Convergence and Stability of Graph Convolutional Networks on Large Random Graphs

Nicolas Keriven[1], Alberto Bietti[2], Samuel Vaiter[3]

[1]CNRS, GIPSA-lab, Grenoble [2]NYU Center for Data Science, New York [3]CNRS, IMB, Dijon

1: Summary

We study the **convergence** of **Graph Convolutional Networks (GCNs)** to their **continuous counterpart** as the number of nodes grows for a **random graph** model, and derive **stability** properties for realistic perturbations of the model.

Classical "isomorphism-based" analyses of GCN [2] or discrete stability bounds [4] are not entirely satisfying on **large** graphs. **How do GCNs interact with (statistical models of) large graphs?**

- We characterize GCNs on **latent position random graphs** as the number of nodes grows;
- Results are non-asymptotic and valid for **relatively sparse** graphs (logarithmic degrees);
- Analyze the stability of GCNs to **small deformations** of the random graph model.

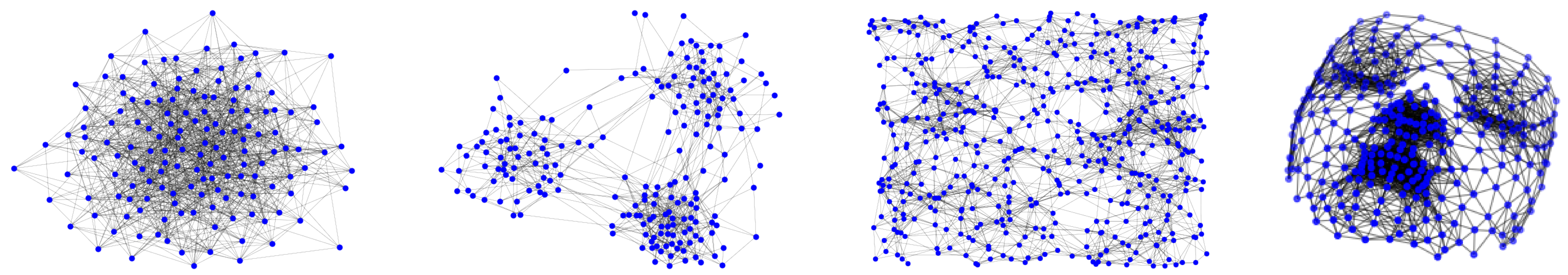
2: *Latent Positions* Random Graphs and Continuous GCN

Graphs	**Random Graphs**
$G = (A, Z)$	$\Gamma = (W, P, f)$
▶ Adjacency matrix $A \in \{0,1\}^{n\times n}$ ▶ Signal over nodes $Z \in \mathbb{R}^{n\times d_0}$	▶ Connectivity kernel $W : \mathcal{X}\times\mathcal{X} \to [0,1]$ ▶ Distribution P over $\mathcal{X} \subset \mathbb{R}^d$ ▶ Function $f : \mathcal{X} \to \mathbb{R}^{d_0}$
▶ Degrees $D = \mathrm{diag}(A1_n)$ ▶ Norm. Laplacian $L = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$	▶ Degree **function** $d = \int W(\cdot,x)\mathrm{d}P(x)$ ▶ Norm. Laplacian $\mathcal{L}f = \int \frac{W(\cdot,x)}{\sqrt{d(\cdot)d(x)}} f(x)\mathrm{d}P(x)$

Generative model

$$x_i \overset{iid}{\sim} P, \quad z_i = f(x_i), \quad a_{ij} \sim \mathrm{Ber}(\alpha_n W(x_i, x_j))$$

- Dense $\alpha_n = O(1)$, Sparse $\alpha_n = O(1/n)$, Relatively sparse $\alpha_n = O(\log n/n)$
- Includes ER, SBM, ε-graphs, Gaussian kernel...



Isomorphism	**Continuous isomorphism**
$(A, Z) \sim (\sigma A \sigma^\top, \sigma Z)$	$(P, W, f) \sim (\varphi_\sharp^{-1} P, W \circ \varphi^{\otimes 2}, f \circ \varphi)$
▶ Permutation matrix $\sigma \in \{0,1\}^{n\times n}$	▶ Bijection $\varphi : \mathcal{X} \to \mathcal{X}$

