LIMITLESS STABILITY FOR GRAPH CONVOLUTIONAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

This work establishes rigorous, novel and widely applicable stability guarantees and transferability bounds for general graph convolutional networks – without reference to any underlying limit object or statistical distribution. Crucially, utilized graph-shift operators are not necessarily assumed to be normal, allowing for the treatment of networks on both directed- and undirected graphs within the developed framework. In the undirected setting, stability to node-level perturbations is related to an 'adequate spectral covering' property of the filters in each layer. Stability to edge-level perturbations is discussed and related to properties of the utilized filters such as their Lipschitz constants. Results on stability to vertex-set non-preserving perturbations are obtained by utilizing recently developed mathematical-physics based tools. As an exemplifying application of the developed theory, it is showcased that general graph convolutional networks utilizing the un-normalized graph Laplacian as graph-shift-operator can be rendered stable to collapsing strong edges in the underlying graph if filters are mandated to be constant at infinity. These theoretical results are supported by corresponding numerical investigations showcasing the response of filters and networks to such perturbations.

1 INTRODUCTION

With increasing interest in the analysis of data on graph-structured domains, (semi-deep) networks generalizing Euclidean convolutional networks to this geometric setting emerged (Kipf & Welling, 2017; Bruna et al., 2014; Defferrard et al., 2016). These networks were termed Graph Convolutional Networks (GCNs) and – if efficiently implemented – replace Euclidean convolutional filters by functional calculus filters; i.e. scalar functions applied to a suitably chosen graph-shift-oprator capturing the geometry of the underlying graph (Kipf & Welling, 2017; Hammond et al., 2011; Defferrard et al., 2016). Similar to Euclidean convolutional networks – which find applications and yield state of the art results in many diverse domains such as image classification, speech recognition or natural language processing – graph convolutional architectures have been successfully applied to a wide variety of tasks such as node clustering (Chen et al., 2019), semi supervised learning (Kipf & Welling, 2017) or graph regression (Kearnes et al., 2016; Gilmer et al., 2017).

A key concept in trying to conceptually understand the underlying reason for such superior numerical performances (as well as a guiding principle for the design of new network architectures with desired properties) is the concept of **stability**. In the Euclidean setting – once a network is fully trained – investigating stability essentially amounts to exploring how the output of a convolutional network varies under non-trivial changes of its input-signals. While a complete 'stability-theory' is still out on the horizon, pioneering work – taking into account a diverse set of input-signal changes – on this subject matter was conducted in Mallat (2012) and further generalized in Wiatowski & Bölcskei (2018) which greatly aided the theoretical understanding of the superior performance of Euclidean convolutional networks. When transcending the established Euclidean setting and considering input-signals supported on general graphs – as opposed to regular grids – additional complications are introduced. Not only may input signals change, but now also the graph shift operators facilitating the convolutions on the graphs may vary. Even worse, there might also occur changes in the topology or vertex sets of the investigated graphs – e.g. when two dissimilar graphs describe the same underlying phenomenon – under which graph convolutional networks should also remain stable. This last stability property is often also referred to as transferability (Levie et al., 2019a).

Partially owing to the additional complexities in the graph setting – which often render traditional approaches of investigation mute – the understanding of stability properties of GCNs is far from being as developed as it is for Euclidean networks. Among the first works investigating stability under changes in graph-shift operators in a single-filter setting was Levie et al. (2019b) which essentially bounded filter responses to a change in graph shift operators by a constant times the change in shift operators for perturbations that are not too large and filters that are functions of the variable $\frac{x-i}{x+i}$. In Kenlay et al. (2021) the stability of GCNs based on normalized graph Laplacians under rewiring is investigated. In **??** stability for a certain class of filters under 'absolute' and 'relative' perturbation (introduced there) is established. Unlike the structural interpretations of stability of the aforementioned works, Bojchevski & Günnemann (2019) takes the approach of establishing robustness certificates for nodes in semi-supervised learning tasks. Taking into account changes in the graph topology, the work (Levie et al., 2019a) investigates the transferability of graph convolutional networks between graphs discretising the same underlying topological space. Similarly, the works (Ruiz et al., 2020; Maskey et al., 2021) consider transferability between graphs discretising the same underlying graphon. Finally there are also approaches to investigating stability on large and random graphs (e.g. Keriven et al. (2020) or Gao et al. (2021)).
Common among all these previous works are two themes limiting practical applicability of the derived results: First and foremost – when investigating stability under perturbations of the graph shift operator – the class of filters to which results are applicable is often severely limited. The same is true for the class of admissible graph shift operators, with non-normal operators (such as they typically appear on directed graphs) being either explicitly or implicitly excluded. Beyond that – when investigating transferability properties – results are almost exclusively available under the assumption that graphs either discretize the same underlying 'continuous' limit object (or are drawn from the same statistical distributions). While these are of course relevant regimes they are inapplicable to non-asymptotic settings. What is more, hardly any work has been done on relating the stability to input-signal perturbations to network properties such as (the interplay of) utilized filters, employed non-linearities or network topologies.
The main focus of this work is to provide alleviation in this situation and develop a 'general theory of stability' for GCNs; agnostic to the types of utilized filters, graph shift operators as well as non-linearities; with practically relevant transferability guarantees independent of any underlying limit objects. To this end Section 2 recapitulates the graph signal processing framework, the notion of functional calculus filters and the (general) architecture of graph convolutional networks. Sections 3 and 4 discuss stability to node- and edge-level perturbations respectively. Section 5 then discusses stability to structural perturbations changing graph topologies (i.e. transferability), for which Section 6 provides further numerical investigations and support.

2 Preliminaries and Framework:

Throughout this work, we will interchangeably use the labels G and $\widetilde{G}$ to denote both graphs and their associated vertex sets. The cardinality of the vertex set G is denoted by $|G|$. Taking a signal processing approach, we consider signals on graphs as opposed to graph embeddings:

Node-Signals: Node-signals on a graph are functions from the node-set G to the complex numbers; i.e. elements of $\mathbb{C}^{|G|}$. We allow nodes $g \in G$ in a given graph to have weights μ_g not necessarily equal to one and equip the space $\mathbb{C}^{|G|}$ with an inner product according to $\langle f, g\rangle = \sum_{g \in G}^{|G|} \overline{f}(g)g(g)\mu_g$ to account for this. We denote the hence created Hilbert space by $\ell^2(G)$.

Characteristic Operators: Fixing an indexing of the vertices, information about connectivity within the graph is encapsulated into the set of edge weights, collected into the adjacency matrix W and (diagonal) degree matrix D. Together with the weight matrix $M := \text{diag}\left(\{\mu_i\}_{i=1}^{|G|}\right)$, various standard geometry capturing characteristic operators – such as weighted adjacency matrix $M^{-1}W$, graph Laplacian $\Delta := M^{-1}(D - W)$ and normalized graph Laplacian $\mathscr{L} := M^{-1}D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$ can then be constructed. If the underlying graph is undirected, all of these operators are self-adjoint, however without this assumption, they need not even be normal. In this work, we shall remain agnostic to the choice of characteristic operator; only differentiating between normal at general characteristic operators in our results.

Functional Calculus Filters: A crucial component of GCNs are functional calculus filters, which arise from applying a function g to an underlying characteristic operator T. This then defines a new operator; denoted by $g(T)$. Various methods of implementations exist, all of which agree if more than one is applicable:

GENERIC FILTERS: If (and only if) T is normal, we may apply generic complex valued functions g to T: Writing normalized eigenvalue-eigenvector pairs of T as $(\lambda_i, \phi_i)_{i=1}^{|G|}$ one defines

$$g(T)f = \sum_{i=1}^{|G|} g(\lambda_i)\langle \phi_i, f\rangle_{\ell^2(G)}\phi_i \tag{1}$$

for any $f \in \ell^2(G)$. In this case one has $\|g(T)\|_{op} = \sup_{\lambda\in\sigma(T)} |g(\lambda)|$, with $\sigma(T)$ denoting the spectrum of T (i.e. its collection of eigenvalues). Furthermore, if g is chosen to be a bounded function, one might thus also obtain the T-independent bound $\|g(T)\|_{op} \leqslant \|g\|_\infty$. As is evident from (1), the function g need not even be defined on all of $\mathbb{C}$; it suffices that it be defined on all eigenvalues in $\sigma(T)$. With this in mind, we now define a space of filters which – as it will turn out – harmonizes well with our concept of transferability discussed in Section 5:

Definition 2.1. Fix $\omega \in \mathbb{C}$ and $C > 0$. Define the space $\mathscr{F}^{cont}_{\omega,C}$ of continuous filters on $\mathbb{C}\backslash\{\omega, \overline{\omega}\}$, to be the space of multilinear power-series' $g(z) = \sum_{\mu,\nu=0}^{\infty} a_{\mu\nu} (\omega - z)^{-\mu} (\overline{\omega} - \overline{z})^{-\mu}$ for which the pseudo-norm $\|g\|_{\mathscr{F}^{cont}_{\omega,C}} := \sum_{\mu,\nu>0}^{\infty} |\mu + \nu|C^{\mu+\nu-1}|a_{\mu\nu}|$ is finite.

Denoting by $B_\epsilon(\omega) \subseteq \mathbb{C}$ the open ball of radius ϵ around ω, one can show that for arbitrary $\delta > 0$ and every continuous function g defined on $\mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$ which is regular at infinity – i.e. satisfies $\lim_{r\to+\infty} g(rz) = c \in \mathbb{C}$ independent of which $z \neq 0$ is chosen – there is a function $f \in \mathscr{F}^{cont}_{\omega,C}$ so that $|f(z) - g(z)| \leqslant \delta$ for all $z \in \mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$. In other words, functions in $\mathscr{F}^{cont}_{\omega,C}$ can approximate a wide class of filters to arbitrary precision. More details are presented in Appendix B.

ENTIRE FILTERS: If T is not necessarily normal one might consistently apply entire (i.e. everywhere complex differentiable) functions to T. More details on the mathematical background are given in Appendix C. Here we simply note that such a function g is representable as an (everywhere convergent) power series $g(z) := \sum_{k=0}^{\infty} a_k^g z^k$ and my simply set

$$g(T) = \sum_{k=0}^{\infty} a_k^g \cdot T^k.$$

For the norm of the derived operator one easily finds $\|g(T)\|_{op} \leqslant \sum_{k=0}^{\infty} |a_k^g| \|T\|_{op}^k$ using the triangle inequality. While entire filters have the advantage that they are easily and efficiently implementable – making use only of matrix multiplication and addition – and are defined even if T is not normal, they suffer from the fact that it is impossible to give a $\|T\|_{op}$-independent bound for $\|g(T)\|_{op}$ as was the case for continuous filters. This behaviour can be traced back to the somewhat deep mathematical fact that no bounded entire function that is non-constant exists (c.f. e.g. Bak & Newman (2017)).

HOLOMORPHIC FILTERS: To define functional calculus filters that are both applicable to non-normal T and boundable somewhat more controlably in terms of T, one may relax the condition that g be entire to demanding that g be complex differentiable (i.e. **holomorphic**) only on an open subset $U \subseteq \mathbb{C}$ of the complex plane. Here we assume that U extends to infinity in each direction (i.e. is the complement of a closed and bounded subset of $\mathbb{C}$). For any g holomorphic on U and regular at infinity we set (with $(zId - T)^{-1}$ the so called reolvent of T at z)

$$g(T) := g(\infty) \cdot Id + \frac{1}{2\pi i} \oint_{\partial D} g(z) \cdot (zId - T)^{-1} dz, \tag{2}$$

for any T whose spectrum $\sigma(T)$ is completely contained in U. Here we have used the notation $g(\infty) = \lim_{r\to+\infty} g(rz)$ and taken D to an open set with nicely behaved boundary ∂D (more precisely a Cauchy domain; c.f. Appendix C for a precise definition) We assume that D completely contains $\sigma(T)$ and that its closure $\overline{D}$ is completely contained in U. The orientation of the boundary ∂D is the usual positive orientation on D (such that D 'is on the left' of ∂D). Using elementary facts

from complex analysis it can be shown (c.f. Gindler (1966)) that the resulting operator $g(T)$ in (2) is independent of the specific choice of D, as long as the above conditions are fulfilled.
While we will present results below in terms of this general definition – remaining agnostic to numerical implementation methods for the most part – it is instructive to consider a specific example-setting with definite and simple numerical implementation of such filters: To this end, chose an arbitrary point $\omega \in \mathbb{C}$ and set $U = \mathbb{C}\backslash\{\omega\}$ in the definitions above. Any function g that is holomorphic on U and regular at ∞ may then be represented by its Laurent series, which is of the form $g(z) = \sum_{k=0}^{\infty} b_k^g (z-\omega)^{-k}$ (Bak & Newman, 2017). For any T with $\sigma(T) \subseteq U$ (i.e. $\omega \notin \sigma(T)$) we then have

$$g(T) = \sum_{k=0}^{\infty} b_k^g \cdot (T-\omega)^{-k}$$

after evaluating the integral in (2). Appendix C contains the corresponding calculation. Such functions have already been used succesfully in graph learning tasks, e.g. in the guise of Cayley filters (c.f. Levie et al.). These filters are polynomials in $\frac{z+i}{z-i} = 1 + \frac{2i}{z-i}$. Cayley filters thus admit an expansion $g(T) = \sum_{k=0}^{\infty} b_k^g (T-i)^{-k}$ (with even only finitely many $\{b_k^g\}$ non-zero) obtained in the above manner (choosing $\omega = i$). We collect such functions into a designated filter space:

Definition 2.2. For a function $g(z) = \sum_{k=0}^{\infty} b_k^g (z-\omega)^{-k}$ on $U := \mathbb{C}\backslash\{\omega\}$ define the pseudo-norm $\|g\|_{\mathscr{F}^{hol}_{\omega,C}} := \sum_{k=1}^{\infty} |b_k^g| k C^{k-1}$ for $C > 0$. Denote the set of such g for which $\|g\|_{\omega,C} < \infty$ by $\mathscr{F}^{hol}_{\omega,C}$.

