A CLT for Polynomial GNNs on Community-Based Graphs

Luciano Vinas and Arash A. Amini

University of California, Los Angeles

Problem Setup

Community-based graph

- **Notation:**
  - Graph: Adjacency matrix $A \in \{0,1\}^{n \times n}$
  - Node features: $X \in \mathbb{R}^{n \times d}$
  - Node labels: $z_i \in [L]$ for $i = 1, \dots, n$
- **Graph generation:**
  - Node label z_i is used to generate X_i
  - Node labels z_i, z_j are used to generate A_{ij}

Consequence: A and X are *conditionally* independent given z.

Motivating neighbor aggregations (AX) for GNNs

- Consider classification: $z_i \in \{1, \ldots, L\}$
- Imagine a mixture model for features

$$X_i = \mu_{z_i} + \varepsilon_i$$

  where $\mu_1, \ldots, \mu_L \in \mathbb{R}^d$ are cluster centers.
- Summing improves signal-to-noise ratio for ε_i iid

$$\phi_i := AX = \sum_{j \in \mathcal{N}(i)} \mu_{z_j} + \sum_{j \in \mathcal{N}(i)} \varepsilon_j$$

  assuming $\{z_j\}_{j \in \mathcal{N}(i)}$ are almost all the same as z_i.

Neighbor aggregation example

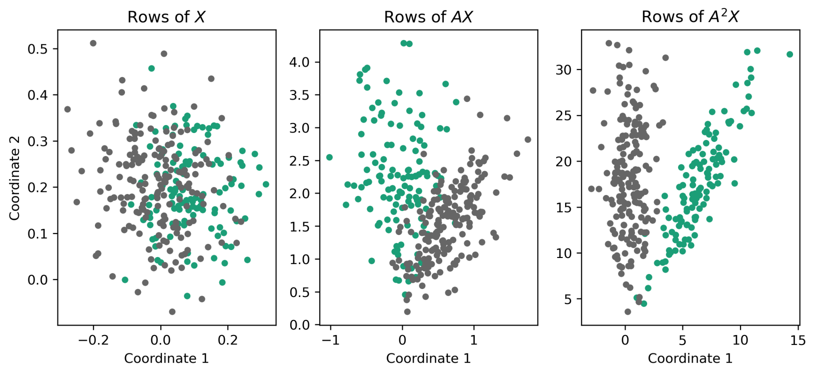

Figure 1: Neighbor aggregation on two cluster example. Network (A, X) is jointly generated where A and X share a common community structure.

Polynomial GNNs (poly-GNNs)

Basic GNN Architecture

$$H^{(k)} = A^k XW$$

- k: network depth (# of layers)
- Simple but effective architecture
- Captures essential feature aggregation mechanism

Contextual Stochastic Block Model (CSBM)

For cluster centers $\mu_1, \ldots, \mu_L \in \mathbb{R}^d$ and connectivity matrix B:

$$\begin{aligned} \text{Features:} \quad X_i \mid z_i &\sim \mu_{z_i} + \varepsilon_i \\ \text{Edges:} \quad A_{ij} \mid z_i, z_j &\sim \text{Bern}(B_{z_i z_j}) \end{aligned}$$

Key Parameters

- $\nu_n = np_{\max}$ (sparsity parameter)
- L: number of clusters
- Feature noise ε_i: sub-Gaussian

Poly-GNN embeddings

- Take GNN embedding $\phi_i^{(k)} = (A^k X)_i$ and normalize:

$$\xi_i^{(k)} = \sqrt{\nu_n}\left(\frac{\phi_i^{(k)}}{\nu_n^k} - \mathbb{E}\left[\frac{\phi_i^{(k)}}{\nu_n^k}\right]\right).$$

- Denote the empirical distribution of $\{\xi_i^{(k)}\}_{i=1}^n$ as

$$\mathbb{P}_n^{(k)} := \frac{1}{n}\sum_{i=1}^n \delta_{\xi_i^{(k)}}.$$

- By a dominant term expansion, we can show

$$\mathbb{E}[\mathbb{P}_n^{(k)}] \rightsquigarrow \mathbb{G} := \sum_{\ell=1}^L \pi_\ell N(0, \Sigma_\ell).$$

