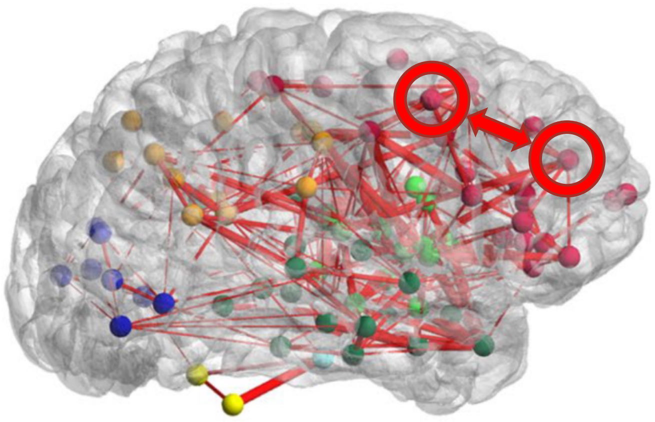
GraphMAD: Graph Mixup for Data Augmentation using Data-Driven Convex Clustering

Madeline Navarro and Santiago Segarra
Department of Electrical and Computer Engineering
Rice University

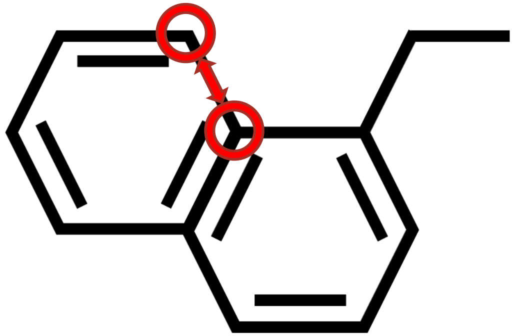
13 Jun 2023
Graph Signal Processing Workshop 2023

Contact:
Email: nav@rice.edu

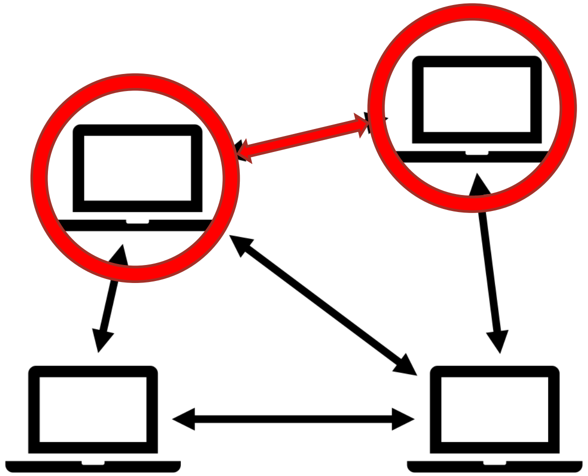
Graph classification required in many fields



Brain connectivity
S.-H. Chu, K. K. Parhi, C. Lenglet,
Scientific Reports 2018

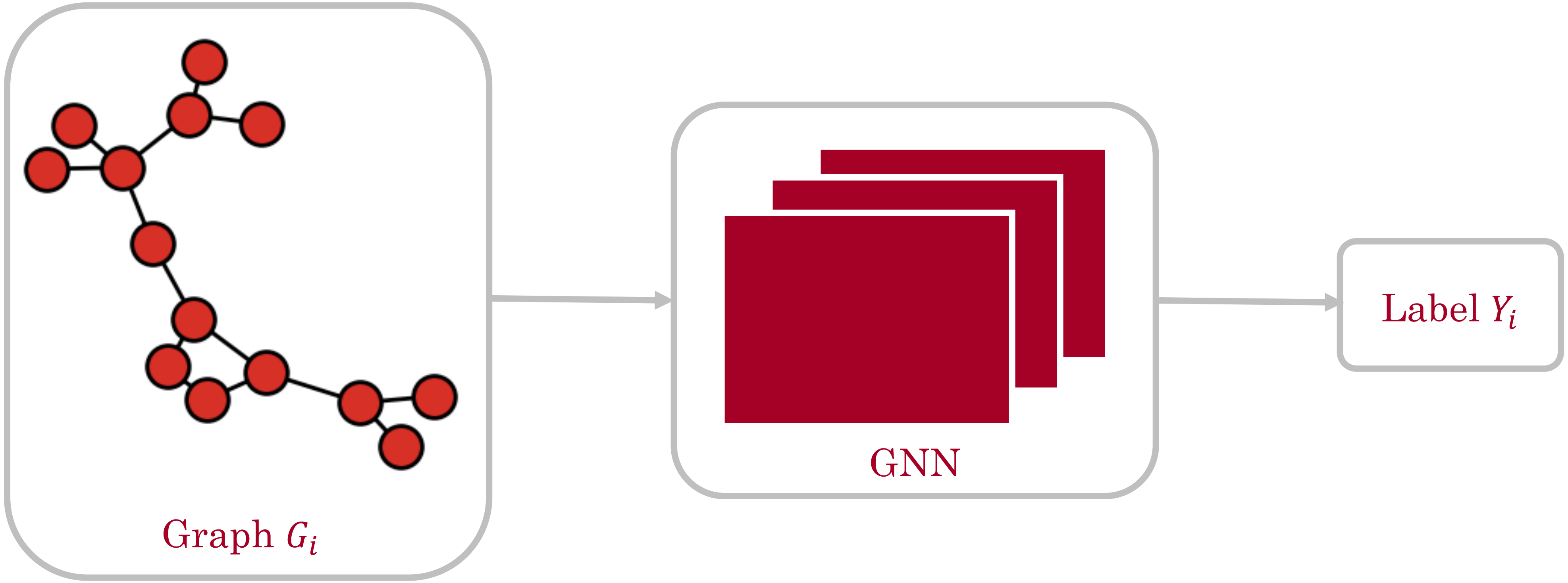


Drug discovery
S.-H. Chu, K. K. Parhi, C. Lenglet,
Scientific Reports 2018



Wireless Communication
A. Chowdhury *et al.*,
IEEE Trans. Wireless Commun. 2021

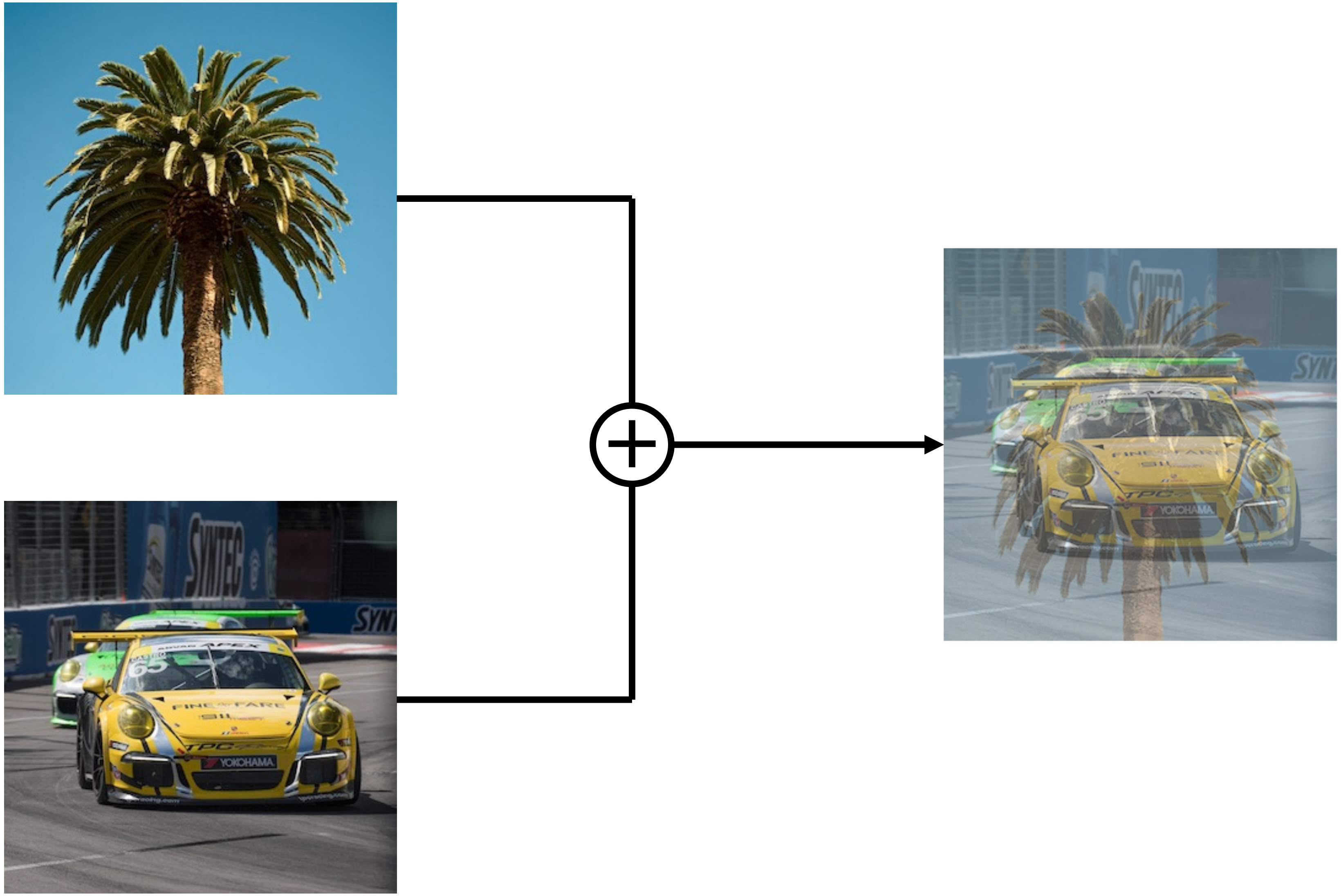
Graph Neural Networks (GNNs) highly successful for graph classification



