A Fractional Graph Laplacian Approach to Oversmoothing

Sohir Maskey[*1], Raffaele Paolino[*1,2], Aras Bacho[1], and Gitta Kutyniok[1,2]

[1] Ludwig-Maximilians-Universität München, Munich, Germany
[2] Munich Center for Machine Learning (MCML), Munich, Germany

Abstract. Graph neural networks (GNNs) have shown state-of-the-art performances in various applications. However, GNNs often struggle to capture long-range dependencies in graphs due to oversmoothing. In this paper, we generalize the concept of oversmoothing from undirected to directed graphs. To this aim, we extend the notion of Dirichlet energy by considering a directed symmetrically normalized Laplacian. As vanilla graph convolutional networks are prone to oversmooth, we adopt a neural graph ODE framework. Specifically, we propose fractional graph Laplacian neural ODEs, which describe non-local dynamics. We prove that our approach allows propagating information between distant nodes while maintaining a low probability of long-distance jumps. Moreover, we show that our method is more flexible with respect to the convergence of the graph's Dirichlet energy, thereby mitigating oversmoothing. We conduct extensive experiments on synthetic and real-world graphs, both directed and undirected, demonstrating our method's versatility across diverse graph homophily levels. Our code is available at https://anonymous.4open.science/r/fLode-1707.

Keywords: Graph representation learning · neural ODE · fractional Laplacians.

1 Introduction

Graph neural networks (GNNs) (Gori et al., 2005; Scarselli et al., 2009; Bronstein et al., 2017) have emerged as a powerful class of machine learning models capable of effectively learning representations of structured data. GNNs have demonstrated state-of-the-art performance in a wide range of applications, including social network analysis (Monti et al., 2019), molecular property prediction (Gilmer et al., 2017), and recommendation systems (Wang et al., 2018; Fan et al., 2019). The majority of existing work on GNNs has focused on undirected graphs (Defferrard et al., 2016; Kipf et al., 2017; Hamilton et al., 2017), where edges have no inherent direction. However, many real-world systems, such as citation networks, transportation systems, and biological pathways, are inherently directed, necessitating the development of methods explicitly tailored to directed graph structures.

---

[*] The first two authors contributed equally to this work

Despite their success, most existing GNN models struggle to capture long-range dependencies, which can be critical for specific tasks, such as node classification and link prediction, and for specific types of graphs, such as heterophilic graphs. This shortcoming also arises from the problem of *oversmoothing*, where increasing the depth of GNNs results in the node features converging to similar values that only convey information about the node's degree (Oono et al., 2019; Cai et al., 2020). Consequently, scaling the depth of GNNs is not sufficient to broaden receptive fields, and other approaches are necessary to address this limitation. While these issues have been extensively studied in the context of undirected graphs (Q. Li et al., 2018; G. Li et al., 2019; Luan et al., 2019; D. Chen et al., 2020; Rusch et al., 2022), their implications for directed graphs remain largely unexplored. Investigating these challenges and developing effective solutions is crucial for applying GNNs to real-world scenarios.

Over-smoothing has been shown to be intimately related to the graph's *Dirichlet energy*, defined as

$$\mathscr{E}(\mathbf{x}) := \frac{1}{4} \sum_{i,j=1}^{N} a_{i,j} \left\| \frac{\mathbf{x}_i}{\sqrt{d_i}} - \frac{\mathbf{x}_j}{\sqrt{d_j}} \right\|_2^2,$$

where $\mathbf{A} = (a_{i,j})_{i,j=1}^{N}$ represents the adjacency matrix of the underlying graph, $\mathbf{x} \in \mathbb{R}^{N \times K}$ denotes the node features, and $d_i \in \mathbb{R}$ the degree of node i. Intuitively, the Dirichlet energy measures the smoothness of nodes' features. Therefore, a GNN that minimizes the Dirichlet energy is expected to perform well on *homophilic graphs*, where similar nodes are likely to be connected. Conversely, a GNN that ensures high Dirichlet energy should lead to better performances on *heterophilic* graphs, for which the nodes' features are less smooth.

This paper aims to bridge the gap in understanding oversmoothing for directed graphs. To this aim, we generalize the concept of Dirichlet energy, providing a rigorous foundation for analyzing oversmoothing. Specifically, we consider the directed symmetrically normalized Laplacian, which accommodates directed graph structures and recovers the usual definition in the undirected case. Even though the directed symmetrically normalized Laplacian has been already used (Zou et al., 2022), its theoretical properties remain widely unexplored.

However, a vanilla graph convolutional network (GCN) (Kipf et al., 2017) implementing this directed Laplacian alone is not able to prevent oversmoothing. For this reason, we adopt a graph neural ODE framework, which has been shown to effectively alleviate oversmoothing in undirected graphs (Bodnar et al., 2022; Di Giovanni et al., 2022; Rusch et al., 2022).

1.1   Graph Neural ODEs

The concept of neural ODE was introduced by Haber et al. (2018) and R. T. Q. Chen et al. (2018), who first interpreted the layers in neural networks as the time variable in ODEs. Building on this foundation, Poli et al. (2021), Chamberlain et al. (2021), and Eliasof et al. (2021) extended the connection to the realm of

graph neural networks, resulting in the development of graph neural ODEs. In this context, each node i of the underlying graph is described by a state variable $\mathbf{x}_i(t) \in \mathbb{R}^K$, representing the node i at time t. We can define the dynamics of $\mathbf{x}(t)$ via the node-wise ODE

$$\mathbf{x}'(t) = f_{\mathbf{w}}(\mathbf{x}(t)), \ t \in [0, T],$$

subject to the initial condition $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^{N \times K}$, where the function $f_{\mathbf{w}} : \mathbb{R}^{N \times K} \to \mathbb{R}^{N \times K}$ is parametrized by the learnable parameters $\mathbf{w}$.

The graph neural ODE can be seen as a continuous learnable architecture on the underlying graph, which computes the final node representation $\mathbf{x}(T)$ from the input nodes' features $\mathbf{x}_0$. Typical choices for $f_{\mathbf{w}}$ include attention-based functions (Chamberlain et al., 2021), which generalize graph attention networks (GATs) (Velickovic et al., 2018), or convolutional-like functions (Di Giovanni et al., 2022) that generalize GCNs (Kipf et al., 2017).

How can we choose the learnable function $f_{\mathbf{w}}$ to accommodate both directed and undirected graphs, as well as different levels of homophily? We address this question in the following subsection.

1.2   Fractional Laplacians

The continuous fractional Laplacian, denoted by $(-\Delta)^{\alpha}$ for $\alpha > 0$, is used to model non-local interactions. For instance, the fractional heat equation $\partial_t u + (-\Delta)^{\alpha} u = 0$ provides a flexible and accurate framework for modeling anomalous diffusion processes. Similarly, the fractional diffusion-reaction, quasi-geostrophic, Cahn-Hilliard, porous medium, Schrödinger, and ultrasound equations are more sophisticated models to represent complex anomalous systems (Pozrikidis, 2018).

Similarly to the continuous case, the fractional graph Laplacian (FGL) (Benzi et al., 2020) models non-local network dynamics. In general, the FGL does not inherit the sparsity of the underlying graph, allowing a random walker to leap rather than walk solely between adjacent nodes. Hence, the FGL is able to build long-range connections, making it well-suited for heterophilic graphs.

1.3   Main Contributions

We present a novel approach to the fractional graph Laplacian by defining it in the singular value domain, instead of the frequency domain (Benzi et al., 2020). This formulation bypasses the need for computing the Jordan decomposition of the graph Laplacian, which lacks reliable numerical methods. We show that our version of the FGL can still capture long-range dependencies, and we prove that its entries remain reasonably bounded.

We then propose two FGL-based neural ODEs: the fractional heat equation and the fractional Schrödinger equation. Importantly, we demonstrate that solutions to these FGL-based neural ODEs offer increased flexibility in terms of the convergence of the Dirichlet energy. Notably, the exponent of the fractional graph Laplacian becomes a learnable parameter, allowing our network to adaptively

determine the optimal exponent for the given task and graph. We show that this can effectively alleviate oversmoothing in undirected and directed graphs.

To validate the effectiveness of our approach, we conduct extensive experiments on synthetic and real-world graphs, with a specific focus on supervised node classification. Our experimental results indicate the advantages offered by fractional graph Laplacians, particularly in non-homophilic and directed graphs.

2 Preliminaries

We denote a graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of edges, and $N = |\mathcal{V}|$ is the number of nodes. For each node $i \in \mathcal{V}$, we define its in- and out-neighbourhood as $\mathcal{N}_i^{\text{in}} = \{j \in \mathcal{V} \mid j \to i\}$, $\mathcal{N}_i^{\text{out}} = \{j \in \mathcal{V} \mid i \to j\}$, respectively. We denote the in- and out-degree of node i as $d_i^{\text{in}} = \left|\mathcal{N}_i^{\text{in}}\right|$ and $d_i^{\text{out}} = |\mathcal{N}_i^{\text{out}}|$.

The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ encodes the edge information, with $a_{i,j} = 1$ if there is an edge directed from node j to i, and 0 otherwise. The in- and out-degree matrices are then defined as

$$\mathbf{D}_1 = \operatorname{diag}(\mathbf{A}\mathbf{1}), \; \mathbf{D}_0 = \operatorname{diag}(\mathbf{A}^\mathsf{T}\mathbf{1}),$$

respectively. The node feature matrix $\mathbf{x} \in \mathbb{R}^{N \times K}$ contains for every node its feature in $\mathbb{R}^K$.

Given any matrix $\mathbf{M} \in \mathbb{C}^{N \times N}$, we denote its spectrum by $\lambda(\mathbf{M}) := \{\lambda_i(\mathbf{M})\}_{i=1}^N$ and sort it in ascending order with respect to the real part, i.e., $\Re\lambda_i(\mathbf{M}) \leq \Re\lambda_j(\mathbf{M})$, $i < j$. Furthermore, we denote by $\|\mathbf{M}\|_2$ and $\|\mathbf{M}\|$ the Frobenius and spectral norm of $\mathbf{M}$, respectively.

3 Dirichlet Energy and Laplacian for (Directed) Graphs

In this section, we introduce the concept of Dirichlet energy and demonstrate its relationship to a directed Laplacian, thereby generalizing well-known results for undirected graphs.

Definition 1. *The* Dirichlet energy *is defined on the node features* $\mathbf{x} \in \mathbb{R}^{N \times K}$ *of a graph* $\mathcal{G}$ *as*

$$\mathscr{E}(\mathbf{x}) := \frac{1}{4} \sum_{i,j=1}^N a_{i,j} \left\| \frac{\mathbf{x}_i}{\sqrt{d_i^{in}}} - \frac{\mathbf{x}_j}{\sqrt{d_j^{out}}} \right\|_2^2. \tag{1}$$

The Dirichlet energy of node features $\mathbf{x} \in \mathbb{R}^{N \times K}$ measures how much the features change over the nodes of $\mathcal{G}$. Intuitively, it measures the disparity between the normalized outflow of information from node j and the normalized inflow of information to node i.

Definition 2. *We define the* symmetrically normalized adjacency (SNA) *as* $\mathbf{L} := \mathbf{D}_1^{-1/2} \mathbf{A} \mathbf{D}_0^{-1/2}$, *and the* symmetrically normalized Laplacian (SNL) *as* $\mathbf{\Delta} := \mathbf{I} - \mathbf{L}$.

Note that $\mathbf{L}$ and $\mathbf{\Delta}$ are symmetric iff $\mathcal{G}$ is undirected; the term "symmetrically" refers to the both-sided normalization, rather than the specific property of the matrix itself. We will now present some properties of the SNA, along with their relation to the Dirichlet energy. The corresponding proofs can be found in Appendix B.

For a connected undirected graph, it is well-known that the spectrum of the SNA lies within $[-1, 1]$, while the spectrum of the SNL is in the range $[0, 2]$ (Chung, 1997). We extend these results to directed graphs, which generally exhibit complex-valued spectra.

Proposition 1. *Let $\mathcal{G}$ be a directed graph with SNA $\mathbf{L}$ and SNL $\mathbf{\Delta}$. For every $\lambda \in \lambda(\mathbf{L})$, it holds that $|\lambda| \leq 1$ and $\lambda(\mathbf{\Delta}) = 1 - \lambda(\mathbf{L})$.*

Proposition 1 provides an upper bound for the largest eigenvalue of any directed graph, irrespective of its size. However, many other spectral properties do not carry over easily from the undirected to the directed case. For example, the SNA may not possess a one-eigenvalue, even if the graph is strongly connected (see, e.g., Figure 5). The one-eigenvalue is of particular interest since its eigenvector $\mathbf{v}$ corresponds to zero Dirichlet energy $\mathscr{E}(\mathbf{v}) = 0$. Therefore, studying when $1 \in \lambda(\mathbf{L})$ is crucial to understand the behavior of the Dirichlet energy. We fully characterize the set of graphs for which $1 \in \lambda(\mathbf{L})$; this is the scope of the following definition.

Definition 3. *A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be* weakly balanced *if there exists $\mathbf{k} \in \mathbb{R}^N$ such that $\mathbf{k} \neq 0$ and*

$$\sum_{j=1}^{N} a_{i,j} \left(\frac{k_j}{\sqrt{d_j^{out}}} - \frac{k_i}{\sqrt{d_i^{in}}} \right) = 0, \ \forall i \in \{1, \ldots, N\} \,,$$

and balanced *if $d_i^{in} = d_i^{out}$ for all $i \in \{1, \ldots, N\}$.*

It is straightforward to see that a balanced graph is weakly balanced since one can choose $k_i = \sqrt{d_i^{\text{in}}}$. Hence, all undirected graphs are also weakly balanced. However, as shown in Figure 5, the set of balanced graphs is a proper subset of the set of weakly balanced graphs.

Proposition 2. *Let $\mathcal{G}$ be a directed graph with SNA $\mathbf{L}$. Then, 1 is an eigenvalue of $\mathbf{L}$ iff the graph is weakly balanced. Suppose the graph is strongly connected, then -1 is an eigenvalue iff the graph is weakly balanced with even period.*

Proposition 2 generalizes a well-known result for undirected graphs: $-1 \in \lambda(\mathbf{L})$ iff the graph is bipartite, i.e., has even period. The next result shows that the Dirichlet energy defined in (1) and the SNL are closely connected.

Proposition 3. *For every $\mathbf{x} \in \mathbb{C}^{N \times K}$, we have*

$$\mathscr{E}(\mathbf{x}) = \frac{1}{2} \Re \left(\text{trace} \left(\mathbf{x}^{\mathsf{H}} \mathbf{\Delta} \mathbf{x} \right) \right) \,.$$

Moreover, there exists $\mathbf{x} \neq \mathbf{0}$ such that $\mathscr{E}(\mathbf{x}) = 0$ iff the graph is weakly balanced.

Proposition 3 generalizes the well-known result from the undirected (see, e.g., (Cai et al., 2020, Definition 3.1)) to the directed case. This result is an important tool for analyzing the evolution of the Dirichlet energy in graph neural networks.
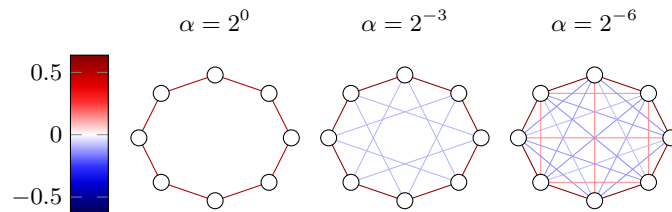
Fig. 1: Visual representation of long-range dependencies generated by the fractional Laplacian on the cycle graph. The values of the fractional Laplacian can also be negative.

4   Fractional Graph Laplacians

We introduce the fractional graph Laplacian through the singular value decomposition (SVD) of a graph's SNA. This approach has two key advantages over the traditional definition (Pozrikidis, 2018; Benzi et al., 2020) in the spectral domain. First, it allows defining the fractional Laplacian based on any choice of graph Laplacian, including those with negative real spectrum such as the SNA. Secondly, the SVD is computationally more efficient and numerically more stable than the calculation of the Jordan decomposition, which would be necessary if the fractional Laplacian was defined in the spectral domain.

Consider a directed graph with SNA $\mathbf{L}$ and its SVD $\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$, where $\mathbf{U}$, $\mathbf{V} \in \mathbb{C}^{N \times N}$ are unitary matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ is a diagonal matrix. Given $\alpha \in \mathbb{R}$, we define the α-*fractional graph Laplacian* of $\mathbf{L}$ (or short α-*FGL*) as

$$\mathbf{L}^{\alpha} := \mathbf{U}\boldsymbol{\Sigma}^{\alpha}\mathbf{V}^{\mathsf{H}}. \tag{2}$$

In undirected graphs, the fractional Laplacian preserves the sign of the eigenvalues λ of $\mathbf{L}$ while modifying their magnitudes, i.e., $\lambda \mapsto \mathrm{sign}(\lambda)\,|\lambda|^{\alpha}$. [*]

The fractional Laplacian is generally less sparse than the original SNA, as it connects nodes that are not adjacent in the underlying graph. The next theorem demonstrates that the weight of such "virtual" edges is relatively low.

Theorem 1. *Let $\mathcal{G}$ be a directed graph with SNA $\mathbf{L}$. For $\alpha > 0$, if the distance $d(i,j)$ between nodes i and j is at least 2, then*

$$\left|(\mathbf{L}^{\alpha})_{i,j}\right| \leq \left(1 + \frac{\pi^2}{2}\right)\left(\frac{\|\mathbf{L}\|}{2\,(d(i,j) - 1)}\right)^{\alpha}. \tag{3}$$

We provide a proof of Theorem 1 in Appendix C. In Figure 1, we visually represent the cycle graph with eight nodes and the corresponding α-FGL entries. We also refer to Figure 7, where we depict the distribution of α-Laplacian entries for the real-world graphs Cora (undirected) and Chameleon (directed) with respect to the distance between the underlying nodes in the original graph. Our empirical findings align with our theoretical results presented in Theorem 1.

---

[*] In fact, this property can be generalized to all directed graphs with a normal SNA matrix, including, among others, directed cycle graphs. See Lemma 1 for more details.

5   FGL Neural ODE

In this study, we explore fractional Laplacian-based graph neural ODEs based on the heat equation,

$$\begin{cases} \mathbf{x}'(t) = -\mathbf{L}^{\alpha}\mathbf{x}(t)\mathbf{W}\,, \\ \mathbf{x}(0) = \mathbf{x}_0\,, \end{cases} \tag{4}$$

where $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$ is the *initial condition*, $\mathbf{x}(t) \in \mathbb{R}^{N \times K}$ for $t > 0$ and $\alpha \in \mathbb{R}$. We assume that the *channel mixing matrix* $\mathbf{W} \in \mathbb{R}^{K \times K}$ is a symmetric matrix. We consider a second FGL-based ODE that is motivated by the Schrödinger equation,

$$\begin{cases} \mathbf{x}'(t) = i\mathbf{L}^{\alpha}\mathbf{x}(t)\mathbf{W}\,, \\ \mathbf{x}(0) = \mathbf{x}_0\,, \end{cases} \tag{5}$$

where now $\mathbf{x}_0, \mathbf{x}(t) \in \mathbb{C}^{N \times K}$ and $\mathbf{W} \in \mathbb{C}^{K \times K}$ is unitary diagonalizable.

We note that both (4) and (5) can be analytically solved. For instance, the solution of (4) is given by $\mathrm{vec}(\mathbf{x}(t)) = \exp(-t\,\mathbf{W} \otimes \mathbf{L}^{\alpha})\mathrm{vec}(\mathbf{x}_0)$, where $\otimes$ denotes the Kronecker product and $\mathrm{vec}(\cdot)$ represents the vectorization operation. However, calculating the exact solution is computationally infeasible, since the memory required to store $\mathbf{W} \otimes \mathbf{L}^{\alpha}$ alone grows as $(NK)^2$. Therefore, we rely on numerical schemes to solve (4) and (5).

Oversmoothing and Frequencies of the SNL $\mathbf{\Delta}$. In the remainder of this section, we analyze the Dirichlet energy for solutions to neural graph ODEs as given by (4) and (5). We begin with the definition of oversmoothing.

Definition 4. *Neural ODE-based GNNs are said to* oversmooth *if the normalized Dirichlet energy decays exponentially fast. That is, for any initial value $\mathbf{x}_0$, the solution $\mathbf{x}(t)$ satisfies for every $t > 0$*

$$\left| \mathscr{E}\left(\frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2} \right) - \lambda_{\min} \right| \le \exp\left(-Ct\right)\,, \tag{6}$$

where $\lambda_{\min}$ is the smallest eigenvalue of the underlying graph's SNL $\mathbf{\Delta}$ and $C > 0$.

