MAT 280: Harmonic Analysis on Graphs & Networks
Lecture 13: Distances on Graphs III: From Commute-Time Distance to Diffusion Distance

Naoki Saito

Department of Mathematics
University of California, Davis

November 7, 2019

Outline

Outline

1. Cautions on Resistance/Commute-Time Distances
2. Diffusion Distances

Problems of Resistance/Commute-Time Distances

- This part is based on the work by U. von Luxburg, A. Radl, and M. Hein (JMLR, 2014) who cautioned the use of commute-time distances for *large* graphs.
- Recall the *average first passage time* (a.k.a. *expected hitting time*) $m(j\,|\,i)$, i.e., the average number of steps that a random walker, starting at v_i, will take to reach v_j for the first time.
- The *commute-time distance* $c(i,j)$ between v_i and v_j was defined as $c(i,j) = m(j\,|\,i) + m(i\,|\,j)$. Recall also that the *resistance distance* $r(i,j) = c(i,j)/\operatorname{vol}(G)$.
- Von Luxburg et al. proved that under mild assumptions, $\forall i \neq j$, $m(j\,|\,i) \to \operatorname{vol}(G)/d_j$, $r(i,j) = c(i,j)/\operatorname{vol}(G) \to 1/d_i + 1/d_j$, as $n \to \infty$, i.e., these do *not* reflect *connectivity* of the graph, just simply reflect the local degree information only.

Problems of Resistance/Commute-Time Distances

- This part is based on the work by U. von Luxburg, A. Radl, and M. Hein (JMLR, 2014) who cautioned the use of commute-time distances for *large* graphs.
- Recall the *average first passage time* (a.k.a. *expected hitting time*) $m(j\,|\,i)$, i.e., the average number of steps that a random walker, starting at v_i, will take to reach v_j for the first time.
- The *commute-time distance* $c(i,j)$ between v_i and v_j was defined as $c(i,j) = m(j\,|\,i) + m(i\,|\,j)$. Recall also that the *resistance distance* $r(i,j) = c(i,j)/\operatorname{vol}(G)$.
- Von Luxburg et al. proved that under mild assumptions, $\forall i \neq j$, $m(j\,|\,i) \to \operatorname{vol}(G)/d_j$, $r(i,j) = c(i,j)/\operatorname{vol}(G) \to 1/d_i + 1/d_j$, as $n \to \infty$, i.e., these do *not* reflect *connectivity* of the graph, just simply reflect the local degree information only.

Problems of Resistance/Commute-Time Distances

- This part is based on the work by U. von Luxburg, A. Radl, and M. Hein (JMLR, 2014) who cautioned the use of commute-time distances for *large* graphs.
- Recall the *average first passage time* (a.k.a. *expected hitting time*) $m(j\,|\,i)$, i.e., the average number of steps that a random walker, starting at v_i, will take to reach v_j for the first time.
- The *commute-time distance* $c(i,j)$ between v_i and v_j was defined as $c(i,j) = m(j\,|\,i) + m(i\,|\,j)$. Recall also that the *resistance distance* $r(i,j) = c(i,j)/\operatorname{vol}(G)$.
- Von Luxburg et al. proved that under mild assumptions, $\forall i \neq j$, $m(j\,|\,i) \to \operatorname{vol}(G)/d_j$, $r(i,j) = c(i,j)/\operatorname{vol}(G) \to 1/d_i + 1/d_j$, as $n \to \infty$, i.e., these do *not* reflect *connectivity* of the graph, just simply reflect the local degree information only.