Given an operator $T : \ell^2(G) \to \ell^2(G)$ and denoting the trace norm by $\|\cdot\|_1$, we also define the continuous function $\gamma_T(z)$ on $\mathbb{C}\backslash\sigma(T)$ by $g_T(z) = 1/\text{dist}(z, \sigma(T))$ if T is normal and $\gamma_T(z) = \exp\left[2\|T\|_1 / d(z,\sigma(T))\right] / d(z, \sigma(T))$ otherwise. With this we find the following result:

Lemma 2.3. For holomorphic g and generic T we have $\|g(T)\|_{op} \leqslant |g(\infty)| + \frac{1}{2\pi} \oint_{\partial D} |g(z)| \gamma_T(z) d|z|$. Furthermore we have for any T with $\gamma_T(\omega) \leqslant C$, that $\|g(T)\|_{op} \leqslant \|g\|_{\mathscr{F}^{hol}_{\omega,C}}$ as long as $g \in \mathscr{F}_{C,\omega}$.

The second estimate of Lemma 2.3 (proved in Appendix D) finally bounds $\|g(T)\|_{op}$ independently of T, as long as g is holomorphic on $U := \mathbb{C}\backslash\{\omega\}$ and the resolvent of T at ω is bounded by C, which – importantly – does not force $\|T\|_{op}$ to be bounded.

Connecting Operators: To account for recently proposed networks where input- and 'processing' graphs are decoupled (see e.g. Alon & Yahav (2021); Topping et al. (2021)), as well as architectures incorporating graph pooling (Lee et al., 2019), we allow signal representations in the hidden network layers n to live in varying graph signal spaces $\ell^2(G_n)$. Connecting operators are then (not necessarily linear) operators $P_n : \ell^2(G_{n-1}) \to \ell^2(G_n)$ connecting the signal utilized of subsequent layers. We assume them to be Lipschitz continuous ($\|P(f) - P(g)\|_{\ell^2(G_{n-1})} \leqslant R\|f-g\|_{\ell^2(G_n)}$) and triviality preserving ($P(0) = 0$). For our original node-signal space we also write $\ell^2(G) \equiv \ell^2(G_0)$.

Non-Linearities: To each layer, we also associate a (possibly) non-linear function $\rho_n : \mathbb{C} \to \mathbb{C}$ acting poinwise on signals in $\ell^2(G_n)$. Similar to connecting operators, we assume ρ_n preserves zero and is Lipschitz-continuous with Lipschitz constant denoted by L_n. This definition allows for the absolute value non-linearity, but also ReLu or – trivially – the identity function.

Graph Convolutional Networks: A GCN with N layers is then constructed as follows: Let us denote the width of the network at layer n by K_n. The collection of hidden signals in this layer can then be thought of a single element of

$$\mathscr{L}_n := \bigoplus_{i \in K_n} \ell^2(G_n).$$

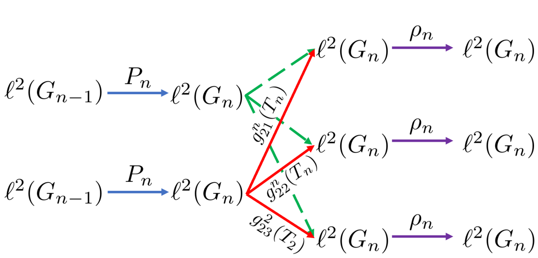

Figure 1: Update Rule for a GCN

Further let us write the collection of functional calculus filters utilized to generate the representation of this layer by $\{g^n_{ij}(\cdot) : 1 \leqslant j \leqslant K_{n-1}; 1 \leqslant i \leqslant K_n\}$. Further denoting the characteristic operator of this layer by T_n, the update rule (c.f. also Fig. 1) from the representation in $\mathscr{L}_{n-1}$ to $\mathscr{L}_n$ is then defined on each constituent in the direct sum $\mathscr{L}_n$ as

$$f_i^{n+1} = \rho_{n+1}\left(\sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(f_j^n)\right), \ \ \forall 1 \leqslant i \leqslant K_n.$$

We also denote the initial signal space by $\mathscr{L}_{\text{in}} := \mathscr{L}_0$ and the final one by $\mathscr{L}_{\text{out}} := \mathscr{L}_N$. The hence constructed map from the initial to the final space is denoted by $\Phi : \mathscr{L}_{\text{in}} \to \mathscr{L}_{\text{out}}$.

3 Stability to Input Signal Perturbations

In order to produce meaningful signal representations, a small change in input signal should produce a small change in the output of our networks. This property is captured by the Lipschitz constant of the map Φ of the GCN, whic is estimated by our first result below; proved in Appendix E.

Theorem 3.1. With the notation of Section 2 let $\Phi_N : \mathscr{L}_{\text{in}} \to \mathscr{L}_{\text{out}}$ be the map associated to an N-layer GCN. We have for all $f, h \in \mathscr{L}_{\text{in}}$ that

$$\|\Phi_N(f) - \Phi_N(h)\|_{\mathscr{L}_{\text{out}}} \leqslant \left(\prod_{n=1}^{N} L_n R_n B_n\right) \cdot \|f - h\|_{\mathscr{L}_{\text{in}}}$$

with $B_n := \sqrt{\sup_{\lambda\in\sigma(T_n)} \sum_{j\in K_{n-1}} \sum_{i\in K_n} |g_{ij}^n(\lambda)|^2}$ if T_n is normal. For general T_n we have for all $\{g_{ij}\}$ entire, holomorphic and in $\mathscr{F}_{\omega,C}$ respectively:

$$B_n := \begin{cases} \sum\limits_{k=0}^{\infty} \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} |(a_{ij}^{g_n})_k|^2} \cdot \|T_n\|_{op}^k \\ \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} \|g_{ij}^n(\infty)\|^2} + \frac{1}{2\pi} \oint_\Gamma \gamma_T(z) \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} |g_{ij}^n(z)|^2} d|z| \\ \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} \|g_{ij}^n\|_{\omega,C}^2} \end{cases}$$

We immediately notice that in the generic setting, where connecting operators are set to the identity and we solely use ReLU-non-linearities throughout the network, the Lipschitz constant of the network is completely controlled by the $\{B_n\}$. If the characteristic operator T_n is normal (e.g. if it is some standard graph Laplacian or adjacency matrix on an undirected graph), B_n in turn is controlled by the interplay of the utilized filters on the spectrum of T_n – as opposed to an estimate solely in terms of their magnitudes $\{\sup_{\lambda\in\sigma(T_n)} |g_{ij}^n(\cdot)|\}$. This for example allows to combine filters with $\sup_{\lambda\in\sigma(T_n)} |g_{ij}^n(\cdot) = \mathcal{O}(1)$, but which are supported on complimentary parts of the spectrum of T_n into a convolutional layer, while still maintaining $B_n = \mathcal{O}(1)$ instead of $\mathcal{O}(\sqrt{K_n \cdot K_{n-1}})$. If T_n is not normal but filters are holomorphic, such an interplay persists – however with filters now evaluated on the points of a curve around the spectrum of T_n and at ∞.

4 Stability to Edge Perturbations

Operators capturing graph-geometries might only be known approximately in real world tasks; e.g. if edge weights are only known to a certain level of precision. Hence it is important that graph convolutional networks be insensitive to small changes in the characteristic operators $\{T_n\}$. Additionally, since we consider graphs with vertex weights $\{\mu_i\}_{i=1}^{|G|}$ that are not necessarily set to unity, we have to consider the possibility that vertex weights are also only known to a certain level of precision; in which case we have signal spaces $\ell^2(G)$, $\ell^2(\widetilde{G})$ with the same vertex set $G = \widetilde{G}$, but only approximately the same vertex weights ($\mu_i \approx \widetilde{\mu}_i$). In this latter case, not only do the characteristic operators T_n, $\widetilde{T}_n$ differ, but also the the spaces $\ell^2(G)$, $\ell^2(\widetilde{G})$ on which they act. To capture this setting mathematically, we assume in this section that there is a linear operator $J : \ell^2(G) \to \ell^2(\widetilde{G})$ facilitating contact between signal spaces. We measure the difference between the two characteristic operators operating in different spaces by considering the norm of the difference $(JT - \widetilde{T}J)$, which as an operator maps from $\ell^2(G)$ to $\ell^2(\widetilde{G})$. Before investigating the stability of entire networks we first comment on single-filter stability. For normal operators we then find the following result, proved in Appendix A building on ideas first developed in (T.P., 2009).

Lemma 4.1. Denote by $\|\cdot\|_F$ the Frobenius norm and let T and $\widetilde{T}$ be normal on $\ell^2(G)$ and $\ell^2(\widetilde{G})$ respectively. Let g be Lipschitz continuous with Lipschitz constant D_g. For any linear $J: \ell^2(G) \to \ell^2(\widetilde{G})$ we have $\|g(\widetilde{T})J - Jg(T)\|_F \leqslant D_g \|\widetilde{T}J - JT\|_F$.

Unfortunately, scalar Lipschitz continuity is only inherited by operator functions if they are applied to normal operators and when using Frobenius norm (as opposed to spectral norm). For general operators we have the following somewhat weaker result, proved in Appendix F:

Lemma 4.2. Let $T, \widetilde{T}$ be operators on on $\ell^2(G)$, $\ell^2(\widetilde{G})$ with $\|T\|_{op}, \|\widetilde{T}\|_{op} \leqslant C$. Let J : $\ell^2(G) \to \ell^2(\widetilde{G})$ be arbitrary but linear. With $K_g = \sum_{k=1}^{\infty} |a_k^g| k C^{k-1}$ for g entire and $K_g = \frac{1}{2\pi} \oint_{\partial D} \frac{1}{z} \gamma_T(z) \gamma_{\widetilde{T}}(z) |g(z)| d|z|$ for g holomorphic, we have

$$\|g(T)J - Jg(\widetilde{T})\|_{op} \leqslant K_g \cdot \|JT - \widetilde{T}J\|_{op}$$

For a complete GCN we then find the following result, proved in Appendix G:

Theorem 4.3. Let $\Phi_N, \widetilde{\Phi}_N$ be the maps associated to N-layer graph convolutional networks with the same non-linearities and functional calculus filters, but based on different graph signal spaces $\ell^2(G), \ell^2(\widetilde{G})$, characteristic operators $T_n, \widetilde{T}_n$ and connecting operators $P_n, \widetilde{P}_n$. Assume $B_n, \widetilde{B}_n \leqslant B$ as well as $R_n, \widetilde{R}_n \leqslant R$ and $L_n \leqslant L$ for some $B, R, L > 0$ and all $n \geqslant 0$. Assume that there are identification operators $J_n : \ell^2(G_n) \to \ell^2(\widetilde{G}_n)$ $(0 \leqslant n \leqslant N)$ commuting with non-linearities and connecting operators in the sense of $\|\widetilde{P}_n J_{n-1} f - J_n P_n f\|_{\ell^2(\widetilde{G}_n)} = 0$ and $\|\rho_n(J_n f) - J_n \rho_n(f)\|_{\ell^2(\widetilde{G}_n)} = 0$. Depending on whether normal or arbitrary characteristic operators are used, define $D_n^2 := \sum_{j \in K_{n-1}} \sum_{i \in K_n} D_{g_{ij}^n}^2$ or $D_n^2 := \sum_{j \in K_{n-1}} \sum_{i \in K_n} K_{g_{ij}^n}^2$. Choose D such that $D_n \leqslant D$ for all n. Finally assume that $\|J_n T_n - \widetilde{T}_n J_n\|_* \leqslant \delta$ and with $* = F$ if both operators are normal and $* = op$ otherwise. Then we have for all $f \in \mathscr{L}_{\text{in}}$ and with $\mathscr{J}_N$ the operator that the K_N copies of J_N induced through concatenation that

$$\|\widetilde{\Phi}(J_0 f) - \mathscr{J}_N \Phi(f)\|_{\widetilde{\mathscr{L}}_{\text{out}}} \leqslant N \cdot D \cdot B^{N-1} \cdot \|f\|_{\mathscr{L}_{\text{in}}} \cdot \delta.$$

The stability result persists with slightly altered stability constants, if identification operators only *almost* commute with non-linearities and/or connecting operators, as Appendix G further elucidates.

5 Stability to Structural Perturbations and Transferability

While the demand that $\|\widetilde{T}J - JT\|$ be small in some norm is well adapted to capture some notions of closeness of graphs – especially to the setting of graph perturbations that do not change the number of vertices – it is too stringent to capture other important concepts of closeness of graphs (and their respective characteristic operators). An illustrative example, further developed in Section 5.2 and numerically investigates in Section 6 below, is given by considering the following setting: Suppose we are given an undirected graph $\widetilde{G}$ with all edge weights of order $\mathcal{O}(1)$ safe one: For this exceptional edge, the weight is of $\mathcal{O}(1/\delta)$ for some small δ. Choosing the characteristic operator as the graph Laplacian (which governs heat-flow in Physics (Cole, 2011)) we might think of the graph as modelling an array of coupled heat reservoirs. In this setting, edge weights correspond to heat-conductivities between different reservoirs and as $\delta \to 0$, the conductivity between the reservoirs coupled through the exceptional edge tends to infinity. Thus heat exchange is instantaneous and both nodes act as a single entity. On physical grounds, we might thus describe this system effectively by collapsing the strong edge and fusing the corresponding vertices into a new vertex hence creating a new graph G with $|G| = |\widetilde{G}| - 1$ vertices. While the numerical investigation of Section 6 strongly suggest that in this setting one might not bound $\|\widetilde{T}J - JT\|$ in terms of δ, Section 5.2 establishes analytically, that the situation is different when considering resolvents of the graph Laplacians: Indeed, Theorem 5.6 below proves that we have $\|(Id + \widetilde{T})^{-1}J - J(Id + T)^{-1}\| \lesssim \delta^{\frac{1}{2}}$. Motivated by this example, Section 5.1 develops a general theory for the difference in outputs of networks evaluated on graphs for which the **resolvents** $R_\omega := (\omega Id - T)^{-1}$ and $\widetilde{R}_\omega := (\omega Id - \widetilde{T})^{-1}$ of the respective characteristic operators are close in some sense. Subsequently, Section 5.2 then picks up the initial physical example of coupled heat reservoirs once more.

