Transitive Distance Clustering: Theories, Algorithms and Applications

Zhiding Yu

Department of Electrical and Computer Eng.
Carnegie Mellon University

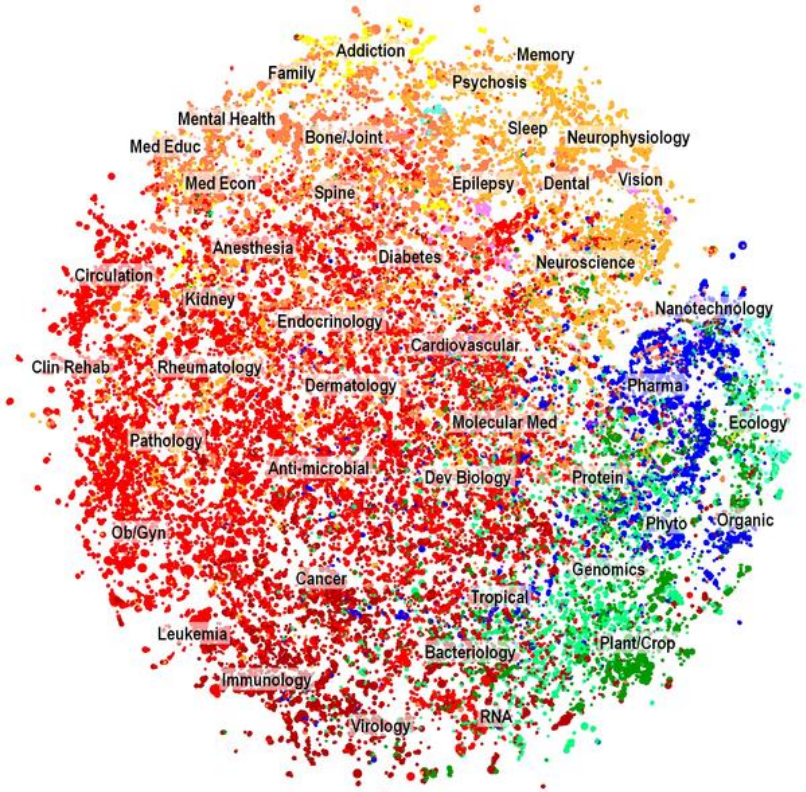
Background

Alyosha Efros tells us the revolution will not be supervised at the ICCV Workshop on Object Understanding from Interactions.

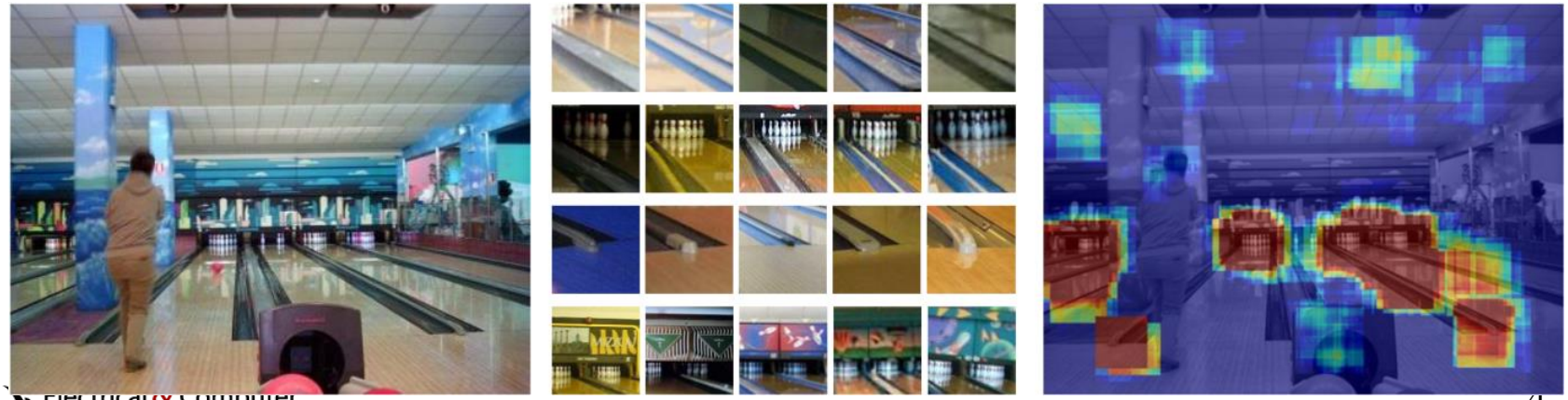
I agree. — Yann LeCun

Wide Applications

Image Segmentation

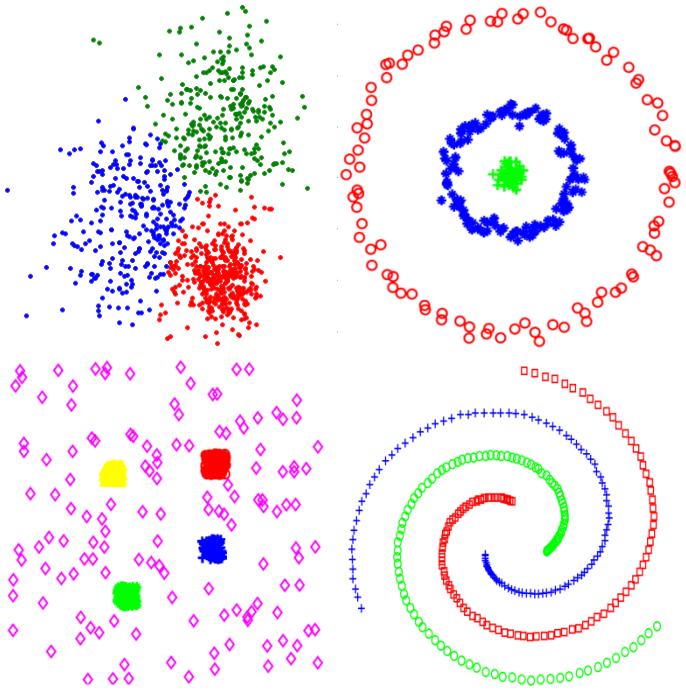


Document & Text Analysis



Mid-level Discriminative Visual Element Discovery

Key Problem Issues



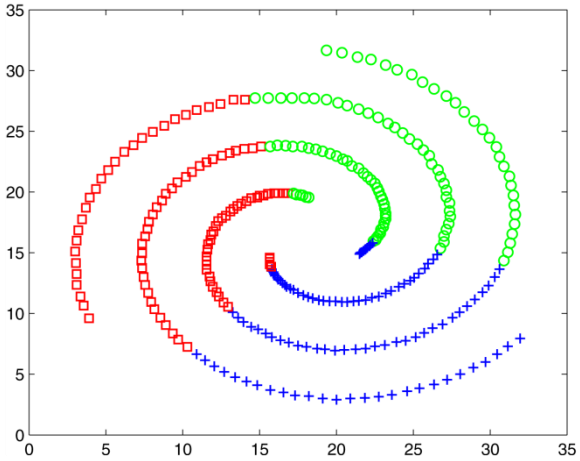
Important Issues:

- Maximally reveal intra-cluster similarity
- Maximally reveal inter-cluster dis-similarity
- Discover clusters with **non-convex shape**
- Consider cluster assumptions & priors
- Robustness

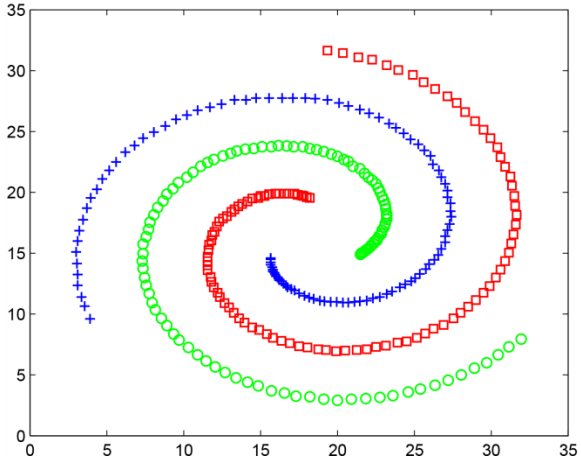
Existing Methods & Literatures

Early Methods	
Centroid-Based	**K-Means** (Lloyd 1982); **Fuzzy Methods** (Bezdek 1981)
Connectivity-Based	**Hierarchical Clustering** (Sibson 1973; Defays 1977)
Distribution-Based	**Mixture Models + EM**
More Recent Developments	
Density-Based	**Mean Shift** (Cheng 1995; Comaniciu and Meer 2002)
Spectral-Based	**Spectral Clustering** (Ng et al. 2002); **Self-Tuning SC** (Zelnik-Manor and Perona 2004); **Normalized Cuts** (Shi and Malik 2000);
Transitive Distance (Path-Based)	**Path-Based Clustering** (Fischer and Buhmann 2003b); **Connectivity Kernel** (Fischer, Roth, and Buhmann 2004); **Transitive Dist Closure** (Ding et al. 2006); **Transitive Affinity** (Chang and Yeung 2005; 2008)
Subspace Clustering	**SSC** (Elhamifar and Vidal 2009); **LSR** (Lu et al. 2012); **LRR** (Liu et al. 2013); **L1-Graph** (Cheng et al., 2010); **L2-Graph** (Peng et al, 2015); **L0-Graph** (Yang et al, 2015); **SMR** (Hu et al., 2014);

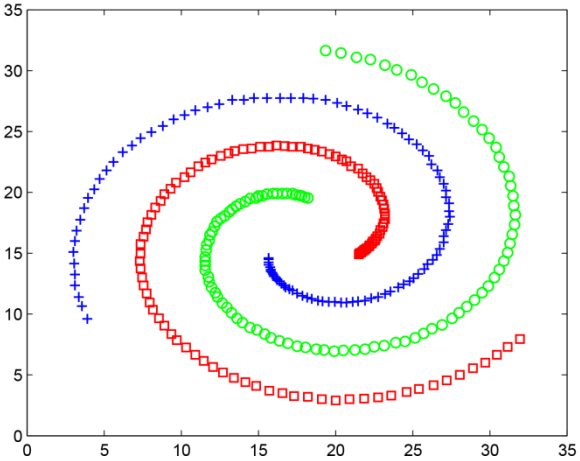
Addressing Non-Convex Clusters



K-means



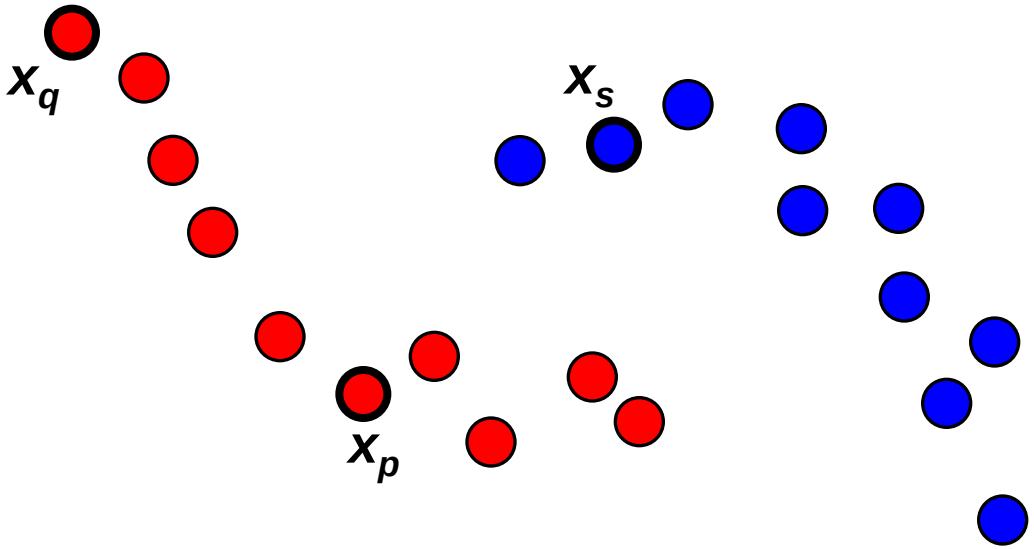
Spectral Clustering



Transitive Distance (Path-based) Clustering

Transitive Dist. (TD) Clustering with K-Means Duality (CVPR14)

