdots & \vdots & & \\
{} \arrow[u] \arrow[r, two heads] & {} \arrow[u, phantom] & \cdots & \\
{} \arrow[u] \arrow[r, two heads] & {} \arrow[u] \arrow[r, two heads] & {} \arrow[r, two heads] & {}
\end{tikzcd}$$

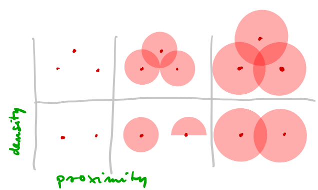
Horizontal maps are surjective!

Does this simplify the picture?

EPIMORPHISMS

- $\mathrm{Rep}(m,n)$: all commutative dgms over $m \times n$ grid
- $\mathrm{Rep}^{\twoheadrightarrow}(m,n)$: epis in horizontal direction
- $\mathrm{Rep}^{\uparrow\!\!\uparrow \twoheadrightarrow}(m,n)$: epis in both directions.

Lemma $\mathrm{Rep}^{\twoheadrightarrow}(m,2)$ is finite type.



$$H_0 \rightsquigarrow \begin{tikzcd} K^3 \arrow[r, two heads, "(1\,1\,1)"] & K \arrow[r, two heads] & K \\ K^2 \arrow[u, "\begin{pmatrix}1&0\\0&1\\0&0\end{pmatrix}"] \arrow[r, two heads] & K^2 \arrow[u, "(1\,1)"'] \arrow[r, two heads, "(1\,1)"'] & K \arrow[u] \end{tikzcd}$$

$$\cong \begin{tikzcd} K \arrow[r, two heads] & K \arrow[r, two heads] & K \\ K \arrow[u] \arrow[r, two heads] & K \arrow[u] \arrow[r, two heads] & K \arrow[u] \end{tikzcd} \oplus \begin{tikzcd} K \arrow[r, two heads] & 0 \arrow[r, two heads] & 0 \\ K \arrow[u] \arrow[r, two heads] & K \arrow[u] \arrow[r, two heads] & 0 \arrow[u] \end{tikzcd} \oplus \begin{tikzcd} K \arrow[r, two heads] & 0 \arrow[r, two heads] & 0 \\ 0 \arrow[u] \arrow[r, two heads] & 0 \arrow[u] \arrow[r, two heads] & 0 \arrow[u] \end{tikzcd}$$

EPIC GRIDS & WILD THINGS

Thm [B, Botnan, Oppermann, Steen 19]

$$\begin{array}{ccc}
\mathrm{Rep}^{\twoheaduparrow \twoheadrightarrow}(m, n) & \sim & \mathrm{Rep}^{\uparrow}(m, n-1) \\
\wr & & \wr \\
\mathrm{Rep}^{\twoheadrightarrow}(m-1, n) & \sim & \mathrm{Rep}(m-1, n-1)
\end{array}$$

same representation type

Corollary $\mathrm{Rep}^{\twoheadrightarrow}(m, n)$ is

- finite type for $n \leq 2$ and $(n = 3,\ m \leq 4)$
- tame for $(n = 3,\ m = 5)$ and $(n = 4,\ m = 3)$
- wild otherwise.

TORSION PAIRS

Example: Abelian groups (fin. generated)

$$G = \underbrace{\mathbb{Z}^{a}}_{F \text{ (free)}} \oplus \underbrace{\mathbb{Z}_{q_1}^{b_1} \oplus \cdots \oplus \mathbb{Z}_{q_n}^{b_n}}_{T \text{ (torsion)}}$$

- canonically: $T \hookrightarrow G \twoheadrightarrow F$ (short exact sequence)
- The only homomorphism $\phi : T \to F$ is $x \mapsto 0$.

Category: $\mathbf{Ab}$ (Abelian groups). Subcategories: $\mathcal{T}$ (torsion), $\mathcal{F}$ (free)

Then $(\mathcal{T}, \mathcal{F})$ is a torsion pair:

- For any $G \in \mathbf{Ab}$, there is a short ex. seq. $\underset{\mathcal{T}}{\underset{\cap}{T}} \hookrightarrow G \twoheadrightarrow \underset{\mathcal{F}}{\underset{\cap}{F}}$
- $\mathrm{Hom}(\mathcal{T}, \mathcal{F}) = 0$

COTORSION PAIRS

$\mathcal{A}$: Abelian category

$\mathcal{C}, \mathcal{D}$: subcategories (closed under direct summands)

$(\mathcal{C}, \mathcal{D})$ is **cotorsion pair** if

- $\mathrm{Ext}^1(\mathcal{C}, \mathcal{D}) = 0$ (any short exact $\underset{\mathcal{D}}{\underset{\cup\!\!\!\shortmid}{D}} \hookrightarrow \underset{\mathcal{A}}{\underset{\cup\!\!\!\shortmid}{E}} \twoheadrightarrow \underset{\mathcal{C}}{\underset{\cup\!\!\!\shortmid}{C}}$ split), and
- For any $A \in \mathcal{A}$, there are short exact sequences
$D \hookrightarrow C \twoheadrightarrow A$ and $A \hookrightarrow \tilde{D} \twoheadrightarrow \tilde{C}$ $(D, \tilde{D} \in \mathcal{D};\ C, \tilde{C} \in \mathcal{C})$.

