



HAL
open science

Convergence of Message Passing Graph Neural Networks with Generic Aggregation On Random Graphs

Matthieu Cordonnier, Nicolas Keriven, Nicolas Tremblay, Samuel Vaiter

► **To cite this version:**

Matthieu Cordonnier, Nicolas Keriven, Nicolas Tremblay, Samuel Vaiter. Convergence of Message Passing Graph Neural Networks with Generic Aggregation On Random Graphs. GSP 2023 - 6th Graph Signal Processing workshop, Jun 2023, Oxford, United Kingdom. pp.1-3. hal-04106511

HAL Id: hal-04106511

<https://hal.science/hal-04106511v1>

Submitted on 25 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

CONVERGENCE OF MESSAGE PASSING GRAPH NEURAL NETWORKS WITH GENERIC AGGREGATION ON RANDOM GRAPHS

Matthieu Cordonnier Nicolas Keriven* Nicolas Tremblay Samuel Vaiter**

GIPSA-lab, CNRS, Univ. Grenoble Alpes, Grenoble INP

*CNRS, IRISA, Rennes

**CNRS and Université Côte d’Azur, Laboratoire J.A. Dieudonné

1. MOTIVATIONS

A classical method to study the expressivity of Graph Neural Networks (GNNs) is to compare them to the Weisfeiler-Lehman test for the graph isomorphism problem. Nevertheless, this approach becomes questionable for very large graphs. Real large graphs are rarely fully known, and their structure can change quickly over time. In such cases, we rather focus on global tendencies for which useful tools are random graphs. Let $(\mathcal{X} \subset \mathbb{R}^d, P)$ be a compact probability space, from a measurable kernel $W : \mathcal{X}^2 \rightarrow [0, 1]$ one can sample random graphs. Specifically, G_n drawn from this model has vertices $\{X_1, \dots, X_n\} \subset \mathcal{X}$ and weights w_{ij} where :

$$X_1, \dots, X_n \stackrel{iid}{\sim} P, \quad w_{ij} = w_{ji} = W(X_i, X_j).$$

Our purpose is to examine whether a message passing GNN (MPGNN) with an unspecified aggregation on a large G_n is close to a continuous counterpart on the underlying random graph model. We give sufficient conditions under which the discrete model is a good approximation of the continuous analogue. Moreover, assuming some regularity on the aggregation, we provide explicit non asymptotic bounds for the deviation between a discrete MPGNN and its continuous counterpart. Our main contribution involves considering a generic aggregation function while previous related work has focused on particular cases, such as Graph Convolutional Network [?, ?], or degree normalized mean [?].

2. MPGNN

A L -layer MPGNN iteratively propagates a signal over a graph. At each step, the current representations of every node’s neighbors are gathered, transformed, and combined to update the node’s representation. Let G be a graph on the vertex set $\{1, \dots, n\}$, and $Z = Z^{(0)} \in \mathbb{R}^{n \times d_0}$ be a signal on it. There are L operators $(F^{(l)})_{1 \leq l \leq L}$ such that, at each layer, the update $Z^{(l+1)}$ of the signal is computed node-wise by:

$$z_i^{(l+1)} = F^{(l+1)} \left(z_i^{(l)}, \left\{ \left(z_j^{(l)}, w_{i,j} \right) \right\}_{j \in \mathcal{N}(i)} \right) \in \mathbb{R}^{d_{l+1}}. \quad (1)$$

The $F^{(l)}$ are referred to as *aggregations*. They are the core of

a MPGNN and their fundamental property is to be invariant to the permutation of the neighbors, so that the full network is consistent with graph isomorphism. This is ensured in (1) since a multiset (double braces) is taken as argument.

