Generalized Dirichlet Energy and Graph Laplacians for Clustering Directed and Undirected Graphs

Anonymous authors
Paper under double-blind review

Abstract

Clustering in directed graphs remains a fundamental challenge due to the asymmetry in edge connectivity, which limits the applicability of classical spectral methods originally designed for undirected graphs. A common workaround is to symmetrize the adjacency matrix, but this often leads to losing critical directional information. In this work, we introduce the *generalized Dirichlet energy* (GDE), a novel energy functional that extends the classical Dirichlet energy to handle arbitrary positive vertex measures and Markov transition matrices. GDE provides a unified framework applicable to both directed and undirected graphs, and is closely tied to the diffusion dynamics of random walks. Building on this framework, we propose the *generalized spectral clustering* (GSC) method that enables the principled clustering of weakly connected digraphs without resorting to the introduction of teleportation to the random walk transition matrix. A key component of our approach is the utilization of a parametrized vertex measure encoding graph directionality and density. Experiments on real-world point-cloud and network datasets show that GSC consistently matches or outperforms existing spectral clustering methods in both clustering accuracy and robustness, making it a strong tool for graph-based data analysis.

Keywords: Dirichlet energy, random walks, graph Laplacian, parametrized graph operators, graph directionality, point-clouds, spectral clustering, graph partitioning.

1 Introduction

Clustering is a fundamental tool broadly used to uncover structure in large and complex datasets. In many modern applications, such as web graphs, citation networks, and information diffusion in social or transportation systems, a dataset is naturally represented as a directed graph (digraph), where edge direction encodes asymmetric relationships between data objects. Among the several graph clustering methods, spectral clustering has become popular due to its conceptual simplicity, efficiency, and strong theoretical underpinnings (Ng et al., 2002; Peng et al., 2015; Boedihardjo et al., 2021). A foundational concept in spectral clustering is the *Dirichlet energy*, which quantifies the smoothness of a function over its domain. For a graph function, defined over the vertices of a graph, this measures the variability of the function across adjacent vertices, hence capturing local regularity with respect to graph connectivity. This notion underlies many successful clustering algorithms that have appeared in machine learning, network science, and other domains.

With data represented as a graph, graph clustering aims at partitioning that graph into subsets with high internal connectivity and sparse external links. This task is typically formulated as a discrete optimization problem, such as minimizing the graph cut or normalized cut, whose relaxed form corresponds to minimizing the Dirichlet energy of a graph function associated with a Laplacian operator (Von Luxburg, 2007; Shi & Malik, 2000). Spectral clustering solves this relaxed problem by computing the leading eigenvectors of the Laplacian, yielding low-dimensional embeddings that can be clustered via standard methods such as k-means.

Most spectral clustering methods are designed for undirected graphs. Many real-world systems, though, such as social and content networks or flow-based systems like transportation, are best modeled as digraphs carrying critical information in edge directionality. A widely adopted practice in handling weakly connected digraphs in graph machine learning tasks, such as spectral clustering (Zhou et al., 2005) or vertex classification

(Peach et al., 2020), is the routine reliance on the teleporting random walk (Page et al., 1999). This process guarantees ergodicity, hence it is often used as a substitute for the natural random walk that may not be ergodic in directed graphs. Despite offering an alignment with classical ergodic-based theoretical frameworks, the teleporting random walk constitutes a convenient workaround rather than a principled modeling decision. Some of the drawbacks it introduces are: First, it adds uniform transitions to all vertices, as it blends the original graph structure with that of a complete graph. This alters its topology and injects non-local interactions, potentially masking critical structural information of the original graph. Imposing ergodicity in cases where it does not naturally hold can lead to misleading conclusions about the true graph dynamics (Schaub et al., 2019). Moreover, by densifying the graph it increases substantially the computational cost in practice. Aside teleportation, other attempts have been made to addressing the lack of ergodicity. Zhou et al. (2005) build on the directed Laplacian of Chung (2005) to develop a spectral clustering method tailored to digraphs. Meilă & Pentney (2007) directly leverage the asymmetric adjacency matrix within a weighted cut framework. More recently, Rohe et al. (2016) proposed a spectral co-clustering approach based on a singular value decomposition, capturing directional information through low-rank approximations.

Point-cloud data commonly generate digraphs via K-nearest neighbor (K-NN) graphs, which are asymmetric by construction (Genton, 2002). Symmetrizing the adjacency matrix before clustering is standard (Satuluri & Parthasarathy, 2011), but this obscures directional patterns and can degrade performance.

A rather different line of work focuses on *flow-based clustering* (Cucuringu et al., 2020; Laenen & Sun, 2020; Coste & Stephan, 2021; Hayashi et al., 2022), aiming at identifying *imbalanced cuts*, i.e. partitions where most edges flow in one direction between clusters. This is particularly suited to domains with strongly asymmetric connectivity, such as migration networks (Cucuringu et al., 2020), food webs, or trade flows (Laenen & Sun, 2020). However, they are distinct from the classical *density-based* clustering, which seeks rather size-balanced clusters with high internal cohesion. More precisely, density-based clustering favors dense subgraphs and penalizes having low-degree or weakly connected vertices, while flow-imbalanced cuts look for sparse boundaries where only a small amount of 'flow' (edges) connects two subgraphs, regardless of how dense each side is internally.

In this work, we revisit the density-based clustering paradigm and we propose a novel principled framework that generalizes to both directed and undirected graphs. Our contributions are as follows:

- ***Generalized Dirichlet energy* (GDE):** We introduce a framework based on a novel energy functional that extends the classical Dirichlet energy (Montenegro et al., 2006). Unlike prior formulations, which are limited to ergodic random walks with specific transition matrices and stationary distributions, GDE is defined for *any* positive vertex measure and Markov transition matrix, without requiring reversibility or ergodicity. GDE stems naturally from a random walk interpretation of graph partitioning, and mitigates the issues appearing when the probability mass of the stationary distribution is concentrated unevenly across the graph. We also propose a parametrized vertex measure that encodes both directionality and edge density.
- ***Generalized spectral clustering* (GSC):** Building on the GDE, we develop a spectral clustering method that applies to both directed and undirected graphs. Unlike earlier spectral approaches that are limited to strongly connected digraphs (Zhou et al., 2005; Palmer & Zheng, 2020), GSC applies to *weakly connected* digraphs without relying on teleporting random walks, such as those used in PageRank (Page et al., 1999). This leads to a principled and flexible clustering method suited for a wide variety of graph-structured data.

2 Preliminaries and background

2.1 Essential concepts

Notation. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ be a weighted directed graph (digraph), where $\mathcal{V}$ is a finite vertex set of size $N = |\mathcal{V}|$, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of directed edges. Each edge (i, j) denotes a directed link from vertex i to vertex j. Digraphs may be given directly or constructed from input point-clouds, i.e. from datapoints $\boldsymbol{X} = \{x_i\}_{i=1}^N$, with $x_i \in \mathbb{R}^d$, using appropriate graph construction techniques. We assume that $\mathcal{G}$ is a *weakly connected* digraph: that is, when the edge directionality is ignored, the obtained graph is connected.

The edge weight function $w : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ assigns a nonnegative value to each vertex pair: $w(i,j) > 0$ if $(i,j) \in \mathcal{E}$, and 0 otherwise. Let $\mathbf{W} = \{w_{ij}\}_{i,j=1}^N \in \mathbb{R}_+^{N\times N}$ be the adjacency matrix, where $w_{ij} = w(i,j)$. Define the out-degree and in-degree of vertex i as $d_{\text{out}}(i) = \sum_j w_{ij}$ and $d_{\text{in}}(i) = \sum_j w_{ji}$, respectively. For a vector ν, define $\mathbf{D}_\nu = \text{diag}(\nu)$. For a subset $S \subseteq \mathcal{V}$, its complement is $\bar{S} = \mathcal{V} \setminus S$, and its characteristic function is $\chi_S \in \{0,1\}^N$, where $\chi_S(i) = 1$ iff $i \in S$. For a singleton set $S = \{v\}$, we use $\delta_v \in \{0,1\}^N$, the Kronecker delta vector. Also, $\mathbf{1}_{N\times M}$ is the $N \times M$ all-ones matrix, and $\mathbb{1}\{\cdot\}$ is the indicator function.

A *positive vertex measure* $\nu : \mathcal{V} \to \mathbb{R}_+^*$ assigns nonnegative values to vertices; when $\sum_{i\in\mathcal{V}} \nu(i) = 1$, it is a *probability vertex measure*. A function $f : \mathcal{V} \to \mathbb{R}$ is represented as $f = [f(i)]_{i\in\mathcal{V}}^\mathsf{T} \in \mathbb{R}^N$. We assume graph functions belong to $\ell^2(\mathcal{V}, \nu)$, a Hilbert space with inner product:

$$\langle f, g\rangle_\nu = f^\mathsf{T}\mathbf{D}_\nu g = \sum_{i\in\mathcal{V}} \nu(i)f(i)g(i), \tag{1}$$

which reduces to the standard dot product $\langle f, g\rangle = f^\mathsf{T} g$ when $\nu = \mathbf{1}_{N\times 1}$. Edge-based functions $q : \mathcal{V}\times\mathcal{V} \to \mathbb{R}_+$, constrained to $q(i,j) = 0$ for $(i,j) \notin \mathcal{E}$, are called *positive edge measures*.

Random walk fundamentals. A natural random walk on $\mathcal{G}$ is defined by a homogeneous Markov chain $\mathcal{X} = (X_t)_{t\geq 0}$ on a state space $\mathcal{V}$, with transition probabilities:

$$p(i,j) = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \frac{w(i,j)}{\sum_z w(i,z)}. \tag{2}$$

The transition matrix $\mathbf{P} = [p(i,j)] \in \mathbb{R}^{N\times N}$ is row-stochastic, and equals: $\mathbf{P} = \mathbf{D}_{\text{out}}^{-1}\mathbf{W}$. Its spectrum lies within the unit disk, i.e. all its eigenvalues have modulus at most 1. A stationary distribution π is a probability vertex measure satisfying $\mathbf{P}^\mathsf{T}\pi = \pi$. Note that π is not necessarily unique, and may not be strictly positive (i.e. $\text{supp}(\pi) \subset \mathcal{V}$). In the case it is not unique, we simply say that it is *stationary*, while when it is unique and strictly positive (i.e. $\text{supp}(\pi) = \mathcal{V}$), we say that the random walk is *ergodic* (Brémaud, 2013; Seabrook & Wiskott, 2023). Ergodicity is verified if $\mathcal{G}$ is strongly connected and aperiodic, but it can also hold in other cases. In this case the distribution $p_t(i,\cdot) = \delta_i^\mathsf{T}\mathbf{P}^t$ converges to a unique stationary distribution $\pi \in \mathbb{R}_+^N$ as $t \to \infty$. For undirected graphs or reversible walks, the stationary distribution satisfies $\pi(i) \propto d(i)$, the degree of vertex i. Reversibility means $\pi(i)p(i,j) = \pi(j)p(j,i)$ for all i,j. This condition always holds in the undirected setting, where $d_{\text{out}}(i) = d_{\text{in}}(i) = d(i)$.

Dirichlet energy and graph Laplacians. The *Dirichlet energy* quantifies the global smoothness of a graph function, penalizing variation across edges. It is central in Dirichlet form theory (Saloff-Coste, 1997; Montenegro et al., 2006), graph signal processing (Shuman et al., 2013), and harmonic analysis on graphs (Sevi et al., 2023). Next, we define the random-walk based definition of the Dirichlet energy.

Definition 2.1 (**Dirichlet energy of a graph function based on random-walks**). *Let $\mathcal{X}$ be a random walk on digraph $\mathcal{G}$ with transition matrix $\mathbf{P}$ and ergodic distribution π* [1]*. Then the Dirichlet energy of a function $f : \mathcal{V} \to \mathbb{R}$ is:*

$$\mathcal{D}^2(f) = \sum_{i,j\in\mathcal{V}} \pi(i)p(i,j)[f(i) - f(j)]^2 \tag{3}$$

$$= 2\,\langle f, \mathbf{L}_{\text{RW}} f\rangle_\pi = 2\,\langle f, \mathbf{L} f\rangle. \tag{4}$$

Here, Eq. 3 encourages smooth functions over high-probability transitions. A feature of the random-walk-based formulation that we exaploit later is that $\pi(i)$ reveals a global importance weight of vertex i. Eq. 4 links the Dirichlet energy to the Laplacian matrices $\mathbf{L}_{\text{RW}}$, $\mathbf{L}$, $\overline{\mathbf{L}}$ (Chung, 2005; Sevi et al., 2023), which are fundamental to spectral methods.

[1] In the case of non-ergodic digraphs, the energy can be defined using stationary distribution.

Definition 2.2 (**Directed Laplacians**). *Given a transition matrix* $\mathbf{P}$ *and its stationary distribution* π*, the directed Laplacians are defined as follows:*

$$\textit{Random walk Laplacian:} \quad \mathbf{L}_{\mathrm{RW}} = \mathbf{I} - \tfrac{1}{2}(\mathbf{P} + \mathbf{D}_\pi^{-1}\mathbf{P}^\mathsf{T}\mathbf{D}_\pi) \tag{5}$$

$$\textit{Unnormalized Laplacian:} \quad \mathbf{L} = \mathbf{D}_\pi - \tfrac{1}{2}(\mathbf{D}_\pi\mathbf{P} + \mathbf{P}^\mathsf{T}\mathbf{D}_\pi) \tag{6}$$

$$\textit{Normalized Laplacian:} \quad \overline{\mathbf{L}} = \mathbf{D}_\pi^{-1/2}\mathbf{L}\mathbf{D}_\pi^{-1/2}. \tag{7}$$

In the undirected case, $\mathbf{D}_\pi \propto \mathbf{D}_d$ and $\mathbf{P} = \mathbf{D}_d^{-1}\mathbf{W}$, implying:

In the particular case of a balanced digraph, where each node has equal in- and out-degree while generally $w(i,j) \neq w(j,i)$, we have that $\pi(i) \propto d_{\mathrm{out}}(i)$, thus the directed Laplacians reduce to the symmetric Laplacian of the symmetrized adjacency matrix (Seabrook & Wiskott, 2023):

$$\mathbf{L} = \mathbf{D}_{\mathrm{out}} - \tfrac{1}{2}(\mathbf{W} + \mathbf{W}^\mathsf{T}). \tag{8}$$

This is a natural extension of the undirected Laplacians to this setting. When the underlying graph is undirected, as it holds $\mathbf{W} = \mathbf{W}^\mathsf{T}$, we recover the classical undirected Laplacians $\mathbf{L} = \mathbf{D}_{\mathrm{out}} - \mathbf{W}$. In general, though, the directed Laplacians differ from their undirected counterparts and they cannot be obtained by simply symmetrizing the adjacency matrix.

2.2 Connection with spectral clustering

The Dirichlet energy yields a Laplacian operator, which is also a link to spectral methods. In spectral clustering, vertex embeddings are obtained from the eigenvectors of a chosen Laplacian. To formalize this connection, we recall the Courant-Fischer min-max theorem (Horn & Johnson, 1990) from spectral theory. For a Hermitian matrix $\mathbf{L}$ with eigenvalues $\lambda_1 \geq ... \geq \lambda_N$, the Rayleigh quotient:

$$\mathcal{R}_{\mathbf{L}}(y) = \frac{y^\mathsf{T}\mathbf{L}y}{y^\mathsf{T}y} = \frac{\langle y, \mathbf{L}y\rangle}{\|y\|^2}, \quad y \in \mathbb{R}^N \setminus \{0_N\}, \tag{9}$$

characterizes the eigenvalues via:

$$\lambda_k = \max_{\substack{M \subset \mathbb{R}^N \\ \dim(M) = N-k+1}} \min_{y \in M} \mathcal{R}_{\mathbf{L}}(y). \tag{10}$$

The k-th eigenvector, v_k, minimizes $\mathcal{R}_{\mathbf{L}}$, subject to orthogonality:

$$\lambda_k = \min_{y \in \mathbb{R}^N} \mathcal{R}_{\mathbf{L}}(y), \quad v_k = \arg\min_{y \in \mathbb{R}^N} \mathcal{R}_{\mathbf{L}}(y) \tag{11}$$

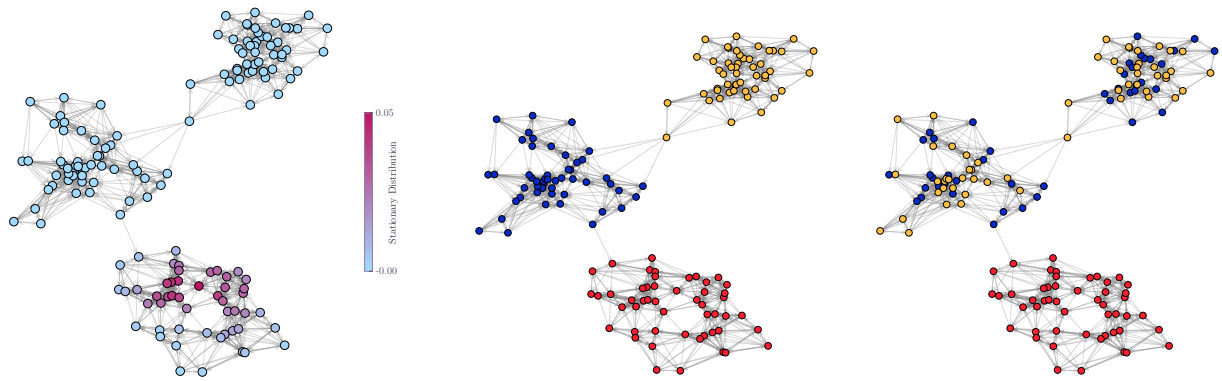
$$\text{subject to } \langle v_j, y\rangle = 0, \text{ for } j = 1, ..., k-1. \tag{12}$$

Remark 2.1. *The Dirichlet energy and the Rayleigh quotient are proportional:*

$$\overline{\mathcal{D}}^2(f) = 2\,\mathcal{R}_{\mathbf{L}}(f), \quad \Rightarrow \quad \min_{f \in \mathbb{R}^N} \mathcal{R}_{\mathbf{L}}(f) \equiv \min_{f \in \mathbb{R}^N} \overline{\mathcal{D}}^2(f), \tag{13}$$

where $\overline{\mathcal{D}}^2(f) = \frac{\mathcal{D}^2(f)}{\|f\|^2}$ *is the* normalized Dirichlet energy.

