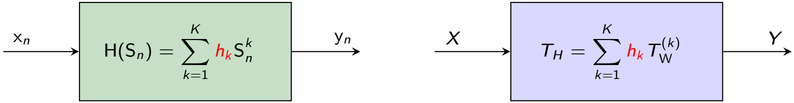
Convergence of Graph Filters in the Spectral Domain

- Convergence of graph filter sequences towards graphon filters for convergent graph signal sequences

Graphon Filters and Sequences of Graph Filters

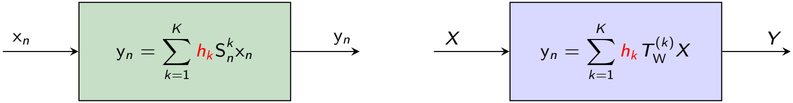
- Given coefficients h_k consider a graph filter sequence and a graphon filter with the same coefficients



- Does the graph filter sequence converge to the graphon filter? $\Rightarrow$ Not the most pertinent question

$\Rightarrow$ Filter convergence is important inasmuch as it implies convergence of filter outputs

- Given coefficients h_k consider a graph filter sequence and a graphon filter with the same coefficients



- Consider a convergent sequence of graph signals $(G_n, x_n) \rightarrow (W, X)$

  $\Rightarrow$ Input graph signal x_n to graph filter $H(S_n)$ to produce output graph signal y_n

  $\Rightarrow$ Input graphon signal X to graphon filter T_H to produce output graphon signal Y

- The graph signal sequence (G_n, y_n) converges to the graphon signal (W, Y) under some conditions

- Given filter coefficients h_k we have five polynomials which are the same except for their variables

- Two polynomials are representations in the node domain

$\Rightarrow$ The graph filter sequence defined on variable $\mathbf{S}_n$ $\Rightarrow \mathbf{H}(\mathbf{S}_n) = \sum_{k=1}^{K} h_k \mathbf{S}_n^k$

$\Rightarrow$ The graphon filter defined on variable $T_{\mathbf{W}}$ $\Rightarrow T_H = \sum_{k=1}^{K} h_k T_{\mathbf{W}}^{(k)}$

- Given filter coefficients h_k we have five polynomials which are the same except for their variables

- Three polynomials are representations in the spectral domain

$\Rightarrow$ The frequency response of the graph and graphon filters with variable λ $\Rightarrow \tilde{h}(\lambda) = \sum_{k=1}^{K} h_k \lambda^{(k)}$

$\Rightarrow$ The frequency representation of the graph filters with variable λ_{nj} $\Rightarrow \tilde{h}(\lambda_{nj}) = \sum_{k=1}^{K} h_k \lambda_{nj}^{(k)}$

$\Rightarrow$ The frequency representation of the graphon filter with variable λ_j $\Rightarrow \tilde{h}(\lambda_j) = \sum_{k=1}^{K} h_k \lambda_j^{(k)}$

⇒ Frequency representation of graph filters ⇒ $\tilde{h}(\lambda_{nj}) = \sum_{k=1}^{K} h_k \lambda_{nj}^k$

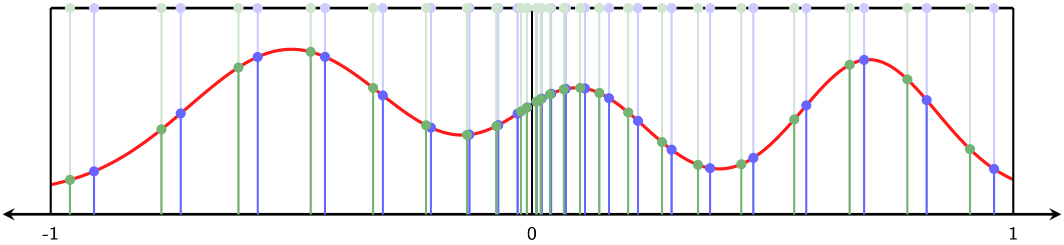
⇒ Frequency representation of graphon filter ⇒ $\tilde{h}(\lambda_j) = \sum_{k=1}^{K} h_k \lambda_j^k$

Theorem (Convergence of graph filter sequences in the frequency domain)

Consider filter coefficients h_k generating a sequence of graph filters $\mathbf{H}(\mathbf{S}_n)$ supported on the graph sequence G_n and a graphon filter T_H supported on the graphon W. If $G_n \to W$

$$\lim_{n\to\infty} \tilde{h}(\lambda_{nj}) = \tilde{h}(\lambda_j)$$

- Graph filter GFT representations converge to graphon filter WFT representation $\Rightarrow \lim_{n\to\infty} \tilde{h}(\lambda_{nj}) = \tilde{h}(\lambda_j)$
- This is true because eigenvalues converge and the frequency responses are the same
- This is not much to say $\Rightarrow$ GFT and WFT are representations. $\Rightarrow$ Filters operate in the node domain

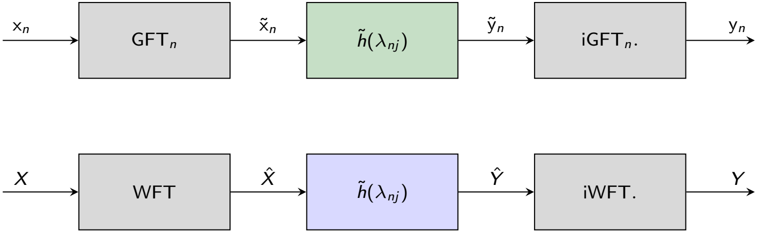


Convergence of Graph Filters in the Node Domain

- We leverage spectral domain convergence to prove convergence of graph filters in the node domain

