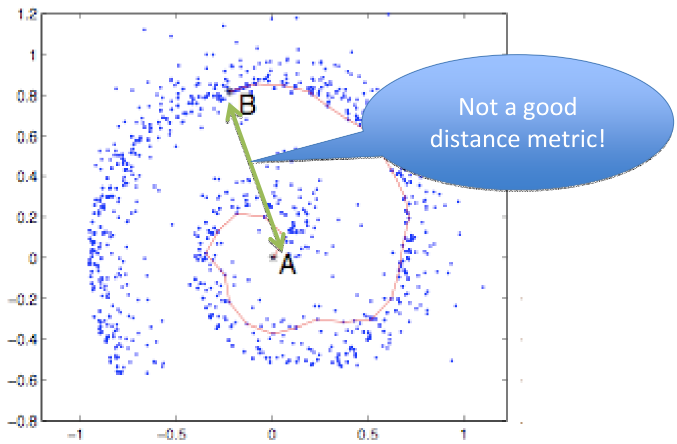
Diffusion Framework and Spectral Transform

Saeed Amizadeh

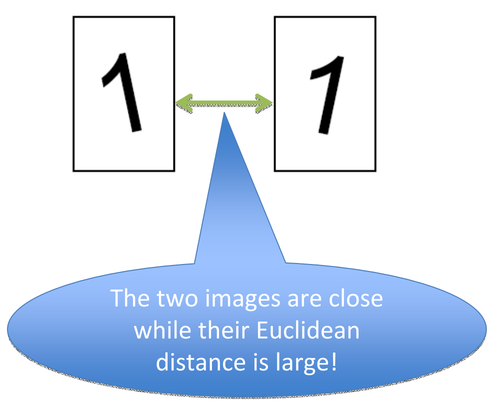
saeed@cs.pitt.edu

Contents

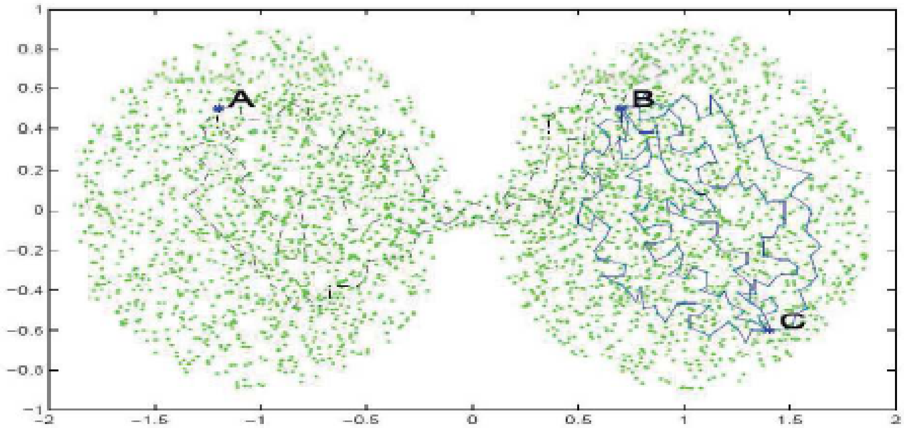
Underestimation of distance



Overestimation of distance

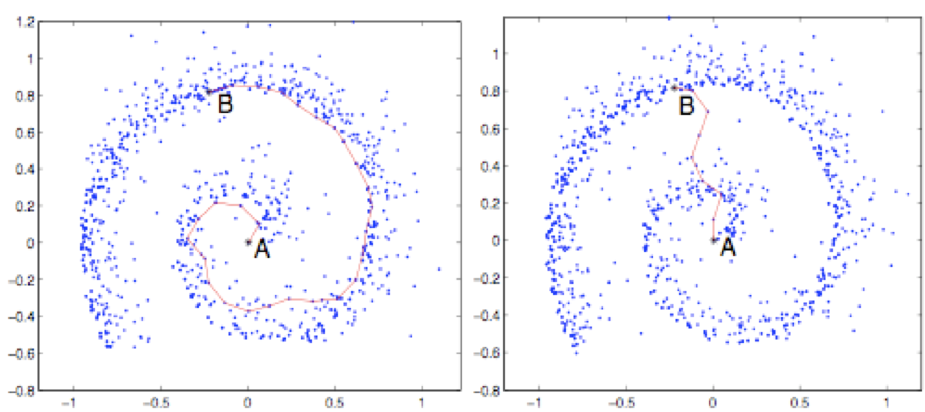


Why diffusion framework?



$d_{geod.}(A, B) \sim d_{geod.}(C, B)$, however $d^{(t)}(A, B) >> d^{(t)}(C, B)$.

Noise sensitivity



Contents

- Introduction
- Random Walk on Graphs
- Multi-scale Random Walk
- Diffusion Distance
- Diffusion Map Approximation
- Spectral Transform
- Continuous-time Diffusion
- Resistance Distance

Random Walk on Graphs

- Let *W* be the similarity matrix on the graph, the transition from node *x* to node *y* is defined as:

$$p_1(y \mid x) = \frac{w(x,y)}{d(x)}$$

$$d(x) = \sum_{z \in N(x)} w(x,z)$$

$$P = \left[p_{ij} = p_1(x_j \mid x_i)\right]_{n \times n}$$

Some properties of RW

- Asymptotic distribution

$$\pi^{\infty}(y) = \lim_{t \to \infty} p_t(y \mid x) = \frac{d(y)}{\sum_z d(z)}$$

- The absolute value of the eigenvalues of *P* are between 0 and 1 with largest one equal to 1.
- The eigenvectors: $P^T \phi_i = \lambda_i \phi_i \quad \text{and} \quad P\psi_i = \lambda_i \psi_i$

$$\psi_i(x) = \frac{\phi_i(x)}{\phi_0(x)}$$

Some properties of RW

