Learning from Heterophilic Graphs: A Spectral Theory Perspective on the Impact of Self-Loops and Parallel Edges

Anonymous authors
Paper under double-blind review

Abstract

Graph heterophily poses a formidable challenge to the performance of Message-passing Graph Neural Networks (MP-GNNs). The familiar low-pass filters like Graph Convolutional Networks (GCNs) face performance degradation, which can be attributed to the blending of the messages from dissimilar neighboring nodes. The performance of the low-pass filters on heterophilic graphs still requires an in-depth analysis. In this context, we update the heterophilic graphs by adding a number of self-loops and parallel edges. We observe that eigenvalues of the graph Laplacian decrease and increase respectively by increasing the number of self-loops and parallel edges. We conduct several studies regarding the performance of GCN on various benchmark heterophilic networks by adding either self-loops or parallel edges. The studies reveal that the GCN exhibited either increasing or decreasing performance trends on adding self-loops and parallel edges. In light of the studies, we established connections between the graph spectra and the performance trends of the low-pass filters on the heterophilic graphs. The graph spectra characterize the essential intrinsic properties of the input graph like the presence of connected components, sparsity, average degree, cluster structures, etc. Our work is adept at seamlessly evaluating graph spectrum and properties by observing the performance trends of the low-pass filters without pursuing the costly eigenvalue decomposition. The theoretical foundations are also discussed to validate the impact of adding self-loops and parallel edges on the graph spectrum.

1 Introduction

Graph Neural Networks (Scarselli et al., 2008) made remarkable strides by achieving impeccable performance in graph-structured data. The key reason behind the immense performance superiority is the message passing (MP) framework, which enables the exchange of messages between the adjacent nodes. Before judging the narrative of the success story of the MP framework, let us first mention that graphs can be broadly categorized into two classes such as (1) *homophilic* graphs where adjacent nodes share identical class labels, and (2) *heterophilic* graphs where adjacent node labels are different from each other. The prowess of MP is mostly observed in homophilic graphs because of the tendency to blend messages from similar types of neighbors. In contrast, several studies (Zhu et al., 2020), (Zhu et al., 2021), (He et al., 2022), (Suresh et al., 2021), (Wang et al., 2022) suggest that the MP framework shows exacerbating performances on heterophilic graphs due to the influence of dissimilar messages received from neighbors.

A well-known study (Nt & Maehara, 2019) reveals that all MP-GNNs such as GCN (Kipf & Welling, 2016), GraphSage (Hamilton et al., 2017), GAT (Veličković et al., 2017), SGC (Wu et al., 2019), etc, which smooth the features of adjacent nodes, are low-pass filters. The low-pass filters successfully convert the features of the connected nodes into more similar ones compared to the features of other non-adjacent nodes. This narrates the key reason behind the successful application of low-pass filters on homophilic graphs. Applying low-pass filters on the heterophilic graphs often leads to a degradation in performance as a result of smoothing the features of the dissimilar adjacent nodes. Therefore, analyzing the performance of low-pass filters on heterophilic graphs requires more in-depth scrutiny. The low-pass filters are designed to amplify the coefficients of lower frequencies in the graph spectrum, representing the eigenvalues of the symmetrically normalized graph Laplacian. In homophilic graphs, the smoothing of node features occurs

due to the amplification of the lower frequencies of the graph spectrum. In the case of heterophilic graphs, high-pass filters sharpen the node features by amplifying the higher frequencies of the graph spectrum.

The graph spectrum entails significant information regarding structural patterns such as connected components, community structures, isolated nodes, sparsity, etc. For instance, if the set of eigenvalues contains sufficient zeros, then the network will contain more connected components or isolated nodes. The network will contain weakly (strongly) connected components if the spectrum has a higher number of low frequencies (high frequencies). Therefore, the dissection of the spectrum yields profound information relating to the spatial properties of the graphs. In this work, we investigate the dependency of the graph structure on the performance of the low-pass filters applied to heterophilic networks. We also attempt to uncover the structural properties of the existing heterophilic graphs from their spectrum. The graph spectrum is typically obtained with expensive eigenvalue decomposition which imposes unnecessary computational burden or often can be infeasible to some real-world scenarios. Therefore, we seek to devise an efficient avenue that significantly addresses the computational overhead of evaluating the graph spectrum.

Table 1: Four possible categories A, B, C, and D are presented. Each category depends on the performance trends of a low-pass filter when either self-loops or parallel edges are added to the heterophilic graph.

	Rewiring		
	Self-loop	Parallel edge	Category
Performance of LPF	Increasing (↑)	Increasing (↑)	A
	Increasing (↑)	Decreasing (↓)	B
	Decreasing (↓)	Increasing (↑)	C
	Decreasing (↓)	Decreasing (↓)	D

We aim to bridge the gap by offering two simple strategies that update the graph topology by incorporating self-loops and parallel edges. After the alteration of edge connections, Graph Convolution Network (GCN), a recognized and well-adopted low-pass filter, is applied to the updated graph. We observe some interesting patterns in the performance trends of GCN when the number of self-loops or parallel edges is gradually increased. The performance either monotonically improves or degrades by adding either self-loops or parallel edges. Therefore, we can have four distinct combinations of performance trends. Each combination is tagged with a category name (Refer to Table 1) and every category delineates a particular set of characteristics regarding the spectrum of the input graph. The performance trend of GCN can be attributed to the underlying parity between the lower and higher frequencies present in the graph spectrum. Each performance trend offers a unique insight into the parity of lower and higher frequencies, which directly links with the spatial edge connectivity of the network. We also observe that with the addition of self-loops in the network, the eigenvalues of the normalized graph Laplacian decrease. Conversely, the addition of parallel edges enhances the eigenvalues of the graph Laplacian. The shrinking or expansion of the graph frequencies leads to a specific performance trend, reflecting the distribution of eigenvalues (or frequencies) in the graph spectrum. In this context, we consider 17 benchmark heterophilic graphs to predict their spectrum characteristics by observing the performance trends of GCN after adding either self-loops or parallel edges. We also offer theoretical underpinnings of the shrinking or expansion of the eigenvalues. The phenomena are also observed empirically on the random Erdős-Rény graphs.

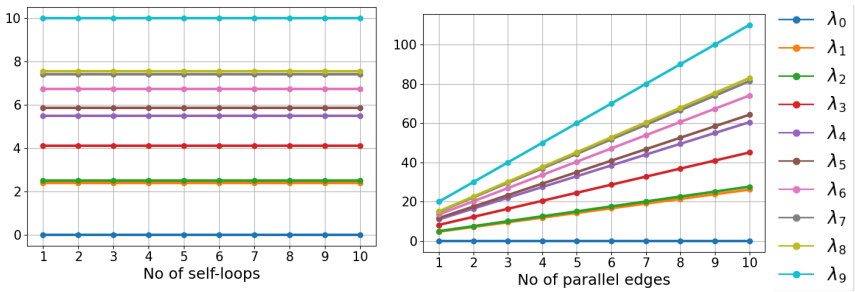
Contribution Our contributions are briefly outlined as follows,

- We provide deeper analyses pertaining to the performance of GCN, a low-pass filter, applied to 17 benchmark heterophilic graphs. We modify the graph structure with the addition of self-loops and parallel edges separately. GCN is applied to the updated graphs and observes the performance trends. We categorize the performance trends into four categories and each graph lies in one of the four categories.

- The categorization of performance trends leads to the identification of characteristics of the graph spectrum. Different performance trends underscore the various patterns of the spectrum which

reveals the properties of the networks like connected components, community structure, sparsity, etc.

- We also observe that the frequencies in the graph spectrum decrease with the addition of self-loops and frequencies increase with the addition of parallel edges. We also establish a connection between the performance trends of GCN and the shrinking or expansion of the frequencies of the graph spectrum.
- We offer extensive theoretical underpinnings for the shrinking or expansion of the frequencies of the graph spectrum with the addition of self-loops and parallel edges in the network. The detailed proofs and derivations are discussed in the Appendix.



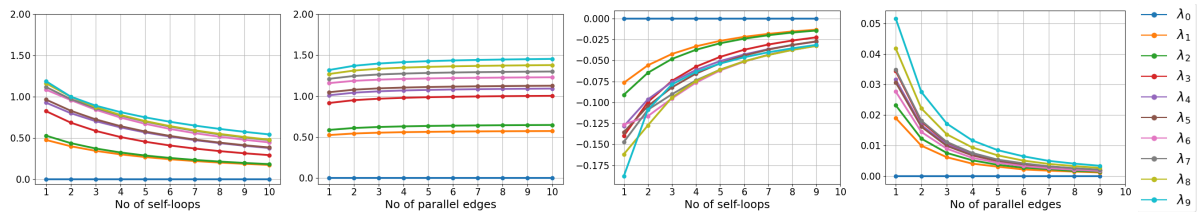
(a) (b)

Figure 1: The changes in the eigenvalues of unnormalized graph Laplacian are presented with the addition of (a) self-loops and (b) parallel edges respectively. The eigenvalues remain unaltered with the addition of self-loops. On the contrary, the unconstrained growth of eigenvalues is observed with the addition of parallel edges.

2 Related Works

The spectral analysis on the graphs gains traction due to its ability to unravel the relationship between frequencies and the spatial connectivity of the networks. Work like (Ortega, 2022) introduces the key ingredients of signal processing like Graph Fourier transforms, frequencies, and the design of the filters for the graph-structured data. Another line of work (Ortega et al., 2018) deals with the intricate details of graph signal processing by shedding light on the spectrum analysis from the perspective of graph Laplacian, extensive real-world applications, and the underlying challenges. The exploitation of spectral analysis in the discrete domain is rigorously harnessed by (Sandryhaila & Moura, 2014). Another mode of work (Tremblay et al., 2018) offers the prospect of designing versatile filter banks and spectral wavelets on graph-structured data. The design of efficient and localized convolutional filters becomes an inevitable area of research which is initiated by (Defferrard et al., 2016).

A large pool of well-adopted GNNs like GCN, GraphSage, GAT, SGC, etc are recognized as potential low-pass filters. The fact is first asserted by (Nt & Maehara, 2019). Chen *et al.* (Chen et al., 2023) established the bridge to fill the gap between the spatial and spectral properties of the prevalent graph neural networks. AutoGCN (Wu et al., 2022) proposes a variant of GCN equipped with low-pass, high-pass, and band-pass filters which automatically adjusts magnitude depending on the homophily or heterophily of the input graph. In this connection, another prominent work AdaGNN (Dong et al., 2021) learns an adaptive filter that spans across multiple layers, capturing the varying node frequencies to improve node embeddings. Taking the cue from above, FAGCN (Bo et al., 2021) firstly proposed one low-pass and one high-pass filter. The proposed filters automatically maintain the balance of collecting information from both homophilic and heterophilic graphs. Designing novel convolutional filters became an enticing avenue which is evident by (Bianchi et al., 2021) which employed an auto-regressive moving average filter (ARMA) filter by replacing polynomial filters to achieve a more robust, flexible frequency response. Another work EGC (Tailor et al., 2021) employed



(a) (b) (c) (d)

Figure 2: (a) Eigenvalues of $\tilde{L}$ exhibit a decreasing trend with the increase in addition of self-loops, and (b) eigenvalues show an increasing trend with the addition of parallel edges. The corresponding changes in the eigenvalues with the addition of self-loops and parallel edges are demonstrated in (c) and (d) respectively.

the effectiveness of isotropic message passing where the message function only depends on the source nodes, over the anisotropic message passing. EGC allows the involvement of multiple learnable filters to achieve a spatially evolving frequency response. (Zhu & Koniusz, 2021) leverage the use of Markov Diffusion Kernel to obtain a filter that maintains a delicate balance between the harnessing information of local and global context for each node across the network. On the other side, (Singh & Chen, 2023) identified the gap for spectral analysis on signed graphs. To mitigate the gap, they proposed two signed GNNs that retain low-pass and high-pass information respectively. Compatible Label Propagation (CLP) (Zhong et al., 2022) combines class compatibility matrix and label propagation to improve node classification on heterophilic graphs. Very recently an inductive spectral filter SLOG (Xu et al.) is proposed which considers real-valued order polynomial filters. SLOG also combines subgraph sampling in the spatial domain and signal processing in the spectral domain.

3 A Deeper Investigation

3.1 Notations

Consider an attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, X)$ where $\mathcal{V}$ denotes the set of vertices with $|\mathcal{V}| = n$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, and $X \in \mathbb{R}^{n \times d}$ is the feature matrix contains d-dimensional feature vectors. We define graph Laplacian $L = D - A$ and symmetrically normalized Laplacian as $\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = \mathrm{I} - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. Also, augmenting self-loops the normalized Laplacian will be $\tilde{L} = \mathrm{I} - \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ where $\tilde{A} = A + \mathrm{I}$ and $\tilde{D} = D + \mathrm{I}$.

