

Multidimensional Persistence and Clustering

Magnus Bakke Botnan (TUM)

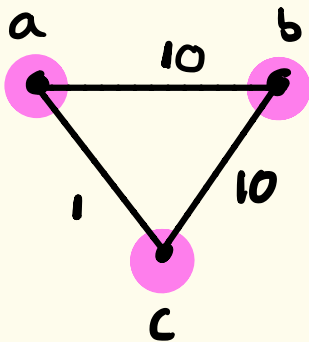
Ulrich Bauer (TUM), Steffen Oppermann (NTNU) and Johan Steen (NTNU)

Max Planck Institute, Leipzig, 19/02 2018

Theorem By Kleinberg

Let S be a finite set with $|S| \geq 2$.

A clustering function takes as input a distance $d: S \times S \rightarrow \mathbb{R}_{\geq 0}$ and returns a partition Γ of S .



$\rightsquigarrow \{ \{a, c\}, \{b\} \}$

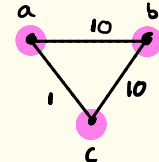
Theorem By Kleinberg

Let S be a finite set with $|S| \geq 2$.

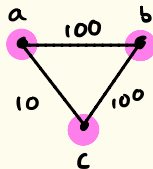
A clustering function f might satisfy:

Scale Invariance

$$\forall \alpha > 0: f(\alpha d) = f(d)$$



$$\rightsquigarrow \{\{a, c\}, \{b\}\}$$



$$\rightsquigarrow \{\{a, c\}, \{b\}\}$$

Theorem By Kleinberg

Let S be a finite set with $|S| \geq 2$.

A clustering function f might satisfy:

Richness

Let Γ be a partition of S .

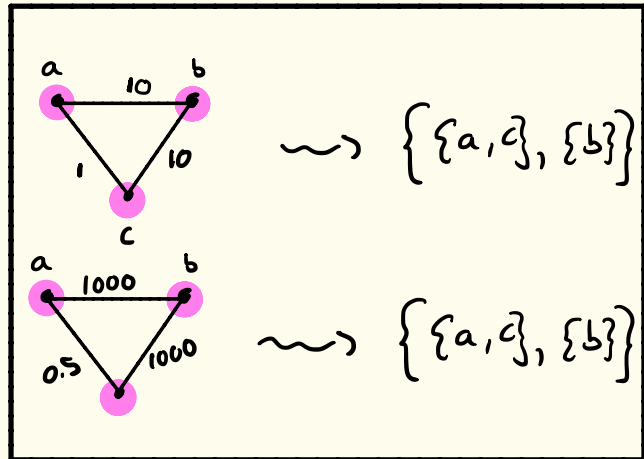
There exists $d: S \times S \rightarrow \mathbb{R}_{\geq 0}$
such that $f(d) = \Gamma$.

Theorem By Kleinberg

Let S be a finite set with $|S| \geq 2$.

A clustering function f might satisfy:

Consistency



Theorem By Kleinberg

Let S be a finite set with $|S| \geq 2$.
There exists **no** clustering
function f on S that satisfies

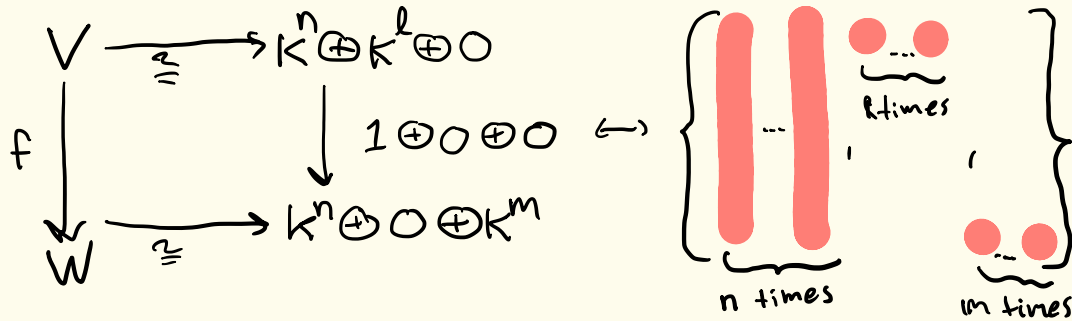
- i) Scale Invariance
- ii) Richness
- iii) Consistency

Consider two finite dimensional vector spaces V and W and a linear morphism $f: V \rightarrow W$

Then f may be reduced to a matrix of the form

$$\begin{array}{c} n \\ m \end{array} \left[\begin{array}{c|c} n & l \\ \hline \mathbb{1} & \mathbb{0} \\ \hline \mathbb{0} & \mathbb{0} \end{array} \right]$$

In particular this means that we have the following commutative diagram



So we see that $f: V \rightarrow W$ decomposes into:

- ① n copies of $K \rightarrow K$
- ② l copies of $K \rightarrow 0$
- ③ m copies of $0 \rightarrow K$

These intervals completely describe $f: V \rightarrow W$

More generally: Using similar arguments one can show that any representation

$$V = V_0 \rightarrow V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_n$$

decomposes into a direct sum of representations of the form

$$0 \rightarrow \dots \rightarrow 0 \rightarrow \underbrace{k \xrightarrow{1} k \xrightarrow{1} k \xrightarrow{1} \dots \xrightarrow{1} k}_{[a] \quad [b]} \rightarrow 0 \rightarrow \dots \rightarrow 0$$

Thm: $V \cong$ collection of intervals.

Special Case: All Surjections.

