

Acronyms

AUROC	Area under the ROC curve
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
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.
$\mathbf{1}$	Constant vector of all 1s.
\mathbf{M}^T	Transpose of \mathbf{M}
\mathbf{M}^*	Conjugate of \mathbf{M}
\mathbf{M}^H	Conjugate transpose of \mathbf{M}
$\ \mathbf{M}\ $	Spectral norm of \mathbf{M}
$\ \mathbf{M}\ _2$	Frobenius norm of \mathbf{M}
$\lambda(\mathbf{M})$	Spectrum of \mathbf{M}
$\sigma(\mathbf{M})$	Singular values of \mathbf{M}
$\mathcal{E}(\mathbf{x})$	Dirichlet energy computed on \mathbf{x}
$\mathcal{H}(\mathcal{G})$	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 \mathbf{M} .

A Implementation Details

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 $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$ 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 (3), 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

$$\mathbf{x}_{t+1} = \mathbf{x}_t - ih \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}.$$

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.

Algorithm 1: fLode

```

% A, x_0 are given.
% Preprocessing
1 D_in = diag(A1)
2 D_out = diag(A^T 1)
3 L = D_in^{-1/2} A D_out^{-1/2}
4 U, Σ, V^H = svd(L)
% The core of the algorithm is very simple
5 def training_step(x_0):
6     x_0 = input_MLP(x_0)
7     for t ∈ {1, ..., T} do
8         x_t = x_{t-1} - ih U Σ^α V^H x_{t-1} W
9     x_T = output_MLP(x_T)
10    return x_T

```

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 $\mathcal{O}(N^2 K)$.

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 the 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 (Kingma et al., 2015) as optimizer with a fixed learning rate. We perform early stopping if the validation metric does not increase for 200 epochs.

A.1 Real-World Graphs

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., 2023). 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 (L. Zhao et al., 2019), GGCN (Yan et al., 2022), Geom-GCN (Pei et al., 2019), H₂GCN (Zhu, Yan, et al., 2020), 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., 2023), and the model GRAFF from (Di Giovanni et al., 2023) itself. Finally, we included GREAD (Choi et al., 2023), GraphCON (Rusch et al., 2022), ACMP (Y. Wang et al., 2022) and GCN and GAT equipped with DropEdge (Rong et al., 2020).

Heterophily-specific Models For heterophily-specific datasets, we use the same models and results as in (Platonov et al., 2023). As baseline models we considered the topology-agnostic ResNet (He et al., 2016) and two graph-aware modifications: ResNet+SGC (F. Wu et al., 2019) where the initial node features are multiplied by powers of the SNA, and ResNet+adj, where rows of the adjacency matrix are used as additional node features; GCN (Kipf et al., 2017), GraphSAGE (Hamilton et al., 2017); GAT (Velickovic et al., 2018) and GT (Shi et al., 2021) as well as their modification GAT-sep and GT-sep which separate ego- and neighbor embeddings; H₂GCN (Zhu, Yan, et al., 2020), CPGNN (Zhu, Rossi, et al., 2021), GPRGNN (Chien et al., 2021), FSGNN (Maurya et al., 2021), GIoGNN (X. Li et al., 2022), FAGCN (Bo et al., 2021), GBK-GNN (Du et al., 2022), and JacobiConv (X. Wang et al., 2022).

The exact hyperparameters for FLODE are provided in Table 5.

A.2 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 and 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, \dots, 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 $i, j \in \{1, \dots, n_c\}$, 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.

Table 4: Test accuracy on node classification: top three models indicated as 1st, 2nd, 3rd.

(a) Undirected graphs.

	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
FAGCN	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
GREAD	37.90 ± 1.17	59.22 ± 1.44	71.38 ± 1.30	77.60 ± 1.81	90.23 ± 0.55	88.57 ± 0.66
GraphCON	35.58 ± 1.24	35.51 ± 1.40	49.63 ± 1.89	76.36 ± 2.67	88.01 ± 0.47	87.22 ± 1.48
ACMP	34.93 ± 1.26	40.05 ± 1.53	57.59 ± 2.09	76.71 ± 1.77	87.79 ± 0.47	87.71 ± 0.95
GCN+DropEdge	29.93 ± 0.80	41.30 ± 1.77	59.06 ± 2.04	76.57 ± 2.68	86.97 ± 0.42	83.54 ± 1.06
GAT+DropEdge	28.95 ± 0.76	41.27 ± 1.76	58.95 ± 2.13	76.13 ± 2.20	86.91 ± 0.45	83.54 ± 1.06
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

(b) Directed graphs.

	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

(c) Heterophily-specific graphs. For Minesweeper, Tolokers and Questions the evaluation metric is the AUROC.

	Roman-empire	Minesweeper	Tolokers	Questions
ResNet	65.88 ± 0.38	50.89 ± 1.39	72.95 ± 1.06	70.34 ± 0.76
ResNet+SGC	73.90 ± 0.51	70.88 ± 0.90	80.70 ± 0.97	75.81 ± 0.96
ResNet+adj	52.25 ± 0.40	50.42 ± 0.83	78.78 ± 1.11	75.77 ± 1.24
GCN	73.69 ± 0.74	89.75 ± 0.52	83.64 ± 0.67	76.09 ± 1.27
GraphSAGE	85.74 ± 0.67	93.51 ± 0.57	82.43 ± 0.44	76.44 ± 0.62
GAT	80.87 ± 0.30	92.01 ± 0.68	83.70 ± 0.47	77.43 ± 1.20
GAT-sep	88.75 ± 0.41	93.91 ± 0.35	83.78 ± 0.43	76.79 ± 0.71
GT	86.51 ± 0.73	91.85 ± 0.76	83.23 ± 0.64	77.95 ± 0.68
GT-sep	87.32 ± 0.39	92.29 ± 0.47	82.52 ± 0.92	78.05 ± 0.93
FAGCN	60.11 ± 0.52	89.71 ± 0.31	73.35 ± 1.01	63.59 ± 1.46
CPGNN	63.96 ± 0.62	52.03 ± 5.46	73.36 ± 1.01	65.96 ± 1.95
H ₂ GCN	64.85 ± 0.27	86.24 ± 0.61	72.94 ± 0.97	55.48 ± 0.91
FSGNN	79.92 ± 0.56	90.08 ± 0.70	82.76 ± 0.61	78.86 ± 0.92
GloGNN	59.63 ± 0.69	51.08 ± 1.23	73.39 ± 1.17	65.74 ± 1.19
FAGCN	65.22 ± 0.56	88.17 ± 0.73	77.75 ± 1.05	77.24 ± 1.26
GBK-GNN	74.57 ± 0.47	90.85 ± 0.58	81.01 ± 0.67	74.47 ± 0.86
JacobiConv	71.14 ± 0.42	89.66 ± 0.40	68.66 ± 0.65	73.88 ± 1.16
FLODE	74.97 ± 0.53	92.43 ± 0.51	84.17 ± 0.58	78.39 ± 1.22

Table 5: Selected hyperparameters, learned exponent, step size, and 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

(c) Heterophily-specific graphs.