3.2 Preliminaries on Spectral Graph Theory

The spectral analysis of graphs (Shuman et al., 2013) revolves around understanding the characteristics of eigenvalues and eigenvectors of the symmetrically normalized graph Laplacian $\tilde{L}$. The analysis entails that eigenvalues and eigenvectors of $\tilde{L}$ represent the Fourier frequency and Fourier modes. Suppose, the eigendecomposition on Laplacian yields us $\tilde{L} = U \Sigma U^\top$ where columns of U represent the eigenvectors and Σ is a diagonal matrix containing the eigenvalues. Let a signal $x \in \mathbb{R}^n$ act on the nodes in the graph, then the Fourier transformation of x is presented as $\hat{x} = U^\top x$. The inverse Fourier transform can be formulated as $x = U\tilde{x}$. Thus, for any filter g, the graph convolution between g and x is estimated as:

$$g * x = U((U^\top g) \odot (U^\top x)) = U \tilde{G} U^\top x, \tag{1}$$

where $\odot$ denotes the element-wise vector multiplication and $\tilde{G} = \mathrm{diag}\{\tilde{g}_1, \cdots, \tilde{g}_n\}$. Each $\tilde{g}_i$ denotes the spectral filter coefficient. A well-known fact is that $\tilde{L}$ has the eigenvalues lie in $[0, 2]$. Let us categorize the set of eigenvalues as $\lambda_{<1}$ or *lower frequencies* which are strictly smaller than 1 and $\lambda_{\geq 1}$ or *higher frequencies*, which are greater than or equal to 1. If a filter amplifies the coefficients of $\lambda_{<1}$, then it acts as a low-pass filter, and on the contrary high-pass filter amplifies the coefficients of $\lambda_{\geq 1}$. For instance, the filter function of GCN is $\tilde{G} = \mathrm{I} - \Sigma$ where the coefficients of $\lambda_{<1}$ increases. Therefore, GCN acts as a low-pass filter.

3.3 Addition of Self-loops

We conduct experiments by adding the self-loops corresponding to each node in the graph. The number of self-loops is denoted by α. After the addition of self-loops α-times, the adjacency matrix will be $\tilde{A}_\alpha = A + \alpha\mathrm{I}$ and the corresponding degree matrix will look like $\tilde{D}_\alpha = D + \alpha\mathrm{I}$. Therefore, the symmetrically normalized graph Laplacian can be presented as $\tilde{L}_\alpha = \mathrm{I} - \tilde{D}_\alpha^{-\frac{1}{2}}\tilde{A}_\alpha\tilde{D}_\alpha^{-\frac{1}{2}}$.

3.4 Addition of Parallel edges

We also perform experiments by adding parallel edges corresponding to every edge in the graph. Assume, γ denotes the number of parallel edges to be added in the network. Therefore, adding γ-times parallel edges, the updated adjacency matrix will be $A_\gamma = (\gamma + 1)A$. Thus, the corresponding degree matrix will be $D_\gamma = (\gamma+1)D$. Adding self-loops the further modified adjacency and degree matrices will be $\tilde{A}_\gamma = (\gamma+1)A+\mathrm{I}$ and $\tilde{D}_\gamma = (\gamma+1)D+\mathrm{I}$. Therefore, we can define the corresponding symmetrically normalized graph Laplacian as $\tilde{L}_\gamma = \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}}\tilde{A}_\gamma\tilde{D}_\gamma^{-\frac{1}{2}}$.

Table 2: Total number of nodes, isolated nodes, edge density, and average degree for 17 heterophilic graphs are presented. The lower values of $\overline{\mathrm{sp}}_{\mathcal{G}}$ and $\overline{d}_{\mathrm{avg}}$ respectively indicate higher density and average degree.

Properties/Datasets	Cornell	Texas	Wisconsin	Chameleon	Squirrel	Actor
# nodes	183	183	251	2277	5201	7600
# isolated nodes	87(47.5%)	73(39.8%)	81(32.3%)	0(0%)	0(0%)	636(8.36%)
density ($\overline{\mathrm{sp}}_{\mathcal{G}}$)	4.02	3.93	4.10	4.27	4.13	6.86
avg. degree ($\overline{d}_{\mathrm{avg}}$)	3.82	3.82	4.13	6.34	7.17	7.54
Properties/Datasets	arxiv-year	snap-patents	Penn94	pokec	twitch-gamers	genius
# nodes	169343	2923922	41554	1632803	168114	421961
# isolated nodes (%)	17440(10.29%)	881754(30.15%)	0(0%)	200110(12.25%)	44596(26.52%)	371870(88.12%)
density ($\overline{\mathrm{sp}}_{\mathcal{G}}$)	9.41	12.63	5.75	10.68	7.63	11.41
avg. degree ($\overline{d}_{\mathrm{avg}}$)	10.65	13.50	9.24	12.91	10.64	11.56
Properties/Datasets	Roman-empire	Amazon-ratings	Minesweeper	Tolokers	Questions	-
# nodes	22662	24492	10000	11758	48921	
# isolated nodes	0(0%)	3495(14.26%)	1(0.01%)	936(7.96%)	17761(36.3%)	
density ($\overline{\mathrm{sp}}_{\mathcal{G}}$)	8.26	8.07	7.14	4.89	8.96	
avg. degree ($\overline{d}_{\mathrm{avg}}$)	8.64	8.71	7.82	7.98	9.41	

3.5 Empirical Evidence on Random Graphs

We conducted experiments on randomly generated Erdős-Rényi graphs to study the effects on the eigenvalues with the addition of self-loops and parallel edges in the graph. A random graph $\mathcal{G}_{er}$ is generated with 10 vertices and having edge probability 0.50. Two individual experiments are performed with (1) the addition of self-loops, and (2) the addition of parallel edges. We varied the number of self-loops or parallel edges ranging from 1 to 10. The eigendecomposition is performed for every stage of addition to monitor the changes that occurred in the corresponding eigenvalues. Refer to Figure 2 for the portrayal of the effect on the eigenvalues and the other corresponding changes in the spectrum. The plots (a) and (b) demonstrate the effect on the of 10 eigenvalues of $\mathcal{G}_{er}$ while the addition of self-loops or parallel edges in the graph. On the other side, plots (c) and (d) depict the changes in the eigenvalues. As we have added 10 self-loops or parallel edges, thus 9 differences are recorded in the plots.

Observation The plots reaffirm that the addition of self-loops leads to the shrinking of the eigenvalues in the spectrum except for the eigenvalue 0 (Refer Figure 2(a)). On the contrary, eigenvalues increase with the addition of parallel edges (Refer Figure 2(b)). Additionally, we also present the change in the eigenvalues in Figures 2(c) and 2(d). Notably, the monotone increase (decrease) of the curves underlines the slower rate with the addition of self-loops (parallel edges).

Table 3: GCN (a low-pass filter) is applied on the 6 standard heterophilic datasets curated by Pie *et al.* (Pei et al., 2020). Performance analyses are demonstrated after adding multiple self-loops and parallel edges respectively in the graphs. The performance trends are also highlighted and corresponding categories are marked for each dataset.

Method	Input adjacency	Chameleon	Squirrel	Actor	Cornell	Texas	Wisconsin
GCN	A + I	71.68 ± 1.92	62.78 ± 1.99	27.40 ± 1.12	40.27 ± 6.44	54.86 ± 5.27	45.29 ± 6.10
	A + 2I	67.69 ± 2.25	58.64 ± 2.27	29.20 ± 1.13	43.78 ± 5.09	56.21 ± 4.95	50.19 ± 6.39
	A + 3I	65.15 ± 1.56	55.35 ± 1.76	30.67 ± 1.25	47.83 ± 8.11	54.05 ± 4.18	56.47 ± 3.80
	A + 4I	63.22 ± 1.22	53.43 ± 1.40	32.08 ± 1.20	46.48 ± 7.81	58.64 ± 6.16	60.98 ± 4.37
	A + 5I	61.90 ± 2.51	51.44 ± 1.51	32.94 ± 0.85	50.27 ± 7.47	57.02 ± 7.49	61.96 ± 5.42
Trend	→	Decreasing (↓)	Decreasing (↓)	Increasing (↑)	Increasing (↑)	Increasing (↑)	Increasing (↑)
GCN	A + I	71.68 ± 1.92	62.70 ± 1.99	27.40 ± 1.12	40.27 ± 6.44	54.86 ± 5.27	45.29 ± 6.10
	2A + I	75.06 ± 1.24	66.43 ± 2.40	25.54 ± 1.30	40.54 ± 6.39	54.59 ± 6.13	47.45 ± 4.94
	3A + I	76.31 ± 0.99	67.79 ± 1.96	25.63 ± 0.69	44.05 ± 7.83	55.67 ± 5.43	47.25 ± 5.22
	4A + I	76.60 ± 0.77	67.92 ± 1.66	24.76 ± 1.19	45.67 ± 8.58	51.89 ± 7.43	46.86 ± 4.24
	5A + I	77.32 ± 1.07	68.59 ± 1.71	24.82 ± 1.34	41.08 ± 5.64	51.08 ± 10.0	46.86 ± 4.75
Trend	→	Increasing (↑)	Increasing (↑)	Decreasing (↓)	Increasing (↑)	Increasing (↑)	Increasing (↑)
Category	→	C	C	B	A	A	A

3.6 Theoretical Analysis: A Spectral Perspective

We perform a deeper theoretical analysis of the effect on the graph spectrum when adding self-loops or parallel edges. The detailed study is provided as follows,

Lemma 1. *Consider a graph G with A and D as the adjacency and degree matrix. Now α-times self-loops are added in G with $\alpha_1 \in \mathbb{Z}^+$. Assume λ_α^{max} is the maximum eigenvalue of symmetrically normalized graph Laplacian $\tilde{L}_\alpha$ of the updated graph. If β_1 is the smallest eigenvalue of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\max_i d_i$ is the maximum degree of G, then $\lambda_\alpha^{max} \leq \frac{\max_i d_i(1-\beta_1)}{\alpha+\max_i d_i}$.*

Lemma 2. *Consider a graph G with A and D as the adjacency and degree matrix. Now γ-times self-loops are added in G with $\gamma \in \mathbb{Z}^+$. Assume λ_γ^{max} is the maximum eigenvalue of symmetrically normalized graph Laplacian $\tilde{L}_\gamma$ of the updated graph. If β_1 is the smallest eigenvalue of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\max_i d_i$ is the maximum degree of G, then $\lambda_\gamma^{max} \leq \frac{(1+\gamma)\max_i d_i(1-\beta_1)}{1+(1+\gamma)\max_i d_i}$.*

Remark 1. *The maximum eigenvalue of $\tilde{L}_\alpha$ decreases with the increasing number of self-loops in the network, signifying the shrinking of the graph spectrum. The maximum eigenvalue of $\tilde{L}_\gamma$ increases with the increasing number of parallel edges in the network, illustrating the expansion of the graph spectrum.*

Lemma 3. *Given a k-regular graph $\mathcal{G}$, the eigenvalues of $\tilde{A}_N^\alpha$ will lie in $[-1, 1]$ $\forall \alpha \geq 1$.*

Lemma 4. *Given a k-regular graph $\mathcal{G}$, the eigenvalues of $\tilde{A}_N^\gamma$ will lie in $[-1, 1]$ $\forall \gamma \geq 1$.*

Remark 2. *The eigenvalues will lie in $[-1, 1]$ whether the self-loops or parallel edges are added in a regular graph. The range of eigenvalues will remain unaffected with the addition of self-loops or parallel edges.*

Theorem 1. *Consider a k-regular graph with $\alpha_1, \alpha_2 \in \mathbb{R}^+$ where $\alpha_1 \leq \alpha_2$, then the inequality will hold $\lambda_{\alpha_1}^i \geq \lambda_{\alpha_2}^i, \forall\ 1 \leq i \leq n$ where $\lambda_{\alpha_1}^i$ and $\lambda_{\alpha_2}^i$ are the ith eigenvalues of $\tilde{L}_{\alpha_1}$ and $\tilde{L}_{\alpha_2}$ respectively.*

Theorem 2. *Consider a k-regular graph with $\gamma_1, \gamma_2 \in \mathbb{R}^+$ with $\gamma_1 \leq \gamma_2$, then the inequality will hold $\lambda_{\gamma_1}^i \leq \lambda_{\gamma_2}^i, \forall\ 1 \leq i \leq n$ where $\lambda_{\gamma_1}^i$ and $\lambda_{\gamma_2}^i$ are the ith eigenvalues of $\tilde{L}_{\gamma_1}$ and $\tilde{L}_{\gamma_2}$ respectively.*

Remark 3. *We have shown that for the regular graphs, the eigenvalues of the symmetrically normalized graph Laplacian decrease when self-loops are added. The addition of self-loops attenuates the frequencies of the graph spectrum. The eigenvalues of the symmetrically normalized graph Laplacian increase when parallel edges are added. The addition of parallel edges amplifies the frequencies of the graph spectrum.*

Corollary 1. *The increase in the eigenvalues of L_γ is independent of the number of self-loop additions in $\mathcal{G}$. On the contrary, The eigenvalues of $\tilde{L}_\gamma$ will increase if at least one self-loop is added per node in $\mathcal{G}$.*

In the following two theorems, we will demonstrate the alteration of the eigenvalues of A_N with the addition of self-loops or parallel edges by the perturbation of the adjacency matrix.

