We will show that graph convolutional filters are equivariant to permutations.
Definition (Permutation matrix)

A square matrix $P$ is a permutation matrix if it has binary entries so that $P \in \{0, 1\}^{n \times n}$ and it further satisfies $P1 = 1$ and $P^T1 = 1$.

- The product $P^T x$ reorders the entries of the vector $x$.

- The product $P^T S P$ is a consistent reordering of the rows and columns of $S$. 
Definition (Permutation matrix)

A square matrix $P$ is a permutation matrix if it has binary entries so that $P \in \{0, 1\}^{n \times n}$ and it further satisfies $P1 = 1$ and $P^T1 = 1$.

- Since $P1 = P^T1 = 1$ with binary entries $\Rightarrow$ Exactly one nonzero entry per row and column of $P$

- Permutation matrices are unitary $\Rightarrow P^TP = I$. Matrix $P^T$ undoes the reordering of matrix $P$
If \((S, x)\) is a graph signal, \((P^TSP, P^Tx)\) is a relabeling of \((S, x)\). Same signal. Different names.

Processing should be label-independent \(\Rightarrow\) Permutation equivariance of graph filters and GNNs.
Graph filter $H(S)$ is a polynomial on shift operator $S$ with coefficients $h_k$. Outputs given by

$$H(S)x = \sum_{k=0}^{K-1} h_k S^k x$$

We consider running the same filter on $(S, x)$ and permuted (relabeled) $(\hat{S}, \hat{x}) = (P^T S P, P^T x)$

$$H(S)x = \sum_{k=0}^{K-1} h_k S^k x \quad \quad \quad \quad \quad H(\hat{S})\hat{x} = \sum_{k=0}^{K-1} h_k \hat{S}^k \hat{x}$$

Filter $H(S)x \Rightarrow$ Coefficients $h_k$. Input signal $x$. Instantiated on shift $S$

Filter $H(\hat{S})\hat{x} \Rightarrow$ Same Coefficients $h_k$. Permutated Input signal $\hat{x}$. Instantiated on permuted shift $\hat{S}$
Theorem (Permutation equivariance of graph filters)

Consider consistent permutations of the shift operator $\hat{\mathbf{S}} = \mathbf{P}^T \mathbf{S} \mathbf{P}$ and input signal $\hat{\mathbf{x}} = \mathbf{P}^T \mathbf{x}$. Then

$$H(\hat{\mathbf{S}})\hat{\mathbf{x}} = \mathbf{P}^T H(\mathbf{S})\mathbf{x}$$

Graph filters are equivariant to permutations $\Rightarrow$ Permute input and shift $\equiv$ Permute output
Proof: Write filter output in polynomial form. Use permutation definitions $\hat{S} = P^T S P$ and $\hat{x} = P^T x$

$$H(\hat{S})\hat{x} = \sum_{k=0}^{K-1} h_k \hat{S}^k \hat{x} = \sum_{k=0}^{K-1} h_k \left( P^T S P \right)^k P^T x$$

- In the powers $\left( P^T S P \right)^k$, $P$ and $P^T$ undo each other ($P^T P = I$) $\Rightarrow \left( P^T S P \right)^k = P^T \left( S \right)^k P$

- Substitute this into filter’s output expression. Cancel remaining $PP^T = I$ product. Factor $P^T$

$$H(\hat{S})\hat{x} = \sum_{k=0}^{K-1} h_k P^T S^k P P^T x = \sum_{k=0}^{K-1} h_k P^T S^k x = P^T \sum_{k=0}^{K-1} h_k S^k x = P^T H(S)x \blacksquare$$
We requested signal processing independent of labeling \( \Rightarrow \) Graph filters fulfill this request

\( \Rightarrow \) Permute input and shift \( \equiv \) Relabel input \( \Rightarrow \) Permute output \( \equiv \) Relabel output

Graph signal \( x \) Supported on \( S \)

Graph signal \( \hat{x} = P^T x \) supported on \( \hat{S} = P^T S \)
We requested signal processing independent of labeling ⇒ Graph filters fulfill this request

⇒ Permute input and shift ≡ Relabel input ⇒ Permute output ≡ Relabel output

Filter's output $H(S)x$ Supported on $S$

Filter's Output $H(\hat{S})\hat{x}$ supported on $\hat{S}$
We requested signal processing independent of labeling \( \Rightarrow \) Graph filters fulfill this request