Therefore, minimizing the Rayleigh quotient is equivalent to minimizing the Dirichlet energy. As minimizing the Dirichlet energy promotes smooth functions over a graph, spectral clustering assigns similar values to well-connected vertices. In particular, for two clusters ($k = 2$), one may use a binary function of the form $f = \chi_{V_1}$, and here more specifically this is the characteristic function for the set V_1. After relaxation, i.e. allowing f to be a non-binary function, minimizing the Dirichlet energy is equivalent to finding the second eigenvector (Fiedler vector) of $\mathbf{L}$. The second eigenvector can then be used to find a 2-partition of the graph, as it minimizes the Dirichlet energy while being orthogonal to the constant vector. This generalizes naturally to multi-way partitioning (see Sec. 4).



(a) Stationary distribution π

(b) Two partitions minimizing the classical Dirichlet energy

Figure 1: **Weak connectivity and the Dirichlet energy.** a) The heterogeneous stationary distribution π in an unweighted digraph with three clusters, where the bottom one is a 'sink' (i.e. it has only incoming edges). In this case, the support of π does not cover all the vertex set. b) Two partitions f, g, both minimizers of the classical Dirichlet energy, i.e. $\mathcal{D}^2(f) = \mathcal{D}^2(g)$. The left one captures the cluster structure, while the other mixes the clusters that are attributed with zero mass of the stationary distribution, which is the reason why the Dirichlet energy is insensitive to function variations in those areas.

3 The generalized Dirichlet energy framework

While the classical Dirichlet energy is a powerful tool for graph-based learning, it has limitations, notably when applied to digraphs, especially those being weakly connected. In this section, we first discuss these limitations and then we introduce the *generalized Dirichlet energy* (GDE), which extends the classical formulation to accommodate arbitrary positive vertex measures and Markov transition matrices. GDE allows capturing a wider range of graph structures and dynamics, particularly in settings where the stationary distribution of a random walk may not be well-defined or may exhibit undesirable properties. We show that GDE gives rise to a family of self-adjoint generalized graph Laplacians, which can be used to develop spectral clustering methods that are applicable to both directed and undirected graphs, without relying on the ergodicity of the underlying random walk.

3.1 Motivation and intuition

In the classical formulation of graph segmentation based on random walks, the Dirichlet energy is defined with respect to the stationary distribution π of the random walk. This approach works well for undirected graphs or strongly connected digraphs where π is well-defined and captures the long-term behavior of the walk. However, in weakly connected digraphs, the random walk is not ergodic, and π cannot be properly defined. In this context, ergodic means that there is a *unique* stationary distribution with a full support over the vertex set, i.e. $\text{supp}(\pi) = \mathcal{V}$ and thus having positive weight for all vertices. Note however that one (or more) stationary distribution with partial support over a vertex subset, i.e. $\text{supp}(\pi) = \mathcal{V}' \subset \mathcal{V}$, might still exist (Seabrook & Wiskott, 2023). Such stationary distributions break the classical Dirichlet energy framework, as they would zero out the energy contribution of vertices outside the support $\mathcal{V}\backslash\mathcal{V}'$.

Fig. 1 shows an example of a weakly connected digraph featuring a stationary distribution with partial support. Any graph functions f and g such that $f(i) = g(i)$ for all $i \in \mathcal{V}'$ and $f(i) \neq g(i)$ for some $i \in \mathcal{V}\backslash\mathcal{V}'$ would lead to $\mathcal{D}^2(f) = \mathcal{D}^2(g)$, since the energy contribution of the vertices in $\mathcal{V}\backslash\mathcal{V}'$ is zero. In the context of clustering, this suggests that partitions assigning different cluster labels to vertices in $\mathcal{V}\backslash\mathcal{V}'$ would be indistinguishable when looking at the Dirichlet energy, regardless of how well they match with the underlying cluster structure of the graph. For instance, in Fig. 1b, both partitions minimize the Dirichlet energy, but only the left one captures the natural cluster structure of the graph, while the right one mixes clusters that are attributed with zero mass of the stationary distribution.

Similar concerns arise even for undirected graphs in the presence of degree heterogeneity, since high-degree vertices can dominate the energy landscape by overshadowing low-degree vertices, and thus lead to biased clustering results. Indeed, the Dirichlet energy is a sum of vertex-wise contributions weighted by π:

$$\mathcal{D}^2(f) = \sum_{i,j\in\mathcal{V}} \pi(i)p(i,j)[f(i)-f(j)]^2 = \sum_{i\in\mathcal{V}} \pi(i) \sum_{j\in\mathcal{V}} p(i,j)[f(i)-f(j)]^2. \tag{14}$$

This tells us that, even if sup $= \mathcal{V}$, when minimizing the Dirichlet energy, a higher $\pi(i)$ induces a higher penalization of function variation at vertex i. Therefore, the distribution of the mass of π across the graph may heavily affect which parts of the cluster structure are easier to detect. For instance, the high concentration of the mass at few individual vertices, or specific clusters, would make the Dirichlet energy more sensitive to function variations at those regions, while hindering the correct detection of the cluster structure in the rest of the graph. Intuitively, π carries global information about the significance of each vertex for the graph, however, when π exhibits high heterogeneity it is reasonable to seek instead a 'smoother' surrogate of that measure with more flat shape; this aspect is developed further in Sec. 3.3.

3.2 Generalized Dirichlet energy and graph Laplacians

We now formalize the connection between Dirichlet energy and graph Laplacians through the introduction of the *generalized Dirichlet energy* (GDE). This functional is defined with respect to a general positive edge measure q, and naturally leads to a broader class of graph Laplacians. We begin with the general definition of GDE, then specialize to cases where q is derived from a vertex measure ν and a transition matrix $\mathbf{P}$. The derived family of Laplacian operators adapts to the graph structure as seen through the random walk dynamics. Although our focus is spectral clustering, the operators defined through the GDE framework can be of use in other graph-based learning problems.

Definition 3.1 (**Generalized Dirichlet Energy of a graph function**). *Let $q : \mathcal{V}\times\mathcal{V} \to \mathbb{R}_+$ be a positive edge measure on a digraph $\mathcal{G}$, and let $\mathbf{Q} = [q(i,j)]_{i,j\in\mathcal{V}}$ denote the corresponding matrix. The* generalized Dirichlet energy *of a function $f : \mathcal{V} \to \mathbb{R}$ is given by:*

$$\mathcal{D}^2_{\mathbf{Q}}(f) = \sum_{i,j\in\mathcal{V}} q(i,j)[f(i)-f(j)]^2. \tag{15}$$

This formulation is generic as the graph structure is captured by the choice of edge measure q. Most typically, the edge weights of the input graph are directly plugged by setting $q(i,j) = w(i,j)$, which –from a minimization point-of-view– essentially encourages smooth functions over edges with high weights. Another instance of the energy functional of Eq. 15 is through the lens of a random walk, by setting $q(i,j) = \pi(i)p(i,j)$:

$$\mathcal{D}^2_{\mathbf{D}_\pi\mathbf{P}}(f) = \sum_{i,j\in\mathcal{V}} \pi(i)p(i,j)[f(i)-f(j)]^2 = \mathcal{D}^2(f), \tag{16}$$

where π is the stationary distribution of a random walk with transition matrix $\mathbf{P}$. This yields the classical random-walk-based Dirichlet energy form ($\mathcal{D}^2(f)$) that we saw earlier in Def. 2.1.

An important feature of the random-walk-based formulation is that it decomposes the edge measure as $q(i,j) = \pi(i)p(i,j)$: $\pi(i)$ expresses the explicit global importance of vertex i as the probability for the random walk to be at that vertex anytime (see Eq. 14 and the discussion in Sec. 3.1), and $p(i,j)$ is a local weight given that the walk is already at vertex i. Relying on this aspect, we use the form $q(i,j) = \nu(i)p(i,j)$, with ν being a tunable positive vertex measure, and obtain the GDE form:

$$\mathcal{D}^2_{\nu}(f) = \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2. \tag{17}$$

This enables the definition of new energy functionals with arbitrary positive vertex measures ν. Since scaling ν simply rescales the energy by $\|\nu\|_1^{-1}$, we may also assume without loss of generality that ν is a probability vertex measure.

Definition 3.2 (**Generalized graph Laplacians**). *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$ with transition matrix $\mathbf{P}$. Let ν be a positive vertex measure, and define the incoming measure $\xi = \mathbf{P}^\mathsf{T}\nu$. Define $\mathbf{D}_{\nu+\xi} = \mathbf{D}_\nu + \mathbf{D}_\xi$. The associated generalized Laplacians are:*

$$\textit{Unnormalized:} \quad \mathbf{L}_\nu = \mathbf{D}_{\nu+\xi} - (\mathbf{D}_\nu \mathbf{P} + \mathbf{P}^\mathsf{T}\mathbf{D}_\nu), \tag{18}$$

$$\textit{Random walk:} \quad \mathbf{L}_{\mathrm{RW},\nu} = \mathbf{D}_{\nu+\xi}^{-1}\mathbf{L}_\nu, \tag{19}$$

$$\textit{Normalized:} \quad \overline{\mathbf{L}}_\nu = \mathbf{D}_{\nu+\xi}^{-1/2}\mathbf{L}_\nu\mathbf{D}_{\nu+\xi}^{-1/2}. \tag{20}$$

Here, $\xi = \mathbf{P}^\mathsf{T}\nu$ corresponds to the Perron–Frobenius operator applied to ν (Ding & Zhou, 2009; Klus & Trower, 2023; Klus & Djurdjevac Conrad, 2023). When $\nu = \pi$, we recover classical Laplacians:

$$\mathbf{L}_\pi = 2\mathbf{D}_\pi - (\mathbf{D}_\pi\mathbf{P} + \mathbf{P}^\mathsf{T}\mathbf{D}_\pi) = 2\mathbf{L}. \tag{21}$$

Proposition 3.1. *Let ν be a positive vertex measure and $\xi = \mathbf{P}^\mathsf{T}\nu$. Then for any graph function f:*

$$\mathcal{D}_\nu^2(f) = \langle f, \mathbf{L}_{\mathrm{RW},\nu} f\rangle_{\nu+\xi} = \langle f, \mathbf{L}_\nu f\rangle.$$

Moreover, the Dirichlet energy can be expressed as follows:

$$\mathcal{D}_\nu^2(f) = \frac{1}{2}\sum_{i,j\in\mathcal{V}} (\nu(i)p(i,j) + \nu(j)p(j,i))\,[f(i) - f(j)]^2.$$

This implies that the generalized Dirichlet energy can be expressed as a quadratic form of the generalized Laplacian $\mathbf{L}_\nu$ in the standard inner product. It also implies that the Laplacians are positive semidefinite and thus admit a real eigenvalue decomposition. This is formalized in the following property.

Proposition 3.2. *The unnormalized and normalized generalized Laplacians, $\mathbf{L}_\nu$ and $\overline{\mathbf{L}}_\nu$, are self-adjoint and positive semi-definite on $\ell^2(\mathcal{V})$. The generalized random walk Laplacian, $\mathbf{L}_{\mathrm{RW},\nu}$, is self-adjoint on $\ell^2(\mathcal{V}, \nu+\xi)$.*

Rayleigh quotient. The normalized generalized Dirichlet energy is given by:

$$\overline{\mathcal{D}}_\nu^2(f) = \frac{\mathcal{D}_\nu^2(f)}{\|f\|^2}. \tag{22}$$

This aligns with the Rayleigh quotient principle, as we have:

$$\mathcal{R}_{\mathbf{L}_\nu}(f) = \frac{\langle f, \mathbf{L}_\nu f\rangle}{\langle f, f\rangle} = \overline{\mathcal{D}}_\nu^2(f).$$

This connection allows us to leverage spectral theory to analyze the properties of the generalized Laplacians, and to develop spectral clustering methods based on the GDE framework.

3.3 Parametrized vertex measure

While the stationary distribution π captures the long-term behavior of a random walk, which summarizes important information about the graph structure, it may not be the best choice for the vertex measure in the Dirichlet energy when π is highly heterogeneous (Sec. 3.1). To address this, we propose vertex measures derived from random walk dynamics that interpolate between the uniform distribution (treating all vertices equally) and the stationary distribution (capturing long-term behavior), providing a more balanced graph representation for clustering.

Vertex measure constructions. In this part, we construct vertex measures, denoted by ν, that capture local and global graph dynamics via random walks. Starting from a uniform measure, we define the following family of measures that depend on the diffusion time t, which generally needs to be short, as well as on the choice of the initializer vertex measure μ:

$$\nu_t = [\mathbf{P}^t]^\mathsf{T}\mu. \tag{23}$$

Proposition 3.3. *Let $\nu_t = [\mathbf{P}^t]^\mathsf{T}\mu$, where $\mathbf{P}$ is ergodic and μ is a probability vertex measure acting as the distribution that initializes the random walk. Then, $\nu_t \xrightarrow{t\to\infty} \pi$ and:*

$$\lim_{t\to\infty} \mathcal{D}^2_{\nu_t}(f) = \mathcal{D}^2_{\pi}(f) = \mathcal{D}^2(f). \tag{24}$$

Note that for any time-dependent vertex measure ν_t such that $\nu_t \xrightarrow{t\to\infty} \nu$, it holds $\mathcal{D}^2_{\nu_t}(f) \xrightarrow{t\to\infty} \mathcal{D}^2_{\nu}(f)$. The original Dirichlet energy (Eq. 3) operates in the stationary regime and uses $\mu = \pi$ as an initializing measure; by the definition of π, this would yield $\nu_t = [\mathbf{P}^t]^\mathsf{T}\pi = \pi$. However, as explained earlier, using π as vertex measure in the energy functional can be problematic when π exhibits high heterogeneity across the graph. To alleviate this problem, we propose the use of the uniform measure $\mu = \frac{1}{N}\mathbf{1}_{N\times 1}$ as initializer to treat fairly all vertices, along with a short random walk (small t) that extracts important structural information about the graph without allowing the ν_t measure to go too close to π. In simple terms, this approach aims at creating a flattened surrogate of the π distribution. The time parameter t controls measure locality and relates closely to mixing times (Levin & Peres, 2017).

On top of the base measure ν_t initialized by $\mu = \mathbf{1}/N$, we can define the following more flexible parametrized measure: given $t \in \mathbb{N}$ and $\alpha \in \mathbb{R}_+$,

$$\nu_{t,\alpha} = \left([\mathbf{P}^t]^\mathsf{T}\tfrac{1}{N}\mathbf{1}_{N\times 1}\right)^{\odot\alpha}, \quad \text{i.e. } \nu_{t,\alpha}(i) = \left([\mathbf{P}^t]^\mathsf{T}\tfrac{1}{N}\mathbf{1}_{N\times 1}\right)^{\alpha}_i. \tag{25}$$

The proposed family of vertex measures $\nu_{t,\alpha}$ is grounded in two key principles. First, the distribution $\nu_{t,\cdot}$ reflects the outcome of a t-step diffusion starting from a uniform distribution. For reversible and ergodic random walks, as $t \to \infty$, $\nu_{t,\alpha}$ converges to the stationary distribution π, enabling a smooth interpolation between the local and the global view of a graph. Second, the exponent α acts as a sharpness parameter, inspired by transformations used in normalized Laplacians and information-theoretic reweightings. Varying α allows the control of the distribution's concentration: $\alpha < 1$ flattens the measure, $\alpha > 1$ sharpens it, and $\alpha = 0$ makes the measure uniform. This formulation offers greater flexibility to adapt to graph heterogeneity. Note that, setting $\alpha \neq 1$ gives a non-probability measure, which however does not affect the essence or validity of the model.

Plugging the proposed $\nu_{t,\alpha}$ into Eq. 17 yields:

$$\mathcal{D}^2_{\nu_{t,\alpha}}(f) = \langle f, \mathbf{L}_{t,\alpha} f\rangle, \tag{26}$$

where $\mathbf{L}_{t,\alpha}$ is the generalized Laplacian associated with the reweighted measure $\nu_{t,\alpha}$. The proposed $\nu_{t,\alpha}$ enables smooth transitions between local/global and sharp/flat configurations via the computation of a family of operators $(\mathbf{L}_{t,\alpha})_{t,\alpha}$. This family of operators supports adaptive, data-driven spectral clustering, with a high focus on flexibility.

Tuning t and α in an unsupervised manner can be challenging, and if not done well may affect the quality of the resulting clustering. When reliable validation criteria are unavailable, a practical alternative is to resort to parameter-free vertex measures that remove this dependency altogether. Natural choices include the degree measure $\nu \propto d_{\text{in}} + d_{\text{out}}$[2], the uniform measure $\nu \propto \mathbf{1}$, or the Perron vector of $\mathbf{P}$ i.e. $\nu = \mathbf{P}^\mathsf{T}\nu$, although, as specified in Sec. 3.1 this measure can be supported only on a subset of vertices. Such alternatives, along with their properties and practical trade-offs, are discussed in App. B.

