Volume 31 (2012), Number 5 pp. 1–14 COMPUTER GRAPHICS *forum*

Multiscale Kernels using Random Walks

A. Sinha and K. Ramani

Purdue University

Abstract

We introduce novel multiscale kernels using the random walk framework and derive corresponding embeddings and pairwise distances. The fractional moments of the rate of continuous time random walk (equivalently diffusion rate) are used to discover higher order kernels (or similarities) between pair of points. The formulated kernels are isometry, scale, and tessellation invariant, can be made globally or locally shape aware, and are insensitive to partial objects and noise based on the moment and influence parameters. Additionally, the corresponding kernel distances and embeddings are convergent and efficiently computable. We introduce dual GMS signatures based on the kernels and discuss the applicability of the multiscale distance and embedding. Collectively, we present a unified view of popular embeddings and distance metrics while recovering intuitive probabilistic interpretations on discrete surface meshes.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems

Key Symbols	
α, β,n	Influence parameters α, β; moment parameter n
Δ, w	Laplace-Beltrami operator Δ with weights w
Δ_M	Scaled Laplace-Beltrami operator
$\Delta_{\alpha M}$	Lazy rate operator
Λ, λ	Eigenvalue matrix Λ with eigenvalues λ
Φ	Eigenvectors
π	Stationary distribution vector
τ	Multiscale kernel (Moment time operator)
Υ, υ	υ is diagonal of $\lvert Q \rvert$ with maximum value Υ
ξ	Vector of node self-weights
A, vol	Area matrix A, volume of graph vol
D, d	Degree matrix D with weights d
$\mathcal{D}$	Kernel distance
$G, \mathcal{G}, \bar{G}$	Green's function of $\Delta_M, \mathcal{L}, L_c$ respectively
H_t, h_t	Heat kernel H_t with values h_t
$\mathcal{L}, L_c$	Normalized and conformal Laplacian
M, p	Transition matrix M with weights p
N	Normalized transition matrix
P_t, p_t	Markov kernel P_t with values p_t
Q, q	Rate matrix Q with weights q
t	Time
u, v, V	Mesh nodes u, v in set V
$\mathcal{Z}$	Fundamental matrix

1. Frame of Reference

Traditional methods for shape analysis resort to a single level approach, where the mesh structure is used to formulate global metrics. However the paradigm is shifting to multiscale methods which help understand a shape over multiple levels and discover salient features on the mesh at global as well as local scales. Current multiscale methods include the heat kernel [SOG09] and multiscale biharmonic kernel [Rus11b] which operate over the mesh domain with a tuning parameter, and have been applied to shape segmentation, correspondence and retrieval [BBGO11, SOCG10, OMMG10]. These kernels differ in weighting of eigenvalues and set up the motivating question for this paper, '*Is there a generalized principle guiding the weighting of eigenvalues and hence the construction of multiscale kernels, while being intuitive in the choice of kernel or scale?*'

We aim to answer this challenging question using the Markov chain framework of random walks. The choice of this framework is motivated by the fact that Markov chains emerge as a generalization of the heat equation and the intimate relationship between the discrete Laplace-Beltrami operator and the rate matrix of a random walk. For a general discussion on random walks, the reader is asked to refer to the excellent survey by Lovász [Lov93]. As meshes are point-sampled and subsequently triangulated, the ran-

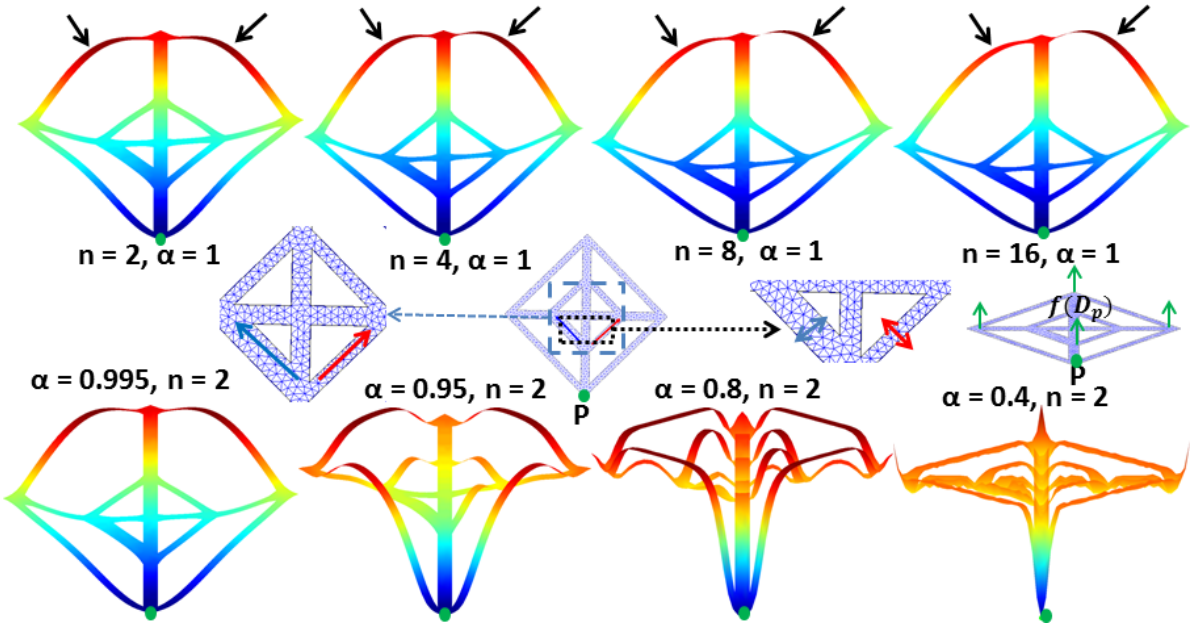


Figure 1: ***Center**: Asymmetric mesh with unequal widths (left strip (blue arrow) > right strip (red arrow)) along with exploded views. (Right) Distances are measured from point P and viewed as a height function $f(D_p)$ above mesh domain (green arrows). Color changes from blue to red as distance increases. **Top Row**: Varying t^n with $n = 2,4,8,16$, respectively, scales underlying asymmetry in path length distribution with n acting as 'frequency filter'. High n separates distance from P to top left and top right strips displayed using black arrows. **Bottom Row**: Varying α with $\alpha = 0.995, 0.95, 0.8, 0.4$, respectively, makes asymmetry disappear with 'scale shift'. At small α all points are at approximately the same distance, apart from points in immediate vicinity*

dom walk framework naturally fits into general mesh analysis. It has already been used in a wide array of applications including mesh segmentation, cutting, denoising etc. [SRML07, LHMR08, ZZC11]. In this paper, we use random walks to construct multiscale kernels with two intuitive parameter choices, which generalize the current approaches under a single framework. All formulated kernels are proven to be convergent and positive definite. Hence they inherit all the '*nice*' properties of popular kernel embeddings (GPS, spectral [Lév06], etc.) and kernel distance metrics (biharmonic, commute-time, etc.) which we briefly review.

The Heat Kernel Signature (HKS), related to diffusion kernels [NLCK05] [KL02], was proposed as a point signature. The idea is to extract information about the object using time t as a scaling parameter. However, choosing a suitable t is not an easy task and most approaches rely on heuristic tuning with no straightforward interpretation [ASC11]. Our approach to this problem comes from the mixing rate of a Markov chain which is intimately connected to the rate of change of the heat kernel, and hence, the diffusion process. As the kernel approaches uniform distribution, we contend its rate of convergence to equilibrium is equally (if not more) informative. This is the central motivation of this paper and key separating factor from other approaches, i.e., to extract information about shapes using the rate of diffusion in a random walk setting. Our work is also inspired by the multiscale biharmonic kernel and biharmonic distance [LRF10]. The biharmonic distance derived from the biharmonic operator (square of Laplacian) is a parameter free distance metric on meshes with inverse square weighting of the eigenvalues. Rustamov's multiscale biharmonic paper advances a novel way to construct general multiscale kernels by minimizing the Laplacian energy subject to the sparsity inducing lasso constraint. On the surface, it is unclear how the inverse exponent weighting of eigenvalues achieves the scaling. As demonstrated in [Rus11a], the answer is interlinked with wavelet transforms on graphs and point clouds. Similar observations have been made in the machine learning community in the context of iterated ranking on manifolds using the Green's function [ZBS11]. Our approach provides an intuitive explanation while generalizing the biharmonic kernel, GPS embedding [Rus07], biharmonic distance and commute time distance [QH07, FPmRS06] on meshed surfaces.

Specifically, the multiscale kernels $\tau^{\alpha n}$ are constructed by integrating the moments of the probability transition rates of a random walk and governed by two parameters n and α:

$$\tau^{\alpha n} = \int_{t=0}^{\infty} t^n [\Delta_{\alpha M} e^{-t\Delta_{\alpha M}}] dt \qquad (1)$$

where t is time, M is the transition matrix and $\Delta_{\alpha M}$ is the lazy rate matrix. Note Section 2 ties M and $\Delta_{\alpha M}$ with the Laplace-Beltrami operator Δ. Figure 1 displays the multiscale distances from the green point P for the asymmetric mesh displayed in the center of the figure. The bottom left strip (blue arrow) is thicker than the bottom right strip (red arrow), and hence, there are more diffusion pathways from point P to nodes on the top left side of the mesh relative to the right. This underlying asymmetry in connecting pathways from P emerge as we increase the moment parameter n from 2 to 16, i.e., distances to nodes on the top right strip relative to the top left strip gradually increase (marked using arrows in top figure), made visible using color plot (red implies large distance) and height field (distance along vertical direction with mesh as base). Decreasing α has the opposite effect of progressively suppressing the asymmetry (see bottom row). For $\alpha = 0.4$, the entire mesh is approximately at the same distance from P, apart from the points in the immediate vicinity. These observations can be explained in terms of equation 1:

