Fine-grained Expressivity of Graph Neural Networks

Jan Böker, Ron Levie, Ningyuan Huang, Soledad Villar and Christopher Morris

Introducción

El paper es muy matemático, y requiere background de cosas que yo (y creo que la mayoría) no tiene:

Introducción

El paper es muy matemático, y requiere background de cosas que yo (y creo que la mayoría) no tiene:

- Análisis real

Introducción

El paper es muy matemático, y requiere background de cosas que yo (y creo que la mayoría) no tiene:

- Análisis real
- Teoría de la medida

Introducción

El paper es muy matemático, y requiere background de cosas que yo (y creo que la mayoría) no tiene:

- Análisis real
- Teoría de la medida
- Topología

Introducción

El paper es muy matemático, y requiere background de cosas que yo (y creo que la mayoría) no tiene:

- Análisis real
- Teoría de la medida
- Topología

A veces voy a hacer explicaciones no necesariamente tan formales matemáticamente, para que sea más sencillo de entender.

Introducción

La motivación del paper es muy clara.

Introducción

La motivación del paper es muy clara.

Muchos trabajos caracterizan la expresividad de las GNNs usando el test de 1-WL.

Introducción

La motivación del paper es muy clara.

Muchos trabajos caracterizan la expresividad de las GNNs usando el test de 1-WL.

Ésto nos permite caracterizar la abilidad de las MPNNs en *distinguir* grafos... pero no de *qué tan distintos* son.

Introducción

Lo que queremos, es una caracterización *continua* de la expresividad.

Introducción

Lo que queremos, es una caracterización *continua* de la expresividad.

Vamos a demostrar un **Teorema de aproximación universal**.

Introducción

Es decir que (casi) toda función definida sobre grafos, es aproximable (lo suficientemente cerca) mediante MPNNs.

Introducción

Es más, en el paper, extienden el resultado no sólo para **grafos**, sino que también para **graphons**.

Graphons

Los Grafos son un subconjunto de los **Graphons**. La diferencia es que los Los graphons es un conjunto más grande, y **completo** (a diferencia de los grafos).

Graphons

Completitud: Los grafos no son un conjunto completo.

Esto significa que podemos tomar una secuencia de grafos, que convergen, pero no a un grafo! (sino que a un graphon).

Graphons

Completitud: Los grafos no son un conjunto completo.

Esto significa que podemos tomar una secuencia de grafos, que convergen, pero no a un grafo! (sino que a un graphon).

Qué significa esto???

Graphons

Completitud: Los grafos no son un conjunto completo.

Esto significa que podemos tomar una secuencia de grafos, que convergen, pero no a un grafo! (sino que a un graphon).

Qué significa esto??? Mejor veamos esto con otros conjuntos.

Graphons

Completitud: Los **racionales** no son un conjunto completo.

Esto significa que podemos tomar una secuencia de **racionales**, que convergen, pero no a un **racional**! (sino que a un **real**)

Graphons

Por ejemplo, definamos $S_n = (1 + \frac{1}{n})^n$.

Graphons

Por ejemplo, definamos $S_n = (1 + \frac{1}{n})^n$.

Sabemos que $lim_{n \to \infty} S_n =$

Graphons

Por ejemplo, definamos $S_n = (1 + \frac{1}{n})^n$.

Sabemos que $lim_{n \to \infty} S_n = e$

Graphons

Por ejemplo, definamos $S_n = (1 + \frac{1}{n})^n$.

Sabemos que $lim_{n \to \infty} S_n = e$

e no es racional, pero cada número dentro de la secuencia sí lo es!

Graphons

Por ejemplo, definamos $S_n = (1 + \frac{1}{n})^n$.

Sabemos que $lim_{n \to \infty} S_n = e$

e no es racional, pero cada número dentro de la secuencia sí lo es!

De la misma manera, existen secuencias de grafos que pueden converger a cosas que no son grafos (sino que graphons).

Graphons

Qué significa que una secuencia de grafos converja?

Graphons

Qué significa que una secuencia de grafos converja?

Podemos representar un grafo como su *matriz de adyacencia*.
Entonces una secuencia de grafos es una secuencia de matrices.

Graphons

Qué significa que una secuencia de grafos converja?