	Dataset			
	Roman-empire	Minesweeper	Tolokers	Questions
learning rate	10^{-3}	10^{-3}	10^{-3}	10^{-2}
weight decay	0	0	0	$5 \cdot 10^{-4}$
hidden channels	512	512	512	128
num. layers	4	4	4	5
encoder layers	2	2	1	2
decoder layers	2	2	2	2
input dropout	0	0	0	0
decoder dropout	0	0	0	0
exponent	0.689 ± 0.038	0.749 ± 0.017	1.053 ± 0.041	1.090 ± 0.046
step size	0.933 ± 0.015	0.984 ± 0.004	0.993 ± 0.009	0.789 ± 0.062
Dirichlet energy	0.059 ± 0.003	0.173 ± 0.019	0.155 ± 0.013	0.092 ± 0.039

Table 6: Node classification accuracy of ordered **DSBM** graphs: top three models as **1st**, **2nd** and **3rd**.

(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
GraphSAGE-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
FLODE	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
GraphSAGE-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
FLODE	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

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 A.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 6 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 6 with exact numerical values in Table 6b.

A.3 Ablation Study

We perform an ablation study on Chameleon and Squirrel (directed, heterophilic), and Citeseer (undirected, homophilic). For this, we sweep over different model options using the same hyperparameters

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	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$1 \cdot 10^{-3}$					
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

via grid search. The test accuracy corresponding to the hyperparameters that yielded maximum validation accuracy is reported in Table 8.

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.

From Theory to Practice. We conduct an ablation study to investigate the role of depth on Chameleon, Citeseer, Cora, and Squirrel datasets. The results, depicted in Figure 8, demonstrate that the neural **ODE** framework enables **GNNs** to scale to large depths (256 layers). Moreover, we see that the fractional Laplacian improves over the standard Laplacian in the heterophilic graphs which is supported by our claims in Section 5.2. We highlight that using only the fractional Laplacian without the neural **ODE** framework oftentimes outperforms the standard Laplacian with the neural **ODE** framework. This indicates the importance of the long-range connections built by the fractional Laplacian.

We further demonstrate the close alignment of our theoretical and experimental results, which enables us to precisely anticipate when the models will exhibit **HFD** or **LFD** behaviors. In this context, we calculate parameters (according to Theorem D.5) and illustrate at each depth the expected and observed behaviors. For Squirrel and Chameleon, which are heterophilic graphs, we observe that both their theoretical and empirical behaviors are **HFD**. Additionally, the learned exponent is small. In contrast, for Cora and Citeseer, we see the opposite.

Finally, we employ the best hyperparameters in Table 5a to solve both fractional heat and Schrödinger graph **ODEs**, further substantiating the intimate link between our theoretical advancements and practical applications.

Table 8: Ablation study on node classification task: top two models are indicated as 1st and 2nd

(a) Chameleon (directed, heterophilic).

	Update Rule	Test Accuracy	Dirichlet Energy
D	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	77.79 ± 1.42	0.213 (t=5)
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L} \mathbf{x}_t \mathbf{W}$	75.72 ± 1.13	0.169 (t=6)
	$\mathbf{x}_{t+1} = -i \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	77.35 ± 2.22	0.177 (t=4)
	$\mathbf{x}_{t+1} = -i \mathbf{L} \mathbf{x}_t \mathbf{W}$	69.61 ± 1.59	0.178 (t=4)
U	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	73.60 ± 1.68	0.131 (t=4)
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L} \mathbf{x}_t \mathbf{W}$	70.15 ± 0.86	0.035 (t=4)
	$\mathbf{x}_{t+1} = -i \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	71.25 ± 3.04	0.118 (t=4)
	$\mathbf{x}_{t+1} = -i \mathbf{L} \mathbf{x}_t \mathbf{W}$	67.19 ± 2.49	0.040 (t=4)
D	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	77.33 ± 1.47	0.378 (t=6)
	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L} \mathbf{x}_t \mathbf{W}$	73.55 ± 0.94	0.165 (t=6)
	$\mathbf{x}_{t+1} = -\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	74.12 ± 3.60	0.182 (t=4)
	$\mathbf{x}_{t+1} = -\mathbf{L} \mathbf{x}_t \mathbf{W}$	68.47 ± 2.77	0.208 (t=4)

(b) Squirrel (directed, heterophilic).

	Update Rule	Test Accuracy	Dirichlet Energy
D	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	74.03 ± 1.58	0.38 ± 0.02
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L} \mathbf{x}_t \mathbf{W}$	64.04 ± 2.25	0.35 ± 0.02
	$\mathbf{x}_{t+1} = -i \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	64.25 ± 1.85	0.46 ± 0.01
	$\mathbf{x}_{t+1} = -i \mathbf{L} \mathbf{x}_t \mathbf{W}$	42.04 ± 1.58	0.29 ± 0.05
U	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	64.23 ± 1.84	0.40 ± 0.02
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L} \mathbf{x}_t \mathbf{W}$	55.19 ± 1.52	0.26 ± 0.03
	$\mathbf{x}_{t+1} = -i \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	61.40 ± 2.15	0.43 ± 0.01
	$\mathbf{x}_{t+1} = -i \mathbf{L} \mathbf{x}_t \mathbf{W}$	41.19 ± 1.95	0.20 ± 0.02
D	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	71.86 ± 1.65	0.50 ± 0.01
	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L} \mathbf{x}_t \mathbf{W}$	59.34 ± 1.78	0.43 ± 0.03
	$\mathbf{x}_{t+1} = -\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	42.91 ± 7.86	0.32 ± 0.08
	$\mathbf{x}_{t+1} = -\mathbf{L} \mathbf{x}_t \mathbf{W}$	35.37 ± 1.69	0.25 ± 0.05
U	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	62.95 ± 2.02	0.61 ± 0.08
	$\mathbf{x}_{t+1} = \mathbf{x}_t - h\mathbf{L} \mathbf{x}_t \mathbf{W}$	52.19 ± 1.17	0.51 ± 0.07
	$\mathbf{x}_{t+1} = -\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	59.04 ± 0.02	0.44 ± 0.02
	$\mathbf{x}_{t+1} = -\mathbf{L} \mathbf{x}_t \mathbf{W}$	39.69 ± 1.54	0.20 ± 0.02

(c) Citeseer (undirected, homophilic).

	Update Rule	Test Accuracy	Dirichlet Energy
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	78.07 ± 1.62	0.021 (t=5)
	$\mathbf{x}_{t+1} = \mathbf{x}_t - ih\mathbf{L} \mathbf{x}_t \mathbf{W}$	77.97 ± 2.29	0.019 (t=4)
	$\mathbf{x}_{t+1} = -i \mathbf{L}^\alpha \mathbf{x}_t \mathbf{W}$	77.27 ± 2.10	0.011 (t=6)
	$\mathbf{x}_{t+1} = -i \mathbf{L} \mathbf{x}_t \mathbf{W}$	77.97 ± 2.23	0.019 (t=4)

Table 9: Learned α and spectrum of \mathbf{W} . According to Theorem 5.3, we denote $\text{FD} := \lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L})) - \lambda_1(\mathbf{W})$ and $\text{FD} := \Im(\lambda_K(\mathbf{W})) f_\alpha(\lambda_1(\mathbf{L})) - \Im(\lambda_1(\mathbf{W}))$ for the fractional heat (H) and Schrödinger (S) graph ODEs, respectively. The heterophilic graphs Squirrel and Chameleon exhibit HFD since $\text{FD} < 0$, while the homophilic Cora, Citeseer, Pubmed exhibit LFD since $\text{FD} > 0$.