- First, replacing $\Delta_{\alpha M}$ with Δ, one immediately establishes that the $\Delta e^{-t\Delta}$ term in the integrand is the differential of the heat operator $e^{-t\Delta}$, hence equal to the negative rate of heat diffusion.
- Second, the n^{th} moment of diffusion rate with respect to time, t^n, scales the rate of diffusion, i.e., say $n = 2$ weights the diffusion rate between two points at large times more heavily than say $n = 1$ over the integral. In effect, diffusions occurring over long time periods are penalized for high values of n and are equally weighted for $n = 0$, hence the moment extracts scale (see Figure 1). From a signal processing viewpoint, the role of n is to filter frequencies similar to parameter t in the heat kernel, i.e., higher frequencies are progressively suppressed by increasing n [ASC11].
- As $\Delta_{\alpha M}$ can be written as $I - \alpha M$ (see Section 2), each element of the operator $e^{-t\Delta_{\alpha M}}$ can be expanded using the Maclaurin series of $e^{-t(I-\alpha M)}$ as:

$$e^{-t\Delta_{\alpha M}}(u,v) = e^{-t}\sum_{k=0}^{\infty}\alpha^k M^k(u,v)\frac{(t)^k}{k!} \qquad (2)$$

  $M^k(u,v)$ can be interpreted as the sum of all random walks of length k joining points u and v [BWH05]. In the series, paths of length k get weight α^k and hence shorter paths (say k_1) will be weighted more than longer ones (say k_2) for $0 < \alpha < 1$ ($\alpha^{k_2} < \alpha^{k_1}$) [New10]. From a signal processing viewpoint, the role of α is to shift the scale of analysis as it is implicitly added to eigenvalues (see Figure 1 and Section 3).
- α^k and t^k are multiplicative parameters in equation 2 and aid each other, i.e., decreasing α suppresses longer path lengths and large t discovers longer paths connecting two points. This is intuitive as time and distance are coupled notions. Allowing dual parameters α and n to control scale provides greater flexibility and reveals different multiscale information (see Section 4).
- Finally, integrating over the entire time interval eliminates the choice of the appropriate time interval and replaces it with the range of moment parameter which we perceive to be more intuitive. It also makes the kernel robust to noise and small topology changes.

The main contribution of this paper is the formulation of the dual-multiscale kernels (over n and α), associated embeddings and distances (Section 3), after formalizing the discrete graph based random walk framework (Section 2). The theoretical properties, relationship with existing kernels and metrics along with applications are established (Section 4). Finally we discuss the conclusions and proposed future work (Section 5).

2. Preliminaries

In this section we explain the construction of probability transition matrix M for discrete random walks, and the construction of rate matrix Q for continuous time random walks on a triangulated surface mesh. We derive relevant spectral decompositions which are used in subsequent sections. The triangulation (discrete state space) assumption is consistent with general analysis on point-sampled meshes.

2.1. Graph Construction

The rate matrix Q is negative of the discrete Laplace-Beltrami operator Δ. This stems from the fact that Brownian motion and heat flow both satisfy the diffusion equation, and that Brownian motion is a generalization of continuous time random walks to continuous state spaces. This relationship can be directly observed by comparing the heat diffusion equation $\frac{\delta H_t}{\delta t} = -\Delta H_t$ (H_t is the heat kernel) and Kolmogorov's backward equation $\frac{\delta P_t}{\delta t} = QP_t$ (P_t is the Markov kernel) for continuous time Markov chains [Law06]. The Brownian interpretation of heat kernel was briefly explained in [SOG09].

It is well known that the Laplace-Beltrami operator Δ is a generalization of the Laplacian from flat spaces to manifolds [MDSB02]. The operator can be constructed using the common cotangent discretization [MDSB02]. The weights w of the Laplace-Beltrami operator are given by

$$w(u,v) = \begin{cases} \sum_v \frac{cota_{u,v}+cotb_{u,v}}{2A_u} & \text{if} \quad v = u \\ -\frac{cota_{u,v}+cotb_{u,v}}{2A_u} & \text{if} \quad v \subseteq N_1(u) \\ -\frac{cota_{u,v}}{2A_u} & \text{if} \quad (u,v) \in E_B \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{N}_1(u)$ is the set of 1-ring neighbours of vertex u, $a_{u,v}, b_{u,v}$ are the two angles supporting the edge connecting vertices u and v, A_u is the associated surface patch or finite volume (usually barycentric or Voronoi) and E_B indicates a boundary edge. The $b_{u,v}$ term drops out for boundary

points imposing the von Neumann boundary conditions. We assume that the triangulation is regular, i.e., $a_{u,v} + b_{u,v} \leq \pi$ so that all weights are positive. Another commonly used operator, the conformal Laplacian L_c [PP93] is related to the Laplace-Beltrami operator by $\Delta = A^{-1}L_c$, where A is a diagonal matrix of the Voronoi, barycentric or mixed area (A_u for point u). As $Q = -\Delta$, the weights q of the rate matrix are given by

$$q(u,v) = \begin{cases} \sum_v \frac{-(cota_{u,v}+cotb_{u,v})}{2A_u} & \text{if} \quad v = u \\ \frac{cota_{u,v}+cotb_{u,v}}{2A_u} & \text{if} \quad v \subseteq N_1(u) \\ \frac{cota_{u,v}}{2A_u} & \text{if} \quad (u,v) \in E_B \\ 0 & \text{otherwise} \end{cases}$$

Verify that $\sum_v q_{uv} = 0 \; \forall v \in V$ where V is the set of all nodes, which means that the sum of all transition rates to or from a node is conserved, i.e., flow conservation. For discrete random walks, the probability transition matrix M satisfies:

Property 1: The single-step transition probability of jumping from any node u to an adjacent node v ($(p_{uv})_{v \in N_1(u)}$) is positive, i.e., $0 \leq p_{uv}$

Property 2: The probability of a random walker to jump from state u to any state is 1, i.e., $\sum_v p_{uv} = 1 \; \forall v \in V$ (conservation property).

We use the uniformization technique of Markov chains to build M satisfying the above properties. Uniformization is a technique to simulate continuous time chains using a discrete chain analog (see Lawler [Law06] for a detailed exposition). Suppose υ_u represents the absolute diagonal values of the rate matrix Q and $\Upsilon = \max_u \upsilon_u$, then the uniformized chain is given by

$$p(u,v) = \begin{cases} \frac{q(u,v)}{\Upsilon} & \text{if} \quad u \neq v \\ 1 - \frac{\upsilon_u}{\Upsilon} & \text{if} \quad u = v \end{cases}$$

Using the above transformation, the elements of the transition matrix M are

$$p(u,v) = \begin{cases} 1 - \sum_v \frac{cota_{u,v}+cotb_{u,v}}{2\Upsilon A_u} & \text{if} \quad v = u \\ \frac{cota_{u,v}+cotb_{u,v}}{2\Upsilon A_u} & \text{if} \quad v \subseteq N_1(u) \\ \frac{cota_{u,v}}{2\Upsilon A_u} & \text{if} \quad (u,v) \in E_B \\ 0 & \text{otherwise} \end{cases}$$

It is easy to verify that regular triangulation and scaling by Υ ensure satisfaction of Property 1 and 2 respectively. Note Property 1 does not hold for general meshes. The diagonal elements of the transition matrix are zero only if u coincides with the index of the maximum element Υ. The non-zero diagonal values ξ_u indicate presence of self-loops wherein a random walker remains in the same state with positive probability and with probability $(1 - \xi_u)$ hops to a 1-ring neighbour (also called the lazy random walk). For notational consistency, we scale the values of the Laplace-Beltrami operator Δ by $1/\Upsilon$ and represent it by Δ_M. Note the relations $Q = -\Upsilon\Delta_M$ and $\Delta_{\alpha M} = I - \alpha M$, where $\Delta_{\alpha M}$ is the lazy rate operator and α is a tuning parameter discussed in Section 3.

2.2. Random Walks

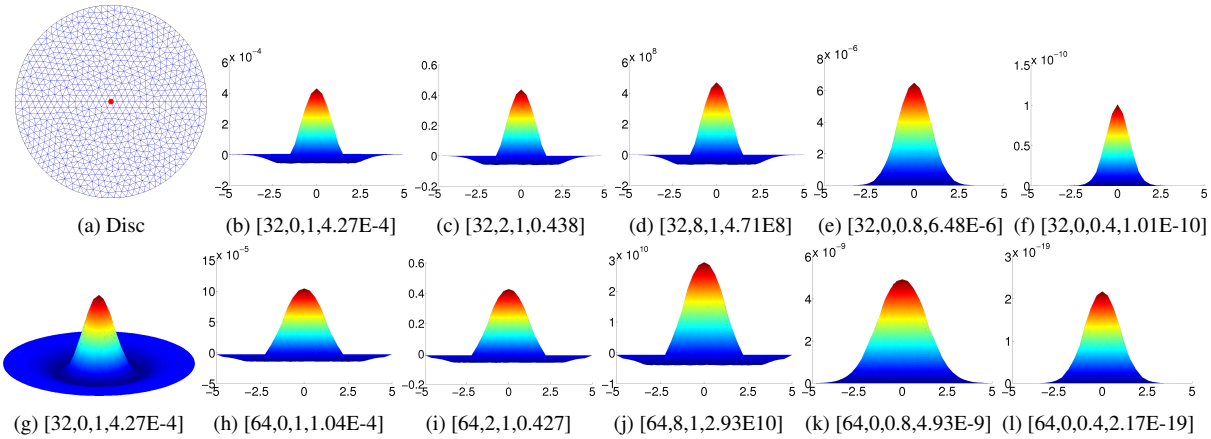
Random walks can be viewed as special cases of finite time-reversible irreducible Markov chains. Let D be a diagonal matrix with elements $d_u = \Upsilon A_u$ called the degree matrix. Note the equivalence of the following relations: $M = I - D^{-1}L_c = I - \Delta_M$. While M is asymmetric, it is easy to bring it to a symmetric form by considering $N = D^{1/2}MD^{-1/2}$ with the corresponding Laplacian denoted by $\mathcal{L}$, and termed the normalized Laplacian as per standard graph terminology. $\mathcal{L} = D^{1/2}\Delta_M D^{-1/2} = D^{-1/2}L_c D^{-1/2}$ which is similar to Δ_M, i.e., the same eigenvalues and u^{th} component of the eigenvector scaled by the inverse square root of its degree $(1/\sqrt{d_u})$. The relationship between the normalized Laplacian and transition matrix is given by $M = D^{-1/2}(I - \mathcal{L})D^{1/2}$. The spectral decomposition of the normalized Laplacian is given by $\mathcal{L} = \Phi\Lambda\Phi^T$ where Λ is a diagonal matrix with ordered eigenvalues, i.e., $0 = \lambda_1 \leq \lambda_2 ... \leq \lambda_u ...$, Φ is a matrix with the corresponding eigenvectors as columns. The eigen decomposition of the normalized transition matrix is $N = \Phi\Lambda'\Phi^T$ with the same set of eigenvectors and eigenvalues related as $\Lambda' = I - \Lambda$ (see [Chu97]). The stationary distribution π or the probability of being at a vertex after the walk has reached equilibrium (after long time or at $t = \infty$) is independent of the initial distribution. It can be proved that $\pi = \mathbf{1}D/vol$ where $\mathbf{1}$ is a vector with all coordinates 1 and $vol = \sum_u d_u$ called the volume of the graph [Chu97]. Intuitively, the stationary distribution is proportional to the area term A_u and it reproduces local information around a node. With these definitions handy, we briefly discuss continuous time random walks.