Mixup for data augmentation via linear combinations of data pairs

Label
Tree: 1
Car: 0

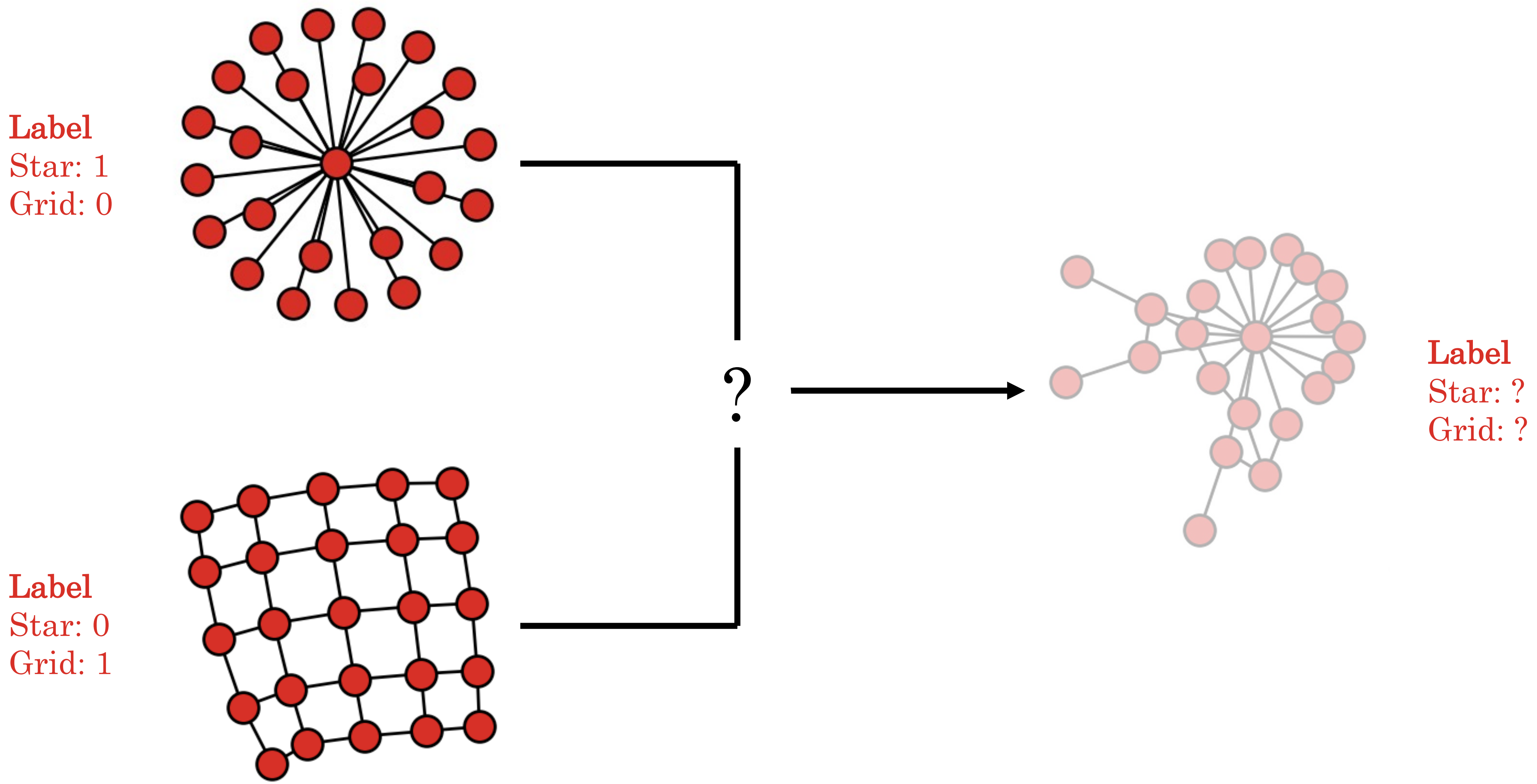
Label
Tree: 0
Car: 1



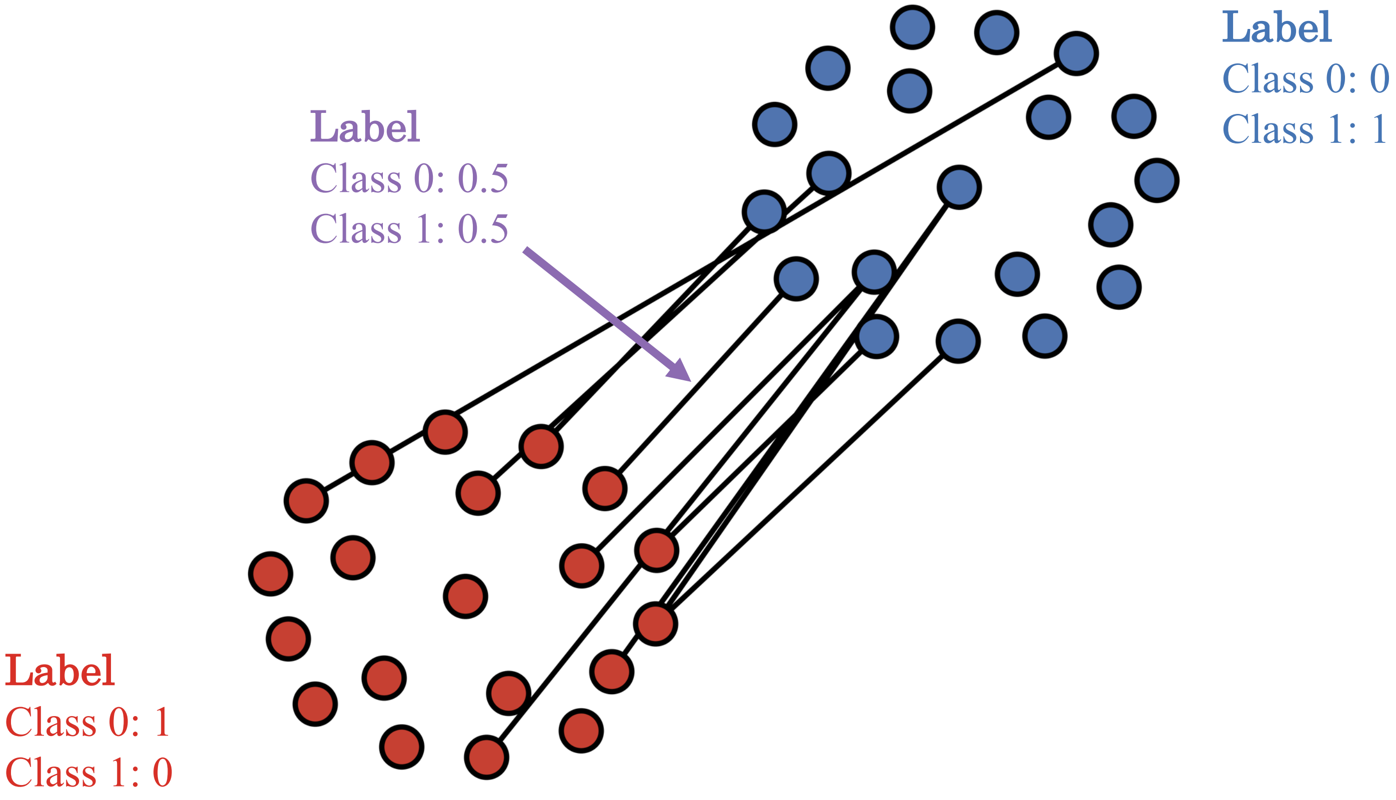
Label
Tree: 0.5
Car: 0.5

H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, ICLR 2017

Non-Euclidean graph data is difficult to mixup

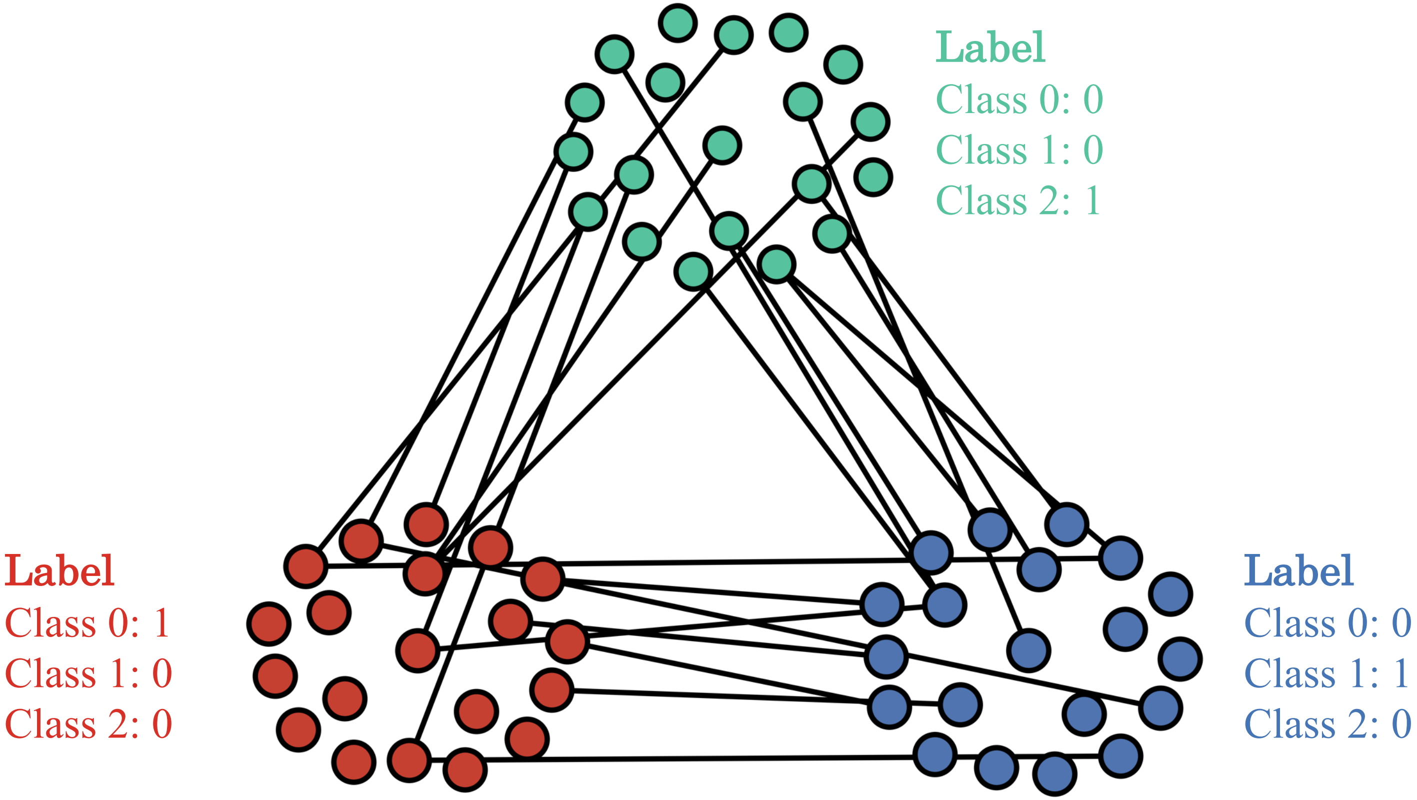


Pairwise linear mixup may ignore useful regions