Definition 4 captures the *actual* smoothness of features by considering the normalized Dirichlet energy, which mitigates the impact of feature amplitude (Cai et al., 2020; Di Giovanni et al., 2022). Additionally, Proposition 3 shows that the normalized Dirichlet energy is intimately related to the numerical range of the SNL $\mathbf{\Delta}$ of the underlying graph. This shows that the Dirichlet energy and eigenvalues (or *frequencies*) of the SNL are intertwined and one can equivalently talk about Dirichlet energy or frequencies (see also Lemma 3). In particular, it holds that

$$0 \le \mathscr{E}\left(\frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2} \right) \le \frac{\|\mathbf{\Delta}\|}{2}\,.$$

As seen in Section 3, the minimal possible value attained by the normalized Dirichlet energy is often strictly greater than 0 for directed graphs. This indicates

that GNNs on general directed graphs inherently cannot oversmooth to the same extent as in undirected. However, we prove that a vanilla GCN implementing the directed SNA oversmooths with respect to Definition 4, see Appendix E.3.

5.1 Frequency Analysis of FGL-based Neural ODE for Undirected Graphs

In this subsection, we focus on the frequency analysis of FGL-based Neural ODEs for undirected graphs. Most classical GNN (Kipf et al., 2017; Velickovic et al., 2018) and also graph neural ODEs (Chamberlain et al., 2021; Eliasof et al., 2021) have been shown to oversmooth. Di Giovanni et al. (2022) proved that the normalized Dirichlet energy for GNNs based on (4) with $\alpha = 1$ can not only converge to its minimal value but also to its maximal possible value. A GNN exhibiting this property is then termed *Highest-Frequency-Dominant (HFD)*.

However, in real-world scenarios, most graphs are not purely homophilic nor purely heterophilic but fall somewhere in between. Intuitively, this suggests that mid-range frequencies might be more suitable. To illustrate this intuition, consider the cycle graph as an example. If we have a homophily of 1, low frequencies are optimal; with a homophily equal to 0, high frequencies are optimal. Interestingly, for a homophily of $1/2$, the mid-range frequency is optimal, even though the eigendecomposition is label-independent. More information on this example can be found in Figure 3 and Appendix G. Based on this observation, we propose the following definition to generalize the concept of HFD, accommodating not only the lowest or highest frequency but all possible frequencies.

Definition 5. *Let $\lambda \geq 0$. Neural ODE-based GNNs initialized at $\mathbf{x}_0$ are λ-Frequency-Dominant (λ-FD) if the solution $\mathbf{x}(t)$ satisfies*

$$\mathscr{E}\left(\frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2}\right) \xrightarrow{t\to\infty} \frac{\lambda}{2}.$$

Suppose λ is the smallest or the largest eigenvalue with respect to the real part. In that case, we call it Lowest-Frequency-Dominant (LFD) *or* Highest-Frequency-Dominant (HFD)*, respectively.*

In the following theorem, we show that (4) and (5) are not limited to being LFD or HFD, but can also be λ-FD for $\lambda \notin \{\lambda_1(\mathbf{\Delta})/2, \lambda_N(\mathbf{\Delta})/2\}$.

Theorem 2. *Let $\mathcal{G}$ be a undirected graph with SNA $\mathbf{L} \in \mathbb{R}^{N\times N}$. Consider the initial value problem in (4) with channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K\times K}$ and $\alpha \in \mathbb{R}$. Then, for almost all initial values $\mathbf{x}_0 \in \mathbb{R}^{N\times K}$ the following holds.*

(i) If $\alpha > 0$, the solution to (4) is HFD if

$$\lambda_K(\mathbf{W})\mathrm{sgn}(\lambda_1(\mathbf{L}))|\lambda_1(\mathbf{L})|^\alpha < \lambda_1(\mathbf{W}), \tag{7}$$

and LFD otherwise.

(ii) Let $\lambda_+(\mathbf{L})$ and $\lambda_-(\mathbf{L})$ be the smallest positive and negative non-zero eigenvalue of $\mathbf{L}$, respectively. If $\alpha < 0$, the solution to (4) *is* $(1 - \lambda_+(\mathbf{L}))$-FD *if*

$$\lambda_1(\mathbf{W})\mathrm{sgn}(\lambda_+(\mathbf{L}))|\lambda_+(\mathbf{L})|^\alpha \tag{8}$$
$$< \lambda_K(\mathbf{W})\mathrm{sgn}(\lambda_-(\mathbf{L}))|\lambda_-(\mathbf{L})|^\alpha\,.$$

Otherwise, the solution to (4) *is* $(1 - \lambda_-(\mathbf{L}))$-FD.

Furthermore, (i)-(ii) also hold for solutions to the Schrödinger equation (5) *if* $\mathbf{W} \in \mathbb{C}^{K \times K}$ *has at least one eigenvalue with non-zero imaginary part.*

Theorem 2 *(i)* generalizes the result by Di Giovanni et al. (2022) for $\alpha = 1$ to arbitrary positive values of α in (4). The convergence speed in Theorem 2 *(i)* depends on the choice of $\alpha \in \mathbb{R}$. By selecting a variable α (e.g., as a learnable parameter), we establish a flexible learning framework capable of adapting the convergence speed of the Dirichlet energy. A slower or more adjustable convergence speed of the Dirichlet energy facilitates broader frequency exploration as it converges more gradually to its maximal or minimal value. Consequently, the frequency component contributions (for finite time, i.e., in practice) are better balanced, which is advantageous for graphs with different homophily levels. Theorem 2 *(ii)* shows that solutions of the fractional neural ODEs in (4) and (5) are not limited to be LFD or HFD. To demonstrate this and the other results of Theorem 2, we solve (4) using an explicit Euler scheme for different choices of α and $\mathbf{W}$ on the Cora and Chameleon graphs. The resulting evolution of the Dirichlet energy with respect to time is illustrated in Figure 4. Finally, we refer to Appendix D.1 for the Proof of Theorem 2.

Remark 1. Theorem 2 is stated for the analytical solutions of (4) and (5), respectively. As noted in Section 5, calculating the analytical solution is infeasible in practice. However, we show in Appendix D.2 and Appendix D.3 that approximations of the solution of (4) and (5) via explicit Euler schemes satisfy the same Dirichlet energy convergence properties if the step size is sufficiently small.

5.2   Frequency Dominance for Directed Graphs

Section 5.2 analyzes the Dirichlet energy in undirected graphs. However, the situation becomes significantly more complex when considering directed graphs. In our experiments (see Appendix A), we observed that the solution to (4) and (5) does not necessarily lead to oversmoothing. On the contrary, the solution can be controlled to exhibit either LFD or HFD for $\alpha > 0$, and mid-frequency-dominance for $\alpha < 0$ as proven for undirected graphs in Theorem 2. We present an initial theoretical result for directed graphs specifically in the case of $\alpha = 1$.

Theorem 3. *Let $\mathcal{G}$ be a directed graph with SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. Consider the initial value problem in* (4) *with diagonal channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ and $\alpha = 1$. Then, for almost all initial values $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$, the solution to* (4) *is* HFD *if*

$$\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L}) < \lambda_1(\mathbf{W})\lambda_N(\mathbf{L}) \tag{9}$$

and $\lambda_1(\mathbf{L})$ is unique. Otherwise, the solution is LFD.

Table 1: Test accuracy on node classification for undirected graphs: top three models colored as first, second, third.

	Film	Squirrel	Chameleon	Citeseer	Pubmed	Cora
GGCN	37.54 ± 1.56	55.17 ± 1.58	71.14 ± 1.84	77.14 ± 1.45	89.15 ± 0.37	87.95 ± 1.05
GPRGNN	34.63 ± 1.22	31.61 ± 1.24	46.58 ± 1.71	77.13 ± 1.67	87.54 ± 0.38	87.95 ± 1.18
H2GCN	35.70 ± 1.00	36.48 ± 1.86	60.11 ± 2.15	77.11 ± 1.57	89.49 ± 0.38	87.87 ± 1.20
GCNII	37.44 ± 1.30	38.47 ± 1.58	63.86 ± 3.04	77.33 ± 1.48	90.15 ± 0.43	88.37 ± 1.25
Geom-GCN	31.59 ± 1.15	38.15 ± 0.92	60.00 ± 2.81	78.02 ± 1.15	89.95 ± 0.47	85.35 ± 1.57
PairNorm	27.40 ± 1.24	50.44 ± 2.04	62.74 ± 2.82	73.59 ± 1.47	87.53 ± 0.44	85.79 ± 1.01
GraphSAGE	34.23 ± 0.99	41.61 ± 0.74	58.73 ± 1.68	76.04 ± 1.30	88.45 ± 0.50	86.90 ± 1.04
GCN	27.32 ± 1.10	53.43 ± 2.01	64.82 ± 2.24	76.50 ± 1.36	88.42 ± 0.50	86.98 ± 1.27
GAT	27.44 ± 0.89	40.72 ± 1.55	60.26 ± 2.50	76.55 ± 1.23	87.30 ± 1.10	86.33 ± 0.48
MLP	36.53 ± 0.70	28.77 ± 1.56	46.21 ± 2.99	74.02 ± 1.90	75.69 ± 2.00	87.16 ± 0.37
CGNN	35.95 ± 0.86	29.24 ± 1.09	46.89 ± 1.66	76.91 ± 1.81	87.70 ± 0.49	87.10 ± 1.35
GRAND	35.62 ± 1.01	40.05 ± 1.50	54.67 ± 2.54	76.46 ± 1.77	89.02 ± 0.51	87.36 ± 0.96
Sheaf (max)	37.81 ± 1.15	56.34 ± 1.32	68.04 ± 1.58	76.70 ± 1.57	89.49 ± 0.40	86.90 ± 1.13
GRAFF-NL	35.96 ± 0.95	59.01 ± 1.31	71.38 ± 1.47	76.81 ± 1.12	89.81 ± 0.50	87.81 ± 1.13
FLODE	37.16 ± 1.42	64.23 ± 1.84	73.60 ± 1.55	78.07 ± 1.62	89.02 ± 0.38	86.44 ± 1.17

We note that the condition on $\lambda_1(\mathbf{L})$ is satisfied in all real-world graphs we considered. The proof of Theorem 3 is given in Appendix E.1. Finally, we refer to Appendix E.2 for the analogous statement and proof when the solution of (4) is approximated via an explicit Euler scheme.

6   Numerical Experiments

This section evaluates the fractional Laplacian ODEs in node classification by approximating (4) and (5) with an Explicit Euler scheme. This leads to the following update rules

$$\mathbf{x}_{t+1} = \mathbf{x}_t - h\,\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}\,, \; \mathbf{x}_{t+1} = \mathbf{x}_t + i\,h\,\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}\,, \tag{10}$$

for the heat and Schrödinger equation, respectively. In both cases, $\mathbf{W}$, α and h are learnable parameters, t is the layer, and $\mathbf{x}_0$ is the initial nodes' feature matrix. In accordance with the results in Section 5, we select $\mathbf{W}$ as a diagonal matrix. The initial features $\mathbf{x}_0$ in (10) are encoded through a MLP, and the output is decoded using a second MLP. We refer to the resulting model as *FLODE* (fractional Laplacian ODE). In Appendix F, we present details on the baseline models, the training setup, the exact hyperparameters; in Appendix F.3, we report an ablation study on the effect of each component of the proposed method, such as learnable vs. fixed α, and the choice of the ODE.

Real-World Graphs. We report results on 6 undirected datasets consisting of both homophilic graphs, such as Cora (McCallum et al., 2000), Citeseer (Sen et al., 2008), and Pubmed (Namata et al., 2012)), and heterophilic graphs, such
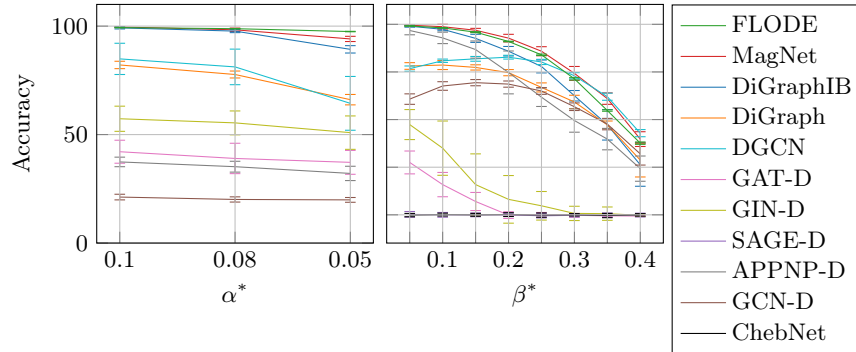
Fig. 2: Experiments on directed stochastic block model. Unlike other models, FLODE's performances do not deteriorate as much when changing the inter-cluster edge density α^*.

as Film (Tang et al., 2009), Squirrel and Chameleon (Rozemberczki et al., 2021). Due to time and memory limitations, we compute only 30% of singular values for Pubmed, which serves as the best low-rank approximation of the original SNA.

We evaluate our method on the directed, as well as undirected version of Squirrel, Film, and Chameleon. In all datasets, we use the standard 10 splits from (Pei et al., 2019). The choice of the baseline models and the corresponding results are taken from (Di Giovanni et al., 2022). The results are in Table 1 and Table 2.

Synthetic Directed Graph. We consider the directed stochastic block model (DSBM) datasets (Zhang et al., 2021). The DSBM divides nodes into 5 clusters and assigns probabilities for interactions between vertices. It considers two sets of probabilities: $\{\alpha_{i,j}\}$ for undirected edge creation and $\{\beta_{i,j}\}$ for assigning edge directions, $i, j \in \{1, \ldots 5\}$. The primary objective is to classify vertices based on their clusters. In the first experiment, $\alpha_{i,j} = \alpha^*$ varies, altering neighborhood information's importance. The second experiment focuses on varying $\beta_{i,j} = \beta^*$ values, which changes directional information. The results are in Figure 2. The splits, the choice of the baseline model, and their results are taken from (Zhang et al., 2021).

Results. The experiments showcase the flexibility of FLODE, as it can accommodate various types of graphs, both directed and undirected, as well as a broad range of homophily levels. While other methods, such as MagNet (Zhang et al., 2021), perform similarly to our approach, they face limitations when applied to certain graph configurations. For instance, when applied to undirected graphs, MagNet reduces to ChebNet, making it unsuitable for heterophilic graphs. Similarly, GRAFF (Di Giovanni et al., 2022) performs well on undirected graphs but falls short on directed graphs. We note that in all our experiments (except for the

Table 2: Test accuracy on node classification for directed graphs: top three models colored as first, second, third.

	Film	Squirrel	Chameleon
ACM	36.89 ± 1.18	54.4 ± 1.88	67.08 ± 2.04
HLP	34.59 ± 1.32	74.17 ± 1.83	77.48 ± 1.50
FSGNN	35.67 ± 0.69	73.48 ± 2.13	78.14 ± 1.25
GRAFF	37.11 ± 1.08	58.72 ± 0.84	71.08 ± 1.75
FLODE	37.41 ± 1.06	74.03 ± 1.58	77.98 ± 1.05

Film dataset), FLODE learned a non-trivial exponent $\alpha \neq 1$, highlighting the advantages of FGL-based GNNs (see Appendix F). Furthermore, we showcase an ablation study in Appendix F.3 that demonstrates the impact of employing FGLs and the neural ODE framework.

7 Conclusion

In this work, we introduce the concepts of Dirichlet energy and oversmoothing for directed graphs and demonstrate the close relationship between these notions and the SNL. Building upon this foundation, we define fractional Laplacians in the singular value domain, resulting in matrices capable of capturing long-range dependencies. To address oversmoothing in directed graphs, we propose fractional Laplacian-based graph ODEs, which are provably not limited to LFD behavior. We demonstrate the flexibility of our method to accommodate various graph structures in node-level tasks. Our method represents a versatile and robust solution for node classification across a wide range of graph structures.

Limitations and Future Work. Approximating the solutions of (4) and (5) poses a challenge. Specifically, the computational cost of the SVD grows cubically in the number of nodes N, while the storage of the singular vectors grows quadratically in N. However, both costs can be significantly reduced by computing only $k \ll N$ singular values via truncated SVD, giving the best k-rank approximation of the SNA. Moreover, the SVD can be computed offline as a preprocessing step.

The frequency analysis of α-FGL neural ODEs in directed graphs is an exciting future direction. It would be interesting, for instance, to investigate the impact of choosing $\alpha \neq 1$ on the convergence speed of the Dirichlet energy. Controlling the speed could facilitate the convergence of the Dirichlet energy to an *optimal* value, which has been shown to exist in synthetic settings (Keriven, 2022; Wu et al., 2023). Another future direction would be to analyze the dynamics when approximating the solution to the FGL neural ODEs using alternative numerical solvers, such as higher-order or adjoint methods.

References

Benzi, M., Bertaccini, D., Durastante, F., and Simunec, I. (2020). "Non-Local Network Dynamics via Fractional Graph Laplacians". In: *Journal of Complex Networks* 8.3.

Bo, D., Wang, X., Shi, C., and Shen, H. (2021). "Beyond Low-frequency Information in Graph Convolutional Networks". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.5, pp. 3950–3957.

Bodnar, C., Di Giovanni, F., Chamberlain, B., Lió, P., and Bronstein, M. (2022). "Neural Sheaf Diffusion: A Topological Perspective on Heterophily and Oversmoothing in GNNs". In: *Advances in Neural Information Processing Systems* 35, pp. 18527–18541.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). "Geometric Deep Learning: Going beyond Euclidean Data". In: *IEEE Signal Processing Magazine* 34.4, pp. 18–42.

Cai, C. and Wang, Y. (2020). *A Note on Over-Smoothing for Graph Neural Networks*. arXiv: 2006.13318 [cs, stat].

Chamberlain, B., Rowbottom, J., Gorinova, M. I., Bronstein, M., Webb, S., and Rossi, E. (2021). "GRAND: Graph Neural Diffusion". In: *Proceedings of the 38th International Conference on Machine Learning*. PMLR, pp. 1407–1418.

Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., and Sun, X. (2020). "Measuring and Relieving the Over-Smoothing Problem for Graph Neural Networks from the Topological View". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.04, pp. 3438–3445.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). "Neural Ordinary Differential Equations". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc.

Chien, E., Peng, J., Li, P., and Milenkovic, O. (2021). "Adaptive Universal Generalized PageRank Graph Neural Network". In: *International Conference on Learning Representations*.

Chung, F. R. K. (1997). *Spectral Graph Theory*. Regional Conference Series in Mathematics no. 92. Providence, R.I: Published for the Conference Board of the mathematical sciences by the American Mathematical Society.

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering". In: *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc.

Di Giovanni, F., Rowbottom, J., Chamberlain, B. P., Markovich, T., and Bronstein, M. M. (2022). *Graph Neural Networks as Gradient Flows: Understanding Graph Convolutions via Energy*. arXiv: 2206.10991 [cs, stat].

Eliasof, M., Haber, E., and Treister, E. (2021). "PDE-GCN: Novel Architectures for Graph Neural Networks Motivated by Partial Differential Equations". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., pp. 3836–3849.

Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., and Yin, D. (2019). "Graph Neural Networks for Social Recommendation". In: *The World Wide Web Conference*. San Francisco CA USA: ACM, pp. 417–426.

Fey, M. and Lenssen, J. E. (2019). *Fast Graph Representation Learning with PyTorch Geometric*. arXiv: 1903.02428 [cs, stat].

Gasteiger, J., Bojchevski, A., and Günnemann, S. (2018). "Predict Then Propagate: Graph Neural Networks Meet Personalized PageRank". In: *International Conference on Learning Representations*.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR, pp. 1263–1272.

Gori, M., Monfardini, G., and Scarselli, F. (2005). "A New Model for Learning in Graph Domains". In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*. Vol. 2, 729–734 vol. 2.

Haber, E. and Ruthotto, L. (2018). "Stable Architectures for Deep Neural Networks". In: *Inverse Problems* 34.1, p. 014004. arXiv: 1705.03341 [cs, math].

Hamilton, W., Ying, Z., and Leskovec, J. (2017). "Inductive Representation Learning on Large Graphs". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Keriven, N. (2022). "Not Too Little, Not Too Much: A Theoretical Analysis of Graph (over)Smoothing". In: *Advances in Neural Information Processing Systems*.

Kingma, D. P. and Ba, J. (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs].

Kipf, T. N. and Welling, M. (2017). "Semi-Supervised Classification with Graph Convolutional Networks". In: *International Conference on Learning Representations*.

Li, G., Muller, M., Thabet, A., and Ghanem, B. (2019). "DeepGCNs: Can GCNs Go As Deep As CNNs?" In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9267–9276.

Li, Q., Han, Z., and Wu, X.-m. (2018). "Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1.

Luan, S., Zhao, M., Chang, X.-W., and Precup, D. (2019). "Break the Ceiling: Stronger Multi-scale Deep Graph Convolutional Networks". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

McCallum, A. K., Nigam, K., Rennie, J., and Seymore, K. (2000). "Automating the Construction of Internet Portals with Machine Learning". In: *Information Retrieval* 3.2, pp. 127–163.