Theorem 3. *Consider a connected graph $\mathcal{G}$ with $A_N = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. Assuming the diagonal of A of G is perturbed by a significantly small $\alpha > 0$, then the updated normalized adjacency matrix will be A_N^{α}. The change in the eigenvalues of A_N^{α} with respect to eigenvalues of A_N will increase when α increases.*

Theorem 4. *Consider a connected graph $\mathcal{G}$ with normalized adjacency matrix $A_N = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. Assuming each element of A except the diagonal multiplied by $1+\gamma$ where $\gamma > 0$ is a significantly small quantity, then the updated normalized adjacency matrix will be A_N^{γ}. The change in the eigenvalues of A_N^{γ} with respect to eigenvalues of A_N will decrease when γ increases.*

Remark 4. *The Theorem 3 suggests the change in the eigenvalues of the normalized adjacency matrix increases on increasing number of self-loops. Conversely, Theorem 4 states the change of eigenvalues of normalized adjacency matrix decreases with the increasing number of parallel edges.*

4 Experiments

4.1 Datasets

Our experiments encompass three categories of datasets (1) Pie *et al.* (Pei et al., 2020) proposed 6 standard heterophilic networks consisting of Cornell, Texas, Wisconsin, Chameleon, Squirrel, and Actor, (2) Lim *et al.* (Lim et al., 2021) curated 6 large-scale heterophilic networks namely arxiv-year, snap-patents, Penn94, pokec, twitch-gamers, and genius, and (3) Platonov *et al.* (Platonov et al., 2023) identified prevailing shortcomings on the existing datasets and developed a set of 5 heterophilic networks viz Roman-empire, Amazon-ratings, Minesweeper, Tolokers, and Questions. The details of all datasets are vividly available in Table 2.

4.2 Experimental Settings

For all graphs from Pie *et al.*, we have considered 10 standard train/valid/test splits with 60%/20%/20% samples. Graphs from Lim *et al.* and Platonov *et al.* train/valid/test splits are fixed as 50%/25%/25% for 5 and 10 splits respectively. We applied a two-layered GCN architecture across all 17 graphs to carry out entire experiments. The dropout rate is fixed at 0.50 and LayerNorm is employed to make training convergence faster. The model parameters are optimized by Adam optimizer. We evaluated the best model on every split and finally reported mean and standard deviations across all splits for each of the datasets. The performance metric is test accuracy and for Minesweeper, Tolokers, and Questions the ROC-AUC is reported. Our Pytorch and Pytorch-geoemtric based implementation is available at https://anonymous.4open.science/r/DIHG-5212/README.md.

4.3 Analysis of Isolated Nodes, Sparsity, and Average Degree in the Heterophilic Networks

We offer an in-depth analysis of the number of isolated nodes, edge density, and the average degree in the context of the heterophilic networks. Refer to Table 2 for illustrating the comparative statistics of the various networks. As per the existing formula, the edge density can be defined as $\text{sp}_{\mathcal{G}} = \frac{2|\mathcal{E}|}{|\mathcal{V}|(|\mathcal{V}|-1)}$. If the input graph is too sparse, the estimated value will be too little to comprehend. Therefore, we devise an alternative solution with the assistance of a logarithmic scale which is defined as $\overline{\text{sp}}_{\mathcal{G}} = -\log(\text{sp} + \epsilon)$ where ϵ is tiny real number is added to avert the numerical instability. The scaling suggests that the lower the $\overline{\text{sp}}_{\mathcal{G}}$, the more dense the graph is. A similar issue is confronted in the estimation of average degree as $d_{\text{avg}} = \frac{2|\mathcal{E}|}{|\mathcal{V}|}$. Therefore, we also pursue similar tricks to tackle the issue. The scaled average degree is defined as $\overline{d}_{\text{avg}} = -\log(d_{\text{avg}} + \epsilon)$ where ϵ is same as defined as earlier. A lower $\overline{d}_{\text{avg}}$ signifies the higher average degree of the underlying network.

4.4 Category of Distribution of Eigenvalues

Suppose we attempt to apply a low-pass filter (like GCN) on any input graph. The graph will be pre-processed either by adding a fixed number of self-loops or parallel edges. Considering all possibilities, the low-pass filter can exhibit four distinct types of performance trends. The various possible trends are vividly portrayed in Table 1. We marked the categories respectively as A, B, C, and D. The different performance trends may

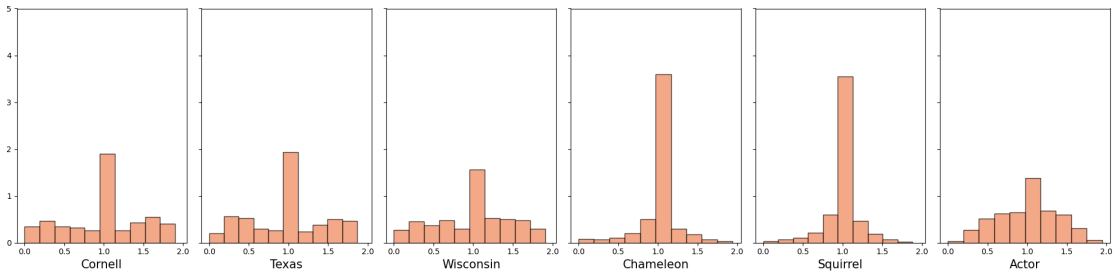


Figure 3: The distribution of eigenvalues of normalized Laplacian matrix presented for six standard heterophilic datasets.

be the manifestations of the underlying edge connectivity of the network. The structure of the networks has an inherent connection with the eigenvalues derived from the normalized graph Laplacian. Therefore, the defined categories may offer to comprehend the parity of lower and higher frequencies (eigenvalues) of the graph spectrum.

Table 4: GCN (a low-pass filter) is applied on the 6 large-scale heterophilic datasets curated by Lim *et al.* (Lim et al., 2021). Performance analyses are demonstrated after adding multiple self-loops and parallel edges respectively in the graphs. The performance trends are also highlighted and corresponding categories are marked for each dataset.

Method	Input Adjacency	arxiv-year	snap-patents	Penn94	pokec	twitch-gamers	genius
GCN	A + I	48.75 ± 0.43	34.96 ± 0.15	77.12 ± 0.43	59.53 ± 0.18	60.83 ± 0.29	80.03 ± 0.66
	A + 2I	48.59 ± 0.53	34.77 ± 0.14	75.82 ± 0.45	59.17 ± 0.20	61.00 ± 0.34	79.81 ± 1.38
	A + 3I	47.34 ± 0.20	34.60 ± 0.04	$74,32 \pm 0.49$	59.77 ± 0.15	61.07 ± 0.33	78.77 ± 0.99
	A + 4I	46.26 ± 0.21	34.57 ± 0.11	72.79 ± 0.55	60.43 ± 0.14	61.18 ± 0.26	77.80 ± 0.51
	A + 5I	45.73 ± 0.48	34.43 ± 0.17	71.39 ± 0.40	60.94 ± 0.17	61.23 ± 0.30	77.35 ± 0.37
Trend	→	Decreasing (↓)	Decreasing (↓)	Decreasing (↓)	Increasing (↑)	Increasing (↑)	Decreasing (↓)
GCN	A + I	48.69 ± 0.37	35.00 ± 0.15	77.13 ± 0.39	59.54 ± 0.23	60.86 ± 0.30	80.09 ± 0.70
	2A + I	48.31 ± 0.54	34.89 ± 0.12	77.64 ± 0.41	61.32 ± 0.14	60.72 ± 0.21	74.41 ± 2.07
	3A + I	47.93 ± 0.64	34.64 ± 0.23	77.81 ± 0.38	62.22 ± 0.10	60.70 ± 0.24	69.88 ± 0.48
	4A + I	47.74 ± 0.52	34.46 ± 0.18	77.88 ± 0.37	62.77 ± 0.09	60.70 ± 0.21	70.17 ± 0.89
	5A + I	47.78 ± 0.38	34.28 ± 0.13	77.81 ± 0.35	63.10 ± 0.10	60.69 ± 0.21	71.13 ± 1.26
Trend	→	Decreasing (↓)	Decreasing (↓)	Increasing (↑)	Increasing (↑)	Decreasing (↓)	Decreasing (↓)
Category	→	D	D	C	A	B	D

4.5 Initial Distribution of Eigenvalues

We estimate the distribution of eigenvalues from the normalized graph Laplacian for six standard heterophilic datasets Cornell, Texas, Wisconsin, Chameleon, Squirrel, and Actor, Refer to Figure 3 for the detailed illustration. The histograms of Cornell, Texas, and Wisconsin are identical. On the other side, the corresponding histograms of Chameleon and Squirrel carry similar patterns, The histogram of the Actor dataset is completely different compared to the rest of the others. Later we will observe that the datasets have histograms of similar patterns that will yield identical trends in the performances.

4.6 Performance Categorisation of Low-pass Filter

We performed semi-supervised node classification on a diverse array of heterophilic graphs to analyze the different performance trends of the low-pass filter. For each graph, separate experiments were conducted to study the effects of the addition of self-loops and parallel edges respectively. We applied GCN, a well-adopted low-pass filter, on 17 heterophilic graphs, and the respective performance trends are demonstrated in Tables 3, 4, and 5 for respectively standard heterophilic graphs, large-scale datasets, and currently proposed heterophilic graphs. The study reveals that each graph exhibits a steady pattern of either increasing or decreasing while adding either the self-loops or the parallel edges. We further marked the categories considering the performance trends for the increasing number of self-loops and parallel edges. For instance,

GCN on Chameleon showed a decreasing trend while increasing the number of parallel edges, and on the contrary, performance increases on adding the parallel edges. Therefore, Chameleon was assumed to be in the category of A as per the rules from Table 1. Note that, the steady patterns are persistent across every dataset considered for the experimentation.

4.7 Observation from Performance Trends

The rationale against the backdrop of the performance trend can be explained through the lens of analyzing the spectrum of the graph. Since GCN performs the low-pass filtering, the coefficients of the lower frequencies will be enhanced, and the coefficients of the higher frequencies will be shrunk. Based on this, our discussion will revolve around analyzing 6 standard heterophilic graphs.

Table 5: GCN (a low-pass filter) is applied on the 5 newly-proposed heterophilic datasets curated by Platonov *et al.* (Platonov et al., 2023). Performance analyses are demonstrated after adding multiple self-loops and parallel edges respectively in the graphs. The performance trends are also highlighted and corresponding categories are marked for each dataset.