Problems of Resistance/Commute-Time Distances

- This part is based on the work by U. von Luxburg, A. Radl, and M. Hein (JMLR, 2014) who cautioned the use of commute-time distances for *large* graphs.
- Recall the *average first passage time* (a.k.a. *expected hitting time*) $m(j \mid i)$, i.e., the average number of steps that a random walker, starting at v_i, will take to reach v_j for the first time.
- The *commute-time distance* $c(i, j)$ between v_i and v_j was defined as $c(i, j) = m(j \mid i) + m(i \mid j)$. Recall also that the *resistance distance* $r(i, j) = c(i, j)/\mathrm{vol}(G)$.
- Von Luxburg et al. proved that under mild assumptions, $\forall i \neq j$, $m(j \mid i) \to \mathrm{vol}(G)/d_j$, $r(i, j) = c(i, j)/\mathrm{vol}(G) \to 1/d_i + 1/d_j$, as $n \to \infty$, i.e., these do *not* reflect *connectivity* of the graph, just simply reflect the local degree information only.

Problems of Resistance/Commute-Time Distances: Electrical Network Intuition

- The effective resistance r_{12} between two vertices connected by two resistors r_1 and r_2 in *series* is $r_{12} = r_1 + r_2$ while that connected by two resistors in *parallel* is $1/r_{12} = 1/r_1 + 1/r_2$.

- Hence, the overall effective resistance between i and j is dominated by the edges adjacent to i and j with contribution $1/d_i + 1/d_j$.

Problems of Resistance/Commute-Time Distances: Electrical Network Intuition

- The effective resistance r_{12} between two vertices connected by two resistors r_1 and r_2 in *series* is $r_{12} = r_1 + r_2$ while that connected by two resistors in *parallel* is $1/r_{12} = 1/r_1 + 1/r_2$.

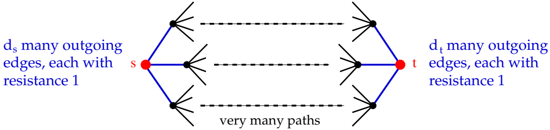

Figure 1: *Electrical network intuition: The effective resistance between s and t is dominated by the edges adjacent to s and t.*

- Hence, the overall effective resistance between i and j is dominated by the edges adjacent to i and j with contribution $1/d_i + 1/d_j$.

Problems of Resistance/Commute-Time Distances: Electrical Network Intuition

- The effective resistance r_{12} between two vertices connected by two resistors r_1 and r_2 in *series* is $r_{12} = r_1 + r_2$ while that connected by two resistors in *parallel* is $1/r_{12} = 1/r_1 + 1/r_2$.

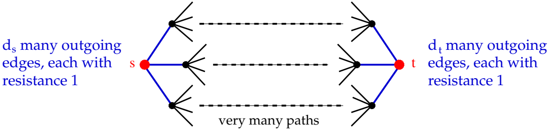

Figure 1: *Electrical network intuition: The effective resistance between s and t is dominated by the edges adjacent to s and t.*

- Hence, the overall effective resistance between i and j is dominated by the edges adjacent to i and j with contribution $1/d_i + 1/d_j$.

Problems of Resistance/Commute-Time Distances: Random Walk Intuition

- Regardless of which vertex i the random walk starts from, the time to hit vertex j just depends on d_j if G gets large. By the time the random walk is close to j, it *has forgotten where it came from:*

- How fast the random walk hits j is inversely proportional to the density of G close to j, i.e., $\approx 1/d_j$.

Problems of Resistance/Commute-Time Distances: Random Walk Intuition

- Regardless of which vertex i the random walk starts from, the time to hit vertex j just depends on d_j if G gets large. By the time the random walk is close to j, it *has forgotten where it came from:*

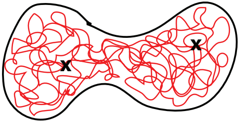

Figure 2: *Random walk intuition: Between its start and target vertex (black crosses), the random walk wanders around so long that by the time it finally arrives at its target it has already "forgotten" where it started from.*

- How fast the random walk hits j is inversely proportional to the density of G close to j, i.e., $\approx 1/d_j$.

