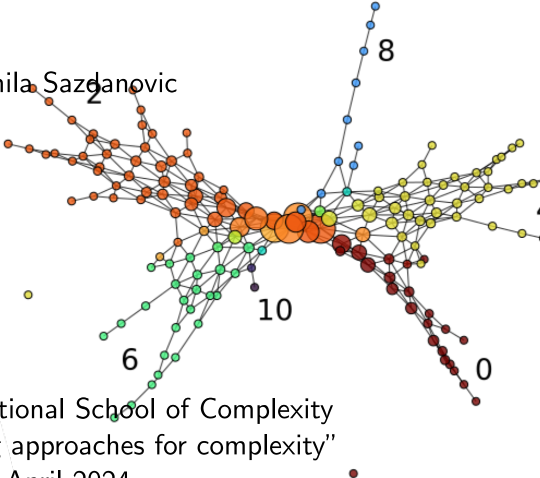
DATA, RELATIONS AND THEIR SHAPE
PART II

Radmila Sazdanovic



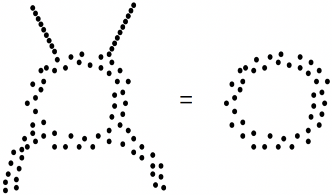
NC STATE UNIVERSITY

18th Erice International School of Complexity
"Machine Learning approaches for complexity"
25 April 2024

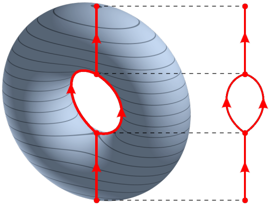
- MAPPER Singh, Mémoli and Carlsson (2007)
- BALL MAPPER Dłotko (2019)
- NEW MAPPER-LIKE TECHNIQUES broaden the scope and applicability of mapper-type algorithms to utilize additional structure of data and visualize the maps between datasets.
- APPLICATIONS; games, materials, knots...

COLLABORATORS

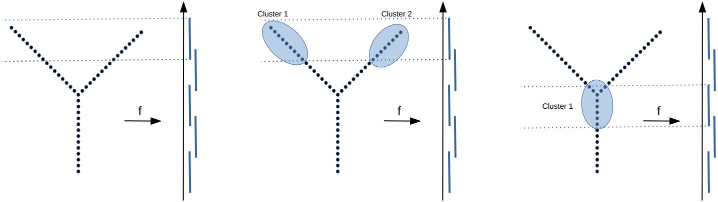
- Mustafa Hajij, Jesse Levitt
- Pawel Dlotko, Davide Gurnari



MAPPER MOTIVATION: REEB GRAPH



- INPUT M, $f : M \to \mathbb{R}$.
- EQUIVALENCE RELATION $x \sim y$, for $x, y \in M$, iff:
  - $f(x) = f(y)$,
  - x and y belong to the same connected component of $f^{-1}(x)$.
- REEB GRAPH quotient space $M/_{\sim}$.



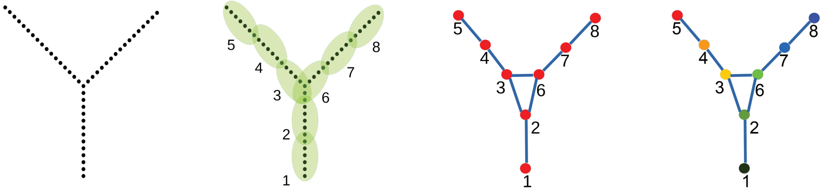
Lens/height function

Elements of a cover

MAPPER: SINGH , MÉMOLI AND CARLSSON (2007)

VISUALIZING HIGH DIMENSIONAL DATA

- INPUT: data, lens function, overlapping cover of $\mathbb{R}$
- MAPPER GRAPH: the nerve of the pullback cover of $\mathbb{R}$ to X
- ADDITION (coloring) function

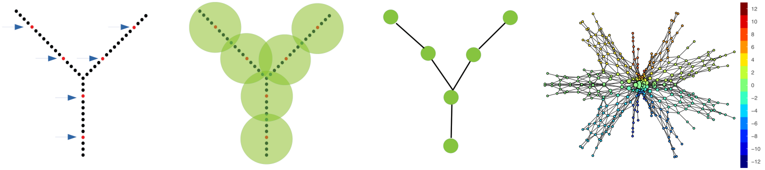


Point cloud   Cover   1-dim nerve   Coloring

BALL MAPPER ALGORITHM DŁOTKO 2019

A MORE EFFICIENT WAY TO OBTAIN A COVER

- INPUT: Point cloud X and a single parameter $\epsilon > 0$ (radius).
- STEP 1: LANDMARKS Construct an ϵ-net of X is $Y \subset X$ such that for every $x \in X$ there exist $y \in Y$ such that $d(x, y) \leq \epsilon$.
- STEP 2: COVER $X \subset \bigcup_{y \in Y} B(y, \epsilon)$, therefore the collection of balls centered at Y of a radius ϵ form a cover of X.

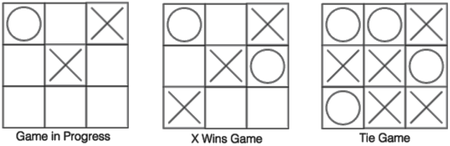


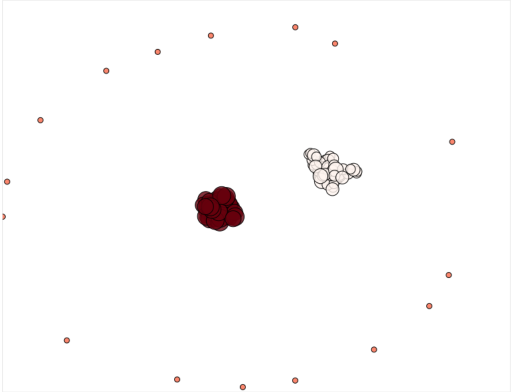
Landmarks Cover Ball Mapper Example

- INPUT Point cloud X in metric space with a group G of isometric automorphism acting on it.
- GOAL Lift this action to the ball mapper graph by insuring that the set of landmarks respects the group action.
- HOW Add the whole orbit to the collection of landmarks.
- OUTPUT Ball mapper graph with inherited symmetries.