1

$$V_0 \xrightarrow{f_0} V_1 \xrightarrow{f_1} \dots \xrightarrow{f_{n-1}} V_n$$

2

Assume we have decomposed
 $V_{i+1} \xrightarrow{f_{i+1}} V_{i+2} \xrightarrow{f_{i+2}} \dots \xrightarrow{f_{n-1}} V_n$

3

Extend this to a decomposition

i) Let $\{a_1, \dots, a_k\}$ be the basis for V_{i+1}

ii) Choose a basis $\{z_1, \dots, z_m\}$ for $\ker f_i \subseteq V_i$

iii) Then $\{z_1, \dots, z_m\} \cup \{z_{m+1} \in f^{-1}(a_1), \dots, z_{m+k} \in f^{-1}(a_k)\}$

is a basis for V_i

4

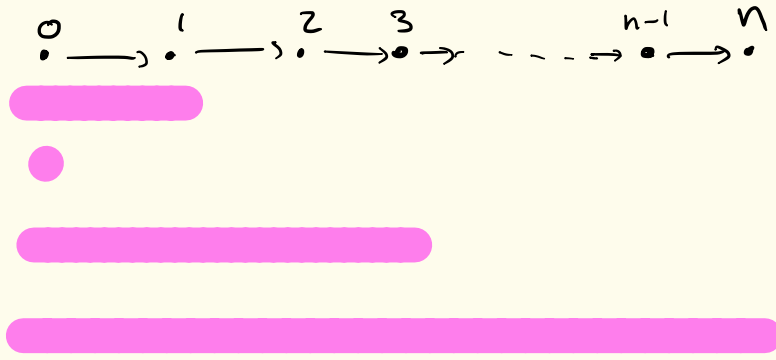
Iterate until $i=0$.

—————> B_{n-2}

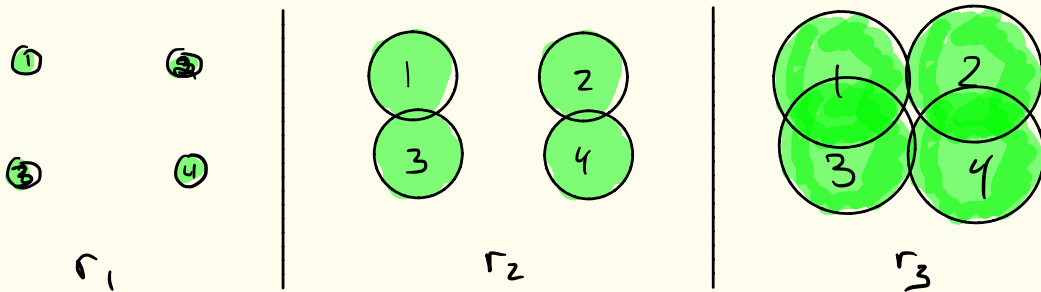
—————> $f_{n-2}^{-1}(B_{n-1})$ —————> B_{n-1}

—————> $f_{n-2}^{-1}(f_{n-1}^{-1}(B_n))$ —————> $f_{n-1}^{-1}(B_n)$ —————> B_n

Observation: every bar is supported on $i=0$.



Single Linkage Clustering



Apply 0-th homology

$$\begin{array}{ccc}
 \mathbb{k}^4 & \xrightarrow{\quad} & \mathbb{k}^2 & \xrightarrow{\quad} & \mathbb{k} \\
 \begin{array}{c} \text{1 2 3 4} \\ \begin{bmatrix} 1 & 1010 \\ 2 & 0101 \end{bmatrix} \end{array} & & \begin{array}{c} \text{1 2} \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{array} & &
 \end{array}$$

Theorem [Carlsson, Mendi]: SLC is uniquely "optimal"

optimal but.....

Chaining Effect:



Single linkage clustering would fail to identify the two clusters



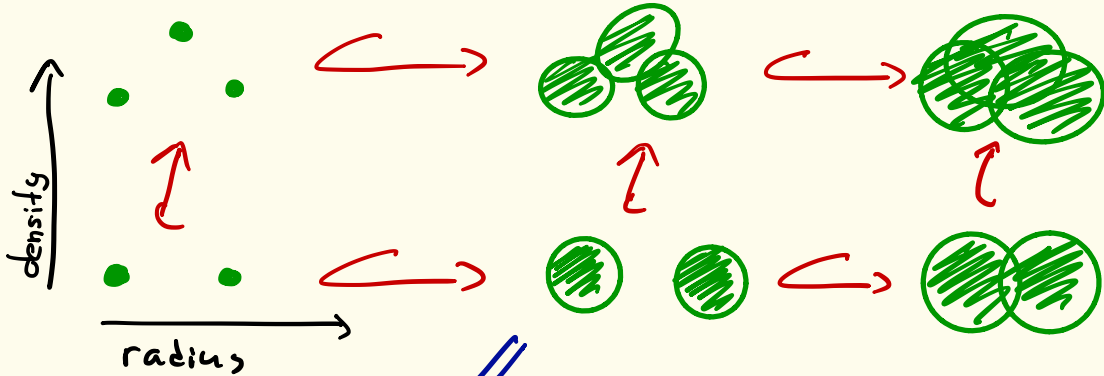
Not much used in practice

no linkage function.