\( \Rightarrow \) Permute input and shift \( \equiv \) Relabel input \( \Rightarrow \) Permute output \( \equiv \) Relabel output

Filter's output \( H(S)x \) Supported on \( S \)

Equivariance theorem \( \Rightarrow H(\hat{S})\hat{x} = P^T H(S)x \)
We will show that graph neural networks inherit the permutation equivariance of graph filters.
Graph Neural Networks and the Permutation of Graph Signals

- $L$ layers recursively process outputs of previous layers. GNN Output parametrized by tensor $\mathcal{H}$

$$x_{\ell} = \sigma \left[ \sum_{k=0}^{K-1} h_{\ell k} S^k x_{\ell-1} \right] = \sigma \left[ H_{\ell}(S) x_{\ell-1} \right]$$

- We consider running the same GNN on $(S, x)$ and permuted (relabeled) $(\hat{S}, \hat{x}) = (P^T S P, P^T x)$

$$\Phi(x; S, \mathcal{H}) \quad \Phi(\hat{x}; \hat{S}, \mathcal{H})$$

- GNN $\Phi(x; S, \mathcal{H}) \Rightarrow$ Tensor $\mathcal{H}$. Input signal $x$. Instantiated on shift $S$

- GNN $\Phi(\hat{x}; \hat{S}, \mathcal{H}) \Rightarrow$ Same Tensor $\mathcal{H}$. Permutated Input signal $\hat{x}$. Instantiated on permuted shift $\hat{S}$
Theorem (Permutation equivariance of graph neural networks)

Consider consistent permutations of the shift operator $\hat{S} = P^T S P$ and input signal $\hat{x} = P^T x$. Then

$$\Phi(\hat{x}; \hat{S}, \mathcal{H}) = P^T \Phi(x; S, \mathcal{H})$$

- GNNs equivariant to permutations $\Rightarrow$ Permute input and shift $\equiv$ Permute output
Proof: GNN Layer $\ell$ recursion on signal $x_{\ell-1}$ and shift $S \Rightarrow x_\ell = \sigma \left[ \sum_{k=0}^{K-1} h_{\ell k} S^k x_{\ell-1} \right] = \sigma \left[ H_\ell(S)x_{\ell-1} \right]$

GNN Layer $\ell$ recursion on signal $\hat{x}_{\ell-1}$ and shift $\hat{S} \Rightarrow \hat{x}_\ell = \sigma \left[ \sum_{k=0}^{K-1} h_{\ell k} \hat{S}^k \hat{x}_{\ell-1} \right] = \sigma \left[ H_\ell(\hat{S})\hat{x}_{\ell-1} \right]$

- **Assume** Layer $\ell$ inputs satisfy $\hat{x}_{\ell-1} = P^T x_{\ell-1}$. Filters are equivariant. Linearity is pointwise

$$\hat{x}_\ell = \sigma \left[ H_\ell(\hat{S})\hat{x}_{\ell-1} \right] = \sigma \left[ P^T H_\ell(S)x_{\ell-1} \right] = P^T \sigma \left[ H_\ell(S)x_{\ell-1} \right] = P^T x_\ell$$

- **This in an induction step** At Layer 1 we have $\hat{x} = P^T x$ by hypothesis. Induction is complete.
GNNs, same as graph filters, perform label-independent processing. The nonlinearity is pointwise.

$$\Rightarrow$$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

Graph signal $x$ Supported on $S$

Graph signal $\hat{x} = P^T x$ supported on $\hat{S} = P^T S$
GNNs, same as graph filters, perform label-independent processing. The nonlinearity is pointwise:

- Permute input and shift \( \equiv \) Relabel input
- Permute output \( \equiv \) Relabel output

GNN output \( \Phi(x; S, \mathcal{H}) \) supported on \( S \)

\[
\begin{align*}
\text{GNN output } & \Phi(x; S, \mathcal{H}) \text{ supported on } S \\
\text{GNN } & \Phi(\hat{x}; \hat{S}, \mathcal{H}) = P^T \Phi(x; S, \mathcal{H}) \text{ on } \hat{S} = P^T S
\end{align*}
\]
Equivariance to Permutations and Signal Symmetries

- Equivariance to permutations allows GNNs to exploit symmetries of graphs and graph signals

- By symmetry we mean that the graph can be permuted onto itself \( S = P^T S P \)