Monti, F., Frasca, F., Eynard, D., Mannion, D., and Bronstein, M. M. (2019). *Fake News Detection on Social Media Using Geometric Deep Learning*. arXiv: 1902.06673 [cs, stat].

Namata, G. M., London, B., Getoor, L., and Huang, B. (2012). "Query-driven Active Surveying for Collective Classification". In: *Workshop on Mining and Learning with Graphs*.

Oono, K. and Suzuki, T. (2019). "Graph Neural Networks Exponentially Lose Expressive Power for Node Classification". In: *International Conference on Learning Representations*.

Paszke, A. et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems.* Vol. 32. Curran Associates, Inc.

Pei, H., Wei, B., Chang, K. C.-C., Lei, Y., and Yang, B. (2019). "Geom-GCN: Geometric Graph Convolutional Networks". In: *International Conference on Learning Representations.*

Perko, L. (2001). *Differential Equations and Dynamical Systems.* Ed. by J. E. Marsden, L. Sirovich, and M. Golubitsky. Vol. 7. Texts in Applied Mathematics. New York, NY: Springer New York.

Poli, M., Massaroli, S., Park, J., Yamashita, A., Asama, H., and Park, J. (2021). *Graph Neural Ordinary Differential Equations.* arXiv: 1911.07532 [cs, stat].

Pozrikidis, C. (2018). *The Fractional Laplacian.* First. Boca Raton : Taylor & Francis, 2016. | "A CRC title.": Chapman and Hall/CRC.

Rozemberczki, B., Allen, C., and Sarkar, R. (2021). "Multi-Scale Attributed Node Embedding". In: *Journal of Complex Networks* 9.2.

Rusch, T. K., Chamberlain, B., Rowbottom, J., Mishra, S., and Bronstein, M. (2022). "Graph-Coupled Oscillator Networks". In: *Proceedings of the 39th International Conference on Machine Learning.* PMLR, pp. 18888–18909.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20.1, pp. 61–80.

Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B., and Eliassi-Rad, T. (2008). "Collective Classification in Network Data". In: *AI Magazine* 29.3, p. 93.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In.

Tang, J., Sun, J., Wang, C., and Yang, Z. (2009). "Social Influence Analysis in Large-Scale Networks". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Paris France: ACM, pp. 807–816.

Tong, Z., Liang, Y., Sun, C., Li, X., Rosenblum, D., and Lim, A. (2020). "Digraph Inception Convolutional Networks". In: *Advances in Neural Information Processing Systems.* Vol. 33. Curran Associates, Inc., pp. 17907–17918.

Tong, Z., Liang, Y., Sun, C., Rosenblum, D. S., and Lim, A. (2020). *Directed Graph Convolutional Network.* arXiv: 2004.13970 [cs, stat].

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). "Graph Attention Networks". In: *International Conference on Learning Representations.*

Wang, J., Huang, P., Zhao, H., Zhang, Z., Zhao, B., and Lee, D. L. (2018). "Billion-Scale Commodity Embedding for E-commerce Recommendation in Alibaba". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* KDD '18. New York, NY, USA: Association for Computing Machinery, pp. 839–848.

Wu, X., Chen, Z., Wang, W., and Jadbabaie, A. (2023). *A Non-Asymptotic Analysis of Oversmoothing in Graph Neural Networks*. arXiv: 2212.10701 [cs, stat].

Xhonneux, L.-P., Qu, M., and Tang, J. (2020). "Continuous Graph Neural Networks". In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR, pp. 10432–10441.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2018). "How Powerful Are Graph Neural Networks?" In: *International Conference on Learning Representations*.

Yan, Y., Hashemi, M., Swersky, K., Yang, Y., and Koutra, D. (2022). "Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks". In: *2022 IEEE International Conference on Data Mining (ICDM)*. Orlando, FL, USA: IEEE, pp. 1287–1292.

Zhang, X., He, Y., Brugnone, N., Perlmutter, M., and Hirn, M. (2021). "MagNet: A Neural Network for Directed Graphs". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., pp. 27003–27015.

Zhao, L. and Akoglu, L. (2019). "PairNorm: Tackling Oversmoothing in GNNs". In: *International Conference on Learning Representations*.

Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., and Koutra, D. (2020). "Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 7793–7804.

Zou, C., Han, A., Lin, L., and Gao, J. (2022). *A Simple Yet Effective SVD-GCN for Directed Graphs*. arXiv: 2205.09335 [cs].

Acronyms

DSBM Directed Stochastic Block Model

FD   Frequency Dominant
FGL   Fractional Graph Laplacian

GAT   Graph Attention Network
GCN   Graph Convolutional Network
GNN   Graph Neural Network

HFD   Highest Frequency Dominant

LCC   Largest Connected Components
LFD   Lowest Frequency Dominant

MLP   Multi-Layer Perceptron

ODE   Ordinary Differential Equation

SNA   Symmetrically Normalized Adjacency
SNL   Symmetrically Normalized Laplacian
SVD   Singular Value Decomposition

Notation

|	
i	Imaginary unit
$\Re(z)$	Real part of $z \in \mathbb{C}$
$\Im(z)$	Imaginary part of $z \in \mathbb{C}$
$\text{diag}(\mathbf{x})$	Diagonal matrix with $\mathbf{x}$ on the diagonal.
$\mathbf{M}^{\mathsf{T}}$	Transpose of the matrix $\mathbf{M}$
$\mathbf{M}^{*}$	Conjugate of the matrix $\mathbf{M}$
$\mathbf{M}^{\mathsf{H}}$	Conjugate transpose of the matrix $\mathbf{M}$
$\|\mathbf{M}\|$	Spectral norm of the matrix $\mathbf{M}$
$\|\mathbf{M}\|_2$	Frobenius norm of the matrix $\mathbf{M}$
$\lambda(\mathbf{M})$	Spectrum of the matrix $\mathbf{M}$
$\sigma(\mathbf{M})$	Singular values of the matrix $\mathbf{M}$
$\mathscr{E}(\mathbf{x})$	Dirichlet energy computed on $\mathbf{x}$
$\mathscr{H}(\mathcal{G})$	Homophily coefficient of the graph $\mathcal{G}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product between $\mathbf{A}$ and $\mathbf{B}$
$\text{vec}(\mathbf{M})$	Vector obtained stacking columns of the matrix $\mathbf{M}$.

A   Additional Images
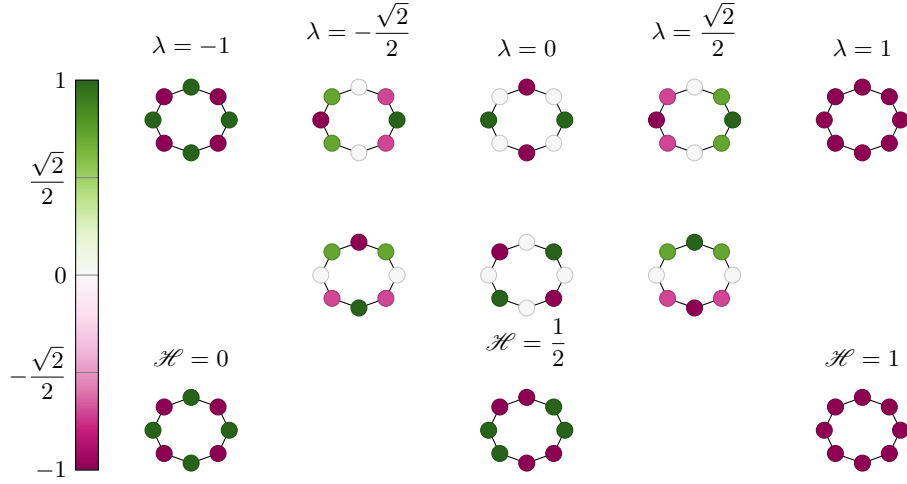
Fig. 3: Eigendecomposition of $\mathbf{L}$ for the cycle graph C_8. The first two rows show the eigenvectors corresponding to the eigenvalues λ. The last row shows how the (label-unaware) eigendecomposition can be used to study homophily, whose definition requires the labels. Additional details can be found in Appendix G.
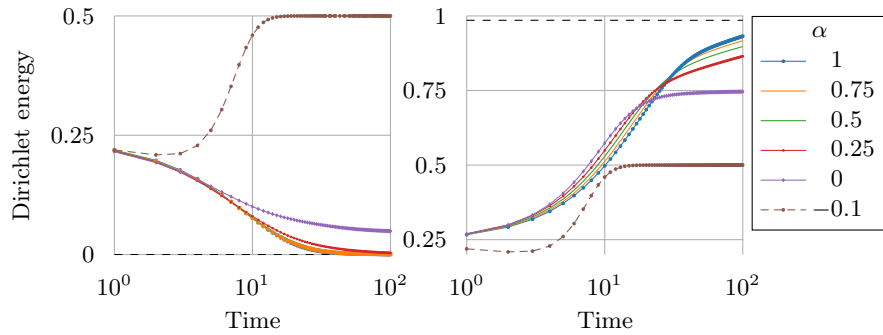
(a) Cora (undirected).
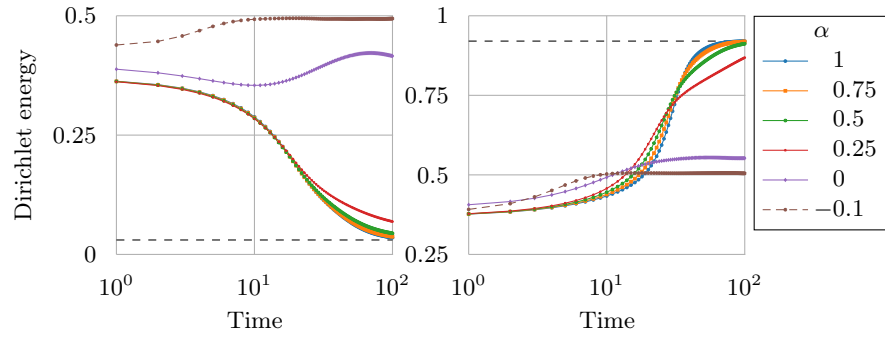
(b) chameleon (directed).

Fig. 4: Convergence of Dirichlet energy for the solution of equation (4) using an explicit Euler scheme with a step size of $h = 10^{-1}$. We consider different α-FGL in (4) and choose $\mathbf{W}$ as a random diagonal matrix. In the left plot, $\mathbf{W}$ has only a negative spectrum, while in the right plot, $\mathbf{W}$ has only a positive spectrum. The black horizontal line represents the theoretical limit based on Theorem 2.
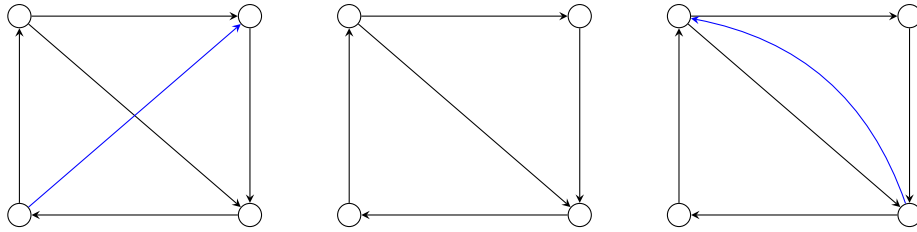
Fig. 5: Examples of non-weakly balanced (left), weakly balanced but not balanced (center), and balanced (right) directed graphs. The plot shows that balanced graphs are a proper subset of weakly balanced graphs and that non-weakly balanced graphs do exist. As the theory suggests, the Perron-Frobenius eigenvalue of the non-weakly balanced graph is $\lambda_{\mathrm{PF}} \approx 0.986652 \neq 1$. The Perron-Frobenius eigenvalue of both weakly balanced and balanced graphs is $\lambda_{\mathrm{PF}} = 1$, with corresponding $\mathbf{k} = (\sqrt{2}, 1, \sqrt{2}, \sqrt{2})^{\mathsf{T}}$ and $\mathbf{k} = (\sqrt{2}, 1, \sqrt{2}, \sqrt{1})^{\mathsf{T}}$, respectively. For this particular weakly balanced graph a stronger condition holds, namely, $a_{i,j} \left(k_j / \sqrt{d_j^{\mathrm{out}}} - k_i / \sqrt{d_i^{\mathrm{in}}} \right) = 0$ for all $i, j \in \{1, \ldots, 4\}$.
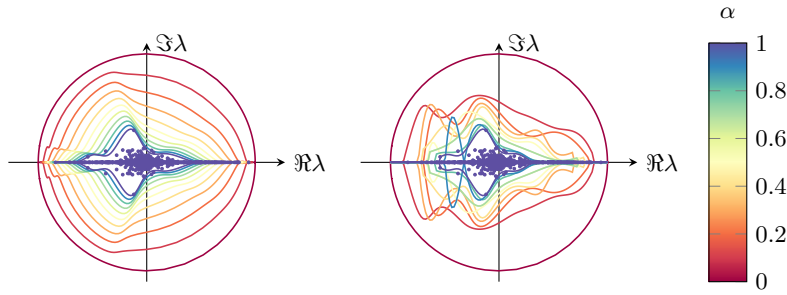
Fig. 6: Spectral evolution described in Lemma 1 for the envelope of Chameleon spectra vs. real dynamics. In the left plot, we computed the envelope of eigenvalues, and we evolved it according to the map $\lambda \mapsto \exp\left(i\operatorname{atan2}\left(\Re\lambda, \Im\lambda\right)\right)|\lambda|^{\alpha}$. In the right plot, we computed for each α the envelope of $\lambda\left(\mathbf{L}^{\alpha}\right) \coloneqq \lambda\left(\mathbf{U}\mathbf{\Sigma}^{\alpha}\mathbf{V}^{\mathsf{H}}\right)$, which describes the real dynamics of the spectrum. The two plots would be roughly equal if chameleon's SNA was a normal matrix.
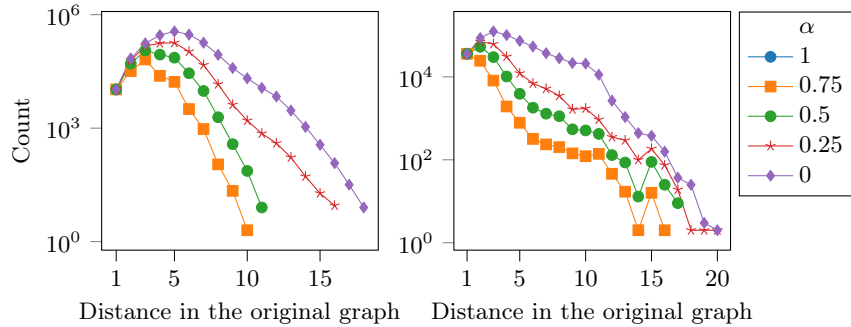
Fig. 7: Number of edges created by the α-FGL for Cora (undirected, left) and chameleon (directed, right) with respect to the distance of nodes in the underlying graph. Specifically, we computed the α-fractional Laplacian $\mathbf{L}^{\alpha} \coloneqq \mathbf{U}\mathbf{\Sigma}^{\alpha}\mathbf{V}^{\mathsf{H}}$ and, for each value of distance $d(i,j) \in \mathbb{N}$ in the original graph, we counted the number of elements $(\mathbf{L}^{\alpha})_{i,j} > 10^{-3}$. Note that for $\alpha = 1$, we have only one point at $x = 1$.
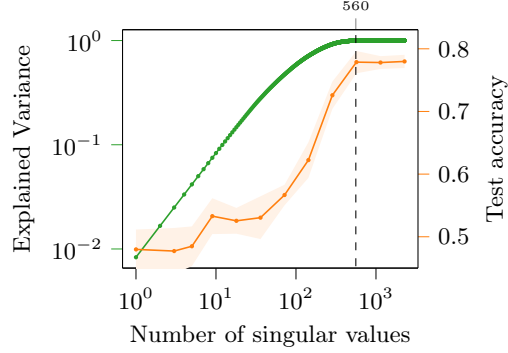
Fig. 8: Effect of different level of sparsity on the test accuracy (orange), chameleon. The explained variance (green) is computed as $\sum_{k=1}^{K} \sigma_k^2 / \sum_{k=1}^{N} \sigma_k^2$ for $K \in \{1, \ldots, N\}$ and it measures the variability explained by the largest K singular values. As shown, the accuracy stabilizes after 560 (24.26%) singular values, corresponding to an explained variance of 0.998: adding more singular values increases the performances only marginally.

B   Appendix for Section 3

Proposition 4 (Proposition 1 in Section 3). *Let $\mathcal{G}$ be a directed graph with symmetrically normalized adjacency $\mathbf{L}$ and Laplacian $\boldsymbol{\Delta}$. It holds $|\lambda(\mathbf{L})| \leq 1$ and $\lambda(\boldsymbol{\Delta}) = 1 - \lambda(\mathbf{L})$.*

Proof. We show that the numerical range $\mathcal{W}(\mathbf{L}) = \{\mathbf{x}^{\mathsf{H}} \mathbf{L} \mathbf{x} : \mathbf{x}^{\mathsf{H}} \mathbf{x} = 1\}$ satisfies $\mathcal{W}(\mathbf{L}) \subset [-1, 1]$. As $\mathcal{W}(\mathbf{L})$ contains all eigenvalues of $\mathbf{L}$ the proposition follows.

Let $\mathbf{A} = (a_{i,j})_{i,j=1}^{N} \in \mathbb{R}^{N \times N}$ be the adjacency matrix of $\mathcal{G}$ and $\mathbf{x} \in \mathbb{C}^N$ with $\mathbf{x}^{\mathsf{H}} \mathbf{x} = 1$. Applying the Cauchy-Schwartz inequality in (2) and (3), we get

$$\left| \mathbf{x}^{\mathsf{H}} \mathbf{L} \mathbf{x} \right| \overset{(1)}{\leq} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j} \frac{|x_i| \, |x_j|}{\sqrt{d_i^{in} d_j^{out}}}$$

$$= \sum_{i=1}^{N} \frac{|x_i|}{\sqrt{d_i^{in}}} \sum_{j=1}^{N} a_{i,j} \frac{|x_j|}{\sqrt{d_j^{out}}}$$

$$\overset{(2)}{\leq} \sum_{i=1}^{N} \frac{|x_i|}{\sqrt{d_i^{in}}} \sqrt{\sum_{j=1}^{N} a_{i,j} \frac{|x_j|^2}{d_j^{out}} \sum_{j=1}^{N} a_{i,j}}$$

$$= \sum_{i=1}^{N} |x_i| \sqrt{\sum_{j=1}^{N} a_{i,j} \frac{|x_j|^2}{d_j^{out}}}$$

$$\overset{(3)}{\leq} \sqrt{\sum_{i=1}^{N} |x_i|^2 \sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j} \frac{|x_j|^2}{d_j^{out}}}$$

$$= \sum_{i=1}^{N} |x_i|^2 \, ,$$

where we used $a_{i,j}^2 = a_{i,j}$. We have $\sum_{i=1}^{N} |x_i|^2 = \mathbf{x}^{\mathsf{H}} \mathbf{x} = 1$ such that $\mathcal{W}(\mathbf{L}) \subset [-1, 1]$ follows. The second claim follows directly by $\mathbf{\Delta} = \mathbf{I} - \mathbf{L}$. □

Proposition 5 (Proposition 2 in Section 3). *Let $\mathcal{G}$ be a directed graph with SNA $\mathbf{L}$. Then 1 is an eigenvalue of $\mathbf{L}$ if and only if the graph is weakly balanced. Suppose the graph is strongly connected; then -1 is an eigenvalue if and only if the graph is weakly balanced with even period.*

Proof. Since the numerical range is only a superset of the set of eigenvalues, we cannot simply consider when the inequalities $(1)-(3)$ in the proof of Proposition 4 are actual equalities. Therefore, we have to find another way to prove the statement. Suppose that the graph is weakly balanced, then

$$\sum_{j=1}^{N} a_{i,j} \left(\frac{k_j}{\sqrt{d_j^{\text{out}}}} - \frac{k_i}{\sqrt{d_i^{\text{in}}}} \right) = 0 \, , \ \forall j \in \{1, \dots, N\} \, .$$

We will prove that $\mathbf{k} = (k_i)_{i=1}^{N}$ is an eigenvector corresponding to the eigenvalue 1,

$$(\mathbf{Lk})_i = \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} k_j$$

$$= \frac{1}{\sqrt{d_i^{\text{in}}}} \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_j^{\text{out}}}} k_j = \frac{1}{\sqrt{d_i^{\text{in}}}} \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_i^{\text{in}}}} k_i = \frac{1}{d_i^{\text{in}}} \left(\sum_{j=1}^{N} a_{i,j} \right) k_i = k_i \, .$$

For the other direction, suppose that there exists $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x} \neq 0$ and $\mathbf{x} = \mathbf{Lx}$. Then for all $i \in \{1, \dots, N\}$

$$0 = (\mathbf{Lx})_i - x_i = \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} x_j - x_i = \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} x_j - \sum_{j=1}^{N} \frac{a_{i,j}}{d_i^{\text{in}}} x_i$$

$$= \sum_{j=1}^{N} \frac{a_{i,j}}{\sqrt{d_i^{\text{in}}}} \left(\frac{x_j}{\sqrt{d_j^{\text{out}}}} - \frac{x_i}{\sqrt{d_i^{\text{in}}}} \right) \, ,$$

hence, the graph is weakly balanced.

