Heat kernel coupling for multiple graph analysis

Michael M. Bronstein

Klaus Glashoff

Institute of Computational Science, Faculty of Informatics,

Università della Svizzera Italiana, Lugano, Switzerland

December 12, 2013

Abstract

In this paper, we introduce heat kernel coupling (HKC) as a method of constructing multimodal spectral geometry on weighted graphs of different size without vertex-wise bijective correspondence. We show that Laplacian averaging can be derived as a limit case of HKC, and demonstrate its applications on several problems from the manifold learning and pattern recognition domain.

1 Introduction

Many problems in computer vision and pattern recognition boil down to constructing a Laplacian operator describing some data manifold and finding its eigenvectors. Notable examples include spectral clustering [26], eigenmaps [4], diffusion maps and distances [12, 25], spectral graph partitioning [14], spectral hashing [36], and image segmentation [31].

Recently, there has been an increased interest in extending spectral geometric constructions to the multimodal setting, involving two or more data spaces. Many data analysis applications involve observations and measurements of data using different modalities, such as multimedia documents [3, 37, 29, 24, 18, 23], audio and video [19, 1, 30], images with different lighting conditions [2], or medical imaging modalities [6].

1

Multimodal (or 'multi-view') clustering was studied in the computer vision and pattern recognition community [13, 22, 33, 8, 21, 15]. Sindhwani et al. [32] used a convex combination of Laplacians in the 'co-regularization' framework. *Manifold alignment* considered multiple manifolds as a single space with 'connections' between points and tries to find an aligned set of eigenvectors [17, 35, 34]. Eynard et al. [16] proposed finding a common eigenbasis of multiple Laplacians by means of joint approximate diagonalization (JADE). Kovnatsky et al. [20] improved this method using subspace parametrization. Bronstein et al. [7] studied the problem of finding closest commuting operators (CCO) and showed its equivalence to joint diagonalization.

One of the main limitations of JADE and CCO problems is the assumption of given bijective correspondence (or more generally, functional correspondence [27]) between the underlying manifolds or graphs. In this paper, we consider the setting where such correspondence is unknown or may not exist, and instead, one is given a set of corresponding functions. We show a problem similar to CCO, wherein we try to minimally modify the Laplacians such that the corresponding heat kernels behave consistently. In the limit case with given bijective correspondence, this heat kernel coupling problem is equivalent to Laplacian averaging.

2 Background

Notation and definitions. Let $\mathbf{A}, \mathbf{B}$ be two $n \times n$ real symmetric matrices. We denote by

$$\|\mathbf{A}\|_{\mathrm{F}} = \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2} = \left(\mathrm{tr}(\mathbf{A}^\top \mathbf{A}) \right)^{1/2}$$

the *Frobenius* norm of $\mathbf{A}$. We say that $\mathbf{A}$ and $\mathbf{B}$ *commute* if $\mathbf{AB} = \mathbf{BA}$, and call $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$ their *commutator*. If there exists a unitary matrix $\hat{\mathbf{U}}$ such that $\hat{\mathbf{U}}^\top \mathbf{A} \hat{\mathbf{U}} = \mathbf{\Lambda}_A$ and $\hat{\mathbf{U}}^\top \mathbf{B} \hat{\mathbf{U}} = \mathbf{\Lambda}_B$ are diagonal, we say that $\mathbf{A}, \mathbf{B}$ are *jointly diagonalizable* and call such $\hat{\mathbf{U}}$ the *joint eigenbasis* of $\mathbf{A}$ and $\mathbf{B}$. Two matrices are jointly diagonalizable iff they commute.

We denote by $\mathrm{diag}(\mathbf{A})$ a column vector containing the diagonal elements of matrix $\mathbf{A}$, and by $\mathrm{diag}(a_1, \ldots, a_n)$ a diagonal matrix containing on the diagonal the elements $a_1, \ldots, a_n$. Furthermore, we use $\mathrm{Diag}(\mathbf{A}) = \mathrm{diag}(\mathrm{diag}(\mathbf{A}))$

to denote a diagonal matrix obtained by setting to zero the off-diagonal elements of $\mathbf{A}$.

Laplacians. Let us be given an undirected weighted graph $G = (V, E, \mathbf{W})$ without loops (i.e., a *simple graph*) with vertice set $V = \{x_1, \ldots, x_n\}$ and edges $E \subseteq, \{1, \ldots, n\}^2$ such that $(i, i) \notin E$ for $i = 1, \ldots, n$. Let $E = \{(i_k, j_k) | k = 1, \ldots, |E|\}$. There are given non-negative weights $w_{ij} \geq 0$, satisfying $w_{ij} = 0$ if x_i, x_j are not connected (i.e., $(i, j) \notin E$). The $n \times n$ matrix $\mathbf{W} = (w_{ij})$ is called the *adjacency matrix* and

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad \mathbf{D} = \text{diag} \left(\sum_{j \neq 1} w_{1j}, \ldots, \sum_{j \neq n} w_{nj} \right) \tag{1}$$

Hereinafter, we denote by $\mathcal{L}(V, E)$ the set of all *valid Laplacian* matrices of a simple graph (V, E), which is defined as follows: $L = (l_{ij}) \in \mathcal{L}(V, E)$ *iff*
(i) $l_{ij} \leq 0$ and $l_{ij} = l_{ji}$ for $i \neq j$; (ii) sparse structure: $l_{ij} = 0$ if $(i, j) \notin E$; and (iii) zero row sum: $\sum_{j=1}^{n} l_{ij} = 0$. Defining the Laplacian according to (1) through the edge weight matrix $\mathbf{W}$, we automatically get properties (i) - (iii) satisfied. The other way round: Any valid Laplacian of a simple graph - in the sense of (i)-(iii) - gives rise to a weight matrix $\mathbf{W}$ of a simple weighted graph by defining $w_{ij} = -l_{ij}, i \neq j$.

