On the Universality of Graph Neural 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, LJAD, Nice

1: Summary

We study the **universality** of **Graph Neural Networks (GNNs)** on **random graphs** (RGs). It is known that GNNs converge to continuous models when the number of nodes grow to infinity. We study their universality on several classical RG models.

Classical "isomorphism-based" analyses of GNN [2] are not entirely satisfying on **large** graphs. **What can GNNs compute on (statistical models of) large graphs? Are recent architectures more powerful than vanilla ones?**

- so-called Structured GNNs (SGNNs) [1] also converge to continuous models on random graphs;
- c-SGNNs are strictly more powerful than c-GNNs;
- c-SGNNs are **universal** on some RG models.

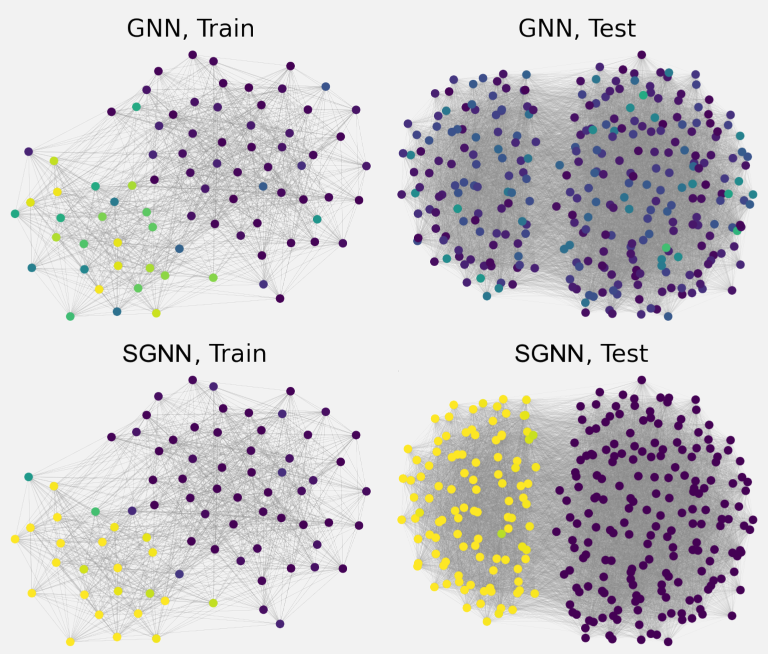
3: c-GNN vs c-SGNN

Approximation power? We look at:

- **Perm.-invariant:** $\{(W,P) \mapsto \Psi_{W,P}\}$ w.r.t. (a subset of) the functions $\mathbb{R}^{\mathcal{W}\times\mathcal{P}}$;
- **Perm.-equivariant:** (W,P) are **fixed**, $\{x \mapsto \Psi_{W,P}(x)\}$ w.r.t. (a subset of) the functions $\mathbb{R}^{\mathcal{X}}$ *(when Ψ parameters vary)*

Theorem c-SGNNs are more powerful than c-GNNs, for perm.-invariant or perm.-equivariant, deterministic or random edges.

SBM with constant expected degree:



[1] Vignac et al. **Building powerful and equivariant graph neural networks with structural message-passing**. *NeurIPS*, 2020.

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

[3] Keriven et al. **Convergence and Stability of Graph Convolutional Networks on Large Random Graphs**. *NeurIPS*, 2020.

