# Spectral-Inspired Graph Neural Networks

Teresa Huang

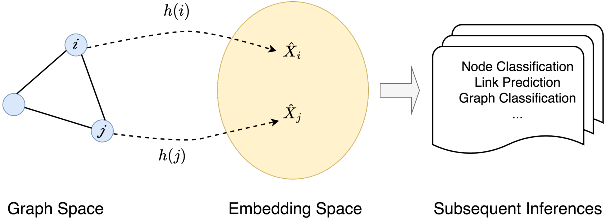
Applied Mathematics and Statistics
Mathematical Institute for Data Science

Mar 2022

*Joint work with Soledad Villar(JHU), Carey Priebe(JHU), Da Zheng(Amazon)*

JOHNS HOPKINS UNIVERSITY

# Graph (Network) Embedding




- Learn a low-dimensional representation for nodes in a given graph while preserving structural information
- Perform subsequent inferences directly on graph embedding

# Methods of Graph (Network) Embedding

- Spectral methods: global eigen-decomposition; unsupervised.
- Graph neural networks: local message passing; semi-supervised.

[2] Cape et al. 2019; Kipf et al. 2016.

# Methods of Graph (Network) Embedding

- Spectral methods: global eigen-decomposition; unsupervised.
- Graph neural networks: local message passing; semi-supervised.

**Definition (Adjacency Spectral Embedding (ASE) [1])**

$h_{ASE} = U_{d'}|\Sigma_{d'}|^{1/2}$, where $A = U\Sigma U^{\top}$, $d'$ is the embedding dimension.

**Definition (Graph Convolutional Network (GCN) [2])**

A $L$-layer GCN embedding is given by $h_{GCN} = \boldsymbol{z}^{(L)}$, where

$$\boldsymbol{z}^{(l)} = \sigma(\bar{A}\boldsymbol{z}^{(l-1)}W^{(l-1)}), \quad \boldsymbol{z}^{0} = \boldsymbol{X} \in \mathbb{R}^{n\times r}$$

with $\bar{A} = \tilde{D}^{-0.5}\tilde{A}\tilde{D}^{-0.5}, \tilde{A} = A + I, \tilde{D} = D + I$, $W^{(l)}$ is the $l$-th layer weight matrix and $\sigma$ is the pointwise activation.

[2] Cape et al. 2019; Kipf et al. 2016.

# Conventional Wisdom

GNNs have advantages over spectral methods because:

1. they use node features;
2. they use (some) label information to optimize end-to-end;
   - Unsupervised GCN fails in some simple generative models whereas semi-supervised ones succeed (Priebe et al. 2021).
3. they perform better on sparse graphs;
4. they enjoy computational advantages.

# A1: Spectral Embedding with Node Features

**Definition (Covariate-Assisted Spectral Embedding (CASE) [3])**

Given graph $A$, node features X and a tuning parameter $\alpha \in [0, 1]$,
$h_{CASE} = \texttt{SVD}(A + \alpha \mathrm{X}\mathrm{X}^\top)$.

**Definition (Multiple Adjacency Spectral Embedding (MASE) [4])**

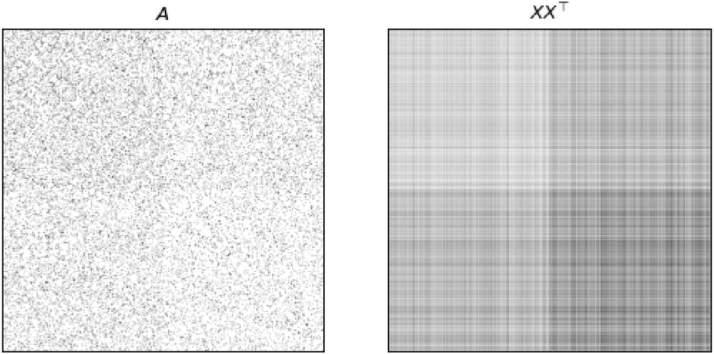
Given graphs $A_1, \cdots, A_J$, $h_{MASE} = \texttt{SVD}([h_{ASE_1}; \cdots, h_{ASE_J}])$.

For example, given a graph $A$ with node features X, we can obtain node similarity graph $A' = \mathrm{X}\mathrm{X}^\top$, and embed $A, A'$ with MASE.

[4] Arroyo et al. 2021; Binkiewicz et al. 2017.

