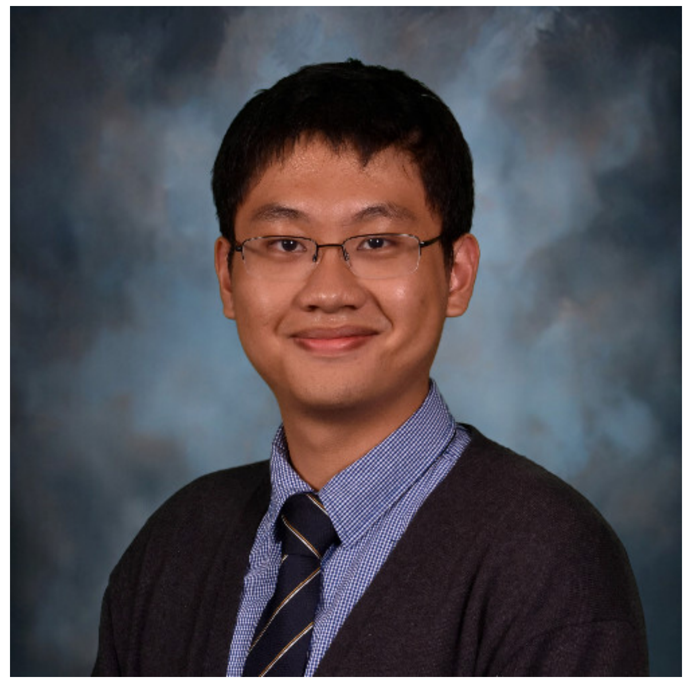
Massachusetts Institute of Technology

JOHNS HOPKINS UNIVERSITY

A Poincaré Inequality and Consistency Results for Signal Sampling on Large Graphs

Thien Le, in joint work with Luana Ruiz and Stefanie Jegelka



November 17th. DeepMath 2023.

Motivation

Graph algorithms scale with number of vertices

- Important data analysis algorithms scale with number of vertices n in a graph:
  - Spectral decomposition: $\Omega(n^2)$ (widely believed: $\Omega(n^{2.376...})$)
  - Graph neural networks (e.g. graph convolutional network): $\Omega(n^2)$
- In most graph-based task, complexity of the task does not scale with n
  - n represents resolution of dataset, not complexity.

Realistic graphs demonstrate intrinsic simplicity

- ‘Small world phenomenon’.
- Scale-free/power-law graphs.