$\Rightarrow$ Provides a first approach to the study of transferability of graph filters

- To prove convergence in the node domain we can go to the frequency domain and back



- Frequency representation of graph filters converge to frequency representation of graphon filter

$\Rightarrow$ But the GFT and the iGFT do not converge $\Rightarrow$ Unless the signals are graphon bandlimited

- Input graph signal sequence $(G_n, \mathbf{x}_n)$ $\Rightarrow$ Generates output sequence $(G_n, \mathbf{y}_n)$ with $\mathbf{y}_n = \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n$
- Input graphon signal (W, X) $\Rightarrow$ Generates output signal (W, Y) with $Y = T_H X$

Theorem (Graph filter convergence for bandlimited inputs)

Given convergent graph signal sequence $(G_n, \mathbf{x}_n) \to (W, X)$ and filters $\mathbf{H}(\mathbf{S}_n)$ and T_H generated by the same coefficients h_k. If the input signals are c-bandlimited

$$(\mathrm{G}_n, \mathrm{y}_n) \to (\mathrm{W}, Y)$$

The sequence of output graph signals converges to the output graphon signal

- Convergence for bandlimited input is easy. Also weak. Therefore cheap. A stronger result is possible

- Lipschitz graphon filters are filters with frequency responses that are Lipschitz in $[-1, 1]$

$$\Big| h(\lambda_1) - h(\lambda_2) \Big| \;\leq\; L \Big| \lambda_1 - \lambda_2 \Big|, \quad \text{for all } \lambda_1, \lambda_2 \in [0, 1]$$

- Claim convergence of graph filter sequence, despite lack of convergence of the GFT and the iGFT

Theorem (Graph filter convergence for Lipschitz continuous filters)

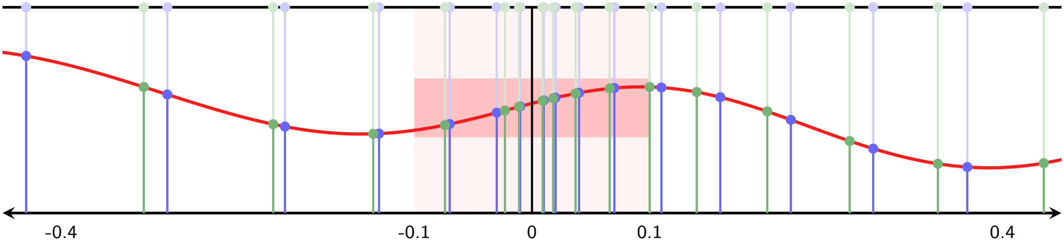
Given convergent graph signal sequence $(G_n, \mathrm{x}_n) \to (W, X)$ and filters $\mathrm{H}(\mathrm{S}_n)$ and T_H generated by the same coefficients h_k. If the frequency response $\tilde{h}(\lambda)$ is Lipschitz

$$(\mathrm{G}_n, \mathrm{y}_n) \to (\mathrm{W}, Y)$$

The sequence of output graph signals converges to the output graphon signal

Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ■

- The challenge of filter convergence comes from the accumulation of eigenvalues around $\lambda = 0$
- Which causes complications with eigenvector convergence.
- Lipschitz continuity renders the effect void. All components are multiplied by similar numbers

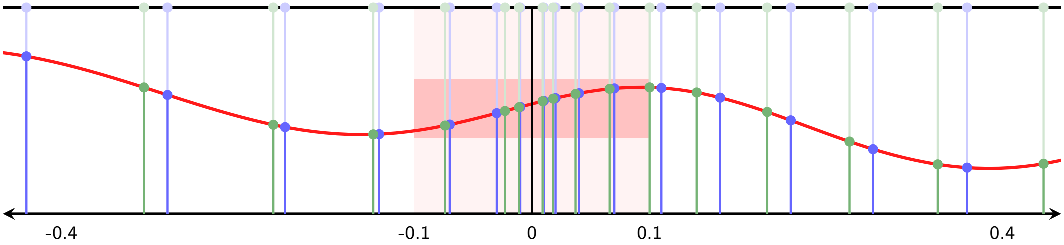


Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ■

- We identify a fundamental issue $\Rightarrow$ Transferability is counter to discriminability

  $\Rightarrow$ If the filter converges, it can't separate eigenvectors associated to eigenvalues close to $\lambda = 0$

- Characterization is just a limit $\Rightarrow$ Work on a finite-n transference bounding



Graphon Filters are Generative Models for Graph Filters

- Graph filters can approximate graphon filters under certain conditions. We discuss them now.