For numerical purposes, we will make use of a proper *parametrization* of the set of valid Laplacians. Let $m = |E|$ denote the number of edges of G. For $\mathbf{u} = (u_1, \ldots, u_m) \in \mathbb{R}^m_{\geq 0}$ we define the weight matrix $\mathbf{W}$ by

$$w_{ij}(\mathbf{u}) = \begin{cases} u_l & i = i_l, j = j_l \text{ or } i = j_l, j = i_l \\ 0 & \text{else.} \end{cases} \tag{2}$$

Defining $\mathbf{L}$ as in (1), the requirements (i)-(iii)of a valid Laplacian are satisfied. In undirected weighted graph, the matrices $\mathbf{W}$ and $\mathbf{L}$ are symmetric. Furthermore, $\mathbf{L}$ is positive semi-definite. Consequently, $\mathbf{L}$ admits the unitary eigendecomposition $\mathbf{L} = \mathbf{\Phi \Lambda \Phi}^\top$ with orthonormal eigenvectors $\mathbf{\Phi} = (\phi_1, \ldots, \phi_n)$ and real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

Heat diffusion on graphs. Let $f : V \to \mathbb{R}$ denote a function defined on the vertex set of the graph. We can identify f with an $|V|$-dimensional vector $\mathbf{f} = (f(x_1), \ldots, f(x_{|V|}))$, and denote by $\mathcal{F}(V)$ the space of such functions.

Similarly to the standard heat diffusion equation, one can define a diffusion process on G, governed by the following equation:

$$\mathbf{L}\mathbf{f}(t) + \frac{\partial}{\partial t}\mathbf{f}(t) = 0, \quad \mathbf{f}(0) = \mathbf{f}_0, \tag{3}$$

where the solution $\mathbf{f}(t) : V \times [0, \infty) \to \mathbb{R}_+$ is the amount of heat at time t at the vertices V. The solution of the heat equation is given by $\mathbf{f}(t) = \mathbf{H}^t \mathbf{f}_0$, where

$$\mathbf{H}^t = e^{-t\mathbf{L}} = \mathbf{\Phi} e^{-t\mathbf{\Lambda}} \mathbf{\Phi}^\top$$

is the *heat operator* (or the *heat kernel*).

The heat kernel gives rise to the *diffusion distance* [12, 11]

$$d_t(x_p, x_q) = \left(\sum_{i=1}^{n} ((\mathbf{H}^t)_{pi} - (\mathbf{H}^t)_{qi})^2 \right)^{1/2} = \left(\sum_{i=1}^{n} e^{-2t\lambda_i}(\phi_{pi} - \phi_{qi})^2 \right)^{1/2}, \tag{4}$$

measuring the 'reachability' of vertex x_q from vertex x_p in time t.

3 Multimodal spectral geometry

Consider two graphs $G_k = (V, E, \mathbf{W}_k)$, $k = 1, 2$ with the same vertices V and edges E with different weights $\mathbf{W}_k$. We denote their respective Laplacians by $\mathbf{L}_k \in \mathcal{L}(V, E)$. Such graphs are referred to as *multi-level graphs* [15], and are used to represent multiple modalities or 'views' of the same data. [1] The main topic of this paper is how to re-define the above spectral geometric constructions (heat kernels, diffusion distances, etc.) in a way that they account for information from both graphs.

3.1 Laplacian averaging

The simplest approach is to define an average weight $\bar{\mathbf{W}} = \frac{1}{2}(\mathbf{W}_1 + \mathbf{W}_2)$ [22]. Equivalently, this problem can be posed as finding a new Laplacian $\bar{\mathbf{L}} \in \mathcal{L}(V, E)$ that is equidistant from the given $\mathbf{L}_1, \mathbf{L}_2$,

$$\min_{\bar{\mathbf{L}} \in L(V,E)} \sum_{i=1}^{2} \|\mathbf{L}_i - \bar{\mathbf{L}}\|_{\mathrm{F}}^2. \tag{5}$$

The solution of (5) is the *average Laplacian* $\bar{\mathbf{L}} = \frac{1}{2}(\mathbf{L}_1 + \mathbf{L}_2)$.

---

[1]For simplicity, we consider only two modalities, though extension to more modalities is straightforward.

3.2 Joint diagonalization

Instead of averaging the Laplacians, Eynard et al. [16] proposed 'averaging' their eigenspaces by means of a joint diagonalization approach: construct a common (approximate) eigenbasis $\hat{\boldsymbol{\Phi}}$ that (approximately) diagonalizes the Laplacians $\mathbf{L}_1, \mathbf{L}_2$, by the following minimization

$$J(\mathbf{L}_1, \mathbf{L}_2) = \min_{\hat{\boldsymbol{\Phi}} \in \mathbb{R}^{n \times n}} \sum_{k=1}^{2} \mathrm{off}(\hat{\boldsymbol{\Phi}}^{\top} \mathbf{L}_k \hat{\boldsymbol{\Phi}}) \quad \text{s.t.} \quad \hat{\boldsymbol{\Phi}}^{\top} \hat{\boldsymbol{\Phi}} = \mathbf{I}, \tag{6}$$

where $\mathrm{off}(\mathbf{A}) = \sum_{i \neq j} a_{ij}^2$ denotes the squared norm of the off-diagonal elements of a matrix [9]. The joint basis $\hat{\boldsymbol{\Phi}}$ obtained in this way satisfies $\hat{\boldsymbol{\Phi}}^{\top} \mathbf{L}_k \hat{\boldsymbol{\Phi}} \approx \mathrm{diag}(\hat{\lambda}_{k,1}, \dots, \hat{\lambda}_{k,|V|})$. The approximate matrices

$$\hat{\mathbf{L}}_k = \hat{\boldsymbol{\Phi}} \mathrm{Diag}(\hat{\boldsymbol{\Phi}}^{\top} \mathbf{L}_k \hat{\boldsymbol{\Phi}}) \hat{\boldsymbol{\Phi}}^{\top} \approx \mathbf{L}_k,$$

obtained by setting to zero the off-diagonal elements of $\hat{\boldsymbol{\Phi}}^{\top} \mathbf{L}_k \hat{\boldsymbol{\Phi}}$ are jointly diagonalizable by $\hat{\boldsymbol{\Phi}}$. Importantly, in most cases $\hat{\mathbf{L}}_k \notin \mathcal{L}(V, E)$, i.e., Laplacian structure does not survive joint diagonalization.