TORSION COTORSION TRIPLES

Theorem [BBOS 19; Beligiannis ~05, unpublished]

$(\mathcal{T}, \mathcal{F})$ torsion pair, $(\mathcal{F}, \mathcal{D})$ cotorsion pair.

Then $\mathcal{T}$ is equivalent to $\dfrac{\mathcal{D}}{\mathcal{F} \cap \mathcal{D}}$.

(We call $(\mathcal{T}, \mathcal{F}, \mathcal{D})$ a torsion cotorsion triple.)

"cotilting subcategory";
already determines $(\mathcal{T}, \mathcal{F}, \mathcal{D})$

APPLICATION TO GRID REPRESENTATIONS

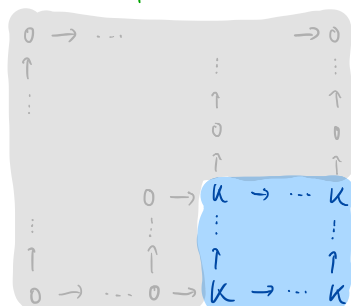
$\mathcal{D}$

Corollary [BBOS 19] $$\frac{\mathrm{Rep}^{\twoheadrightarrow}(m,n)}{\mathrm{Rep}^{\hookrightarrow\twoheadrightarrow}(m,n)} \simeq \mathrm{Rep}(m,n-1)$$

$\mathrm{Rep}^{\hookrightarrow}(m,n) = \mathcal{F} \cap \mathcal{D}$

$\mathcal{F}$

indecomposables are of the form



$\Rightarrow$ finite type

THE EQUIVALENCE, MADE EXPLICIT

$$
\begin{tikzcd}
\boxed{V_{n1} \twoheadrightarrow \cdots \twoheadrightarrow V_{nm}} & = & \boxed{I_n} & & 0 \\
 & & \boxed{I_{n-1}} \arrow[u, "f_{n-1}"] & & \ker(f_{n-1}) \arrow[u] \\
\vdots \quad \vdots \quad \vdots & & \vdots \arrow[u] & & \vdots \arrow[u] \\
\boxed{V_{21} \twoheadrightarrow \cdots \twoheadrightarrow V_{2m}} \arrow[u] & = & \boxed{I_2} \arrow[u, "f_2"] & \rightsquigarrow & \ker(f_{n-1} \circ \cdots \circ f_2) \arrow[u, hook] \\
\boxed{V_{11} \twoheadrightarrow \cdots \twoheadrightarrow V_{1m}} \arrow[u] & = & \boxed{I_1} \arrow[u, "f_1"] & & \ker(f_{n-1} \circ \cdots \circ f_2 \circ f_1) \arrow[u, hook]
\end{tikzcd}
$$

REALIZATIONS

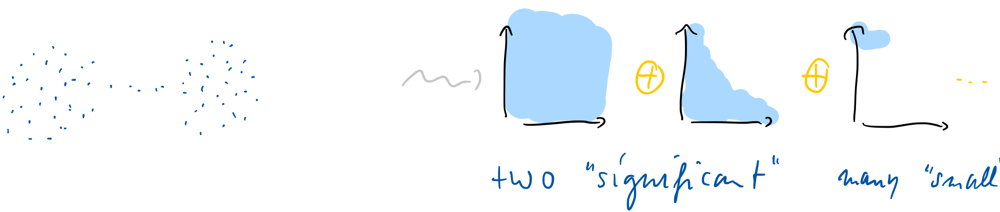
Thm [Carlsson, Zomorodian] Any $\mathrm{Rep}(m,n)$ ($m \times n$ diagram of $\mathbb{Z}_2$-vector spaces) can be realized as pth-homology (H_p of an $m \times n$ diagram of top. spaces), for any $p > 0$.

What about H_0?

Thm [BBOS]
Not every $\mathrm{Rep}^{\uparrow \twoheadrightarrow}(m,n)$ can be realized as $\tilde{H}_0$, not even as a summand (counterexample).

INDECOMPOSABLES VS CLUSTERS

- Indecomposables of H_0 from density/proximity do not correspond directly to clusters
- Rather: "linear combinations of components"
- Topological features, help to survey parameter space
- Typical example:



CHALLENGES

- COMPUTATION (OF INDECOMPOSABLES)
- CLASSIFICATION (OF REPS ARISING FROM CLUSTERING)
- INFERENCE (OF COMPONENTS IN THE PRESENCE OF NOISE)

FURTHER READING

U. Bauer, M. Botnan, S. Oppermann, J. Steen

Cotorsion Torsion Triples and the Representation Theory of Filtered Hierarchical Clustering

arXiv 1904.07322

Funding: DFG TR-109 Discretization in Geometry & Dynamics