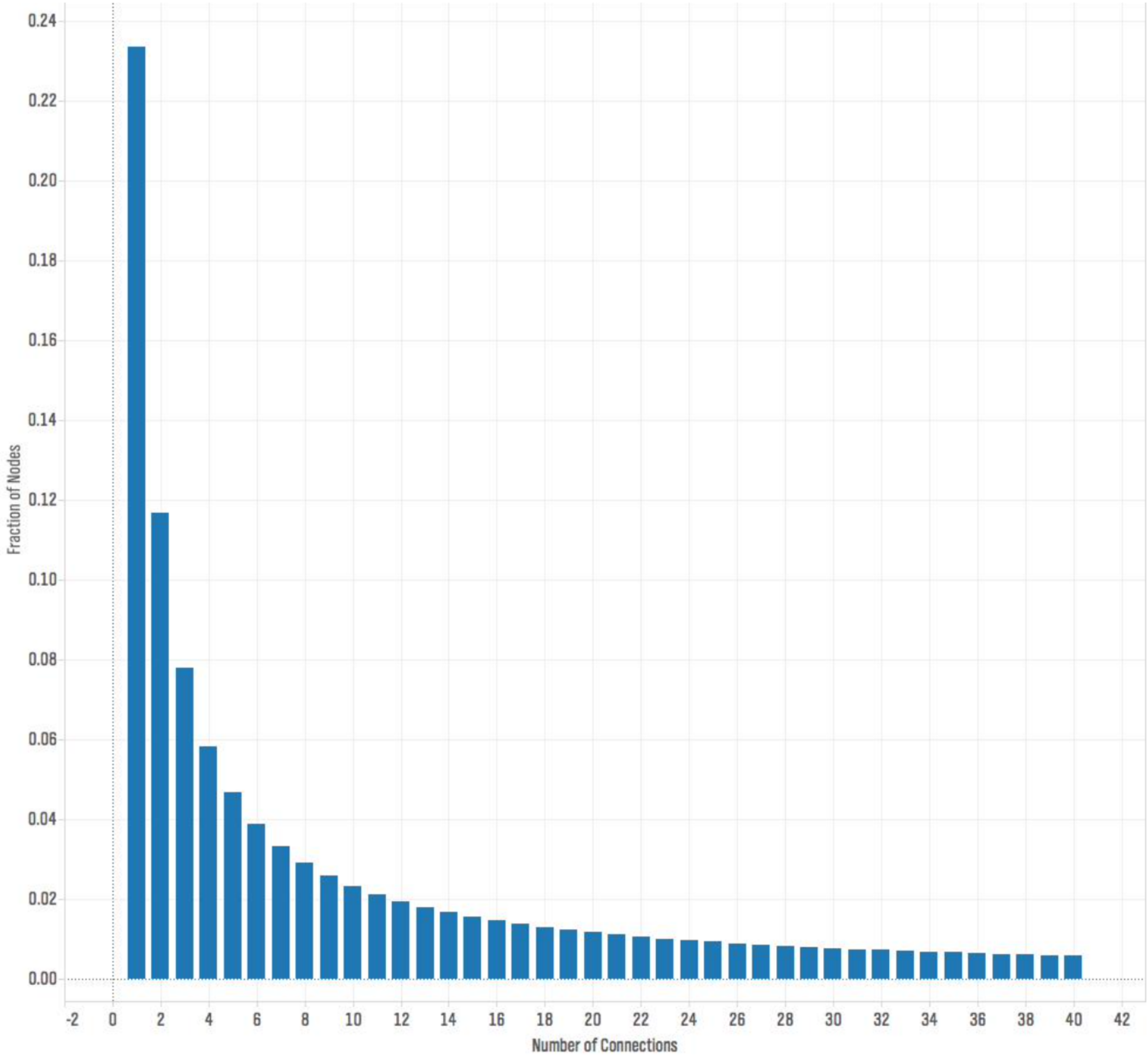


Figure by PJ Lamberson - UCLA

Reference: Kleinberg 2004, The small world phenomenon and decentralized search;
Milgram 1967, “The small world problem”;
Barabasi, Albert 1999, Emergence of scaling in random networks.

Solution: Sample small subset of nodes!

Two criteria

- Subsampling a small subset of k vertices, $k \ll n$, scales graph algorithms:

1. Find an "informative" subset.

   Strategy: optimal sampling

2. Find a "transferable" subset that is robust when the graph grows in size.

   Strategy: graph limit

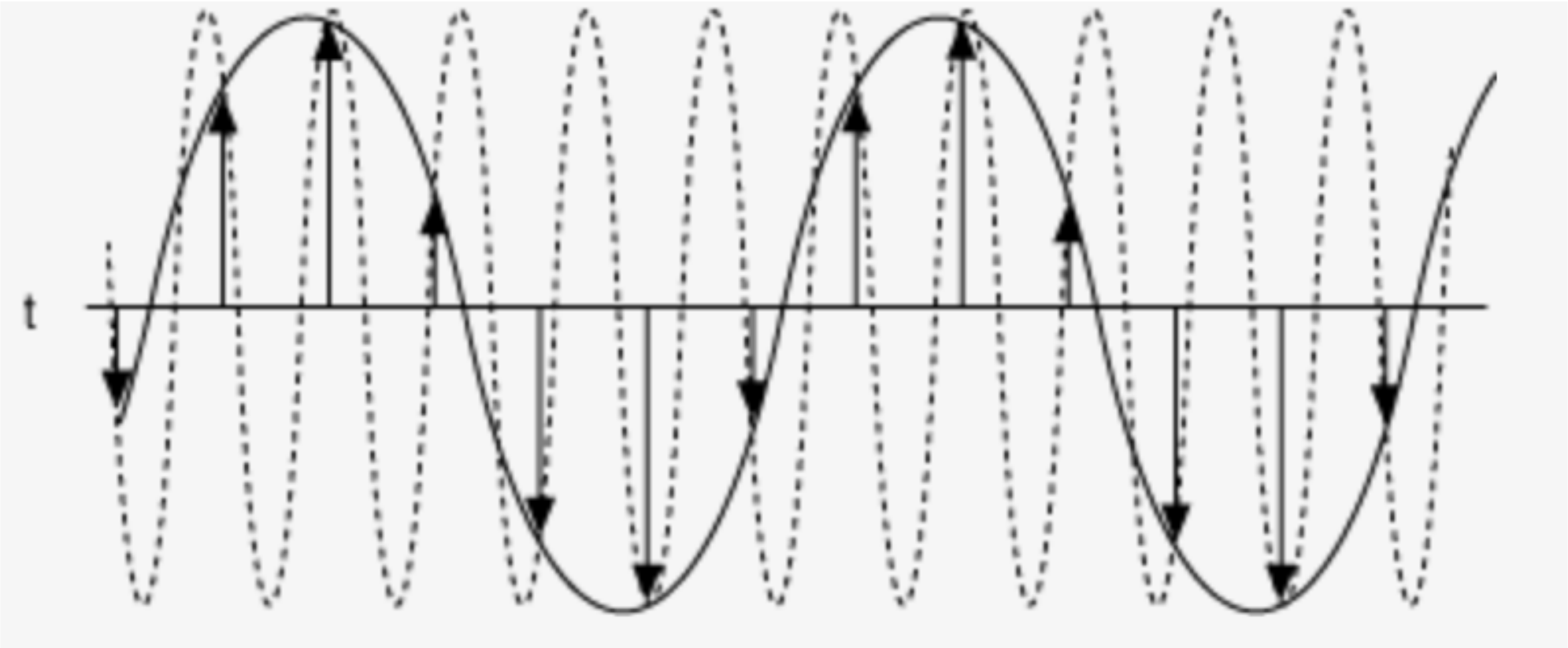
Look ahead: What can we do with such a subset?

- Faster existing algorithms:
  - Spectral decomposition (e.g. eigenvector positional encoding): $\cancel{\Omega(n^2)} \gg \Omega(k^2)$
  - Graph neural networks (e.g. graph convolutional network): $\cancel{\Omega(n^2)} \gg \Omega(k^2)$
- Theory: consistency/transferability results implies small decay in performance
- Empirical: node classification on citation network datasets

Informative subset: Sampling theory

Warm up: $L^1(\mathbb{R})$

Theorem (Shannon-Nyquist theorem): *An analog signal $f \in L^1(\mathbb{R})$ with bandwidth in $(0,2k)$ is uniquely determined by uniform discrete samples at rate k.*



Reference: Shannon, C.E., 1949. Communication in the Presence of Noise.
Figure by National Instruments