Continuous Time Random Walk : Discrete random walks take place over discrete time steps starting from $t = 0$ over the positive integer domain. The difference between the discrete and the continuous setting of a random walk is the waiting time between hops of a random walker, i.e., constant 1 for the discrete walk (as $t = 0, 1, 2 \ldots$) and an exponential distribution for continuous time. Recall, the exponential distribution characterizes the waiting time in a Poisson process, and hence, the number of jumps completed by a random walker at time t is a Poisson distribution. We solve the Kolmogorov backward equation to determine the probability distribution over the state spaces at time t and the solution is naturally the heat kernel

$$\frac{\delta H_t}{\delta t} = -\Delta_M H_t \text{ with solution, } H_t = e^{-t\Delta_M}$$

Using the orthogonality of the eigenvectors ($\Phi_i^T\Phi_j = \delta_{ij}$), the relation $\Delta_M = D^{-1/2}\Phi\Lambda\Phi^T D^{1/2}$ and noticing that the normalized Laplacian matrix is diagonalizable, the matrix exponential is calculated as $H_t = D^{-1/2}\Phi e^{-t\Lambda}\Phi^T D^{1/2}$. Hence, the spectral decomposition of the heat kernel at time t is

$$h_{uv}^t = \pi_v + \sum_{i \geq 2} e^{-t\lambda_i}\Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{3}$$



(a) Disc (b) [32,0,1,4.27E-4] (c) [32,2,1,0.438] (d) [32,8,1,4.71E8] (e) [32,0,0.8,6.48E-6] (f) [32,0,0.4,1.01E-10]

(g) [32,0,1,4.27E-4] (h) [64,0,1,1.04E-4] (i) [64,2,1,0.427] (j) [64,8,1,2.93E10] (k) [64,0,0.8,4.93E-9] (l) [64,0,0.4,2.17E-19]

Figure 2: *(a) Disc of radius 5. (g) 3D plot of rate of diffusion from center (red point) of disc viewed as height function and color coded (rate decreases as color changes from red to blue). (b)-(f) and (h)-(l): 2D plot of function* $t^n[\Delta_{\alpha M} e^{-t\Delta_{\alpha M}}]$ *for different values of* t, n, α*, viewed as height field (function value) along vertical axis and distance from center along horizontal axis. The quadruple indicates* $[t, n, \alpha, \varsigma]$ *values where* ς *is the maximum value of function indicative of weightage in the integral*

The uv^{th} entry h^t_{uv} of the heat kernel H_t evaluates the probability of transition from u to v in time t. This probability tends to the stationary distribution (π_v) in the long time limit as $\Lambda > 0$. Hence, H_t becomes uninformative for large t values as the equilibrium value at nodes are proportional to the corresponding area. Also, the conservation property (Property 2) holds for all non-negative values of t, i.e., $\sum_v h^t_{uv} = 1$ $\forall v \in V$ [KL02].

3. Formulation

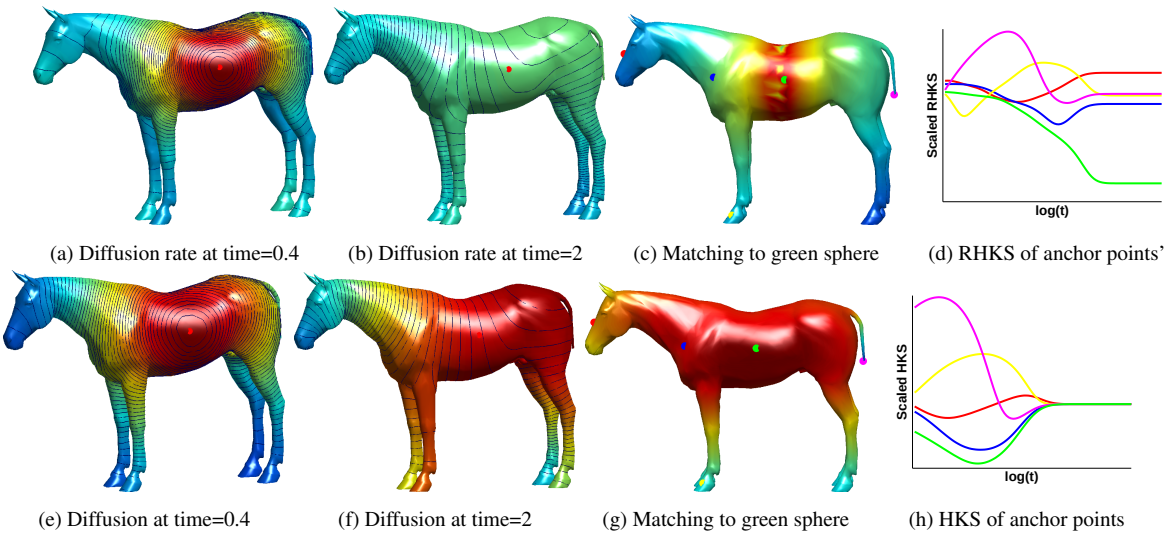
The formulation of the multiscale kernels is based on two parameters- a moment parameter n and an influence parameter α, given by equation 1

$$\tau^{\alpha n} = \int_{t=0}^{\infty} t^n [\Delta_{\alpha M} e^{-t\Delta_{\alpha M}}] dt$$

Although t^n can be replaced by general functions of t, we restrict our attention to powers in t (n is a positive fraction), and term $\tau^{\alpha n}$ as the moment time operator. The implication of this functional form was discussed in Section 1. Loosely speaking, n and α produce similar effect of exaggerating/suppressing '*faraway*' nodes. However, they do so in very different ways, which we illustrate by plotting the values of $t^n[\Delta_{\alpha M} e^{-t\Delta_{\alpha M}}]$ for different values of $[t, n, \alpha]$, from the center of a disc in Figure 2. Plot 2g shows the rate of diffusion values to all points from the center as a color plot (red implies higher values) and height field from the base of the disc. Note, the flow conservation property holds $\forall t > 0$. Formally, $\sum_v \Delta_M e^{-t\Delta_M}(u,v) = 0$ $\forall v \in V$. This can be verified by writing $\Delta_M e^{-t\Delta_M} dt$ as $(e^{-t\Delta_M} - e^{-(t+dt)\Delta_M})$. As all rows of $e^{-t\Delta_M}, e^{-(t+dt)\Delta_M}$ sum to 1 (Property 2), flow conservation follows. It does not hold in general for $\alpha \neq 1$. Plots 2b, 2h display the rate of diffusion at times $t = [32, 64]$ with almost equal maximum values $\approx [4 \cdot 10^{-4}, 10^{-4}]$. Weighting the rate of diffusion by t^n produces similarly shaped scaled plots for the same time, as t^n is a multiplicative factor to the term in parenthesis. It can be visually verified by comparing Plot 2b to Plots 2c, 2d and Plot 2h to Plots 2i, 2j. However, t^8 weighting produces two orders of magnitude difference in the maximum value $\approx [4.7 \cdot 10^8, 2.9 \cdot 10^{10}]$ at the two times. This validates the claim that parameter n penalizes the rate of diffusion at long time scales more than short ones in the integral, and large values of n provide global information about the shape. The influence parameter α affects the diffusion rate on the disc in two ways. Firstly, it suppresses the diffusion to nodes further away from the center, consequently suppressing the rate (notice the progressive squeezing in Plots 2b to 2e to 2d and Plots 2h to 2k to 2l as α decreases from 1 to 0.8 to 0.4). Secondly, α decreases the magnitude of rate of diffusion, affecting the magnitude at long time scales more than short ones. The maximum value drops from $\approx 6 \cdot 10^{-6}$ (Plot 2e) to $\approx 10^{-10}$ (Plot 2f) for $t = 32$, but drops from $\approx 5 \cdot 10^{-9}$ (Plot 2k) to $\approx 2 \cdot 10^{-19}$ (Plot 2l) at larger time $t = 64$. This can be understood by considering the equivalent formulation for discrete random walks (equation 14) and that the term in parenthesis can be written as $\alpha^t[M^{t-1}/\alpha - (M)^t]$, with α appearing as a multiplicative factor with time t as exponent. Hence, the notion of '*faraway*' nodes for parameters n, α can be stated as:

- For the moment parameter n, '*faraway*' is related to the rate of diffusion before converging to the equilibrium distribution, and hence, (loosely) tied to time scales.
- For the influence parameter α, '*faraway*' is related to the separation between the source and destination nodes, and hence, (loosely) tied to distance scales.