Graphon Filters are Generative Models for Graph Filters

- For a converging graph sequence, graph filters converge asymptotically to graphon filters

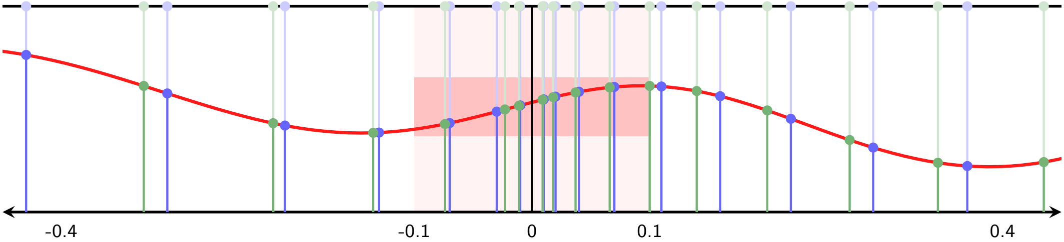
- Thus, as n grows, the graph filters become more similar to the graphon filter

$$\mathbf{x}_n \longrightarrow \boxed{\mathbf{y}_n = \sum_{k=1}^{K} h_k \mathbf{S}_n^k \mathbf{x}_n} \longrightarrow \mathbf{y}_n \qquad\qquad X \longrightarrow \boxed{\mathbf{y}_n = \sum_{k=1}^{K} h_k T_{\mathbf{W}}^{(k)} X} \longrightarrow Y$$

- And we can then use a graph filter as a surrogate for the graphon filter

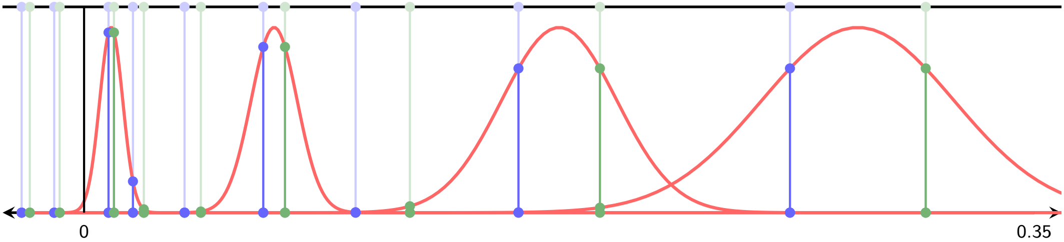
- We now want to quantify the quality of that approximation for different values of n

- Graphon eigenvalues accumulate at $\lambda = 0$
- Making it hard to match graph eigenvalues to the corresponding graphon eigenvalues if λ is small

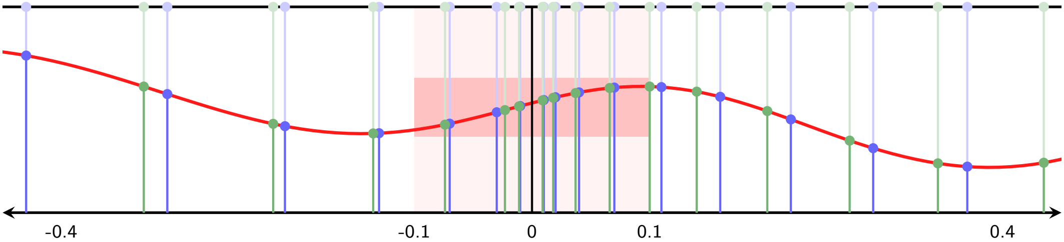


Small Eigenvalues are Hard to Discriminate

- Which in turn makes it hard to discriminate consecutive eigenvalues in that range
- If the filter changes rapidly near zero, it may modify the graph and graphon eigenvalues differently
- To obtain good approximations, we must then assume filters do not change much around $\lambda = 0$

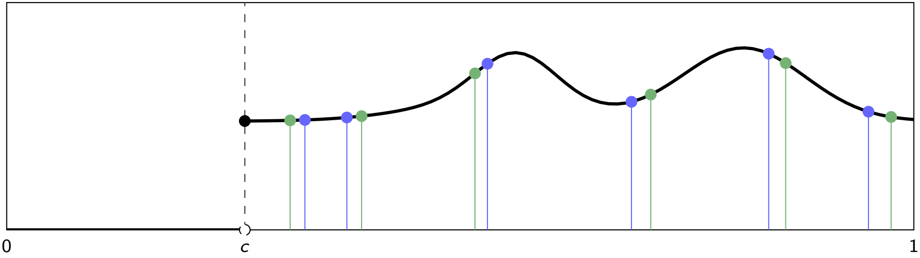


- Which in turn makes it hard to discriminate consecutive eigenvalues in that range
- If the filter changes rapidly near zero, it may modify the graph and graphon eigenvalues differently
- To obtain good approximations, we must then assume filters do not change much around $\lambda = 0$



- Graphon eigenvalues tend to zero as the index i grows $\Rightarrow \lim_{i\to\infty} \lambda_i = \lim_{i\to\infty} \lambda_{-i} = 0$

- Low-pass graphon filters must thus be zero for $\lambda < c$. Constant c determines the filter's band.



- The filter removes high frequency components. But low-frequency components are not affected.

