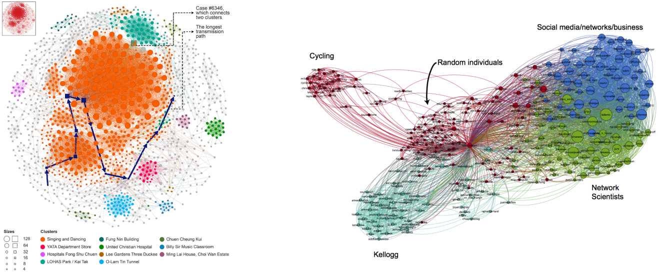
On the Stability of Low Pass Graph Filter with A Large Number of Edge Rewires

HS. Nguyen[1], Y. He[2], HT. Wai[2]

[1]Department of Computer Science & Engineering, CUHK
[2]Department of Systems Engineering & Engineering Management, CUHK

May. 2022

Graph data



(a) COVID-19 cases (link to source)

(b) User communities on Twitter (link to source)

Figure: Examples of network data

- Complex data can be represented as graphs of relationship between entities, e.g., epidemics, social networks, power networks,...
- Graph Signal Processing develops rigorous tools for processing data on graphs, e.g. graph filters.

Motivation

- Recently, graph neural networks (GNNs) have shown great empirical performance in network inference on various benchmark dataset
- Graph filters are used as building blocks for many GNNs; see [Defferard et al., 2016], [Kipf et al., 2017], [Gama et al., 2020]
- Specifically, GNN is a cascade of *layers*; each *layer* applies a bank of *graph filters*, followed by a pointwise *nonlinearity*

Motivation

- Since retraining a GNN is expensive, it is desirable to utilize the pre-trained model.
- This motivates the stability analysis of GNNs subject to perturbations of graph edges, graph nodes, graph signals,...
  - How much the output would change?
  - What kinds of perturbations greatly affect the output?
  - etc.
- As stability of graph filters characterizes stability of GNNs, our goal is to investigate the former.

Scope: We focus on stability analysis of graph filters subjecting to perturbations in graph topology.

Graph

- Consider undirected and connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $|\mathcal{V}| = n$ with adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and degree matix $\boldsymbol{D} = Diag(\boldsymbol{A}\mathbf{1})$.
- A graph shift operator (GSO) is a symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ such that $[\boldsymbol{S}]_{ij} \neq 0$ iff $i = j$ or $(i, j) \in \mathcal{E}$
  - Admit an eigendecomposition $\boldsymbol{S} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^T$.
  - Denote $\boldsymbol{\Lambda}_k$ and $\boldsymbol{V}_k$ as the matrices of smallest-k eigenvalues and eigenvectors of $\boldsymbol{S}$.
- In this work, we consider two common GSOs:
  - Unnormalized Laplacian $\boldsymbol{L}_U := \boldsymbol{D} - \boldsymbol{A}$.
  - Normalized Laplacian $\boldsymbol{L}_{\texttt{norm}} := \boldsymbol{D}^{-1/2} \boldsymbol{L}_U \boldsymbol{D}^{-1/2}$.

Graph Signal and Filter

- A graph signal on $\mathcal{G}$ is a scalar function $x : \mathcal{V} \to \mathbb{R}$, often represented by a vector $\boldsymbol{x} \in \mathbb{R}^n$.
- A graph filter $\mathcal{H}(\boldsymbol{S}) \in \mathbb{R}^{n \times n}$ maps the input signal $\boldsymbol{x} \in \mathbb{R}^{n \times n}$ to the output signal $\boldsymbol{y} = \mathcal{H}(\boldsymbol{S})\boldsymbol{x}$.
- Consider linear graph filter:

$$\mathcal{H}(\boldsymbol{S}) = \sum_{t=0}^{T-1} h_t \boldsymbol{S}^t.$$

- By setting $h(\lambda) = \sum_{t=0}^{T-1} h_t \lambda^t$, we can rewrite the filter as:

$$\mathcal{H}(\boldsymbol{S}) = \boldsymbol{V} h(\boldsymbol{\Lambda}) \boldsymbol{V}^T$$

where $h(\boldsymbol{\Lambda}) = Diag(h(\lambda_1), ..., h(\lambda_n))$.

