Convergence of Graph Neural Networks with generic aggregation functions on random graphs

Matthieu CORDONNIER[1] Nicolas KERIVEN[2] Nicolas TREMBLAY[1] Samuel VAITER[3]

[1]CNRS, GIPSA-lab, UGA, Grenoble INP 11 rue des Mathématiques 38400 St-Martin-d'Hères

[2]CNRS, IRISA, Univ. Rennes 1, 263 Av. Général Leclerc, 35000 Rennes

[3]CNRS, LJAD, Univ. Côte d'Azur, 28 Avenue Valrose 06108 Nice

Résumé – On s'intéresse au comportement limite des Réseaux de Neurones sur Graphes appliqués aux grands graphes aléatoires. On démontre que sous certaines hypothèses de régularité, un GNN converge vers un homologue «continu». L'originalité de ce travail est de considérer des fonctions d'agrégation abstraites et générales tandis que les études antérieures traitent de cas particuliers. On quantifie les convergences à l'aide de bornes non asymptotiques en probabilité basées sur l'inégalité de McDiarmid pour des agrégations ayant une régularité de type lipschitzienne. Le cas du maximum, non inclus dans cette catégorie, est traité à part.

Abstract – We study the limit behavior of GNNs on large random graphs. We consider GNNs with a generic aggregation function and show that under mild regularity conditions, they converge to a "continuous" counterpart. We provide some non asymptotic bounds with high probability for this convergence which encompass several cases of aggregation such as, the mean, or the maximum.

1 Introduction

A classical approach to study the properties of Graph Neural Networks [1] (GNNs) is to compare them to the Weisfeiler-Lehman test for the graph isomorphism problem [10]. Nevertheless, this approach becomes questionable for very large graphs, where we would rather focus on global tendencies. The latter are traditionally modelled with *random graphs* [2]: since the early Erdös Rényi model, they have become classical tools in statistics, statistical learning and statistical physics, for instance to study clustering problems [4] or limits of large graphs [5] and their properties. In the case of GNNs, this approach has shed light on their generalization properties [7] or stability to deformations [3, 9]. In this paper, our purpose is to examine whether a GNN on a large random graph is close to a "continuous" limit on the random graph model. We aim to address Message-Passing GNN (MPGNN) with *generic aggregation function*, including non-smooth ones like max, while previous work [3, 7, 8] focused on particular cases such as Graph Convolutional Network or degree normalized mean. We give sufficient conditions under which the discrete model is a good approximation of the continuous analogue, and provide non asymptotic deviation bounds. Depending on the regularity of the aggregation function, we obtain different rates of convergence. Due to space constraints, we refer the reader to the full version of the paper [6] for the mathematical proofs.

2 Notations and definitions

We fix a positive integer d and $\mathcal{X}$ a compact subset of $\mathbb{R}^d$, endowed with the infinite norm $\|x\|_\infty = \max_i |x_i|$ as well as its Borel sigma algebra. Except when specified differently, all topological concepts will be considered relatively to $\|\cdot\|_\infty$.

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).

The space of continuous functions from $\mathcal{X}$ to $\mathbb{R}^p$ is written $\mathcal{C}(\mathcal{X}, \mathbb{R}^p)$, and it is also equipped with the supremum norm $\|f\|_\infty = \sup_{x\in\mathcal{X}} \|f(x)\|_\infty$.
The group of permutations of $\{1, \dots, n\}$ is denoted as S_n. If $x = (x_1, \dots, x_n)$ is an n-tuple and σ an element of S_n, we define the n-tuple $\sigma \cdot x$ as $\sigma \cdot x = (x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(n)})$. The set of bijections ϕ of $\mathcal{X}$ such that both ϕ and ϕ^{-1} are measurable is a group for the composition of functions. We call this group the group of automorphisms of $\mathcal{X}$ and denote it as $\mathrm{Aut}(\mathcal{X})$. We denote as $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. For a measure $P \in \mathcal{P}(\mathcal{X})$ and a bijection $\phi \in \mathrm{Aut}(\mathcal{X})$, the push forward measure of P through ϕ is defined as $\phi_\# P(A) = P(\phi^{-1}(A))$ for all open sets A. Since this makes $\mathrm{Aut}(\mathcal{X})$ act on the set of probability measures, we also use the notation $\phi \cdot P = \phi_\# P$, which is standard for a (left) group action. For the same reason, we shall use the notation $\phi \cdot f = f \circ \phi^{-1}$ and $\phi \cdot W = W(\phi^{-1}(\cdot), \phi^{-1}(\cdot))$ whenever f is a measurable function on $\mathcal{X}$ and W is a bivariate measurable function on $\mathcal{X} \times \mathcal{X}$.
Sets are represented between braces $\{\cdot\}$, whereas multisets, sets in which an element is allowed to appear twice or more, are represented by double braces $\{\!\{\cdot\}\!\}$. If $\mathcal{M}$ and $\mathcal{M}'$ are two multisets of the same size, say n, containing elements from a metric space $(\mathcal{E}, d)$, we define their distance by:

$$\delta(\mathcal{M}, \mathcal{M}') = \min_{\sigma\in S_n} \max_{x_i\in\mathcal{M}, x'_i\in\mathcal{M}'} d(x_i, x'_{\sigma(i)}). \tag{1}$$

We define the sampling operator the following way. If $f : \mathcal{E} \to \mathcal{E}'$ and $X = (x_1, \dots, x_n) \in \mathcal{E}^n$:

$$S_X f = (f(x_1), \dots, f(x_n)) \in \mathcal{E}'^n. \tag{2}$$

Graph. An undirected weighted graph G with n vertices is defined by a triplet (V, E, w), where $V = \{v_1, \dots, v_n\}$ is a finite set called the set of vertices (or nodes) and E is the set of edges. The set of neighbors of a vertex v_i in G is referred to as $\mathcal{N}(i)$. The weight function w assigns a nonnegative number to each edge. It is often represented by a symmetric function

$w : V^2 \to \mathbb{R}^+$ and the abbreviation $w_{i,j}$ is used to denote the weight $w(v_i, v_j) = w(v_j, v_i)$ where $\{v_i, v_j\} \in E$. The set of graphs defined on the vertex set V is denoted as $\mathcal{G}(V)$.

Graph signal. Given a graph $G \in \mathcal{G}(V)$, where $|V| = n$, a signal on G is a map from the set of vertices V to $\mathbb{R}^d$ that assigns a d-dimensional vector z_i to each vertex v_i. The images from all vertices are stacked into a tensor Z of size $n \times d$. Abusing notations, we may not distinguish between the map and its image Z, the latter being also named the signal.

Random Graph Model. A random graph model is a couple (W, P) where P is a probability measure on $\mathcal{X}$ and $W : \mathcal{X}^2 \mapsto [0, 1]$ is a *similarity kernel*, *i.e.* a symmetric measurable function. We generate a random graph of size n from (W, P) as follows:

$$X_1, \dots, X_n \overset{iid}{\sim} P, \quad w_{i,j} = w_{j,i} = W(X_i, X_j)\,. \quad (3)$$

When convenient, we will use the short notation $X = (X_1, \dots, X_n)$ for the tuple of the vertices of a random graph. We call $\mathcal{G}_n(W, P)$ the distribution from which random graphs with n nodes are drawn. We bring the reader's attention to the fact that in the above definition, a random graph is always fully connected and edge may have a weight equal to zero. A common model [3] is to add a Bernoulli distribution to the connectivity, which is not done here for the sake of simplicity.

Graph and Random Graph Model isomorphism. Two graphs G_1 and G_2 in $\mathcal{G}(V)$ are said to be isomorphic if there is a permutation $\sigma \in S_n$ such that $E_2 := \{\{v_{\sigma^{-1}(i)}, v_{\sigma^{-1}(j)}\} |\ \{v_i, v_j\} \in E_1\}$ and $w_1(v_i, v_j) = w_2(v_{\sigma^{-1}(i)}, v_{\sigma^{-1}(j)})$. This permutation is called a graph isomorphism. In this case we note $G_2 = \sigma \cdot G_1$. Moreover, if Z is a signal on G_1 and $\sigma \in S_n$, $\sigma \cdot Z$ is an isomorphic signal on the graph $\sigma \cdot G_1$. Two probability measures P_1 and P_2 on $\mathcal{X}$ are said isomorphic if there is some ϕ in $\mathrm{Aut}(\mathcal{X})$ such that $P_2 = \phi \cdot P_1$. Similarly, two random graph models (W_1, P_1) and (W_2, P_2) on $\mathcal{X}$ are said to be isomorphic if there is a ϕ in $\mathrm{Aut}(\mathcal{X})$ such that $(W_2, P_2) = (\phi \cdot W_1, \phi \cdot P_1)$, in this case, we will note $(W_2, P_2) = \phi \cdot (W_1, P_1)$.

