



JHU vision lab

# Understanding Doubly Stochastic Clustering

Tianjiao Ding<sup>†</sup>, Derek Lim<sup>‡</sup>, René Vidal<sup>†</sup>, Benjamin D. Haeffele<sup>†</sup>

<sup>†</sup>Mathematical Institute for Data Science, Johns Hopkins University

<sup>‡</sup>Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology



THE DEPARTMENT OF BIOMEDICAL ENGINEERING

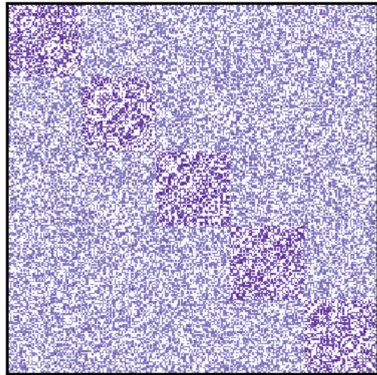
The Whitaker Institute at Johns Hopkins



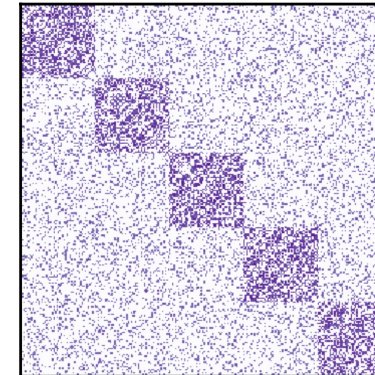
JOHNS HOPKINS  
MATHEMATICAL INSTITUTE  
for DATA SCIENCE