We do not treat time and distance as independents (both pa-



(a) Diffusion rate at time=0.4 (b) Diffusion rate at time=2 (c) Matching to green sphere (d) RHKS of anchor points'

(e) Diffusion at time=0.4 (f) Diffusion at time=2 (g) Matching to green sphere (h) HKS of anchor points

Figure 3: *Comparison of Rate of Diffusion (Top) vs Diffusion (Bottom). (a)-(b),(e)-(f) show isocontours and color plots of the magnitude of diffusion/ rate of diffusion emanating from red sphere at different time values. Magnitude decreases as color changes from red to blue. At small times isolines are localized and at large times the color plot is intuitive for rate of diffusion (see description in Section 3). (c),(g) display color plots of difference between scaled signatures for point marked by green sphere and other points on model for same time range $[t_1, t_2]$. Difference increases as color changes from red to blue. Signatures derived from diffusion rate are informative. (d),(h) displays scaled HKS (diffusion) and RHKS (rate of diffusion) vs logarithmic time scale for 5 color coded points on model in (c)/(g). RHKS is discriminative at all scales*

rameters influence time and distance) but attach the notion of time to n and distance to α, as it makes intuitive sense and for ease of comprehension. We demonstrate in Section 4 that each reveal useful multiscale information about the triangulated mesh. An equivalent operator (namely the Pochhammer time operator) for discrete random walks is formulated in the appendix. To sum it up, the basic intuition in defining these operators is that the '*rate of diffusion is informative for understanding shapes*'. This point is illustrated further in Figure 3 where we compare the diffusion kernel to the rate of diffusion kernel for $\alpha = 1$. The key separating feature is that the smallest eigenvalue and its corresponding (uninformative) constant eigenvector does not appear in the rate kernel. This is true as $\Delta_M e^{-t\Delta_M} dt$ can be written as $(e^{-t\Delta_M} - e^{-(t+dt)\Delta_M})$ and using equation 3. This is beneficial as it suppresses the fastest decaying eigenvalue (λ_1) and most global eigenvector (Φ_1). Hence the rate kernel is '*more local*' relative to the heat kernel and does not converge to the uniform distribution, as the constant eigenvector which dominates at long time is removed. The isocontours of diffusion rate are more intuitive, as they follow the shape locally and the derived signatures are discriminative at all scales, unlike the HKS. These claims are validated in Figure 3. The Rate of Heat Kernel Signature (RHKS) is defined similar to HKS on $\Delta_M e^{-t\Delta_M}$ instead of $e^{-t\Delta_M}$. Plots 3c, 3g indicate robust matching using RHKS to green anchor point and plots 3d, 3h indicate discriminative signatures at all scales unlike HKS. The isolines and color plots of the rate of diffusion from the anchor point are more localized at small times (plot 3a vs. 3e); also the isolines follow similar pattern for time $t = 2$. The color plots, plot 3b vs. 3f, are starkly different and can be explained by considering $1/\lambda_2$ value which is ≈ 2. $1/\lambda_2$ can be interpreted as the mixing time of diffusion (see Section 4) and plot 3b clearly indicates that the rate of diffusion is small on the entire horse model, hence diffusion has nearly converged to equilibrium. Thus, we validate our claim that the rate of diffusion is equally informative and more discriminative, compared to the diffusion kernel. Having defined the operator, we establish desired symmetric and positive-definite kernels, and hence, develop the distance metric and the corresponding embedding.

3.1. Kernel Computation

We compute the multiscale operator using the eigen spectrum of the transition matrix and show that the result is intimately connected to the discrete Green's function, and hence the commute time and biharmonic kernel. Further, we use the kernel distance to evaluate pairwise distances and characterize the associated embedding.

Moment Time Operator : The term in the parenthesis in equation 1 can be written as $[e^{-t\Delta_{\alpha M}} - e^{-(t+dt)\Delta_{\alpha M}}]$. Hence, the null space drops out for $\alpha = 1$ by substituting the spectral representation of $e^{-t\Delta_M}$ (Recall in equation 3, eigenvalue 0 and the corresponding eigenvector appear as π_v). Using the

orthogonality of eigenvectors and evaluating the integral, $\tau^{\alpha n}$ simplifies to,

$$\tau^n_{uv} = \Gamma(n+1)\sum_{i\geq 2}\frac{1}{\lambda_i{}^n}\Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{4}$$

or $\tau^n = \Gamma(n+1)D^{-1/2}\hat{\Phi}\hat{\Lambda}^{-n}\hat{\Phi}^T D^{1/2}$ for $\alpha = 1$ where Γ is the gamma function and $\hat{\Lambda}$ and $\hat{\Phi}$ indicate the removal of the 0 eigenvalue and the corresponding eigenvector. For $0 < \alpha < 1$, $\Delta_{\alpha M}$ (equivalently $I - \alpha M$) becomes invertible (no 0 eigenvalue). To see this, write $I - \alpha M$ as $\alpha(\beta I + \Delta_m)$ where $\beta = \frac{1-\alpha}{\alpha}$. Using the relation that eigenvalues of $(\beta I + \Delta_m)$ are $(\beta + \Lambda)$ and $\Phi^T\Phi = I$, the representation is

$$\tau^{\alpha n}_{uv} = \frac{\Gamma(n+1)}{\alpha^n}\sum_{i\geq 1}\frac{1}{(\beta+\lambda_i)^n}\Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{5}$$

or $\tau^{\alpha n} = \Gamma(n+1)\alpha^{-n}D^{-1/2}\Phi(\beta+\Lambda)^{-n}\Phi^T D^{1/2}$ for $0 < \alpha < 1$. Note the starting index i is different for the two equations above, based on the value of α. Henceforth, general results are derived for α and one appropriately changes the limits in the summand when $\alpha = 1$. Recall that the eigenvectors are in decreasing smoothness order for the Laplacian operator as we arrange the eigenvalues in increasing order, and hence, the order n scales the smoothness while filtering out smaller '*frequencies*'. Thus, t^n interprets the scale on the mesh surface with larger n revealing larger scales on the mesh surface. Following the same reasoning, the smaller the α value, the larger the value of β and hence α performs a '*spectra or scale shift* ' while killing the null space of the transition matrix [ZBS11]. From these equations, it is easy to verify that the symmetric matrix $D^{1/2}\tau^{\alpha n}D^{-1/2}$ is positive-definite and so is the matrix $\tau^{\alpha n}(diag(\pi))^{-1}$ obtained by pre and post multiplying by $(diag(\pi))^{-1/2}$, where $diag(\pi)$ is a diagonal matrix of stationary probabilities. We use this kernel to define kernel distances, as other mesh based distance metrics naturally arise from it.

Fundamental Matrix and Green's Function : We define the derived kernels in terms of standard operators and show the generic nature of our method in reference to other popular kernels. In the Markov chain theory, the fundamental matrix $\mathcal{Z}^0$ for continuous time random walk is defined as

$$\mathcal{Z}^0_{uv} = \int_0^\infty (h^t_{uv} - \pi_v)dt \tag{6}$$

where h^t_{uv} is an element of the heat kernel. This matrix is very useful for a wide array of calculations like expected number of visits to a state, expected return and access times [Kei79]. The Green's function introduced in 1828 by George Green and its discrete counterpart associated with the Laplacian, introduced in a 2000 paper [CY00], have received much attention. Formally, it is the pseudo-inverse of the defined Laplace operator and can be more explicitly stated as

$$G_{uv}\Delta_{M(uv)} = \delta_{uv} - \frac{d_v}{vol} \tag{7}$$

$$\mathcal{G}_{uv}\mathcal{L}_{uv} = \delta_{uv} - \frac{\sqrt{d_u d_v}}{vol} \tag{8}$$

where G and $\mathcal{G}$ are the left inverse operators of Δ_M and $\mathcal{L}$ respectively. It can be verified that $\mathcal{Z}^0$ and G reduce to the same spectral form as τ^1 (see [CY00] and page 107 [Kei79]). Hence the relationship between the defined kernel, fundamental matrix and Green's function using $G^n = \mathcal{G}^n\sqrt{\frac{d_v}{d_u}}$ can be verified to be: $\tau^n = \Gamma(n+1)(\mathcal{Z}^0)^n = \Gamma(n+1)G^n$. From here on, in order to maintain consistency, we use the Green's function for representation.

3.2. Distance Computation

We use the kernel distance to evaluate pairwise distances between points on a mesh. The kernel distance is defined for positive definite matrices as $(\mathcal{D}^{\alpha n}_{uv})^2 = (\kappa_{uu} - 2\kappa_{uv} + \kappa_{vv})$, where κ is a positive definite kernel. Hence, the general expressions using the discrete Green's function for distance between two vertices u and v for the generated kernels are

$$(\mathcal{D}^{\alpha n}_{uv})^2 = vol^n\left(\frac{\mathcal{G}^{\alpha n}_{uu}}{d_u} + \frac{\mathcal{G}^{\alpha n}_{vv}}{d_v} - 2\frac{\mathcal{G}^{\alpha n}_{uv}}{\sqrt{d_u d_v}}\right) \tag{9}$$

$$(\mathcal{D}^{\alpha n}_{uv})^2 = vol^n\left(\frac{G^{\alpha n}_{uu}}{d_u} + \frac{G^{\alpha n}_{vv}}{d_v} - \frac{G^{\alpha n}_{uv}}{d_v} - \frac{G^{\alpha n}_{vu}}{d_u}\right) \tag{10}$$

where $vol^n = \Gamma(n+1)\alpha^{-n}vol$, $\mathcal{G}^{\alpha n} = \Phi(\beta+\Lambda)^{-n}\Phi^T$ or $\mathcal{G}^n = \hat{\Phi}\hat{\Lambda}^{-n}\hat{\Phi}^T$ as required, based on the parameters n and α. We use the relation $G^{\alpha n} = \mathcal{G}^{\alpha n}\sqrt{\frac{d_v}{d_u}}$ to derive the second relation from the first. In terms of spectral representation this can be written as

$$(\mathcal{D}^{\alpha n}_{uv})^2 = vol^n\sum_{i\geq 1}\frac{1}{(\beta+\lambda_i)^n}\left(\frac{\Phi_i(u)}{\sqrt{d_u}} - \frac{\Phi_i(v)}{\sqrt{d_v}}\right)^2 \tag{11}$$

Note that the sum starts from $i = 2$ for $\mathcal{D}^n_{uv}$ and $i = 1$ for $\mathcal{D}^{\alpha n}_{uv}$. Thus to calculate the exact pairwise distances, we need to find the full set of eigenvectors and corresponding eigenvalues for the normalized Laplacian. Using the decreasing smoothness property of the eigenvectors, multiscale distances can be approximated using only the first K eigenvectors and provide a trade-off between accuracy and computational time. We shall return to computational complexity in the next section.

3.3. Kernel Embedding

Having derived the kernel distances, we see that the embedding that preserves the n^{th} distance moment with parameter α is given by:

$$\Theta^{\alpha n} = \sqrt{vol^n}(\beta+\Lambda)^{-n/2}\Phi^T D^{-1/2} \tag{12}$$

Note that the $D^{-1/2}$ scaling appears due the scaling of the conformal Laplacian by area terms. To drive home this

point, if the kernels were derived starting from the conformal Laplacian instead of the normalized Laplacian, the embedding would be: $\tilde{\Theta}^{\alpha n} = \sqrt{vol^n}(\tilde{\beta}+\tilde{\Lambda})^{-n/2}\tilde{\Phi}^T$, where $\tilde{\lambda}$ and $\tilde{\Phi}$ are the eigenvalues and eigenvectors of L_c and $\tilde{\beta}$ is the corresponding influence parameter. A similar embedding has recently been proposed in [ZBS11] for understanding manifolds, though with a very different approach. The difference in the embedding is due to the scaling of the operator. This can be understood by noting the relation between the Green's function of the induced subgraphs of $L_c, \mathcal{L}$ and Δ_M is $\tilde{G} = GD^{-1} = D^{-1/2}\mathcal{G}D^{-1/2}$ where $\tilde{G}$ is the Green's function of the conformal Laplacian [CY00]. As the Green's function of the induced subgraphs differ only by a scaling matrix, it is not that surprising that the two embeddings are equivalent upto a scaling matrix. In the next section we state properties of the kernels, distances, embeddings and show related applications.