5.1 General Theory

Throughout this section, fix a complex number $\omega \in \mathbb{C}$ and for each appearing operator T assume $\omega, \overline{\omega} \notin \sigma(T)$. This can always be achieved by choosing ω such that $|\omega| \geqslant \|T\|_{op}$, but is especially easy to fulfil if T is self adjoint (choose e.g. $\omega = i$) or a non-negative operator (choose e.g. $\omega = (-1)$ such as the graph Laplacian and its normalized cousin.
As a first step, we then note that the conclusion of Lemma 4.1 can always be satisfied if we chose $J \equiv 0$. To exclude this case – where the application of J corresponds to losing too much information – we make the following definition; following Post (2012):

Definition 5.1. Let $J : \ell^2(G) \to \ell^2(\widetilde{G})$ and $\widetilde{J} : \ell^2(\widetilde{G}) \to \ell^2(G)$ be linear, and let $T, \widetilde{T}$ be operators on $\ell^2(G)$ and $\ell^2(\widetilde{G})$ respectively. We say that J and $\widetilde{J}$ are quasi unitary with respect to $T, \widetilde{T}$ and ω if

$$\|Jf\|_{\ell^2(\widetilde{G})} \leqslant 2\|f\|_{\ell^2(G)}, \quad \|(J - \widetilde{J}^*)f\|_{\ell^2(\widetilde{G})} \leqslant \delta\|f\|_{\ell^2(G)},$$
$$\|(Id - \widetilde{J}J)R_\omega f\|_{\ell^2(G)} \leqslant \delta\|f\|_{\ell^2(G)}, \quad \|(Id - J\widetilde{J})\widetilde{R}_\omega u\|_{\ell^2(\widetilde{G})} \leqslant \delta\|u\|_{\ell^2(\widetilde{G})}. \tag{3}$$

The motivation to include the resolvents in the norm estimates (3) is twofold: On the one hand, it facilitates contact between the geometry of the graph (captured by T and hence also its resolvent) and the identification operators. On the other hand, when $T = \Delta$ is the (positive) graph Laplacian, the left equation in (3 is for example equivalent to demanding that $\|(Id - \widetilde{J}J)Rf\|^2_{\ell^2(G)}$ be smaller than $\delta(\|f\|^2 + \mathcal{E}_\Delta(f))^{\frac{1}{2}}$, with $\mathcal{E}_\Delta(\cdot) = \langle \cdot, \Delta \cdot \rangle_{\ell^2(G)}$ the (positive) energy form induced by the Laplacian Δ (Post, 2012). This can thus be interpreted as a (natural) relaxation of the standard demand $\|(Id - \widetilde{J}J)\|_{op} \leqslant \delta$. Relaxing the demand that $\|\widetilde{T}J - JT\| \lesssim \delta$ of Section 4, we now demand closeness of resolvents instead of closeness of operators:

Definition 5.2. If, for some $w \in \mathbb{C}$ and linear mapping $J : \ell^2(G) \to \ell^2(\widetilde{G})$ the resolvents R_ω and $\widetilde{R}_\omega$ satisfy $\|(\widetilde{R}_\omega J - JR_\omega)f\|_{\ell^2(\widetilde{G})} \leqslant \delta\|f\|_{\ell^2(G)}$ for all $f \in \ell^2(G)$, we say that T and $\widetilde{T}$ are **ω-δ-close** with identification operator J. If additonally $\|(\widetilde{R}^*_\omega J - JR^*_\omega)f\|_{\ell^2(\widetilde{G})} \leqslant \delta\|f\|_{\ell^2(G)}$ holds, we say that T and $\widetilde{T}$ are **doubly ω-δ-close**.

Section 6 gives empirical evidence that such a relaxation is indeed useful to capture naturally appearing and practically relevant settings. Here we will focus on establishing that operators being (doubly-)ω-δ-close has useful consequences. We then find the following result:

Lemma 5.3. Let T and $\widetilde{T}$ be characteristic operators on $\ell^2(G)$ and $\ell^2(\widetilde{G})$ be respectively. If these operators are ω-δ-close with identification operator J, and $\|R_\omega\|_{op}, \overline{R}_\omega\|_{op} \leqslant C$ we have

$$\|Jg(T) - g(\widetilde{T})J\|_{op} \leqslant K_g \cdot \|(\widetilde{R}_\omega J - JR_\omega)\|_{op}$$

with $K_g = \frac{1}{2\pi}\oint_{\partial D}(1 + |z-\omega|\gamma_T(z))(1 + |z-\omega|\gamma_{\widetilde{T}}(z))|g(z)|d|z|$ if g is holomorphic and $K_g = \|g\|_{\mathscr{F}^{hol}_{\omega,C}}$ if $g \in \mathscr{F}^{hol}_{\omega,C}$. If T and $\widetilde{T}$ are normal as well as doubly ω-δ-close and $g \in \mathscr{F}^{cont}_{\omega,C}$, we have $K_g = \|g\|_{\mathscr{F}^{cont}_{\omega,C}}$.

Single filter results may be extended to entire networks, as detailed in Theorem 5.4 below. This result as well as the previous Lemma are proved and further discussed in Appendix H.

Theorem 5.4. Let $\Phi, \widetilde{\Phi}$ be the maps associated to N-layer graph convolutional networks with the same non-linearities and functional calculus filters, but based on different graph signal spaces $\ell^2(G_n), \ell^2(\widetilde{G}_n)$, characteristic operators $T_n, \widetilde{T}_n$ and connecting operators $P_n, \widetilde{P}_n$. Assume $B_n, \widetilde{B}_n \leqslant B$ as well as $R_n, \widetilde{R}_n \leqslant R$ and $L_n \leqslant L$ for some $B, R, L > 0$ and all $n \geqslant 0$. Assume that there are identification operators $J_n : \ell^2(G_n) \to \ell^2(\widetilde{G}_n)$ $(0 \leqslant n \leqslant N)$ commuting with non-linearities and connecting operators in the sense of $\|\widetilde{P}_n J_{n-1} f - J_n P_n f\|_{\ell^2(\widetilde{G}_n)} = 0$ and $\|\rho_n(J_n f) - J_n\rho_n(f)\|_{\ell^2(\widetilde{G}_n)} = 0$. define $D_n^2 := \sum_{j \in K_{n-1}} \sum_{i \in K_n} K^2_{g^n_{ij}}$ with $K_{g^n_{ij}}$ as in Lemma 5.3. Choose D such that $D_n \leqslant D$ for all n. Finally assume that $\|J_n(\omega Id - T_n)^{-1} - (\omega Id - \widetilde{T}_n)^{-1}J_n\|_{op} \leqslant \delta$. If filters in $\mathscr{F}^{cont}_{\omega,C}$ are used, assume additionally that $\|J_n((\omega Id - T_n)^{-1})^* - ((\omega Id - \widetilde{T}_n)^{-1})^* J_n\|_{op} \leqslant \delta$. Then we have for all $f \in \mathscr{L}_{\text{in}}$ and with $\mathscr{J}_N$ the operator that the K_N copies of J_N induced through concatenation that

$$\|\widetilde{\Phi}(J_0 f) - \mathscr{J}_N \Phi(f)\|_{\widetilde{\mathscr{L}}_{\text{out}}} \leqslant N \cdot D \cdot B^{N-1} \cdot \|f\|_{\mathscr{L}_{\text{in}}} \cdot \delta.$$

5.2 COLLAPSING STRONG EDGES

We shall now once more pick up the example of a graph modelling a system of connected heat reservoirs. To be more precise, we consider two graphs with vertex sets G and $\widetilde{G}$. We consider G to be a subset of the vertex set $\widetilde{G}$ and think of the graph corresponding to G as arising in a collapsing procedure from the 'larger' graph $\widetilde{G}$. More precisely, we assume that the vertex set $\widetilde{G}$ can be split into three disjoint subsets $\widetilde{G} = \widetilde{G}_{Latin} \bigcup \widetilde{G}_{Greek} \bigcup \{\star\}$ (c.f. also Fig. 2). We assume that the adjacency matrix $\widetilde{W}$ when restricted to Latin vertices or a Latin vertex and the exceptional node '$\star$' is of order unity $(\widetilde{W_{ab}}, \widetilde{W}_{a\star} = \mathcal{O}(1), \forall a, b \in \widetilde{G}_{Latin})$. For Greek indices, we assume that we may write $\widetilde{W}_{\alpha\beta} = \frac{\omega_{\alpha\beta}}{\delta}$ and $\widetilde{W}_{\alpha\star} = \frac{\omega_{\alpha\star}}{\delta}$ such that $(\omega_{\alpha\beta}, \omega_{\alpha\star} = \mathcal{O}(1)$ for all $\alpha, \beta \in \widetilde{G}_{Greek}$. In other words, we assume that $\widetilde{W}_{\alpha\beta}$ and $\widetilde{W}_{\alpha\star}$ are of order $1/\delta$ whenever they are non-zero. We also assume that the sub-graph corresponding to vertices in $\widetilde{G}_{Greek} \bigcup \{\star\}$ is connected.

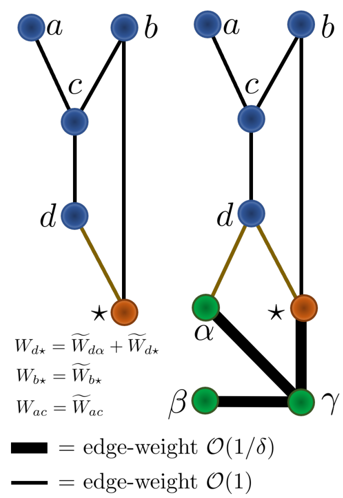

Figure 2: Collapsed (left) and uncollapsed (right) Graphs

From the graph corresponding to $\widetilde{G}$, we construct a smaller graph through 'collapsing strong edges'. That is, we take $G = \widetilde{G}_{Latin} \bigcup \{\star\}$ (c.f. again Fig. 2). The adjacency matrix W on this graph is constructed by defining $W_{ab} = \widetilde{W}_{ab}, \forall a, b \in \widetilde{G}_{Latin}$ and setting

$$W_{\star a} := \widetilde{W}_{a\star} + \sum_{\beta \in \widetilde{G}_{Greek}} \widetilde{W}_{a\beta} \qquad \left(\forall a \in \widetilde{G}_{Latin}\right),$$

with $W_{a\star} \equiv W_{\star a}$. We also allow our graph $\widetilde{G}$ to posses node-weights $\{\widetilde{\mu}_{\widetilde{g}}\}_{\widetilde{g} \in \widetilde{G}}$ that are not necessarily equal to one. The Laplace operator $\Delta_{\widetilde{G}}$ acting on the graph signal space $\ell^2(\widetilde{G})$ induces a positive definite and convex energy form on this signal space via

$$E_{\widetilde{G}}(u) := \langle u, \Delta_{\widetilde{G}} u \rangle_{\ell^2(\widetilde{G})} = \sum_{g,h \in \widetilde{G}} W_{gh} |u(g) - u(h)|^2.$$

Using this energy form, we now define a set comprised of $|G|$ signals, all of which live in $\ell^2(\widetilde{G})$. These signals will then in turn be used to facilitate contact between the respective graph signal spaces $\ell^2(G)$ and $\ell^2(\widetilde{G})$. They will also be used to define appropriate weights on the smaller graph G.

Definition 5.5. For each $g \in G$, define the signal $\psi_g \in \ell^2(\widetilde{G})$ as the unique solution to the convex optimization program

$$\min E_{\widetilde{G}}(u) \quad \textit{subject to} \;\; u(h) = \delta_{hg} \;\; \textit{for all } h \in \widetilde{G}_{Latin} \bigcup \{\star\}. \tag{4}$$

With the given boundary conditions, what is left to determine in the above optimization program are the 'Greek entries' $\psi_g(\alpha)$ of each ψ_g. As Appendix I further elucidates, these can be calculated explicitly and purely in terms of the inverse of $\Delta_{\widetilde{G}}$ restricted to Greek indices as well as certain row vectors of the adjacency matrix. Node-weights on G are then defined as $\mu_g := \sum_{\widetilde{h} \in \widetilde{G}} \psi_g(\widetilde{h}) \cdot \widetilde{\mu}_{\widetilde{h}}$. We denote the corresponding signal space by $\ell^2(G)$. Importantly, one has $\mu_a \to \widetilde{\mu}_a$ for any Latin index and $\mu_\star \to \widetilde{\mu}_\star + \sum_{\alpha \in \widetilde{G}_{Greek}} \widetilde{\mu}_\alpha$ as $\delta \to 0$; recovering our physical intuition from the beginning of the section. To translate signals from $\ell^2(G)$ space to the 'larger' space $\ell^2(\widetilde{G})$ and back, we define two identification operators $J : \ell^2(G) \to \ell^2(\widetilde{G})$ and $\widetilde{J} : \ell^2(G) \to \ell^2(G)$ via $Jf := \sum_{g \in G} f(g) \cdot \psi_g$ and $(J'u)(g) := \langle u, \psi_x \rangle_{\ell^2(\widetilde{G})} / \mu_g$ for all $f \in \ell^2(G)$, $u \in \ell^2(\widetilde{G})$ and $g \in G$. Our main theorem in this section then states the following:

Theorem 5.6. With definitions and notation as above, there are constants $K_1, K_2 \geqslant 0$ such that the operators J and $\widetilde{J}$ are $(K_1\sqrt{\delta})$-close with respect to $\Delta_{\widetilde{G}}$, Δ_G and $\omega = (-1)$. Furthermore, the operators $\Delta_{\widetilde{G}}$ and Δ_G are (-1)-$(K_2\sqrt{\delta})$ close with identification operator J.

The (fairly involved) proof of this Theorem is contained in Appendix I. Importantly (as evident from the proof), the constants K_1, K_2 are of order $\mathcal{O}(\sum_{g \in \widetilde{G}_{Greek}} \widetilde{\mu}_g)$ as opposed to $\mathcal{O}(\sum_{g \in \widetilde{G}} \widetilde{\mu}_g)$.

6 NUMERICAL RESULTS

In this section we complement the analytical results of the previous sections through numerical investigations. As the results and discussion of Section 5 are conceptionally the most involved and also furthest removed from other approaches to transferability, we focus on numerically illustrating the derived claims. To this end, we consider the setting introduced in Section 5.2 and consider a generic fully connected graph $\widetilde{G}$ with $|\widetilde{G}| = 8$. We consider a spliting into $\widetilde{G} = \widetilde{G}_{Latin} \bigcup \widetilde{G}_{Greek} \bigcup \{\star\}$ with $|\widetilde{G}_{Latin}| = 3$ and $|\widetilde{G}_{Greek}| = 4$. As described in Section 5.2 we assume $\widetilde{W_{ab}}, \widetilde{W}_{a\star} = \mathcal{O}(1), \forall a, b \in \widetilde{G}_{Latin}$ and $\widetilde{W}_{\alpha\beta} = \frac{\omega_{\alpha\beta}}{\delta}$ and $\widetilde{W}_{\alpha\star} = \frac{\omega_{\alpha\star}}{\delta}$ such that $(\omega_{\alpha\beta}, \omega_{\alpha\star} = \mathcal{O}(1)$ for all $\alpha, \beta \in \widetilde{G}_{Greek}$. For completeness and reproducibility, the full adjacency matrix $\widetilde{W}$ can be found in Appendix J. We set node weight on $\widetilde{G}$ to one and – as discussed – construct a graph G with $|G| = 4$ through 'collapsing strong edges'. Figure 3 then compares various quantities on these two graphs with each other.