Method	Input Adjacency	Roman-empire	Amazon-ratings	Minesweeper	Tolokers	Questions
GCN	A + I	76.70 ± 0.63	42.01 ± 0.58	89.51 ± 0.54	80.10 ± 1.11	73.96 ± 1.44
	A + 2I	76.28 ± 0.58	41.58 ± 0.56	89.46 ± 0.53	80.13 ± 1.06	73.69 ± 0.67
	A + 3I	75.99 ± 0.76	41.56 ± 0.53	89.41 ± 0.55	80.03 ± 1.02	72.54 ± 1.64
	A + 4I	75.72 ± 0.70	41.49 ± 0.40	89.33 ± 0.54	79.76 ± 1.02	72.56 ± 1.23
	A + 5I	75.26 ± 0.55	41.21 ± 0.60	89.22 ± 0.53	79.70 ± 0.98	72.53 ± 1.15
Trend	→	Decreasing (↓)	Decreasing (↓)	Decreasing (↓)	Decreasing (↓)	Decreasing (↓)
GCN	A + I	76.71 ± 0.62	42.00 ± 0.58	89.51 ± 0.54	80.10 ± 1.12	73.30 ± 2.04
	2A + I	76.75 ± 0.80	41.93 ± 0.50	89.57 ± 0.51	79.95 ± 1.08	74.51 ± 1.23
	3A + I	76.99 ± 0.67	41.93 ± 0.90	89.57 ± 0.52	79.87 ± 0.93	75.05 ± 0.99
	4A + I	76.90 ± 0.59	42.02 ± 0.53	89.57 ± 0.51	79.87 ± 0.99	74.93 ± 1.18
	5A + I	76.95 ± 0.65	42.18 ± 0.77	89.60 ± 0.49	79.75 ± 0.94	74.73 ± 1.58
Trend	→	Increasing (↑)	Increasing (↑)	Increasing (↑)	Decreasing (↓)	Increasing (↑)
Category	→	C	C	C	D	C

Addition of Self-loops We observed that the addition of self-loops improves the performance of GCN on Cornell, Texas, Wisconsin, and Actor (Refer Table 3). Aligning to the theoretical analyses, adding self-loops will create a decreasing trend of the eigenvalues of $\tilde{L}$. Consequently, the number of lower frequencies will increase in the graph spectrum, which is beneficial for the performance boost of the GCN. Refer to Figure 4((a), (b), (c), (f)) to visualize the change of distribution of eigenvalues of $\tilde{L}$ with the addition of self-loops. A deep observation reveals that the number of lower frequencies gradually increases while increasing the number of self-loops in the graph. This transformation offers a conducive environment for the operation of the low-pass filter (here GCN). Finally, the shape of the eigenvalue distribution resembles when $\alpha = 5$ for all four datasets.

A stark contrast was observed in the performance trends of the Chameleon and Squirrel datasets. Both of the datasets demonstrated degrading performances with the addition of an increasing number of self-loops. As mentioned earlier the addition of self-loops will shrink the spectrum of the graph, leading to the increase in the lower frequencies. Refer 4((d), (e)) for the nuanced view of the alteration of eigenvalue distribution of $\tilde{L}$ for both datasets. Both histograms depict that the lower frequencies increase but eigenvalues are mostly centering towards 1 which might not be beneficial for the operation of GCN. Therefore, in this scenario, GCN witnessed deteriorating performances with the increasing number of self-loops.

Addition of Parallel edges The addition of parallel edges improves the performance of Cornell, Texas, Wisconsin, Chameleon, and Squirrel. The theoretical analysis illustrates that eigenvalues increase with the increasing number of parallel edges. Consequently, the number of higher frequencies will increase which seems to impede the performance of GCN. Refer to Figure 5((a), (b), (c), (d), (e)) to visualize the change in the distribution of eigenvalues of $\tilde{L}$ with the addition of the parallel edges. One common point should be mentioned that despite the increasing number of higher frequencies, many lower frequencies are still retained

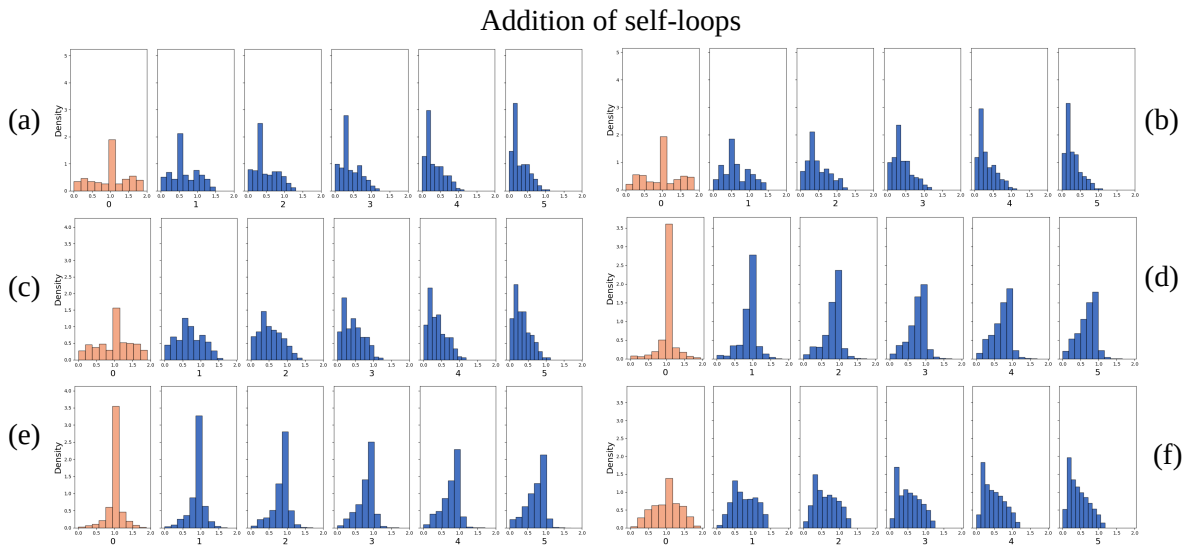


Figure 4: The distribution of eigenvalues of normalized graph Laplacian for the datasets (a) Cornell, (b) Texas, (c) Wisconsin, (d) Chameleon, (e) Squirrel, and (f) Actor is demonstrated after adding self-loops. The initial distribution is shown on the left side of each diagram. The number of self-loops varies from 1 to 5 and corresponding changes are recorded.

in the spectrum. This phenomenon performs the balance of the parity of lower and higher frequencies in the graph spectrum, eventually improving the performance of GCN.

On the contrary, GCN exhibited deteriorating performances on the Actor dataset with the increasing number of parallel edges. Refer to Figure 5(f) for the histogram asserting that the number of lower frequencies almost diminished. As a consequence, GCN witnessed a degrading performance.

Key Takeaway The crux of the experiments lies in the identification of trends (either increasing or decreasing) in the performance of low-pass filters with the addition of self-loops or parallel edges. Every input graph can be uniquely mapped to one of the four pre-defined categories depending on the performance trends. We marked the respective categories of all 17 heterophilic graphs involved in the experimentation. The assigned category characterizes the specific eigenvalue distribution of the normalized graph Laplacian for the corresponding graph. The distribution of the eigenvalues offers significant insights into the structural patterns of the graphs like connected components, community structures, expansion properties, clustering, robustness, etc. The unraveling of such properties is often accomplished by pursuing computationally expensive eigenvalue decomposition or time-consuming prevailing graph algorithms. Amid the growing size of the graph resorting to such strategies leads to a labyrinth of the methodologies.

The size of the graphs like Cornell, Texas, or Chameleon is moderate and we performed eigenvalue decomposition to study the eigenvalue distribution of $\tilde{L}$. The initial distributions of the six graphs are presented in Figure 3. One can easily contemplate the high-level structural properties by observing the initial patterns. For instance, Cornell, Texas, and Wisconsin have a distribution mostly symmetrical around 1 in the spectrum, asserting the balanced parity of lower and higher frequencies. The observation underscores the presence of weakly connected components or isolated nodes. This is also validated from Table 2 as Cornell contains 47.5% isolated nodes. A similar argument applies to both Texas and Wisconsin. Conversely, Chameleon and Squirrel both have higher frequencies which signifies that graphs are dense, have strongly connected components, and contain no isolated nodes as confirmed from Table 2. In a different vein, Actor has an almost balanced parity of lower and higher frequencies characterizing the lower number of isolated nodes with a moderate average degree and density compared to the other five previously mentioned graphs.

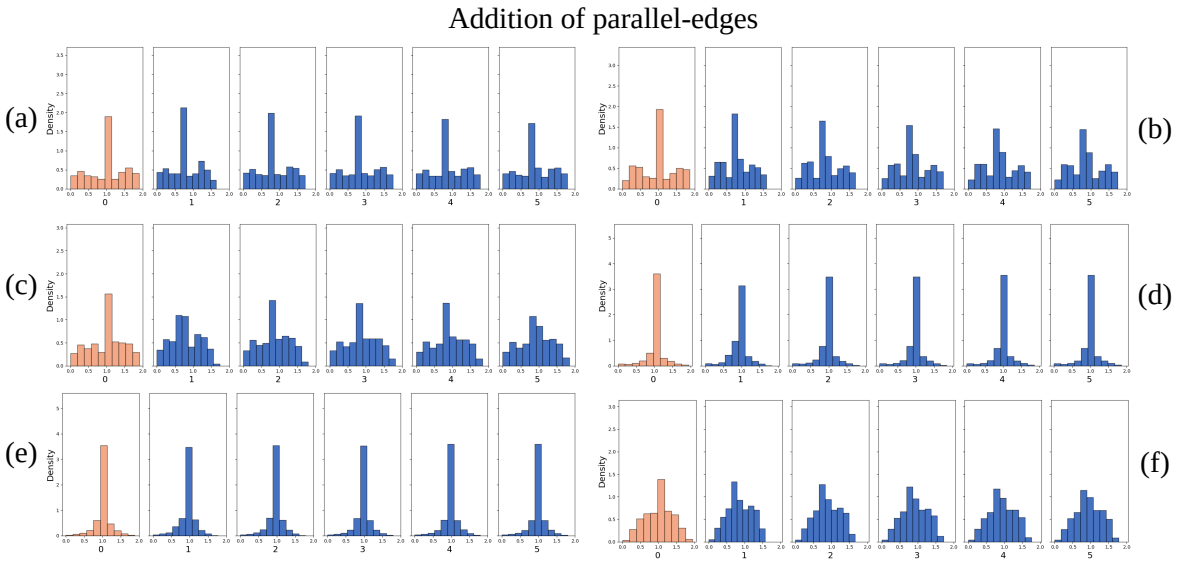


Figure 5: The distribution of eigenvalues of normalized graph Laplacian for the datasets (a) Cornell, (b) Texas, (c) Wisconsin, (d) Chameleon, (e) Squirrel, and (f) Actor are demonstrated after adding parallel edges. The initial distribution is shown on the left side of each diagram. The number of parallel edges varies from 1 to 5 and corresponding changes are recorded.

4.8 Inferring Characteristics of Spectrum for Large-scale Graphs

Evaluating the eigenvalue distribution of $\tilde{L}$ for a large-scale graph is critically challenging due to the potential computational overhead. Exploration of eigenvalue distribution is possible by separately observing the performance trends of the low-pass filters, with the addition of self-loops and parallel edges. This approach drastically reduces the computational budget and offers deeper insights into the intricate structural patterns in the given networks.

arxiv-year, snap-patents, and genius As per empirical evidence, GCN witnessed decreasing performance trends on arxiv-year, snap-patents, and genius in both cases of self-loops and parallel edges. The results emphasized that three networks contain a significantly lower number of non-zero frequencies with a substantial number of zero eigenvalues, asserting that the networks contain a large number of isolated nodes. The edge density and average degree of the graphs are also lower for the three graphs which can also be verified by referring to Table 2. Therefore, we can predict the weak connectivity and the presence of isolated nodes in those graphs by only observing the performance trends with the addition of self-loops and parallel edges.

Penn94 Penn94 was categorized in "C" resembling the category of Chameleon and Squirrel (Refer Figure 3). The initial frequency distribution of $\tilde{L}$ suggests a greater number of higher frequencies resulting in the densely connected network compared to the other five graphs (Refer Table 2). Also, Penn94 contains no isolated nodes sharing properties similar to those of Chameleon and Squirrel.

twitch-gamers The performance trends made twitch-gamers posited in category "B" which is identical to Actor. Thus, the eigenvalue distribution of twitch-gamers also resembles to Actor, having balanced centering to eigenvalue 1 and almost equal parity of lower and higher frequencies. Like Actor, twitch-gamers also have moderately dense and average node degrees with a lesser number of isolated nodes compared to the other five heterophilic graphs in this group.

pokec The category of pokec marked as A shares identical characteristics with Cornell, Texas, and Wisconsin. The performance trend indicated the presence of the eigenvalues centering around eigenvalue 1 having balanced parity of lower and higher frequencies. Identically, pokec may contain isolated nodes and also weakly connected components.

Table 6: Analysis of the performance of GCN on Chameleon dataset is presented with the variation of number of self-loops and parallel edges simultaneously across the network.