By Perron-Frobenius theorem for irreducible non-negative matrices, one gets that $\mathbf{L}$ has exactly h eigenvalues with maximal modulus corresponding to the h roots of the unity, where h is the period of $\mathbf{L}$. Hence, -1 is an eigenvalue of $\mathbf{L}$ if and only if the graph is weakly balanced and h is even. □

Proposition 6 (Proposition 3 in Section 3). *For every* $\mathbf{x} \in \mathbb{C}^{N \times K}$, *we have*

$$\Re\left(\text{trace}\left(\mathbf{x}^{\mathsf{H}} \boldsymbol{\Delta} \mathbf{x}\right)\right) = \frac{1}{2} \sum_{i,j=1}^{N} a_{i,j} \left\| \frac{\mathbf{x}_i}{\sqrt{d_i^{in}}} - \frac{\mathbf{x}_j}{\sqrt{d_j^{out}}} \right\|_2^2,$$

Moreover, there exists $\mathbf{x} \neq 0$ *such that* $\mathscr{E}(\mathbf{x}) = 0$ *if and only if the graph is weakly balanced.*

Proof. By direct computation, it holds

$$\frac{1}{2} \sum_{i,j=1}^{N} a_{i,j} \left\| \frac{x_{i,:}}{\sqrt{d_i^{\text{in}}}} - \frac{x_{j,:}}{\sqrt{d_j^{\text{out}}}} \right\|_2^2$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} a_{i,j} \sum_{k=1}^{K} \left| \frac{x_{i,k}}{\sqrt{d_i^{\text{in}}}} - \frac{x_{j,k}}{\sqrt{d_j^{\text{out}}}} \right|^2$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} a_{i,j} \sum_{k=1}^{K} \left(\frac{x_{i,k}}{\sqrt{d_i^{\text{in}}}} - \frac{x_{j,k}}{\sqrt{d_j^{\text{out}}}} \right)^* \left(\frac{x_{i,k}}{\sqrt{d_i^{\text{in}}}} - \frac{x_{j,k}}{\sqrt{d_j^{\text{out}}}} \right)$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{|x_{i,k}|^2}{d_i^{\text{in}}} + \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{|x_{j,k}|^2}{d_j^{\text{out}}}$$

$$- \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k}^* x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k} x_{j,k}^*}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K} |x_{i,k}|^2 + \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{K} |x_{j,k}|^2 - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k}^* x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k} x_{j,k}^*}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} |x_{i,k}|^2 - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{(\mathbf{x}^{\mathsf{H}})_{k,i} x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k} (\mathbf{x}^{\mathsf{H}})_{k,j}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} |x_{i,k}|^2 - \frac{1}{2} \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{(\mathbf{x}^{\mathsf{H}})_{k,i} x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} - \frac{1}{2} \left(\sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{(\mathbf{x}^{\mathsf{H}})_{k,i} x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} \right)^*$$

$$= \Re\left(\sum_{i=1}^{N} \sum_{k=1}^{K} |x_{i,k}|^2 - \sum_{i,j=1}^{N} \sum_{k=1}^{K} a_{i,j} \frac{x_{i,k}^* x_{j,k}}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} \right)$$

$$= \Re\left(\text{trace}\left(\mathbf{x}^{\mathsf{H}} \boldsymbol{\Delta} \mathbf{x} \right) \right).$$

The last claim can be proved as follows. For simplicity, suppose $\mathbf{x} \in \mathbb{R}^N$. The " $\Longleftarrow$ " is clear since one can choose $\mathbf{x}$ to be $\mathbf{k}$. To prove the " $\Longrightarrow$ ", we reason by contradiction. Suppose there exists a $\mathbf{x} \neq 0$ such that $\mathscr{E}(\mathbf{x}) = 0$ and the

underlying graph is not weakly connected, i.e.,

$$\forall \tilde{\mathbf{x}} \neq \mathbf{0}, \; \left| \sum_{j=1}^{N} a_{i,j} \left(\frac{\tilde{x}_j}{\sqrt{d_j^{\text{out}}}} - \frac{\tilde{x}_i}{\sqrt{d_i^{\text{in}}}} \right) \right| > 0, \; \forall i \in \{1, \ldots, N\},$$

Then, since $\mathbf{x} \neq 0$,

$$
\begin{aligned}
0 = \mathscr{E}(\mathbf{x}) &= \frac{1}{4} \sum_{i,j=1}^{N} a_{i,j} \left| \frac{x_i}{\sqrt{d_i^{\text{in}}}} - \frac{x_j}{\sqrt{d_j^{\text{out}}}} \right|^2 \\
&\geq \frac{1}{4} \sum_{i=1}^{N} \frac{1}{d_i^{\text{in}}} \left(\sum_{j=1}^{N} a_{i,j} \left| \frac{x_i}{\sqrt{d_i^{\text{in}}}} - \frac{x_j}{\sqrt{d_j^{\text{out}}}} \right|^2 \right) \left(\sum_{j=1}^{N} a_{i,j} \right) \\
&\geq \frac{1}{4} \sum_{i=1}^{N} \frac{1}{d_i^{\text{in}}} \left(\sum_{j=1}^{N} a_{i,j} \left| \frac{x_i}{\sqrt{d_i^{\text{in}}}} - \frac{x_j}{\sqrt{d_j^{\text{out}}}} \right| \right)^2 \\
&\geq \frac{1}{4} \sum_{i=1}^{N} \frac{1}{d_i^{\text{in}}} \left| \sum_{j=1}^{N} a_{i,j} \left(\frac{x_i}{\sqrt{d_i^{\text{in}}}} - \frac{x_j}{\sqrt{d_j^{\text{out}}}} \right) \right|^2 \\
&> 0,
\end{aligned}
$$

where we used Cauchy-Schwartz and triangle inequalities. $\qquad\qquad\square$

We give the following simple corollary.

Corollary 1. *For every $x \in \mathbb{R}^{N \times K}$, it holds*

$$\mathscr{E}(\mathbf{x}) = \frac{1}{2} \Re \left(\text{vec}(\mathbf{x})^{\mathsf{H}} (\mathbf{I} \otimes \boldsymbol{\Delta}) \text{vec}(\mathbf{x}) \right).$$

C   Appendix for Section 4

In this section, we provide some properties about FGLs. The first statement shows that the FGL of a normal SNA $\mathbf{L}$ only changes the magnitude of the eigenvalues of $\mathbf{L}$.

Lemma 1. *Let $\mathbf{M}$ be a normal matrix with eigenvalues $\lambda_1, \ldots, \lambda_N$ and corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_N$. Suppose $\mathbf{M} = \mathbf{L}\boldsymbol{\Sigma}\mathbf{R}^{\mathsf{H}}$ is its singular value decomposition. Then it holds*

$$\boldsymbol{\Sigma} = |\boldsymbol{\Lambda}|, \; \mathbf{L} = \mathbf{V}, \; \mathbf{R} = \mathbf{V} \exp(i\boldsymbol{\Theta}), \; \boldsymbol{\Theta} = \text{diag}\left(\{\theta_i\}_{i=1}^{N}\right), \; \theta_i = \text{atan2}\left(\Re\lambda_i, \Im\lambda_i\right).$$

Proof. By hypothesis, there exist a unitary matrix $\mathbf{V}$ such that $\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{H}}$, then

$$\mathbf{M}^{\mathsf{H}}\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}^{*}\mathbf{V}^{\mathsf{H}}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{H}} = \mathbf{V}|\boldsymbol{\Lambda}|^2\mathbf{V}^{\mathsf{H}},$$

$$\mathbf{M}^{\mathsf{H}}\mathbf{M} = \mathbf{R}\boldsymbol{\Sigma}\mathbf{L}^{\mathsf{H}}\mathbf{L}\boldsymbol{\Sigma}\mathbf{R}^{\mathsf{H}} = \mathbf{L}\boldsymbol{\Sigma}^2\mathbf{L}^{\mathsf{H}}.$$

Therefore, $\boldsymbol{\Sigma} = |\boldsymbol{\Lambda}|$ and $\mathbf{L} = \mathbf{V}$

$$\mathbf{M} = \mathbf{R}\,|\boldsymbol{\Lambda}|\,\mathbf{V}^{\mathsf{H}}$$

Finally, $\mathbf{R} = \mathbf{V}\exp(i\boldsymbol{\Theta})$ where $\Theta = \mathrm{diag}\left(\{\mathrm{atan2}(\Re\lambda_i, \Im\lambda_i)\}_{i=1}^N\right)$ and atan2 is the 2-argument arctangent. $\qquad\square$

We proceed by proving Theorem 1, which follows the proof of a similar result given in (Benzi et al., 2020) for the fractional Laplacian defined in the spectral domain of an in-degree normalized graph Laplacian. However, our result also holds for directed graphs and inparticular for fractional Laplacians that are defined via the SVD of a graphs SNA. We begin with

Lemma 2. *Let* $\mathbf{M} \in \mathbb{R}^{n\times n}$ *with singular values* $\sigma(\mathbf{M}) \subset [a,b]$. *For* $f : [a,b] \to \mathbb{R}$, *define* $f(\mathbf{M}) = \mathbf{U}f(\boldsymbol{\Sigma})\mathbf{V}^{\mathsf{H}}$, *where* $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$ *is the singular value decomposition of* $\mathbf{M}$. *If* f *has modulus of continuity* ω *and* $d(i,j) \geq 2$, *it holds*

$$|f(\mathbf{M})|_{i,j} \leq \left(1 + \frac{\pi^2}{2}\right)\omega\left(\frac{b-a}{2}\,|d(i,j) - 1|^{-1}\right).$$

Proof. Let $g : [a,b] \to \mathbb{R}$ be any function, then

$$\begin{aligned}
\|f(\mathbf{M}) - g(\mathbf{M})\|_2 &= \left\|\mathbf{U}f(\boldsymbol{\Sigma})\mathbf{V}^{\mathsf{H}} - \mathbf{U}g(\boldsymbol{\Sigma})\mathbf{V}^{\mathsf{H}}\right\|_2 \\
&= \|f(\boldsymbol{\Sigma}) - g(\boldsymbol{\Sigma})\|_2 \\
&= \|f(\lambda) - g(\lambda)\|_{\infty,\sigma(M)}.
\end{aligned}$$

The second equation holds since the 2-norm is invariant under unitary transformations. By Jackson's Theorem, there exists for every $m \geq 1$ a polynomial p_m of order m such that

$$\|f(\mathbf{M}) - p_m(\mathbf{M})\|_2 \leq \|f - p_m\|_{\infty,[a,b]} \leq \left(1 + \frac{\pi^2}{2}\right)\omega\left(\frac{b-a}{2m}\right).$$

Fix $i,j \in \{1,\ldots,n\}$. If $d(i,j) = m + 1$, then any power of $\mathbf{M}$ up to order m has a zero entry in (i,j), i.e., $(\mathbf{M}^m)_{i,j} = 0$. Hence, $f(\mathbf{M})_{i,j} = f(\mathbf{M})_{i,j} - p_m(\mathbf{M})_{i,j}$, and we get

$$\begin{aligned}
|f(\mathbf{M})_{i,j}| \leq \|f(\mathbf{M}) - p_m(\mathbf{M})\|_2 &\leq \omega\left(1 + \frac{\pi^2}{2}\right)\left(\frac{b-a}{2m}\right) \\
&= \omega\left(1 + \frac{\pi^2}{2}\right)\left(\frac{b-a}{2}|d(i,j) - 1|^{-1}\right)
\end{aligned}$$

from which the thesis follows. $\qquad\square$

Finally, we give a proof of Theorem 1, which is a consequence of the previous statement.

Proof of Theorem 1. The eigenvalues of $\mathbf{L}$ are in the unit circle, i.e., $\|\mathbf{L}\| \leq 1$. Hence, $\|\mathbf{L}\mathbf{L}^{\mathsf{H}}\| \leq 1$, and the singular values of $\mathbf{L}$ are in $[0,1]$. By Lemma 2 and the fact that $f(x) = x^\alpha$ has modulus of continuity $\omega(t) = t^\alpha$ the thesis follows. $\qquad\square$

D   Appendix for Section 5

In this section, we provide the appendix for Section 5. We begin by analyzing the solution of linear matrix ODEs. For this, let $\mathbf{M} \in \mathbb{C}^{N \times N}$. For $\mathbf{x}_0 \in \mathbb{C}^N$, consider the initial value problem

$$\mathbf{x}'(t) = -\mathbf{M}\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{11}$$

Theorem 4 (Existence and uniqueness of linear ODE solution). *The initial value problem given by* (11) *has a unique solution* $\mathbf{x}(t) \in \mathbb{C}^N$ *for any initial condition* $\mathbf{x}_0 \in \mathbb{C}^N$.

The solution of (11) can be expressed using matrix exponentials, even if $\mathbf{M}$ is not symmetric. The matrix exponential is defined as:

$$\exp(-\mathbf{M}t) = \sum_{k=0}^{\infty} \frac{(-\mathbf{M})^k t^k}{k!},$$

where $\mathbf{M}^k$ is the k-th power of the matrix $\mathbf{M}$. The solution of (11) can then be written as

$$\mathbf{x}(t) = \exp(-\mathbf{M}t)\mathbf{x}_0. \tag{12}$$

D.1   Appendix for Section 5.1

In this section, we analyze the solution to (4) and (5). We further provide a proof for Theorem 2. We begin by considering the solution to the fractional heat equation (4). The analysis for the Schrödinger equation (5) follows analogously.

The fractional heat equation $\mathbf{x}'(t) = -\mathbf{L}^\alpha \mathbf{x} \mathbf{W}$ can be vectorized and rewritten via the Kronecker product as

$$\text{vec}(\mathbf{x})'(t) = -\mathbf{W} \otimes \mathbf{L}^\alpha \text{vec}(\mathbf{x})(t). \tag{13}$$

In the undirected case $\mathbf{L}$ and $\mathbf{\Delta}$ are both symmetric, and the eigenvalues satisfy the relation $\lambda_i(\mathbf{\Delta}) = 1 - \lambda_i(\mathbf{L})$. The corresponding eigenvectors $\psi_i(\mathbf{L})$ and $\psi_i(\mathbf{\Delta})$ can be chosen to be the same for $\mathbf{L}$ and $\mathbf{\Delta}$. In the following, we assume that these eigenvectors are orthonormalized.

If $\mathbf{L}$ is symmetric, we can decompose it via the spectral theorem into $\mathbf{L} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U} = [\psi_1(\mathbf{L}), \ldots, \psi_N(\mathbf{L})]$ is an orthogonal matrix containing the eigenvectors of $\mathbf{L}$, and $\mathbf{D}$ is the diagonal matrix of eigenvalues.

Due to Lemma 1, the fractional Laplacian $\mathbf{L}^\alpha$ can be written as $\mathbf{L}^\alpha = \mathbf{U}\,\mathrm{f}_\alpha(\mathbf{D})\mathbf{U}^T$, where $\mathrm{f}_\alpha : \mathbb{R} \to \mathbb{R}$ is the map $x \mapsto \text{sign}(x)\,|x|^\alpha$ and is applied element-wise. Clearly, the eigendecomposition of $\mathbf{L}^\alpha$ is given by the eigenvalues $\{\mathrm{f}_\alpha(\lambda_1(\mathbf{L})), \ldots, \mathrm{f}_\alpha(\lambda_N(\mathbf{L}))\}$ and the corresponding eigenvectors $\{\psi_1(\mathbf{L}), \ldots, \psi_N(\mathbf{L})\}$.

Now, by well-known properties of the Kronecker product, one can write the eigendecomposition of $\mathbf{W} \otimes \mathbf{L}^\alpha$ as

$$\left\{\lambda_r(\mathbf{W})\,\mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right)\right\}_{r \in \{1,\ldots,K\},\, l \in \{1,\ldots,N\}}, \ \left\{\psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L})\right\}_{r \in \{1,\ldots,K\},\, l \in \{1,\ldots,N\}}.$$

Note that $1 \in \lambda(\mathbf{L})$ and, since $\operatorname{trace}(\mathbf{L}) = 0$, the SNA has at least one negative eigenvalue. This property is useful since it allows to retrieve of the indices (r, l) corresponding to eigenvalues with minimal real (or imaginary) parts in a simple way.

The initial condition $\operatorname{vec}(\mathbf{x}_0)$ can be decomposed as

$$\operatorname{vec}(\mathbf{x}_0) = \sum_{r=1}^{K} \sum_{l=1}^{N} c_{r,l}\, \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}), \ c_{r,l} = \langle \operatorname{vec}(\mathbf{x}_0),\ \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{W}) \rangle . \ (14)$$

Then, the solution $\operatorname{vec}(\mathbf{x})(t)$ of (13) can be written as

$$\operatorname{vec}(\mathbf{x})(t) = \sum_{r=1}^{K} \sum_{l=1}^{N} c_{r,l}\, \exp\left(-t\lambda_r(\mathbf{W})\, \mathrm{f}_\alpha(\lambda_l(\mathbf{L}))\right) \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}). \qquad (15)$$

The following result shows the relationship between the frequencies of $\boldsymbol{\Delta}$ and the Dirichlet energy and serves as a basis for the following proofs.

Lemma 3. *Let $\mathcal{G}$ be a graph with SNL $\boldsymbol{\Delta} \in \mathbb{R}^{N \times N}$. Consider $\mathbf{x}(t) \in \mathbb{C}^{N \times K}$ such that there exists $\boldsymbol{\varphi} \in \mathbb{C}^{N \times K} \setminus \{0\}$ with*

$$\frac{\operatorname{vec}(\mathbf{x})(t)}{\|\operatorname{vec}(\mathbf{x})(t)\|_2} \xrightarrow{t \to \infty} \operatorname{vec}(\boldsymbol{\varphi}),$$

and $(\mathbf{I} \otimes \boldsymbol{\Delta})\operatorname{vec}(\boldsymbol{\varphi}) = \lambda \operatorname{vec}(\boldsymbol{\varphi})$. Then,

$$\mathscr{E}\left(\frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2}\right) \xrightarrow{t \to \infty} \frac{\Re(\lambda)}{2} .$$

Proof. As $\operatorname{vec}(\boldsymbol{\varphi})$ is the limit of unit vectors, $\operatorname{vec}(\boldsymbol{\varphi})$ is a unit vector itself. We calculate its Dirichlet energy,

$$\mathscr{E}(\operatorname{vec}(\boldsymbol{\varphi})) = \frac{1}{2}\Re\left(\operatorname{vec}(\boldsymbol{\varphi})^{\mathsf{H}}(\mathbf{I} \otimes \boldsymbol{\Delta})\operatorname{vec}(\boldsymbol{\varphi})\right) = \frac{1}{2}\Re\left(\lambda\,\operatorname{vec}(\boldsymbol{\varphi})^{\mathsf{H}}\operatorname{vec}(\boldsymbol{\varphi})\right) = \frac{1}{2}\Re(\lambda) .$$

Since $\mathbf{x} \mapsto \mathscr{E}(\mathbf{x})$ is continuous, the thesis follows. $\qquad\square$

Another useful result that will be extensively used in proving Theorem 2 is presented next.