3 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 $G \in G(V)$ with $|V| = n$, and $Z = Z^{(0)} \in \mathbb{R}^{n \times d_0}$ be a signal on it. There are L operators $(F^{(l)})_{1 \le l \le L}$ such that, at each layer, the update $Z^{(l+1)}$ of the signal is computed node-wise by:

$$z_i^{(l+1)} = F^{(l+1)}\left(z_i^{(l)}, \{\!\{\left(z_j^{(l)}, w_{i,j}\right)\}\!\}_{j \in \mathcal{N}(i)}\right) \in \mathbb{R}^{d_{l+1}}. \quad (4)$$

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. As a result, the full network is *equivariant* to permutations of the graph, that is, consistent with graph isomorphism.

Proposition 3.1. *Call $\Theta_G(Z) = Z^{(L)}$. For all $Z \in \mathbb{R}^{n \times d_0}$, $\sigma \in S_n$, we have $\Theta_{\sigma \cdot G}(\sigma \cdot Z) = \sigma \cdot \Theta_G(Z)$.*

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

$$c_{i,j}^{(l+1)} = c^{(l+1)}\left(z_i^{(l)}, z_j^{(l)}, w_{ij}\right) \quad (5)$$

such that (4) can be rewritten as

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

Examples.

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. *Degree Normalized Convolution:* $z_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \frac{w_{i,j}}{\sum_{k \in \mathcal{N}(i)} w_{i,k}} \psi^{(l+1)}(z_j^{(l)})$.
3. *Graph Attention (GAT): the $c^{(l)}$ from (5) are learnable:* $z_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \frac{c_{i,j}^{(l+1)}}{\sum_{k \in \mathcal{N}(v_i)} c_{i,k}^{(l+1)}} \psi^{(l+1)}(z_j^{(l)})$.
4. *Max Convolution:* $z_i^{(l+1)} = \max_{j \in \mathcal{N}(i)} w_{i,j} \psi^{(l+1)}(z_j^{(l)})$.

4 Continuous MPGNN

As we will see in the next section, when the number of nodes grows, MPGNNs will often converge to "continuous models" on random graphs. Let (W, P) be a random graph model. A L-layer *continuous* MPGNN (cMPGNN) propagates a function over the latent space $\mathcal{X}$, using operators $\mathcal{F}_P^{(l+1)}$ that take a vector and a function as input and output a vector. For an input $f = f^{(0)} \in \mathcal{C}(\mathcal{X}, \mathbb{R}^{d_0})$, $f^{(l+1)}$ is computed by:

$$f^{(l+1)}(x) = \mathcal{F}_P^{(l+1)}\left(f^{(l)}(x), \left(f^{(l)}, W(x, \cdot)\right)\right) \in \mathbb{R}^{d_{l+1}}. \quad (6)$$

For notational convenience, we overload the notations $\mathcal{F}_p^{(l)}$ as

$$\mathcal{F}_P^{(l)}(f, W) : x \mapsto \mathcal{F}_P^{(l)}(f(x), (f, W(x, \cdot))). \quad (7)$$

Such, Eq. (6) now writes as

$$f^{(l+1)}(x) = \mathcal{F}_P^{(l+1)}(f^{(l)}, W)(x). \quad (8)$$

Naturally, we also demand the cMPGNN to be equivariant to random graph model isomorphisms. To that extent, we need the following assumption on the operators $\mathcal{F}_P^{(l)}$:

Assumption 4.1. *There is a subgroup $H \subset \mathrm{Aut}(\mathcal{X})$ such that $\forall 1 \le l \le L$, $\forall f \in \mathcal{C}(\mathcal{X}, \mathbb{R}^{d_l})$, $\forall \phi \in H$:*

$$\mathcal{F}_{\phi \cdot P}^{(l)}(\phi \cdot f, \phi \cdot W) = \phi \cdot \mathcal{F}_P^{(l)}(f, W).$$

Ass. 4.1 is inspired by the classical change of variable formula in Lebesgue integration: for any ϕ bijective and measurable and any measurable map f, $\int f\, dP = \int \phi \cdot f\, d(\phi \cdot P)$.

Proposition 4.1. *Call $\Theta_{W,P}(f) = f^{(L)}$. Under Ass. 4.1, for any f and any $\phi \in H$, $\Theta_{\phi \cdot (W,P)}(\phi \cdot f) = \phi \cdot \Theta_{W,P}(f)$.*

Ideally, one would like $H = \mathrm{Aut}(\mathcal{X})$. However, we will see with Ex. 4 that this is not always possible and one may have to restrict equivariance to a subgroup of $\mathrm{Aut}(\mathcal{X})$.