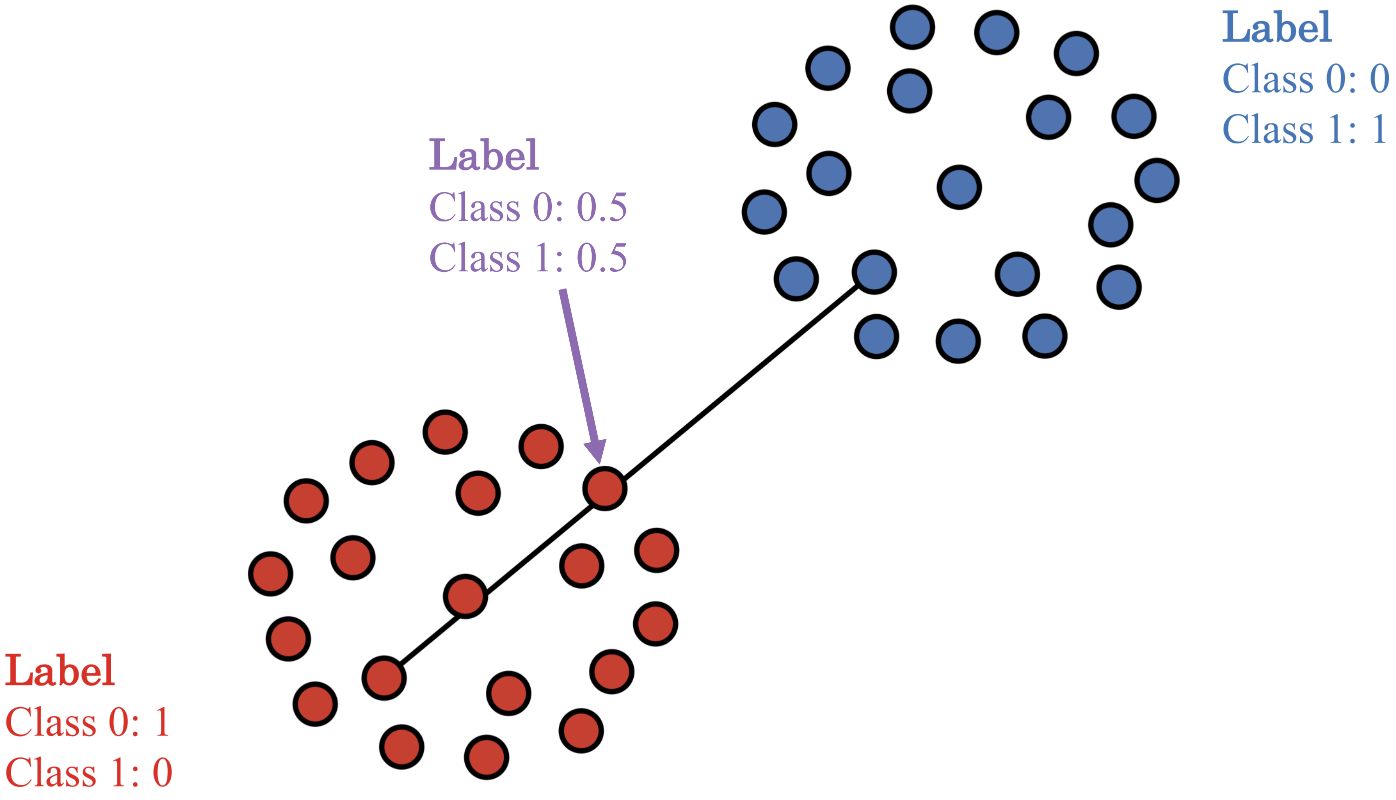
Pairwise mixup only considers space between pairs of classes

Pairwise linear mixup may ignore useful regions



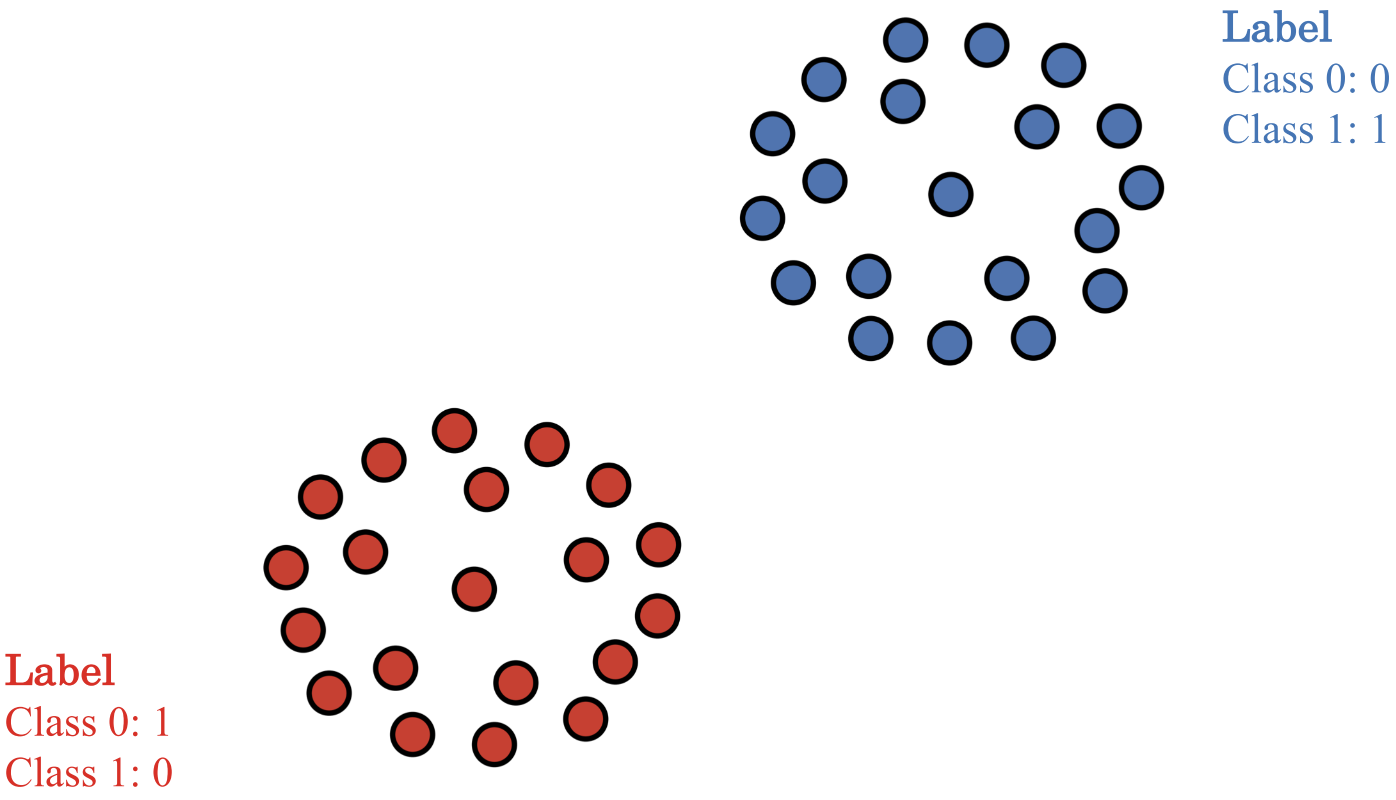
Pairwise mixup only considers space between pairs of classes

Traditional mixup only considers linear behavior between pairs of samples



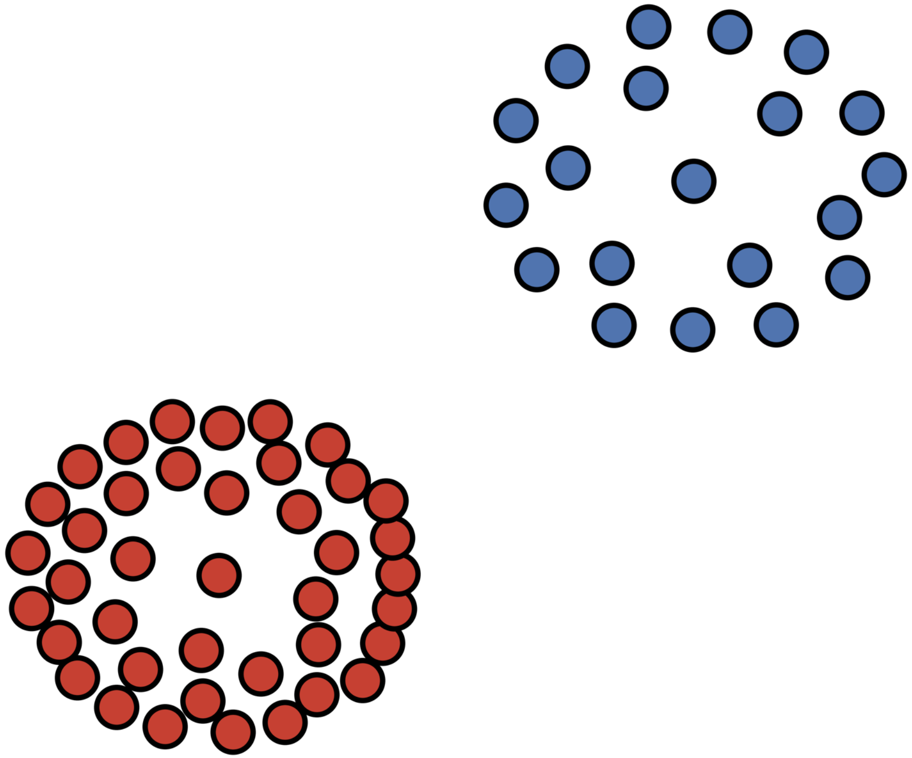
Linear mixup may add uncertainty in ways that are unhelpful