Lemma 4. *Suppose $\mathbf{x}(t)$ can be expressed as*

$$\mathbf{x}(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} c_{k,n}\, \exp\left(-t\,\lambda_{k,n}\right) \mathbf{v}_k \otimes \mathbf{w}_n ,$$

for some choice of $c_{k,n}$, $\lambda_{k,n}$, $\{\mathbf{v}_k\}$, $\{\mathbf{w}_n\}$. Let (a, b) be the unique index of $\lambda_{k,n}$ with minimal real part and corresponding non-null coefficient $c_{k,n}$, i.e.

$$(a, b) := \underset{(k,n) \in [K] \times [N]}{\arg\min}\ \{\Re(\lambda_{k,n}) : c_{k,n} \neq 0\} .$$

Then

$$\frac{x(t)}{\|x(t)\|_2} \xrightarrow{t \to \infty} \frac{c_{a,b}\, \mathbf{v}_a \otimes \mathbf{w}_b}{\|c_{a,b}\, \mathbf{v}_a \otimes \mathbf{w}_b\|_2} .$$

Proof. The key insight is to separate the addend with index (a, b). It holds

$$\mathbf{x}(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} c_{k,n} \exp\left(-t\,\lambda_{k,n}\right) \mathbf{v}_n \otimes \mathbf{w}_m$$

$$= \exp\left(-t\,\lambda_{a,b}\right) \left(c_{a,b}\mathbf{v}_a \otimes \mathbf{w}_b + \sum_{\substack{(k,n)\in[K]\times[N] \\ (k,n)\neq(a,b)}} c_{k,n} \exp\left(-t\,(\lambda_{k,n}-\lambda_{a,b})\right) \mathbf{v}_k \otimes \mathbf{w}_n \right).$$

We note that

$$\lim_{t\to\infty} \left|\exp\left(-t\,(\lambda_{k,n}-\lambda_{a,b})\right)\right| = \lim_{t\to\infty} \left|\exp\left(-t\,\Re\left(\lambda_{k,n}-\lambda_{a,b}\right)\right)\exp\left(-i\,t\,\Im\left(\lambda_{k,n}-\lambda_{a,b}\right)\right)\right|$$

$$= \lim_{t\to\infty} \exp\left(-t\,\Re\left(\lambda_{k,n}-\lambda_{a,b}\right)\right)$$

$$= 0\,,$$

for all $(k,n) \neq (a,b)$, since $\Re\left(\lambda_{k,n}-\lambda_{a,b}\right) > 0$. Therefore, one gets

$$\frac{x(t)}{\|x(t)\|_2} \xrightarrow{t\to\infty} \frac{c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b}{\|c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b\|_2}\,,$$

where the normalization removes the dependency on $\exp\left(-t\,\lambda_{a,b}\right)$ □

When $\lambda_{a,b}$ is not unique, it is still possible to derive a convergence result. In this case, $\mathbf{x}$ will converge to an element in the span generated by vectors corresponding to $\lambda_{a,b}$, i.e.,

$$\frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2} \xrightarrow{t\to\infty} \frac{\displaystyle\sum_{(a,b)\in\mathcal{A}} c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b}{\left\|\displaystyle\sum_{(a,b)\in\mathcal{A}} c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b\right\|_2}\,,$$

where $\mathcal{A} := \{(k,n) : \Re(\lambda_{k,n}) = \Re(\lambda_{a,b})\,,\ c_{k,n} \neq 0\}$.

A similar result to Lemma 4 holds for a slightly different representation of $\mathbf{x}(t)$.

Lemma 5. *Suppose* $\mathbf{x}(t)$ *can be expressed as*

$$\mathbf{x}(t) = \sum_{k=1}^{K} \sum_{n=1}^{N} c_{k,n} \exp\left(i\,t\,\lambda_{k,n}\right) \mathbf{v}_k \otimes \mathbf{w}_n\,,$$

for some choice of $c_{k,n}$, $\lambda_{k,n}$, $\{\mathbf{v}_k\}$, $\{\mathbf{w}_n\}$. *Let* (a,b) *be the unique index of* $\lambda_{k,n}$ *with minimal imaginary part and corresponding non-null coefficient* $c_{k,n}$, *i.e.*

$$(a,b) := \operatorname*{arg\,min}_{(k,n)\in[K]\times[N]} \left\{\Im\left(\lambda_{k,n}\right) : c_{k,n} \neq 0\right\}.$$

Then

$$\frac{x(t)}{\|x(t)\|_2} \xrightarrow{t\to\infty} \frac{c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b}{\|c_{a,b}\,\mathbf{v}_a \otimes \mathbf{w}_b\|_2}\,.$$

Proof. The proof follows the same reasoning as in the proof of Lemma 4. The difference is that the dominating frequency is the one with the minimal imaginary part, since

$$\Re\left(i\,\lambda_{k,n}\right) = -\Im\left(\lambda_{k,n}\right)\,,$$

and, consequently,

$$\arg\max_{(k,n)\in[K]\times[N]}\left\{\Re\left(i\,\lambda_{k,n}\right)\right\} = \arg\min_{(k,n)\in\in[K]\times[N]}\left\{\Im\left(\lambda_{k,n}\right)\right\}\,.$$

$\square$

Proof of Theorem 2 We denote the eigenvalues of $\mathbf{L}$ closest to 0 from above and below as

$$\lambda_{+}(\mathbf{L}) := \arg\min_{l}\left\{\lambda_{l}(\mathbf{L})\ :\ \lambda_{l}(\mathbf{L}) > 0\right\}\,,$$
$$\lambda_{-}(\mathbf{L}) := \arg\max_{l}\left\{\lambda_{l}(\mathbf{L})\ :\ \lambda_{l}(\mathbf{L}) < 0\right\}\,. \tag{16}$$

We assume that the channel mixing $\mathbf{W} \in \mathbb{R}^{K \times K}$ and the graph Laplacians $\mathbf{L}, \mathbf{\Delta} \in \mathbb{R}^{N \times N}$ are real matrices. Finally, we suppose the eigenvalues of a generic matrix $\mathbf{M}$ are sorted in ascending order, i.e., $\lambda_i(\mathbf{M}) \leq \lambda_j(\mathbf{M})$, $i < j$.

We now reformulate Theorem 2 for the fractional heat equation (4) and provide its full proof, which follows a similar frequency analysis to the one in (Di Giovanni et al., 2022, Theorem B.3)

Theorem 5. *Let $\mathcal{G}$ be an undirected graph with SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. Consider the initial value problem in (4) with channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ and $\alpha \in \mathbb{R}$. Then, for almost all initial conditions $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$ the following is satisfied.*

(i) If $\alpha > 0$, the solution to (4) is HFD if

$$\lambda_K(\mathbf{W})\,\mathrm{f}_{\alpha}\left(\lambda_1(\mathbf{L})\right) < \lambda_1(\mathbf{W})\,,$$

   and LFD otherwise.
(ii) If $\alpha < 0$, the solution to (4) is $(1 - \lambda_{-}(\mathbf{L}))$-FD if

$$\lambda_K(\mathbf{W})\,\mathrm{f}_{\alpha}\left(\lambda_{-}(\mathbf{L})\right) < \lambda_1(\mathbf{W})\,\mathrm{f}_{\alpha}\left(\lambda_{+}(\mathbf{L})\right)\,,$$

   and $(1 - \lambda_{+}(\mathbf{L}))$-FD otherwise.

Proof of (i).. As derived in (15), the solution of (4) with initial condition $\mathbf{x}_0$ can be written in a vectorized form as

$$\mathrm{vec}(\mathbf{x})(t) = \exp\left(-t\,\mathbf{W}^{\mathsf{T}} \otimes \mathbf{L}^{\alpha}\right)\mathrm{vec}(\mathbf{x}_0)$$
$$= \sum_{r=1}^{K}\sum_{l=1}^{N} c_{r,l}\,\exp\left(-t\,\lambda_r(\mathbf{W})\,\mathrm{f}_{\alpha}\left(\lambda_l(\mathbf{L})\right)\right)\psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}), \tag{17}$$

where $\lambda_r(\mathbf{W})$ are the eigenvalues of $\mathbf{W}$ with corresponding eigenvectors $\psi_r(\mathbf{W})$, and $\lambda_l(\mathbf{L})$ are the eigenvalues of $\mathbf{L}$ with corresponding eigenvectors $\psi_l(\mathbf{L})$. The coefficients $c_{r,l}$ are the Fourier coefficients of $\mathbf{x}_0$, i.e.,

$$c_{r,l} := \langle \mathrm{vec}(\mathbf{x}_0),\ \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}) \rangle .$$

The key insight is to separate the eigenprojection corresponding to the most negative frequency. By Lemma 4, this frequency component dominates for t going to infinity.

Suppose

$$\lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right) < \lambda_1(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_N(\mathbf{L})\right) = \lambda_1(\mathbf{W}) .$$

In this case, $\lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right)$ is the most negative frequency. Assume for simplicity that $\lambda_K(\mathbf{W})$ has multiplicity one; the argument can be applied even if this is not the case, since the corresponding eigenvectors are orthogonal for higher multiplicities.

For almost all initial conditions $\mathbf{x}_0$, the coefficient $c_{K,1}$ is not null; hence

$$\frac{\mathrm{vec}(\mathbf{x})(t)}{\|\mathrm{vec}(\mathbf{x})(t)\|_2} \xrightarrow{t \to \infty} \frac{c_{K,1}\, \psi_K\left(\mathbf{W}\right) \otimes \psi_1\left(\mathbf{L}\right)}{\|c_{K,1}\, \psi_K(\mathbf{W}) \otimes \psi_1\left(\mathbf{L}\right)\|_2} .$$

By standard properties of the Kronecker product, we have

$$\begin{aligned}
(\mathbf{I} \otimes \mathbf{L})\left(\psi_K\left(\mathbf{W}\right) \otimes \psi_1\left(\mathbf{L}\right)\right) &= \left(\mathbf{I}\, \psi_K\left(\mathbf{W}\right)\right) \otimes \left(\mathbf{L}\, \psi_1\left(\mathbf{L}\right)\right) \\
&= \lambda_1(\mathbf{L})\, \psi_K\left(\mathbf{W}\right) \otimes \psi_1\left(\mathbf{L}\right) ,
\end{aligned} \tag{18}$$

i.e., $\psi_K\left(\mathbf{W}\right) \otimes \psi_1\left(\mathbf{L}\right)$ is an eigenvector of $\mathbf{I} \otimes \mathbf{L}$ corresponding to the eigenvalue $\lambda_1(\mathbf{L})$. Then, by Proposition 1, $\psi_K\left(\mathbf{W}\right) \otimes \psi_1\left(\mathbf{L}\right)$ is also an eigenvector of $\mathbf{I} \otimes \boldsymbol{\Delta}$ corresponding to the eigenvalue $1 - \lambda_1(\mathbf{L}) = \lambda_N(\boldsymbol{\Delta})$. An application of Lemma 3 finishes the proof.

Similarly, we can show that if $\alpha > 0$ and $\lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1\left(\mathbf{L}\right)\right) > \lambda_1\left(\mathbf{W}\right)$ the lowest frequency component $\lambda_1(\boldsymbol{\Delta})$ is dominant.

Proof of (ii). In this case either $\mathrm{f}_\alpha\left(\lambda_+\left(\mathbf{L}\right)\right)\lambda_1\left(\mathbf{W}\right)$ or $\mathrm{f}_\alpha\left(\lambda_-\left(\mathbf{L}\right)\right)\lambda_K\left(\mathbf{W}\right)$ are the most negative frequency components. Hence, if

$$\mathrm{f}_\alpha\left(\lambda_-\left(\mathbf{L}\right)\right)\lambda_K\left(\mathbf{W}\right) > \mathrm{f}_\alpha\left(\lambda_+\left(\mathbf{L}\right)\right)\lambda_1\left(\mathbf{W}\right)$$

the frequency $\mathrm{f}_\alpha\left(\lambda_+\left(\mathbf{L}\right)\right)\lambda_1\left(\mathbf{W}\right)$ is dominating and otherwise the frequency $\mathrm{f}_\alpha\left(\lambda_-\left(\mathbf{L}\right)\right)\lambda_K\left(\mathbf{W}\right)$. We can see this by following the exact same reasoning of *(i)*. $\qquad\square$

Remark 2. In the proof of *(ii)*, we are tacitly assuming that $\mathbf{L}$ has only non-zero eigenvalues. If not, we can truncate the SVD and remove all zeros singular values (which correspond to zeros eigenvalues). In doing so, we obtain the best invertible approximation of $\mathbf{L}$ to which the theorem can be applied.

In the following, we provide the full statement and proof for the claims made in Theorem 2 when the underlying ODE is the Schrödinger equation as presented in (5).

Theorem 6. *Let $\mathcal{G}$ be a undirected graph with SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. Consider the initial value problem in (5) with channel mixing matrix $\mathbf{W} \in \mathbb{C}^{K \times K}$ and $\alpha \in \mathbb{R}$. Suppose that $\mathbf{W}$ has at least one eigenvalue with non-zero imaginary part and sort the eigenvalues of $\mathbf{W}$ in ascending order with respect to their imaginary part. Then, for almost initial values $\mathbf{x}_0 \in \mathbb{C}^{N \times K}$ the following is satisfied.*

(i) If $\alpha > 0$, solutions of (5) are HFD if

$$\Im\left(\lambda_K(\mathbf{W})\right) f_\alpha\left(\lambda_1(\mathbf{L})\right) < \Im\left(\lambda_1(\mathbf{W})\right) ,$$

and LFD otherwise.

(ii) Let $\lambda_+(\mathbf{L})$ and $\lambda_-(\mathbf{L})$ be the smallest positive and biggest negative non-zero eigenvalue of $\mathbf{L}$, respectively. If $\alpha < 0$, solutions of (5) are $(1 - \lambda_-(\mathbf{L}))$-FD if

$$\Im\left(\lambda_K(\mathbf{W})\right) f_\alpha\left(\lambda_-(\mathbf{L})\right) < \Im\left(\lambda_1(\mathbf{W})\right) f_\alpha\left(\lambda_+(\mathbf{L})\right) .$$

Otherwise, solutions of (5) are $(1 - \lambda_+(\mathbf{L}))$-FD.

Proof. The proof follows the same reasoning as the proof for the heat equation in Theorem 5. The difference is that we now apply Lemma 5 instead of Lemma 4.

Therefore, if $\alpha > 0$ the dominating frequency is either $\lambda_K(\mathbf{W}) f_\alpha\left(\lambda_1(\mathbf{L})\right)$ or $\lambda_1(\mathbf{W}) f_\alpha\left(\lambda_N(\mathbf{L})\right)$, and $\lambda_K(\mathbf{W}) f_\alpha\left(\lambda_-(\mathbf{L})\right)$ or $\lambda_1(\mathbf{W}) f_\alpha\left(\lambda_+(\mathbf{L})\right)$ if $\alpha < 0$. $\qquad\square$

D.2 Frequency Dominance for Numerical Approximations of the Heat Equation

For $n \in \mathbb{N}$ and $h \in \mathbb{R}$, $h > 0$, the solution of (4) at time $nh > 0$ can be approximated with an explicit Euler scheme

$$\text{vec}(\mathbf{x})(n\,h) = \sum_{k=0}^{n} \binom{n}{k} h^k (-\mathbf{W} \otimes \mathbf{L}^\alpha)^k \text{vec}(\mathbf{x}_0) ,$$

which can be further simplified via the binomial theorem as

$$\text{vec}(\mathbf{x})(n\,h) = \left(\mathbf{I} - h\left(\mathbf{W} \otimes \mathbf{L}^\alpha\right)\right)^n \text{vec}(\mathbf{x}_0) .$$

Hence, it holds the representation formula

$$\text{vec}(\mathbf{x})(n\,h) = \sum_{r,l} c_{r,l} \left(1 - h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right)\right)^n \psi_r\left(\mathbf{W}\right) \otimes \psi_l\left(\mathbf{L}\right) .$$

In this case, the dominating frequency maximizes $\left|1 - h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right)\right|$. When $h < \|\mathbf{W}\|^{-1}$, the product $h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right)$ is guaranteed to be in $[-1, 1]$, and

$$\left|1 - h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right)\right| = 1 - h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right) \in [0, 2] .$$

Therefore, the dominating frequency minimizes $h\,\lambda_r\left(\mathbf{W}\right) f_\alpha\left(\lambda_l(\mathbf{L})\right)$. This is the reasoning behind the next result.

Proposition 7. *Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional heat equation* (4) *with $\alpha \in \mathbb{R}$. Let $\{\mathbf{x}(n\,h)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating* (4) *with an explicit Euler scheme with step size h. Suppose $h < \|\mathbf{W}\|^{-1}$. Then, for almost all initial values $\mathbf{x}_0$*

$$\mathscr{E} \left(\frac{\mathbf{x}(n\,h)}{\|\mathbf{x}(n\,h)\|_2} \right) \xrightarrow{n \to \infty} \begin{cases} \dfrac{\lambda_N(\boldsymbol{\Delta})}{2}, & \text{if } \lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right) < \lambda_1(\mathbf{W}), \\ 0, & \text{otherwise}. \end{cases}$$

Proof. Define

$$(\lambda_a, \lambda_b) \coloneqq \arg\max_{r,l} \left\{ |1 - h\lambda_r(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right)| : r \in \{1, \dots, K\}, l \in \{1, \dots, N\} \right\}.$$

By the hypothesis on h, this is equivalent to

$$(\lambda_a, \lambda_b) = \arg\min_{r,l} \left\{ \lambda_r(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right) : r \in \{1, \dots, K\}, l \in \{1, \dots, N\} \right\}.$$

Therefore, (λ_a, λ_b) is either $(\lambda_1(\mathbf{W}), \lambda_N(\mathbf{L}))$ or $(\lambda_K(\mathbf{W}), \lambda_1(\mathbf{L}))$. Hence,

$$\frac{\mathrm{vec}(\mathbf{x})(n\,h)}{\|\mathrm{vec}(\mathbf{x})(n\,h)\|} \xrightarrow{n \to \infty} \frac{c_{a,b}\, \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})}{\|c_{a,b}\, \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})\|}.$$

If the condition $\lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right) < \lambda_1(\mathbf{W})$ is satisfied, we have $b = 1$. Then by (18), the normalized $\mathrm{vec}(\mathbf{x})$ converges to the eigenvector of $\mathbf{I} \otimes \boldsymbol{\Delta}$ corresponding to the largest frequency $1 - \lambda_1(\mathbf{L}) = \lambda_N(\boldsymbol{\Delta})$. An application of Lemma 3 finishes the proof.

If $\lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right) < \lambda_1(\mathbf{W})$ is not satisfied, we have $b = N$, and the other direction follows with the same argument. $\square$

Similarly to Proposition 7 one can prove the following results for negative fractions.

Proposition 8. *Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional heat equation* (4) *with $\alpha < 0$. Let $\{\mathbf{x}(n\,h)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating the solution of* (4) *with an explicit Euler scheme with step size h. Suppose that $h < \|\mathbf{W}\|^{-1}$. The approximated solution is $(1 - \lambda_-(\mathbf{L}))$-FD if*

$$\lambda_1(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_+(\mathbf{L})\right) < \lambda_K(\mathbf{W})\, \mathrm{f}_\alpha\left(\lambda_-(\mathbf{L})\right),$$

and $(1 - \lambda_+(\mathbf{L}))$-FD otherwise.

Proof. The proof follows the same reasoning as the proof of Proposition 7 by realizing that the dominating frequencies (λ_a, λ_b) are either given by $(\lambda_1(\mathbf{W}), \lambda_+(\mathbf{L}))$ or $(\lambda_K(\mathbf{W}), \lambda_-(\mathbf{L}))$. $\square$

D.3 Frequency Dominance for Numerical Approximations of the Schrödinger Equation

For $n \in \mathbb{N}$ and $h \in \mathbb{R}$, $h > 0$, the solution of (5) at time $n\,h > 0$ can be approximated with an explicit Euler scheme as well. Similarly to the previous section, we can write

$$\mathrm{vec}(\mathbf{x})(n\,h) = \left(\mathbf{I} + i\,h\left(\mathbf{W} \otimes \mathbf{L}^\alpha\right)\right)^n \mathrm{vec}(\mathbf{x}_0)\,. \tag{19}$$

and

$$\mathrm{vec}(\mathbf{x})(n\,h) = \sum_{r,l} c_{r,l}\,\left(1 + i\,h\,\lambda_r\left(\mathbf{W}\right) \mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right)\right)^n \psi_r\left(\mathbf{W}\right) \otimes \psi_l\left(\mathbf{L}\right)\,.$$

The dominating frequency will be discussed in the following theorem.