4. Properties and Applications

In this section, we describe how random walks and specifically the derived kernels present a unified view of embeddings and distance evaluation on discrete meshes. We discuss possible applications of the kernel, embedding and distance metric. We first state the theoretical properties.

4.1. Theoretical Properties

The formulated kernels are isometry, scale and tessellation invariant, can be made shape aware, insensitive to partial objects and noise based on the moment function and an additional influence parameter. Additionally, the corresponding kernel distances and embeddings have all the important properties, i.e., are metrics, convergent and fast to compute. We give a brief summary of the properties.

Multiscale: The central theme of the last section was the dual formulation of the multiscale operator using parameter α and moment function t^n. An intuitive interpretation of the α value is that the remainder $(1-\alpha)M$ probability translates to a random surfer model, where with probability $(1-\alpha)$, a random walker leaves the current node and teleports to another node, without regard to mesh connectivity. Hence, random contributions due to the teleportation cancel out in the kernel evaluation, resulting in increasingly local affinities with decrease in α. It is interesting to note that the α value is a fundamental component in the celebrated PageRank algorithm and precisely performs the same function, albeit for ranking search queries [BP98]. On the other hand, t^n penalizes pairwise nodes which have a slow rate of convergence to the equilibrium distribution, i.e., nodes separated by longer path lengths as a function of time. Thus the kernels are dual multiscale, (loosely) over distance as well as over time scales with different multiscale behaviour, and is a key differentiation from other multiscale kernels which define multiscale behaviour over either time (heat kernel) or distance scales

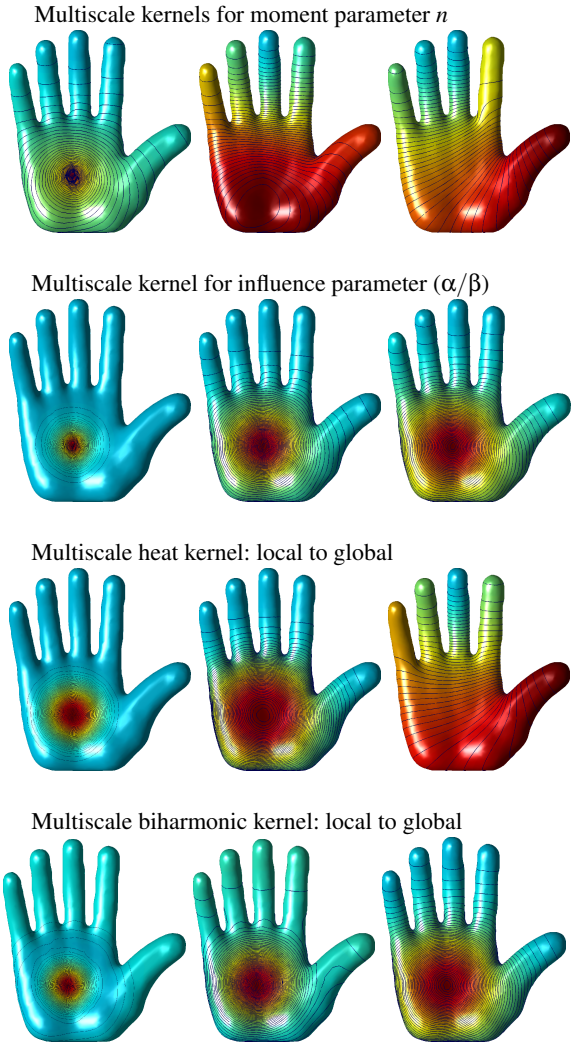


Figure 4: *Comparison of four multiscale kernels- moment, influence, heat and biharmonic kernels with local (left) to global (right) support for point at center of palm*

(multiscale biharmonic kernel). Figure 4 shows the multiscale behavior over moment n, influence α, the heat kernel over time and multiscale biharmonic kernel over the lasso constraint. We observe that the multiscale behavior over α is similar to the multiscale biharmonic kernel and the behavior over moment n is similar to the behavior over t in heat kernel. Multiscale over n has an interesting behavior. For large moments, the maximum affinity from the center of hand becomes skewed towards the thumb, which has an intuitive explanation. The operator weighs the change in transition probability and visual inspection of affinity gradient indicates that the diffusion from center of palm towards the thumb occurs over short time interval, whereas a random

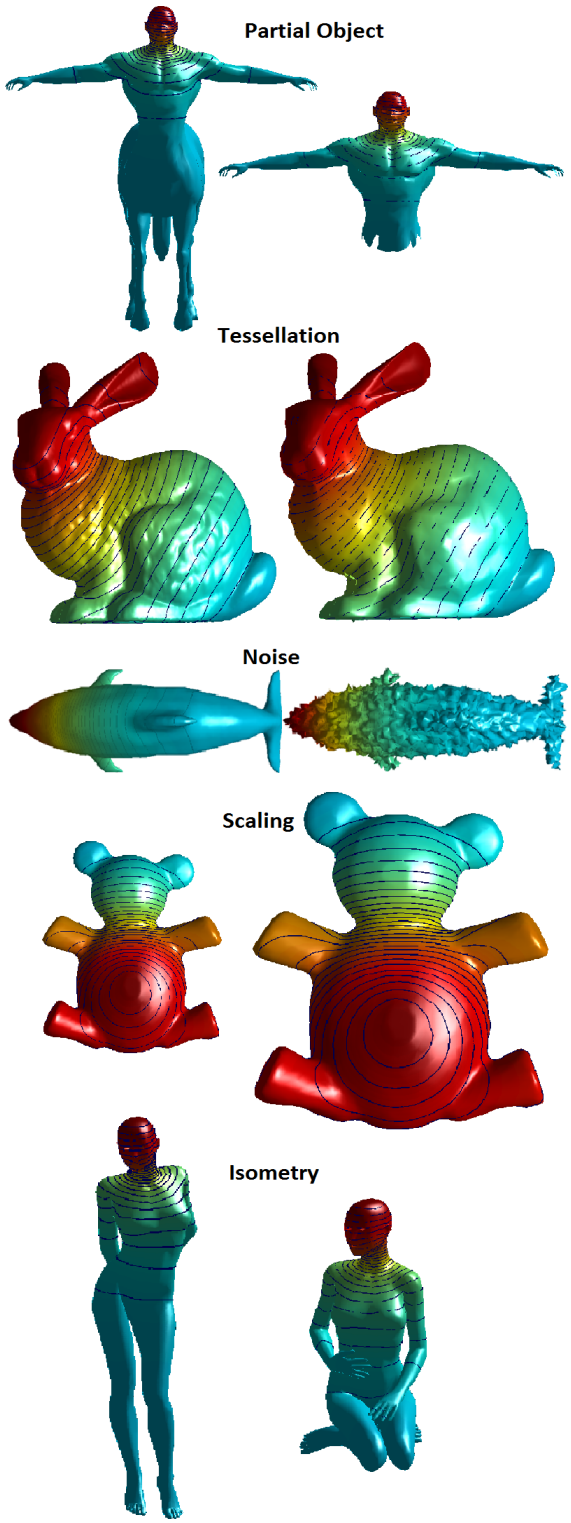


Figure 5: *Insensitivity of Multiscale Kernels to partial objects, tessellation, noise ,scaling and deformation for different mesh models from fixed anchor point*

walker in the central regions and other four fingers tend to get '*lost*', and hence diffusion occurs over a large time interval. It also indicates there are shorter pathways connecting the palm point to the thumb and relatively longer pathways connecting to the rest of the hand. The flip in the least affinity (blue color) from the middle finger to ring finger can also be explained similarly.

Scale, Isometry and Tessellation Invariant: For scale invariance, the kernel weighting function on the eigenvectors must satisfy $K(\gamma/\omega^2) = \omega^{2-2/p}K(\gamma)$ [BB11]. It can be verified that the weighting function $K(\beta+\lambda) = (\beta+\lambda)^{-1/p}$ satisfies this property and hence the kernels are scale invariant. In Figure 5 the teddy model is scaled 100 times, however the coloring on the mesh is unaffected, validating this property. As the transition/rate matrices are derived from the discrete Laplace-Beltrami operator, calculated using the isometry invariant curvature normals, the derived kernels are also isometry invariant. Isometric deformation of the Victoria model in Figure 5 does not change the isocontours confirming this theoretical property. The defined transition (M) and rate matrices (Δ_M) implicitly account for area terms and hence tessellation invariant. The bunny model in Figure 5 shows the invariance of the isolines of the kernel as the model is simplified from 35000 to 1500 vertices. Additionally, invariance to noise is indicated in the dolphin model by adding gaussian noise (200%) proportional to the average edge length. The partial and full centaur models in Figure 5 display similar isocontours from fixed anchor point on the top of the human head. The partial centaur model was created by spectral bipartition into 2 segments, which is more natural than passing a cutting plane through the object.

Convergence: The convergence of the multiscale kernels over finite graphs for t^n where n is any positive real number can be verified using the property that the integral $\int_0^\infty e^{-t}t^{z-1}dt$ (Gamma function) is absolutely convergent for all z with $Re(z) > 0$ after doing a variable change $t' = t/\lambda_i$ (Note $\lambda_i > 0$). As each term in the spectral representation converges and the representation is finite, it follows that the kernel converges. However for continuous state space, the Green's function of a 2D surface would have logarithmic singularity along the diagonal and not defined [LRF10].