Low pass Graph Filter

- Consider low pass graph filter, which retains low frequency components of input while suppressing high frequency components.
- The graph filter $\mathcal{H}(\cdot)$ is $(\underline{\lambda}, \overline{\lambda})$-low pass if

$$\eta = \frac{\max_{\lambda \in [\overline{\lambda}, \infty)} |h(\lambda)|}{\min_{\lambda \in [0, \underline{\lambda}]} |h(\lambda)|} < 1.$$

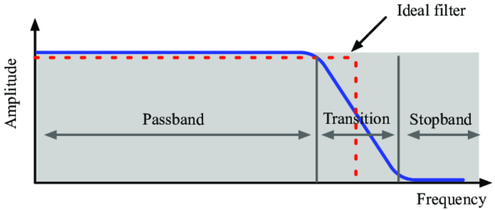


Figure: Visualization of low pass graph filter (`link to source`)

Measure of stability

We adopt the following measure of stability used by previous works such as [Gama et al., 2020] and [Kenlay et al., 2021]:

Graph filter distance

Consider a graph topology perturbation, e.g., via edge rewiring, of $\mathcal{G}$ (with GSO $\boldsymbol{S}$) into $\hat{\mathcal{G}}$ (with new GSO $\hat{\boldsymbol{S}}$). The graph filter distance between two filters $\mathcal{H}(\boldsymbol{S})$ and $\mathcal{H}(\hat{\boldsymbol{S}})$ is

$$\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}}) := ||\mathcal{H}(\boldsymbol{S}) - \mathcal{H}(\hat{\boldsymbol{S}})||_2.$$

- We ignore node permutations for sake of simplicity.
- Our goal is to upperbound the graph filter distance by quantities that are related to macroscopic structure (e.g., community) of the graph.

Prior works

- Existing works focus on the case of relatively small and arbitrary perturbations:
  - Given that $||\boldsymbol{S} - \hat{\boldsymbol{S}}||_2 \leq \varepsilon$, most previous works showed that graph filters are linearly stable, i.e. $\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}}) = \mathcal{O}(\varepsilon)$.
  - Notably, [Kenlay et al., 2021] gave a linear stability bound when the graph topology is subjected to a type of structural perturbation (double-edge rewires).
- On the other hand, [Keriven et al., 2020] gave a stability bound for very large graphs where ε can be arbitrarily large, *but only for normalized Lapacian matrix*.

Our contribution

- We give a stability bound on graph filter distance which depends on *community structure* and *low pass filtering capability*:
  - independent of the size of perturbations.
  - works for both unnormalized and normalized Laplacian matrix.
- We show that when graph filter is sufficiently low pass,

$$\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}}) = O(\eta) + o(1)$$

  where η is a constant depending on choice of filter.
- Numerical experiments support our findings.

Some possible approaches

There are several approaches to analyze graph filter distance $\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})$:

- Algebraic: $||\mathcal{H}(\boldsymbol{S}) - \mathcal{H}(\hat{\boldsymbol{S}})||_2 = ||\sum_{t=0}^{T-1} h_t(\boldsymbol{S}^t - \hat{\boldsymbol{S}}^t)||_2$.
  - Used by [Levie et al., 2019], [Gama et al., 2020], [Kenlay et al., 2021].
  - The bound would depend on the size of perturbations, i.e. $||\boldsymbol{S} - \hat{\boldsymbol{S}}||_2$.
- **Spectral**: $||\mathcal{H}(\boldsymbol{S}) - \mathcal{H}(\hat{\boldsymbol{S}})||_2 = ||\boldsymbol{V} h(\Lambda) \boldsymbol{V}^T - \hat{\boldsymbol{V}} h(\hat{\Lambda}) \hat{\boldsymbol{V}}^T||_2$.
  - Enable us to break free from dependence on $||\boldsymbol{S} - \hat{\boldsymbol{S}}||_2$.
  - Principal eigenvalues/eigenvectors of $\boldsymbol{S}$ and $\hat{\boldsymbol{S}}$ contain information on community structures.

Bounding graph filter distance

Assumptions

- (H1) There exists a constant $\mathbb{H}_{max}$ such that $\sup_{\lambda\in[0,\bar{\lambda}]} |h(\lambda)| \leq \mathbb{H}_{max}$.
- (H2) There exists a constant $L_{\mathbb{H}}$ such that $|h(\lambda) - h(\lambda')| \leq L_{\mathbb{H}}|\lambda - \lambda'| \,, \forall \lambda, \lambda' \in [0, \bar{\lambda}]$.

