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 datasets demonstrate that GSC consistently outperforms existing spectral clustering approaches in terms of clustering accuracy and robustness, offering a powerful new 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 node 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.

Another prominent source of digraphs arises from point-cloud data, where K-nearest neighbor (K-NN) graphs are commonly used. These graphs are typically asymmetric due to the directional nature of distance- or kernel-based construction methods (Genton, 2002). In this context, it is common to symmetrize the adjacency matrix and apply spectral clustering on the resulting undirected graph Laplacian (Satuluri & Parthasarathy, 2011). While convenient, this transformation can obscure directional patterns and 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 dynamics, 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 balanced clusters with high internal cohesion.

In this work, we revisit the density-based clustering paradigm and we propose a novel framework that generalizes classical spectral methods 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. It arises naturally from a random walk interpretation of graph partitioning. Moreover, we 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 $\mathbf{D}_\nu = \mathrm{diag}(\nu)$. 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.

Any graph function $\nu : \mathcal{V} \to \mathbb{R}_+$ assigning nonnegative values to vertices can be seen as a *positive vertex measure*, when $\sum_{i\in\mathcal{V}} \nu(i) = 1$, ν 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: $\mathbf{P} = \mathbf{D}_{\mathrm{out}}^{-1}\mathbf{W}$. Its spectrum lies within the unit disk, i.e. all its eigenvalues have absolute value at most 1. If $\mathcal{G}$ is strongly connected and aperiodic, then the random walk is ergodic, and 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 $t \to \infty$ (Brémaud, 2013). 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 $d_{\mathrm{out}}(i) = d_{\mathrm{in}}(i) = d(i)$.

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).

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

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

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

Eq. 3 encourages smooth functions over high-probability transitions. Eq. 4 links the Dirichlet energy to the Laplacian matrices $\mathbf{L}_{\mathrm{RW}}$, $\mathbf{L}$, $\overline{\mathbf{L}}$ (Chung, 2005; Sevi et al., 2023), which are fundamental to spectral methods.

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:

$$\mathbf{L}_{\mathrm{RW}} = \mathbf{I} - \mathbf{P}, \quad \mathbf{L} = \mathbf{D}_d - \mathbf{W}. \tag{8}$$

For balanced graphs (equal in- and out-degrees), these Laplacians reduce to those for undirected graphs using the symmetrized matrix $\mathbf{W}_{\mathrm{sym}} = \frac{1}{2}(\mathbf{W} + \mathbf{W}^\mathsf{T})$.

2.2 Connection with spectral clustering

The Dirichlet energy yields a Laplacian operator, connecting it to spectral methods. In spectral clustering, node 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 Dirichlet energy. This also reveals that minimizing the Dirichlet energy promotes smooth functions over a graph, assigning similar values to well-connected nodes. For binary clustering ($k = 2$), one may use $f = \chi_{V_1}$, the indicator of cluster V_1. This generalizes naturally to a multi-way partitioning (see Sec. 4).

3 The generalized Dirichlet energy framework

3.1 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 $\mathbf{P}$. This leads to a family of Laplacian operators that adapt to the graph structure and 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:*

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

This formulation is flexible: the graph structure is captured by the choice of edge measure q. For instance, setting $q(i,j) = \pi(i)p(i,j)$ where π is the stationary distribution of a random walk with transition matrix $\mathbf{P}$, is a special case yielding the classical Dirichlet energy:

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

More generally, we adopt the factorized form $q(i,j) = \nu(i)p(i,j)$, with ν being a tunable vertex measure. With this we define:

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

which enables new energy functionals with arbitrary positive vertex measures ν. Since scaling ν simply rescales the energy by $\|\nu\|_1^{-1}$, we may 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:*

$$\textit{Unnormalized:} \quad \mathbf{L}_\nu = \mathbf{D}_{\nu+\xi} - (\mathbf{D}_\nu\mathbf{P} + \mathbf{P}^\mathsf{T}\mathbf{D}_\nu), \tag{17}$$

$$\textit{Random walk:} \quad \mathbf{L}_{\mathrm{RW},\nu} = \mathbf{D}_{\nu+\xi}^{-1}\mathbf{L}_\nu, \tag{18}$$

$$\textit{Normalized:} \quad \overline{\mathbf{L}}_\nu = \mathbf{D}_{\nu+\xi}^{-1/2}\mathbf{L}_\nu\mathbf{D}_{\nu+\xi}^{-1/2}. \tag{19}$$