- Equivariance theorem implies \( \Phi(P^T x; S, \mathcal{H}) = \Phi(P^T x; P^T S P, \mathcal{H}) = P^T \Phi(x; S, \mathcal{H}) \)

From observing \( x \) supported on \( S \)

Learn to process \( P^T x \) supported on \( S = P^T S P \)
Symmetry is Rare but Quasi-Symmetry is Common

- Graph not symmetric but close to symmetric $\Rightarrow$ perturbed version of a permutation of itself

- We will show conditions for stability to deformations $\Rightarrow$ Approximate (close to) equivariance
Definition (Operator Distance Modulo Permutation)

For operators $\Psi$ and $\hat{\Psi}$, the operator distance modulo permutation is defined as

$$
\|\Psi - \hat{\Psi}\|_P = \min_{P \in \mathcal{P}} \max_{x : \|x\| = 1} \|P^T \Psi(x) - \hat{\Psi}(P^T x)\|
$$

where $\mathcal{P}$ is the set of $n \times n$ permutation matrices and where $\| \cdot \|$ stands for the $\ell_2$-norm.

- Equivariance to permutations of graph filters $\Rightarrow$ If $\|\hat{S} - S\|_P = 0$. Then $\|H(\hat{S}) - H(S)\|_P = 0$
- Equivariance to permutations GNNs $\Rightarrow$ If $\|\hat{S} - S\|_P = 0$. Then $\|\Phi(\cdot; \hat{S}, \mathcal{H}) - \Phi(\cdot; S, \mathcal{H})\|_P = 0$
- When distance $\|\hat{S} - S\|_P$ is small? (not zero) $\Rightarrow$ Stability properties of graph filters and GNNs
Lipschitz and Integral Lipschitz Filters

- Classes of filters to study discriminability of GNNs \(\Rightarrow\) Lipschitz and integral Lipschitz graph filters
Graph filters are polynomials on shift operators $S$ with given coefficients $h_k \Rightarrow H(S) = \sum_{k=0}^{\infty} h_k S^k$.

Filter’s frequency response is the same polynomial with scalar variable $\lambda \Rightarrow \tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k$.

Frequency response determined by filter coefficients $h_k$. Independent of particular given graph.
Definition (Lipschitz Filter)

Given a graph filter with coefficients $h = \{h_k\}_{k=1}^{\infty}$, and graph frequency response

$$\tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k,$$

we say that the filter is Lipschitz if there exists a constant $C > 0$ such that for $\lambda_1$ and $\lambda_2$

$$|\tilde{h}(\lambda_2) - \tilde{h}(\lambda_1)| \leq C |\lambda_2 - \lambda_1|.$$

- Change in values of frequency response is at most linear with rate $C \implies$ Derivative $\tilde{h}'(\lambda) \leq C$
Discriminability of Lipschitz Filters

- Frequency response $\tilde{h}(\lambda)$ of Lipschitz filter is Lipschitz continuous $\Rightarrow$ Maximum slope is $\tilde{h}'(\lambda) \leq C$

- Lipschitz constant determines discriminability $\Rightarrow$ Small / Large $C \equiv$ Low / High discriminability
Discriminability of Lipschitz Filters

- Frequency response $\tilde{h}(\lambda)$ of Lipschitz filter is Lipschitz continuous $\Rightarrow$ Maximum slope is $\tilde{h}'(\lambda) \leq C$

- Lipschitz constant determines discriminability $\Rightarrow$ Small / Large $C \equiv$ Low / High discriminability
A Lipschitz frame with constant $C$ is made up of Lipschitz filters with constant $C$.

Larger $C$ allows for sharper filters, that can discriminate more signals. Tighter packing.

The discriminability of the frame is (or can be) the same at all frequencies.
Definition (Integral Lipschitz Filter)

Consider graph filter with coefficients $h_k$ and graph frequency response $\tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k$. The filter is said integral Lipschitz if there exists constant $C > 0$ such that for all $\lambda_1$ and $\lambda_2$,

$$|\tilde{h}(\lambda_2) - \tilde{h}(\lambda_1)| \leq C \frac{|\lambda_2 - \lambda_1|}{|\lambda_1 + \lambda_2|/2}.$$

- Lipschitz with a constant that is inversely proportional to the interval’s midpoint $\Rightarrow 2C/|\lambda_1 + \lambda_2|$.
- Letting $\lambda_2 \to \lambda_1$ we get that $\lambda \tilde{h}'(\lambda) \leq C \Rightarrow$ The filter can’t change for large $\lambda$. 