Typically, the aggregation first transforms the $z_j^{(l)}$ through a learnable transformation $\psi^{(l+1)}$. Then combine them using a weighted mean \bigoplus (in a broad sense, such as arithmetic mean, point-wise maximum, *etc.*) with weights

$$c_{i,j}^{(l+1)} = c^{(l+1)} \left(z_i^{(l)}, z_j^{(l)}, w_{ij} \right) \quad (2)$$

such that (1) can be rewritten as

$$z_i^{(l+1)} = \bigoplus \left(\left\{ \left(\psi^{(l+1)}(z_j^{(l)}), c_{i,j}^{(l+1)} \right) \right\}_{j \in \mathcal{N}(i)} \right).$$

Examples of different aggregations

1. Convolution :

$$z_i^{(l+1)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} w_{i,j} \psi^{(l+1)}(z_j^{(l)}).$$

2. Max Convolution :

$$z_i^{(l+1)} = \max_{j \in \mathcal{N}(i)} w_{i,j} \psi^{(l+1)}(z_j^{(l)}).$$

3. Degree Normalized Convolution :

$$z_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \frac{w_{i,j}}{\sum_{k \in \mathcal{N}(i)} w_{i,k}} \psi^{(l+1)}(z_j^{(l)}).$$

4. Graph Attention (GAT): the $c^{(l)}$ from (2) are learnable

$$: z_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \frac{c_{i,j}^{(l+1)}}{\sum_{k \in \mathcal{N}(i)} c_{i,k}^{(l+1)}} \psi^{(l+1)}(z_j^{(l)}).$$

3. CONTINUOUS MPGNN

Let (W, P) be a random graph model, a L -layer continuous MPGNN (cMPPGNN) propagates a function over the latent space \mathcal{X} . For an input $f = f^{(0)} \in \mathcal{C}(\mathcal{X}, \mathbb{R}^{d_0})$, $f^{(l+1)}$ is computed by:

$$f^{(l+1)}(x) = \mathcal{F}_P^{(l+1)} \left(f^{(l)}(x), \left(f^{(l)}, W(x, \cdot) \right) \right) \in \mathbb{R}^{d_{l+1}}. \quad (3)$$

Examples (continuous equivalents of 1, 2, 3 and 4) :

- a. Convolution :

$$f^{(l+1)}(x) = \int_y W(x, y) \psi^{(l+1)}(f^{(l)}(y)) dP.$$

b. Max Convolution :

$$f^{(l+1)}(x) = \text{esssup}_P W(x, \cdot) \psi^{(l+1)}(f^{(l)}(\cdot)).$$

c. Degree Normalized Convolution :

$$f^{(l+1)}(x) = \int_y \frac{W(x,y)}{\int_t W(x,t) dP(t)} \psi^{(l+1)}(f^{(l)}(y)) dP.$$

d. GAT :

$$f^{(l+1)}(x) = \int_y \frac{c^{(l+1)}(x,y)}{\int_t c^{(l+1)}(x,t) dP(t)} \psi^{(l+1)}(f^{(l)}(y)) dP.$$

4. CONVERGENCE ON DENSE RANDOM GRAPHS

We are interested in a MPGNN on a random graph G_n drawn from (W, P) . For such a discrete MPGNN, we shall define its continuous counterpart on (W, P) . Omitting technical details

Definition 4.1. We say that \mathcal{F} is the continuous counterpart of F if for any f , for any P ,

$$x \mapsto \mathcal{F}_P(f(x), (f, W(x, \cdot)))$$

is the uniform limit, as n tends to infinity, of

$$x \mapsto \mathbb{E}[F(f(x), \{(f(X_k), W(x, X_k))\}_{1 \leq k \leq n})]$$

where $X_k \stackrel{iid}{\sim} P$. The rate of this convergence is called a_n .

For each example calculate 1, 2, 3 or 4, we prove that the continuous equivalent is indeed a, b, c or d, and we calculate the corresponding a_n .

Denote $X = (X_1, \dots, X_n)$ the vertices of G_n and S_X the sampling operator at X . For an input function f , $S_X(f)$ is a signal on G_n . To compare the output of this signal through the discrete network to the output of f through the continuous counterpart, we shall use the Maximum Absolute Error (MAE) defined by :

$$MAE(f) = \max_i \left\| (S_X(f))_i^{(L)} - f^{(L)}(X_i) \right\|_{\infty}.$$

Our main result is a sufficient condition for this MAE to be arbitrary small with high probability. We provide a non asymptotic bound with high probability that relies on the so called McDiarmid inequality.

