Topological Structures in the Analysis of Images and Data

Chao Chen

City University of New York (CUNY)

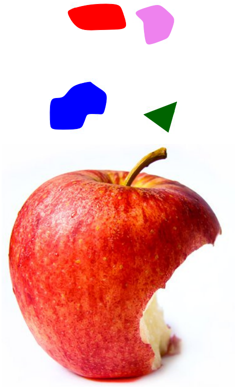
Oct. 2016

Outline

1. Topological Structures
2. High Dimensional Data
   - Algorithms
   - Applications

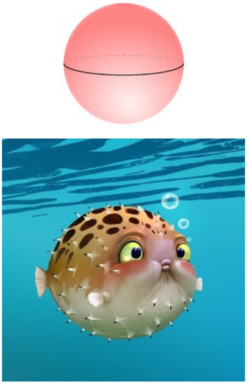
Topological Structures

- global, multi-scale, independent to geometry



0 dim

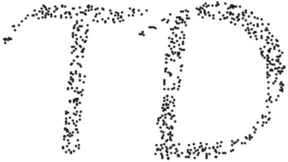
1 dim



2 dim

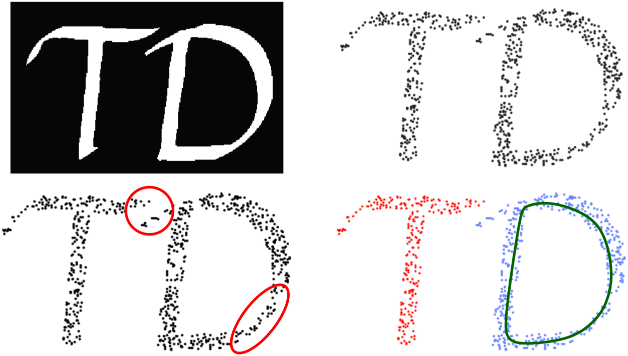
Topological Structures of Data

- For a dataset, what are the components and loops of the data?
- TDA: detect these structures in a robust way.

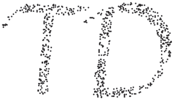


Topological Structures of Data

- For a dataset, what are the components and loops of the data?
- TDA: detect these structures in a robust way.

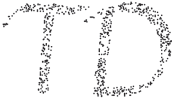


Persistent Homology: A Robust Way to Extract Topological Structures



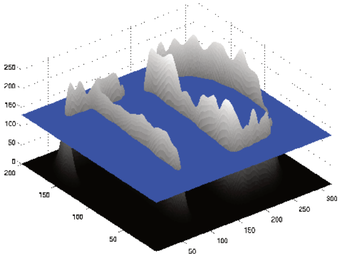
- Input: a (density) function, f
- Output:
  topological structures & their persistence

Persistent Homology: A Robust Way to Extract Topological Structures



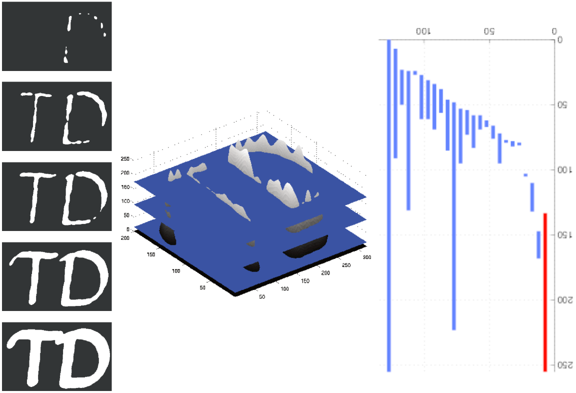
- Input: a (density) function, f
- Output:
  topological structures & their persistence

- Def: given threshold t, the superlevel set $f^{-1}[t, +\infty) := \{x | f(x) \geq t\}$



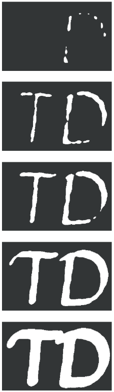
Persistent Homology (continued)

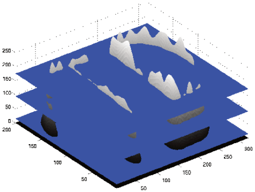
- the true structures are hidden in superlevel sets
- consider the whole stack of superlevel sets
- identify structures that often appear (high persistence)
- Output: persistence diagram – dots representing all structures



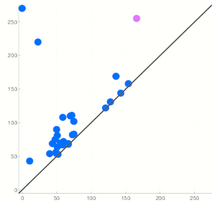
Persistent Homology (continued)

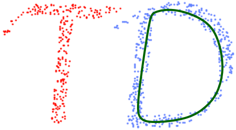
- the true structures are hidden in superlevel sets
- consider the whole stack of superlevel sets
- identify structures that often appear (high persistence)
- Output: persistence diagram – dots representing all structures



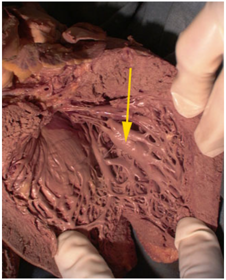


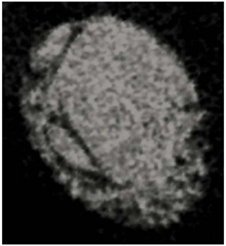
Diagram



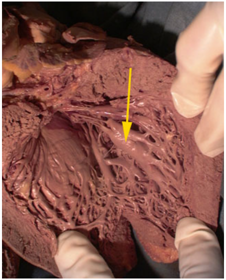


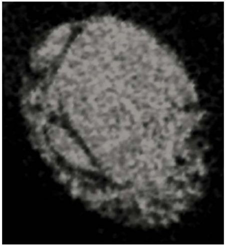
Why Topological Structures: Cardiac data (**Demo**)



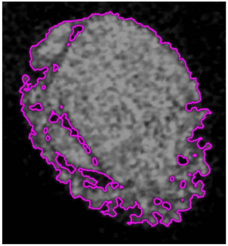


Why Topological Structures: Cardiac data (**Demo**)