Discriminability of Integral Lipschitz Filters

- At medium frequencies, integral Lipschitz filters are akin to Lipschitz filters. Roughly speaking

- At low frequencies integral Lipschitz filters can be arbitrarily thin ⇒ arbitrary discriminability

- At high frequencies integral Lipschitz filters have to be flat ⇒ They lose discriminability

\[ \tilde{h}(\lambda) \]
As Lipschitz frames, integral Lipschitz frames are more discriminative for larger $C$. Tighter packing

Except that around $\lambda = 0$, filters can be thin no matter $C \Rightarrow$ High discriminability

But for large $\lambda$ filters have to be wide no matter $C \Rightarrow$ No discriminability
Stability of Graph Filters to Scaling

- Scaling of shift operators is a perturbation form that illustrates proof techniques and insights

- We show that graph filters are stable with respect to scaling
Graphs are subject to estimation error and changes ⇒ Running filters on similar graphs

- We scale edges by $(1 + \epsilon)$. Scaling deformation of the shift operator ⇒ $\hat{S} = (1 + \epsilon)S$

- Deformation model is reasonable ⇒ Edges change proportional to their values

- Also unrealistic ⇒ All of the edges change by the same proportion

⇒ Illuminating for discussions. Stability proof contains essential arguments of more generic proof.
Theorem (Integral Lipschitz Graph Filters are Stable to Scaling)

Given graph shift operators $S$ and $\hat{S} = (1 + \epsilon)S$ and an integral Lipschitz filter with constant $C$.

The operator norm difference between filters $H(S)$ and $H(\hat{S})$ is bounded as

$$\| H(\hat{S}) - H(S) \| \leq C \epsilon + O(\epsilon^2).$$

Stability to scaling is possible. ⇒ But it requires a restriction to the use of integral Lipschitz filters.
The key arguments of the proof are in the GFT domain. We provide two preliminary spectral facts.

**Fact 1:**

If $\tilde{x} = V^H x$ is the GFT of $x$ we can write $x = \sum_{i=1}^{n} \tilde{x}_i v_i$, where $v_i$ are the eigenvectors of $S$.

**Proof:** Write $x$ using the inverse GFT $\Rightarrow x = V \tilde{x} = [v_1, \ldots, v_n] \times \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = \tilde{x}_1 v_1 + \ldots + \tilde{x}_n v_n$.
The key arguments of the proof are in the GFT domain. We provide two preliminary spectral facts.

**Fact 2:**

The frequency response derivative is \( \tilde{h}'(\lambda) = \sum_{k=0}^{\infty} k h_k \lambda^{k-1} \). Consequently \( \lambda \tilde{h}'(\lambda) = \sum_{k=0}^{\infty} k h_k \lambda^k \).

**Proof:** Frequency response is the series \( \tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k \). The summands' derivatives are \( k h_k \lambda^{k-1} \).
Proof Step 1: From Shift Perturbations to Filter Perturbations

**Proof:** Filter difference given by graph filter definition $H(S) = \sum_{k=0}^{\infty} h_k S^k$. Further write $\hat{S} = (1 + \epsilon) S$

$$H(\hat{S}) - H(S) = \sum_{k=0}^{\infty} h_k \hat{S}^k - \sum_{k=0}^{\infty} h_k S^k = \sum_{k=0}^{\infty} h_k \left[ \left( (1 + \epsilon) S \right)^k - \hat{S}^k \right]$$

- Expand binomial $\left( (1 + \epsilon) S \right)^k$ to first order only. Group all high order terms in matrix $O_k(\epsilon)$

$$\left( (1 + \epsilon) S \right)^k = (1 + k \epsilon) S^k + O_k(\epsilon)$$

- Upon substitution the terms $S^k$ cancel out $\Rightarrow H(\hat{S}) - H(S) = \sum_{k=0}^{\infty} h_k k \epsilon S^k + O(\epsilon)$

- The matrix $O(\epsilon)$ satisfies $0 < \lim_{\epsilon \to 0} \frac{\|O(\epsilon)\|}{\epsilon^2} < \infty$ because filter is analytic. Term is of order $O(\epsilon^2)$
Proof Step 2: Evaluating and Reducing the Operator Norm

- Have reduced the filter difference to \( H(\hat{S}) - H(S) = \sum_{k=0}^{\infty} h_k \epsilon S^k + O(\epsilon) = \Delta(S) + O(\epsilon) \)