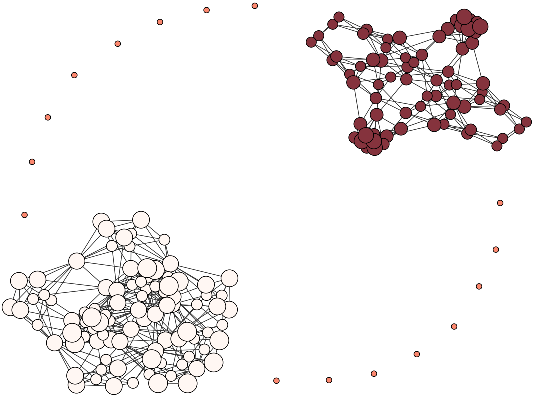
TIC TAC TOE ENDGAME DATA

958 vectors in $\mathbb{R}^9$ representing 3 by 3 matrices with ± 1 for x and o and 0 for empty spots.

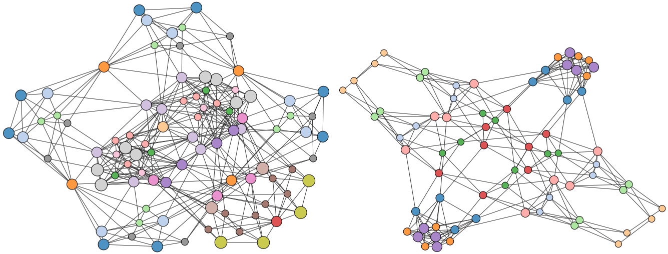




- Data: Tic-Tac-Toe endgame data set consisting of 958 vectors in $\mathbb{R}^9$ representing 3 by 3 matrices.
- Ball mapper for $\epsilon = 2.5$ colored by the wins of the first player (red), loses (white), disjoint clusters (ties).



- Equivariant BM colored by the wins of the first player (red), loses (white), disjoint clusters (ties).
- Win, tie, and losses clusters are disjoint for $\epsilon < 3$ as symmetries of the board do not change the outcome.



- The wins cluster (left) and loses cluster (right) with color denoting the orbits.
- Different orbits might have different lengths. Asymmetric configurations have length 8 orbits.
- The maximally symmetric configuration has an orbit of length 1 -the only red node (left).

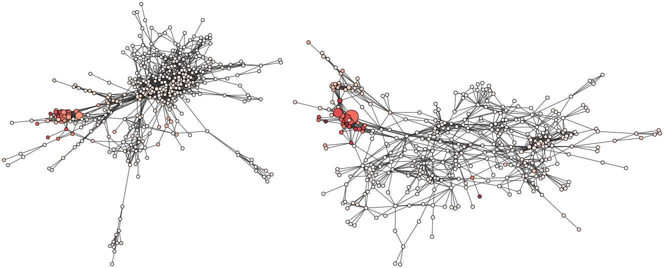
- Mapper is a tool to build a model of a space X along with visualizing a function $f : X \to \mathbb{R}$ by coloring the output.
- INPUT $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ and a relation $f \subset X \times Y$
- STEP 1 Construct BM graphs $BM_x(V_X, E_X)$, $BM_Y(V_Y, E_Y)$
- STEP 2 Define a map $\tilde{f} : V_X \to V_Y^{[0,1]}$, where $V_Y^{[0,1]}$ denotes a set of functions from V_Y to $[0, 1]$ for every vertex $w \in V_Y$

$$\tilde{f}(w) = \frac{|B(w) \cap f(B(v))|}{|B(w)|}$$

is the percentage of points in w that are in the image of the points covered by the vertex v in $G(X)$.
- OUTPUT coloring function on vertices of BM_Y for every vertex $v \in BM_X$.

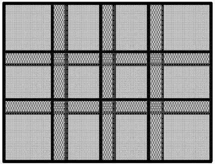
- Superconductor data, UC Irvine ML Repository. Size: 21263

1. Characteristics dataset: 81 features scaled, with a radius 2
2. Chemical composition dataset: sparse vectors in $\mathbb{R}^{86}$. ϵ is 0.25 for cosine similarity measure. Concentration of measure.



- Hamidieh: map that relates two datasets and sends one cluster in Ball Mapper of each dataset to multiple disconnected in the other. Sets likely provide distinct and unrelated information.

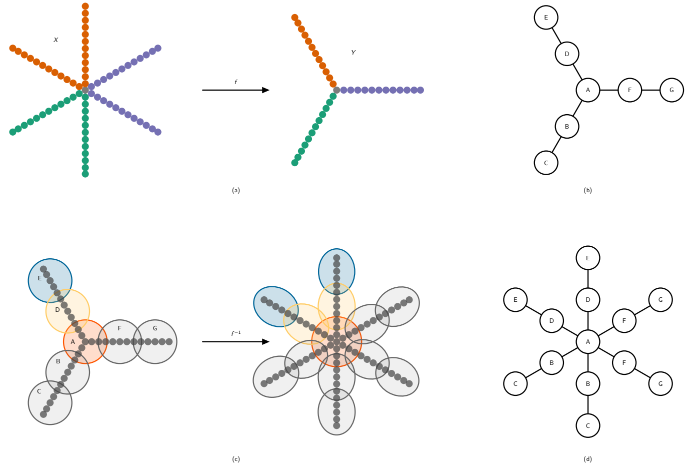
n=1



n=2

- Let $f : X \to Y$ where $X \subset \mathbb{R}^N$, $Y \subset \mathbb{R}^n$, and $0 \ll n \ll N$.
- MAPPER Lenses with domains in $\mathbb{R}^n$ are more likely to preserve essential information about the point cloud, and the fact that having k intervals in each of n directions requires k^n cover elements
- MAPPER ON BALL MAPPER Construct Mapper of X using
  - f as a lens function and
  - Ball Mapper as an overlaping cover of $f(X) \subset Y \subset \mathbb{R}^n$.