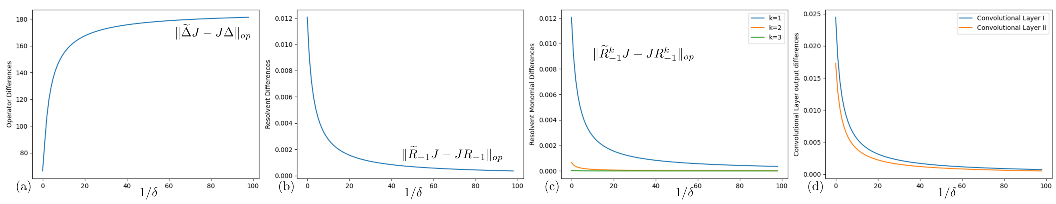

Figure 3: (a) Differences in operators, (b) Differences in Resolvents, (c) Differences in monomials of operators ($k = 1, 2, 3$), (d) Differences in graph-convolutional-layer outputs

Comparing Fig. 3 (a) and (b), we can clearly see that – at least while utilizing the (natural) identification operator J – the difference (in the sense of Section 4) between the two graph Laplacians does not decay but rather increases with the scale $1/\delta$ of the edge-weights in the 'Greek sector'. The resolvents of the respective Laplacians however do approach each other (in the sense of Definition 5.2) as is evident from Fig. 3 (b). We speculate that this behaviour is due to the fact that the resolvent suppresses large eigenvalues as $\propto (1 + \lambda_{large})^{-1}$. We hypothesize that the eigenspaces corresponding to these large eigenvalues eventually concentrate precisely on 'Greek indices' and are thus 'suppressed when applying $(Id + \widetilde{\Delta})^{-1}$, while dominating when applying $\widetilde{\Delta}$. Figure 3 (c) depicts the difference of monomials in the resolvents (from which filters in $\mathscr{F}^{cont}_{(-1),1}$ and $\mathscr{F}^{hol}_{(-1),1}$ are then constructed). It is found that contrary to the theoretical bound in Lemma 5.3, differences seem to decrease as the power k of the monomial increases. Finally Fig. 3(d) depicts the transferability of full convolutional network layers. In both layers the same filters (in total 12) are utilized; however the network topologies are different. Convolutional Layer I is depicted in Fig. 1. It maps (in the notation of Section 2) from a hidden signal representation with width $K_n = 2$ to one with width $K_{n+1} = 3$. Convolutional Layer II maps from a a hidden signal representation with width $K_n = 2$ to one with width $K_{n+1} = 6$ and is sparsely connected. Occuring filters are the same in both convolutional Layer, so that both should have the same stability constant in the sense of Theorem 5.4. We see that the sparsely connected network is slightly more transferable, having lower transferability error at the same value of $1/\delta$. This showcases the importance of accounting for network topologies when aiming to provide tight-as-possible architecture-specific stability bounds as opposed to our derived topology-agnostic bounds.

7 DISCUSSION

We have developed a mathematically well founded framework for the analysis of stability properties of general graph convolutional networks. For undirected graphs, we related node-level stability to to spectral covering properties and edge-level stability to Lipschitz constants of utilized filters. For directed graphs (more generally non-normal characteristic operators), tools from complex analysis provided grounds for derived stability properties. We also introduced a new notion of stability to structural perturbations (i.e. transferability) and detailed how the developed line of thought captures relevant settings of structural changes in graphs. We obtained that the 'size of the conceptual change in description' when collapsing a subgraph to a node is given by a constant describing the size of the collapsed sub-graph multiplied by the inverse of the square root of the characteristic coupling strength within this sub-graph. Numerical investigations supported the theoretical analysis.

References

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

Joseph Bak and Donald J. Newman. *Complex analysis*. Springer, 2017.

Oscar F. Bandtlow. Estimates for norms of resolvents and an application to the perturbation of spectra. *Mathematische Nachrichten*, 267, 2004.

Aleksandar Bojchevski and Stephan Günnemann. Certifiable robustness to graph perturbations. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 8317–8328, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/e2f374c3418c50bc30d67d5f7454a5b4-Abstract.html.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann Lecun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.

Zhengdao Chen, Lisha Li, and Joan Bruna. Supervised community detection with line graph neural networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=H1g0Z3A9Fm.

Kevin D. Cole. *Heat conduction using Green's functions*. Taylor and Francis, 2011.

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.

Zhan Gao, Elvin Isufi, and Alejandro Ribeiro. Stability of graph convolutional neural networks to stochastic perturbations. *Signal Process.*, 188:108216, 2021. doi: 10.1016/j.sigpro.2021.108216. URL https://doi.org/10.1016/j.sigpro.2021.108216.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Herbert A. Gindler. An operational calculus for meromorphic functions. *Nagoya Mathematical Journal*, 26:31–38, 1966. doi: 10.1017/S0027763000011600.

David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8): 595–608, 2016.

Henry Kenlay, Dorina Thanou, and Xiaowen Dong. On the stability of graph convolutional neural networks under edge rewiring. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, ON, Canada, June 6-11, 2021*, pp. 8513–8517. IEEE, 2021. doi: 10.1109/ICASSP39728.2021.9413474. URL https://doi.org/10.1109/ICASSP39728.2021.9413474.

Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. Convergence and stability of graph convolutional networks on large random graphs. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f5a14d4963acf488e3a24780a84ac96c-Abstract.html.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3734–3743. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/lee19c.html.

Ron Levie, Federico Monti, Xavier Bresson, and Michael M. Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. URL https://arxiv.org/abs/1705.07664.

Ron Levie, Michael M. Bronstein, and Gitta Kutyniok. Transferability of spectral graph convolutional neural networks. *CoRR*, abs/1907.12972, 2019a. URL http://arxiv.org/abs/1907.12972.

Ron Levie, Elvin Isufi, and Gitta Kutyniok. On the transferability of spectral graph filters. *CoRR*, abs/1901.10524, 2019b. URL http://arxiv.org/abs/1901.10524.

Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012. doi: https://doi.org/10.1002/cpa.21413. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21413.

Sohir Maskey, Ron Levie, and Gitta Kutyniok. Transferability of graph neural networks: an extended graphon approach. *CoRR*, abs/2109.10096, 2021. URL https://arxiv.org/abs/2109.10096.

Barry Simon Michael Reed. *Methods of modern mathematical physics, Volume 1*. Academic Press, 1981.

Olaf. Post. *Spectral Analysis on Graph-like Spaces / by Olaf Post.* Lecture Notes in Mathematics, 2039. Springer Berlin Heidelberg, Berlin, Heidelberg, 1st ed. 2012. edition, 2012. ISBN 3-642-23840-8.

Olaf Post and Jan Simmer. Approximation of fractals by discrete graphs: norm resolvent and spectral convergence, 2017. URL https://arxiv.org/abs/1704.00064.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon neural networks and the transferability of graph neural networks. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/12bcd658ef0a540cabc36cdf2b1046fd-Abstract.html.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature, 2021. URL https://arxiv.org/abs/2111.14522.

Wihler T.P. On the hölder continuity of matrix functions for normal matrices. *Journal of inequalities in pure and applied mathematics*, 10(4), Dec 2009. ISSN 1443-5756. URL https://www.emis.de/journals/JIPAM/images/276_09_JIPAM/276_09_www.pdf.

Thomas Wiatowski and Helmut Bölcskei. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory*, 64:1845–1866, 2018.

A Some Concepts in Linear Algebra

In the interest of self-containedness, we provide a brief review of some concepts from linear algebra utilized in this work that might potentially be considered more advanced. Presented results are all standard; a very thorough reference is Michael Reed (1981).

Hilbert Spaces: To us, a Hilbert space — often denoted by $\mathcal{H}$ — is a vector space over the complex numbers which also has an inner product — often denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Prototypical examples are given by the Euclidean spaces $\mathbb{C}^d$ with inner product $\langle x, y \rangle_{\mathbb{C}^d} := \sum_{i=1}^{d} \overline{x}_i y_i$. Associated to an inner product is a norm, denoted by $\| \cdot \|_{\mathcal{H}}$ and defined by $\|x\|_{\mathcal{H}} := \sqrt{\langle x, x \rangle_{\mathcal{H}}}$ for $x \in \mathcal{H}$.

Direct Sums of Spaces: Given two potentially different Hilbert spaces $\mathcal{H}$ and $\widehat{\mathcal{H}}$, one can form their direct sum $\mathcal{H} \oplus \widehat{\mathcal{H}}$. Elements of $\mathcal{H} \oplus \widehat{\mathcal{H}}$ are vectors of the form (a, b), with $a \in \mathcal{H}$ and $b \in \widehat{\mathcal{H}}$. Addition and scalar multiplication are defined in the obvious way by

$$(a, b) + \lambda(c, d) := (a + \lambda c, b + \lambda d)$$

for $a, c \in \mathcal{H}$, $b, d \in \widehat{\mathcal{H}}$ and $\lambda \in \mathbb{C}$. The inner product on the direct sum is defined by

$$\langle (a, b), (c, d) \rangle_{\mathcal{H} \oplus \widehat{\mathcal{H}}} := \langle a, c \rangle_{\mathcal{H}} + \langle b, d \rangle_{\widehat{\mathcal{H}}}.$$

As is readily checked, this implies that the norm $\| \cdot \|_{\mathcal{H} \oplus \widehat{\mathcal{H}}}$ on the direct sum is given by

$$\|(a, b)\|^2_{\mathcal{H} \oplus \widehat{\mathcal{H}}} := \|a\|^2_{\mathcal{H}} + \|b\|^2_{\widehat{\mathcal{H}}}.$$

Standard examples of direct sums are again the Euclidean spaces, where one has $\mathbb{C}^d = \mathbb{C}^n \oplus \mathbb{C}^m$ if $m + n = d$, as is easily checked. One might also consider direct sums with more than two summands, writing $\mathbb{C}^d = \oplus_{i=1}^{d} \mathbb{C}$ for example. In fact, one might also consider infinite sums of Hilbert spaces: The space $\oplus_{i=1}^{\infty} \mathcal{H}_i$ is made up of those elements $a = (a_1, a_2, a_3, ...)$ with $a_i \in \mathcal{H}_i$ for which the norm

$$\|a\|^2_{\oplus_{i=1}^{\infty} \mathcal{H}_i} := \sum_{i=1}^{\infty} \|a_i\|^2_{\mathcal{H}_i}$$

is finite. This means for example that the vector $(1, 0, 0, 0, ...)$ is in $\oplus_{i=1}^{\infty} \mathbb{C}$, while $(1, 1, 1, 1, ...)$ is not.

Direct Sums of Maps: Suppose we have two collections of Hilbert spaces $\{\mathcal{H}_i\}_{i=1}^{\Gamma}$, $\{\widetilde{\mathcal{H}}_i\}_{i=1}^{\Gamma}$ with $\Gamma \in \mathbb{N}$ or $\Gamma = \infty$. Suppose further that for each $i \leqslant \Gamma$ (resp. $i < \Gamma$) we have a (not necessarily linear) map $J_i : \mathcal{H}_i \to \widetilde{\mathcal{H}}_i$. Then the collection $\{J_i\}_{i=1}^{\Gamma}$ of these 'component' maps induce a 'composite' map

$$\mathscr{J} : \oplus_{i=1}^{\Gamma} \mathcal{H}_i \longrightarrow \oplus_{i=1}^{\Gamma} \widetilde{\mathcal{H}}_i$$

between the direct sums. Its value on an element $a = (a_1, a_2, a_3, ...) \in \oplus_{i=1}^{\Gamma} \mathcal{H}_i$ is defined by

$$\mathscr{J}(a) = (J_1(a_1), J_2(a_2), J_3(a_3), ...) \in \oplus_{i=1}^{\Gamma} \widetilde{\mathcal{H}}_i.$$

Strictly speaking, one has to be a bit more careful in the case where $\Gamma = \infty$ to ensure that $\|\mathscr{J}(a)\|_{\oplus_{i=1}^{\infty} \widetilde{\mathcal{H}}_i} \neq \infty$. This can however be ensured if we have $\|J_i(a_i)\|_{\widetilde{\mathcal{H}}_i} \leqslant C \|a_i\|_{\mathcal{H}_i}$ for all $1 \leqslant i$ and some C independent of all i, since then $\|\mathscr{J}(a)\|_{\oplus_{i=1}^{\infty} \widetilde{\mathcal{H}}_i} \leqslant C \|a\|_{\oplus_{i=1}^{\infty} \mathcal{H}_i} \leqslant \infty$. If each J_i is a linear operator, such a C exists precisely if the operator norms (defined below) of all J_i are smaller than some constant.

Operator Norm: Let $J : \mathcal{H} \to \widetilde{\mathcal{H}}$ be a linear operator between Hilbert spaces. We measure its 'size' by what is called the operator norm, denoted by $\| \cdot \|_{op}$ and defined by

$$\|J\|_{op} := \sup_{\psi \in \mathcal{H}, \|\psi\|_{\mathcal{H}} = 1} \frac{\|A\psi\|_{\widetilde{\mathcal{H}}}}{\|\psi\|_{\mathcal{H}}}.$$

Adjoint Operators Let $J : \mathcal{H} \to \widetilde{\mathcal{H}}$ be a linear operator from the Hilbert space $\mathcal{H}$ to the Hilbert space $\widetilde{\mathcal{H}}$. Its adjoint $J^* : \widetilde{\mathcal{H}} \to \mathcal{H}$ is an operator mapping in the opposite direction. It is uniquely determined by demanding that

$$\langle Jf, u \rangle_{\widetilde{\mathcal{H}}} = \langle f, J^* u \rangle_{\mathcal{H}}$$

holds true for arbitrary $f \in \mathcal{H}$ and $u \in \widetilde{\mathcal{H}}$.