		Film	Squirrel	Chameleon	Citeseer	Pubmed	Cora
	$\lambda_1(\mathbf{L})$	-0.9486	-0.8896	-0.9337	-0.5022	-0.6537	-0.4826
H	α	1.008 ± 0.007	0.19 ± 0.05	0.37 ± 0.14	0.89 ± 0.06	1.15 ± 0.08	0.89 ± 0.01
	$\lambda_1(\mathbf{W})$	-2.774 ± 0.004	-1.62 ± 0.03	-1.81 ± 0.02	-1.76 ± 0.01	-1.66 ± 0.06	-1.81 ± 0.01
	$\lambda_K(\mathbf{W})$	2.858 ± 0.009	2.21 ± 0.03	2.29 ± 0.05	2.28 ± 0.06	1.1 ± 0.3	2.32 ± 0.01
	FD	0.367 ± 0.001	-0.54 ± 0.02	-0.42 ± 0.04	0.52 ± 0.02	0.97 ± 0.09	0.60 ± 0.01
S	α	1.000 ± 0.002	0.17 ± 0.03	0.34 ± 0.11	0.90 ± 0.07	0.76 ± 0.07	0.90 ± 0.02
	$\Im(\lambda_1(\mathbf{W}))$	-2.795 ± 0.001	-1.68 ± 0.01	-1.79 ± 0.01	-1.70 ± 0.04	-1.74 ± 0.01	-1.78 ± 0.01
	$\Im(\lambda_K(\mathbf{W}))$	2.880 ± 0.002	2.21 ± 0.03	2.46 ± 0.02	2.29 ± 0.07	0.98 ± 0.09	2.30 ± 0.02
	FD	0.4945 ± 0.0001	-0.48 ± 0.03	-0.62 ± 0.03	0.46 ± 0.06	1.03 ± 0.05	0.59 ± 0.01

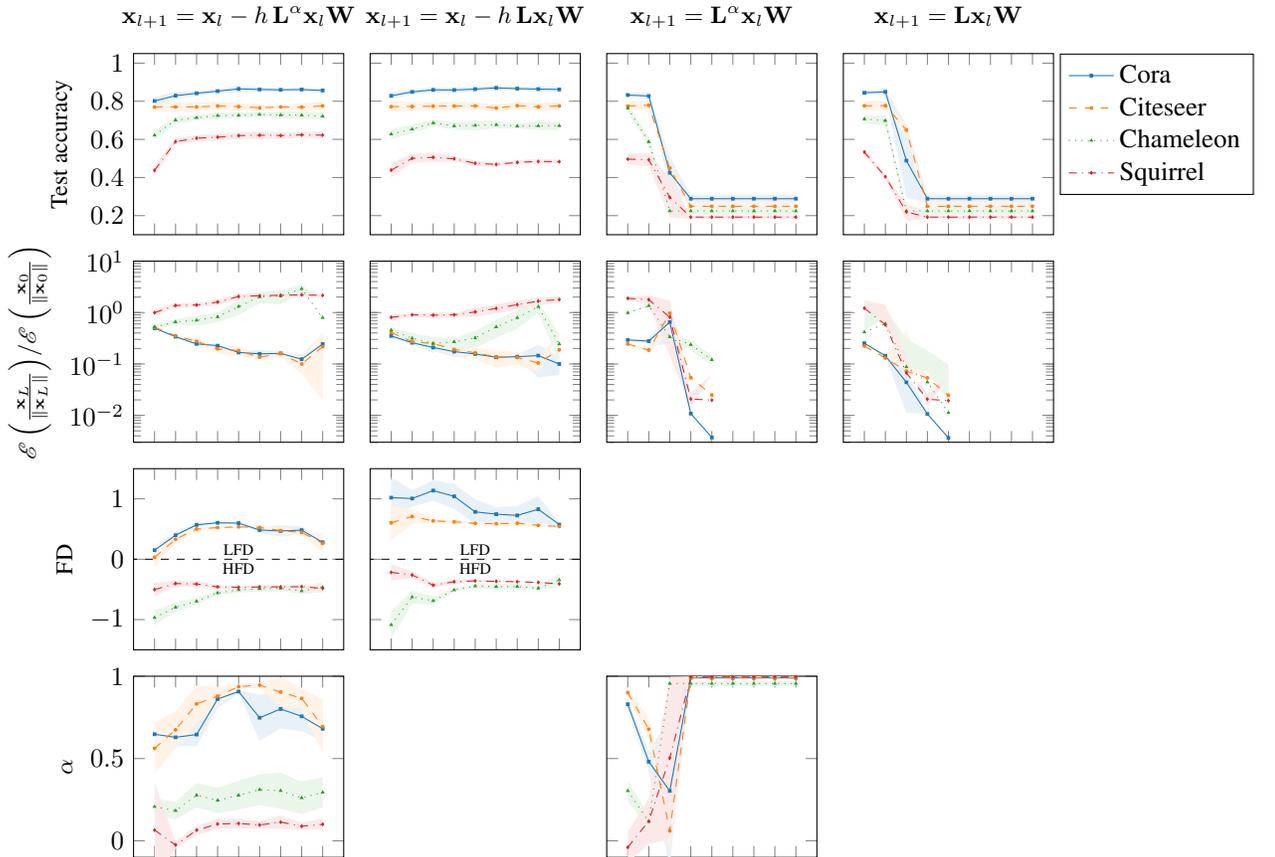


Figure 8: Ablation study on the effect of different update rules and different number of layers on undirected datasets. The x-axis shows the number of layers 2^L for $L \in \{0, \dots, 8\}$. FD is calculated according to Theorem 5.3.

B Appendix for Section 3

Proposition 3.3. *Let \mathcal{G} be a directed graph with SNA \mathbf{L} . For every $\lambda \in \lambda(\mathbf{L})$, it holds $|\lambda| \leq 1$ and $\lambda(\mathbf{I} - \mathbf{L}) = 1 - \lambda(\mathbf{L})$.*

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

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

$$\begin{aligned}
|\mathbf{x}^H \mathbf{L} \mathbf{x}| &\stackrel{(1)}{\leq} \sum_{i=1}^N \sum_{j=1}^N a_{i,j} \frac{|x_i| |x_j|}{\sqrt{d_i^{\text{in}} d_j^{\text{out}}}} \\
&= \sum_{i=1}^N \frac{|x_i|}{\sqrt{d_i^{\text{in}}}} \sum_{j=1}^N a_{i,j} \frac{|x_j|}{\sqrt{d_j^{\text{out}}}} \\
&\stackrel{(2)}{\leq} \sum_{i=1}^N \frac{|x_i|}{\sqrt{d_i^{\text{in}}}} \sqrt{\sum_{j=1}^N a_{i,j} \frac{|x_j|^2}{d_j^{\text{out}}}} \sum_{j=1}^N a_{i,j} \\
&= \sum_{i=1}^N |x_i| \sqrt{\sum_{j=1}^N a_{i,j} \frac{|x_j|^2}{d_j^{\text{out}}}} \\
&\stackrel{(3)}{\leq} \sqrt{\sum_{i=1}^N |x_i|^2 \sum_{i=1}^N \sum_{j=1}^N a_{i,j} \frac{|x_j|^2}{d_j^{\text{out}}}} \\
&= \sum_{i=1}^N |x_i|^2,
\end{aligned}$$