MAPPER ON BALL MAPPER: ILLUSTRATION



(A) The map between point clouds $f : X \to Y$

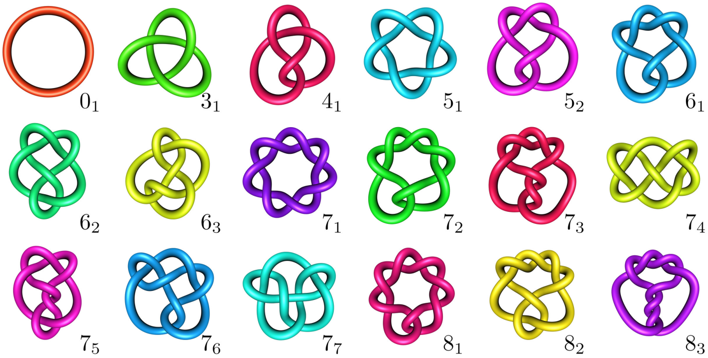
(B) Ball Mapper for Y

(C) Cover of X obtained as a pullback of the cover of Y

(D) MonBM for X, labeled by cover elements in Y.

BALL MAPPER: SUMMARY

- BALL MAPPER provides a way to visualize high dimensional data in any metric space that preserves proximity and depends on data density.
- EQUIVARIANT BALL MAPPER provides tools for visualizing data with observed or assumed isometries. Challenging/impossible to achieve on Mapper.
- MAPPER ON BALL MAPPER (MoBM) hybrid between the two algorithms. Extends the Mapper algorithm from 1-dimensional to high dimensional lens functions in a computable way more likely to preserve and reveal information about the input data.
- MAPPINGMAPPERS allows to visualize maps between high dimensional data sets and enables comparison



KNOT is an equivalence class of smooth embeddings $f : S^1 \to \mathbb{R}^3$
Two knots are equivalent if they can be connected by an isotopy.

HOW TO DISTINGUISH KNOTS?

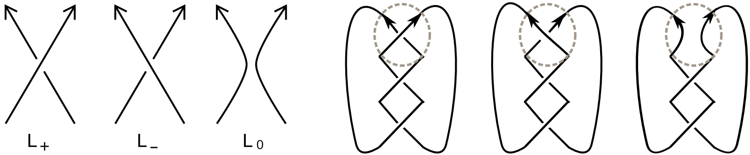
Knot invariants: have the same value on isotopic knots but might fail to distinguish all of them (incomplete).

- There are more than 50 knot invariants of various types
  - numerical: components, linkings, colorings
  - polynomial: Alexander, Jones, Kauffman, etc.
  - algebraic: group of colorings, knot group, homology theories such as Khovanov link, knot Floer, etc.
- Software: SnapPy, Knot Theory, KnotPlot, etc.
- Databases: Knot Atlas, KnotInfo, etc.

WHY

- Related to a number of questions in 3- and 4–dim topology
- Applications: biology, materials science, physics in S^3.

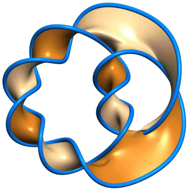
POLYNOMIAL KNOT INVARIANTS



- Alexander $(q^{1/2} - q^{-1/2}) \cdot \Delta_{L_0}(q) = \Delta_{L_+}(q) - \Delta_{L_-}(q)$
- Jones polynomial:
  - Skein relation $(q^{1/2} - q^{-1/2}) \cdot J_{L_0}(q) = q^{-1} \cdot J_{L_+}(q) - q \cdot J_{L_-}(q)$
  - Hecke algebra of the braid group
  - Quantum field theory as the unknot normalized vacuum expectation value of the Wilson loop operator in $SU(2)$ Chern–Simons gauge theory
- HOMFLY-PT: $z \cdot H_{L_0}(q) = a \cdot H_{L_+}(q) - a^{-1} \cdot H_{L_-}(q)$
- Khovanov homology – categorification of the Jones polynomial
- Bar Natan and Van der Veen: ρ-polynomial

- Minimal crossing number
- The signature $\sigma(K)$: computable from the Alexander module

  $\sigma(K)$ is the signature of $V + V^T$ for V Seifert form whose entries are linking numbers of pushoffs of generators of H_1 of the Seifert surface.
  Combinatorial formula for alternating knots by Traczyk



$$\sigma(K) = s_A(D) - n_+(D) - 1 = -s_B(D) + n_-(D) + 1$$

- Rasmussen s-invariant: defined using Lee spectral sequence on Khovanov homology.

- DISTINGUISHING KNOTS
  Compare their invariants until you find an invariant that distinguishes them!
- CHARACTERIZING INVARIANTS
  Given a knot invariant which is not complete, which knots can it distinguish
- DOES IT DISTINGUISH THE UNKNOT?
  Is there a non-trivial knot with the same value of the invariant as the unknot?
  - genus, knot Floer homology, Khovanov homology can
  - it is still an open question for the Jones polynomial

- The unknot: Kronheimer–Mrowka (2010)
- The unlink Hedden–Ni (2013), Batson–Seed (2015)
- The trefoils Baldwin–Sivek (2018)
- The Hopf link Baldwin–Sivek–Xie (2018)
- The connected sum of two Hopf links, the torus link $T(2,4)$ Xie–Zhang (2019)
- Split links Lipshitz–Sarkar (2019)
- the torus link $T(2,6)$ Martin (2020)
- $L6n1$ Xie–Zhang (2020)
- $L7n1$, the connected sum of a trefoil and Hopf link Li–Xie–Zhang (2020)

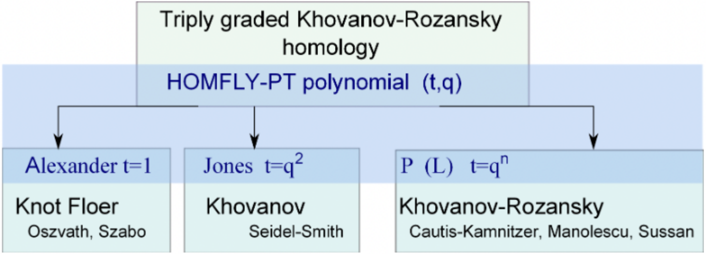
THEOREM (BALDWIN, DOWLIN, LEVINE, LIDMAN, S.)