*A more principled way of taking density into account, that does not depend on ad hoc constructions which destroy the stability property, would be to explicitly build the density into the method. In Carlsson and Mémoli (2009) we study **multiparameter clustering methods**, which are similar to HC methods but we track connected components in a multiparameter landscape. We also study the classification and stability properties of multiparameter clustering methods.*

2D Clustering

Carlsson + Memoli '10
Multiparameter Hierarchical
Clustering schemes



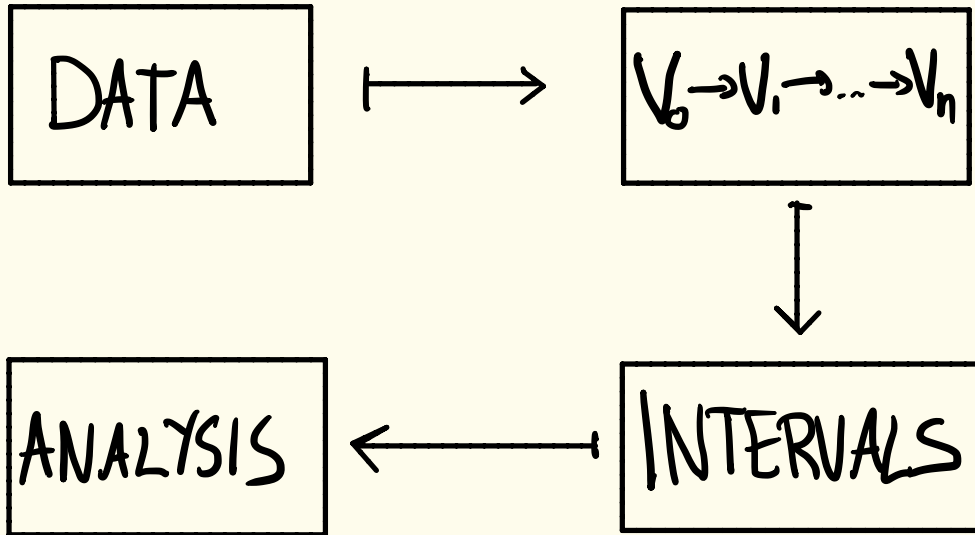
H_0

$$\begin{array}{ccccc}
 k^3 & \xrightarrow{(111)} & k & \xrightarrow{1} & k \\
 \uparrow & & \uparrow & & \uparrow 1 \\
 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & & & & \\
 k^2 & \xrightarrow{1} & k^2 & \xrightarrow{(11)} & k
 \end{array}$$

Note!

In this setup every horizontal morphism is an **epimorphism**.

TDA PIPELINE:



Poset Representations:

K alg. closed.

Let \mathcal{P} be a finite poset.

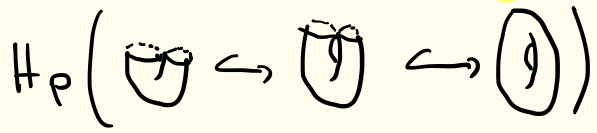
A representation V of \mathcal{P} is a collection of K -vector spaces $\{V_p\}_{p \in \mathcal{P}}$ and linear morphisms $V_p \rightarrow V_q$ whenever $p \leq q$, such that

① $V_p \rightarrow V_p$ is the identity.

② $V_p \rightarrow V_q \rightarrow V_{p'}$ commutes $\forall p \leq q \leq p'$

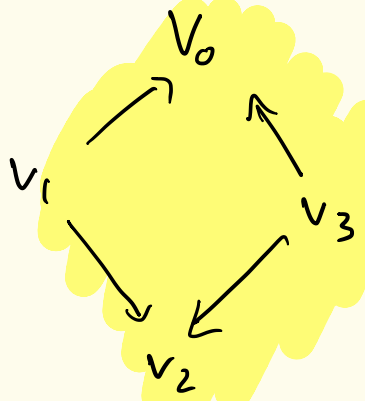
Linear

$$V_0 \longrightarrow V_1 \longrightarrow V_2$$



Zigzag

$$V_0 \longleftarrow V_1 \longrightarrow V_2$$



Circular



$$\begin{array}{cccc}
 V_{20} & \longrightarrow & V_{21} & \longrightarrow & V_{22} & \longrightarrow & V_{32} \\
 \uparrow & & \uparrow & & \uparrow & & \uparrow \\
 V_{10} & \longrightarrow & V_{11} & \longrightarrow & V_{21} & \longrightarrow & V_{31} \\
 \uparrow & & \uparrow & & \uparrow & & \uparrow \\
 V_{00} & \longrightarrow & V_{10} & \longrightarrow & V_{20} & \longrightarrow & V_{30}
 \end{array}$$

2D

Fact: Every V can be written as

$$V \cong \bigoplus_{i=1, \dots, k} V^i$$

where V^i is an **indecomposable** representation of P , in an **essentially unique** way

PROBLEM! What are the indecomposables?