Podemos representar un grafo como su *matriz de adyacencia*.
Entonces una secuencia de grafos es una secuencia de matrices.

Pero necesitamos una noción de distancia.

Graphons

Qué significa que una secuencia de grafos converja?

Podemos representar un grafo como su *matriz de adyacencia*. Entonces una secuencia de grafos es una secuencia de matrices.

Pero necesitamos una noción de distancia. Para esto está la *cut distance*.

Graphons

Cut distance

Graphons

Cut distance

Queremos tener una noción de distancia entre 2 grafos distintos.
La vamos hacer usando *cortes de los grafos*.

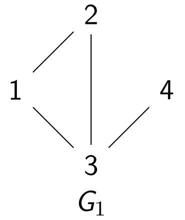
Graphons

Cut distance
Queremos tener una noción de distancia entre 2 grafos distintos.
La vamos hacer usando *cortes de los grafos*.

G_1 G_2

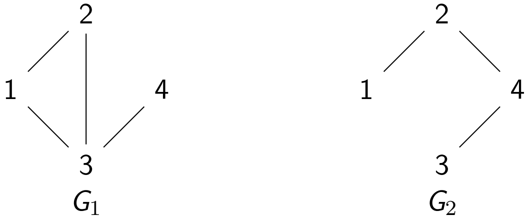
- Podemos reordenar los nodos como queramos.
- Queremos encontrar el órden tal que cualquier *corte* en el grafo, sea similar en ambos.

Graphons



Los dos grafos son iguales! Pues existe un reordenamiento de los nodos tal que no podemos hacer un *corte* que distinga entre los dos.

Graphons



Pero en este ejemplo, con cualquier órden, podemos encontrar un corte que distingue los grafos. La *cut distance* formaliza esta idea.

Graphons

Lo interesante, es que podemos computar el **cut distance** únicamente a partir de las matrices de adyacencia.

Graphons

Lo interesante, es que podemos computar el **cut distance** únicamente a partir de las matrices de adyacencia.

Y la definición de **cut distance** va a poder ser generalizada para *graphons* también.

Medidas

Vamos a también necesitar la noción de una *medida*.

Medidas

Vamos a también necesitar la noción de una *medida*.

Intuitivamente, una medida es una función que mapea un conjunto de valores, a un valor numérico (que sigue ciertas propiedades de consitencia).

Medidas

Vamos a también necesitar la noción de una *medida*.

Intuitivamente, una medida es una función que mapea un conjunto de valores, a un valor numérico (que sigue ciertas propiedades de consitencia).

Si nuestro espacio es X, una medida sobre los elementos de X es:

$$f : 2^X \rightarrow \mathbb{R}^+$$

Medidas

Por ejemplo, digamos que estamos en el espacio de $\mathbb{R}$.

Entonces, la medida intuitiva en este espacio funciona como esperaríamos ($f([0,2]) = 2$ por ejemplo).

Pero podríamos diseñar otras medidas que también "hacen sentido".

Medidas

El paper introduce el concepto de *iterated degree measures* (IDMs). Es una secuencia de medidas en grafos, definidas *recursivamente*.

Medidas

El paper introduce el concepto de *iterated degree measures* (IDMs). Es una secuencia de medidas en grafos, definidas *recursivamente*.

- $\mathbb{M}_0 := \{1\}$

Medidas

El paper introduce el concepto de *iterated degree measures* (IDMs). Es una secuencia de medidas en grafos, definidas *recursivamente*.

- $\mathbb{M}_0 := \{1\}$
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$

Medidas

El paper introduce el concepto de *iterated degree measures* (IDMs). Es una secuencia de medidas en grafos, definidas *recursivamente*.

- $\mathbb{M}_0 := \{1\}$
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$
  - Son todas las medidas de peso total menor o igual a 1, sobre el conjunto anterior.

Medidas

El paper introduce el concepto de *iterated degree measures* (IDMs). Es una secuencia de medidas en grafos, definidas *recursivamente*.

- $\mathbb{M}_0 := \{1\}$
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$
  - Son todas las medidas de peso total menor o igual a 1, sobre el conjunto anterior.