Problems of Resistance/Commute-Time Distances: Random Walk Intuition

- Regardless of which vertex i the random walk starts from, the time to hit vertex j just depends on d_j if G gets large. By the time the random walk is close to j, it *has forgotten where it came from:*

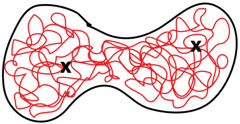

Figure 2: *Random walk intuition: Between its start and target vertex (black crosses), the random walk wanders around so long that by the time it finally arrives at its target it has already "forgotten" where it started from.*

- How fast the random walk hits j is inversely proportional to the density of G close to j, i.e., $\approx 1/d_j$.

Problems of Resistance/Commute-Time Distances: Setup

- Recall the properties of the transition matrix $P := D^{-1}A$, and the symmetrically-normalized graph Laplacian matrix $L_{\mathrm{sym}} := D^{-1/2}LD^{-1/2} = I_n - D^{-1/2}AD^{-1/2}$:
  - λ is an eigenvalue of L_{sym} iff $1-\lambda$ is an eigenvalue of P.
  - Let $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$ be the eigenvalues of P, i.e., $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$.
  - The *spectral gap* of P is defined as $1 - \max\{\mu_1, |\mu_{n-1}|\}$.
  - Using the definition of L_{sym}, $m(j\,|\,i)$, r_{ij}, and c_{ij} can be written as:

$$m(j\,|\,i) = \mathrm{vol}(G)\left\langle \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\mathrm{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_j}}\boldsymbol{e}_j - \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i\right)\right\rangle$$

$$r_{ij} = \frac{1}{\mathrm{vol}(G)}c_{ij} = \left\langle \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\mathrm{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j\right)\right\rangle.$$

Problems of Resistance/Commute-Time Distances: Setup

- Recall the properties of the transition matrix $P := D^{-1}A$, and the symmetrically-normalized graph Laplacian matrix $L_{\text{sym}} := D^{-1/2}LD^{-1/2} = I_n - D^{-1/2}AD^{-1/2}$:
- λ is an eigenvalue of L_{sym} iff $1-\lambda$ is an eigenvalue of P.
- Let $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$ be the eigenvalues of P, i.e., $\lambda_j^{\text{sym}} = 1 - \mu_j$.
- The *spectral gap* of P is defined as $1 - \max\{\mu_1, |\mu_{n-1}|\}$.
- Using the definition of L_{sym}, $m(j\,|\,i)$, r_{ij}, and c_{ij} can be written as:

$$m(j\,|\,i) = \mathrm{vol}(G)\left\langle \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_j}}\boldsymbol{e}_j - \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i\right)\right\rangle$$

$$r_{ij} = \frac{1}{\mathrm{vol}(G)}c_{ij} = \left\langle \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j\right)\right\rangle.$$

Problems of Resistance/Commute-Time Distances: Setup

- Recall the properties of the transition matrix $P := D^{-1}A$, and the symmetrically-normalized graph Laplacian matrix $L_{\text{sym}} := D^{-1/2}LD^{-1/2} = I_n - D^{-1/2}AD^{-1/2}$:
- λ is an eigenvalue of L_{sym} iff $1-\lambda$ is an eigenvalue of P.
- Let $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$ be the eigenvalues of P, i.e., $\lambda_j^{\text{sym}} = 1 - \mu_j$.
- The *spectral gap* of P is defined as $1 - \max\{\mu_1, |\mu_{n-1}|\}$.
- Using the definition of L_{sym}, $m(j\,|\,i)$, r_{ij}, and c_{ij} can be written as:

$$m(j\,|\,i) = \text{vol}(G)\left\langle \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_j}}\boldsymbol{e}_j - \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i\right)\right\rangle$$

$$r_{ij} = \frac{1}{\text{vol}(G)}c_{ij} = \left\langle \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j\right)\right\rangle.$$