4 Generalized spectral clustering via GDE minimization

The *generalized spectral clustering* (GSC) method, based on GDE minimization, is applicable to any graph, directed or not. The method stems from the random walk viewpoint of graph partitioning. We start by revisiting the random walk perspective of graph 2-partitioning and we then pass to the k-partitioning.

[2]Due to the constraint $\nu(i) > 0$, the degree measure should include both in- and out-degrees to be well-defined in weakly-connected graphs.

4.1 GDE of a graph 2-partition

For an arbitrary probability vertex measure $\nu : \mathcal{V} \mapsto \mathbb{R}_+$, let $\nu(S)$ be its evaluation over a subset $S \subseteq \mathcal{V}$: $\nu(S) = \sum_{i \in S} \nu(i)$. Let also $q : \mathcal{E} \mapsto \mathbb{R}_+$ be the composite edge measure such that $q(i,j) = \nu(i)p(i,j)$. Respectively, we can define $q(S,U)$ to be the edge measure between two disjoint vertex subsets $S, U \subseteq \mathcal{V}$, $S \cap U = \emptyset$, and generally $S \cup U \neq \mathcal{V}$:

$$\begin{aligned} q(S,U) &= \sum_{i \in S, j \in U} q(i,j) \;=\; \sum_{i \in S, j \in U} \nu(i)p(i,j), \\ &= \sum_{i \in S, j \in U} \mathbb{P}(X_t = i)\mathbb{P}(X_{t+1} = j \,|\, X_t = i), \\ &= \mathbb{P}(X_t \in S, X_{t+1} \in U), \quad \text{for any } t \geq 0. \end{aligned} \tag{27}$$

The generic measure $q(S,U)$ is related to an ergodic Markov chain at the equilibrium, i.e. when $\nu = \pi$ (Sinclair, 1992; Levin & Peres, 2017). We generalize $q(S,U)$ for any transition matrix and vertex measure to quantify the escape probability from S to U in one step, when the starting vertex is drawn from the measure ν. When considering $U = \bar{S}$, this discussion becomes very interesting for graph partitioning. In essence, $q(S,\bar{S})$ offers a probabilistic point of view over the graph cut between the set S and the rest of the graph.

Note that Meilă & Shi (2001) have presented a similar measure in the ergodic setting $\nu = \pi$. Next, we establish the connection between GDE and the edge measure $q(S,\bar{S})$ (the proof is in App. C.4).

Proposition 4.1. *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$. Let ν be a positive vertex measure, and q be its associated positive edge measure, both on $\mathcal{G}$. Let $S \subseteq \mathcal{V}$ and $\bar{S} = \mathcal{V} \backslash S$. Consider the characteristic function χ_S, associated with the set S, as a graph function. The composite edge measure $q(S,\bar{S})$ and the GDE are related as follows:*

$$q(S,\bar{S}) + q(\bar{S},S) = \mathcal{D}^2_\nu(\chi_S). \tag{28}$$

In the case of a 2-partitioning, the function χ_S serves as an indicator (partition) function that separates the vertex set into two disjoint parts, S and $\bar{S}$ (see also Sec. 2.2). Eq. 28 thus provides a natural interpretation of the generalized Dirichlet energy for graph partitions: the quantity $\mathcal{D}^2_\nu(\chi_S)$ measures how easily a random walk with transition probabilities q escapes from S to $\bar{S}$, or vice versa. This energy is symmetric by construction, since $\mathcal{D}^2_\nu(\chi_S) = \mathcal{D}^2_\nu(\chi_{\bar{S}})$. Under mild assumptions[3], one can further show that $\mathcal{D}^2_\nu(\chi_S)$ corresponds to the normalized cut (N-Cut) of a graph with edge weights $q(i,j)$. In this context, the vertex measure ν plays the role of a regularizer in the graph cut objective, shaping the structure of the optimal partition.

When the vertex measure is chosen as the stationary distribution $\nu = \pi$ of the random walk, the following corollary holds (the proof is in App. C.5).

Corollary 4.1 (of Prop. 4.1)**.** *Let $\mathcal{X}$ be an ergodic random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$ and ergodic (stationary) distribution π. For any subset $S \subseteq \mathcal{V}$, let $\bar{S} = \mathcal{V} \setminus S$, and let χ_S be the indicator function of S. Then, the composite edge measure $q(S,\bar{S})$ associated with π and the generalized Dirichlet energy are related as:*

$$q(S,\bar{S}) = \frac{1}{2}\mathcal{D}^2_\pi(\chi_S).$$

This corollary reveals a key insight: minimizing the Dirichlet energy $\mathcal{D}^2_\pi(\chi_S)$ is equivalent to minimizing the total ergodic flow crossing between S and its complement $\bar{S}$. In other words, $\mathcal{D}^2_\pi(\chi_S)$ serves as a normalized cut-like objective under the stationary distribution of the random walk. More generally, in this work we establish the idea that if a probability measure ν is used in the place of π (e.g. a measure that approximates or regularizes π) then minimizing $\mathcal{D}^2_\nu(\chi_S)$ can still be interpreted as finding a well-separated 2-partition of the graph under the dynamics induced by ν.

[3]The graph defined by the transition matrix $q(i,j)$ must be connected and aperiodic (non-bipartite).

Algorithm 1 Generalized Spectral Clustering (GSC)

Input: Weighted adjacency matrix $\mathbf{W}$; number of clusters k; diffusion time t; reweighting exponent α.
Output: $\boldsymbol{V}_{t,\alpha}$: graph partition into k clusters.

1: Compute the generalized Laplacian $\mathbf{L}_{t,\alpha}$ (Eq. 18).
2: Compute $\mathbf{U}_{t,\alpha} \in \mathbb{R}^{N\times k}$, whose columns are the eigenvectors corresponding to the k smallest eigenvalues of $\mathbf{L}_{t,\alpha}$.
3: Embed each vertex i into $\mathbb{R}^k$ using the i-th row of $\mathbf{U}_{t,\alpha}$, and apply a clustering algorithm to the embedded points.
4: Construct the vertex partition $\boldsymbol{V}_{t,\alpha} = \{V_{t,\alpha}^{(\kappa)}\}_{\kappa=1}^k$ based on clustering results.
5: **return** $\boldsymbol{V}_{t,\alpha}$

4.2 GDE minimization for graph k-partitioning

The goal of graph k-partitioning is to divide the vertex set of a digraph into k disjoint subsets such that the edge density across subsets is minimized. Let $\boldsymbol{V} = \{V_\kappa\}_{\kappa=1}^k$ denote such a partition, where each $V_\kappa \subset \mathcal{V}$ is a subset of the vertex set of the graph, and $V_\kappa \cap V_{\kappa'} = \emptyset$ for $\kappa \neq \kappa'$, with $\bigcup_{\kappa=1}^k V_\kappa = \mathcal{V}$. Associated with each subset V_κ, there is a partition function $\chi_{V_\kappa} \in \{0,1\}^N$ that serves as an indicator vector over the vertices. The partition functions define a set of binary signals $\{\chi_{V_\kappa}\}_{\kappa=1}^k$, each highlighting one cluster versus the rest.

We define the *Dirichlet energy of the k-partition* under a vertex measure ν as:

$$\mathcal{D}_\nu^2(\boldsymbol{V}) = \sum_{\kappa=1}^k \mathcal{D}_\nu^2(\chi_{V_\kappa}), \tag{29}$$

where each term measures the generalized Dirichlet energy associated with a single cluster.

Let $\mathbf{U} = [u_1 \ u_2 \ \cdots \ u_k] \in \mathbb{R}^{N\times k}$ be the matrix whose columns are the indicator vectors $u_\kappa = \chi_{V_\kappa}$. The generalized spectral clustering (GSC) method, based on the GDE framework, casts the graph partitioning task as the following optimization problem:

$$\min_{\boldsymbol{V}=\{V_1,\ldots,V_k\}} \mathcal{D}_\nu^2(\boldsymbol{V}) = \min_{\mathbf{U}} \ \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}_\nu\mathbf{U}) \quad \text{s.t.} \ \ u_\kappa = \chi_{V_\kappa}, \ \forall \kappa \in \{1,\ldots,k\}. \tag{30}$$

As with classical spectral clustering, this problem is NP-hard due to the discrete constraints on $\mathbf{U}$ (Von Luxburg, 2007). To make it tractable, we adopt the standard relaxation: instead of binary indicator functions, we allow $\mathbf{U}$ to be any real-valued matrix with orthonormal columns. This leads to the relaxed optimization problem:

$$\min_{\mathbf{U}} \ \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}_\nu\mathbf{U}) \quad \text{s.t.} \ \ \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_k, \tag{31}$$

whose solution is given by the k eigenvectors of $\mathbf{L}_\nu$ corresponding to its smallest eigenvalues.

The novelty of the proposed method lies in the generalized Laplacian $\mathbf{L}_\nu$, which is defined with respect to an arbitrary positive vertex measure ν. This flexibility stems from the GDE framework and captures a rich family of partitioning behaviors grounded in random walk dynamics. When the measure is chosen as the stationary distribution, $\nu = \pi$, and the normalized Dirichlet energy (see Eq. 22) is minimized, the resulting method recovers the classical approach for strongly connected digraphs (Zhou et al., 2005).

Clustering algorithm. The GSC framework relies on a parametrized vertex measure $\nu := \nu_{t,\alpha}$ introduced in Sec. 3.3. Given a pair of parameter values ($t \in \mathbb{N}$, $\alpha \in \mathbb{R}_+$), Eq. 25 defines the vertex measure $\nu_{t,\alpha}$ that induces a generalized Dirichlet energy $\mathcal{D}_{\nu_{t,\alpha}}^2(\boldsymbol{V})$ (Eq. 29). This, in turn, leads to the generalized spectral graph partitioning objective of Eq. 31, based on the generalized Laplacian $\mathbf{L}_{t,\alpha}$ (see Def. 19). In practice, the GSC algorithm follows the same computational pipeline of the classical spectral clustering (see Alg. 1), with the difference that it substitutes the standard Laplacian with the generalized one $\mathbf{L}_{t,\alpha}$, and computes its leading eigenvectors $\mathbf{U}_{t,\alpha} \in \mathbb{R}^{N\times k}$ for downstream clustering.

4.3 Cheeger-type inequalities for GDE minimization

In the context of spectral clustering, Cheeger-type inequalities provide a fundamental link between the optimal value of a clustering objective (e.g. the normalized cut) and the spectral properties of the associated

graph Laplacian. These inequalities establish that if the second smallest eigenvalue of the Laplacian is small, then there exists a partition of the graph with low conductance, and conversely, if there exists a partition with low conductance, then the second smallest eigenvalue must also be small. This connection is crucial for understanding the theoretical guarantees of spectral clustering methods. In this subsection, we extend the Cheeger-type inequalities to the GDE minimization problem, providing theoretical insights into the quality of the partitions obtained by the GSC method. More specifically, for subset $S \subseteq \mathcal{V}$, we establish a conductance-like quantity defined with respect to the generalized edge measure $q(i,j) = \nu(i)p(i,j)$:

$$\phi_\nu(S) = \frac{q(S,\bar{S}) + q(\bar{S},S)}{\min\{\mathrm{vol}_{\nu+\xi}(S), \mathrm{vol}_{\nu+\xi}(\bar{S})\}} = \frac{\mathcal{D}^2_\nu(\chi_S)}{\min\{\mathrm{vol}_{\nu+\xi}(S), \mathrm{vol}_{\nu+\xi}(\bar{S})\}}. \tag{32}$$

Above, $\mathrm{vol}_{\nu+\xi}(S) = \sum_{i\in S} \nu(i) + \xi(i)$ is the volume of the set S with respect to the vertex measure $\nu + \xi$, such that $\mathrm{vol}_{\nu+\xi}(\mathcal{V}) = 2$. The quantity $\phi_\nu(S)$ measures the normalized bi-directional flow across the cut defined by S and $\bar{S}$, where the normalization is done with respect to the vertex measure ν. The bi-directional flow, $q(S,\bar{S}) + q(\bar{S},S)$, captures the total interaction between the two subsets, which is crucial for understanding the quality of a partition in terms of the dynamics induced by ν.

Eq. 32 is a natural generalization of the classical conductance, which is defined with respect to the stationary distribution π (Chung, 2005). Using Prop. 4.1 and its corollary, we see that when $\nu = \pi$, this formulation recovers the classical conductance, up to a multiplicative constant. This generalized conductance can be used to establish Cheeger-type inequalities that relate the optimal value of the GDE minimization problem to the quality of a partition in terms of ϕ_ν. We can thus define the vertex-measure-based Cheeger constant as:

$$\phi_\nu = \min_{S\subset\mathcal{V}} \phi_\nu(S),$$

which allows to characterize the 'quality' of the best possible partition of the graph in terms of ϕ_ν.

Theorem 4.1 (**Cheeger-type inequalities for the GDE minimization problem**). *Let λ_2 be the second smallest eigenvalue of the generalized Laplacian $\overline{\mathbf{L}}_\nu$. Then, the generalized conductance ϕ_ν and λ_2 satisfy the following inequalities:*

$$\frac{\lambda_2}{2} \le \phi_\nu \le \sqrt{2\lambda_2}. \tag{33}$$

The proof is given in App. C.6 and relies on standard techniques from spectral graph theory, adapted to the generalized conductance. Note that, using $\nu = \pi$ recovers the classical Cheeger inequalities for directed graphs (Chung, 2005), while using other vertex measures allows to establish new Cheeger-type inequalities that are tailored to the specific dynamics induced by ν. The lower bound indicates that if the second smallest eigenvalue (λ_2) is small, then there exists a partition of the graph with low generalized conductance ϕ_ν, which implies that the graph can be partitioned into two subsets with weak interaction under the dynamics induced by ν. The upper bound indicates that if there exists a partition with low generalized conductance ϕ_ν, then the second smallest eigenvalue must also be small, which implies that the GDE minimization problem has a low optimal value. These inequalities provide theoretical guarantees for the performance of the GSC method regarding the partition quality it produces. Note that, using $\nu = \pi$ recovers the classical Cheeger inequalities for directed graphs (Chung, 2005), while using other vertex measures allows to establish new Cheeger-type inequalities that are tailored to the specific dynamics induced by ν.

5 Experiments

5.1 Experimental setup

The proposed GSC method is empirically evaluated on clustering digraphs constructed from benchmark point-cloud data. The performance is compared against established spectral clustering baselines, under both fully unsupervised and label-informed evaluation protocols.

Compared methods. We evaluate two GSC variants based on generalized Laplacians from Sec. 3.3:

- $\mathrm{GSC}_{\mathrm{un}}$ (t,α): using the unnormalized generalized Laplacian $\mathbf{L}_\nu$ (Eq. 18),

- $\mathrm{GSC_n}$ (t, α): using the normalized generalized Laplacian $\overline{\mathbf{L}}_\nu$ (Eq. 19),

with $\nu := \nu_{t,\alpha}$ being the parametrized vertex measure (Eq. 25). The grid search for tuning the hyperparameters is over the ranges: $t \in \{0, 1, ..., 25\}$ and $\alpha \in \{0.0, 0.1, ..., 1.5\}$, and the upper limits are both set by taking into account the scale of the graph sizes we deal with in the experiments. Additionally, a comparison of the proposed GSC variants with some natural alternatives to the $\nu_{t,\alpha}$ vertex measure is provided in App. B.

We compare against the following spectral clustering methods from the literature, the first four of which have been developed for directed graphs:

- DSC+(γ) (Zhou et al., 2005; Palmer & Zheng, 2020), extended via teleporting random walks (Page et al., 1999) with teleportation parameter γ.
- $\mathrm{DI\text{-}SIM_L}(\tau)$, $\mathrm{DI\text{-}SIM_R}(\tau)$, and $\mathrm{DI\text{-}SIM_C}(\tau)$ (Rohe et al., 2016), which apply spectral clustering to the left, right, or concatenated singular vectors of a regularized transition matrix. For a graph with n vertices and average degree $\bar{d}$: $\tau \in \left\{\mathrm{round}\left(\bar{d} \cdot 10^s\right) \,\middle|\, s \in \{-1, -0.5, ..., 1\}\right\}$.
- $\mathrm{SC_{un}}$ and $\mathrm{SC_n}$: spectral clustering using unnormalized and normalized Laplacians on the symmetrized adjacency matrix (Von Luxburg, 2007).

All methods follow a standard spectral clustering pipeline: a graph embedding is computed via the eigenvectors of a Laplacian-type operator, followed by k-means++ (100 restarts) clustering (Arthur & Vassilvitskii, 2007) on the rows of the embedding, where the number of clusters k is assumed to be known. The best result is retained, according to on an internal evaluation index.