Let $K \subset S^3$ be a knot whose reduced Khovanov homology over $\mathbb{Q}$ is 5-dimensional and is supported in a single δ-grading d. Then:

1. *If $d = 0$, then K is the figure-eight knot.*
2. *If $d \neq 0$, then, up to mirroring, $d = 2$ and K is a*
    - *genus 2*
    - *fibered*
    - *strongly quasipositive knot*

    whose bigraded knot Floer homology over $\mathbb{Q}$ is isomorphic to that of the torus knot $T(2,5)$.

COMPARING KNOT INVARIANTS



WHAT ARE THEIR DISCRIMINATIVE POWERS?

- HOMFLYPT specializes to Jones and Alexander
- Link homology theories are stronger than the polynomials
- Khovanov homology of alternating knots is determined by the Jones polynomial and signature (Shumakovitch)

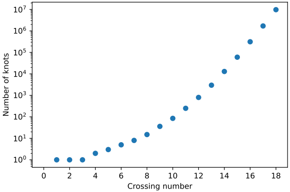
KNOTS ARE BIG DATA

Theorem (Ernst, Sumners '87)

The number of distinct knots grows exponentially with the crossing number.

crossings	up to 17	18	19	20
no. of knots	1,701, 936	9,755,329	350×10^6	1,847,319,428

- Rolfsen's table; Hoste, Thistlethwaite, Weeks, Burton
- 10,718,938,763,889 knot diagrams for up to 23 crossings (Sikora, Tuzun).



2019 Hughes: Neural network approach to predicting and computing knot invariants.

2019 Jejjalaa, Karb, Parrikarb: Hyperbolic Volume of a Knot

2020 Ward, Rawdon: Data mining and deep learning

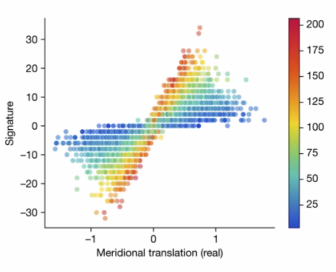
2021 Gukov, Halverson, Ruehle, Sułkowski: Learning to unknot.

2021 Davies, Velickovic, Buesing, Blackwell, Zheng, Tomasev, Tanburn, Battaglia, Blundell, Juhasz, Lackenby, Williamson, Hassabis, Kohli: Advancing mathematics by guiding human intuition with AI.

2022 Craven, Hughes, Jejjala, Kar: Illuminating new and known relations between knot invariants

2023 Gukov, Halverson, Manolescu, Ruehle: Searching for ribbons with machine learning.

- Predicting values and finding relations between knot invariants
  - An inequality that relates the signature, slope, volume, and injectivity radius of hyperbolic knots was discovered using ML.
  - 200,000 experiments trying to uncover pairwise and triple correlations between invariants measured by the accuracy of the neural network prediction.



- Proving conjectures

2021 Detecting the unknot

1962-2023 Fox's slice-ribbon conjecture: Every slice knot is ribbon.

INPUT The coefficients of the one-variable polynomial invariant $\mathcal{I}$ of a finite collection of knots

STEP 1 Given a knot K and its single variable polynomial $I(K)$ extract a vector of the coefficients

STEP 2 Compute the minimal and maximal powers min_t, max_t of the variable denoted by t among all knots in $\mathcal{K}$. Then the maximal length of all such vectors is $d = max_t - min_t + 1$.

STEP 3 Add zeros on both sides of each vector of coefficients to obtain a vector $I(K)_v \in \mathbb{R}^d$ to ensure a correct alignment of corresponding powers.

Shift vectors so q^0 is in the same position in every vector.

	q^{-3}	q^{-2}	q^{-1}	q^0	q^1	q^2	q^3	q^4	q^5	q^6	q^7
$J(0_1)$	0	0	0	1	0	0	0	0	0	0	0
$J(\text{mir}(3_1))$	0	0	0	0	1	0	1	-1	0	0	0
$J(4_1)$	0	1	-1	1	-1	1	0	0	0	0	0
$J(\text{mir}(5_1))$	0	0	0	0	0	1	0	1	-1	1	-1
$J(\text{mir}(5_2))$	0	0	0	0	1	-1	2	-1	1	-1	0
$J(\text{mir}(6_1))$	0	1	-1	2	-2	1	-1	1	0	0	0
$J(\text{mir}(6_2))$	0	0	1	-1	2	-2	2	-2	1	0	0
$J(6_3)$	-1	2	-2	3	-2	2	-1	0	0	0	0

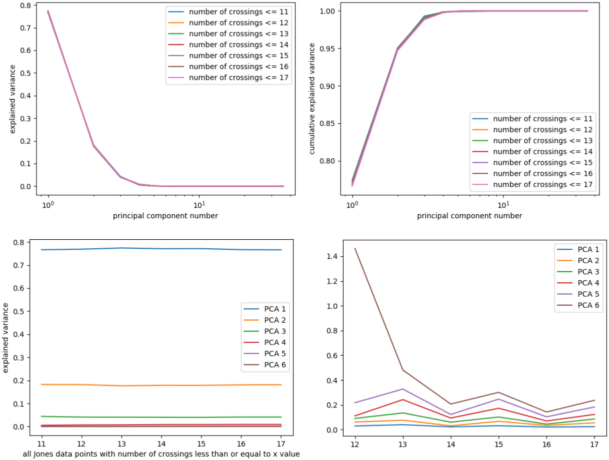
Invariant	Unknot	Trefoil	Data vector
Alexander	1	$t^{-1} - 1 + t$	(0,1,-1,1,0)
Jones	1	$t + t^3 - t^4$	(0,0,0,0,0,1,0,1,-1)
HOMFLYPT	1	$-a^4 + 2a^2 + a^2z^2$	(-1,0,0,0,0,0,2,0,0,**0,0,0,0,0,0,1,0,0**)
Khovanov Q	$q^{-1} + q$	$q^{-9}t^{-3} + q^{-5}t^{-2}$ $+q^{-3} + q^{-1}$	(0,0,0,1,0,0,0,**0,0,0,1,0,0,0**,0,1,0,0,0,0,0, **0,0,0,0,0,0,0**,1,0,0,0,0,0,0)