Informative subset: Sampling theory

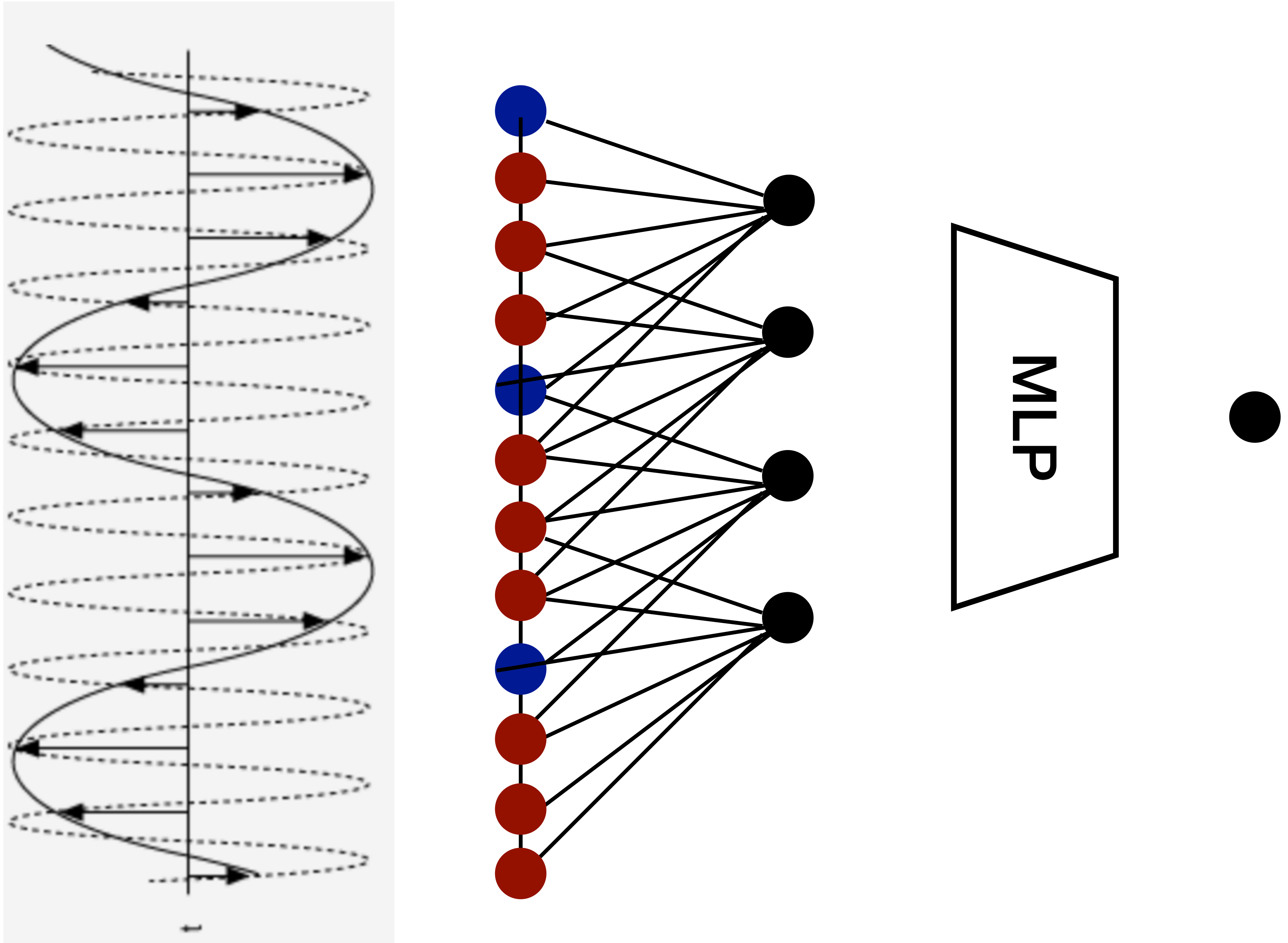
Corollary: 1D-CNN, path graph

Theorem (Shannon-Nyquist theorem): *An analog signal $f \in L^1(\mathbb{R})$ with bandwidth in $(0,2k)$ is uniquely determined by uniform discrete samples at rate k.*

Corollary (Sampling for 1D-CNN): *Given an analog 1D-image $f \in L^1([0,1])$ with bandwidth in $(0,2k)$ and n uniform pixels, only $k \ll n$ pixels is needed to uniquely determine f.*

Informative subset: Sampling theory

Corollary: 1D-CNN, path graph



blue nodes contain enough information to train the CNN

Corollary (Sampling for 1D-CNN): *Given an analog 1D-image $f \in L^1([0,1])$ with bandwidth in $(0,2k)$ and n uniform pixels, only $k \ll n$ pixels is needed to uniquely determine f.*

Informative subset: Sampling theory

Finite graph $G = (V, E)$ sampling

$f : V \to \mathbb{R}$

Theorem (Pesenson, 2008, informal): *A graph signal $f \in \ell^2(G)$ with limited bandwidth λ is uniquely determined by a uniqueness set $U \subseteq V$*

