Approximation theory of graph neural networks

Thien Le, in joint work with Stefanie Jegelka
MIT, TU Munich

Joint Math Meetings (JMM) 2025 - Seattle, WA

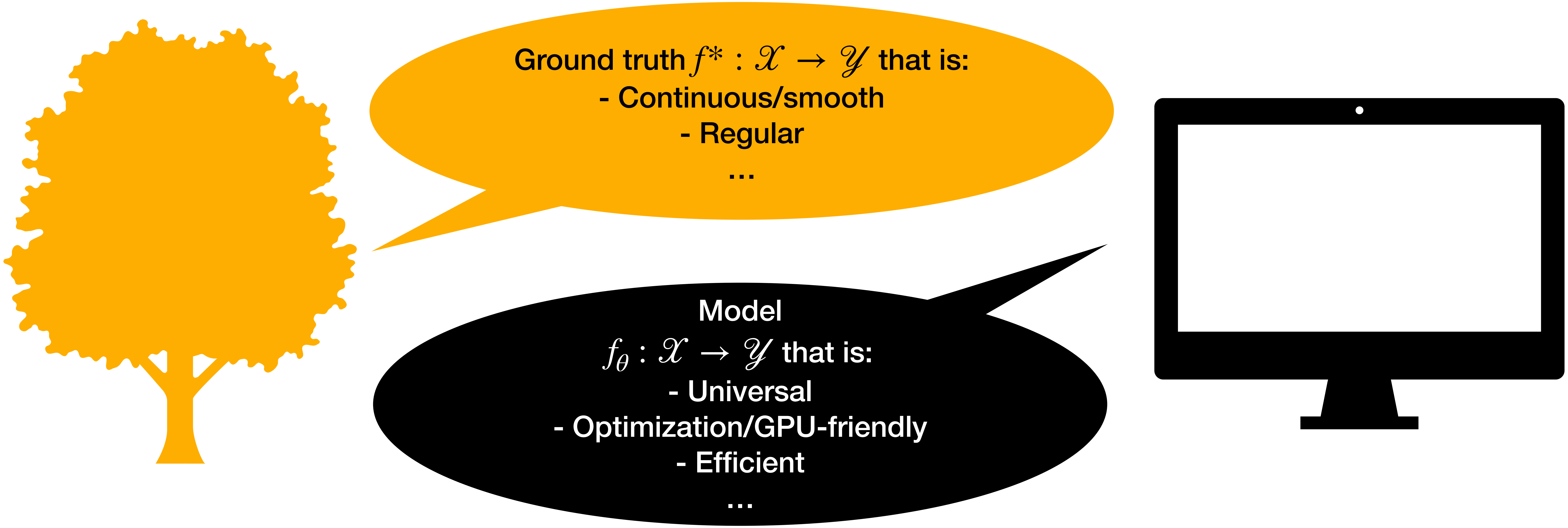
Roadmap and summary

1. Approximation theory
   a. Neural network approximation
   b. Challenges to graph input
2. Graph embeddings and graph limits
   a. Graphons and the limit of dense graph sequences
   b. Challenges to sparse graph sequences

Roadmap and summary

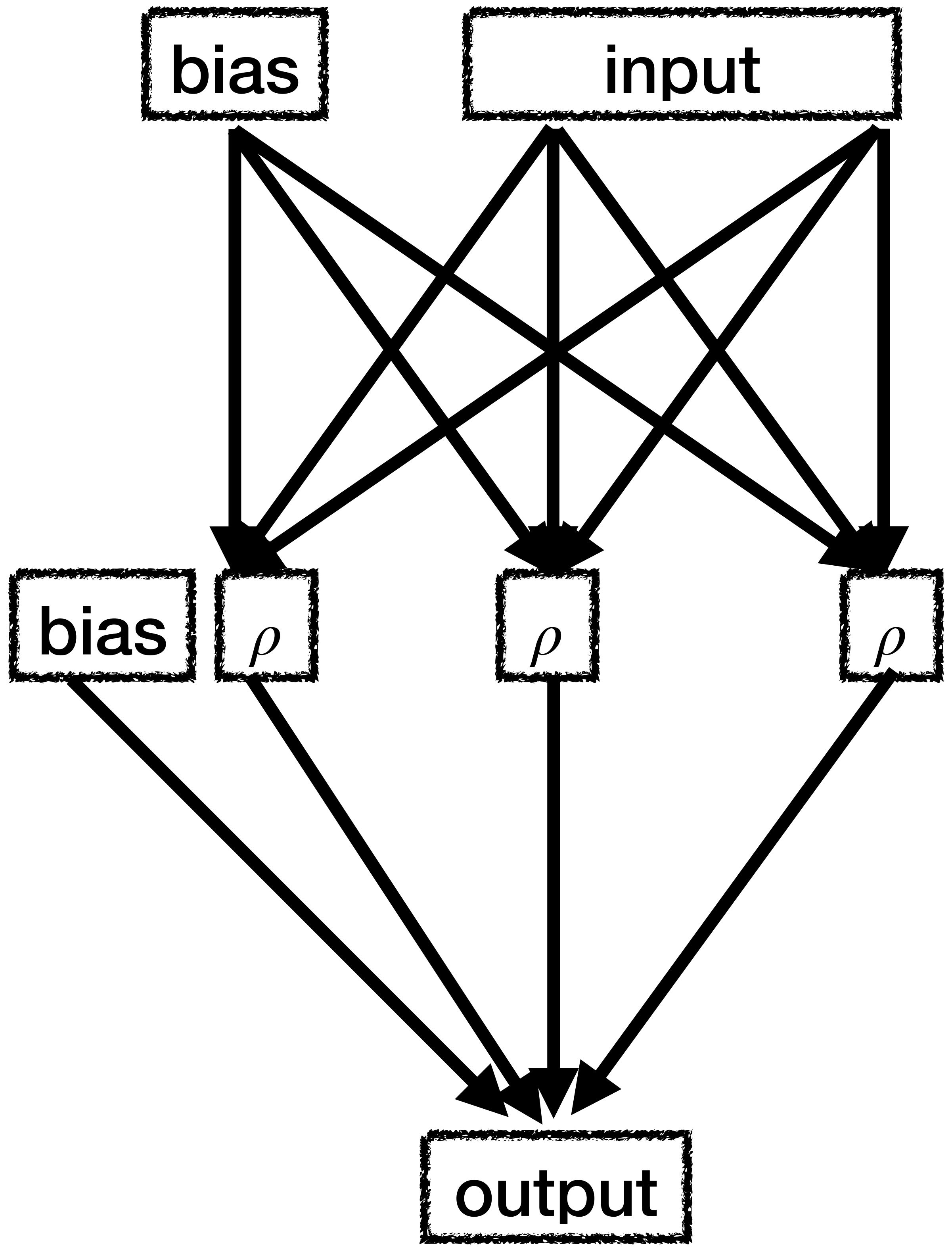
1. Approximation theory
    a. Neural network approximation
    b. Challenges to graph input
2. Graph embeddings and graph limits
    a. Graphons and the limit of dense graph sequences
    b. Challenges to sparse graph sequences

Deep learning framework



Function approximation

How do you design a family of universal models



$f_{W,b}(x) = W_2\rho(W_1x + b_1) + b_2$

Theorem (Weierstrass, 1885): For any continuous function f on a real interval $[a, b]$, for any $\epsilon > 0$, there exists a polynomial p such that $\|f - p\|_\infty < \epsilon$

universality

Theorem (Hornik, Stinchcombe, White 1989): For any continuous function f on a real hypercube $[0,1]^d$, for any $\epsilon > 0$, there exists a two-layer neural net p such that $\|f - p\|_\infty < \epsilon$

optimization/GPU-friendly

Theorem (Telgarsky, 2015): There exists a family of classification problems indexed by k, such that to achieve error $< 1/6$, two-layer neural nets require $2^{\Omega(k)}$ nodes while $2k$-layer neural nets require $O(k)$

efficiency

Roadmap and summary

1. Approximation theory
    a. Neural network approximation
    b. Challenges to graph input
2. Graph embeddings and graph limits
    a. Graphons and the limit of dense graph sequences
    b. Challenges to sparse graph sequences

Deep learning approaches to learning on graphs

Non-Euclidean graph space poses unique challenges

- Graph neural networks (GNNs) (Gilmer *et al.*, 2017)
    - Direct parameterization of functions on space of graphs (e.g. $\mathbb{R}^{n\times n}/S_n$) and embellishments (vertex/edge features)
    - Exact invariance (e.g. $\mathbb{R}^{n\times n}/S_n \to R$)

Graph (convolutional) neural networks

Deep learning architectures with built-in graph symmetry

- G(C)NN with parameter h, graph G and node features X:

$$\mathrm{GNN}(h, G, X) = \mathrm{X_L}(h, G, X)$$

$$\mathrm{X}_l(h, G, X) = \rho\left(\mathrm{A}_l(h, G)\mathrm{X}_{l-1}\right), \qquad \mathrm{A}_l(h, G) := \sum_{k=0}^{K} h_{l,k}\,\mathrm{Adj}(G)^k$$

$$\mathrm{X}_0(h, G, X) = X$$

- h does not depend on (size of) G

GNN as operator on node feature

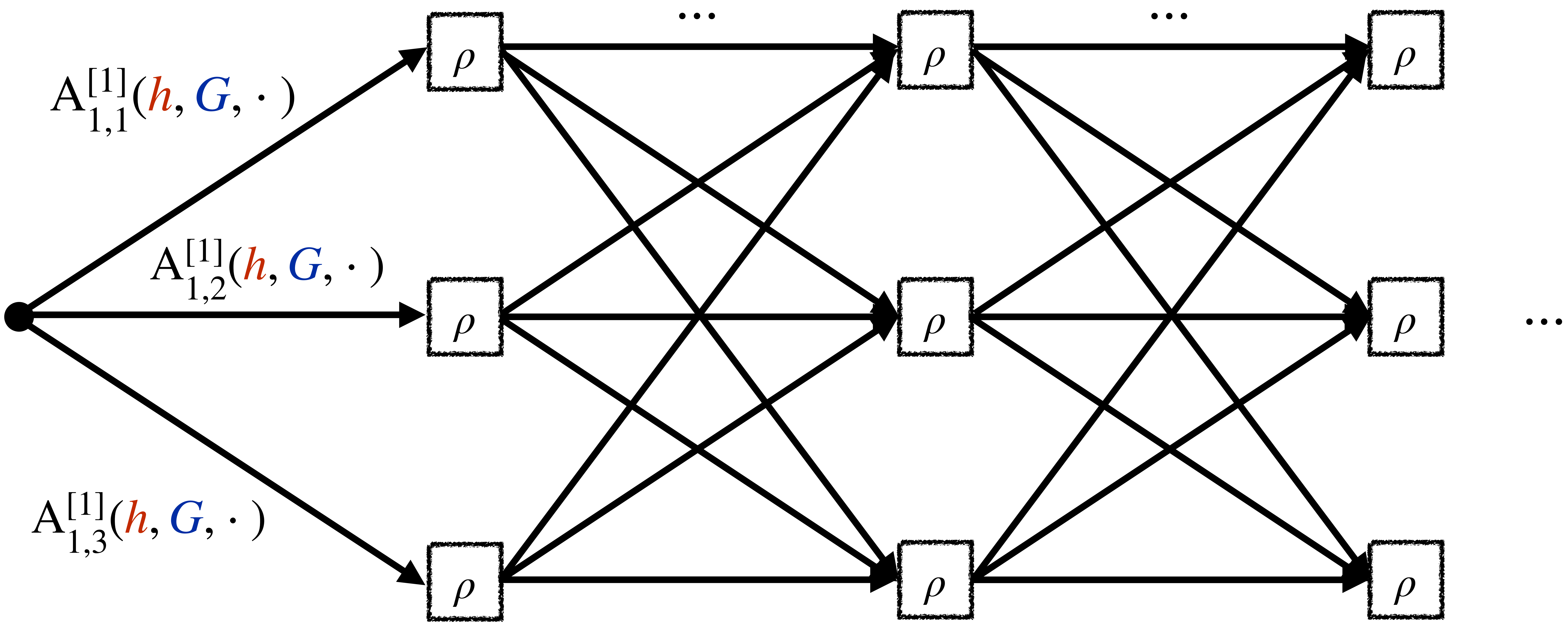
$$\mathrm{GNN}(h, G, \cdot\,) : \ell^2([n]) \to \ell^2([n]), \qquad n = |\,V(G)\,|$$