Traditional mixup only considers linear behavior between pairs of samples



Pairwise mixup ignores most of the dataset when mixing two samples

Traditional mixup only considers linear behavior between pairs of samples

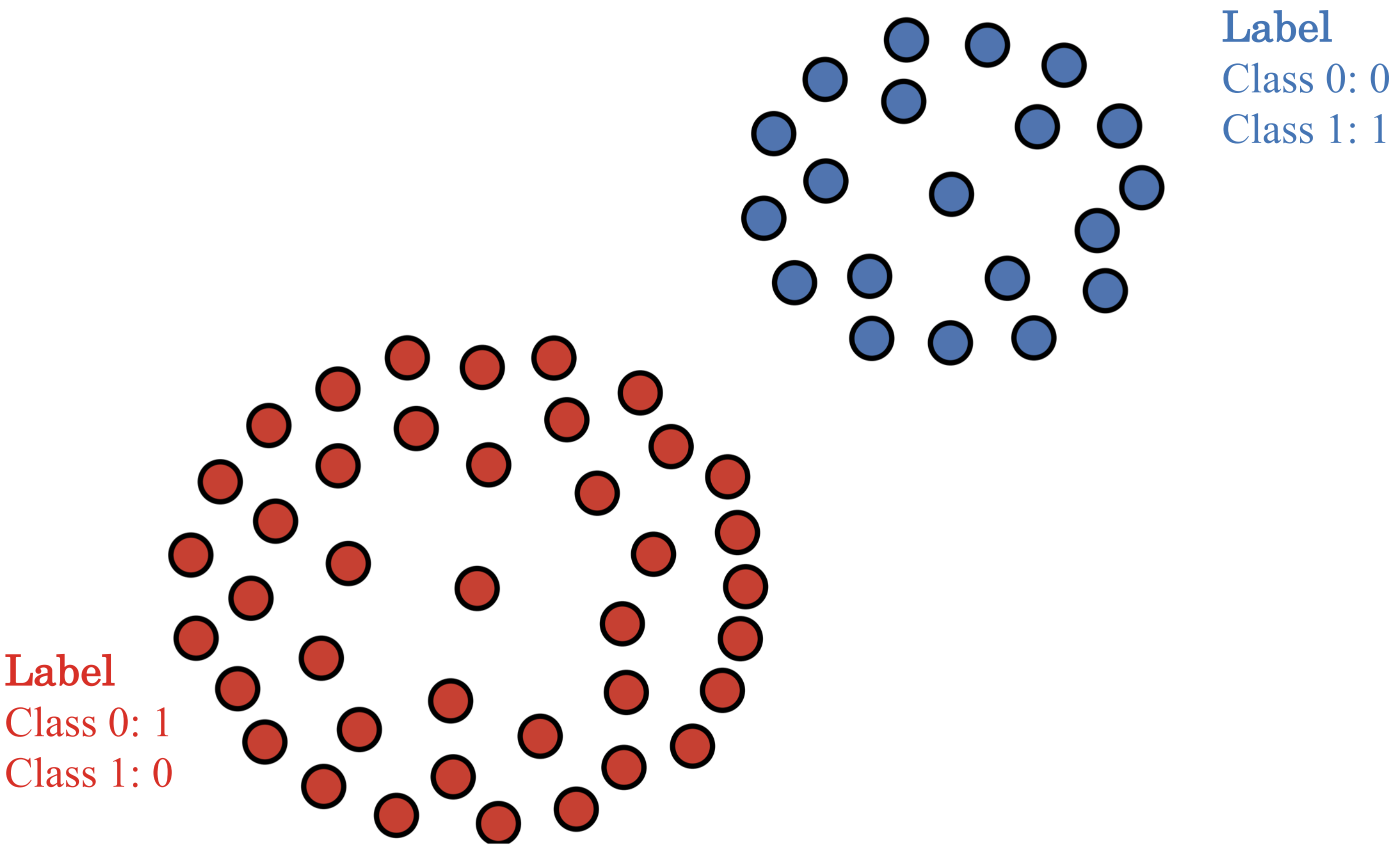


Label
Class 0: 0
Class 1: 1

Label
Class 0: 1
Class 1: 0

Pairwise mixup ignores most of the dataset when mixing two samples

Traditional mixup only considers linear behavior between pairs of samples



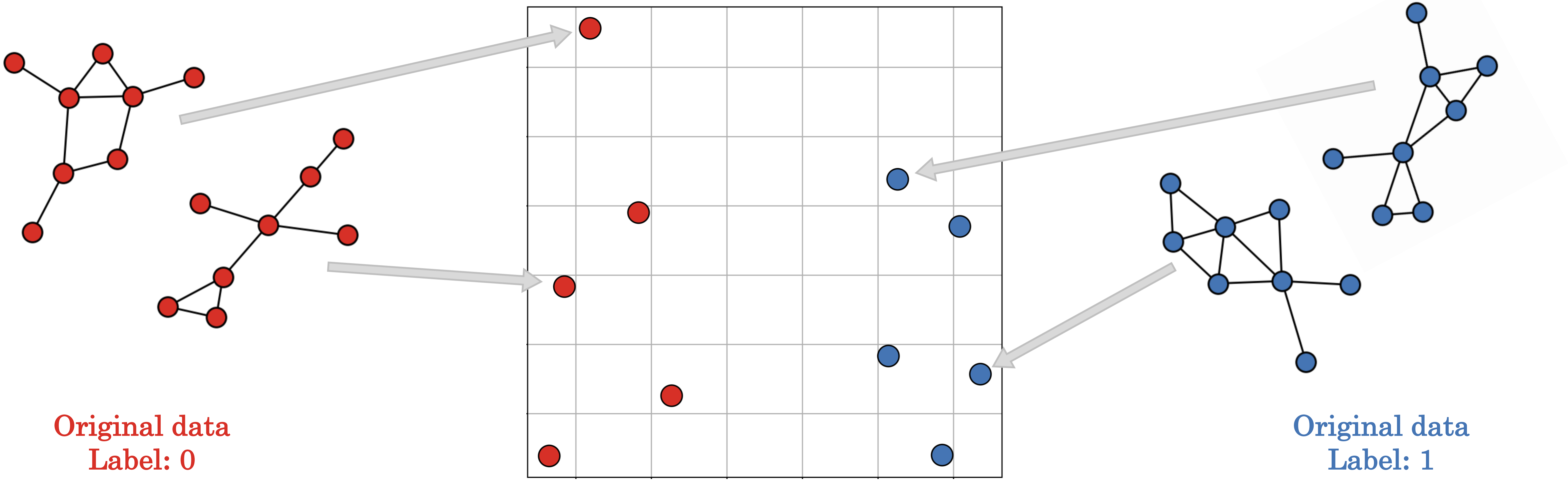
Pairwise mixup ignores most of the dataset when mixing two samples

What kind of new samples should we add to the dataset?

Limitations of pairwise linear mixup:

- Convex combinations are nontrivial for graph data
- Pairwise mixup may ignore useful sample regions
- Linear mixup may add uncertainty in ways that are unhelpful
- Sampling for linear mixup does not take the data into account

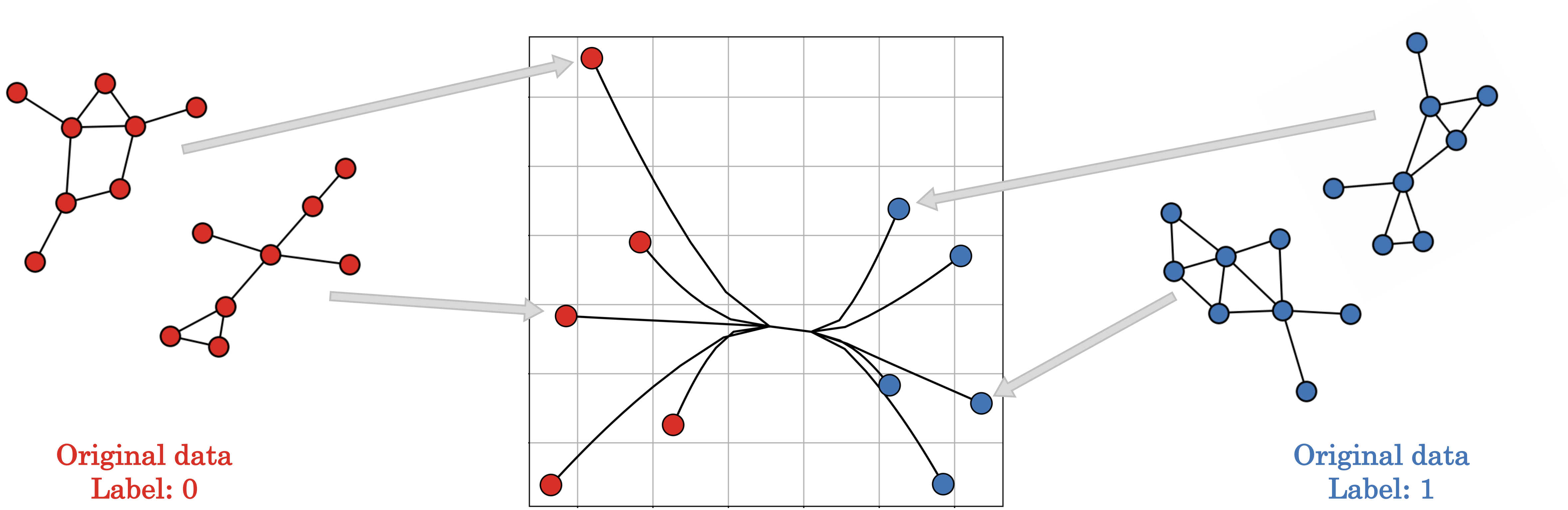
Graph Mixup for Augmenting Data (GraphMAD) – Concept