3.3 Closest commuting operators

In [7], we considered a different problem of finding a pair $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2$ of commuting matrices (referred to as *closest commuting operators* or CCOs) that are closest to the given $\mathbf{L}_1, \mathbf{L}_2$,

$$C(\mathbf{L}_1, \mathbf{L}_2) = \min_{\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_1 \in \mathbb{R}^{n \times n}} \sum_{i=1}^{2} \|\tilde{\mathbf{L}}_i - \mathbf{L}_i\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \tilde{\mathbf{L}}_1 \tilde{\mathbf{L}}_2 = \tilde{\mathbf{L}}_2 \tilde{\mathbf{L}}_1; \tag{7}$$

and showed that this problem is equivalent to JADE (6) in the following sense:

Theorem 3.1. *Let* $\mathbf{A}, \mathbf{B}$ *be symmetric matrices. Then:*
1. $C(\mathbf{A}, \mathbf{B}) = J(\mathbf{A}, \mathbf{B})$.
2. Let $\hat{\mathbf{U}}$ *be the approximate joint eigenbasis of* $\mathbf{A}, \mathbf{B}$ *obtained by solving the JADE problem (6). Then,* $\tilde{\mathbf{A}} = \hat{\mathbf{U}} \mathrm{Diag}(\hat{\mathbf{U}}^{\top} \mathbf{A} \hat{\mathbf{U}}) \hat{\mathbf{U}}^{\top}$ *and* $\tilde{\mathbf{B}} = \hat{\mathbf{U}} \mathrm{Diag}(\hat{\mathbf{U}}^{\top} \mathbf{B} \hat{\mathbf{U}}) \hat{\mathbf{U}}^{\top}$ *are the closest commuting matrices to* $\mathbf{A}, \mathbf{B}$ *solving the CCO problem (7).*
3. Let $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}$ *be the closest commuting matrices solving the CCO problem (7). By virtue of their commutativity,* $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}$ *are jointly diagonalizable, and their joint eigenbasis* $\hat{\mathbf{U}}$ *solves the JADE problem (6).*

A big advantage of this approach compared to JADE is the possibility to demand that the closest commuting matrices define valid Laplacians, i.e., restrict the search space to $\mathcal{L}(V, E)$:

$$C_{\mathrm{L}}(\mathbf{L}_1, \mathbf{L}_2) \quad = \quad \min_{\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2 \in \mathcal{L}(V, E)} \sum_{i=1}^{2} \|\tilde{\mathbf{L}}_i - \mathbf{L}_i\|_{\mathrm{F}}^2 \ \text{ s.t. } \tilde{\mathbf{L}}_1 \tilde{\mathbf{L}}_2 = \tilde{\mathbf{L}}_2 \tilde{\mathbf{L}}_1. \qquad (8)$$

4 Heat kernel coupling

The methods described in Section 3 rely on the assumption of graphs with equal vertex set, which may be too restrictive in many cases. More generally, we are given two different graphs $G_k = (V_k, E_k, \mathbf{W}_k)$, $k = 1, 2$, where $|V_1| \neq |V_2|$. The correspondence between the vertices is not bijective anymore, but one can consider *functional correspondence* $T : \mathcal{F}(V_1) \to \mathcal{F}(V_2)$, represented by the $|V_2| \times |V_1|$ matrix $\mathbf{T}$ [27].

Let us consider the heat equation (3) on the graphs G_1, G_2. We say that the corresponding heat kernels are *strongly coupled* if the solution of the heat equation on G_1 with some initial condition $\mathbf{f} \neq$ const and the solution of the heat equation on G_2 with the corresponding initial condition $\mathbf{Tf}$ coincide under the correspondence:

$$\mathbf{T}\mathbf{H}_1^t \mathbf{f} = \mathbf{H}_2^t \mathbf{T}\mathbf{f}, \qquad (9)$$

for $t \geq 0$. The strong coupling condition implies that the structure of the graphs is similar, in the sense that heat flows on them in the same way. [2]

If the correspondence $\mathbf{T}$ is further assumed to be unknown, we have to replace the strong coupling condition (9) with a *weak coupling* condition

$$\mathbf{f}^\top \mathbf{H}_1^t \mathbf{f} = \mathbf{f}^\top \mathbf{T}^\top \mathbf{H}_2^t \mathbf{T}\mathbf{f}, \qquad (10)$$

requiring that the projection of the solution on the corresponding functions is equal. Note that while condition (9) compares vectors (which requires the knowledge of correspondence $\mathbf{T}$), condition (10) compares scalars, which

---

[2]If the strong coupling condition holds for a set of functions that span the whole $\mathcal{F}(V_1)$, it is equivalent to commutativity of the heat- and the functional correspondence operators, $\mathbf{T}\mathbf{H}_1^t = \mathbf{H}_1^t \mathbf{T}$. In the case of bijective correspondence ($|V_1| = |V_2|$ and w.l.o.g. $\mathbf{T} = \mathbf{I}$), having the strong coupling condition hold for one value of $t > 0$ implies that $\mathbf{L}_1 = \mathbf{L}_2$ and thus the weighted graphs are isometric [28].

does not require the knowledge of the correspondence $\mathbf{T}$ but rather the pair of corresponding functions $\mathbf{f}$ and $\mathbf{Tf}$. Obviously, condition (9) implies (10), but not vice versa.