Normalized Laplacian $I - (D^\dagger)^{1/2} A (D^\dagger)^{1/2}$

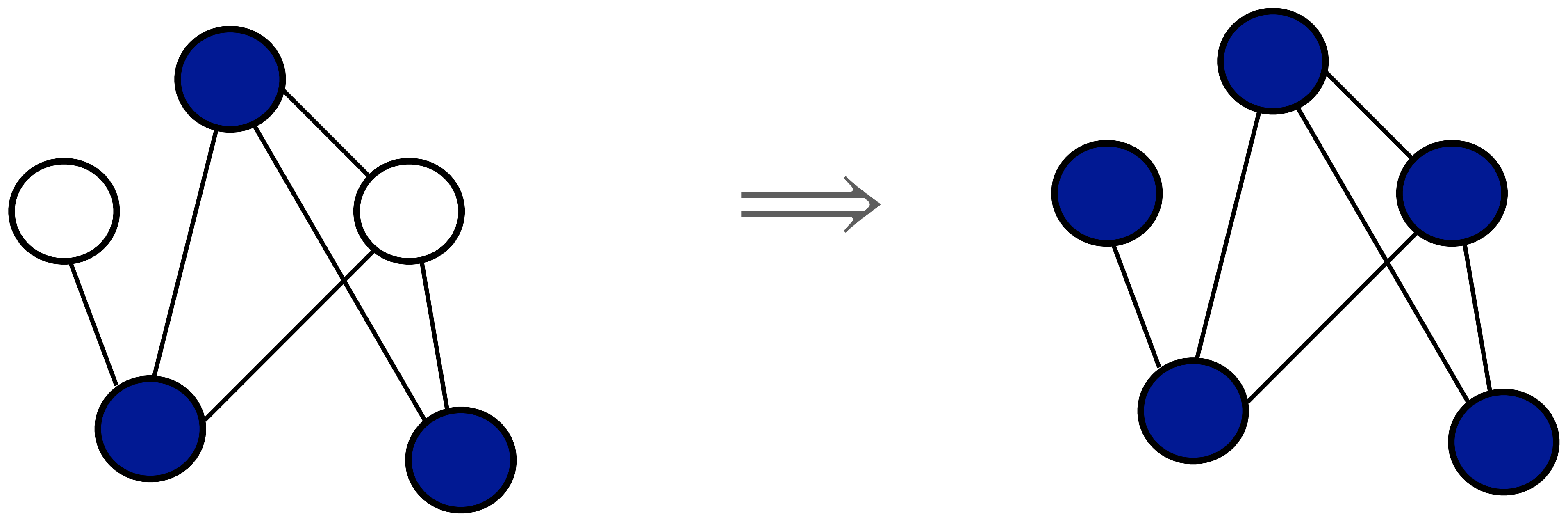
Graph Fourier transform

Eigenvalues/frequencies $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

Bandwidth cutoff at λ

Paley-Wiener space $PW_\lambda(G)$

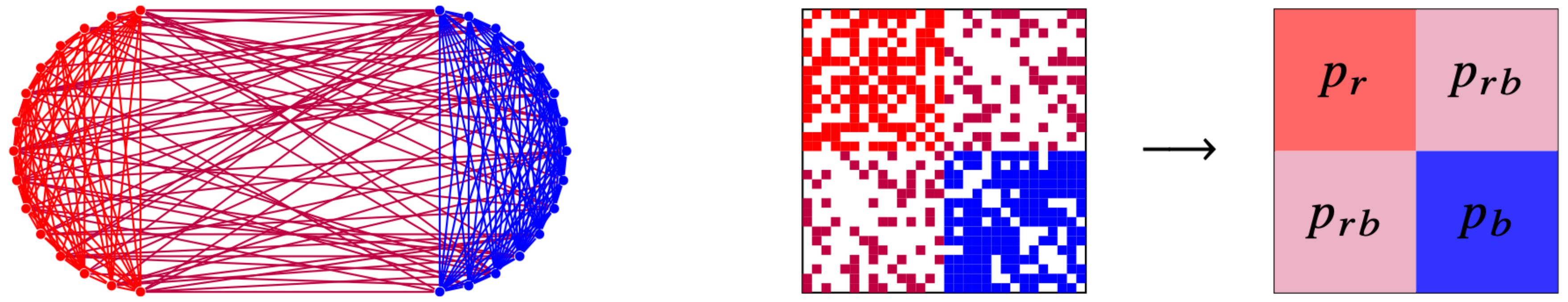
$\|f_1 - f_2\|_{\ell^2(U)} = 0 \implies \|f_1 - f_2\|_{\ell^2(G)} = 0$



Reference: Pesenson 2008. Sampling in Paley-Wiener spaces on combinatorial graphs.

Towards transferable subset: graph limit

Graphons and limit of dense graphs



- Graphons are symmetric, measurable functions $\mathbf{W} : [0,1] \times [0,1] \to \mathbb{R}$
- 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 $(v_i, v_j) \overset{\text{iid}}{\sim} \text{Bern}(\mathbf{W}(v_i, v_j))$

Reference: Lovasz, 2012, Large network and graph limits;
Borgs, Chayes, Lovasz, Sos and Vesztergombi 2006 - 2008: Convergent sequences of dense graphs I, II: Subgraph frequencies, metric properties and testing,
Figure: Zhao, Graph Theory and Additive Combinatorics pg 134

Reference: Ruiz, Chamon, and Ribeiro, 2023. Transferability properties of graph neural networks.

Towards transferable subset: graph limit

Transferability of graphs sampled from graphons

sampled frome same $\mathbf{W}$

- Transferability between graphs with similar structures: for some rate $c > 0$,

$$d_{?}(\mathrm{GNN}(\,\cdot\,, G_n, \theta), \mathrm{GNN}(\,\cdot\,, G_m, \theta)) \leq O(1/m^c + 1/n^c)$$

appropriate metric

input node features $\mathbb{R}^n$

input graph: n nodes

learnable parameter

Theorem (Ruiz et al. 2023; informal): *RHS = small number +* $O(1/m + 1/n)$

Transferable subset: graph limit

Putting together…

- Uniqueness set for a finite graph = informative subset
- Finite graphs sampled from the same limit graphons are structurally similar

$\implies$ Study how to sample informative subset from limit graphon and get transferability for free