# Spectral Clustering

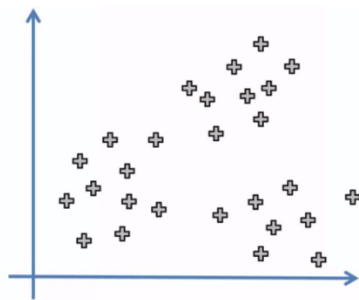
Define affinity matrix  $K$



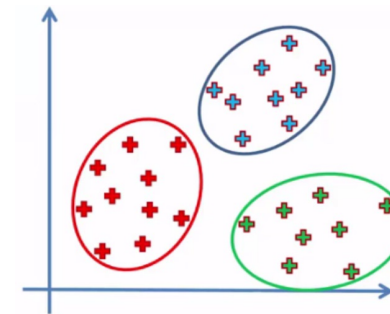
Compute normalized affinity  $A$



Compute embedding of the data from eigenvectors of Laplacian  $L = D_A - A$

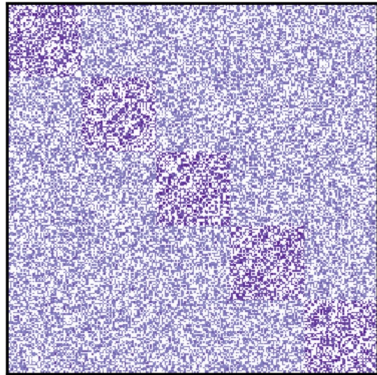


Cluster the embedding

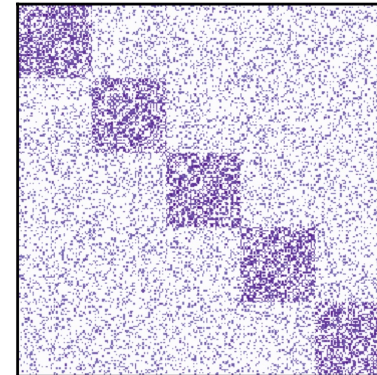


# Spectral Clustering

Define affinity matrix  $K$

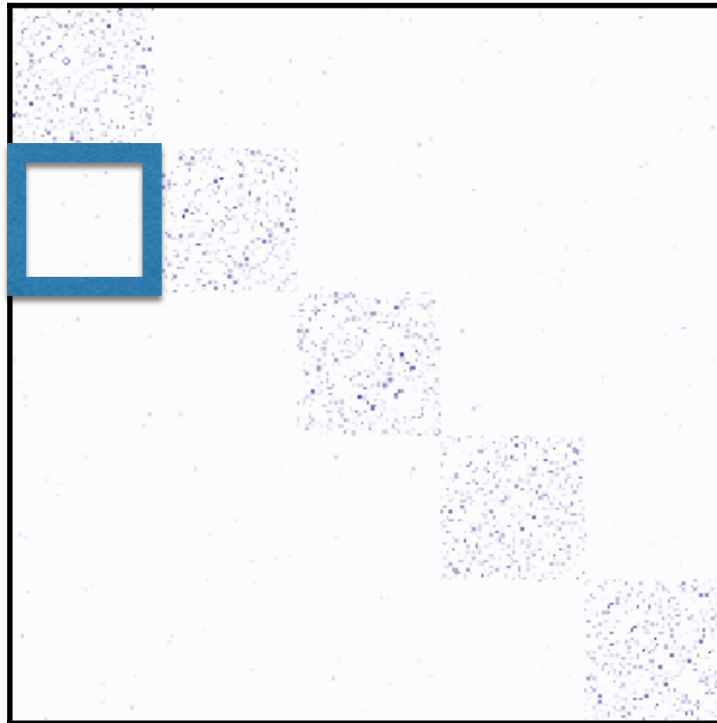


Compute normalized affinity  $A$



# Ideal Affinity for Clustering

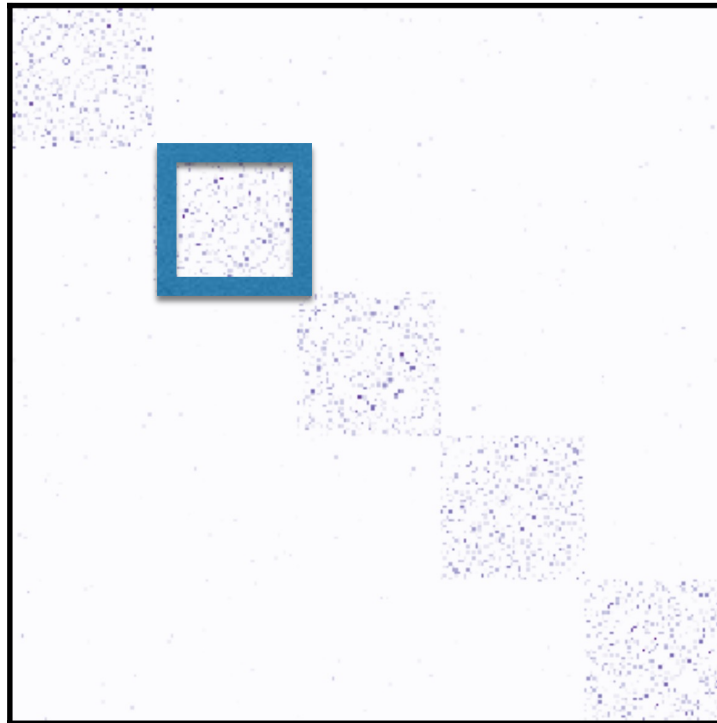
- No False Connections (NFC) for points between different clusters





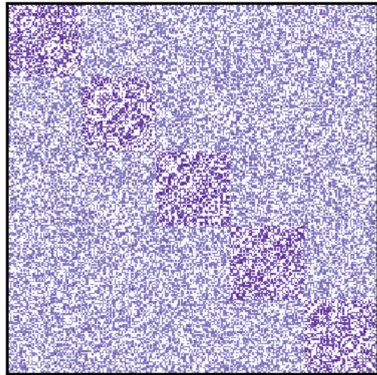
# Ideal Affinity for Clustering

- **Connectivity:** points within the same cluster are well connected



# How to get good affinity for clustering

Define affinity matrix  $K$



- Design the affinity based on **data geometry**
  - Clusters are centroids
  - Clusters are linear subspaces [1-4]