Convergence is to a scale-mixture of Gaussians

A CLT from simulated data

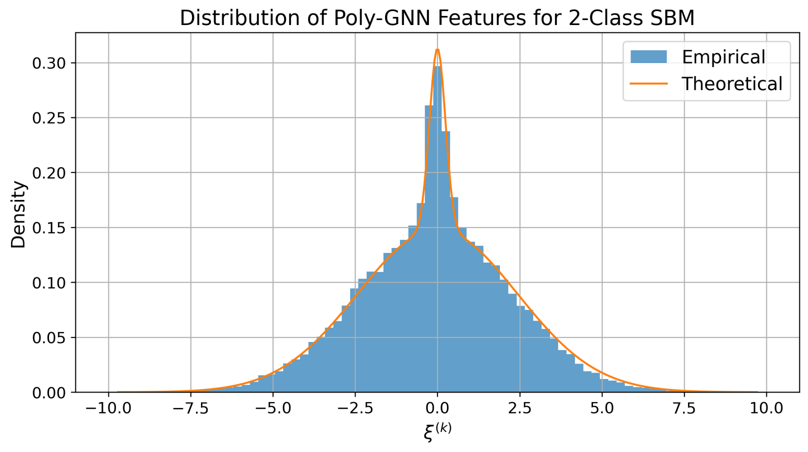

Figure 2: Histogram and theoretical density of $(\xi_i^{(k)})_i$ with $k = 3$ and (A, X) generated by a two class CSBM with 1-D features. **Note:** Entries $\xi_i^{(k)}$ and $\xi_j^{(k)}$ are not independent under the common aggregation of A.

Cross-entropy gradient descent paths

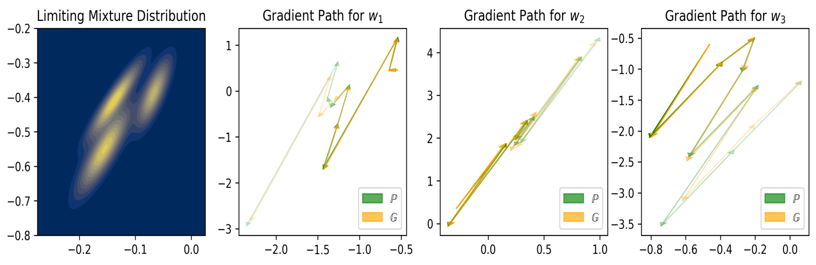

Figure 3: Ten gradients steps of cross-entropy optimization problem for (A, X) drawn from a 3-class CSBM. Shown on right are gradient paths for samples drawn from empirical and theoretical distributions for $\bar{\phi}^{(k)}$.

Insights into oversmoothing

- With $\bar{\phi}_i^{(k)} := \phi_i^{(k)} / \nu_n^k$ and noting $\bar{\phi}_i^{(k)} \to \mu_{z_i} := e_\ell^T J^k M$:

$$\bar{\phi}_i^{(k)} \approx \sum_{\ell=1}^{L} \pi_\ell N\Big(e_\ell^T J^k M, \frac{(J^{k-1}M)^T \mathrm{diag}(e_\ell^T J)(J^{k-1}M)}{\nu_n}\Big)$$

- For our case, $J^k \to \lambda_1^k u_1 v_1^T$ as k increases

Consequence: $\bar{\phi}_i^{(k)}$ mean and covariance collapse into same subspace

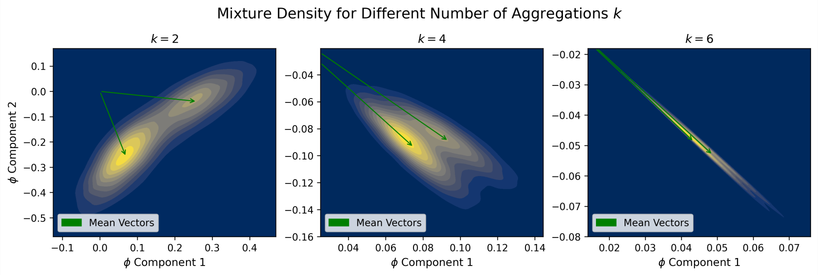

Figure 4: Kernel density estimates of $\bar{\phi}^{(k)}$ at moderate k for a CSBM.

Summary

- Poly-GNNs embeddings $\phi^{(k)}$ have a CLT despite their component dependence
- Strong convergence guarantees makes our result applicable to a large class of optimization problems
- Gives insight into why deep GNNs perform worse than their shallow counterparts (class mean-covariance collapse)