Graph Convolution for Semi-Supervised Classification: Improved Linear Separability and Out-of-Distribution Generalization

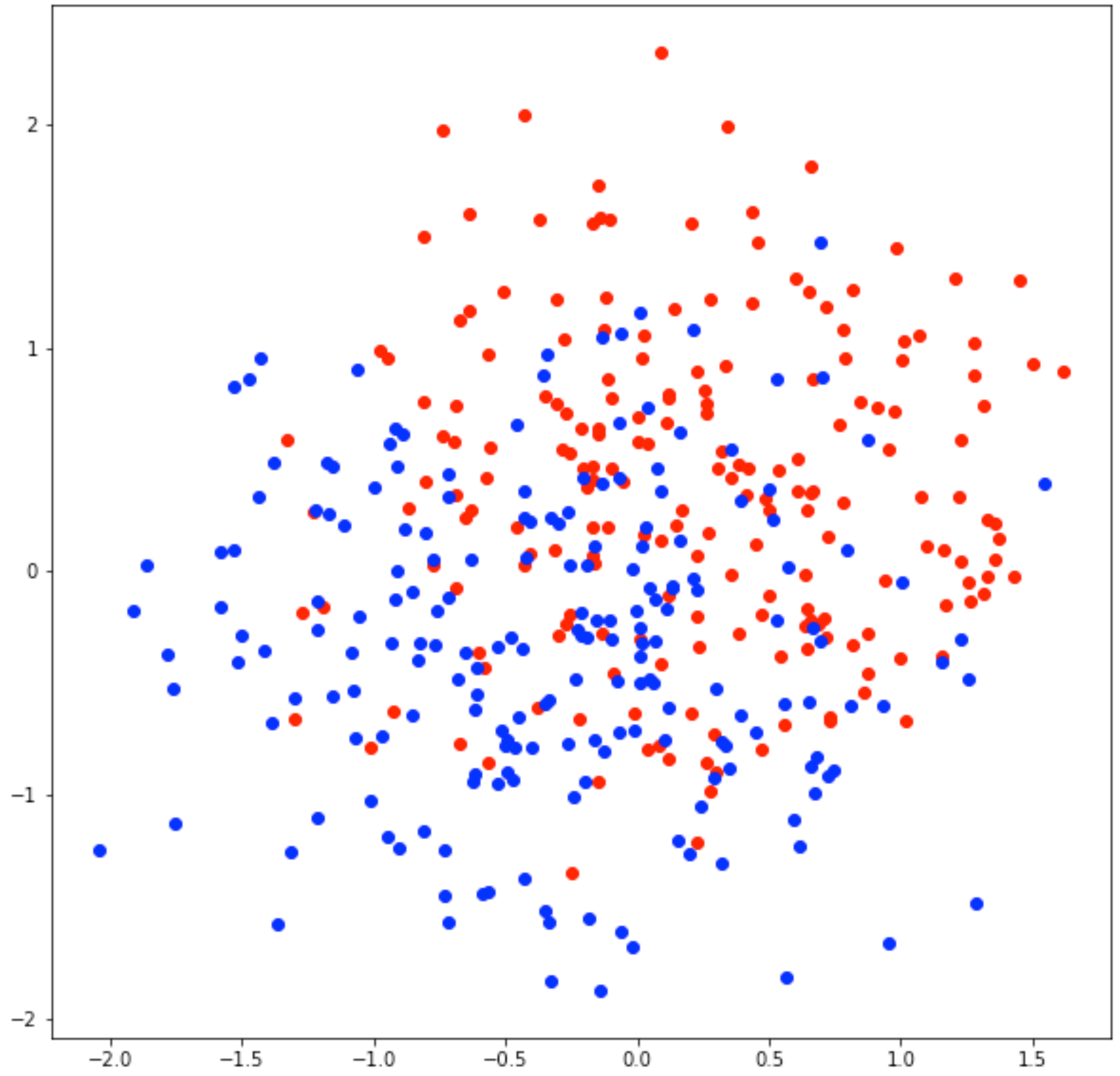
Aseem Baranwal, Kimon Fountoulakis, Aukosh Jagannath

UNIVERSITY OF WATERLOO | DAVID R. CHERITON SCHOOL OF COMPUTER SCIENCE

Contributions

- Study the effect of graph convolution on linear separability of a GMM
- Analyze the generalization potential of the linear classifier obtained by minimizing the cross-entropy loss
- Extensive experiments in various settings to illustrate our results