$$\bullet \xrightarrow{\mathrm{A}_1(h, G, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_2(h, G, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_3(h, G, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_4(h, G, \cdot\,)} \boxed{\mathrm{Id}}$$

where $\mathrm{A}_l(h, G) := \sum_{k=0}^{K} h_{l,k} \, \mathrm{Adj}(G)^k$

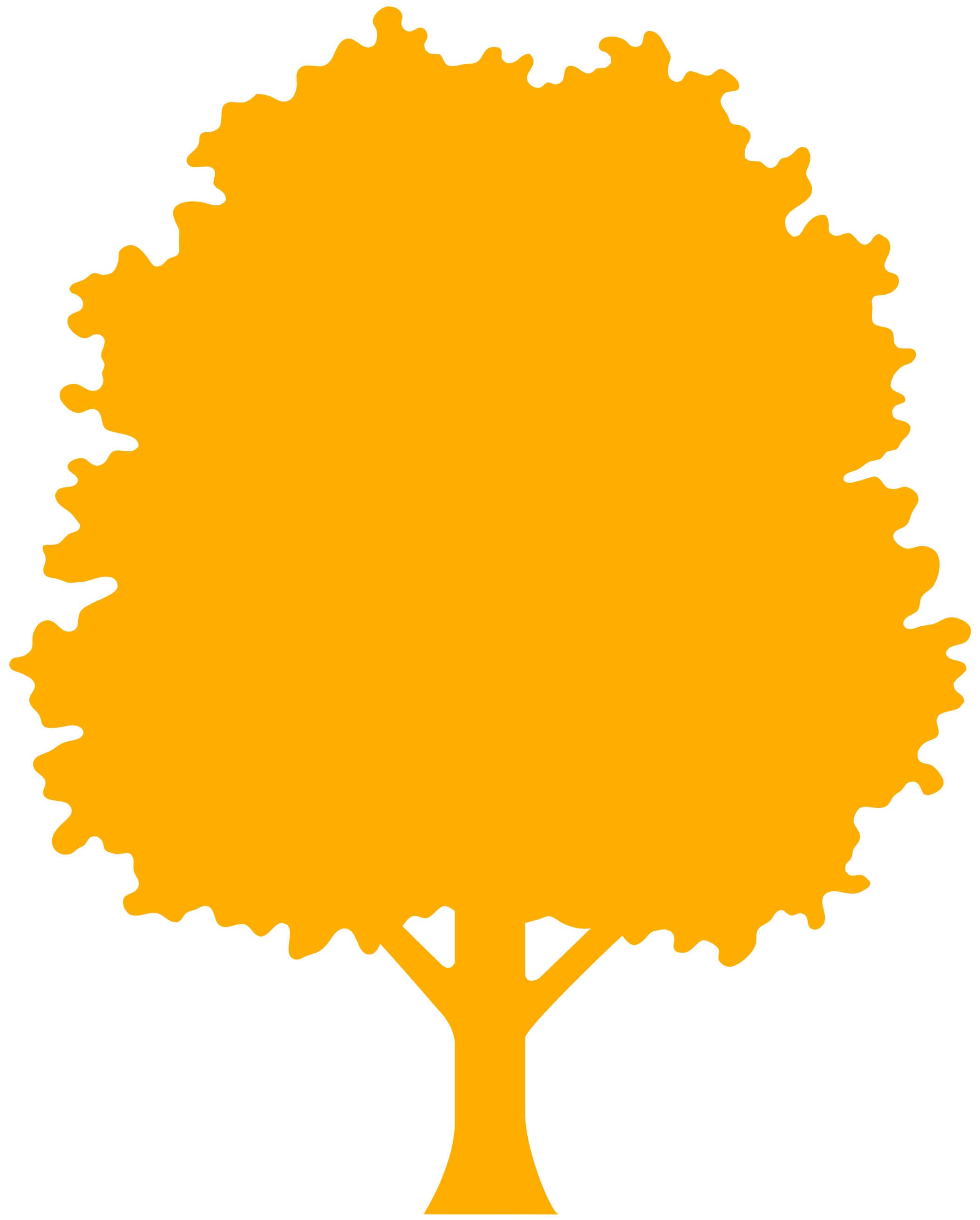
GNN as operator on node feature

$$\mathrm{GNN}(h, G, \cdot\,) : \ell^2([n]) \to \ell^2([n]), \qquad n = |G| := |V(G)|$$



$$\mathrm{A}^{[l]}_{\mathrm{f,g}}(h, G, \cdot\,) := \sum_{k=0}^{K} h^{[l]}_{f,g,k}\, \mathrm{Adj}(G)^k$$

Is GNN a ‘good’ model?



Ground truth

$f^* : \mathscr{G} \times \mathscr{X} \to \mathscr{Y}$ that is:

- Continuous/smooth
- Regular

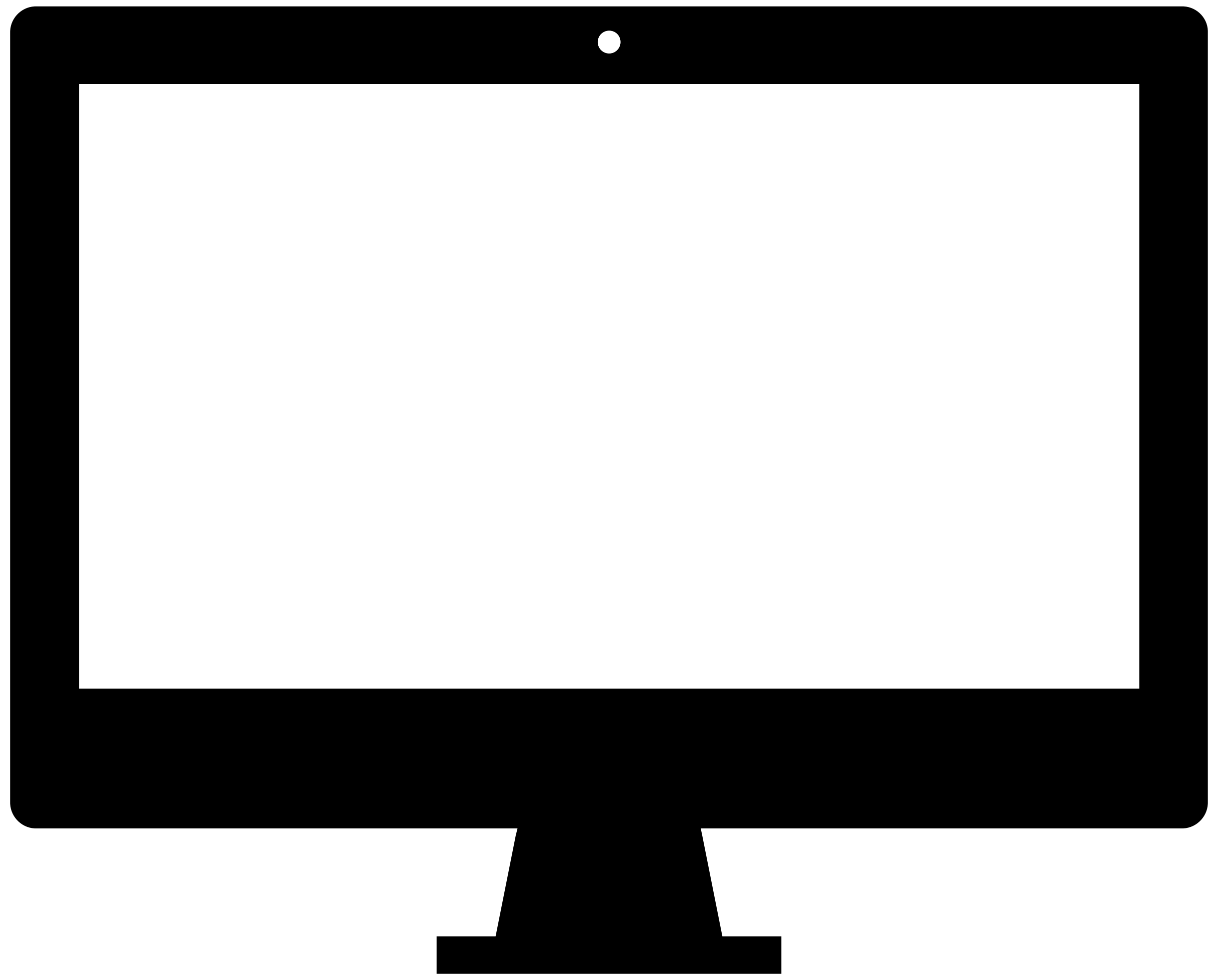
...

Model

$f_\theta : \mathscr{G} \times \mathscr{X} \to \mathscr{Y}$ that is:

- Universal
- Optimization/GPU-friendly
- Efficient

...