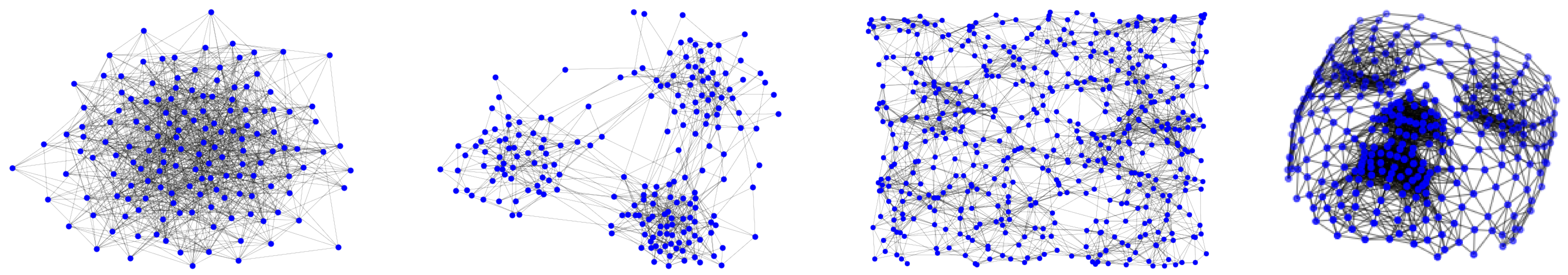
2: Random Graphs, GNNs, SGNNs

Latent Position Random Graphs

- Compact latent space $\mathcal{X} \subset \mathbb{R}^d$, connectivity kernel $W : \mathcal{X}\times\mathcal{X} \to [0,1]$, distrib. $P \in \mathcal{P}(\mathcal{X})$

$$x_1,\ldots,x_n \overset{iid}{\sim} P, \quad a_{ij} = \begin{cases} W(x_i,x_j) & \text{deterministic edges case} \\ \alpha_n^{-1}\mathrm{Ber}(\alpha_n W(x_i,x_j)) & \text{random edges case (NB: normalized)}\end{cases}$$

- Two cases, some results simpler (and/or only valid) in the deterministic case
- Dense $\alpha_n = O(1)$, Sparse $\alpha_n = O(1/n)$, Relatively sparse $\alpha_n = O(\log n/n)$
- Includes ER, SBM, ε-graphs, geometric graphs...



(Spectral) GNNs

- Propagate node signal $Z^{(\ell)} \in \mathbb{R}^{n\times d_\ell}$
- Poly. filters $h(A) = \sum_k \beta_k A^k$

$$Z_{:,j}^{(\ell+1)} = \rho\left(\sum_i h_{ij}^{(\ell)}(A/n) Z_{:,i}^{(\ell)} + b_j^{(\ell)}\right)$$

Output:

$$\Phi_A(Z^{(0)}) = \begin{cases} Z^{(M)} & \text{perm.-equi.} \\ g_{\mathrm{MLP}}(\frac{1}{n}\sum_i Z_{i,:}^{(M)}) & \text{perm.-inv.}\end{cases}$$

Continuous-GNNs (c-GNNs)

- Propagate function $f^{(\ell)} : \mathcal{X} \to \mathbb{R}^{d_\ell}$
- Poly. filters with $T_{W,P}f = \int W(\cdot,x)f(x)dP(x)$

$$f_j^{(\ell+1)} = \rho\left(\sum_i h_{ij}^{(\ell)}(T_W) f_i^{(\ell)} + b_j^{(\ell)}\right)$$

Output:

$$\Phi_{W,P}(f^{(0)}) = \begin{cases} f^{(M)} & \text{perm.-equi.} \\ g_{\mathrm{MLP}}(\int f^{(M)} dP) & \text{perm.-inv.}\end{cases}$$

No input signal?