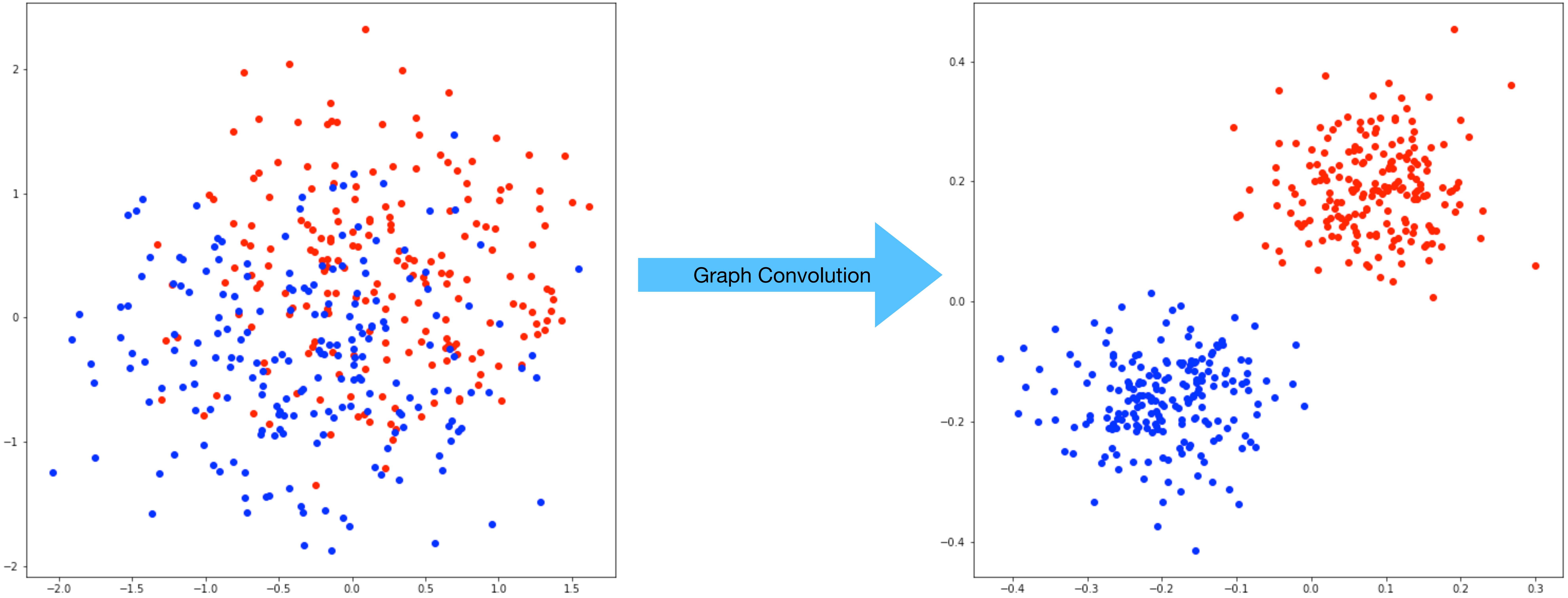
What can graph convolution do?



Original Data

- Consider distributions with 2D features
- We cannot separate this data linearly
- Can graph convolution help?

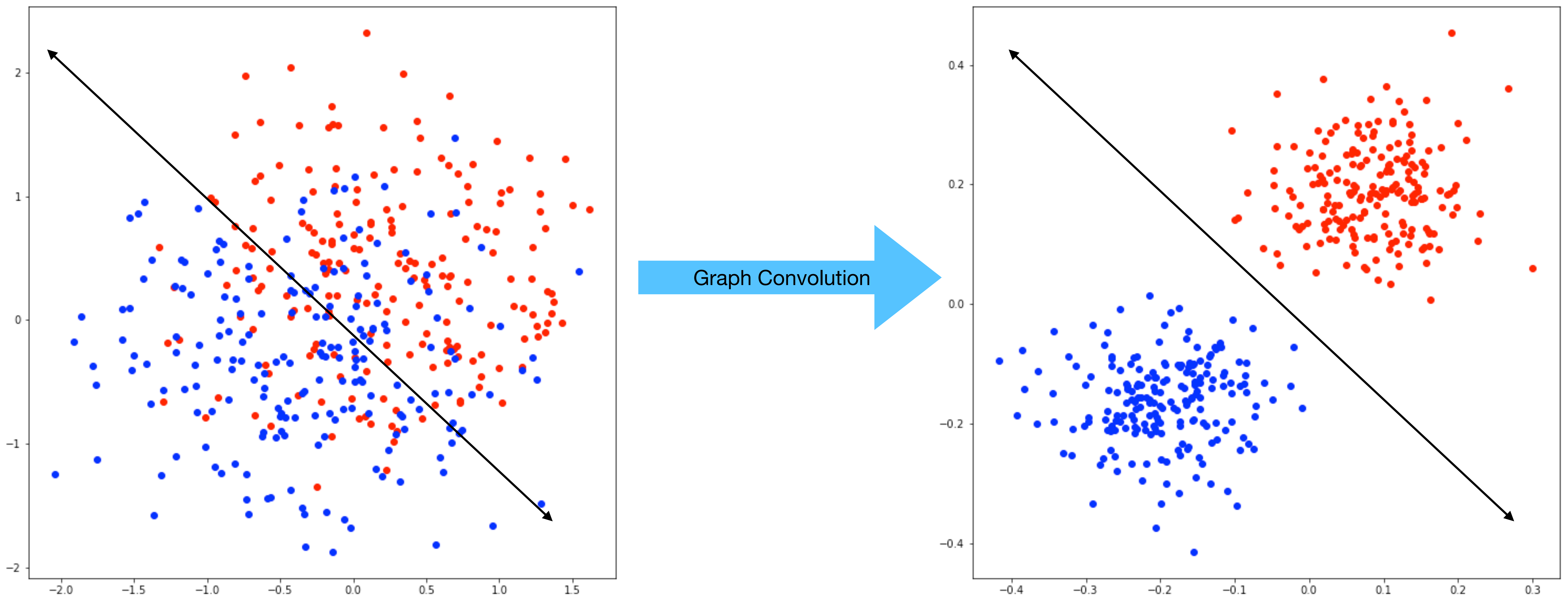
What can graph convolution do?



