Generalized Spectral Clustering for Directed and Undirected Graphs

Harry Sevi [1] **Matthieu Jonckeere** [2] **Argyris Kalogeratos** [1]

Abstract

Spectral clustering is a popular approach for clustering undirected graphs, but its extension to directed graphs (digraphs) is much more challenging. A typical workaround is to naively symmetrize the adjacency matrix of the directed graph, which can however lead to discarding valuable information carried by edge directionality. In this paper, we present a *generalized spectral clustering* framework that can address both directed and undirected graphs. Our approach is based on the spectral relaxation of a new functional that we introduce as the generalized Dirichlet energy of a graph function, with respect to an arbitrary positive regularizing measure on the graph edges. We also propose a practical parametrization of the regularizing measure constructed from the iterated powers of the natural random walk on the graph. We present theoretical arguments to explain the efficiency of our framework in the challenging setting of unbalanced classes. Experiments using directed K-NN graphs constructed from real datasets show that our graph partitioning method performs consistently well in all cases, while outperforming existing approaches in most of them.

1. Introduction

Clustering is one of the most popular techniques in analyzing large datasets and has widespread applications in machine learning, network analysis, and biology. Typically, when viewing the data as a graph, the problem of clustering is to partition the graph into several weakly interconnected clusters. This notion is formalized as a discrete optimization problem aiming to minimize a functional such as the graph cut or the normalized cut (Von Luxburg, 2007; Shi & Malik, 2000). The spectral relaxation of this minimization problem leads to finding the eigenvectors of a certain graph Laplacian and using these as input features for the k-means algorithm. This forms the backbone of spectral clustering.

Over the last three decades, spectral clustering has become one of the most widely used clustering methods due to its simplicity, efficiency, and strong theoretical background (Ng et al., 2002; Peng et al., 2015; Boedihardjo et al., 2021). Unfortunately, although many graphs carry valuable information in their directed edges, the vast majority of spectral clustering algorithms only operate on undirected graphs. Typical examples of directed graphs (digraphs) are social or content networks, as well as networks with flows (e.g. roads, electrical networks, rivers). Another fundamental source of digraphs in data processing comes from the representation of points clouds in d-dimensions through, e.g. K-nearest neighbors (K-NN) graphs or other kernel-based procedures.

Therefore, on the one hand, the information encoded in the edge directionality of digraphs should be used. On the other hand, the extension of the spectral clustering in the digraph setting is not straightforward. The adjacency matrix of a digraph is non-symmetric. It thus seems to be no obvious definition of a symmetric and real-valued graph Laplacian with a full set of real eigenvalues that uniquely encodes any digraph. The commonly used approach for clustering digraphs is to build a symmetrized adjacency matrix from the original non-symmetric one, and then to apply spectral clustering techniques to the graph Laplacian of it (Satuluri & Parthasarathy, 2011). As explained, this potentially discards valuable information.

In an attempt to overcome this challenging problem, a slew of works have been proposed in the last two decades. In (Zhou et al., 2005), they use the Laplacian on digraphs defined in (Chung, 2005) to propose spectral clustering on digraphs. In (Meilă & Pentney, 2007), they attack the clustering problem on digraphs from the original asymmetric adjacency matrix of a given digraph through the weighted cut formulation. In (Rohe et al., 2016), they propose a novel spectral co-clustering on digraphs based on the singular value decomposition of a modified adjacency matrix. In the recent years, some clustering approaches based on Hermitian operators on digraphs have been investigated (Cucuringu et al., 2020; Laenen & Sun, 2020). Note that all the approaches cited above are based on the construction of symmetric operators on digraphs.

[1]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190 Gif-sur-Yvette, France [2]CNRS, LAAS, 7 Av. du Colonel Roche, 31400 Toulouse,France. Correspondence to: Harry Sevi <harry.sevi@ens-paris-saclay.fr>.

,

In this paper, we present a unifying spectral clustering framework on directed and undirected graphs based on the spectral relaxation of novel energy functional, which in turn allows us to generalize graph Laplacians. This functional is termed *generalized Dirichlet energy* (GDE) as it extends the well-known notion of Dirichlet energy. In particular GDE is defined with respect to any positive regularizing measure and any Markov transition matrix. We propose for practical use a parametrized family of such measures. The resulting *generalized spectral clustering* (GSC) approach extends standard spectral clustering, usually applied on strongly connected digraphs (Zhou et al., 2005; Palmer & Zheng, 2020), to any digraphs. More importantly, it achieves that without involving the Pagerank's teleporting random walk (Page et al., 1999) as a surrogate of the natural random walk.

The rest of the paper is organized as follows. In Sec. 2, we present basic concepts of graph theory and Markov chains. In Sec. 3, we introduce the generalized Dirichlet energy and the new generalized graph Laplacians. In Sec. 4, we present the formulation of the GSC and an algorithmic scheme. In Sec. 5, we provide theoretical results proving the efficiency of GSC in the challenging setting of unbalanced clusters compared to classical spectral clustering. Sec. 6 includes our extensive experimental study on a toy dataset and various real-world UCI datasets. Technical proofs are provided in the Appendix.

2. Preliminaries and background

Essential concepts. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ be a weighted directed graph (digraph) where $\mathcal{V}$ is the finite set of $N = |\mathcal{V}|$ vertices, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the finite set of edges. Each edge (x, y) is an ordered vertex pair representing the direction of a link from vertex x to vertex y.

Any function $\nu : \mathcal{V} \to \mathbb{R}_+$, associating a nonnegative value to each graph vertex, can be regarded as a positive vertex measure; respectively any function $q : \mathcal{E} \to \mathbb{R}_+$ can be regarded as a positive edge measure.

The edge weight function $w : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ associates a nonnegative real value to every vertex pair: $w(x, y) \geq 0$, iff $(x, y) \in \mathcal{E}$, otherwise $w(x, y) = 0$.

The graph $\mathcal{G}$ can be entirely represented by its weighted adjacency matrix $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{N} \in \mathbb{R}_+^{N \times N}$, where $w_{ij} = w(x_i, x_j)$ is the weight of the edge (x_i, x_j). We define the out-degree and the in-degree of the i-th vertex by $d_i^+ = \sum_{j=1}^{N} w_{ij}$ and $d_i^- = \sum_{j=1}^{N} w_{ji}$, respectively. Also, the function $\mathrm{diag}(\nu)$ returns a square diagonal matrix with the elements of the input vector ν in its diagonal.

Given a subset of vertices $S \subseteq \mathcal{V}$, we denote its complement by $\bar{S} = \mathcal{V} \backslash S$. Also we denote the characteristic function of a set S by $\chi_S \in \{0, 1\}^N$, which gives $\chi_S(x) = 1$, iff $x \in S$, and $\chi_S(x) = 0$ otherwise.

Consider a graph function f mapping all of its vertices to an N-dimensional complex column vector :

$f = [f(x)]_{x \in \mathcal{V}}^{\mathsf{T}} \in \mathbb{C}^N$.

We assume that graph functions are defined in $\ell^2(\mathcal{V}, \nu)$, which is the Hilbert space of functions defined over the vertex set $\mathcal{V}$ of $\mathcal{G}$, endowed with the inner product, and ν is a positive measure. Hence, for all $f, g \in \ell^2(\mathcal{V}, \nu)$ it holds:

$$\langle f, g \rangle_\nu = \sum_{x \in \mathcal{V}} \overline{f(x)} g(x) \nu(x),$$

where $\overline{f(x)}$ denotes the complex conjugate of $f(x)$.

What we call in short as random walk on a weighted graph $\mathcal{G}$, is defined more formally as a natural random walk on the graph as a homogeneous Markov chain $\mathcal{X} = (X_t)_{t \geq 0}$ with a finite state space $\mathcal{V}$, and with state transition probabilities proportional to the edge weights.

The entries of the transition matrix $\mathbf{P} = [p(x, y)]_{x,y \in \mathcal{V}}$ are defined by:

$$p(x, y) = \mathbb{P}(X_{t+1} = y \,|\, X_t = x) = \frac{w(x, y)}{\sum_{z \in \mathcal{V}} w(x, z)}.$$

For a directed and strongly connected $\mathcal{G}$, the random walk $\mathcal{X}$ is irreducible. Under mild conditions, $\mathcal{X}$ is also ergodic, and therefore as $t \to \infty$, the measures $p^t(x, \cdot)$, $\forall x \in \mathcal{V}$, converge towards the *unique* stationary distribution denoted by the row vector $\pi \in \mathbb{R}_+^N$ (Brémaud, 2013). Within the undirected setting: $d_i^+ = d_i^- = d_i$, where $d \in \mathbb{R}_+^{N \times 1}$ is the vector of the vertex degrees; moreover, the ergodic distribution is proportional to the vertex degree distribution, i.e. $\pi \propto d$.

The emphasis of our presentation is put on digraphs, however, *our theoretical framework applies to* ***any*** *type of graph with nonnegative weights*.

Dirichlet energy and graph Laplacians.

In the literature of Dirichlet forms (Saloff-Coste, 1997; Montenegro et al., 2006) or harmonic analysis on graphs (Sevi et al., 2018), the definition of the Dirichlet energy of a graph function f is usually as follows.

Definition 2.1. ***Dirichlet energy of a graph function.*** *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$. Let also be the ergodic distribution π of the random walk, and $\pi(x)$ referring to vertex $x \in \mathcal{V}$. The Dirichlet energy of a graph function f is defined by:*

$$\mathcal{D}(f) = \sum_{x,y \in \mathcal{V}} \pi(x) p(x, y) |f(x) - f(y)|^2. \qquad (1)$$

This quantity can be also expressed in its quadratic form:

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

In this form, the Dirichlet energy reveals the random walk Laplacian $\mathbf{L}_{\text{RW}}$ and equivalently the unnormalized Laplacian $\mathbf{L}$ on directed graphs (Chung, 2005; Sevi et al., 2018). These matrices are formally defined as follows:

$$\mathbf{L}_{\text{RW}} = \mathbf{I} - \frac{1}{2}(\mathbf{P} + \mathbf{\Pi}^{-1}\mathbf{P}^{\mathsf{T}}\mathbf{\Pi}), \tag{2}$$

$$\mathbf{L} = \mathbf{\Pi} - \frac{1}{2}(\mathbf{\Pi}\mathbf{P} + \mathbf{P}^{\mathsf{T}}\mathbf{\Pi}), \tag{3}$$

where $\mathbf{I}$ is the identity matrix of suitable size (here N), and $\mathbf{\Pi} = \text{diag}(\pi)$ is the diagonal matrix associated with an ergodic measure π.

It is worth mentioning that $\mathbf{L}_{\text{RW}}$ and $\mathbf{L}$ on directed graphs, are the counterpart of the random walk Laplacian and unnormalized Laplacian on undirected graphs. For undirected graphs it holds $\mathbf{\Pi} \propto \mathbf{D} = \text{diag}(d)$ and $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$; therefore, in that case the random walk and unnormalized Laplacians become respectively: $\mathbf{L}_{\text{RW}} = \mathbf{I} - \mathbf{P}$ and $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

3. Generalized Dirichet energy and Laplacians on graphs

We have seen in the previous section the conventional way certain concepts appear in the literature. In this section, we introduce the generalized Dirichlet energy (GDE), which is defined under an arbitrary positive measure q over the graph edges, and the associated generalized graph Laplacians. These concepts constitute the foundation of our framework.

Definition 3.1. ***Generalized Dirichlet Energy of a graph function.*** *Let q be an arbitrary positive edge measure on a digraph $\mathcal{G}$, and $\mathbf{Q} = \{q(x,y)\}_{x,y\in\mathcal{V}}$ the edge measure operator. The generalized Dirichlet energy of a graph function f associated with the edge measure q on $\mathcal{G}$ is expressed as:*

$$\mathcal{D}^2_{\mathbf{Q}}(f) = \sum_{x,y\in\mathcal{V}} q(x,y)|f(x) - f(y)|^2. \tag{4}$$

The broad generality of this definition stems from the fact that it integrates all the graph-related information into the arbitrary positive edge measure q, thus its operator $\mathbf{Q}$. As can be noted, Definition 2.1 is a particular case of our generalized form, as $\mathcal{D}^2_{\mathbf{Q}}(f) = \mathcal{D}(f)$, when $q(x,y) = \pi(x)p(x,y)$. More generally, $q(x,y)$ can be a function combining an arbitrary vertex measure ν and an edge measure based on the transition matrix $\mathbf{P} = [p(x,y)]_{x,y\in\mathcal{V}}$ of the random walk $\mathcal{X}$ on $\mathcal{G}$. We refine accordingly the GDE of a graph function f associated with a random walk as:

$$\mathcal{D}^2_{\nu,\mathbf{P}}(f) = \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)|f(x) - f(y)|^2. \tag{5}$$

This formulation suggests that we can go further and derive interesting energy functionals by replacing the stationary distribution π of the random walk with other more sophisticated or better adapted vertex measures for digraphs. Note that, although ν can be an arbitrary measure, it is easy to see that replacing it by its ℓ^1-normalized counterpart $\nu' = \frac{\nu}{\|\nu\|_1}$ would merely scale the GDE of Eq. (5) by $\frac{1}{\|\nu\|_1}$. Therefore, we could safely restrict ν to be a probability vertex measure.

We are now ready to introduce the generalized graph Laplacians that rely on the GDE of Eq. (5).

Definition 3.2. ***Generalized graph Laplacians.*** *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$. Under an arbitrary positive vertex measure ν on $\mathcal{G}$, consider the positive vertex measure:*

$$\xi(y) = \sum_{x\in\mathcal{V}} \nu(x)p(x,y), \quad \forall y \in \mathcal{V}. \tag{6}$$

*Let be the diagonal matrices $\mathbf{N} = \text{diag}(\nu)$ and $\mathbf{\Xi} = \text{diag}(\xi)$. The **generalized random walk Laplacian** and the **unnormalized generalized Laplacian** on $\mathcal{G}$ are defined by:*

$$\mathbf{L}_{\text{RW}}(\nu) = \mathbf{I} - (\mathbf{I} + \mathbf{N}^{-1}\mathbf{\Xi})^{-1}(\mathbf{P} + \mathbf{N}^{-1}\mathbf{P}^{\mathsf{T}}\mathbf{N}), \tag{7}$$

$$\mathbf{L}(\nu) = \mathbf{N} + \mathbf{\Xi} - (\mathbf{N}\mathbf{P} + \mathbf{P}^{\mathsf{T}}\mathbf{N}). \tag{8}$$

$\mathbf{L}_{\text{RW}}(\nu)$ and $\mathbf{L}(\nu)$ extend the graph Laplacians defined in Eq. (2) and Eq. (3). Moreover, $\mathbf{L}(\nu)$ has the property of being self-adjoint in $\ell^2(\mathcal{V})$, i.e. $\mathbf{L}(\nu) = \mathbf{L}(\nu)^{\mathsf{T}}$.

Next, we establish the connection between the GDE and the generalized random walk Laplacian.

Proposition 3.1. *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$. Let ν be an arbitrary positive vertex measure on $\mathcal{G}$, and ξ be the vertex measure defined by Eq. (6). The generalized Dirichlet energy of a graph function f and the generalized random walk Laplacian $\mathbf{L}_{\text{RW}}(\nu)$ of Eq. (7) are associated as follows:*

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

As we can appreciate, $\mathbf{L}_{\text{RW}}(\nu)$ is self-adjoint in $\ell^2(\mathcal{V}, \nu+\xi)$.

Finally, we introduce the normalized GDE (also known as Rayleigh quotient) of a graph function f:

$$\overline{\mathcal{D}^2_{\nu,\mathbf{P}}}(f) = \frac{\mathcal{D}^2_{\nu,\mathbf{P}}(f)}{\|f\|^2_{\nu+\xi}}. \tag{9}$$

Given a positive vertex measure μ, we can introduce a parametrized vertex measure ν_t, $t \geq 0$, derived from the iterated powers of the natural random walk on a given graph:

$$\nu_t(x) = \mu^{\mathsf{T}}\mathbf{P}^t\delta_x\,, \quad \mu \in \mathbb{R}^N_+\,, \tag{10}$$

where $\delta_x \in \{0,1\}^{N\times 1}$ is the vector output of the Kronecker delta function at $x \in \mathcal{V}$. We can now derive the following proposition for the ergodic setting.

Proposition 3.2. *Let $\mathcal{X}$ be an ergodic random walk on a digraph $\mathcal{G}$, whose transition matrix is $\mathbf{P}$ with stationary distribution π. At $t \to \infty$, we have:*

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

This interesting result indicates that, as $t \to \infty$, the GDE associated with the transition matrix $\mathbf{P}$ of a graph function f, under a parametrized vertex measure ν_t, coincides with the respective energy of an ergodic random walk under the usual unnormalized Laplacian $\mathbf{L}(\pi)$.

4. Generalized spectral clustering on graphs

This section presents our general spectral clustering formulation for any type of graphs, which is based on the GDE and the generalized graph Laplacians.

4.1. Graph partitioning functional based on GDE

We commence with some reminders and additional preliminary concepts. Recall that $\mathcal{G}$ is a digraph of $N = |\mathcal{V}|$ vertices, and $\mathcal{X} = (X_t)_{t\geq 0}$ is a natural random walk on $\mathcal{G}$, with transition matrix $\mathbf{P} = [p(x,y)]_{x,y\in\mathcal{V}}$. In the general setting, $\mathcal{X}$ may be transient, thus not having an ergodic distribution. Let $\nu : \mathcal{V} \to \mathbb{R}_+$ be a vertex measure, and $\nu(S)$ be its evaluation over a subset $S \subseteq \mathcal{V}$: $\nu(S) = \sum_{x\in S} \nu(x)$.

Now, let $q : \mathcal{E} \to \mathbb{R}_+$ be a composite edge measure such that $q(x,y) = \nu(x)p(x,y)$. Respectively, consider the edge measure between two disjoint vertex subsets $S, U \subseteq \mathcal{V}$ by:

$$\begin{aligned} q(S,U) &= \sum_{x\in S, y\in U} q(x,y) \;=\; \sum_{x\in S, y\in U} \nu(x)p(x,y) \\ &= \mathbb{P}(X_t \in S, X_{t+1} \in U), \quad \text{for any } t \geq 0. \end{aligned} \tag{11}$$

$q(S,U)$ is a generic measure related to Markov chains (Sinclair, 1992; Levin & Peres, 2017). In our setting, 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 a set S and the rest of the graph (Meilă & Shi, 2001).

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 a 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 generalized Dirichlet energy $\mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S)$ are related as follows:*

$$q(S,\bar{S}) + q(\bar{S},S) = \mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S). \tag{12}$$

To bring this discussion closer to the clustering setting, we can imagine χ_S to be a *decision function* produced by some algorithm that aims to partition the graph into two parts. In that sense, Eq. (12) offers a meaningful interpretation of the GDE of any graph partitioning decision vector: $\mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S)$ quantifies how difficult it is for a random walk with transitions q to escape from S and reach $\bar{S}$, or vice versa.

Note also the symmetricity , i.e. $\mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S) = \mathcal{D}^2_{\nu,\mathbf{P}}(\chi_{\bar{S}})$.

The multiway graph partitioning problem aims to partition a digraph into a given number of disjoint subgraphs, such that the edge density among them is minimal. Given a k-partition of the graph vertices, denoted by $\boldsymbol{V} = \{V_i\}_{i=1}^k$, where $\bigcup_{i=1}^k V_i = \mathcal{V}$, then, under an arbitrary vertex measure ν, we define the *partition's Dirichlet energy* by:

$$\mathcal{D}^2_{\nu,\mathbf{P}}(\boldsymbol{V}) = \sum_{i=1}^k \mathcal{D}^2_{\nu,\mathbf{P}}(\chi_{V_i}). \tag{13}$$

This makes concrete our interest in finding a k-partition of the digraph $\mathcal{G}$ that has minimal GDE $\mathcal{D}^2_{\nu,\mathbf{P}}(\boldsymbol{V})$. Let us rewrite the energy of the k-partition as:

$$\mathcal{D}^2_{\nu,\mathbf{P}}(\boldsymbol{V}) = \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}(\nu)\mathbf{U}),$$

where $\mathbf{U} = [u_i]_{i=1}^k \in \mathbb{R}^{N\times k}$ is a matrix whose i-th column is $u_i = \chi_{V_i}$. Therefore, the *generalized Dirichlet graph partitioning problem* can be formulated as:

$$\min_{\boldsymbol{V}=\{V_1,\ldots,V_k\}} \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}(\nu)\mathbf{U}) \quad \text{s.t.}\, u_i = \chi_{V_i}, \forall i \in \{1,\ldots,k\}.$$

As mentioned earlier, the generalized Dirichlet graph partitioning problem is NP-hard. We thus proceed as in spectral clustering, by relaxing the combinatorial constraint of $\mathbf{U}$ and seeking instead a solution among all matrices $\mathbf{U}$ with orthonormal columns. For a given arbitrary measure ν, the relaxed problem becomes:

$$\min_{\mathbf{U}} \mathrm{tr}(\mathbf{U}^\mathsf{T}\mathbf{L}(\nu)\mathbf{U}) \quad \text{s.t.}\, \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_k, \tag{14}$$

whose solution $\mathbf{U}$ can be shown to be the eigenvectors corresponding to the k smallest eigenvalues of the unnormalized generalized Laplacian $\mathbf{L}(\nu)$.

The novelty of our framework lies in the definition of a generalized Laplacian associated with an arbitrary positive measure, and its connection to a general formulation of spectral clustering. In the case where ν is the stationary measure π, the spectral relaxation of the normalized Dirichlet energy Eq. (9) leads to the approach for strongly connected digraphs proposed by (Zhou et al., 2005).

4.2. The GSC algorithm

The framework we have presented so far led to Eq. (14), which relies on an arbitrary positive vertex measure ν to compute a set of k decision functions $\{\chi_{V_i}\}_{i=1}^{k}$ of minimal GDE. Each χ_{V_i} indicates one of the k pairwise disjoint vertex clusters versus the rest of the graph.

To render our framework more flexible for practical use, we extend what was previously described in Eq. (10) by introducing a parametrized vertex measure ν derived from the iterated powers of the natural random walk on a graph. Specifically, we consider three optional points of parametrization: i) the number of iterations $t \in \mathbb{N}$ of the random walk, ii) a uniform mixing parameter $\gamma \in [0,1]$ for the transition matrix, and iii) an exponent $\alpha \in \mathbb{R}$. Therefore, at a given vertex x, the proposed measure is given by:

$$\nu^{\alpha}_{(t,\gamma)}(x) = \left(\tfrac{1}{N}\mathbf{1}^{\mathsf{T}}_{N\times 1}\tilde{\mathbf{P}}^{t}_{\gamma}\delta_x\right)^{\alpha}, \tag{15}$$

where $\mathbf{1}_{N\times 1}$ is the all-ones vector , $\delta_x \in \{0,1\}^{N\times 1}$ is the vector output of the Kronecker delta function at $x \in \mathcal{V}$, and

$$\tilde{\mathbf{P}}_{\gamma} = \gamma\mathbf{P} + (1-\gamma)\tfrac{1}{N}\mathbf{1}_{N\times N}, \quad \gamma \in [0,1]. \tag{16}$$

Note that $\tilde{\mathbf{P}}_{\gamma}$ is a dense matrix, since its a convex combination of the original transition matrix and a uniform edge measure (in the form of a complete graph $\frac{1}{N}\mathbf{1}_{N\times N}$). Moreover, $\lim_{\gamma\to 1}\tilde{\mathbf{P}}_{\gamma} = \mathbf{P}$. Interestingly, $\tilde{\mathbf{P}}_{\gamma}$ makes us recall the teleporting random walk (Page et al., 1999); this connection is discussed in Sec. 5.1.

Plugging $\nu^{\alpha}_{(t,\gamma)}$ to Eq. (5), gives us the expression of the GDE of one decision function χ_{V_i} (seen as a graph function), under this new composite edge measure:

$$\mathcal{D}^{2}_{\nu^{\alpha}_{(t,\gamma)},\mathbf{P}}(\chi_{V_i}) = \langle \chi_{V_i}, \mathbf{L}^{(\alpha)}_{t,\gamma}\chi_{V_i}\rangle. \tag{17}$$

For a given α, and for any $t \in \mathbb{N}$, our derived *generalized spectral graph partitioning problem*, associated with the GDE of all the cluster-related decision functions, that we expressed earlier as $\mathcal{D}^{2}_{\nu^{\alpha}_{(t,\gamma)},\mathbf{P}}(\boldsymbol{V})$, is:

$$\min_{\mathbf{U}} \operatorname{tr}(\mathbf{U}^{\mathsf{T}}\mathbf{L}^{(\alpha)}_{t,\gamma}\mathbf{U}) \quad \text{s.t.}\, \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I}_k. \tag{18}$$

Simply put, the GSC algorithm employs the same optimization procedure (see Alg. 1) as the classical spectral clustering. The novelty is that we rely on a generalized graph Laplacian $\mathbf{L}^{(\alpha)}_{t,\gamma}$ and compute its eigenvectors $\mathbf{U}^{(\alpha)}_{t,\gamma} \in \mathbb{R}^{N\times k}$. Letting the random walk iteration (time) parameter vary, provides a set of eigenmaps $\{\mathbf{U}^{(\alpha)}_{t,\gamma}\}_{t=1}^{t_{\max}}$, for each of which we obtain a suggested graph k-partition using k-means. The different partitions can be compared according to suitable quality metrics to select the best one. It is also possible to obtain partitions of the same quality, which means there exists a subset of generalized Laplacian whose embeddings produce similar graph partitions results.

Algorithm 1 Generalized Spectral Clustering (GSC)

Input: $\mathbf{W}$: weighted adjacency matrix; k: number of clusters, γ: the uniform mixing parameter (see Eq. (16)); α: power (see Eq. (15)); $t_{\max}$: maximum number of power iterations (representing time) to perform over the transition matrix of the natural random walk.
Output: $\{\boldsymbol{V}^{(\alpha)}_{t,\gamma}\}_{t=1}^{t_{\max}}$: the graph k-partition for each time t.

1: **for** $t = 0$ **to** $t_{\max}$ **do**
2: Compute the generalized Laplacian $\mathbf{L}^{(\alpha)}_{t,\gamma}$, see Eq. (8).
3: Compute $\mathbf{U}^{(\alpha)}_{t,\gamma} \in \mathbb{R}^{N\times k}$ whose columns are the eigenvectors corresponding to the k smallest eigenvalues of $\mathbf{L}^{(\alpha)}_{t,\gamma}$.
4: Consider each $x_i \in \mathbb{R}^k$, $i = 1,...,N$, to be the embedding of the i-th vertex, represented by the i-th row of $\mathbf{U}^{(\alpha)}_{t,\gamma}$, and apply a clustering method (k-means) to all these vectors.
5: Obtain the k-partition $\boldsymbol{V}^{(\alpha)}_{t,\gamma} = \{V^{(\alpha)}_{j,t,\gamma}\}_{j=1}^{k}$ of the graph vertices based on the clustering result of Step 4.
6: **end for**
7: **return** $\{\boldsymbol{V}^{(\alpha)}_{t,\gamma}\}$, for all $t \in [0,...,t_{\max}]$.

5. Discussion

5.1. Misconception about the use of teleporting random walk for non-strongly connected digraphs

A frequently encountered misconception about how to deal with non-strongly connected digraphs in graph machine learning tasks, such as spectral clustering (Zhou et al., 2005) or node classification (Peach et al., 2020), concerns the use of the teleporting random walk (Page et al., 1999). This particular type of random walk is ergodic, however it is used as a substitute for the natural random walk that is generally non-ergodic in the digraph setting. In this sense, we realize that the teleporting random walk has been seen mainly as a trick to overcome non-ergodicity and be consistent with the standard ergodic theoretical framework.

Nevertheless, the use of the teleporting random walk as a direct proxy for the natural random walk may potentially bring disadvantages. Firstly, introducing teleportation, from any vertex to any other vertex, is equivalent to mixing with an unweighted complete graph. Consequently, this may modify drastically the graph topology and cause non-local perturbations to the random walk dynamics. Secondly, teleportation imposes ergodicity despite that may not be the case for the natural random walk on a given graph. Hence, the conclusions drawn when using this approach may be questionable (Schaub et al., 2019).

In our framework, we rather propose to incorporate teleportation as a regularizing measure of the GDE (see the involvement of $\tilde{\mathbf{P}}_{\gamma}$ in Eq. (15)) without changing the structure of the random walk itself (see that Eq. (18) minimizes $\mathcal{D}^{2}_{\nu^{\alpha}_{(t,\gamma)},\mathbf{P}}(\boldsymbol{V})$ that still depends on the original $\mathbf{P}$).

5.2. Why using measure regularized graph operators: the special case of unbalanced data

To explain why our generalized graph Laplacian operator can be instrumental to improve the performance of vanilla spectral clustering (VSC), we consider a toy model where the latter might fail depending on how unbalanced and separable the data classes are, whereas a reasonably tuned regularized operator will be successful.

We first analyze a mathematical caricature of a K-NN graph representing two clusters, and we then provide a simple experimental validation on a toy dataset. For simplicity, we consider an undirected graph (since the underlying stationary measure is explicit there and leads to easy explicit computations), but the argument could be generalized for directed graphs. Here we use as regularizing measure a simplified version of our proposal: $\nu(x)=\pi(x)^\alpha$, where π is the stationary measure proportional to vertex degrees.

Proposition 5.1. *Consider a two-cluster graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ of $|\mathcal{V}|=N$ vertices whose the ground truth is indicated by the sets V_1^* and V_2^* with cardinality $|V_1^*|=N_1^*$ and $|V_2^*|=N_2^*$, respectively, such that $V_1^*\cup V_2^*=\mathcal{V}, V_1^*\cap V_2^*=\emptyset$. Given a set $V\subset\mathcal{V}$, we define the internal frontier set of V as*

$$\partial^-(V)=\{x\in V:\exists y\in\bar{V}\,\text{s.t}\,(x,y)\in\mathcal{E}\}.$$

Let $N_1=N_1^*-\left|\partial^-(V_1^*)\right|$ and $N_2=N_2^*-\left|\partial^-(V_2^*)\right|$ be respectively the interior points of V_1^* and V_2^*. There is a frontier of $c_N=c\omega_N$ points, such that $N_1^*+N_2^*+c_N=N$. Assume further that the number of neighbors are constant, equal to ϵ_N inside the clusters, and equal to $\rho\epsilon_N$, $\rho<1$, along the frontier (this amounts to a separability assumption). We also assume that cutting inside the clusters leads to a frontier of order ω_N points. In the case where this latter is fulfilled, this leads to consider the subsets V_1 and V_2 with respectively N_1 and N_2 interior points such that $N_1+N_2+\omega_N=N$. Simple computations lead to the following property.

Assume $\frac{\omega_N}{N}\to 0$, $\frac{N_1^*}{N}\to b$, $\frac{N_2^*}{N}\to 1-b$.

Then if $c>b$, there exists a non empty set $V\subset V_2^*$ such that:

$$\overline{\mathcal{D}^2_{\pi,\mathbf{P}}}(\chi_{V_1^*})>\overline{\mathcal{D}^2_{\pi,\mathbf{P}}}(\chi_{V_1^*\cup V}).$$

On the other hand, if $\alpha>\frac{\log(b/c)}{\log(\rho)}$, then for all $V\subset V_2^*$

$$\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*})<\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V}).$$

We have hence shown that unbalanced data can reveal the inefficiency of the usual VSC, and that this can be corrected by a sufficient measure regularization. In Sec. 6.1 we validate this finding empirically using a relevant toy dataset.

Note that there are prior works motivating the operator regularization (Qin & Rohe, 2013; Amini et al., 2013; Zhang & Rohe, 2018), but they concern mainly the stochastic block model and not specifically the problem of unbalanced datasets. What we intended to stress in this subsection is that our theoretical GDE-based framework offers a new viewpoint to see and analyze such difficult data aspects.

6. Experiments

General setup. The experimental study concentrates on directed graphs naturally arising when processing point clouds. We consider input data of the form $X=\{x_i\}_{i=1}^N, \forall x_i\in\mathbb{R}^d$. Graph construction is a core phase of the graph partitioning pipeline and can affect the whole process. Based on X and pairwise point instances, we need to construct a generally sparse graph that is the first step towards representing the data. There are several options for this step, with different levels of sophistication and complexity. For instance, one could consider the simple yet natural approach of truncating a distance measure to create edges only for points that are sufficiently close to each other, e.g. via a K-NN or ε-graph. Alternatively, one could also employ parametrized kernels (e.g. a RBF) to create a similarity matrix.

Since the main focus of this work is on determining the right graph Laplacian operator, we rather choose a simple graph construction approach. We construct an unweighted directed K-NN graph that is represented by its non-symmetric adjacency matrix $\mathbf{W}=\{w_{ij}\}_{i,j=1}^N$, with entries:

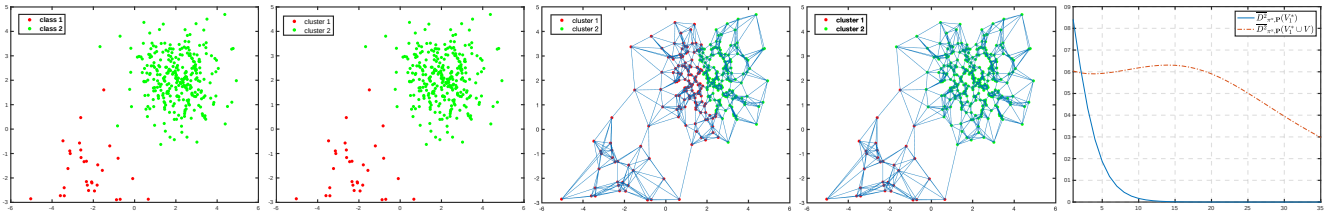
$$w_{ij}=\mathbb{1}\left\{\frac{|x_i-x_j|^2}{\text{dist}_K(x_i)^2}\leq 1\right\}. \tag{19}$$

In the above, $x_i\in\mathbb{R}^d$ represents the original coordinates of the data point corresponding to the i-th vertex, $\text{dist}_K(x)$ is the distance between x and its K-th-NN, and $\mathbb{1}\{\cdot\}\in\{0,1\}$ is the indicator function that evaluates the truth of the input condition. We always fix $K=\lceil\log(N)\rceil$, which makes the constructed graphs relatively sparse and not strongly connected.

6.1. Demonstration on a synthetic toy dataset

We first demonstrate empirically the insights discussed in Sec. 5.2. We generate a point could $X=\{x_i\}_{i=1}^N, \forall x_i\in\mathbb{R}^2$, of $N=330$ data points drawn independently from two classes using two Gaussians of different centers and same unit variance: $V_1^*: x_{i1},x_{i2}\sim\mathcal{N}(-2,1)$ or $V_2^*: x_{i1},x_{i2}\sim\mathcal{N}(2,1)$.

Let $N_1=30$ and $N_2=300$ be the number of points drawn from V_1^* and V_2^*, respectively. Here $K=\lceil\log(N)\rceil=6$. Exclusively for this case, we symmetrize the adjacency matrix $\mathbf{W}$ with $\tilde{\mathbf{W}}=\frac{1}{2}(\mathbf{W}+\mathbf{W}^\mathsf{T})$, in order to meet the conditions of the setting described in Sec. 5.2. Recall that $\chi_{V_1^*}\in\{0,1\}^N$ is the decision function deciding which vertices belong to the first cluster V_1^*. Let us set respectively



(a) Ground truth (b) k-means result (c) VSC result (d) GSC result (e) $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1})$ -vs- $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$

Figure 1. Comparison of VSC and GSC on an easy synthetic toy dataset with unbalanced classes. (a) Ground truth; (b) clustering result from k-means; (c) VCS result; (d) the result of the proposed GSC; (e) comparison of the quantity $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*})$ and $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$.

α_{th} and α_{xp} the theoretical and experimental exponent α we are looking to determine.

Fig. 1a shows the ground truth data classes. Fig. 1b shows the clustering result of the data obtained with k-means, which makes only one mistake w.r.t the ground truth and confirms that this is an easy scenario. Fig. 1c shows the clustering result obtained by VSC. We refer to the clusters obtained by VSC as sets V_1 and V_2, respectively. As we can observe, $V_1 = V_1^* \cup V$ where $V \subset V_2^*$. Consequently, let us define $\chi_{V_1^*\cup V}$ the decision function for the vertices belonging to V_1. Finally, Fig. 1d shows the clustering result obtained by GSC based on the generalized Laplacian $\mathbf{L}(\pi^\alpha)$ (see Eq. (8)). As observed, VSC fails completely to find the correct partition. On the other end, GSC recovers the same partition as the results from k-means applied directly to the data. To put this result in perspective to the toy model of Sec. 5.2, we compute the necessary parameters to get α. For $b = \frac{N_1}{N} \approx 0.08, \rho \approx 0.75, c \approx 0.29$, we obtain $\alpha_{\text{th}} \approx 4.5$. In order to validate α_{th}, we introduce Fig. 1e that compares the values of the generalized normalized Dirichlet energies $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*})$ and $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$ associated respectively of the decision functions $\chi_{V_1^*}$ and $\chi_{V_1^*\cup V}$ for increasing values of α (x-axis). We note that for $\alpha < 2.5$, $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*}) > \overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$. From the spectral clustering perspective, this means to choosing between the sets V_1^* and $V_1^* \cup V$ which minimizes the normalized Dirichlet energy i.e. $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$. When $\alpha > 2.5$, $\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*}) < \overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*\cup V})$ which is the equivalent of choosing V_1^*. Consequently, $\alpha_{\text{xp}} \approx 2.5 < \alpha_{\text{th}}$. We are able thus to recover a partition close to V_1^*, and this is what we achieve with GSC in Fig. 1d.

6.2. Results on benchmark datasets

This section reports the results of experiments we conducted to evaluate the performance of the proposed GSC method on 11 benchmark datasets from the UCI repository (Dheeru & Karra Taniskidou, 2017). We use three variants of the GSC method. The first one, denoted by $\text{GSC}_1(\gamma = 1, \alpha, t)$ is associated with the GDE defined in Eq. (17) with $\alpha \in [0, \infty), t \geq 0$ and $\gamma = 1$. The second one, denoted by $\text{GSC}_2(\gamma, \alpha, t)$ is also associated with the same GDE, but uses with $\alpha \in [0, \infty), t \geq 0$ and $\gamma \in [0, 1)$. The third one, denoted by $\text{GSC}_3(\gamma, \alpha, t)$ is the normalized version of $\text{GSC}_2(\gamma, \alpha, t)$ obtained thanks to Eq. (9).

GSC and all the competitors we mention below follow the spectral clustering setting but use the eigenvectors of different graph operators to finally apply k-means clustering (we report the best score out of 100 restarts). To ensure fair evaluations, we select for each method the optimal parameter values, obtained through cross-validation over a grid search, yielding to the closest partition to the ground truth. The grid used for each parameter of GSC is: $\alpha \in \{0, 0.1, ..., 1\}$, $t \in \{0, 1, ..., 100\}$, and $\gamma \in \{0, 0.05, ..., 0.95\}$.

Competitors. We compare against the following methods:

- DSC+(γ) (Zhou et al., 2005; Palmer & Zheng, 2020) spectral clustering on strongly connected digraphs. To extend this method to the graphs used in our experiments, we employ the teleporting random walk (Page et al., 1999) defined in Eq. (16) endowed with the parameter $\gamma \in [0, 1)$. We use the same cross-validation for γ as what mentioned earlier for GSC.
- DI-SIM$_{\text{L}}(\tau)$ and DI-SIM$_{\text{R}}(\tau)$ (Rohe et al., 2016) are two variants that are based on the left and the right singular vectors, respectively, of a given regularized and normalized operator whose regularization is denoted by the parameter $\tau \geq 0$. We use cross-validation to search the optimal parameter with a grid search over $\tau \in \{1, 2, ..., 20\}$.
- SC-SYM$_1$ and SC-SYM$_2$, two variants of the vanilla spectral clustering (Von Luxburg, 2007) based on the unnormalized and the normalized graph Laplacian obtained from the symmetrization of the adjacency matrix $\mathbf{W}$, respectively.

Results. The obtained partitions are first evaluated by the normalized mutual information (NMI) (Strehl & Ghosh, 2002) and the adjusted Rand index (ARI) (Hubert & Arabie, 1985). Both are supervised cluster evaluation measures that make use of the ground truth labels of the data. Also for both, larger values are better. Tab. 1 summarizes the comparative results based on NMI, while the ARI results are provided in the Appendix. In nearly all cases, we observe that the proposed GSC_1, GSC_2 and GSC_3 outperform significantly the other methods and GSC_1 gives the best result in average. Our approach performs much better than

Table 1. Clustering performance (NMI) on UCI datasets with optimal parameters in brackets.

DATASET	N	d	k	SC-SYM$_1$	SC-SYM$_2$	DI-SIM$_L$	DI-SIM$_R$	DSC+(γ)	GSC$_1$(γ=1,α,t)	GSC$_2$(γ,α,t)	GSC$_3$(γ,α,t)
IRIS	150	3	4	80.58	80.58	74.98 (1)	66.57 (1)	68.63 (0.80)	**90.11 (0.9,4)**	**90.11 (0.95,0.7,20)**	**90.11 (0.95,0.7,3)**
GLASS	214	9	6	38.59	38.92	38.95 (1)	36.41 (1)	39.72 (0.80)	45.73 (0.1,42)	**45.96 (0.95,1,14)**	38.56 (0.85, 0.1,73)
WINE	178	13	3	86.33	86.33	83.66 (1)	85.62 (1)	**91.09 (0.80)**	86.33 (0.1,1)	86.33 (0.95,0.1,53)	**91.09 (0.95,0.8,2)**
WBDC	569	30	2	67.73	69.47	68.54 (2)	53.43 (1)	61.12 (0.10)	72.02 (1,5)	**73.24 (0.95,0.8,3)**	71.45 (0.95,0.3,8)
CONTROL CHART	600	60	6	81.17	81.17	82.94 (1)	77.72 (1)	79.45 (0.90)	**85.62 (0.1,90)**	82.79 (0.95,0.3,65)	82.82 (0.90,0.7,96)
PARKINSON	185	22	2	21.96	19.13	28.89 (1)	27.36 (13)	25.82 (0.95)	32.65 (0.2,19)	**36.08 (0.95,0.4,10)**	31.93 (0.95,0.1,23)
VERTEBRAL	310	6	3	39.26	39.26	52.06 (2)	41.76 (2)	56.63 (0.80)	**64.26 (1,5)**	59.37 (0.95,1,1)	51.50 (0.85,1,15)
BREAST	106	9	6	54.03	54.43	54.04 (2)	49.33 (2)	51.64 (0.20)	56.66 (0.1,40)	**58.64 (0.95,1,56)**	56.40 (0.95,0.6,88)
SEEDS	210	7	3	73.90	73.90	76.29 (1)	73.06 (1)	74.80 (0.80)	**80.10 (1,4)**	**80.10 (0.95,0.9,4)**	**80.10 (0.95,0.9,4)**
IMAGE SEG.	2310	19	7	67.06	67.41	67.42 (1)	64.77 (1)	31.83 (0.99)	**73.40 (0.2,50)**	68.11 (0.95,1,64)	68.05 (0.95,0.1,56)
YEAST	1484	8	10	30.58	31.11	31.37 (2)	28.89 (1)	27.50 (0.90)	**37.46 (0.5,9)**	35.59 (0.95,1,40)	31.70 (0.95,0.4,67)
AVERAGE	–	–	–	58.29	58.34	59.92	54.77	56.37	**65.85**	65.12	63.06

Table 2. Clustering performance (NMI) on UCI datasets with optimal parameters in brackets using Calinski-Harabasz index.

DATASET	N	d	k	SC-SYM$_1$	SC-SYM$_2$	DI-SIM$_L$(τ)	DI-SIM$_R$(τ)	DSC+(γ)	GSC$_1$($\gamma=1,\alpha,t$)
IRIS	150	3	4	80.58	80.58	74.98 (1)	68.57 (1)	68.63 (0.85)	**83.66 (0.2,31)**
GLASS	214	9	6	38.59	38.92	37,39 (2)	35.87 (1)	36.58 (0.85)	**43.15 (0.1,38)**
WINE	178	13	3	86.33	86.33	83.66 (1)	82.02 (1)	63.16 (0.85)	86.33 (0.1,28)
WBDC	569	30	2	67.73	69.47	64.77 (1)	53.43 (1)	61.12 (0.10)	**69.47 (0.1,46)**
CONTROL CHART	600	60	6	81.17	81.17	82.94 (2)	77.44 (1)	79.45 (0.90)	**85.62 (0.4,17)**
PARKINSON	185	22	2	21.96	19.13	28.89 (2)	27.36 (13)	22.97 (0.30)	**31.10 (0.1,45)**
VERTEBRAL	310	6	3	39.26	39.26	45.89 (1)	39.62 (1)	**54.24 (0.80)**	51.83 (0.2,34)
BREAST TISSUE	106	9	6	54.03	54.43	54.04 (2)	49.27 (1)	51.64 (0.20)	**55.16 (0.1,24)**
SEEDS	210	7	3	73.90	73.90	76.26 (1)	73.06 (1)	74.80 (0.80)	**77.44 (0.8,2)**
IMAGE SEG.	2310	19	7	67.06	67.41	67.42 (1)	64.77 (1)	31.46 (0.99)	**69.60 (0.1,73)**
YEAST	1484	8	10	30.58	31.11	31.22 (1)	28.89 (1)	27.47 (0.90)	**32.16 (0.2,2)**
AVERAGE	–	–	–	58.29	58.34	58.86	54.57	51.95	**62.32**

SC-SYM$_1$ and SC-SYM$_2$, or VSC on the symmetric version of the directed K-NN graph. This allows us to state that our GSC associated with the GDE of Eq. (17), defined with respect to the vertex measure of Eq. (15), brings indeed real added value in the spectral clustering problem. It also allows obtaining better graph embeddings.

The case related to the WINE dataset is interesting to analyze. The highest NMI score is achieved by DSC+(γ), which outperforms GSC$_1$. Nevertheless, we remark that GSC$_3$ also achieves the highest NMI score on this dataset. This indicates that using the teleporting random walk as a regularizing measure is beneficial even without affecting the graph topology, unlike what DSC+(γ) does. Moreover, DSC+(γ) gives on average results below the symmetrized version, which suggests that DSC+(γ) can also indeed deteriorate the spectral clustering performance by considering the teleportation random walk instead of the natural one.

To further validate the efficiency of GSC, we also evaluated our framework without using the ground truth labels as an input. For this setting, we restrict the comparison of the proposed methods to GSC$_1$. Since our framework constructs a list of graph partitions, we use the Calinski-Harabasz (CH) (Caliński & Harabasz, 1974) as a measure of the quality of a partition of a dataset corresponding to the normalized ratio between the overall inter-cluster variance and the overall intra-cluster variance. We estimate the parameters α and t that maximize the CH index, to select a solution among all the obtained partitions. The results of the comparison are shown in Tab. 2. As noticed, GSC$_1$ outperforms significantly the other methods in nearly all cases, and on average outperforms significantly the other methods. Compared to the unsupervised evaluation reported in Tab. 1, here the NMI of GSC$_1$ stays lower by few percent. This indicates that the fully unsupervised version offers us comparable graph partition qualities to the case where we have the ground truth.

7. Conclusion

We have proposed the *generalized spectral clustering* (GSC) framework that applies to both directed and undirected graphs. First, we introduced the *generalized Dirichlet energy* (GDE), associated with an arbitrary positive edge measure as an extension of the classical Dirichlet energy for graph functions. Through the GDE formulation, we have proposed generalized Laplacian operators on graphs associated with an arbitrary positive vertex measure. We then provided a random walk interpretation of the GDE essential to our framework. Our proposal comes with an algorithm for our framework, where the vertex measure corresponds to the iterated powers of the natural random walk on the graph. We demonstrated theoretically that our framework is efficient in the unbalanced setting. Finally, numerical results showed that the GSC approach outperforms on several datasets existing approaches for directed graphs.

References

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, 41(4):2097–2122, 2013.

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

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

Caliński, T. and Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

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

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

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

Hubert, L. and Arabie, P. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

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

Levin, D. A. and Peres, Y. *Markov chains and mixing times*, volume 107. American Mathematical Society, 2017.

Meilă, M. and Pentney, W. Clustering by weighted cuts in directed graphs. In *Proceedings of the SIAM international Conference on Data Mining*, pp. 135–144, 2007.

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

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

Ng, A. Y., Jordan, M. I., and Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pp. 849–856, 2002.

Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

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

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

Peng, R., Sun, H., and Zanetti, L. Partitioning well-clustered graphs: Spectral clustering works! In *Conference on Learning Theory*, pp. 1423–1455. PMLR, 2015.

Qin, T. and Rohe, K. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *preprint arXiv:1309.4111*, 2013.

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

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

Satuluri, V. and Parthasarathy, S. Symmetrizations for clustering directed graphs. In *Proceedings of the International Conference on Extending Database Technology*, pp. 343–354, 2011.

Schaub, M. T., Delvenne, J.-C., Lambiotte, R., and Barahona, M. Multiscale dynamical embeddings of complex networks. *Physical Review E*, 99(6):062308, 2019.

Sevi, H., Rilling, G., and Borgnat, P. Harmonic analysis on directed graphs and applications: from Fourier analysis to wavelets. *preprint arXiv:1811.11636*, 2018.

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

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

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

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

Zhang, Y. and Rohe, K. Understanding regularized spectral clustering via graph conductance. *preprint arXiv:1806.01468*, 2018.

Zhou, D., Huang, J., and Schölkopf, B. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the International Conference on Machine learning*, pp. 1036–1043, 2005.

A. Technical proofs

Proposition 3.1. *Let $\mathcal{X}$ be a random walk on a digraph $\mathcal{G}$, with transition matrix $\mathbf{P}$. Let ν be an arbitrary positive vertex measure on $\mathcal{G}$, and ξ be the vertex measure defined by Eq. (6). The generalized Dirichlet energy of a graph function f and the generalized random walk Laplacian $\mathbf{L}_{\text{RW}}(\nu)$ of Eq. (7) are associated as follows:*

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

Proof.

$$\begin{aligned}
\mathcal{D}^2_{\nu,\mathbf{P}}(f) &= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)|f(x)-f(y)|^2 \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)|f(x)|^2 + \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)|f(y)|^2 - 2\sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)|f(x)||f(y)| \\
&= \langle f, \mathbf{N}f\rangle + \langle f, \mathbf{\Xi} f\rangle - \langle f, (\mathbf{NP}+\mathbf{P}^{\mathsf{T}}\mathbf{N})f\rangle \\
&= \langle f, \big(\mathbf{N}+\mathbf{\Xi}-(\mathbf{NP}+\mathbf{P}^{\mathsf{T}}\mathbf{N})\big)f\rangle \\
&= \langle f, \big(\mathbf{I}-(\mathbf{I}+\mathbf{N^{-1}\Xi})^{-1}(\mathbf{P}+\mathbf{N^{-1}P^{\mathsf{T}}N})\big)f\rangle_{\nu+\xi} \\
&= \langle f, \mathbf{L}_{\text{RW}}(\nu) f\rangle_{\nu+\xi}.
\end{aligned}$$

□

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 a 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 generalized Dirichlet energy $\mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S)$ are related as follows:*

$$q(S,\bar{S}) + q(\bar{S},S) = \mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S). \tag{12}$$

Proof.

$$\begin{aligned}
q(S,\bar{S}) &= \sum_{x\in S, y\in\bar{S}} \nu(x)p(x,y) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)\chi_{\bar{S}}(y) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)(1-\chi_S(y)) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x) - \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)\chi_S(y) \\
\Rightarrow\ q(S,\bar{S}) &= \sum_{x\in\mathcal{V}} \nu(x)\chi_S(x) - \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)\chi_S(y)
\end{aligned} \tag{20}$$

$$\begin{aligned}
q(\bar{S},S) &= \sum_{x\in\bar{S}, y\in S} \nu(x)p(x,y) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_{\bar{S}}(x)\chi_S(y) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)(1-\chi_S(x))\chi_S(y) \\
&= \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(y) - \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)\chi_S(y) \\
\Rightarrow\ q(\bar{S},S) &= \sum_{y\in\mathcal{V}} \Big(\sum_{x\in\mathcal{V}} \nu(x)p(x,y)\Big)\chi_S(y) - \sum_{x,y\in\mathcal{V}} \nu(x)p(x,y)\chi_S(x)\chi_S(y)
\end{aligned} \tag{21}$$

$$\begin{aligned}\mathcal{D}^2_{\nu,\mathbf{P}}(\chi_S) &= \sum_{(x,y)\in\mathcal{E}} \nu(x)p(x,y)|\chi_S(x)-\chi_S(y)|^2.\\ &= \sum_{x\in\mathcal{V}}\nu(x)\chi_S(x) + \sum_{y\in\mathcal{V}}\left(\sum_{x\in\mathcal{V}}\nu(x)p(x,y)\right)\chi_S(y) - 2\sum_{x,y\in\mathcal{V}}\nu(x)p(x,y)\chi_S(x)\chi_S(y)\\ &\Rightarrow (20)+(21)\end{aligned}$$

□

Proposition 3.2. *Let $\mathcal{X}$ be an ergodic random walk on a digraph $\mathcal{G}$, whose transition matrix is $\mathbf{P}$ with stationary distribution π. At $t\to\infty$, we have:*

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

Proof. $\mathbf{P}$ admits the eigen-decomposition $\mathbf{P}=\sum_{j=1}^N \lambda_j\phi_j\psi_j^*$ with $1=|\lambda_1|>|\lambda_2|\geq|\lambda_3|\geq\ldots\geq|\lambda_{N-1}|\geq -1$. Let μ be a positive vertex measure such that $\langle\mu,\phi_1\rangle=1$. The Dirichlet energy of a graph function f with respect to the measure $\nu_{(t)}$ (see Eq. (10)) has the following expression:

$$\mathcal{D}^2_{\nu_t,\mathbf{P}}(f) = \mathcal{D}^2_{\pi,\mathbf{P}}(f) + \langle f, \mathbf{E}_t f\rangle = \langle f, \big(\mathbf{L}(\pi)+\mathbf{E}_t\big)f\rangle,$$

where

$$\begin{aligned}\mathbf{E}_t &= \sum_{j\geq 2} c_{j,t}(\mu)\mathbf{Z}_j,\\ \mathbf{L}(\pi) &= \mathbf{\Pi} - \frac{1}{2}(\mathbf{\Pi P}+\mathbf{P}^\mathsf{T}\mathbf{\Pi}),\\ \mathbf{Z}_j &= \mathbf{\Psi}_j - \frac{1}{1+\lambda_j}(\mathbf{\Psi}_j\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{\Psi}_j)\\ \mathbf{\Psi}_j &= \mathrm{diag}(\psi_j)\\ c_{j,t}(\mu) &= (1+\lambda_j)\hat{\vartheta}_{j,t}(\mu)\\ \hat{\vartheta}_{j,t}(\mu) &= \langle\mu,\phi_j\rangle\lambda_j^t.\end{aligned}$$

$$\begin{aligned}\nu_t(x) &= \mu^\mathsf{T}\mathbf{P}^t\delta_x\\ &= \mu^\mathsf{T}\left(\sum_{j=1}^N \lambda_j^t\phi_j\psi_j^*\right)\delta_x\\ &= \mu^\mathsf{T}\left(\phi_1\pi^\mathsf{T} + \sum_{j\geq 2}^N \lambda_j^t\phi_j\psi_j^*\right)\delta_x\\ &= \langle\mu,\phi_1\rangle\pi(x) + \sum_{j\geq 2}^N \lambda_j^t\langle\phi_j,\mu\rangle\psi_j(x)\\ \nu_t(x) &= \pi(x) + \sum_{j\geq 2}^N \lambda_j^t\langle\phi_j,\mu\rangle\psi_j(x) \qquad (22)\end{aligned}$$

Remplacing the explicit form of $\nu_t(x)$ from Eq. (22) in $\mathcal{D}^2_{\nu_t,\mathbf{P}}(f)$ yields :

$$\begin{aligned}
\mathcal{D}^2_{\nu_t,\mathbf{P}}(f) &= \sum_{x,y\in\mathcal{V}} \nu(x,t)p(x,y)|f(x)-f(y)|^2, \\
&= \sum_{x,y\in\mathcal{V}} \left(\pi(x)+\sum_{j\geq 2}^{N}\lambda_j^t\langle\phi_j,\mu\rangle\psi_j(x)\right)p(x,y)|f(x)-f(y)|^2, \\
&= \sum_{x,y\in\mathcal{V}} \pi(x)p(x,y)|f(x)-f(y)|^2+\sum_{j\geq 2}^{N}\lambda_j^t\langle\phi_j,\mu\rangle\left(\sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)-f(y)|^2\right), \\
\mathcal{D}^2_{\nu_t,\mathbf{P}}(f) &= \mathcal{D}^2_{\pi,\mathbf{P}}(f)+\sum_{j\geq 2}^{N}\hat{\vartheta}_{j,t}(\mu)\mathcal{D}^2_{\psi_j,\mathbf{P}}(f). \qquad (23)
\end{aligned}$$

$$\begin{aligned}
\mathcal{D}^2_{\psi_j,\mathbf{P}}(f) &= \sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)-f(y)|^2 \qquad (24) \\
&= \sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)|^2+\sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(y)|^2-2\sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)||f(y)| \\
&= \sum_{x\in\mathcal{V}}\psi_j(x)|f(x)|^2+\lambda_j\sum_{y\in\mathcal{V}}\psi_j(x)|f(y)|^2-2\sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)||f(y)| \\
&= (1+\lambda_j)\sum_{x\in\mathcal{V}}\psi_j(x)|f(x)|^2-2\sum_{x,y\in\mathcal{V}}\psi_j(x)p(x,y)|f(x)||f(y)| \\
&= (1+\lambda_j)\langle f,\left(\mathbf{\Psi}_j-\frac{1}{(1+\lambda_j)}(\mathbf{\Psi}_j\mathbf{P}+\mathbf{P}^\mathsf{T}\mathbf{\Psi}_j)\right)f\rangle, \\
\mathcal{D}^2_{\psi_j,\mathbf{P}}(f) &= (1+\lambda_j)\langle f,\mathbf{Z}_j f\rangle. \qquad (25)
\end{aligned}$$

By putting Eq. (25) into Eq. (26), it yields

$$\begin{aligned}
\mathcal{D}^2_{\nu_t,\mathbf{P}}(f) &= \mathcal{D}^2_{\pi,\mathbf{P}}(f)+\sum_{j\geq 2}^{N}\hat{\vartheta}_{j,t}(\mu)\mathcal{D}^2_{\psi_j,\mathbf{P}}(f) \qquad (26) \\
&= \mathcal{D}^2_{\pi,\mathbf{P}}(f)+\sum_{j\geq 2}^{N}\hat{\vartheta}_{j,t}(\mu)(1+\lambda_j)\langle f,\mathbf{Z}_j f\rangle, \\
&= \mathcal{D}^2_{\pi,\mathbf{P}}(f)+\sum_{j\geq 2}^{N}c_{j,t}(\mu)\langle f,\mathbf{Z}_j f\rangle \\
\mathcal{D}^2_{\nu_t,\mathbf{P}}(f) &= \mathcal{D}^2_{\pi,\mathbf{P}}(f)+\langle f,\mathbf{E}_t f\rangle. \qquad (27)
\end{aligned}$$

At $t\to\infty$, we have

$$\lim_{t\to\infty}\mathbf{E}_t = 0.$$

Consequently, we have

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

Therefore, this result indicates that the GDE of a graph function f, associated with the transition matrix $\mathbf{P}$ and under a parametrized measure ν_t, is the sum of a quadratic form involving the usual unnormalized Laplacian $\mathbf{L}(\pi)$ and an operator $\mathbf{E}_t$ that tends to 0 as $t\to\infty$. □

Proposition 5.1. *Consider a two-cluster graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ of $|\mathcal{V}|=N$ vertices whose the ground truth is indicated by the sets V_1^* and V_2^* with cardinality $|V_1^*|=N_1^*$ and $|V_2^*|=N_2^*$, respectively, such that $V_1^*\cup V_2^*=\mathcal{V}, V_1^*\cap V_2^*=\emptyset$. Given a set $V\subset\mathcal{V}$, we define the internal frontier set of V as*

$$\partial^-(V)=\{x\in V:\exists\, y\in\bar{V}\ \text{s.t}\,(x,y)\in\mathcal{E}\}.$$

Let $N_1 = N_1^* - \left|\partial^-(V_1^*)\right|$ and $N_2 = N_2^* - \left|\partial^-(V_2^*)\right|$ be respectively the interior points of V_1^* and V_2^*. There is a frontier of $c_N = c\omega_N$ points, such that $N_1^* + N_2^* + c_N = N$. Assume further that the number of neighbors are constant, equal to ϵ_N inside the clusters, and equal to $\rho\epsilon_N$, $\rho < 1$, along the frontier (this amounts to a separability assumption). We also assume that cutting inside the clusters leads to a frontier of order ω_N points. In the case where this latter is fulfilled, this leads to consider the subsets V_1 and V_2 with respectively N_1 and N_2 interior points such that $N_1 + N_2 + \omega_N = N$. Simple computations lead to the following property.

Assume $\frac{\omega_N}{N} \to 0$, $\frac{N_1^*}{N} \to b$, $\frac{N_2^*}{N} \to 1-b$.

Then if $c > b$, there exists a non empty set $V \subset V_2^*$ such that:

$$\overline{\mathcal{D}^2_{\pi,\mathbf{P}}}(\chi_{V_1^*}) > \overline{\mathcal{D}^2_{\pi,\mathbf{P}}}(\chi_{V_1^* \cup V}).$$

On the other hand, if $\alpha > \frac{\log(b/c)}{\log(\rho)}$, then for all $V \subset V_2^*$

$$\overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^*}) < \overline{\mathcal{D}^2_{\pi^\alpha,\mathbf{P}}}(\chi_{V_1^* \cup V}).$$

Proof. Given the assumption on our toy model graph:

$$\overline{\mathcal{D}^2_{\pi^\alpha}}(\chi_{V_1^*}) = \frac{c_N(\rho\epsilon_N)^\alpha}{N_1\epsilon_N^\alpha + o(N_1\epsilon_N^\alpha)} = \frac{\omega_N}{N} c\frac{\rho^\alpha}{a} + o\Big(\frac{\omega_N}{N}\Big). \tag{28}$$

Now, with $V = F \cup H$, where $H \in V_2^*$ and $|V_1^* \cup V| = \tilde{a}N + o(N)$:

$$\overline{\mathcal{D}^2_{\pi^\alpha}}(\chi_{V_1^* \cup V}) = \frac{\omega_N(\epsilon_N)^\alpha}{N\tilde{a}\epsilon_N^\alpha + o(N\epsilon_N^\alpha)} = \frac{\omega_N}{N}\frac{1}{\tilde{a}} + o\Big(\frac{\omega_N}{N}\Big). \tag{29}$$

Hence if $c\rho > b$, $\overline{\mathcal{D}^2_{\pi}}(\chi_{V_1^* \cup V}) < \overline{\mathcal{D}^2_{\pi}}(\chi_{V_1^*})$ for some V, whereas if $c\rho^\alpha < b$, $\overline{\mathcal{D}^2_{\pi^\alpha}}(\chi_{V_1^* \cup V}) > \overline{\mathcal{D}^2_{\pi^\alpha}}(\chi_{V_1^*})$, $\forall V$. □

B. Additional experimental results

Table 3. Clustering performance (ARI) on UCI datasets.

DATASET	N	d	k	SC-SYM$_1$	SC-SYM$_2$	DI-SIM$_L$	DI-SIM$_R$	DSC+(γ)	GSC$_1$($\gamma=1,\alpha,t$)	GSC$_2$(γ,α,t)	GSC$_3$(γ,α,t)
IRIS	150	3	4	75.92	75.92	69.41(1)	58.44 (1)	52.96 (0.80)	**92.22 (0.9,4)**	**92.22 (0.95,0.7,20)**	**92.22 (0.95,0.7,3)**
GLASS	214	9	6	23.12	24.80	22.05 (1)	18.89 (1)	20.93 (0.80)	28.00 (0.1,42)	**28.01 (0.95,0.9,22)**	26.85 (0.85,0.7,52)
WINE	178	13	3	87.82	87.82	84.98 (1)	89.74 (1)	**92.95 (0.80)**	87.82 (0.1,1)	87.92 (0.95,0.1,53)	**92.95 (0.95,0.8,2)**
WBDC	569	30	2	76.69	77.30	77.95 (2)	57.75 (1)	64.58 (0.30)	80.48 (1,4)	**81.12 (0.95,0.8,3)**	78.56 (0.95,0.3,8)
CONTROL	600	60	6	62.25	62.25	66.79 (1)	59.83 (1)	60.04 (0.90)	**71.91 (0.1,90)**	64.77 (0.95,0.3,65)	64.78 (0.90.0.7,96)
PARKINSON	185	22	2	35.42	32.91	28.90 (1)	24.82 (1)	24.82 (0.95)	42.26 (1,3)	**43.73 (0.95,4,10)**	40.28 (0.95,0.2, 60)
VERTEBRAL	310	6	3	29.70	29.70	38.85 (2)	31.03 (2)	54.64 (0.80)	**63.14 (1,5)**	58.40 (0.95,1,1)	38.56 (0.85,1,9)
BREAST TISSUE	106	9	6	36.96	38.41	41.94 (2)	30.60 (2)	21.76 (0.90)	41.01 (0.1,40)	38.02 (0.95,0.6,77)	**44.69 (0.95,0.4,96)**
SEEDS	210	7	3	78.46	78.46	81.09 (1)	74.41 (1)	77.64 (0.80)	**83.52 (1,5)**	**83.52 (0.95,0.9,4)**	**83.52 (0.95,0.9,4)**
IMAGE SEG.	2310	19	7	47.83	51.75	50.83 (1)	36.54 (1)	08.89 (0.99)	52.15 (0.2,50)	**61.17 (0.95,1,78)**	55.81 (0.95,0.4,41)
YEAST	1484	8	10	19.41	21.17	19.97 (2)	19.49 (2)	16.62 (0.90)	**28.48 (0.5,9)**	26.85 (0.95,0.9,93)	21.78 (0.95,0.4,19)
AVERAGE	–	–	–	52.14	52.77	52.98	45.59	48.70	59.95	**60.52**	58.18