Thresholding

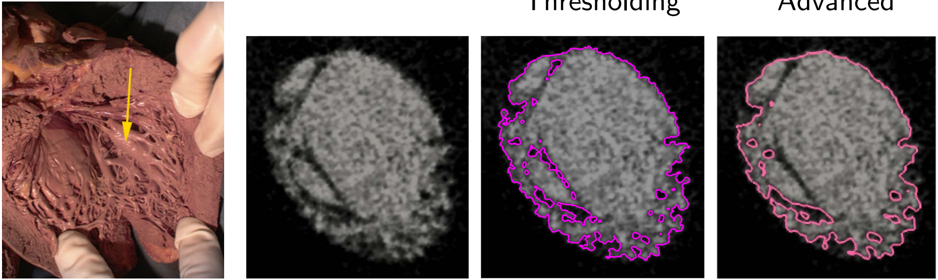


- Thresholding: local evidence, minimize energy $E(\boldsymbol{y})$

$$E(\boldsymbol{y}) = \sum_{v} E_v(y_v), \quad y_v \in \{BG, FG\}$$

Why Topological Structures: Cardiac data (**Demo**)

Thresholding Advanced



- Thresholding: local evidence, minimize energy $E(\boldsymbol{y})$

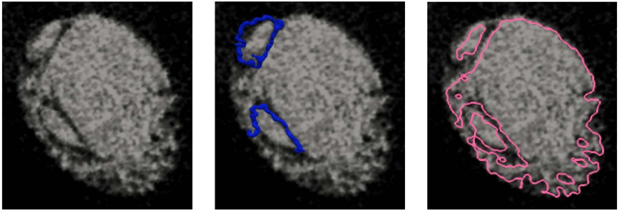
$$E(\boldsymbol{y}) = \sum_{v} E_v(y_v), \quad y_v \in \{BG, FG\}$$

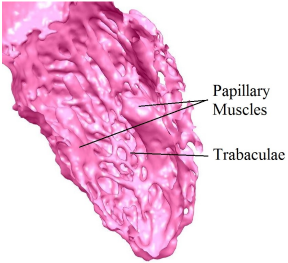
- Advanced: pairwise local evidence

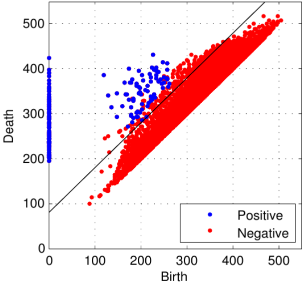
$$E(\boldsymbol{y}) = \sum_{v} E_v(y_v) + \sum_{(u,v)} E_{u,v}(y_u, y_v)$$

Why Topology Data Analysis?

Recovering missing trabeculae:



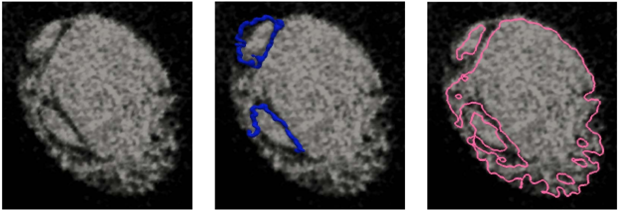


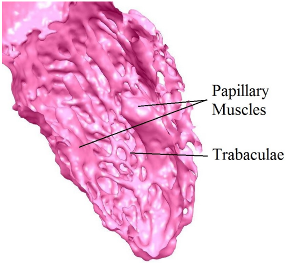


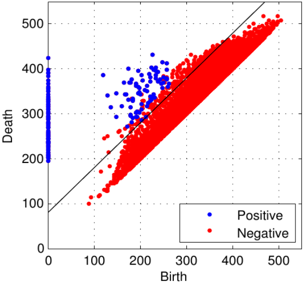
[Gao, **Chen** , *et al.* IPMI'13]

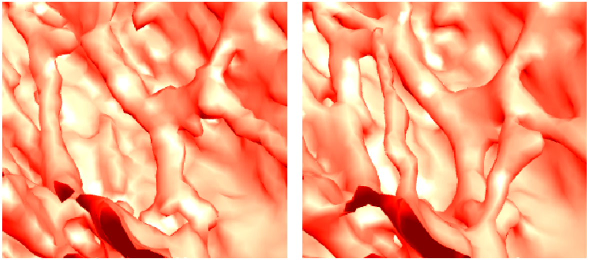
Why Topology Data Analysis?

Recovering missing trabeculae:





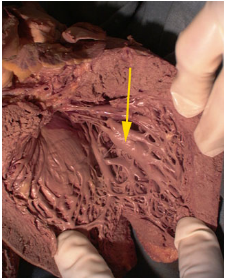


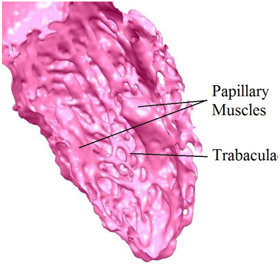


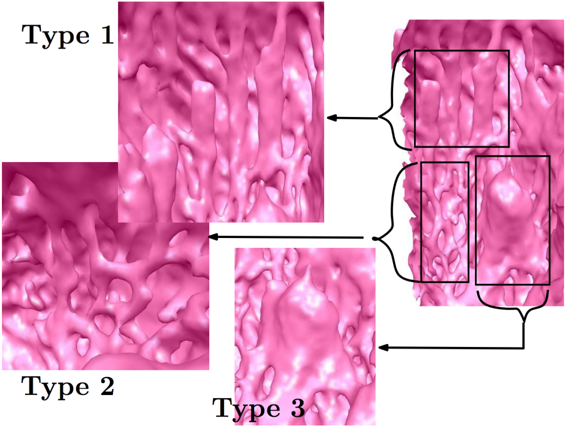
[Gao, **Chen** , *et al.* IPMI'13]