- The eigenvectors of *P* are bi-orthogonal:

$$\phi_k^T \psi_\ell = \delta(k, \ell)$$

- Normalization:

$$\|\phi_\ell\|_{1/\phi_0}^2 = \sum_x \frac{\phi_\ell^2(x)}{\phi_0(x)} = 1, \quad \|\psi_\ell\|_{\phi_0}^2 = \sum_x \psi_\ell^2(x) \phi_0(x) = 1$$

Forward Diffusion Process

- If the probability vector $\pi^t(.)$ represents the distribution of random walker at time *t,* we have:

$$\pi^{t+1}(x) = \sum_y p_1(x \mid y)\pi^t(y)$$

$$\pi^{t+1} = P^T \pi^t$$

- Corollary:

$$\pi^\infty = P^T \pi^\infty \Rightarrow \pi^\infty = \phi_0$$

Backward Diffusion Process

- If $g^t(.)$ is a real-valued function defined on the graph at time *t*, $g^{t+1}(.)$ is the average of $g^t(.)$ at time *t+1*:

$$g^{t+1}(x) = \sum_y p_1(y \mid x) g^t(y)$$

$$g^{t+1} = P g^t$$

- Corollary: $g^\infty(.)$ is the smoothest function (i.e. the constant function)

$$g^\infty = P g^\infty \Rightarrow g^\infty = \psi_0 = 1$$

Link to Normalized Laplacian

- The transition matrix is closely related to the asymmetric normalized Laplacian:

$$d_{ii} = \sum_{y \in N(x)} w(x,y)$$

$$P = D^{-1}W,$$

$$L_{rw} = I - D^{-1}W = I - P$$

Contents

- Introduction
- Random Walk on Graphs
- Multi-scale Random Walk
- Diffusion Distance
- Diffusion Map Approximation
- Spectral Transform
- Continuous-time Diffusion
- Resistance Distance