(A1) The graphon W is L_1-Lipschitz $\Rightarrow$ For all arguments (u_1, v_1) and (u_2, v_2), it holds

$$\left| W(u_2, v_2) - W(u_1, v_1) \right| \leq L_1 \Big(\left| u_2 - u_1 \right| + \left| v_2 - v_1 \right| \Big)$$

(A2) The filter's response is L_2-Lipschitz and normalized $\Rightarrow$ For all λ_1, λ_2 and λ we have

$$\left| \tilde{h}(\lambda_2) - \tilde{h}(\lambda_1) \right| \leq L_2 \left| \lambda_2 - \lambda_1 \right| \quad \text{and} \quad \left| h(\lambda) \right| \leq 1$$

(A3) The graphon signal X is L_3-Lipschitz $\Rightarrow$ For all u_1 and u_2

$$\left| X(u_2) - X(u_1) \right| \leq L_3 \left| u_2 - u_1 \right|$$

- We fix a bandwidth $c > 0$ to separate eigenvalues not close to $\lambda = 0$ and define

(D1) The c-band cardinality of G_n is the number of eigenvalues with absolute value larger than c

$$B_{nc} = \#\left\{ \lambda_{ni} \,:\, |\lambda_{ni}| > c \right\}$$

(D2) The c-eigenvalue margin of graph G_n is the

$$\delta_{nc} = \min_{i, j \neq i} \left\{ |\lambda_{ni} - \lambda_j| \,:\, |\lambda_{ni}| > c \right\}$$

- Where λ_{ni} are eigenvalues of the shift operator S_n and λ_j are eigenvalues of graphon W

Theorem (Graphon filter approximation by graph filter for low-pass filters)

Consider a graphon filter $Y = \Phi(X; h, W)$ and a graph filter $y_n = \Phi(x_n; h, S_n)$ instantiated from Y. With Definitions **(D1)** - **(D2)**, Assumptions **(A1)** - **(A3)**, and

(A4) $h(\lambda)$ is zero for $|\lambda| < c$

The difference between Y and $Y_n = \Phi(X_n; h, W_n)$ (graph filter induced by y_n) is bounded by

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1}\left(L_2 + \frac{\pi n_c}{\delta_{nc}}\right) n^{-\frac{1}{2}} \|X\|_{L_2} + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}}$$

Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ■

- High-pass filters have null frequency response for $|\lambda| > c$, removing low-frequency components
- Moreover, we consider filters that have low variability around $\lambda = 0$

- This makes it easier to match graph eigenvalues to graphon eigenvalues around $\lambda = 0$

Graph-Graphon Filter Approximation Theorem for High-Pass Filters

Theorem (Graphon filter approximation by graph filter for high-pass filters)

Consider a graphon filter $Y = \Phi(X; h, W)$ and a graph filter $y_n = \Phi(x_n; h, S_n)$ instantiated from Y. With Definitions **(D1)** - **(D2)**, Assumptions **(A1)** - **(A3)**, and

(A4) $h(\lambda)$ is zero for $|\lambda| > c$

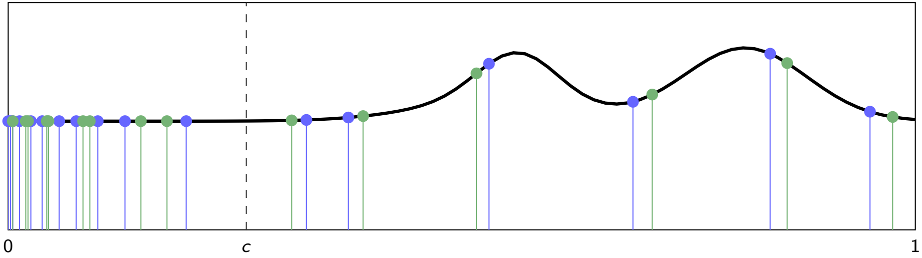
The difference between Y and $Y_n = \Phi(X_n; h, W_n)$ (graph filter induced by y_n) is bounded by

$$\|Y - Y_n\|_{L_2} \leq L_2 c \|X\|$$

Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ■

Lipschitz Filters with Variable Band

- Filter response has low variability for $|\lambda| < c$. Where the eigenvalues of the graphon accumulate
- For $|\lambda| > c$, graphon eigenvalues are countable. And easier to match to those of the graph



- A Lipschitz filter with variable band is the composition of a low-pass filter and a high-pass one

Theorem (Graphon filter approximation by graph filter)

Consider a graphon filter $Y = \Phi(X; h, W)$ and a graph filter $y_n = \Phi(x_n; h, S_n)$ instantiated from Y. With Definitions **(D1)** - **(D2)**, Assumptions **(A1)** - **(A3)**, and

(A4) $h(\lambda)$ has low variability for $|\lambda| < c$

The difference between Y and $Y_n = \Phi(X_n; h, W_n)$ (graph filter induced by y_n) is bounded by

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1}\left(L_2 + \frac{\pi n_c}{\delta_{nc}}\right) n^{-\frac{1}{2}}\|X\|_{L_2} + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} + L_2 c\|X\|$$

- Filter with variable band is the sum of an L_2-Lipschitz filter $h_1(\lambda)$ with $h_1(\lambda) = 0$ for $|\lambda| < c$
- And a high-pass filter $h_2(\lambda)$ with $h_2(\lambda)$ showing low variability for $|\lambda| < c$ and 0 otherwise
- Thus, by the triangle inequality

$$\|Y - Y_n\|_{L_2} = \|T_{\mathrm{H}}X - T_{\mathrm{H}_n}\|_{L_2} \leq \|T_{\mathrm{H}_1}X - T_{\mathrm{H}_{1_n}}X_n\|_{L_2} + \|T_{\mathrm{H}_2}X - T_{\mathrm{H}_{2_n}}X_n\|_{L_2}$$