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

Proposition 3.5. *Let \mathcal{G} be a directed graph with SNA \mathbf{L} . Then $1 \in \lambda(\mathbf{L})$ if and only if the graph is weakly balanced. Suppose the graph is strongly connected; then $-1 \in \lambda(\mathbf{L})$ if and only if the graph is weakly balanced with an 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 previous proof are actual equalities. Therefore, we have to find another way to prove the statement. Suppose that the graph is weakly balanced, then

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

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

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

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

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

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

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. \square

Proposition 3.6. *For every $\mathbf{x} \in \mathbb{C}^{N \times K}$, we have*

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

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

Proof. By direct computation, it holds

$$\begin{aligned} & \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}}} \\ &\quad - \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}^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}^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}^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}^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(\text{trace}(\mathbf{x}^H (\mathbf{I} - \mathbf{L}) \mathbf{x})). \end{aligned}$$

The last claim can be proved as follows. For simplicity, suppose $\mathbf{x} \in \mathbb{R}^N$. The “ \Leftarrow ” is clear since one can choose \mathbf{x} to be \mathbf{k} . To prove the “ \Rightarrow ”, we reason by contradiction. Suppose there exists a $\mathbf{x} \neq 0$ such that $\mathcal{E}(\mathbf{x}) = 0$ and the underlying graph is not weakly connected, i.e.,

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

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

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

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

We give the following simple corollary.

Corollary B.1. *For every $\mathbf{x} \in \mathbb{R}^{N \times K}$, it holds $\mathcal{E}(\mathbf{x}) = \frac{1}{2} \Re(\text{vec}(\mathbf{x})^H (\mathbf{I} \otimes (\mathbf{I} - \mathbf{L})) \text{vec}(\mathbf{x}))$.*

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

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

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

$$\begin{aligned}
\mathbf{M}^H\mathbf{M} &= \mathbf{V}\Lambda^*\mathbf{V}^H\mathbf{V}\Lambda\mathbf{V}^H = \mathbf{V}|\Lambda|^2\mathbf{V}^H, \\
\mathbf{M}^H\mathbf{M} &= \mathbf{R}\Sigma\mathbf{L}^H\mathbf{L}\Sigma\mathbf{R}^H = \mathbf{L}\Sigma^2\mathbf{L}^H.
\end{aligned}$$

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

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

Finally, we note that it must hold $\mathbf{R} = \mathbf{V} \exp(i\Theta)$ where $\Theta = \text{diag}(\{\text{atan2}(\Re\lambda_i, \Im\lambda_i)\}_{i=1}^N)$ and atan2 is the 2-argument arctangent. \square

We proceed by proving Theorem 4.1, 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 in particular for fractional Laplacians that are defined via the SVD of a graph SNA.

Lemma C.2. *Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ with singular values $\sigma(\mathbf{M}) \subset [a, b]$. For $f : [a, b] \rightarrow \mathbb{R}$, define $f(\mathbf{M}) = \mathbf{U}f(\Sigma)\mathbf{V}^H$, where $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^H$ 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 $g : [a, b] \rightarrow \mathbb{R}$ be any function, then

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

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 $i, j \in \{1, \dots, n\}$. 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

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

from which the thesis follows. \square

Finally, we give a proof of Theorem 4.1, which is a consequence of the previous statement.

Proof of Theorem 4.1. The eigenvalues of \mathbf{L} are in the unit circle, i.e., $\|\mathbf{L}\| \leq 1$. Hence, $\|\mathbf{L}\mathbf{L}^H\| \leq 1$, and the singular values of \mathbf{L} are in $[0, 1]$. By Lemma C.2 and the fact that $f(x) = x^\alpha$ has modulus of continuity $\omega(t) = t^\alpha$ the thesis follows. \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

$$\mathbf{x}'(t) = -\mathbf{M}\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0. \quad (5)$$

Theorem D.1 (Existence and uniqueness of linear ODE solution). *The initial value problem given by (5) 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 (5) 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 (5) can then be written as

$$\mathbf{x}(t) = \exp(-\mathbf{M}t)\mathbf{x}_0. \quad (6)$$

D.1 Appendix for Section 5.1

In this section, we analyze the solution to (2) and (3). We further provide a proof for Theorem 5.3. We begin by considering the solution to the fractional heat equation (2). The analysis for the Schrödinger equation (3) 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

$$\text{vec}(\mathbf{x})'(t) = -\mathbf{W} \otimes \mathbf{L}^\alpha \text{vec}(\mathbf{x})(t). \quad (7)$$

In the undirected case \mathbf{L} and $\mathbf{I} - \mathbf{L}$ are both symmetric, and the eigenvalues satisfy the relation $\lambda_i(\mathbf{I} - \mathbf{L}) = 1 - \lambda_i(\mathbf{L})$. The corresponding eigenvectors $\psi_i(\mathbf{L})$ and $\psi_i(\mathbf{I} - \mathbf{L})$ can be chosen to be the same for \mathbf{L} and $\mathbf{I} - \mathbf{L}$. 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 $\mathbf{U} = [\psi_1(\mathbf{L}), \dots, \psi_N(\mathbf{L})]$ is an orthogonal matrix containing the eigenvectors of \mathbf{L} , and \mathbf{D} is the diagonal matrix of eigenvalues.

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

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

$$\{\lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))\}_{r \in \{1, \dots, K\}, l \in \{1, \dots, N\}}, \{\psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L})\}_{r \in \{1, \dots, K\}, l \in \{1, \dots, N\}}.$$

Note that $1 \in \lambda(\mathbf{L})$ and, since $\text{trace}(\mathbf{L}) = 0$, 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 $\text{vec}(\mathbf{x}_0)$ can be decomposed as

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

Then, the solution $\text{vec}(\mathbf{x})(t)$ of (7) can be written as

$$\text{vec}(\mathbf{x})(t) = \sum_{r=1}^K \sum_{l=1}^N c_{r,l} \exp(-t \lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))) \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}). \quad (8)$$

The following result shows the relationship between the frequencies of $\mathbf{I} - \mathbf{L}$ and the Dirichlet energy and serves as a basis for the following proofs.

Lemma D.2. *Let \mathcal{G} be a graph with SNA \mathbf{L} . 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*

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

and $(\mathbf{I} \otimes (\mathbf{I} - \mathbf{L})) \text{vec}(\boldsymbol{\varphi}) = \lambda \text{vec}(\boldsymbol{\varphi})$. Then,

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

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

$$\mathcal{E}(\text{vec}(\boldsymbol{\varphi})) = \frac{1}{2} \Re(\text{vec}(\boldsymbol{\varphi})^H (\mathbf{I} \otimes (\mathbf{I} - \mathbf{L})) \text{vec}(\boldsymbol{\varphi})) = \frac{1}{2} \Re(\lambda \text{vec}(\boldsymbol{\varphi})^H \text{vec}(\boldsymbol{\varphi})) = \frac{1}{2} \Re(\lambda).$$

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

Another useful result that will be extensively used in proving Theorem 5.3 is presented next.