Proposition 9. *Let $h \in \mathbb{R}$, $h > 0$. Let $\{\mathbf{x}(n\,h)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating (5) with an explicit Euler scheme with sufficiently small step size h. Sort the eigenvalues of $\mathbf{W}$ in ascending order with respect to their imaginary part. Then, for almost all initial values $\mathbf{x}_0$*

$$\mathscr{E}\left(\frac{\mathbf{x}(n\,h)}{\|\mathbf{x}(n\,h)\|_2}\right) \xrightarrow{n \to \infty} \begin{cases} \dfrac{\lambda_N(\boldsymbol{\Delta})}{2}\,, & \text{if } \mathrm{f}_\alpha\left(\lambda_1(\mathbf{L})\right)\Im\left(\lambda_K(\mathbf{W})\right) < \mathrm{f}_\alpha\left(\lambda_N(\mathbf{L})\right)\Im\left(\lambda_1(\mathbf{W})\right) \\ 0\,, & \text{otherwise.} \end{cases}$$

Proof. Define

$$(\lambda_a, \lambda_b) \coloneqq \arg\max_{r,l}\left\{\left|1 + i\,h\lambda_r\left(\mathbf{W}\right) \mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right)\right| \,:\, r \in \{1, \dots, K\}\,,\, l \in \{1, \dots, N\}\right\}\,.$$

By definition of a and b, for all r and l it holds

$$\left|1 + i\,h\,\lambda_a\left(\mathbf{W}\right) \mathrm{f}_\alpha\left(\lambda_b(\mathbf{L})\right)\right| > \left|1 + i\,h\,\lambda_r\left(\mathbf{W}\right) \mathrm{f}_\alpha\left(\lambda_l(\mathbf{L})\right)\right|\,. \tag{20}$$

Hence,

$$\frac{\mathrm{vec}(\mathbf{x})(t)}{\|\mathrm{vec}(\mathbf{x})(t)\|} \xrightarrow{t \to \infty} \frac{c_{a,b}\psi_a\left(\mathbf{W}\right) \otimes \psi_b\left(\mathbf{L}\right)}{\|c_{a,b}\psi_a\left(\mathbf{W}\right) \otimes \psi_b\left(\mathbf{L}\right)\|}\,.$$

We continue by determining the indices a and b. To do so, we note that (20) is equivalent to

$$\mathrm{f}_\alpha\left(\lambda_l\left(\mathbf{L}\right)\right)\Im\left(\lambda_r\left(\mathbf{W}\right)\right) - \mathrm{f}_\alpha\left(\lambda_b\left(\mathbf{L}\right)\right)\Im\left(\lambda_a\left(\mathbf{W}\right)\right)$$
$$> \frac{h}{2}\left(\mathrm{f}_\alpha\left(\lambda_l\left(\mathbf{L}\right)\right)^2 \left|\lambda_r\left(\mathbf{W}\right)\right|^2 - \mathrm{f}_\alpha\left(\lambda_b(\mathbf{L})\right)^2 \left|\lambda_a\left(\mathbf{W}\right)\right|^2\right)$$

for all r, l. Denote by ε the gap

$$0 < \varepsilon \coloneqq \min_{(r,l) \neq (a,b)}\left\{\mathrm{f}_\alpha\left(\lambda_l\left(\mathbf{L}\right)\right)\Im\left(\lambda_r\left(\mathbf{W}\right)\right) - \mathrm{f}_\alpha\left(\lambda_b\left(\mathbf{L}\right)\right)\Im\left(\lambda_a\left(\mathbf{W}\right)\right)\right\}\,.$$

Noting that

$$\begin{cases} \dfrac{h}{2} \left(f_\alpha \left(\lambda_l \left(\mathbf{L} \right) \right)^2 \left| \lambda_r \left(\mathbf{W} \right) \right|^2 - f_\alpha \left(\lambda_b (\mathbf{L}) \right)^2 \left| \lambda_a \left(\mathbf{W} \right) \right|^2 \right) \le h \left\| \mathbf{W} \right\|^2 \left\| \mathbf{L} \right\|^{2\alpha} = h \left\| \mathbf{W} \right\|^2, \\ \dfrac{h}{2} \left(f_\alpha \left(\lambda_l \left(\mathbf{L} \right) \right)^2 \left| \lambda_r \left(\mathbf{W} \right) \right|^2 - f_\alpha \left(\lambda_b (\mathbf{L}) \right)^2 \left| \lambda_a \left(\mathbf{W} \right) \right|^2 \right) < \varepsilon \end{cases}$$

one gets that (20) is satisfied for $h < \varepsilon \left\| \mathbf{W} \right\|^{-2}$. Therefore, for sufficiently small h, the dominating frequencies are the ones with minimal imaginary part, i.e., either $f_\alpha \left(\lambda_1(\mathbf{L}) \right) \Im \left(\lambda_K(\mathbf{W}) \right)$ or $f_\alpha \left(\lambda_N(\mathbf{L}) \right) \Im \left(\lambda_1(\mathbf{W}) \right)$. If $f_\alpha \left(\lambda_1(\mathbf{L}) \right) \Im \left(\lambda_K(\mathbf{W}) \right) < f_\alpha \left(\lambda_N(\mathbf{L}) \right) \Im \left(\lambda_1(\mathbf{W}) \right)$, then $b = 1$, and the normalized $\mathrm{vec}\left(\mathbf{x} \right)$ converges to the eigenvector corresponding to the smallest frequency $\lambda_1(\mathbf{L})$. By (18), this is also the eigenvector of $\mathbf{I} \otimes \boldsymbol{\Delta}$ corresponding to the largest frequency $1 - \lambda_1(\mathbf{L}) = \lambda_N(\boldsymbol{\Delta})$. An application of Lemma 3 finishes the proof. $\square$

Finally, we present a similar result for negative powers.

Proposition 10. *Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional Schrödinger equation* (5) *with $\alpha < 0$. Let $\{\mathbf{x}(n\,h)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating the solution of* (5) *with an explicit Euler scheme with step size h. Suppose that h is sufficiently small. Sort the eigenvalues of $\mathbf{W}$ in ascending order with respect to their imaginary part. The approximated solution is $(1 - \lambda_+(\mathbf{L}))$-FD if*

$$\lambda_1(\mathbf{W}) \, f_\alpha \left(\lambda_+(\mathbf{L}) \right) < \lambda_K(\mathbf{W}) \, f_\alpha \left(\lambda_-(\mathbf{L}) \right),$$

and $(1 - \lambda_-(\mathbf{L}))$-FD otherwise.

Proof. Similar to Proposition 9, we can prove the statement by realizing that the dominating frequencies (λ_a, λ_b) in (20) are either given by $(\lambda_1(\mathbf{W}), \lambda_+(\mathbf{L}))$ or $(\lambda_K(\mathbf{W}), \lambda_-(\mathbf{L}))$. $\square$

E   Appendix for Section 5.2

We begin this section by describing the solution of general linear matrix ODEs of the form (12) in terms of the Jordan decomposition of $\mathbf{M}$. This is required when $\mathbf{M}$ is not diagonalizable. For instance, the SNA of a directed graph is not in general a symmetric matrix, hence, not guaranteed to be diagonalizable. We then proceed in Appendix E.1 with the proof of Theorem 3.

For a given matrix $\mathbf{M} \in \mathbb{C}^{N \times N}$, the Jordan normal form is given by

$$\mathbf{M} = \mathbf{P} \mathbf{J} \mathbf{P}^{-1},$$

where $\mathbf{P} \in \mathbb{C}^{N \times N}$ is an invertible matrix whose columns are the generalized eigenvectors of $\mathbf{M}$, and $\mathbf{J} \in \mathbb{C}^{N \times N}$ is a block-diagonal matrix with Jordan blocks along its diagonal. Denote with $\lambda_1, \dots, \lambda_m$ the eigenvalues of $\mathbf{M}$ and with $\mathbf{J}_1, \dots, \mathbf{J}_m$ the corresponding Jordan blocks. Let k_l be the algebraic multiplicity of the eigenvalue λ_l, and denote with $\left\{ \psi_l^i(\mathbf{M}) \right\}_{i \in \{1, \dots, k_l\}}$ the generalized eigenvectors of the Jordan block $\mathbf{J}_l$.

We begin by giving the following well-known result, which fully characterizes the frequencies for the solution of a linear matrix ODE.

Lemma 6. *Let* $\mathbf{M} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1} \in \mathbb{C}^{N \times N}$ *be the Jordan normal form of* $\mathbf{M}$. *Let* $\mathbf{x} : [0, T] \to \mathbb{R}^n$ *be a solution to*

$$\mathbf{x}'(t) = \mathbf{M}\mathbf{x}(t), \ \mathbf{x}(0) = \mathbf{x}_0.$$

Then, $\mathbf{x}$ *is given by*

$$\mathbf{x}(t) = \sum_{l=1}^{m} \exp\left(\lambda_l(\mathbf{M})t\right) \sum_{i=1}^{k_l} c_l^j \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} \psi_l^j(\mathbf{M}),$$

where

$$\mathbf{x}_0 = \sum_{l=1}^{m} \sum_{i=1}^{k_l} c_l^i \mathbf{P}\mathbf{e}_l^i,$$

and $\left\{\mathbf{e}_l^i : i \in \{1, \ldots k_l\}, \ l \in \{1, \ldots, m\}\right\}$ *is the standard basis satisfying* $\mathbf{P}\mathbf{e}_l^i = \psi_l^i(\mathbf{M})$.

Proof. By (Perko, 2001, Section 1.8), the solution can be written as

$$\exp\left(\mathbf{M}\,t\right)\mathbf{x}_0 = \mathbf{P}\exp\left(\mathbf{J}\,t\right)\mathbf{P}^{-1}\left(\sum_{l=1}^{m}\sum_{i=1}^{k_l} c_l^i \mathbf{P}\mathbf{e}_l^i\right) = \mathbf{P}\exp\left(\mathbf{J}\,t\right)\left(\sum_{l=1}^{m}\sum_{i=1}^{k_l} c_l^i \mathbf{e}_l^i\right),$$

where $\exp\left(\mathbf{J}\,t\right) = \mathrm{diag}\left(\{\exp\left(\mathbf{J}_l\,t\right)\}_{l=1}^{m}\right)$ and

$$\exp\left(\mathbf{J}_l\,t\right) = \exp\left(\lambda_l(\mathbf{M})\,t\right)\begin{bmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{k_l}}{(k_l-1)!} \\ & 1 & t & & \vdots \\ & & 1 & \ddots & \frac{t^2}{2!} \\ & & & \ddots & t \\ & & & & 1 \end{bmatrix}.$$

Since $\exp\left(\mathbf{J}\,t\right) = \bigoplus_{l=1}^{m} \exp\left(\mathbf{J}_l\,t\right)$, we can focus on a single Jordan block. Fix $l \in \{1, \ldots, m\}$, it holds

$$\mathbf{P}\exp\left(\mathbf{J}_l\,t\right)\left(\sum_{i=1}^{k_l} c_l^i \mathbf{e}_l^i\right)$$

$$= \mathbf{P}\exp\left(\lambda_l(\mathbf{M})\,t\right)\left(c_l^1 \mathbf{e}_l^1 + c_l^2 \left(t\,\mathbf{e}_l^1 + \mathbf{e}_l^2\right) + c_l^3 \left(\frac{t^2}{2!}\mathbf{e}_l^1 + t\,\mathbf{e}_l^2 + \mathbf{e}_l^3\right) + \ldots\right)$$

$$= \exp\left(\lambda_l(\mathbf{M})\,t\right)\left(c_l^1 \psi_l^1(\mathbf{M}) + c_l^2 \left(t\,\psi_l^1(\mathbf{M}) + \psi_l^2(\mathbf{M})\right)\right.$$

$$\left. + c_l^3 \left(\frac{t^2}{2!}\psi_l^1(\mathbf{M}) + t\,\psi_l^2(\mathbf{M}) + \psi_l^3(\mathbf{M})\right) + \ldots\right)$$

$$= \exp\left(\lambda_l(\mathbf{M})\,t\right)\sum_{i=1}^{k_l} c_l^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!}\psi_l^j(\mathbf{M}).$$

Bringing the direct sums together, we get

$$\exp\left(\mathbf{M}\,t\right)\mathbf{x}_0 = \sum_{l=1}^{m}\exp\left(\lambda_l(\mathbf{M})\,t\right)\sum_{i=1}^{k_l}c_l^i\sum_{j=1}^{i}\frac{t^{i-j}}{(i-j)!}\psi_l^j(\mathbf{M})\,,$$

from which the thesis follows. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the following, we derive a formula for the solution of ODEs of the form

$$\mathbf{x}'(t) = \mathbf{M}\mathbf{x}(t)\mathbf{W}\,,\ \mathbf{x}(0) = \mathbf{x}_0\,, \qquad\qquad (21)$$

for a diagonal matrix $\mathbf{W}\in\mathbb{C}^{K\times K}$ and a general square matrix $\mathbf{M}\in\mathbb{C}^{N\times N}$ with Jordan normal form $\mathbf{PJP}^{-1}$. By vectorizing (21), we obtain the equivalent linear system

$$\operatorname{vec}(\mathbf{x})'(t) = \mathbf{W}\otimes\mathbf{M}\operatorname{vec}(\mathbf{x})(t)\,,\ \operatorname{vec}(\mathbf{x})(0) = \operatorname{vec}(\mathbf{x}_0)\,. \qquad (22)$$

Then, by properties of the Kronecker product, there holds

$$\mathbf{W}\otimes\mathbf{M} = \mathbf{W}\otimes(\mathbf{PJP}^{-1}) = (\mathbf{I}\otimes\mathbf{P})(\mathbf{W}\otimes\mathbf{J})(\mathbf{I}\otimes\mathbf{P}^{-1}) = (\mathbf{I}\otimes\mathbf{P})(\mathbf{W}\otimes\mathbf{J})(\mathbf{I}\otimes\mathbf{P})^{-1}.$$

Note that $(\mathbf{I}\otimes\mathbf{P})(\mathbf{W}\otimes\mathbf{J})(\mathbf{I}\otimes\mathbf{P})^{-1}$ is not the Jordan normal form of $\mathbf{D}\otimes\mathbf{M}$. However, we can characterize the Jordan form of $\mathbf{W}\otimes\mathbf{M}$ as follows.

Lemma 7. *The Jordan decomposition of* $\mathbf{W}\otimes\mathbf{J}$ *is given by* $\mathbf{W}\otimes\mathbf{J} = \tilde{\mathbf{P}}\tilde{\mathbf{J}}\tilde{\mathbf{P}}^{-1}$ *where* $\tilde{\mathbf{J}}$ *is a block diagonal matrix with blocks*

$$\tilde{\mathbf{J}}_{j,l} = \begin{bmatrix} w_j\lambda_l(\mathbf{J}) & 1 & & & \\ & w_j\lambda_l(\mathbf{J}) & 1 & & \\ & & \ddots & & \\ & & & w_j\lambda_l(\mathbf{J}) & 1 \\ & & & & w_j\lambda_l(\mathbf{J}) \end{bmatrix},$$

and $\tilde{\mathbf{P}}$ *is a diagonal matrix obtained by concatenating* $\tilde{\mathbf{P}}_{j,l} = \operatorname{diag}\left(\{w_j^{-n+1}\}_{n=1}^{k_l}\right)$.

Proof. As $\mathbf{J} = \bigoplus_{l=1}^{m}\mathbf{J}_l$, we can focus on a single Jordan block. Fix $l\in\{1,\dots,m\}$. We have

$$\mathbf{W}\otimes\mathbf{J}_l = \operatorname{diag}\left(\{w_j\,\mathbf{J}_l\}_{j=1}^{K}\right) = \bigoplus_{j=1}^{K}w_j\mathbf{J}_l\,,$$

hence, we can focus once again on a single block. Fix $j\in\{1,\dots,K\}$; the Jordan decomposition of $w_j\mathbf{J}_l$ is given by $\tilde{\mathbf{P}}_l = \operatorname{diag}\left(\{w_j^{-n+1}\}_{n=1}^{k_l}\right)$ and

$$\tilde{\mathbf{J}}_l = \begin{bmatrix} w_j\lambda_l(\mathbf{J}) & 1 & & & \\ & w_j\lambda_l(\mathbf{J}) & 1 & & \\ & & \ddots & \ddots & \\ & & & w_j\lambda_l(\mathbf{J}) & 1 \\ & & & & w_j\lambda_l(\mathbf{J}) \end{bmatrix}.$$

To verify it, compute the (n, m) element

$$\left(\tilde{\mathbf{P}}_l \tilde{\mathbf{J}}_l \tilde{\mathbf{P}}_l^{-1}\right)_{n,m} = \sum_{i,k} \left(\tilde{\mathbf{P}}_l\right)_{n,i} \left(\tilde{\mathbf{J}}_l\right)_{i,k} \left(\tilde{\mathbf{P}}_l^{-1}\right)_{k,m} .$$

Since $\tilde{\mathbf{P}}_l$ is a diagonal matrix, the only non-null entries are on the diagonal; therefore, $i = n$ and $k = m$

$$= \left(\tilde{\mathbf{P}}_l\right)_{n,n} \left(\tilde{\mathbf{J}}_l\right)_{n,m} \left(\tilde{\mathbf{P}}_l^{-1}\right)_{m,m}$$

and the only non-null entries of $\tilde{\mathbf{J}}_l$ are when $m = n$ or $m = n + 1$, hence

$$= \begin{cases} \left(\tilde{\mathbf{P}}_l\right)_{n,n} \left(\tilde{\mathbf{J}}_l\right)_{n,n} \left(\tilde{\mathbf{P}}_l^{-1}\right)_{n,n} = w_j \lambda_l \left(\mathbf{J}\right) , & m = n \, , \\ \left(\tilde{\mathbf{P}}_l\right)_{n,n} \left(\tilde{\mathbf{J}}_l\right)_{n,n+1} \left(\tilde{\mathbf{P}}_l^{-1}\right)_{n+1,n+1} = w_j^{-n+1} w_j^n = w_j \, , & m = n + 1 \, . \end{cases}$$

The thesis follows from assembling the direct sums back. $\qquad\square$

Lemma 7 leads to the following result that fully characterizes the solution of (22) in terms of the generalized eigenvectors and eigenvalues of $\mathbf{M}$ and $\mathbf{W}$.

Proposition 11. *Consider* (22) *with* $\mathbf{M} = \mathbf{PJP}^{-1}$ *and* $\mathbf{W} \otimes \mathbf{J} = \tilde{\mathbf{P}} \tilde{\mathbf{J}} \tilde{\mathbf{P}}^{-1}$, *where* $\tilde{\mathbf{J}}$ *and* $\tilde{\mathbf{P}}$ *are given in Lemma* 7. *The solution of* (22) *is*

$$\text{vec}(\mathbf{x})(t) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \exp\left(\lambda_{l_1}(\mathbf{W}) \lambda_{l_2}(\mathbf{M}) t\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} \left(\lambda_{l_1}(\mathbf{W})\right)^{1-j} \mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{M}) \, ,$$

(23)

where the coefficients c_{l_1,l_2}^i *are given by*

$$\text{vec}(\mathbf{x}_0) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i (\mathbf{I} \otimes \mathbf{P}) \tilde{\mathbf{P}} \mathbf{e}_{l_1} \otimes \mathbf{e}_{l_2}^i \qquad (24)$$

where $\left\{\mathbf{e}_{l_2}^i \ : \ l_2 \in \{1, \ldots, m\} \, , \ i \in \{1, \ldots, k_{l_2}\}\right\}$ *is the standard basis satisfying* $\mathbf{P} \mathbf{e}_{l_2}^i = \psi_{l_2}^i(\mathbf{M})$.

Proof. By Lemma 7, the eigenvalues of $\mathbf{W} \otimes \mathbf{M}$ and the corresponding eigenvectors and generalized eigenvectors are

$$\lambda_{l_1}(\mathbf{W}) \lambda_{l_2}(\mathbf{M}) \, , \ \mathbf{e}_{l_1} \otimes \psi_{l_2}^1(\mathbf{M}) \, , \ \left(\lambda_{l_1}(\mathbf{W})\right)^{-i+1} \mathbf{e}_{l_1} \otimes \psi_{l_2}^i(\mathbf{M})$$

for $l_1 \in \{1, \ldots, K\}$, $l_2 \in \{1, \ldots, m\}$ and $i \in \{2, \ldots, k_l\}$. Hence, by Lemma 6, the solution of (22) is given by

$$\text{vec}(\mathbf{x})(t) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \exp\left(\lambda_{l_2}(\mathbf{M}) \lambda_{l_1}(\mathbf{W}) t\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} \left(\lambda_{l_1}(\mathbf{W})\right)^{1-j} \left(\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})\right) \, ,$$

where the coefficients c_{l_1,l_2}^i are given by

$$\text{vec}(\mathbf{x}_0) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i (\mathbf{I} \otimes \mathbf{P}) \tilde{\mathbf{P}} (\mathbf{e}_{l_1} \otimes \mathbf{e}_{l_2}^i(\mathbf{M})) \,.$$

$\square$

E.1   Proof of Theorem 3

In the following, we reformulate and prove Theorem 3.

Corollary 2. *Let $\mathcal{G}$ be a strongly connected directed graph with SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. Consider the initial value problem in (4) with diagonal channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ and $\alpha = 1$. Then, for almost all initial values $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$, the solution to (4) is HFD if*

$$\lambda_K(\mathbf{W}) \Re \lambda_1(\mathbf{L}) < \lambda_1(\mathbf{W}) \lambda_N(\mathbf{L})$$

and $\lambda_1(\mathbf{L})$ is the unique eigenvalue that minimizes the real part among all eigenvalues of $\mathbf{L}$. Otherwise, the solution is LFD.