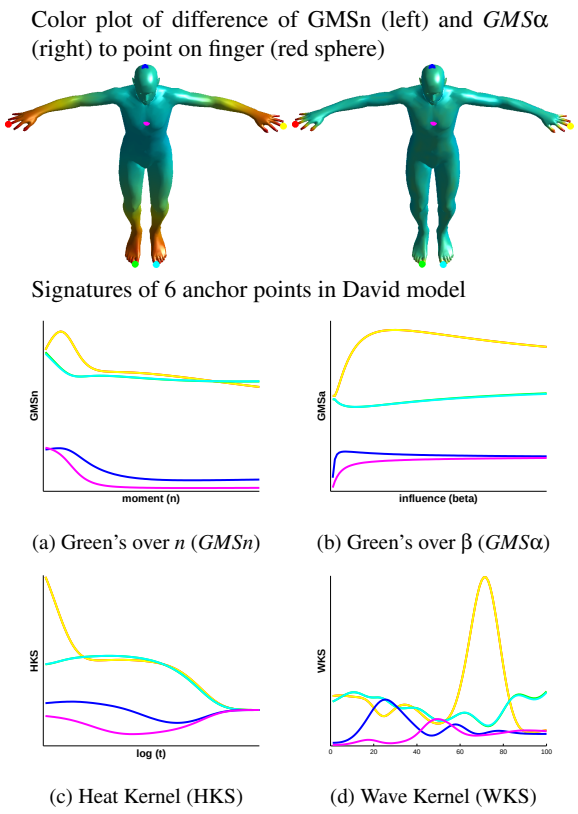
Metric: The kernel $\tau^{\alpha n}(diag(\pi))^{-1}$ is positive definite as $D^{1/2}\tau^{\alpha n}D^{-1/2}$ is positive-definite and we pre and post multiply with a diagonal matrix $(diag(\pi))^{-1/2}$ which is a positive diagonal matrix. Because the multiscale distances are derived using a strictly positive definite kernel, it satisfies all the necessary conditions for it to be a valid metric, i.e., it is non-negative and the diagonal values vanish. The fact that it is symmetric follows from the observation that interchanging u and v does not change the distance calculation. Also $\mathcal{D}^{\alpha n}_{uv} = 0$ *iff* $u = v$ because if this were not true then $\Phi_i(u) = \Phi_i(v)\ \forall i$, as the eigenvectors form an orthonormal basis, all functions f over the surface would take the same value, reaching a contradiction.

Complexity: The multiscale kernels (Green's function) and hence distances can be computed either using the complete eigen spectrum $\mathcal{O}(|V|^2)$ or using the first K eigenvectors $\mathcal{O}(|V|^{3/2})$). An alternative approach is proposed in [LRF10] for calculating on a set of vertices by solving a set of linear equations($\mathcal{O}(S|V|)$ where S is the size of the subset of vertices). The subset calculation is exact and the complexity can be reduced to almost linear time for a small subset ($\mathcal{O}(|V|)$ when $S << V$). However, such an approach suffers from the disadvantage that we require independent computations to find multiscale kernels with a different set of parameters. The eigenvectors and eigenvalues of the Laplace operator offer a common orthonormal basis and all multiscale kernels can be computed by changing the weightage (exponent and parameter β) of eigenvalues. We use the spectral approach and use the first 300 eigenvectors to approximate the kernel matrix. Hence the computation time is the same as for the approximate heat kernel evaluation.

4.2. General Discussion with Applications

The multiscale kernels consolidate popular metrics under a single framework and we discuss possible applications.

Multiscale Kernel: In Section 3, we stressed that the intuition driving the formulation was that of using the rate of change instead of the probability values, and hence, is fundamentally different from the heat kernel. However, the resulting kernel is functionally equivalent to heat kernel, i.e., exponential with parameter n (instead of t) with logarithmic weighting of eigenvalues ($\lambda^{-n} = e^{-n\ln\lambda}$ with $\lambda > 0$). Our first observation is that as we average over entire time, the resulting kernel possesses robust multiscale characteristics. Second, the logarithmic weighting of eigenvalues (instead of linear in heat kernel) makes the multiscale kernel discriminative at all scales, unlike the heat kernel which converges to a uniform distribution. To see this, logarithm of eigenvalues in the interval $(0,1)$ are negative, hence contribute exponentially to the Green's function and for very large scales, i.e, as $n \to \infty$, the sum will be dominated by the normalized Fielder vector and second eigenvalue. Specifically, the multiscale kernel embedding will tend to $\sqrt{vol^n}\lambda_2^{-n/2}\Phi_2^T D^{-1/2}$ in the limit . Inspired by the remarkable informative property of the heat kernel and these observations, we define the dual Green's Mean Signature (*GMS*) of vertex u to be $|(\frac{\mathcal{G}^n_{uu}}{\sum_v \mathcal{G}^n_{vv}})|$ and $|(\frac{\mathcal{G}^\alpha_{uu}}{\sum_v \mathcal{G}^\alpha_{vv}})|$, and name them *GMSn* and *GMS*α, to indicate the multiscale nature over moment and influence parameter respectively. The α parameter used in our multiscale kernels has a direct semblance with the parameter t used to derive multiscale biharmonic kernels. Note the parameter α (hence β) performs a frequency shift of the eigenvalues, and hence, indirectly enforces partial support of the eigenfunctions. Interestingly, the wave kernel signature (*WKS*) [ASC11] can be linked to random walks, specifically maximal entropy



(a) Green's over n (*GMSn*) (b) Green's over β (*GMS*α)

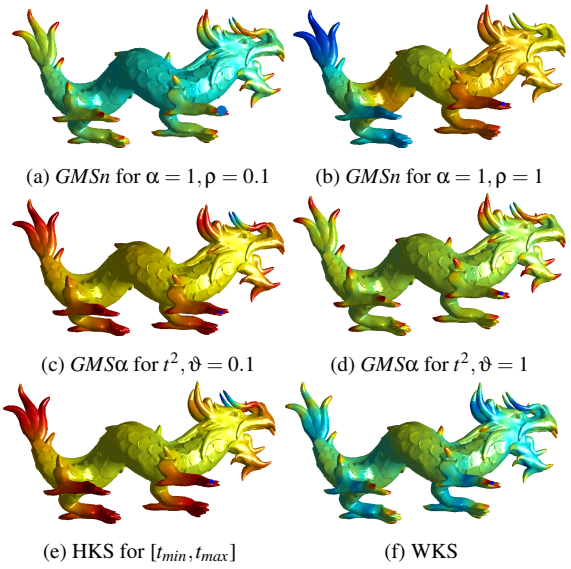
(c) Heat Kernel (HKS) (d) Wave Kernel (WKS)

Figure 6: *Green's Mean Signature difference (GMSn and GMS*α*) between finger and other points on David model (top) based on entire range of parameters. In color plot red indicates similar and cyan indicates dissimilar,* $\alpha = 1$ *for GMSn and* $n = 2$ *for GMS*α*. (a)-(d) represent the 4 signatures for 6 anchor points. Signatures for 2 fingers and 2 feet are indistinguishable because of symmetry for all signatures. Note GMSn and GMS*α *are discriminative at all scales (unlike HKS) and multiscale (unlike WKS)*

random walks (MERW) [SGGnL*11]. The transition matrix (M_Ψ) for MERW is given by

$$M_\Psi(u,v) = \frac{\bar{M}(u,v)\bar{\Phi}_{0v}}{\bar{\lambda}_0\bar{\Phi}_{0u}}$$

where $\bar{M} = DM = D - L_c$ and $\bar{\lambda}, \bar{\Phi}$ are associated eigenvalues and eigenvectors respectively. The unique stationary distribution is $(\bar{\Phi}_{0i})^2$ and can be interpreted as the probability of finding a particle in the ground state of the operator $(-\bar{M})$ with wave function $\bar{\Phi}_{0i}$ [Dud12]. The particular choice of log-normal energy distribution in *WKS* is derived from a perturbation-theoretical analysis. *GMSn* emphasizes on global features and has higher '*specificity (detect false negative errors)*' while *GMS*α suppresses large-scale features and has higher '*sensitivity (detect false positive errors)*', similar to *HKS* and *WKS* respectively, as argued in [Bro11]. Having shown the interrelationship to HKS, WKS



(a) *GMSn* for $\alpha = 1, \rho = 0.1$ (b) *GMSn* for $\alpha = 1, \rho = 1$

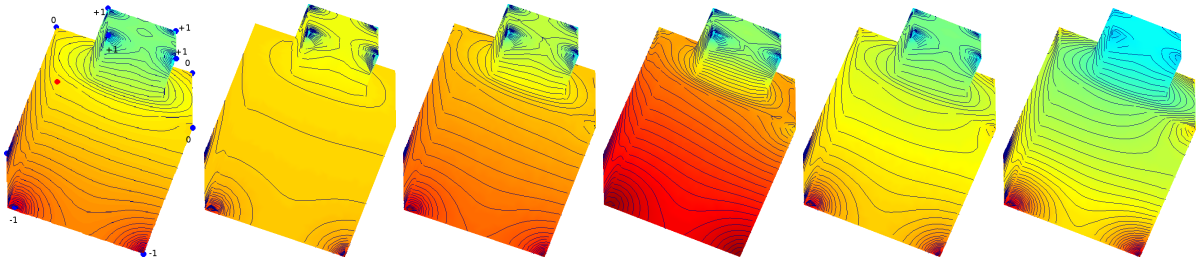
(c) *GMS*α for $t^2, \vartheta = 0.1$ (d) *GMS*α for $t^2, \vartheta = 1$

(e) HKS for $[t_{min}, t_{max}]$ (f) WKS

Figure 7: *Signature similarity to point marked on front feet with blue sphere in (a) with difference increasing as color ranges from red to cyan to blue. (a)-(b) GMSn signature for two values of* ρ. *(c)-(d) GMS*α *signature for two values of* ϑ. *Increasing* ρ *produces global signature while increasing* ϑ *produces local signature. (e) HKS and (c); (f) WKS and (a) have similar color plots*

and multiscale biharmonic kernels, we now discuss computation and compare our signatures. As mentioned earlier, we find the 300 smallest eigenvalues and associated eigenvectors of $\mathcal{L}$, or alternatively solve for the generalized eigenvalue problem $L_c\Phi = D\lambda\Phi$. Using the formulae in the previous sections, we calculate the *GMSn* or *GMS*α as required by fixing the range of n to $[0, \rho log(1/\lambda_2)]$ and range of β to $[0, \vartheta\lambda_{300}]$ where ρ, ϑ are scale tuning parameters with 100 linear increments over the range. $1/\lambda_2$ is the mixing time of a random walk and as GMSn kernels are logarithmically weighted over eigenvalues, hence transforming the logarithmic time scaling in heat kernel to linear scaling over n, we empirically take the logarithm of the mixing time. Figure 6 shows the comparison between $GMSn, GMS\alpha, HKS, WKS$ for David model and we confirm the theoretical claim that the set of GMS signatures are discriminative at all scales unlike the *HKS*, which decay to equilibrium distribution. Also, the ρ, ϑ parameters provide a multiscale setting which is not possible with WKS. Figure 6 shows matching between the node on finger marked by red sphere and the rest of the shape, by calculating the L_2-norm of the difference between point signatures. It is represented using color plot, with similarity decreasing as color changes from red to cyan and $\rho = 1, \vartheta = 1$. We observe that the *GMSn* provides a global understanding of the shape while the *GMS*α provides a local understanding hence matching the fingers. These characteristics are further exemplified using dragon model in Figure 7 where we show multiscale matching by varying the parameters ρ and ϑ. The specificity of *GMSn* in matching the legs and local sensitivity of *GMS*α in identifying protrusions, can be coherently combined in a learning framework to construct optimal signatures, using the approach in [Bro11], which we leave to future work.