Most works only guarantees  
NFC but **not** connectivity

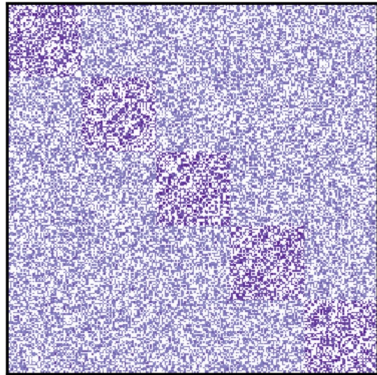
We give guarantees on both  
NFC **and** connectivity

[1] E. Elhamifar and R. Vidal, Sparse Subspace Clustering, 2009.  
[2] G. Liu et al, Robust Subspace Segmentation by Low-Rank Representation, 2010.  
[3] C. Lu et al, Robust and Efficient Subspace Segmentation via Least Squares Regression, 2012.  
[4] C. You et al, Oracle based active set algorithm for scalable elastic net subspace clustering, 2016.

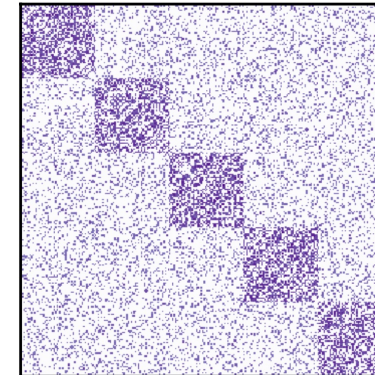


# How to get good affinity for clustering

Define affinity matrix  $K$



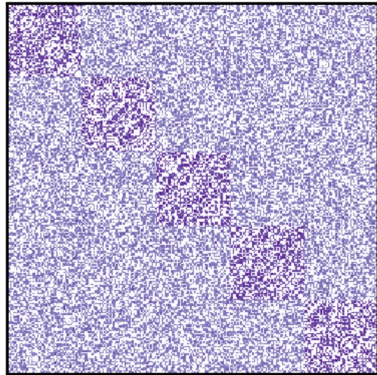
Normalized  $A$  by thresholding  $K$



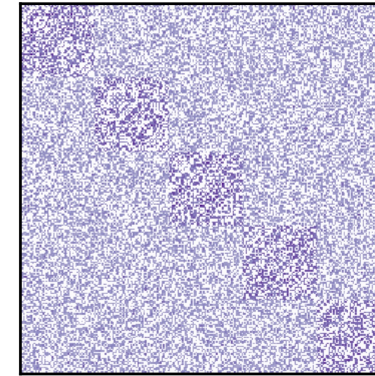
- Use **normalization** to clean up the affinity?
- ✗ Sparsify the affinity by thresholding

# How to get good affinity for clustering

Define affinity matrix  $K$



Symmetric normalized affinity  $A$



- Use **normalization** to clean up the affinity?
  - ✗ Sparsify the affinity by thresholding
  - ✗ Normalized cut [1]

[1] Shi, J. and Malik, J. Normalized cuts and image segmentation, 2000.

[2] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.

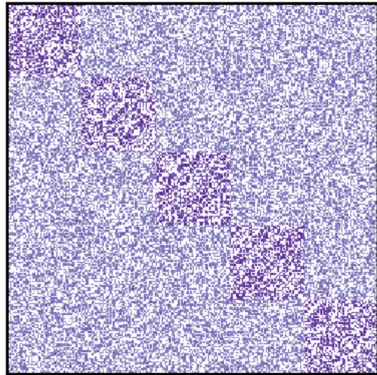
[3] Lim et al, Doubly stochastic subspace clustering, 2020.