Clustering evaluation protocol. The clustering quality is assessed via the following metrics:

- *Internal evaluation*: We use an internal evaluation index, the Calinski-Harabasz index (CH), that computes the ratio between the sum of between-cluster variance and the within-cluster variance, for which it utilizes cluster centers. Higher CH values are better. The CH index for a partition $\mathcal{C} = \{C_1, ..., C_k\}$ of N datapoints is computed with respect to a distance or divergence function div:

$$\mathrm{CH}(\mathcal{C}) = \frac{N-k}{k-1} \cdot \frac{\sum_{j=1}^{k} |C_j| \,\mathrm{div}(c_j, c)}{\sum_{j=1}^{k} \sum_{i \in C_j} \mathrm{div}(x_i, c_j)}, \quad \text{where} \quad c_j = \frac{1}{|C_j|} \sum_{i \in C_j} x_i, \quad c = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

  Above, div() is a divergence; for Euclidean point-cloud data, div() is the squared Euclidean distance, for distribution based data, div() can be the KL-divergence. For network data, we can first represent each vertex i by the random walk distribution $x_i = \mathbf{P}\delta_i$, and row-wise differences dissimilarities by a divergence div() that can capture the local connectivity; when really needed, we call this variant *Graph CH* (GCH). To capture richer spatial context, longer diffusion times can be used, i.e. $x_i = \mathbf{P}^t\delta_i$, or mixtures of multiple different diffusion times to balance local and global information. In our experiments, we employ the standard CH for Euclidean data and the GCH for network data, using random walk representations at diffusion times $t \in \{2, 3\}$.

- *External evaluation*: Adjusted Mutual Information (AMI) (Strehl & Ghosh, 2002) is an external evaluation measure computed post hoc and assesses the statistical agreement (adjusted for agreement by chance) between the found clusters and the ground-truth labels. Higher AMI values indicate better clustering.

A *fully unsupervised clustering evaluation protocol* is employed, which emulates real-world conditions where no data labeling is available during training or parameter selection:

- *Model selection via internal validation*: Each method explores its respective hyperparameter space: GSC sweeps over (t, α), DSC+ over γ, and DI-SIM over τ. Each method performs 100 k-means++ runs per configuration of its hyperparameters, and the CH index is computed for each clustering solution. By design, CH prefers more 'regular' cluster shapes, meaning spherical (convex and compact) density shapes. We present experimental results that focus on the clustering methods by selecting to use mostly datasets in which CH is suitable. A workaround to this can be to make a more informed choice of the internal evaluation index and use with the same evaluation pipeline, or to go further and find ways to combine multiple evaluation indices.

- *Evaluation and reporting*: The solution yielding the best internal score (CH), measured with respect to the original data representation space, is retained along with the associated optimal hyperparameters; e.g. for GSC, those would be $(t, \alpha)^*_{\text{CH}} := (t^*_{\text{CH}}, \alpha^*_{\text{CH}})$. The AMI of that best solution is reported post hoc to better interpret the results using an external evaluation index.

The above protocol is designed to be fully unsupervised; the key point is that the model selection process relies solely on an internal validation index, without any access to ground-truth labels. The choice of the internal evaluation index influences the results. In all cases, we employ the generic and well-understood CH index (and the GCH variant for network data), however, this is not a hardwired feature of the proposed framework or pipeline. Thus, a user might consider selecting an internal index that aligns well with the expected cluster structure of the data. As a sanity check, we also report the best AMI obtained by each method when the hyperparameters are optimized directly on the AMI, which is a supervised evaluation protocol that serves as an upper bound for the unsupervised results. As a summary, we report the *Performance Ratio to Best* (PRB) of each method, defined as the mean of the ratios between the AMI of each method and the best AMI across all methods, for each dataset. PRB quantifies how close a method is to the best performing method in each dataset, with values closer to 1 indicating better performance:

$$\text{PRB}(m) = \frac{1}{|\mathcal{D}|} \sum_{j \in \mathcal{D}} \frac{\text{AMI}_j(m)}{\max_{m'} \text{AMI}_j(m')},$$

where $\mathcal{D}$ is the collection of all datasets, $\text{AMI}_j(m)$ is the AMI score of the method m on dataset j, and $\max_{m'} \text{AMI}_j(m')$ is the maximum AMI score achieved by any method on dataset j.

Graph construction from point-clouds. We test the clustering performance across different real-world datasets from the UCI repository (Dheeru & Karra Taniskidou, 2017). Given a dataset $\boldsymbol{X} = \{x_i\}_{i=1}^N$ with $x_i \in \mathbb{R}^d$, we build a sparse directed graph using an unweighted, directed K-nearest neighbor (KNN) construction, where $K = \lceil \log(N) \rceil$. This process yields sparse and typically weakly connected digraphs. The unweighted and asymmetric adjacency matrix $\mathbf{W} = \{w_{ij}\}$ is defined as: $w_{ij} = \mathbb{1}\{\frac{\|x_i - x_j\|^2}{\text{dist}_K(x_i)^2} \leq 1\}$, where $\text{dist}_K(x_i)$ denotes the distance of x_i to its K-th nearest neighbor.

Network data. We evaluate the clustering performance also on real-world networks, including Polblogs, Email-Eu-Core, Polbooks, College Football, a synthetic directed stochastic block model (DiSBM-Chain) with 3 communities and $N = 500$ vertices, organized in a chain structure, and a Degree-Corrected DiSBM, where one community have higher average degree than the others. For these datasets, we use the original adjacency matrices as input to the clustering methods, without any additional preprocessing or symmetrization.

5.2 Results

This section presents extensive empirical results following the described protocol. We start with clustering results on a number of benchmark point-cloud and network datasets, for which we also perform a sensitivity analysis over the parametrization of the proposed vertex measure, and reassuring runtime results that the computational cost of GSC remains reasonable. Then, we explore three challenging scenarios, using a controlled protocol and synthetic data. Those scenarios concern the impact of directionality in clustering through the usage of a chain-structured DiSBM as well as clustering under cluster size-imbalance and degree imbalance (inhomogeneity). Although that dealing with these scenarios is not among the main aims of the present work, they constitute characteristic challenges in clustering for which we demonstrate that GSC has a more flexible and robust behavior compared to the other methods. Finally, we summarize the results in the last part of the section.

5.2.1 Experiments on benchmark datasets

Clustering performance. Tab. 1 and Tab. 2 present clustering results across UCI datasets and networks. Each table consists of three subtables: the first refers to the unsupervised optimization of each method with respect to the CH (or GCH) criterion, for which the precise objective values are reported; the second subtable contains the AMI scores corresponding to the same clustering results of the first subtable; the third subtable shows the clustering results for when the methods are optimized with respect to ground-truth labels. As

explained in Sec. 5.1, the latter results are provided only as a reference to demonstrate the capacity of each model without the effect of the chosen internal validation index.

Supervised results. $\mathrm{GSC_n}$ achieves the top performance on 8 of the 9 UCI datasets (all except Wine). On networks, $\mathrm{GSC_n}$ wins decisively on DiSBM-Chain (AMI 1.0 vs. 0.53 for $\mathrm{SC_N}$) and Email Eu Core. Across both domains, $\mathrm{GSC_n}$'s PRB (resp. 1.00 and 0.93) indicates consistently strong relative performance. $\mathrm{GSC_{un}}$ also performs well, although generally underperforms compared to $\mathrm{GSC_n}$, suggesting that the normalized formulation captures better the data structure. $\mathrm{SC_n}$ and all DI-SIM variants show competitive performance on the UCI datasets, but underperform on the networks: both SC variants achieve 0 AMI on Polblogs, while $\mathrm{SC_{un}}$ gives 0 AMI also on Deg-Corr and Email-EU. DSC+ tend to overall underperform, with 0.58 and 0.54 PRB, respectively. $\mathrm{DI\text{-}SIM_R}$ underperforms on networks, with 0.52 PRB.

Unsupervised results. Under CH/GCH-based tuning, both GSC variants remain strong. On UCI datasets, $\mathrm{GSC_n}$ and $\mathrm{GSC_{un}}$ achieve PRB scores of 0.99 and 0.96, respectively; on networks, they achieve 0.84 and 0.90. $\mathrm{SC_n}$ and $\mathrm{SC_{un}}$ remain competitive on UCI, but on networks only $\mathrm{SC_n}$ stays relatively competitive (PRB 0.82), whereas $\mathrm{SC_{un}}$ drops to 0.64, notably affected by PolBlogs, Deg-Corr, and DiSBM-Chain. DSC+ and DI-SIM variants are generally weaker in this setting, especially on networks, with PRB around 0.70 for DSC+ and around 0.60 for DI-SIM variants.

As expected, unsupervised hyperparameter selection lowers performance for most parameterized methods compared to supervised tuning. The drop is smaller for $\mathrm{GSC_{un}}$, which often favors the uniform measure ($\alpha = 0$), a regime that is well aligned with CH/GCH index. The normalized $\mathrm{GSC_n}$ variant is more sensitive to unsupervised tuning, but remains promising when longer diffusion times are beneficial in the tuning of the vertex measure.

Sensitivity analysis for the parametrized vertex measure. Here, we analyze how the GSC clustering quality depends on the vertex measure, and we validate that an internal index such as CH can be used as a surrogate index for unsupervised model selection, through hyperparameter tuning. Fig. 2 illustrates the sensitivity of the vertex measure $\nu_{t,\alpha}$ (Eq. 25) to its hyperparameters. For each dataset, two heatmaps are shown with the landscapes produced for the CH and the AMI indices. Shades of orange indicate clustering results of higher quality. A white star in each plot indicates the hyperparameter configuration that yields the best respective evaluation index. Although some variability arises from the k-means++ initialization, for those datasets poor-quality regions can be identified in the CH landscape, which aligns well with the AMI landscape in high-quality regions. For datasets with well-separated clusters that are easier to detect, the clustering quality is relatively insensitive to the choice of $\nu_{t,\alpha}$; often the measure relies on α or t being equal to 0. Conversely, in more challenging cases, the choice of measure becomes crucial, and identifying the best hyperparameter becomes a task of interest. These results support using a suitable internal index as a reliable proxy for guiding model selection in fully unsupervised settings.

Runtimes. The computational efficiency of GSC is evaluated by measuring the runtime of the clustering pipeline across datasets of varying sizes. In particular, we show both the runtime of the different algorithms across the UCI datasets, and the runtime of $\mathrm{GSC_{un}}$ with and without hyperparameter tuning, compared to the runtime of $\mathrm{SC_{un}}$ in context where the number of points varies. The results (Fig. 3) indicate that $\mathrm{GSC_{un}}$ without tuning has a runtime close to that of $\mathrm{SC_{un}}$, demonstrating that the core algorithmic steps are as scalable as standard spectral clustering. Indeed, the main computational steps of GSC involve the construction of the graph Laplacian and the eigen-decomposition, which are similar to those in vanilla spectral clustering. The additional overhead from computing the vertex measure $\nu_{t,\alpha}$ is negligible compared to these steps, as it can be computed efficiently using power iterations or closed-form expressions for certain choices of t and α. The main overhead arises from the tuning of hyperparameters (t, α), which involves multiple runs of the clustering pipeline. However, even with tuning, the runtime remains reasonable in the benchmark datasets. In larger scale datasets, the tuning overhead can be mitigated by employing more efficient search strategies (e.g., random search, Bayesian optimization), by leveraging parallel computing resources or by using a non-tunable vertex measure (App. B). Note that DSC+ is generally quite close to GSC in terms of runtimes, while having a smaller grid-search size, this is due to the teleportation random walk, which implies a dense matrix, making the eigen-decomposition step more expensive. The un-normalized GSC and SC variants are generally faster than their normalized counterparts.

Table 1: **Comparison of clustering methods on UCI datasets.** (a-b) Unsupervised evaluation by optimizing the CH criterion: the first table shows CH scores, which is the pure unsupervised objective, while the second shows the corresponding AMI scores for those clustering solutions. (c) Supervising evaluation by optimizing directly the supervised AMI scores. The optimized hyperparameters are in parentheses $(\cdot)$.

(a) unsupervised evaluation (ch scores | ch-optimized)

Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(s)$	**DI-SIM**$_{\mathbf{L}}(s)$	**DI-SIM**$_{\mathbf{C}}(s)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Breast Tissue	50.66	47.79	27.23 (0.05)	54.54 (-1.00)	54.39 (-1.00)	54.31 (-0.50)	**54.78 (1, 0.4)**	52.04 (10, 0.4)
Iris	555.67	555.67	22.71 (0.05)	501.63 (-1.00)	501.63 (-1.00)	501.63 (-1.00)	555.67 (0, 0.0)	**558.06 (7, 0.1)**
Mnist64	139.69	139.69	56.40 (0.20)	143.06 (-0.50)	159.48 (0.00)	149.34 (-0.50)	139.69 (0, 0.0)	**161.54 (3, 0.5)**
Olivetti Faces	14.08	14.33	10.68 (0.10)	11.55 (-0.50)	13.95 (0.00)	12.57 (0.00)	14.08 (0, 0.0)	**14.83 (9, 0.2)**
Ph Recognition	431.49	431.49	8.71 (0.05)	42.53 (-1.00)	46.45 (-1.00)	41.65 (-1.00)	431.49 (0, 0.0)	**438.12 (1, 0.2)**
Seeds	247.31	247.31	192.67 (0.80)	230.54 (-1.00)	242.24 (0.50)	231.10 (-1.00)	247.31 (0, 0.0)	**248.04 (3, 0.1)**
Segmentation	418.28	435.75	10.90 (0.20)	436.24 (-0.50)	451.27 (-1.00)	448.89 (-1.00)	418.28 (0, 0.0)	**507.94 (12, 0.6)**
Wdbc	257.69	257.69	110.25 (0.85)	219.76 (-1.00)	257.29 (-1.00)	256.08 (-1.00)	257.69 (0, 0.0)	**258.74 (22, 0.6)**
Wine	70.37	70.37	70.43 (0.15)	68.54 (-1.00)	69.82 (0.50)	70.43 (0.50)	70.37 (0, 0.0)	**70.44 (22, 0.1)**
PRB	0.95	0.95	0.43	0.81	0.86	0.84	0.96	**0.99**

(b) unsupervised evaluation (ami scores | ch-optimized)

Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(s)$	**DI-SIM**$_{\mathbf{L}}(s)$	**DI-SIM**$_{\mathbf{C}}(s)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Breast Tissue	0.48	0.45	0.32 (0.05)	0.48 (-1.00)	0.47 (-1.00)	**0.51 (-0.50)**	0.48 (1, 0.4)	0.48 (10, 0.4)
Iris	**0.80**	**0.80**	0.35 (0.05)	0.75 (-1.00)	0.75 (-1.00)	0.75 (-1.00)	**0.80 (0, 0.0)**	0.77 (7, 0.1)
Mnist64	0.91	0.91	0.48 (0.20)	0.88 (-0.50)	**0.96 (0.00)**	0.96 (-0.50)	0.91 (0, 0.0)	0.95 (3, 0.5)
Olivetti Faces	0.68	0.69	0.64 (0.10)	0.65 (-0.50)	0.70 (0.00)	0.65 (0.00)	0.68 (0, 0.0)	**0.71 (9, 0.2)**
Ph Recognition	0.55	0.55	0.16 (0.05)	0.30 (-1.00)	0.34 (-1.00)	0.33 (-1.00)	0.55 (0, 0.0)	**0.56 (1, 0.2)**
Seeds	**0.74**	**0.74**	0.62 (0.80)	0.71 (-1.00)	0.69 (0.50)	0.70 (-1.00)	**0.74 (0, 0.0)**	0.74 (3, 0.1)
Segmentation	0.64	**0.64**	0.05 (0.20)	0.61 (-0.50)	0.64 (-1.00)	0.64 (-1.00)	0.64 (0, 0.0)	0.61 (12, 0.6)
Wdbc	**0.68**	**0.68**	0.33 (0.85)	0.64 (-1.00)	0.66 (-1.00)	**0.68 (-1.00)**	**0.68 (0, 0.0)**	0.62 (22, 0.6)
Wine	0.86	0.86	0.86 (0.15)	**0.88 (-1.00)**	0.83 (0.50)	0.86 (0.50)	0.86 (0, 0.0)	0.85 (22, 0.1)
PRB	**0.98**	0.97	0.57	0.90	0.93	0.93	**0.98**	0.97

(c) supervised evaluation (ami scores | ami-optimized)

Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(s)$	**DI-SIM**$_{\mathbf{L}}(s)$	**DI-SIM**$_{\mathbf{C}}(s)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Breast Tissue	0.48	0.45	0.32 (0.05)	0.48 (-1.00)	0.51 (-0.50)	0.51 (-0.50)	0.51 (21, 0.1)	**0.51 (1, 0.3)**
Iris	0.80	0.80	0.35 (0.05)	0.76 (-0.50)	0.75 (-1.00)	0.75 (-1.00)	0.80 (0, 0.0)	**0.88 (3, 1.2)**
Mnist64	0.91	0.91	0.48 (0.20)	0.88 (-0.50)	0.96 (0.00)	0.96 (-0.50)	0.91 (0, 0.0)	**0.97 (15, 0.9)**
Olivetti Faces	0.68	0.69	0.64 (0.30)	0.65 (-0.50)	0.70 (0.00)	0.67 (0.50)	0.69 (2, 0.1)	**0.74 (2, 0.8)**
Ph Recognition	0.55	0.55	0.16 (0.50)	0.30 (-1.00)	0.34 (-1.00)	0.33 (-1.00)	0.55 (0, 0.0)	**0.56 (2, 0.2)**
Seeds	0.74	0.74	0.62 (0.55)	0.71 (-1.00)	0.69 (0.50)	0.70 (0.00)	0.74 (0, 0.0)	**0.77 (2, 1.0)**
Segmentation	0.64	0.64	0.07 (0.10)	0.61 (-0.50)	0.64 (-0.50)	0.64 (-1.00)	0.64 (0, 0.0)	**0.70 (13, 1.4)**
Wdbc	0.68	0.68	0.46 (0.10)	0.64 (-1.00)	0.69 (-0.50)	0.69 (-0.50)	0.68 (0, 0.0)	**0.73 (5, 0.8)**
Wine	0.86	0.86	0.87 (0.05)	**0.90 (0.00)**	0.83 (-1.00)	0.86 (0.50)	0.86 (0, 0.0)	0.86 (0, 0.0)
PRB	0.94	0.93	0.58	0.87	0.90	0.90	0.95	**1.00**

Table 2: **Comparison of clustering methods on network datasets.** (a-b) Unsupervised evaluation by optimizing the GCH criterion: the first table shows GCH scores, which is the pure unsupervised objective, while the second shows the corresponding AMI scores for those clustering solutions. (c) Supervising evaluation by optimizing directly the supervised AMI scores. The optimized hyperparameters are in parentheses $(\cdot)$.

