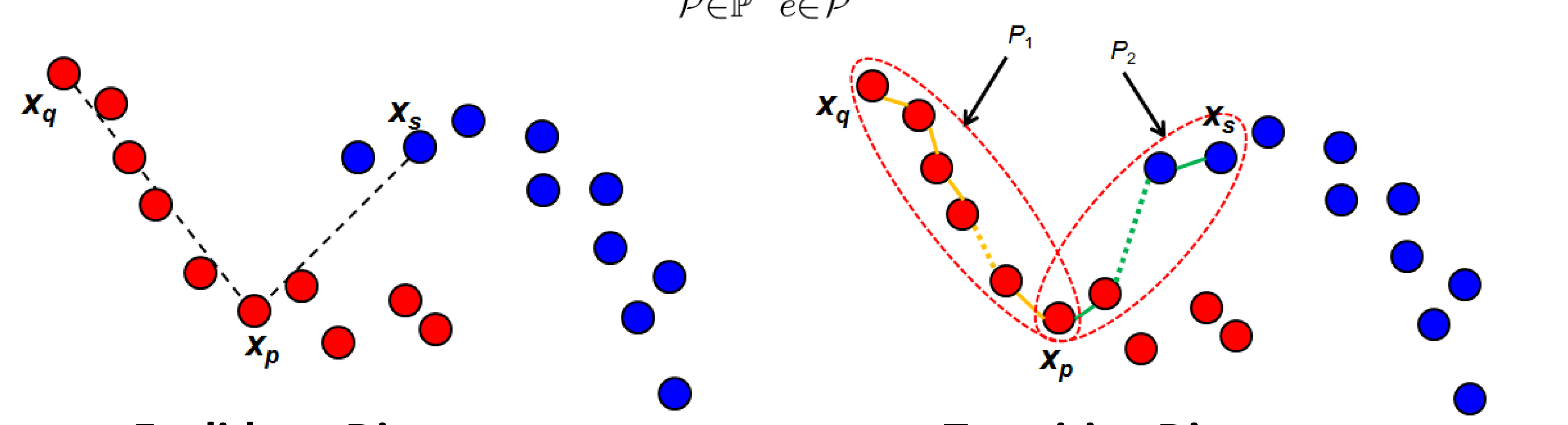
# On Order-Constrained Transitive Distance Clustering

Zhiding Yu, Weiyang Liu, Wenbo Liu, Yingzhen Yang, Ming Li, B. V. K. Vijaya Kumar

yzhiding@andrew.cmu.edu, kumar@ece.cmu.com

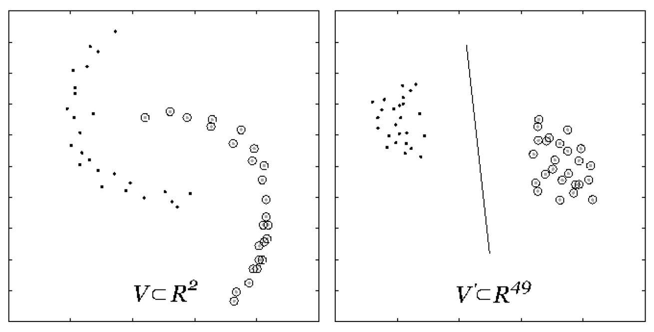
## Introduction I: Transitive Distance (TD)

- **Math Definition:** $D_T(x_p, x_q) = \min_{\mathcal{P} \in \mathbb{P}} \max_{e \in \mathcal{P}} \{d(e)\}$




**Euclidean Distance** **Transitive Distance**

- **Theorem 1:** TD is an ultra-metric.
- **Theorem 2:** Every finite ultrametric space with $n$ distinct points can be isometrically embedded into an $n$−1 dim Euclidean space.




- **Theorem 3:** Given a weighted graph with edge weights, each transitive edge lies on the minimum spanning tree (MST).

## Introduction II: TD Clustering

- Under TD embedding, data from the same cluster becomes compact. It is therefore desirable to perform clustering in the embedded space.
- Intuitively, TD clustering can be regarded as an approximate spectral clustering (SC) where TD embedding is similar to eigen decomposition.