T-step Diffusion Processes

- We can generalize the forward and backward processes:

$$\pi^{t_0+t} = P^T P^T \ldots P^T \pi^{t_0} = (P^{(t)})^T \pi^{t_0}$$

$$g^{t_0+t} = P.P.\ldots.P.g^{t_0} = P^{(t)} g^{t_0}$$

$$\Psi = [\psi_0 \mid \psi_1 \mid \ldots \mid \psi_{n-1}], \quad \Phi = [\phi_0 \mid \phi_1 \mid \ldots \mid \phi_{n-1}]$$

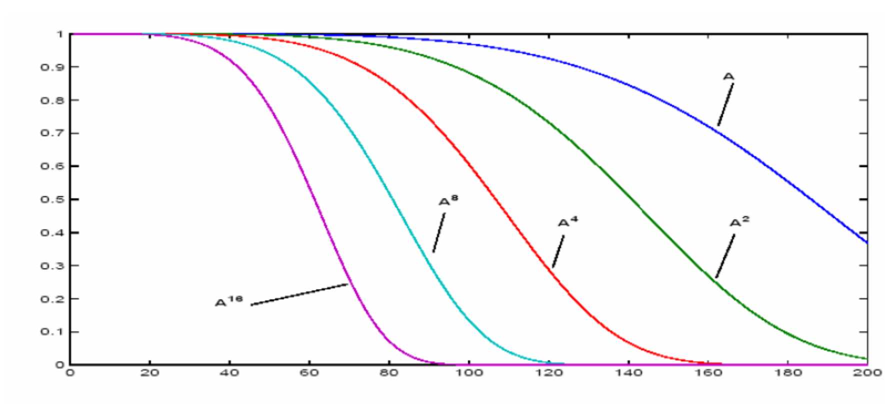
$$\Lambda = [\Lambda_{ii} = \lambda_i]_{n \times n}$$

$$P = \Psi \Lambda \Phi^T \Rightarrow P^{(t)} = \Psi \Lambda^{(t)} \Phi^T$$

$$P^{(t)} = \sum_x \lambda_x^t \phi_x^T . \psi_x$$

What does it do?

- The smaller eigenvalues decay:



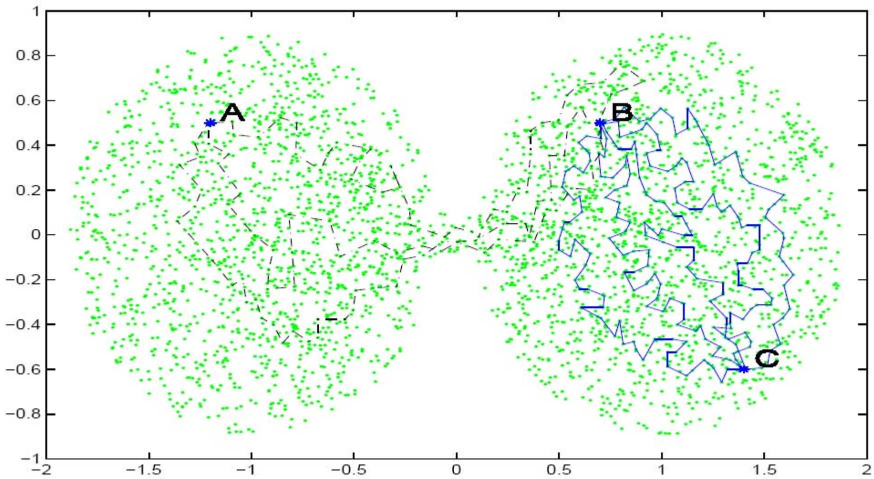
Contents

Diffusion Distance

- The diffusion distance between nodes *x* and *z* at scale *t* is defined as:

$$D_t^2(x,z) = \left\| p_t(.\,|\,x) - p_t(.\,|\,z) \right\|_{1/\phi_0}^2 = \sum_y \frac{\left(p_t(y\,|\,x) - p_t(y\,|\,z)\right)^2}{\phi_0(y)}$$

Interpretation



Contents

Diffusion Map

- The diffusion map of point x at scale with dimensionality q is defined as:

$$\Psi_f : x \rightarrow \begin{pmatrix} \lambda_1^t \psi_1(x) \\ \lambda_2^t \psi_2(x) \\ \\ \lambda_q^t \psi_q(x) \end{pmatrix}$$

Approximating Diffusion Dist.