Transitive Distance: Concept

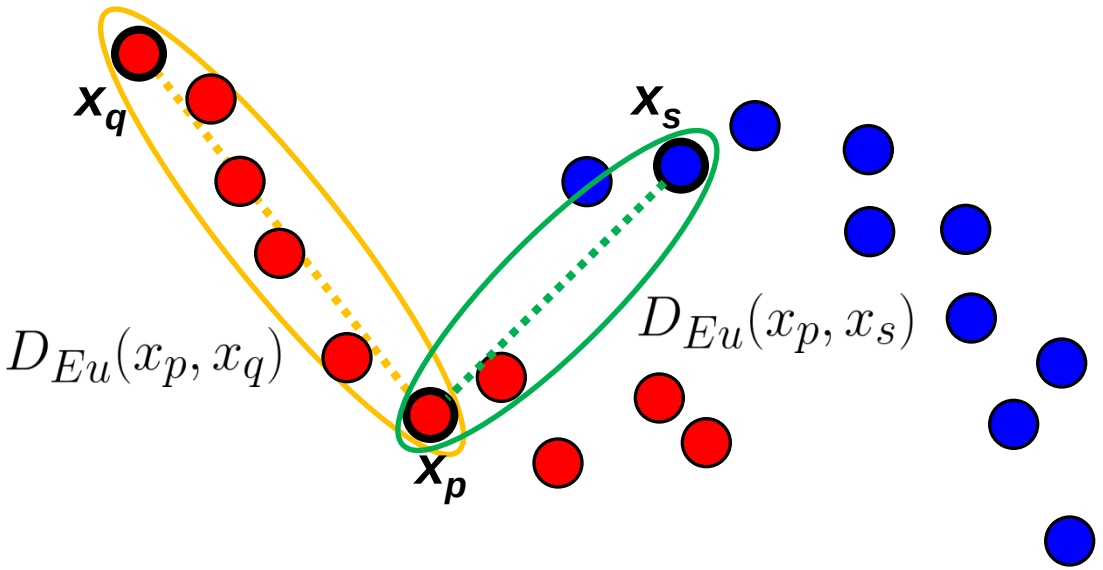


Ideally, we want:

$$D(x_p, x_q) < D(x_p, x_s)$$

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Transitive Distance: Concept

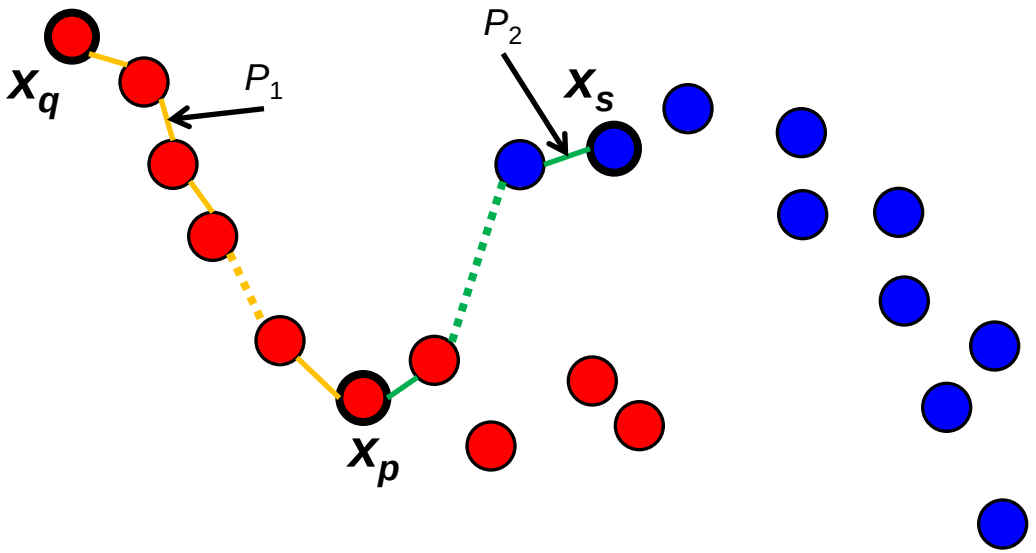


Euclidean Distance:

$$D_{Eu}(x_p, x_q) > D_{Eu}(x_p, x_s)$$

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

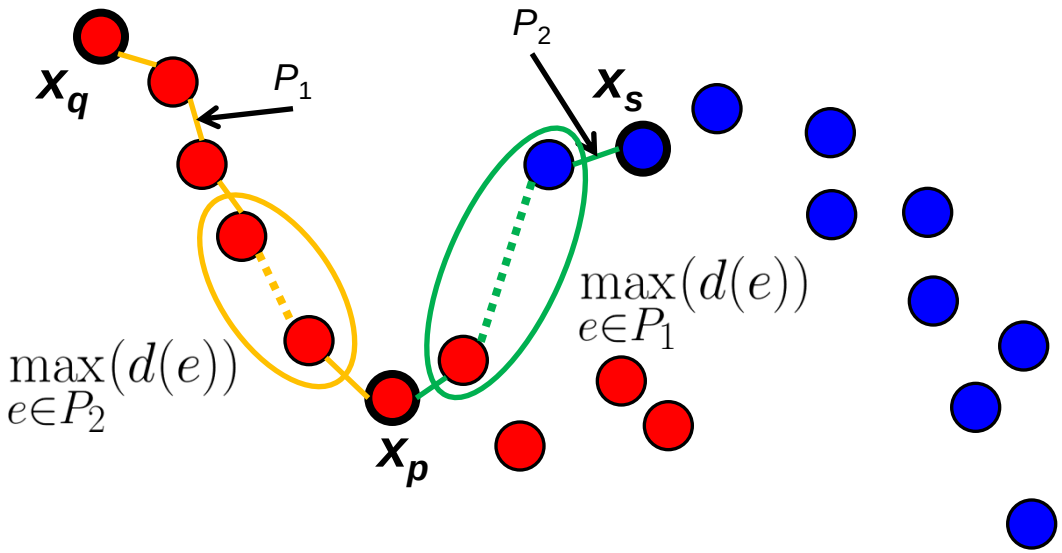
Transitive Distance: Concept



Intuition: Far away points can belong to the same class, because there is strong evidence of a path connecting them

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

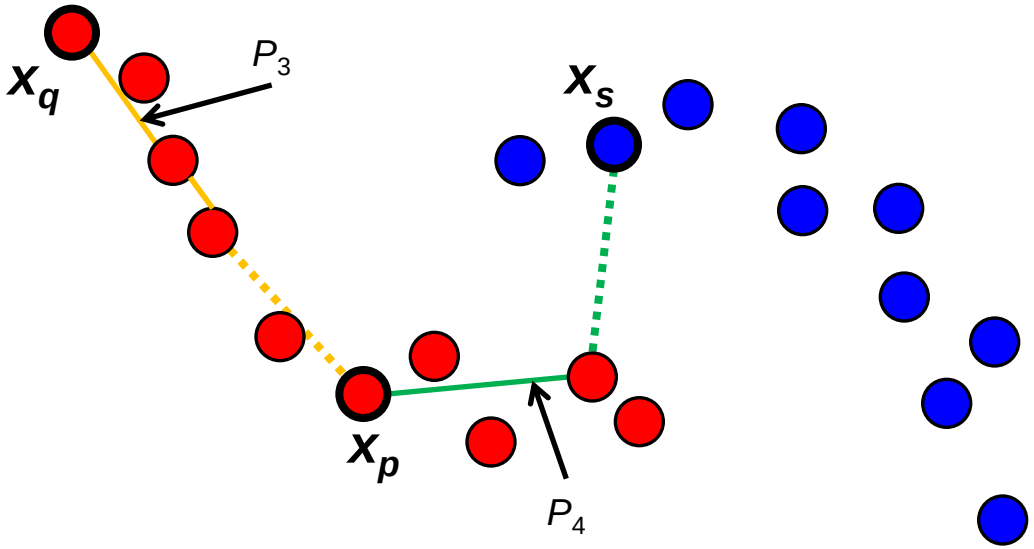
Transitive Distance: Concept



The size of the maximum gap on the path decides how strong the path evidence is. It is therefore a better measure of point distances than Euclidean distance