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

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

d. Max Convolution:
$f^{(l+1)}(x) = \operatorname*{ess\,sup}_P W(x,.)\psi^{(l+1)}(f^{(l)}(.)).$

5 Construction of the cMPGNN

Let (W,P) be a random graph model and $f \in \mathcal{C}(\mathcal{X},\mathbb{R}^d)$. In this subsection, we consider a single layer of a MPGNN applied on a random graph $G_n \sim \mathcal{G}_n(W,P)$ and input matrix $S_X f$ as node features. We present a canonical way of defining a limit corresponding cMPGNN layer on (W,P) with input map f, under some convergence assumptions that will be satisfied in our examples. Since there is no multi layers in this section, we drop the superscript indexation.

Definition 5.1. *Let F be a MPGNN layer and (W,P) a random graph model. For $f \in \mathcal{C}(\mathcal{X},\mathbb{R}^d)$ we define the sequence*

$$F_{P,n}(f,W): \\ x \mapsto \mathbb{E}_{X_1,\dots,X_n}\left[F\left(f(x), \{\!\{(f(X_k), W(x,X_k))\}\!\}_{1\le k\le n}\right)\right] \tag{9}$$

where the expected value is taken over all the $X_1,\dots,X_n \overset{iid}{\sim} P$. Let $\mathcal{F}$ be an operator of the form (8) *taking value in $\mathcal{C}(\mathcal{X},\mathbb{R}^{d'})$ and suppose we have a non-trivial subgroup $H \subset \mathrm{Aut}(\mathcal{X})$ such that for any $f \in \mathcal{C}(\mathcal{X},\mathbb{R}^d)$, for any $\phi \in H$,*

$$F_{\phi\cdot P,n}(\phi\cdot f, \phi\cdot W) \xrightarrow{\|\cdot\|_\infty} \mathcal{F}_{\phi\cdot P}(\phi\cdot f, \phi\cdot W). \tag{10}$$

*Then we say that $\mathcal{F}$ is the **continuous counterpart** of F.*

This $\mathcal{F}$ is a good candidate to be a cMPGNN. Indeed, it satisfies 4.1 on the ϕ for which it is well defined.

Proposition 5.1. *Let $\mathcal{F}$ be the continuous counterpart of F as defined in Def. 5.1. Then it satisfies Ass. 4.1 for any $\phi \in H$.*

We prove that the continuous equivalents of our previous examples 1, 2, 3 and 4 are respectively a, b, c and d:

Examples.

1-a. a. is the continuous counterpart of 1. for $H = \mathrm{Aut}(\mathcal{X})$

2-b. If ψ is bounded and $W \ge \alpha$ for some $\alpha > 0$. Then b. is the continuous counterpart of 2. for the full $H = \mathrm{Aut}(\mathcal{X})$.

3-c. Call $V(x,y) = c(f(x), f(y), W(x,y))$, if ψ is bounded and $\alpha \le V \le \beta$ a. s for some $0 < \alpha \le \beta$, then c. is the continuous counterpart of 3. for the full $H = \mathrm{Aut}(\mathcal{X})$.

4-d. Suppose that W and ψ, are continuous and that the measure P is strictly positive on $\mathcal{X}$ i.e*, any nonvoid relative open of $\mathcal{X}$ has a strictly positive measure by P. Then d. is the continuous counterpart of 4. for $H = \mathrm{Hom}(\mathcal{X})$: the subgroup of $\mathrm{Aut}(\mathcal{X})$ made of the $\phi \in \mathrm{Aut}(\mathcal{X})$ that are homeomorphisms.*

The rates of convergence are given in the next section.

6 Convergence of MPGNN to cMPGNN

Let (W,P) be a random graph model, $(G_n)_{n\ge1}$ be a sequence of random graphs drawn from $\mathcal{G}_n(W,P)$. We consider a MPGNN $(F^{(l)})_{1\le l\le L}$ and its continuous counterparts $(\mathcal{F}^{(l)})_{1\le l\le L}$. For an $f \in \mathcal{C}(\mathcal{X},\mathbb{R}^{d_0})$, does the MPGNN on G_n with input signal $S_X f$ actually converge to the cMPGNN on (W,P) with input signal f? If yes, at which speed? In this section we provide non asymptotic bounds with high probability to quantify this convergence.
Our main theorem states that, under mild regularity condition, with high probability, $(S_X(f))_i^{(L)}$ is close to $f^{(L)}(X_i)$. To compare the output of this signal through the discrete network to the output of f through the continuous counterpart, we shall use the following Maximum Absolute Error (MAE):

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

6.1 Bounded differences method

Our main result relies on the so called McDiarmid concentration inequality, which relies on the following property.