- Where we have defined the filter variation \( \Delta(S) = \epsilon \sum_{k=0}^{\infty} k h_k S^k \) to simplify notation

- Triangle inequality \( \Rightarrow \|H(\hat{S}) - H(S)\| \leq \|\Delta(S)\| + O(\epsilon) = \|\Delta(S)\| + O(\epsilon^2) \)

- Since \( \|\Delta(S)\| = \max_{\|x\|=1} \|\Delta(S)x\| \) theorem follows if we prove \( \|\Delta(S)x\| \leq C\epsilon \) for all \( x \) with \( \|x\| = 1 \)
Proof Step 3: Shifting to the GFT Domain

- Product of filter variation with unit norm $x$. Write the iGFT of the input $x = \sum_{i=1}^{n} \tilde{x}_i v_i$ ($Sv_i = \lambda_i v_i$)

$$\Delta(S)x = \epsilon \sum_{k=0}^{\infty} k h_k S^k x = \epsilon \sum_{k=0}^{\infty} k h_k S^k \left[ \sum_{i=1}^{n} \tilde{x}_i v_i \right] = \sum_{i=1}^{n} \tilde{x}_i \epsilon \sum_{k=0}^{\infty} k h_k S^k v_i$$

- Since the $v_i$ are eigenvectors of $S \Rightarrow S^k v_i = \lambda_i^k v_i$. With $\lambda_i$ the associated eigenvalue

$$\Delta(S)x = \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k h_k S^k v_i = \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k h_k \lambda_i^k v_i = \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k h_k \lambda_i^k v_i$$

- The derivative of the filter’s response appears $\Rightarrow \sum_{k=0}^{\infty} k h_k \lambda_i^k = \lambda_i \tilde{h}'(\lambda_i)$
Proof Step 4: Leveraging the Integral Lipschitz Condition

- End up with remarkably simple equation $\Rightarrow \Delta(S)x = \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k h_k \lambda_i^k v_i = \epsilon \sum_{i=1}^{n} \tilde{x}_i \left( \lambda_i \tilde{h}'(\lambda_i) \right) v_i$

- Which involves the quantity we bound with the integral Lipschitz condition $\Rightarrow |\lambda_i \tilde{h}'(\lambda_i)| \leq C$

- Compute energy. Use integral Lipschitz bound. Recall that signal has unit energy, $\|x\|^2 = \|\tilde{x}\|^2 = 1$

$$\|\Delta(S)x\|^2 = \epsilon^2 \sum_{i=1}^{n} \tilde{x}_i^2 \left( \lambda_i \tilde{h}'(\lambda_i) \right)^2 \leq \epsilon^2 \sum_{i=1}^{n} \tilde{x}_i^2 C^2 = (C\epsilon)^2$$

- Take square root
Integral Lipschitz filters are necessary for stability to deformations of the supporting graph.

This is not an artifact of the analysis. The result is tight. The term $\sum_{k=0}^{\infty} k h_k \lambda_i^k = \lambda_i h'(\lambda_i)$ appears.
One would expect a **stability vs discriminability tradeoff**. But in a sense, we get a non-tradeoff.

Integral Lipschitz filters **have to be flat at high frequencies**. \(\Rightarrow\) They can’t discriminate

It is **impossible to separate signals with high frequency features and be stable** to deformations.
One would expect a **stability vs discriminability tradeoff**. But in a sense, we get a non-tradeoff.

- **Integral Lipschitz filters** have to be flat at high frequencies. ⇒ They can’t discriminate

- It is **impossible to separate** signals with high frequency features and be **stable** to deformations
Stability of Graph Neural Networks to Scaling

- Scaling of shift operators is a perturbation form that illustrates proof techniques and insights

- We show that Graph Neural Networks are stable with respect to scaling
Normalizations

▶ To avoid appearance of meaningless constants we normalize the filters and the nonlinearity.