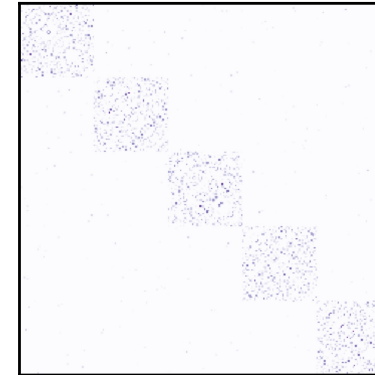


# How to get good affinity for clustering

Define affinity matrix  $K$



Doubly stochastic normalized  $A$



- Use **normalization** to clean up the affinity?
  - ✗ Sparsify the affinity by thresholding
  - ✗ Normalized cut [1]
  - ✓ Doubly stochastic projection [2,3]

[1] Shi, J. and Malik, J. Normalized cuts and image segmentation, 2000.

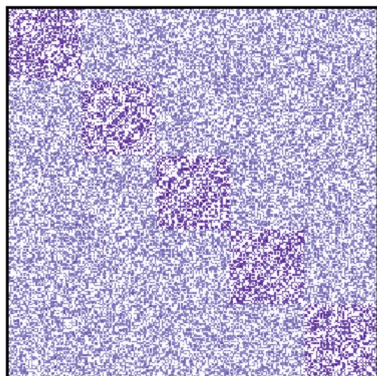
[2] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.

[3] Lim et al, Doubly stochastic subspace clustering, 2020.

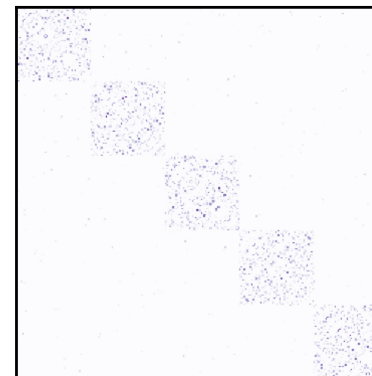


# Why doubly stochastic projection

Define affinity matrix  $K$



Doubly stochastic normalized  $A$



- $\mathcal{A}$  : the set of doubly stochastic matrices
  - Normalized cut  $\approx$  the closest matrix in  $\mathcal{A}$  to  $K$  under KL divergence [1]
  - **Doubly stochastic projection** := the closest matrix in  $\mathcal{A}$  to  $K$  under  $\ell_2$  metric
- Doubly stochastic projection achieves SOTA clustering performance [2]
  - E.g., 98.4% clustering accuracy on COIL-100, 99% on MNIST

Theoretical understanding? This paper: 

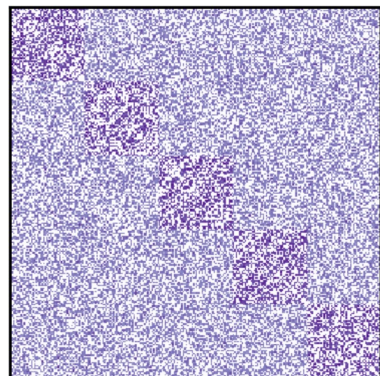
[1] Zass, and Shashua, Doubly Stochastic Normalization for Spectral Clustering, 2006.

[2] Lim et al, Doubly stochastic subspace clustering, 2020.

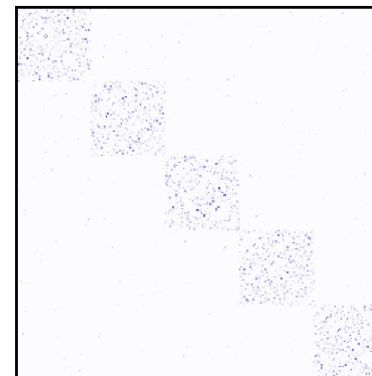


# Contributions

Define affinity matrix  $K$



Doubly stochastic normalized  $A$



- ✓ Provable guarantees for  $A$  being NFC and well-connected
- ✓ Additional guarantees for subspace clustering (each cluster is a subspace)
  - Guarantees depend on interpretable quantities (angles between subspaces, etc.)