- By replacing

$$p^{(t)}(x,y) = \sum_z \lambda_z^t \phi_z^T(x).\psi_z(y)$$

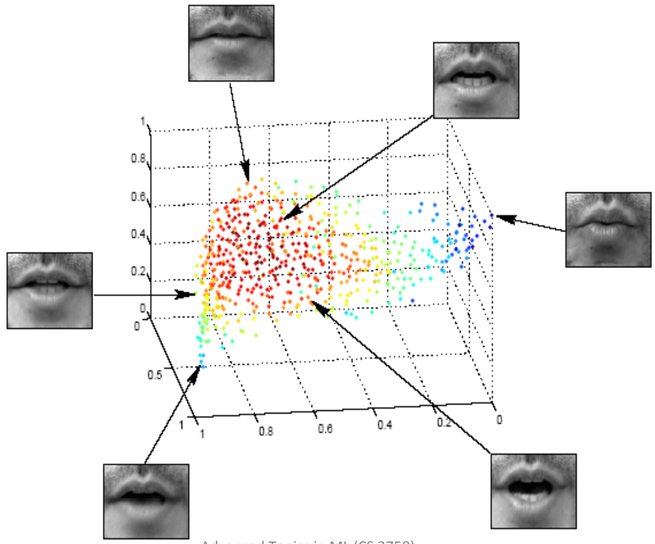
- in $D_t^2(x,z) = \left\| p_t(.|x) - p_t(.|z) \right\|_{1/\phi_0}^2 = \sum_y \frac{\left(p_t(y|x) - p_t(y|z)\right)^2}{\phi_0(y)}$

- we get: $D_t^2(x,z) = \sum_y \lambda_y^{2t} (\psi_y(x) - \psi_y(z))^2$

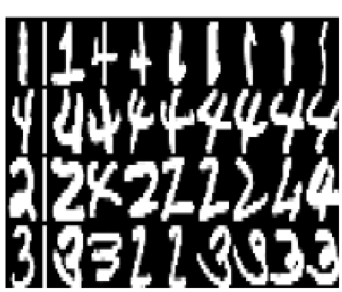
$$\approx \sum_{y=1}^{q} \lambda_y^{2t} (\psi_y(x) - \psi_y(z))^2 = \left\| \Psi_t(x) - \Psi_t(z) \right\|^2$$