Dataset	# Parallel edges		1	2	3	4	5
Chameleon	# Self-loops	1	65.00 ± 1.32	68.61 ± 1.52	68.94 ± 1.53	69.14 ± 1.35	69.75 ± 0.98
		2	59.84 ± 1.97	65.06 ± 2.07	66.71 ± 1.35	68.50 ± 1.70	68.64 ± 1.77
		3	58.55 ± 2.15	68.61 ± 2.26	65.00 ± 1.95	66.88 ± 2.01	67.41 ± 1.39
		4	58.04 ± 2.78	59.53 ± 1.63	62.82 ± 2.03	64.42 ± 1.81	67.03 ± 1.90
		5	57.93 ± 2.31	59.16 ± 1.54	60.24 ± 2.55	63.70 ± 2.00	65.06 ± 1.49

Roman-empire, Amazon-ratings, Minesweeper, and Questions The four graphs have demonstrated declining performance trends while self-loops are added and performance is improved with the addition of parallel edges. The graphs thereby belonged to the category of "C" similar to Chameleon and Squirrel. This predicament has underscored that graphs contain higher frequencies than lower frequencies. The graph spectra are indicative of the higher sparsity and lower average degree of the graphs, compared to the same of the Tolokers.

Tolokers Tolokers is categorized as "D" having similarity with arxiv-year, snap-patents, and genius. The performance trends signal the presence of a higher number of zero eigenvalues in the spectrum, containing a good number of isolated nodes (Refer to Table 2).

4.9 Visualization of Spectrum for Heterophilic Graphs

We conducted a comparative study on the lower and higher frequencies of the six heterophilic graphs Chameleon, Squirrel, Texas, Cornell, Wisconsin, and Actor. The eigenvalue decomposition is performed on $\tilde{L}$ of individual datasets. The corresponding eigenvalues are divided into two sets $\lambda_{<1}$ and $\lambda_{\geq 1}$ as mentioned earlier. We estimated the percentage of the low frequencies and high frequencies of each dataset. Refer to Figure 6 for the detailed illustration. The figure delineates that the number of lower and higher

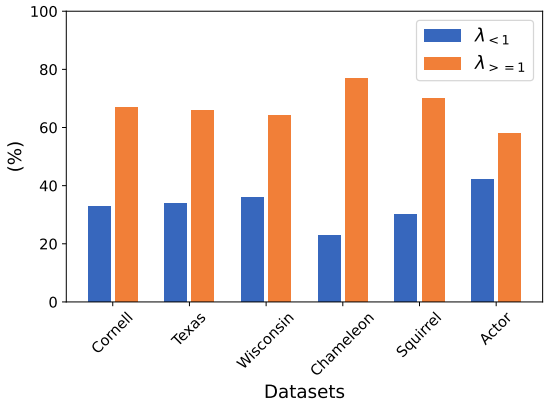
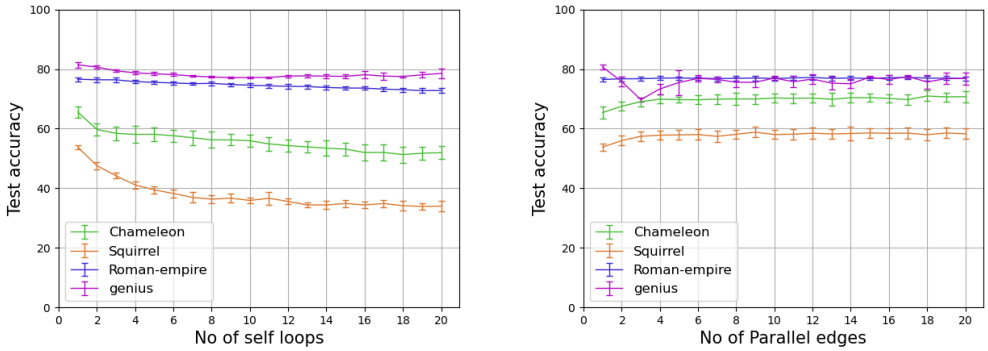


Figure 6: A comparative study on the number of low frequencies and high frequencies of the graph spectrum for six standard heterophilic graphs.

frequencies for Cornell, Texas, and Wisconsin are identical. In the experiments, their category is also similar which is "A". Concurrently, the Chameleon and Squirrel have similar patterns of the parity of eigenvalues and they also belong to a similar category "C". Furthermore, the pattern for the Actor is completely different from the rest of the others, and ends up marked as "B". The study established the unique interconnectedness of the parity of eigenvalues, frequency distribution of spectrum, and performance trends of low-pass filters.



(a) (b)

Figure 7: The effect on the performance of GCN on Chameleon, Squirrel, Roman-empire, and genius when (a) self-loops and (b) parallel edges are added a higher number of times are presented.

4.10 Variation of both Self-loops and Parallel edges

A study is conducted to observe the effect on the performance of GCN with the variation of both self-loops and parallel edges. We consider the Chameleon as our candidate graph to serve the purpose. The number of self-loops and parallel edges is both varied from 1 to 5 taking into account 25 combinations. GCN is applied to the modified Chameleon graph for each combination. The mean test accuracy with standard deviation estimated over 10 splits is reported against each combination. Refer to Table 6 to get a detailed illustration of the results. The uptick trend is noted for the increasing number of parallel edges in the network. While increasing the number of self-loops, a degrade in the model performance is observed. For a fixed number of self-loops, the test accuracy increases with the addition of parallel edges irrespective of the beginning of the initial performance. Every combination of (α_i, γ_j) produces a filter with the adjusted lower and higher number of frequencies. Notably, the lowest performance is achieved when the number of self-loops is highest ($\alpha = 5$) and the number of parallel edges is lowest ($\gamma = 1$). The best performance is obtained with just the reverse settings like the lowest number of self-loops ($\alpha = 1$) and the highest number of parallel edges ($\gamma = 5$).

4.11 Effect on Performance with Very Large α and γ

We performed a study on the Chameleon, Squirrel, Roman-empire, and genius by increasing α and γ to the higher values to monitor the performance of GCN. The number of both α, γ are varied from 1 to 20, and corresponding test accuracy with standard deviations are demonstrated in Figure 7. The study indicates the maintenance of the performance trends of the different datasets. The trend mostly depends on the input graph. For example, GCN on Chameleon showed decreasing (increasing) in the increasing number of self-loops (parallel edges). Conversely, GCN on genius exhibited a downtrend in the performance by adding both self-loops and parallel edges. It is noteworthy that performance stabilized with the higher value of α or γ. Empirical observation points out that the change in eigenvalues will become comparably negligible when the value of α or γ exceeds a certain limit. This phenomenon can be attributed to the saturating performance trends observed in the experiment.

5 Conclusion & Future Works

We conduct detailed studies regarding the performance of low-pass filters like GCN on the heterophilic graphs. The studies revealed that GCN showed monotone performance trends when the input graph is equipped with an increasing number of self-loops or parallel edges. We categorize the input graphs into four distinct categories based on these performance trends. We further observed that eigenvalues of normalized graph Laplacian decrease and increase when self-loops or parallel edges are respectively added. Consequently,

each category entails a significant amount of information pertaining to the characteristics of the graph spectrum. Therefore, the performance trends depend solely on the distribution of the eigenvalues in the entire spectrum. The graph spectrum patterns reveal the graph's intrinsic characteristics such as connected components, community structure, etc. Our work manifests the cost-effective pathway for estimating and understanding intrinsic properties and intricate patterns in the graph data, refraining from performing expensive computations. The designing of effective application of GNNs to replace the prevailing costly algorithmic computations can be the potential avenue for future research directions.

References

Filippo Maria Bianchi, Daniele Grattarola, Lorenzo Livi, and Cesare Alippi. Graph neural networks with convolutional arma filters. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3496–3507, 2021.

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 3950–3957, 2021.

Zhiqian Chen, Fanglan Chen, Lei Zhang, Taoran Ji, Kaiqun Fu, Liang Zhao, Feng Chen, Lingfei Wu, Charu Aggarwal, and Chang-Tien Lu. Bridging the gap between spatial and spectral domains: A unified framework for graph neural networks. *ACM Computing Surveys*, 56(5):1–42, 2023.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Yushun Dong, Kaize Ding, Brian Jalaian, Shuiwang Ji, and Jundong Li. Adagnn: Graph neural networks with adaptive frequency response filter. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pp. 392–401, 2021.

William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 1025–1035, 2017.

Dongxiao He, Chundong Liang, Huixin Liu, Mingxiang Wen, Pengfei Jiao, and Zhiyong Feng. Block modeling-guided graph convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4022–4029, 2022.

Kedar Karhadkar, Pradeep Kr Banerjee, and Guido Montúfar. Fosr: First-order spectral rewiring for addressing oversquashing in gnns. *arXiv preprint arXiv:2210.11790*, 2022.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. *Advances in Neural Information Processing Systems*, 34:20887–20902, 2021.

Hoang Nt and Takanori Maehara. Revisiting graph neural networks: All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*, 2019.

Antonio Ortega. *Introduction to graph signal processing.* Cambridge University Press, 2022.

Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, 2018.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287*, 2020.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? *arXiv preprint arXiv:2302.11640*, 2023.

Aliaksei Sandryhaila and Jose MF Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on signal processing*, 62(12):3042–3054, 2014.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

Rahul Singh and Yongxin Chen. Signed graph neural networks: A frequency perspective. *Transactions on Machine Learning Research*, 2023.

Susheel Suresh, Vinith Budde, Jennifer Neville, Pan Li, and Jianzhu Ma. Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns. *arXiv preprint arXiv:2106.06586*, 2021.

Shyam A Tailor, Felix L Opolka, Pietro Lio, and Nicholas D Lane. Do we need anisotropic graph neural networks? *arXiv preprint arXiv:2104.01481*, 2021.

Nicolas Tremblay, Paulo Gonçalves, and Pierre Borgnat. Design of graph filters and filterbanks. In *Cooperative and Graph Signal Processing*, pp. 299–324. Elsevier, 2018.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Tao Wang, Di Jin, Rui Wang, Dongxiao He, and Yuxiao Huang. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4210–4218, 2022.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pp. 6861–6871. PMLR, 2019.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Beyond low-pass filtering: Graph convolutional networks with automatic filtering. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6687–6697, 2022.

Haobo Xu, Yuchen Yan, Dingsu Wang, Zhe Xu, Zhichen Zeng, Tarek F Abdelzaher, Jiawei Han, and Hanghang Tong. Slog: An inductive spectral graph neural network beyond polynomial filter. In *Forty-first International Conference on Machine Learning*.

Zhiqiang Zhong, Sergei Ivanov, and Jun Pang. Simplifying node classification on heterophilous graphs with compatible label propagation. *Transactions on Machine Learning Research (TMLR)*, 2022.

Hao Zhu and Piotr Koniusz. Simple spectral graph convolution. In *International conference on learning representations*, 2021.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in Neural Information Processing Systems*, 33:7793–7804, 2020.

Jiong Zhu, Ryan A Rossi, Anup Rao, Tung Mai, Nedim Lipka, Nesreen K Ahmed, and Danai Koutra. Graph neural networks with heterophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11168–11176, 2021.

A Appendix

A.1 Proofs

This section will offer the detailed proofs and necessary derivations for Lemma 1, Lemma 2, Lemma 3, Lemma 4, Theorem 1, Theorem 2, Theorem **??**, Theorem **??**, and Corollary 2. Before delving into the detailed proofs and derivations, the following Lemma will assist in proving the following theorems. The following proposition is adopted from (Wu et al., 2019).

Proposition 1. *Let us assume $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ are the eigenvalues of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_n$ are the eigenvalues of $\tilde{D}^{-\frac{1}{2}}A\tilde{D}^{-\frac{1}{2}}$ where $\tilde{D} = D + \alpha I$, then we have the following inequalities*

$$\delta_1 \geq \frac{max_i d_i}{\alpha + max_i d_i}\beta_1, \qquad \delta_n \leq \frac{min_i d_i}{\alpha + min_i d_i}. \tag{2}$$

Proof. Recall that $L_{\text{sym}} = \text{I} - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and a well-known fact is that 0 is an eigenvalue of L_{sym}. Therefore, we have $\beta_n = 1$. As A is free of self-loops then $\text{Tr}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}) = 0 = \sum_i \beta_i$ which implies $\beta_1 < 0$.

Choose x such that $||x|| = 1$ and consider $y = D^{\frac{1}{2}}\tilde{D}^{-\frac{1}{2}}x$. Now, $||y||^2 = \sum_i \frac{d_i}{d_i+\alpha}x_i^2$. Also, we have $\frac{\min_i d_i}{\alpha+\min_i d_i} \leq ||y||^2 \leq \frac{\max_i d_i}{\alpha+\max_i d_i}$.