Morphological Analysis

Endocardial Surface [ISBI'14]

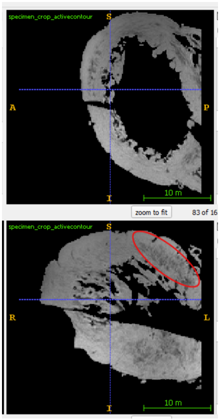




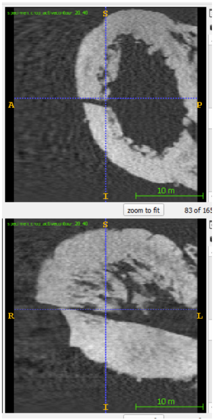


Follow-up Questions (Ongoing)

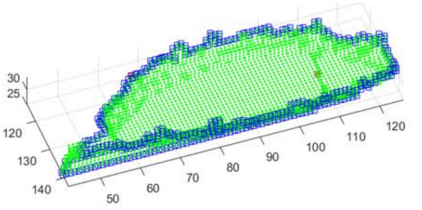
- Validation on a specimen
- Homology localization problem



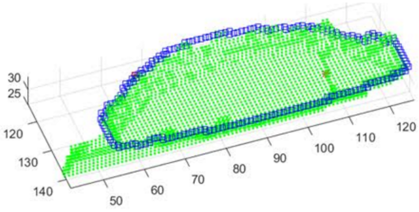
Ground Truth



Simulation

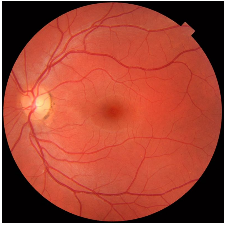
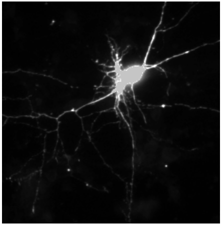


Bad Generator



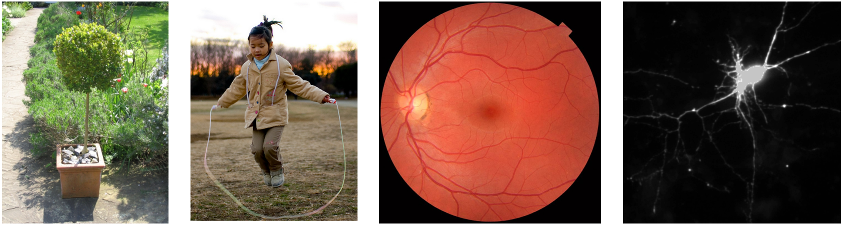
Good Generator

Topological Information as Constraints in Segmentation

[**Chen** *et al.* CVPR 2011]

Topological Information as Constraints in Segmentation



[**Chen** *et al.* CVPR 2011]

Input Stencil 1 2 3 4 Final

[Jain, **Chen** , *et al.* , Computer & Graphics, 2015]

Additional Application: Multi-Layer Stencil Creation

Canvas/wall result:

- Website, interactive

[Jain, **Chen** , *et al.* , Computer & Graphics, 2015]

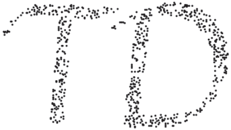
Outline

1. Topological Structures

2. High Dimensional Data
   - Algorithms
   - Applications

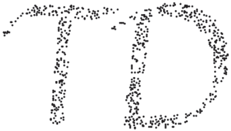
Topological Structures for High Dimensional Data

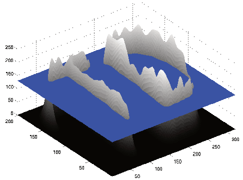
- Plenty have been done: data centric, simplicial complex, mapper, etc.



Topological Structures for High Dimensional Data

- Plenty have been done: data centric, simplicial complex, mapper, etc.

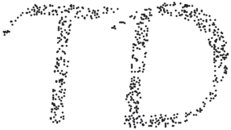


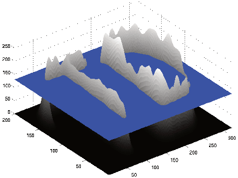


- My focus: density function.
  - Need a good model: high dim, flexibility, computation
    – graphical model

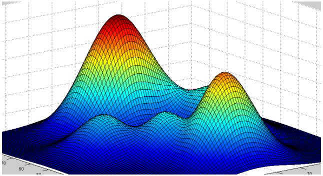
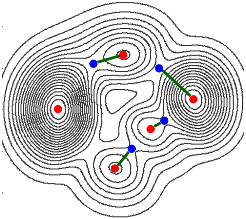
Topological Structures for High Dimensional Data

- Plenty have been done: data centric, simplicial complex, mapper, etc.





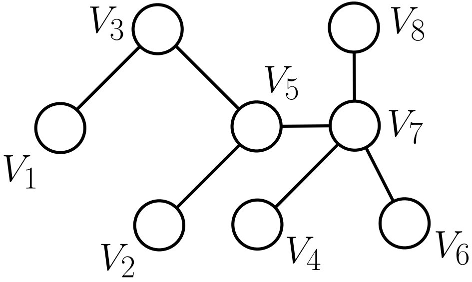
- My focus: density function.
  - Need a good model: high dim, flexibility, computation – graphical model
  - Locations that contribute to major topological events, critical points

Graphical Model

Markov Random Field (MRF)

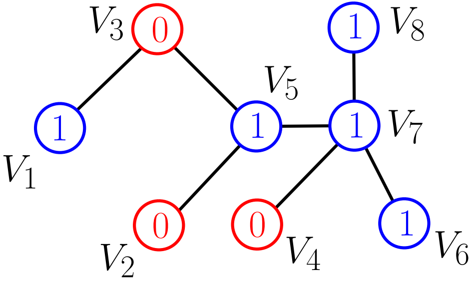
- D dimension; values/labels $\mathcal{L} = \{1, \ldots, L\}$
- configurations/labelings: $\mathcal{X} = \mathcal{L}^D = \{1, \cdots, L\}^D$



Graphical Model

Markov Random Field (MRF)

- D dimension; values/labels $\mathcal{L} = \{1, \ldots, L\}$
- configurations/labelings: $\mathcal{X} = \mathcal{L}^D = \{1, \cdots, L\}^D$



Binary Potentials $\theta_{ij}(x_i, x_j)$

x_i \ x_j	0	1
0	$\theta_{ij}(0,0)$	$\theta_{ij}(0,1)$
1	$\theta_{ij}(1,0)$	$\theta_{ij}(1,1)$

- Energy: $E(x) = \sum_{(i,j)\in\mathcal{E}} \theta_{ij}(x_i, x_j)$
- Probability: $P(x) = \exp(-E(x))/Z$

What can we do with a graphical model?