In this weaker setting, we assume that correspondence $\mathbf{T}$ is unknown, but we have a set of q *corresponding functions* on V_1 and V_2 represented as columns of matrices $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_q)$ and $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_q)$ such that $\mathbf{TF} = \mathbf{G}$. Writing the weak coupling condition (10) for every pair $\mathbf{f}_i, \mathbf{g}_j$, we get the condition on the equality of $q \times q$ matrices

$$\mathbf{F}^\top \mathbf{H}_1^t \mathbf{F} = \mathbf{G}^\top \mathbf{H}_2^t \mathbf{G}, \tag{11}$$

which we consider on a finite set of the values $t \in \{t_1, \ldots, t_p\}$.

Typically, two different graphs will have their heat kernels uncoupled, violating the coupling conditions (see example in Figure 1 (top), where different behavior of the heat equation stems from topological noise). The problem of *heat kernel coupling* (HKC) treated in this paper is how to *minimally* modify the Laplacians of the graphs to make the respective heat kernels (approximately) satisfy the weak coupling condition by enforcing (11); in Figure 1 (bottom) such a modification amounts to disconnecting the rings in both graphs.

Our HKC problem bears resemblance to the CCO problem described in Section 3: we are looking for new graphs $\tilde{G}_k = (V_k, E_k, \tilde{\mathbf{W}}_k)$ with respective adjacency matrices $\tilde{\mathbf{W}}_k$, such that the new Laplacians $\tilde{\mathbf{L}}_k$ are as close as possible to the original $\mathbf{L}_k$, and the corresponding new heat operators $\tilde{\mathbf{H}}_k^t = e^{-t\tilde{\mathbf{L}}_k}$ are as coupled as possible,

$$\min_{\tilde{\mathbf{L}}_k \in \mathcal{L}(V_k, E_k)} \sum_{k=1}^2 \|\tilde{\mathbf{L}}_k - \mathbf{L}_k\|_{\mathrm{F}}^2 + \alpha \sum_{m=1}^p \|\mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1} \mathbf{F} - \mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2} \mathbf{G}\|_{\mathrm{F}}^2. \tag{12}$$

It is important to observe the following limit case: for graphs $G_k = (V, E, \mathbf{W}_k)$ with equal vertex and edge sets discussed in Section 3, we have bijective correspondence between the entries of the heat operators $\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2$, implying $\mathbf{F} = \mathbf{G} = \mathbf{I}$. In the limit $\alpha \to \infty$, we have $\tilde{\mathbf{H}}_1^t = \tilde{\mathbf{H}}_2^t$, from which it follows that $\tilde{\mathbf{L}}_1 = \tilde{\mathbf{L}}_2 = \tilde{\mathbf{L}}$. Thus, the HKC problem (12) boils down to the simple Laplacian averaging (5), and can be considered an extension of this technique to the setting where one cannot straightforwardly average Laplacians since the correspondence between the graphs is not given.
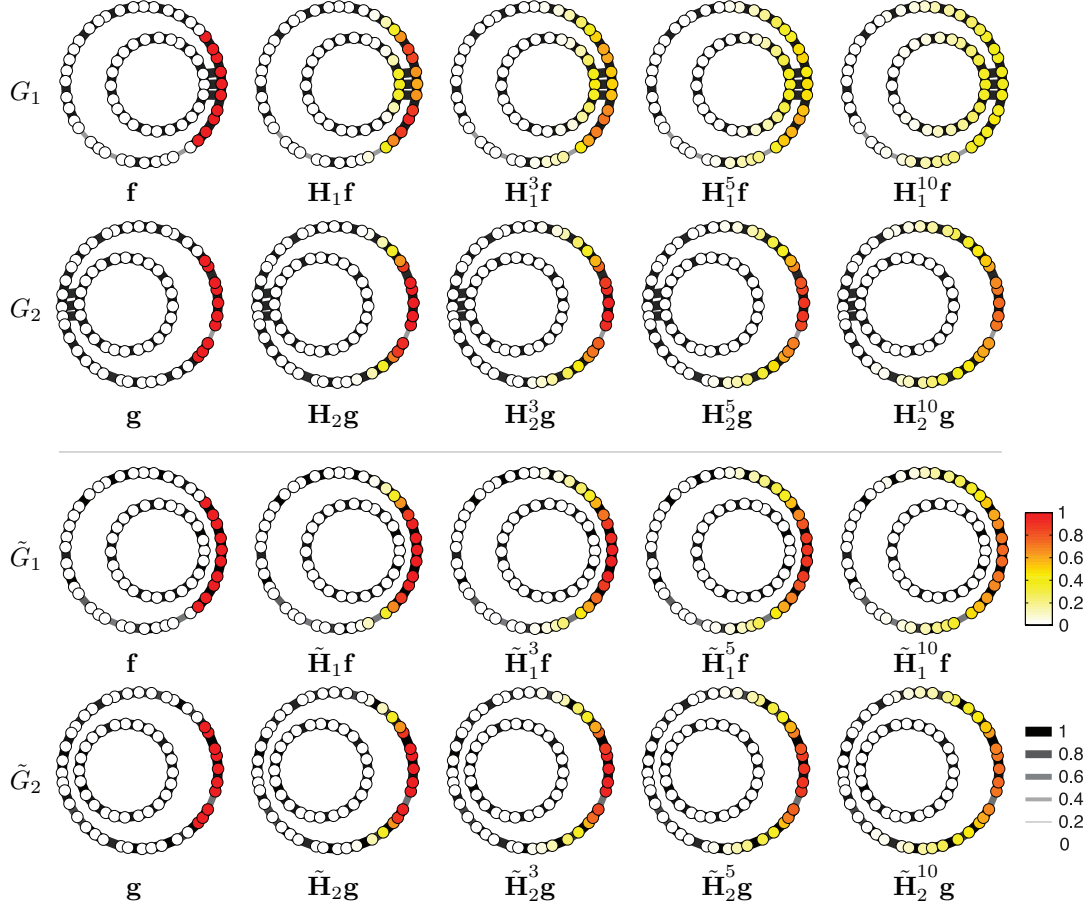
Figure 1: Solution of the heat equation on the original graphs G_1, G_2 (first and second rows) using corresponding initial conditions $\mathbf{f}$ and $\mathbf{g} = \mathbf{Tf}$ (left) does not satisfy the coupling condition, $\mathbf{H}_1^t \mathbf{f} \neq \mathbf{TH}_2^t \mathbf{g}$. On graphs $\tilde{G}_1, \tilde{G}_2$ modified using our HKC procedure (third and fourth rows), the heat kernels are approximately coupled, $\tilde{\mathbf{H}}_1^t \mathbf{f} \approx \mathbf{T}\tilde{\mathbf{H}}_2^t \mathbf{g}$. In this and in following figures, the values of a function defined on the graph vertices are shown in color; the graph edge weights are shown in line width and gray shade.