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

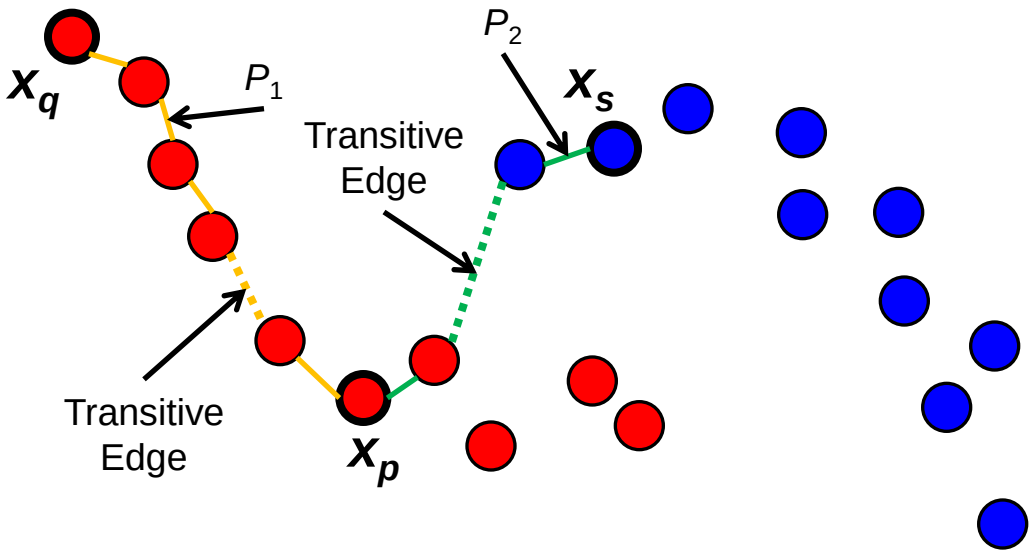
Transitive Distance: Concept



But there could exist many other path combinations…

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

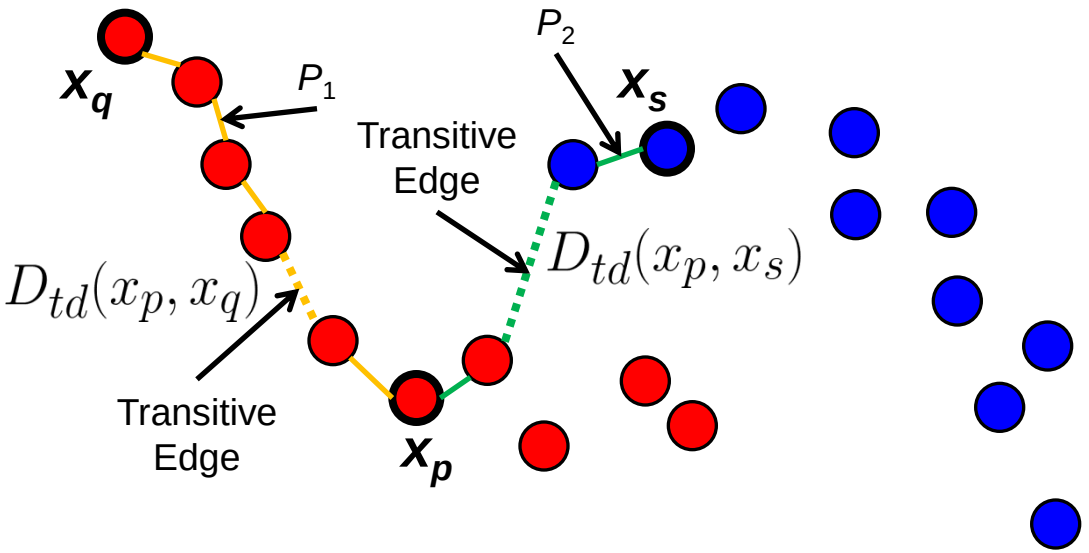
Transitive Distance: Concept



Just select the path with the minimum max gap from all possible paths. The max gaps on the selected path are called transitive edges and defines the final distance

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Transitive Distance: Concept



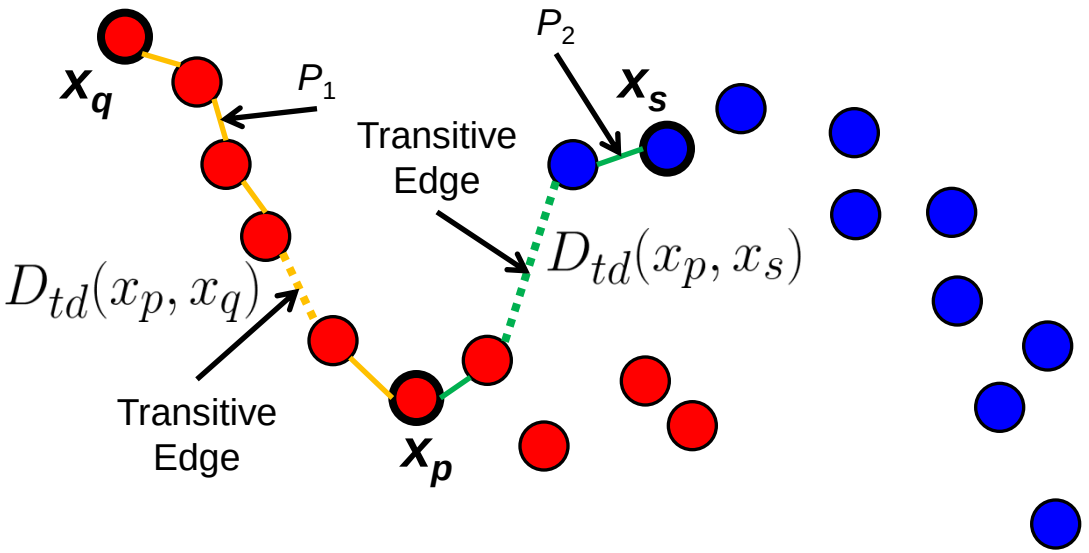
Transitive Distance:

$$D_{td}(x_p, x_q) < D_{td}(x_p, x_s)$$

Transitive Distance: $D_{td}(x_p, x_q) = \min_{P \in \mathbf{P}} \max_{e \in P}(d(e))$

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Transitive Distance: Concept



Transitive Distance:

$$D_{td}(x_p, x_q) < D_{td}(x_p, x_s)$$

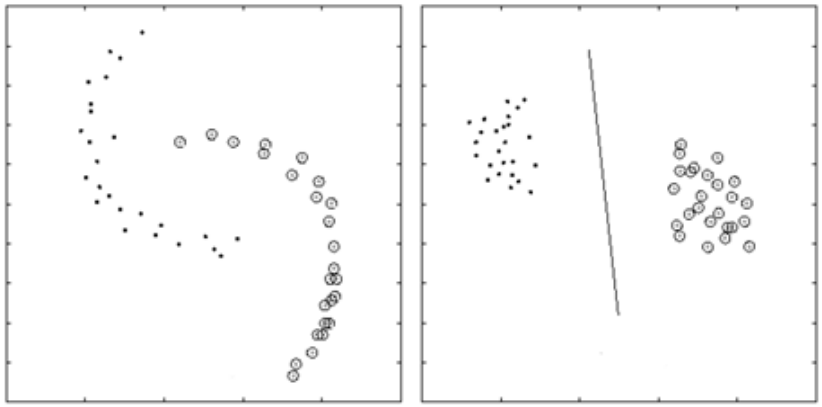
Transitive Distance: $D_{td}(x_p, x_q) = \min_{P \in \mathbf{P}} \max_{e \in P}(d(e))$

Theorem 1:

Given a weighted graph with edge weights, each transitive edge lies on the ***minimum spanning tree (MST)****.*

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Transitive Distance Embedding



Original Space Projected Space

Lemma 1:

*The Transitive Distance is an **ultrametric** (metric with strong triangle property).*

Lemma 2:

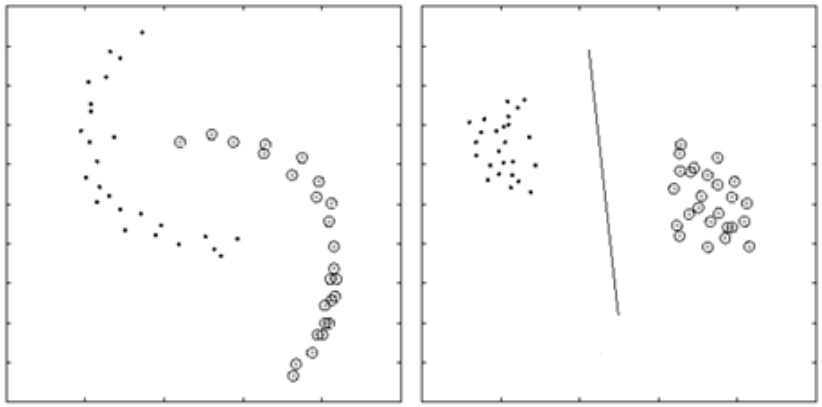
Every finite ultrametric space with n distinct points can be embedded into an n−1 dim Euclidean space.

Theorem 2:

If a labeling scheme of a dataset is consistent with the original distance, then given the derived transitive distance, the convex hulls of the projected images in the TD embedded space do not intersect with each other.

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Transitive Distance Embedding



Original Space Projected Space

Lemma 1:

*The Transitive Distance is an **ultrametric** (metric with strong triangle property).*

Lemma 2:

Every finite ultrametric space with n distinct points can be embedded into an n−1 dim Euclidean space.

Remarks:

- TD can be **embedded** into an Euclidean space.
- Intuitively, for manifold or path cluster structures, TD drags far away intra-cluster data to be closer. The projected data show nice and compact clusters.
- It is very desirable to perform k-means clustering in the embedded space.
- Here, TD is doing a similar job as **spectral embedding**.

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

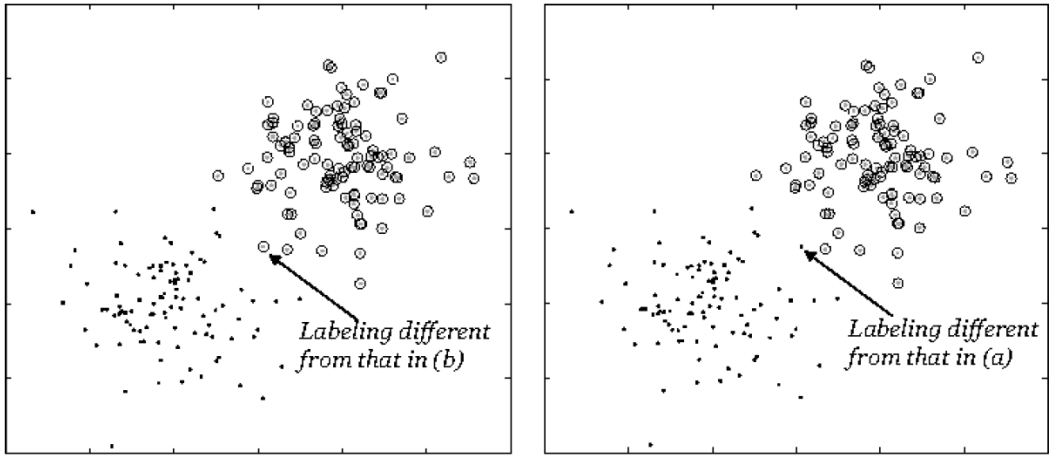
K-Means Duality

Denote: V the set of data. **E** the corresponding Euclidean dist matrix of **V**.

$$\mathbf{E} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}$$

Property: (K-Means Duality)

The k-means clustering result on the rows of **E** (treating each row of **E** like data) *is very similar to the result of k-means directly on* **V**.



K-means on V **K-means on rows of E**

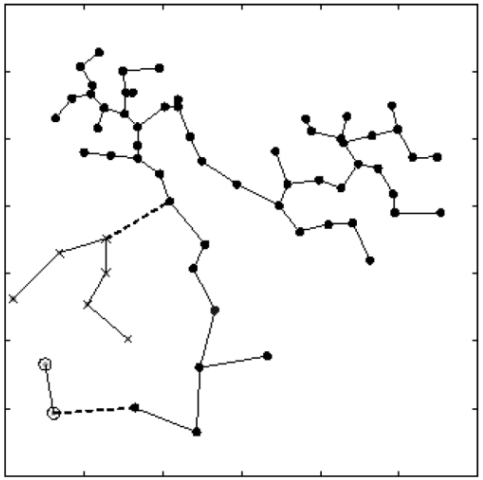
Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Clustering with K-Means Duality

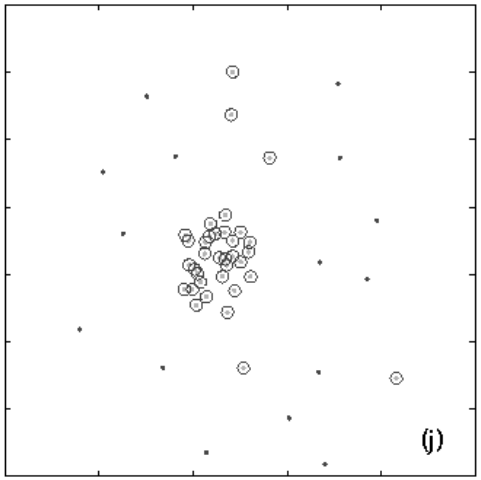
- Given a set of data, construct a weighted complete graph.
- Extract an MST from the graph.
- Compute the transitive distance between pair-wise data by referring to the path edge with largest weight.
- Perform k-means on the rows of transitive distance matrix.