Lemma D.3. *Suppose $\mathbf{x}(t)$ can be expressed as*

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

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.

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

Then

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

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

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=1}^K \sum_{n=1}^N c_{k,n} \exp(-t \lambda_{k,n}) \mathbf{v}_n \otimes \mathbf{w}_m \\ &= \exp(-t \lambda_{a,b}) \left(c_{a,b} \mathbf{v}_a \otimes \mathbf{w}_b + \sum_{\substack{(k,n) \in [K] \times [N] \\ (k,n) \neq (a,b)}} c_{k,n} \exp(-t (\lambda_{k,n} - \lambda_{a,b})) \mathbf{v}_k \otimes \mathbf{w}_n \right). \end{aligned}$$

We note that

$$\begin{aligned} \lim_{t \rightarrow \infty} |\exp(-t (\lambda_{k,n} - \lambda_{a,b}))| &= \lim_{t \rightarrow \infty} |\exp(-t \Re(\lambda_{k,n} - \lambda_{a,b})) \exp(-i t \Im(\lambda_{k,n} - \lambda_{a,b}))| \\ &= \lim_{t \rightarrow \infty} \exp(-t \Re(\lambda_{k,n} - \lambda_{a,b})) \\ &= 0, \end{aligned}$$

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

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

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

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.,

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

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

A similar result to Lemma D.3 holds for a slightly different representation of $\mathbf{x}(t)$.

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

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

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.

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

Then

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

Proof. The proof follows the same reasoning as in the proof of Lemma D.3. The difference is that the dominating frequency is the one with the minimal imaginary part, since

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

and, consequently,

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

□

D.1.1 Proof of Theorem 5.3

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

$$\begin{aligned}\lambda_+(\mathbf{L}) &:= \arg \min_l \{ \lambda_l(\mathbf{L}) : \lambda_l(\mathbf{L}) > 0 \} , \\ \lambda_-(\mathbf{L}) &:= \arg \max_l \{ \lambda_l(\mathbf{L}) : \lambda_l(\mathbf{L}) < 0 \} .\end{aligned}\tag{9}$$

We assume that the channel mixing $\mathbf{W} \in \mathbb{R}^{K \times K}$ and the graph Laplacians $\mathbf{L}, \mathbf{I} - \mathbf{L} \in \mathbb{R}^{N \times N}$ 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 5.3 for the fractional heat equation (2) and provide its full proof, which follows a similar frequency analysis to the one in (Di Giovanni et al., 2023, Theorem B.3)

Theorem D.5. *Let \mathcal{G} be an undirected graph with SNA \mathbf{L} . Consider the initial value problem in (2) 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.*

($\alpha > 0$) The solution to (2) is *HFD* if

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

and *LFD* otherwise.

($\alpha < 0$) The solution to (2) is $(1 - \lambda_-(\mathbf{L}))$ -*FD* if

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

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

Proof of ($\alpha > 0$). As derived in (8), the solution of (2) with initial condition \mathbf{x}_0 can be written in a vectorized form as

$$\begin{aligned}\text{vec}(\mathbf{x})(t) &= \exp(-t \mathbf{W}^\top \otimes \mathbf{L}^\alpha) \text{vec}(\mathbf{x}_0) \\ &= \sum_{r=1}^K \sum_{l=1}^N c_{r,l} \exp(-t \lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))) \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}),\end{aligned}$$

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.,

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

The key insight is to separate the eigenprojection corresponding to the most negative frequency. By Lemma D.3, this frequency component dominates for t going to infinity.

Suppose

$$\lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L})) < \lambda_1(\mathbf{W}) f_\alpha(\lambda_N(\mathbf{L})) = \lambda_1(\mathbf{W}) .$$

In this case, $\lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L}))$ 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

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

By standard properties of the Kronecker product, we have

$$(\mathbf{I} \otimes \mathbf{L})(\psi_K(\mathbf{W}) \otimes \psi_1(\mathbf{L})) = (\mathbf{I} \psi_K(\mathbf{W})) \otimes (\mathbf{L} \psi_1(\mathbf{L})) = \lambda_1(\mathbf{L}) \psi_K(\mathbf{W}) \otimes \psi_1(\mathbf{L}) , \tag{10}$$

i.e., $\psi_K(\mathbf{W}) \otimes \psi_1(\mathbf{L})$ is an eigenvector of $\mathbf{I} \otimes \mathbf{L}$ corresponding to the eigenvalue $\lambda_1(\mathbf{L})$. Then, by Proposition 3.3, $\psi_K(\mathbf{W}) \otimes \psi_1(\mathbf{L})$ is also an eigenvector of $\mathbf{I} \otimes \mathbf{I} - \mathbf{L}$ corresponding to the eigenvalue $1 - \lambda_1(\mathbf{L}) = \lambda_N(\mathbf{I} - \mathbf{L})$. An application of Lemma D.2 finishes the proof.

Similarly, we can show that if $\alpha > 0$ and $\lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L})) > \lambda_1(\mathbf{W})$ the lowest frequency component $\lambda_1(\mathbf{I} - \mathbf{L})$ is dominant.

Proof of $(\alpha < 0)$. In this case either $f_\alpha(\lambda_+(\mathbf{L}))\lambda_1(\mathbf{W})$ or $f_\alpha(\lambda_-(\mathbf{L}))\lambda_K(\mathbf{W})$ are the most negative frequency components. Hence, if $f_\alpha(\lambda_-(\mathbf{L}))\lambda_K(\mathbf{W}) > f_\alpha(\lambda_+(\mathbf{L}))\lambda_1(\mathbf{W})$ the frequency $f_\alpha(\lambda_+(\mathbf{L}))\lambda_1(\mathbf{W})$ is dominating and otherwise the frequency $f_\alpha(\lambda_-(\mathbf{L}))\lambda_K(\mathbf{W})$. We can see this by following the exact same reasoning of (i). \square

Remark D.6. In the proof of $(\alpha < 0)$, 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.

We now generalize the previous result to all directed graphs with normal SNA.

Theorem D.7. Let \mathcal{G} be a strongly connected directed graph with normal SNA \mathbf{L} such that $\lambda_1(\mathbf{L}) \in \mathbb{R}$. Consider the initial value problem in (2) with channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ and $\alpha > 0$. Then, for almost all initial values $\mathbf{x}_0 \in \mathbb{R}^{N \times K}$ the solution to (2) is HFD if

$$\lambda_K(\mathbf{W})|\lambda_1(\mathbf{L})|^\alpha < \lambda_1(\mathbf{W})|\lambda_N(\mathbf{L})|^\alpha,$$

and LFD otherwise.