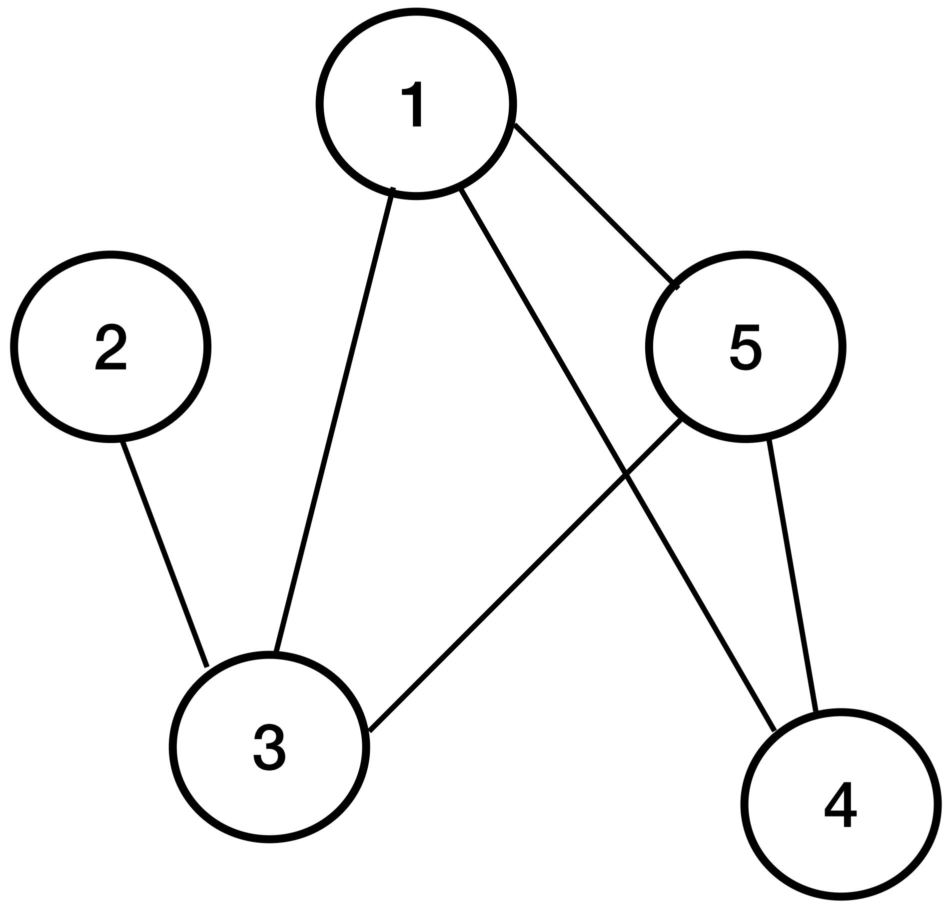
Universality - Efficiency tradeoff due to hardness of GI

Roadmap and summary

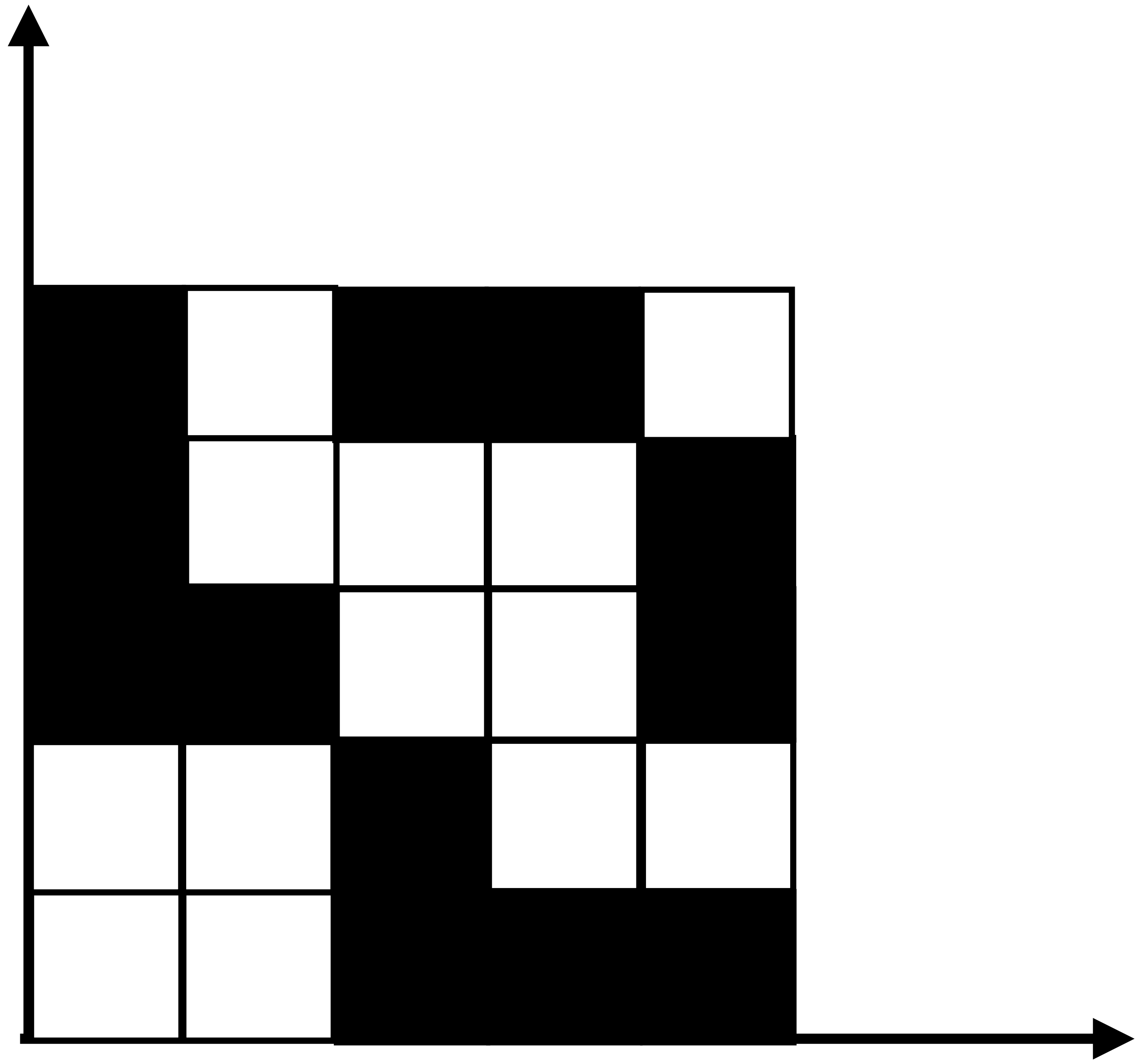
1. Approximation theory
    a. Neural network approximation
    b. Challenges to graph input
2. Graph embeddings and graph limits
    a. Graphons and the limit of dense graph sequences
    b. Challenges to sparse graph sequences

Graphons and the limit of dense graph sequences

- Graphons are symmetric, Lebesgue-measurable functions $\mathbf{W} : [0,1] \times [0,1] \rightarrow [0,1]$



$\Longrightarrow$



Graphons and the limit of dense graph sequences

Chayes, Borgs, Lovász, Sós, Vesztergombi ~2007

- Graphons are symmetric, Lebesgue-measurable functions $\mathbf{W} : [0,1] \times [0,1] \to [0,1]$
- Interpretation:
  1. $(V = [0,1], E = \{(u,v) : \mathbf{W}(u,v) \neq 0\})$ (informal)
  2. Finite graph sampler:
     1. Given number of nodes n, sample $v_1, \ldots, v_n \overset{\text{iid}}{\sim} \text{Unif}[0,1]$
     2. For each $i < j \in [n]$, sample $(v_i, v_j) \overset{\text{iid}}{\sim} \text{Bern}(\mathbf{W}(v_i, v_j))$

Most realistic graphs are not dense

Social networks

Name	Type	Nodes	Edges	Description
ego-Facebook	Undirected	4,039	88,234	Social circles from Facebook (anonymized)
ego-Gplus	Directed	107,614	13,673,453	Social circles from Google+
ego-Twitter	Directed	81,306	1,768,149	Social circles from Twitter
soc-Epinions1	Directed	75,879	508,837	Who-trusts-whom network of Epinions.com
soc-LiveJournal1	Directed	4,847,571	68,993,773	LiveJournal online social network
soc-Pokec	Directed	1,632,803	30,622,564	Pokec online social network
soc-Slashdot0811	Directed	77,360	905,468	Slashdot social network from November 2008
soc-Slashdot0922	Directed	82,168	948,464	Slashdot social network from February 2009
wiki-Vote	Directed	7,115	103,689	Wikipedia who-votes-on-whom network
wiki-RfA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)
gemsec-Deezer	Undirected	143,884	846,915	Gemsec Deezer dataset
gemsec-Facebook	Undirected	134,833	1,380,293	Gemsec Facebook dataset
soc-RedditHyperlinks	Directed, Signed, Temporal, Attributed	55,863	858,490	Hyperlinks between subreddits on Reddit
soc-sign-bitcoin-otc	Weighted, Signed, Directed, Temporal	5,881	35,592	Bitcoin OTC web of trust network
soc-sign-bitcoin-alpha	Weighted, Signed, Directed, Temporal	3,783	24,186	Bitcoin Alpha web of trust network
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993	Dynamic face-to-face interaction network between group of people
musae-twitch	Undirected	34,118	429,113	Social networks of Twitch users.
musae-facebook	Undirected	22,470	171,002	Facebook page-page network with

SNAP

Stanford Large Network Dataset Collection

Approximation for sparse graph sequences

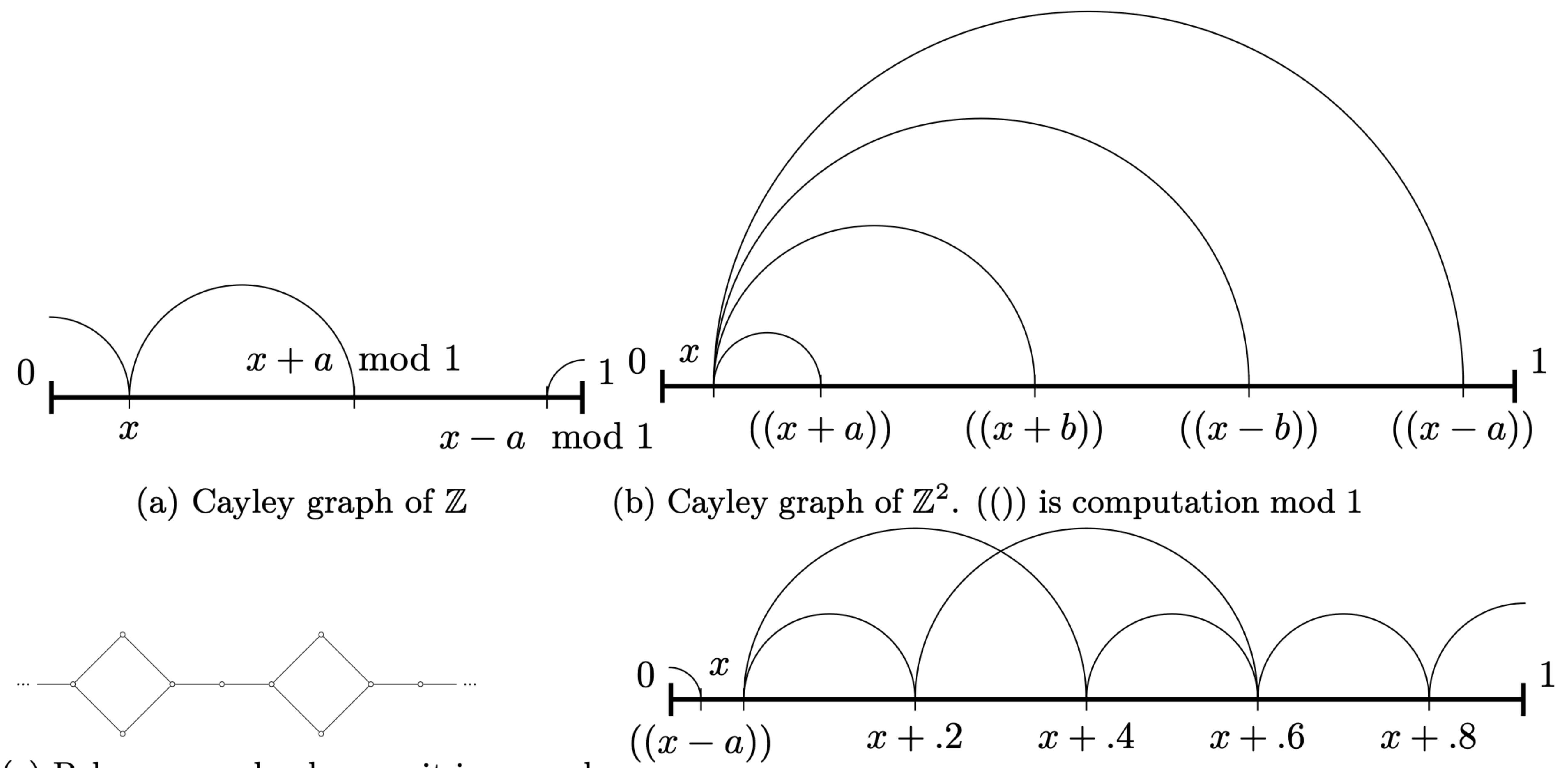
T.L., S. Jegelka (2023)

- d_M metric compares graphop operators (Backhausz and Szegedy, 2022 + modification to work with continuity assumptions)
- For A_n a discretization of limit object A

$$d_M(\text{GNN}(h, A_n, \cdot\,), \text{GNN}(h, A, \cdot\,)) \leq O\left(n^{-\frac{1}{2}}\right)$$

Example of sparse graph limit: graphings

Difficult to apply existing spectral techniques.