Multiscale Distance: The 0^{th} moment operator is $I - \pi\mathbf{1}^T$ and corresponding distance is $vol(\frac{1}{d_u} + \frac{1}{d_v})$. This can be understood as $\frac{vol}{d_u}$ is the distance from any node on the graph to vertex u, because the expected value for a random walker would depend only on the degree (scaled area) of the destination vertex and by symmetry. We only recover local connectivity, thus naturally at the lowest scale in our set of multiscale kernels. The commute time distance is the expected time for a random walker to travel from one vertex to another and back, i.e., it is the first moment. The commute distance derived in [QH07] is precisely the multiscale distance for $n = 1$ and $\alpha = 1$. Considering $n = 2$ and $\alpha = 1$, the distance metric is equivalent to biharmonic distance [LRF10] upto a constant scaling factor. The biharmonic distance is thus the second moment of the rate of heat diffusion which we discern to be more intuitive, i.e., it weights diffusion rate between 2 nodes by t^2 and hence is a more global metric relative to the commute time distance. The diffusion distance between 2 nodes separated by time t is defined as $(\mathcal{D}_{uv}^{Diff})^2 = \sum_{i \geq 1} e^{-t\tilde{\lambda}_i} \left(\tilde{\Phi}_i(u) - \tilde{\Phi}_i(v)\right)^2$. These multiscale distances can be viewed as performing a Mellin transform on $e^{-t\Lambda}$ (Λ is a constant for given mesh). It replaces the exponential over t weighting scheme by zeta function ($\zeta(n) = \sum_{\lambda_i \neq 0} \lambda_i^{-n}$) which is exponential over moment n and infers scale akin to time. Such an approach has been applied to the trace of the heat kernel to successfully extract graph characteristics [XHW09]. Briefly, we mention that the geodesic distance for a graph is the shortest distance between 2 points and translates into finding the smallest t for which p_{uv}^t is non-zero. Hence it is connected to the floor function of the multiscale kernel distances. Thus, these kernels can be viewed as generalization of existing distance metrics and with additional parameter β, opens up a host of possibilities. We introduce scale sensitive variant of Shepard interpolation [She68] demonstrated in [LRF10]. The Shepard interpolant with zeroth order precision for point u on the mesh is defined as $f(u) = \frac{\sum_i w_i(u) f_i}{\sum_i w_i(u)}$ where we choose $w_i(u)$ to be scale sensitive weights proportional to $1/\mathcal{D}_{uv_i}^{\alpha n}$. Figure 8 shows the interpolant on the mesh along with value (ς) at fixed anchor point (red sphere) for different β values at 11 anchor points and 2 multiscale kernels ($n = 2, 4$). The β values for bottom, middle and top anchor points (blue sphere) indicated by triple in parentheses [bottom, middle, top] are constants multiplied with λ_2 and fix the scale sensitive weights w_i. The values f_i at anchor points are set at -1,0,1 for bottom, middle and top anchor points respectively. We see that changing the weighting scheme, allows the anchor points to increase or

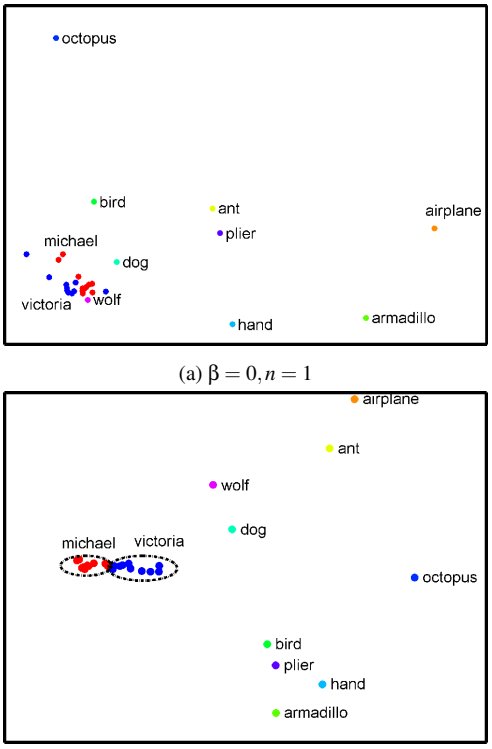


(a) $\beta = [0,0,0]\lambda_2$ $n = 2, \varsigma = -0.051$ (b) $\beta = [5,5,5]\lambda_2$ $n = 2, \varsigma = -0.005$ (c) $\beta = [0,0.5,1]\lambda_2$ $n = 2, \varsigma = 0.108$ (d) $\beta = [0,0.5,1]\lambda_2$ $n = 4, \varsigma = 0.319$ (e) $\beta = [1,0.5,0]\lambda_2$ $n = 2, \varsigma = -0.184$ (f) $\beta = [1,0.5,0]\lambda_2$ $n = 4, \varsigma = -0.439$

Figure 8: *Scale Sensitive Function Interpolation: Function values ς for point marked with red sphere with 4 bottom, 3 middle and 4 top row anchor (blue) points set at -1, 0, +1 respectively for varying n and β (hence α) weighting. ς can vary from negative to positive value based on β weighting and magnitude can be tuned using n*

decrease dominance over the function values at other points on the mesh surface. A large value of β will only allow local influence as shown in Plot 8b, while reversing the set of weights as shown in Plots 8c,8e have very different effects on the value at the red anchor point. Changing n from 2 to 4 in Plots 8d,8f, affects the magnitude at the red anchor point.

Multiscale Embedding: The first moment embedding is equivalent to the GPS embedding upto scaling Υ [Rus07]. We discuss some insights of these embeddings useful for shape segmentation. Figure 9 shows the segmentation of the ant model using naive kmeans for different moments. The 0^{th} moment embedding or the eigenvectors without eigenvalue scaling correctly segments the head, torso and abdomen of the ant model. However, increasing the moment results in skewed segmentations. As kmeans minimizes distortion, central regions which have highest affinity to other regions tend to be clustered together (the n^{th} embedding preserves the n^{th} distance moment). Indeed, larger values of k are required as input to kmeans at higher scales, to identify salient segments and the torso region suffers from oversegmentation. Consequently, the 0^{th} embedding of the kernel is best suited for segmentation, but it provides little information as a distance metric. As a second example, we illustrate the advantage of using scaled G2-distributions for calculating the shape similarity and subsequent MDS projection by varying β in equation 12 as introduced in [Rus07]. Figure 10 displays the MDS projection for 11 different shapes which include 11 isometric models of Victoria and Michael for two values of β; $\beta_1 = 0$ which corresponds to using G2-distributions ($n = 1$) and G2α-distributions for $\beta_2 = 100\lambda_2$. We see that β_2 is able to better discriminate between the models: dog, wolf, Michael and Victoria, and is also able to separate the 11 isometric models of Victoria (blue spheres) and Michael (red spheres) into two tight clusters, unlike the MDS projection using the original G2 distribution (β_1). This is because β acts as a *'scale-shift'* parameter and the embedding with β_2 is more localized, and hence, neglects contribution from distant (possibly noisy) nodes. We can create a

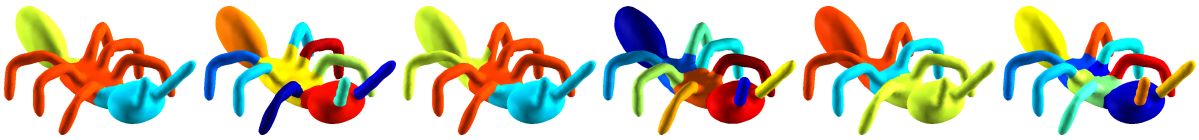


(a) $\beta = 0, n = 1$

(b) $\beta = 100\lambda_2, n = 1$

Figure 10: *Scale sensitive G2-distributions: Classical MDS projection for shape similarity by summing L2 distances between 36 G2α-distribution histograms with 50 dimensions for 11 shape models (see [Rus07] for details) (b) $\beta_2 = 100\lambda_2$ is more discriminative between shapes relative to (a) $\beta_1 = 0$*

family of distributions similar to G2 by varying n in equation 12 and these distributions can have interesting yet intuitive implications based on the application.



(a) $K = 3$, $n = 0, \beta = 0$ (b) $K = 11$, $n = 0, \beta = 0$ (c) $K = 3$, $n = 1, \beta = 0$ (d) $K = 12$, $n = 1, \beta = 0$ (e) $K = 3$, $n = 4, \beta = 0$ (f) $K = 14$, $n = 4, \beta = 0$

Figure 9: *Sensitivity of multiscale embeddings to shape segmentation: Increasing moment from $n = 0$ (a)-(b) to $n = 1$ (c)-(d) to $n = 4$ (e)-(f), leads to oversegmentation of torso region into 1,2 and 4 segments for $n = 0, 1, 4$ respectively, while ensuring all 6 legs and 2 antennas are segmented. K is input to kmeans and $\beta = 0$*

5. Conclusions and Future Work

This work presents a general framework to construct multiscale kernels on discrete meshes using random walks and in doing so presents a consolidated view of current state of the art shape signatures, distance metrics and embeddings. It also offers an intuitive explanation not offered by current frameworks. These kernels possess all necessary characteristics, hence immediately applicable for shape analysis including and not limited to shape retrieval, robust segmentation, correspondence propagation and symmetry detection. The notion of dual multiscale kernels is to our knowledge the first and the experiments show promising potential for *GMS*. One current limitation is regular triangulation of the mesh as defined in [WMKG07], so as to ensure positive weights (Property 1). For future work, we wish to investigate continuous state space random walks, i.e., Brownian motion on Riemannian manifolds and recover the same set of intuitive interpretations. We will also focus on combining the *GMSn* and *GMS*α signatures in a coherent way so as to construct optimal descriptors. Another line of research is to investigate different functional forms of t instead of t^n suited for different applications. Our work serves as a gateway from Markovian processes to discrete geometry and it is our belief that it only constitutes the tip of the proverbial iceberg in realizing the full scope of such an integrated approach.