Normal Operators: If a linear operator $\Delta : \mathcal{H} \to \mathcal{H}$ maps from and to the same Hilbert space, we can compare it directly with its adjoint. If $\Delta\Delta^* = \Delta^*\Delta$, we say that the operator Δ is normal. Special instances of normal operators are self-adjoint operators, for which we have the stronger property $\Delta = \Delta^*$. If an operator is normal, there are unitary maps $U : \mathcal{H} \to \mathcal{H}$ diagonalizing Δ as

$$U^*\Delta U = \text{diag}(\lambda_1, ...\lambda_n),$$

with eigenvalues in $\mathbb{C}$. We call the collection of eigenvalues the spectrum $\sigma(\Delta)$ of Δ. If $\dim \mathcal{H} = d$, we may write $\sigma(\Delta) = \{\lambda\}_{i=1}^d$. It is a standard exercise to verify that each eigenvalue satisfies $|\lambda_i| \leqslant \|\Delta\|_{op}$. Associated to each eigenvalue is an eigenvector ϕ_i. The collection of all (normalized) eigenvectors forms an orthonormal basis of $\mathcal{H}$. We may then write

$$\Delta f = \sum_{i=1}^{d} \lambda_i \langle \phi_i, f \rangle_{\mathcal{H}} \phi_i.$$

Resolvent of an Operator: Given an operator T on some Hilbert space $\mathcal{H}$, we have by definition that the operator $(T - z) : \mathcal{H} \to \mathcal{H}$ is invertible precisely if $z \neq \sigma(T)$. In this case we write

$$R_z(T) = (zId - T)^{-1}$$

and call this operator the **resolvent** of T at z.
If T is normal it can be proved that the norm of the resolvent satisfies

$$\|R_z(T)\|_{op} = \frac{1}{\textit{dist}(z, \sigma(\Delta))},$$

where $\textit{dist}(z, \sigma(\Delta))$ denotes the minimal distance between z and any eigenvalue of Δ. In general, one can prove

$$\|R_z(T)\|_{op} \leqslant \gamma_T(z)$$

with

$$\gamma_T(z) = \exp\left[2\|T\|_1 / d(z, \sigma(T))\right] / d(z, \sigma(T))$$

as is proved in Bandtlow (2004).

Frobenius Norm: Given two finite dimensional Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ with orthonormal bases $\{\phi_i^1\}_{i=1}^{d_1}$ and $\{\phi_i^1\}_{i=1}^{d_1}$, the Frobenius norm $\|\cdot\|_F$ of an operator $A : \mathcal{H}_1 \to \mathcal{H}_2$ may be defined as

$$\|A\|_2^2 := \sum_{i=1}^{d_2} \sum_{j=1}^{d_1} |A_{ij}|^2$$

with A_{ij} the matrix representation of A with respect to the bases $\{\phi_i^1\}_{i=1}^{d_1}$ and $\{\phi_i^1\}_{i=1}^{d_1}$. It is a standard exercise to verify that this norm is indeed independent of any choice of basis and hence invariant under multiplying A with a unitary on either the left or the right side. More precisely, if $U : \mathcal{H}_2 \to \mathcal{H}_2$ and $V : \mathcal{H}_1 \to \mathcal{H}_1$ are unitary, we have

$$\|UAV\|_F^2 = \|A\|_F^2.$$

Frobenius norms can be used to transfer Lipschitz continuity properties of complex functions to the setting of functions applied to normal operators:

Lemma A.1. Let $g : \mathbb{C} \to \mathbb{C}$ be Lipschitz continuous with Lipschitz constant D_g. This implies

$$\|g(X)J - Jg(Y)\|_F \leqslant D_g \cdot \|X - Y\|_F.$$

for normal operators X on $\mathcal{H}_2$, Y on $\mathcal{H}_1$ and any linear map $J : \mathcal{H}_1 \to \mathcal{H}_2$.

Proof. This proof is a modified version of the proof in T.P. (2009). Let U, W be unitary (with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$) operators diagonalizing the normal operators X and Y as

$$V^*XV = \text{diag}(\lambda_1, ...\lambda_{d_2}) =: D(X)$$
$$W^*YW = \text{diag}(\mu_1, ...\mu_{d_1}) =: D(Y).$$

Since the Frobenius norm is invariant under unitary transformations we find

$$\begin{aligned}
\|g(X)J - Jg(Y)||_F^2 &= ||g(VD(X)V^*) - g(WD(Y)W^*)\|_F^2 \\
&= \|Vg(D(X))V^*J - JWg(D(Y))W^*\|_F^2 \\
&= \|g(D(X))V^*JW - V^*JWg(D(Y))\|_F^2 \\
&= \sum_{i,j} |(g(D(X))V^*JW - V^*JWg(D(Y)))_{ij}|^2 \\
&= \sum_{i,j} \left| \sum_k [g(D(X))]_{ik}[V^*JW]_{kj} - [V^*JW]_{ik}[g(D(Y))]_{kj} \right|^2 \\
&= \sum_{i,j} |[V^*W]_{ij}|^2 \, |g(\lambda_j) - g(\mu_i)|^2 \\
&\leqslant \sum_{i,j} |[V^*W]_{ij}|^2 \, D_g^2 |\lambda_j - \mu_i|^2 \\
&= D_g^2 \|X - Y\|_F^2.
\end{aligned}$$

□

B Approximating bounded continuous filters

Let us recall Definition 2.1:

Definition B.1. Fix $\omega \in \mathbb{C}$ and $C > 0$. Define the space $\mathscr{F}^{cont}_{\omega,C}$ of continuous filters on $\mathbb{C}\backslash\{\omega, \overline{\omega}\}$, to be the space of multilinear power-series' $g(z) = \sum_{\mu,\nu=0}^{\infty} a_{\mu\nu} (\omega - z)^{-\mu} (\overline{\omega} - \overline{z})^{-\mu}$ for which the norm $\|g\|_{\mathscr{F}^{cont}_{\omega,C}} := \sum_{\mu,\nu=0}^{\infty} |\mu + \nu| C^{\mu+\nu} |a_{\mu\nu}|$ is finite.

We now prove that upon denoting by $B_\epsilon(\omega) \subseteq \mathbb{C}$ the open ball of radius ϵ around ω, one can show that for arbitrary $\delta > 0$ and every continuous function g defined on $\mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$ which is regular at infinity – i.e. satisfies $\lim_{r\to+\infty} g(rz) = c \in \mathbb{C}$ independent of which $z \neq 0$ is chosen – there is a function $f \in \mathscr{F}^{cont}_{\omega,C}$ so that $|f(z) - g(z)| \leqslant \delta$ for all $z \in \mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$.
Making use of the Stone-Weierstrass theorem for complex functions, it suffices to prove that for every point z in $\mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$ there are functions f and g in $\mathscr{F}^{cont}_{\omega,C}$ for which

$$f(z) \neq g(z).$$

But this is obvious since $(\omega - z)^{-1}$ is injective on $\mathbb{C}\backslash(B_\epsilon(\omega) \cup B_\epsilon(\overline{\omega}))$.

C Complex Analysis

A general reference for topics discussed in this section is Bak & Newman (2017).
For a complex valued function f of a single complex variable, the derivative of f at a point $z_0 \in \mathbb{C}$ in its domain of definition is defined as the limit

$$f' z_0 := \lim_{z\to z_0} \frac{f(z) - f(z_0)}{z - z_0}.$$

For this limit to exist, it needs to be independent of the 'direction' in which z approaches z_0, which is a stronger requirement than being real-differentiable. A function is called holomorphic on an open set U os ot is complex differentiable at every point in U. It is called entire if it is complex differentiable at every point in $\mathbb{C}$. Every entire function has an everywhere convergent power series representation

$$g(z) = \sum_{k=0}^{\infty} a^g z^k. \tag{5}$$

If a function g is analytic (i.e. can be expanded into a power series), we have

$$g(\lambda) = -\frac{1}{2\pi i} \oint_S \frac{g(z)}{\lambda - z} dz \tag{6}$$

for any circle $S \subseteq \mathbb{C}$ encircling λ by Cauchy's integral formula.
In fact, the integration contour need not be a circle S, but may be the boundary of any so called Cauchy domain containing λ:

Definition C.1. A subset D of the complex plane $\mathbb{C}$ is called a Cauchy domain if D is open, has a finite number of components (the closure of two of which are disjoint) and the boundary of ∂D of D is composed of a finite number of closed rectifiable Jordan curves, no two of which intersect.

Equation (6) forms the backbone of complex analysis. Since the integral

$$I := -\frac{1}{2\pi i} \oint_{\partial D} g(z)(zId - T)^{-1} dz \tag{7}$$

is well defined for holomorphic $g(\cdot)$ and any operator T for which $\sigma(T)$ and ∂D are disjoint (c.f. e.g. Post (2012) for details), we can essentially take (7) as a defining equation through which one might apply holomorphic functions to operators.

While functions that are everywhere complex differentiable have a series representation according to (5), complex functions that are holomorphic only on $\mathbb{C}\backslash\{\omega\}$ have a series representation (called Laurent series) according to

$$g(z) = \sum_{k=-\infty}^{\infty} a_k (z-\omega)^k.$$

If these functions are assumed to be regular at infinity, no positive terms are permitted and (changing the indexing) we may thus write

$$g(z) = \sum_{k=0}^{\infty} a_k (z-\omega)^{-k}.$$

Motivated by this, we now prove the following consistency result:

Lemma C.2. With the notation of Section 2 we have for any $k \geqslant 1$ and $\omega \notin \sigma(T)$ that

$$(\omega - T)^{-k} := \frac{1}{2\pi i} \oint_{\partial D} (\omega - z)^{-k} \cdot (zId - T)^{-1} dz,$$

where we interpret the left hand side of the equation in terms of inversion and matrix powers.

Proof. We first note that we may write

$$R_\lambda(T) = \sum_{n=0}^{\infty} (\lambda - \omega)^n (-1)^n R_\omega(t)^{n+1}$$

for $|\lambda - \omega| \leqslant \|R_\omega(T)\|$ using standard results in matrix analysis (namely 'Neumann Characterisation of the Resolvent' which is obtained by repeated application of a resolvent identity; c.f. Post (2012) for more details). We thus find

$$\frac{1}{2\pi i} \oint_{\partial D} \left(\frac{1}{\omega - z}\right)^k \frac{1}{zId - T} dz = \frac{1}{2\pi i} \oint_{\partial D} \left(\frac{1}{\omega - z}\right)^k \sum_{n=0}^{\infty} (\omega - z)^n R_\omega(T)^{n+1}.$$

Using the fact that

$$\frac{1}{2\pi i} \oint_{\partial D} (z - \omega)^{n-k-1} dz = \delta_{nk}$$

then yields the claim. $\square$

D Proof of Lemma 2.3

We want to prove the following:

Lemma D.1. For holomorphic g and generic T we have $\|g(T)\|_{op} \leqslant |g(\infty)| + \frac{1}{2\pi} \oint_{\partial D} |g(z)| \gamma_T(z) d|z|$. Furthermore we have for any T with $\gamma_T(\omega) \leqslant C$, that $\|g(T)\|_{op} \leqslant \|g\|_{\mathscr{F}^{hol}_{\omega,C}}$ as long as $g \in \mathscr{F}_{C,\omega}$.

Proof. We first note

$$\left\| g(\infty)\cdot Id + \frac{1}{2\pi i}\oint\limits_{\partial D} g(z)\cdot(zId-T)^{-1}dz \right\| \leqslant \|g(\infty)\cdot Id\|_{op} + \left\| \frac{1}{2\pi i}\oint\limits_{\partial D} g(z)\cdot(zId-T)^{-1}dz \right\|_{op}$$
$$\leqslant |g(\cdot)| + \frac{1}{2\pi}\oint\limits_{\partial D} |g(z)| \left\|\cdot(zId-T)^{-1}\right\|_{op} d|z|.$$

The first claim thus follows together with $\|R_z(T)\|_{op} \leqslant \gamma_T(z)$. The second claim can be derived as follows:

$$\|g(T)\|_{op} = \left\| \sum_{k=0}^{\infty} b_k^g (T-\omega)^{-k} \right\|_{op} \leqslant \sum_{k=0}^{\infty} |b_k^g| \left\|(T-\omega)^{-k}\right\|_{op} \leqslant \sum_{k=0}^{\infty} |b_k^g| \gamma_T(\omega)^k \leqslant \sum_{k=0}^{\infty} |b_k^g| C^k.$$

□

E Proof of Theorem 3.1

. We want to prove the following:

Theorem E.1. With the notation of Section 2 let $\Phi_N : \mathscr{L}_{\text{in}} \to \mathscr{L}_{\text{out}}$ be the map associated to an N-layer GCN. We have

$$\|\Phi_N(f) - \Phi_N(h)\|_{\mathscr{L}_{\text{out}}} \leqslant \left(\prod_{n=1}^{N} L_n R_n B_n\right) \cdot \|f-h\|_{\mathscr{L}_{\text{in}}}$$

with $B_n := \sqrt{\sup_{\lambda\in\sigma(T_n)} \sum_{j\in K_{n-1}} \sum_{i\in K_n} |g_{ij}^n(\lambda)|^2}$ if T_n is normal. For general T_n we have for all $\{g_{ij}\}$ entire, holomorphic and in $\mathscr{F}_{\omega,C}$ respectively:

$$B_n := \begin{cases} \sum\limits_{k=0}^{\infty} \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} |(a_{ij}^{g_n})_k|^2} \cdot \|T_n\|_{op}^k \\ \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} \|g_{ij}^n(\infty)\|^2} + \frac{1}{2\pi}\oint_\Gamma \gamma_T(z) \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} |g_{ij}^n(z)|^2} d|z| \\ \sqrt{\sum_{j\in K_{n-1}} \sum_{i\in K_n} \|g_{ij}^n\|_{\omega,C}^2} \end{cases}$$

Proof. Given input signals $f, h^n \in \mathscr{L}_{\text{in}}$, let us – sticking to the notation introduced in Section 2 – denote the intermediate signal representations in the intermediate layers $\mathscr{L}_n$ by $f^n, h^n \in \mathscr{L}_n$. With the update rule described in Section 2 and the norm induced on each $\mathscr{L}_n$ as described in Appendix A, we then have

$$\begin{aligned}
&\|f^{n+1} - h^{n+1}\|_{\mathscr{L}_{n+1}}^2 \\
=&\sum_{i=1}^{K_{n+1}} \left\| \rho_{n+1}\left(\sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(f_j^n)\right) - \rho_{n+1}\left(\sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(h_j^n)\right) \right\|_{\ell^2(G_{n+1})}^2 \\
\leqslant& L_{n+1}^2 \sum_{i=1}^{K_{n+1}} \left\| \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(f_j^n) - \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(h_j^n) \right\|_{\ell^2(G_{n+1})}^2 \\
=& L_{n+1}^2 \sum_{i=1}^{K_{n+1}} \left\| \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})\left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \right\|_{\ell^2(G_{n+1})}^2 .
\end{aligned}$$

We next note

$$\begin{aligned}
&\sum_{i=1}^{K_{n+1}} \left\| \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1}) \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \right\|^2_{\ell^2(G_{n+1})} \\
\leqslant &\sum_{i=1}^{K_{n+1}} \left(\sum_{j=1}^{K_n} \|g_{ij}^{n+1}(T_{n+1})\|_{op} \| \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \|_{\ell^2(G_{n+1})} \right)^2 \\
\leqslant &\left(\sum_{i=1}^{K_{n+1}} \sum_{j=1}^{K_n} \|g_{ij}^{n+1}(T_{n+1})\|^2_{op} \right) \sum_{j=1}^{K_n} \|\| \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \|^2_{\ell^2(G_{n+1})} \\
\leqslant &R^2_{n+1} \left(\sum_{i=1}^{K_{n+1}} \sum_{j=1}^{K_n} \|g_{ij}^{n+1}(T_{n+1})\|^2_{op} \right) \|\|f^n - h_j^n\|^2_{\mathscr{L}_n}
\end{aligned}$$

where the second to last step is an application of the Cauchy Schwarz inequality.
Proceeding inductively and using our previously established estimates, this proves the claim for all settings in which T_n is nor normal (using an additional application of the triangle inequality for the case of holomorphic filters).