(a) Cayley graph of $\mathbb{Z}$

(b) Cayley graph of $\mathbb{Z}^2$. (()) is computation mod 1

(c) Polymer graph whose unit is a graph with 5 vertices. The pattern extends ad infinitum on both sides.

(d) Graphing of (c). (()) is computation mod 0.2. Edges drawn are from a single unit in the polymer.

Figure 2: Examples of limit objects. The vertex set is the interval $[0, 1]$. Example edges are the arcs connecting points on the intervals. a and b are distinct irrational numbers. In each graph, edges that miss an endpoint are identified as a single edge connecting the two existing endpoints.

Eigengap may not be continuous at the limit!

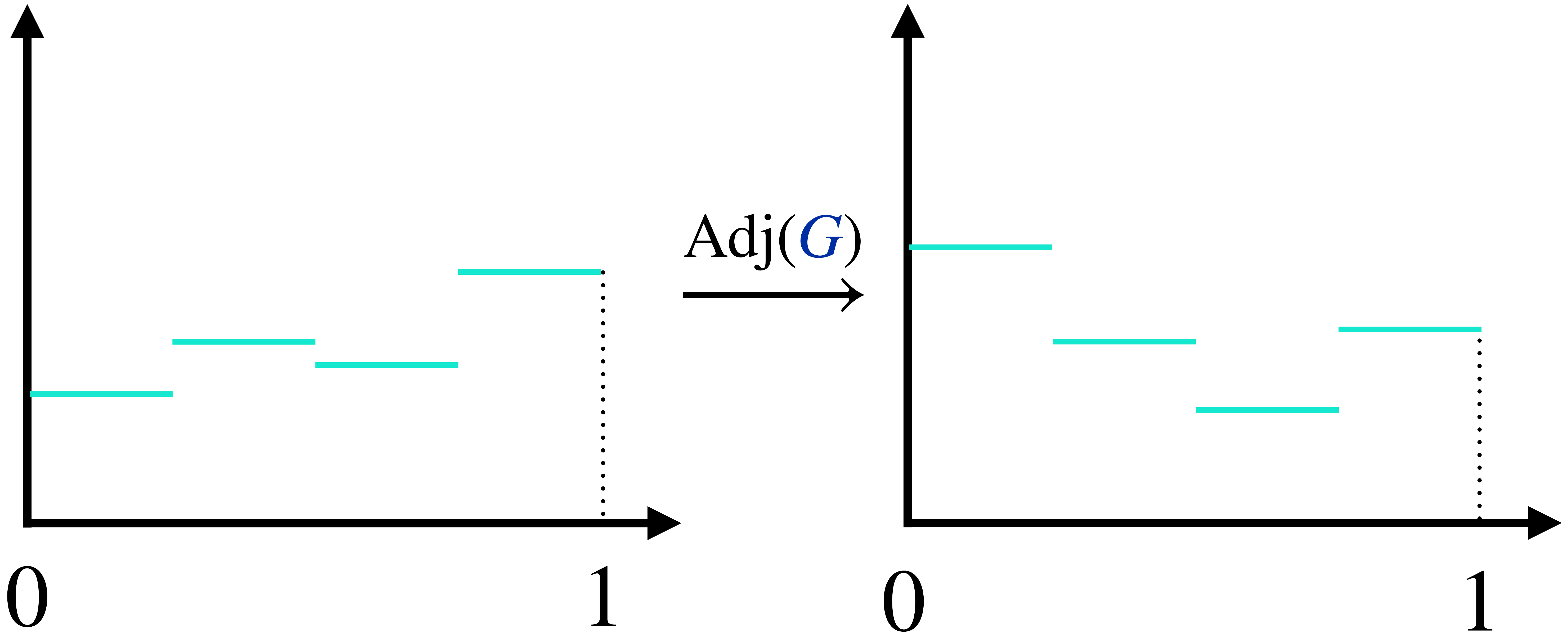
Assumptions

Structural assumptions on $\mathrm{Adj}(G)$ instead of spectral

$\mathrm{Adj}(G)$ sends (n-**piece) piece-wise constant** function to (n-**piece) piece-wise constant** function, for all n in resolution set $\mathcal{N}$

OR

$\mathrm{Adj}(G)$ sends **Lipschitz** function to **Lipschitz** function

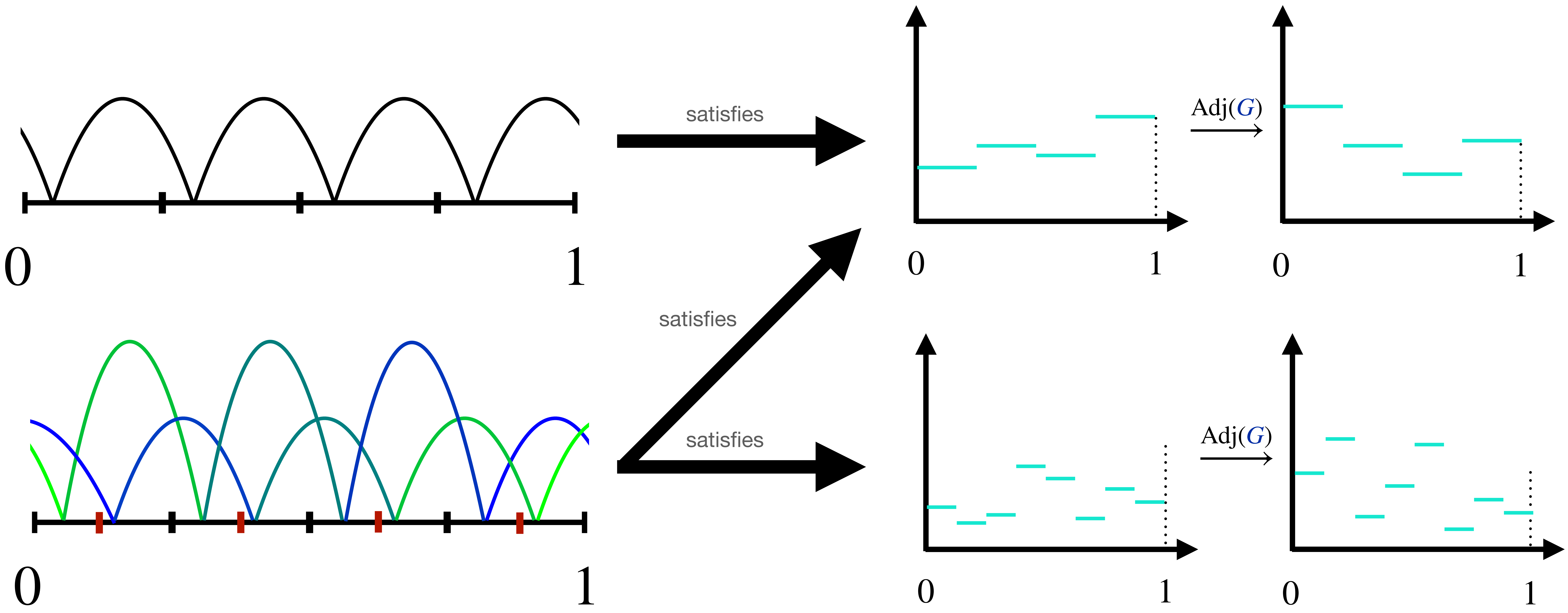


Other assumptions:

- $\mathrm{Adj}(G)$ is a Lipschitz operator
- $|h| \leq 1$ point-wise
- ρ is 1-Lipschitz

Example: cycle graphs

C_4, C_8 **embedded in** $[0,1]$**, resolution set** $\mathcal{N} = \{4,8\}$



Other examples

Satisfying structural assumptions

Paths/Grids/k-D grids

Polymer graphs

Hypercube/Hamming graph

satisfies

$\mathrm{Adj}(G)$ sends **piecewise constant** function
to **piecewise constant** function

Lipschitz graphons

satisfies

$\mathrm{Adj}(G)$ sends **Lipschitz** function
to **Lipschitz** function

Approximation for sparse graph sequences

T.L., S. Jegelka (2023)

- d_M metric compares graphop operators (Backhausz and Szegedy, 2022 + modification to work with continuity assumptions)

- For A_n a discretization of limit object A

$$d_M(\text{GNN}(h, A_n, \cdot\,), \text{GNN}(h, A, \cdot\,)) \leq O\left(n^{-\frac{1}{2}}\right)$$

Summary

- To obtain approximation theory results, additional structures are needed on the space of graphs
- We prove an approximation result for GNNs by graph limit
- Unlike dense graphs, sparse graph limits can be pathological, which was circumvented by enforcing structural assumptions

Thank you

Q&A

Machine learning on graph structures

- Social networks
  - Community detection
  - Link prediction
- Molecular graphs
  - Property prediction
  - Geometry prediction
- Traditional tasks in vision (pixel graphs) or NLP (path graphs)

SNAP

Stanford Large Network Dataset Collection

OGB

OPEN GRAPH BENCHMARK

Deep learning approaches to graph learning

- Graph neural networks (GNNs) (Gilmer *et al.*, 2017)
    - Direct parameterization of functions on space of graphs (e.g. $\mathbb{R}^{n\times n}/S_n$) and embellishments (vertex/edge features)
    - Exact invariant (e.g. $\mathbb{R}^{n\times n}/S_n \to R$) or equivariant (e.g. $\mathbb{R}^{n\times n}/S_n \to R^n/S_n$)
- Graph transformers (Veličković *et al.*, 2018)
    - Graph information added to input of transformer (self attention layers)
- Mixed architectures