- Classic strategy: **constant input** $\Phi_A = \Phi_A(1_n)$, $\Phi_{W,P} = \Phi_{W,P}(1)$
- Just as Weisfeiler-Lehman [2], limited power on constant degree function (for instance)

Structured GNN (SGNN) [1]

- **Node ids** $E_q \in \mathbb{R}^n$ with **arbitrary ordering**, like e_q or $A^k e_q$
- Perm.-Equi. Φ_A, perm.-inv./equi. Φ'_A

$$\Psi_A = \Phi'_A(\tfrac{1}{n}\textstyle\sum_q \Phi_A(E_q))$$

- Still perm.-inv./equi. as Φ'_A
- Proved to be more powerful than WL [1]

Continuous-SGNN

- **Bivariate function** $\eta : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$
- E.g., $\delta(x-y)$ or $T_W^k\delta(x-y)$

$$\Psi_{W,P} = \Phi'_{W,P}(\textstyle\int \Phi_{W,P}(\eta(\cdot,x))dP(x))$$

- We need $E_q \approx [\eta(x_i,x_q)]_i \to k \geqslant 1$ for deterministic edges, $k \geqslant 2$ for random edges

Theorem GNNs converge to c-GNNs [4], and **SGNNs converge to c-SGNNs**.

4: Universality of c-SGNN

Stochastic Block Models

$$\mathcal{X} = \{1,\ldots,K\}$$

- **Assume: no P_k is a sum/difference of the others.**

Proposition

- **Inv.:** fix P, c-SGNNs $\{W \to \Psi_{W,P}\}$ are universal.
- **Equi.:** fix W,P. If additionally $[W_{kl}]$ is invertible, c-SGNNs can distinguish communities.

"Additive" kernels

$$W(x,y) = u(v(x)+v(y))$$

- When u,v can be anything, **universal approximators** of kernels
- **Assume: $u,v : \mathbb{R}\to\mathbb{R}$ are injective**

Proposition

- **Inv.:** Fix W: c-SGNNs $\{P \to \Psi_{W,P}\}$ are universal.
- **Equi.:** Fix W,P: c-SGNNs are universal.

Radial kernels

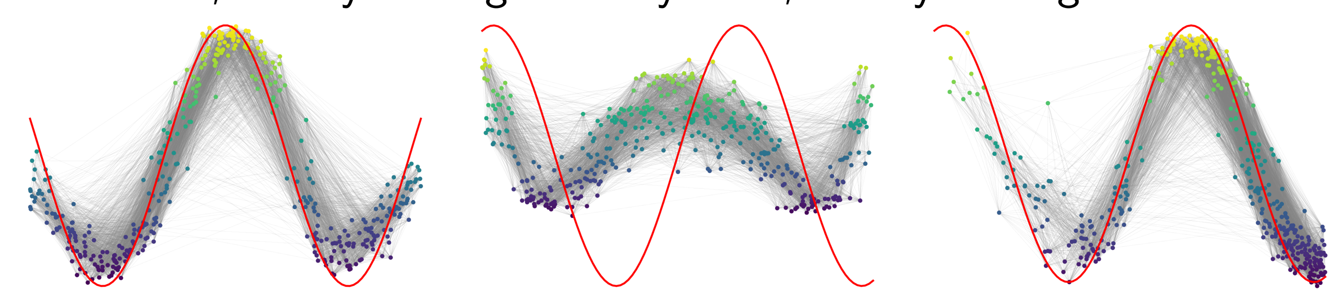
$$W(x,y) = w(\|x-y\|)$$

- **Assume: $\mathcal{X} = [-1,1]$ and injective w**

Proposition

- **Inv.:** Fix W. c-SGNNs are universal when P vary.
- **Equi.:** Fix W and centered P. When P is symmetric, c-SGNNs are universal **in symmetric functions**; when P is non-symmetric, c-SGNNs are universal.

Latent position on x-axis, output on y-axis. From left to right: sym. target and P; non-sym. target but sym. P, non-sym. target and P.



Spherical kernels

$$W(x,y) = w(x^\top y), \quad \mathcal{X} = \mathbb{S}^{d-1}$$

- **Assume: w injective, P has density f with "some injectivity condition" (w.r.t. spherical harmonics)**

Proposition

- **Equi.:** Fix W,P. c-SGNNs are universal.

nicolas.keriven@cnrs.fr, alberto.bietti@nyu.edu, samuel.vaiter@cnrs.fr

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