Theorem 1

Let $\mathcal{H}(\cdot)$ be a $(\underline{\lambda}, \overline{\lambda})$-low pass filter with ratio η and the GSOs $\boldsymbol{S}$, $\hat{\boldsymbol{S}}$ satisfy $\lambda_k, \hat{\lambda}_k \leq \underline{\lambda} < \overline{\lambda} \leq \lambda_{k+1}, \hat{\lambda}_{k+1}$. Then,

$$\mathcal{D}_{\mathbb{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}}) \leq 2\mathbb{H}_{max}\eta + L_{\mathbb{H}}||\boldsymbol{\Lambda}_k - \hat{\boldsymbol{\Lambda}}_k||_2 + 2\mathbb{H}_{max}||\boldsymbol{V}_k - \hat{\boldsymbol{V}}_k||_2$$

Interpretation: The bound on $\mathcal{D}_{\mathbb{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})$ depends on

- η, which is dependent on frequency response
- $||\boldsymbol{\Lambda}_k - \hat{\boldsymbol{\Lambda}}_k||_2$ and $||\boldsymbol{V}_k - \hat{\boldsymbol{V}}_k||_2$, capturing similarity in community structure

Application to PPM

- For simplicity, consider a planted partition model $PPM(n, k, a_n, b_n)$ where n nodes are equally divided into k clusters:
  - probability of edge between nodes from same cluster is $a_n + b_n$
  - probability of edge between nodes from different clusters is b_n
- Recall from Theorem 1 that the graph filter distance is bounded by $||\boldsymbol{V}_k - \hat{\boldsymbol{V}}_k||_2$ and $||\boldsymbol{\Lambda}_k - \hat{\boldsymbol{\Lambda}}_k||_2$, which are known to contain information about community structure

Intuition: if original $\mathcal{G}$ and perturbed $\hat{\mathcal{G}}$ have the same community structure (e.g. having the same PPM), then $\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})$ is small

Perturbation scheme

For example of a perturbation scheme that does not alter the generating PPM of the original graph, consider the following:

Edge rewiring scheme: $\mathcal{G} \to \hat{\mathcal{G}}$

1. It is given that $\mathcal{G} \sim PPM(n, k, a_n, b_n)$.
2. For each inter/intra-cluster block (i, j):
   - Delete a portion of $p_{re} \in [0, 1]$ edges uniformly.
   - Then add edges to the node pairs without edges with probability $[b_{ij}^{-1} - (1 - p_{re})]^{-1} p_{re}$ independently.

The resultant $\hat{\mathcal{G}}$ is considered to have the same underlying PPM as $\mathcal{G}$.

Stability with Unnormalized Laplacian as GSO

Corollary 1 (adapted from [Deng et al., 2021])

Let $\alpha, \beta \in \mathbb{R}^+$, consider $\mathcal{G}, \hat{\mathcal{G}} \sim PPM(n, 2, \alpha \log n/n, \beta \log n/n)$. Suppose that $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$, then with probability at least $1 - o(1)$,

$$||\boldsymbol{V}_2 - \hat{\boldsymbol{V}}_2||_2 = o(1);\ ||\boldsymbol{\Lambda}_2 - \hat{\boldsymbol{\Lambda}}_2||_2 = \mathcal{O}\left(\frac{\log n}{n}\right)$$

Stability bound with Unnormalized Laplacian as GSO

Under the conditions of Theorem 1 and Corollary 1, with number of blocks $k = 2$

$$\mathbb{D}_{\mathcal{H}}(\boldsymbol{L}_U, \hat{\boldsymbol{L}}_U) \leq 2\eta \mathbb{H}_{max} + \mathbb{H}_{max} o(1) + L_{\mathbb{H}} \mathcal{O}\left(\frac{\log n}{n}\right)$$

with high probability. $\mathbb{D}_{\mathcal{H}}(\boldsymbol{L}_U, \hat{\boldsymbol{L}}_U)$ is small when $\eta \ll 1$ and $n \to \infty$.

- The case of $k > 2$ is left for future works.

Stability with Normalized Laplacian as GSO

Corollary 2 (adapted from [Deng et al., 2021])

Let $\alpha, \beta \in \mathbb{R}^+$, consider $\mathcal{G}, \hat{\mathcal{G}} \sim PPM(n, k, \alpha \log n/n, \beta \log n/n)$. Suppose that $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$, then with probability at least $1 - o(1)$,

$$||\boldsymbol{L}_{norm} - \hat{\boldsymbol{L}}_{norm}||_2 = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$$

Stability bound with Normalized Laplacian as GSO

Under the conditions of Theorem 1 and Corollary 2,

$$\mathbb{D}_{\mathcal{H}}(\boldsymbol{L}_{norm}, \hat{\boldsymbol{L}}_{norm}) \leq 2\eta\mathbb{H}_{max} + (\mathbb{H}_{max} + L_{\mathbb{H}})\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$$

with high probability. $\mathbb{D}_{\mathcal{H}}(\boldsymbol{L}_{norm}, \hat{\boldsymbol{L}}_{norm})$ is small when $\eta \ll 1$ and $n \to \infty$.