Graph (convolutional) neural networks

Deep learning architectures with built-in graph symmetry

- G(C)NN with parameter h, graph G and node features X:

$$\mathrm{GNN}(h, G, X) = \mathrm{X_L}(h, G, X)$$

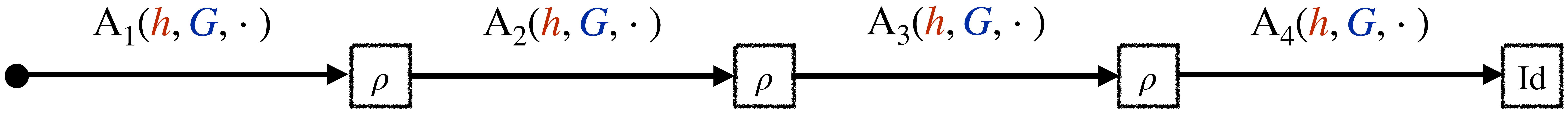
$$\mathrm{X}_l(h, G, X) = \rho\left(\mathrm{A}_l(h, G)\mathrm{X}_{l-1}\right), \qquad \mathrm{A}_l(h, G) := \sum_{k=0}^{K} h_{l,k}\,\mathrm{Adj}(G)^k$$

$$\mathrm{X}_0(h, G, X) = X$$

- h does not depend on (size of) G

GNN as operator on node feature

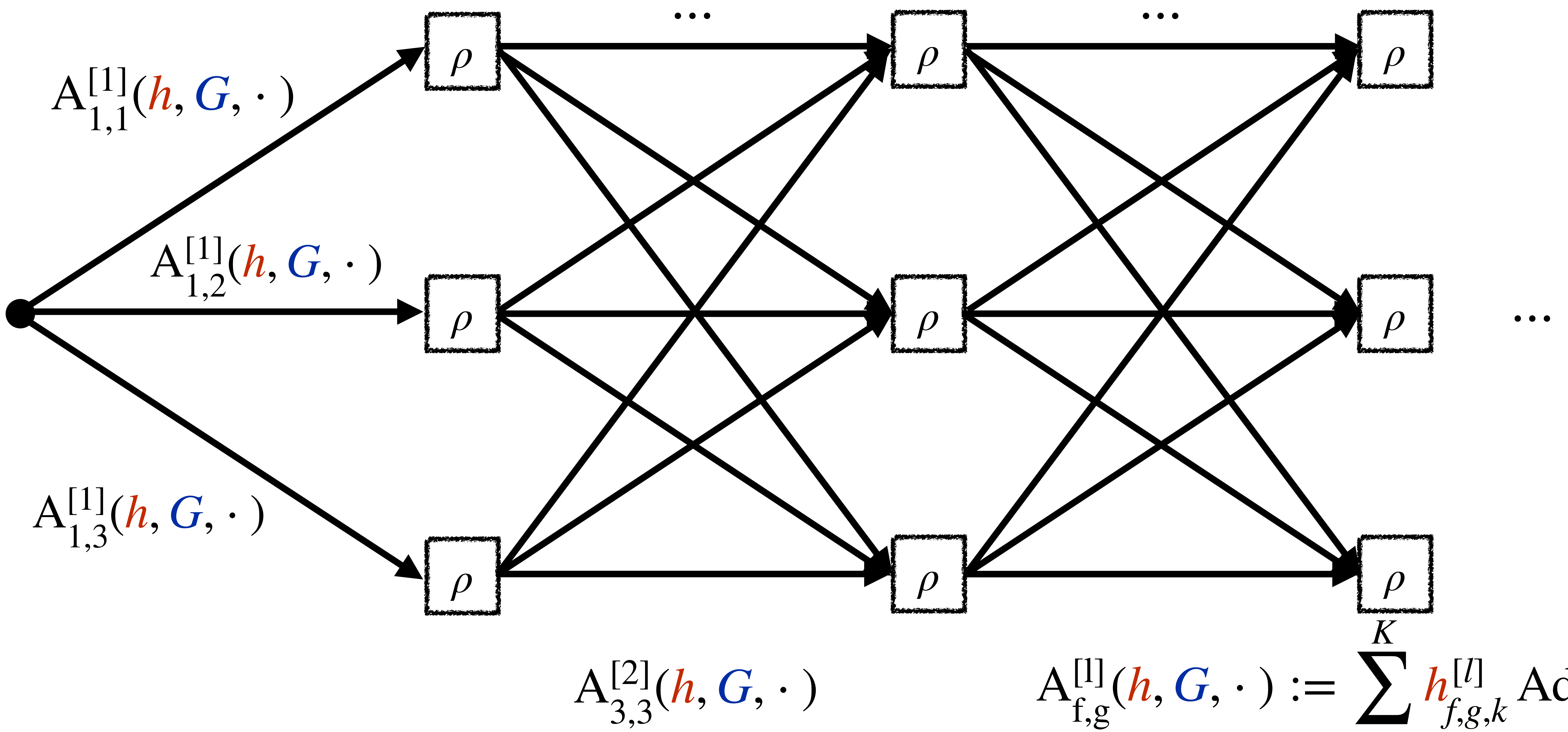
$$\mathrm{GNN}(h, G, \cdot\,) : \ell^2([n]) \to \ell^2([n]), \qquad n = |V(G)|$$



where $\mathrm{A}_l(h, G) := \sum_{k=0}^{K} h_{l,k}\, \mathrm{Adj}(G)^k$

GNN as operator on node feature

$\mathrm{GNN}(h, G, \cdot) : \ell^2([n]) \to \ell^2([n]), \qquad n = |G| := |V(G)|$



$$\mathrm{A}^{[l]}_{\mathrm{f,g}}(h, G, \cdot) := \sum_{k=0}^{K} h^{[l]}_{f,g,k} \operatorname{Adj}(G)^k$$

Transferability of GNNs

GNNs on ‘similar’ graphs are also ‘similar’

- Train on small graphs and test on large graphs
- Motivations:
    - Explanation for pretraining performance
    - Training vs evaluation compute difference
    - Out-of-distribution generalization/extrapolation

Transferability of GNNs

GNNs on ‘similar’ graphs are also ‘similar’

$$d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_1, \cdot\,), \mathrm{GNN}(h, G_2, \cdot\,)) \qquad \text{vs} \qquad d_G(G_1, G_2)$$

Size transferability of GNNs

GNNs on 'similar' graphs of different sizes are also 'similar'

Sequence of graphs $F_1, F_2, F_3, F_4, \ldots$ of increasing size $n_1, n_2, n_3, n_4, \ldots$

$d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_1, \cdot\,), \mathrm{GNN}(h, G_2, \cdot\,))$ vs $\mathrm{decreasing\_fn}(\,|G_1|\,, (\,|G_2|\,)$

e.g. $1/\,|G_1|^c + 1/\,|G_2|^c, c > 0$

Size transferability of GNNs

A limit approach ($F_n \xrightarrow{n\to\infty} G^\infty$ in some limit)

$$d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_1, \cdot\,), \mathrm{GNN}(h, G_2, \cdot\,))$$

$$\leq$$

$$d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_1, \cdot\,), \mathrm{GNN}(h, G^\infty, \cdot\,)) \quad \text{vs} \quad O(1/|G_1|^c)$$

$$+$$

approximation theorems

$$d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_2, \cdot\,), \mathrm{GNN}(h, G^\infty, \cdot\,)) \quad \text{vs} \quad O(1/|G_2|^c)$$

This paper

We show size transferability for different graph sequences, in particular sparse graphs

Via approximation bound: $d_{\mathrm{GNN}}(\mathrm{GNN}(h, G_1, \cdot), \mathrm{GNN}(h, G^{\infty}, \cdot)) \leq O(1/|G_1|^c)$

	Sparse		Dense
	Bounded-degree	Relatively-sparse	
Number of edges	$\Theta(n)$	$\Theta(n \log n)$	$\Theta(n^2)$
Examples covered under our assumptions	infinite grids, polymer graphs	hypercubes, Hamming graphs	graphons
Graphons (Ruiz et al., 2023a)			$O(n^{-1})$
Unbounded graphons (Maskey et al., 2023)			inexplicit
Random graph model (Keriven et al., 2020)		$O((\log n)^{-1/2})$	$O(n^{-1/2})$
Spectral methods (1 layer) (Levie et al., 2022)		inexplicit	inexplicit
Graphings (1 layer) (Roddenberry et al., 2022)	inexplicit		

Table 1: Summary of our results compared to related work. Quantitative results (e.g. $O(n^{-1})$) upper-bound the distance between GNNs on sampled graphs of size n and the limiting object in term of n (in an appropriate metric and limit notion). Empty cells are graph models where the approaches in the corresponding papers do not apply to or give trivial bounds (e.g. bounds that compare to a constant-0 graphon). "Inexplicit" refers to asymptotic results where rates of convergence is not explicit.

Graph limit: graphons

Chayes, Borgs, Lovász, Sós, Vesztergombi ~2007

- Graphons are symmetric, Lebesgue-measurable functions $\mathbf{W} : [0,1] \times [0,1] \to [0,1]$
- Interpretation:
    1. $(V = [0,1], E = \{(u, v) : \mathbf{W}(u, v) \neq 0\})$ (informal)
    2. Finite graph sampler:
        1. Given number of nodes n, sample $v_1, \ldots, v_n \overset{\text{iid}}{\sim} \text{Unif}[0,1]$
        2. For each $i < j \in [n]$, sample $(v_i, v_j) \overset{\text{iid}}{\sim} \text{Bern}(\mathbf{W}(v_i, v_j))$

Transferability via graphons

Ruiz, Chamon, Ribeiro, 2023

$$\| \mathrm{GNN}(h, G_1, \cdot\,) - \mathrm{GNN}(h, G_2, \cdot\,))\|_{L_2} \le O(\,|\,G_1\,|^{-1} + |\,G_2\,|^{-1}\,) + \epsilon$$

Note: It is possible to optimize Ruiz *et al.*'s bound to get rid of the ϵ but incur a slower rate