(Spectral) GCNs	**Continuous-GCNs (c-GCNs)**
▶ Propagate **signal over nodes** ▶ Poly. filters $h(L) = \sum_k \beta_k L^k$	▶ Propagate **function over latent space** ▶ Poly. filters with $\mathcal{L}^k = \mathcal{L} \circ \ldots \circ \mathcal{L}$
$z_j^{(\ell+1)} = \rho\left(\sum_i h_{ij}^{(\ell)}(L) z_i^{(\ell)} + b_j^{(\ell)} 1_n\right)$	$f_j^{(\ell+1)} = \rho\left(\sum_i h_{ij}^{(\ell)}(\mathcal{L}) f_i^{(\ell)} + b_j^{(\ell)}\right)$
Equivariant output	**Equivariant output**
$\Phi_A(Z) = Z^{(M)}\theta + 1_n b^\top$	$\Phi_{W,P}(f) = f^{(M)}\theta + b$
▶ $\Phi_{\sigma A \sigma^\top}(\sigma Z) = \sigma \Phi_A(Z)$	▶ $\Phi_{W\circ\varphi^{\otimes 2}, \varphi_\sharp^{-1}P}(f\circ\varphi) = \Phi_{W,P}(f)\circ\varphi$
Invariant output	**Invariant output**
$\bar\Phi_A(Z) = 1_n^\top \Phi_A(Z)$	$\bar\Phi_{W,P}(f) = \int \Phi_{W,P}(f)\mathrm{d}P$
▶ $\bar\Phi_{\sigma A \sigma^\top}(\sigma Z) = \bar\Phi_A(Z)$	▶ $\bar\Phi_{W\circ\varphi^{\otimes 2}, \varphi_\sharp^{-1}P}(f\circ\varphi) = \bar\Phi_{W,P}(f)$

3: Convergence to continuous GCNs

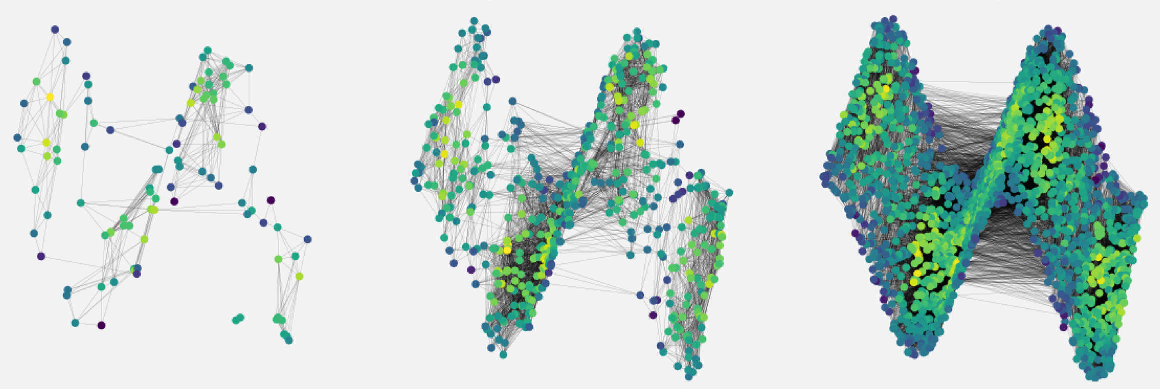
Theorem Let $(A, Z) \sim \Gamma$ with n nodes be drawn from Γ. When $\alpha_n \gtrsim \log n/n$, with probability $1 - n^{-r}$ for some $r > 0$, we have

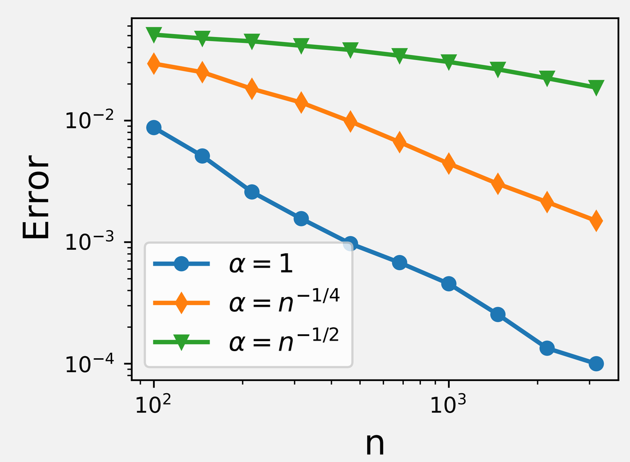
$$\left.\begin{array}{r}\sqrt{\frac{1}{n}\sum_i (\Phi_A(Z)_i - \Phi_{W,P}(f)(x_i))^2} \\ \|\bar\Phi_A(Z) - \bar\Phi_{W,P}(f)\|_2\end{array}\right\} \leqslant R_n$$

where $R_n = O(dn^{-\frac{1}{2}} + (n\alpha_n)^{-\frac{1}{2}})$.

Numerical illustration

Equivariant GCN output for constant input $f = 1$ with growing number of nodes and convergence with different sparsity levels α_n





4: Stability of GCNs to model deformations

For $i = 1, 2$, assume (A_i, Z_i) drawn from models (W_i, P_i, f_i).