Problems of Resistance/Commute-Time Distances: Setup

- Recall the properties of the transition matrix $P := D^{-1}A$, and the symmetrically-normalized graph Laplacian matrix $L_{\text{sym}} := D^{-1/2}LD^{-1/2} = I_n - D^{-1/2}AD^{-1/2}$:
- λ is an eigenvalue of L_{sym} iff $1-\lambda$ is an eigenvalue of P.
- Let $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$ be the eigenvalues of P, i.e., $\lambda_j^{\text{sym}} = 1 - \mu_j$.
- The *spectral gap* of P is defined as $1 - \max\{\mu_1, |\mu_{n-1}|\}$.
- Using the definition of L_{sym}, $m(j\,|\,i)$, r_{ij}, and c_{ij} can be written as:

$$m(j\,|\,i) = \text{vol}(G)\left\langle \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_j}}\boldsymbol{e}_j - \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i\right)\right\rangle$$

$$r_{ij} = \frac{1}{\text{vol}(G)}c_{ij} = \left\langle \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\text{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j\right)\right\rangle.$$

Problems of Resistance/Commute-Time Distances: Setup

- Recall the properties of the transition matrix $P := D^{-1}A$, and the symmetrically-normalized graph Laplacian matrix $L_{\mathrm{sym}} := D^{-1/2}LD^{-1/2} = I_n - D^{-1/2}AD^{-1/2}$:
- λ is an eigenvalue of L_{sym} iff $1-\lambda$ is an eigenvalue of P.
- Let $1 = \mu_0 \ge \mu_1 \ge \cdots \ge \mu_{n-1} > -1$ be the eigenvalues of P, i.e., $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$.
- The *spectral gap* of P is defined as $1 - \max\{\mu_1, |\mu_{n-1}|\}$.
- Using the definition of L_{sym}, $m(j\,|\,i)$, r_{ij}, and c_{ij} can be written as:

$$m(j\,|\,i) = \mathrm{vol}(G)\left\langle \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\mathrm{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_j}}\boldsymbol{e}_j - \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i\right)\right\rangle$$

$$r_{ij} = \frac{1}{\mathrm{vol}(G)}c_{ij} = \left\langle \frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j, L_{\mathrm{sym}}^{\dagger}\left(\frac{1}{\sqrt{d_i}}\boldsymbol{e}_i - \frac{1}{\sqrt{d_j}}\boldsymbol{e}_j\right)\right\rangle.$$

Problems of Resistance/Commute-Time Distances: Setup

- Von Luxberg et al. deal with the so-called *random geometric graphs*.
- '*Random*' implies that the underlying data vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ are n i.i.d. realizations of a stochastic process with some pdf on $\mathbb{R}^d$.
- '*Geometric*' implies that graphs under consideration are either: 1) k-NN graphs with $\|\cdot\|_2$; 2) ε-neighborhood graphs with $\|\cdot\|_2$; or 3) complete graphs with the weights $\exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / h^2\right)$.

Definition (Valid Region)

Let p be any pdf on $\mathbb{R}^d$. We call a connected subset $\mathscr{X} \subset \mathbb{R}^d$ a *valid region* if the following properties are satisfied:

(i) p on $\mathscr{X}$ is bounded away from 0: $\forall \boldsymbol{x} \in \mathscr{X}$, $\exists p_{\min}$, s.t., $p(\boldsymbol{x}) \geq p_{\min} > 0$.

(ii) $\mathscr{X}$ has "bottleneck" larger than some value $h > 0$; the set $\{\boldsymbol{x} \in \mathscr{X} \mid \mathrm{dist}(\boldsymbol{x}, \partial\mathscr{X}) > h/2\}$ is connected.

(iii) $\partial\mathscr{X}$ is regular: $\exists \alpha > 0, \varepsilon_0 > 0$ s.t. if $\varepsilon < \varepsilon_0$, then $\forall \boldsymbol{x} \in \partial\mathscr{X}$, $\mathrm{vol}(B_\varepsilon(\boldsymbol{x}) \cap \mathscr{X}) \geq \alpha\, \mathrm{vol}(B_\varepsilon(\boldsymbol{x}))$. In other words, $\partial\mathscr{X}$ cannot contain arbitrarily thin spikes.