- We know the first-term on the right-hand side. It's the bound for low-pass filters
- And the second-term on the right-hand side is the bound for constant filters
- Summing up the two bounds, we then prove our result for Lipschitz filters with variable band

Theorem (Graphon filter approximation by graph filter)

The difference between Y and $Y_n = \Phi(X_n; \mathrm{h}, \mathrm{W}_n)$ (graph filter induced by y_n) is bounded by

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1}\left(L_2 + \frac{\pi n_c}{\delta_{nc}}\right) n^{-\frac{1}{2}} \|X\|_{L_2} + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} + L_2 c \|X\|$$

- Bound depends on the filter transferability constant and on the difference between X and X_n
- Transferability constant depends on the graphon via L_1 which also affects the graphon variability
- As n grows, the transferability constant dominates the bound

Filter Response Determines the Approximation Bound

Theorem (Graphon filter approximation by graph filter)

The difference between Y and $Y_n = \Phi(X_n; \mathrm{h}, \mathrm{W}_n)$ (graph filter induced by y_n) is bounded by

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1}\left(L_2 + \frac{\pi n_c}{\delta_{nc}}\right) n^{-\frac{1}{2}} \|X\|_{L_2} + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} + L_2 c \|X\|$$

- Transferability constant depends on the filter parameters L_2, n_c and δ_{nc}
- Filter's Lipschitz constant L_2 and filter's band $[c, 1]$ determine variability of the spectral response
- Number of eigenvalues in the passing band has to be limited: $n_c < \sqrt{n}$
- This ensures eigenvalues of W_n converge to those of W. And thus so does the filter approximation

Discriminability - Approximation Trade-Off

- We identify a fundamental issue $\Rightarrow$ Good approximations are counter to discriminability

$\Rightarrow$ Tight approximation bounds require filters with low variability around $\lambda = 0$

$\Rightarrow$ But then the filter can't discriminate components associated to eigenvalues close to $\lambda = 0$

- That is less of an issue for larger graphs. Filter approximation requires $n_c < \sqrt{n}$

$\Rightarrow$ As n grows, we can afford a larger number of eigenvalues n_c in the passing band

$\Rightarrow$ Improving discriminability without penalizing the approximation bound

Transferability of Graph Filters: Theorem

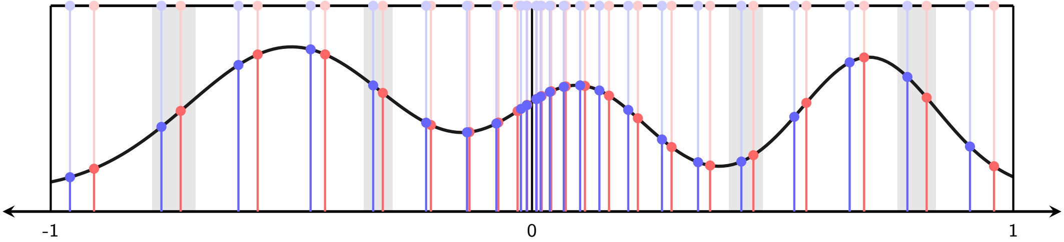
- We show that graph filters are transferable across graphs that are drawn from a common graphon

Comparing Graph Filters through their Generating Graphon Filter

- Have not proven transferability $\Rightarrow$ Have proven that graph filters are close to graphon filters

$\Rightarrow$ Graph G_n with n nodes sampled from graphon W

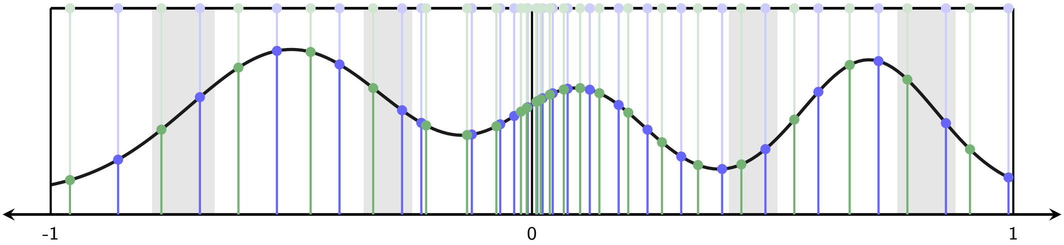
$\Rightarrow$ Have shown that graph filter $H(S_n)$ running on G_n is close to the graphon filter T_H



- Transferability means that two different graphs with different number of nodes are close

$\Rightarrow$ Graph G_n and graph G_m with $n \neq m$ nodes. Both sampled from graphon W

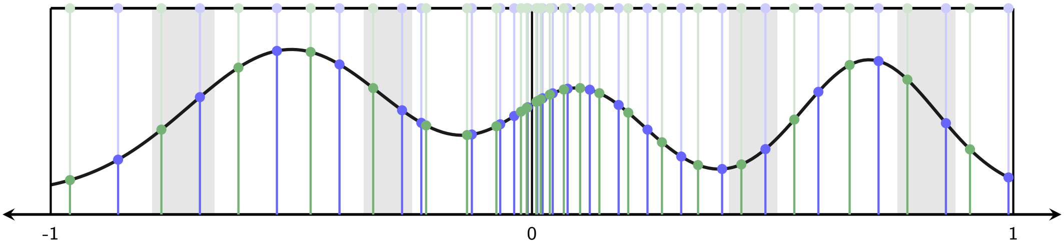
$\Rightarrow$ Want to show that graph filter $\mathrm{H}(\mathrm{S}_n)$ and graph filter $\mathrm{H}(\mathrm{S}_m)$ are close