Applying the Rayleigh quotient we have the following bound for the smallest eigenvalue α_1,

$$\begin{aligned}
\delta_1 &= \min_{||x||=1}(x^\top \tilde{D}^{-\frac{1}{2}}A\tilde{D}^{-\frac{1}{2}}x) \\
&= \min_{||x||=1}(y^\top D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y) \qquad \text{(by variable substitution)} \\
&= \min_{||x||=1}(\frac{y^\top D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y}{||y||^2}||y||^2) \\
&\geq \min_{||x||=1}(\frac{y^\top D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y}{||y||^2}) \max_{||x||=1}(||y||^2) \\
&(\because \min(f(z)g(z)) \geq \min(f(z))\max(g(z)) \text{ if} \\
&\min(f(z)) < 0, \forall g(z) > 0 \\
&\text{and } \min_{||x||=1}(\frac{y^\top D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y}{||y||^2}) = \beta_1 < 0) \\
&= \beta_1 \max_{||x||=1}||y||^2 \\
&\geq \frac{\max_i d_i}{\alpha + \max_i d_i}\beta_1
\end{aligned} \tag{3}$$

Similarly, the upper bound for δ_n can be proved as $\delta_n \leq \frac{\min_i d_i}{\alpha+\min_i d_i}$. □

Lemma 1 Consider a graph G with A and D as the adjacency and degree matrix. Now α-times self-loops are added in G with $\alpha_1 \in \mathbb{Z}^+$. Assume $\lambda_\alpha^{\max}$ is the maximum eigenvalue of symmetrically normalized graph Laplacian $\tilde{L}_\alpha$ of the updated graph. If β_1 is the smallest eigenvalue of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\max_i d_i$ is the maximum degree of G, then $\lambda_\alpha^{\max} \leq \frac{\max_i d_i(1-\beta_1)}{\alpha+\max_i d_i}$.

Proof. After applying α-times self-loops the symmetrically normalized Laplacian is presented as:

$$\begin{aligned}
\tilde{L}_\alpha &= \text{I} - \tilde{D}_\alpha^{-\frac{1}{2}}\tilde{A}_\alpha\tilde{D}_\alpha^{-\frac{1}{2}} \\
&= \text{I} - \tilde{D}_\alpha^{-\frac{1}{2}}(A + \alpha\text{I})\tilde{D}_\alpha^{-\frac{1}{2}} \\
&= \text{I} - \tilde{D}_\alpha^{-\frac{1}{2}}A\tilde{D}_\alpha^{-\frac{1}{2}} - \alpha\tilde{D}_\alpha^{-1}
\end{aligned} \tag{4}$$

Let $\lambda_\alpha^{\max}$ is the maximum eigenvalue of $\tilde{L}_\alpha$ and applying Rayleigh quotient the following can be obtained

$$\begin{aligned}
\lambda_\alpha^{\max} &= \max_{||x||=1} x^\top \tilde{L}_\alpha x \\
&= \max_{||x||=1} x^\top (\mathrm{I} - \tilde{D}_\alpha^{-\frac{1}{2}} A \tilde{D}_\alpha^{-\frac{1}{2}} - \alpha \tilde{D}_\alpha^{-1}) x \\
&\leq (1 - \min_{||x||=1} x^\top \tilde{D}_\alpha^{-\frac{1}{2}} A \tilde{D}_\alpha^{-\frac{1}{2}} x - \min_{||x||=1} \alpha x^\top \tilde{D}_\alpha^{-1} x) \\
&= 1 - \delta_1 - \frac{\alpha}{\alpha + \max_i d_i} \qquad \text{(from Proposition 1)} \\
&\leq 1 - \frac{\max_i d_i}{\alpha + \max_i d_i} \beta_1 - \frac{\alpha}{\alpha + \max_i d_i} \\
&= \frac{\max_i d_i (1 - \beta_1)}{\alpha + \max_i d_i}
\end{aligned} \tag{5}$$

When α increases the upper bound of $\lambda_\alpha^{\max}$ decreases which indicates the possible shrinking of the maximum eigenvalue of the graph spectrum.

□

Lemma 2 Consider a graph G with A and D as the adjacency and degree matrix. Now γ-times self-loops are added in G with $\gamma \in \mathbb{Z}^+$. Assume $\lambda_\gamma^{\max}$ is the maximum eigenvalue of symmetrically normalized graph Laplacian $\tilde{L}_\gamma$ of the updated graph. If β_1 is the smallest eigenvalue of $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ and $\max_i d_i$ is the maximum degree of G, then $\lambda_\gamma^{\max} \leq \frac{(1+\gamma)\max_i d_i (1-\beta_1)}{1+(1+\gamma)\max_i d_i}$.

Proof. After applying γ-times parallel edges in the graph, the symmetrically normalized graph Laplacian will be

$$\begin{aligned}
\tilde{L}_\gamma &= \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}} \\
&= \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}} ((\gamma+1)A + \mathrm{I}) \tilde{D}_\gamma^{-\frac{1}{2}} \\
&= \mathrm{I} - (\gamma+1) \tilde{D}_\gamma^{-\frac{1}{2}} A \tilde{D}_\gamma^{-\frac{1}{2}} - \tilde{D}_\gamma^{-1}
\end{aligned} \tag{6}$$

Let $\lambda_\gamma^{\max}$ is the maximum eigenvalue of $\tilde{L}_\gamma$ and applying Rayleigh quotient the following can be obtained

$$\begin{aligned}
\lambda_\gamma^{\max} &= \max_{||x||=1} x^\top \tilde{L}_\gamma x \\
&= \max_{||x||=1} x^\top (\mathrm{I} - (\gamma+1) \tilde{D}_\gamma^{-\frac{1}{2}} A \tilde{D}_\gamma^{-\frac{1}{2}} - \tilde{D}_\gamma^{-1}) x \\
&\leq (1 - (\gamma+1) \min_{||x||=1} x^\top \tilde{D}_\gamma^{-\frac{1}{2}} A \tilde{D}_\gamma^{-\frac{1}{2}} x - \min_{||x||=1} x^\top \tilde{D}_\gamma^{-1} x) \\
&= 1 - (\gamma+1)\delta_1 - \frac{1}{1 + \gamma \max_i d_i} \text{ (from Proposition 1)} \\
&\leq 1 - \frac{(\gamma+1)\max_i d_i}{1 + (1+\gamma)\max_i d_i} \beta_1 - \frac{1}{1 + (1+\gamma)\max_i d_i} \\
&= \frac{(1+\gamma)\max_i d_i (1 - \beta_1)}{1 + (1+\gamma)\max_i d_i}
\end{aligned} \tag{7}$$

When γ increases the upper bound of $\lambda_\gamma^{\max}$ increases which shows the possible expansion of the maximum eigenvalue of the graph spectrum.

□

Lemma 3 Given a k-regular graph $\mathcal{G}$, the eigenvalues of $\tilde{A}_N^\alpha$ will lie in $[-1, 1]$ $\forall \alpha \geq 1$.

Proof. Consider a k-regular graph $\mathcal{G}$ where each node has a degree k with the normalized adjacency matrix is $\tilde{A}_N = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ where $\tilde{A} = A + I, \tilde{D} = D + I$. If α-times (with $\alpha \geq 1$) self-loops are added, then the updated normalized adjacency matrix is $\tilde{A}_N^{\alpha} = \tilde{D}_{\alpha}^{-\frac{1}{2}}\tilde{A}_{\alpha}\tilde{D}_{\alpha}^{-\frac{1}{2}}$ where $\tilde{A}_{\alpha} = A + \alpha I, \tilde{D}_{\alpha} = D + \alpha I$. As the graph is regular then $\tilde{A}_N$ can be presented as:

$$\begin{aligned}\tilde{A}_N &= \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{k+1}}(A+I)\frac{1}{\sqrt{k+1}} \\ &= \frac{1}{k+1}(A+I)\end{aligned} \tag{8}$$

In a similar fashion, we can express $\tilde{A}_N^{\alpha} = \frac{1}{k+\alpha}(A + \alpha I)$. Since, both $\tilde{A}_N$ and $\tilde{A}_N^{\alpha}$ are the linearly scaled transformations of A, then both will share a similar set of eigenvectors with different scaled eigenvalues. Assume v is an eigenvector of $\tilde{A}_N$ associated with any eigenvalue λ_1. Therefore,

$$\begin{aligned}&\tilde{A}_N v = \lambda_1 v \\ &\frac{1}{k+1}(A+I)v = \lambda_1 v \\ &Av = ((k+1)\lambda_1 - 1)v\end{aligned} \tag{9}$$

We can say that v is an eigenvector of A with corresponding eigenvalue $\lambda_A = ((k+1)\lambda_1 - 1)$. Consider the eigenvector v of $\tilde{A}_N^{\gamma}$ with the with the eigenvalue λ_2. Then, we have the following,

$$\begin{aligned}&\tilde{A}_N^{\alpha} v = \lambda_2 v \\ &\frac{1}{k+\alpha}(A+\alpha I)v = \lambda_2 v \\ &\lambda_2 = \frac{\lambda_A + \alpha}{k+\alpha}\end{aligned} \tag{10}$$

As the range of eigenvalues of $\tilde{A}_N$ is $[-1, 1]$, thus we have $\lambda_1 \leq 1$. The following inequality can be expressed,

$$\begin{aligned}&\lambda_1 \leq 1 \\ &(k+1)\lambda_1 \leq k+1 \\ &(k+1)\lambda_1 - 1 \leq k+1-1 \\ &\lambda_A \leq k\end{aligned} \tag{11}$$

Using the inequality we will prove the next stage as

$$\begin{aligned}&\lambda_A \leq k \\ &\alpha + \lambda_A \leq \alpha + k \\ &\frac{\alpha + \lambda_A}{k+\alpha} \leq \frac{\alpha + k}{\alpha + k} \\ &\lambda_2 \leq 1\end{aligned} \tag{12}$$

In this way, we showed that any eigenvalue of $\tilde{A}_N^{\alpha}$ is 1. For the lower bound, we can write,

$$\begin{aligned}&\lambda_1 \geq -1 \\ &(k+1)\lambda_1 \geq -k-1 \\ &(k+1)\lambda_1 - 1 \geq -k-1-1 \\ &\lambda_A \geq -k-2\end{aligned} \tag{13}$$

Using the inequality we will prove the next stage as,

$$\begin{aligned}
&\lambda_A \geq -k-2 \\
&\alpha + \lambda_A \geq \alpha - k - 2 \\
&\frac{\alpha+\lambda_A}{k+\alpha} \geq \frac{\alpha-k-2}{\alpha+k} \\
&\lambda_2 \geq 1 - \frac{2(k+1)}{k+\alpha}
\end{aligned} \tag{14}$$

The degree $k > 1$, then we have $\lambda_2 \geq -1$. Therefore, the addition of self-loops will not alter the range of the eigenvalues of the symmetrically normalized adjacency matrix. □

Lemma 4 Given a k-regular graph $\mathcal{G}$, the eigenvalues of $\tilde{A}_N^{\gamma}$ will lie in $[-1, 1]$ $\forall \gamma \geq 1$.

Proof. Consider a k-regular graph $\mathcal{G}$ where each node has a degree k with the normalized adjacency matrix is $\tilde{A}_N = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ where $\tilde{A} = A + I, \tilde{D} = D + I$. If γ-times (with $\gamma \geq 1$) parallel edges are added, then the updated normalized adjacency matrix is $\tilde{A}_N^{\gamma} = \tilde{D}_{\gamma}^{-\frac{1}{2}}\tilde{A}_{\gamma}\tilde{D}_{\gamma}^{-\frac{1}{2}}$ where $\tilde{A}_{\gamma} = (1+\gamma)A + I, \tilde{D}_{\gamma} = (1+\gamma)D + I$. As the graph is regular then A_N can be presented as:

$$\begin{aligned}
\tilde{A}_N &= \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} \\
&= \frac{1}{\sqrt{k+1}}(A+I)\frac{1}{\sqrt{k+1}} \\
&= \frac{1}{k+1}(A+I)
\end{aligned} \tag{15}$$

In a similar fashion, we can express $\tilde{A}_N^{\gamma} = \frac{1}{1+(1+\gamma)k}((1+\gamma)A + I)$. Since, both $\tilde{A}_N$ and $\tilde{A}_N^{\gamma}$ are the linearly scaled transformations of A, then both will share a similar set of eigenvectors with different scaled eigenvalues. Assume v is an eigenvector of $\tilde{A}_N$ associated with any eigenvalue λ_1. Therefore,

$$\begin{aligned}
&\tilde{A}_N v = \lambda_1 v \\
&\frac{1}{k+1}(A+I)v = \lambda_1 v \\
&Av = ((k+1)\lambda_1 - 1)v
\end{aligned} \tag{16}$$