Original Data

After graph convolution

What can graph convolution do?



Original Data

After graph convolution

Graph convolution makes the data linearly separable

Model

- Two-component Gaussian Mixture Model (GMM) coupled with a Stochastic Block Model (SBM)

- Two classes C_0, C_1
  n data points with features $(X_i)_{i=1}^n \in \mathbb{R}^d$

- $X_i \sim \mathcal{N}(\mu, \sigma^2 I)$ if $i \in C_0$
  $X_i \sim \mathcal{N}(\nu, \sigma^2 I)$ if $i \in C_1$

- $A \sim SBM(p, q)$

$$\mathbb{P}(A_{ij} = 1) = \begin{cases} p & \text{if } i, j \text{ are in the same class} \\ q & \text{otherwise} \end{cases}$$

Graph improves linear separability

- Without the graph, no hyperplane can separate a binary GMM if means are $\mathcal{O}(\sigma)$ apart, i.e.,

$$\|\mu - \nu\|_2 = \mathcal{O}(\sigma)$$

- With graph convolution, this threshold changes to

$$\|\mu - \nu\| = \mathcal{O}\left(\frac{\sigma}{\sqrt{\mathbb{E}D}}\right)$$

Expected degree of a node

Bounds on training loss

- We use binary cross entropy loss to learn the classifier
- Without graph convolution, the loss is lower bounded by

$$Loss \geq (2 \log 2)\Phi\left(-\frac{\|\mu - \nu\|}{2\sigma}\right)$$

- When the convolved data is separable, the loss is upper bounded by

$$Loss(A, X) \leq C \exp\left(-d\|\mu - \nu\|\Gamma(p, q)\right),$$

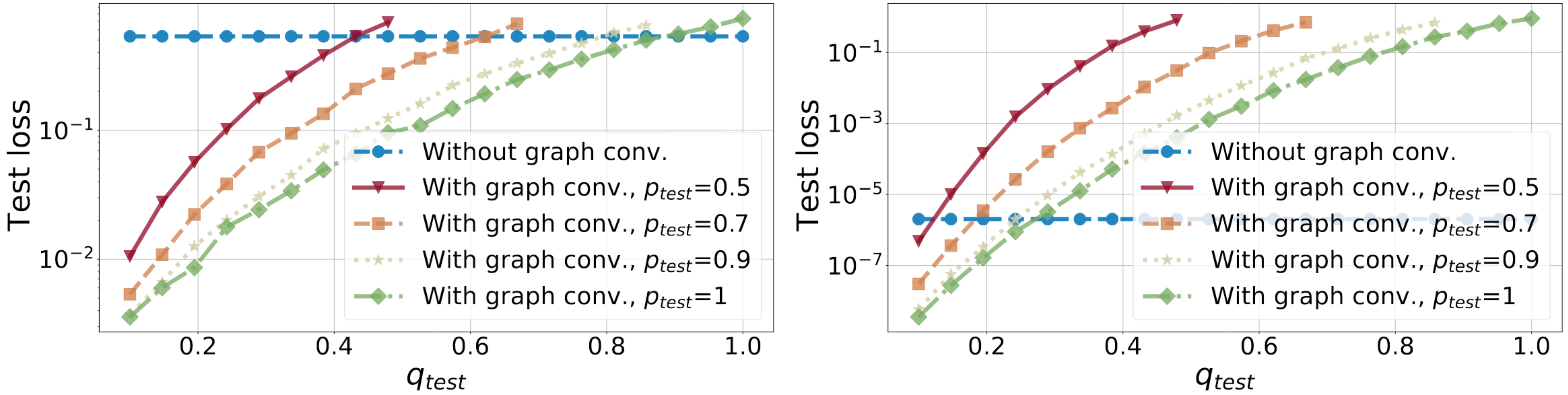
where $\Gamma(p, q) = \dfrac{p - q}{p + q}$

Generalization

- If the graph is not sparse, then for any new dataset A, X with different n, p, q, the loss is bounded above

$$Loss(A, X) \leq C \exp\left(-d\|\mu - \nu\|\Gamma(p, q)\right)$$

- Loss increases with inter-class edge probability (noisy graph)



(a) $\|\mu_d - \nu_d\| = 2/\sqrt{d}$ (b) $\|\mu_d - \nu_d\| = 16/\sqrt{d}$

What’s next?

- Is graph convolution helpful in the following settings:
    - Multiple classes — how does the distance between means matter here?
    - If the data is non-linearly separable — what does the optimal classifier look like?
- What about generalization in the above scenarios?
- Analysis of deeper GCNs