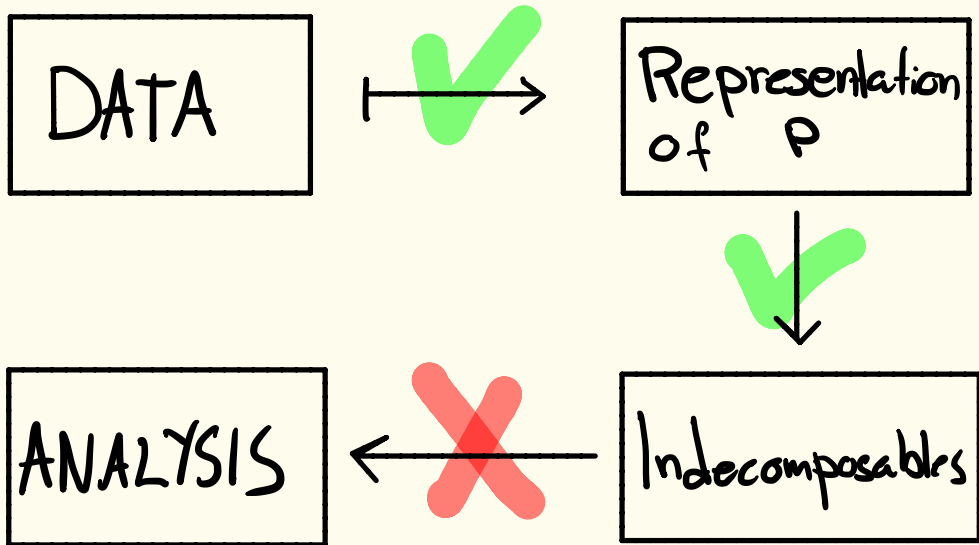
DECOMPOSITIONS

$$\begin{array}{ccc}
 K_{e_1} \xrightarrow{1} K_{e_2} & \begin{array}{l} \nearrow (1) \\ \searrow \end{array} & \begin{array}{l} K^2_{e_3, e_4} \\ K_{e_5} \end{array} \\
 \rightsquigarrow & & \\
 K_{e_1} \xrightarrow{1} K_{e_2} & \begin{array}{l} \nearrow (0) \\ \searrow \end{array} & \begin{array}{l} K^2_{e_3+e_4, e_4} \\ K_{e_5} \end{array}
 \end{array}$$

$$= \boxed{
 \begin{array}{ccc}
 K & \xrightarrow{1} & K \\
 & \nearrow 1 & \nearrow K \\
 & \searrow 1 & \searrow K
 \end{array}
 } \oplus \boxed{
 \begin{array}{ccc}
 0 & \rightarrow & 0 \\
 & \nearrow & \nearrow K \\
 & \searrow & \searrow 0
 \end{array}
 }$$

↑
Indecomposable

TDA PIPELINE for general posets



Theorem (Drozd)

Let P be a finite poset. Then (up to iso.)

① There exists a finite number of } indecomposable representations of P . } **Finite type**

or

② There exists a "canonical" way to } parametrize the isomorphism classes in } **Tame**
a sensible finite way.

or

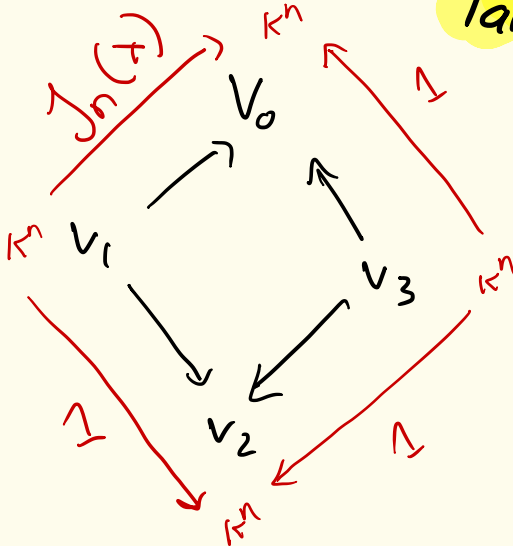
③ Impossible to understand the } indecomposable representations } **Wild**

$$K \xrightarrow{1} K \xrightarrow{0} 0$$

$$V_0 \longrightarrow V_1 \longrightarrow V_2$$

Finite type

Tame

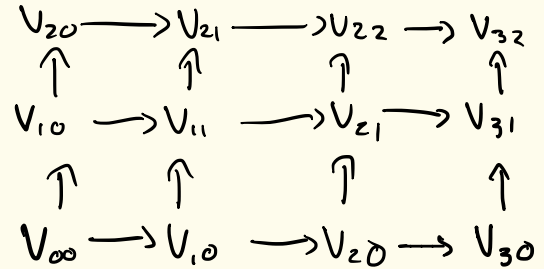


$$K \xleftarrow{1} K \xrightarrow{0} 0$$

$$V_0 \longleftarrow V_1 \longrightarrow V_2$$

Finite type

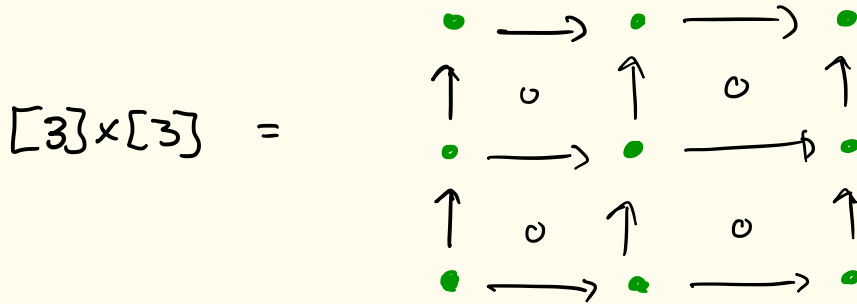
Wild



??

Multi-D

Let $P = [m] \times [n]$ be the grid poset on $m \cdot n$ vertices:



Lemma (Carlsson + Zomorodian): For every $m, n \geq 1$ and any $V \in \text{Rep}([m] \times [n])$ there exists a bifiltration of a CW-complex which yields V in p -th homology, $p \geq 0$.