Proof. Any normal matrix is unitary diagonalizable, i.e., there exist eigenvalues $\lambda_1, \dots, \lambda_N$ and corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_N$ such that $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$. Then, by Lemma C.1, the singular value decomposition of \mathbf{L} is given by $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where

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

Hence,

$$\mathbf{L}^\alpha = \mathbf{U}\mathbf{\Sigma}^\alpha\mathbf{V}^H = \mathbf{V}|\mathbf{\Lambda}|^\alpha \exp(i\mathbf{\Theta})\mathbf{V}^H.$$

Then, equivalent to the derivation of (8), the solution to the vectorized fractional heat equation

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

is given by

$$\text{vec}(\mathbf{x})(t) = \sum_{r=1}^K \sum_{l=1}^N c_{r,l} \exp(-t\lambda_r(\mathbf{W})f_\alpha(\lambda_l(\mathbf{L}))) \psi_r(\mathbf{W}) \otimes \psi_l(\mathbf{L}).$$

with

$$f_\alpha(\lambda_l(\mathbf{L})) = |\lambda(\mathbf{L})_l|^\alpha \exp(i\theta_l).$$

Now, equivalent to the proof of Theorem 5.3, we apply Lemma D.3. Therefore, the dominating frequency is given by the eigenvalue of $\mathbf{W} \otimes \mathbf{L}^\alpha$ with the most negative real part. The eigenvalues of $\mathbf{W} \otimes \mathbf{L}^\alpha$ are given by $\lambda_r(\mathbf{W})f_\alpha(\lambda_l(\mathbf{L}))$ for $r = 1, \dots, K, l = 1, \dots, N$. The corresponding real parts are given by

$$\Re(\lambda_r(\mathbf{W})f_\alpha(\lambda_l(\mathbf{L}))) = \lambda_r(\mathbf{W})|\lambda(\mathbf{L})_l|^\alpha \cos(\theta_l) = \lambda_r(\mathbf{W})|\lambda(\mathbf{L})_l|^{\alpha-1} \Re(\lambda(\mathbf{L})_l).$$

By Perron-Frobenius, the eigenvalue of \mathbf{L} with the largest eigenvalues is given by $\lambda_N(\mathbf{L}) \in \mathbb{R}$. Hence, for all $l = 1, \dots, N$,

$$|\lambda(\mathbf{L})_l|^\alpha \cos(\theta_l) \leq |\lambda(\mathbf{L})_N|^\alpha.$$

Similarly, for all $l = 1, \dots, N$ with $\Re(\lambda(\mathbf{L})_l) < 0$,

$$-|\lambda(\mathbf{L})_l|^\alpha \cos(\theta_l) \leq -|\lambda(\mathbf{L})_1|^\alpha.$$

Thus, the frequency with the most negative real part is either given by $\lambda_K(\mathbf{W})f_\alpha(\lambda_1(\mathbf{L}))$ or $\lambda_1(\mathbf{W})f_\alpha(\lambda_N(\mathbf{L}))$. The remainder of the proof is analogous to the proof of Theorem D.7. \square

In the following, we provide the complete statement and proof for the claims made in Theorem 5.3 when the underlying ODE is the Schrödinger equation as presented in (3).

Theorem D.8. Let \mathcal{G} be a undirected graph with **SNA** \mathbf{L} . Consider the initial value problem in (3) 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.

($\alpha > 0$) Solutions of (3) are **HFD** if

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

and **LFD** otherwise.

($\alpha < 0$) Let $\lambda_+(\mathbf{L})$ and $\lambda_-(\mathbf{L})$ be the smallest positive and biggest negative non-zero eigenvalue of \mathbf{L} , respectively. Solutions of (3) are $(1 - \lambda_-(\mathbf{L}))$ -**FD** if

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

Otherwise, solutions of (3) are $(1 - \lambda_+(\mathbf{L}))$ -**FD**.

Proof. The proof follows the same reasoning as the proof for the heat equation in Theorem D.5. The difference is that we now apply Lemma D.4 instead of Lemma D.3.

Therefore, the dominating frequency is either $\lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L}))$ or $\lambda_1(\mathbf{W}) f_\alpha(\lambda_N(\mathbf{L}))$ if $\alpha > 0$, and $\lambda_K(\mathbf{W}) f_\alpha(\lambda_-(\mathbf{L}))$ or $\lambda_1(\mathbf{W}) f_\alpha(\lambda_+(\mathbf{L}))$ if $\alpha < 0$. \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 (2) at time $nh > 0$ can be approximated with an explicit Euler scheme

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

which can be further simplified via the binomial theorem as

$$\text{vec}(\mathbf{x})(nh) = (\mathbf{I} - h(\mathbf{W} \otimes \mathbf{L}^\alpha))^n \text{vec}(\mathbf{x}_0) . \quad (11)$$

Hence, it holds the representation formula

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

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

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

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

Proposition D.9. Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional heat equation (2) with $\alpha \in \mathbb{R}$. Let $\{\mathbf{x}(nh)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating (2) with an explicit Euler scheme with step size h . Suppose $h < \|\mathbf{W}\|^{-1}$. Then, for almost all initial values \mathbf{x}_0

$$\mathcal{E} \left(\frac{\mathbf{x}(nh)}{\|\mathbf{x}(nh)\|_2} \right) \xrightarrow{n \rightarrow \infty} \begin{cases} \frac{\lambda_N(\mathbf{I} - \mathbf{L})}{2} , & \text{if } \lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L})) < \lambda_1(\mathbf{W}) , \\ 0 , & \text{otherwise} . \end{cases}$$

Proof. Define

$$(\lambda_a, \lambda_b) := \arg \max_{r,l} \{|1 - h \lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))| : r \in \{1, \dots, K\}, l \in \{1, \dots, N\}\} .$$

By the hypothesis on h , this is equivalent to

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

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,

$$\frac{\text{vec}(\mathbf{x})(nh)}{\|\text{vec}(\mathbf{x})(nh)\|_2} \xrightarrow{n \rightarrow \infty} \frac{c_{a,b} \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})}{\|c_{a,b} \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})\|_2}.$$

If the condition $\lambda_K(\mathbf{W}) f_\alpha(\lambda_1(\mathbf{L})) < \lambda_1(\mathbf{W})$ is satisfied, we have $b = 1$. Then by (10), the normalized $\text{vec}(\mathbf{x})$ converges to the eigenvector of $\mathbf{I} \otimes \mathbf{I} - \mathbf{L}$ corresponding to the largest frequency $1 - \lambda_1(\mathbf{L}) = \lambda_N(\mathbf{I} - \mathbf{L})$. An application of Lemma D.2 finishes the proof.

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

Similarly to Proposition D.9 one can prove the following results for negative fractions.

Proposition D.10. *Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional heat equation (2) with $\alpha < 0$. Let $\{\mathbf{x}(nh)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating the solution of (2) 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*

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

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

Proof. The proof follows the same reasoning as the proof of Proposition D.9 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 (3) at time $nh > 0$ can be approximated with an explicit Euler scheme as well. Similarly to the previous section, we can write

$$\text{vec}(\mathbf{x})(nh) = (\mathbf{I} + ih(\mathbf{W} \otimes \mathbf{L}^\alpha))^n \text{vec}(\mathbf{x}_0).$$

and

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

The dominating frequency will be discussed in the following theorem.