Definition 6.1 (Bounded Differences Property)**.** *Let $f : \mathcal{E}^n \to \mathbb{R}^p$ be a function of n variables. We say that f has the bounded differences property if there exist n nonnegative constants $c_1,\dots,c_n$ such that for any $1 \le i \le n$:*

$$\|f(x_1\dots,x_i,\dots,x_n) - f(x_1,\dots,x_i',\dots,x_n)\|_\infty \le c_i, \tag{11}$$

for all $x_1,\dots,x_i,x_i',\dots,x_n$.

Fix, $x_1 \in \mathcal{X}$, we are interested at the bounded differences of

$$F^{(l)}(f^{(l-1)}(x_1), \{\!\{(f^{(l-1)}(x_k), W(x_1,x_k))\}\!\}_{k\ge2}) \tag{12}$$

as a map of the $n-1$ variables $x_2,\dots,x_n$. If $c_1,\dots,c_n$ satisfy (11), since (12) is invariant to the permutation of $x_2,\dots,x_n$, they can be taken all equal. We call $D_n^{(l)}(x_1) = c_1 = \cdots = c_n$. Moreover, since (12) is continuous as a function of x_1, it is also bounded by compactness, we define

$$D_n^{(l)} = \sup_{x_1\in\mathcal{X}} D_n^{(l)}(x_1) \tag{13}$$

Since for all l, $\mathcal{F}^{(l)}$ is the continuous counterpart of $F^{(l)}$, using the notations of Def. 5.1, we let $(a_n^{(l)})$ be a sequence of positive reals such that $a_n^{(l)} \to 0$ and for all n

$$\|F_{P,n}^{(l)}(f,W) - \mathcal{F}_P^{(l)}(f,W)\|_\infty \le a_n^{(l)}. \tag{14}$$

Next, we suppose, that (13), (14) and Def. 5.1 are satisfied. Plus that the $F^{(l)}$ have some "Lipschitz-like" smoothness.

Assumption 6.1. *(i) For all $1 \leq l \leq L$, we endow $\mathbb{R}^{d_{l-1}} \times [0,1]$ with the norm $\|(y,t)\|_1 = \|y\|_\infty + |t|$ and call δ_1 the corresponding distance on multisets as defined in* (1)*. Let $x, x' \in \mathbb{R}^{d_{l-1}}$ and $\mathcal{M}, \mathcal{M}'$ be two multisets of same cardinal n containing elements of $\mathbb{R}^{d_{l-1}} \times [0,1]$, then there exist two constants $\mu^{(l)} \geq 0$ and $\lambda^{(l)}_{F,n} > 0$ such that:*

$$\begin{aligned} &\left\| F^{(l)}(x, \mathcal{M}) - F^{(l)}(x', \mathcal{M}') \right\|_\infty \\ &\leq \mu^{(l)}_F \|x - x'\|_\infty + \lambda^{(l)}_{F,n} \delta_1(\mathcal{M}, \mathcal{M}'). \end{aligned} \tag{15}$$

(ii) The sequence $(\lambda^{(l)}_{F,n})$ is bounded over n.

(iii) We have some $D^{(l)}_n$ and $a^{(l)}_n$ as defined in (13) *and* (14)*.*

(iv) The $\mathcal{F}^{(l)}$ are the continuous counterparts *of the $F^{(l)}$).*

Theorem 6.1. *Under Ass. 6.1 for any $\rho > 0$,*

$$\mathrm{MAE}_X(f) \lesssim L D_n \sqrt{n \ln\left(\frac{n 2^L d_{max}}{\rho}\right)} + L a_{n-1}. \tag{16}$$

with probability at least $1 - \rho$. In (16)*, $D_n = \max_l D^{(l)}_n$, $d_{max} = \max_l d_l$, $a_n = \max_l a^{(l)}_n$ and $\lesssim$ hides some multiplicative constants which do not depend on 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.

Corollary 6.1.1. *If $D_n = o(1/\sqrt{n \ln n})$ then $\mathrm{MAE}_X(f)$ converges to zero in probability.*

This table sums up the results on the examples.