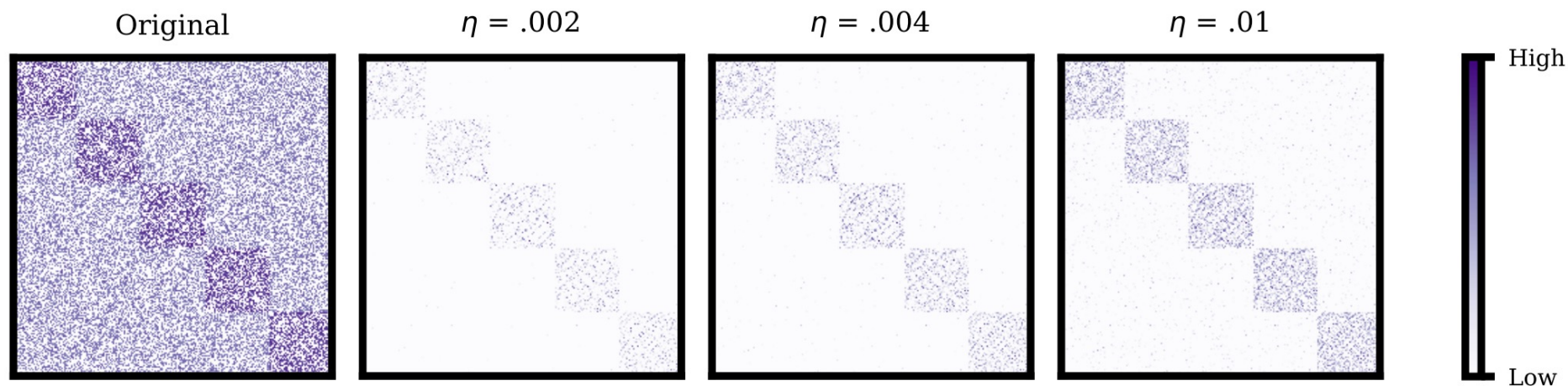
# Doubly Stochastic Normalization

- Given input affinity  $\mathbf{K} = \mathbf{K}^\top \in \mathbb{R}^{n \times n}$ , doubly stochastic normalized  $\mathbf{A}^*$  is given by

$$\mathbf{A}^* = \underset{\mathbf{A} \in \mathcal{A}}{\operatorname{argmin}} \left\| \mathbf{A} - \frac{1}{\eta} \mathbf{K} \right\|_F$$

the set of doubly stochastic matrices

$\eta > 0$ : parameter such that  $\mathbf{A}^*$  is sparser as  $\eta \downarrow$



✓ Compute the spectral embedding directly from  $\mathbf{A}^*$



# Theorem: Optimality Conditions

$$\min_{A \in \mathcal{A}} \left\| A - \frac{1}{\eta} K \right\|_F$$

$$\alpha^* \in \mathbb{R}^n: \frac{1}{n} \sum_{j=1}^n \max(K_{ij} - \alpha_i^* - \alpha_j^*, 0) = \eta, \forall i \quad (\#)$$