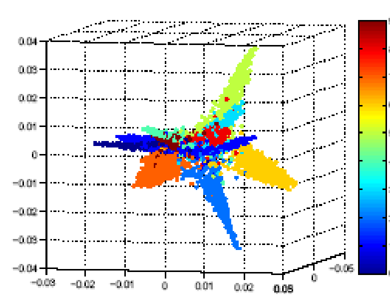
The Euclidean distance

Embedded Space for Image Data

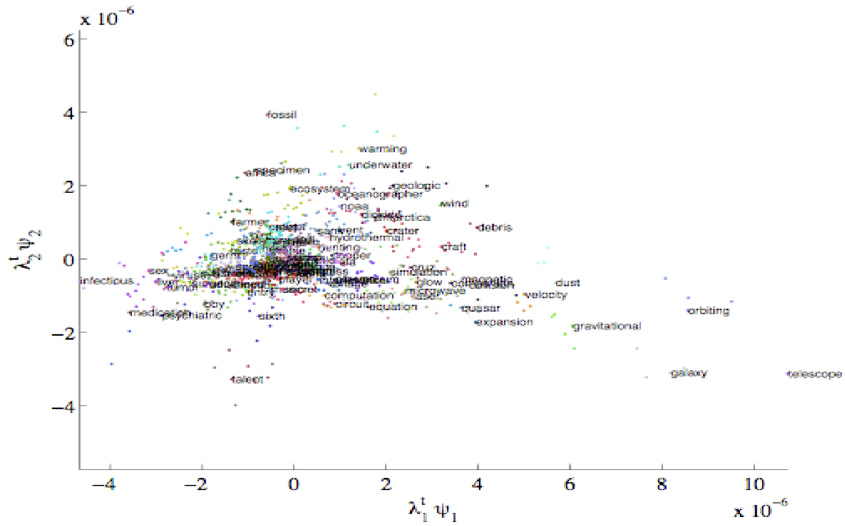


Handwritten Digits

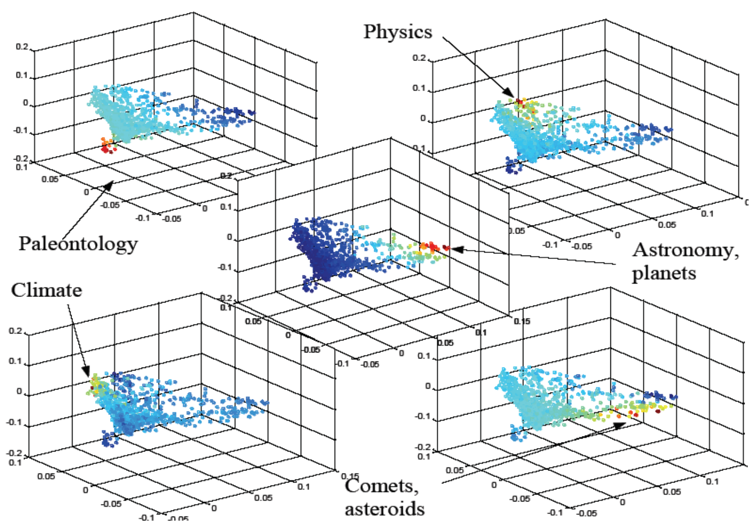




Words in 2D



Clusters = Topics



Contents

- Introduction
- Random Walk on Graphs
- Multi-scale Random Walk
- Diffusion Distance
- Diffusion Map Approximation
- Spectral Transform
- Continuous-time Diffusion
- Resistance Distance

T-step Random Walk Kernel

- The kernel in the embedded space is:

$$K_t(x,y) = \langle \Psi_t(x), \Psi_t(y) \rangle = \sum_z \lambda_z^{2t} \psi_z(x) \psi_z(y)$$

$$= \Psi \Lambda^{2t} \Psi^T = \Psi f(\Lambda) \Psi^T$$

- One can define more general kernel by letting *f* to be any increasing function

$$K_f(x,y) = \langle \Psi_f(x), \Psi_f(y) \rangle = \sum_z f(\lambda_z) \psi_z(x) \psi_z(y) = \Psi f(\Lambda) \Psi^T$$

The General Mapping

- The induced mapping is:

$$\Psi_f : x \to \begin{pmatrix} \sqrt{f(\lambda_1)}\psi_1(x) \\ \sqrt{f(\lambda_2)}\psi_2(x) \\ \\ \sqrt{f(\lambda_q)}\psi_q(x) \end{pmatrix}$$

- The corresponding distance is:

$$D_f^2(x,z) = (e_x - e_z)^T K_f (e_x - e_z) = \left\| \Psi_f(x) - \Psi_f(z) \right\|^2$$

Contents

- Introduction
- Random Walk on Graphs
- Multi-scale Random Walk
- Diffusion Distance
- Diffusion Map Approximation
- Spectral Transform
- Continuous-time Diffusion
- Resistance Distance

Diffusion Kernel

- The Diffusion (Heat) Kernel reads:

$$\frac{\partial}{\partial t}K_t(x,y) = H.K_t(x,y)$$
$$K_0(x,y) = \delta(x,y)$$

- Informally, this process diffuses the local similarity *H* to obtain the global similarity *K* after time *t*.