Proof. Using the notation from Proposition 11 and its proof, we can write the solution of the vectorized form of (4) as

$$\text{vec}(\mathbf{x})(t) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \exp\left(-\lambda_{l_1}(\mathbf{W}) \lambda_{l_2}(\mathbf{L}) t\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})).$$

As done extensively, we separate the terms corresponding to the frequency with minimal real part. This frequency dominates as the exponential converges faster than polynomials for t going to infinity. Consider the case $\lambda_K(\mathbf{W}) \Re(\lambda_1(\mathbf{L})) < \lambda_1(\mathbf{W}) \Re(\lambda_N(\mathbf{L}))$. As $\lambda_1(\mathbf{L})$ is unique, the product $\lambda_K(\mathbf{W}) \Re(\lambda_1(\mathbf{L}))$ is the unique most negative frequency. Assume without loss of generality that $\lambda_K(\mathbf{W})$ has multiplicity one. The argument does not change for higher multiplicities as the corresponding eigenvectors are orthogonal since $\mathbf{W}$ is diagonal. Then, $\lambda_K(\mathbf{W}) \lambda_1(\mathbf{L})$ has multiplicity one, and we calculate $\text{vec}(\mathbf{x})(t)$ as

$$\sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \exp\left(-\lambda_{l_1}(\mathbf{W}) \lambda_{l_2}(\mathbf{L}) t\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L}))$$

$$= c_{K,1}^{k_1} \exp\left(-t \lambda_K(\mathbf{W}) \lambda_1(\mathbf{L})\right) \frac{t^{k_1-1}}{(k_1-1)!} (\mathbf{e}_K \otimes \psi_1^1(\mathbf{L}))$$

$$+ c_{K,1}^{k_1} \exp\left(-t \lambda_K(\mathbf{W}) \lambda_1(\mathbf{L})\right) \sum_{j=2}^{k_1} \frac{t^{k_1-j}}{(k_1-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \exp\left(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})\right) \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} (\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=2}^{m} \exp\left(-\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L})t\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j}(\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L}))$$

$$= \exp\left(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})\right) t^{k_1-1} \Bigg(c_{K,1}^{k_1} \frac{1}{(k_1-1)!}(\mathbf{e}_K \otimes \psi_1^1(\mathbf{L}))$$

$$+ c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!}(\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^{i} \frac{1}{(i-j)!} t^{i-j-k_1+1}(\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=2}^{m} \exp\left(-t(\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j-k_1+1}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j}(\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \Bigg).$$

We can then write the normalized solution as

$$\Bigg(\frac{c_{K,1}^{k_1}}{(k_1-1)!}(\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) + c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!}(\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^{i} \frac{t^{i-j-k_1+1}}{(i-j)!}(\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=2}^{m} e^{-t(\lambda_K(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))} \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j-k_1}}{(i-j)!}(\lambda_{l_1}(\mathbf{W}))^{1-j}(\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \Bigg)$$

$$\cdot \Bigg\| \frac{c_{K,1}^{k_1}}{(k_1-1)!}\left(\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})\right) + c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!}(\lambda_K(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^{i} \frac{t^{i-j-k_1+1}}{(i-j)!}(\lambda_{l_1}(\mathbf{W}))^{1-j}(\mathbf{e}_K \otimes \psi_1^j(\mathbf{L}))$$

$$+ \sum_{l_1=1}^{K} \sum_{l_2=2}^{m} \exp\left(-t(\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))\right) \sum_{i=1}^{k_{l_2}} c_{l_1,l_2}^i \sum_{j=1}^{i} \frac{t^{i-j-k_1}}{(i-j)!}(\lambda_{l_1}(\mathbf{W}))^{1-j}(\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \Bigg\|_2^{-1}.$$

All summands, except the first, converge to zero for t going to infinity. Hence,

$$\frac{\mathrm{vec}(\mathbf{x})(t)}{\|\mathrm{vec}(\mathbf{x})(t)\|_2} \xrightarrow{t \to \infty} \Bigg\| \frac{c_{K,1}^{k_1}}{(k_1-1)!}(\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) \Bigg\|_2^{-1} \left(\frac{c_{K,1}^{k_1}}{(k_1-1)!}(\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) \right).$$

We apply Lemma 3 to finish the proof for the HFD case. Note that $\psi_1^1(\mathbf{L})$ is an eigenvector corresponding to $\lambda_1(\mathbf{L})$. The LFD case is equivalent. By Perron

Frobenius for irreducible non-negative matrices, there is no other eigenvalue with the same real part as $1 - \lambda_N(\mathbf{L}) = \lambda_1(\mathbf{\Delta})$. □

Remark 3. If the hypotheses are met, the convergence result also holds for $\mathbf{L}^\alpha$. With the same reasoning, we can prove that the normalized solution converges to the eigenvector corresponding to the eigenvalue of $\mathbf{L}^\alpha$ with minimal real part. It suffices to consider the eigenvalues and generalized eigenvectors of $\mathbf{L}^\alpha$. However, we do not know the relationship between the singular values of $\mathbf{L}^\alpha$, where we defined the fractional Laplacian, and the eigenvalues of $\mathbf{L}$. Hence, it is much more difficult to draw conclusions on the Dirichlet energy.

E.2   Explicit Euler

In this subsection, we show that the convergence properties of the Dirichlet energy from Corollary 2 are also satisfied when (4) is approximated via an explicit Euler scheme.

As noted in Appendix D.2, the vectorized solution to (4) can be written as

$$\mathrm{vec}(\mathbf{x})(n\,h) = \left(\mathbf{I} - h\left(\mathbf{W} \otimes \mathbf{L}\right)\right)^n \mathrm{vec}(\mathbf{x}_0),$$

when $\alpha = 1$. We thus aim to analyze the Jordan decomposition of $\mathbf{L}^n$ for $\mathbf{L} \in \mathbb{C}^{n \times n}$ and $n \in \mathbb{N}$. Let $\mathbf{L} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$, where $\mathbf{J}$ is the Jordan form, and $\mathbf{P}$ is a invertible matrix of generalized eigenvectors.

Consider a Jordan block $\mathbf{J}_i$ associated with the eigenvalue $\lambda_i(\mathbf{M})$. For a positive integer n, the n-th power of the Jordan block can be computed as:

$$\mathbf{J}_l^n = \lambda_l(\mathbf{L})^n \begin{bmatrix} 1 & \binom{n}{1}\lambda_l(\mathbf{L})^{-1} & \binom{n}{2}\lambda_l(\mathbf{L})^{-2} & \cdots & \binom{n}{k_l-1}\lambda_l(\mathbf{L})^{-k_l+1} \\ & 1 & \binom{n}{1}\lambda_l(\mathbf{L})^{-1} & & \binom{n}{k_l-2}\lambda_l(\mathbf{L})^{-k_l+2} \\ & & 1 & & \vdots \\ & & & \ddots & \binom{n}{1}\lambda_l(\mathbf{L})^{-1} \\ & & & & 1 \end{bmatrix}$$

We compute the n-th power of $\mathbf{L}$ as $\mathbf{L}^n = (\mathbf{P}\mathbf{J}\mathbf{P}^{-1})^n = \mathbf{P}\mathbf{J}^n\mathbf{P}^{-1}$, and we expand $\mathbf{x}_0$ as

$$\mathbf{x}_0 = \sum_{l=1}^{m} \sum_{i=1}^{k_l} c_l^i \mathbf{P} \mathbf{e}_l^i,$$

where $\left\{ \mathbf{e}_l^i : i \in \{1, \ldots k_l\}, l \in \{1, \ldots, m\} \right\}$ is the standard basis and $\mathbf{P}\mathbf{e}_l^i = \psi_l^i(\mathbf{L})$ are the generalized eigenvectors of $\mathbf{L}$. It is easy to see that

$$\mathbf{L}^n \mathbf{x}_0 = \mathbf{P}\mathbf{J}^n\mathbf{P}^{-1} \left(\sum_{l=1}^{m} \sum_{i=1}^{k_l} c_l^i \mathbf{P} \mathbf{e}_l^i \right) = \mathbf{P}\mathbf{J}^n \left(\sum_{l=1}^{m} \sum_{i=1}^{k_l} c_l^i \mathbf{e}_l^i \right).$$

As $\mathbf{J}^n = \bigoplus_{l=1}^m \mathbf{J}_l^n$, we can focus on a single Jordan block. Fix $l \in \{1, \ldots, m\}$, and compute

$$
\mathbf{P}\mathbf{J}_l^n \left(\sum_{i=1}^{k_l} c_l^i \mathbf{e}_l^i \right) = \mathbf{P} \left(\lambda_l(\mathbf{M})^n c_l^1 \mathbf{e}_l^1 \right) + \mathbf{P} \left(\binom{n}{1} \lambda_l(\mathbf{M})^{n-1} c_l^1 \mathbf{e}_l^1 + \lambda_l(\mathbf{M})^n c_l^2 \mathbf{e}_l^2 \right)
$$

$$
+ \mathbf{P} \left(\binom{n}{2} \lambda_l(\mathbf{M})^{n-2} c_l^1 \mathbf{e}_l^1 + \binom{n}{1} \lambda_l(\mathbf{L})^{n-1} c_l^2 \mathbf{e}_l^2 + \lambda_l(\mathbf{M})^n c_l^3 \mathbf{e}_l^3 \right)
$$

$$
+ \ldots.
$$

We can summarize our findings in the following lemma.

Lemma 8. *For any* $\mathbf{L} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1} \in \mathbb{R}^{N \times N}$ *and* $\mathbf{x}_0 = \sum_{l=1}^m \sum_{i=1}^{k_l} c_l^i \psi_l^i(\mathbf{L})$, *we have*

$$
\mathbf{L}^n \mathbf{x}_0 = \sum_{l=1}^m \sum_{i=1}^{\min\{k_l, n-1\}} \sum_{j=1}^i \binom{n}{i-j} \lambda_l(\mathbf{L})^{n-i+j} c_l^j \psi_l^j(\mathbf{L}).
$$

We proceed with the main result of this subsection.

Proposition 12. *Let* $\mathcal{G}$ *be a strongly connected directed graph with* SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. *Consider the initial value problem in* (4) *with diagonal channel mixing matrix* $\mathbf{W} \in \mathbb{R}^{K \times K}$ *and* $\alpha = 1$. *Approximate the solution to* (4) *with an explicit Euler scheme with a sufficiently small step size h. Then, for almost all initial values* $\mathbf{x}_0 \in \mathbb{C}^{N \times K}$ *the following holds. If* $\lambda_1(\mathbf{L})$ *is unique and*

$$
\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L}) < \lambda_1(\mathbf{W})\Re\lambda_N(\mathbf{L}), \tag{25}
$$

the approximated solution is HFD. *Otherwise, the solution is* LFD.

Proof. As noted in Appendix D.2, the vectorized solution to (4) with $\alpha = 1$, can be written as
$$
\mathrm{vec}(\mathbf{x})(n\,h) = (\mathbf{I} - h\,(\mathbf{W} \otimes \mathbf{L}))^n \mathrm{vec}(\mathbf{x}_0).
$$

Consider the Jordan decomposition of $\mathbf{L} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ and the Jordan decomposition of $\mathbf{W} \otimes \mathbf{J} = \tilde{\mathbf{P}}\tilde{\mathbf{J}}\tilde{\mathbf{P}}^{-1}$, where $\tilde{\mathbf{J}}$ and $\tilde{\mathbf{P}}$ are specified in Lemma 7. Then,

$$
\begin{aligned}
\mathrm{vec}(\mathbf{x})(n\,h) &= \left(\mathbf{I} + h\mathbf{W} \otimes (\mathbf{P}\mathbf{J}\mathbf{P}^{-1})\right)^n \mathrm{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})(\mathbf{I} - h\mathbf{W} \otimes \mathbf{J})^n (\mathbf{I} \otimes \mathbf{P})^{-1} \mathrm{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})(\mathbf{I} - h\tilde{\mathbf{P}}\tilde{\mathbf{J}}\tilde{\mathbf{P}}^{-1})^n (\mathbf{I} \otimes \mathbf{P})^{-1} \mathrm{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}}(\mathbf{I} - h\tilde{\mathbf{J}})^n \tilde{\mathbf{P}}^{-1}(\mathbf{I} \otimes \mathbf{P})^{-1} \mathrm{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}}(\mathbf{I} - h\tilde{\mathbf{J}})^n ((\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}})^{-1}\mathrm{vec}(\mathbf{x}_0).
\end{aligned}
$$

Now, decompose $\mathbf{x}_0$ into the basis of generalized eigenvectors, i.e.,

$$
\mathbf{x}_0 = \sum_{l_1=1}^K \sum_{l_2=1}^m \sum_{i=1}^{k_{l_2}} c_l^i ((\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}})(\mathbf{e}_{l_1} \otimes \mathbf{e}_{l_2}^i(\mathbf{L})).
$$

Then, by Lemma 8, we have

$$\text{vec}(\mathbf{x})(n\,h) = \sum_{l_1=1}^{K}\sum_{l_2=1}^{m}\sum_{i=1}^{\min\{k_{l_2},t-1\}}\sum_{j=1}^{i}\binom{n}{i-j}(1-h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{n-i+j}\,c_{l_1,l_2}^{j}$$
$$(\lambda_{l_1}(\mathbf{W}))^{1-j}\,\psi_{l_1}(\mathbf{W})\otimes\psi_{l_2}^{j}(\mathbf{L}).$$

Now, consider the maximal frequency, i.e.,

$$L_1, L_2 = \arg\max_{l_1,l_2}\left\{|1-h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L})|\right\}.$$

Then, the solution $\text{vec}(\mathbf{x})(n\,h)$ can be written as

$$\sum_{l_1=1}^{K}\sum_{l_2=1}^{m}\sum_{i=1}^{\min\{k_{l_2},n-1\}}\sum_{j=1}^{i}\binom{n}{i-j}(1-h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{n-i+j}\,c_{l_1,l_2}^{j}\psi_{l_1}(\mathbf{W})\otimes\psi_{l_2}^{j}(\mathbf{L})$$
$$= (1-h\lambda_{L_1}(\mathbf{W})\lambda_{L_2}(\mathbf{L}))^{n}$$
$$\cdot\sum_{l_1=1}^{K}\sum_{l_2=1}^{m}\sum_{i=1}^{\min\{k_{l_2},t-1\}}\sum_{j=1}^{i}\binom{n}{i-j}\frac{(1-h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{n-i+j}}{(1-h\lambda_{L_1}(\mathbf{W})\lambda_{L_2}(\mathbf{L}))^{n}}\,c_{l_1,l_2}^{j}\psi_{l_1}(\mathbf{W})\otimes\psi_{l_2}^{j}(\mathbf{L}).$$
$$(26)$$

With a similar argument as in the proof of Theorem 3, we can then see that

$$\frac{\text{vec}(\mathbf{x})(n\,h)}{\|\text{vec}(\mathbf{x})(n\,h)\|_2}\xrightarrow{t\to\infty}\frac{c_{L_1,L_2}^{1}\,\psi_{L_1}(\mathbf{W})\otimes\psi_{L_2}^{1}(\mathbf{L})}{\left\|c_{L_1,L_2}^{1}\,\psi_{L_1}(\mathbf{W})\otimes\psi_{L_2}^{1}(\mathbf{L})\right\|_2},$$

where $\psi_{L_2}^{1}(\mathbf{L})$ is the eigenvector corresponding to $\lambda_{L_2}(\mathbf{L})$. Note that for almost all $\mathbf{x}$, we have $c_{L_1,L_2}^{1}\neq 0$. Then $\psi_{L_2}^{1}(\mathbf{L})$ is also an eigenvector of $\mathbf{\Delta}$ corresponding to the eigenvalue $1-\lambda_{L_2}(\mathbf{L})$. By Lemma 3, we have that the approximated solution is $(1-\lambda_{L_2}(\mathbf{L}))$-FD.

We finish the proof by showing that $L_2 = 1$ if (25) is satisfied, and $L_2 = N$ otherwise. First, note that either $\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L})$ or $\lambda_1(\mathbf{W})\Re\lambda_N(\mathbf{L})$ are the most negative real parts among all $\{\lambda_l(\mathbf{W})\Re\lambda_r(\mathbf{L})\}_{l\in\{1,\dots,K\},r\in\{1\dots,N\}}$. Assume first that $\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L})$ has the most negative real part, i.e., (25) holds. Then, define

$$\varepsilon := \max_{l,r}|\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L}) - \lambda_l(\mathbf{W})\Re\lambda_r(\mathbf{L})|\,,$$

and assume $h < \varepsilon\|\mathbf{W}\|^2$. Now it is easy to see that

$$2\lambda_K(\mathbf{W})\Re\lambda_1(\mathbf{L}) - h\lambda_K(\mathbf{W})^2|\lambda_1(\mathbf{L})|^2 < 2\lambda_l(\mathbf{W})\Re\lambda_r(\mathbf{L}) - h\lambda_l(\mathbf{W})^2|\lambda_r(\mathbf{L})|^2,$$

which is equivalent to $(K,1) = (L_1, L_2)$. Hence, the dynamics are $(1-\lambda_1(\mathbf{L}))$-FD. As $(1-\lambda_1(\mathbf{L}))$ is highest frequency of $\mathbf{\Delta}$, we get HFD dynamics. Similary, we can show that if $\lambda_1(\mathbf{W})\Re\lambda_N(\mathbf{L})$ is the most negative frequency, we get LFD dynamics. Note that for the HFD argument, we must assume that $\lambda_1(\mathbf{L})$ is the unique eigenvalue with the smallest real part. For the LFD argument, it is already given that $\lambda_N(\mathbf{L})$ has multiplicity one by Perron-Frobenius Theorem. $\qquad\square$

E.3 GCN oversmooths

Proposition 13. *Let $\mathcal{G}$ be a strongly connected and aperiodic directed graph with SNA $\mathbf{L} \in \mathbb{R}^{N \times N}$. A GCN with the update rule*

$$\mathbf{x}_{t+1} = \mathbf{L}\mathbf{x}_t\mathbf{W},$$

where $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$ are the input node features, always oversmooths.

Proof. The proof follows similarly to the proof of Proposition 12. The difference is that instead of (26), we can write the node features after t layers as

$$\text{vec}(\mathbf{x}_t) = \sum_{l_1=1}^{K} \sum_{l_2=1}^{m} \sum_{i=1}^{\min\{k_{l_2}, t-1\}} \sum_{j=1}^{i} \binom{t}{i-j} (\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{t-i+j} c_{l_1,l_2}^{j} \psi_{l_1}(\mathbf{W}) \otimes \psi_{l_2}^{j}(\mathbf{L}).$$

Now note that by Perron-Frobenius the eigenvalue $\lambda_N(\mathbf{L})$ with the largest absolute value is real and has multiplicity one. Then, $\max_{l_1,l_2} |\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L})|$ is attained at either $\lambda_1(\mathbf{W})\lambda_N(\mathbf{L})$ or $\lambda_K(\mathbf{W})\lambda_N(\mathbf{L})$. Equivalently to the proof of Proposition 12, we can show that the corresponding GCN is $1 - \lambda_N(\mathbf{L})$-FD. Now $1 - \lambda_N(\mathbf{L}) = \lambda_1(\boldsymbol{\Delta})$, and $\lambda_1(\boldsymbol{\Delta})$-FD corresponds to LFD, hence the GCN oversmooths.   □

F   Appendix for Section 6

In this section, we give the details on the numerical results in Section 6. We begin by describing the exact model.

Model architecture. Let $\mathcal{G}$ be a directed graphs and $\mathbf{x}_0 \in \mathbb{R}^{N \times d}$ the node features. Our architecture first embeds the input node features $\mathbf{x}$ via a multi-layer perceptron (MLP). We then evolve the features $\mathbf{x}_0$ according to a slightly modified version of (5), i.e, $\mathbf{x}'(t) = -i\,\mathbf{L}^\alpha\mathbf{x}(t)\mathbf{W}$ for some time $t \in [0, T]$. In our experiments, we approximate the solution with an Explicit Euler scheme with step size $h > 0$. This leads to the following update rule

$$\mathbf{x}_{n+1} = \mathbf{x}_n - ih\mathbf{L}^\alpha\mathbf{x}_n\mathbf{W}\,.$$

The channel mixing matrix is a diagonal learnable matrix $\mathbf{W} \in \mathbb{C}^{K \times K}$, and $\alpha \in \mathbb{R}$, $h \in \mathbb{C}$ are also learnable parameters. The features at the last time step $\mathbf{x}(T)$ are then fed into a second MLP, whose output is used as the final output. Both MLPs use LeakyReLU as non-linearity and dropout (Srivastava et al., 2014). On the contrary, the graph layers do not use any dropout nor non-linearity. A sketch of the algorithm is reported in fLode.

Complexity. The computation of the SVD is $\mathcal{O}(N^3)$. However, one can compute only the first $p \ll N$ singular values: this cuts down the cost to $\mathcal{O}(N^2 p)$. The memory required to store the singular vectors is $\mathcal{O}(N^2)$, since they are not sparse in general. Each training step has a cost of $\mathcal{O}(N^2 K)$.