- But graph filters are close because they are both close to the graphon filter

$\Rightarrow$ Graph filter $H(S_n)$ close to graphon filter T_H. Graph filter $H(S_m)$ close to graphon filter T_H

$\Rightarrow$ Graph filter $H(S_n)$ is close to graph filter $H(S_m)$ $\Rightarrow$ This is just the triangle inequality



Running the Same Filter on Different Graphs

- Consider graph signals (S_n, x_n) and (S_m, x_m) sampled from the graphon signal (W, X)

- Given filter coefficients h_k we process signals on their respective graphs

$\Rightarrow$ Run filter with coefficients h_k on graph S_n to process x_n $\Rightarrow y_n = H(S_n)x_n = \sum_{k=1}^{K} h_k S_n^k x_n$

$\Rightarrow$ Run filter with coefficients h_k on graph S_m to process x_m $\Rightarrow y_m = H(S_m)x_m = \sum_{k=1}^{K} h_k S_m^k x_n$

- Since they have different number of components we compare induced graphon signals Y_n and Y_m

(A1) The graphon W is L_1-Lipschitz $\Rightarrow$ For all arguments (u_1, v_1) and (u_2, v_2), it holds

$$\Big| W(u_2, v_2) - W(u_1, v_1) \Big| \;\leq\; L_1 \Big(\big| u_2 - u_1 \big| \;+\; \big| v_2 - v_1 \big| \Big)$$

(A2) The filter's response is L_2-Lipschitz and normalized $\Rightarrow$ For all λ_1, λ_2 and λ we have

$$\big| \tilde{h}(\lambda_2) - \tilde{h}(\lambda_1) \big| \;\leq\; L_2 \big| \lambda_2 - \lambda_1 \big| \qquad \text{and} \qquad \big| h(\lambda) \big| \;\leq\; 1$$

(A3) The graphon signal X is L_3-Lipschitz $\Rightarrow$ For all u_1 and u_2

$$\big| X(u_2) - X(u_1) \big| \;\leq\; L_3 \big| u_2 - u_1 \big|$$

- We fix a bandwidth $c > 0$ to separate eigenvalues not close to $\lambda = 0$ and define

(D1) The c-band cardinality of G_n is the number of eigenvalues with absolute value larger than c

$$B_{nc} = \#\left\{ \lambda_{ni} \,:\, |\lambda_{ni}| > c \right\}$$

(D2) The c-eigenvalue margin of of graph G_n is the

$$\delta_{nc} = \min_{i,j \neq i} \left\{ |\lambda_{ni} - \lambda_j| \,:\, |\lambda_{ni}| > c \right\}$$

- Where λ_{ni} are eigenvalues of the shift operator S_n and λ_j are eigenvalues of graphon W

Theorem (Graph filter transferability)

Consider graph signals $(\mathbf{S}_n, \mathbf{x}_n)$ and $(\mathbf{S}_m, \mathbf{x}_m)$ sampled from graphon signal (W, X) along with filter outputs $\mathbf{y}_n = \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n$ and $\mathbf{y}_m = \mathbf{H}(\mathbf{S}_m)\mathbf{x}_m$. With Assumptions (A1)-(A3) and Definitions (D1)-(D2) the difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \leq \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ■

Transferability of Graph Filters: Remarks

- We present remarks on the transferability theorem of graph filters sampled from a graphon filter

Theorem (Graph filter transferability)

The difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \leq \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

Thing 1: A term that comes from the discretization of the graphon signal $\Rightarrow$ Not very important

Thing 2: A term coming from filter variability at eigenvalues $|\lambda| > c$ $\Rightarrow$ The easy components

Thing 3: A term coming from filter variability at eigenvalues $|\lambda| \leq c$ $\Rightarrow$ The difficult components

Theorem (Graph filter transferability)

The difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \le \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

- As $(n, m) \to \infty$ most of the transferability error decreases with the square root of the graph sizes
- We can also afford smaller bandwidth limit c $\Rightarrow$ Transfer filters closer to $\lambda = 0$
- Sharper filter responses (larger Lipschitz constant L_2) $\Rightarrow$ Transfer more discriminative filters

Theorem (Graph filter transferability)

The difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \ \leq \ \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

- Graph signals and graphons with rapid variability make filter transference more difficult
- This is because of sampling approximation error $\Rightarrow$ Not fundamental
- The constants can be sharpened with modulo-permutation Lipschitz constants

Theorem (Graph filter transferability)

The difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \leq \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

- Filters that are more discriminative are more difficult to transfer

  $\Rightarrow$ True in the part of the bound related to easy components associated with eigenvalues $|\lambda| > c$

  $\Rightarrow$ True in the part of the bound related to difficult components associated with eigenvalues $|\lambda| \leq c$