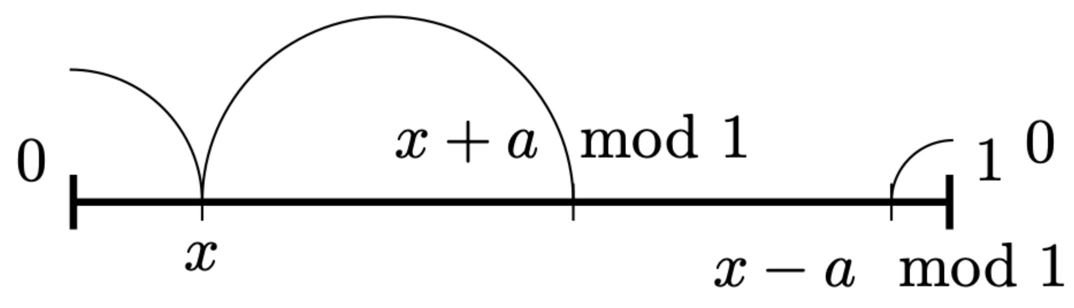
Most realistic graphs are not dense

Social networks

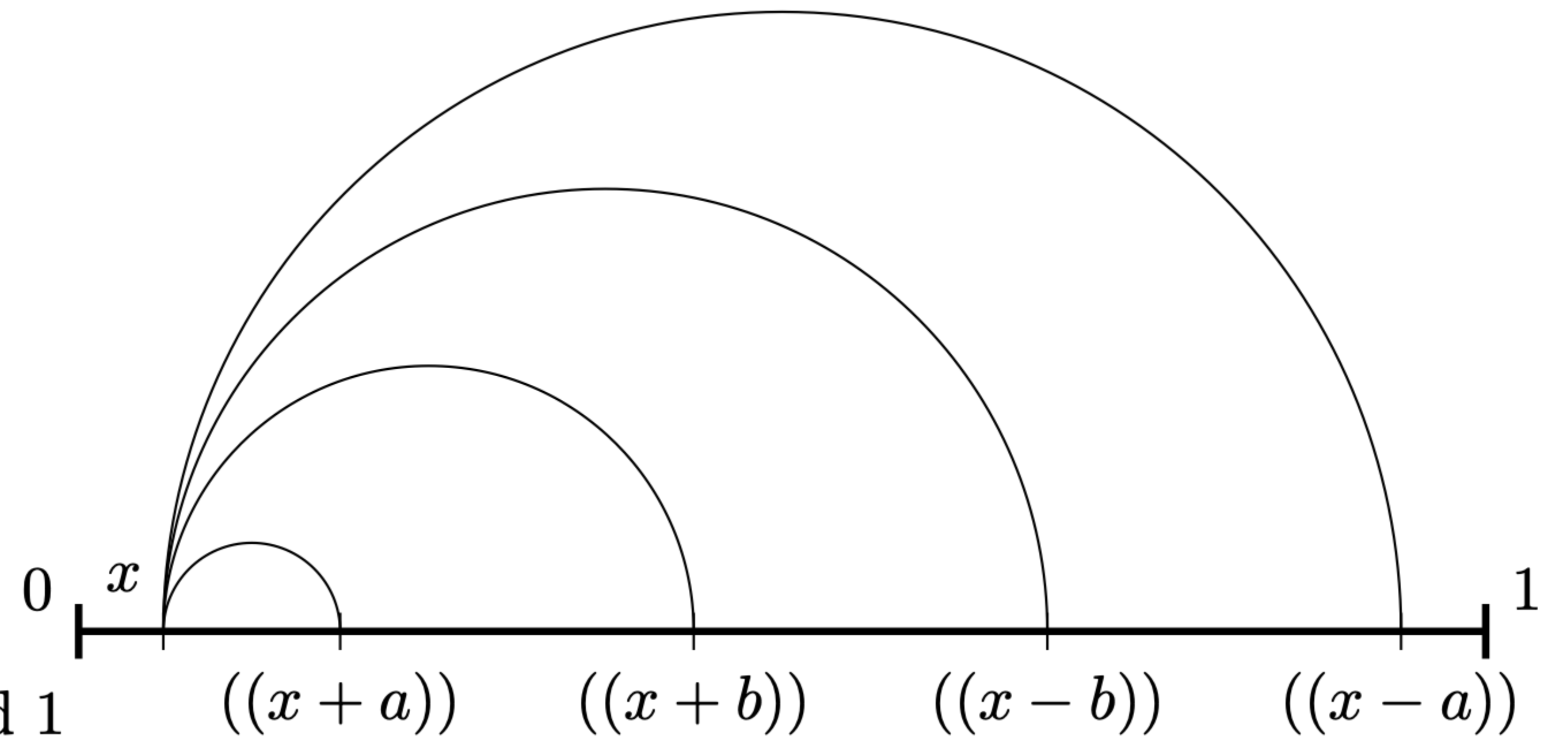
Name	Type	Nodes	Edges	Description
ego-Facebook	Undirected	4,039	88,234	Social circles from Facebook (anonymized)
ego-Gplus	Directed	107,614	13,673,453	Social circles from Google+
ego-Twitter	Directed	81,306	1,768,149	Social circles from Twitter
soc-Epinions1	Directed	75,879	508,837	Who-trusts-whom network of Epinions.com
soc-LiveJournal1	Directed	4,847,571	68,993,773	LiveJournal online social network
soc-Pokec	Directed	1,632,803	30,622,564	Pokec online social network
soc-Slashdot0811	Directed	77,360	905,468	Slashdot social network from November 2008
soc-Slashdot0922	Directed	82,168	948,464	Slashdot social network from February 2009
wiki-Vote	Directed	7,115	103,689	Wikipedia who-votes-on-whom network
wiki-RfA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)
gemsec-Deezer	Undirected	143,884	846,915	Gemsec Deezer dataset
gemsec-Facebook	Undirected	134,833	1,380,293	Gemsec Facebook dataset
soc-RedditHyperlinks	Directed, Signed, Temporal, Attributed	55,863	858,490	Hyperlinks between subreddits on Reddit
soc-sign-bitcoin-otc	Weighted, Signed, Directed, Temporal	5,881	35,592	Bitcoin OTC web of trust network
soc-sign-bitcoin-alpha	Weighted, Signed, Directed, Temporal	3,783	24,186	Bitcoin Alpha web of trust network
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993	Dynamic face-to-face interaction network between group of people
musae-twitch	Undirected	34,118	429,113	Social networks of Twitch users.
musae-facebook	Undirected	22,470	171,002	Facebook page-page network with

Example of sparse graph limit: graphings

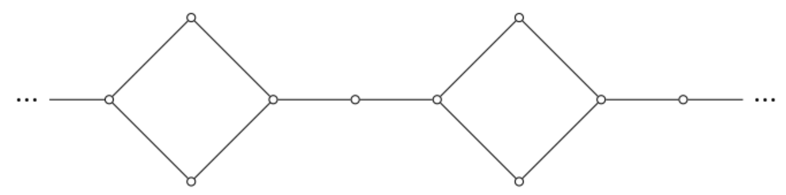
Difficult to apply existing spectral techniques.



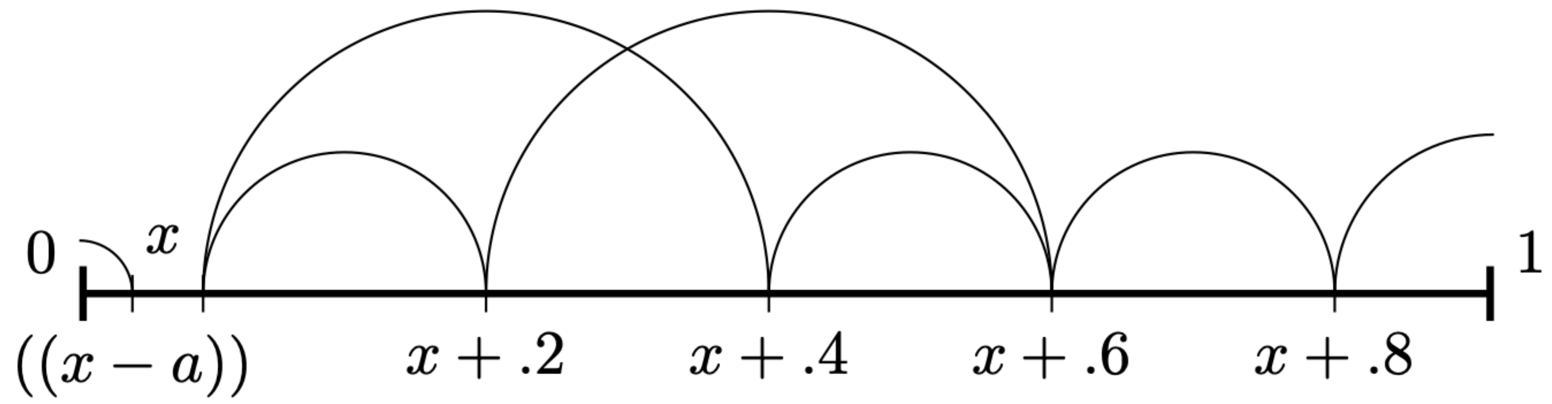
(a) Cayley graph of $\mathbb{Z}$



(b) Cayley graph of $\mathbb{Z}^2$. $(())$ is computation mod 1



(c) Polymer graph whose unit is a graph with 5 vertices. The pattern extends ad infinitum on both sides.



(d) Graphing of (c). $(())$ is computation mod 0.2. Edges drawn are from a single unit in the polymer.

Figure 2: Examples of limit objects. The vertex set is the interval $[0, 1]$. Example edges are the arcs connecting points on the intervals. a and b are distinct irrational numbers. In each graph, edges that miss an endpoint are identified as a single edge connecting the two existing endpoints.

Eigengap may not be continuous at the limit!

GNN as operator on node feature

$$\mathrm{GNN}(h, G, \cdot\,) : L^2([0,1]) \to L^2([0,1]), \qquad [0,1] = V(G)$$

$$\bullet \xrightarrow{\mathrm{A}_1(h,\, G,\, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_2(h,\, G,\, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_3(h,\, G,\, \cdot\,)} \boxed{\rho} \xrightarrow{\mathrm{A}_4(h,\, G,\, \cdot\,)} \boxed{\mathrm{Id}}$$

where $\mathrm{A}_l(h, G) := \sum_{k=0}^{K} h_{l,k} \,\mathrm{Adj}(G)^k, \qquad \mathrm{Adj}(G) : L^2[0,1] \to L^2[0,1]$

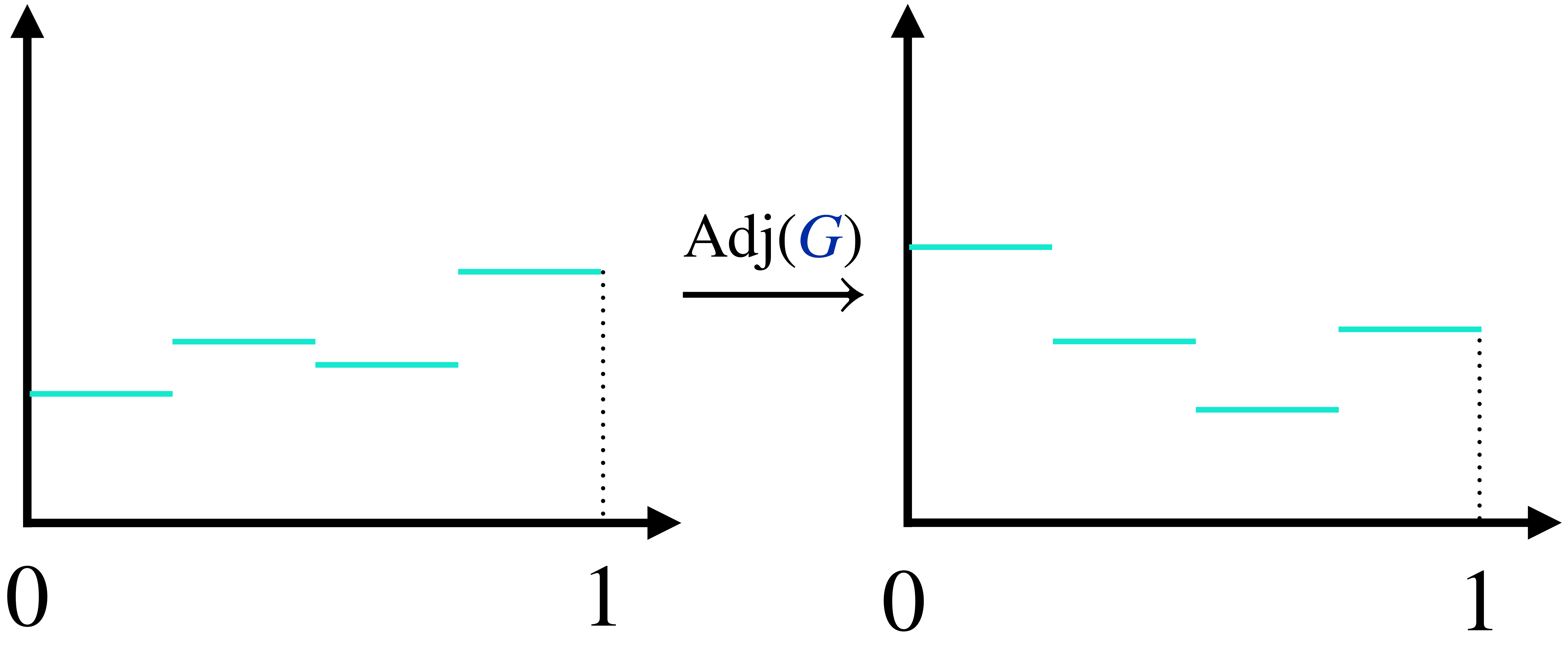
Assumptions

Structural assumptions on $\mathrm{Adj}(G)$ instead of spectral

$\mathrm{Adj}(G)$ sends (n-**piece) piece-wise constant** function
to (n-**piece) piece-wise constant** function,
for all n in resolution set $\mathcal{N}$