Problems of Resistance/Commute-Time Distances: Setup

- Von Luxberg et al. deal with the so-called *random geometric graphs*.
- '*Random*' implies that the underlying data vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ are n i.i.d. realizations of a stochastic process with some pdf on $\mathbb{R}^d$.
- '*Geometric*' implies that graphs under consideration are either: 1) k-NN graphs with $\|\cdot\|_2$; 2) ε-neighborhood graphs with $\|\cdot\|_2$; or 3) complete graphs with the weights $\exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / h^2\right)$.

Definition (Valid Region)

Let p be any pdf on $\mathbb{R}^d$. We call a connected subset $\mathscr{X} \subset \mathbb{R}^d$ a *valid region* if the following properties are satisfied:

(i) p on $\mathscr{X}$ is bounded away from 0: $\forall \boldsymbol{x} \in \mathscr{X}$, $\exists p_{\min}$, s.t., $p(\boldsymbol{x}) \geq p_{\min} > 0$.

(ii) $\mathscr{X}$ has "bottleneck" larger than some value $h > 0$; the set $\{\boldsymbol{x} \in \mathscr{X} \mid \operatorname{dist}(\boldsymbol{x}, \partial\mathscr{X}) > h/2\}$ is connected.

(iii) $\partial\mathscr{X}$ is regular: $\exists \alpha > 0, \varepsilon_0 > 0$ s.t. if $\varepsilon < \varepsilon_0$, then $\forall \boldsymbol{x} \in \partial\mathscr{X}$, $\operatorname{vol}(B_\varepsilon(\boldsymbol{x}) \cap \mathscr{X}) \geq \alpha \operatorname{vol}(B_\varepsilon(\boldsymbol{x}))$. In other words, $\partial\mathscr{X}$ cannot contain arbitrarily thin spikes.

Problems of Resistance/Commute-Time Distances: Setup

- Von Luxberg et al. deal with the so-called *random geometric graphs*.
- '*Random*' implies that the underlying data vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ are n i.i.d. realizations of a stochastic process with some pdf on $\mathbb{R}^d$.
- '*Geometric*' implies that graphs under consideration are either: 1) k-NN graphs with $\|\cdot\|_2$; 2) ε-neighborhood graphs with $\|\cdot\|_2$; or 3) complete graphs with the weights $\exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / h^2\right)$.

Definition (Valid Region)

Let p be any pdf on $\mathbb{R}^d$. We call a connected subset $\mathscr{X} \subset \mathbb{R}^d$ a *valid region* if the following properties are satisfied:

(i) p on $\mathscr{X}$ is bounded away from 0: $\forall \boldsymbol{x} \in \mathscr{X}$, $\exists p_{\min}$, s.t., $p(\boldsymbol{x}) \geq p_{\min} > 0$.

(ii) $\mathscr{X}$ has "bottleneck" larger than some value $h > 0$; the set $\{\boldsymbol{x} \in \mathscr{X} \mid \operatorname{dist}(\boldsymbol{x}, \partial\mathscr{X}) > h/2\}$ is connected.

(iii) $\partial\mathscr{X}$ is regular: $\exists \alpha > 0, \varepsilon_0 > 0$ s.t. if $\varepsilon < \varepsilon_0$, then $\forall \boldsymbol{x} \in \partial\mathscr{X}$, $\operatorname{vol}(B_\varepsilon(\boldsymbol{x}) \cap \mathscr{X}) \geq \alpha \operatorname{vol}(B_\varepsilon(\boldsymbol{x}))$. In other words, $\partial\mathscr{X}$ cannot contain arbitrarily thin spikes.

Problems of Resistance/Commute-Time Distances: Setup

- Von Luxberg et al. deal with the so-called *random geometric graphs*.
- '*Random*' implies that the underlying data vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ are n i.i.d. realizations of a stochastic process with some pdf on $\mathbb{R}^d$.
- '*Geometric*' implies that graphs under consideration are either: 1) k-NN graphs with $\|\cdot\|_2$; 2) ε-neighborhood graphs with $\|\cdot\|_2$; or 3) complete graphs with the weights $\exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / h^2\right)$.