Our results

Finite graph $G = (V, E)$ sampling

Informative subset: Sampling theory

$f : V \to \mathbb{R}$

Theorem (Pesenson, 2008, informal): *A graph signal $f \in \ell^2(G)$ with limited bandwidth λ is uniquely determined by a uniqueness set $U \subseteq V$*

Normalized Laplacian $I - (D^\dagger)^{1/2} A (D^\dagger)^{1/2}$

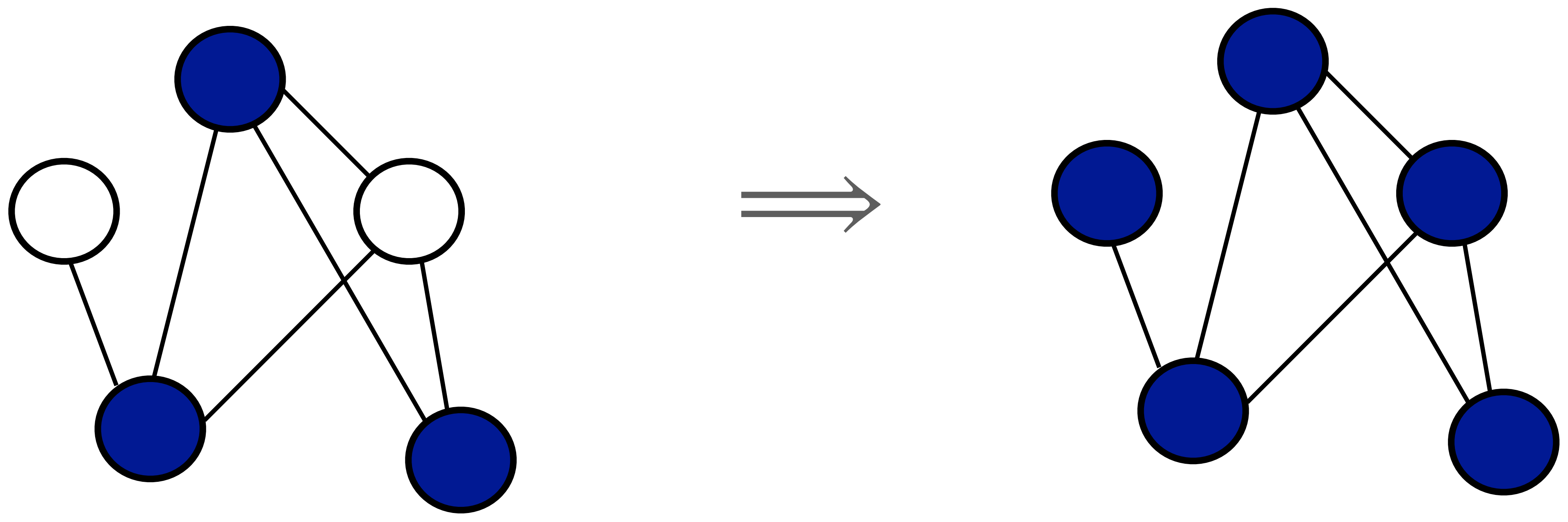
Graph Fourier transform

Eigenvalues/frequencies $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

Bandwidth cutoff at λ

Paley-Wiener space $PW_\lambda(G)$

$\|f_1 - f_2\|_{\ell^2(U)} = 0 \implies \|f_1 - f_2\|_{\ell^2(G)} = 0$



Reference: Pesenson 2008. Sampling in Paley-Wiener spaces on combinatorial graphs.

Graphon $\mathbf{W} : [0,1]^2 \to \mathbb{R}$ sampling (via Poincaré inequality)

Informative subset: Sampling theory

$f : [0,1] \to \mathbb{R}$

Theorem 2,3 (L., Ruiz, Jegelka, 2023; informal): *A graphon signal $f \in L^2([0,1])$ with limited bandwidth λ is uniquely determined by a msrbl uniqueness set s*

Normalized graphon Laplacian

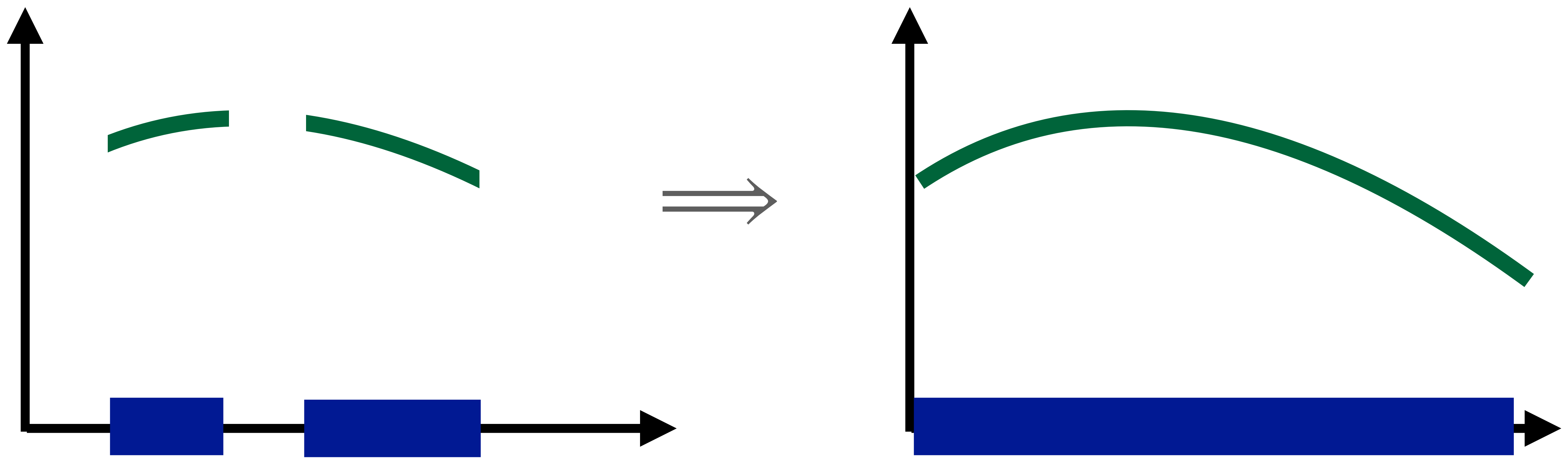
↓ Graphon Fourier transform

Eigenvalues/frequencies $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{-1}$