Realization Example

$$\begin{array}{ccccc}
 k & \rightarrow & 0 & \rightarrow & 0 \\
 \uparrow & & \uparrow & & \uparrow \\
 (0 \ 1) & & & & \\
 \uparrow & & \uparrow & & \uparrow \\
 k^2 & \xrightarrow{(1 \ 1)} & k & \rightarrow & 0 \\
 \parallel & & \uparrow (1 \ 1) & & \uparrow \\
 k^2 & = & k^2 & \xrightarrow{(1 \ 0)} & k
 \end{array}$$

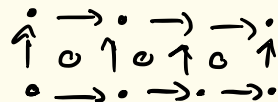
Indecomposable



Realization by a bifiltration

Not all wild!

- Hiroaka + Emerson proved using Auslander-Reiten theory that $[3] \times [2]$ and $[4] \times [2]$ are of finite type.

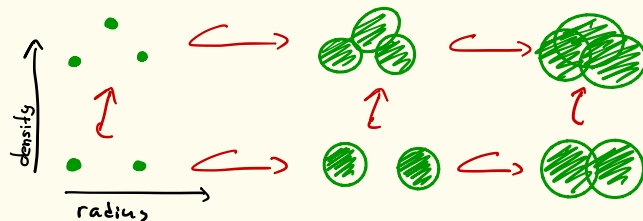


This was applied to a problem in material science.

- The case $[5] \times [2]$ is known to be tame.
- For $n \geq 6$, $[n] \times [2]$ is wild.
- $[3] \times [3]$ is also tame.

Epimorphisms

Returning to our initial problem of clustering!



This type of clustering yields representations of

the form:

$$\begin{array}{ccccccccccc} v_{1,2} & \longrightarrow & v_{2,2} & \longrightarrow & v_{3,2} & \longrightarrow & \dots & \longrightarrow & v_{n,2} \\ \uparrow & \circ & \uparrow & \circ & \uparrow & \circ & \uparrow & \circ & \uparrow \\ v_{1,1} & \longrightarrow & v_{2,1} & \longrightarrow & v_{3,1} & \longrightarrow & \dots & \longrightarrow & v_{n,1} \end{array}$$

All horizontal morphisms are epimorphisms!

Maybe restricting to such representations will help?

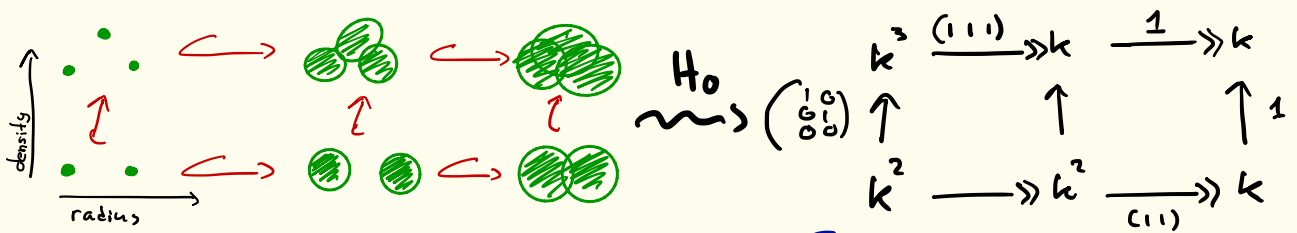
Epimorphisms

Define the following full subcategories of $\text{Rep}(\mathbb{Z}_m \times \mathbb{Z}_n)$

- $\text{Rep}^{\rightarrow}([m] \times [n])$, all horizontal morphisms epi
- $\text{Rep}^{\uparrow}([m] \times [n])$, all vertical morphisms epi
- $\text{Rep}^{\uparrow \rightarrow}([m] \times [n]) =$ Both vertical & horizontal epi.

Lemma! $\text{Rep}^{\rightarrow}([m] \times 2)$ is of finite type.

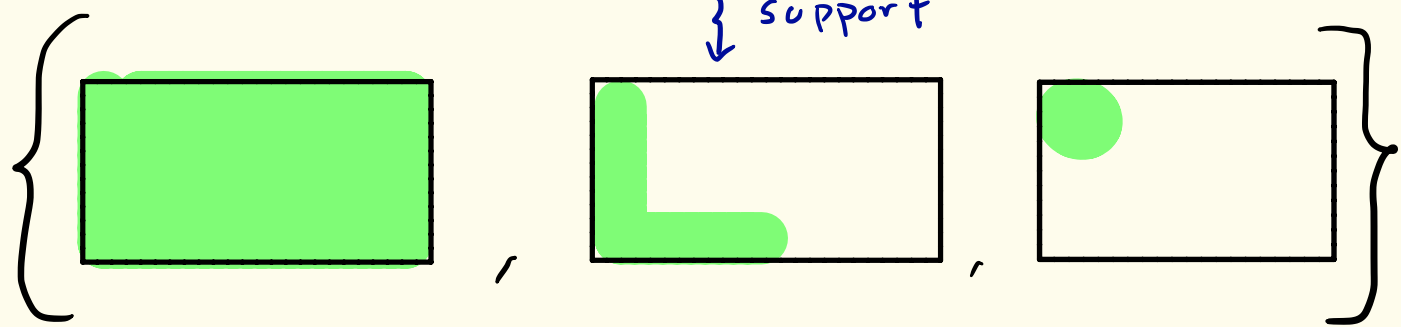
$$\begin{array}{ccc}
 \bullet & \twoheadrightarrow & \bullet \\
 \uparrow & & \uparrow \\
 \bullet & \twoheadrightarrow & \bullet
 \end{array}
 \sim
 \begin{array}{ccccc}
 k & \rightarrow & k & \rightarrow & 0 \\
 \uparrow & & \uparrow & & \uparrow \\
 k & \twoheadrightarrow & k & \twoheadrightarrow & k
 \end{array}
 \sim
 \boxed{\text{diagonal lines}}$$