**Algorithm-1 (Non-SVD):**

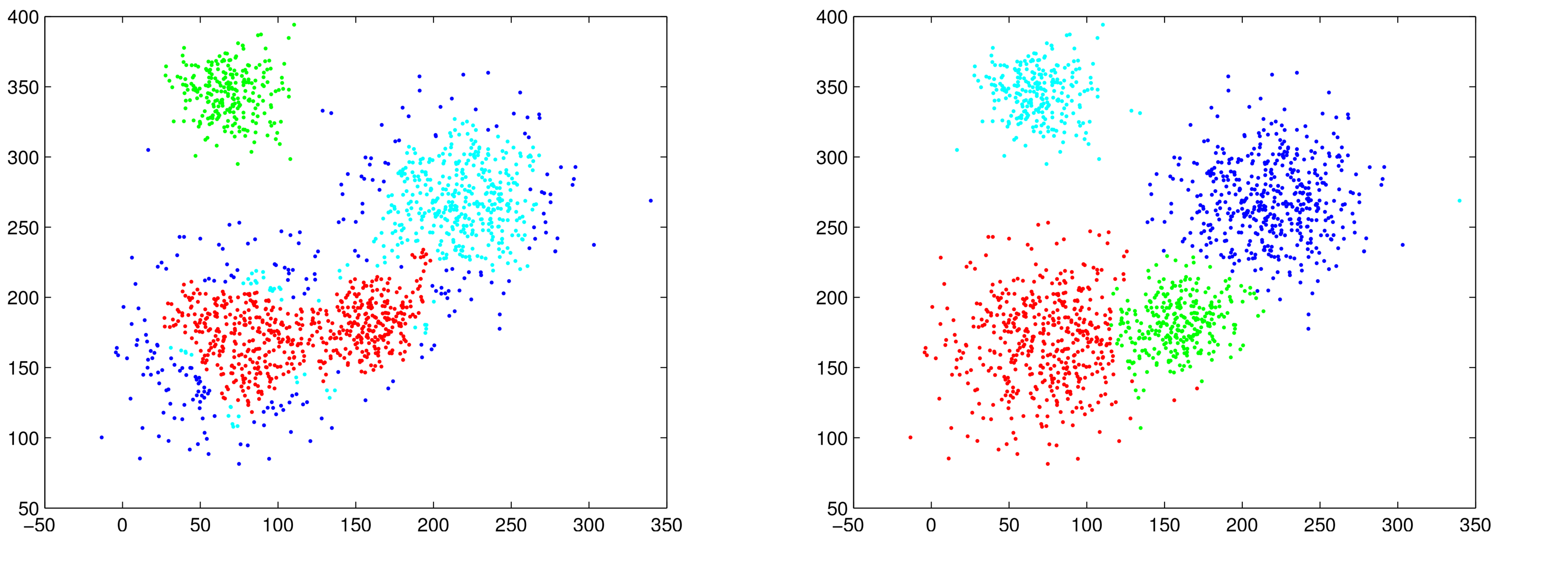
- Treat each row of TD matrix as embedded data and apply k-means.
- Produces similar clustering results as directly performing k-means in the embedded space. (**K-means Duality**)

**Algorithm-2 (SVD):**

- Given a computed TD matrix D, perform SVD: $D = U\Sigma V^*$
- Treat each row from the top several columns with largest eigenvalues as data samples, and perform k-means.

## Order-Constrained TD (OCTD)

- Math Definition: $D_{octd}(x_p, x_q) = \min_{\substack{\mathcal{P} \in \mathbb{P}, \\ \mathcal{O}(\mathcal{P}) < L}} \max_{e \in \mathcal{P}} \{d(e)\}$
- TD is sensitive to short links (see the following left figure). The additional constraint on path order is able to introduce more robustness.