We can say that v is an eigenvector of A with corresponding eigenvalue $\lambda_A = ((k+1)\lambda_1 - 1)$. Consider the eigenvector v of $\tilde{A}_N^{\gamma}$ with the with the eigenvalue λ_2. Then, we have the following,

$$\begin{aligned}
&\tilde{A}_N^{\gamma} v = \lambda_2 v \\
&\frac{1}{1+(1+\gamma)k}((1+\gamma)A + I)v = \lambda_2 v \\
&\lambda_2 = \frac{(1+\gamma)\lambda_A + 1}{(1+\gamma)k + 1}
\end{aligned} \tag{17}$$

As the range of eigenvalues of $\tilde{A}_N$ is $[-1, 1]$, thus we have $\lambda_1 \leq 1$. The following inequality can be expressed,

$$\begin{aligned}
&\lambda_1 \leq 1 \\
&(k+1)\lambda_1 \leq k+1 \\
&(k+1)\lambda_1 - 1 \leq k+1-1 \\
&\lambda_A \leq k
\end{aligned} \tag{18}$$

Using the inequality we will prove the next stage as

$$\begin{aligned}
&\lambda_A \leq k \\
&(1+\gamma)\lambda_A \leq (1+\gamma)k \\
&(1+\gamma)\lambda_A \leq (1+\gamma)k \\
&(1+\gamma)\lambda_A + 1 \leq (1+\gamma)k + 1 \\
&\frac{(1+\gamma)\lambda_A + 1}{(1+\gamma)k+1} \leq 1 \\
&\lambda_2 \leq 1
\end{aligned} \tag{19}$$

In this way, we have shown the maximum eigenvalue of $\tilde{A}_N^{\gamma}$ is 1. The lower bound of the eigenvalues can be shown as

$$\begin{aligned}
&\lambda_1 \geq -1 \\
&(k+1)\lambda_1 \geq -k-1 \\
&(k+1)\lambda_1 - 1 \geq -k-1-1 \\
&\lambda_A \leq -k-2
\end{aligned} \tag{20}$$

Using the inequality we will prove the next stage as

$$\begin{aligned}
&\lambda_A \leq -k-2 \\
&(1+\gamma)\lambda_A \geq (1+\gamma)(-k-2) \\
&(1+\gamma)\lambda_A + 1 \geq (1+\gamma)(-k-2) + 1 \\
&\frac{(1+\gamma)\lambda_A + 1}{(1+\gamma)k+1} \geq \frac{(1+\gamma)(-k-2)+1}{(1+\gamma)k+1} \\
&\lambda_2 \geq 1 - \frac{2(k+1)(\gamma+1)}{(1+\gamma)k+1}
\end{aligned} \tag{21}$$

As $k, \gamma > 0$, then we have $\lambda_2 \geq -1$. Therefore, the addition of parallel edges will not alter the range of the eigenvalues of the symmetrically normalized adjacency matrix. □

Theorem 1 Consider a k-regular graph with $\alpha_1, \alpha_2 \in \mathbb{R}^+$ with $\alpha_1 \leq \alpha_2$, then $\lambda^i_{\alpha_1} \geq \lambda^i_{\alpha_2}, \forall\ 1 \leq i \leq n$, where $\lambda^i_{\alpha_1}$ and $\lambda^i_{\alpha_2}$ are the ith eigenvalues of $\tilde{L}_{\alpha_1}$ and $\tilde{L}_{\alpha_2}$ respectively.

Proof. Consider a k-regular graph $\mathcal{G}$ where each node of degree k with the normalized adjacency matrix is $A_N = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ where $\tilde{A} = A + I, \tilde{D} = D + I$. If α-times (with $\alpha \geq 1$) self-loops are added, then the updated normalized adjacency matrix is $\tilde{A}^{\alpha}_N = \tilde{D}_{\alpha}^{-\frac{1}{2}}\tilde{A}_{\alpha}\tilde{D}_{\alpha}^{-\frac{1}{2}}$ where $\tilde{A}_{\alpha} = A + \alpha I, \tilde{D}_{\alpha} = D + \alpha I$. As the graph is regular then we can have the following

$$\begin{aligned}
A^{\alpha}_N &= \tilde{D}_{\alpha}^{-\frac{1}{2}}\tilde{A}_{\alpha}\tilde{D}_{\alpha}^{-\frac{1}{2}} \\
&= \frac{1}{\sqrt{k+\alpha}}(A+\alpha I)\frac{1}{\sqrt{k+\alpha}} \\
&= \frac{1}{(k+\alpha)}(A+\alpha I)
\end{aligned} \tag{22}$$

Therefore, we can express $A^{\alpha_1}_N = \frac{1}{k+\alpha_1}(A+\alpha_1 I)$ and $A^{\alpha_2}_N = \frac{1}{k+\alpha_2}((A+\alpha_2 I)$. Since, $A^{\alpha_1}_N$ and $A^{\alpha_2}_N$ are the linear transformation of A, both will have the same set of eigenvectors but with different eigenvalues. Assume v is the eigenvector of A and its corresponding eigenvalue is λ_{α_1}. Therefore, the following can be written as

$$\begin{aligned}
&A^{\alpha_1}_N v = \lambda_{\alpha_1} v \\
&\frac{1}{(k+\alpha_1)}(A+\alpha_1 I)v = \lambda_{\alpha_1} v \\
&Av = ((k+\alpha_1)\lambda_{\alpha_1} - \alpha_1)v
\end{aligned} \tag{23}$$

We can say that v is the eigenvector of A with the corresponding eigenvalue $\lambda_A = ((k+\alpha_1)\lambda_{\alpha_1} - \alpha_1)$. For $A_N^{\alpha_2}$, for eigenvector v, the corresponding eigenvalue is λ_{α_2}. Now, we have the following:

$$\begin{aligned} A_N^{\alpha_2} v &= \lambda_{\alpha_2} v \\ \frac{1}{(k+\alpha_2)}(A+\alpha_2 I)v &= \lambda_{\alpha_2} v \\ Av &= ((k+\alpha_2)\lambda_{\alpha_2} - \alpha_2)v \end{aligned} \tag{24}$$

Similarly, we can also express $\lambda_A = ((k+\alpha_2)\lambda_{\alpha_2} - \alpha_2)$. Now, equating the two different expressions of λ_A, we can express the following

$$\begin{aligned} &(k+\alpha_1)\lambda_{\alpha_1} - \alpha_1 = (k+\alpha_2)\lambda_{\alpha_2} - \alpha_2 \\ &k\lambda_{\alpha_1} - \alpha_1(1-\lambda_{\alpha_1}) = k\lambda_{\alpha_2} - \alpha_2(1-\lambda_{\alpha_2}) \\ &k(\lambda_{\alpha_1} - \lambda_{\alpha_2}) = \alpha_1(1-\lambda_{\alpha_1}) - \alpha_2(1-\lambda_{\alpha_2}) \\ &\text{As per provided condition, we have } \alpha_1 < \alpha_2 \\ &\alpha_1(1-\lambda_{\alpha_1}) < \alpha_2(1-\lambda_{\alpha_1}) \\ &\alpha_1(1-\lambda_{\alpha_1}) - \alpha_2(1-\lambda_{\alpha_2}) < \alpha_2(1-\lambda_{\alpha_1}) - \alpha_2(1-\lambda_{\alpha_2}) \\ &k(\lambda_{\alpha_1} - \lambda_{\alpha_2}) < \alpha_2(1-\lambda_{\alpha_1}) - \alpha_2(1-\lambda_{\alpha_2}) \\ &k(\lambda_{\alpha_1} - \lambda_{\alpha_2}) < \alpha_2(\lambda_{\alpha_2} - \lambda_{\alpha_1}) \\ &(k+\alpha_2)(\lambda_{\alpha_!} - \lambda_{\alpha_2}) < 0 \\ &\text{As } k, \alpha_2 > 0 \text{ then} \\ &\lambda_{\alpha_!} - \lambda_{\alpha_2} < 0 \\ &\lambda_{\alpha_!} < \lambda_{\alpha_2} \end{aligned}$$

The above equation holds for all $|\lambda_{\alpha_1}|, |\lambda_{\alpha_2}| \leq 1$ which is ensured from Lemma 3. The eigenvalue of $A_N^{\alpha_2}$ became greater than that of $A_N^{\alpha_1}$ when self-loops are added to the graph. We know that $\tilde{L}_\alpha = \mathrm{I} - \tilde{D}_\alpha^{-\frac{1}{2}} \tilde{A}_\alpha \tilde{D}_\alpha^{-\frac{1}{2}}$, indicates if the eigenvalue of A_N^α increases then the corresponding eigenvalue of $\tilde{L}_\alpha$ decreases. Therefore, it can be concluded that the eigenvalue of $\tilde{L}_\alpha$ decreases with the addition of self-loops. □

Theorem 2 Consider a k-regular graph with $\gamma_1, \gamma_2 \in \mathbb{R}^+$ with $\gamma_1 \leq \gamma_2$, then $\lambda_{\gamma_1}^i \leq \lambda_{\gamma_2}^i, \forall\ 1 \leq i \leq n$, where $\lambda_{\gamma_1}^i$ and $\lambda_{\gamma_2}^i$ are the ith eigenvalues of $\tilde{L}_{\gamma_1}$ and $\tilde{L}_{\gamma_2}$ respectively.

Proof. Consider a k-regular graph $\mathcal{G}$ where each node has a degree k with the normalized adjacency matrix is $A_N = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ where $\tilde{A} = A + I, \tilde{D} = D + I$. If γ-times (with $\gamma \geq 1$) parallel edges are added, then the updated normalized adjacency matrix is $\tilde{A}_N^\gamma = \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}}$ where $\tilde{A}_\gamma = (1+\gamma)A + I, \tilde{D}_\gamma = (1+\gamma)D + I$. As the graph is regular then we can have the following

$$\begin{aligned} A_N^\gamma &= \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{1+(1+\gamma)k}}((1+\gamma)A + I)\frac{1}{\sqrt{1+(\gamma+1)k}} \\ &= \frac{1}{1+(1+\gamma)k}((1+\gamma)A + I) \end{aligned} \tag{25}$$

Therefore, we can express $A_N^{\gamma_1} = \frac{1}{1+(1+\gamma_1)k}((1+\gamma_1)A+I)$ and $A_N^{\gamma_2} = \frac{1}{1+(1+\gamma_2)k}((1+\gamma_2)A+I)$. Since, $A_N^{\gamma_1}$ and $A_N^{\gamma_2}$ are the scalar transformation of A, both will have the same set of eigenvectors with different eigenvalues. Assume v is the eigenvector of A and its corresponding eigenvalue is λ_{γ_1}. Therefore, the following can be written as

$$\begin{aligned} A_N^{\gamma_1} v &= \lambda_{\gamma_1} v \\ \frac{1}{1+(1+\gamma_1)k}((1+\gamma_1)A + I)v &= \lambda_{\gamma_1} v \\ Av &= \frac{(1+(1+\gamma_1)k)\lambda_{\gamma_1} - 1}{1+\gamma_1} v \end{aligned} \tag{26}$$

We can say that v is the eigenvector of A with the corresponding eigenvalue $\lambda_A = \frac{(1+(1+\gamma_1)k)\lambda_{\gamma_1}-1}{1+\gamma_1}$. For A_N^{γ}, the eigenvector is v with the eigenvalue λ_2. Then, we have the following

$$\begin{aligned} &A_N^{\gamma_2} v = \lambda_{\gamma_2} v \\ &\frac{1}{1+(1+\gamma_2)k}((1+\gamma_2)A + I)v = \lambda_{\gamma_2} v \\ &Av = \frac{(1+(1+\gamma_2)k)\lambda_{\gamma_2} - 1}{1+\gamma_2} v \end{aligned} \tag{27}$$