Decomposition

$$\begin{array}{c}
 k \xrightarrow{1} k \xrightarrow{1} k \\
 \uparrow \quad \uparrow \quad \uparrow \\
 k \xrightarrow{1} k \xrightarrow{1} k
 \end{array}
 \oplus
 \begin{array}{c}
 k \xrightarrow{1} 0 \xrightarrow{1} 0 \\
 \uparrow \quad \uparrow \quad \uparrow \\
 k \xrightarrow{1} k \xrightarrow{1} 0
 \end{array}
 \oplus
 \begin{array}{c}
 k \xrightarrow{1} 0 \xrightarrow{1} 0 \\
 \uparrow \quad \uparrow \quad \uparrow \\
 0 \xrightarrow{1} 0 \xrightarrow{1} 0
 \end{array}$$

support



Wild Things

"Same representation type"

Theorem:

↑
special
case

$$\text{Rep}^{\hat{\rightarrow}}([m] \times [n]) \sim \text{Rep}^{\hat{\rightarrow}}([m] \times [n-1])$$

$$\text{Rep}^{\rightarrow}([m-1] \times [n]) \sim \text{Rep}([m-1] \times [n-1])$$

Corollaries:

• $\text{Rep}^{\rightarrow}([m] \times [2]) \sim \text{Rep}([m] \times 1) \sim \text{Rep}(A_m)$

↑ finite
type

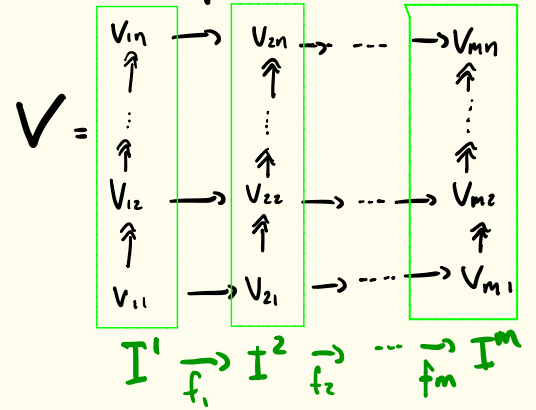
• $\text{Rep}^{\rightarrow}([m] \times [3]) \sim \text{Rep}([m] \times [2])$ which is
finite type for $m \leq 4$.

• $\text{Rep}^{\rightarrow}([m] \times [n])$ is wild for $n \geq 3$ and $m \geq 6$.

Note: Dual statements for monomorphisms.

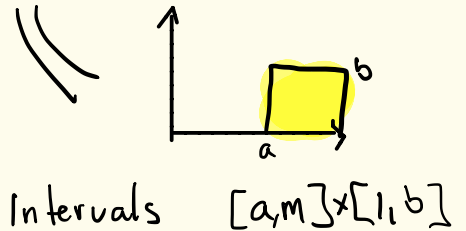
$$F: \text{Rep} \hat{\text{Rep}}([m] \times [n]) / \Pi \downarrow \cong \text{Rep}([m-1] \times [n])$$

$$V \in \text{Rep} \hat{\text{Rep}}([m] \times [n]) \Leftrightarrow$$



$$F(V) = F(I^1 \xrightarrow{f_1} \dots \xrightarrow{f_m} I^m) = \ker f_1 \rightarrow \ker(f_2 \circ f_1) \rightarrow \dots \rightarrow \ker(f_m \circ \dots \circ f_1)$$

$$\Pi \subseteq \text{Rep} \hat{\text{Rep}}([m] \times [n])$$



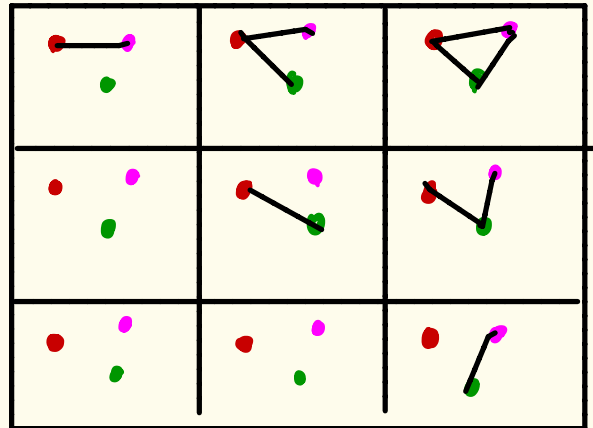
Realizations

$V \in \text{Rep}^{\mathbb{A}}([m] \times [n])$. $\exists S: [m] \times [n] \rightarrow \text{CW}$ such that

$$V \cong \mathbb{H}_0 \circ S \quad ?$$

$$\begin{array}{ccccc}
 k & \longrightarrow & 0 & \longrightarrow & 0 \\
 \uparrow \binom{1}{0} & & \uparrow & & \uparrow \\
 k^2 & \xrightarrow{\binom{1}{1}} & k & \longrightarrow & 0 \\
 \uparrow \binom{1}{1} & & \uparrow \binom{1}{1} & & \uparrow \\
 k^2 & \xrightarrow{\binom{1}{0}} & k & & k
 \end{array}$$

$$\cong \mathbb{H}_0$$



Realizations: Not always!