**Clustering with TD** **Clustering with (Approx.) OCTD**

## Intuition: Path-Order Constraint

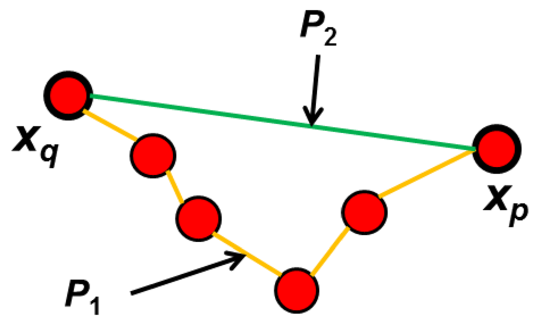
**Euclidean Distance**

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


**Transitive Distance**

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

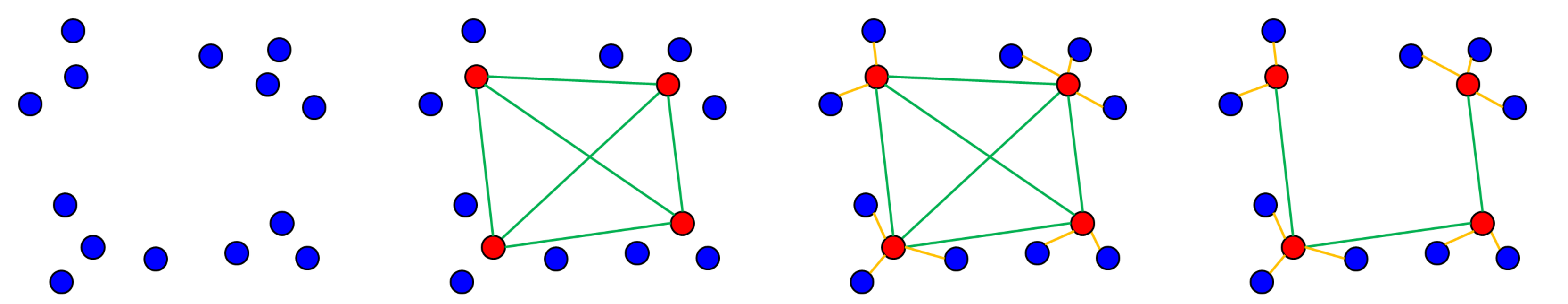
**Path Order:**




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

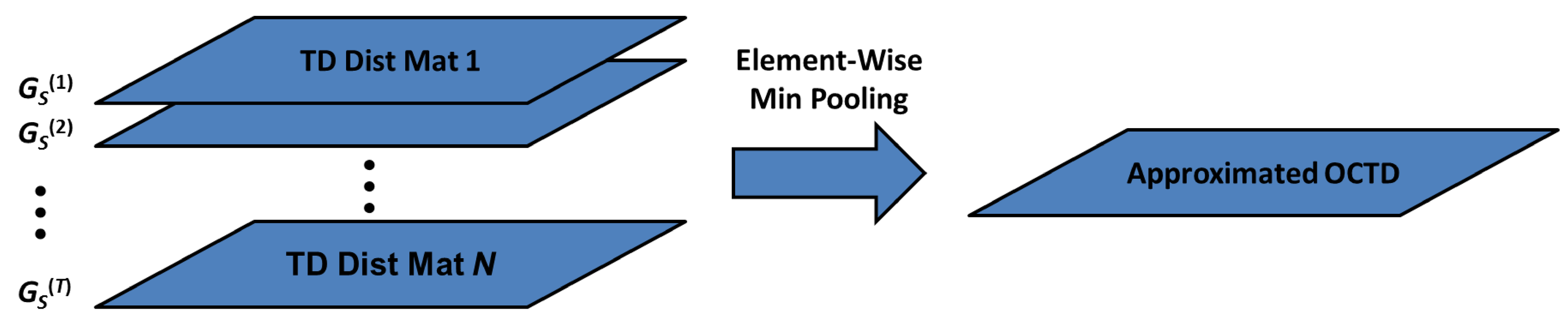
## Approximating OCTD with Random Subsampling and Diversified Spanning Graphs

- **Theorem 3** gives a practical solution to the computation of TD through MST.
- No similar theorems apply for OCTD because of the path order constraint. Finding the exact OCTD becomes infeasible when dataset size is large.
- As a result, finding a method to approximate OCTD is necessary.



**Approximation flow: 1. Original dataset. 2. Subsampled data and the constructed clique. 3. Connecting non-sampled data to nearest sampled ones to form a spanning graph. 4. Obtain TD from the spanning graph using MST.**