Proposition D.11. *Let $h \in \mathbb{R}$, $h > 0$. Let $\{\mathbf{x}(nh)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating (3) 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*

$$\mathcal{E} \left(\frac{\mathbf{x}(nh)}{\|\mathbf{x}(nh)\|_2} \right) \xrightarrow{n \rightarrow \infty} \begin{cases} \frac{\lambda_N(\mathbf{I} - \mathbf{L})}{2}, & \text{if } f_\alpha(\lambda_1(\mathbf{L})) \Im(\lambda_K(\mathbf{W})) < f_\alpha(\lambda_N(\mathbf{L})) \Im(\lambda_1(\mathbf{W})) \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Define

$$(\lambda_a, \lambda_b) := \arg \max_{r,l} \{|1 + ih \lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))| : r \in \{1, \dots, K\}, l \in \{1, \dots, N\}\}.$$

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

$$|1 + ih \lambda_a(\mathbf{W}) f_\alpha(\lambda_b(\mathbf{L}))| > |1 + ih \lambda_r(\mathbf{W}) f_\alpha(\lambda_l(\mathbf{L}))|. \quad (12)$$

Hence,

$$\frac{\text{vec}(\mathbf{x})(t)}{\|\text{vec}(\mathbf{x})(t)\|_2} \xrightarrow{t \rightarrow \infty} \frac{c_{a,b} \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})}{\|c_{a,b} \psi_a(\mathbf{W}) \otimes \psi_b(\mathbf{L})\|_2}.$$

We continue by determining the indices a and b . To do so, we note that (12) is equivalent to

$$\begin{aligned} & f_\alpha(\lambda_l(\mathbf{L})) \Im(\lambda_r(\mathbf{W})) - f_\alpha(\lambda_b(\mathbf{L})) \Im(\lambda_a(\mathbf{W})) \\ & > \frac{h}{2} \left(f_\alpha(\lambda_l(\mathbf{L}))^2 |\lambda_r(\mathbf{W})|^2 - f_\alpha(\lambda_b(\mathbf{L}))^2 |\lambda_a(\mathbf{W})|^2 \right) \end{aligned}$$

for all r, l . Denote by ε the gap

$$0 < \varepsilon := \min_{(r,l) \neq (a,b)} \{f_\alpha(\lambda_l(\mathbf{L})) \Im(\lambda_r(\mathbf{W})) - f_\alpha(\lambda_b(\mathbf{L})) \Im(\lambda_a(\mathbf{W}))\}.$$

Noting that

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

one gets that (12) is satisfied for $h < \varepsilon \|\mathbf{W}\|^{-2}$. Therefore, for sufficiently small h , the dominating frequencies are the ones with minimal imaginary part, i.e., either $f_\alpha(\lambda_1(\mathbf{L})) \Im(\lambda_K(\mathbf{W}))$ or $f_\alpha(\lambda_N(\mathbf{L})) \Im(\lambda_1(\mathbf{W}))$. If $f_\alpha(\lambda_1(\mathbf{L})) \Im(\lambda_K(\mathbf{W})) < f_\alpha(\lambda_N(\mathbf{L})) \Im(\lambda_1(\mathbf{W}))$, then $b = 1$, and the normalized $\text{vec}(\mathbf{x})$ converges to the eigenvector corresponding to the smallest frequency $\lambda_1(\mathbf{L})$. By (10), this is also the eigenvector of $\mathbf{I} \otimes \mathbf{I} - \mathbf{L}$ corresponding to the largest frequency $1 - \lambda_1(\mathbf{L}) = \lambda_N(\mathbf{I} - \mathbf{L})$. An application of Lemma D.2 finishes the proof. \square

Finally, we present a similar result for negative powers.

Proposition D.12. *Let $h \in \mathbb{R}$, $h > 0$. Consider the fractional Schrödinger equation (3) with $\alpha < 0$. Let $\{\mathbf{x}(nh)\}_{n \in \mathbb{N}}$ be the trajectory of vectors derived by approximating the solution of (3) 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*

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

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

Proof. Similar to Proposition D.11, we can prove the statement by realizing that the dominating frequencies (λ_a, λ_b) in (12) 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 (6) 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 5.6.

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 $\{\psi_l^i(\mathbf{M})\}_{i \in \{1, \dots, k_l\}}$ 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 E.1. *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 $\mathbf{x} : [0, T] \rightarrow \mathbb{R}^n$ be a solution to*

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

Then, \mathbf{x} is given by

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

where

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

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

E.1 Proof of Theorem 5.6

In the following, we reformulate and prove Theorem 5.6.

Corollary E.4. *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 (2) 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 (2) 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 E.3 and its proof, we can write the solution of the vectorized form of (2) as

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

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

$$\begin{aligned} & \sum_{l_1=1}^K \sum_{l_2=1}^m \exp(-\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L})t) \sum_{i=1}^{k_{l_2}} c_{l_1, l_2}^i \sum_{j=1}^i \frac{t^{i-j}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \\ &= c_{K,1}^{k_1} \exp(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})) \frac{t^{k_1-1}}{(k_1-1)!} (\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) \\ &+ c_{K,1}^{k_1} \exp(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})) \sum_{j=2}^{k_1} \frac{t^{k_1-j}}{(k_1-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\ &+ \exp(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})) \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^i \frac{t^{i-j}}{(i-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\ &+ \sum_{l_1=1}^K \sum_{l_2=2}^m \exp(-\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L})t) \sum_{i=1}^{k_{l_2}} c_{l_1, l_2}^i \sum_{j=1}^i \frac{t^{i-j}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \\ &= \exp(-t\lambda_K(\mathbf{W})\lambda_1(\mathbf{L})) t^{k_1-1} \left(c_{K,1}^{k_1} \frac{1}{(k_1-1)!} (\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) \right. \\ &+ c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\ &+ \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^i \frac{1}{(i-j)!} t^{i-j-k_1+1} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\ &+ \sum_{l_1=1}^K \sum_{l_2=2}^m \exp(-t(\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))) \sum_{i=1}^{k_{l_2}} c_{l_1, l_2}^i \\ &\cdot \left. \sum_{j=1}^i \frac{t^{i-j-k_1+1}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \right). \end{aligned}$$

We can then write the normalized solution as

$$\begin{aligned}
& \left(\frac{c_{K,1}^{k_1}}{(k_1-1)!} (\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) + c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \right. \\
& + \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^i \frac{t^{i-j-k_1+1}}{(i-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\
& + \sum_{l_1=1}^K \sum_{l_2=2}^m e^{-t(\lambda_K(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))} \sum_{i=1}^{k_{l_2}} c_{l_1, l_2}^i \sum_{j=1}^i \frac{t^{i-j-k_1}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \left. \right) \\
& \cdot \left\| \frac{c_{K,1}^{k_1}}{(k_1-1)!} (\mathbf{e}_K \otimes \psi_1^1(\mathbf{L})) + c_{K,1}^{k_1} \sum_{j=2}^{k_1} \frac{t^{1-j}}{(k_1-j)!} (\lambda_K(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \right. \\
& + \sum_{i=1}^{k_1-1} c_{K,1}^i \sum_{j=1}^i \frac{t^{i-j-k_1+1}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_K \otimes \psi_1^j(\mathbf{L})) \\
& + \sum_{l_1=1}^K \sum_{l_2=2}^m \exp(-t(\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}) - \lambda_K(\mathbf{W})\lambda_1(\mathbf{L}))) \\
& \cdot \left. \sum_{i=1}^{k_{l_2}} c_{l_1, l_2}^i \sum_{j=1}^i \frac{t^{i-j-k_1}}{(i-j)!} (\lambda_{l_1}(\mathbf{W}))^{1-j} (\mathbf{e}_{l_1} \otimes \psi_{l_2}^j(\mathbf{L})) \right\|_2^{-1}.
\end{aligned}$$

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

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

We apply Lemma D.2 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 $1 - \lambda_N(\mathbf{L}) = \lambda_1(\mathbf{I} - \mathbf{L})$. \square

Remark E.5. *If the hypotheses are met, the convergence result also holds for \mathbf{L}^α . With the same reasoning, we can prove that the normalized solution converges to the eigenvector corresponding to the eigenvalue of \mathbf{L}^α with minimal real part. It suffices to consider the eigenvalues and generalized eigenvectors of \mathbf{L}^α . However, we do not know the relationship between the singular values of \mathbf{L}^α , where we defined the fractional Laplacian, and the eigenvalues of \mathbf{L} . Hence, it is much more challenging 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 Theorem 5.6 are also satisfied when (2) is approximated via an explicit Euler scheme.