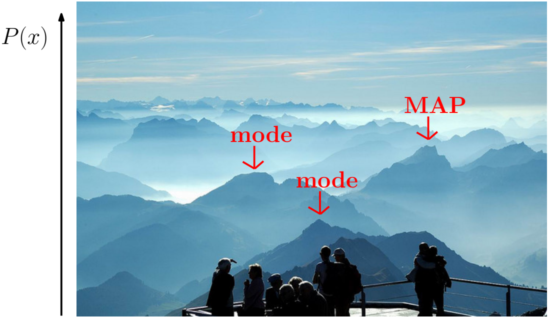
Previously:

- Computing the maximum a posteriori (MAP): $\text{argmax}_{x \in \mathcal{X}} P(x) = \text{argmin}_{x \in \mathcal{X}} E(x)$
- marginals, sampling, etc.

What can we do with a graphical model?

Previously:

- Computing the maximum a posteriori (MAP): $\text{argmax}_{x \in \mathcal{X}} P(x) = \text{argmin}_{x \in \mathcal{X}} E(x)$
- marginals, sampling, etc.

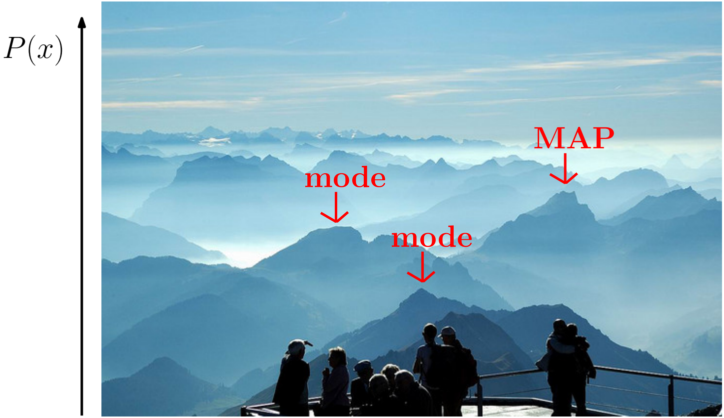


New Question:

- How about modes (local maxima)?

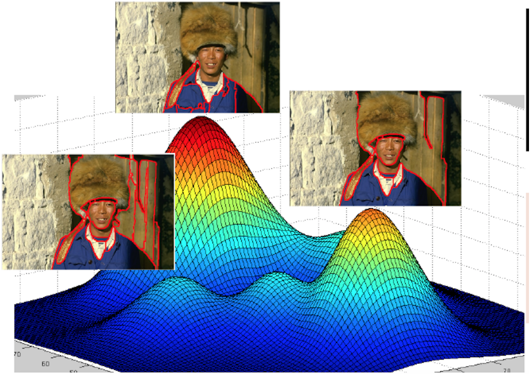
Why modes?

- A concise description of the probabilistic landscape



Why modes?

- A concise description of the probabilistic landscape
- Multiple predictions
  - model is not perfect, ambiguity
  - multiple hypotheses, diverse, highly possible



Other applications: biology, NLP

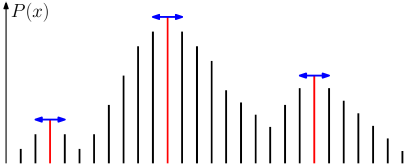
Previous: mean-shift

Definitions

- Given a distance function $d(\cdot, \cdot)$ and a scalar δ
  - Neighborhood: $N_\delta(x) = \{x' \mid d(x, x') \leq \delta\}$
  - x is a mode if it has a bigger prob. than all its neighbors
  - $\mathcal{M}^\delta$: the set of all modes for a given scale δ

Definitions

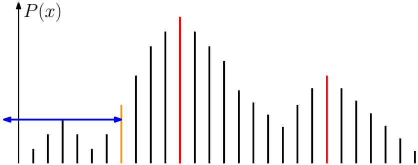
- Given a distance function $d(\cdot, \cdot)$ and a scalar δ
  - Neighborhood: $N_\delta(x) = \{x' \mid d(x, x') \leq \delta\}$
  - x is a mode if it has a bigger prob. than all its neighbors
  - $\mathcal{M}^\delta$: the set of all modes for a given scale δ



$\delta = 1$

Definitions

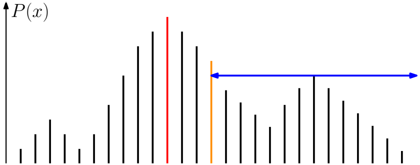
- Given a distance function $d(\cdot, \cdot)$ and a scalar δ
  - Neighborhood: $N_\delta(x) = \{x' \mid d(x, x') \le \delta\}$
  - x is a mode if it has a bigger prob. than all its neighbors
  - $\mathcal{M}^\delta$: the set of all modes for a given scale δ



$\delta = 4$

Definitions

- Given a distance function $d(\cdot,\cdot)$ and a scalar δ
  - Neighborhood: $N_\delta(x) = \{x' \mid d(x,x') \leq \delta\}$
  - x is a mode if it has a bigger prob. than all its neighbors
  - $\mathcal{M}^\delta$: the set of all modes for a given scale δ