(a) unsupervised evaluation (gch scores | gch-optimized)

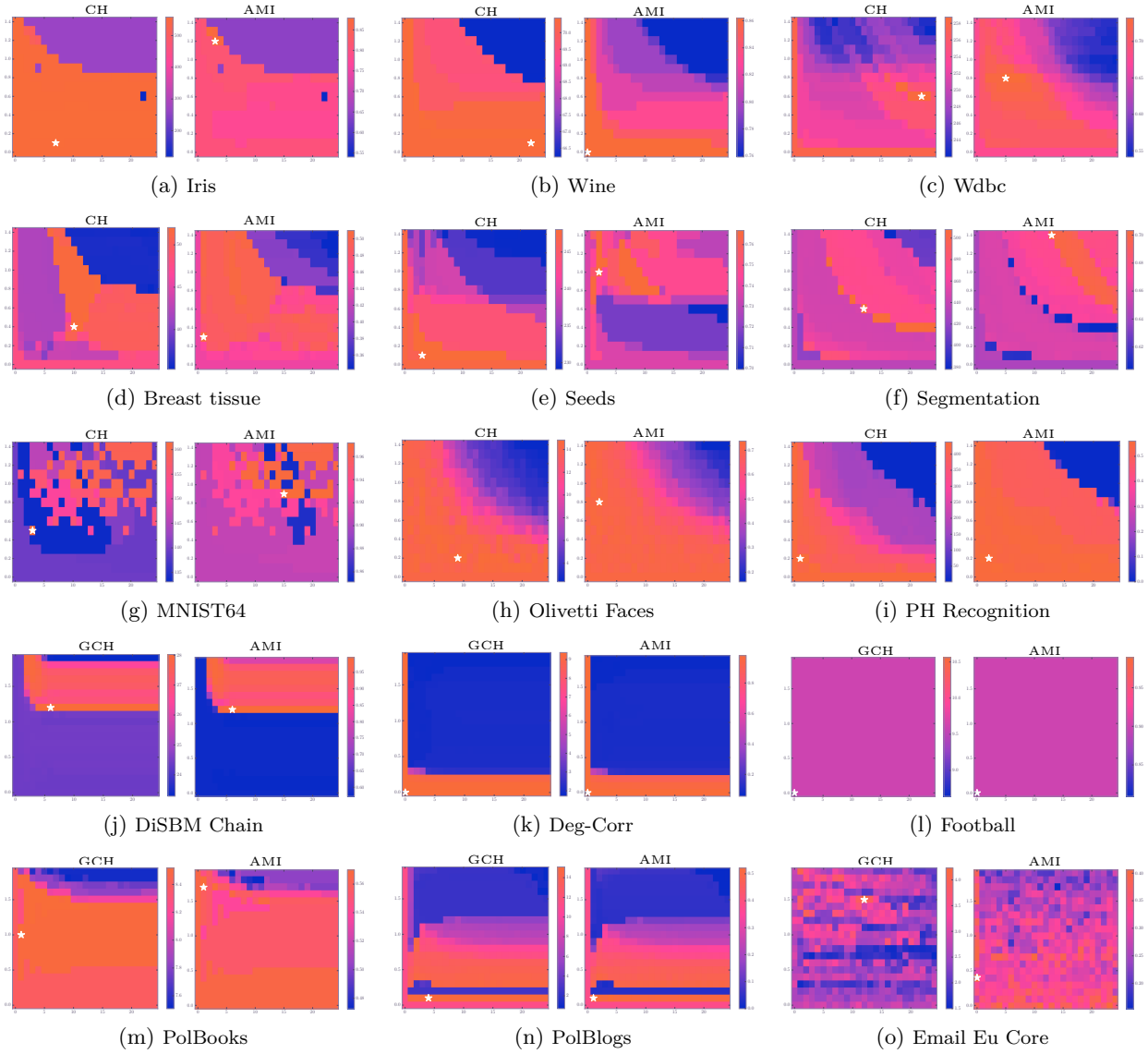
Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(\tau)$	**DI-SIM**$_{\mathbf{L}}(\tau)$	**DI-SIM**$_{\mathbf{C}}(\tau)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Deg-Corr	3.85	9.37	4.46 (0.05)	1.80	4.44	4.50	9.37 (0, 0.0)	**9.38 (0, 0.0)**
Disbm Chain	16.90	23.49	14.64 (0.05)	4.27	13.94	13.88	23.55 (3, 0.2)	**28.02 (6, 1.2)**
Email Eu Core	4.63	**5.43**	2.16 (0.15)	2.42	1.62	2.21	4.62 (0, 1.2)	3.95 (2, 1.3)
Football	**9.60**	**9.60**	9.14 (0.05)	**9.60**	**9.60**	**9.60**	**9.60 (0, 0.0)**	**9.60 (0, 0.0)**
Polblogs	5.22	5.22	**51.53 (0.05)**	35.44	24.52	25.08	36.57 (1, 0.2)	15.09 (4, 0.1)
Polbooks	7.20	8.38	7.14 (0.10)	7.41	7.16	7.54	8.51 (1, 0.3)	**8.52 (1, 1.0)**
PRB	0.64	0.82	0.70	0.56	0.60	0.63	**0.90**	0.84

(b) unsupervised evaluation (ami scores | gch-optimized)

Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(\tau)$	**DI-SIM**$_{\mathbf{L}}(\tau)$	**DI-SIM**$_{\mathbf{C}}(\tau)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Deg-Corr	0.00	**0.99**	0.44 (0.05)	0.01	0.45	0.40	0.99 (0, 0.0)	0.98 (0, 0.0)
Disbm Chain	0.73	0.53	0.31 (0.05)	0.11	0.32	0.37	0.73 (3, 0.2)	**1.00 (6, 1.2)**
Email Eu Core	0.00	**0.52**	0.32 (0.15)	0.18	0.30	0.32	0.35 (0, 1.2)	0.29 (2, 1.3)
Football	**0.90**	**0.90**	0.86 (0.05)	**0.90**	**0.90**	**0.90**	**0.90 (0, 0.0)**	**0.90 (0, 0.0)**
Polblogs	0.00	0.00	0.02 (0.05)	0.02	0.54	**0.57**	0.04 (1, 0.2)	0.52 (4, 0.1)
Polbooks	**0.64**	0.57	0.55 (0.10)	0.53	0.53	0.53	0.55 (1, 0.3)	0.54 (1, 1.0)
PRB	0.46	0.74	0.54	0.39	0.69	0.70	0.72	**0.88**

(c) supervised evaluation (ami scores | ami-optimized)

Dataset	$\mathbf{SC_{un}}$	$\mathbf{SC_{n}}$	**DSC+**(γ)	**DI-SIM**$_{\mathbf{R}}(\tau)$	**DI-SIM**$_{\mathbf{L}}(\tau)$	**DI-SIM**$_{\mathbf{C}}(\tau)$	$\mathbf{GSC_{un}}(t,\alpha)$	$\mathbf{GSC_{n}}(t,\alpha)$
Deg-Corr	0.00	**0.99**	0.44 (0.05)	0.09	0.45	0.40	0.99 (0, 0.0)	0.98 (0, 0.0)
Disbm Chain	0.73	0.53	0.31 (0.05)	0.11	0.32	0.37	0.73 (0, 0.0)	**1.00 (6, 1.2)**
Email Eu Core	0.00	**0.52**	0.34 (0.40)	0.32	0.32	0.37	0.38 (12, 0.1)	0.41 (5, 0.0)
Football	**0.90**	**0.90**	0.86 (0.05)	**0.90**	**0.90**	**0.90**	**0.90 (0, 0.0)**	**0.90 (0, 0.0)**
Polblogs	0.00	0.00	0.02 (0.05)	0.26	0.54	**0.57**	0.54 (0, 0.0)	0.52 (1, 0.1)
Polbooks	**0.64**	0.57	0.55 (0.10)	0.53	0.53	0.53	0.58 (8, 0.3)	0.57 (0, 0.0)
PRB	0.46	0.74	0.54	0.52	0.70	0.72	0.88	**0.93**



(a) Iris (b) Wine (c) Wdbc

(d) Breast tissue (e) Seeds (f) Segmentation

(g) MNIST64 (h) Olivetti Faces (i) PH Recognition

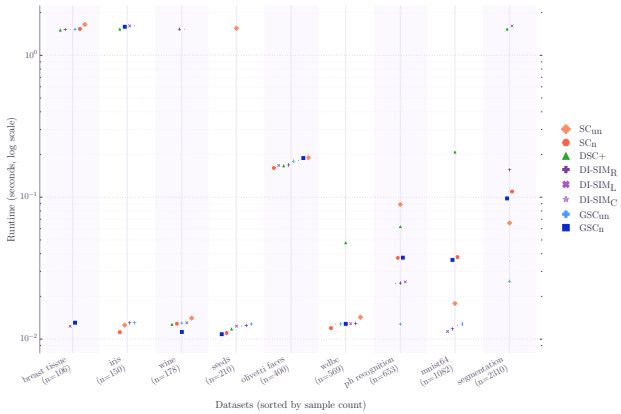
(j) DiSBM Chain (k) Deg-Corr (l) Football

(m) PolBooks (n) PolBlogs (o) Email Eu Core

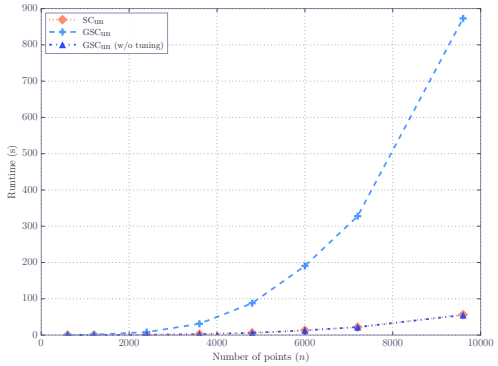
Figure 2: **$\mathbf{GSC_n}$ sensitivity to hyperparameters.** For each dataset, CH and AMI heatmaps are shown over the space of vertex measure parametrization (t: x-axis, α: y-axis). Shades of orange indicate better clustering, and a white star marks the best configuration for each index, i.e. $(t, \alpha)^*_{\text{CH}}$ vs $(t, \alpha)^*_{\text{AMI}}$. In most cases, the CH landscape aligns well with that of AMI, supporting CH-based unsupervised model selection. (a-i) Heatmaps for the UCI point-cloud datasets; (j-o) heatmaps for the networks datasets. For Wdbc, Seeds and Segmentation, the CH and AMI landscapes differ, highlighting the difficulty of unsupervised model selection when the internal index does not align well with the cluster structure of the data. In datasets like MNIST64, the behavior of both CH and AMI exhibits high sensitivity to the choice of vertex measure, with very sporadic regions. The correspondence between the two appears to be still relevant in this context.

5.2.2 Exploration of challenging synthetic scenarios

We investigate three synthetic scenarios that are challenging for vanilla SC, and therefore informative for assessing the benefits of GSC. Following the intuition in Sec. 3.1, we focus on cases where skewed stationary distributions hinder the recovery of the cluster structure by SC. We compare GSC with vanilla SC and DSC+, a strong directed-graph baseline based on teleportation random walks.



(a) Runtime of the algorithms across the UCI datasets.



(b) Runtime of GSC_{un} with and without hyperparameter tuning, compared to SC_{un}.

Figure 3: **Runtimes.** (a) Runtimes for all methods across the UCI datasets. (b) Runtimes of GSC_{un} with and without hyperparameter tuning, compared to SC_{un}, as the number of points increases. The results demonstrate that GSC is computationally as efficient as vanilla SC, while hyperparameter tuning brings reasonable overhead. The un-normalized GSC and SC variants are generally faster than their normalized counterparts.

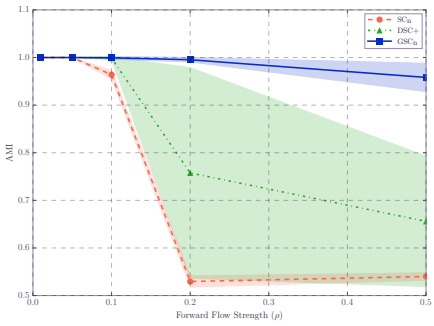
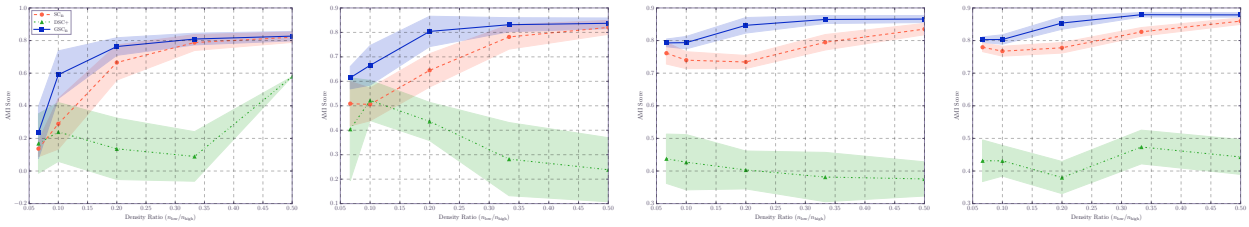


Figure 4: **Clustering performance in the chain-structured digraph setting.** GSC_{n}, SC_{n} and DSC+ are evaluated on a synthetic chain-structured directed graph. We vary the forward edge probability ρ and measure the AMI score of the clustering solutions.

Chain-structured digraphs. In this setting, the graph is composed of 3 clusters of 500 points, arranged in a chain structure, with forward edges from cluster 1 to cluster 2, and from cluster 2 to cluster 3 (see details in Appendix. A). The forward edge probability ρ controls the strength of the directed structure, with higher ρ leading to more pronounced directionality. The backward edges probability is fixed to 0.01. This yields a setting where the stationary distribution of the natural random walk is skewed, with most of the mass concentrated in the last cluster. The results in Fig. 4 show that GSC significantly outperforms SC and DSC+ across all values of ρ. In particular, as ρ increases, the performance of SC degrades significantly, while GSC remains robust and maintains a high AMI score. This demonstrates that GSC is able to effectively leverage the directed structure of the graph to recover meaningful clusters, even in cases where SC_{n} fails due to skewed stationary distributions. DSC+ also struggles in this setting, despite being better than SC_{n}.

Cluster size-imbalance. As highlighted in Sec. 3.1, the case of skewed stationary distribution is of particular interest, as it can lead to poor clustering for classical SC methods. To investigate this, we construct synthetic datasets with varying size imbalance, and evaluate the performance of GSC against SC_{n} and DSC+. The datasets are constructed as follows: we place K clusters on a squared grid, with alternating sizes. The smaller clusters contain n_{low} points, while the larger clusters contain $n_{high} = 300$ points. Datapoints are sampled from Gaussian distributions centered at the cluster centroids, with a fixed variance $\sigma^2 = 0.3$, and cluster



(a) 2x1 grid (b) 2x2 grid (c) 3x3 grid (d) 4x4 grid

Figure 5: **Clustering performance in the size-unbalance setting.** GSC_n, SC_n and DSC+ are evaluated on synthetic datasets with varying levels of cluster size imbalance. The x-axis represents the ratio of points in smaller clusters to larger clusters, while the y-axis shows the AMI score.

centroids are spaced-out by 3. We vary the number of clusters K and the size of the smaller clusters n_{low} to create different levels of imbalance. The results (Fig. 5) show that GSC is more effective at recovering small clusters compared to vanilla SC, which tend to merge them into larger clusters. In particular, as the number of clusters grows, we can see that the ratio needed to obtain full performance becomes bigger. In every case, GSC recovers the structure for smaller ratios of density than Vanilla SC. While in every cases considered, DSC+ performs poorly, due to the fact that the stationary distribution is very skewed, and that it relies on the teleportation random walk, which may hinder the information contained in the original graph. While not shown in the figures, DI-SIM variants performs similarly to SC_n.