We can also express $\lambda_A = \frac{(1+(1+\gamma_2)k)\lambda_{\gamma_2}-1}{1+\gamma_2}$. Now, equating the two different values of λ_A, we can have the following

$$\begin{aligned} &\frac{(1+(1+\gamma_1)k)\lambda_{\gamma_1} - 1}{1+\gamma_1} = \frac{(1+(1+\gamma_2)k)\lambda_{\gamma_2} - 1}{1+\gamma_2} \\ &(1+\gamma_2)((1+(1+\gamma_1)k)\lambda_{\gamma_1} - 1) = \\ &\qquad (1+\gamma_1)(1+(1+\gamma_2)k)\lambda_{\gamma_2} - 1) \\ &(1+\gamma_2)(1+(1+\gamma_1)k)\lambda_{\gamma_1} - (1+\gamma_2) = \\ &\qquad (1+\gamma_1)(1+(1+\gamma_2)k)\lambda_{\gamma_2} - (1+\gamma_1) \\ &(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)\lambda_{\gamma_1} - 1 - \gamma_2 = \\ &\qquad (1+\gamma_1+(1+\gamma_1)(1+\gamma_2)k)\lambda_{\gamma_2} - 1 - \gamma_1 \\ &(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)\lambda_{\gamma_1} = \\ &\qquad (\gamma_2-\gamma_1) + (1+\gamma_1+(1+\gamma_1)(1+\gamma_2)k)\lambda_{\gamma_2} \\ &\lambda_{\gamma_1} = \frac{\gamma_2-\gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)} + \\ &\qquad \frac{(1+\gamma_1+(1+\gamma_1)(1+\gamma_2)k)}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)}\lambda_{\gamma_2} \\ &= \frac{\gamma_2-\gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)} + \\ &\qquad \frac{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k) - \gamma_2 + \gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)}\lambda_{\gamma_2} \\ &= \frac{\gamma_2-\gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)} + \\ &\qquad (1 - \frac{\gamma_2-\gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)})\lambda_{\gamma_2} \\ &= \frac{\gamma_2-\gamma_1}{(1+\gamma_2+(1+\gamma_1)(1+\gamma_2)k)}(1-\lambda_{\gamma_2}) + \lambda_{\gamma_2} \\ &\text{According to the condition provided } \gamma_2 > \gamma_1 \text{ we can say} \\ &\lambda_{\gamma_1} > \lambda_{\gamma_2} \end{aligned} \tag{28}$$

The eigenvalue of $A_N^{\gamma_2}$ became lesser than that of $A_N^{\gamma_1}$ with the addition of parallel edges. The Eq. 28 holds for $|\lambda_{\gamma_2}| \leq 1$ which is assured from Lemma 4. We know that $\tilde{L}_\gamma = \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}}$, indicates if the eigenvalue of A_N^{γ} decreases then the corresponding eigenvalue of $\tilde{L}_\gamma$ increases. Therefore, it can be concluded that the eigenvalue of $\tilde{L}_\gamma$ increases with the addition of parallel edges for the regular graphs. □

Corollary 2. *The increase in the eigenvalues of L_γ is independent of the number of self-loop additions in $\mathcal{G}$. On the contrary, The eigenvalues of $\tilde{L}_\gamma$ will increase if at least one self-loop is added per node in $\mathcal{G}$.*

Proof. Let us prove the statement by contradiction. We know $L = D - A$ and after adding α-times self-loops and γ-times parallel edges, the $L_\gamma = D_\gamma - A_\gamma$ where $\tilde{A}_\gamma = (1+\gamma)A + \alpha I, \tilde{D}_\gamma = (\gamma+1)D + \alpha I$. Then,

$$\begin{aligned} L_\gamma &= (\tilde{D}_\gamma - \tilde{A}_\gamma) \\ &= (((\gamma+1)D + \alpha I) - ((\gamma+1)A + \alpha I)) \\ &= (\gamma+1)(D - A) \\ &= (\gamma+1)L \end{aligned} \tag{29}$$

The equation is independent of the number of self-loops we confirm that the effect of adding parallel edges prevails. Similarly, let us also prove the second part by contradiction. We know that $\tilde{L}_\gamma = \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}}$ and consider the expression without self-loops with the addition of γ-times parallel edges as $\tilde{D} = (1+\gamma)D, \tilde{A} = (1+\gamma)A$.

$$\begin{aligned} \tilde{L}_\gamma &= \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}} \tilde{A}_\gamma \tilde{D}_\gamma^{-\frac{1}{2}} \\ &= \mathrm{I} - \frac{1}{\sqrt{(1+\gamma)}} D^{-\frac{1}{2}} (1+\gamma) A \frac{1}{\sqrt{(1+\gamma)}} D^{-\frac{1}{2}} \\ &= \mathrm{I} - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \\ &= \tilde{L} \end{aligned} \tag{30}$$

Therefore, adding parallel edges without at least one self-loop per node does not change the normalized graph Laplacian. Hence, the result is proved. □

Theorem 3 Consider a connected graph $\mathcal{G}$ with $A_N = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. Assuming the diagonal of A of G is perturbed by a significantly small $\alpha > 0$, then the updated normalized adjacency matrix will be A_N^α. The change in the eigenvalues of A_N^α with respect to eigenvalues of A_N will increase when α increases.

Proof. If the diagonal of A is perturbed by α, the symmetrically normalized graph Laplacian of G will be,

$$\tilde{L}_\alpha = \mathrm{I} - \tilde{D}_\alpha^{-\frac{1}{2}} \tilde{A}_\alpha \tilde{D}_\alpha^{-\frac{1}{2}}, \tag{31}$$

where $\tilde{A}_\alpha = A + \alpha\mathrm{I}$ and $\tilde{D}_\alpha = D + \alpha\mathrm{I}$. Let us denote the normalized adjacency matrix without self-loops as $A_N = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. The element of A_N is represented as:

$$A_N[i][j] = \begin{cases} \frac{A_{ij}}{\sqrt{d_i d_j}}, & i \neq j \\ 0, & i = j \end{cases} \tag{32}$$

After perturbed by α, the normalized adjacency will be $A_N^\alpha = \tilde{D}_\alpha^{-\frac{1}{2}} \tilde{A}_\alpha \tilde{D}_\alpha^{-\frac{1}{2}}$. The elements of $\tilde{A}_N^\alpha$ can be represented as:

$$A_N^\alpha[i][j] = \begin{cases} \frac{A_{ij}}{\sqrt{\alpha+d_i}\sqrt{\alpha+d_j}} & i \neq j \\ \frac{\alpha}{\alpha+d_i} & i = j \end{cases} \tag{33}$$

The entry-wise change in the normalized adjacency matrix is presented as:

$$\delta A_N[i][j] = \begin{cases} \frac{A_{ij}}{\sqrt{\alpha+d_i}\sqrt{\alpha+d_j}} - \frac{A_{ij}}{\sqrt{d_i d_j}}, & i \neq j \\ \frac{\alpha}{\alpha+d_i}, & i = j \end{cases} \tag{34}$$

Following the notion of the Theorem 4 stated in (Karhadkar et al., 2022) we can assume x is a normalized eigenvector of A_N with corresponding eigenvalue λ. Therefore, the first-order change in the corresponding

eigenvalue can be represented as:

$$
\begin{aligned}
x^\top(\delta A_N)x &= \sum_{i\neq j}\delta A_N[i][j]x_ix_j + \sum_{i=j}\delta A_N[i][j]x_i^2 \\
&= \sum_{i\neq j}\left(\frac{A_{ij}}{\sqrt{\alpha+d_i}\sqrt{\alpha+d_j}} - \frac{A_{ij}}{\sqrt{d_id_j}}\right)x_ix_j \\
&\quad + \sum_{i=j}\frac{\alpha}{\alpha+d_i}x_i^2 \\
&= \sum_{i\neq j}\left(\frac{1}{\sqrt{\alpha+d_i}\sqrt{\alpha+d_j}} - \frac{1}{\sqrt{d_id_j}}\right)A_{ij}x_ix_j \\
&\quad + \sum_{i=j}\frac{\alpha}{\alpha+d_i}x_i^2
\end{aligned}
\tag{35}
$$

Consider,

$$
\begin{aligned}
F_{ij}^{(1)}(\alpha) &= \left(\frac{1}{\sqrt{\alpha+d_i}\sqrt{\alpha+d_j}} - \frac{1}{\sqrt{d_id_j}}\right) \\
F_i^{(2)}(\alpha) &= \frac{\alpha}{\alpha+d_i}x_i^2
\end{aligned}
\tag{36}
$$

If $d_i, d_j \gg \alpha$, then $F_{ij}^{(1)}(\alpha) \approx 0$ which lead to

$$
x^\top(\delta A_N)x \approx F_i^{(2)}(\alpha) \approx \sum_{i=j}\frac{\alpha}{\alpha+d_i}x_i^2. \tag{37}
$$

If α increases then $F_i^{(2)}(\alpha)$ also increases. This reflects the change in the eigenvalue of A_N^α decreases indicating the decreases in the eigenvalues of the $\tilde{L}_\alpha$. □

Theorem 4 Consider a connected graph $\mathcal{G}$ with normalized adjacency matrix $A_N = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. Assuming each element of A except the diagonal multiplied by $1+\gamma$ where $\gamma > 0$ is a significantly small quantity, then the updated normalized adjacency matrix will be A_N^γ. The change in the eigenvalues of A_N^γ with respect to eigenvalues of A_N will decrease when γ increases.

Proof. If the non-diagonal elements of A are multiplied by $1+\gamma$, then symmetrically normalized graph Laplacian will be,

$$
\tilde{L}_\gamma = \mathrm{I} - \tilde{D}_\gamma^{-\frac{1}{2}}\tilde{A}_\gamma\tilde{D}_\gamma^{-\frac{1}{2}}, \tag{38}
$$

where $\tilde{A}_\gamma = (\gamma+1)A + \mathrm{I}$ and $\tilde{D}_\gamma = (\gamma+1)D + \mathrm{I}$. Let us denote the normalized adjacency matrix without self-loops as $A_N = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. The element of A_N is represented as:

$$
A_N[i][j] = \begin{cases} \frac{A_{ij}}{\sqrt{d_id_j}}, & i \neq j \\ 0, & i = j \end{cases} \tag{39}
$$

After multiplying $(1+\gamma)$ to the non-diagonal elements of A and adding one self-loops, the normalized adjacency will be $A_N^\gamma = \tilde{D}_\gamma^{-\frac{1}{2}}\tilde{A}_\gamma\tilde{D}_\gamma^{-\frac{1}{2}}$. The elements of $\tilde{A}_N^\gamma$ can be represented as:

$$
A_N^\gamma[i][j] = \begin{cases} \frac{(\gamma+1)A_{ij}}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} & i \neq j \\ \frac{1}{1+(1+\gamma)d_i} & i = j \end{cases} \tag{40}
$$

The entry-wise change in the normalized adjacency matrix is presented as:

$$\delta A_N[i][j] = \begin{cases} \frac{(\gamma+1)A_{ij}}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} - \frac{A_{ij}}{\sqrt{d_i d_j}}, & i \neq j \\ \frac{1}{1+(1+\gamma)d_i}, & i = j \end{cases} \tag{41}$$

Following the notion of the Theorem 4 stated in (Karhadkar et al., 2022) we can assume x is a normalized eigenvector of A_N with corresponding eigenvalue λ. Therefore, the first-order change in the spectral gap can be represented as:

$$\begin{aligned} x^\top(\delta A_N)x &= \sum_{i\neq j} \delta A_N[i][j]x_i x_j + \sum_{i=j} \delta A_N[i][j]x_i^2 \\ &= \sum_{i\neq j}\left(\frac{(\gamma+1)A_{ij}}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} - \frac{A_{ij}}{\sqrt{d_i d_j}}\right)x_i x_j \\ &\quad + \sum_{i=j}\frac{1}{1+(1+\gamma)d_i}x_i^2 \\ &= \sum_{i\neq j}\left(\frac{(\gamma+1)}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} - \frac{1}{\sqrt{d_i d_j}}\right)A_{ij}x_i x_j \\ &\quad + \sum_{i=j}\frac{1}{1+(1+\gamma)d_i}x_i^2 \end{aligned} \tag{42}$$

Consider the following,

$$\begin{aligned} F_{ij}^{(1)}(\gamma) &= \left(\frac{(\gamma+1)}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} - \frac{1}{\sqrt{d_i d_j}}\right) \\ F_i^{(2)}(\gamma) &= \frac{1}{1+(1+\gamma)d_i}x_i^2 \end{aligned} \tag{43}$$

Now we can rewrite,

$$\begin{aligned} F_{ij}^{(1)}(\gamma) &= \left(\frac{(\gamma+1)}{\sqrt{1+(\gamma+1)d_i}\sqrt{1+(\gamma+1)d_j}} - \frac{1}{\sqrt{d_i d_j}}\right) \\ &= \left(\frac{1}{\sqrt{\frac{1}{1+\gamma}+d_i}\sqrt{\frac{1}{1+\gamma}+d_j}} - \frac{1}{\sqrt{d_i d_j}}\right) \end{aligned} \tag{44}$$

As we mentioned γ is a sufficiently small quantity and with $d_i, d_j \gg 1$, then $F_{ij}^{(1)}(\gamma) \approx 0$. Now we have,

$$x^\top(\delta A_N)x \approx F_i^{(2)}(\gamma) \approx \frac{1}{1+(1+\gamma)d_i}x_i^2 \tag{45}$$

If γ increases then $F_i^{(2)}(\gamma)$ decreases reflecting the change in the eigenvalues of $\tilde{A}_N^\gamma$ decreases. This indicates the increase in the eigenvalues of $\tilde{L}_N^\gamma$. □