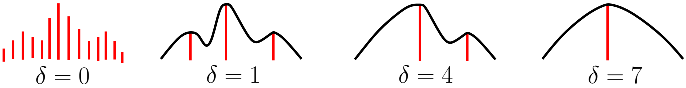


$\delta = 7$

Definitions

- D the dimension; $\mathcal{L} = \{1, \dots, L\}$ the label set; $\mathcal{X} = \mathcal{L}^D$ the domain
- Given a distance function $d(\cdot, \cdot)$ and a scalar δ
  - Neighborhood: $N_\delta(x) = \{x' \mid d(x, x') \leq \delta\}$
  - x is a mode if it has a bigger prob. than all its neighbors
  - $\mathcal{M}^\delta$: the set of all modes for a given scale δ

$$\mathcal{X} = \mathcal{M}^0 \supseteq \mathcal{M}^1 \supseteq \cdots \supseteq \mathcal{M}^\infty = \{\text{global maximum (MAP)}\}$$



Problem

Problem (MModes)

Given a scale δ, compute the top M elements in $\mathcal{M}^{\delta}$.

Challenge: exponential domain, exponential neighborhood

Problem

Problem (MModes)

Given a scale δ, compute the top M elements in $\mathcal{M}^\delta$.

Challenge: exponential domain, exponential neighborhood Contributions

- Algorithms (chains, trees):
  - Dynamic programming (DP)
  - Heuristic search
  - Local neighborhood search
- Applications

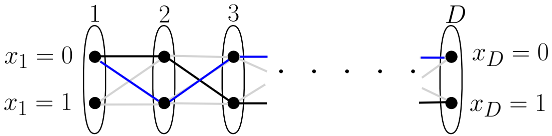
[AISTATS 2013, NIPS 2014, IJCAI 2016, ICML 2016]

Outline

1 Topological Structures

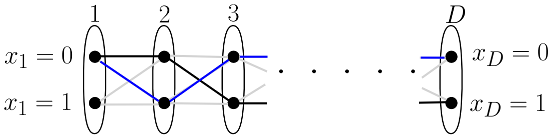
2 High Dimensional Data
- Algorithms
- Applications

Algorithm: Chains



- configurations/labelings = paths
- MAP: the optimal path (dynamic programming), $O(DL^2)$
  - from right to left
  - each step: best energy for subchain $[i, D]$ with given label on i

Algorithm: Chains



- configurations/labelings = paths
- MAP: the optimal path (dynamic programming), $O(DL^2)$
  - from right to left
  - each step: best energy for subchain $[i, D]$ with given label on i
- MBest: best M configurations/labelings

$$x^1 = \text{argmin}_{x \in \mathcal{X}} E(x)$$
$$x^m = \text{argmin}_{x \in \mathcal{X} \setminus \{x^1, \cdots, x^{m-1}\}} E(x)$$

Nilsson'98 (fancy DP)
$O(DL^2 + MDL + MD\log(MD))$

Algorithm: Modes on Chains [AISTATS 2013]

Key Idea

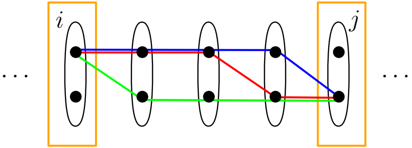
- The whole chain $[1, D] \rightarrow$ subchains $[i, j]$ of a fixed length
- Global modes $\rightarrow$ local modes

Algorithm: Modes on Chains [AISTATS 2013]

Key Idea

- The whole chain $[1, D] \to$ subchains $[i, j]$ of a fixed length
- Global modes $\to$ local modes

- A partial labeling $x_{i:j}$

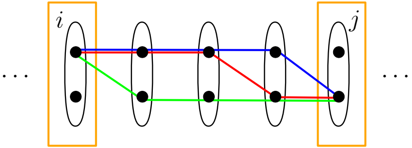


- $x_{i:j}$ is a local mode iff for any $y_{i:j}$ s.t. $y_i = x_i$, $y_j = x_j$ $E(x_{i:j}) < E(y_{i:j})$

Algorithm: Modes on Chains [AISTATS 2013]

Key Idea

- The whole chain $[1, D] \rightarrow$ subchains $[i, j]$ of a fixed length
- Global modes $\rightarrow$ local modes
- A partial labeling $x_{i:j}$



- $x_{i:j}$ is a local mode iff for any $y_{i:j}$ s.t. $y_i = x_i$, $y_j = x_j$ $E(x_{i:j}) < E(y_{i:j})$

Lemma

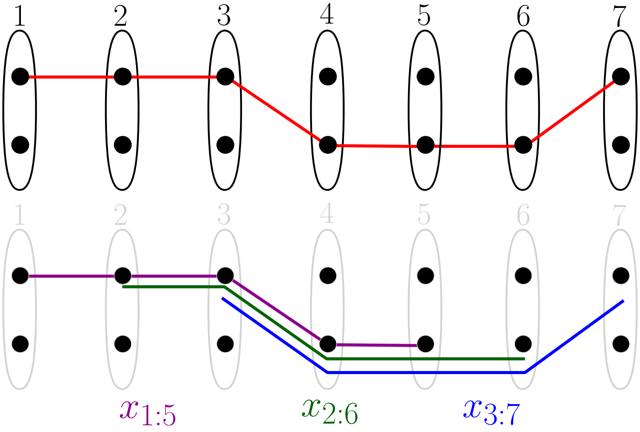
any $[i, j]$ has L^2 local modes, computable in polynomial time

Algorithm: Modes on Chains

Theorem (local-global)

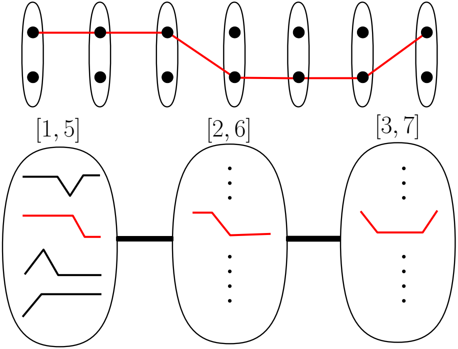
x is a mode iff any length $\delta + 2$ partial labeling $x_{i:j}$ is a local mode