5   Numerical optimization

We parametrize our problem through the adjacency matrix $\tilde{\mathbf{W}}_k(\mathbf{u}_k)$, where the edge weights $\mathbf{u}_k = (u_1^k, \ldots, u_{|E_k|}^k)$ are defined according to (2). Problem (12) can be rewritten as

$$\min_{\mathbf{u}_1, \mathbf{u}_2 \geq 0} \sum_{k=1}^{2} \|\tilde{\mathbf{L}}_k(\mathbf{u}_k) - \mathbf{L}_k\|_{\mathrm{F}}^2 + \alpha \sum_{m=1}^{p} \|\mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1(\mathbf{u}_1)} \mathbf{F} - \mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2(\mathbf{u}_2)} \mathbf{G}\|_{\mathrm{F}}^2. \quad (13)$$

Its solution is carried out using standard optimization techniques, requiring the gradient of the cost function.

We differentiate the cost function (13) w.r.t. the edge weights $\tilde{w}_{ij}^k : (i,j) \in E_k$ constituting the vectors $\mathbf{u}_k$, accounting for the symmetric structure of $\tilde{\mathbf{W}}_k$. The gradient of the distance term is given by

$$\frac{\partial \|\tilde{\mathbf{L}}_k - \mathbf{L}_k\|_{\mathrm{F}}^2}{\partial \tilde{w}_{ij}^k} = 2(\mathbf{O}_k + \mathbf{O}_k^\top - 2(\tilde{\mathbf{L}}_k - \mathbf{L}_k))_{ij}$$

where $\mathbf{O}_k = (\mathrm{diag}(\tilde{\mathbf{L}}_k - \mathbf{L}_k), \ldots, \mathrm{diag}(\tilde{\mathbf{L}}_k - \mathbf{L}_k))$ is an $|V_k| \times |V_k|$ matrix with equal columns containing the diagonal of $\tilde{\mathbf{L}}_k - \mathbf{L}_k$.

The gradient of the coupling term is computed by applying the chain rule several times, as follows. First, let

$$\frac{\partial}{\partial w_{ij}^k} \tilde{\mathbf{L}}_k \;=\; \begin{pmatrix} . & . & . & . & . \\ . & +1 & . & -1 & . \\ . & . & . & . & . \\ . & -1 & . & +1 & . \\ . & . & . & . & . \end{pmatrix}$$

be a $|V_k| \times |V_k|$ matrix containing only four non-zero elements in its ith and jth row and column. Second, for each $k = 1, 2$ and $m = 1, \ldots, p$ compute the $2|V_k| \times 2|V_k|$ matrix exponent

$$\mathbf{H}_{k,m}^{ij} \;=\; \exp \begin{pmatrix} -t_m \tilde{\mathbf{L}}_k & -t_m \frac{\partial}{\partial w_{ij}^k} \tilde{\mathbf{L}}_k \\ 0 & -t_m \tilde{\mathbf{L}}_k \end{pmatrix}$$

and extract its $|V_k| \times |V_k|$ upper right block, which we denote by $\hat{\mathbf{H}}_{k,m}^{ij}$. Finally,

$$\frac{\partial}{\tilde{w}_{ij}^1} \|\mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1} \mathbf{F} - \mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2} \mathbf{G}\|_{\mathrm{F}}^2 \;=\; 2\mathrm{tr}\,(\mathbf{F}(\mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1} \mathbf{F} - \mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2} \mathbf{G})\mathbf{F}^\top \hat{\mathbf{H}}_{1,m}^{ij});$$

$$\frac{\partial}{\tilde{w}_{ij}^2} \|\mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1} \mathbf{F} - \mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2} \mathbf{G}\|_{\mathrm{F}}^2 \;=\; 2\mathrm{tr}\,(\mathbf{G}(\mathbf{G}^\top e^{-t_m \tilde{\mathbf{L}}_2} \mathbf{G} - \mathbf{F}^\top e^{-t_m \tilde{\mathbf{L}}_1} \mathbf{F})\mathbf{G}^\top \hat{\mathbf{H}}_{2,m}^{ij}).$$

6 Results

In this section, we demonstrate our HKC approach on several synthetic and real datasets coming from shape analysis, manifold learning, and pattern recognition problems. The experiments closely follow our previous work [7], and their leitmotif is, given two datasets representing similar objects in somewhat different ways, to reconcile the information of the two modalities producing a single consistent representation. We should stress that though we know the groundtruth correspondence between the vertices of the graphs representing different modalities, we are not using it in our HKC problem. Instead, we only assume to be given few corresponding functions $\mathbf{F}, \mathbf{G}$ that are used to couple the heat kernels.

In all the experiments, we used unnormalized Laplacians (1) constructed with Gaussian weights. We used $\alpha = 10^6$ in the cost (13). Optimization was performed using MATLAB optimization toolbox.