OR

$\mathrm{Adj}(G)$ sends **Lipschitz** function
to **Lipschitz** function

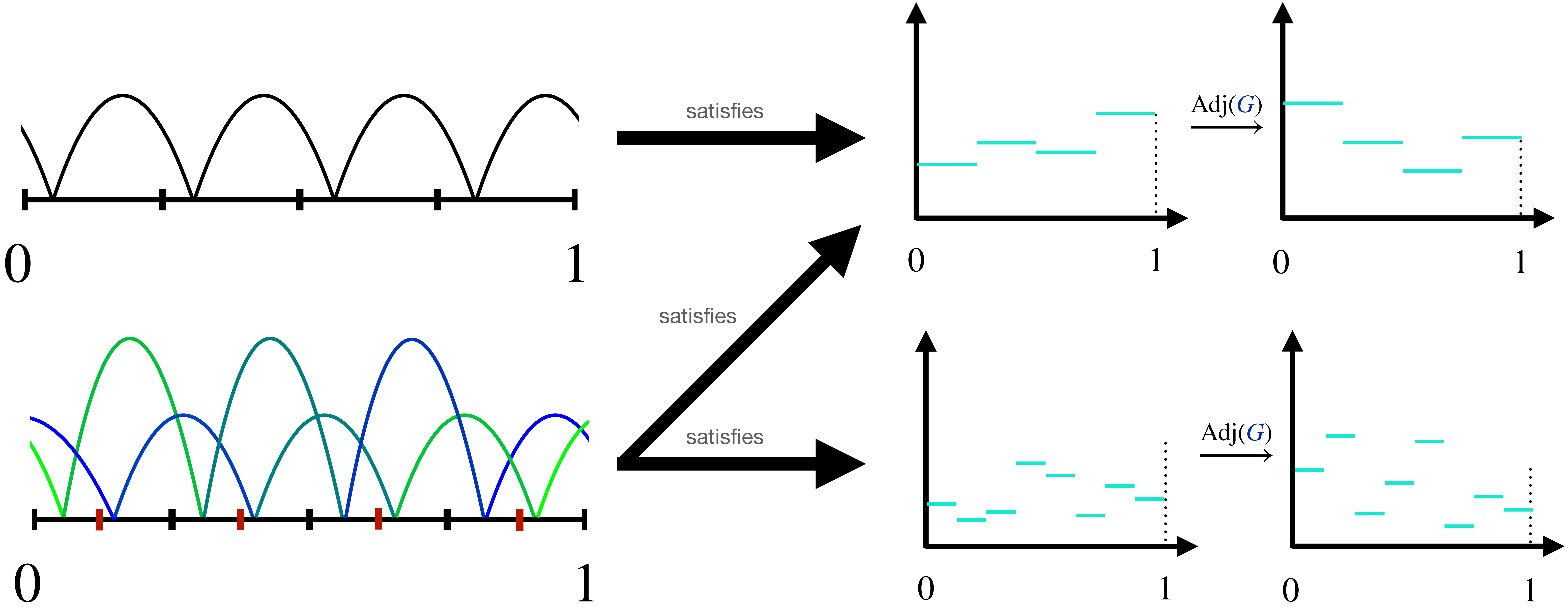


Other assumptions:

- $\mathrm{Adj}(G)$ is a Lipschitz operator
- $|h| \leq 1$ point-wise
- ρ is 1-Lipschitz

Example: cycle graphs

C_4, C_8 **embedded in** $[0,1]$**, resolution set** $\mathcal{N} = \{4,8\}$



Other examples

Satisfying structural assumptions

Paths/Grids/k-D grids

Polymer graphs

Hypercube/Hamming graph

$\xrightarrow{\text{satisfies}}$

$\mathrm{Adj}(G)$ sends **piecewise constant** function to **piecewise constant** function

Lipschitz graphons $\xrightarrow{\text{satisfies}}$ $\mathrm{Adj}(G)$ sends **Lipschitz** function to **Lipschitz** function

Main theorem: Approximation and Size transferability

via graphop (Backhausz and Szegedy 2022: action convergence)

- d_M metric compares operators (Backhausz and Szegedy, 2022 + modification to work with continuity assumptions)

- For A_n a discretization of limit object A

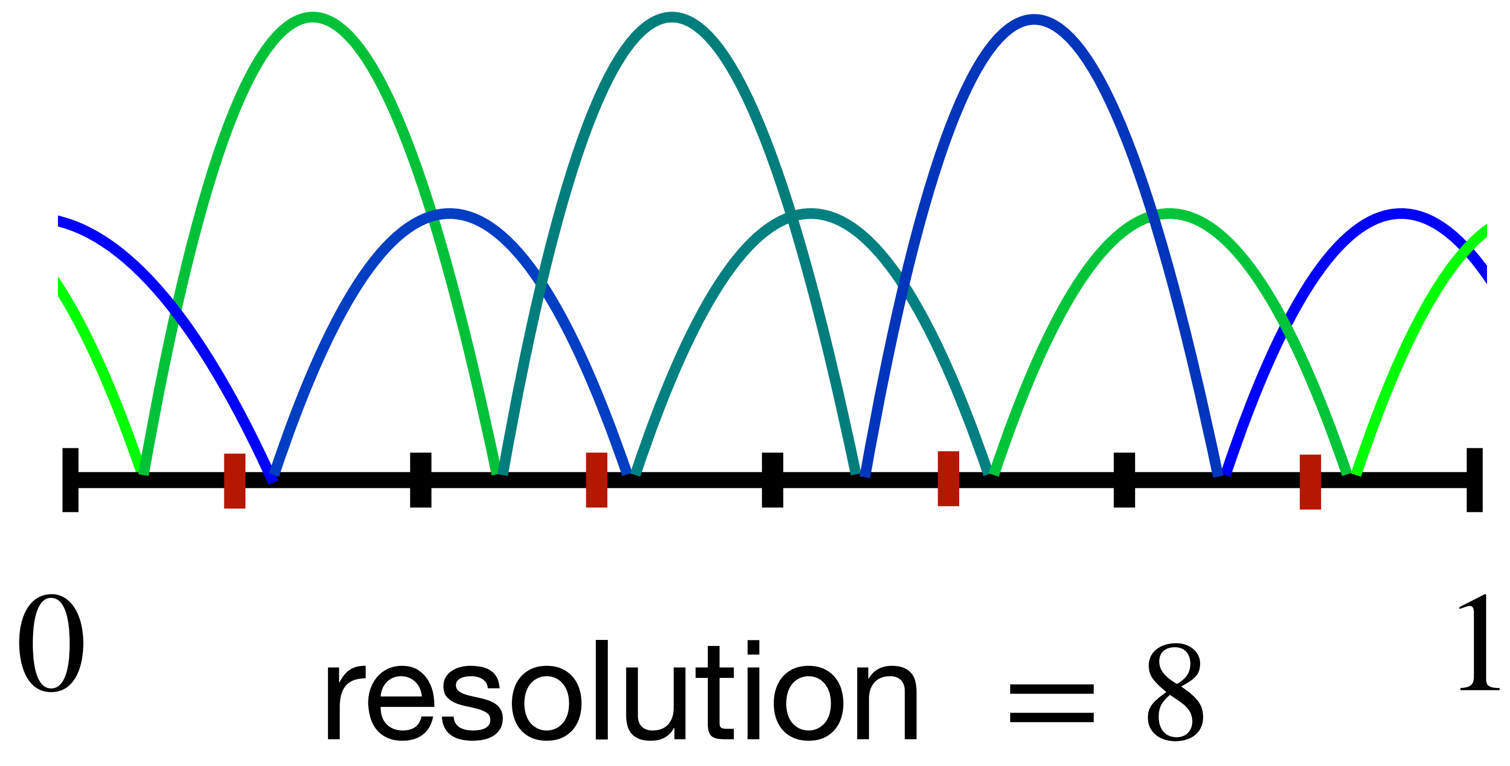
$$d_M(\text{GNN}(h, A_n, \cdot\ ), \text{GNN}(h, A, \cdot\ )) \leq O\left(n^{-\frac{1}{2}}\right)$$

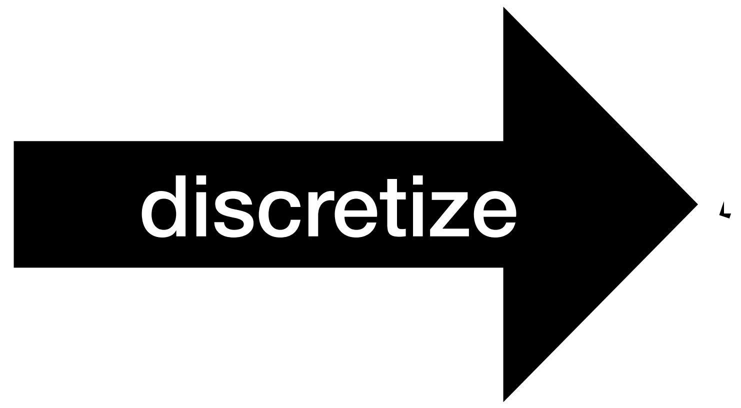
- For A_m, A_n two different discretizations of the same limit object of size m, n