Lemma:

1) There are $V \in \text{Rep}^{\hat{\rightarrow}}([m] \times [n])$ for which there is no $S: [m] \times [n]$ such that $V \cong \tilde{H}_0 \circ S$.

2) There exist indecomposables $V \in \text{Rep}([m] \times [n])$ of arbitrary high dimension vector, such that $V \cong \tilde{H}_0 \circ S$

3) We can find an arbitrary number of non-isomorphic indecomposable $V_i \in \text{Rep}^{\rightarrow}([m] \times [n])$ with the same dimension vector for which

$$V_i \cong \tilde{H}_0 \circ S_i \quad \forall i$$

Realization

open problem

- Let $V \in \widehat{\text{Rep}}([m] \times [n])$.
- Does there exist $S: [m] \times [n] \rightarrow CW$ such that $H_0 \circ S \cong \bigoplus_i X_i$ and $V \cong X_j$ for some j ?

In Practice

$V \in \text{Rep}(\mathcal{Q})$ is thin if $\dim V_q \leq 1 \quad \forall q \in \mathcal{Q}$

$$\begin{array}{ccc} k & \rightarrow & 0 \\ \uparrow & & \uparrow \\ k & \xrightarrow{1} & k \end{array}$$

Thin \checkmark

$$\begin{array}{ccc} k^2 & \rightarrow & 0 \\ \uparrow & & \uparrow \\ k^2 & \rightarrow & k \end{array}$$

Not Thin \times

$$k \xrightarrow{1} k \xrightarrow{1} k \xrightarrow{1} k$$

Thin \checkmark

Decompose

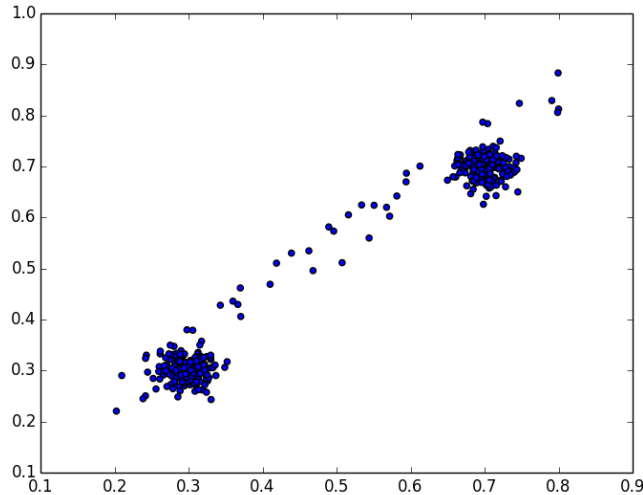
$V \cong \bigoplus V_i$. If all V_i are thin (indecomposable) then V is completely described* by a collection of intervals in \mathcal{Q} .

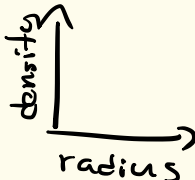
Problem:

Given a random process which generates $V \in \text{Rep}(\mathcal{Q})$. What is the probability that V has a non-thin summand?

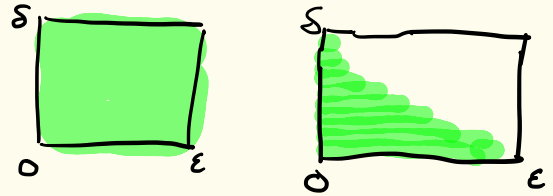
* In Multi-D

Clusters in \mathbb{R}^2

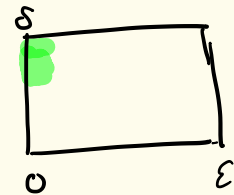


Filter by 

- All summands thin
- Two clear persistent intervals.



- Many small

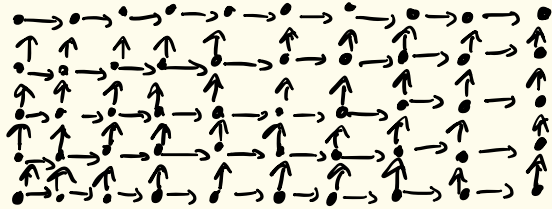


Random Points

20 points in \mathbb{R}^d

10 distance scales

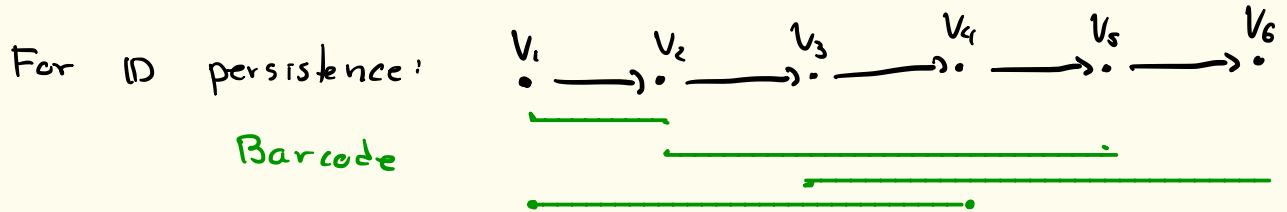
5 density scales



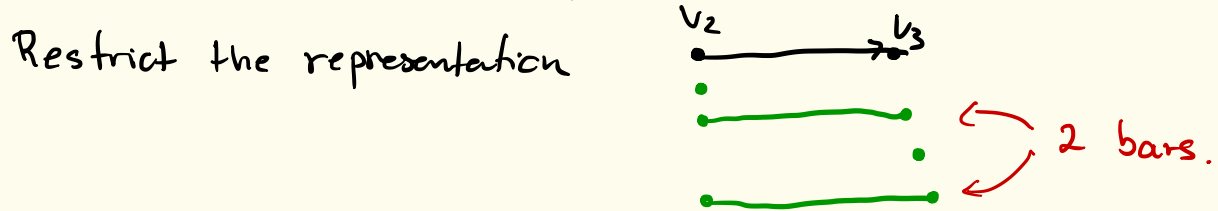
100 Runs

dimension d	2	50	200	1000
Non-thin M	14	42	56	50
Non-thin summands	15/1788	54/1742	60/1732	64/1372

Counting Bars



How many bars are supported on the whole of $[2,3]$?