↓ Bandwidth cutoff at λ

graphon Paley-Wiener space $PW_\lambda(\mathbf{W})$

$$\|f_1 - f_2\|_{L^2(U)} = 0 \implies \|f_1 - f_2\|_{L^2([0,1])} = 0$$



Towards a sampling algorithm

Input: large graph G_N. Output: small uniqueness set for all $G_n, n \geq N$

1. Obtain an approximation of the limit graphon with G_N.

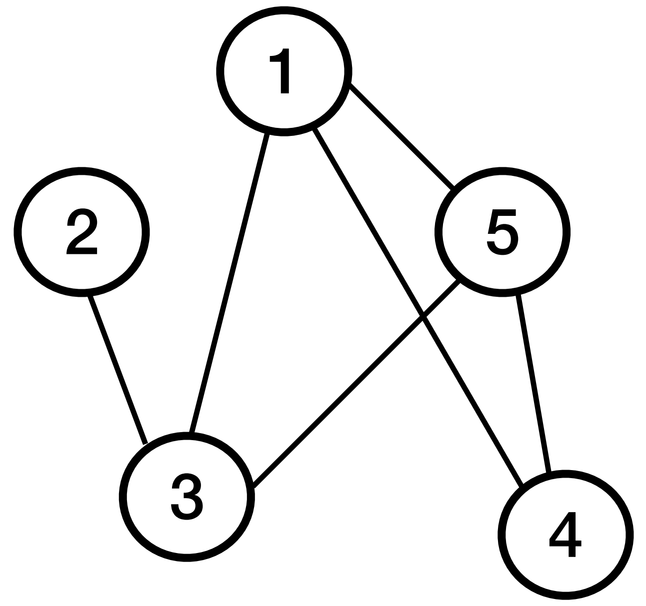
2. Approximate uniqueness set $U \subseteq [0,1]$ by G_N uniqueness set $U_N \subseteq [N]$.

3. Map $U_N \subseteq [N]$ to nodes $U_n \subseteq [n]$ of finite graph $G_n, n \geq N$.

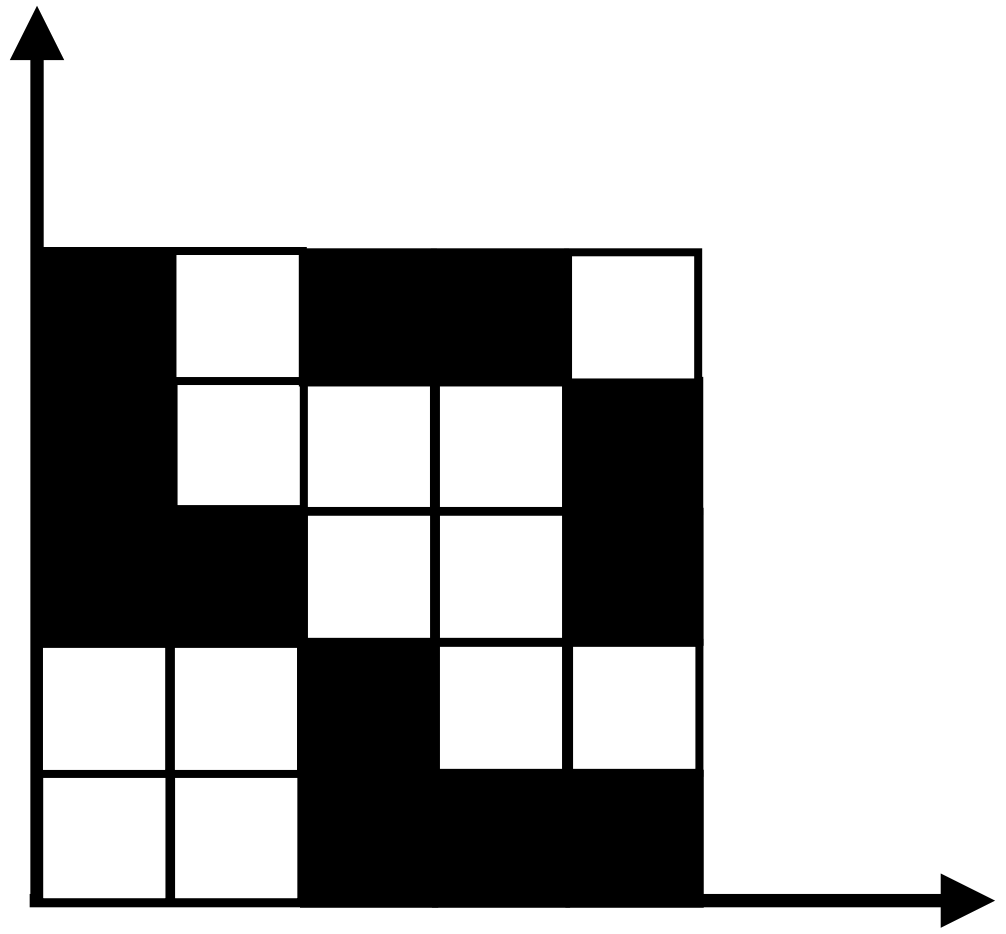
1. Obtain an approximation of limit graphon