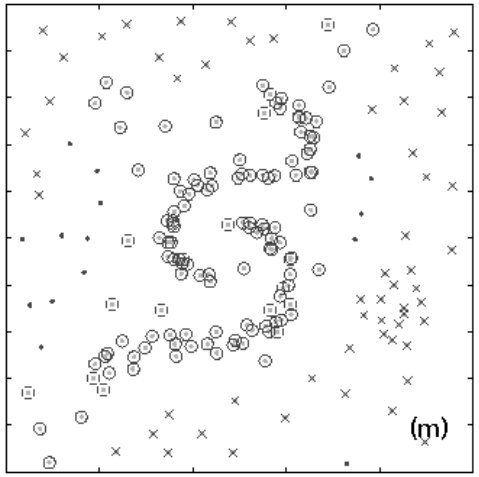
Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Experiment: Synthetic Data

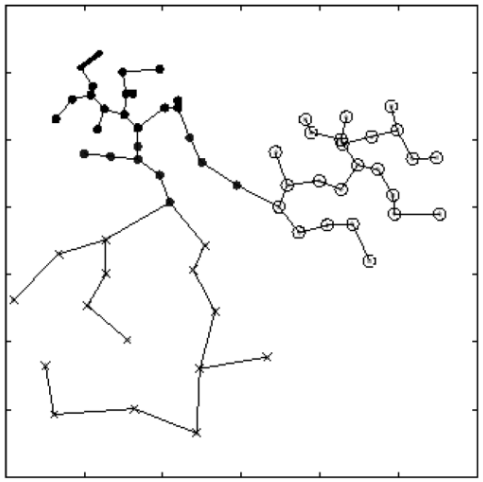


SL

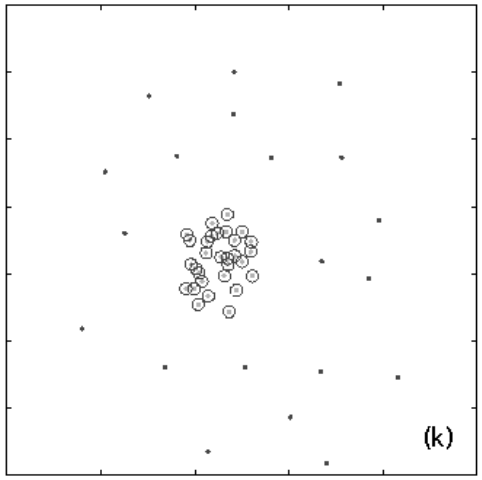


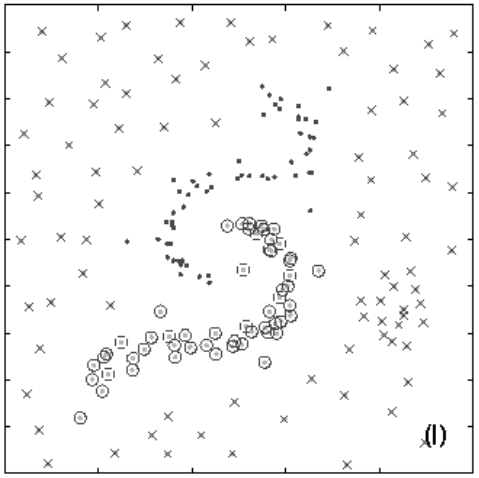


SC



TD





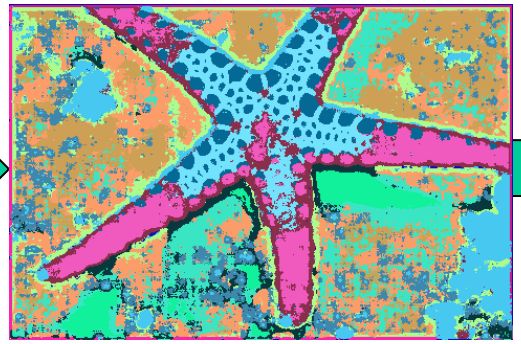
TD

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Image Segmentation Algorithm

Input

Superpixelization



Texton Feature

RAG

TD Mat

TD Clust

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Experiment: Image Segmentation

Qualitative result on BSDS300

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Experiment: Image Segmentation

Quantitative result on BSDS300

	PRI	VoI	GCE	BDE
MGD	0.7559	2.4701	0.1925	15.10
NTP	0.7521	2.4954	0.2373	16.30
Ncut	0.7853	2.1031	0.1947	**12.9703**
PRIF	**0.8006**	–	–	–
Ours	0.7926	**2.0871**	**0.1835**	13.1707

MGD: T. Cour et al.. Spectral Segmentation with Multiscale Graph Decomposition. *CVPR* 2005.

NTP: J.Wang et al.. Normalized Tree Partitioning for Image Segmentation. *CVPR* 2008

PRIF: M. Mignotte. A label field fusion Bayesian model and its penalized maximum rand estimator for image segmentation. *IEEE Trans. on Image Proc.*, 2010.

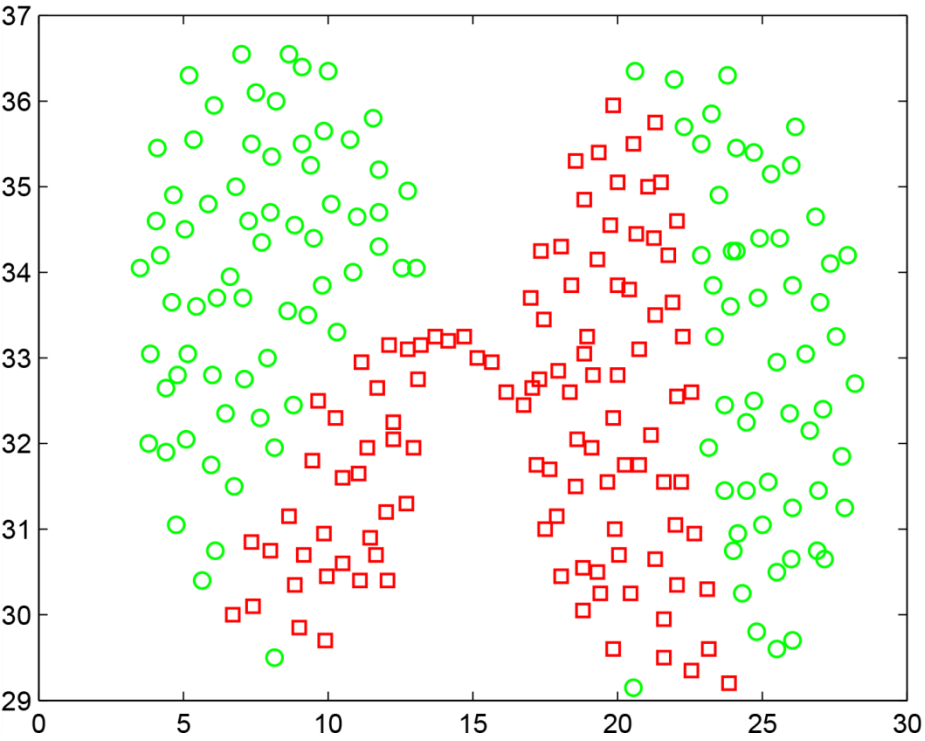
Conclusions

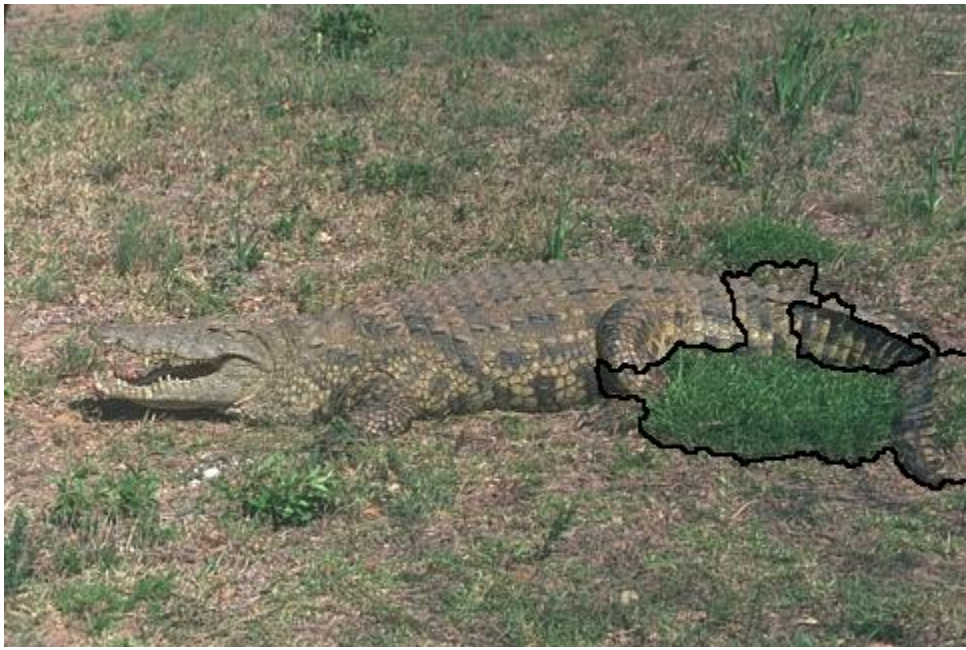
- Proposed a top-down clustering method.
- An approximate spectral clustering method without eigen-decomposition.
- Transitive distance vs. eigen-decomposition
- Able to handle arbitrary cluster shapes
- Application to image segmentation with good performance

Zhiding Yu et al., **Transitive Distance Clustering with K-Means Duality**, *CVPR* 2014.