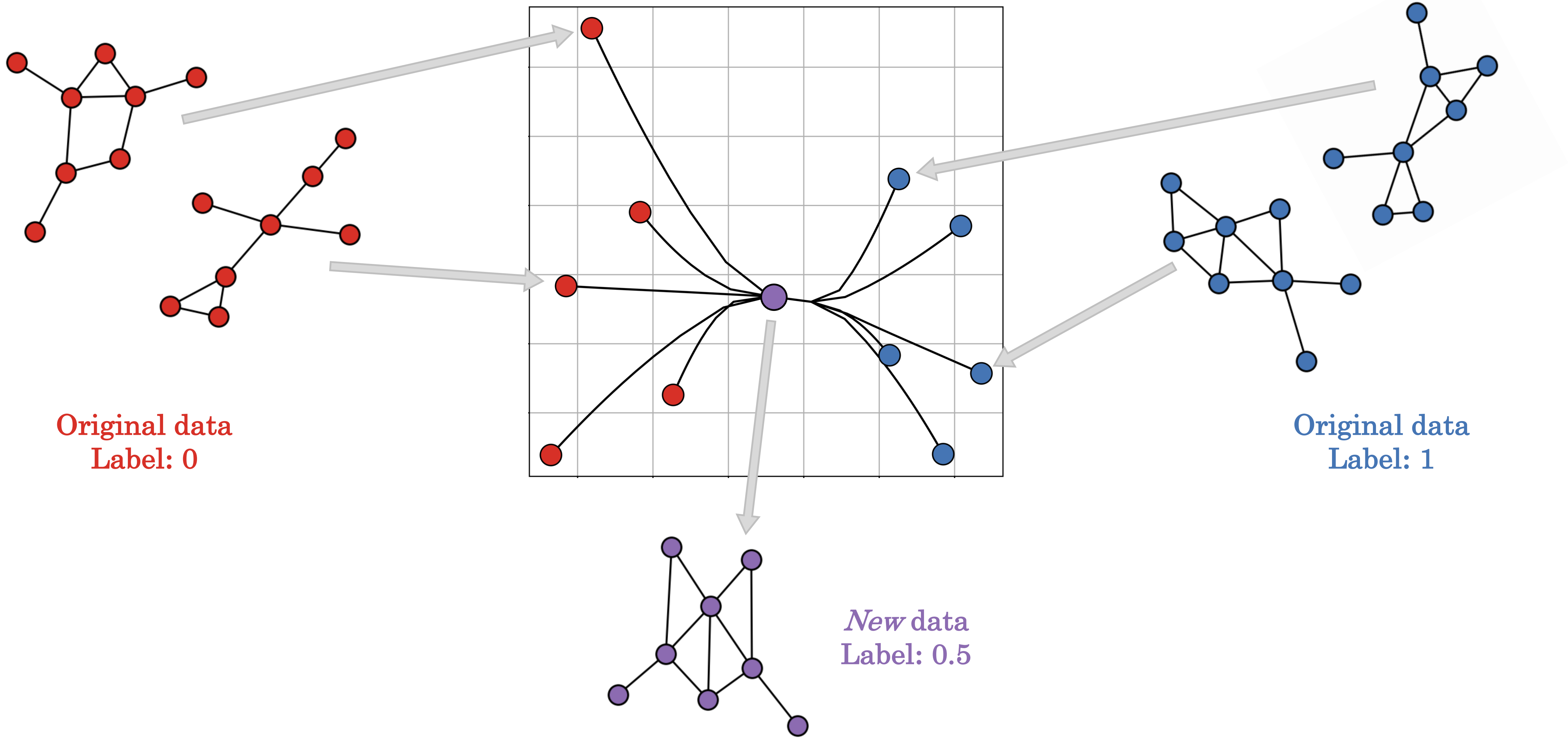
Original data
Label: 0

Original data
Label: 1

Graph Mixup for Augmenting Data (GraphMAD) – Concept



Graph Mixup for Augmenting Data (GraphMAD) – Concept



Graph Mixup for Augmenting Data (GraphMAD) – Details

Given set of labeled graphs $\{(G_i, \mathbf{y}_i)\}_{i=1}^{T}$:

Step 1: Convert each graph G_i to a continuous descriptor $\boldsymbol{\theta}(G_i)$

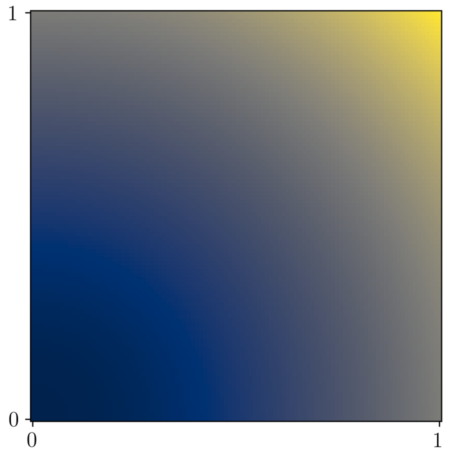
Step 2: Obtain nonlinear mixup function $\hat{\mathbf{u}}(\lambda)$ from descriptors $\{(\boldsymbol{\theta}(G_i), \mathbf{y}_i)\}_{i=1}^{T}$

Step 3: Sample new graphs G_i^{new} from a point in mixup function $\boldsymbol{\theta}^{\text{new}} = \hat{\mathbf{u}}(\lambda)$

Then perform graph classification on original dataset + new labeled graphs

GraphMAD Step 1: Convert each graph to graphon

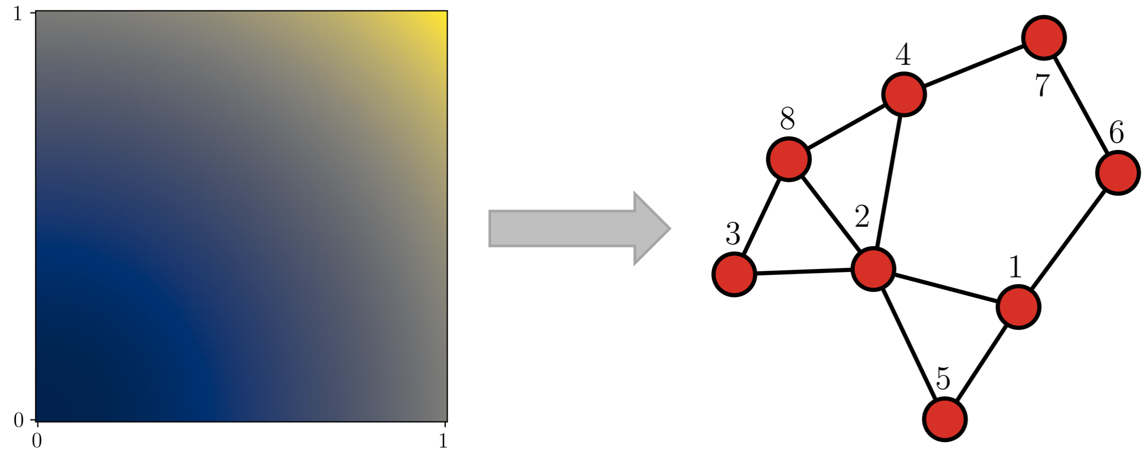
Convert each graph G_i to a continuous descriptor $\boldsymbol{\theta}(G_i)$, *graphon*, a bounded, continuous symmetric function



Graphon
$\mathcal{W}: [0,1]^2 \rightarrow [0,1]$

GraphMAD Step 1: Convert each graph to graphon

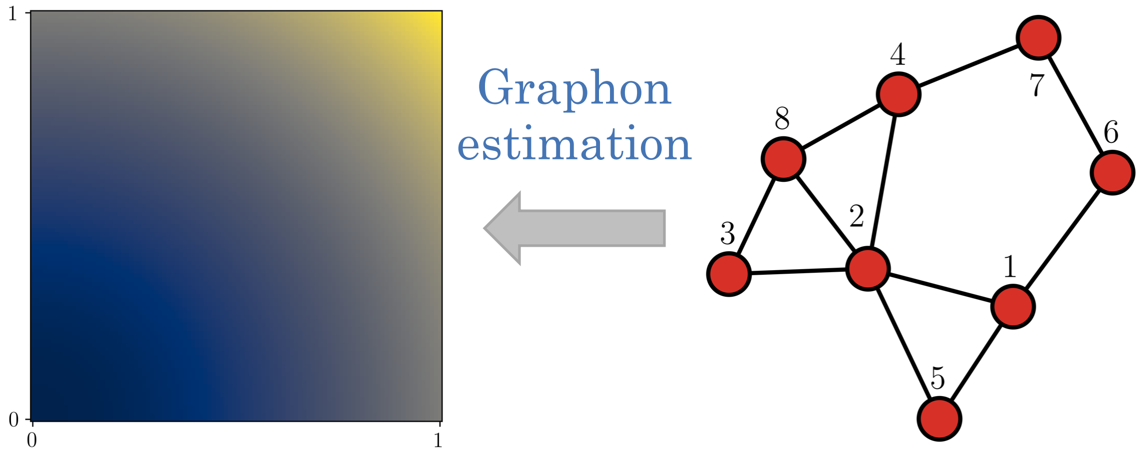
Convert each graph G_i to a continuous descriptor $\boldsymbol{\theta}(G_i)$, *graphon*, a bounded, continuous symmetric function



Graphon
$\mathcal{W}: [0,1]^2 \to [0,1]$

Sampled graph
$G \sim \mathcal{W}$

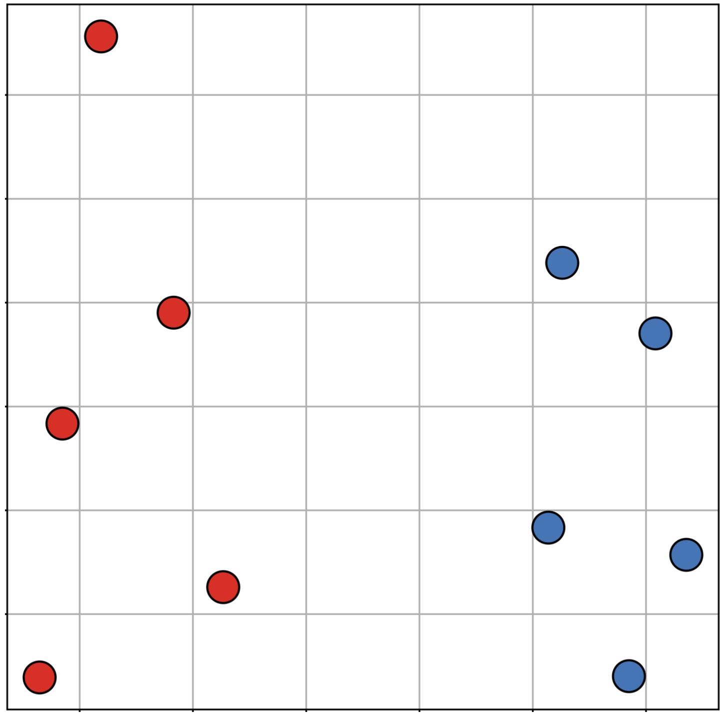
Convert each graph G_i to a continuous descriptor $\boldsymbol{\theta}(G_i)$, *graphon*, a bounded, continuous symmetric function