An example: $D = 7$, $\delta = 3$



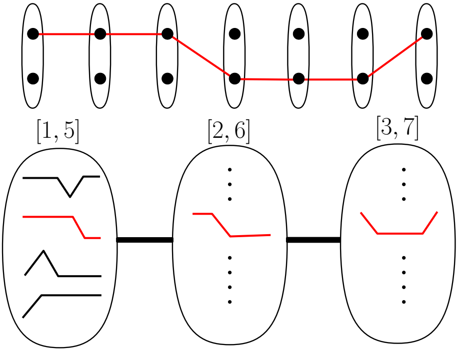
Algorithm: Modes on Chains

- Intuition
  - Combinations of local modes → global modes
  - Consistent: agree at common vertices



Algorithm: Modes on Chains

- Intuition
  - Combinations of local modes $\rightarrow$ global modes
  - Consistent: agree at common vertices

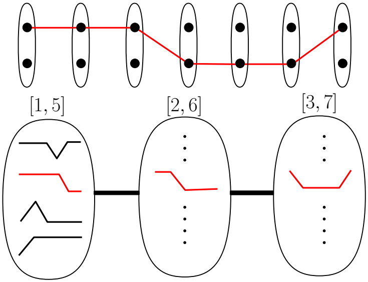


- Step 1: construct a new chain,
  - supernodes $[i, j]$
  - labels {local modes of $[i, j]$}
  - feasible only if consistent
  - preserve the energy of the original graph

Fact

New chain labeling space: $\widehat{\mathcal{X}} = \mathcal{M}^{\delta}$

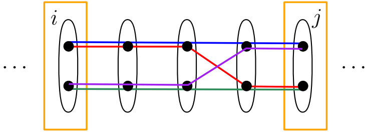
Algorithm: Modes on Chains



- Step 1: construct a new chain,
  - Configuration space: $\widehat{\mathcal{X}} = \mathcal{M}^{\delta}$
  - Energy: $\widehat{E}(\widehat{x}) = E(x)$
- Step 2: M-Modes is reduced to M-Best in the new chain
  - M-Best: compute the top M configurations
  - Use Nilsson'98
- Total Complexity $O(DL^3\delta + MDL^2 + MD\log(MD))$

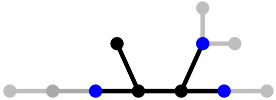
Trees

Chains



- Subchain of δ plus two adjacent nodes
- Local modes (L^2)

Trees



- Subtree of size δ plus all adjacent nodes
- Local modes (exponential to the number of adjacent nodes)

Theorem (local-global)

x is a mode iff within any subchain/subtree it is a local mode.

Can extend to any graph!

General Situations

Extending the Algorithm

- Trees (DP) [NIPS'14]
- Systematic search [IJCAI'16]
- Local neighborhood search [ICML'16]

General Situations

Extending the Algorithm

- Trees (DP) [NIPS'14]
- Systematic search [IJCAI'16]
- Local neighborhood search [ICML'16]

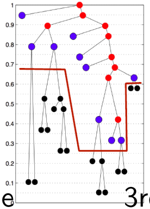
Model Unknown

- Input: samples
- Algorithm:
  - Step 1: estimate a tree distribution (Chow-Liu algorithm)
  - Step 2: compute modes
- Theoretical guarantee $P(\widehat{\mathcal{M}^\delta} = \mathcal{M}^\delta) \to 1$ as $S \to \infty$

Outline

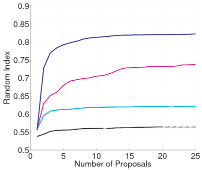
1. Topological Structures
2. High Dimensional Data
   - Algorithms
   - Applications

Application: Multiple Predictions

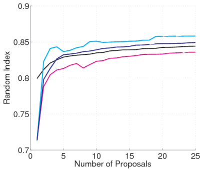


- High Probability; Diversity
- Image Partitioning Task (Berkeley, Stanford Datasets)

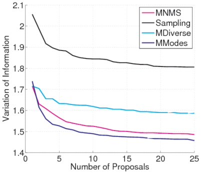
Ground Truth | 1st Mode | 2nd Mode | 3rd Mode



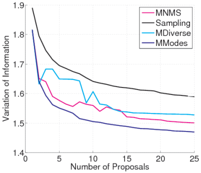
Standford RI



Berkeley RI



Standford VOI



Berkeley VOI

Application: Video Analysis

Gesture recognition:

[**Chen** *et al.* AISTATS]

Pic from [Liu, **Chen** *et al.* CVIU]

Clustering Discrete Data [ICML 2016]

- Start from each data, local search until stops at a mode

Synthetic data: $D = 110$, $L = 4$, 4 clusters
randomly perturb 5% and 10% attributes

- Visualized in 2D (using MDS)

GT/Ours ROCK AP kmodes

- Performance (in NMI)

Synthetic Data									
	K-Means	DPGMM	AP	SC	MPD	TMode	K-Modes	ROCK	Ours
5% Corrupted	0.75	0.75	0.73	0.08	0.72	0.63	0.74	0.47	**1.00**
10% Corrupted	0.75	0.74	0.72	0.05	0.70	0.63	0.74	0.47	**0.90**

Clustering Discrete Data [ICML 2016]

- DNA Barcoding data ([Kuksa & Pavlovic BMC Bioinformatics])
- 600 to 900 dimension
- Alignment free