Generalized TD with Minimum Spanning Random Forest (IJCAI15)

Robustness: Short Link Problem

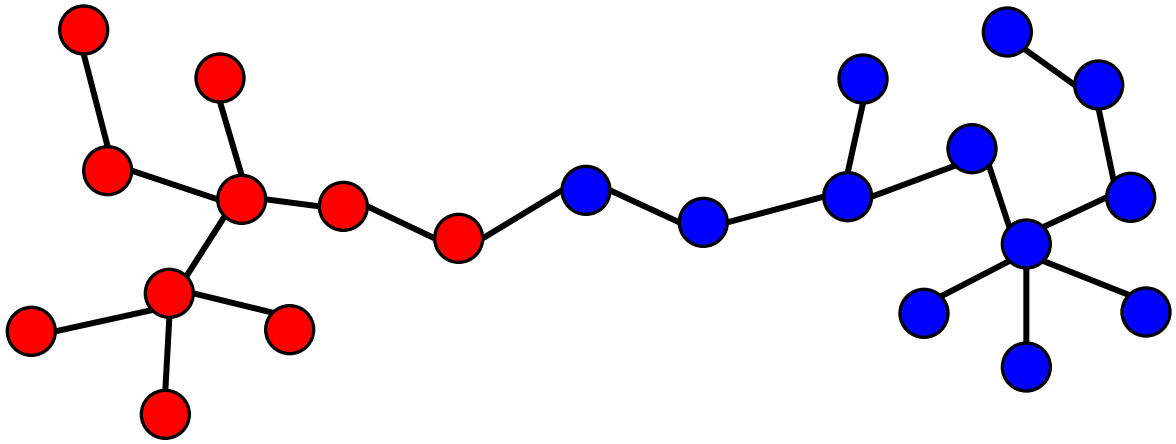




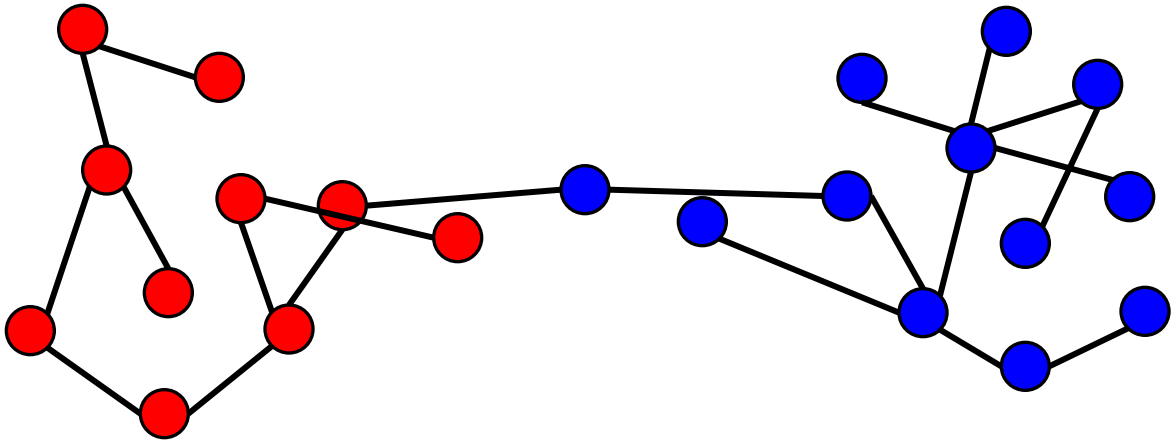
MST is an over-simplified representation of data. Therefore, TD clustering can be sensitive to noise. (but still much better than single linkage algorithm)

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Intuition: Consider Linkage Thickness



MST1



MST2

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Generalized TD (GTD): Definition

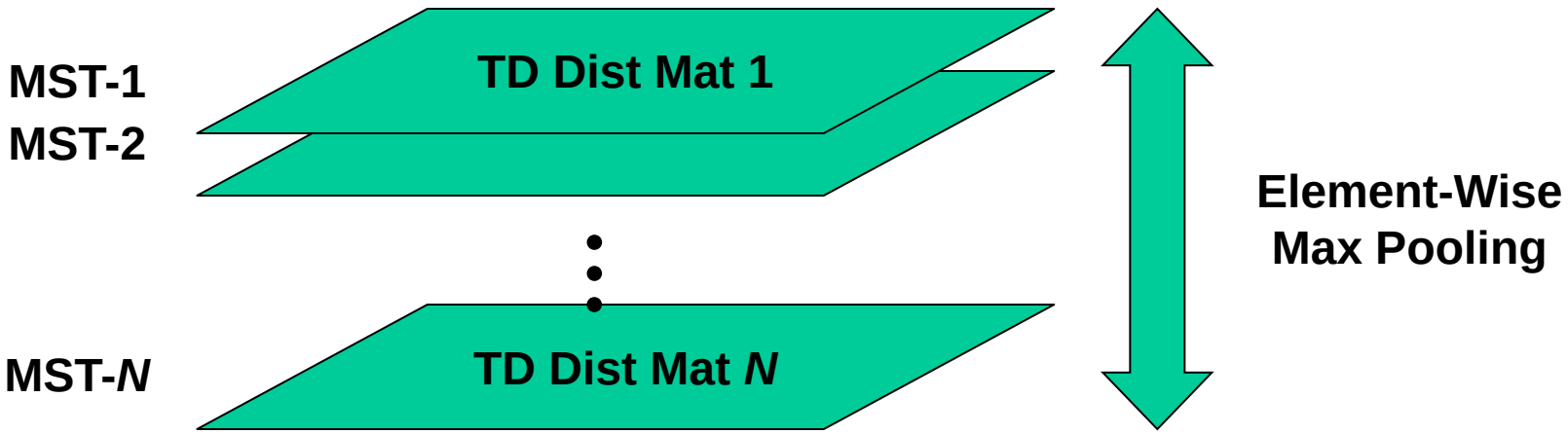
Definition: $$D_{gtd}(x_p, x_q) = \max_{t} \operatorname*{gmin}_{\substack{\mathcal{P}_t \in \mathbb{P}_t, \\ \forall t \in \{1,\ldots,T\}}} \max_{e \in \mathcal{P}_t} \{d(e)\}$$

Notes:

- Function "**gmin**" denotes the **generalized min** returning a set of minimum values from multiple sets.
- $\mathbb{P}_t$ denotes multiple sets of paths, each containing a set of all possible paths from one configuration (realization) of perturbed graph.

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Generalized TD (GTD): Definition



Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Theoretical Properties

Theorem 1:

The generalized transitive distance is also an ultrametric, and can also be embedded into a finite dimensional Euclidean space.

Theorem 2:

Given a set of bagged graphs, the transitive distance edges lie on the ***minimum spanning random forest (MSRF)*** *formed by MSTs extracted from these bagged graphs.*

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Perturbation Algorithm I

Algorithm 1 Extended Sequential Kruskal's Algorithm

1: Initialize $G_1 = G = (V, E)$, where G is a weighted graph and E is the set of available edges.
2: Extract MST from G_t using the Kruskal's algorithm and return the $n \times n$ pairwise transitive distance matrix.
3: Remove the set of MST edges P_t from G_t and update: $G_{t+1} = (V, E_t - P_t)$.
4: Repeat 2 to 4 for T times.
5: Perform element wise max pooling over the stack of transitive distance matrices.

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Top-Down Clustering

Algorithm 1: (Non-SVD)

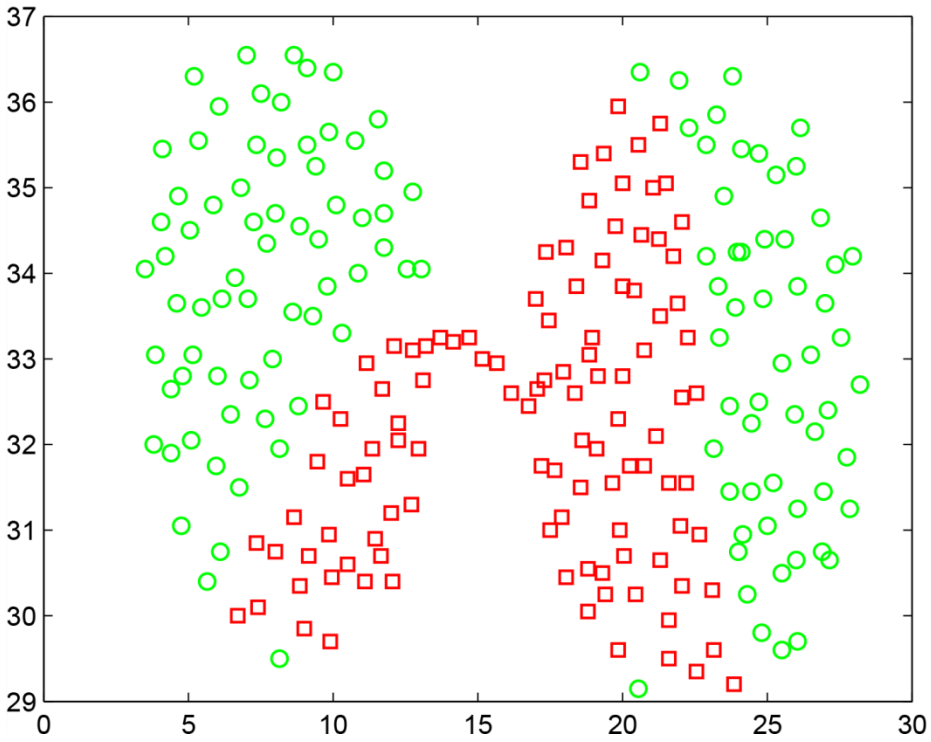
- Given a computed GTD pairwise distance matrix *D*, treat each row as a data sample
- Perform k-means on the rows to generate final clustering labels. (K-means Duality)

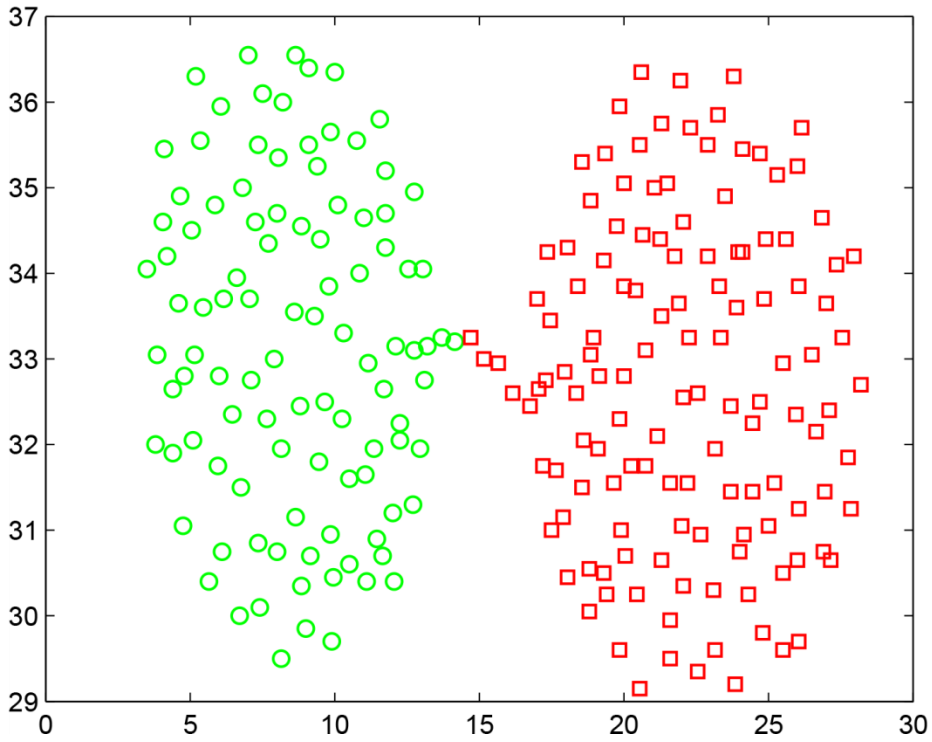
Algorithm 2: (SVD)

- Given a computed GTD pairwise distance matrix *D*, perform **SVD**: $D = U\Sigma V^*$
- Extract the top several columns of *U* with the largest singular values.
- Treat each row of the columns a data sample.
- Perform k-means on the rows to generate final clustering labels.