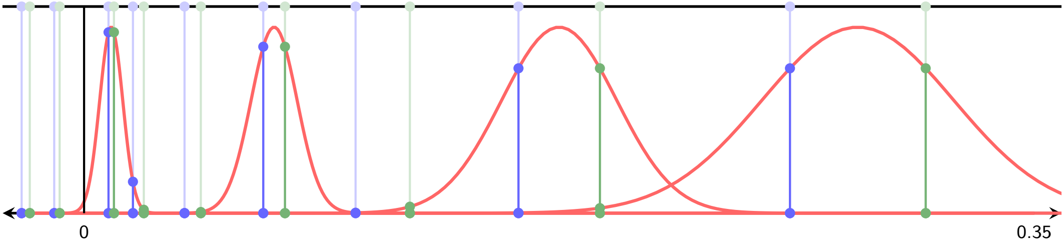
Theorem (Graph filter transferability)

The difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \leq \sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{2L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + L_2 c\|X\|$$

- Bound is parametric on the bandwidth c $\Rightarrow$ Different c result in different values for the bound
- Increase c-band cardinality or decrease c-eigenvalue margin $\Rightarrow$ More challenging transferability
- A property of the graphon $\Rightarrow$ Since eigenvalues converge B_{nc} and δ_{nc} converge

- If we fix n and m we observe emergence of a transferability vs discriminability non-tradeoff
- Discriminating around $\lambda = 0$ needs large Lipschitz constant L_2 $\Rightarrow$ Useless transferability bound
- To make transferability and discriminability compatible $\Rightarrow$ Graph Neural Networks



Transferability of GNNs

- We define graphon neural networks and discuss their interpretation as generative models for GNNs
- We show that graph neural networks inherit the transferability properties of graph filters

- Graph filters are transferable $\Rightarrow$ we can expect GNNs to inherit transferability from graph filters
- To analyze GNN transferability, we we first define Graphon Neural Networks (WNNs)
- The lth layer of a WNN composes a graphon convolution with parameters h and a nonlinearity σ

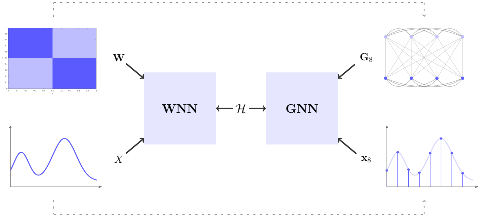
$$X_l^f = \sigma\left(\sum_{g=1}^{F_{l-1}} h_{kl}^{fg} T_W^{(k)} X_{l-1}^g\right)$$

L layers, $1 \leq f \leq F_l$ output features per layer. WNN input is $X_0 = X$. Output is $Y = X_L$

- Can be represented as $Y = \Phi(\mathcal{H}; W; X)$ with coefficients $\mathcal{H} = \{\mathrm{h}_{kl}^{fg}\}_{k,l,f,g}$. Just like the GNN

WNNs as Generative Models for GNNs

- As in the GNN map $\Phi(\mathcal{H}; S; x)$, in the WNN $\Phi(\mathcal{H}; W; X)$, the set $\mathcal{H}$ doesn't depend on the graphon
- Therefore, we can use WNNs to instantiate GNNs $\Rightarrow$ the WNN is a generative model for GNNs



- We will consider GNNs $\Phi(\mathcal{H}; S_n; x_n)$ instantiated from $\Phi(\mathcal{H}; W; X)$ on weighted graphs G_n

$$[S_n]_{ij} = W(u_i, u_j) \qquad [x_n]_i = X(u_i)$$

- Consider a graph signal $(\mathbf{S}_n, \mathbf{x}_n)$ sampled from the graphon signal (W, X)

- Given WNN coefficients $\mathcal{H}$ for L layers, width $F_l = F$ for $1 \leq l < L$, and $F_0 = F_L = 1$

  $\Rightarrow$ Run WNN with coefficients $\mathcal{H}$ on graphon W to process X $\Rightarrow Y = \Phi(\mathcal{H}; W, X)$

  $\Rightarrow$ Run GNN with coefficients $\mathcal{H}$ on graph S_n to process x_n $\Rightarrow \mathbf{y}_n = \Phi(\mathcal{H}; S_n, x_n)$

- Since one is a vector and the other a function we consider the induced graphon signal Y_n

(A1) The graphon W is L_1-Lipschitz $\Rightarrow$ For all arguments (u_1, v_1) and (u_2, v_2), it holds

$$\left| W(u_2, v_2) - W(u_1, v_1) \right| \leq L_1 \Big(\big| u_2 - u_1 \big| + \big| v_2 - v_1 \big| \Big)$$

(A2) The filter's response is L_2-Lipschitz and normalized $\Rightarrow$ For all λ_1, λ_2 and λ we have

$$\big| \tilde{h}(\lambda_2) - \tilde{h}(\lambda_1) \big| \leq L_2 \big| \lambda_2 - \lambda_1 \big| \quad \text{and} \quad \big| h(\lambda) \big| \leq 1$$

(A3) The graphon signal X is L_3-Lipschitz $\Rightarrow$ For all u_1 and u_2

$$\big| X(u_2) - X(u_1) \big| \leq L_3 \big| u_2 - u_1 \big|$$

(A4) The nonlinearities σ are normalized Lipschitz and $\sigma(0) = 0$ $\Rightarrow$ For all x and y

$$\big| \sigma(x) - \sigma(y) \big| \leq |x - y|$$

- We fix a bandwidth $c > 0$ to separate eigenvalues not close to $\lambda = 0$ and define