$$A_{ij}^* = \frac{1}{\eta} \text{ReLU}(K_{ij} - \alpha_i^* - \alpha_j^*)$$

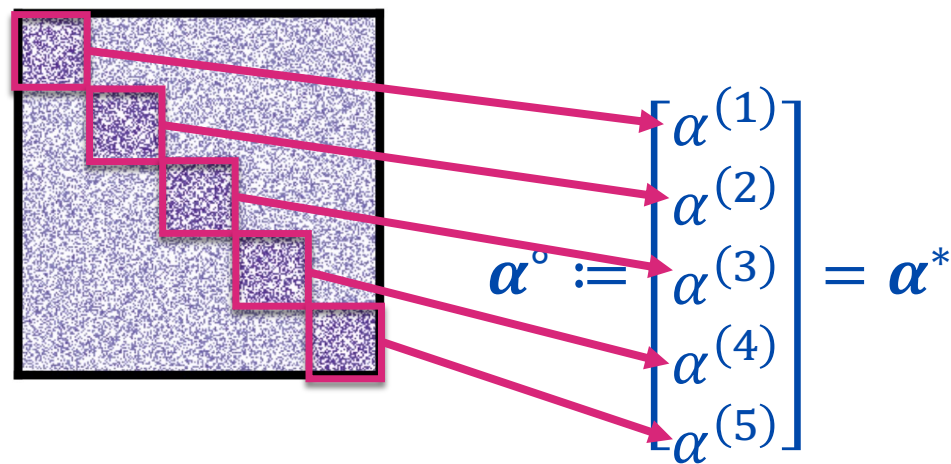
- $A^*$  has **no false connections**  
 $\Leftrightarrow K_{ij} \leq \alpha_i^* + \alpha_j^*$  for inter cluster connections  $i, j$
- $A^*$  is **well connected**  $\Leftrightarrow K_{ij} > \alpha_i^* + \alpha_j^*$  for intra cluster connections  $i, j$
- **Problem:** Hard to bound  $\alpha^*$  due to (#) coupling among **different** clusters



# Theorem: Decoupling is equivalent to NFC

Given input affinity  $\mathbf{K} = \mathbf{K}^\top \in \mathbb{R}^{n \times n}$  and sparsity parameter  $\eta$ ,

- $\mathbf{A}^*$  has no false connections  $\Leftrightarrow$



- $\alpha^{(l)}$  is a solution to (#) of the submatrix of intra-cluster connections of cluster  $l$

✓ Bounding  $\alpha^\circ$  is much easier

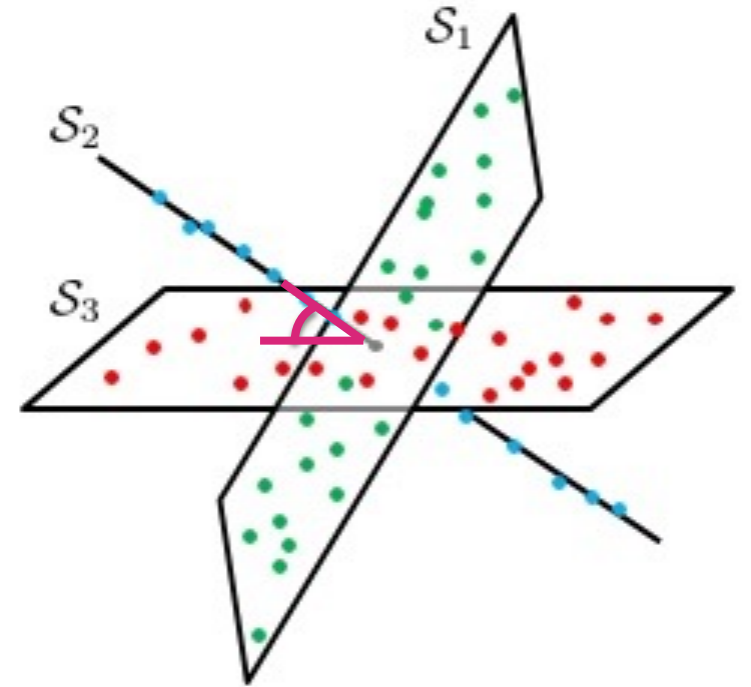
# Theorem: Subspace Clustering (Informal)

- Subspace clustering
  - clustering data from a union of low-dimensional linear subspaces
  - each subspace defines a cluster

✓ We prove:  $A^*$  has no false connection if the subspaces

- are sufficiently separated in angle, or
- have sufficiently low dimensions, or
- are well balanced in terms of number of points.

✓ We also guarantee connectivity!



# More Information

Research supported by NSF grant 1704458 and Northrop Grumman Mission Systems Research in Applications for Learning Machines (REALM) initiative.

Vision Lab @ JHU  
<http://www.vision.jhu.edu>

Center for Imaging Science @ JHU  
<http://www.cis.jhu.edu>

Mathematical Institute for Data Science @ JHU  
<http://www.minds.jhu.edu>

# Thank You!