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Result on Toy Example





Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Perturbation Algorithm II

Algorithm 2 Random Perturbation Algorithm

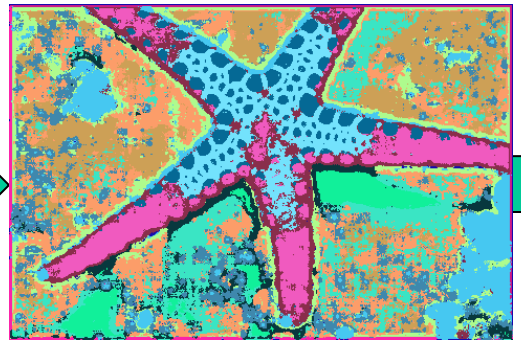
1: Initialize $G_1 = G = (V, E)$, where G is a weighted graph and E is the set of available edges.
2: If $t \neq 1$, obtain G_t by randomly perturbate the edge length of G with a random number $\epsilon * rand(1)$.
3: Extract MST from G_t using the Kruskal's algorithm and return the $n \times n$ pairwise transitive distance matrix.
4: Repeat 2 to 4 for T times.
5: Perform element wise max pooling over the stack of transitive distance matrices.

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

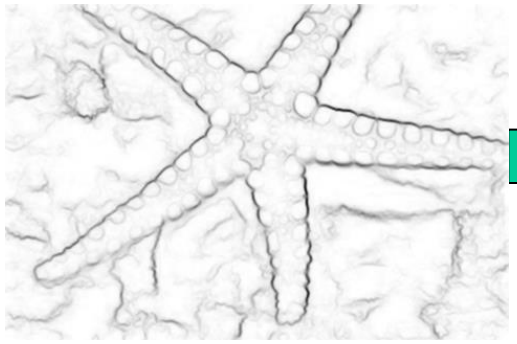
Image Segmentation Algorithm

Input

Superpixelization



Texton Feature



Structured Edge Det.

RAG

GTD Mat

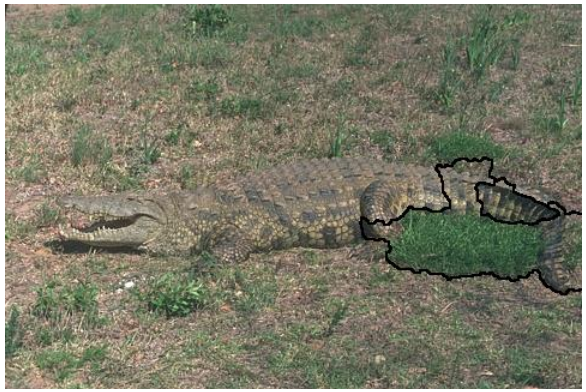
GTD+
Non-SVD

Experiment: Image Segmentation

Qualitative result on BSDS300





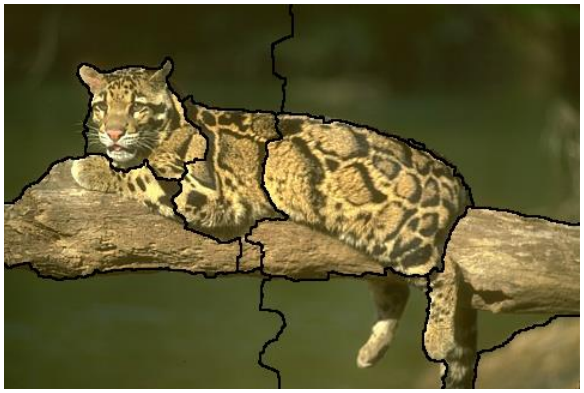
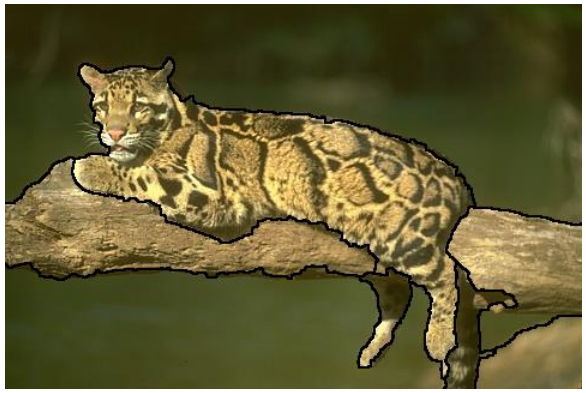
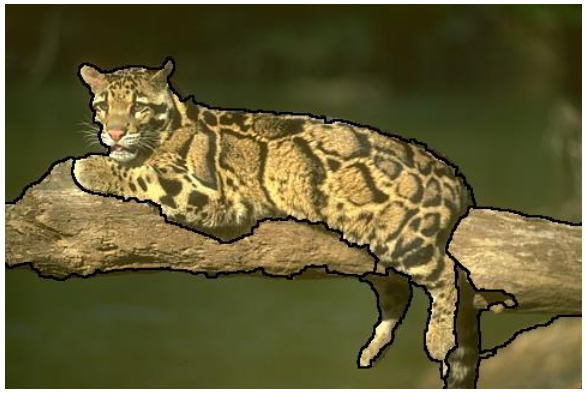



Normalized Cuts | **TD + Non-SVD** | **GTD + Non-SVD**

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Experiment: Image Segmentation

Qualitative result on BSDS300


Normalized Cuts | **TD + Non-SVD** | **GTD + Non-SVD**

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Experiment: Image Segmentation

Quantitative result on BSDS300

Method	PRI	VoI	GCE	BDE
[Cour *et al.*, 2005]	0.7559	2.47	0.1925	15.10
[Wang *et al.*, 2008]	0.7521	2.495	0.2373	16.30
[Mignotte, 2010]	0.8006	–	–	–
[Li *et al.*, 2011]	0.8205	1.952	0.1998	12.09
[Kim *et al.*, 2013]	0.8146	1.855	0.1809	12.21
[Li *et al.*, 2012]	0.8319	1.685	0.1779	11.29
[Arbelaez *et al.*, 2011]	0.81	1.65	–	–
[Yu *et al.*, 2014]	0.7926	2.087	0.1835	13.171
[Wang *et al.*, 2014]	0.8039	2.021	0.2066	13.77
Baseline: Ncut	0.7607	2.108	0.2217	14.608
Baseline: Transitive	0.8295	1.645	0.1688	10.568
GTD (Perturb.)	**0.8331**	**1.639**	**0.1655**	**10.372**

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

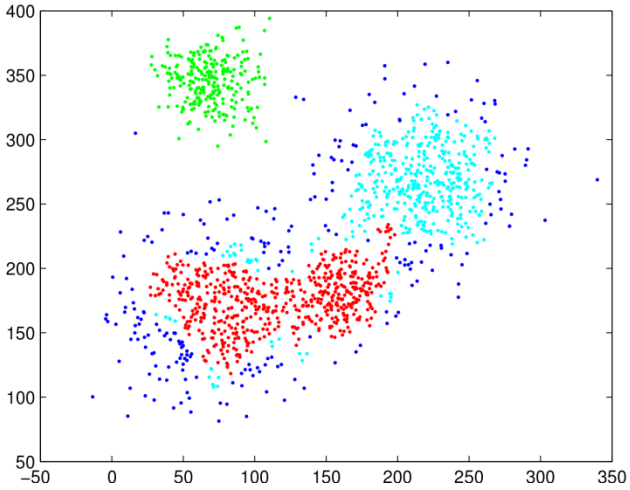
Conclusions

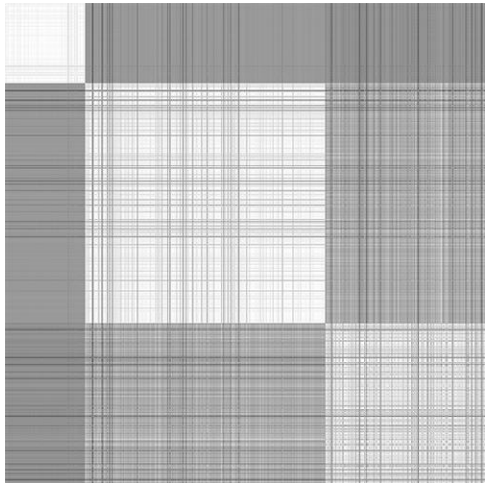
- Extending TD to GTD with minimum spanning random forest and max pooling
- Partially addresses the short link problem in data clustering and weak object boundaries in image segmentation
- Application to image segmentation with good performance

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

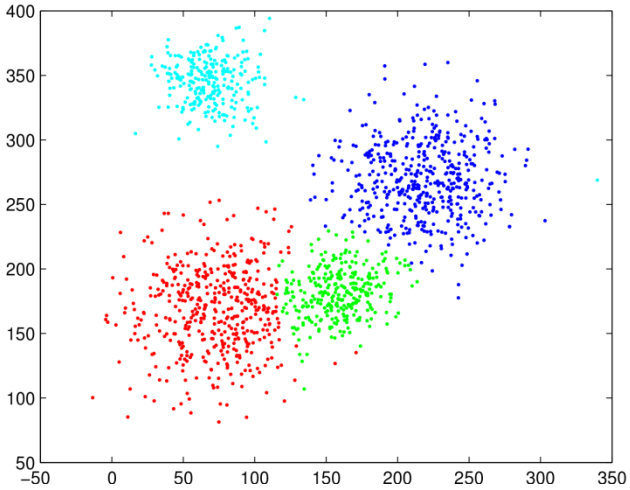
On Order-Constrained Transitive Distance (OCTD) Clustering (AAAI16)

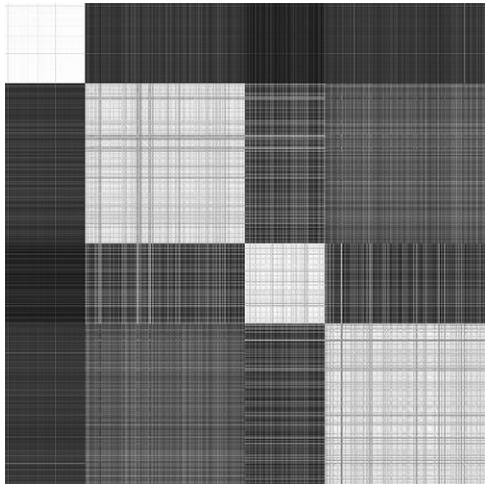
Robustness: Clustering Ambiguity





TD+SVD





OCTD+SVD

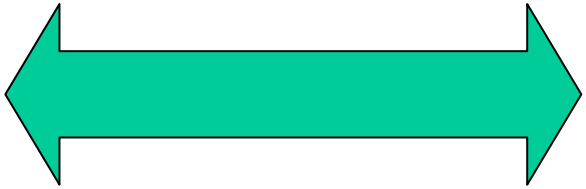
Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Intuition: Path Order Constraint

Euclidean Distance

- Weak cluster flexibility
- Strong cluster shape prior
- More robustness against clustering ambiguity
- Path order = 2

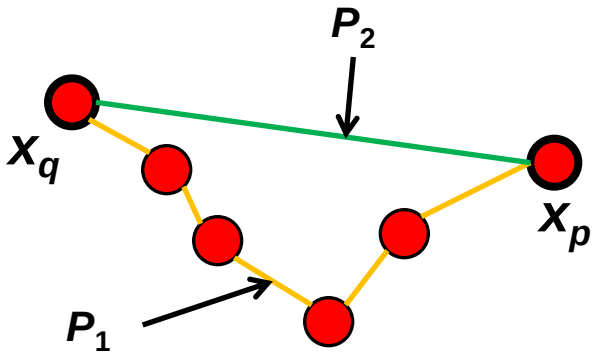
Trade-Off ?