# Contextual Stochastic Block Model (C-SBM)

- Symmetric two blocks;
- $A \sim P = SBM(B; n),\ B = [[p^2, pq]; [pq, q^2]]$ (*rank-1*).
- Node features $X_i | Y_i = 0 \sim \mathcal{N}(q, \sigma_q); X_i | Y_i = 1 \sim \mathcal{N}(p, \sigma_p)$



# Task: Node Classification

Generative Model:

- $A_{ij} \overset{ind}{\sim} Bernoulli(P), P = SBM; X_i|Y_i = k \overset{ind}{\sim} \mathcal{N}_k.$

Data:

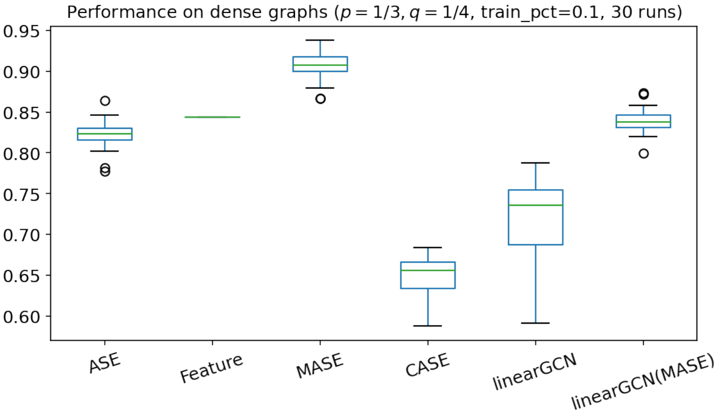
- Adjacency matrix $A \in \mathbb{R}^{n \times n}$, node features $\mathrm{X} \in \mathbb{R}^{n \times d}$
- $m$ out of $n$ labels: $Y_1, \cdots, Y_m$

Method:

- Unsupervised learning:
  - Obtain graph embedding $h(A; X)$ without label supervision;
  - Train a linear classifier $W^L$ based on $\hat{\mathcal{T}}_{m,n} = \{(\hat{X}_i, Y_i)\}_{i \in \{1, \cdots, m\}}$
- Semi-supervised learning:
  - Learn jointly a linear classifier $W^L$ and the graph embedding $h(A; X; Y_1 \cdots Y_m)$ with (partial) label supervision.

Evaluation: evaluate the classification accuracy on the test set.

# Results




- A2: semi-supervised GCN on node feature does not outperform spectral methods (except CASE, by design)

# Injecting global spectral information to local GNN

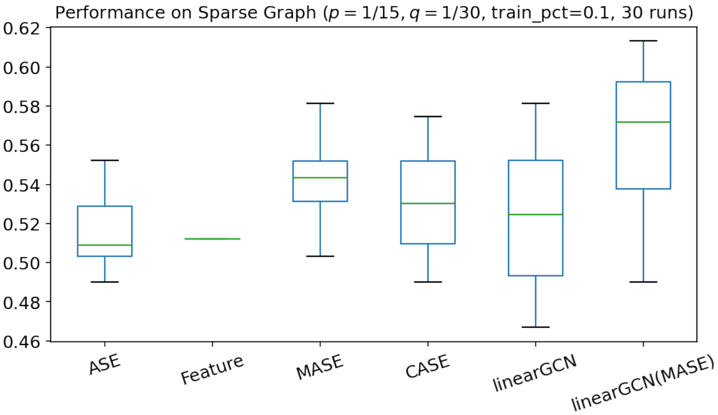
**Definition (Spectral-inspired GNN)**

$h_{GCN(MASE)} = [h_{GCN}; h_{MASE}]$.

Many other techniques in the wild:

- Concatenate $A \in \mathbb{R}^{n \times n}$ and $\mathrm{X} \in \mathbb{R}^{n \times d}$ (Buffelli et al. 2022);
- Positional Encoding using spectral embeddings in Transformer-based GNNs (Kreuzer et al. 2021; Ying et al. 2021).

# A3: GNNs perform better on sparse graphs




- Q3: Can GNNs learn from biased spectral information on sparse graphs?

Arroyo, Jesús et al. (2021). "Inference for multiple heterogeneous networks with a common invariant subspace". In: *Journal of Machine Learning Research* 22.142, pp. 1–49.

Binkiewicz, Norbert, Joshua T Vogelstein, and Karl Rohe (2017). "Covariate-assisted spectral clustering". In: *Biometrika* 104.2, pp. 361–377.

Buffelli, Davide and Fabio Vandin (2022). "The Impact of Global Structural Information in Graph Neural Networks Applications". In: *Data* 7.1, p. 10.

Cape, Joshua, Minh Tang, and Carey E. Priebe (Oct. 2019). "The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics". In: *Annals of Statistics* 47.5, pp. 2405–2439. DOI: 10.1214/18-AOS1752.

Kipf, Thomas N and Max Welling (2016). "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907*.

Kreuzer, Devin et al. (2021). "Rethinking Graph Transformers with Spectral Attention". In: *Advances in Neural Information Processing Systems*.

Priebe, Carey E et al. (2021). "A Simple Spectral Failure Mode for Graph Convolutional Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ying, Chengxuan et al. (2021). "Do Transformers Really Perform Badly for Graph Representation?" In: *Advances in Neural Information Processing Systems* 34.