To prove the claim for normal T_n as well, we note that in this setting we have (writing $(\phi_\alpha, \lambda_\alpha)_{\alpha=1}^{|G|}$ for a normalozed eigenvalue-eigenvector sequence of T_{n+1}) that we have

$$\begin{aligned}
&\sum_{i=1}^{K_{n+1}} \left\| \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1}) \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \right\|^2_{\ell^2(G_{n+1})} \\
= &\sum_{i=1}^{K_{n+1}} \left\| \sum_{j=1}^{K_n} \sum_\alpha g_{ij}^{n+1}(\lambda_\alpha) \langle \phi_\alpha, \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \rangle_{\ell^2(G_{n+1})} \phi_\alpha \right\|^2_{\ell^2(G_{n+1})} \\
= &\sum_{i=1}^{K_{n+1}} \sum_{j=1}^{K_n} \sum_\alpha |g_{ij}^{n+1}(\lambda_\alpha)|^2 |\langle \phi_\alpha, \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \rangle_{\ell^2(G_{n+1})}|^2 \\
\leqslant &\sum_\alpha \left(\sum_{i,j} |g_{ij}(\lambda_\alpha)|^2 \right) \sum_{j=1}^{K_n} |\langle \phi_\alpha, \left[P_{n+1}(f_j^n) - P_{n+1}(h_j^n)\right] \rangle_{\ell^2(G_{n+1})}|^2 \\
\leqslant &B_{n+1} R_{n+1} \|\|f^n - h_j^n\|^2_{\mathscr{L}_n}.
\end{aligned}$$

Here we applied Cauchy Schwarz once more in the second to last step and bounded

$$\left(\sum_{i,j} |g_{ij}(\lambda_\alpha)|^2 \right) \leqslant \left(\sup_{\lambda \in \sigma(T)} \sum_{i,j} |g_{ij}(\lambda)|^2 \right).$$

□

F Proof of Lemma 4.2

We want to prove the following:

Lemma F.1. Let $T, \widetilde{T}$ be operators on on $\ell^2(G)$, $\ell^2(\widetilde{G})$ with $\|T\|_{op}, \|\widetilde{T}\|_{op} \leqslant C$. Let $J : \ell^2(G) \to \ell^2(\widetilde{G})$ be arbitrary but linear. With $K_g = \sum_{k=1}^\infty |a_k^g| k C^{k-1}$ for g entire and $K_g = \frac{1}{2\pi} \oint_{\partial D} \frac{1}{z} \gamma_T(z) \gamma_{\widetilde{T}}(z) |g(z)| d|z|$ for g holomorphic, we have

$$\|g(T)J - Jg(\widetilde{T})\|_{op} \leqslant K_g \cdot \|JT - \widetilde{T}J\|_{op}$$

Proof. Let us first verify the claim for entire g. We first note that

$$\begin{aligned}
&\widetilde{T}^k J - JT^k = \widetilde{T}^{k-1}(\widetilde{T}J - JT) + (\widetilde{T}^{k-1}J - JT^{k-1})T \\
= &\widetilde{T}^{k-1}(\widetilde{T}J - JT) + \widetilde{T}^{k-2}(\widetilde{T}J - JT)T + (\widetilde{T}^{k-2}J - JT^{k-2})T^2.
\end{aligned}$$

Thus, with $\|T\|_{op}, \|\widetilde{T}\|_{op} \leqslant C$ we find

$$\|\widetilde{T}^k J - JT^k\|_{op} \leqslant kC^{k-1}\|\widetilde{T}J - JT\|_{op}.$$

The claim now follows from applying the triangle inequality.
Now let us prove the bound for holomorphic g. We first note the following:

$$\begin{aligned}
&\frac{1}{\widetilde{T}-z}(\widetilde{T}J - JT)\frac{1}{T-z}\\
=&\frac{1}{\widetilde{T}-z}\widetilde{T}J\frac{1}{T-z} - \frac{1}{\widetilde{T}-z}JT\frac{1}{T-z}\\
=&\left[\frac{1}{\widetilde{T}-z}(\widetilde{T}-z)J + \frac{z}{\widetilde{T}-z}\right]\frac{1}{T-z} - \frac{1}{\widetilde{T}-z}\left[\frac{1}{T-z}(T-z)J + \frac{z}{T-z}\right]\\
=&z\left(J\frac{1}{T-z} - \frac{1}{\widetilde{T}-z}J\right).
\end{aligned}$$

Thus we have

$$\|g(\widetilde{T})J - Jg(T)\|_{op} \leqslant \frac{1}{2\pi}\oint_{\partial D}\frac{1}{z}\|R_z(T)\|_{op}\|R_z(\widetilde{T})\|_{op}|g(z)|d|z| \leqslant \frac{1}{2\pi}\oint_{\partial D}\frac{1}{z}\gamma_T(z)\gamma_{\widetilde{T}}(z)|g(z)|d|z|.$$

□

G Proof of Theorem 4.3

We prove the following generalization of Theorem 4.3:

Theorem G.1. Let $\Phi_N, \widetilde{\Phi}_N$ be the maps associated to N-layer graph convolutional networks with the same non-linearities and functional calculus filters, but based on different graph signal spaces $\ell^2(G), \ell^2(\widetilde{G})$, characteristic operators $T_n, \widetilde{T}_n$ and connecting operators $P_n, \widetilde{P}_n$. Assume $B_n, \widetilde{B}_n \leqslant B$ as well as $R_n, \widetilde{R}_n \leqslant R$ and $L_n \leqslant L$ for some $B, R, L > 0$ and all $n \geqslant 0$. Assume that there are identification operators $J_n : \ell^2(G_n) \to \ell^2(\widetilde{G}_n)$ $(0 \leqslant n \leqslant N)$ almost commuting with non-linearities and connecting operators in the sense of $\|\widetilde{P}_n J_{n-1} f - J_n P_n f\|_{\ell^2(\widetilde{G}_n)} \leqslant \delta_2 \|f\|_{\ell^2(G_n)}$ and $\|\rho_n(J_n f) - J_n\rho_n(f)\|_{\ell^2(\widetilde{G}_n)} \leqslant \delta_1\|f\|_{\ell^2(G_n)}$. Depending on whether normal or arbitrary characteristic operators are used, define $D_n^2 := \sum_{j\in K_{n-1}}\sum_{i\in K_n} D^2_{g^n_{ij}}$ or $D_n^2 := \sum_{j\in K_{n-1}}\sum_{i\in K_n} K^2_{g^n_{ij}}$. Choose D such that $D_n \leqslant D$ for all n. Finally assume that $\|J_nT_n - \widetilde{T}_nJ_n\|_* \leqslant \delta$ and with $* = F$ if both operators are normal and $* = op$ otherwise. Then we have for all $f \in \mathscr{L}_{\text{in}}$ and with $\mathscr{J}_N$ the operator that the K_N copies of J_N induced through concatenation that

$$\|\widetilde{\Phi}(J_0 f) - \mathscr{J}_N\Phi(f)\|_{\widetilde{\mathscr{L}}_{\text{out}}} \leqslant N\cdot[RLD\delta + \delta_1 BR + \delta_2 BL]\cdot B^{N-1}\cdot\|f\|_{\mathscr{L}_{\text{in}}}.$$

Proof. For simplicity in notation, let us denote the hidden representation of $J_0 f$ in $\widetilde{\mathscr{L}_n}$ by $\widetilde{f}^n$. We then note the following

$$\begin{aligned}
&\|\mathscr{J}_{n+1}f^{n+1} - \widetilde{f}^{n+1}\|_{\widetilde{\mathscr{L}}_{n+1}}\\
=&\left(\sum_{i=1}^{K_{n+1}}\left\|J_{n+1}\rho_{n+1}\left(\sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})P_{n+1}(f^n_j)\right) - \rho_{n+1}\left(\sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})\widetilde{P}_{n+1}(\widetilde{f}^n_j)\right)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}}\\
\leqslant&\left(\sum_{i=1}^{K_{n+1}}\left\|J_{n+1}\rho_{n+1}\left(\sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})P_{n+1}(f^n_j)\right) - \rho_{n+1}\left(J_{n+1}\sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})P_{n+1}(f^n_j)\right)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}}\\
&+L\left(\sum_{i=1}^{K_{n+1}}\left\|J_{n+1}\sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})P_{n+1}(f^n_j) - \sum_{j=1}^{K_n} g^{n+1}_{ij}(T_{n+1})\widetilde{P}_{n+1}(\widetilde{f}^n_j)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}}
\end{aligned}$$

We can bound the first term by $\delta_1 B \cdot R \cdot (BRL)^n \cdot \|f\|_{\mathscr{L}_{\text{in}}}$. For the second term we find

$$
\begin{aligned}
&L\left(\sum_{i=1}^{K_{n+1}}\left\|J_{n+1}\sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})P_{n+1}(f_j^n) - \sum_{j=1}^{K_n} g_{ij}^{n+1}(T_{n+1})\widetilde{P}_{n+1}(\widetilde{f}_j^n)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}} \\
\leqslant &L\left(\sum_{i=1}^{K_{n+1}}\left\|\sum_{j=1}^{K_n}(J_{n+1}g_{ij}^{n+1}(T_{n+1}) - g_{ij}^{n+1}(\widetilde{T}_{n+1})J_{n+1})P_{n+1}(f_j^n)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}} \\
+&LB\left(\sum_{j=1}^{K_n}\left\|J_{n+1}P_{n+1}(f_j^n) - \widetilde{P}_{n+1}(\widetilde{f}_j^n)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}}
\end{aligned}
$$

Arguing as in the proof of 3.1 we can bound the first term by $LD \cdot \delta R \cdot (BRL)^n \|f\|_{\mathscr{L}_{\text{in}}}$. For the second term we find,

$$
\begin{aligned}
&LB\left(\sum_{j=1}^{K_n}\left\|J_{n+1}P_{n+1}(f_j^n) - \widetilde{P}_{n+1}(\widetilde{f}_j^n)\right\|^2_{\ell^2(G_{n+1})}\right)^{\frac{1}{2}} \\
\leqslant\ &LB\delta_2(BRL)^n + \|\mathscr{J}_n f^n - \widetilde{f}^n\|_{\widetilde{\mathscr{L}}_n}
\end{aligned}
$$

arguing as above. Iterating from $n = N$ to $n = 0$ then yields the claim. □

H Proofs of Lemma 5.3 and Theorem 5.4

Lemma H.1. Let T and $\widetilde{T}$ be characteristic operators on $\ell^2(G)$ and $\ell^2(\widetilde{G})$ be respectively. If these operators are ω-δ-close with identification operator J, and $\|R_\omega\|_{op}, \overline{R}_\omega\|_{op} \leqslant C$ we have

$$\|Jg(T) - g(\widetilde{T})J\|_{op} \leqslant K_g \cdot \|(\widetilde{R}_\omega J - JR_\omega)\|_{op}$$

with $K_g = \oint_{\partial D}(1 + |z-\omega|\gamma_T(z))(1 + |z-\omega|\gamma_{\widetilde{T}}(z))|g(z)|d|z|$ if g is holomorphic and $K_g = \|g\|_{\mathscr{F}^{hol}_{\omega,C}}$ if $g \in \mathscr{F}^{hol}_{\omega,C}$. If T and $\widetilde{T}$ are normal as well as doubly ω-δ-close and $g \in \mathscr{F}^{cont}_{\omega,C}$, we have $K_g = \|g\|_{\mathscr{F}^{cont}_{\omega,C}}$.

Proof. We first deal with the statement concerning holomorphic g. To this end we note that Lemma 4.5.9 of Post (2012) proves

$$\|\widetilde{R}_z J - JR_z\|_{op} \leqslant (1 + |z-\omega|\gamma_T(z))(1 + |z-\omega|\gamma_{\widetilde{T}}(z)) \cdot \|\widetilde{R}_\omega J - JR_\omega\|_{op}.$$

The claim then follows from

$$\|Jg(T) - g(\widetilde{T})J\|_{op} \leqslant \frac{1}{2\pi}\oint_{\partial D}|g(z)|\|\widetilde{R}_z J - JR_z\|_{op} d|z|.$$

For $g \in \mathscr{F}^{hol}_{\omega,C}$ the claim is proved exactly as in the proof of Lemma 2.3.
For $g \in \mathscr{F}^{cont}_{\omega,C}$ we note that

$$(\widetilde{R}_\omega)^\mu(\widetilde{R}^*_\omega)^\nu J - J\,(R_\omega)^\mu\,(R^*_\omega)^\nu = (\widetilde{R}_\omega)^\mu\left[(\widetilde{R}^*_\omega)^\nu J - J\,(R^*_\omega)^\nu\right] + \left[(\widetilde{R}_\omega)^\mu J - J(R_\omega)^\mu\right](R^*_\omega)^\nu.$$

Together with the result

$$\|\widetilde{T}^k J - JT^k\|_{op} \leqslant kC^{k-1}\|\widetilde{T}J - JT\|_{op}.$$

established in the proof of Lemma 4.2, the claim then follows from the triangle inequality together with the definition of the pseudo-norm $\|g\|_{\mathscr{F}^{cont}_{\omega,C}}$.

