Can graph neural networks count substructures?

Soledad Villar

Applied Mathematics and Statistics
Mathematical Institute for Data Science

Johns Hopkins University

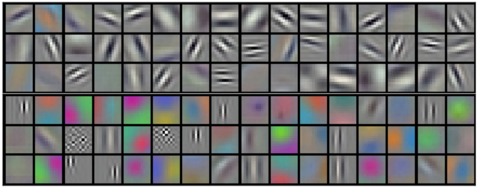
Based on joint work with Zhengdao Chen, Lei Chen and Joan Bruna

String-data 2020

Motivation: deep learning beyond images

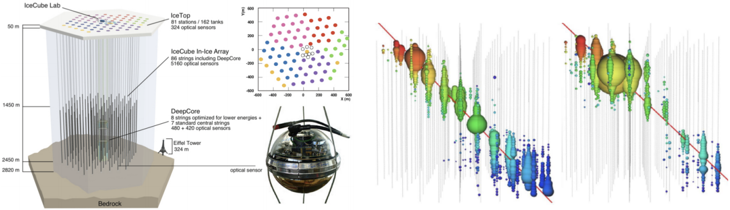
Deep learning is extremely successful at certain tasks.

CNNs use filters that seem to exploit intrinsic symmetries of images.



source: Jonathan Hui

Not all data are images. Learning on graphs.



source: Choma et al '18. Neutrino detection with GNNs

Neural networks on graphs

Input: $G = (V, E, x_v, x_e)$. Output: $f_\Theta(G)$ embedding of the graph.

Fundamental property: equivariance with respect to permutations.

Neural networks on graphs

Input: $G = (V, E, x_v, x_e)$. Output: $f_\Theta(G)$ embedding of the graph.

Fundamental property: equivariance with respect to permutations.

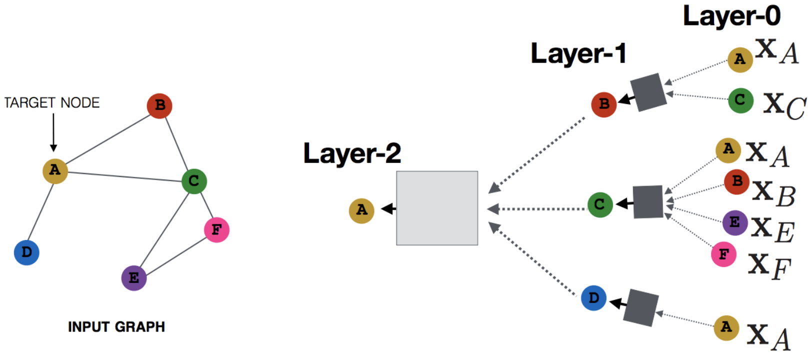
Many types of GNN architectures

- Message passing neural networks
  [Gilmer et al '16, Hamilton et at '17]
- Graph convolutional networks
  [Duvenaud et al '15, Defferrard et al '16, Kipf & Welling, '17]
- Spectral graph neural networks [Chen et al '19]
- k-invariant networks [Maron et al '19]
- ...

Message passing neural networks

$$a_v^{(k)} = \text{AGGREGATE}^{(k)}\left(\{h_u^{(k-1)} : u \in \mathcal{N}(u)\}\right)$$

$$h_v^{(k)} = \text{COMBINE}^{(k)}\left(h_v^{(k-1)}, a_v^{(k)}\right)$$

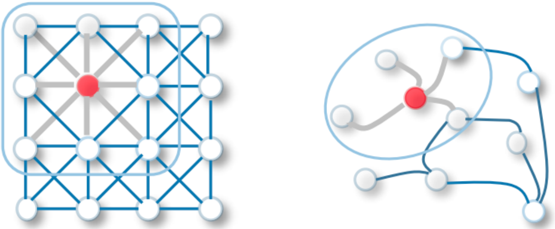


Gilmer et al '16, Hamilton et at '17, image: Leskovec '18

Graph convolutional networks

Classical convolutions: $H^{t+1} = \sigma(W^t H^t + b^t)$

Graph convolution: $H^{t+1} = \sigma(W^t H^t A)$
(A adjacency matrix of graph)



Duvenaud et al '15, Defferrard et al '16, Kipf & Welling, '17, image: Wu et. al '19

Spectral GNNs

Originally motivated by "Bethe Hessian" for clustering the stochastic block model

Graph with adjacency matrix A. Set $\mathcal{M} = \{I_n, D, A, A^2, \ldots, A^J\}$,

Combine graph operators $\mathcal{M}$ to produce a "good spectral method"

Scarselli, Tsoi, Hagenbuchner, Monfardini, 2009

Chen, Li, Bruna, 2017

Spectral GNNs

Originally motivated by "Bethe Hessian" for clustering the stochastic block model