Circles. We used two graphs shaped as two eccentric circles , containing 64 points and having different connectivity (Figure 1, top). We used four corresponding functions in the HKC optimization. The closest Laplacians that produce coupled heat kernels result in edge weights shown in Figure 1 (bottom): the optimization performs a 'surgery' disconnecting the inconsistent connections and producing two connected components.

Ring. We used a ring and a cracked ring sampled at 70 points and connected using four nearest neighbors (Figure 2, top and bottom). Three functions only were used for coupling (Figure 2, three leftmost columns). Because of the topological difference, the behavior of the heat flow differs dramatically (Figure 2, fourth column from left) The HKC optimization cuts the connections in the first graph, making the two rings topologically equivalent and resulting in the same heat flow (Figure 2, rightmost)

Man. We used two poses of the human shape from the TOSCA dataset [5], uniformly sampled at 500 points and connected using five nearest neighbors. The resulting graphs have different topology (the hands are connected or disconnected, compare Figure 3 top and bottom), resulting in a very different heat flow. Two functions were used for coupling (Figure 3, two leftmost columns) in our HKC problem; our optimization disconnects these links (Figure 3, right) making the heat flow in both cases behave similarly.

NUS. We used a subset of the NUS-WIDE dataset [10] containing images (represented by 64-dimensional color histograms) and their text annotations (represented by 1000-dimensional distributions of most frequent tags) from

seven classes. The classes were selected on purpose in order to be ambiguous in different modalities: for example, in the Tags modality *underwater tigers* can be similar both to *tigers* and *water animals*, as they share many tags. On the other hand, in the Color modality *tigers* may be similar to the class of *nature*, containing images with orange-yellow autumn colors [16].

In each modality, we used Laplacians with Gaussian weights and 25 nearest neighbors, computed with self-tuning scales. Seven functions were used for coupling. We computed diffusion distances (4) on the original and the modified graphs, and used them to rank the dataset entries in a leave-one-out retrieval experiment. Retrieval performance was evaluated using *mean average precision* $mAP = \sum_{r=1}^{R} P(r) \cdot rel(r)$, where $rel(r)$ is the relevance of a given rank (one if it belongs to the same class of the query and zero otherwise), R is the number of retrieved results, and $P(r)$ is *precision at r*, defined as the percentage of relevant results in the first r top-ranked retrieved matches. Respectively, *recall $R(r)$* is defined as the percentage of relevant results in the first r top-ranked retrieved matches out of all items belonging to the query class.

Figure 4 shows the precision-recall curve of different methods, and Table 1 summarizes the mean average precision. We can see that after HKC optimization, performance increases significantly, outperforming each modality on its own. Figure 5 shows examples of first matches corresponding to ambiguous queries. For reference only, we show the performance of Laplacian averaging, which however relies on bijective correspondence between the graphs (which is not used in our HKC problem) and is thus not directly comparable.

7   Conclusions

We showed the heat kernel coupling problem, whereby we seek to minimally modify a pair of Laplacians to make the corresponding heat kernels to become coupled, such that the solution of a heat equation on two graphs behaves consistently. This problem generalizes simple Laplacian averaging to the setting when the correspondence between the two graphs is unknown.
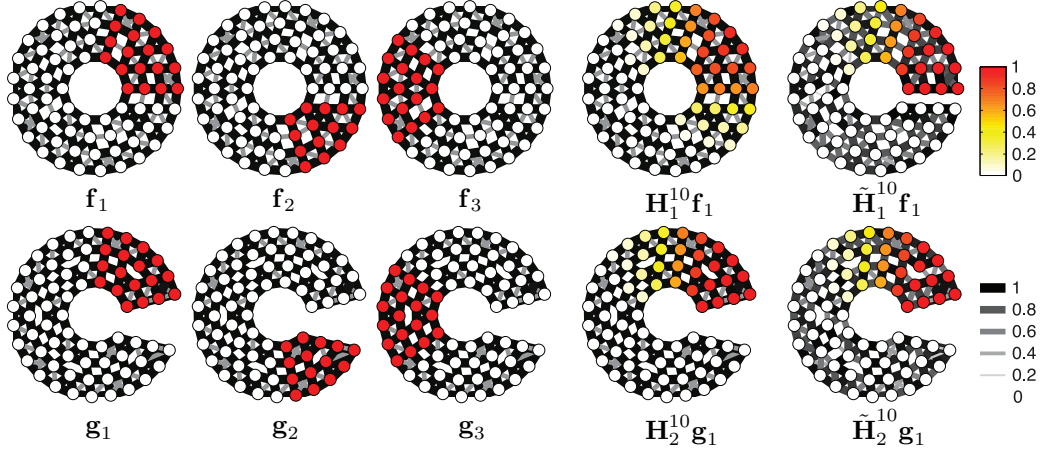
Figure 2: *Ring* experiment. Left: coupling functions; Right: Solution of the heat equation at $t = 10$ on the original graphs G_1, G_2 (fourth column) and the modified graphs $\tilde{G}_1, \tilde{G}_2$ (fifth column) using $\mathbf{f}_1, \mathbf{g}_1$ as initial conditions. The modification amounts to cutting the ring G_1 (top right).