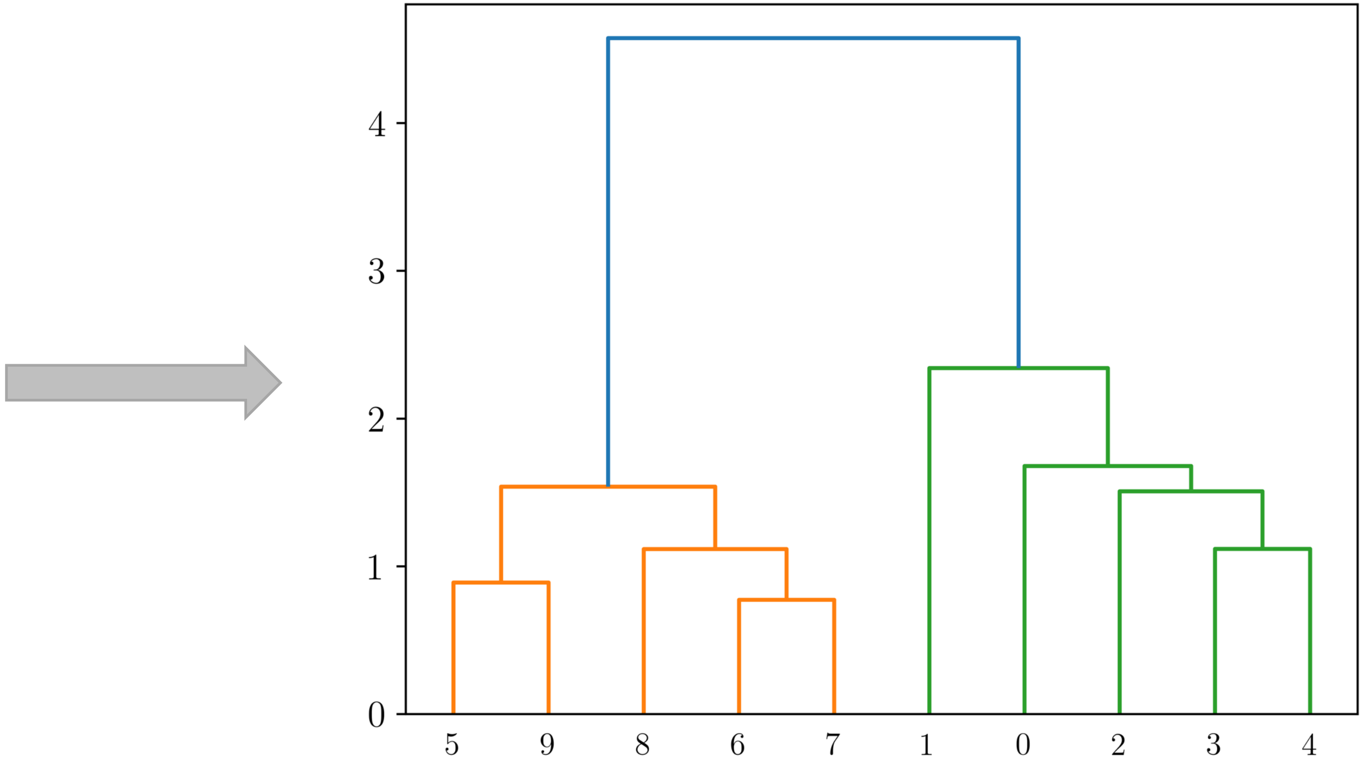


Estimated graphon
$\widehat{\mathcal{W}}: [0,1]^2 \rightarrow [0,1]$

Sampled graph
$G \sim \mathcal{W}$

GraphMAD Step 2: Nonlinear mixup function from graphons using convex clustering





Clustering methods such as hierarchical clustering use relationships among data to assign data to groups

GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering

Convex clustering tradeoff between fusing clusters and fitting to samples

K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, PASCAL 2005

GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering

$$\{\hat{\mathbf{u}}_j(\lambda)\}_{j=1}^{T} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{j=1}^{T} \left\| \mathbf{u}_j - \boldsymbol{\theta}(G_j) \right\|_2^2 + \frac{\lambda}{1-\lambda} \sum_{i<j} w_{ij} \left\| \mathbf{u}_i - \mathbf{u}_j \right\|_1$$

- $\boldsymbol{\theta}(G_j)$: Each graphon
- $\hat{\mathbf{u}}_j(\lambda)$: Cluster centroid for each graphon at $\lambda \in [0,1]$

Convex clustering tradeoff between fusing clusters and fitting to samples

K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, PASCAL 2005

GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering

$$\{\hat{\mathbf{u}}_j(\lambda)\}_{j=1}^{T} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{j=1}^{T} \left\| \mathbf{u}_j - \boldsymbol{\theta}(G_j) \right\|_2^2 + \frac{\lambda}{1-\lambda} \sum_{i<j} w_{ij} \left\| \mathbf{u}_i - \mathbf{u}_j \right\|_1$$

- $\boldsymbol{\theta}(G_j)$: Each graphon
- $\hat{\mathbf{u}}_j(\lambda)$: Cluster centroid for each graphon at $\lambda \in [0,1]$
- λ: Fusion parameter

Convex clustering tradeoff between fusing clusters and fitting to samples

K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, PASCAL 2005

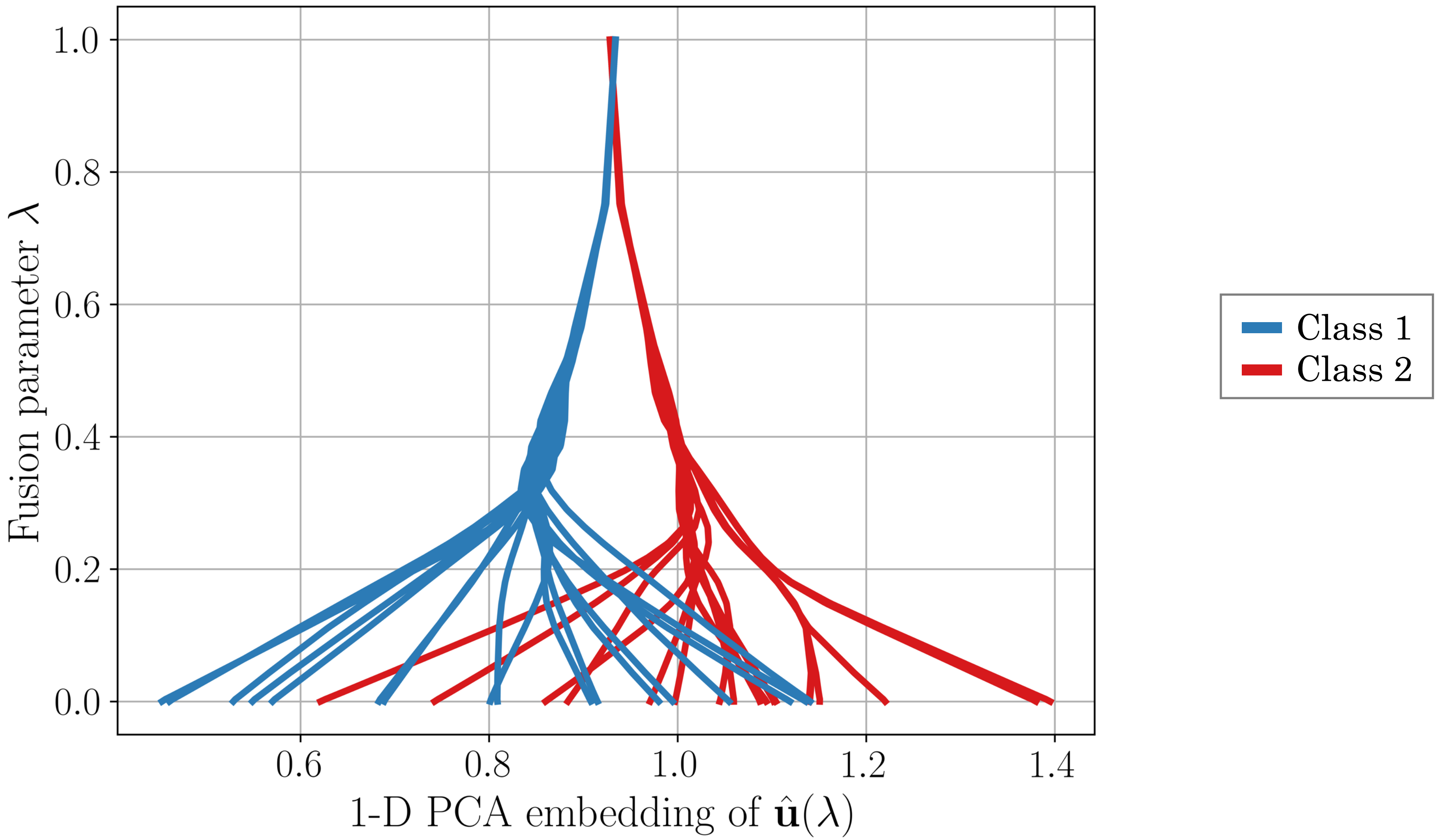
GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering

$$\{\hat{\mathbf{u}}_j(\lambda)\}_{j=1}^{T} = \underset{\mathbf{u}}{\operatorname{argmin}} \sum_{j=1}^{T} \left\| \mathbf{u}_j - \boldsymbol{\theta}(G_j) \right\|_2^2 + \frac{\lambda}{1-\lambda} \sum_{i<j} w_{ij} \left\| \mathbf{u}_i - \mathbf{u}_j \right\|_1$$