Graph with adjacency matrix A. Set $\mathcal{M} = \{I_n, D, A, A^2, \ldots, A^J\}$,

Combine graph operators $\mathcal{M}$ to produce a "good spectral method"

Unroll as power method: $v^{t+1} = Mv^t \quad t = 1, \ldots, T$.
And overparametrize:

$$v^{t+1} = \left(\sum_{M \in \mathcal{M}} M v^t \theta_M \right) ,$$

with $v^t \in \mathbb{R}^{n \times d_t}$,
$\Theta = \{\theta_1^t, \ldots, \theta_{|\mathcal{M}|}^t\}_t$, $\theta_M^t \in \mathbb{R}^{d_t \times d_{t+1}}$ trainable parameters.

Scarselli, Tsoi, Hagenbuchner, Monfardini, 2009

Chen, Li, Bruna, 2017

Spectral GNNs

Originally motivated by "Bethe Hessian" for clustering the stochastic block model

Graph with adjacency matrix A. Set $\mathcal{M} = \{I_n, D, A, A^2, \ldots, A^J\}$,

Combine graph operators $\mathcal{M}$ to produce a "good spectral method"

Unroll as power method: $v^{t+1} = Mv^t \quad t = 1, \ldots, T$.
And overparametrize:

$$v_l^{t+1} = \left(\sum_{M \in \mathcal{M}} M v^t \theta_{M,l} \right), \; l = 1, \ldots, d_{t+1}$$

with $v^t \in \mathbb{R}^{n \times d_t}$,
$\Theta = \{\theta_1^t, \ldots, \theta_{|\mathcal{M}|}^t\}_t$, $\theta_M^t \in \mathbb{R}^{d_t \times d_{t+1}}$ trainable parameters.

Scarselli, Tsoi, Hagenbuchner, Monfardini, 2009

Chen, Li, Bruna, 2017

Spectral GNNs

Originally motivated by "Bethe Hessian" for clustering the stochastic block model

Graph with adjacency matrix A. Set $\mathcal{M} = \{I_n, D, A, A^2, \ldots, A^J\}$,

Combine graph operators $\mathcal{M}$ to produce a "good spectral method"

Unroll as power method: $v^{t+1} = Mv^t \quad t = 1, \ldots, T$.
And overparametrize:

$$v_l^{t+1} = \left(\sum_{M \in \mathcal{M}} M v^t \theta_{M,l}{}^t \right), \; l = 1, \ldots, d_{t+1}$$

with $v^t \in \mathbb{R}^{n \times d_t}$,
$\Theta = \{\theta_1^t, \ldots, \theta_{|\mathcal{M}|}^t\}_t$, $\theta_M^t \in \mathbb{R}^{d_t \times d_{t+1}}$ trainable parameters.

Scarselli, Tsoi, Hagenbuchner, Monfardini, 2009

Chen, Li, Bruna, 2017

Spectral GNNs

Originally motivated by "Bethe Hessian" for clustering the stochastic block model

Graph with adjacency matrix A. Set $\mathcal{M} = \{I_n, D, A, A^2, \ldots, A^J\}$,

Combine graph operators $\mathcal{M}$ to produce a "good spectral method"

Unroll as power method: $v^{t+1} = Mv^t \quad t = 1, \ldots, T$.
And overparametrize:

$$v_l^{t+1} = \rho\left(\sum_{M \in \mathcal{M}} M v^t \theta_{M,l}{}^t\right) \ , l = 1, \ldots, d_{t+1}$$

with $v^t \in \mathbb{R}^{n \times d_t}$,
$\Theta = \{\theta_1^t, \ldots, \theta_{|\mathcal{M}|}^t\}_t$, $\theta_M^t \in \mathbb{R}^{d_t \times d_{t+1}}$ trainable parameters.

Scarselli, Tsoi, Hagenbuchner, Monfardini, 2009
Chen, Li, Bruna, 2017

Invariant graph networks

- Linear case:
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}$ invariant, then $vec(L) = \pi^{\otimes k} vec(L)$.
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}^{n^k}$ equivariant, then $vec(L) = \pi^{\otimes 2k} vec(L)$

  - The space of invariant [equivariant] linear functions on k-tensors has dimension $b(k)$ [$b(2k)$].

    ($b(k)$ denotes Bell Number: number of partitions of a size k set).

Maron, Ben-Hamu, Shamir, Lipman, 2019
Maron, Fetaya, Segol, Lipman, 2019
Keriven, Peyré, 2019

Invariant graph networks

- Linear case:
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}$ invariant, then $vec(L) = \pi^{\otimes k} vec(L)$.
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}^{n^k}$ equivariant, then $vec(L) = \pi^{\otimes 2k} vec(L)$

  - The space of invariant [equivariant] linear functions on k-tensors has dimension $b(k)$ [$b(2k)$].