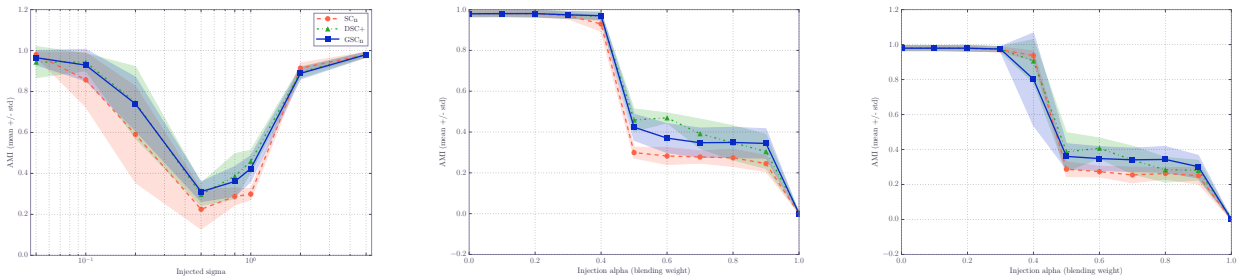
Node degree-imbalance. In degree-imbalanced graphs, the stationary distribution of the natural random walk is skewed, which can lead to poor clustering for classical SC methods. To investigate this, we construct synthetic directed graphs with varying degrees of degree imbalance and evaluate the performance of GSC against SC_n and DSC+. The graphs are constructed as follows: starting from 3 Gaussian clusters of 300 points, we construct a "Gaussian injection" using:

$$K(i,j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2s^2}\right) \quad \text{and} \quad \tilde{K}(i,j) = \exp\left(-\frac{\|x_i - c\|^2 + \|x_j - c\|^2}{2\sigma^2}\right),$$

where c is the fixed center of the Gaussian injection (here, the center of the first cluster). The parameter σ controls the level of degree imbalance, with smaller σ leading to more skewed degree distributions. We then construct the adjacency matrix A, by combining the base kernel K and the bias kernel $\tilde{K}$, as $A_\alpha = (1-\alpha)K + \alpha\tilde{K}$. Results are shown in Fig. 6. In Fig. 6a, we investigate how, for alpha fixed ($= 0.5$) the different methods manage for different σ. In Fig. 6b and 6c, we investigate how, for σ fixed, the different methods manage for different α values. In all cases, we see that GSC is more robust to degree imbalance compared to vanilla SC_n, which tends to be biased towards high-degree vertices. In particular, as the degree imbalance increases, GSC_n and DSC+ maintains better performance than SC_n. While not shown in the figures, DI-SIM variants performs similarly to SC_n.

5.2.3 Summary of the results

As a whole, the reported experiments highlight the flexibility and robustness of the GSC framework. The tunable parameters (t, α) enable adaptation to diverse graph topologies, from sparse and directed to dense and symmetric. The observed alignment between internal (CH) and external (AMI) metrics across a variety of datasets supports the practical value of CH-based model selection in fully unsupervised settings, provided that CH is a suitable index for the cluster structure of a dataset. By generalizing Dirichlet energy through flexible vertex measures, GSC provides a principled and scalable approach to graph clustering that performs reliably across a wide range of conditions. GSC_{un} requires no tuning at $\alpha = 0$ and remains competitive with baselines in many cases. However, optimizing (t, α) yields substantial improvements, especially for GSC_n and on complex graphs with skewed degree properties. Finally, the disagreement between unsupervised and supervised evaluation is not negligible, but it rather highlights the difficulty of unsupervised model selection, and more specifically the mismatch between a considered internal evaluation index and the cluster structure



(a) $\alpha = 0.5$ (b) $\sigma = 1$ (c) $\sigma = 0.8$.

Figure 6: **Gaussian noise injection.** $\mathrm{GSC_n}$, $\mathrm{SC_n}$ and DSC+ are evaluated on synthetic directed graphs with varying levels of degree imbalance. In (a) The x-axis represents the standard deviation σ of the Gaussian noise added to the edge weights, in (b-c), the x-axis represents the hyperparameter α. The y-axis shows the mean AMI score obtained by the methods, over multiple seeds. The results indicate that $\mathrm{GSC_n}$ is more robust to degree imbalance compared to vanilla SC, which tends to be biased towards high-degree vertices. In particular, as the degree imbalance increases, $\mathrm{GSC_n}$ and DSC+ maintains better performance than $\mathrm{SC_n}$.

present in a dataset. This finding leaves to for future work the investigation of other more flexible internal indices and optimization strategies.

6 Conclusion

This work introduced the *Generalized Dirichlet Energy* (GDE) framework, which extends classical notions of graph smoothness to directed graphs through the use of arbitrary vertex measures and random walk dynamics. This leads to a principled construction of parametrized graph Laplacians that can capture both directional and structural properties without relying on teleportation or artificial symmetrization. Built on this foundation, we developed the *Generalized Spectral Clustering* (GSC), a method that leverages a parametrized vertex measure $\nu_{t,\alpha}$ to adapt to the geometry of directed graphs. The resulting model enables a smooth interpolation between local/global behavior and sparse/dense connectivity, offering a flexible and theoretically grounded approach to clustering in asymmetric and weakly connected settings.

Empirical evaluations on real-world point-cloud datasets show that GSC consistently performs well, often outperforming classical and teleportation-based baselines, under both internal and external evaluation metrics. Moreover, the proposed method has empirically demonstrated flexibility and robustness when dealing with various scenarios of structural biases that typically challenge spectral clustering methods. These results highlight the value of modeling diffusion via non-ergodic, structure-aware dynamics driven by carefully chosen vertex measures.

The versatility of the GDE framework opens up multiple avenues for future research, including its integration into semi-supervised learning, graph signal processing, and operator-based methods in non-reversible or directed settings. An especially promising direction is the automated or data-driven selection of (t, α), enabling a fully adaptive graph inference across tasks and domains.

References

David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.

March Boedihardjo, Shaofeng Deng, and Thomas Strohmer. A performance guarantee for spectral clustering. *SIAM Journal on Mathematics of Data Science*, 3(1):369–387, 2021.

Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2013.

Fan Chung. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.

Simon Coste and Ludovic Stephan. A simpler spectral approach for clustering in directed networks. *preprint arXiv:2102.03188*, 2021.

Mihai Cucuringu, Huan Li, He Sun, and Luca Zanetti. Hermitian matrices for clustering directed graphs: insights and applications. In *International Conference on Artificial Intelligence and Statistics*, pp. 983–992, 2020.

D Dheeru and E Karra Taniskidou. UCI repository of learnin learning databases. *University of California, Irvine, School of Information and Computer Sciences*, 2017.

Jiu Ding and Aihui Zhou. *Frobenius-Perron Operators*, pp. 62–91. Springer, Berlin, Heidelberg, 2009.

Marc G. Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, 2:299–312, mar 2002.

Koby Hayashi, Sinan G Aksoy, and Haesun Park. Skew-symmetric adjacency matrices for clustering directed graphs. *Preprint arXiv:2203.01388*, 2022.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

Stefan Klus and Nataša Djurdjevac Conrad. Koopman-based spectral clustering of directed and time-evolving graphs. *Journal of Nonlinear Science*, 33(1):1–22, 2023.

Stefan Klus and Maia Trower. Transfer operators on graphs: Spectral clustering and beyond. *Journal of Physics: Complexity*, 5(1):015014, feb 2023.

Steinar Laenen and He Sun. Higher-order spectral clustering of directed graphs. *preprint arXiv:2011.05080*, 2020.

David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2017.

Marina Meilă and William Pentney. Clustering by weighted cuts in directed graphs. In *SIAM International Conference on Data Mining*, pp. 135–144, 2007.

Marina Meilă and Jianbo Shi. A random walks view of spectral segmentation. In *International Workshop on Artificial Intelligence and Statistics*, pp. 203–208, 2001.

Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2002.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The Pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

William R Palmer and Tian Zheng. Spectral clustering for directed networks. In *International Conference on Complex Networks and Their Applications*, pp. 87–99. Springer, 2020.

Robert L. Peach, Alexis Arnaudon, and Mauricio Barahona. Semi-supervised classification on graphs using explicit diffusion dynamics. *Foundations of Data Science*, 2(1):19–33, 2020.

Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In *Conference on Learning Theory*, pp. 1423–1455, 2015.

Karl Rohe, Tai Qin, and Bin Yu. Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences*, 113(45):12679–12684, 2016.

Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on Probability Teory and Statistics*, pp. 301–413. Springer, 1997.

Venu Satuluri and Srinivasan Parthasarathy. Symmetrizations for clustering directed graphs. In *International Conference on Extending Database Technology*, pp. 343–354, 2011.

Michael T Schaub, Jean-Charles Delvenne, Renaud Lambiotte, and Mauricio Barahona. Multiscale dynamical embeddings of complex networks. *Physical Review E*, 99(6):062308, 2019.

Eddie Seabrook and Laurenz Wiskott. A Tutorial on the Spectral Theory of Markov Chains. *Neural Computation*, 35(11), 2023.

Harry Sevi, Gabriel Rilling, and Pierre Borgnat. Harmonic analysis on directed graphs and applications: From fourier analysis to wavelets. *Applied and Computational Harmonic Analysis*, 62:390–440, 2023.

Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

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

Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(Dec):583–617, 2002.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *International Conference on Machine Learning*, pp. 1036–1043, 2005.

A Additional Information on the used datasets

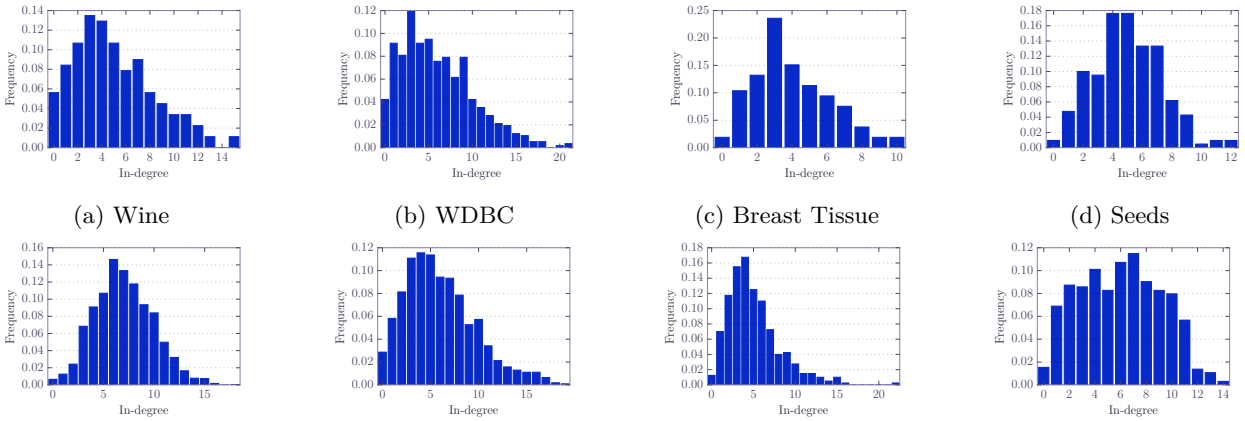
We provide in Tab. 3 some statistics on the datasets used in the experiments: the number of samples N, the number of features E, the number of data classes K, as well as statistics on the graph structure, such as the Gini coefficient of the in-degree distribution, the Reciprocity of edge weights, and the Cluster-Level reciprocity. Moreover, their connectivity properties are also provided (i.e. number of weakly-connected components and number of strongly-connected components). The latter is a measure of how reciprocal the connections are between pairs of ground-truth clusters, and can provide insights into the suitability of the different clustering methods. For binary edge weights, all these measures lie in $[0, 1]$ and are defined as follows:

$$\begin{aligned}
\text{Gini}(x) &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|}{2N\sum_{i=1}^{N} x_i}, \\
\text{Reciprocity}(\mathbf{W}) &= \frac{\sum_{i,j}\mathbf{W}_{ij}\cdot\mathbf{W}_{ji}}{\sum_{i,j}\mathbf{W}_{ij}}, \\
\text{C-L Reciprocity}(\mathbf{W}, y) &= \frac{1}{K(K-1)}\sum_{c\neq c'}\frac{2\cdot\min\big(E(c,c'), E(c',c)\big)}{E(c,c') + E(c',c)},
\end{aligned}$$

where $E(c, c') = \sum_{i\in c, j\in c'}\mathbf{W}_{ij}$ is the number of directed edges from cluster c to cluster c', y is a ground-truth cluster assignment, and K is the total number of clusters. In addition, we provide in Fig. 7 the in-degree distributions of a subset of the K-NN graphs constructed from the different datasets. Due to their construction, those graphs have a fixed out-degree equal to K, but their in-degree distribution can be quite skewed. This motivates the use of flexible vertex measures that can adapt to such heterogeneity.

Table 3: **Dataset statistics.** For each dataset, we report the number of samples N, the number of edges E, the number of classes K, as well as some statistics on the graph structure, such as the Gini coefficient of the out-degree distribution, the reciprocity, Cluster-Level reciprocity and number of both weakly and strongly connected components.

Dataset	N	$\lvert E \rvert$	K	**Gini**	**Reciprocity**	**C-L Reciprocity**	**#WCC**	**#SCC**
Breast Tissue	106	424	6	0.308	0.646	0.418	1	8
Email-EUcore	1,005	25,571	42	0.539	0.718	0.542	20	203
Football	115	1,226	12	0.040	1.000	0.894	1	1
Iris	150	750	3	0.321	0.635	0.283	2	13
Mnist64	1,082	6,492	6	0.341	0.615	0.275	2	47
Olivetti Faces	400	2,000	40	0.331	0.661	0.039	4	38
Ph Recognition	653	3,918	15	0.303	0.646	0.297	2	132
PolBlogs	814	15,939	2	0.676	0.290	0.860	2	12
PolBooks	105	882	3	0.328	1.000	1.000	1	1
Seeds	210	1,050	3	0.257	0.690	0.589	1	6
Segmentation	2,310	16,170	7	0.232	0.725	0.258	2	57
Wdbc	569	3,414	2	0.382	0.478	0.796	1	52
Wine	178	890	3	0.371	0.575	0.196	1	18



(a) Wine (b) WDBC (c) Breast Tissue (d) Seeds

(e) Segmentation (f) MNIST64 (g) Olivetti Faces (h) PH Recognition

Figure 7: **In-degree distributions for the K-NN graphs.** Histograms of the node in-degrees for different datasets. Due to their construction as (directed) K-NN graphs, the distributions are not uniform.

Directed Stochastic Block Models. In addition to the real-world datasets, we also evaluate our method on synthetic datasets generated from directed stochastic block models (DiSBM). These models allow us to control the level of reciprocity and degree imbalance in the graph, providing a more systematic evaluation of the clustering methods under different structural conditions. The DiSBM is defined by a block structure with specified probabilities of edge formation between and within blocks, which can be tuned to create various levels of connectivity and directionality. By varying these parameters, we can simulate scenarios ranging from highly reciprocal to highly asymmetric graphs, and assess how well GSC and other methods can recover the underlying cluster structure in each case. In particular, we consider a 'chain' DiSBM, where 3 clusters are connected in a chain-like structure, with edges predominantly directed from one cluster to the next. This setup allows us to evaluate the ability of the methods to capture directional relationships and recover the correct clustering in a non-reciprocal setting, it is defined as follows:

$$N = [500, 500, 500], \quad \boldsymbol{Q}_\rho = \begin{bmatrix} 0.1 & \rho & 0.01 \\ 0.01 & 0.1 & \rho \\ 0.01 & 0.01 & 0.1 \end{bmatrix},$$

where N is the vector of cluster sizes and $\boldsymbol{Q}_\rho$ is the matrix of edge probabilities, with ρ controlling the level of directionality. By default, ρ is set to 0.1 creating a strong directional bias to the graph.

We also generate degree-corrected directed SBMs with block-specific power-law distributions. For each block k, we sample node propensities θ_i by drawing from a Pareto(α_k) distribution, normalizing to unit mean, then scaling by σ_k. The edge probability between nodes i and j is $p_{ij} = \min(p_{z_i z_j} \cdot \theta_i \theta_j, 1)$, where z_i is node i's block label and $p_{z_i z_j}$ is the baseline probability between blocks z_i and z_j. For our experiments, we use three balanced blocks of 500 nodes with intra-block probability $p_{\text{intra}} = 0.05$ and inter-block probability $p_{\text{inter}} = 0.01$. Pareto exponents $(\alpha_1, \alpha_2, \alpha_3) = (1.8, 3.5, 3.5)$ yield a heavy-tailed first block, while scaling factors $(\sigma_1, \sigma_2, \sigma_3) = (2.5, 0.7, 0.7)$ give it substantially higher expected degree.

B Different approaches for vertex measures

In this section, we provide additional details on the different approaches for vertex measures that we have considered in our experiments. In particular, those vertex measures are:

- $\nu_{t,\alpha}$: the proposed vertex measure (see Sec. 4).
- $\nu_{\text{deg}}(\gamma)$: a degree-based vertex measure defined as $\nu_{\text{deg}}(\gamma)(i) = \gamma d_{\text{out}}(i) + (1-\gamma) d_{\text{in}}(i)$. This measure interpolates between out-degree and in-degree, with $\gamma \in [0, 1]$ controlling the balance. Generally, gamma is set to 0.5.
- ν_{unif}: the uniform vertex measure, where $\nu_{\text{unif}}(i) = 1/|\mathcal{V}|$ for all $i \in \mathcal{V}$.
- ν_{Perron}[4]: the Perron vector of the random walk transition matrix P, which corresponds to the stationary distribution of the natural random walk on the graph.

As stated in Sec. 3.3, the proposed vertex measure $\nu_{t,\alpha}$ recovers the uniform measure when $\alpha = 0$. Using ν_{Perron} (the stationary distribution) recovers the Directed Laplacians of Chung (2005). Unlike DSC+ (which uses teleportation), the Perron vector may be supported on a subset of nodes in weakly connected graphs (Sec. 3.1), degrading performance. Results in Tab. 4 show: $\nu_{t,\alpha}$ achieves best supervised performance and is competitive unsupervised; $\nu_{\text{deg}}(\gamma)$ and ν_{unif} also perform well; Perron measures perform poorly, especially unnormalized. These results confirm the importance of flexible, adaptive vertex measures and the advantages of $\nu_{t,\alpha}$, while supporting alternatives when (t, α) tuning is costly or non-parametric measures are preferred.

[4]In practice, having zero entries makes the definition of the normalized generalized Lpalacians untractable, hence, we set those entries to a small positive value.

Table 4: **Comparison of different vertex measures.**(a-b) Unsupervised evaluation by optimizing the CH criterion: the first table shows CH scores, which is the pure unsupervised objective, while the second shows the corresponding AMI scores for those clustering solutions. (c) Supervising evaluation by optimizing directly the supervised AMI scores. The optimized hyperparameters are in parentheses $(\cdot)$.