Not hard to see: given an interval $[i,j]$,
the number of bars supported on the whole interval
 $[i,j]$ equals $\text{Rank}(V_i \rightarrow V_j)$.

Counting Thin Summands

Let \mathcal{Q} be a connected poset and $V \in \text{Rep}(\mathcal{Q})$ be indecomposable.

Note: V can be considered as a functor

$$V : \mathcal{Q} \longrightarrow \text{Vec}$$

Theorem:

$$\text{Rank}(\lim V \rightarrow \text{colim} V) =$$

$$\left\{ \begin{array}{l} 1 \text{ if } \dim V_i = 1 \ \forall i \\ \text{and every non-zero} \\ \text{morphism is 1.} \\ \\ 0 \text{ otherwise} \end{array} \right.$$

Example

i) $\lim = k \xrightarrow{1} k \begin{array}{l} \nearrow 1 \\ \searrow 1 \end{array} k$ $\text{colim} \cong k$

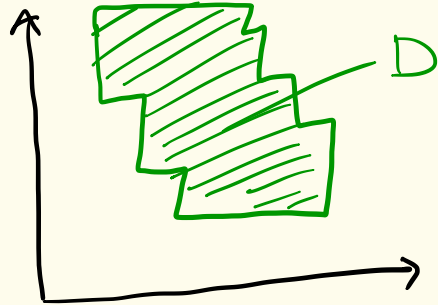
$$\text{Rank} = \underline{1}$$

ii) $\lim = k \xrightarrow{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} k \begin{array}{l} \xrightarrow{\begin{pmatrix} 0 \\ 1 \end{pmatrix}} k \\ \searrow \begin{pmatrix} 1 \\ 1 \end{pmatrix} k \end{array}$ $\text{colim} \cong 0$

$$\text{Rank} = 0$$

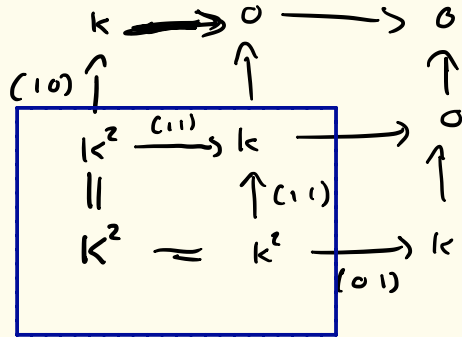
Thin Summands in Multi-D

- Let $V \in \text{Rep}(m \times n)$
and D a connected subposet.

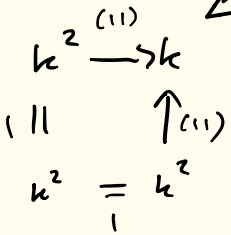


- restricts to $M|_D \in \text{Rep}(D)$.
- $V|_D \cong \bigoplus_j V_j$
- The number of j 's such that $\dim(V_j|_i) = 1 \forall i \in Q_0$
 $= \text{rk}(\lim V|_D \rightarrow \text{colim } V|_D)$.
- Can be efficiently computed using zigzag persistence.
- New interpretation of the rank invariant.

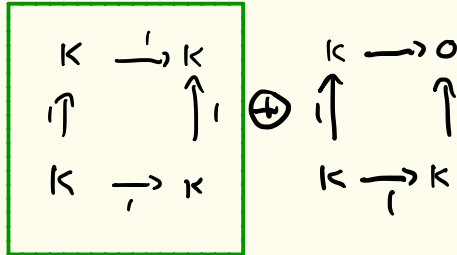
Thin Example



Indecomposable.
 $\text{Rank}(k^2 \rightarrow 0) = 0.$



\cong



$$\lim \longrightarrow \text{colim}$$

$$= k^2 \xrightarrow{(1,1)} k$$

$$\text{Rank } k = 1$$

In interleaving Distance

The interleaving distance is the most discriminative stable distance on multi-dimensional persistence modules

Lesnick
(IK = IFP)

open problem: Computational complexity of computing $d_I(M, N)$.

partial Result: • NP-Hard for a generalization

• At least as hard as:

$$n \begin{bmatrix} 0 & 0 & 0 & ? \\ ? & ? & ? & ? \\ ? & ? & 0 & ? \\ 0 & ? & ? & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & ? & 0 \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & 0 & ? & ? \end{bmatrix} = I_n$$

S/w \uparrow

Håvard

Bjerkevik

Example

i)
$$\begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \circ & \bullet \\ \bullet & \bullet & \circ \end{bmatrix} \begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \circ \\ \bullet & \circ & \bullet \end{bmatrix} = \begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix} \quad \checkmark$$

ii)
$$\begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix} \begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix} = \begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix} \quad \times$$