    ($b(k)$ denotes Bell Number: number of partitions of a size k set).

- Universal approximation:
  - Invariant networks constructed by composition of linear invariant layers $L_t : \mathbb{R}^{n^k \times a} \to \mathbb{R}^b$ with ReLU or sigmoid activation functions universally approximate the space of invariant functions.
  - Extension to equivariant functions.

Maron, Ben-Hamu, Shamir, Lipman, 2019

Maron, Fetaya, Segol, Lipman, 2019

Keriven, Peyré, 2019

Invariant graph networks

- Linear case:
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}$ invariant, then $vec(L) = \pi^{\otimes k} vec(L)$.
  - If $L : \mathbb{R}^{n^k} \to \mathbb{R}^{n^k}$ equivariant, then $vec(L) = \pi^{\otimes 2k} vec(L)$

  - The space of invariant [equivariant] linear functions on k-tensors has dimension $b(k)$ [$b(2k)$].

    ($b(k)$ denotes Bell Number: number of partitions of a size k set).

- Universal approximation:
  - Invariant networks constructed by composition of linear invariant layers $L_t : \mathbb{R}^{n^k \times a} \to \mathbb{R}^b$ with ReLU or sigmoid activation functions universally approximate the space of invariant functions.
  - Extension to equivariant functions.

    Arbitrary high order tensors are needed.

    Rates of convergence are not known.

Maron, Ben-Hamu, Shamir, Lipman, 2019

Maron, Fetaya, Segol, Lipman, 2019

Keriven, Peyré, 2019

Approximation power of MPNN

How powerful are graph neural networks? [Xu et al' 19, Morris et al' '19]

Expressive power of GNNs $\longleftrightarrow$ distinguish non-isomorphic graphs

Approximation power of MPNN

How powerful are graph neural networks? [Xu et al' 19, Morris et al' '19]

Expressive power of GNNs $\longleftrightarrow$ distinguish non-isomorphic graphs

GNN $\equiv \{f_\Theta : \Theta \in \mathbb{R}^N$ parameters$\}$
Q: Let G_1, G_2 not isomorphic. Does there exists Θ such that

$$f_\Theta(G_1) \neq f_\Theta(G_2)?$$

Approximation power of MPNN

How powerful are graph neural networks? [Xu et al' 19, Morris et al' '19]

Expressive power of GNNs $\longleftrightarrow$ distinguish non-isomorphic graphs

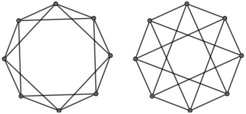
GNN $\equiv \{f_\Theta : \Theta \in \mathbb{R}^N$ parameters$\}$
Q: Let G_1, G_2 not isomorphic. Does there exists Θ such that

$$f_\Theta(G_1) \neq f_\Theta(G_2)?$$

A: If and only if the Weisfeler-Lehman test (1968) can distinguish them.

W-L test is as powerful as the LP relaxation [Ullman et al '94].

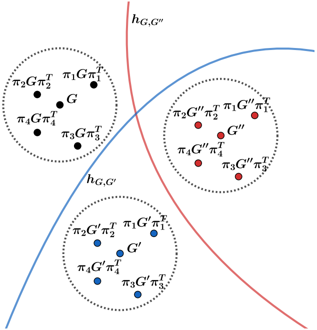
In particular MPNN cannot distinguish between non-isomorphic regular graphs with the same degree.



Graph isomorphism equivalence to universal approximation

GNN architecture $\equiv \{f_\Theta : \Theta \in \mathbb{R}^N \text{ parameters}\}$

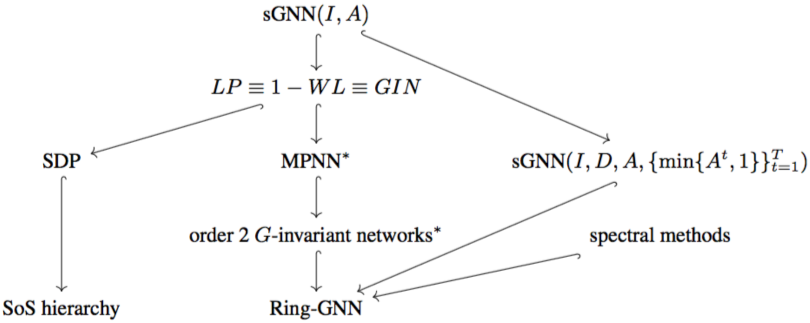
Universal approximation of invariant functions $\longleftrightarrow$ Distinguish all pairs of non-isomorphic graphs



Chen, V., Chen, Bruna, NeurIPS 2019

Comparison of architectures through GIso

$\mathcal{C} \subseteq \mathcal{C}'$ if for all pairs of non-isomorphic graphs G_1, G_2, if there exists $h \in \mathcal{C}$ so that $h(G_1) \neq h(G_2)$ then there exists $h' \in \mathcal{C}'$ so that $h'(G_1) \neq h'(G_2)$.