As noted in (11), the vectorized solution to (2) can be written as

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

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 $\{\mathbf{e}_l^i : i \in \{1, \dots, k_l\}, l \in \{1, \dots, m\}\}$ 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 $l \in \{1, \dots, m\}$, and compute

$$\begin{aligned} \mathbf{P}\mathbf{J}_l^n \left(\sum_{i=1}^{k_l} c_l^i \mathbf{e}_l^i \right) &= \mathbf{P} \left(\lambda_l(\mathbf{M})^n c_l^1 \mathbf{e}_l^1 \right) + \mathbf{P} \left(\binom{n}{1} \lambda_l(\mathbf{M})^{n-1} c_l^1 \mathbf{e}_l^1 + \lambda_l(\mathbf{M})^n c_l^2 \mathbf{e}_l^2 \right) \\ &\quad + \mathbf{P} \left(\binom{n}{2} \lambda_l(\mathbf{M})^{n-2} c_l^1 \mathbf{e}_l^1 + \binom{n}{1} \lambda_l(\mathbf{L})^{n-1} c_l^2 \mathbf{e}_l^2 + \lambda_l(\mathbf{M})^n c_l^3 \mathbf{e}_l^3 \right) \\ &\quad + \dots \end{aligned}$$

We can summarize our findings in the following lemma.

Lemma E.6. 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 E.7. Let \mathcal{G} be a strongly connected directed graph with $\mathbf{L} \in \mathbb{R}^{N \times N}$. Consider the initial value problem in (2) with diagonal channel mixing matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ and $\alpha = 1$. Approximate the solution to (2) 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

$$\lambda_K(\mathbf{W}) \Re \lambda_1(\mathbf{L}) < \lambda_1(\mathbf{W}) \Re \lambda_N(\mathbf{L}), \quad (15)$$

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

Proof. As noted in (11), the vectorized solution to (2) with $\alpha = 1$, can be written as

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

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 E.2. Then,

$$\begin{aligned} \text{vec}(\mathbf{x})(nh) &= (\mathbf{I} + h\mathbf{W} \otimes (\mathbf{P}\mathbf{J}\mathbf{P}^{-1}))^n \text{vec}(\mathbf{x}_0) \\ &= (\mathbf{I} \otimes \mathbf{P})(\mathbf{I} - h\mathbf{W} \otimes \mathbf{J})^n (\mathbf{I} \otimes \mathbf{P})^{-1} \text{vec}(\mathbf{x}_0) \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{I} \otimes \mathbf{P})(\mathbf{I} - h\tilde{\mathbf{P}}\tilde{\mathbf{J}}\tilde{\mathbf{P}}^{-1})^n(\mathbf{I} \otimes \mathbf{P})^{-1}\text{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}}(\mathbf{I} - h\tilde{\mathbf{J}})^n\tilde{\mathbf{P}}^{-1}(\mathbf{I} \otimes \mathbf{P})^{-1}\text{vec}(\mathbf{x}_0) \\
&= (\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}}(\mathbf{I} - h\tilde{\mathbf{J}})^n((\mathbf{I} \otimes \mathbf{P})\tilde{\mathbf{P}})^{-1}\text{vec}(\mathbf{x}_0).
\end{aligned}$$

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

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

Then, by Lemma E.6, we have

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

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

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

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

$$\begin{aligned}
&\sum_{l_1=1}^K \sum_{l_2=1}^m \sum_{i=1}^{\min\{k_{l_2}, n-1\}} \sum_{j=1}^i \binom{n}{i-j} (1 - h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{n-i+j} c_{l_1, l_2}^j \psi_{l_1}(\mathbf{W}) \otimes \psi_{l_2}^j(\mathbf{L}) \\
&= (1 - h\lambda_{L_1}(\mathbf{W})\lambda_{L_2}(\mathbf{L}))^n \\
&\quad \cdot \sum_{l_1=1}^K \sum_{l_2=1}^m \sum_{i=1}^{\min\{k_{l_2}, t-1\}} \sum_{j=1}^i \binom{n}{i-j} \frac{(1 - h\lambda_{l_1}(\mathbf{W})\lambda_{l_2}(\mathbf{L}))^{n-i+j}}{(1 - h\lambda_{L_1}(\mathbf{W})\lambda_{L_2}(\mathbf{L}))^n} c_{l_1, l_2}^j \psi_{l_1}(\mathbf{W}) \otimes \psi_{l_2}^j(\mathbf{L}).
\end{aligned} \tag{16}$$

With a similar argument as in the proof of Theorem 5.6, we can then see that

$$\frac{\text{vec}(\mathbf{x})(nh)}{\|\text{vec}(\mathbf{x})(nh)\|_2} \xrightarrow{t \rightarrow \infty} \frac{c_{L_1, L_2}^1 \psi_{L_1}(\mathbf{W}) \otimes \psi_{L_2}^1(\mathbf{L})}{\|c_{L_1, L_2}^1 \psi_{L_1}(\mathbf{W}) \otimes \psi_{L_2}^1(\mathbf{L})\|_2},$$

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 $\mathbf{I} - \mathbf{L}$ corresponding to the eigenvalue $1 - \lambda_{L_2}(\mathbf{L})$. By Lemma D.2, 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 (15) 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., (15) 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 $\mathbf{I} - \mathbf{L}$, we get HFD dynamics. Similarly, 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. \square

E.3 GCN oversmooths

Proposition E.8. 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 E.7. The difference is that instead of (16), 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 E.7, we can show that the corresponding GCN is $1 - \lambda_N(\mathbf{L})$ -FD. Now $1 - \lambda_N(\mathbf{L}) = \lambda_1(\mathbf{I} - \mathbf{L})$, and $\lambda_1(\mathbf{I} - \mathbf{L})$ -FD corresponds to LFD, hence the GCN oversmooths. \square

F 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

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

associated to the eigenvalue $\lambda_j = \cos(2\pi j/N)$. First, we can note that $\lambda_j = \lambda_{N-j}$ for all $j \in \{1, \dots, N/2\}$; 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

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

are two eigenvectors of the same eigenvalue λ_j . To 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 $n \notin \{0, N/2\}$

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

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

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

which is equivalent to

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

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

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

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

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

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

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

Their sum is then equivalent to

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

hence, the mid eigenvector of \mathbf{L} guarantees homophily $1/2$. A visual explanation is shown in Figure 4.