(D1) The c-band cardinality of G_n is the number of eigenvalues with absolute value larger than c

$$B_{nc} = \#\left\{ \lambda_{ni} \ : \ |\lambda_{ni}| > c \right\}$$

(D2) The c-eigenvalue margin of of graph G_n is the

$$\delta_{nc} = \min_{i,j\neq i}\left\{ |\lambda_{ni} - \lambda_j| \ : \ |\lambda_{ni}| \ > c \right\}$$

- Where λ_{ni} are eigenvalues of the shift operator S_n and λ_j are eigenvalues of graphon W

Theorem (GNN-WNN approximation)

Consider the graph signal (S_n, x_n) sampled from the graphon signal (W, X) along with the GNN output $y_n = \Phi(\mathcal{H}; S_n, x_n)$ and WNN output $Y = \Phi(\mathcal{H}; W, X)$. With Assumptions (A1)-(A4) and Definitions (D1)-(D2) the norm difference $\|Y_n - Y\|$ is bounded by

$$\|Y - Y_n\| \leq LF^{L-1}\sqrt{L_1}\left(L_2 + \pi\frac{B_{nc}}{\delta_{nc}}\right)\left(\frac{1}{\sqrt{n}}\right)\|X\| + \frac{L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}}\right) + LF^{L-1}L_2 c\|X\|$$

Proof: See course webpage https://gnn.seas.upenn.edu/lectures/lecture-10/ ∎

- The error incurred when using a GNN to approximate a WNN can be upper bounded
- Same comments as for graph and graphon filters apply. With additional dependence on L and F
- Distances between GNNs and WNN can be combined to calculate distance between GNNs
- GNNs $Y_n = \Phi(\mathcal{H}; W_n, x_n)$ and $Y_m = \Phi(\mathcal{H}; W_m, x_m)$ instantiated from WNN $Y = \Phi(\mathcal{H}; W, X)$

$$\|Y_n - Y_m\| = \|Y_n - Y + Y - Y_m\| \leq \|Y_n - Y\| + \|Y - Y_m\|$$

- The inequality follows from the triangle inequality. By which we have proved GNN transferability

- Consider graph signals $(\mathbf{S}_n, \mathbf{x}_n)$ and $(\mathbf{S}_m, \mathbf{x}_m)$ sampled from the graphon signal (W, X)

- Given GNN coefficients $\mathcal{H}$ for L layers, width $F_l = F$ for $1 \leq l < L$, and $F_0 = F_L = 1$

  $\Rightarrow$ Run GNN with coefficients $\mathcal{H}$ on graph S_n to process x_n $\Rightarrow y_n = \Phi(\mathcal{H}; S_n, x_n)$

  $\Rightarrow$ Run filter with coefficients $\mathcal{H}$ on graph S_m to process x_m $\Rightarrow y_m = \Phi(\mathcal{H}; S_m, x_n)$

- Since they have different number of components we compare induced graphon signals Y_n and Y_m

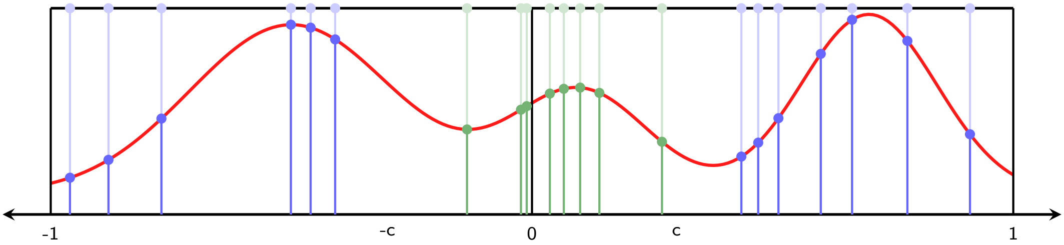
Theorem (GNN transferability)

Consider graph signals $(\mathbf{S}_n, \mathbf{x}_n)$ and $(\mathbf{S}_m, \mathbf{x}_m)$ sampled from graphon signal (W, X) along with GNN outputs $\mathbf{y}_n = \Phi(\mathcal{H}; \mathbf{S}_n, \mathbf{x}_n)$ and $\mathbf{y}_m = \Phi(\mathcal{H}; \mathbf{S}_m, \mathbf{x}_m)$. With Assumptions (A1)-(A4) and Definitions (D1)-(D2) the difference norm of the respective graphon induced signals is bounded by

$$\|Y_n - Y_m\| \leq LF^{L-1}\sqrt{L_1}\left(L_2 + \pi\frac{\max(B_{nc}, B_{mc})}{\min(\delta_{nc}, \delta_{mc})}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)\|X\| + \frac{L_3}{\sqrt{3}}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) + LF^{L-1}L_2 c\|X\|$$

- Same comments as in the case of graph filter transferability. With additional dependence on L, F

- The transferability-discriminability trade-off looks the same. But it is helped by the nonlinearities
- At each layer of the GNN, the nonlinearities σ scatter eigenvalues from $|\lambda| \leq c$ to $|\lambda| > c$



- Nonlinearities allows $\downarrow c$ and $\uparrow L_2$ $\Rightarrow$ increasing discriminability while retaining transferability
- For the same level of discriminability, GNNs are more transferable than graph filters