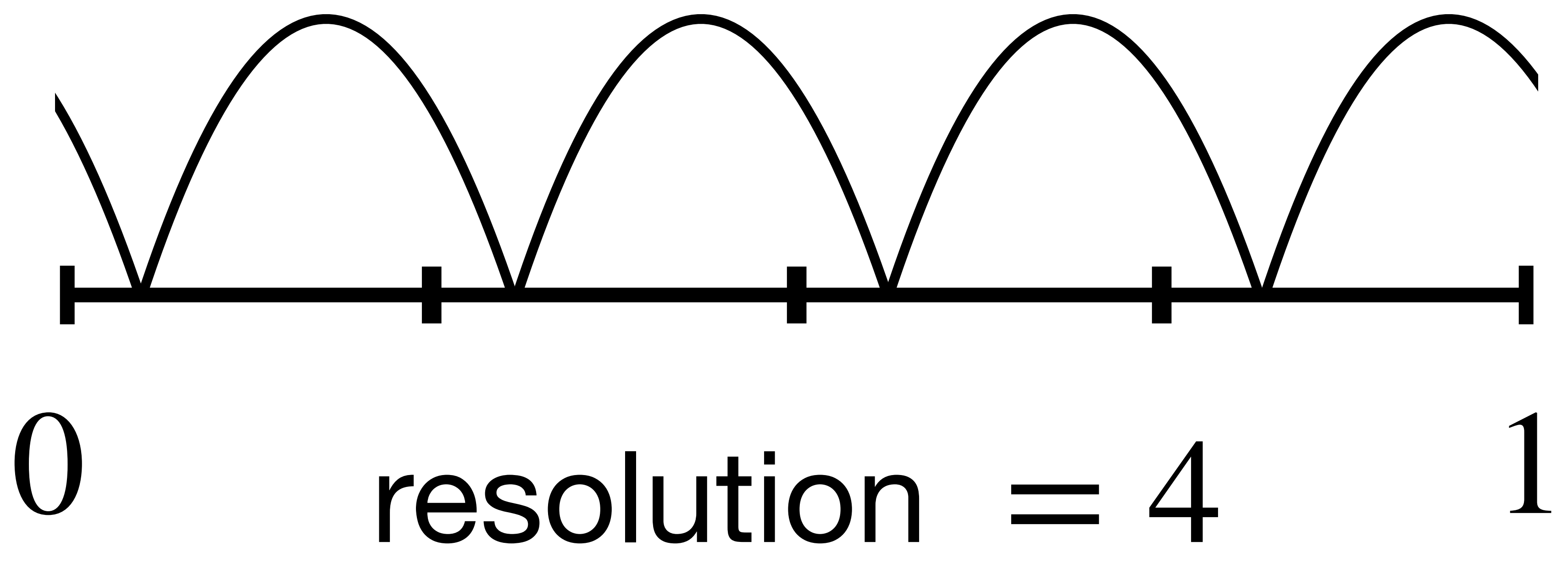
$$d_M(\text{GNN}(h, A_m, \cdot\ ), \text{GNN}(h, A_n, \cdot\ )) \leq O\left(m^{-\frac{1}{2}} + n^{-\frac{1}{2}}\right)$$

Discretizing adjacency operator/GNNs

‘Averaging connections’







More generally,

$$A_m X(v) := m \int_{v-\frac{1}{m}}^{v} (A\tilde{X}) d\lambda, \text{ for all } v \in [m]/m, \qquad \mathrm{GNN}_m(h, A, X) := \mathrm{GNN}(h, A_m, X),$$

Summary

- We prove an approximation and size transferability result for GNNs by graph limit.
- Unlike dense graphs, sparse graph limits can be pathological.
- By enforcing structural assumptions, our result works for sparse and dense graph limits.

Future directions

- Relaxing assumptions
  - Warning: unconditional approximation theorem solves group theoretic open questions (e.g. existence of non-sofic groups) - Backhausz and Szegedy.
  - But other than that?
- Graph transformer
- Optimization of sequence transformer via Kuramoto model

Q & A

- Thank you for your attention.

Back up slides

Optimization of transformers via Kuramoto model

A MATHEMATICAL PERSPECTIVE ON TRANSFORMERS

BORJAN GESHKOVSKI, CYRIL LETROUIT, YURY POLYANSKIY,
AND PHILIPPE RIGOLLET

ABSTRACT. Transformers play a central role in the inner workings of large language models. We develop a mathematical framework for analyzing Transformers based on their interpretation as interacting particle systems, with a particular emphasis on long-time clustering behavior. Our study explores the underlying theory and offers new perspectives for mathematicians as well as computer scientists.

Part 3: Further questions. We propose potential avenues for future research, largely in the form of open questions substantiated by numerical observations. We first focus on the case $d = 2$ (Section 7) and elicit a link to Kuramoto oscillators. We

Graphop Mean-Field Limits for Kuramoto-Type Models

Marios-Antonios Gkogkas* and Christian Kuehn *

December 16, 2020

For dense graphs converging to graphon limits, one also knows that mean-field approximation holds for certain classes of models, e.g., for the Kuramoto model on graphs. Yet, the space of intermediate density and sparse graphs is clearly extremely relevant. Here we prove that the Kuramoto model can be be approximated in the mean-field limit by far more general graph limits than graphons.

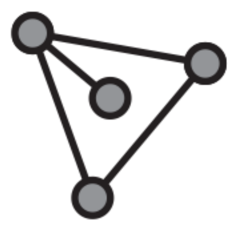
Graphings and Benjamini-Schramm convergence

Local convergence of bounded-degree graphs

- $(X, \mathscr{B}, \nu)$ Borel probability measure on topological space X. Graphings are graphs with vertex set X and Borel edge set $E \subset X \times X$ with a symmetric constraint.
- Rooted distance: $d(G_1, G_2) = 1/k$ where k is the smallest number such that the k -neighborhood of G_1 and G_2 around their roots are isomorphic.
  - Theorem: space of rooted graph with rooted distance is compact
  - Cor: space of Borel probability measures on rooted graphs is compact in weak topology
  - BS convergence: embed graphs as Borel measures that is uniform over root, for each radius

Graphop

Backhausz and Szegedy, 2022



$$\Rightarrow \begin{pmatrix} 0&1&1&1\\ 1&0&1&0\\ 1&1&0&0\\ 1&0&0&0 \end{pmatrix} \Rightarrow \begin{matrix} (\,0.3\;\; 0.2\;\; 1\;\; -0.8\,) \\ \downarrow \\ (\,0.4\;\; 1.3\;\; 0.5\;\; 0.3\,) \end{matrix} \Rightarrow \frac{1}{4}\Big(\delta_{(0.3,0.4)} + \delta_{(0.2,1.3)} + \delta_{(1,0.5)} + \delta_{(-0.8,0.3)}\Big)$$

Figure: Backhausz and Szegedy

Figure 2: Graph ⇒ operator ⇒ action ⇒ measure (computing an element in the 1-profile of a graph).

- P-operators: linear bounded (in operator norm) operators
- k-profile: the set $\mathcal{S}_k(A)$ of all possible probability measures of the form $\mathcal{D}(v_1, \ldots, v_k, Av_1, \ldots, Av_k) = \frac{1}{n}\sum_{j=1}^{n} \delta_{(v_{1,j}, \ldots, v_{k,j}, [Av_1]_j, \ldots, [Av_k]_j)}$
- Hausdorff distance between closed sets of probability measures: $d_H(X, Y) = \max(\sup_x \inf_y d(x, y), \sup_y \inf_x d(x, y))$
- Action convergence metric: $d_M(A, B) = \sum_{k=1}^{\infty} 2^{-k} d_H(\mathcal{S}_k(A), \mathcal{S}_k(B))$

Small experiment for rate

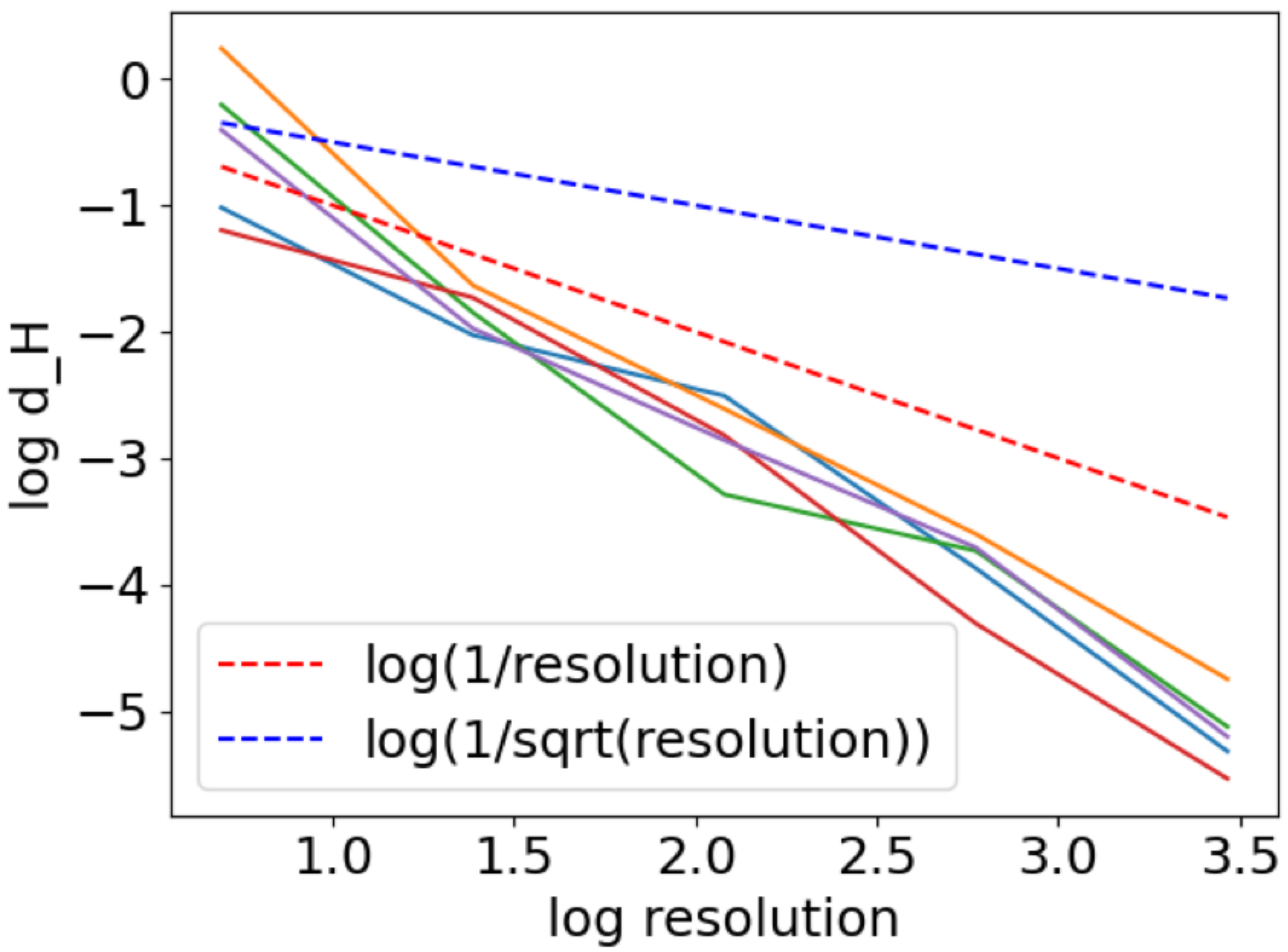


Figure 1: Hausdorff metric between samples from 1-profiles of 2-hidden-layer GNN on finite polymer graphs vs on large polymer graphs (see Appendix A for polymer graphs). The GNN uses GSO $A_n^2 + A_n$ where A_n is the normalized adjacency matrix on n nodes and ReLU nonlinearties at each layer. Different solid lines are different random draws of functions that make up the estimated 1-profile. See Appendix A for details.

Graphop neural network as P-operator

- There is no requirement for P-operators to be linear!
- Action convergence can be defined for nonlinear operators

Conjecture 1 (Action convergence of graphop neural networks)**.** *Let* $(A_n)_{n\in\mathbb{N}}$ *be an action convergent sequence of graphops. Then* $(\Phi(h, A_n, \cdot))_{n\in\mathbb{N}}$ *is an action convergent sequence of* P*-operators.*