- Values of several knot polynomials for the unknot and the trefoil and the corresponding data vector.
- Note that for the 2-variable polynomial the matrix is flattened into a vector; for the HOMFLYPT variable z is in rows, a in columns, and for the Khovanov q is in rows and t in columns.

dataset	*rho*	Alexander	Jones	HOMFLY-PT	Khovanov
dimension	15	17	51	152	3003

- Dimensionality of the point cloud: The smallest value **d** for which the normalized explained variance of the first **d** principal components sums to more than 95% across all considered orders (joint with M. Hajij and J. Levitt)
- PCA projection of the Jones polynomial data for up to 17 crossings to 3 dimensions colored by the knot signature.

EXPLAINED VARIANCE FOR THE JONES POLYNOMIAL



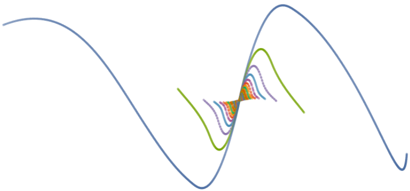
x-axis: the log of the bounding norm of r_j

y axis: $\theta_{i,j}$

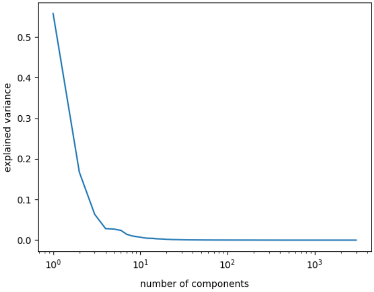
i: colors.

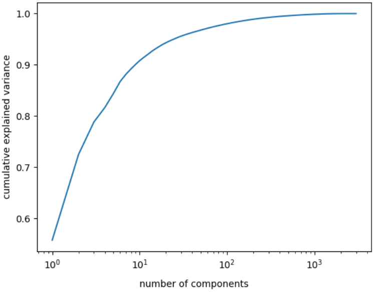
- Explained variance plotted against the crossing filtration is remarkably level (left).
- Cumulative version showing that the more significant the component, the more stable it is (right).

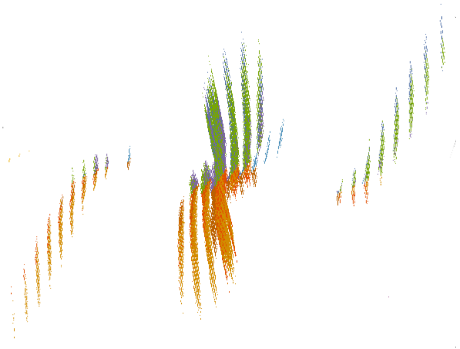
Explained variance for torus Knots



Explained variance plotted against the index of the principal component (left). The incremental summation plotted against the index of the principal component (right).

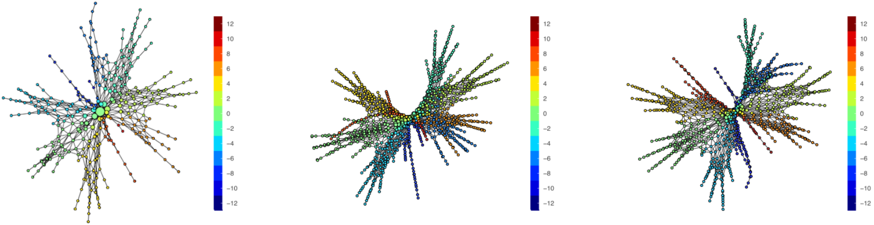




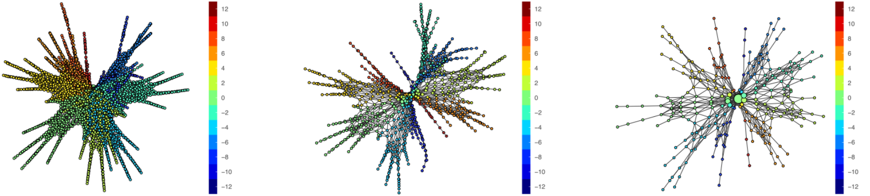


- PCA projection of the Alexander polynomial data for up to 17 crossings to 2 dimensions colored by the knot signature to highlight internal structure. Approximates 1D manifold.

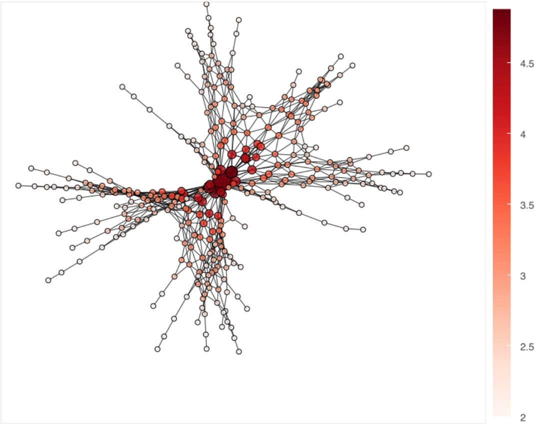
BALL MAPPER ON JONES DATA: STABILITY



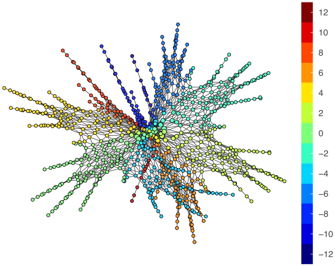
Knots up to 15, 16 and 17 crossings

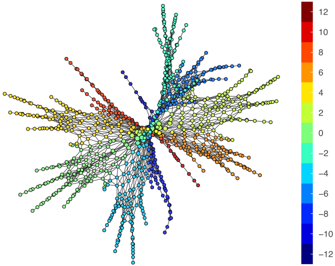


Zooming in: Knots up to 17 crossings for $\epsilon = 50, 100, 200$.