Example	D_n	a_n	Cvrg by Th. 6.1
1-a	$O(1/n)$	0	✓
2-b	$O(1/n)$	$O(1/\sqrt{n})$	✓
3-c	$O(1/n)$	$O(1/\sqrt{n})$	✓
4-d	$\Omega(1)$	–	✗

6.2 The case of max

It turns out that MPGNNs with max aggregation do not have sharp bounded differences. Nevertheless, we provide a bound for its convergence based on other concentration inequalities.

Definition 6.2 (Volume retaining property)**.** *We say that the probability space $(\mathcal{X}, P)$ has the (r_0, κ)-volume retaining property if for any $r \geq r_0$ and for any $x \in \mathcal{X}$,*

$$P(\mathcal{B}(x,r) \cap \mathcal{X}) \geq \kappa m(\mathcal{B}(x,r)) \tag{17}$$

Where $\mathcal{B}(x,r)$ is the ball of center x and radius r and m is the classical d-dimensional Lebesgue measure in $\mathbb{R}^d$

Clearly, volume-retention implies strict positiveness of the measure. This property ensures that the measure of a relative small ball centered in a point of $\mathcal{X}$ is at least a portion of the volume of that ball in $\mathbb{R}^d$. As an example, it is easy to check that $([0,1]^d, m)$ has the $(1, 1/2^d)$-volume retaining property. For a volume retaining probability space, we obtain a new concentration inequality

Lemma 6.1. *Let $g : \mathcal{X}^2 \to \mathbb{R}^p$ be λ_g Lipschitz and $(\mathcal{X}, P)$ have the (r, κ)-volume retaining property for some $r, \kappa > 0$.*

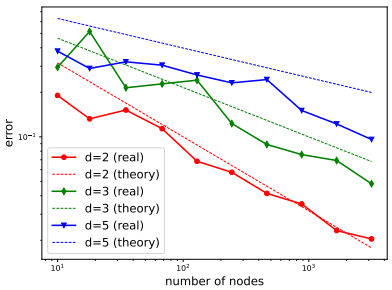

Figure 1: Log-scaled illustration of (18) for $d = 2, 3, 5$. The dashed line represents $(1/n)^{1/d}$ and the plain line the MAE. $L = 3$, $\mathcal{X} = [0,1]^d$ with the Lebesgue measure, $W(x,y) = \exp(-\|x - y\|^2)$ and $f = 1$.

Then for any $\rho > 0$, for any random variables $X_1, \ldots, X_n \overset{iid}{\sim} P$, with probability at least $1 - \rho$:

$$\| \max_{1 \leq i \leq n} g(X_i) - \sup_{x \in \mathcal{X}} g(x) \|_\infty \leq \frac{\lambda_g}{2} \left(\frac{\ln(p/\rho)}{n\kappa} \right)^{1/d}.$$

Armed with this lemma, we are ready to state the non asymptotic bound for a MPGNN with max aggregation:

Theorem 6.2. *For a Max Conv. MPGNN, assume that $(\mathcal{X}, P)$ has the (r, κ)-volume retaining property. Then for any $\rho > 0$, we have with probability at least $1 - \rho$,*

$$MAE(f) \lesssim L \left(\frac{1}{n-1} \ln\left(\frac{2^{L-1} n d_{\max}}{\rho} \right) \right)^{1/d}. \tag{18}$$

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 (16) for a large d.

References

[1] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021. arXiv:2104.13478.

[2] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2009.

[3] 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*, 2020.

[4] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, 2015. arXiv: 1312.2050.

[5] László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, Providence, Rhode Island, December 2012.

[6] N. Tremblay S. Vaiter M. Cordonnier, N. Keriven. Convergence of graph neural networks with generic aggregation functions on random graphs. 2023. `https://hal.science/hal-04059402`.

[7] Sohir Maskey, Yunseok Lee, Ron Levie, and Gitta Kutyniok. Stability and Generalization Capabilities of Message Passing Graph Neural Networks. 2022. arXiv: 2202.00645.

[8] Luana Ruiz, Luiz Chamon, and Alejandro Ribeiro. Graphon Neural Networks and the Transferability of Graph Neural Networks. In *Advances in Neural Information Processing Systems*, 2020.

[9] Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon Signal Processing. *IEEE Trans. Signal Process.*, 69:4961–4976, 2021. arXiv: 2003.05030.

[10] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks? 2019. arXiv: 1810.00826.