Method	t	Precision@5	mAP
	0.75	82.3 %	78.4 %
Tags only	1.0	81.2 %	77.0 %
	1.25	79.7 %	76.2 %
	0.75	61.8 %	55.7 %
Color only	1.0	61.6 %	53.5 %
	1.25	59.6 %	51.5 %
	0.75	87.3 %	82.2 %
HKC Tags	1.0	86.3 %	81.5 %
	1.25	84.4 %	80.3 %
	0.75	83.2 %	76.2 %
HKC Color	1.0	82.3 %	75.6 %
	1.25	80.6 %	74.6 %
	0.75	68.7 %	64.2 %
Average	1.0	66.5 %	61.0 %
	1.25	63.6 %	58.8 %

Table 1: Retrieval performance on the NUS dataset using diffusion distances with different time scale t.
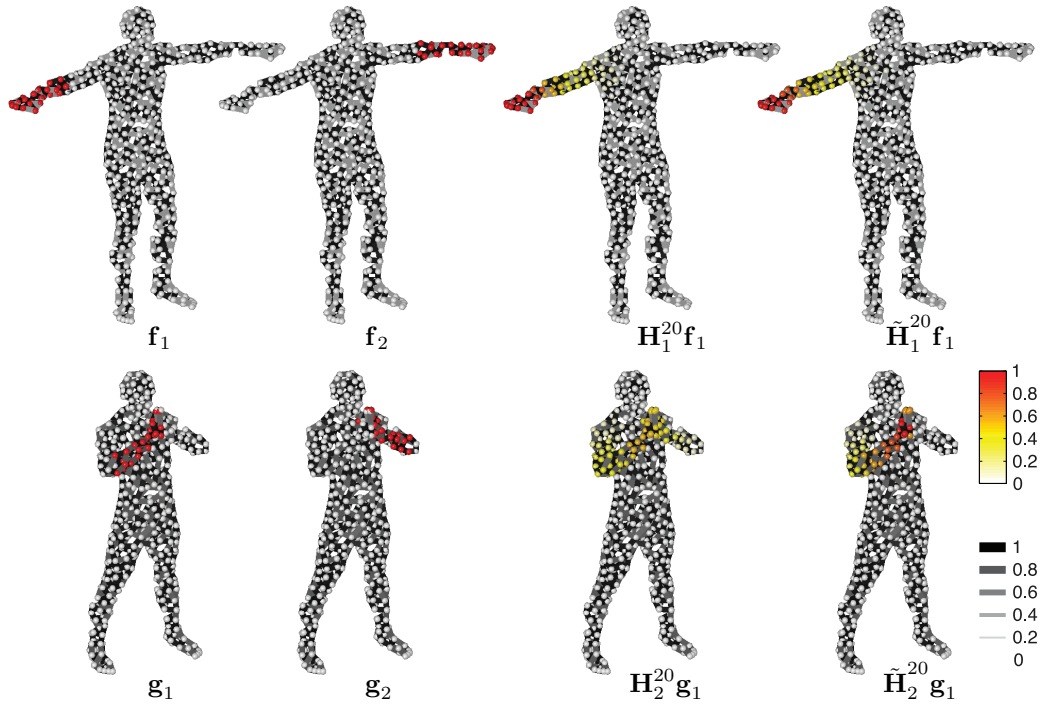
Figure 3: *Man* experiment. Left: coupling functions; Right: Solution of the heat equation at $t = 20$ on the original graphs G_1, G_2 (third column) and the modified graphs $\tilde{G}_1, \tilde{G}_2$ (fourth column) using $\mathbf{f}_1, \mathbf{g}_1$ as the initial conditions.
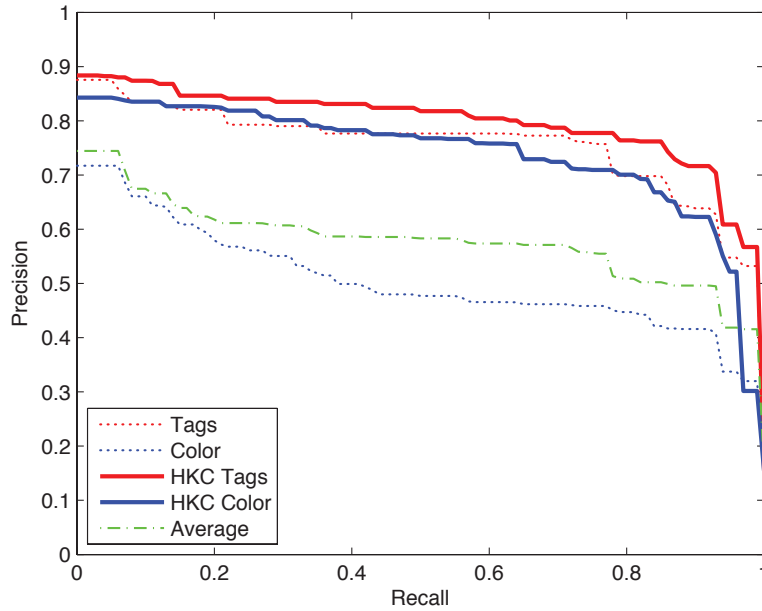
Figure 4: Precision-recall curve of the *NUS* experiment, using diffusion distances d_1 and d_2 computed from uni-modal Laplacians (Tags modality (dotted red) and Colors modality, (dotted blue), respectively), diffusion distances $\tilde{d}_1$ and $\tilde{d}_2$ computed from the HKC-modified Laplacians (solid red and blue, respectively). Result of Laplacian averaging (green dash-dot) requires bijective correspondence between the data in the two modalities and is shown for reference only. In all cases, time scale $t = 1.25$ is used.
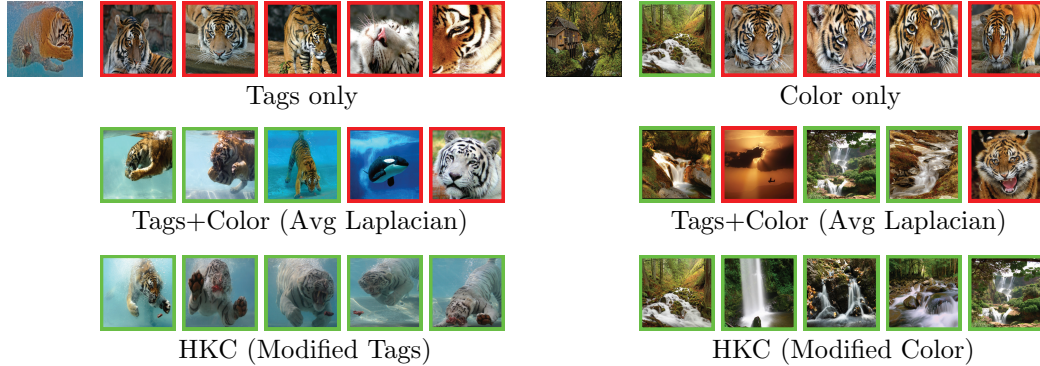
Figure 5: *NUS* experiment. Five nearest neighbors to two queries measured using diffusion distance with $t = 1.25$, using single modality (first row; Tags: left, Color: right), multiple modality by Laplacian averaging (second row) and HK coupling (third row). Matches belonging to the same class of the query are marked with green. Using Tags modality only, the *swimming tiger* query (tagged as *underwater, tiger, animal, zoo*) is confused with *tigers* (tagged as *animal, zoo, cat, tiger, nature*). Using Color modality only, the orange-yellow colored *nature* query is confused with similarly colored *tigers*.