4.5
4
3.5
3
2.5
2

- Mirror $mir(K)$ of a knot K is obtained by switching all positive and negative crossings in the knot diagram of K.
- $A(mir(K))(t) = A(K)(t)$ and $J(mir(K))(q) = J(K)(q^{-1})$
- $\sigma(mir(K)) = -\sigma(K)$

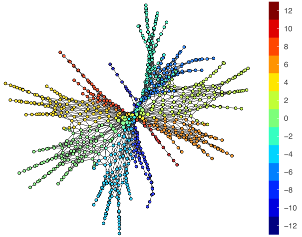




Theorem (Garoufalidis '03)

For all simple knots up to 8 crossings and for all torus knots, the colored Jones polynomial determines the signature of the knot.

- Conjectured for all simple knots.
- A knot K is simple, if all the roots $\alpha \in \Delta(K)$ of the Alexander polynomial where $|\alpha| = 1$ have multiplicity 1.

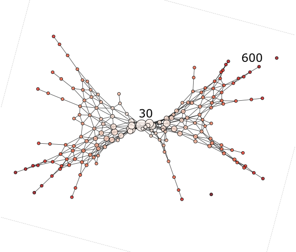


EXAMINE RELATIONS BETWEEN JONES AND SIGNATURE

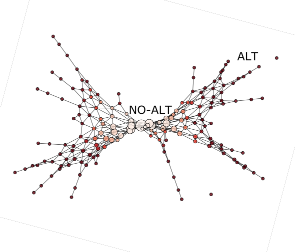
Jones polynomial is colored for N=1.

- They can be small
  $\{4_1, \text{mir}(11n19)\}$ $0, 4$ $\sqrt{5}$
  $\{\text{mir}(7_2), 11n88\}$ $2, 6$ $\sqrt{17}$
- There can be multiple in a grouping
  $\{\text{mir}(5_2), 11n57, 13n3082\}$ $2, 6$ 3
  $\{13n137, \mathbf{13n627}, \mathbf{13n716}, 13n1539, \text{mir}(13n1560), \ldots\}$ with signature either $0, 4$ and $L2$-norm equal to $\sqrt{509}$.[1]
- They can have the same number of crossings
  $\{11n28, 11n64\}$ $0, 4$
  $\{12n107, 12n171\}$ $2, 6$
- They can both be alternating
  $\{12a802, 12a1242\}$ $2, 6$ $\sqrt{215}$

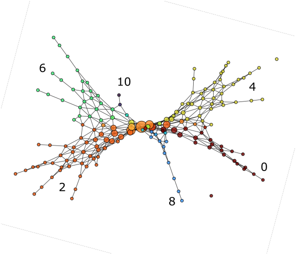
[1] **13n627**, **13n716** have identical signature, determinant, Alexander, HOMFLYPT and Kauffman polynomials but different DT-codes