Es enredado de entender sin tanto background en teoría de la medida 😬

Medidas

Pero la inspiración de la necesidad de las IDMs es para generalizar la noción del test 1-WL (y sus caracterizaciones) a los graphons.

Medidas

Conectan las medidas en $\mathbb{M}_L$ con los embeddings obtenidos de cada nodo en una MPNN de L layers.

Medidas

Conectan las medidas en $\mathbb{M}_L$ con los embeddings obtenidos de cada nodo en una MPNN de L layers.

La interpretación que podemos hacer:

Medidas

Conectan las medidas en $\mathbb{M}_L$ con los embeddings obtenidos de cada nodo en una MPNN de L layers.

La interpretación que podemos hacer:

- Podemos interpretar a una MPNNs como una especie de medida sobre el conjunto de nodos del grafo

Medidas

Conectan las medidas en $\mathbb{M}_L$ con los embeddings obtenidos de cada nodo en una MPNN de L layers.

La interpretación que podemos hacer:

- Podemos interpretar a una MPNNs como una especie de medida sobre el conjunto de nodos del grafo
- A cierta capa L, esta MPNNs puede expresar toda medida en $\mathbb{M}_L$.

Aproximación Universal

Vamos a demostrar un teorema de aproximación universal para MPNNs en IDMs.

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$ and $\mathcal{N}\mathcal{N}_L^1$ is dense in $C(\mathscr{P}(\mathbb{M}_L), \mathbb{R})$.

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$ and $\mathcal{N}\mathcal{N}_L^1$ is dense in $C(\mathscr{P}(\mathbb{M}_L), \mathbb{R})$.

Vamos paso a paso.

Aproximación Universal

Empecemos con definiciones.

Aproximación Universal

Empecemos con definiciones.

$C(\mathbb{M}_L, \mathbb{R}^n)$ es el conjunto de funciones continuas $f : \mathbb{M}_L \to \mathbb{R}^n$.

Aproximación Universal

Empecemos con definiciones.

$C(\mathbb{M}_L, \mathbb{R}^n)$ es el conjunto de funciones continuas $f : \mathbb{M}_L \to \mathbb{R}^n$.

En donde $\mathbb{M}_L$ es el espacio de IDMs después de L layers (que en mi intuición lo podemos ver como las embeddings que podemos formar después de L layers).

Aproximación Universal

Sea $\mathcal{N}_L^n \subseteq C(\mathbb{M}_L, \mathbb{R}^n)$, el subconjunto de funciones continuas que efectivamente un MPNNs con L layers puede formar.

Aproximación Universal

Vamos a acotar el teorema que demuestran a:

Aproximación Universal

Vamos a acotar el teorema que demuestran a:

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$

Aproximación Universal

Vamos a acotar el teorema que demuestran a:

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$

Es decir, sólo nos interesa la demostración a MPNNs que mapean a los reales (y no vectores).

Aproximación Universal

Qué significa que un conjunto A sea denso en un conjunto B?

Aproximación Universal

Qué significa que un conjunto A sea denso en un conjunto B?

Significa que con elementos $a \in A$, podemos estar arbitrariamente cerca de cualquier elemento $b \in B$. Sería algo como:

Aproximación Universal

Qué significa que un conjunto A sea denso en un conjunto B?

Significa que con elementos $a \in A$, podemos estar arbitrariamente cerca de cualquier elemento $b \in B$. Sería algo como:

$$\forall \epsilon > 0 \;\; \forall b \in B \;\; \exists a \in A \;\; dist(a, b) < \epsilon$$

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$

Efectivamente esto implica que las MPNNs pueden ser aproximadoras universales de funciones continuas de $f : \mathbb{M}_L \to R$.

Aproximación Universal

Cómo se demuestra? En el paper lo dejan casi que como ejercicio para el lector.

Aproximación Universal

Cómo se demuestra? En el paper lo dejan casi que como ejercicio para el lector.

Dicen que es una aplicación directa del teorema de Stone-Weierstrass.

Aproximación Universal

Stone-Weierstrass:

Aproximación Universal

Stone-Weierstrass: cualquier función continua aplicada sobre un espacio *compacto*, puede ser aproximada *uniformemente* por subálgebras que distinguen puntos.

Aproximación Universal

Stone-Weierstrass: cualquier función continua aplicada sobre un espacio *compacto*, puede ser aproximada *uniformemente* por subálgebras que distinguen puntos.