Graphon is the closure of graphs (under cut norm)

- Embed G_N vertices into $[0,1]$ - induced graphon



$\Longrightarrow$



Theorem (Lovász and Szegedy 2007): *Graphon space is compact. Finite graphs are dense.*

Towards a sampling algorithm

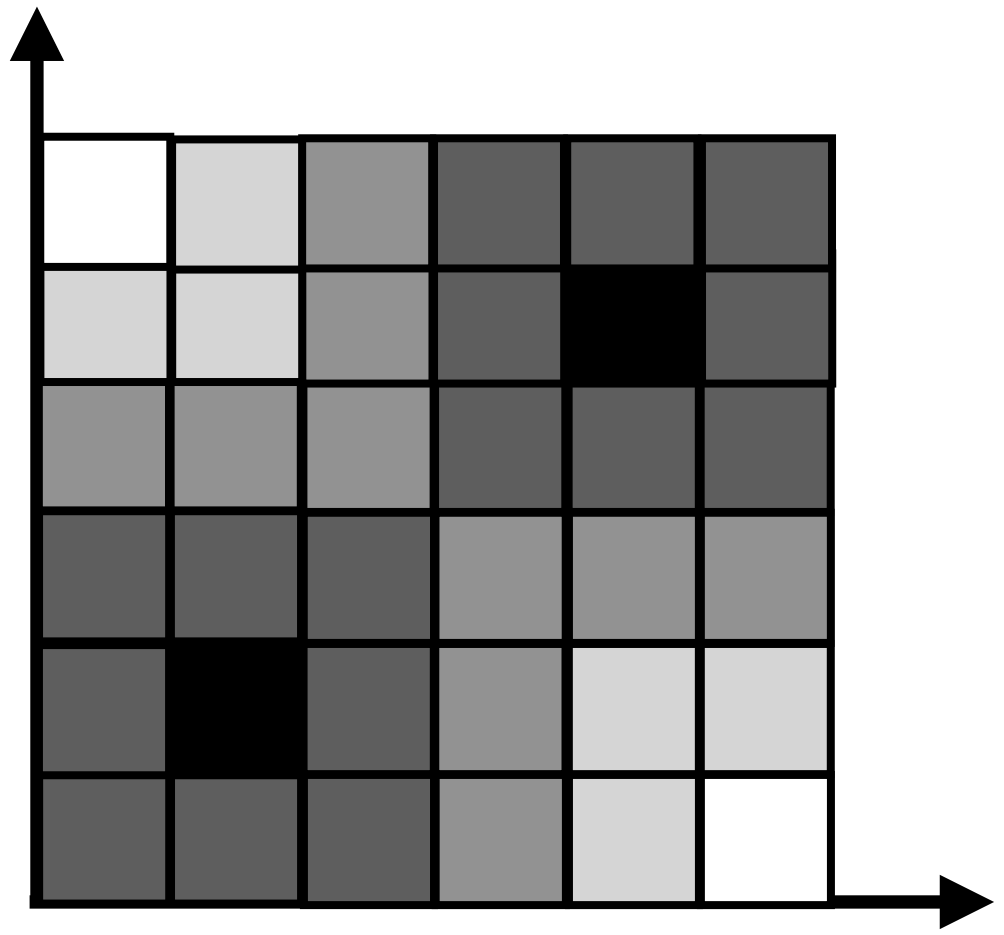
Input: large graph G_N. Output: small uniqueness set for all $G_n, n \geq N$

1. Obtain an approximation of the limit graphon with G_N.
2. Approximate uniqueness set $U \subseteq [0,1]$ by G_N uniqueness set $U_N \subseteq [N]$.
3. Map $U_N \subseteq [N]$ to nodes $U_n \subseteq [n]$ of finite graph $G_n, n \geq N$.

2. Approximate uniqueness set of graphons

Graphon $\mathbf{W} \equiv$ mixture model of random graphs $\mathbf{K} = \left(\Omega, \sum_i \mathbb{P}_i, \mathbf{k}\right)$

Proposition 2 (L., Ruiz, Jegelka, 2023; informal): *A graphon random model is equivalent to mixture models of random graphs. Well-fittedness measured by a difficulty function $\varphi(\mathbf{W}; \mathbf{k}, \mathbb{P}_i)$ (Schiebinger et al. 2015)*



$$\equiv d\mathbb{P} = \frac{p_{\mathcal{N}(0.25,\sigma^2)} + p_{\mathcal{N}(0.,75,\sigma^2)}}{2}$$

2. Approximate uniqueness set of graphons

Gaussian elimination find uniqueness set with high probability

Theorem (L., Ruiz, Jegelka, 2023; informal) - Corollary of Theorem 2 (Schiebinger et. al, 2015): *If $\varphi \ll 1$, GE <u>with pivoting</u> find uniqueness set with high probability.*