Theorem 4.1. Call $D_n^{(l)}$ the bounded differences of the $F^{(l)}$ and D_n their maximum over l . Call d_{max} the maximum of the d_l over l . Then, up to some regularity assumptions on the $F^{(l)}$, for any $\rho > 0$, with probability at least $1 - \rho$:

$$MAE(f) \lesssim LD_n \sqrt{n \ln \left(\frac{n 2^L d_{max}}{\rho} \right)} + a_n. \quad (4)$$

Where \lesssim hides multiplicative constants independent of n .

This bound suggests that if the bounded differences are sharp enough, typically $D_n = o(1/\sqrt{n \ln n})$, the MAE is small.

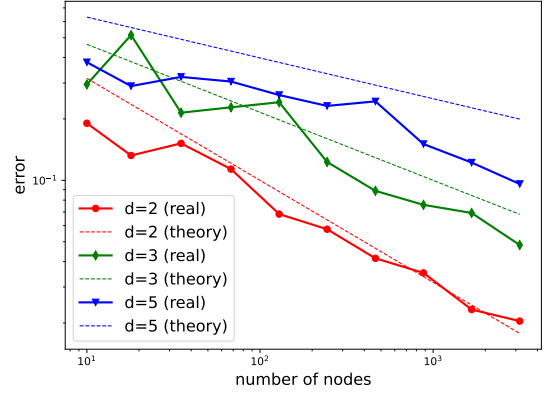


Fig. 1: illustration of (5) for different values of d , the dashed line represented $(1/n)^{1/d}$ and the plain line the MAE. The plot is log scaled.

Corollary 4.1.1. If $D_n = o(1/\sqrt{n \ln n})$ then

$$\max_i \left\| (S_X(f))_i^{(L)} - f^{(L)}(X_i) \right\|_{\infty} \rightarrow 0 \text{ in probability.}$$

We check this condition on the examples. It appears that it is verified by 1, 3 and 4 with a $D_n = O(1/n)$ but not by 2. Nevertheless, the convergence remains true for that case with a different bound.

Theorem 4.2. For a Max Conv. MPGNN, assume that $P \geq C\lambda$ for some $C > 0$ where λ is the Lebesgue measure. Then for any $\rho > 0$, we have with probability at least $1 - \rho$,

$$MAE(f) \lesssim L \left(\frac{1}{n} \ln \left(\frac{2^{L-1} n d_{max}}{\rho} \right) \right)^{1/d}. \quad (5)$$

This bound indicates a different behaviour when the aggregation is a maximum. It depends on the input space's dimension and decreases significantly slower than (4) for a large d .

5. REFERENCES

- [1] Nicolas Keriven, Alberto Bietti, and Samuel Vaiter, "Convergence and Stability of Graph Convolutional Networks on Large Random Graphs," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 21512–21523, Curran Associates, Inc.
- [2] Luana Ruiz, Luiz Chamon, and Alejandro Ribeiro, "Graphon Neural Networks and the Transferability of Graph Neural Networks," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 1702–1712, Curran Associates, Inc.

Extended abstract presented at the Graph Signal Processing workshop 2023. This work was partially supported by the French National Research Agency in the framework of the « France 2030 » program (ANR-15-IDEX-0002), the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01), and the ANR grants GRANDMA (ANR-21-CE23-0006), GRANOLA (ANR-21-CE48-0009) and GRAVA (ANR-18-CE40-0005).

- [3] Sohir Maskey, Yunseok Lee, Ron Levie, and Gitta Kutyniok, “Stability and Generalization Capabilities of Message Passing Graph Neural Networks,” *arXiv:2202.00645 [cs, math]*, Feb. 2022, arXiv: 2202.00645.