(a) UNSUPERVISED EVALUATION (CH SCORES | CH-OPTIMIZED)

Dataset	$\nu_{t,\alpha}$		$\nu_{\text{deg}}(\gamma)$		ν_{unif}		ν_{Perron}	
	GSC$_{\text{un}}(t,\alpha)$	**GSC$_{\text{n}}(t,\alpha)$**	**deg-GSC$_{\text{un}}(\gamma)$**	**deg-GSC$_{\text{n}}(\gamma)$**	**uniform-GSC$_{\text{un}}$**	**uniform-GSC$_{\text{n}}$**	**perron-GSC$_{\text{un}}$**	**perron-GSC$_{\text{n}}$**
Breast Tissue	**54.78** (1, 0.4)	52.04 (10, 0.4)	54.73 (0.9)	50.57 (0.6)	50.66	47.79	17.43	35.29
Iris	555.67 (0, 0.0)	**558.06** (7, 0.1)	555.67 (0.0)	555.67 (0.0)	555.67	555.67	2.21	**558.06**
Mnist64	139.69 (0, 0.0)	**161.54** (3, 0.5)	141.14 (1.0)	141.14 (1.0)	139.69	139.69	6.02	70.42
Olivetti Faces	14.08 (0, 0.0)	**14.83** (9, 0.2)	14.74 (0.2)	14.70 (0.4)	14.08	14.23	2.19	4.27
Ph Recognition	431.49 (0, 0.0)	**438.12** (1, 0.2)	436.85 (0.2)	436.85 (0.1)	431.49	431.49	80.18	10.00
Seeds	247.31 (0, 0.0)	248.04 (3, 0.1)	247.31 (0.0)	**248.60** (0.3)	247.31	247.31	2.10	247.31
Segmentation	418.28 (0, 0.0)	**507.94** (12, 0.6)	471.20 (0.7)	473.96 (0.6)	418.28	435.75	175.48	52.27
Wdbc	257.69 (0, 0.0)	**258.74** (22, 0.6)	257.69 (0.0)	257.69 (0.0)	257.69	257.69	0.91	251.28
Wine	70.37 (0, 0.0)	**70.44** (22, 0.1)	70.37 (0.0)	**70.44** (1.0)	70.37	70.37	2.98	69.97
PRB	0.96	0.99	0.98	0.97	0.95	0.95	0.12	0.61

(b) UNSUPERVISED EVALUATION (AMI SCORES | CH-OPTIMIZED)

Dataset	$\nu_{t,\alpha}$		$\nu_{\text{deg}}(\gamma)$		ν_{unif}		ν_{Perron}	
	GSC$_{\text{un}}(t,\alpha)$	**GSC$_{\text{n}}(t,\alpha)$**	**deg-GSC$_{\text{un}}(\gamma)$**	**deg-GSC$_{\text{n}}(\gamma)$**	**uniform-GSC$_{\text{un}}$**	**uniform-GSC$_{\text{n}}$**	**perron-GSC$_{\text{un}}$**	**perron-GSC$_{\text{n}}$**
Breast Tissue	0.48 (1, 0.4)	0.48 (10, 0.4)	**0.49** (0.9)	**0.49** (0.6)	0.48	0.45	0.41	0.43
Iris	**0.80** (0, 0.0)	0.77 (7, 0.1)	**0.80** (0.0)	**0.80** (0.0)	**0.80**	**0.80**	0.03	0.77
Mnist64	0.91 (0, 0.0)	**0.95** (3, 0.5)	0.86 (1.0)	0.86 (1.0)	0.91	0.91	0.05	0.64
Olivetti Faces	0.68 (0, 0.0)	0.71 (9, 0.2)	0.69 (0.2)	0.69 (0.4)	0.68	**0.72**	0.14	0.32
Ph Recognition	0.55 (0, 0.0)	**0.56** (1, 0.2)	**0.56** (0.2)	**0.56** (0.1)	0.55	0.55	0.42	0.18
Seeds	0.74 (0, 0.0)	0.74 (3, 0.1)	0.74 (0.0)	**0.75** (0.3)	0.74	0.74	0.01	0.74
Segmentation	**0.64** (0, 0.0)	0.61 (12, 0.6)	0.61 (0.7)	0.61 (0.6)	0.64	**0.64**	0.35	0.34
Wdbc	0.68 (0, 0.0)	0.62 (22, 0.6)	0.68 (0.0)	0.68 (0.0)	0.68	0.68	0.00	**0.71**
Wine	**0.86** (0, 0.0)	0.85 (22, 0.1)	**0.86** (0.0)	0.85 (1.0)	**0.86**	**0.86**	0.01	0.82
PRB	0.98	0.97	0.97	0.97	0.98	0.98	0.27	0.75

(c) SUPERVISED EVALUATION (AMI SCORES | AMI-OPTIMIZED)

Dataset	$\nu_{t,\alpha}$		$\nu_{\text{deg}}(\gamma)$		ν_{unif}		ν_{Perron}	
	GSC$_{\text{un}}(t,\alpha)$	**GSC$_{\text{n}}(t,\alpha)$**	**deg-GSC$_{\text{un}}(\gamma)$**	**deg-GSC$_{\text{n}}(\gamma)$**	**uniform-GSC$_{\text{un}}$**	**uniform-GSC$_{\text{n}}$**	**perron-GSC$_{\text{un}}$**	**perron-GSC$_{\text{n}}$**
Breast Tissue	0.51 (21, 0.1)	**0.51** (1, 0.3)	0.49 (0.4)	0.50 (0.7)	0.48	0.45	0.41	0.43
Iris	0.80 (0, 0.0)	**0.88** (3, 1.2)	0.80 (0.0)	0.80 (0.0)	0.80	0.80	0.03	0.77
Mnist64	0.91 (0, 0.0)	**0.97** (15, 0.9)	0.91 (0.9)	0.91 (0.9)	0.91	0.91	0.05	0.64
Olivetti Faces	0.69 (2, 0.1)	**0.74** (2, 0.8)	0.73 (0.6)	0.73 (0.7)	0.68	0.72	0.14	0.32
Ph Recognition	0.55 (0, 0.0)	**0.56** (2, 0.2)	**0.56** (0.2)	**0.56** (0.1)	0.55	0.55	0.42	0.18
Seeds	0.74 (0, 0.0)	0.77 (2, 1.0)	**0.79** (1.0)	0.77 (0.9)	0.74	0.74	0.01	0.74
Segmentation	0.64 (0, 0.0)	**0.70** (13, 1.4)	0.66 (0.9)	0.66 (0.8)	0.64	0.64	0.35	0.34
Wdbc	0.68 (0, 0.0)	**0.73** (5, 0.8)	0.71 (0.8)	0.71 (0.6)	0.68	0.68	0.00	0.71
Wine	**0.86** (0, 0.0)	**0.86** (0, 0.0)	**0.86** (0.0)	**0.86** (0.0)	**0.86**	**0.86**	0.01	0.82
PRB	0.95	1.00	0.97	0.97	0.94	0.94	0.26	0.72

C Proofs

C.1 Proof of Prop. 3.1

Proof. We begin by expanding the Dirichlet form:

$$\begin{aligned}
\mathcal{D}_\nu(f) &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2 \\
&= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\left(f(i)^2+f(j)^2-2f(i)f(j)\right) \\
&= \sum_{i\in\mathcal{V}} \nu(i)f(i)^2 + \sum_{j\in\mathcal{V}}\left(\sum_{i\in\mathcal{V}}\nu(i)p(i,j)\right)f(j)^2 - 2\sum_{i,j\in\mathcal{V}}\nu(i)p(i,j)f(i)f(j) \\
&= \sum_{i\in\mathcal{V}} \nu(i)f(i)^2 + \sum_{j\in\mathcal{V}}\xi(j)f(j)^2 - \left(\sum_{i,j\in\mathcal{V}}\nu(i)p(i,j)f(i)f(j) + \sum_{i,j\in\mathcal{V}}\nu(j)p(j,i)f(i)f(j)\right) \\
&= \langle f, \mathbf{D}_\nu f\rangle + \langle f, \mathbf{D}_\xi f\rangle - \langle f, (\mathbf{D}_\nu\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{D}_\nu)f\rangle \\
&= \langle f, (\mathbf{D}_{\nu+\xi} - \mathbf{D}_\nu\mathbf{P}-\mathbf{P}^\mathsf{T}\mathbf{D}_\nu)f\rangle \\
&= \langle f, \mathbf{L}_\nu f\rangle \\
&= \langle f, \mathbf{D}_{\nu+\xi}^{-1}\mathbf{L}_\nu f\rangle_{\nu+\xi} \\
&= \langle f, \mathbf{L}_{\mathrm{RW}}(\nu) f\rangle_{\nu+\xi}
\end{aligned}$$

where we used the identity $\langle f, g\rangle_{\nu+\xi} = \langle f, (\mathbf{D}_\nu + \mathbf{D}_\xi)g\rangle$, and the definition of the generalized random walk Laplacian $\mathbf{L}_{\mathrm{RW}}(\nu)$ (see Def. 3.2).

As for the second equation, we can see that

$$\begin{aligned}
\mathcal{D}_\nu^2(f) &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2 \\
&= \frac{1}{2}\left(\sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2 + \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2\right) \\
&= \frac{1}{2}\left(\sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)[f(i)-f(j)]^2 + \sum_{i,j\in\mathcal{V}} \nu(j)p(j,i)[f(j)-f(i)]^2\right) \\
&= \frac{1}{2}\sum_{i,j\in\mathcal{V}} (\nu(i)p(i,j)+\nu(j)p(j,i))[f(i)-f(j)]^2.
\end{aligned}$$

Where we used the fact that $[f(j)-f(i)]^2 = [f(i)-f(j)]^2$ to symmetrize the expression. This shows that the GDE can be expressed as a sum over pairs of nodes, weighted by the symmetrized vertex-edge measure $\nu(i)p(i,j)+\nu(j)p(j,i)$, which captures the mutual influence between nodes i and j in both directions. □

C.2 Proof of Prop. 3.2

Proof. Recall that $\mathbf{L}_\nu = \mathbf{D}_{\nu+\xi} - \mathbf{D}_\nu\mathbf{P} - \mathbf{P}^\mathsf{T}\mathbf{D}_\nu$. For self-adjointness in $\ell^2(\mathcal{V})$, we compute:

$$\begin{aligned}
\langle f, \mathbf{L}_\nu g\rangle &= \langle f, (\mathbf{D}_{\nu+\xi} - \mathbf{D}_\nu\mathbf{P}-\mathbf{P}^\mathsf{T}\mathbf{D}_\nu)g\rangle \\
&= \langle f, \mathbf{D}_{\nu+\xi} g\rangle - \langle f, [\mathbf{D}_\nu\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{D}_\nu]g\rangle.
\end{aligned}$$

The first term is symmetric in f and g since $\mathbf{D}_{\nu+\xi}$ is a diagonal operator. For the second term, we have:

$$\langle f, [\mathbf{D}_\nu\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{D}_\nu]g\rangle = \sum_{i,j\in\mathcal{V}} f(i)g(j)(\nu(i)p(i,j)+\nu(j)p(j,i)) = \langle g, [\mathbf{D}_\nu\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{D}_\nu]f\rangle,$$

where we used that $\nu(i)p(i,j)+\nu(j)p(j,i)$ is symmetric in i and j. Combining both, gives $\langle f, \mathbf{L}_\nu g\rangle = \langle g, \mathbf{L}_\nu f\rangle$, which proves self-adjointness.

For positive semi-definiteness, we use the fact that $\mathbf{L}_\nu$ is self-adjoint and that the GDE is defined as $\mathcal{D}^2_\nu(f) = \langle f, \mathbf{L}_\nu f\rangle$. Since $\mathcal{D}^2_\nu(f) \geq 0$ for all f, it follows that $\langle f, \mathbf{L}_\nu f\rangle \geq 0$ for all f, which implies that $\mathbf{L}_\nu$ is positive semi-definite.

For the self-adjointness of $\mathbf{L}_{\mathrm{RW}}(\nu)$ in $\ell^2(\mathcal{V}, \nu+\xi)$, we can see that

$$\begin{aligned}
\langle f, \mathbf{L}_{\mathrm{RW}}(\nu) g\rangle_{\nu+\xi} &= \langle f, \mathbf{D}^{-1}_{\nu+\xi}\mathbf{L}_\nu g\rangle_{\nu+\xi} \\
&= \langle f, \mathbf{L}_\nu g\rangle \\
&= \langle g, \mathbf{L}_\nu f\rangle \\
&= \langle g, \mathbf{D}^{-1}_{\nu+\xi}\mathbf{L}_\nu f\rangle_{\nu+\xi} \\
&= \langle g, \mathbf{L}_{\mathrm{RW}}(\nu) f\rangle_{\nu+\xi}.
\end{aligned}$$

The same way, the quadratic form of $\mathbf{L}_{\mathrm{RW}}(\nu)$ is given by the Generalized Dirichlet energy, implying the positive semi-definiteness of $\mathbf{L}_{\mathrm{RW}}(\nu)$.

As for the normalized Laplacian $\overline{\mathbf{L}}_\nu = \mathbf{D}^{-1/2}_{\nu+\xi}\mathbf{L}_\nu\mathbf{D}^{-1/2}_{\nu+\xi}$, we can see that $\overline{\mathbf{L}}^{\mathsf{T}}_\nu = \overline{\mathbf{L}}_\nu$, which proves self-adjointness. For the semi-positive definiteness, we can see that $\langle f, \overline{\mathbf{L}}_\nu f\rangle = \langle \mathbf{D}^{-1/2}_{\nu+\xi} f, \mathbf{L}_\nu \mathbf{D}^{-1/2}_{\nu+\xi} f\rangle$, which is non-negative since $\mathbf{L}_\nu$ is positive semi-definite. □

C.3 Proof of Prop. 3.3.

Proof. As $\mathbf{P}$ is ergodic and μ is assumed to be a probability distribution, we have that $\nu_t = [\mathbf{P}^t]^{\mathsf{T}}\mu \to \pi$, let $\varepsilon_t = \pi - \nu_t$ be the distance of each element of ν_t to π. We have that $\|\varepsilon_t\| \xrightarrow[t\to\infty]{} 0$. Now, we can express the GDE of a graph function f as follows:

$$\begin{aligned}
\mathcal{D}^2_{\nu_t}(f) &= \sum_{i,j\in\mathcal{V}} \nu_t(i)p(i,j)[f(i)-f(j)]^2 \\
&= \sum_{i,j\in\mathcal{V}} (\pi(i)-\varepsilon_t(i))p(i,j)[f(i)-f(j)]^2 \\
&= \mathcal{D}^2_\pi(f) - \sum_{i,j\in\mathcal{V}} \varepsilon_t(i)p(i,j)[f(i)-f(j)]^2
\end{aligned} \tag{34}$$

Since $\|\varepsilon_t\|_1 \to 0$, it is clear that the second term vanishes. This result indicates that the GDE of a graph function f, associated to the transition matrix $\mathbf{P}$ and a parametrized measure ν_t, is the sum of a quadratic form involving the usual unnormalized Laplacian $\mathbf{L}_\pi$ (the Dirichlet energy $\mathcal{D}^2_\pi(f)$) and a vanishing term. Note that this vanishing term can also be seen as a quadratic form. □

C.4 Proof of Prop. 4.1

Proof. We begin by expressing the flow from S to $\bar{S}$:

$$\begin{aligned} q(S,\bar{S}) &= \sum_{i\in S,\, j\in\bar{S}} \nu(i)p(i,j) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_{\bar{S}}(j) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)(1-\chi_S(j)) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i) - \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_S(j) \\ \Rightarrow \quad q(S,\bar{S}) &= \sum_{i\in\mathcal{V}} \nu(i)\chi_S(i) - \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_S(j). \end{aligned} \tag{35}$$

Similarly, the flow from $\bar{S}$ to S is:

$$\begin{aligned} q(\bar{S},S) &= \sum_{i\in\bar{S},\, j\in S} \nu(i)p(i,j) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_{\bar{S}}(i)\chi_S(j) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)(1-\chi_S(i))\chi_S(j) \\ &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(j) - \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_S(j) \\ \Rightarrow \quad q(\bar{S},S) &= \sum_{j\in\mathcal{V}} \left(\sum_{i\in\mathcal{V}} \nu(i)p(i,j)\right)\chi_S(j) - \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_S(j). \end{aligned} \tag{36}$$

Now consider the Dirichlet form evaluated at the indicator vector χ_S:

$$\begin{aligned} \mathcal{D}_\nu(\chi_S) &= \sum_{(i,j)\in\mathcal{E}} \nu(i)p(i,j)\left|\chi_S(i)-\chi_S(j)\right|^2 \\ &= \sum_{i\in\mathcal{V}} \nu(i)\chi_S(i) + \sum_{j\in\mathcal{V}} \left(\sum_{i\in\mathcal{V}} \nu(i)p(i,j)\right)\chi_S(j) - 2\sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)\chi_S(i)\chi_S(j) \\ \Rightarrow \quad \mathcal{D}_\nu(\chi_S) &= q(S,\bar{S}) + q(\bar{S},S), \qquad \text{(by Eqs. 35 and 36).} \end{aligned}$$

□

C.5 Proof of Corr. 4.1

Proof. Let π be the stationary distribution of the natural random walk on $\mathcal{G}$, i.e. $\pi(j) = \sum_{i\in\mathcal{V}} \pi(i)\mathbf{P}(i,j)$ for all j. From the proof of Prop. 4.1, we have:

$$\begin{aligned} q(S,\bar{S}) - q(\bar{S},S) &= \sum_{i\in\mathcal{V}} \pi(i)\chi_S(i) - \sum_{j\in\mathcal{V}} \left(\sum_{i\in\mathcal{V}} \pi(i)\mathbf{P}(i,j)\right)\chi_S(j) \\ &= \sum_{i\in\mathcal{V}} \pi(i)\chi_S(i) - \sum_{j\in\mathcal{V}} \pi(j)\chi_S(j) = 0. \end{aligned}$$

Thus, $q(S,\bar{S}) = q(\bar{S},S)$, and by Prop. 4.1, we conclude that:

$$\mathcal{D}_\pi^2(\chi_S) = q(S,\bar{S}) + q(\bar{S},S) = 2q(S,\bar{S}).$$

□

C.6 Proof of Thm. 4.1

Proof. This proof is divided into two parts. First, we show that $\phi_\nu \geq \lambda_2/2$ and $\Phi_\nu(S) \leq \sqrt{2\lambda_2}$.