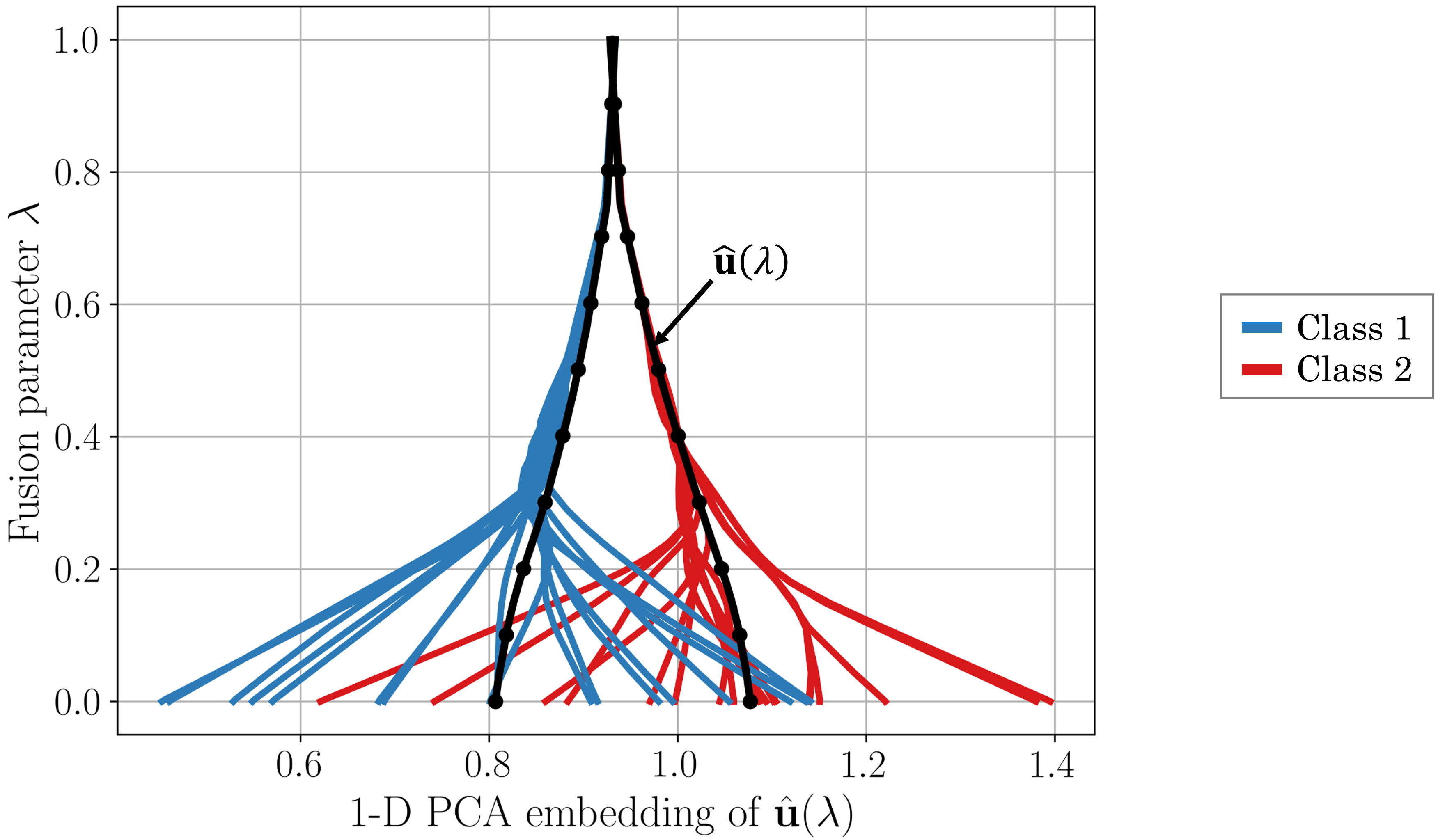
- λ tunes between original dataset and total fusion (dataset mean)
  - $\lambda = 0$: T singleton clusters
  - $\lambda \in (0,1)$: Data samples begin to fuse into clusters
  - $\lambda = 1$: All samples in one cluster

Convex clustering tradeoff between fusing clusters and fitting to samples

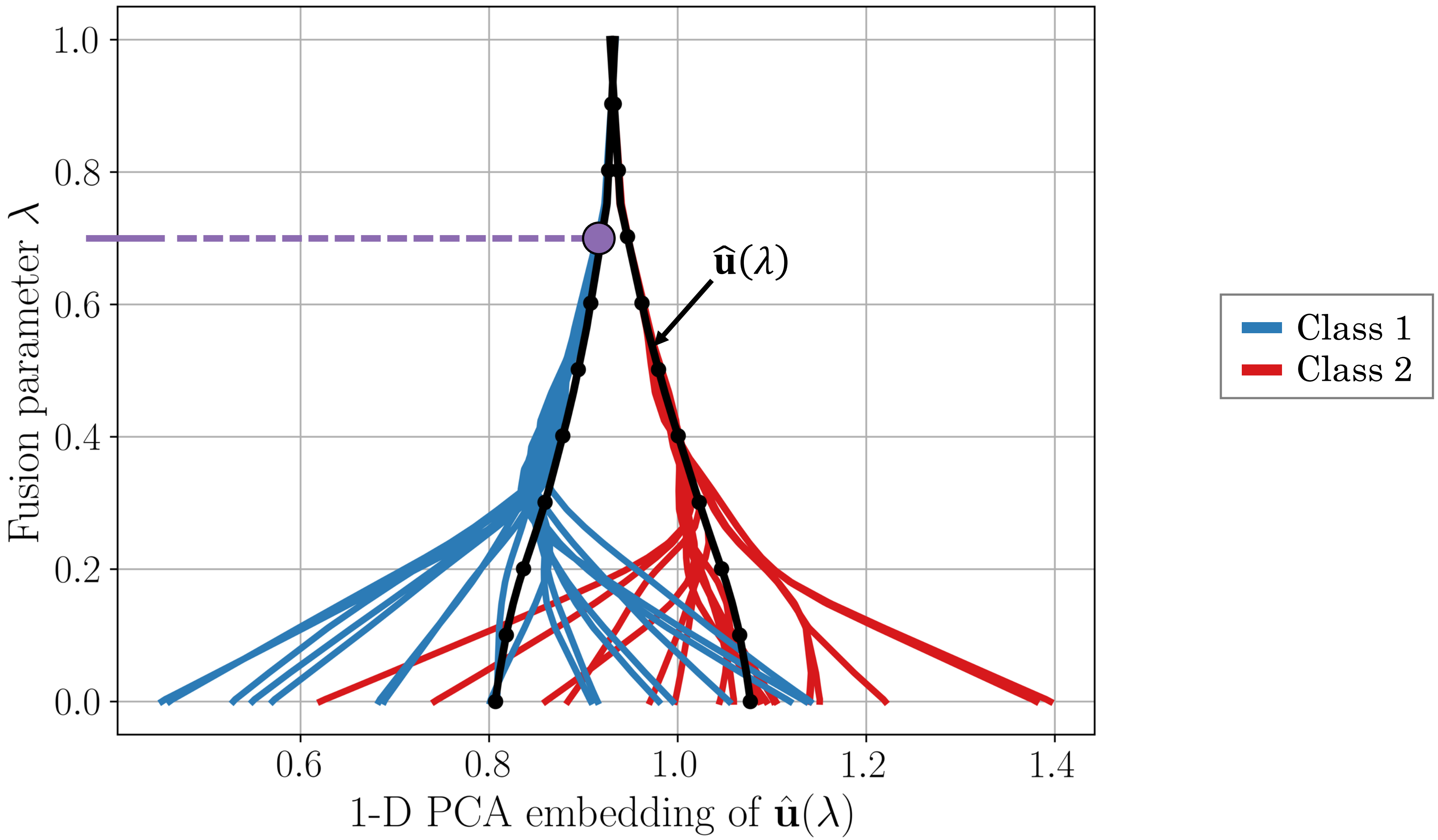
K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, PASCAL 2005


Fusion parameter λ
1.0
0.8
0.6
0.4
0.2
0.0
0.6
0.8
1.0
1.2
1.4
1-D PCA embedding of $\hat{\mathbf{u}}(\lambda)$
Class 1
Class 2

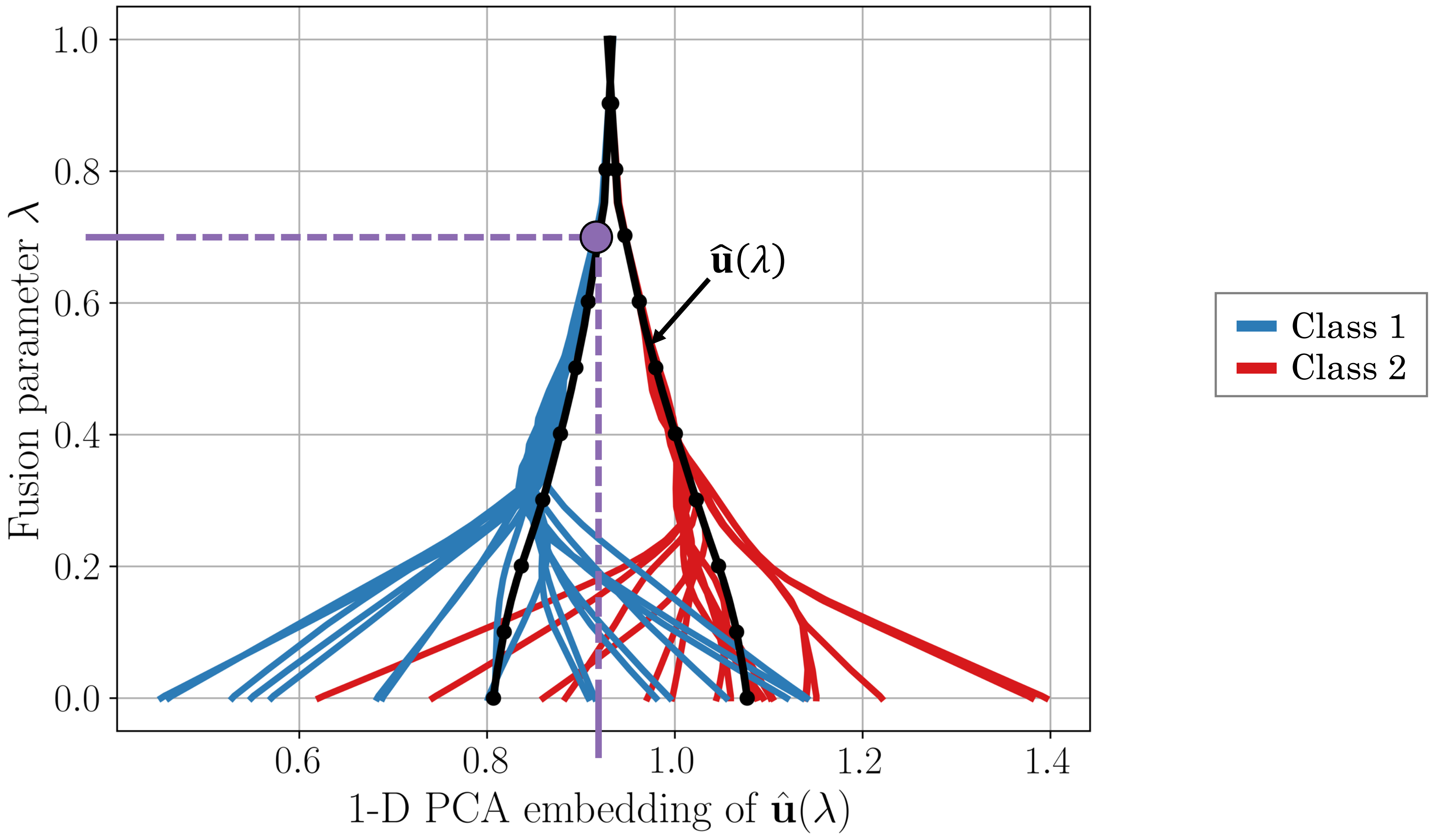
GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering



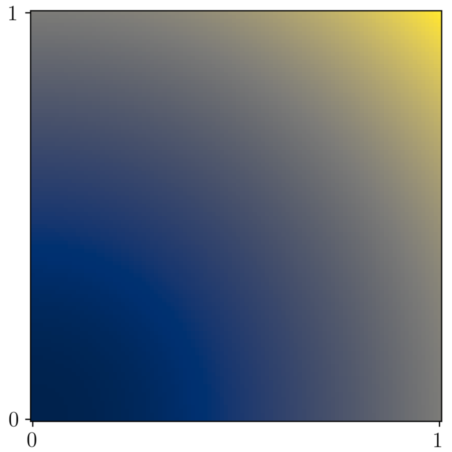
GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering



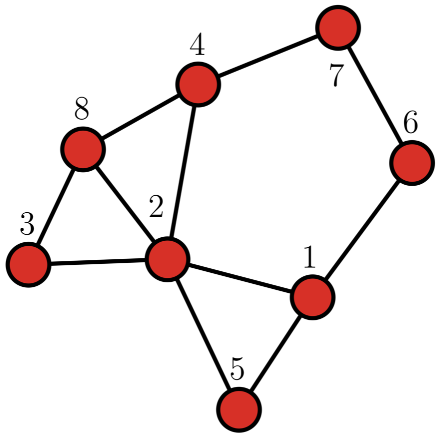
GraphMAD Step 2: Compute nonlinear mixup function from graphons using convex clustering



GraphMAD Step 3: Sample new graphs given mixup function

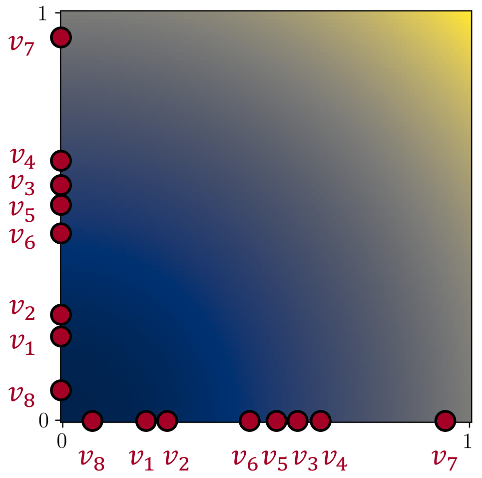


Graphon
$\mathcal{W}: [0,1]^2 \rightarrow [0,1]$



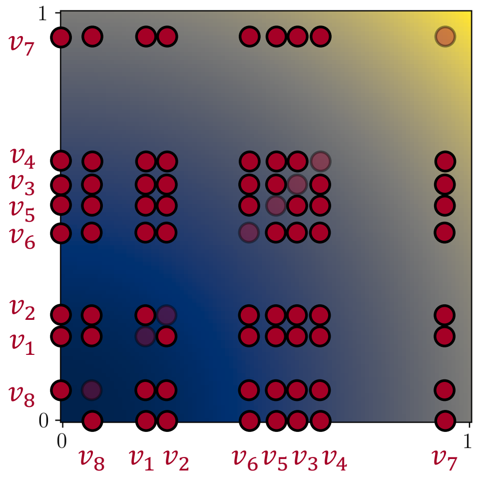
Sampled graph
$G \sim \mathcal{W}$

GraphMAD Step 3: Sample new graphs given mixup function



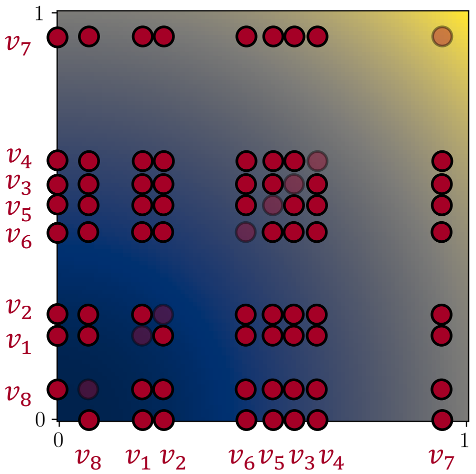
Graphon
$\mathcal{W}: [0,1]^2 \to [0,1]$

GraphMAD Step 3: Sample new graphs given mixup function

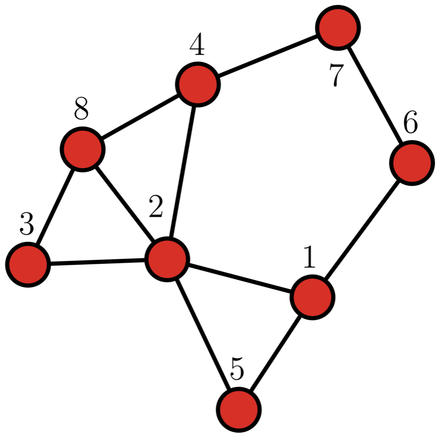


Graphon
$\mathcal{W}: [0,1]^2 \to [0,1]$

GraphMAD Step 3: Sample new graphs given mixup function



Graphon
$\mathcal{W}: [0,1]^2 \rightarrow [0,1]$



Sampled graph
$G \sim \mathcal{W}$

GraphMAD improves performance and outperforms linear mixup on all datasets

Graph classification accuracy on molecule and bioinformatics datasets

Method	
Data mixup	**Label mixup**
None	None
Linear	Linear
	Sigmoid
	Logit
	Clusterpath
Clusterpath	Linear
	Sigmoid
	Logit
	Clusterpath

GraphMAD improves performance and outperforms linear mixup on all datasets

Graph classification accuracy on molecule and bioinformatics datasets

Method		DD	PROTEINS	ENZYMES	AIDS	MUTAG	NCI109
Data mixup	**Label mixup**	**2 classes**	**2 classes**	**6 classes**	**2 classes**	**2 classes**	**2 classes**
None	None						
Linear	Linear						
	Sigmoid						
	Logit						
	Clusterpath						
Clusterpath	Linear						
	Sigmoid						
	Logit						
	Clusterpath						

GraphMAD improves performance and outperforms linear mixup on all datasets

Graph classification accuracy on molecule and bioinformatics datasets

Method		DD	PROTEINS	ENZYMES	AIDS	MUTAG	NCI109
Data mixup	Label mixup	2 classes	2 classes	6 classes	2 classes	2 classes	2 classes
None	None	68.77 ± 2.35	**69.51 ± 1.20**	26.43 ± 2.55	96.18 ± 2.57	84.59 ± 5.53	68.23 ± 2.13
Linear	Linear	67.01 ± 1.72	65.15 ± 2.53	24.88 ± 3.38	96.82 ± 1.39	85.71 ± 7.15	68.16 ± 2.72
	Sigmoid	64.89 ± 1.49	68.42 ± 3.94	24.76 ± 4.10	96.07 ± 1.42	85.71 ± 4.63	65.96 ± 2.34
	Logit	66.22 ± 3.82	69.25 ± 2.94	25.95 ± 5.48	96.07 ± 1.27	80.08 ± 5.60	66.81 ± 4.07
	Clusterpath	68.22 ± 3.71	69.38 ± 2.04	24.64 ± 2.39	95.86 ± 1.88	**87.22 ± 4.96**	65.01 ± 3.07
Clusterpath	Linear	67.11 ± 1.56	67.51 ± 2.62	**26.67 ± 6.49**	**97.15 ± 1.00**	**87.24 ± 4.21**	**68.61 ± 1.41**
	Sigmoid	68.23 ± 3.61	64.60 ± 5.07	**32.62 ± 6.35**	97.07 ± 1.35	85.20 ± 3.53	67.50 ± 2.06
	Logit	**70.07 ± 2.51**	67.26 ± 2.84	25.71 ± 4.26	95.87 ± 1.47	80.10 ± 14.77	65.33 ± 3.35
	Clusterpath	**70.44 ± 3.79**	**71.18 ± 3.98**	24.52 ± 3.30	**97.22 ± 0.54**	85.71 ± 5.40	**68.54 ± 3.16**

Data augmentation with GraphMAD consistently outperforms linear mixup,
and different label mixup functions can improve accuracy

GraphMAD improves performance and outperforms linear mixup on all datasets

Graph classification accuracy on social datasets

Method		COLLAB	IMDB-B	IMDB-M
Data mixup	**Label mixup**	**3 classes**	**2 classes**	**3 classes**
None	None	**80.00 ± 0.96**	73.14 ± 3.15	47.71 ± 4.25
Linear	Linear	77.60 ± 1.53	72.07 ± 2.06	47.24 ± 4.21
	Sigmoid	78.21 ± 1.16	**74.00 ± 2.14**	**49.67 ± 2.15**
	Logit	78.19 ± 1.61	72.64 ± 1.73	47.43 ± 2.45
	Clusterpath	78.41 ± 0.99	71.43 ± 3.25	47.29 ± 5.21
Clusterpath	Linear	78.93 ± 2.63	70.57 ± 4.89	45.52 ± 4.09
	Sigmoid	77.89 ± 1.30	**75.00 ± 5.13**	44.48 ± 2.78
	Logit	**80.39 ± 1.20**	73.43 ± 4.75	48.76 ± 2.43
	Clusterpath	79.55 ± 2.29	71.43 ± 4.72	**49.71 ± 4.33**

Data augmentation with GraphMAD consistently outperforms linear mixup,
and different label mixup functions can improve accuracy

GraphMAD provides data-driven graph generation with learned regions of interest

Conclusion

Nonlinear mixup that considers all samples instead of pairs

Uncertainty dependent on existing data instead of enforced linear behavior

GraphMAD provides data-driven graph generation with learned regions of interest

Conclusion

Nonlinear mixup that considers all samples instead of pairs

Uncertainty dependent on existing data instead of enforced linear behavior

Next steps

Replace graphons as descriptors with learned descriptors

Applicable beyond graphs: higher-order networks, image data, etc.