□

As in the previous section, we state a slightly more general version of our main theorem of this section:

Theorem H.2. Let $\Phi, \widetilde{\Phi}$ be the maps associated to N-layer graph convolutional networks with the same non-linearities and functional calculus filters, but based on different graph signal spaces $\ell^2(G_n), \ell^2(\widetilde{G}_n)$, characteristic operators $T_n, \widetilde{T}_n$ and connecting operators $P_n, \widetilde{P}_n$. Assume $B_n, \widetilde{B}_n \leqslant B$ as well as $R_n, \widetilde{R}_n \leqslant R$ and $L_n \leqslant L$ for some $B, R, L > 0$ and all $n \geqslant 0$. Assume that there are identification operators $J_n : \ell^2(G_n) \to \ell^2(\widetilde{G}_n)$ $(0 \leqslant n \leqslant N)$ almost commuting with non-linearities and connecting operators in the sense of $\|\widetilde{P}_n J_{n-1} f - J_n P_n f\|_{\ell^2(\widetilde{G}_n)} \leqslant \delta_2 \|f\|_{\ell^2(G_n)}$ and $\|\rho_n(J_n f) - J_n \rho_n(f)\|_{\ell^2(\widetilde{G}_n)} \delta_1 \|f\|_{\ell^2(G_n)}$. define $D_n^2 := \sum_{j \in K_{n-1}} \sum_{i \in K_n} K^2_{g^n_{ij}}$ with $K_{g^n_{ij}}$ as in Lemma 5.3. Choose D such that $D_n \leqslant D$ for all n. Finally assume that $\|J_n(\omega Id - T_n)^{-1} - (\omega Id - \widetilde{T}_n)^{-1} J_n\|_{op} \leqslant \delta$. If filters in $\mathscr{F}^{cont}_{\omega,C}$ are used, assume additionally that $\|J_n((\omega Id - T_n)^{-1})^* - ((\omega Id - \widetilde{T}_n)^{-1})^* J_n\|_{op} \leqslant \delta$. Then we have for all $f \in \mathscr{L}_{\text{in}}$ and with $\mathscr{J}_N$ the operator that the K_N copies of J_N induced through concatenation that

$$\|\widetilde{\Phi}(J_0 f) - \mathscr{J}_N \Phi(f)\|_{\widetilde{\mathscr{L}}_{\text{out}}} \leqslant N \cdot [RLD\delta + \delta_1 BR + \delta_2 BL] \cdot B^{N-1} \cdot \|f\|_{\mathscr{L}_{\text{in}}}.$$

Proof. The proof proceeds in complete analogy to the one of Theorem 4.3. ☐

I COLLAPSING STRONG EDGES: PROOFS AND FURTHER DETAILS

We use the utilize the notation introduced in Section 5.2. Beyond this, we write denote the inner product induced by the energy functional $E_{\widetilde{G}}$ by

$$E_{\widetilde{G}}(u, v) := \langle u, \Delta_G v \rangle_{\ell^2(\widetilde{G})}.$$

We further use the notation $E_{\widetilde{G}}(u) := E_{\widetilde{G}}(u, u)$. Similar considerations apply for G.

Solving the convex optimization program (4), one finds that the 'Greek entries' of the vector ψ_g are given explicitly by

$$\begin{pmatrix} \psi_g(\alpha) \\ \psi_g(\beta) \\ \vdots \end{pmatrix} = \begin{pmatrix} \widetilde{d}_\alpha & -\widetilde{W}_{\alpha\beta} & \dots \\ -\widetilde{W}_{\beta\alpha} & \widetilde{d}_\beta & \vdots \\ \vdots & \dots & \ddots \end{pmatrix}^{-1} \cdot \begin{pmatrix} \widetilde{W}_{g\alpha} \\ \widetilde{W}_{g\beta} \\ \vdots \end{pmatrix},$$

with degrees in $\widetilde{G}$ denoted by $\widetilde{d}_\alpha$.

To see that all entries of Ψ_g are non-negative, we note the following: Let "u" be a graph signal in $\ell^2(\widetilde{G})$. Denote by captial indices those indices that correspond to negative entries $(u_A)_A$. Denote by lowercase indices those indices that correspond to non-negative entries $(u_a)_a$. Then we have

$$E_{\widetilde{G}}(u) = \sum_{ab} \widetilde{W}_{ab} |u(a) - u(b)|^2 + \sum_{ab} \widetilde{W}_{AB} |u(A) - u(B)|^2 + 2 \sum_{Aa} \widetilde{W}_{Aa} |u(A) - u(a)|^2.$$

Under the mapping $u(g) \to |u(g)|$ the first two terms are invariant while final one is decreasing. Since ψ_g is obtained in a minimization procedure, all entries are thus non-negative.
In fact, one can even prove that the ψ_gs form a partition of unity in the sense of

$$\sum_{g \in G} \psi_g = \mathbb{1}_{\widetilde{G}},$$

with $\mathbb{1}_{\widetilde{G}}$ the vector of all ones in $\ell^2(\widetilde{G})$ (Post & Simmer, 2017).
Given a degree $\widetilde{d}_\alpha$ corresponding to a Greek index, we decompose it as

$$\widetilde{d}_\alpha = \widetilde{d}^r_\alpha + \widetilde{W}_{\alpha\star} + V_\alpha$$

with $\widetilde{d}^r_\alpha$ accounting for edges from α to other greek vertices

$$\widetilde{d}^r_\alpha = \sum_{\beta \in \widetilde{G}_{Greek}} \widetilde{W}_{\alpha\beta} = \frac{1}{\delta} \sum_{\beta \in \widetilde{G}_{Greek}} \omega_{\alpha\beta},$$

and V_α accounting for edges from α to Latin vertices

$$V_\alpha = \sum_{a \in \widetilde{G}_{Latin}} \widetilde{W}_{a\alpha}.$$

Recall that we also may write

$$\widetilde{W}_{\alpha\star} = \frac{1}{\delta}\omega_{\alpha\star}.$$

We may then write

$$\begin{aligned}\begin{pmatrix} \widetilde{d}_\alpha & -\widetilde{W}_{\alpha\beta} & \dots \\ -\widetilde{W}_{\beta\alpha} & \widetilde{d}_\beta & \vdots \\ \vdots & \dots & \ddots \end{pmatrix} &= \begin{pmatrix} \widetilde{d}^r_\alpha & -\widetilde{W}_{\alpha\beta} & \dots \\ -\widetilde{W}_{\beta\alpha} & \widetilde{d}^r_\beta & \vdots \\ \vdots & \dots & \ddots \end{pmatrix} + \frac{1}{\delta}\begin{pmatrix} \omega_{\alpha\star} & 0 & \dots \\ 0 & \omega_{\beta\star} & \vdots \\ \vdots & \dots & \ddots \end{pmatrix} + \begin{pmatrix} V_\alpha & 0 & \dots \\ 0 & V_\beta & \vdots \\ \vdots & \dots & \ddots \end{pmatrix} \\ &=: \frac{1}{\delta}\mathscr{L} + \frac{1}{\delta}diag(\omega_\star) + V,\end{aligned}$$

where we made the obvious definitions for the matrices $\mathscr{L}$ and V and denoted by $\omega_\star$ the vector with entries $\omega_{\alpha\star}$.
We note that

$$(\mathscr{L} + diag(\omega_\star))\eta = \omega_\star$$

is uniquely solved by $\eta = (1, 1, 1, ...)$. Let us use the notation

$$h := \mathscr{L} + diag(\omega_\star).$$

If we denote the restrictions of ψ_g to Greek indices by η_g, we have

$$(h + \delta V)^{-1}\eta_\star = \omega_\star.$$

Since the solution $\eta_\star$ of this equation is known for $\delta = 0$ (where it is given by $\eta_\star = (1, 1, 1, ...)$) and we assume $\delta << 1$, we can find the solution for non-zero δ through perturbation theory. Writing $\eta_\star = \mathbb{1}_{\widetilde{G}_{Greek}} + \Delta$, we find

$$\Delta = (h + \delta V)^{-1}(-\delta V \mathbb{1}_{\widetilde{G}_{Greek}})$$

by standard arguments from perturbation theory. Importantly, one finds

$$\|\Delta\| \leqslant \|(Id + h^{-1})^{-1}\|_{op}\|h^{-1}\|_{op}\|V\|_F \cdot \delta.$$

For δ sufficiently small, one finds

$$\|(Id + h^{-1})^{-1}\|_{op}\|h^{-1}\|_{op} \leqslant \|h^{-1}\|_{op}\frac{1}{1 - \delta\|h^{-1}V\|_{op}} \leqslant 2\|h^{-1}\|_{op}.$$

We define the constant

$$K := 2\|h^{-1}\|^2_{op}\|V\|_F. \tag{8}$$

Having set the scene, we are now ready to prove Theorem 5.4. Following Post & Simmer (2017), instead of checking the conditions of Definition 5.1 and Definition 5.2 it is instead sufficient to check the following, with J $\widetilde{J}$ as defined in Section 5.2 to establish Theorem 5.6:
There are additional operators $J^1 : \ell^2(G) \to \ell^2(\widetilde{G})$ and $\widetilde{J}^1 : \ell^2(\widetilde{G}) \to \ell^2(G)$ so that the following set of equations is satisfied with $\epsilon = \mathcal{O}(\delta^{\frac{1}{2}})$ (for increased readability we henceforth drop the subscripts $\ell^2(\widetilde{G})$ and $\ell^2(G)$ for norms and inner products. Below, we always have $u \in \ell^2(\widetilde{G})$ and $f \in \ell^2(G)$):

$$\|Jf\| \leqslant (1 + \epsilon)\|f\|, \quad |\langle Jf, u\rangle - \langle f, \rangle\widetilde{J}u\rangle| \leqslant \epsilon\|f\| \tag{9}$$

$$\|f - \widetilde{J}Jf\| \leqslant \epsilon\sqrt{\|f\|^2 + E_G(f)}, \quad \|u - J\widetilde{J}u\| \leqslant \epsilon\sqrt{\|u\|^2 + E_{\widetilde{G}}(u)} \tag{10}$$

$$\|J^1 f - Jf\| \leqslant \epsilon\sqrt{\|f\|^2 + E_G(f)}, \quad \|\widetilde{J}u - \widetilde{J}^1 u\| \leqslant \delta\epsilon\sqrt{\|u\|^2 + E_{\widetilde{G}}(u)} \tag{11}$$

$$|E_{\widetilde{G}}(J^1 f, u) - E_G(f, \widetilde{J}^1 u)| \leqslant \epsilon \cdot \sqrt{\|f\|^2 + E_G(f)} \cdot \sqrt{\|u\|^2 + E_{\widetilde{G}}(u)}.$$

We set $J^1 f = Jf$ and $(\widetilde{J}^1 u)(x) = u(x)$.
For the left hand side of (9) we note (using $2ab \leqslant a^2 + b^2$ and the fact that the ψ_g form a partition of unity):

$$\begin{aligned}
\|Jf\|^2 &= \sum_{h,g\in G} \langle \psi_h, \psi_g \rangle \overline{f}(h) f(g) \\
&\leqslant \frac{1}{2}\sum_{h\in G} |f(h)|^2 \sum_{g\in G} \langle \psi_h, \psi_g \rangle + \frac{1}{2}\sum_{g\in G} |f(g)|^2 \sum_{h\in G} \langle \psi_h, \psi_g \rangle \quad = \frac{1}{2}\sum_{h\in G} |f(h)|^2 \langle \psi_h, \mathbb{1}\rangle + \frac{1}{2}\sum_{g\in G} |f(g)|^2 \langle \mathbb{1}, \psi_g \rangle \\
&= \sum_{g\in G} |f(g)|^2 \mu_g \rangle \\
&= \|f\|^2.
\end{aligned}$$

Here the second to last inequality follows from the definition of the weights μ_g. Thus the left hand side of (9) holds with $\epsilon = 0$.
The right hand side of (9) holds trivially, also with $\epsilon = 0$ since we have chose $J^* = \widetilde{J}$.
Now let us check the l.h.s. of (10). We have:

$$(f - \widetilde{J}Jf)(y) = f(y) - \sum_{g\in G} f(g) \frac{\langle \psi_g, \psi_y \rangle}{\mu_y}.$$

Using the constant K defined in (8) and the discussion preceeding this deformation we find together with the definition of the weights μ_g that we have

$$\widetilde{\mu}_g \leqslant \mu_g \leqslant \widetilde{\mu}_g + \delta K \sum_{\alpha \in \widetilde{G}_{Greek}} \widetilde{\mu}_\alpha$$

if $g \neq \star$. We also write $\widetilde{\mu}(\widetilde{G}_{Greek}) := \sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{\mu}_\alpha$. If $G = \star$, we have

$$\widetilde{\mu}_\star + (1-\delta)\widetilde{\mu}(\widetilde{G}_{Greek}) \leqslant \mu_\star \leqslant \widetilde{\mu}_\star + 1\widetilde{\mu}(\widetilde{G}_{Greek}).$$

We next note

$$\langle \psi_x, \psi_y \rangle = \widetilde{\mu}_x \delta_{xy} + \delta^2 \langle (h+\delta V)^{-1} \widetilde{W}_x, \langle (h+\delta V)^{-1}\widetilde{W}_y \rangle$$

with $\widetilde{W}_x$ the vector with entries $\widetilde{W}_x(\alpha) = \widetilde{W}_{x\alpha}$.
Thus for $y \neq$ we find

$$|(f - \widetilde{J}Jf)(y)| \leqslant (1 - \frac{\widetilde{\mu}_y}{\mu_y})|f(y)| + \left(\langle (h+\delta V)^{-1}\widetilde{W}_x, \langle (h+\delta V)^{-1}\widetilde{W}_y \rangle \right).$$