The Exponential Family

- Solving the differential equation, we get:

$$\frac{\partial}{\partial t}K_t(x,y) = H.K_t(x,y) \qquad K_t = e^{tH} = \lim_{n\to\infty}\left(I + \frac{tH}{n}\right)^n$$
$$K_0(x,y) = \delta(x,y)$$

- Example: for the continuous Laplacian operator *H=Δ*, *k* is the Gaussian kernel.

Continuous Diffusion on Discrete Graphs

- For discrete graphs, we have

$$H = -L$$

- Such that:

$$L = U(I-\Lambda)U^T \Rightarrow K_t = e^{tH} = e^{-tL} = Ue^{-t(I-\Lambda)}U^T$$

$$\mu_i = e^{-t(1-\lambda_i)} = f(\lambda_i)$$

This is a spectral transform!

How related to random walk?

- Continuous Diffusion Kernel is the limit of Lazy Random Walk:

$$p'_{ij} = t_0 \Delta t . p_{ij}, \quad p'_{ii} = 1 - \sum_{j \in N(i)} t_0 \Delta t . p_{ij} = 1 - t_0 \Delta t, \ \ N = 1/\Delta t$$

$$P'^N = \left((1 - t_0 \Delta t) I + t_0 \Delta t . P\right)^{1/\Delta t} = \left(I + t_0 \Delta t . (-L_{rw})\right)^{1/\Delta t}$$

$$\text{where } L_{rw} = I - P$$

$$\lim_{\Delta t \to 0} P'^N = \lim_{\Delta t \to 0} \left(I + \frac{t_0(-L_{rw})}{1/\Delta t} \right)^{1/\Delta t} = \exp(-t_0 L_{rw})$$

Contents

The Circuit Analogy

- Let G be a circuit with:

$$w_{ij} = c_{ij} = \frac{1}{r_{ij}}$$

similarity $\leftrightarrow$ conductance

distance $\leftrightarrow$ resistance

function on graph $\leftrightarrow$ node potential

The Kirchhoff's laws

- If y_{ij} is the current from node *i* to *j* and *Y* is the total current from source *a* to sink *b*, then

$$\sum_{j \in N(i)} y_{ij} = \begin{cases} Y & \text{if} \quad i = a \\ -Y & \text{if} \quad i = b \\ 0 & \text{otherwise} \end{cases}$$

- If C is a cycle in the circuit with ordered edges i➔j

$$\sum_{i \to j} y_{ij} r_{ij} = 0 = \sum_{i \to j} (v_i - v_j)$$

The Effective Resistance

- The effective resistance between *a* and *b* with the total current *Y* is defined as:

$$R_{ab} = \frac{v_a - v_b}{Y}$$

- Theorem: $R_{ab} = (e_a - e_b)^T L^+ (e_a - e_b)$

$$\text{if } L = U \Lambda U^T \Rightarrow L^+ = U f(\Lambda) U^T$$

$$\text{where } f(x) = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Resistance as Distance

- Comparing

$$R_{ab} = (e_a - e_b)^T L^+ (e_a - e_b)$$
$$D_f^2(a,b) = (e_a - e_b)^T K_f (e_a - e_b)$$

- The spectral transform is:

$$f(x) = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Some properties

- *R* is a distance metric
- *R* is a non-increasing function of edge weights
- *R* is a lower bound on the geodesic distance

$$R_{ab} \leq d_{ab}$$

Relation to Random Walk

- Let T_{ab} denote the number of transitions (i.e. discrete time) that take random walker from *a* to *b.* The *average commute time* between *a* and *b* is defined as:

$$C_{ab} = E(T_{ab}) + E(T_{ba})$$

- Theorem: $$R_{ab} = \frac{C_{ab}}{\sum_i d_i}$$

Thank You!