8   Acknowledgement

References

[1] X. Alameda-Pineda, V. Khalidov, R. Horaud, and F. Forbes. Finding audio-visual events in informal social gatherings. In *Proc. ICMI*, 2011.

[2] M. Bansal and K. Daniilidis. Joint spectral correspondence for disparate image matching. In *Proc. CVPR*, 2013.

[3] R. Bekkerman, R. El-Yaniv, and A. McCallum. Multi-way distributional clustering via pairwise interactions. In *Proc. ICML*, 2005.

[4] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2002.

[5] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer, 2008.

[6] M. M. Bronstein, A. M. Bronstein, F. Michel, and N. Paragios. Data fusion through cross-modality metric learning using similarity-sensitive hashing. In *Proc. CVPR*, 2010.

[7] M. M Bronstein, K. Glashoff, and T. A. Loring. Making Laplacians commute. *ArXiv:1307.6549*, 2013.

[8] X. Cai, F. Nie, H. Huang, and F. Kamangar. Heterogeneous image feature integration via multi-modal spectral clustering. In *Proc. CVPR*, 2011.

[9] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Analysis Appl.*, 17:161–164, 1996.

[10] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. NUS-WIDE: A real-world web image database from National University of Singapore. In *Proc. CIVR*, 2009.

[11] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

[12] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, F. Warner, and S. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426–7431, 2005.

[13] V.R. de Sa. Spectral clustering with two views. In *Proc. ICML Workshop on Learning with Multiple Views*, 2005.

[14] C. H. Q. Ding, X. He, H. Zha, M. Gu, and H. D. Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Proc. ICDM*, 2001.

[15] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov. Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. *ArXiv:1303.2221*, 2013.

[16] D. Eynard, K. Glashoff, M.M. Bronstein, and A.M. Bronstein. Multimodal diffusion geometry by joint diagonalization of Laplacians. *ArXiv:1209.2295*, 2012.

[17] J. Ham, D. Lee, and L. Saul. Semisupervised alignment of manifolds. In *Proc. Conf. Uncertainty in Artificial Intelligence*, 2005.

[18] G. Irie, D. Liu, Z. Li, and S.-F. Chang. A Bayesian approach to multimodal visual dictionary learning. In *Proc. CVPR*, 2013.

[19] E. Kidron, Y. Y. Schechner, and M. Elad. Pixels that sound. In *Proc. CVPR*, 2005.

[20] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, K. Glashoff, and R. Kimmel. Coupled quasi-harmonic bases. *Computer Graphics Forum*, 32:439–448, 2013.

[21] A. Kumar, P. Rai, and H. Daumé III. Co-regularized multi-view spectral clustering. In *Proc. NIPS*, 2011.

[22] C. Ma and C.-H. Lee. Unsupervised anchor shot detection using multi-modal spectral clustering. In *Proc. ICASSP*, 2008.

[23] J. Masci, M. M. Bronstein, A. M. Bronstein, and J. Schmidhuber. Multimodal similarity-preserving hashing. *Trans. PAMI*, 2014.

[24] B. McFee and G. R. G. Lanckriet. Learning multi-modal similarity. *JMLR*, 12:491–523, 2011.

[25] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In *Proc. NIPS*, 2005.

[26] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Proc. NIPS*, 2001.

[27] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. J. Guibas. Functional maps: A flexible representation of maps between shapes. *Trans. Graphics*, 31(4), 2012.

[28] M. Ovsjanikov, Q. Mérigot, F. Mémoli, and L. Guibas. One point isometric matching with the heat kernel. *Computer Graphics Forum*, 29(5):1555–1564, 2010.

[29] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. G. Lanckriet, R. Levy, and N. Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proc. ICM*, 2010.

[30] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *Proc. CVPR*, 2012.

[31] J. Shi and J. Malik. Normalized cuts and image segmentation. *Trans. PAMI*, 22:888–905, 2000.

[32] V. Sindhwani, P. Niyogi, and M. Belkin. A co-regularization approach to semi-supervised learning with multiple views. In *Proc. ICML Workshop on Learning with Multiple Views*, 2005.

[33] W. Tang, Z. Lu, and I. S. Dhillon. Clustering with multiple graphs. In *Proc. Data Mining*, 2009.

[34] C. Wang and S. Mahadevan. Manifold alignment using procrustes analysis. In *Proc. ICML*, 2008.

[35] C. Wang and S. Mahadevan. A general framework for manifold alignment. In *Proc. Symp. Manifold Learning and its Applications*, 2009.

[36] Y. Weiss, A. Torralba, and R. Fergus. Spectral hashing. In *Proc. NIPS*, 2008.

[37] J. Weston, S. Bengio, and N. Usunier. Large scale image annotation: learning to rank with joint word-image embeddings. *Machine Learning*, 81(1):21–35, 2010.