Thus we find

$$\begin{aligned}
\sum_{\star\neq g\in G} |(f - \widetilde{J}Jf)(y)|^2 &\lesssim \mathcal{O}(\delta^2)\|f\| + \delta^2 \sum_{y\neq\star} \frac{1}{\mu_y} \sum_x \langle (h+\delta V)^{-1}\widetilde{W}_x, \langle (h+\delta V)^{-1}\widetilde{W}_y \rangle \\
&\lesssim \mathcal{O}(\delta^2)\|f\|.
\end{aligned}$$

For $y = \star$ we find

$$\begin{aligned}
|(f - \widetilde{J}Jf)(\star)| &= \left| f(\star) - \frac{1}{\mu_\star}\langle \psi_\star, \psi_\star \rangle + \delta \sum_{x\neq\star} f(x) \langle (h+\delta V)^{-1}\widetilde{W}_x, \psi_\star \rangle \right| \\
&\leqslant |f(\star)| \left| 1 - \frac{\langle \psi_\star, \psi_\star \rangle}{\mu(\star)} \right\| + \frac{1}{\mu}\|f\|\delta \left| \langle (h+\delta V)^{-1}\widetilde{W}_x, \psi_\star \rangle \right| \\
&= \mathcal{O}(\delta)\|f\|.
\end{aligned}$$

Thus the left hand side of (10) holds with $\epsilon = \mathcal{O}(\delta)$.
The left hand side of (11) is true with $\epsilon = 0$ by definition.
Let us thus check the right hand side of (11): We have

$$(\widetilde{J} - \widetilde{J}^1)(x) = \frac{1}{\mu_x}\langle u, \psi_x \rangle - u(x).$$

For $x = *$ we then have in the limit $\delta \to 0$

$$\begin{aligned}\frac{1}{\mu_\star}\langle u, \psi_\star\rangle - u(\star) &\longrightarrow \frac{1}{\widetilde{\mu}_\star + \widetilde{\mu}(\widetilde{G}_{Greek})} \sum_{g \in \widetilde{G}_{Greek} \cup \{\star\}} |u(g) - u(\star)| \\ &\lesssim \delta \sqrt{E_{\widetilde{G}}(u)}.\end{aligned}$$

Let us denote quantities obtained in the limit $\delta \to 0$ by a superscript of ∞ (e.g. $\psi_\star^\infty$, $\mu_\star^\infty$,...). Hence we need to bound

$$\begin{aligned}&\left|\frac{1}{\mu_x}\langle u, \psi_\star\rangle - \frac{1}{\mu_x^\infty}\langle u, \psi_\star^\infty\rangle\right| \leqslant \|u\| \cdot \|\psi_\star/\mu_\star - \psi_\star^\infty/\mu_\star^\infty\| \\ &\leqslant \|u\| \left|1/\mu_\star - 1/\mu_\star^\infty\right| \sqrt{\mu_\star} + \|u\|\mathcal{O}(\delta) \\ &\lesssim \|u\|\mathcal{O}(\delta).\end{aligned}$$

For $x \neq \star$: a similar calculation yields

$$\sum_x \left|\frac{1}{\mu(x)}\langle u, \psi_x\rangle - u(x)\right|^2 \mu_x \leqslant \left(\max\left(\mu_x^{-\frac{1}{2}}\right)\right) \|u\| \widetilde{\mu}(\widetilde{G}_{Greek}) \cdot k \cdot \delta.$$

Thus (11) is true with $\epsilon = \mathcal{O}(\sqrt{\delta})$.

Hence let us now check the right hand side of (10). We have the following:

$$(u - J\widetilde{J}u) = u - \sum_{x \in G} \frac{\langle \psi_x, u\rangle}{\mu_x}\psi_x.$$

Carefully investigating this equation, one finds that the matrix representation of the operator $Id - J\widetilde{J}$ in the limit $\delta \to 0$ is given by a matrix

$$Id - J\widetilde{J} \longrightarrow \begin{pmatrix} 0 & 0 \\ 0 & M^\infty \end{pmatrix}$$

with

$$M^\infty_{\alpha\beta} = \widetilde{\mu}_\alpha \delta_{\alpha\beta} - \frac{\widetilde{\mu}_\alpha \widetilde{\mu}_\beta}{\widetilde{\mu}(\widetilde{G}_{Greek})}.$$

Squaring this matrix yields a matrix $\Delta = (M^\infty)^2$ that is a graph Laplacian belonging to the adjacency matrix

$$C_{\alpha\beta} = \frac{\widetilde{\mu}_\alpha \widetilde{\mu}_\beta^2 + \widetilde{\mu}_\alpha^2 \widetilde{\mu}_\beta}{\widetilde{\mu}(\widetilde{G}_{Greek})} + \frac{\widetilde{\mu}_\alpha \widetilde{\mu}_\beta}{\widetilde{\mu}(\widetilde{G}_{Greek})^2}\left(\sum_{\alpha \in \widetilde{G}_{Greek}} \widetilde{\mu}_\alpha^2\right).$$

Thus in the limit $\delta \to 0$ we have

$$\begin{aligned}\|(Id - J\widetilde{J})u\|^2 &= \sum_{\alpha\beta \in \widetilde{G}_{Greek}} C_{\alpha\beta}|u(\alpha) - u(\beta)| \\ &\lesssim \mathcal{O}(\delta)\sqrt{E_{\widetilde{G}}(u)}.\end{aligned}$$

For the remainder term; establishing

$$\left\|\sum_{x \in G} \frac{\langle \psi_x, u\rangle}{\mu_x}\psi_x - \sum_{x \in G} \frac{\langle \psi_x^\infty, u\rangle}{\mu\infty_x}\psi_x^\infty\right\| \lesssim \delta\|u\| \tag{12}$$

proceeds by noticing that the matrix representation of the operator applied to u in (12) is given as a matrix

$$\begin{pmatrix} C & A \\ A* & M - M^\infty \end{pmatrix}$$

and investigating the behaviour of its constituents independently:
The matrix C is easily seen to be diagonal with entries $\propto(1/\mu_x - 1/\widetilde{\mu}_x)$ so that we may argue as before to establish that $\|C\|_{op}\|u\| = \mathcal{O}(\delta)\|u\|$. The object $(M - M^\infty)$ is explicitly given as

$$M - M^\infty = \sum_{x\neq\star} \frac{|\psi_x|_{\widetilde{G}_{Greek}\cup\{\star\}}\rangle\langle\psi_x|_{\widetilde{G}_{Greek}\cup\{\star\}}|}{\mu_x} + \frac{|\psi_\star|_{\widetilde{G}_{Greek}\cup\{\star\}}\rangle\langle\psi_\star|_{\widetilde{G}_{Greek}\cup\{\star\}}|}{\mu_\star} - \frac{|\psi^\infty_\star|_{\widetilde{G}_{Greek}\cup\{\star\}}\rangle\langle\psi^\infty_\star|_{\widetilde{G}_{Greek}\cup\{\star\}}|}{\mu^\infty_\star}$$

The second term can be bounded (in operator norm) in terms of δ and $\|u\|$ as before.
Entries in the first term are given as $\sum_{g\neq\star}\frac{1}{\mu(x)}\psi_g(\alpha)\psi_g(\beta)$ which are of order δ.
It remains to investigate the matrix A. We find that in its matrix representation the matrix A is given as

$$A = \begin{pmatrix} 0 & \psi_a(\alpha)/\mu_a & \psi_a(\beta)/\mu_a & \dots \\ 0 & \psi_b(\alpha)/\mu_b & \psi_a(\beta)/\mu_a & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Taking the column sums yields results of $\mathcal{O}\delta$ so that we also in total find $\|A\|_{op} = \mathcal{O}(\delta)$.

It finally only remains to prove that

$$|E_{\widetilde{G}}(J^1 f, u) - E_G(f, \widetilde{J}^1 u)| \leqslant \epsilon \cdot \sqrt{\|f\|^2 + E_G(f)} \cdot \sqrt{\|u\|^2 + E_{\widetilde{G}}(u)}$$

with $\epsilon = \mathcal{O}(\delta)$. We note that the operator associated to the energy E_G is given by

$$(\Delta_G f)(x) = \frac{1}{\mu(x)} \sum_{y\sim_G x} W_{xy}(f(x) - f(y)).$$

Here the notation "$y \sim_G x$" signifies that nodes x and y are connected **within** G through edges with **positive** edge-weights.
Similarly the operator associated to $E_{\widetilde{G}}$ is given by

$$(\Delta_{\widetilde{G}} u)(x) = \frac{1}{\widetilde{\mu}(x)} \sum_{y\sim_{\widetilde{G}} x} \widetilde{W}_{xy}(u(x) - u(y)).$$

Denote as above the restriction of ψ_g to G by $\psi_g\big|_G$.
Then we have by linearity :

$$E_{\widetilde{G}}(J^1 f, u) - E_G(f, \widetilde{J}^1 u) = \sum_{g\in G} \overline{f}(g) \cdot \left[E_{\widetilde{G}}(\psi_g, u) - E_G(\psi_g\big|_G, u\big|_G)\right].$$

We note that we have

$$E_G(\psi_g\big|_G, u\big|_G) = \sum_{y\sim_G x} W_{xy}(u(x) - u(y))$$

on the smaller graph G. On the larger graph $\widetilde{G}$ we have

$$E_{\widetilde{G}}(\psi_g, u) = \sum_{y\sim_{\widetilde{G}} x} \widetilde{W}_{xy}(u(x) - u(y)) + \sum_{\alpha\in\widetilde{G}_{Greek}} \sum_{y\sim_{\widetilde{G}}} \widetilde{W}_{\alpha y}(u(\alpha) - u(y)).$$

Going through tedious algebra, one finds for $x \neq \star$ that

$$E_G(\psi_x\big|_G, u\big|_G) - E_{\widetilde{G}}(\psi_x, u) = \sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{x\alpha}(u(\alpha) - u(\star)) - \sum_{\alpha\in\widetilde{G}_{Greek}} \psi_x(\alpha) \sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha) - u(y)).$$

We may bound the first term as

$$\left|\sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{x\alpha}(u(\alpha)-u(\star))\right| \lesssim \delta\left[\sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{x\alpha}\right] E_{\widetilde{G}}(u)$$

For the second term we find

$$\begin{aligned}\left|\sum_{\alpha\in\widetilde{G}_{Greek}} \psi_x(\alpha)\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha)-u(y))\right| &\leqslant \sum_{\alpha\in\widetilde{G}_{Greek}} \psi_x(\alpha)\left[\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}\right]^{\frac{1}{2}}\cdot\left[\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}|u(\alpha)-u(y)|^2\right]^{\frac{1}{2}}\\ &\leqslant K\delta\left[\sum_{\alpha\in\widetilde{G}_{Greek}} \sqrt{d_\alpha}\right]\cdot\sqrt{E_{\widetilde{G}}(u)}\\ &\lesssim \sqrt{\delta}\sqrt{E_{\widetilde{G}}(u)}.\end{aligned}$$

It thus remains to bound the $x=\star$ case. To this end we note

$$E_G(\psi_\star\big|_G, u\big|_G) = \sum_{y\sim_G\star} W_{\star y}(u(\star)-u(y))$$

and

$$\begin{aligned}E_{\widetilde{G}}(\psi_\star, u) = &\sum_{y\sim_{\widetilde{G}}\star} W_{\star y}(u(\star)-u(y))\\ &+\sum_{\alpha\in\widetilde{G}_{Greek}} \psi_\star(\alpha)\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha)-u(y)).\end{aligned}$$

Going through a lot of tedious algebra once more, one finds

$$\begin{aligned}E_G(\psi_\star\big|_G, u\big|_G) - E_{\widetilde{G}}(\psi_\star, u) = &\left[\sum_{\alpha\in\widetilde{G}_{Greek}} (1-\psi_\star(\alpha))\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha)-u(y))\right]\\ &+\sum_{y\sim_G\star}\left(\sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{y\alpha}(u(\star)-u(y))\right).\end{aligned}$$

For the first term we find

$$\begin{aligned}\left|\sum_{\alpha\in\widetilde{G}_{Greek}} (1-\psi_\star(\alpha))\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha)-u(y))\right| &\leqslant \delta\cdot K|\widetilde{G}_{Greek}|\cdot\left|\sum_{\alpha\in\widetilde{G}_{Greek}} (1-\psi_\star(\alpha))\sum_{y\sim_{\widetilde{G}}\alpha} \widetilde{W}_{y\alpha}(u(\alpha)-u(y))\right|\\ &\leqslant \delta\cdot K|\widetilde{G}_{Greek}|\cdot\left[\sum_{\alpha\in\widetilde{G}_{Greek}} \sqrt{d_\alpha}\right]\sqrt{E_{\widetilde{G}}(u)}\\ &\lesssim \sqrt{\delta}\sqrt{E_{\widetilde{G}}(u)}.\end{aligned}$$

For the second term we find with similar arguments as before

$$\left|\sum_{y\sim_G\star}\left(\sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{y\alpha}(u(\star)-u(y))\right)\right| \lesssim \delta\left[\sum_{y\sim_{\widetilde{G}}\star}\sum_{\alpha\in\widetilde{G}_{Greek}} \widetilde{W}_{y\alpha}\right]\sqrt{E_{\widetilde{G}}(u)}.$$

J ADDITIONAL DETAILS ON EXPERIMENTAL SETUP

The adjacency matrix of the larger 'un-collapsed' graph $\widetilde{G}$ we consider in Section 6 is given as follows

$$\widetilde{W} = \begin{pmatrix} 0 & 4 & 2 & 10 & 4 & 5 & 6 & 7 \\ 4 & 0 & 17 & 9 & 8 & 9 & 10 & 11 \\ 2 & 17 & 0 & 42 & 12 & 13 & 14 & 15 \\ 10 & 9 & 42 & 0 & 16/\delta & 7/\delta & 18/\delta & 19/\delta \\ 4 & 8 & 12 & 16/\delta & 0 & 6/\delta & 22/\delta & 3/\delta \\ 5 & 9 & 13 & 7/\delta & 6/\delta & 0 & 1/\delta & 90/\delta \\ 6 & 10 & 14 & 18/\delta & 22/\delta & 1/\delta & 0 & 23/\delta \\ 7 & 11 & 15 & 19/\delta & 3/\delta & 90/\delta & 23/\delta & 0 \end{pmatrix}$$

The exceptional vertex $\star$ here carries index "4" ("$\star = 4$"). Node weights are set to unity.