Chen, V., Chen, Bruna, NeurIPS 2019

Counting substructures

Given G, M graphs. How many embeddings $i : G \hookrightarrow M$ exist?

- as a subgraph
- as an induced subgraph

Motivations:

- More natural expressive power measurement than graph isomorphism
- Identify structures in social networks
- Compute similarities between molecules

Chen, Chen, V., Bruna, NeurIPS 2020

Can graph neural networks count substructures?

- MPNN [Gilmer et al '17] and 2-IGNs [Maron et al '19] cannot count connected induced subgraphs of more than 2-nodes.
- MPNNs and 2-IGNs can count star-shaped subgraphs of any size.
- k-WL and k-IGNs can count induced subgraphs of size k
- Upper bound on the size of subgraphs that k-WL can count after T iterations: $(k+1)2^T$.
- Propose a Local relation pooling architecture designed to count substructures.

Chen, Chen, V., Bruna, 2020

Local relational pooling

Inspired by [Murphy et al '19]
At each layer we consider all permutations on neighborhoods of size k for each node:

$$f(G) = \sum_{i \in V} \sum_{\Pi \in S_k} \hat{f}(\Pi B_i)$$

Where B_i is the (cropped) neighborhood of i in G represented by a $k \times k$ matrix.

	Erdős-Renyi				Random Regular			
	Triangle		3-Star		Triangle		3-Star	
	top 1	top 3	top 1	top 3	top 1	top 3	top 1	top 3
LRP-1-4	1.56E-4	2.49E-4	2.17E-5	5.23E-5	2.47E-4	3.83E-4	1.88E-6	2.81E-6
LRP-1-4 (dp)†	2.81E-5	4.77E-5	1.12E-5	3.78E-5	1.30E-6	5.16E-6	2.07E-6	4.97E-6
2-IGN	9.83E-2	9.85E-1	5.40E-4	5.12E-2	2.62E-1	5.96E-1	1.19E-2	3.28E-1
Powerful-IGN	5.08E-8	2.51E-7	4.00E-5	6.01E-5	1.40E-6	3.71E-5	8.49E-5	9.50E-5
GIN	1.23E-1	1.25E-1	1.62E-4	3.44E-4	4.70E-1	4.74E-1	3.73E-4	4.65E-4
GCN	6.78E-1	8.27E-1	4.36E-1	4.55E-1	1.82	2.05	2.63	2.80
sGNN	9.25E-2	1.13E-1	2.36E-3	7.73E-3	3.92E-1	4.43E-1	2.37E-2	1.41E-1

Extensions - Future work

- Design expressive architectures:
  GNN architecture depends on the task.

- Optimization landscape of GNNs:
  Current analysis of optimization landscape relies in simplified models to show that all local minima are confined in low-energy configurations.

- Connection with SoS:
  For some classes of "detecting hidden structures problems" existence of degree-d SoS refutations implies success of certain (typically non-explicit) spectral methods.
  - Can we express such class of spectral methods with GNNs.
  - Can we learn them?

Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer, 2017

TRIPODS Winter School & Workshop on Graph Learning and Deep Learning

Winter School: January 6-8th. Workshops: January 13-15.

Edo Airoldi
Associate Professor
Harvard University

Taco Cohen
Researcher
Qualcomm

William Hamilton
Assistant Professor
McGill University

Stefanie Jegelka
Associate Professor
MIT

Eric Kolaczyk
Professor
Boston University

Gitta Kutyniok
Professor
LMU

Elizaveta Levina
Professor
University of Michigan

Jason Lee
Assistant Professor
Princeton University

Jure Leskovec
Associate Professor
Stanford University

Guido Montúfar
Assistant Professor
UCLA

Hanie Sedghi
Research Scientist
Google Brain

Alex Smola
Director
Amazon Web Services

Matus Telgarsky
Assistant Professor
UIUC

Johns Hopkins University
`https://www.minds.jhu.edu/2020/12/08/tripods-winter-school-workshop/`

References

A Survey on The Expressive Power of Graph Neural Networks
R. Sato, 2019

On the equivalence between graph isomorphism testing and function approximation with GNNs
Z. Chen, S. Villar, L. Chen, J. Bruna, NeurIPS 2019

Can graph neural networks count substructures?
Z. Chen, S. Villar, L. Chen, J. Bruna, NeurIPS 2020

SIMONS FOUNDATION