6. Acknowledgements

We would like to thank Project TOSCA and Princeton Benchmark for sharing the 3D models. We thank the anonymous reviewers who helped improve the article with constructive suggestions.

References

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In *ICCV Workshops* (2011), IEEE, pp. 1626–1633. 2, 3, 10

[BB11] BRONSTEIN M. M., BRONSTEIN A. M.: Shape recognition with spectral distances. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 5 (2011), 1065–1071. 9

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph. 30*, 1 (Feb. 2011), 1:1–1:20. 1

[BP98] BRIN S., PAGE L.: The anatomy of a large-scale hypertextual web search engine. *Computer Networks 30*, 1-7 (1998), 107–117. 8

[Bro11] BRONSTEIN A. M.: Spectral descriptors for deformable shapes. *CoRR abs/1110.5015* (2011). 10, 11

[BWH05] BAI X., WILSON R., HANCOCK E.: Manifold embedding of graphs using the heat kernel. In *IMA* (Loughborough, UK, 2005), vol. 3604, Springer, pp. 34–49. 3

[Chu97] CHUNG F. R. K.: *Spectral Graph Theory*. American Mathematical Society, 1997. 4

[CY00] CHUNG F. R. K., YAU S.-T.: Discrete green's functions. *J. Comb. Theory, Ser. A 91*, 1-2 (2000), 191–214. 7, 8

[Dud12] DUDA J.: From maximal entropy random walk to quantum thermodynamics. *Journal of Physics: Conference Series 361*, 1 (2012), 012039. 10

[FPmRS06] FOUSS F., PIROTTE A., MICHEL RENDERS J., SAERENS M.: Random-walk computation of similarities between nodes of a graph, with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering 19* (2006), 2007. 2

[Kei79] KEILSON J.: *Markov chain models-Rarity and Exponentiality*. No. 28 in Applied mathematical sciences. Springer, New York, USA, 1979. 7

[KL02] KONDOR R. I., LAFFERTY J. D.: Diffusion kernels on graphs and other discrete input spaces. In *ICML* (Sydney, Australia, 2002), Morgan Kaufmann, pp. 315–322. 2, 5, 14

[Law06] LAWLER G.: *Introduction to Stochastic Processes*. Chapman and Hall/CRC Probability Series. Chapman & Hall/CRC, 2006. 3, 4

[Lév06] LÉVY B.: Laplace-beltrami eigenfunctions towards an algorithm that "understands" geometry. In *SMI* (College Station, USA, 2006), IEEE, p. 13. 2

[LHMR08] LAI Y.-K., HU S.-M., MARTIN R. R., ROSIN P. L.: Fast mesh segmentation using random walks. In *SPM* (New York, USA, 2008), ACM, pp. 183–191. 2

[Lov93] LOVÁSZ L.: Random walks on graphs: A survey. *Combinatorics, Paul Erdos is Eighty 2*, 1 (1993), 1–46. 1

[LRF10] LIPMAN Y., RUSTAMOV R., FUNKHOUSER T.: Biharmonic distance. *ACM Transactions on Graphics 29*, 3 (June 2010). 2, 9, 10, 11

[MDSB02] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential-geometry operators for triangulated 2-manifolds. In *Proc. VisMath* (2002), pp. 35–57. 3

[New10] NEWMAN M.: *Networks: An Introduction*. Oxford University Press, Inc., New York, USA, 2010. 3, 14

[NLCK05] NADLER B., LAFON S., COIFMAN R. R., KEVREKIDIS I. G.: Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators. In *NIPS* (Vancouver, Canada, 2005), MIT Press, pp. 955–962. 2

[OMMG10] OVSJANIKOV M., MÉRIGOT Q., MÉMOLI F., GUIBAS L. J.: One point isometric matching with the heat kernel. *Comput. Graph. Forum 29*, 5 (2010), 1555–1564. 1

[PP93] PINKALL U., POLTHIER K.: Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics 2* (1993), 15–36. 4

[QH07] QIU H., HANCOCK E. R.: Clustering and embedding using commute times. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 11 (2007), 1873–1890. 2, 11

[Rus07] RUSTAMOV R. M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *SGP* (Barcelona,Spain, 2007), Eurographics, pp. 225–233. 2, 12

[Rus11a] RUSTAMOV R. M.: Average interpolating wavelets on point clouds and graphs. *CoRR abs/1110.2227* (2011). 2

[Rus11b] RUSTAMOV R. M.: Multiscale biharmonic kernels. *Comput. Graph. Forum 30*, 5 (2011), 1521–1531. 1

[SGGnL*11] SINATRA R., GÓMEZ-GARDEÑES J., LAMBIOTTE R., NICOSIA V., LATORA V.: Maximal-entropy random walks in complex networks with limited information. *Phys. Rev. E 83* (Mar 2011), 030103. 10

[She68] SHEPARD D.: A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 23rd ACM national conference* (New York, USA, 1968), ACM, pp. 517–524. 11

[SOCG10] SKRABA P., OVSJANIKOV M., CHAZAL F., GUIBAS L.: Persistence-based segmentation of deformable shapes. In *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment* (June 2010). 1

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L. J.: A concise and provably informative multi-scale signature based on heat diffusion. *Comput. Graph. Forum 28*, 5 (2009), 1383–1392. 1, 3

[SRML07] SUN X., ROSIN P. L., MARTIN R. R., LANGBEIN F. C.: Random walks for mesh denoising. In *SPM* (Beijing, China, 2007), ACM, pp. 11–22. 2

[WMKG07] WARDETZKY M., MATHUR S., KÄLBERER F., GRINSPUN E.: Discrete laplace operators: no free lunch. In *SGP* (Barcelona, Spain, 2007), Eurographics, pp. 33–37. 13

[XHW09] XIAO B., HANCOCK E. R., WILSON R. C.: Graph characteristics from the heat kernel trace. *Pattern Recognition 42*, 11 (2009), 2589–2606. 11

[ZBS11] ZHOU X., BELKIN M., SREBRO N.: An iterated graph laplacian approach for ranking on manifolds. In *KDD* (San Diego, USA, 2011), ACM, pp. 877–885. 2, 7, 8

[ZZC11] ZHANG J., ZHENG J., CAI J.: Interactive mesh cutting using constrained random walks. *IEEE Transactions on Visualization and Computer Graphics 17*, 3 (May 2011), 357–367. 2

APPENDIX

Pochhammer Time Operator: The rule of walk on the discrete mesh can be expressed as $P_{t+1} = M^T P_t$, hence iterating, $P_t = (M^T)^t P_0$ where P_t is the probability distribution starting from the initial distribution P_0. The spectral decomposition of the transition matrix is $M = D^{-1/2}\Phi\Lambda'\Phi^T D^{1/2}$ and hence, $M^t = D^{-1/2}\Phi\Lambda'^t\Phi^T D^{1/2}$. It can be written as

$$p^t_{uv} = \pi_v + \sum_{i\geq 2} \lambda_i'^t \Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{13}$$

where p^t_{uv} is the probability of starting at u and reaching v in t steps and is the uv^{th} entry of the matrix M^t. p^t_{uv} tends to π_v in the long time limit as $\Lambda' < 1$. The formulation equivalent to equation 1 for discrete walks is:

$$\hat{\tau}^{\alpha n} = \sum_{t=1}^{\infty} (x)_n [(\alpha M)^{t-1} - (\alpha M)^t] \tag{14}$$

where $\hat{\tau}^{\alpha n}$ denotes the operator defined over the function $(x)_n$, also called the Pochhammer symbol or rising factorial and represented as $(x)_n = x(x+1)(x+2)...(x+n-1)$. As M^t denotes the transition matrix after time t, $M^{t-1} - M^t$ denotes the change in transition probability after one time step and equivalent to rate of diffusion in the continuous time limit. More rigorously, the term in parenthesis in equation 14 can be written as $[e^{-t\Delta_{\alpha M}} - e^{-(t+dt)\Delta_{\alpha M}}]$, where we have used the fact that $(\alpha M)^t$ or $(I - \Delta_{\alpha M})^t$ converges to $e^{-t\Delta_{\alpha M}}$ for continuous time walks (see [KL02] for derivation). Taylor series expansion of the derived $I - e^{-\Delta_{\alpha M} dt}$ term in $e^{-t\Delta_{\alpha M}}(I - e^{-\Delta_{\alpha M} dt})$ results in the rate of diffusion. An alternative interpretation of the uv^{th} term in $M^{t-1} - M^t$ is that it denotes the probability that a random walker is at node v in the t^{th} time-step starting from node u at time $t = 0$ (see page 159 [New10] for case $(x)_1$). Parameters n in $(x)_n$ and α play a similar role of exaggerating/suppressing pathways (or '*faraway*' nodes), as in the continuous time case. Recall in equation 13, eigenvalue 1 and the corresponding eigenvector appear as π_v in both M^t as well as M^{t-1}. Substituting the spectral representation of M in the equation 14 for $\alpha = 1$, removes the eigenvalue 1 (of M) and the corresponding eigenvector from the summand, enabling us to use $I - M$ in an invertible setting (removal of eigenvalue 1 from M implies removal of eigenvalue 0 from $I - M$, hence killing its null space). Using the sum of infinite series $\sum_{k=1}^{\infty} k(k+1)...(k+n-1)[x^{k-1} - x^k] = n!/(1-x)^n$ and spectral decomposition of $(I - M)^{-n}$ or Δ_M^{-n}, the operator is represented as

$$\hat{\tau}^n_{uv} = n! \sum_{i\geq 2} \frac{1}{\lambda_i^n} \Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{15}$$

Following a similar derivation as in the continuous time case, the representation for $0 < \alpha < 1$ is

$$\hat{\tau}^{\alpha n}_{uv} = \frac{n!}{\alpha^n} \sum_{i\geq 1} \frac{1}{(\beta + \lambda_i)^n} \Phi_i(u)\Phi_i(v)\sqrt{\frac{d_v}{d_u}} \tag{16}$$

Thus, the use of Pochhammer symbol in lieu of t^n results in the same resultant kernel as in the continuous case.