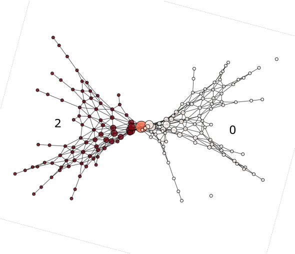
determinant



alt vs. nonalt

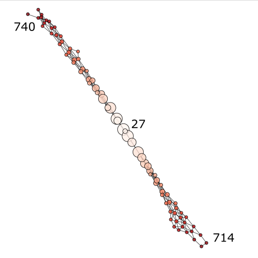


signature

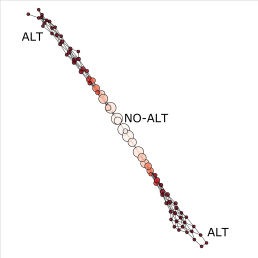


signature mod 4

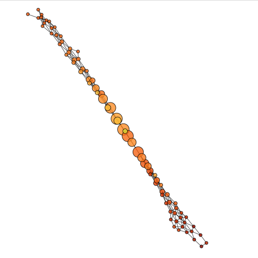
ALEXANDER DATA



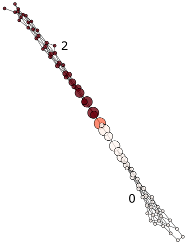
determinant



alt vs. nonalt

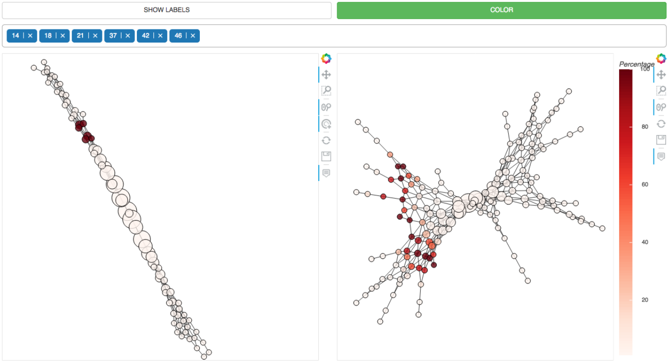


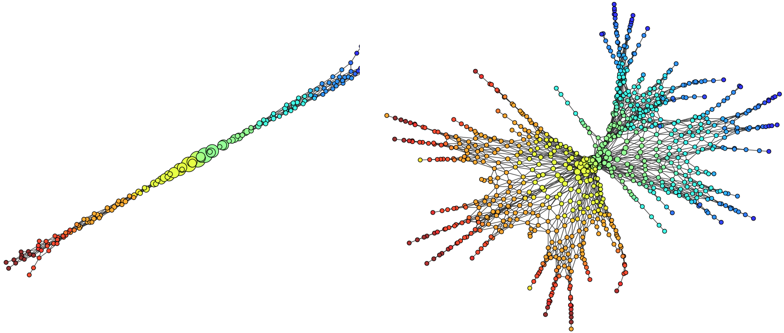
signature



signature mod 4

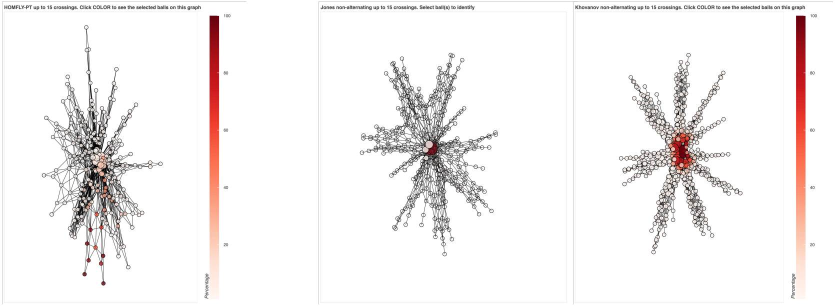
- Mapper is a tool to build a model of a space X along with visualizing a function $f : X \to \mathbb{R}$ by coloring the output.
- Use MappingMappers to compare polynomial knot invariants. Custom colored Alexander Ball Mapper (left) and Jones ball mapper (right) region correspondence





- Correspondence between the Alexander polynomial point cloud in $\mathbb{R}^{17}$ and the space of Jones polynomial point cloud $\mathbb{R}^{51}$.
- PCA explained: shared direction baased on the value of the determinant of a knot.

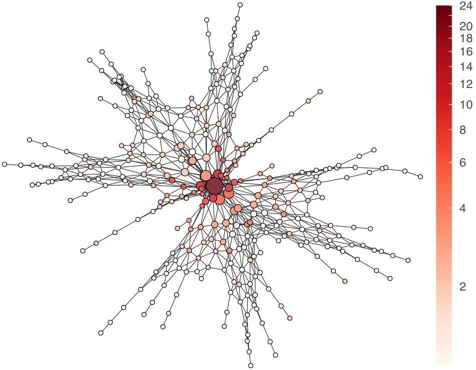
MAPPINGMAPPERS: COMPARISON



(A) HOMFLYPT (B) Jones (C) Khovanov

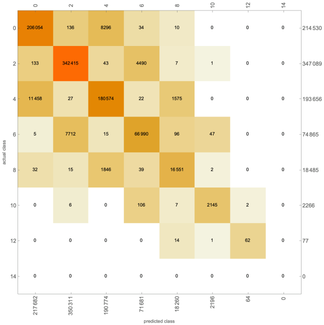
- HOMFLYPT and Khovanov ball mapper graphs colored by the percentage of knots contained in the selected in the Jones ball mapper graph.
- Analyze distrubitions of knots in star-like structures of ball mapper graphs.

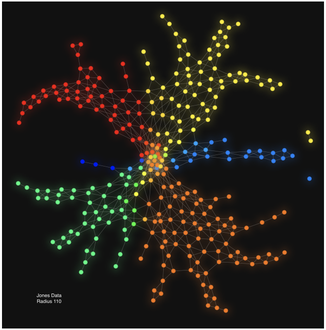
MAPPER ON BALL MAPPER: HOMFLYPT FROM JONES



- Jones Ball Mapper graph colored by the cardinality of the fiber
- Color depends on the number of clusters of the Mapper on Ball Mapper HOMFLYPT graph in the preimage of each node.

ML vs. Ball Mapper on Jones and Signature



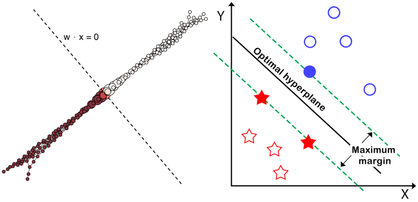


Theorem (Hajij-Levitt-S.)

Given a knot K with the span of the Jones polynomial equal to s, and K_{nn} is the knot with Jones span less than s is the nearest neighbor to K under the L^2-norm, then the probability

$$P(\sigma(K) = \sigma(K_{nn})) > 95\%$$

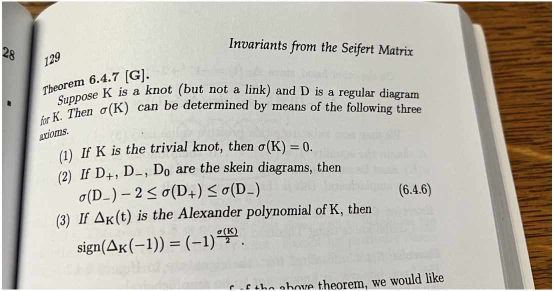
THEOREMS (RE)DISCOVERED



SVM (Support machine vector) algorithm that finds a hyperplane in an N-dimensional feature space that distinctly classifies the data points.

- The Alexander Ball Mapper graph colored by signature mod 4.
- LINEAR SVM CLASSIFIER trained on the Alexander polynomial data, get perfect separation with $w = [1, -1, 1, \ldots, -1, 1]$.
- SVM fails to converge on the Jones polynomial data
- MAPPING MAPPERS suggests same holds for the Jones polynomial

- Folklore theorem attributed to many (Alexander, Conway, Giller, written in paper of Przytycki and Traczyk) that the Jones and Alexander polynomial essentially determine signature mod 4.
- $sign(\Delta_K(-1)) = sign(J_K(-1)) = (-1)^{\frac{\sigma(K)}{2}}$


28

129 *Invariants from the Seifert Matrix*

Theorem 6.4.7 [G]. *Suppose* K *is a knot (but not a link) and* D *is a regular diagram for* K. *Then* $\sigma(K)$ *can be determined by means of the following three axioms.*

(1) *If* K *is the trivial knot, then* $\sigma(K) = 0$.