Transitive Distance

- Strong cluster flexibility
- Weak cluster shape prior
- Less robustness against clustering ambiguity
- Large path order

Path Order:



- $O(P_1) = 6$
- $O(P_2) = 2$
- Euclidean dist. can be viewed as a special case of TD with order = 2.

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Order-Constrained TD: Definition

Definition: $$D_{octd}(x_p, x_q) = \min_{\substack{\mathcal{P}\in\mathbb{P}, \\ \mathcal{O}(\mathcal{P})<L}} \max_{e\in\mathcal{P}}\{d(e)\}$$

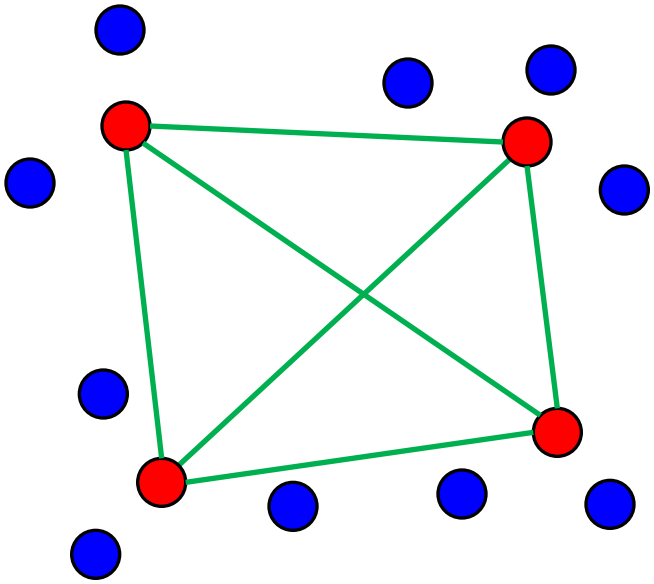
Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.

Computing OCTD

- Computing OCTD seems to be easier than TD because the set of candidate path is only a subset of TD (high order paths not considered).
- Remember the following theorem for TD:
  Given a weighted graph with edge weights, each transitive edge lies on the minimum spanning tree.
- The same theorem does not hold on OCTD!
- Finding the true OCTD is actually very hard.

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

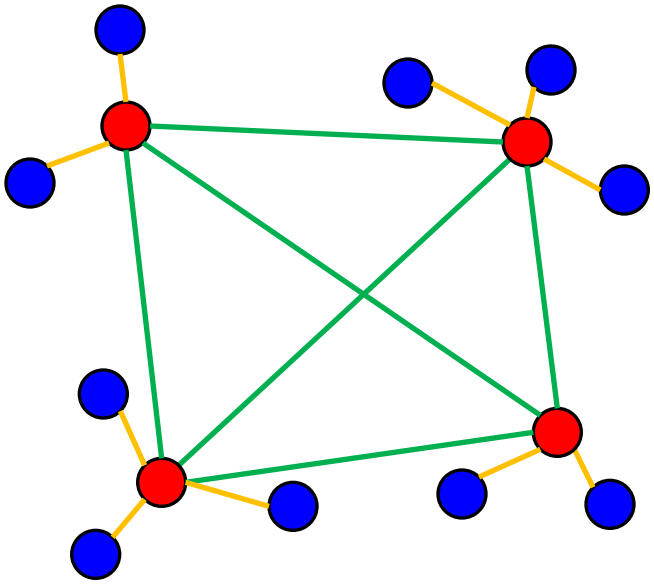
Approximating OCTD with Randomized Samplings



The sampled data forms a clique G_C

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

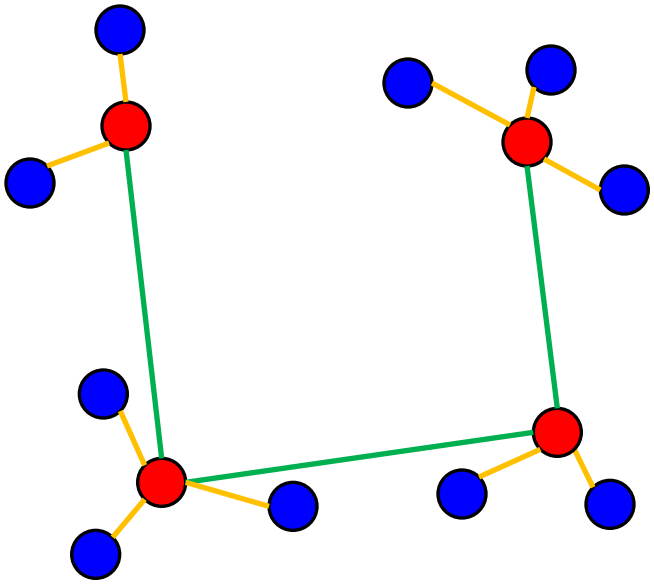
Approximating OCTD with Randomized Samplings



The rest of the data links to nearest sampled data and form a spanning graph G_S together with the clique G_C.

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Approximating OCTD with Randomized Samplings



Compute a pairwise TD matrix on G_S by extracting an MST

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Approximating OCTD with Randomized Samplings

Theorem 1:

The maximum possible path order on the spanning graph G_C is upper bounded by $|S| + 2$.

Theorem 2:

For any pair of nodes, the number of connecting paths on the spanning graph is upper bounded by (|S|-2)!

Theorem 3:

The transitive distance obtained on lower-bounded by the order-constrained transitive distance obtained on the original fully connected graph G

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Sampling Strategy

Kernel Density Estimation:

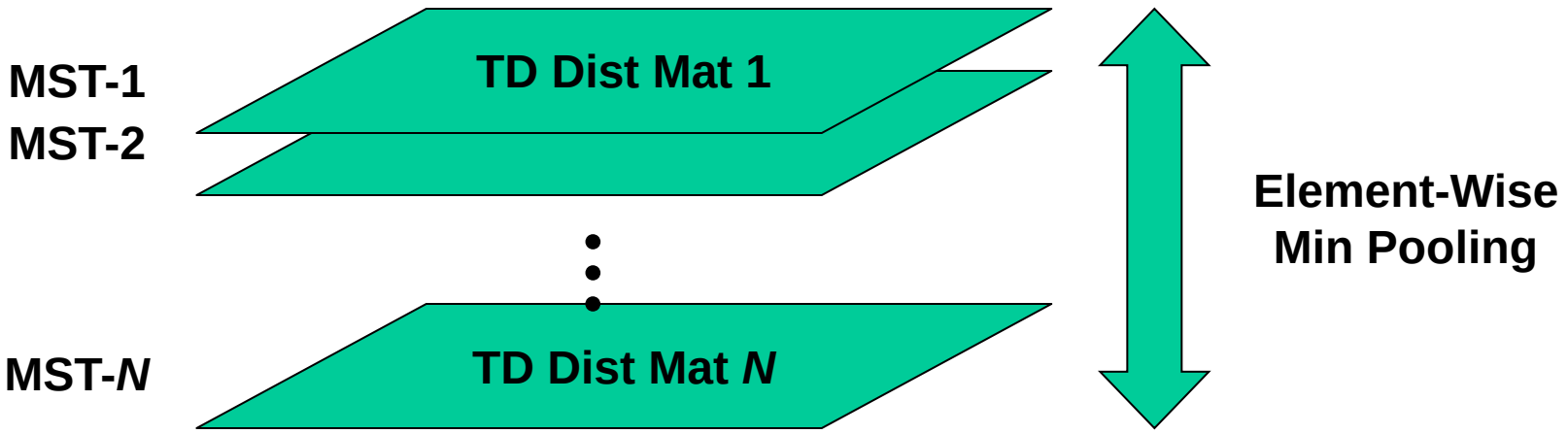
$$\hat{p}(\mathbf{x}_i) = C \sum_{j=1}^{N} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$$

Bandwidth Estimation:

$$\hat{\sigma} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathrm{knn}(\mathbf{x}_i, k)\|_2$$

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Ensemble with Min Pooling



Theorem 4:

Given the set of randomly sampled OCTD distances, min pooling gives the optimal approximation of the true OCTD from the fully connected graph *G*

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Ensemble with Mean Pooling

- Unfortunately, OCTD (Min) is not a metric.
- We can use mean pooling instead of min pooling to return OCTD (Mean) which sub-optimally approximates OCTD but holds metricity.
- **Theorem 5:** OCTD (Mean) is a metric.

Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Experiment: Toy Example Datasets