▶ At each layer of the GNN, the filters have unit operator norm \( \| H_\ell(S) \| = 1 \)  
  \( \Rightarrow \) Easy to achieve with scaling \( \Rightarrow \) Equivalent to \( \max_\lambda \tilde{h}_\ell(\lambda) = 1 \)

▶ The nonlinearity \( \sigma \) is Lipschitz and normalized so that \( \| \sigma(x_2) - \sigma(x_1) \| \leq \| x_2 - x_1 \| \)  
  \( \Rightarrow \) Easy to achieve with scaling. True of ReLU, hyperbolic tangent, and absolute value

▶ Joining both assumptions \( \Rightarrow \) If input energy is \( \| x \| \leq 1 \), all layer outputs have energy \( \| x_\ell \| \leq 1 \)
Theorem (Integral Lipschitz GNNs are Stable to Scaling)

Given shift operators $S$ and $\hat{S} = (1 + \epsilon)S$ and a GNN operator $\Phi(\cdot; S, \mathcal{H})$ with $L$ single-feature layers. The filters at each layer have unit operator norms and are integral Lipschitz with constant $C$. The nonlinearity $\sigma$ is normalized Lipschitz. Then

$$\| \Phi(\cdot; S, \mathcal{H}) - \Phi(\cdot; \hat{S}, \mathcal{H}) \| \leq C L \epsilon + O(\epsilon^2).$$

▶ GNNs inherit the stability of graph filters. It’s the same bound. Propagated through $L$ layers
Proof Step 1: Eliminating the Pointwise Nonlinearity

Proof: The theorem is true because the nonlinearity is pointwise. It is unaware of the graph.

- Formally \( \Rightarrow \) Let \( x_\ell \) be the Layer \( \ell \) output of GNN \( \Phi(x; S, H) \)
  \[ \Rightarrow \text{Let } \hat{x}_\ell \text{ be the Layer } \ell \text{ output of GNN } \Phi(\hat{x}; \hat{S}, H) \]

- Layer \( \ell \) is a perceptron with filter \( H_\ell \) \( \Rightarrow \) \( \| x_\ell - \hat{x}_\ell \| = \| \sigma [H_\ell(S)x_{\ell-1}] - \sigma [H_\ell(\hat{S})\hat{x}_{\ell-1}] \| \)

- Nonlinearity is normalized Lipschitz \( \Rightarrow \) \( \| x_\ell - \hat{x}_\ell \| \leq \| H_\ell(S)x_{\ell-1} - H_\ell(\hat{S})\hat{x}_{\ell-1} \| \)

- This is the critical step of the proof. The rest of the proof is just algebra.
Proof Step 2: Implementing Norm Manipulations

- In last bound, add and subtract $H_\ell(\hat{S})x_{\ell-1}$. Triangle inequality. Submultiplicative property of norms

$$\left\| x_\ell - \hat{x}_\ell \right\| \leq \left\| H_\ell(S)x_{\ell-1} - H_\ell(\hat{S})\hat{x}_{\ell-1} + H_\ell(\hat{S})x_{\ell-1} - H_\ell(\hat{S})x_{\ell-1} \right\|$$

$$\leq \left\| H_\ell(S) - H_\ell(\hat{S}) \right\| \times \left\| x_{\ell-1} \right\| + \left\| H_\ell(\hat{S}) \right\| \times \left\| x_{\ell-1} - \hat{x}_{\ell-1} \right\|$$

- Since filters are normalized $\Rightarrow$ Filter norm $\left\| H_\ell(\hat{S}) \right\| = 1$. Signal norm $\Rightarrow \left\| x_{\ell-1} \right\| \leq 1$

- The theorem on stability of filters to scaling holds $\Rightarrow \left\| H_\ell(S) - H_\ell(\hat{S}) \right\| \leq \epsilon C + O(\epsilon^2)$

- Put all bounds together $\Rightarrow \left\| x_\ell - \hat{x}_\ell \right\| \leq \epsilon C \times 1 + 1 \times \left\| x_{\ell-1} - \hat{x}_{\ell-1} \right\| + O(\epsilon^2)$

- Apply recursively from Layer $L$ back to Layer 1. The $L$ factor appears
GNNs have the same stability properties of graph filters. They need integral Lipschitz filters.

Integral Lipschitz filters have to be flat at high frequencies. ⇒ They can’t discriminate

It is impossible to separate signals with high frequency features and be stable to deformations
GNNs have the same stability properties of graph filters. They need integral Lipschitz filters.

On the flip side, integral Lipschitz filter can be very sharp at low frequencies.

We can be very discriminative at low frequencies. And at the same very stable to deformations.
GNNs use low-pass nonlinearities to demodulate high frequencies into low frequencies.

Where they can be discriminated sharply with a stable filter at the next layer.

Thus, they can be stable and discriminative. Something that linear graph filters can’t be.