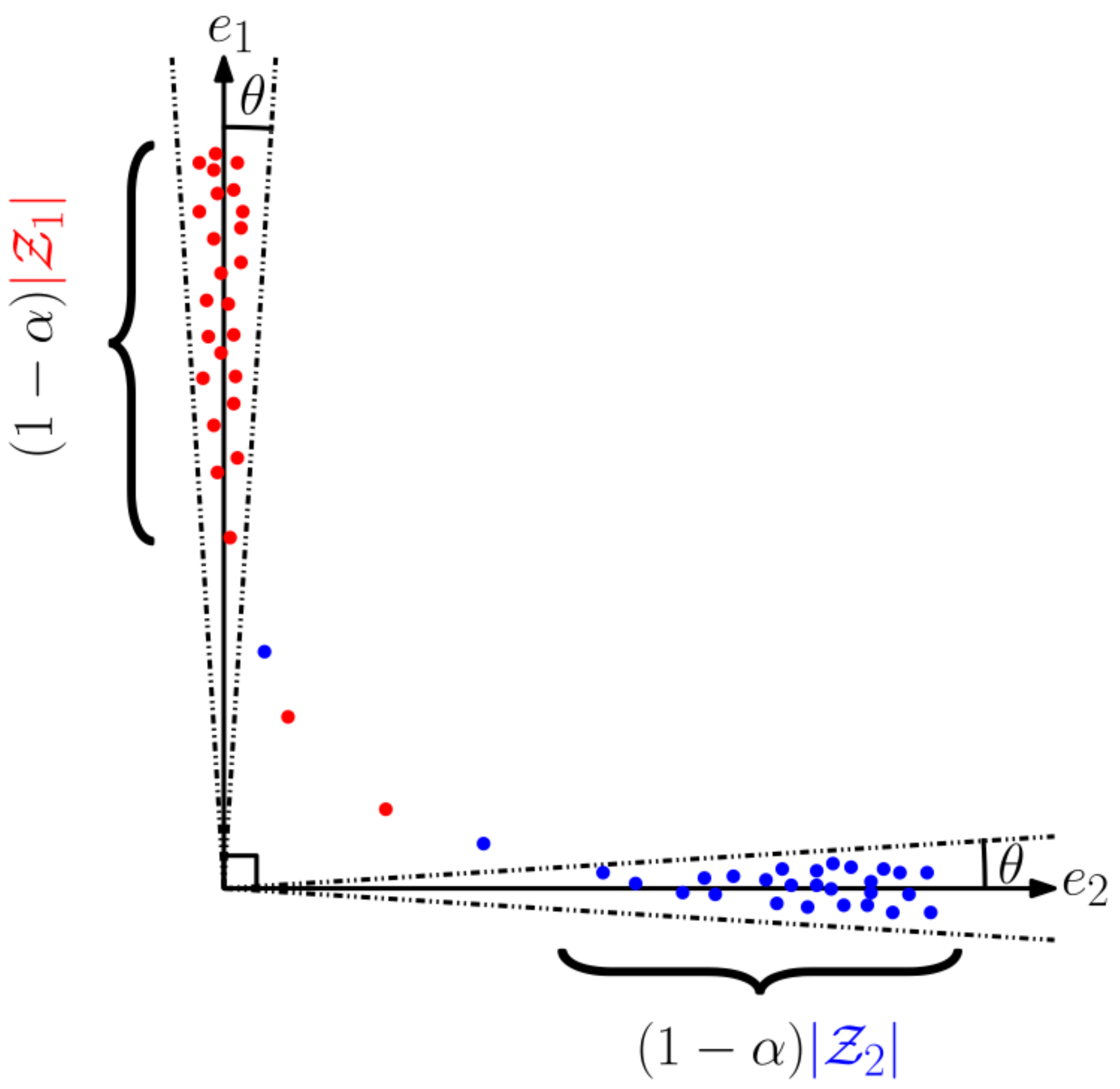


Figure from Schiebinger et al.
Reference: Schiebinger, Wainwright, Yu 2015. The geometry of kernelized spectral clustering

Towards a sampling algorithm

Input: large graph G_N. Output: small uniqueness set for all $G_n, n \geq N$

1. Obtain an approximation of the limit graphon with G_N.
2. Approximate uniqueness set $U \subseteq [0,1]$ by G_N uniqueness set $U_N \subseteq [N]$.
3. Map $U_N \subseteq [N]$ to nodes $U_n \subseteq [n]$ of finite graph $G_n, n \geq N$.

3. Map graphon uniqueness set to nodes in graph

Consistency result

Proposition 6 (L., Ruiz, Jegelka, 2023; informal): *If $\varphi \ll 1$, for a sequence of graphs $G_N \to \mathbf{W}$, there exists a number of node N such that for all $n > N$, uniqueness set of G_N is also a uniqueness set of G_n.*

Empirical results

Experiments

Transferability

Table 2: Classification accuracy on the MalNet-Tiny dataset, (i) w/o positional encodings (PEs), (ii) w/ PEs computed on the full graph, (iii) w/ PEs computed on a graphon-sampled subgraph (removing or not isolated nodes), and (iv) w/ PEs computed on a subgraph with randomly sampled nodes (removing or not isolated nodes).

	no PEs	full graph PEs	graphon sampled PEs		randomly sampled PEs	
			w/ isolated	w/o	w/ isolated	w/o
mean	0.26±0.03	0.43±0.07	**0.29±0.06**	**0.33±0.06**	0.28±0.07	0.27±0.07
max	0.30	0.51	**0.40**	**0.42**	0.35	0.37

Conclusion

- Subsampling a small subset of k vertices, $k \ll n$, scales graph algorithms:
  - Find an `informative' subset: sampling theory
  - Find a `transferable' subset: graph limit
- We derived a consistent algorithm that approximately samples from the limiting graphon uniqueness set, under theoretical guarantees.
- We tested our approach in real-world datasets, showing improvement over uniformly random sampling.

Thank you!

Q&A

Massachusetts Institute of Technology

- Major thanks to my collaborators Luana Ruiz and Stefanie Jegelka
- Contact: thienle@mit.edu
- Full paper: