Prospects for Applied Mathematics and Data Analysis

♦ ISSN: 2836-4449 Vol: 06, No. 01, 2026 ♦ PP. 07–14
♦ Article DOI: https://doi.org/10.54216/PAMDA.060102

FULL LENGTH ARTICLE

Heat-Semigroup Persistence on Data Graphs: A Multiscale Extension of Normalized Cut for Class-Separability Analysis

Nader Taffach[1,*] **Mohammad Al-Shiekh**[1]

[1] Department of Mathematics, Idlib University, Idlib, Syria

Emails: nader_mahmoud_taffach@idlib.edu.sy; mohammad_al_sheikh@idlib.edu.sy

Received: October 24, 2025 Revised: November 16, 2025 Accepted: January 02, 2026 ⋆ Corresponding author

ABSTRACT

Let $G = (V,W)$ be a weighted data graph with symmetric normalized Laplacian $L = I - D^{-1/2}WD^{-1/2}$, and let u denote the degree-balanced signal associated with a binary partition $C \cup \bar{C} = V$. Instead of reducing the partition geometry to the single Rayleigh quotient $u^\top Lu/\|u\|_2^2$, we study the heat-semigroup persistence

$$P_u(t) = \frac{\|e^{-tL}u\|_2^2}{\|u\|_2^2}, \qquad H_T(u) = \frac{1}{T}\int_0^T P_u(t)\,dt.$$

Writing $L\phi_j = \lambda_j\phi_j$ and $\omega_j = |\langle u,\phi_j\rangle|^2/\|u\|_2^2$ yields $P_u(t) = \sum_j \omega_j e^{-2t\lambda_j}$, so the complete curve is the Laplace transform of the label spectral measure $\nu_u = \sum_j \omega_j \delta_{\lambda_j}$. We prove four identities that give this construction a cut-theoretic interpretation. First, P_u is completely monotone. Second, $-P_u'(0)/2 = \mathrm{Ncut}(C,\bar{C})$. Third, for the instantaneous leakage rate $\kappa_u(t) = -\frac{1}{2}\,d\log P_u(t)/dt$, one has $\kappa_u(0) = \mathrm{Ncut}$ and $\kappa_u'(t) = -2\,\mathrm{Var}_{\nu_{u,t}}(\lambda) \le 0$ under the exponentially tilted spectral measure. Fourth, when $u \perp \ker L$, $\int_0^\infty P_u(t)dt = u^\top L^\dagger u/(2\|u\|_2^2)$. Hence normalized cut is only the zero-time slope of a multiscale diffusion object whose higher derivatives recover all spectral moments. A perturbation bound $|H_T(L) - H_T(\widetilde{L})| \le T\|L - \widetilde{L}\|_2$ is also established for a fixed partition signal. Numerical evaluation on a 1,797-sample, 64-variable handwritten-digit benchmark uses all 45 class pairs and ten repeated stratified train/test splits. With graphs formed exclusively from training observations, mean H_1 has Spearman correlation -0.924 with held-out pairwise error (95% bootstrap interval $[-0.957, -0.850]$); normalized cut gives 0.927, and the second spectral central moment gives 0.930. The comparable predictive rankings are material: the proposed functional is not presented as a replacement for normalized cut, but as its multiscale completion, retaining spectral information that a first moment necessarily discards.

Keywords: Ggraph Laplacian ▪ Heat semigroup ▪ Normalized cut ▪ Spectral measure ▪ Diffusion geometry ▪ Class separability

1. INTRODUCTION

For a weighted graph with nonnegative symmetric adjacency matrix $W = [w_{ij}]$, degree matrix $D = \mathrm{diag}(d_1,\dots,d_n)$, and normalized Laplacian

$$L = I - D^{-1/2}WD^{-1/2}, \qquad 0 \preceq L \preceq 2I, \tag{1}$$

a binary partition is commonly summarized by a cut energy. If u is the degree-balanced partition signal defined below, the normalized-cut value is exactly the Rayleigh quotient

$$\mathscr{R}_L(u) = \frac{u^\top L u}{u^\top u}. \tag{2}$$

This scalar is mathematically natural, but it is also a severe compression: in the eigendecomposition $L = \sum_j \lambda_j \phi_j \phi_j^\top$, equation (2) retains only the first moment of the energy distribution $|\langle u, \phi_j \rangle|^2$. Two partitions can therefore have the same Rayleigh quotient while placing their spectral mass very differently over $[0,2]$.

The present work replaces the one-moment description by the heat evolution of the same signal. For $t \geq 0$, define

$$S_t = e^{-tL}, \qquad P_u(t) = \frac{\|S_t u\|_2^2}{\|u\|_2^2}. \tag{3}$$

The quantity $P_u(t)$ measures the fraction of squared partition energy remaining after diffusion time t. The construction introduces no learned classifier, loss function, or auxiliary embedding. It is determined entirely by (L,u) and therefore permits exact analysis. If

$$\omega_j = \frac{|\langle u, \phi_j \rangle|^2}{\|u\|_2^2}, \qquad \sum_{j=1}^n \omega_j = 1, \tag{4}$$

then

$$P_u(t) = \sum_{j=1}^n \omega_j e^{-2t\lambda_j}. \tag{5}$$

Consequently, P_u is not merely a diffusion score. It is a Laplace transform of a probability measure on the graph spectrum. Its derivatives at the origin recover spectral moments; its logarithmic derivative describes a time-dependent leakage rate; and its integral is connected to the Laplacian pseudoinverse.

This perspective is useful because modern graph constructions are increasingly treated as approximations of geometric operators rather than as arbitrary affinity matrices. Convergence results now cover kernel graphs, ε-graphs, k-nearest-neighbor graphs, and self-tuned graph Laplacians under increasingly precise assumptions [1, 2, 3, 4]. Regularity and continuum-limit analyses further connect graph Laplacians to partial differential equations and variational learning problems [5, 6, 7]. In this setting, it is natural to ask not only whether a partition has low cut energy, but how its label signal is distributed across the approximating operator's spectrum.

The mathematical development below establishes that normalized cut is the initial slope of P_u, while the rate at which this slope changes is exactly a spectral variance under an exponential tilt. More generally,

$$\frac{(-1)^m}{2^m} P_u^{(m)}(0) = \sum_j \omega_j \lambda_j^m, \qquad m = 0, 1, 2, \ldots, \tag{6}$$

so the complete heat-persistence curve contains every moment of the label spectral measure. Because the measure has compact support, these moments determine it uniquely. A finite-horizon area,

$$H_T(u) = \frac{1}{T} \int_0^T P_u(t)\, dt, \tag{7}$$

is then used as a stable scalar summary when a full curve is inconvenient.

The numerical part is intentionally subordinate to the analysis. It asks whether the quantities derived from P_u reflect empirical pairwise class difficulty on a real data graph. Training-only graphs are formed for every pair of handwritten digits under repeated stratified splits; the resulting graph quantities are compared with an external held-out 5-nearest-neighbor error. This design avoids using test labels to construct the graph and separates the proposed functional from the classifier used only for evaluation. The results show that H_1, normalized cut, and the second spectral central moment all track difficult class pairs closely. That near-equivalence in rank prediction is reported rather than obscured: the mathematical gain is spectral resolution, not an inflated claim of predictive dominance.

2. RELATED WORK

Spectral graph partitioning converts combinatorial separation into an operator problem. Recent analysis has sharpened performance guarantees for spectral clustering and its scalable variants [8, 9]. Extensions based on nonlinear or p-spectral objectives seek more flexible cut geometries [10], while Cheeger-type limits quantify how discrete graph cuts approach continuum isoperimetric problems [11]. Graph regularization has likewise developed beyond the ordinary Dirichlet energy, including ℓ_p formulations and fractional powers of graph-Laplacian operators [12, 7]. These results motivate treating the graph spectrum as an analytical object rather than using an eigenspace only as a preprocessing step.

The graph itself is a source of approximation error. Self-tuned kernels, Gaussian-kernel eigensystems, doubly stochastic affinity learning, and robust manifold scaling have all been studied as ways to control density bias, heterogeneous noise, clustering structure, and finite-sample effects [13, 3, 4, 14, 15]. Sharp spectral convergence on neighborhood graphs and regularity of graph-Laplacian eigenvectors provide complementary justification for finite-sample spectral calculations [2, 5]. The graph construction used here follows the self-tuned-neighborhood principle, but the proposed quantity does not depend on a particular kernel family.

Diffusion offers a different use of the same spectrum. Diffusion-map theory studies semigroup-like smoothing and spectral embeddings of manifold data [1, 16], while time-inhomogeneous diffusion has been used to expose multiscale geometry and topology [17]. Heat dynamics also enter signal recovery through space–time sampling [18]. The present construction differs in purpose: the initial condition is a degree-balanced class signal, and the object of interest is its decay profile under a fixed normalized Laplacian.

Spectral summaries other than eigenvalue locations can also carry structural information. Matrix spectral entropy, for example, compresses an entire spectrum through a nonlinear functional [19]. Active and semi-supervised graph methods similarly exploit operator geometry to quantify uncertainty or propagate sparse labels [20, 6]. Here, no labels are propagated. Instead, the labels define a signal whose spectral measure is examined directly. The resulting identities connect a familiar cut functional to the complete heat response of that signal.

3. MATHEMATICAL FORMULATION

3.1 Degree-balanced partition signal

Let $G = (V,W)$ be an undirected weighted graph with $V = \{1,\ldots,n\}$, $W = W^\top \geq 0$, $w_{ii} = 0$, and $d_i = \sum_j w_{ij} > 0$. For $A \subset V$, write

$$\mathrm{vol}(A) = \sum_{i\in A} d_i, \qquad \mathrm{cut}(A,A^c) = \sum_{i\in A,\, j\in A^c} w_{ij}.$$

For a nontrivial partition $C \cup \bar{C} = V$, define

$$z_i = \begin{cases} \mathrm{vol}(C)^{-1}, & i \in C, \\ -\mathrm{vol}(\bar{C})^{-1}, & i \in \bar{C}, \end{cases} \qquad u = D^{1/2}z. \tag{8}$$

Then $u^\top D^{1/2}\mathbf{1} = 0$. If G is connected, $D^{1/2}\mathbf{1}$ spans $\ker L$, so $u \perp \ker L$.

Proposition 1 (Normalized cut as a Rayleigh quotient)**.** *For the signal in* (8),

$$\frac{u^\top Lu}{\|u\|_2^2} = \mathrm{cut}(C,\bar{C})\left(\frac{1}{\mathrm{vol}(C)} + \frac{1}{\mathrm{vol}(\bar{C})}\right) = \mathrm{Ncut}(C,\bar{C}). \tag{9}$$

Proof. The normalized-Laplacian quadratic form gives

$$u^\top Lu = \frac{1}{2}\sum_{i,j} w_{ij}\left(\frac{u_i}{\sqrt{d_i}} - \frac{u_j}{\sqrt{d_j}}\right)^2.$$

Only edges crossing the partition contribute. Hence

$$u^\top Lu = \mathrm{cut}(C,\bar{C})\left(\frac{1}{\mathrm{vol}(C)} + \frac{1}{\mathrm{vol}(\bar{C})}\right)^2.$$

Moreover,

$$\|u\|_2^2 = \sum_{i\in C}\frac{d_i}{\mathrm{vol}(C)^2} + \sum_{i\in\bar{C}}\frac{d_i}{\mathrm{vol}(\bar{C})^2} = \frac{1}{\mathrm{vol}(C)} + \frac{1}{\mathrm{vol}(\bar{C})}.$$

Division yields (9). □

3.2 Heat persistence and label spectral measure

Let $0 \leq \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of L and $\{\phi_j\}$ an orthonormal eigenbasis. Define

$$\omega_j = \frac{|\langle u,\phi_j\rangle|^2}{\|u\|_2^2}, \qquad \nu_u = \sum_{j=1}^n \omega_j \delta_{\lambda_j}. \tag{10}$$

The measure ν_u is a probability measure on $[0,2]$. Equations (3) and (10) imply

$$P_u(t) = \int_{[0,2]} e^{-2t\lambda}\, d\nu_u(\lambda). \tag{11}$$

Thus the partition is represented by a scalar completely monotone function rather than by one cut value.

Theorem 1 (Complete monotonicity and spectral moments)**.** *For every integer $m \geq 0$ and $t \geq 0$,*

$$(-1)^m P_u^{(m)}(t) = 2^m \int \lambda^m e^{-2t\lambda}\, d\nu_u(\lambda) \geq 0. \tag{12}$$

In particular,

$$-\frac{1}{2}P_u'(0) = \mathrm{Ncut}(C,\bar{C}), \qquad \frac{1}{4}P_u''(0) = \int \lambda^2 d\nu_u(\lambda). \tag{13}$$

Moreover, the full function $P_u(t)$ for $t \geq 0$ uniquely determines ν_u.

Proof. Differentiation of the finite sum (11) yields (12). At $t = 0$, the first moment is $\int \lambda d\nu_u = u^\top Lu/\|u\|_2^2$, so Proposition 1 gives the first identity in (13); the second follows directly. Finally, all moments satisfy

$$\int \lambda^m d\nu_u(\lambda) = \frac{(-1)^m}{2^m} P_u^{(m)}(0).$$

Two probability measures on the compact interval $[0,2]$ with the same moments integrate every polynomial identically. By uniform density of polynomials in $C[0,2]$, they integrate every continuous function identically and hence are the same measure. □

The theorem makes precise what is lost by a single cut score. Ncut is $m_1 = \int \lambda d\nu_u$, while P_u contains $m_1, m_2, \ldots$ and therefore the complete label spectrum.

3.3 Instantaneous leakage and spectral variance

The logarithmic slope of persistence isolates the frequencies still present at time t.

Definition 1 (Instantaneous leakage rate)**.** Whenever $P_u(t) > 0$, define

$$\kappa_u(t) = -\frac{1}{2}\frac{d}{dt}\log P_u(t). \tag{14}$$

Also define the tilted probability weights

$$\omega_j(t) = \frac{\omega_j e^{-2t\lambda_j}}{P_u(t)}. \tag{15}$$

Theorem 2 (Leakage-rate identity)**.** *For every $t \geq 0$,*

$$\kappa_u(t) = \sum_j \omega_j(t)\lambda_j, \qquad \kappa_u'(t) = -2\,\mathrm{Var}_{\omega(t)}(\lambda) \leq 0. \tag{16}$$

Consequently,

$$\kappa_u(0) = \mathrm{Ncut}(C,\bar{C}), \qquad \kappa_u'(0) = -2\mathscr{V}_u, \tag{17}$$

where

$$\mathscr{V}_u = \int \lambda^2 d\nu_u - \left(\int \lambda d\nu_u\right)^2 = \frac{1}{4}P_u''(0) - \frac{1}{4}P_u'(0)^2. \tag{18}$$

Proof. From (11),

$$-\frac{1}{2}\frac{P_u'(t)}{P_u(t)} = \frac{\sum_j \omega_j\lambda_j e^{-2t\lambda_j}}{\sum_j \omega_j e^{-2t\lambda_j}} = \sum_j \omega_j(t)\lambda_j.$$

Differentiate the quotient, or equivalently use $\omega_j'(t) = -2\omega_j(t)(\lambda_j - \kappa_u(t))$, to obtain

$$\kappa_u'(t) = -2\sum_j \omega_j(t)(\lambda_j - \kappa_u(t))^2.$$

The identities at $t=0$ then follow from Theorem 1 and Proposition 1. □

Equation (16) shows that a broad label spectrum forces the leakage rate to decrease rapidly with diffusion time. A label spectrum concentrated near one eigenvalue behaves approximately as a single exponential, for which the variance term vanishes.

3.4 Finite-horizon persistence and pseudoinverse energy

For numerical comparison we use the normalized area H_T from (7). Spectrally,

$$H_T(u) = \sum_{j=1}^{n} \omega_j \psi_T(\lambda_j), \tag{19}$$

$$\psi_T(\lambda) = \begin{cases} 1, & \lambda = 0, \\ \dfrac{1-e^{-2T\lambda}}{2T\lambda}, & \lambda > 0. \end{cases} \tag{20}$$

Because ψ_T is decreasing, H_T favors label energy at low graph frequencies.

Proposition 2 (Infinite-horizon identity). *If $u \perp \ker L$, then*

$$\int_0^\infty P_u(t)\,dt = \frac{1}{2\|u\|_2^2} u^\top L^\dagger u. \tag{21}$$

Proof. Under $u \perp \ker L$, all nonzero weights in (10) correspond to $\lambda_j > 0$. Hence

$$\int_0^\infty P_u(t)dt = \sum_j \omega_j \int_0^\infty e^{-2t\lambda_j}dt = \frac{1}{2}\sum_j \frac{\omega_j}{\lambda_j},$$

which is exactly (21) by the spectral representation of $L^\dagger$. □

If the graph is disconnected, a partition signal may have nonzero projection onto $\ker L$, in which case $P_u(t)$ has a nondecaying component and the infinite-horizon integral diverges. The finite-horizon H_T remains well defined and correctly records perfect or near-perfect graph separation through values close to one.

3.5 Bounds and operator perturbations

For a connected graph with $u \perp \ker L$, let λ_2 denote the spectral gap and λ_n the largest eigenvalue. Since all active spectral mass lies in $[\lambda_2, \lambda_n]$,

$$e^{-2\lambda_n t} \le P_u(t) \le e^{-2\lambda_2 t}, \tag{22}$$

and therefore

$$\psi_T(\lambda_n) \le H_T(u) \le \psi_T(\lambda_2). \tag{23}$$

The next result controls graph perturbations without assuming eigenvector-by-eigenvector matching.

Theorem 3 (Stability for a fixed partition signal). *Let L and $\widetilde{L}$ be symmetric positive semidefinite matrices and let $u \neq 0$ be fixed. Then*

$$|P_{u,L}(t) - P_{u,\widetilde{L}}(t)| \le 2t\|L-\widetilde{L}\|_2, \tag{24}$$

and

$$|H_{T,L}(u) - H_{T,\widetilde{L}}(u)| \le T\|L-\widetilde{L}\|_2. \tag{25}$$

Proof. The Duhamel identity gives

$$e^{-tL} - e^{-t\widetilde{L}} = -\int_0^t e^{-(t-s)L}(L-\widetilde{L})e^{-s\widetilde{L}}ds.$$

Both semigroups are contractions, hence $\|e^{-tL} - e^{-t\widetilde{L}}\|_2 \le t\|L-\widetilde{L}\|_2$. Using $|\|a\|_2^2 - \|b\|_2^2| \le (\|a\|_2 + \|b\|_2)\|a-b\|_2$ and $\|e^{-tL}u\|_2, \|e^{-t\widetilde{L}}u\|_2 \le \|u\|_2$ proves (24). Integration over $[0,T]$ proves (25). □

The theorem isolates operator perturbation from signal perturbation. In data-dependent graphs the degree-balanced vector also changes with W; the numerical repetitions below therefore measure total empirical variability rather than invoking (25) as a direct sampling bound.

4. MATERIALS AND METHODS

4.1 Dataset and preprocessing

The numerical study used a handwritten-digit benchmark containing $n = 1{,}797$ observations. Each observation is an 8×8 grayscale digit represented by 64 numerical variables and one of ten class labels. Class counts range from 174 to 183, so no class rebalancing was applied. The exact matrix used in the analysis is included with the reproducibility package.

Ten stratified random partitions were generated with 70% training and 30% test observations. For each split, standardization parameters were estimated from the training data only. A 20-dimensional principal-component representation was then fitted on the standardized training matrix and applied unchanged to the corresponding test matrix. No test observation entered a graph construction, bandwidth estimate, eigenproblem, or parameter choice.

4.2 Pairwise graph construction

For each training split and each unordered class pair (a,b), the corresponding training observations formed a separate graph. There are $\binom{10}{2} = 45$ pairs and therefore 450 independently reconstructed split–pair graphs. For a pair containing m training observations with transformed coordinates $x_1, \ldots, x_m \in \mathbb{R}^{20}$, let r_i be the Euclidean distance from x_i to its tenth nearest neighbor. A locally scaled directed affinity was assigned by

$$\widetilde{w}_{ij} = \begin{cases} \exp\left[-\dfrac{\|x_i - x_j\|_2^2}{r_i r_j}\right], & j \in \mathscr{N}_{10}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{26}$$

followed by union symmetrization $w_{ij} = \max\{\widetilde{w}_{ij}, \widetilde{w}_{ji}\}$. This construction follows the self-tuned-neighborhood principle analyzed in recent graph-Laplacian convergence work [3] while avoiding a single global bandwidth.

For each graph, the normalized Laplacian (1) and degree-balanced class signal (8) were computed. The eigendecomposition supplied ν_u, $P_u(t)$, H_1, normalized cut, and the spectral variance $\mathscr{V}_u$. The finite horizon was fixed at $T = 1$ before the repeated-split evaluation.

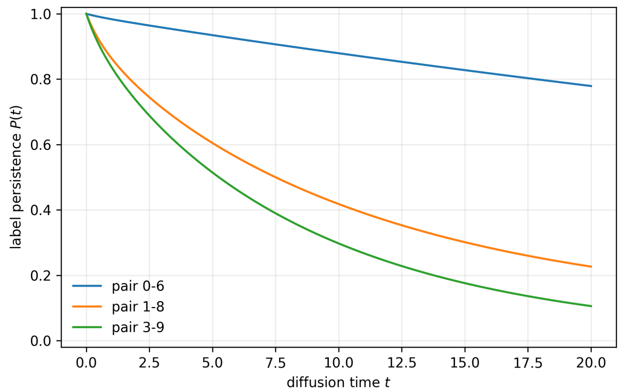

Figure 1. Representative heat-semigroup persistence curves $P_u(t)$. A slowly decaying curve indicates that the degree-balanced class signal is concentrated at low graph frequencies.

4.3 External difficulty criterion

Graph separability was compared with a quantity not used in its construction. A distance-weighted 5-nearest-neighbor classifier was fitted to each pair's 20-dimensional training representation and evaluated on the two corresponding held-out classes. If $\widehat{y}_i$ denotes the prediction for held-out pair observation i, pairwise error was

$$E_{ab} = \frac{1}{n_{ab}^{\text{test}}} \sum_{i=1}^{n_{ab}^{\text{test}}} \mathbf{1}\{\widehat{y}_i \neq y_i\}. \tag{27}$$

The classifier is therefore an external probe of pair difficulty; it is not an ingredient of P_u, H_T, Ncut, or $\mathscr{V}_u$.

4.4 Statistical analysis

For each class pair, the ten split-specific values were averaged. Monotone association between graph quantities and held-out error was measured by Spearman's rank coefficient over the 45 pair means. A common-resample bootstrap with 5,000 replicates resampled the 45 pairs with replacement and recomputed each coefficient; percentile 2.5% and 97.5% quantiles provide the reported intervals. Split-level associations over all 450 observations were also retained as a check against averaging artifacts.

5. RESULTS

5.1 Multiscale persistence separates easy and difficult pairs

Figure 1 shows representative persistence curves from the first split. Easy digit pairs retain nearly all degree-balanced label energy over the displayed diffusion range, which corresponds to little graph coupling between classes. Difficult pairs decay more rapidly and exhibit visible curvature, reflecting nonzero spectral variance through Theorem 2. The curve is therefore richer than its initial tangent: normalized cut fixes $P_u'(0)$, whereas the subsequent trajectory depends on the distribution of spectral mass.

The pairwise matrix of mean H_1 values in Figure 2 exposes a coherent block of comparatively low persistence involving digits 8 and 9. The smallest mean values occur for pairs 3–9, 1–8, 8–9, and 3–8. These are not imposed by the classifier: the matrix is computed entirely from training graphs and class

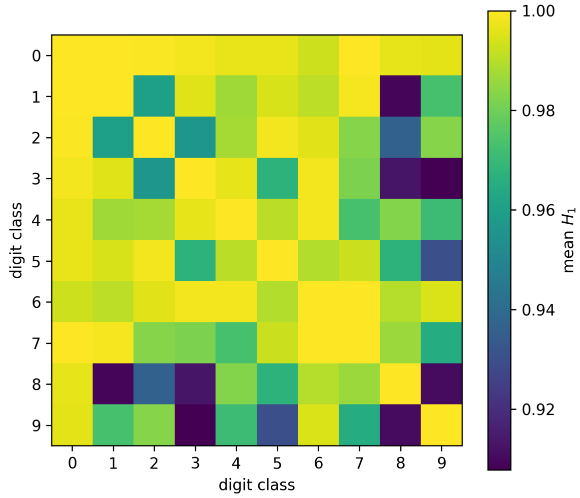

Figure 2. Mean finite-horizon persistence H_1 over ten training splits for all 45 digit pairs. Higher values indicate stronger low-frequency retention of the balanced label signal.

Table 1. Ten class pairs with largest mean held-out error.

Pair	H_1	Ncut	$\mathscr{V}_u$	Error
8–9	0.911	0.121	0.0555	0.0236
5–9	0.931	0.096	0.0510	0.0211
1–8	0.909	0.128	0.0673	0.0206
3–8	0.913	0.116	0.0516	0.0178
7–9	0.965	0.047	0.0241	0.0157
3–9	0.908	0.119	0.0463	0.0156
2–8	0.936	0.085	0.0392	0.0143
4–7	0.973	0.038	0.0233	0.0102
3–5	0.967	0.042	0.0202	0.0100
2–3	0.956	0.057	0.0275	0.0083

memberships.

Table 1 lists the ten pairs with the largest held-out error. The values illustrate why a single ranking statistic should not be overinterpreted. Pair 1–8 has lower H_1 and larger normalized cut than 8–9, yet its mean external error is slightly smaller. Pair 7–9 has comparatively high persistence but remains among the ten hardest pairs. The second spectral moment carries complementary information on the spread of leakage frequencies.

5.2 Association with held-out pair difficulty

Across the 45 pair means, H_1 has Spearman correlation -0.9236 with held-out 5-NN error (Table 2); the negative sign is expected because larger persistence denotes greater class separation. Normalized cut has correlation 0.9265, and $\mathscr{V}_u$ has correlation 0.9301. All three associations are strong, and their bootstrap intervals overlap substantially. The evidence therefore does not support a claim that H_1 is a superior predictor of pair difficulty. It supports the narrower statement that the multiscale heat functional preserves the separability information captured by normalized cut while exposing additional spectral structure.

At the split–pair level ($n = 450$), the corresponding correlations are -0.6082 for H_1 and 0.6062 for normalized cut, with $p < 2 \times 10^{-46}$ in both cases. The reduction from pair-mean

Table 2. Rank association with held-out pair error over 45 class-pair means.

Quantity	ρ	95% bootstrap interval	p
H_1	-0.924	$[-0.957, -0.850]$	1.63×10^{-19}
Ncut	0.927	$[0.856, 0.960]$	7.25×10^{-20}
$\mathscr{V}_u$	0.930	$[0.863, 0.959]$	2.57×10^{-20}

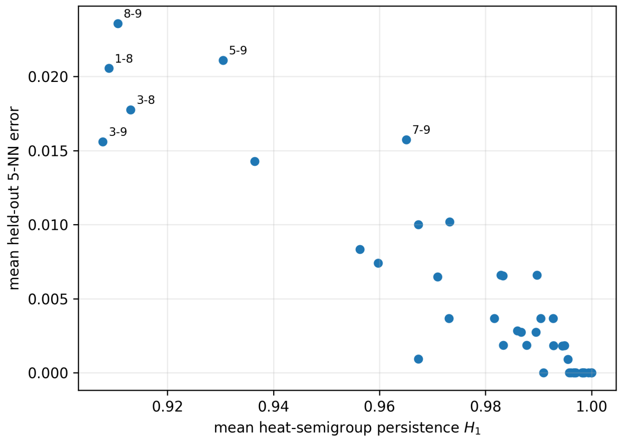

Figure 3. Mean H_1 against mean held-out pairwise 5-NN error. Labels identify six pairs with the largest errors.

correlations is expected because each individual held-out pair contains relatively few observations and its error is discrete. Averaging across repeated splits primarily stabilizes pair-specific difficulty.

Figure 3 displays the pair-mean relation. The lowest-persistence region is populated by the high-error pairs 1–8, 3–8, 3–9, and 8–9. The relation is monotone rather than linear, which is why rank correlation is used as the primary numerical summary.

5.3 The spectral measure explains curvature beyond normalized cut

Figure 4 plots the cumulative label spectral measure for three representative pairs. For a highly separable pair, most mass lies extremely close to zero, so $P_u(t)$ remains near one. For harder pairs, mass spreads to larger eigenvalues. Equation (16) then gives two distinct effects: a larger first moment raises the initial leakage $\kappa_u(0)$, while larger variance causes the leakage rate itself to decrease more rapidly with t as high-frequency components are preferentially damped.

This distinction can be expressed through the short-time expansion

$$P_u(t) = 1 - 2m_1 t + 2m_2 t^2 - \frac{4}{3} m_3 t^3 + O(t^4), \qquad (28)$$

where $m_q = \int \lambda^q d\nu_u(\lambda)$. Normalized cut supplies only m_1. The empirical correlation of $\mathscr{V}_u = m_2 - m_1^2$ with held-out error (0.9301) shows that the second central moment is not numerically degenerate on these graphs. It is nevertheless strongly related to the same class geometry, so it should be interpreted as a refinement of the cut description rather than an independent accuracy mechanism.

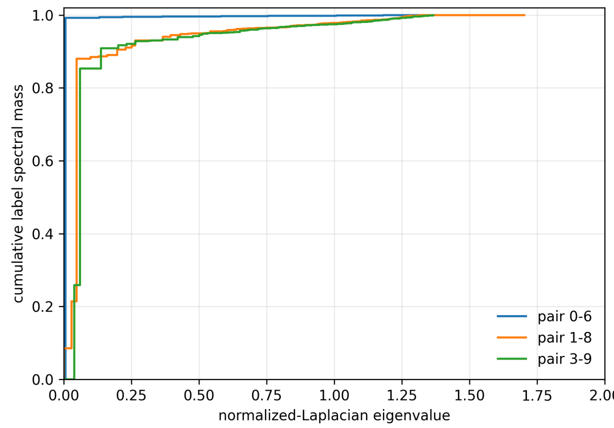

Figure 4. Cumulative label spectral measures ν_u for representative digit pairs. The normalized-cut value is the first moment of each distribution; $P_u(t)$ retains the full measure.

6. DISCUSSION

The principal mathematical result is a change in representation. A binary graph partition is usually mapped to one cut energy; here it is mapped to the probability measure ν_u or, equivalently, to its heat-persistence transform P_u. Proposition 1 and Theorem 1 show that this representation contains normalized cut exactly, not approximately: $\mathrm{Ncut} = -P_u'(0)/2$. Theorem 2 then identifies a second level of structure. The logarithmic leakage rate $\kappa_u(t)$ is the mean eigenvalue under a continuously tilted label spectrum, and its derivative is minus twice the corresponding variance. Diffusion time therefore orders spectral information without introducing an arbitrary sequence of unrelated statistics.

The finite-horizon area H_T is useful when a scalar is required. Its kernel $\psi_T(\lambda)$ in (19) is a low-pass spectral weighting, so T controls resolution: small T keeps H_T close to $1 - T\,\mathrm{Ncut} + O(T^2)$, whereas larger T increasingly emphasizes near-null spectral mass. The full curve avoids choosing one horizon and is injective at the level of the compactly supported label measure. For applications where two partitions have similar normalized cut but different higher-order spectra, comparing $P_u(t)$ or $\kappa_u(t)$ can therefore reveal a distinction that the Rayleigh quotient cannot express.

The numerical study was designed to test this interpretation without turning the paper into a classification benchmark. Graphs were reconstructed on training observations alone, and the external 5-NN model served only to estimate pair difficulty on unseen observations. The strong association between graph separability and held-out error is reassuring, but the near-equality of the H_1 and normalized-cut rank correlations deserves emphasis. On this dataset, a first spectral moment already orders class pairs remarkably well. The added value of P_u is consequently analytical and diagnostic: it supplies curvature, higher moments, a leakage trajectory, and a pseudoinverse connection while retaining the familiar cut quantity as a boundary derivative.

Several limitations follow directly from the formulation. First, P_u is conditional on the graph construction. Self-tuned k-nearest-neighbor graphs reduce dependence on a global bandwidth, but different neighborhood rules or robust normalizations can alter finite-sample spectra; recent convergence and robustness results make this an active issue rather than a settled preprocessing detail [13, 14]. Second, Theorem 3 fixes u while perturbing L. In data graphs, degrees change with the affinity matrix and therefore the balanced signal also changes. Extending (25) to the joint map $(W,C) \mapsto (L,u)$ is a natural next step. Third, disconnected graphs produce zero-eigenvalue label mass. This is not a numerical defect–it represents exact separation–but it precludes the infinite-horizon pseudoinverse identity unless the signal is orthogonal to the complete nullspace. Finally, the empirical experiment is intentionally small enough to permit exact eigendecomposition for every split–pair graph. Large graphs would require polynomial, rational-Krylov, stochastic-trace, or low-rank approximations to semigroup quantities; numerical spectral-function methods such as those studied by [19] provide relevant tools.

The framework also suggests extensions that remain mathematical rather than model-specific. For multiclass partitions, one may replace u by a degree-orthogonal indicator matrix U and study $\operatorname{tr}(U^\top e^{-2tL}U)/\operatorname{tr}(U^\top U)$. For signed or directed operators, contractivity and spectral positivity require separate treatment. Time-inhomogeneous diffusion, already studied in geometric data analysis [17], raises a different question because the propagator is no longer a single matrix exponential and the label spectral measure may not be stationary. These directions preserve the central premise: partition quality can be studied as an operator evolution, not only as a static cut.

7. CONCLUSION

For a degree-balanced binary partition signal u, heat-semigroup persistence

$$P_u(t) = \|e^{-tL}u\|_2^2/\|u\|_2^2$$

provides a multiscale completion of normalized cut. The complete curve is the Laplace transform of the label spectral measure; normalized cut is its first spectral moment and equals $-P_u'(0)/2$; the initial curvature determines the second moment; and the logarithmic leakage rate obeys $\kappa_u'(t) = -2\,\mathrm{Var}_{\nu_{u,t}}(\lambda)$. Under nullspace orthogonality, the integrated curve is a Laplacian-pseudoinverse energy, while finite-horizon persistence admits a direct operator perturbation bound. These identities place cut energy, spectral variance, diffusion decay, and effective resistance-type structure in one calculation.

The repeated handwritten-digit experiment supports the resulting separability interpretation without suggesting an unwarranted predictive advantage over normalized cut. Across 45 class pairs, H_1 and normalized cut have almost equal magnitude of rank association with held-out pairwise error, while the second spectral central moment is similarly informative. The distinction is therefore structural: a cut is a first-moment statistic, whereas the heat curve determines the entire label spectrum. That additional information is available without changing the graph or fitting a new predictive model.

REFERENCES

[1] C. L. Wormell and S. Reich, "Spectral convergence of diffusion maps: Improved error bounds and an alternative normalization," *SIAM Journal on Numerical Analysis*, vol. 59, no. 3, pp. 1687–1734, 2021.

[2] J. Calder and N. García Trillos, "Improved spectral convergence rates for graph laplacians on ε-graphs and k-NN graphs," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 123–175, 2022.

[3] X. Cheng and H.-T. Wu, "Convergence of graph laplacian with kNN self-tuned kernels," *Information and Inference: A Journal of the IMA*, vol. 11, no. 3, pp. 889–957, 2022.

[4] X. Cheng and N. Wu, "Eigen-convergence of gaussian kernelized graph laplacian by manifold heat interpolation," *Applied and Computational Harmonic Analysis*, vol. 61, pp. 132–190, 2022.

[5] J. Calder, N. García Trillos, and M. Lewicka, "Lipschitz regularity of graph laplacians on random data clouds," *SIAM Journal on Mathematical Analysis*, vol. 54, no. 1, pp. 1169–1222, 2022.

[6] T. Roith and L. Bungert, "Continuum limit of lipschitz learning on graphs," *Foundations of Computational Mathematics*, vol. 23, pp. 393–431, 2023.

[7] M. M. Dunlop, D. Slepčev, A. M. Stuart, and M. Thorpe, "Large data and zero noise limits of graph-based semi-supervised learning algorithms," *Applied and Computational Harmonic Analysis*, vol. 49, no. 2, pp. 655–697, 2020.

[8] M. Boedihardjo, S. Deng, and T. Strohmer, "A performance guarantee for spectral clustering," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 369–387, 2021.

[9] H.-L. de Kergorlay and D. J. Higham, "Consistency of anchor-based spectral clustering," *Information and Inference: A Journal of the IMA*, vol. 11, no. 3, pp. 801–822, 2022.

[10] D. Pasadakis, C. L. Alappat, O. Schenk, and G. Wellein, "Multiway p-spectral graph cuts on grassmann manifolds," *Machine Learning*, vol. 111, pp. 791–829, 2022.

[11] N. García Trillos, R. Murray, and M. Thorpe, "From graph cuts to isoperimetric inequalities: Convergence rates of cheeger cuts on data clouds," *Archive for Rational Mechanics and Analysis*, vol. 244, pp. 541–598, 2022.

[12] M. Flores, J. Calder, and G. Lerman, "Analysis and algorithms for ℓ_p-based semi-supervised learning on graphs," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 77–122, 2022.

[13] B. Landa, R. R. Coifman, and Y. Kluger, "Doubly stochastic normalization of the gaussian kernel is robust to heteroskedastic noise," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 388–413, 2021.

[14] B. Landa and X. Cheng, "Robust inference of manifold density and geometry by doubly stochastic scaling," *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 3, pp. 589–614, 2023.

[15] J. Ah-Pine, "Learning doubly stochastic and nearly idempotent affinity matrix for graph-based clustering," *European Journal of Operational Research*, vol. 299, no. 3, pp. 1069–1078, 2022.

[16] H. Antil, T. Berry, and J. Harlim, "Fractional diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 54, pp. 145–175, 2021.

[17] G. Huguet, A. Tong, B. Rieck, J. Huang, M. Kuchroo, M. Hirn, G. Wolf, and S. Krishnaswamy, "Time-inhomogeneous diffusion geometry and topology," *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 2, pp. 346–372, 2023.

[18] A. Aldroubi, I. A. Krishtal, and S. Tang, "Phaseless reconstruction from space–time samples," *Applied and Computational Harmonic Analysis*, vol. 48, no. 1, pp. 395–414, 2020.

[19] M. Benzi, M. Rinelli, and I. Simunec, "Computation of the von neumann entropy of large matrices via trace estimators and rational krylov methods," *Numerische Mathematik*, vol. 155, pp. 377–414, 2023.

[20] K. Miller and J. Calder, "Poisson reweighted laplacian uncertainty sampling for graph-based active learning," *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 4, pp. 1160–1190, 2023.