(2) *If* D_+, D_-, D_0 *are the skein diagrams, then*

$$\sigma(D_-) - 2 \leq \sigma(D_+) \leq \sigma(D_-) \qquad (6.4.6)$$

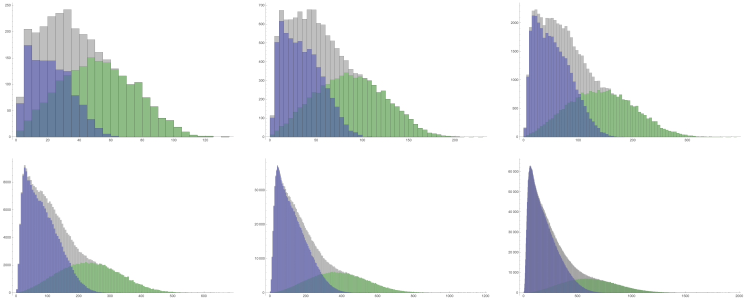
(3) *If* $\Delta_K(t)$ *is the Alexander polynomial of* K, *then*

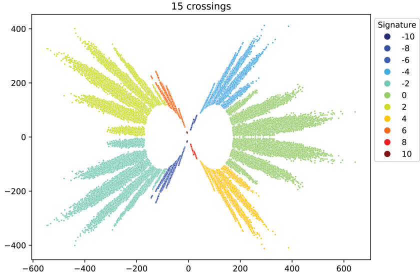
$$\text{sign}(\Delta_K(-1)) = (-1)^{\frac{\sigma(K)}{2}}.$$

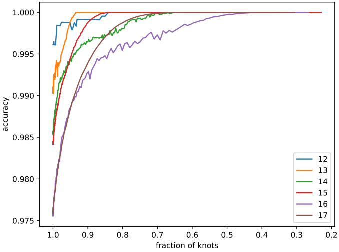
... of the above theorem, we would like

SAMPLING ISSUES

- The full data set is of infinite size
- Special families to check extrapolation or random generator
- Filtrations help extrapolation: norm filtration suffers from batch size issues but crossing filtration is consistent
- The distribution of the l_2-norms (total count vs. norm) for the alternating (green), nonalternating (blue), and all (grey) knots up to 12, 13, 14, 15, 16, and then 17 crossings.

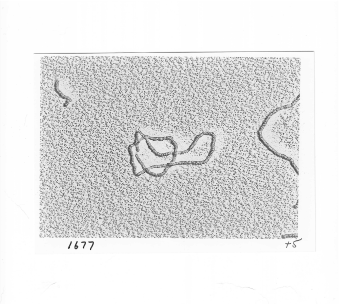


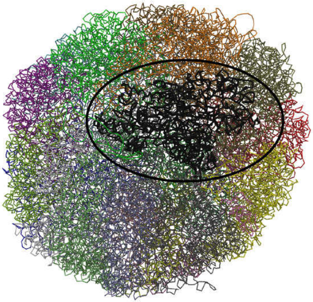




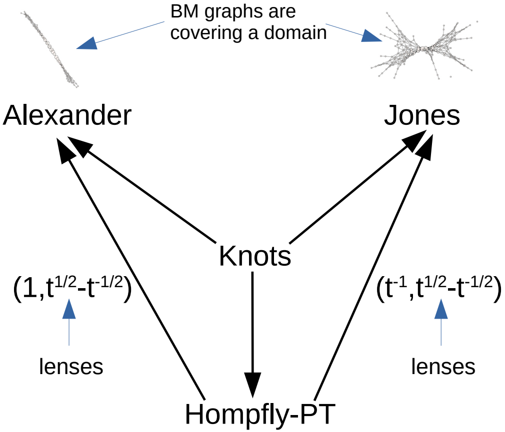
2D PCA projection of the coefficients of Jones polynomials for alternating knots with 15 crossing with determinant greater than 553.

Accuracy of the linear SVM classifier in predicting the knot's signature from its Jones polynomial as a function of the fraction of considered knots.



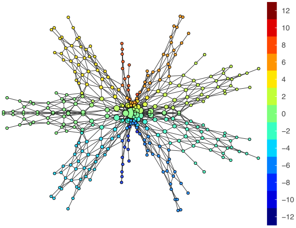


- Credits: Nancy Chrimson and Virnau et al.
- Input: Knot from an experiment
- Compute polynomial time invariants such as signature and determinant
- Find the region it belongs to in Jones for further relations
- Klotz, Andreson: writhe of the tight knot relates with signature and s-invariant

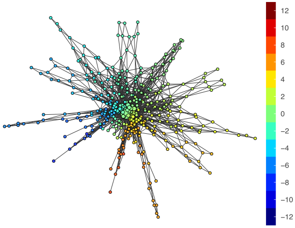


Point cloud for conventional Mapper construction

MAPPER ON BALL MAPPER: JONES VS. HOMFLYPT

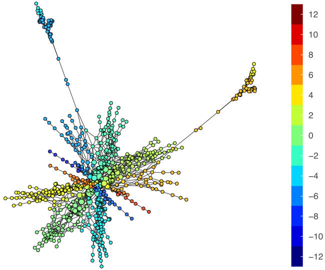


(A) Jones BM



(B) HOMFLY-PT BM

- Pullback from Jones to HOMFLYPT with the two long flares emerging indicating knots with the same s-invariant but not signature



- Statistical nature of the point clouds that arise from invariants should reveal structures and relations that are difficult to see using traditional theoretical and computational approaches.
- Which types of results can be (re)discovered, improved, or illuminated using ML, Ball Mapper, and other big data tools?
- Random data and choice of data specific embedding and metric
- Ball Mapper provides a set of TDA tools for exploring high dimensional data, visualizing scalar–valued functions, and comparing high dimensional data descriptors of a given dataset based on proximity and density (local to global).

QUESTIONS?