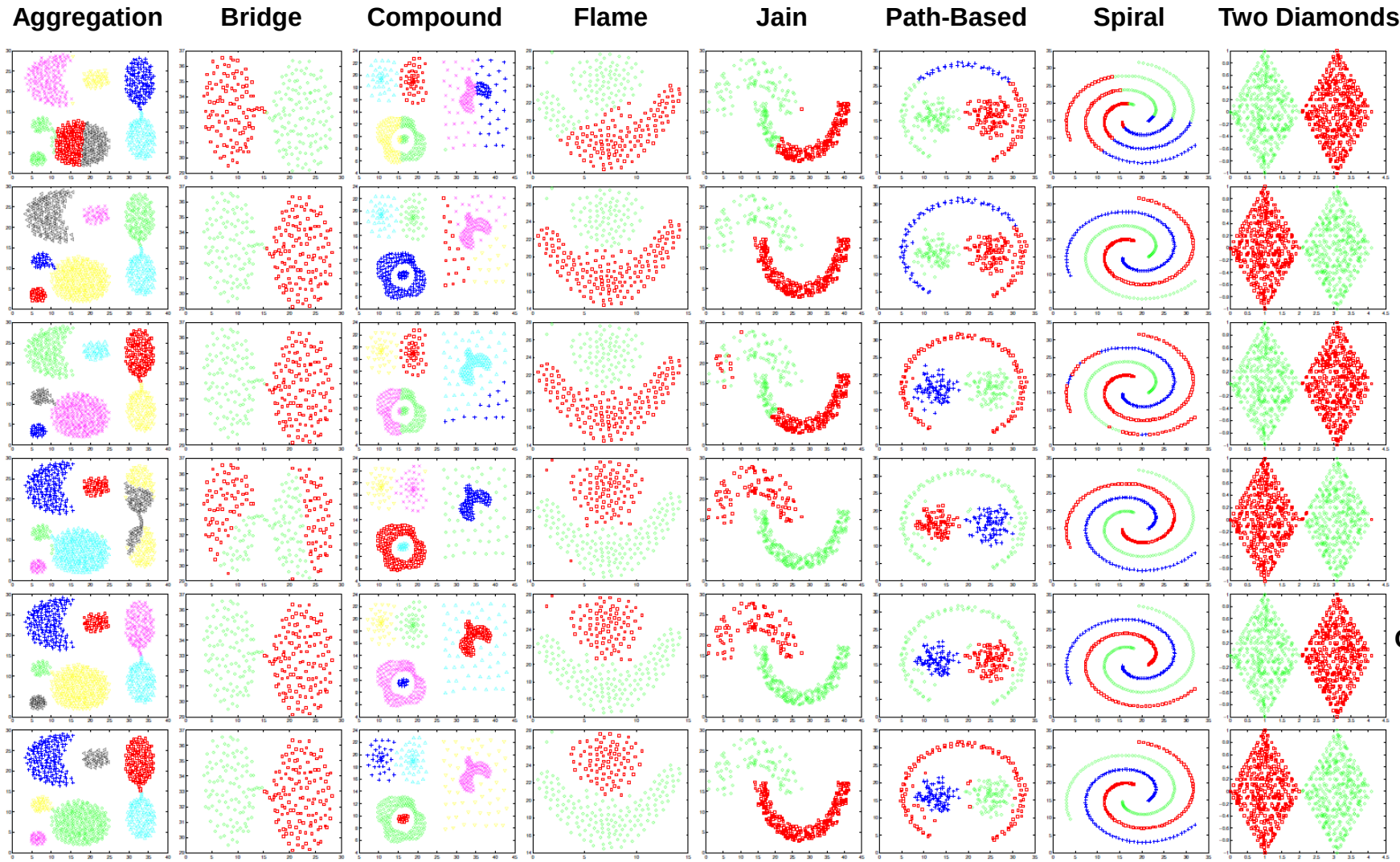


Figure 2: Results of comparing methods on toy examples with varying cluster shapes (Best viewed in color). Row 1-6 respectively correspond to Kms (Euc), SC, Ncut, TD+SVD, OCTD (Min) and OCTD (Mean). Names of examples are respectively "Aggregation", "Bridge", "Compound", "Flame", "Jain", "Pathbased", "Spiral" and "Two Diamonds".

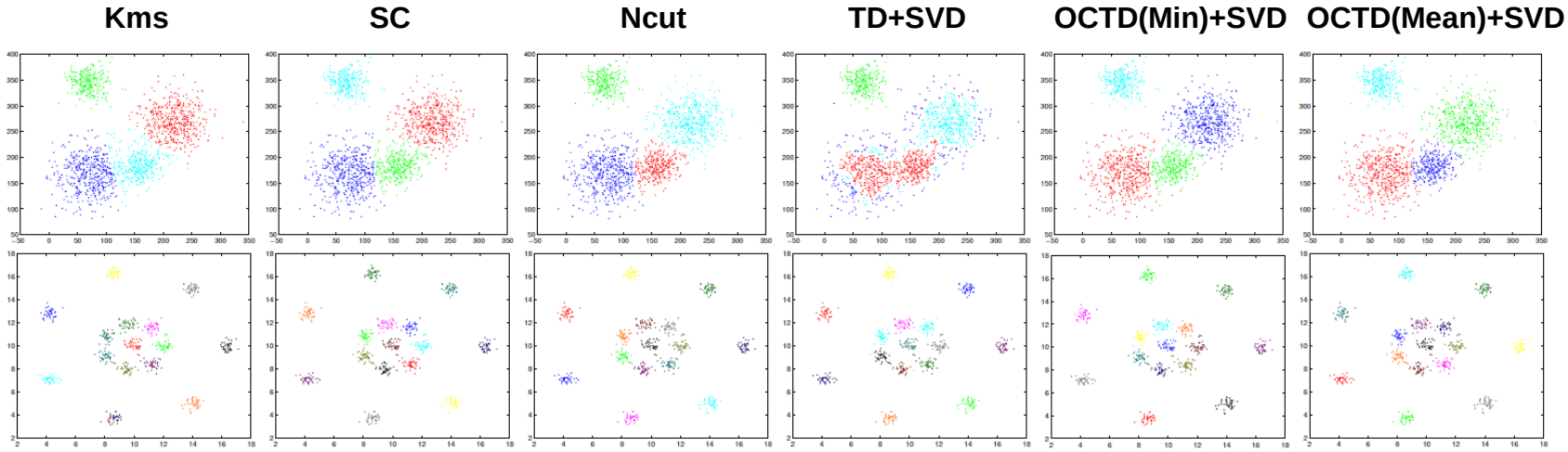


Figure 3: Results of comparing methods on toy examples with densely aligned Gaussian distributions (Best viewed in color). Column 1-6 respectively correspond to K-Means, SC, Ncut, TD+SVD, OCTD (Min) and OCTD (Mean). Names of examples are respectively "Gaussian" and "R15".

Table 1: Quantitative results of comparing methods on toy datasets. Accuracies are measured with %.

Method	Aggregation	Bridge	Compound	Flame	Jain	Path.	Spiral	TwoDiam.	Gaussian	R15
Kms (Euc)	93.91	99.14	83.21	83.75	78.28	74.58	33.97	**100**	93.13	92.5
SC	99.37	99.14	91.73	97.92	**100**	87.63	100	**100**	95.2	**99.67**
Ncut	99.37	99.14	86.72	98.75	77.48	**98.66**	87.18	**100**	**95.8**	**99.67**
TD+SVD	87.94	60.78	99.5	98.75	**100**	96.99	**100**	99.25	78.6	92.33
OCTD (Min)	**99.87**	**99.57**	**99.75**	**100**	**100**	96.66	**100**	**100**	95.33	99
OCTD (Mean)	99.75	**99.57**	**99.75**	98.33	**100**	96.32	**100**	**100**	**95.8**	**99.67**

Experiment: Image Datasets

Extended Yale B Dataset (ExYB)

- 2414 frontal-faces (192 x 168) of 38 subjects.
- Resize images to 55 x 48
- PCA whitening with 99% of energy

AR Face Dataset (AR)

- 50 male and 50 female subjects, 1400 cropped faces
- Resize images to 55 x 40
- PCA whitening with 98% of energy

USPS Dataset

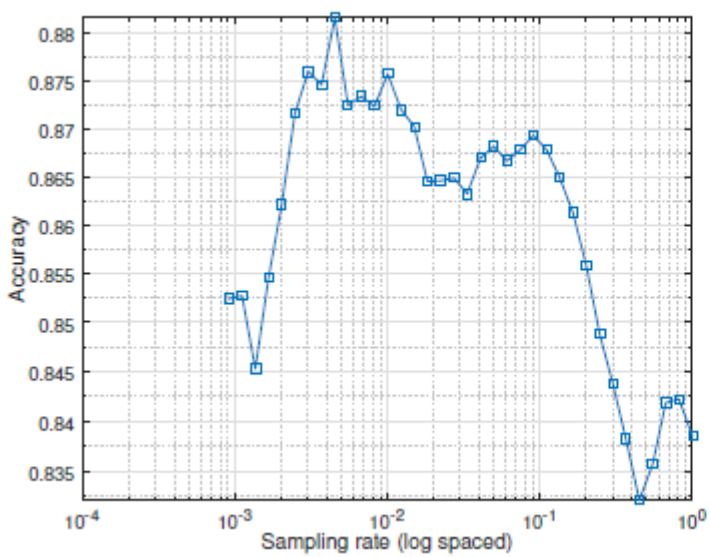
- 9298 16 x 16 hand written digit images
- PCA whitening with 98.5% of energy

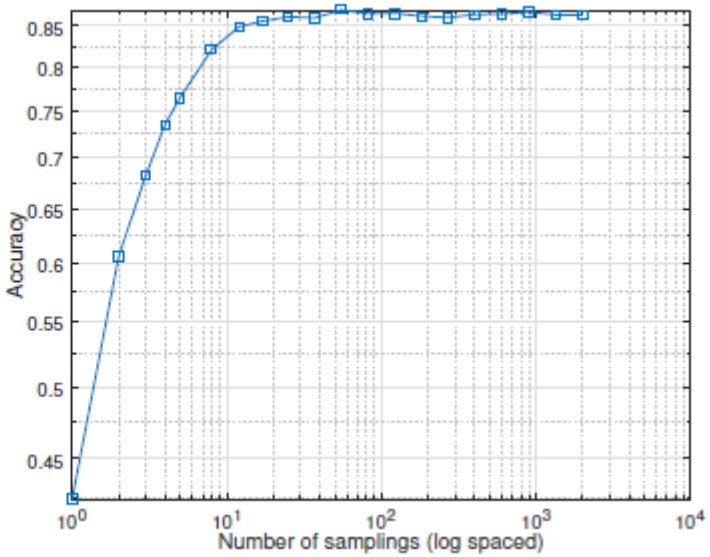
Experiment: Image Datasets

Clustering Accuracies (%)

Method	Kms	SC	Ncut	TD	OCTD (Min)
ExYB	44.74	87.28	83.76	82.81	**90.64**
AR	64.29	80.64	87.29	83.85	**88.28**
USPS	64.38	82.94	82.38	54.31	**85.13**

Parameter Experiment





Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Experiment: Large-Scale Speech Data

Table 3: Quantitative results of comparing methods on speech datasets. Accuracies are measured with %.

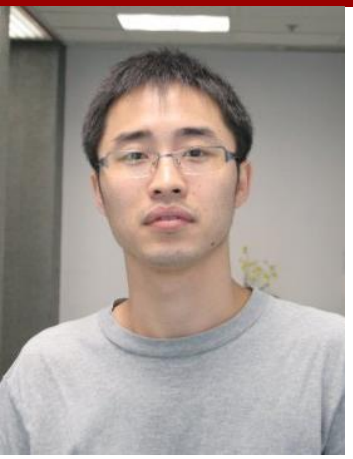
Method	Kms (Euclid)	Kms (Cos)	SC	Ncut	TD+SVD	OCTD (Min)	OCTD (Mean)
NIST 04	66.32	81.49	83.32	80.49	77.17	**84.9**	84.51
NIST 05	72.99	77.08	74.3	76.1	72.86	**77.87**	73.04
NIST 06	79.84	86.43	80.72	84.4	87.07	**88.29**	83.47
NIST 08	74.52	78.58	**81.51**	62.65	74.13	77.91	78.81
NIST Combined	70.85	78.97	76.21	71.66	72.07	**80.89**	77.24
Switch Board	86.03	90.80	87.79	80.83	78.73	87.53	**90.88**

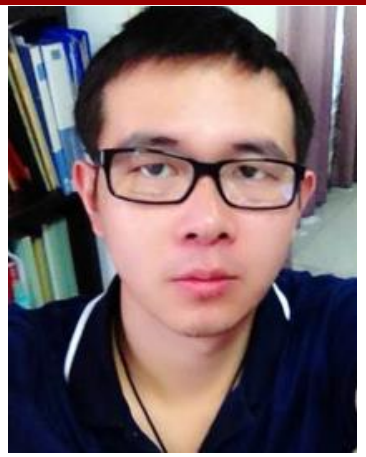
Zhiding Yu et al., **On Order-Constrained Transitive Distance Clustering**, *AAAI* 2016.

Conclusions

- Extending TD to OCTD with random sampling and min pooling
- Significantly improved the algorithm robustness against clustering ambiguity
- Application to both image data and large scale speech data clustering with good performance.

Zhiding Yu et al., **Generalized Transitive Distance with Minimum Spanning Random Forest**, *IJCAI* 2015.




Thank You!