Los autores mencionan que se puede aplicar inductivamente Stone-Weierstrass sobre la definición de las IDMs para llegar a este resultado.

Aproximación Universal

Por qué es eso?

Aproximación Universal

Por qué es eso?

- El conjunto $\mathbb{M}_0 := \{1\}$ es trivialmente compacto.

Aproximación Universal

Por qué es eso?

- El conjunto $\mathbb{M}_0 := \{1\}$ es trivialmente compacto.
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$

Aproximación Universal

Por qué es eso?

- El conjunto $\mathbb{M}_0 := \{1\}$ es trivialmente compacto.
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$
  - Son todas las medidas de peso total menor o igual a 1, sobre el conjunto anterior. Esto parece mantener la compacidad?

Aproximación Universal

Por qué es eso?

- El conjunto $\mathbb{M}_0 := \{1\}$ es trivialmente compacto.
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$
  - Son todas las medidas de peso total menor o igual a 1, sobre el conjunto anterior. Esto parece mantener la compacidad?

Y con ello ya tenemos todo, pues implica que un $\mathbb{M}_L$ arbitrario es compacto, y por Stone-Weierstrass, cualquier función continua aplicada acá se puede ser aproximada

Aproximación Universal

Por qué es eso?

- El conjunto $\mathbb{M}_0 := \{1\}$ es trivialmente compacto.
- $\mathbb{M}_{h+1} := \mathcal{M}_{\leq 1}(\mathbb{M}_h)$
  - Son todas las medidas de peso total menor o igual a 1, sobre el conjunto anterior. Esto parece mantener la compacidad?

Y con ello ya tenemos todo, pues implica que un $\mathbb{M}_L$ arbitrario es compacto, y por Stone-Weierstrass, cualquier función continua aplicada acá se puede ser aproximada ~~*uniformemente* por subálgebras que distinguen puntos~~.

Aproximación Universal

Sólo mostramos la primera parte del teorema.

Aproximación Universal

Sólo mostramos la primera parte del teorema.

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$ and $\mathcal{N}\mathcal{N}_L^1$ is dense in $C(\mathscr{P}(\mathbb{M}_L), \mathbb{R})$.

Aproximación Universal

$$\mathcal{N}\mathcal{N}_L^n \coloneqq \{\mathbf{h}_- \mid \boldsymbol{\varphi}\ L\text{-layer MPNN model}, \psi\colon \mathbb{R}^{d_L} \to \mathbb{R}^n \text{ Lipschitz}\} \subseteq C(\mathscr{P}(\mathbb{M}_L), \mathbb{R}^n)$$

Aproximación Universal

$$\mathcal{NN}_L^n := \{\mathbf{h}_- \mid \boldsymbol{\varphi}\ L\text{-layer MPNN model}, \psi\colon \mathbb{R}^{d_L} \to \mathbb{R}^n \text{ Lipschitz}\} \subseteq C(\mathscr{P}(\mathbb{M}_L), \mathbb{R}^n)$$

$\mathcal{P}(\mathbb{M}_L)$ es el espacio de todas las medidas de probabilidad de $\mathbb{M}_L$.

Aproximación Universal

$$\mathcal{NN}_L^n \coloneqq \{\mathbf{h}_- \mid \boldsymbol{\varphi}\ L\text{-layer MPNN model, } \psi\colon \mathbb{R}^{d_L} \to \mathbb{R}^n \text{ Lipschitz}\} \subseteq C(\mathscr{P}(\mathbb{M}_L), \mathbb{R}^n)$$

$\mathcal{P}(\mathbb{M}_L)$ es el espacio de todas las medidas de probabilidad de $\mathbb{M}_L$.

Son las MPNNs sobre probabilidades de embeddings a valores en $\mathbb{R}^n$.

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$ and $\mathcal{NN}_L^1$ is dense in $C(\mathscr{P}(\mathbb{M}_L), \mathbb{R})$.

Aproximación Universal

Theorem 4. Let $0 \leq L \leq \infty$. Then, $\mathcal{N}_L^1$ is dense in $C(\mathbb{M}_L, \mathbb{R})$ and $\mathcal{N}\mathcal{N}_L^1$ is dense in $C(\mathscr{P}(\mathbb{M}_L), \mathbb{R})$.