We start by noting that the Rayleigh quotient of a function f with respect to the normalized generalized Laplacian $\overline{\mathbf{L}}_\nu$ can be expressed as:

$$\mathcal{R}_{\overline{\mathbf{L}}_\nu}(f) = \frac{\langle f, \overline{\mathbf{L}}_\nu f\rangle}{\langle f, f\rangle} = \frac{\langle f, \mathbf{D}_{\nu+\xi}^{-1/2}\mathbf{L}_\nu\mathbf{D}_{\nu+\xi}^{-1/2} f\rangle}{\langle f, f\rangle} = \frac{\langle \mathbf{D}_{\nu+\xi} f, \mathbf{L}_\nu\mathbf{D}_{\nu+\xi} f\rangle}{\langle f, f\rangle}.$$

If we define $g = \mathbf{D}_{\nu+\xi}^{-1/2} f$, we can rewrite the Rayleigh quotient as:

$$\mathcal{R}_{\overline{\mathbf{L}}_\nu}(f) = \frac{\langle \mathbf{D}_{\nu+\xi}^{-1/2} f, \mathbf{L}_\nu\mathbf{D}_{\nu+\xi}^{-1/2} f\rangle}{\langle \mathbf{D}_{\nu+\xi}^{-1/2} f, \mathbf{D}_{\nu+\xi}^{-1/2} f\rangle} = \frac{\langle g, \mathbf{L}_\nu g\rangle}{\langle g, g\rangle_{\nu+\xi}},$$

where $\langle g, g\rangle_{\nu+\xi} = \langle g, \mathbf{D}_{\nu+\xi} g\rangle$. In particular, the second smallest eigenvalue λ_2 of $\overline{\mathbf{L}}_\nu$ can be characterized as:

$$\lambda_2 = \min_{f\perp\mathbf{1}} \mathcal{R}_{\overline{\mathbf{L}}_\nu}(f) = \min_{g\perp_{\nu+\xi}\mathbf{1}} \mathcal{R}_{\mathbf{L}_\nu}(g).$$

where the minimum is taken over all functions orthogonal to the constant function $\mathbf{1}$ for the inner product $\langle\cdot,\cdot\rangle_{\nu+\xi}$, i.e $\sum_{i\in\mathcal{V}}(\nu_i+\xi_i)g(i) = 0$. Throughout the proof, will denote by $\mathrm{vol}(S)$ the volume of a set S with respect to the measure $\nu+\xi$, i.e. $\mathrm{vol}(S) = \sum_{i\in S}(\nu(i)+\xi(i))$.

Lower-bound. We start by showing that $\phi_\nu \geq \lambda_2/2$. Let $S^* = \arg\min_{S\subset\mathcal{V}} \phi_\nu(S)$ be the optimal partition that achieves the minimum conductance, so that $\mathrm{vol}(S^*) \leq \mathrm{vol}(\mathcal{V})/2$. Consider the indicator function χ_{S^*} associated with this partition. Define the function $f = \chi_{S^*}\,\mathrm{vol}(\overline{S}^*) - \mathrm{vol}(S^*)\chi_{\overline{S}^*}$, which is orthogonal to the constant function $\mathbf{1}$ (in the weighted inner product space), indeed:

$$\langle f, \mathbf{1}\rangle_{\nu+\xi} = \langle \chi_{S^*}\,\mathrm{vol}(\overline{S}^*) - \mathrm{vol}(S^*)\chi_{\overline{S}^*}, \mathbf{1}\rangle_{\nu+\xi} = \mathrm{vol}(S^*)\,\mathrm{vol}(\overline{S}^*) - \mathrm{vol}(S^*)\,\mathrm{vol}(\overline{S}^*) = 0.$$

In particular by the Rayleigh quotient characterization, f is a valid candidate for the second eigenvector of $\mathbf{L}_\nu$. Notice that when i, j are part of the same set (either $i, j \in S^*$ or $i, j \in \overline{S}^*$), we have $f(i) - f(j) = 0$, and when i, j are part of different sets (i.e. $i \in S^*$ and $j \in \overline{S}^*$, or vice versa), the difference is non-zero. Therefore, the generalized Dirichlet energy of f can be computed as:

$$\begin{aligned}
\mathcal{D}_\nu^2(f) &= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)(f(i)-f(j))^2 \\
&= \sum_{i,j\in\mathcal{V}} \nu(i)p(i,j)(\chi_{S^*}(i)-\chi_{S^*}(j))^2(\mathrm{vol}(S^*)+\mathrm{vol}(\overline{S}^*))^2 \\
&= \mathcal{D}_\nu^2(\chi_{S^*})(\mathrm{vol}(S^*)+\mathrm{vol}(\overline{S}^*))^2 \\
&= (q(S^*,\bar{S}^*)+q(\bar{S}^*,S^*))(\mathrm{vol}(S^*)+\mathrm{vol}(\overline{S}^*))^2 && \text{(by Prop. 4.1)} \\
&= (q(S^*,\bar{S}^*)+q(\bar{S}^*,S^*))\,\mathrm{vol}(\mathcal{V})^2.
\end{aligned}$$

By definition of the conductance, it holds that $q(S^*,\bar{S}^*)+q(\bar{S}^*,S^*) = \phi_\nu(S^*)\,\mathrm{vol}(S^*)$. On the other end, the norm of f:

$$\begin{aligned}
\langle f, f\rangle_{\nu+\xi} &= \langle \chi_{S^*}\,\mathrm{vol}(\overline{S}^*) - \mathrm{vol}(S^*)\chi_{\overline{S}^*}, \chi_{S^*}\,\mathrm{vol}(\overline{S}^*) - \mathrm{vol}(S^*)\chi_{\overline{S}^*}\rangle_{\nu+\xi} \\
&= \sum_{i\in S^*}(\nu(i)+\xi(i))\,\mathrm{vol}(\overline{S}^*)^2 + \sum_{i\in\overline{S}^*}(\nu(i)+\xi(i))\,\mathrm{vol}(S^*)^2 \\
&= \mathrm{vol}(S^*)\,\mathrm{vol}(\overline{S}^*)(\mathrm{vol}(S^*)+\mathrm{vol}(\overline{S}^*)) \\
&= \mathrm{vol}(S^*)\,\mathrm{vol}(\overline{S}^*)\,\mathrm{vol}(\mathcal{V}).
\end{aligned}$$

Plugging these two expressions into the Rayleigh quotient of f, we have:

$$\mathcal{R}_{\mathbf{L}_\nu}(f) = \frac{\mathcal{D}_\nu^2(f)}{\langle f, f\rangle_{\nu+\xi}} = \frac{(q(S^*, \bar{S^*}) + q(\bar{S^*}, S^*))\,\mathrm{vol}(\mathcal{V})^2}{\mathrm{vol}(S^*)\,\mathrm{vol}(\overline{S}^*)\,\mathrm{vol}(\mathcal{V})} = \frac{\phi_\nu(S^*)\,\mathrm{vol}(\mathcal{V})}{\mathrm{vol}(\overline{S}^*)}.$$

Since $\mathrm{vol}(\overline{S}^*) \geq \mathrm{vol}(\mathcal{V})/2$, we have $\mathcal{R}_{\mathbf{L}_\nu}(f) \leq \phi_\nu(S^*)\,\mathrm{vol}(\mathcal{V})/(\mathrm{vol}(\mathcal{V})/2) = 2\phi_\nu(S^*)$. Since λ_2 is the minimum Rayleigh quotient over all functions orthogonal to the constant function, we have

$$\lambda_2 \leq \mathcal{R}_{\mathbf{L}_\nu}(f) \leq 2\phi_\nu(S^*) = 2\phi_\nu.$$

Upper bound. Let f_2 be the second eigenvector of $\mathbf{L}_\nu$, without loss of generality, we assume that the vertices are ordered so that $f_2(1) \geq f_2(2) \geq \cdots \geq f_2(n)$, by definition it holds that $\langle f_2, \mathbf{1}\rangle_{\nu+\xi} = 0$. We will assume that the graph has no self-loops. This implies that we can define level sets as:

$$S_k = \{i \in \mathcal{V} : f_2(i) \geq f_2(k)\} = \{1, 2, ..., k\}.$$

We define r to be the largest index such that $\mathrm{vol}(S_r) \leq \mathrm{vol}(\mathcal{V})/2$. And define g and h by:

$$g(i) = \begin{cases} f_2(i) - f_2(r) & \text{if } f_2(i) \geq f_2(r), \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad h(i) = \begin{cases} f_2(r) - f_2(i) & \text{if } f_2(i) < f_2(r), \\ 0 & \text{otherwise.} \end{cases}$$

Note that g and h are non-negative functions, and that $f_2 = g - h$. Moreover, g is supported on S_r, while h is supported on $\overline{S_r}$. Using the fact that $\lambda_2 = \mathcal{R}_{\mathbf{L}_\nu}(f_2)$, we have:

$$\begin{aligned}
\lambda_2 &= \frac{\langle f_2, \mathbf{L}_\nu f_2\rangle}{\langle f_2, f_2\rangle_{\nu+\xi}} \\
&= \frac{\langle g, \mathbf{L}_\nu g\rangle + \langle h, \mathbf{L}_\nu h\rangle}{\langle g, g\rangle_{\nu+\xi} + \langle h, h\rangle_{\nu+\xi}} && \text{(by the definition of } g \text{ and } h\text{)} \\
&\geq \min\left\{\frac{\langle g, \mathbf{L}_\nu g\rangle}{\langle g, g\rangle_{\nu+\xi}}, \frac{\langle h, \mathbf{L}_\nu h\rangle}{\langle h, h\rangle_{\nu+\xi}}\right\} && \text{(by the convexity of the Rayleigh quotient).}
\end{aligned}$$

Without loss of generality, we assume that $\mathcal{R}_{\mathbf{L}_\nu}(g) \leq \mathcal{R}_{\mathbf{L}_\nu}(h)$. In particular, we have that, $\lambda_2 \geq \mathcal{R}_{\mathbf{L}_\nu}(g)$. Now, we can use the fact that g is supported on S_r to upper bound $\mathcal{R}_{\mathbf{L}_\nu}(g)$ as follows:

$$\begin{aligned}
\mathcal{R}_{\mathbf{L}_\nu}(g) &= \frac{\langle g, \mathbf{L}_\nu g\rangle}{\langle g, g\rangle_{\nu+\xi}} \\
&= \frac{\sum_{i,j} \nu(i)p(i,j)(g(i) - g(j))^2}{\sum_i (\nu(i) + \xi(i))g(i)^2} \\
&= \frac{\sum_{i,j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) - g(j))^2}{2(\sum_i (\nu(i) + \xi(i))g(i)^2)} \\
&= \frac{\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) - g(j))^2}{\sum_i (\nu(i) + \xi(i))g(i)^2} \\
&= \frac{\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) - g(j))^2}{\sum_i (\nu(i) + \xi(i))g(i)^2} \times \frac{\sum_{i>j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) + g(j))^2}{\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) + g(j))^2} \\
&\geq \frac{\left(\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))|g(i)^2 - g(j)^2|\right)^2}{\sum_i (\nu(i) + \xi(i))g(i)^2 \times \sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) + g(j))^2},
\end{aligned}$$

Where the last inequality is due to the Cauchy-Schwarz inequality, and the factor 2 comes from Prop. 3.1. For the denominator, we can use the fact that $(g(i) + g(j))^2 \leq 2(g(i)^2 + g(j)^2)$, so that:

$$\begin{aligned}
\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i) + g(j))^2 &\leq 2\sum_{i<j} (\nu(i)p(i,j) + \nu(j)p(j,i))(g(i)^2 + g(j)^2) \\
&= 2\sum_{i,j} (\nu(i)p(i,j) + \nu(j)p(j,i))g(i)^2 \\
&= 2\sum_i (\nu(i) + \xi(i))g(i)^2.
\end{aligned}$$

As for the numerator, we can see that it can be expressed using the conductance of the level sets of f_2. Indeed, we have that:

$$\begin{aligned}\sum_{i<j}(\nu(i)p(i,j)+\nu(j)p(j,i))|g(i)^2-g(j)^2| &= \sum_{j=1}^{n}\sum_{i=1}^{j-1}(\nu(i)p(i,j)+\nu(j)p(j,i))|g(i)^2-g(j)^2| \\ &= \sum_{j=1}^{n}\sum_{i=1}^{j-1}(\nu(i)p(i,j)+\nu(j)p(j,i))\sum_{k=i}^{j-1}(g(k)^2-g(k+1)^2),\end{aligned}$$

Where the last equality is due to the fact that we can write $g(i)^2 = \sum_{k=1}^{i-1}(g(k)^2 - g(k+1)^2)$, and that $g(1) = 0$. Now, we can exchange the order of the summation to get:

$$\begin{aligned}\sum_{i<j}(\nu(i)p(i,j)+\nu(j)p(j,i))|g(i)^2-g(j)^2| &= \sum_{k=1}^{n-1}(g(k)^2-g(k+1)^2)\sum_{j=k+1}^{n}\sum_{i=1}^{k}(\nu(i)p(i,j)+\nu(j)p(j,i)) \\ &= \sum_{k=1}^{n-1}(g(k)^2-g(k+1)^2)(q(S_k,\bar{S_k})+q(\bar{S_k},S_k)).\end{aligned}$$

By definition of the generalized conductance, we have that for all k,

$$q(S_k,\bar{S_k})+q(\bar{S_k},S_k) \geq \phi_\nu(S_k)\min\{\mathrm{vol}(S_k),\mathrm{vol}(\bar{S_k})\}, \quad \text{and } \phi_\nu(S_k) \geq \phi_\nu.$$

Plugging this into the previous equation, we get:

$$\lambda_2 \geq \frac{\left(\sum_{k=1}^{n-1}(g(k)^2-g(k+1)^2)\phi_\nu(S_k)\min\{\mathrm{vol}(S_k),\mathrm{vol}(\bar{S_k})\}\right)^2}{2\langle g,g\rangle_{\nu+\xi}^2}.$$

By the fact that g is supported on S_r, we have that $\min\{\mathrm{vol}(S_k),\mathrm{vol}(\bar{S_k})\} = \mathrm{vol}(S_k)$ for all $k \leq r$ and that $g(k) = 0$ for all $k > r$. Moreover, since $\phi_\nu(S_k) \geq \phi_\nu$ for all k, we have that:

$$\lambda_2 \geq \frac{\left(\sum_{k=1}^{r}(g(k)^2-g(k+1)^2)\phi_\nu(S_k)\,\mathrm{vol}(S_k)\right)^2}{2\langle g,g\rangle_{\nu+\xi}^2} \geq \phi_\nu^2\frac{\left(\sum_{k=1}^{r}(g(k)^2-g(k+1)^2)\,\mathrm{vol}(S_k)\right)^2}{2\langle g,g\rangle_{\nu+\xi}^2}.$$

First, using the fact that $g(r+1) = 0$ (by definition of g), and defining $\mathrm{vol}(S_0) = 0$, we can rewrite the numerator as:

$$\begin{aligned}\sum_{k=1}^{r}(g(k)^2-g(k+1)^2)\,\mathrm{vol}(S_k) &= g(1)^2\,\mathrm{vol}(S_1)+\sum_{k=2}^{r}g(k)^2(\mathrm{vol}(S_k)-\mathrm{vol}(S_{k-1}))-g(r+1)^2\,\mathrm{vol}(S_r) \\ &= \sum_{k=1}^{r}g(k)^2(\mathrm{vol}(S_k)-\mathrm{vol}(S_{k-1}))\end{aligned}$$

Moreover, by the definition of the volume, we have that $\mathrm{vol}(S_k) - \mathrm{vol}(S_{k-1}) = \nu(k) + \xi(k)$, so that:

$$\sum_{k=1}^{r}(g(k)^2-g(k+1)^2)\,\mathrm{vol}(S_k) = \sum_{k=1}^{r}g(k)^2(\nu(k)+\xi(k)) = \langle g,g\rangle_{\nu+\xi}.$$

Putting this together, we have that:

$$\lambda_2 \geq \phi_\nu^2\frac{\left(\sum_{k=1}^{r}(g(k)^2-g(k+1)^2)\,\mathrm{vol}(S_k)\right)^2}{2\langle g,g\rangle_{\nu+\xi}^2} = \phi_\nu^2\frac{\langle g,g\rangle_{\nu+\xi}^2}{2\langle g,g\rangle_{\nu+\xi}^2} = \frac{\phi_\nu^2}{2}.$$

Implying that $\phi_\nu \leq \sqrt{2\lambda_2}$, which concludes the proof of the upper bound. □