Synthetic experiment: Objective & Setup

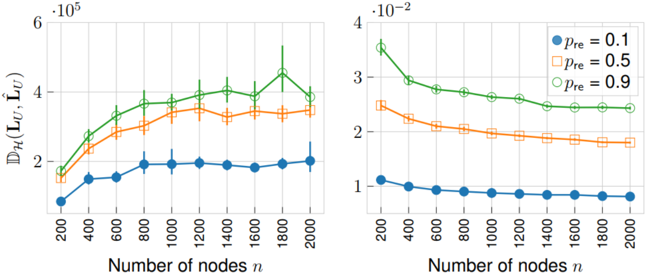
- Compare $\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})$ against number of nodes n under the following configurations of graph filters

	Unnormalized Laplacian[1]	Normalized Laplacian
$\mathcal{H}_{LP}(\cdot)$	$exp(-(1/\log n)\boldsymbol{L}_U)$	$exp(-\boldsymbol{L}_{norm})$
$\mathcal{H}_{HP}(\cdot)$	$exp((1/\log n)\boldsymbol{L}_U)$	$exp(\boldsymbol{L}_{norm})$

- We compare high-pass filters $\mathcal{H}_{\rm HP}$ with low-pass filters $\mathcal{H}_{\rm LP}$ to demonstrate the need of low-pass property for graph filter stability
- Generate original graph $\mathcal{G}$ from a pre-defined PPM, and generated perturbed graph $\hat{\mathcal{G}}$ using defined edge rewiring scheme
- For every n, perform Monte-Carlo simulations with 100 trials to estimate $\mathbb{E}[\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})]$ for each of the graph filter configurations

[1] For $\boldsymbol{L}_U$, $\pm(1/\log n)$ ensures low-pass property is insensitive to spectrum growth

Synthetic experiment: Result



(a) Unnormalized Laplacian $\mathbf{L}_U$ as GSO.

Figure: Comparing high pass (left) vs. low pass (right) filter for $\boldsymbol{L}_U$

Real data experiment

- Original graph $\mathcal{G}$ is `email-Eu-core` network with 1005 nodes, 25571 edges and 42 communities.
- We perform a simplified edge rewiring process on $\mathcal{G}$ to obtain the perturbed graph $\hat{\mathcal{G}}$.
- We compare $\mathbb{D}_{\mathcal{H}}(\boldsymbol{S}, \hat{\boldsymbol{S}})$ between high pass filter and low pass filter against rewiring ratios $p_{\texttt{re}}$.

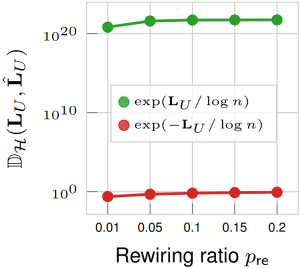


Figure: Comparing high pass filter (green) vs. low pass filter (red) for $\boldsymbol{L}_U$

Conclusion

- We study stability of low pass graph filters subjecting to a large number of edge rewires
- We propose a stability bound w.r.t. frequency response, instead of the size of perturbations
- Our bound shows that if the underlying community structure is unchanged, the low pass graph filters are stable
- Numerical experiments support our hypotheses

References

- T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
- M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," Advances in neural information processing systems, vol. 29, pp. 3844–3852, 2016.
- F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," IEEE Transactions on Signal Processing, vol. 68, pp. 5680–5695, 2020.
- H. Kenlay, D. Thanou, and X. Dong, "On the stability of polynomial spectral graph filters," in ICASSP, 2020.
- H. Kenlay, D. Thanou, and X. Dong, "Interpretable stability bounds for spectral graph filters," in ICML, 2021.

References (cont.)

- R. Levie, E. Isufi, and G. Kutyniok, "On the transferability of spectral graph filters," in 2019 13th International conference on Sampling Theory and Applications (SampTA). IEEE, 2019, pp. 1–5.
- N. Keriven, A. Bietti, and S. Vaiter, "Convergence and stability of graph convolutional networks on large random graphs," in Neural Information Processing Systems, 2020.
- K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," The Annals of Statistics, vol. 39, no. 4, 8 2011.
- J. Lei and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," The Annals of Statistics, vol. 43, no. 1, 2 2015.
- S. Deng, S. Ling, and T. Strohmer, "Strong consistency, graph laplacians, and the stochastic block model," Journal of Machine Learning Research, vol. 22, no. 117, pp. 1–44, 2021.