En vez de MPNNs a nivel de grafos, esto lo muestra a nivel de *graphons* (que es una conclusión más fuerte).

Recap: principales resultados

Esos eran los principales resultados del paper, que después mencionan como corolarios:

Theorem 1 (informal)**.** The following are equivalent for all graphons U and W:

1. U and W are close in δ_{P} (or alternatively δ_{W}).
2. U and W are close in $\delta_{\square}^{\mathcal{T}}$.
3. MPNN outputs on U and W are close for all MPNNs with Lipschitz constant C and L layers.
4. Homomorphism densities in U and W are close for all trees up to order k.

Más resultados

El paper tiene otros resultados que valen la pena mencionar.

Más resultados

Definieron métricas para grafos/graphons que están bounded por el *cut distance* y demostraron que las MPNNs no pueden "separar" demasiado dos puntos (grafos) bajo estas métricas.

Más resultados

Definieron métricas para grafos/graphons que están bounded por el *cut distance* y demostraron que las MPNNs no pueden "separar" demasiado dos puntos (grafos) bajo estas métricas.

- Mostraron que sus métricas están bien correlacionadas con las distancias entre embeddings de dos grafos en la práctica.

Más resultados

Un resultado interesante es que mostraron que las MPNNs ganan expresividad a medida que ganan más capas, **sin la necesidad de afirmar ningún tipo de entrenamiento sobre ellas**.

Más resultados

Un resultado interesante es que mostraron que las MPNNs ganan expresividad a medida que ganan más capas, **sin la necesidad de afirmar ningún tipo de entrenamiento sobre ellas**.

- Puede una MPNN iniciada aleatoriamente ser útil sin entrenarla (y sólo entrenar la capa de clasificación sobre los embeddings finales)?

Más resultados

Accuracy ↑	MUTAG	IMDB-BINARY	IMDB-MULTI	NCI1	PROTEINS	REDDIT-BINARY
GIN-m (trained)	79.01 ± 2.24	69.96 ± 1.43	46.29 ± 0.76	**78.61 ± 0.34**	**73.51 ± 0.47**	**89.73 ± 0.37**
GIN-m (untrained)	**82.56 ± 3.12**	**70.70 ± 0.60**	**47.59 ± 0.95**	77.82 ± 0.55	73.45 ± 0.30	82.32 ± 0.45
GraphConv-m (trained)	**81.62 ± 2.08**	59.14 ± 1.93	38.75 ± 1.62	**63.28 ± 0.6**	71.49 ± 0.67	**82.4 ± 0.19**
GraphConv-m (untrained)	78.03 ± 1.57	**65.77 ± 1.32**	**43.29 ± 0.96**	62.36 ± 0.45	**71.83 ± 0.42**	77.15 ± 0.29

Más resultados

Accuracy ↑	MUTAG	IMDB-BINARY	IMDB-MULTI	NCI1	PROTEINS	REDDIT-BINARY
GIN-m (trained)	79.01 ± 2.24	69.96 ± 1.43	46.29 ± 0.76	**78.61 ± 0.34**	**73.51 ± 0.47**	**89.73 ± 0.37**
GIN-m (untrained)	**82.56 ± 3.12**	**70.70 ± 0.60**	**47.59 ± 0.95**	77.82 ± 0.55	73.45 ± 0.30	82.32 ± 0.45
GraphConv-m (trained)	**81.62 ± 2.08**	59.14 ± 1.93	38.75 ± 1.62	**63.28 ± 0.6**	71.49 ± 0.67	**82.4 ± 0.19**
GraphConv-m (untrained)	78.03 ± 1.57	**65.77 ± 1.32**	**43.29 ± 0.96**	62.36 ± 0.45	**71.83 ± 0.42**	77.15 ± 0.29

- Impresionantemente, funciona bastante bien!

Más resultados

Hay muchos detalles que no mencioné, y que tampoco me para entender en profundidad. El paper es increíblemente completo y denso en resultados.

Fine-grained Expressivity of Graph Neural Networks

Jan Böker, Ron Levie, Ningyuan Huang, Soledad Villar and Christopher Morris