Finite-sample stability in the equivariant case

Theorem Denote $Q_i = \Phi_{W_i,P_i}(f_i)_\sharp P_i$. With prob. $1 - n^{-r}$:

$$\min_{\sigma\in\Sigma_n} \sqrt{\tfrac{1}{n}\sum_i ((\Phi_{A_1})_i - (\Phi_{A_2})_{\sigma(i)})^2} \leqslant \mathcal{W}_2(Q_1, Q_2) + R_n + O(n^{-1/d}),$$

where $\mathcal{W}_2$ is the Wasserstein-2 distance and $d = \dim(\mathcal{X})$.

Deformation of a translation-invariant model

- Translation-invariant kernel $W(x, y) = w(x - y)$
- Smooth diffeomorphism [3] $\tau : \mathcal{X} \to \mathcal{X}$ ("size" $\|\nabla\tau\|_\infty$)

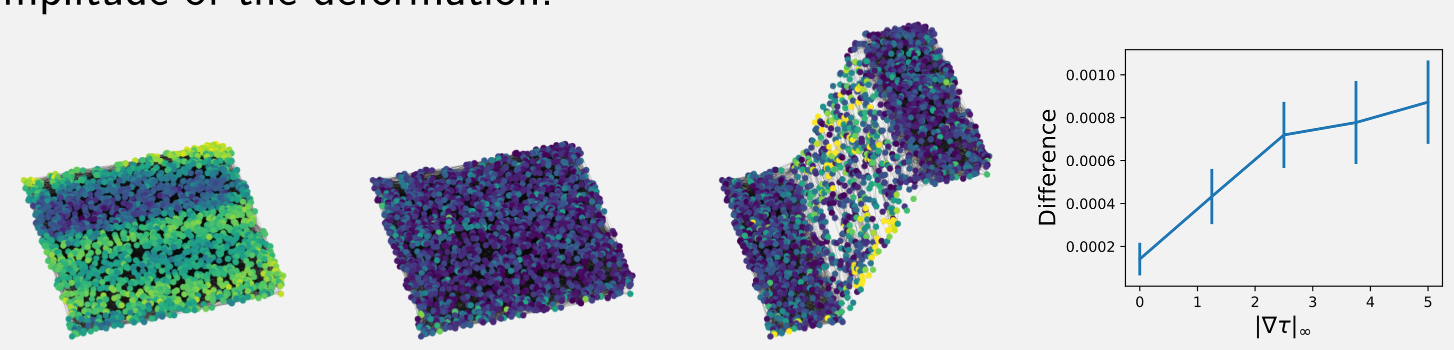
Theorem (Deformations of W, P, or f.)

- $W(x, x') \to W_\tau(x, x') \overset{\text{def.}}{=} W(x - \tau(x), x' - \tau(x'))$
  $$\|\bar\Phi_{W_\tau} - \bar\Phi_W\| \lesssim \|\nabla\tau\|_\infty$$
- $P \to P_\tau = (Id - \tau)_\sharp P$, and $f' = f \circ (Id - \tau)$, or degree functions as inputs $(f, f') = (d_P, d_{P_\tau})$
  $$\|\bar\Phi_P(f) - \bar\Phi_{P_\tau}(f')\| \lesssim \|\nabla\tau\|_\infty$$
- $f \to f_\tau = f \circ (Id - \tau)$
  $$\|\bar\Phi(f_\tau) - \bar\Phi(f)\| \lesssim \|\nabla\tau\|_\infty.$$

Similar bounds hold for the equivariant case on $\mathcal{W}_2(Q_1, Q_2)$ with $Q_i = \Phi_{W_i,P_i}(f_i)_\sharp P_i$.

Numerical illustration (Random graph with 3D latent positions)

From left to right: output signal; new drawing of the random edges; deterministically deformed latent positions; invariant GCN with respect to the amplitude of the deformation.



[1] Bruna et al. **Spectral Networks and Locally Connected Networks on Graphs**. *ICLR*, 2014.

[2] Xu et al. **How Powerful are Graph Neural Networks?**. *ICLR*, 2020.

[3] Mallat. **Group Invariant Scattering**. *Comm. Pure Appl. Math.*, 2012.

[4] Gama et al. **Stability Properties of Graph Neural Networks**. *IEEE Trans. Sig. Proc.*, 2020.

nicolas.keriven@gipsa-lab.grenoble-inp.fr, alberto@bietti.me, samuel.vaiter@u-bourgogne.fr

This work was supported by ANR GraVa ANR-18-CE40-0005 and ANER RAGA G048CVCRB2018ZZ.