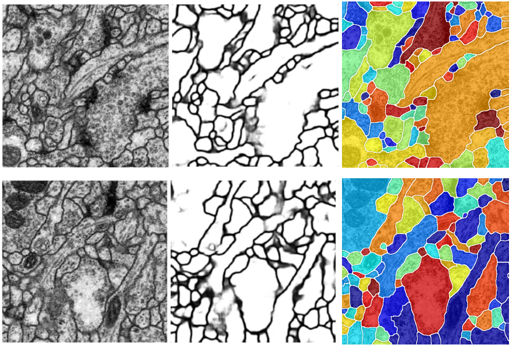
DNA Barcoding									
	K-Means	DPGMM	AP	SC	MPD	TMode	K-Modes	ROCK	Ours
ACG G	0.60	0.49	0.53	0.42	0.63	0.42	0.75	0.76	**0.79**
ACG S	0.80	0.50	0.61	0.62	0.84	0.49	0.86	0.88	**0.89**
Bats G	0.81	0.79	**0.82**	0.39	0.84	0.49	**0.82**	0.72	**0.82**
Bats S	0.91	0.79	0.89	0.48	**0.92**	0.79	0.87	0.80	0.89
Birds G	0.61	0.35	0.48	0.40	0.78	0.16	0.80	**0.82**	**0.82**
Birds S	0.79	0.45	0.58	0.56	0.82	0.19	0.88	**0.90**	0.89
Fish G	0.88	0.44	0.89	0.59	0.84	0.77	**0.90**	0.84	0.88
Fish S	**0.94**	0.44	0.75	0.89	0.91	0.81	0.91	0.92	**0.94**
Hesp. G	0.61	0.43	0.45	0.47	0.70	0.13	0.75	0.78	**0.81**
Hesp. S	0.80	0.48	0.57	0.61	0.87	0.15	0.87	0.89	**0.90**

Average Running Time (seconds)									
	K-Means	DPGMM	AP	SC	MPD	TMode	K-Modes	ROCK	Ours
	30.2	138.9	112.4	1689.7	1872.5	829.5	473.9	2013.0	540.6

Also UCI datasets.

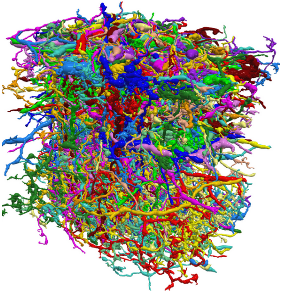
Application: User Interaction (Ongoing)

- Electron Microscopy (EM) Images of Fly/Mouse Brains
- Input: 2D or 3D EM images; boundary likelihood map
- Output: partitioning of the image



EM Images Likelihood Results

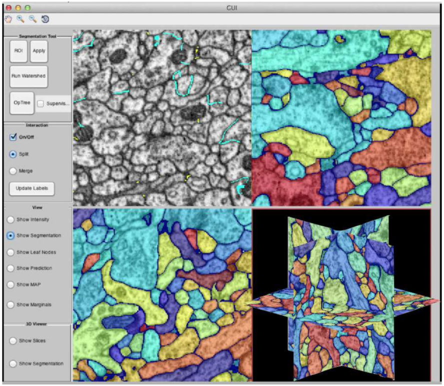
Pic from Takemura *et al.* Nature'13

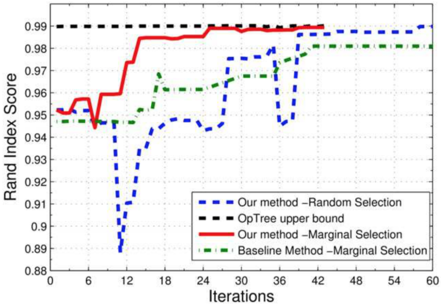


[Uzunbaș, **Chen** and Metaxas, MICCAI'14, MedIA'15]

Application: User Interaction (Ongoing)

- Multiple proposals for user to select and modify





Conclusion

Topological structures: global structure/prior/information

- Individual data/images
- Whole dataset
  - New perspective to the model: inference and more

Thank You! Questions?

Appendix

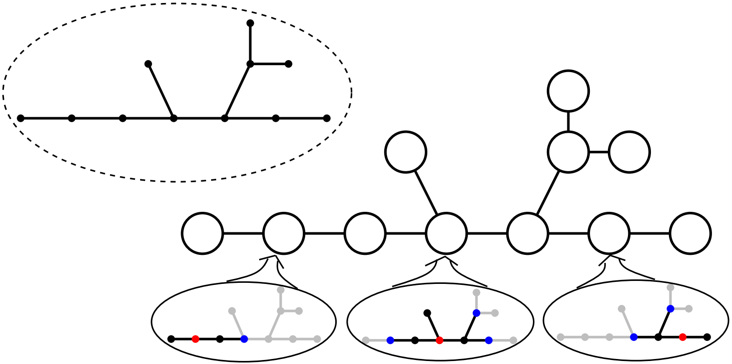
- Convergence rate for modes estimation

$$\mathbb{P}(\widehat{\mathcal{M}}^\delta = \mathcal{M}^\delta) \geq 1 - L^2 d(d-1)\exp\{-n\Delta_{\min}^2/(18L^4c_0^2)\} - 3(d-1)L^2\exp\{-n\Delta_{\mathcal{M}}^2/(18d^2c_1^2)\}.$$

d dimension, *L* label set size, *n* sample size

Trees: Idea 1, Fancy DP [NIPS 2014]

- Build a new tree
  - Supernodes ← subtrees
  - Labels ← local modes
  - M-Best configurations ← M-Modes
- Issue: number of local modes can be exponential to the tree-degree, even for small δ



Trees: Idea 1, Fancy DP [NIPS 2014]

Complexity

$$O\left(D^2 dL\delta^2(L+\delta)(L^d+\lambda^d)+D\lambda^2+MD\lambda+MD\log(MD)\right)$$

- d tree degree
- λ max # of local modes for any ball

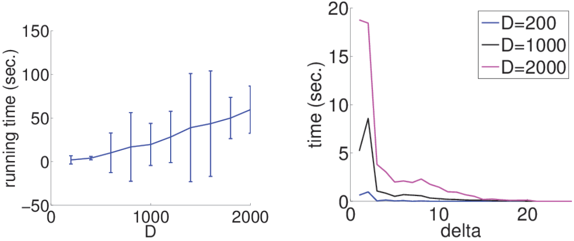
Trees: Idea 1, Fancy DP [NIPS 2014]