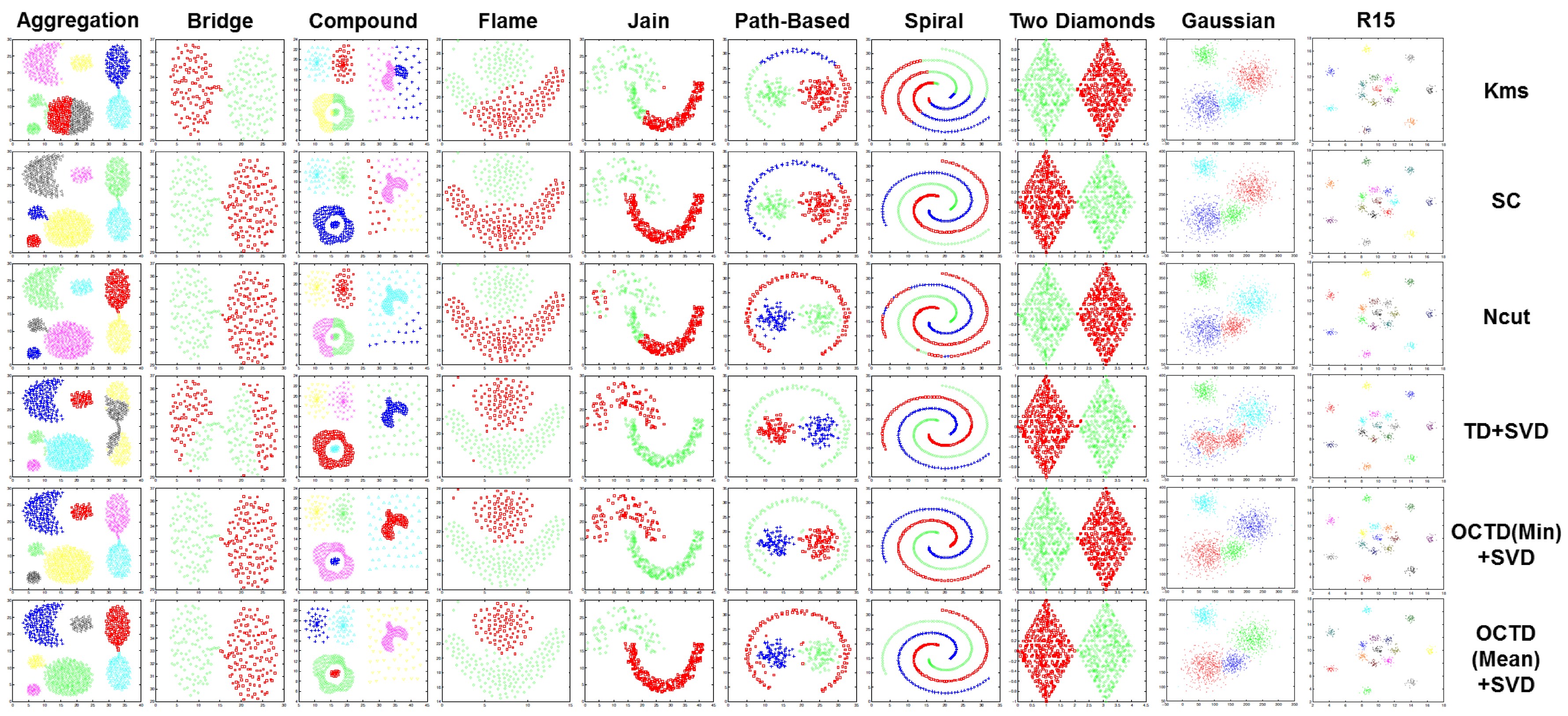
- **Theorem 4:** The maximum possible path order on the spanning graph $G_S$ is upper bounded by $|S| + 2$.
- **Remark:** The pairwise TD matrix obtained on $G_S$ is order-constrained and can be used to approximate the true OCTD.
- Repeat $T$ times diversify spanning graphs, each returning a single TD matrix. Perform element-wise min pooling on the stack of $T$ TD matrices.




- Unfortunately, OCTD (Min) is not a metric. One can also use mean pooling instead of min pooling to sub-optimally approximate OCTD but hold metricity.

## Experimental Results

### Toy Example Datasets




| 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** |

### Image Datasets

**Extended Yale B Dataset (ExYB)**

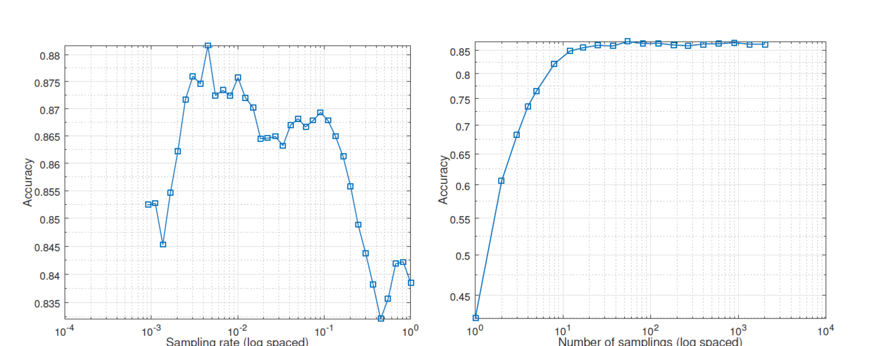
- 2414 frontal-faces (192 x 168) of 38 subjects.
- Resize images to 55 x 48
- PCA whitening with 99% of energy

**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




**Parameter experiment. Left: Varying the sampling rate and fixing T = 500. Right: Varying T and fixing the sampling rate to be 0.06.**

| 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** |

### Large-Scale Speech Datasets

| 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** |