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:

$$\mathbf{L}_\pi = 2\mathbf{D}_\pi - (\mathbf{D}_\pi\mathbf{P} + \mathbf{P}^\mathsf{T}\mathbf{D}_\pi) = 2\mathbf{L}. \tag{20}$$

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

$$\mathcal{D}_\nu^2(f) = \langle f, \mathbf{L}_{\mathrm{RW},\nu} f\rangle_{\nu+\xi} = \langle f, \mathbf{L}_\nu f\rangle. \tag{21}$$

This implies that $\mathbf{L}_{\mathrm{RW},\nu}$ is self-adjoint in $\ell^2(\mathcal{V}, \nu+\xi)$, while $\mathbf{L}_\nu$ and $\overline{\mathbf{L}}_\nu$ are self-adjoint in $\ell^2(\mathcal{V},\nu)$.

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

$$\overline{\mathcal{D}}_\nu^2(f) = \frac{\mathcal{D}_\nu^2(f)}{\|f\|_{\nu+\xi}^2}, \quad \text{where } \|f\|_{\nu+\xi}^2 = \langle f, \mathbf{D}_{\nu+\xi} f\rangle. \tag{22}$$

This aligns with the Rayleigh quotient principle:

$$\mathcal{R}_{\mathbf{L}_\nu}(f) = \frac{\langle f, \mathbf{L}_\nu f\rangle}{\langle f, \mathbf{D}_{\nu+\xi} f\rangle}. \tag{23}$$

3.2 Parametrized vertex measure

We now construct vertex measures, denoted by ν, that capture local and global graph dynamics via random walks. The underlying idea is that, even when the stationary distribution π exists, other parametrizations can capture better the graph geometry. Starting from a uniform distribution, we define the following family of measures that depend on the diffusion time t:

$$\nu_t = [\mathbf{P}^t]^\mathsf{T}\tfrac{1}{N}\mathbf{1}_{N\times 1}. \tag{24}$$

Proposition 3.2. *Let $\nu_t = [\mathbf{P}^t]^\mathsf{T}\frac{1}{N}\mathbf{1}_{N\times 1}$, where $\mathbf{P}$ is ergodic. Then:*

$$\lim_{t\to\infty} \mathcal{D}_{\nu_t}^2(f) = \mathcal{D}_\pi^2(f) = \mathcal{D}^2(f). \tag{25}$$

Note that the above property holds for any probability measure $\nu = [\mathbf{P}^t]^\mathsf{T}\mu$, where μ is a probability vertex measure; in Eq. 24, though, μ is taken to be the uniform measure. Moreover, for any time-based vertex measure ν_t such that $\nu_t \xrightarrow{t\to\infty} \nu$, it holds that $\mathcal{D}_{\nu_t}^2(f) \xrightarrow{t\to\infty} \mathcal{D}_\nu^2(f)$.

On the top of the base measure ν_t, we can define the following more flexible parametrized measure: Given $t \in \mathbb{N}$ and $\alpha \in \mathbb{R}_+$,

$$\nu_{t,\alpha} = \left([\mathbf{P}^t]^\mathsf{T} \tfrac{1}{N}\mathbf{1}_{N\times 1}\right)^{\odot\alpha}, \quad \text{i.e. } \nu_{t,\alpha}(i) = \left([\mathbf{P}^t]^\mathsf{T} \tfrac{1}{N}\mathbf{1}_{N\times 1}\right)_i^{\alpha}. \tag{26}$$

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. 16 yields:

$$\mathcal{D}^2_{\nu_{t,\alpha}}(f) = \langle f, \mathbf{L}_{t,\alpha} f\rangle, \tag{27}$$

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.

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.

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 $S \cup U \neq \mathcal{V}$:

$$\begin{aligned} q(S,U) &= \sum_{i\in S, j\in U} q(i,j) \;= \sum_{i\in S, j\in U} \nu(i)p(i,j), \\ &= \sum_{i\in S, j\in U} \mathbb{P}(X_t = i)\mathbb{P}(X_{t+1} = j \,|\, X_t = i), \\ &= \mathbb{P}(X_t \in S, X_{t+1} \in U), \quad \text{for any } t \geq 0. \end{aligned} \tag{28}$$

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). In our setting, we generalize $q(S,U)$ for any transition matrix and any vertex measure, such that it quantifies the probability that the random walk escapes from the set S to U in one step, when the starting vertex of the walk is drawn according to the arbitrary vertex 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 Appendix A.3).

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:

$$q(S, \bar{S}) + q(\bar{S}, S) = \mathcal{D}_\nu^2(\chi_S). \tag{29}$$

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. 29 thus provides a natural interpretation of the generalized Dirichlet energy for graph partitions: the quantity $\mathcal{D}_\nu^2(\chi_S)$ 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 $\mathcal{D}_\nu^2(\chi_S) = \mathcal{D}_\nu^2(\chi_{\bar{S}})$. Under mild assumptions[1], one can further show that $\mathcal{D}_\nu^2(\chi_S)$ 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 Appendix A.4).

Corollary 4.1 (of Proposition 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:*

$$q(S, \bar{S}) = \frac{1}{2}\mathcal{D}_\pi^2(\chi_S). \tag{30}$$

This corollary reveals a key insight: minimizing the Dirichlet energy $\mathcal{D}_\pi^2(\chi_S)$ is equivalent to minimizing the total ergodic flow crossing between S and its complement $\bar{S}$. In other words, $\mathcal{D}_\pi^2(\chi_S)$ 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 $\mathcal{D}_\nu^2(\chi_S)$ can still be interpreted as finding a well-separated 2-partition of the graph under the dynamics induced by ν.

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:

$$\mathcal{D}_\nu^2(\boldsymbol{V}) = \sum_{\kappa=1}^k \mathcal{D}_\nu^2(\chi_{V_\kappa}), \tag{31}$$

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:

$$\min_{\boldsymbol{V}=\{V_1,\ldots,V_k\}} \mathcal{D}_\nu^2(\boldsymbol{V}) = \min_{\mathbf{U}} \ \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}_\nu\mathbf{U}) \quad \text{s.t.} \ \ u_\kappa = \chi_{V_\kappa}, \ \forall \kappa \in \{1, ..., k\}. \tag{32}$$

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:

$$\min_{\mathbf{U}} \ \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}_\nu\mathbf{U}) \quad \text{s.t.} \ \ \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_k, \tag{33}$$

[1]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. 17).
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** $\boldsymbol{V}_{t,\alpha}$

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. 3.2. Given a pair of parameter values ($t \in \mathbb{N}$, $\alpha \in \mathbb{R}_+$), Eq. 26 defines the vertex measure $\nu_{t,\alpha}$ that induces a generalized Dirichlet energy $\mathcal{D}^2_{\nu_{t,\alpha}}(\boldsymbol{V})$ (Eq. 31). This, in turn, leads to the generalized spectral graph partitioning objective of Eq. 33, based on the generalized Laplacian $\mathbf{L}_{t,\alpha}$ (see Def. 18). 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.

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. 3.2:

- GSC_{un} (t, α): using the unnormalized generalized Laplacian $\mathbf{L}_\nu$ (Eq. 17),
- GSC_{n} (t, α): using the normalized generalized Laplacian $\overline{\mathbf{L}}_\nu$ (Eq. 18),

with $\nu := \nu_{t,\alpha}$ being the parametrized vertex measure (Eq. 26). 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.

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.
- $\text{DI-SIM}_{\text{L}}(\tau)$, $\text{DI-SIM}_{\text{R}}(\tau)$, and $\text{DI-SIM}_{\text{C}}(\tau)$ (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 nodes and average degree $\bar{d}$: $\tau \in \left\{\text{round}\left(\bar{d}\cdot 10^s\right) \,\middle|\, s \in \{-1, -0.5, ..., 1\}\right\}$.
- SC_{un} and 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.
- *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 clusterings.

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

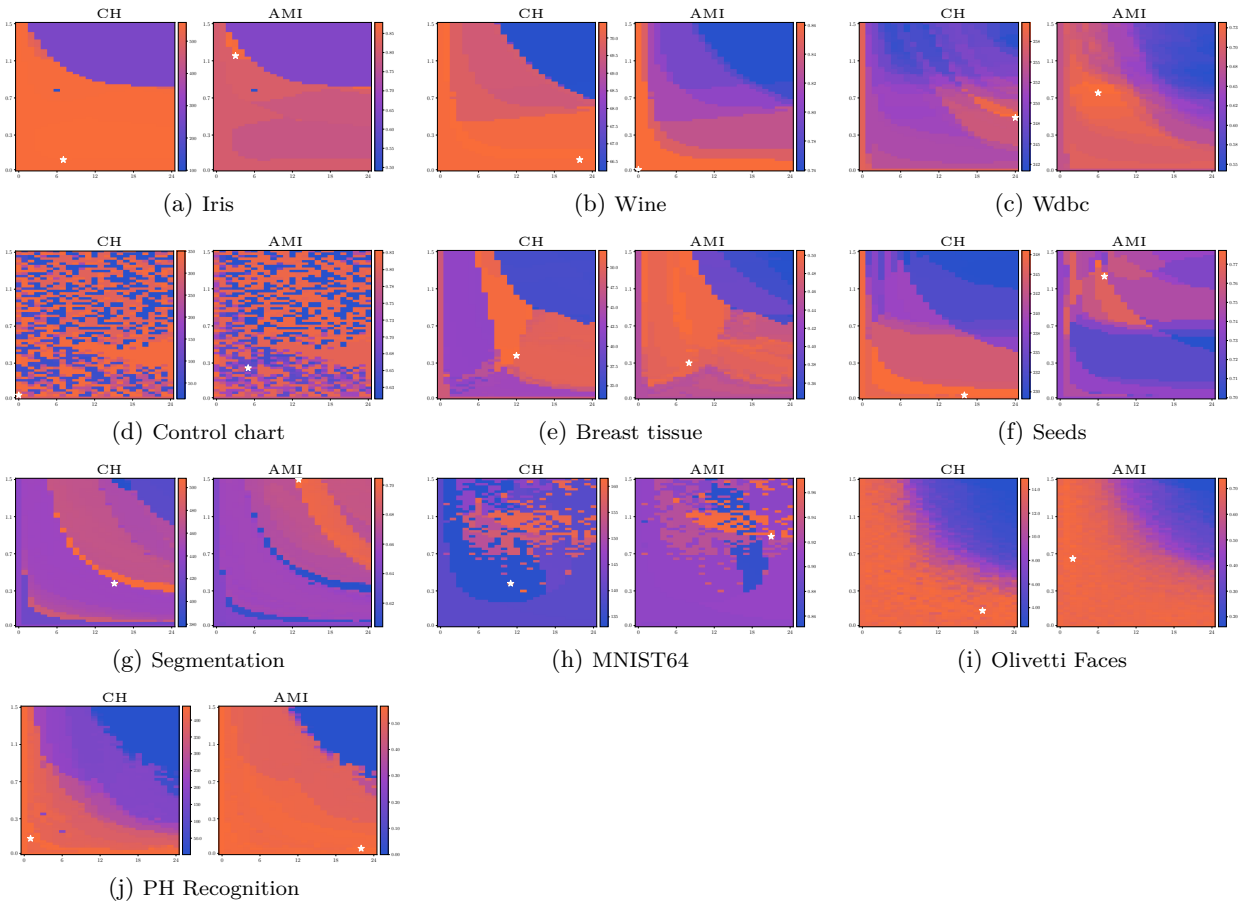
- *Model selection via internal validation*: Each method explores its respective hyperparameter space: GSC sweeps over (t, α), DSC+ over γ, and DI-SIM over τ. Each method performs 100 k-means++ runs per configuration of its hyperparameters, and the CH index is computed for each clustering solution. By 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 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 interpret better the results using an external evaluation index.

Graph construction from point-clouds. We test the clustering performance across 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.

5.2 Results

Sensitivity analysis for the parametrized vertex measure. The objective of the first step in our empirical evaluation is two-fold: we demonstrate how the quality of GSC clustering depends on the vertex measure, we also validate that an internal index such as CH can be used as a surrogate index for unsupervised model selection, through hyperparameter tuning. Fig. 1 illustrates the sensitivity of the vertex measure $\nu_{t,\alpha}$ (Eq. 26) 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. We can also see that 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. Contrary, 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 a fully unsupervised settings.

Model selection and internal evaluation. We employ a CH-based model selection to provide a fully unsupervised comparison of the different clustering algorithms. Numerical results are displayed in Tab. 1, which reports the best CH index of the clustering solutions found by all the compared methods. This is a purely unsupervised evaluation of the clustering quality: in the absence of label information, it is a protocol a user could use to choose the optimal hyperparameters for each method, and the overall 'best' solution, by relying to an internal index of choice. The average ranking of the methods across all datasets is also reported, where lower ranks indicate better performance. The competitive index quantifies how close a method is to the best performing method in each dataset, i.e. a value closer to 1 means that they are on average closer the best solutions. The results show that GSC_n can be efficiently guided to maximize a chosen internal measure,



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

(d) Control chart (e) Breast tissue (f) Seeds

(g) Segmentation (h) MNIST64 (i) Olivetti Faces

(j) PH Recognition

Figure 1: **GSC 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.

as it achieves the best average rank (1.5) in terms of CH. This highlights the strength of the normalized generalized Laplacian paired with the flexible vertex measure $\nu_{t,\alpha}$. GSC_{un} and SC_{un} both follow at an average rank of 3.30. SC_{n} and DI-SIM_{L} follow with average ranks around 3.9 and 4, respectively. DI-SIM_{C} follows closely with an average rank of 4.3. In contrast, DSC+ and DI-SIM_{R} tend to underperform, and are ranked on average lower than 5-th. The performance gap between GSC_{n} (rank 1.5) and GSC_{un} (rank 3.3) suggests that the normalized formulation allows for capturing better the data, in particular that the information in the proposed vertex measure might be better highlighted using the normalized variant of GSC.

Consistency with external evaluation. To improve the interpretability of the results, the clustering solutions identified as best by the previous unsupervised protocol (Tab. 1) can be evaluated post hoc using the ground-truth labels that come with each datasets. Tab. 2 shows the corresponding AMI scores. In those selected datasets where CH is a suitable proxy for AMI, GSC variants achieve the best or near-best AMI scores, validating the unsupervised model selection. On smaller, well-separated datasets (e.g. Iris, Wine), symmetrized spectral methods remain competitive, showing their adequacy in symmetric, low-noise settings. However, on graphs with asymmetry, imbalance, or noise (e.g. Control Chart), GSC's adaptive vertex measure and unnormalized formulation offer clear benefits. Overall, GSC_{un} achieves the best average AMI ranking (2), followed by SC variants (2.2 and 3.1), GSC_{n} (3.7), DI-SIM_{C} (3.80), DI-SIM_{L} (4.7), DI-SIM_{R} and lastly DSC+. This confirms again the advantages of employing the GDE framework in capturing complex graph structures.

Table 1: **Internal CH-based clustering evaluation.** The CH index of each method's solution. The best result per dataset appears bold and shaded; the estimated hyperparameters are in parentheses.

DATASETS				COMPETITORS						GSC	
Name	N	d	k	$\mathbf{SC_{un}}$	$\mathbf{SC_n}$	$\mathbf{DSC{+}}(\gamma)^*_{\mathrm{CH}}$	$\mathbf{DI\text{-}SIM_L}(\tau)^*_{\mathrm{CH}}$	$\mathbf{DI\text{-}SIM_R}(\tau)^*_{\mathrm{CH}}$	$\mathbf{DI\text{-}SIM_C}(\tau)^*_{\mathrm{CH}}$	$\mathbf{GSC_{un}}(t,\alpha)^*_{\mathrm{CH}}$	$\mathbf{GSC_n}(t,\alpha)^*_{\mathrm{CH}}$
Iris	150	4	3	555.67	555.67	22.71 (0.05)	501.63 (0.00)	501.63 (0.00)	501.63 (0.00)	555.67 (00, 0.00)	**558.06** (07, 0.10)
Wine	178	13	3	70.37	70.37	70.43 (0.15)	69.82 (16.00)	68.54 (0.00)	70.43 (16.00)	70.37 (00, 0.00)	**70.44** (22, 0.10)
WDBC	569	30	2	257.69	257.69	110.25 (0.85)	257.29 (1.00)	219.76 (1.00)	256.08 (1.00)	257.69 (00, 0.00)	**258.74** (22, 0.60)
Control Chart	600	60	6	341.50	290.57	166.90 (0.85)	341.50 (1.00)	335.25 (1.00)	341.50 (1.00)	275.10 (03, 0.20)	**350.42** (00, 0.10)
Breast Tissue	106	9	6	50.66	47.79	27.23 (0.05)	54.39 (0.00)	54.54 (0.00)	54.31 (1.00)	**54.78** (01, 0.40)	52.04 (10, 0.40)
Seeds	210	7	3	247.31	247.31	192.67 (0.80)	242.24 (16.00)	230.54 (0.00)	231.10 (0.00)	247.31 (00, 0.00)	**248.04** (03, 0.10)
Segmentation	2310	19	7	418.28	435.75	10.90 (0.20)	451.27 (1.00)	436.24 (2.00)	448.89 (1.00)	418.28 (00, 0.00)	**507.94** (12, 0.60)
MNIST64	1082	64	6	139.69	139.69	56.40 (0.20)	**159.48** (6.00)	143.06 (2.00)	149.34 (2.00)	139.69 (00, 0.00)	159.46 (20, 0.90)
Olivetti Faces	400	4096	5	14.41	14.33	10.68 (0.10)	13.95 (5.00)	11.08 (2.00)	12.62 (5.00)	14.41 (00, 0.00)	**14.83** (09, 0.20)
PH Recognition	653	3	6	436.85	436.85	8.08 (0.05)	46.45 (1.00)	38.30 (1.00)	41.65 (1.00)	436.85 (00, 0.00)	**438.91** (04, 0.10)
Average rank ▷				3.30	3.90	7.40	4.00	5.60	4.30	3.30	**1.50**
Competitiveness index ▷				0.955	0.938	0.433	0.876	0.819	0.855	0.944	**0.995**

Table 2: **External AMI-based clustering evaluation.** The CH index of each method's solution. The best result per dataset appears bold and shaded; the estimated hyperparameters are in parentheses.

DATASETS				COMPETITORS						GSC	
Name	N	d	k	$\mathbf{SC_{un}}$	$\mathbf{SC_n}$	$\mathbf{DSC{+}}(\gamma)^*_{\mathrm{AMI}}$	$\mathbf{DI\text{-}SIM_L}(\tau)^*_{\mathrm{AMI}}$	$\mathbf{DI\text{-}SIM_R}(\tau)^*_{\mathrm{AMI}}$	$\mathbf{DI\text{-}SIM_C}(\tau)^*_{\mathrm{AMI}}$	$\mathbf{GSC_{un}}(t,\alpha)^*_{\mathrm{AMI}}$	$\mathbf{GSC_n}(t,\alpha)^*_{\mathrm{AMI}}$
Iris	150	4	3	**0.803**	**0.803**	0.352 (0.05)	0.753 (0.00)	0.753 (0.00)	0.753 (0.00)	**0.803** (00, 0.00)	0.775 (07, 0.10)
Wine	178	13	3	0.862	0.862	0.860 (0.15)	0.833 (16.00)	**0.877** (0.00)	0.860 (16.00)	0.862 (00, 0.00)	0.846 (22, 0.10)
WDBC	569	30	2	**0.677**	**0.677**	0.328 (0.85)	0.662 (1.00)	0.639 (1.00)	**0.677** (1.00)	**0.677** (00, 0.00)	0.624 (22, 0.60)
Control Chart	600	60	6	0.806	0.767	0.629 (0.85)	0.806 (1.00)	0.796 (1.00)	0.806 (1.00)	**0.840** (03, 0.20)	0.824 (00, 0.10)
Breast Tissue	106	9	6	0.476	0.446	0.325 (0.05)	0.474 (0.00)	0.485 (0.00)	**0.506** (1.00)	0.478 (01, 0.40)	0.484 (10, 0.40)
Seeds	210	7	3	**0.737**	**0.737**	0.622 (0.80)	0.693 (16.00)	0.712 (0.00)	0.696 (0.00)	**0.737** (00, 0.00)	0.736 (03, 0.10)
Segmentation	2310	19	7	0.639	**0.643**	0.053 (0.20)	0.639 (1.00)	0.614 (2.00)	0.639 (1.00)	0.639 (00, 0.00)	0.606 (12, 0.60)
MNIST64	1082	64	6	0.907	0.907	0.483 (0.20)	0.961 (6.00)	0.880 (2.00)	0.956 (2.00)	0.907 (00, 0.00)	**0.972** (20, 0.90)
Olivetti Faces	400	4096	5	0.702	0.690	0.640 (0.10)	0.696 (5.00)	0.635 (2.00)	0.686 (5.00)	0.702 (00, 0.00)	**0.706** (09, 0.20)
PH Recognition	653	3	6	0.560	0.560	0.160 (0.05)	0.340 (1.00)	0.296 (1.00)	0.335 (1.00)	0.560 (00, 0.00)	**0.562** (04, 0.10)
Average rank ▷				2.30	3.10	7.60	4.70	5.30	3.80	**2.00**	3.70
Competitiveness index ▷				0.980	0.968	0.591	0.928	0.904	0.937	**0.985**	0.973

In summary, these 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. Finally, the disagreement between Tab. 1 and Tab. 2 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 present in a dataset.

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 *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. 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 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*, volume 107. 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.

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 Appendix

A.1 Proof of Proposition 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). □

A.2 Proof of Proposition 3.2.

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 $\epsilon_t = \pi - \nu_t$ be the distance of each element of ν_t to π.

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

Since $\|\epsilon_t\|_1 \to 0$, 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. □

A.3 Proof of Proposition 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:

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

□

A.4 Proof of Corollary 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:

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

□