Complexity

$$O\left(D^2 dL\delta^2(L+\delta)(L^d+\lambda^d)+D\lambda^2+MD\lambda+MD\log(MD)\right)$$

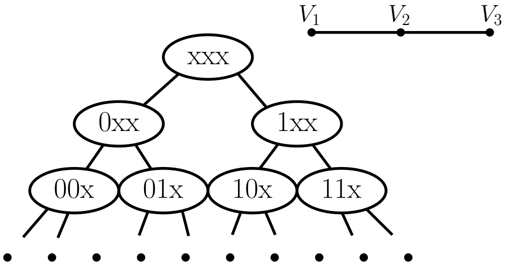
- d tree degree
- λ max # of local modes for any ball

In practice (bounded tree degree)



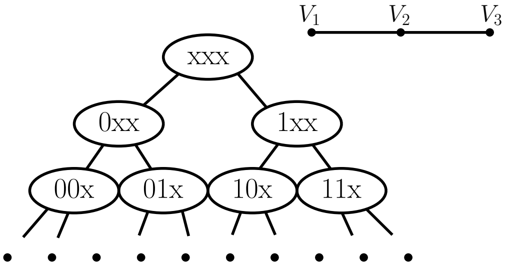
Trees: Idea 2, Heuristic Search [IJCAI 2016]

- Compute all local modes → only compute when necessary
- Heuristic search:
  - For each state, verify whether one local pattern is a local mode
    - ⋆ if not, prune the whole subtree
  - Many states (and thus local modes) may never be reached
  - A*, Death First Search Branch and Bound (DFBnB)



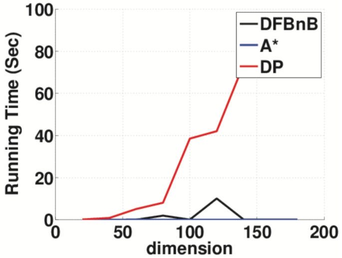
Trees: Idea 2, Heuristic Search [IJCAI 2016]

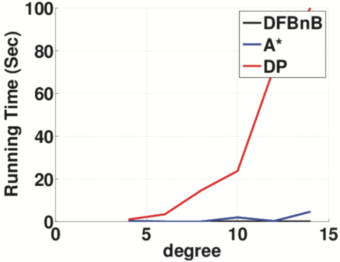
- Compute all local modes → only compute when necessary
- Heuristic search:
  - For each state, verify whether one local pattern is a local mode
    - ⋆ if not, prune the whole subtree
  - Many states (and thus local modes) may never be reached
  - A*, Death First Search Branch and Bound (DFBnB)



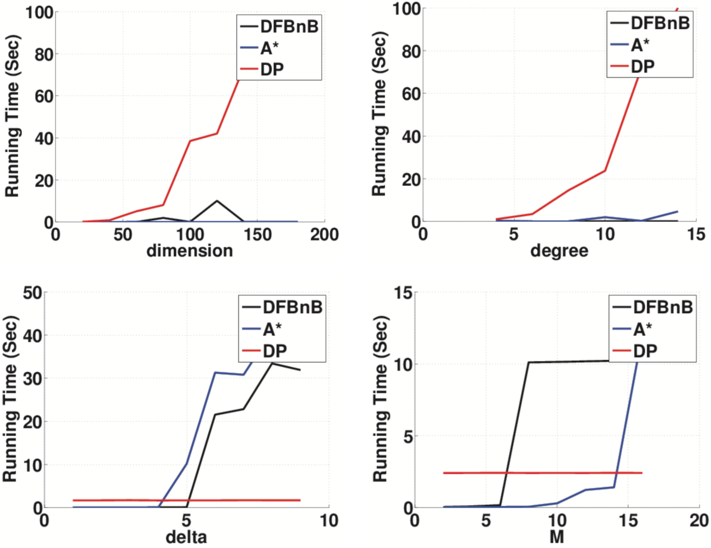
- Caveat:
  - Not any cheaper in the worst case senario
  - Needs the MAP computation

Trees: Idea 2, Heuristic Search [IJCAI 2016]



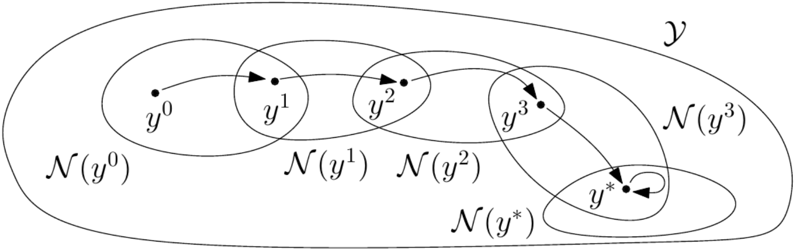


Trees: Idea 2, Heuristic Search [IJCAI 2016]



Also UCI datasets.

Trees: Idea 3, Local Search



- $\mathcal{N}_t : \mathcal{Y} \to 2^{\mathcal{Y}}$, neighborhood system
- Optimization with respect to $\mathcal{N}_t(y)$ must be tractable:

$$y^{t+1} = \operatorname*{argmax}_{y \in \mathcal{N}_t(y^t)} g(x, y)$$

Pic from Nowozin and Lampert

Trees: Idea 3, Local Search

Each step: to compute the best neighbor in $\mathcal{N}_\delta(y)$,

$$\mathrm{argmin}_{z \in \mathcal{N}_\delta(y) \setminus \{y\}} E(z)$$

- Complexity $O(DdL\delta^2(L+\delta))$

$$\mu_{i\to j}(\ell_i, \tau) = \min_{z_{T_i}: z_i = \ell_i, \rho(z_{T_i}, y) \le \tau} f(z_{T_i}),$$