Algorithm 1: fLode

```
% A, x₀ are given.
% Preprocessing
```
1   $\mathbf{D}_{\text{in}}$ $= \text{diag}(\mathbf{A}\mathbf{1})$
2   $\mathbf{D}_{\text{out}}$ $= \text{diag}(\mathbf{A}^{\mathsf{T}}\mathbf{1})$
3   $\mathbf{L}$ $= \mathbf{D}_{\text{in}}^{-1/2}\mathbf{A}\mathbf{D}_{\text{out}}^{-1/2}$
4   $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^{\mathsf{H}}$ $= \texttt{svd}(\mathbf{L})$
```
% The core of the algorithm is very simple
```
5   `def training_step(`$\mathbf{x}_0$`):`
6     $\mathbf{x}_0 = \texttt{input\_MLP}(\mathbf{x}_0)$
7   `for` $n \in \{1, \ldots, N\}$ `do`
8       $\mathbf{x}_n = \mathbf{x}_{n-1} - i\,h\,\mathbf{U}\mathbf{\Sigma}^{\alpha}\mathbf{V}^{\mathsf{H}}\mathbf{x}_{n-1}\mathbf{W}$
9     $\mathbf{x}_N = \texttt{output\_MLP}(\mathbf{x}_N)$
10   `return` $\mathbf{x}_N$

Experimental details. Our model is implemented in `PyTorch` (Paszke et al., 2019), using `PyTorch geometric` (Fey et al., 2019). The computation of the SVD for the fractional Laplacian is implemented using the library `linalg` provided by `PyTorch`. In case of truncated SVD, we use the function `randomized_svd` provided by the library `extmath` from `sklearn`. The code and instructions to reproduce the experiments are available on `GitHub`. Hyperparameters were tuned using grid search. All experiments were run on an internal cluster with `NVIDIA GeForce RTX 2080 Ti` and `NVIDIA TITAN RTX` GPUs with 16 and 24 GB of memory, respectively.

Training details. All models were trained for 1000 epochs using Adam as optimizer (Kingma et al., 2017) with a fixed learning rate. We perform early stopping if the validation accuracy does not increase for 100 epochs.

F.1   Details on Experiments on Real-World Undirected Graphs

We conducted 10 repetitions using data splits obtained from (Pei et al., 2019). For each split, 48% of the nodes are used for training, 32% for validation and 20% for testing. In all datasets, we considered the largest connected component (LCC). Chameleon, Squirrel, and Film are directed graphs; hence, we converted them to undirected. Cora, Citeseer, and Pubmed are already undirected graphs: to these, we added self-loops. We normalized the input node features for all graphs.

As baseline models, we considered the same models as in (Di Giovanni et al., 2022). The results were provided by (Pei et al., 2019) and include standard GNNs, such as GAT (Velickovic et al., 2018), GCN (Kipf et al., 2017), and GraphSAGE (Hamilton et al., 2017). We also included models designed to address oversmoothing and heterophilic graphs, such as PairNorm (Zhao et al., 2019), GGCN (Zhu et al., 2020), Geom-GCN (Yan et al., 2022), H2GCN (Bo et al., 2021), GPRGNN (Chien et al., 2021), and Sheaf (Bodnar et al., 2022). Furthermore,

we included the graph neural ODE-based approaches, CGNN (Xhonneux et al., 2020) and GRAND (Chamberlain et al., 2021), as in (Di Giovanni et al., 2022), and the model GRAFF from (Di Giovanni et al., 2022) itself.

The exact hyperparameters for FLODE, are provided in Appendix F.1.

Table 5: Selected hyperparameters, learned exponent, learned step size, and the Dirichlet energy in the last layer for real-world datasets.

(a) Undirected.

	Dataset					
	Film	Squirrel	Chameleon	Citeseer	Pubmed	Cora
learning rate	10^{-3}	$2.5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	10^{-2}	10^{-2}	10^{-2}
weight decay	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	10^{-3}	$5 \cdot 10^{-3}$	10^{-3}	$5 \cdot 10^{-3}$
hidden channels	256	64	64	64	64	64
num. layers	1	6	4	2	3	2
encoder layers	3	1	1	1	3	1
decoder layers	2	2	2	1	1	2
input dropout	0	$1.5 \cdot 10^{-1}$	0	0	$5 \cdot 10^{-2}$	0
decoder dropout	10^{-1}	10^{-1}	0	0	10^{-1}	0
exponent	1.001 ± 0.003	0.17 ± 0.03	0.35 ± 0.15	0.92 ± 0.03	0.82 ± 0.07	0.90 ± 0.02
step size	0.991 ± 0.002	1.08 ± 0.01	1.22 ± 0.03	1.04 ± 0.02	1.12 ± 0.02	1.06 ± 0.01
dirichlet energy	0.246 ± 0.006	0.40 ± 0.02	0.13 ± 0.03	0.021 ± 0.001	0.015 ± 0.001	0.0227 ± 0.0006

(b) Directed.

	Dataset		
	Film	Squirrel	Chameleon
learning rate	10^{-3}	$2.5 \cdot 10^{-3}$	10^{-2}
weight decay	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	10^{-3}
hidden channels	256	64	64
num. layers	1	6	5
encoder layers	3	1	1
decoder layers	2	2	2
input dropout	0	10^{-1}	0
decoder dropout	0.1	10^{-1}	0
exponent	1.001 ± 0.005	0.28 ± 0.06	0.30 ± 0.11
step size	0.990 ± 0.002	1.22 ± 0.02	1.22 ± 0.05
dirichlet energy	0.316 ± 0.005	0.38 ± 0.02	0.27 ± 0.04

F.2   Details on Experiments on Synthetic Directed Graphs

The dataset and code are taken from (Zhang et al., 2021). As baseline models, we considered the ones in (Zhang et al., 2021) for which we report the corresponding results. The baseline models include standard GNNs, such as ChebNet (Defferrard et al., 2016), GCN (Kipf et al., 2017), GraphSAGE (Hamilton et al., 2017), APPNP (Gasteiger et al., 2018), GIN (Xu et al., 2018), GAT (Velickovic et al., 2018), but also models specifically designed for directed graphs, such as DGCN (Tong, Liang, Sun, Rosenblum, et al., 2020), DiGraph, DiGraphIB (Tong, Liang, Sun, X. Li, et al., 2020), MagNet (Zhang et al., 2021)).

The DSBM dataset. The directed stochastic block model (DSBM) is described in detail in (Zhang et al., 2021, Section 5.1.1). To be self-contained, we include a short explanation.

The DSBM model is defined as follows. There are N vertices, which are divided into n_c clusters $(C_1, C_2, ...C_{n_c})$, each having an equal number of vertices. An interaction is defined between any two distinct vertices, u and v, based on two sets of probabilities: $\{\alpha_{i,j}\}_{i,j=1}^{n_c}$ and $\{\beta_{i,j}\}_{i,j=1}^{n_c}$.

The set of probabilities $\{\alpha_{i,j}\}$ is used to create an undirected edge between any two vertices u and v, where u belongs to cluster C_i and v belongs to cluster C_j. The key property of this probability set is that $\alpha_{i,j} = \alpha_{j,i}$, which means the chance of forming an edge between two clusters is the same in either direction.

The set of probabilities $\{\beta_{i,j}\}$ is used to assign a direction to the undirected edges. For all $i, j \{1, \ldots, n_c\}$, we assume that $\beta_{i,j} + \beta_{j,i} = 1$ holds. Then, to the undirected edge (u, v) is assigned the direction from u to v with probability $\beta_{i,j}$ if u belongs to cluster C_i and v belongs to cluster C_j, and the direction from v to u with probability $\beta_{j,i}$.

The primary objective here is to classify the vertices based on their respective clusters.

There are several scenarios designed to test different aspects of the baseline models and our model. In the experiments, the total number of nodes is fixed at $N = 2500$ and the number of clusters is fixed at $n_c = 5$. In all experiments, the training set contains 20 nodes per cluster, 500 nodes for validation, and the rest for testing. The results are averaged over 5 different seeds and splits.

Following Zhang et al. (2021), we train our model in both experiments for 3000 epochs and use early-stopping if the validation accuracy does not increase for 500 epochs. We select the best model based on the validation accuracy after sweeping over a few hyperparameters. We give exact numerical values for the experiments with the standard error in Table 6a and refer to Appendix F.2 for the chosen hyperparameters.

DSBM with varying edge density. In the first experiment, the model is evaluated based on its performance on the DSBM with varying $\alpha_{i,j} = \alpha^*$, $\alpha^* \in \{0.1, 0.08, 0.05\}$ for $i \neq j$, which essentially changes the density of edges between different clusters. The other probabilities are fixed at $\alpha_{i,i} = 0.5$, $\beta_{i,i} = 0.5$ and $\beta_{i,j} = 0.05$ for $i > j$. The results are shown in Figure 2 with exact numerical values in Table 6a.

DSBM with varying net flow. In the other scenario, the model is tested on how it performs when the net flow from one cluster to another varies. This is achieved by keeping $\alpha_{i,j} = 0.1$ constant for all i and j, and allowing $\beta_{i,j}$ to vary from 0.05 to 0.4. The other probabilities are fixed at $\alpha_{i,i} = 0.5$ and $\beta_{i,i} = 0.5$. The results are shown in Figure 2 with exact numerical values in Table 6b.

Table 6: Node classification accuracy of ordered DSBM graphs: the top three models are coloured as first, second, third.

(a) Varying edge density.

	α^*		
	0.1	0.08	0.05
ChebNet	19.9 ± 0.6	20.0 ± 0.7	20.0 ± 0.7
GCN-D	68.9 ± 2.1	67.6 ± 2.7	58.5 ± 2.0
APPNP-D	97.7 ± 1.7	95.9 ± 2.2	90.3 ± 2.4
SAGE-D	20.1 ± 1.1	19.9 ± 0.8	19.9 ± 1.0
GIN-D	57.3 ± 5.8	55.4 ± 5.5	50.9 ± 7.7
GAT-D	42.1 ± 5.3	39.0 ± 7.0	37.2 ± 5.5
DGCN	84.9 ± 7.2	81.2 ± 8.2	64.4 ± 12.4
DiGraph	82.1 ± 1.7	77.7 ± 1.6	66.1 ± 2.4
DiGraphIB	99.2 ± 0.5	97.7 ± 0.7	89.3 ± 1.7
MagNet	99.6 ± 0.2	98.3 ± 0.8	94.1 ± 1.2
Ours	99.3 ± 0.1	98.8 ± 0.1	97.5 ± 0.1

(b) Varying net flow.

	β^*							
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
ChebNet	19.9 ± 0.7	20.1 ± 0.6	20.0 ± 0.6	20.1 ± 0.8	19.9 ± 0.9	20.0 ± 0.5	19.7 ± 0.9	20.0 ± 0.5
GCN-D	68.6 ± 2.2	74.1 ± 1.8	75.5 ± 1.3	74.9 ± 1.3	72.0 ± 1.4	65.4 ± 1.6	58.1 ± 2.4	45.6 ± 4.7
APPNP-D	97.4 ± 1.8	94.3 ± 2.4	89.4 ± 3.6	79.8 ± 9.0	69.4 ± 3.9	59.6 ± 4.9	51.8 ± 4.5	39.4 ± 5.3
SAGE-D	20.2 ± 1.2	20.0 ± 1.0	20.0 ± 0.8	20.0 ± 0.7	19.6 ± 0.9	19.8 ± 0.7	19.9 ± 0.9	19.9 ± 0.8
GIN-D	57.9 ± 6.3	48.0 ± 11.4	32.7 ± 12.9	26.5 ± 10.0	23.8 ± 6.0	20.6 ± 3.0	20.5 ± 2.8	19.8 ± 0.5
GAT-D	42.0 ± 4.8	32.7 ± 5.1	25.6 ± 3.8	19.9 ± 1.4	20.0 ± 1.0	19.8 ± 0.8	19.6 ± 0.2	19.5 ± 0.2
DGCN	81.4 ± 1.1	84.7 ± 0.7	85.5 ± 1.0	86.2 ± 0.8	84.2 ± 1.1	78.4 ± 1.3	69.6 ± 1.5	54.3 ± 1.5
DiGraph	82.5 ± 1.4	82.9 ± 1.9	81.9 ± 1.1	79.7 ± 1.3	73.5 ± 1.9	67.4 ± 2.8	57.8 ± 1.6	43.0 ± 7.1
DiGraphIB	99.2 ± 0.4	97.9 ± 0.6	94.1 ± 1.7	88.7 ± 2.0	82.3 ± 2.7	70.0 ± 2.2	57.8 ± 6.4	41.0 ± 9.0
MagNet	99.6 ± 0.2	99.0 ± 1.0	97.5 ± 0.8	94.2 ± 1.6	88.7 ± 1.9	79.4 ± 2.9	68.8 ± 2.4	51.8 ± 3.1
Ours	99.3 ± 0.1	98.5 ± 0.1	96.7 ± 0.2	92.8 ± 0.1	87.2 ± 0.3	77.1 ± 0.5	63.8 ± 0.3	50.1 ± 0.5

Table 7: Selected hyperparameters for DSBM dataset.

(a) Varying edge density.

	α^*		
	0.1	0.08	0.05
learning rate	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$
decay	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-4}$
input dropout	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1}$	$1 \cdot 10^{-1}$
decoder dropout	$1 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$1 \cdot 10^{-1}$
hidden channels	256	256	256

(b) Varying net flow.

	β^*							
	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
learning rate	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
decay	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$
input dropout	$1 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$2 \cdot 10^{-1}$
decoder dropout	$1 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$5 \cdot 10^{-2}$	$5 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1}$	$1 \cdot 10^{-1}$
hidden channels	256	256	256	256	256	256	256	256

F.3   Ablation Study

We perform ablation studies on Chameleon (directed, heterophilic) and Citeseer (undirected, homophilic). For this, we sweep over different model options using the same hyperparameters via grid search. The test accuracy corresponding to the hyperparameters that yielded maximum validation accuracy is reported in Table Table 8."

Whenever a box in the table is marked green, it means that the specific model property was used. The keyword ODE describes whether the node features were updated following (4) or (5). The keyword GCN describes whether the node features were updated following $\mathbf{x}_{t+1} = \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$ for real, and $\mathbf{x}_{t+1} = i\,\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$ for complex, which correspond to a vanilla GCN implementing the fractional Laplacian.

The ablation study on Chameleon demonstrates that all the components of the model (learnable exponent, ODE framework with the Schrödinger equation, and directionality via the SNA) contribute to the performance of FLODE. The fact that performance drops when any of these components are not used suggests

that they all play crucial roles in the model's ability to capture the structure and evolution of heterophilic graphs. It is important to note that the performance appears to be more dependent on the adjustable fraction in the FGL than on the use of the ODE framework, illustrating that the fractional Laplacian alone can effectively capture long-range dependencies. However, when the ODE framework is additionally employed, a noticeable decrease in variance is observed.

Table 8: Ablation study on node classification task.

(a) Chameleon (directed, heterophilic).

α learnable	α 1	ODE Heat	ODE Schrödinger	GCN Real	GCN Complex	Directed	Test Accuracy	Energy
✓			✓			✓	77.79 ± 1.42	0.213 (t=5)
	✓		✓			✓	75.72 ± 1.13	0.169 (t=6)
✓				✓		✓	77.35 ± 2.22	0.177 (t=4)
	✓			✓		✓	69.61 ± 1.59	0.178 (t=4)
✓			✓				73.60 ± 1.68	0.131 (t=4)
	✓		✓				70.15 ± 0.86	0.035 (t=4)
✓				✓			71.25 ± 3.04	0.118 (t=4)
	✓			✓			67.19 ± 2.49	0.040 (t=4)
✓		✓				✓	77.33 ± 1.47	0.378 (t=6)
	✓	✓				✓	73.55 ± 0.94	0.165 (t=6)
✓				✓		✓	74.12 ± 3.60	0.182 (t=4)
	✓			✓		✓	68.47 ± 2.77	0.208 (t=4)

(b) Citeseer (undirected, homphilic).

α learnable	α 1	ODE Heat	ODE Schrödinger	GCN Real	GCN Complex	Test Accuracy	Energy
✓			✓			78.07 ± 1.62	0.021 (t=5)
	✓		✓			77.97 ± 2.29	0.019 (t=4)
✓					✓	77.27 ± 2.10	0.011 (t=6)
	✓				✓	77.97 ± 2.23	0.019 (t=4)

G   Appendix for the Cycle Graph Example

Consider the cycle graph with N nodes numbered from 0 to $N - 1$. Since each node has degree 2, the SNA $\mathbf{L} = \mathbf{A}/2$ is a circulant matrix produced by the

vector $\mathbf{v} = (\mathbf{e}_1 + \mathbf{e}_{N-1})/2$. Denote $\omega = \exp(2\pi i/N)$, the eigenvectors can be computed as

$$\mathbf{v}_j = \frac{1}{\sqrt{N}} \left(1, \omega^j, \omega^{2j}, \ldots, \omega^{(N-1)j}\right)$$

asociated to the eigenvalue $\lambda_j = \cos(2\pi j/N)$. First, we can note that $\lambda_j = \lambda_{N-j}$ for all $j \in \{1, \ldots, N/2\}$; therefore, the multiplicity of each eigenvalue is 2 except λ_0 and, if N is even, $\lambda_{N/2}$. Since the original matrix is symmetric, there exists a basis of real eigenvectors. A simple calculation

$$\mathbf{L}\Re\mathbf{v}_j + i\mathbf{L}\Im\mathbf{v}_j = \mathbf{L}\mathbf{v}_j = \lambda_j\mathbf{v}_j = \lambda_j\Re\mathbf{v}_j + i\lambda_j\Im\mathbf{v}_j$$

shows that $\Re\mathbf{v}_j$ and $\Im\mathbf{v}_j$, defined as

$$\Re\mathbf{v}_j = \frac{1}{\sqrt{N}} \left(\cos\left(\frac{2\pi j\, n}{N}\right)\right)_{n=0}^{N-1}, \ \Im\mathbf{v}_j = \frac{1}{\sqrt{N}} \left(\sin\left(\frac{2\pi j\, n}{N}\right)\right)_{n=0}^{N-1}$$

are two eigenvectors of the same eigenvalue λ_j. Tho show that they are linearly independent, we compute under which conditions

$$0 = a\Re\mathbf{v}_j + b\Im\mathbf{v}_j\,.$$

We note that the previous condition implies that for all $n \notin \{0, N/2\}$

$$0 = a\cos\left(\frac{2\pi j\, n}{N}\right) + b\sin\left(\frac{2\pi j\, n}{N}\right)$$
$$= \sqrt{a^2 + b^2}\sin\left(\frac{2\pi j\, n}{N} + \arctan\left(\frac{b}{a}\right)\right)$$

Suppose $a, b \neq 0$, then it must be

$$\frac{2\pi j\, n}{N} + \arctan\left(\frac{b}{a}\right) = k\pi\,, \ k \in \mathbb{Z}$$

which is equivalent to

$$2j\, n = \left(k - \frac{\arctan\left(\frac{b}{a}\right)}{\pi}\right) N\,, \ k \in \mathbb{Z}$$

The left-hand side is always an integer, while the right-hand side is an integer if and only if $b = 0$. This reduces the conditions to

$$\begin{cases} a\cos\left(\dfrac{2\pi j\, n}{N}\right) = 0 \\ |a|\sin\left(\dfrac{2\pi j\, n}{N}\right) = 0 \end{cases}$$

which is true if and only if $a = 0$. Consider now an even number of nodes N; the eigenspace of $\lambda_{N/2} = -1$ is

$$\mathbf{v}_{N/2} = \frac{1}{\sqrt{N}} \left((-1)^n \right)_{n=0}^{N-1}$$

hence, the maximal eigenvector of $\mathbf{\Delta}$ guarantees homophily 0. Consider now a number of nodes N divisible by 4; the eigenspace of $\lambda_{N/4} = 0$ has basis

$$\Re\mathbf{v}_{N/4} = \frac{1}{\sqrt{N}} \left(\cos\left(\frac{\pi n}{2}\right) \right)_{n=0}^{N-1} \, , \; \Im\mathbf{v}_{N/4} = \frac{1}{\sqrt{N}} \left(\sin\left(\frac{\pi n}{2}\right) \right)_{n=0}^{N-1}$$

Their sum is then equivalent to

$$\begin{aligned}
\Re\mathbf{v}_{N/4} + \Im\mathbf{v}_{N/4} &= \frac{1}{\sqrt{N}} \left(\cos\left(\frac{\pi n}{2}\right) + \sin\left(\frac{\pi n}{2}\right) \right)_{n=0}^{N-1} \\
&= \frac{\sqrt{2}}{\sqrt{N}} \left(\sin\left(\frac{\pi n}{2} + \frac{\pi}{4}\right) \right)_{n=0}^{N-1} \\
&= \sqrt{\frac{2}{N}} \left(\sin\left(\frac{\pi}{4}(2n+1)\right) \right)_{n=0}^{N-1} \\
&= \frac{1}{\sqrt{N}} \left(1, 1, -1, -1, \ldots \right)
\end{aligned}$$

hence, the mid eigenvector of $\mathbf{\Delta}$ guarantees homophily $1/2$. A visual explanation is shown in Figure 3.