Definition (Valid Region)

Let p be any pdf on $\mathbb{R}^d$. We call a connected subset $\mathscr{X} \subset \mathbb{R}^d$ a *valid region* if the following properties are satisfied:

(i) p on $\mathscr{X}$ is bounded away from 0: $\forall \boldsymbol{x} \in \mathscr{X}$, $\exists p_{\min}$, s.t., $p(\boldsymbol{x}) \geq p_{\min} > 0$.

(ii) $\mathscr{X}$ has "bottleneck" larger than some value $h > 0$; the set $\{\boldsymbol{x} \in \mathscr{X} \mid \operatorname{dist}(\boldsymbol{x}, \partial\mathscr{X}) > h/2\}$ is connected.

(iii) $\partial\mathscr{X}$ is regular: $\exists \alpha > 0, \varepsilon_0 > 0$ s.t. if $\varepsilon < \varepsilon_0$, then $\forall \boldsymbol{x} \in \partial\mathscr{X}$, $\operatorname{vol}(B_\varepsilon(\boldsymbol{x}) \cap \mathscr{X}) \geq \alpha \operatorname{vol}(B_\varepsilon(\boldsymbol{x}))$. In other words, $\partial\mathscr{X}$ cannot contain arbitrarily thin spikes.

General Limitations of Theorems of Von Luxburg et al.

- Their approximation results only hold if the graph is "reasonably strongly" connected and does not have too extreme bottlenecks. In other words, no single edge dominates the commute time behavior.
- Their results only hold if $d_{\min}(G)$ (the minimum degree of G) is "reasonably large" compared to $n = |V(G)|$, e.g., $d_{\min}(G) \approx \log n$. In other words, no single vertex dominates the commute time behavior.
- Hence, their results do *not* hold for power-law graphs/scale-free networks where $d_{\min}(G)$ is constant.

General Limitations of Theorems of Von Luxburg et al.

- Their approximation results only hold if the graph is "reasonably strongly" connected and does not have too extreme bottlenecks. In other words, no single edge dominates the commute time behavior.
- Their results only hold if $d_{\min}(G)$ (the minimum degree of G) is "reasonably large" compared to $n = |V(G)|$, e.g., $d_{\min}(G) \approx \log n$. In other words, no single vertex dominates the commute time behavior.
- Hence, their results do *not* hold for power-law graphs/scale-free networks where $d_{\min}(G)$ is constant.

General Limitations of Theorems of Von Luxburg et al.

- Their approximation results only hold if the graph is "reasonably strongly" connected and does not have too extreme bottlenecks. In other words, no single edge dominates the commute time behavior.
- Their results only hold if $d_{\min}(G)$ (the minimum degree of G) is "reasonably large" compared to $n = |V(G)|$, e.g., $d_{\min}(G) \approx \log n$. In other words, no single vertex dominates the commute time behavior.
- Hence, their results do *not* hold for power-law graphs/scale-free networks where $d_{\min}(G)$ is constant.

Main Results of Von Luxburg, Radl, & Hein

Theorem (Commute-distance on unweighted ε-neighborhood graphs)

Let $\mathscr{X}$ be a valid region with bottleneck h and minimal density value $p_{\min}$. For $\varepsilon \le h$, consider an unweighted ε-neighborhood graph built from $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ that have been drawn i.i.d. from the pdf p. Fix i and j. Assume that $\operatorname{dist}(\boldsymbol{x}_\ell, \partial\mathscr{X}) \ge h$, $\ell = i, j$ and that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \ge 8\varepsilon$. Then, there exist constants $c_\ell > 0$, $\ell = 1, \dots, 7$ (depending on the dimension and geometry of $\mathscr{X}$) such that with probability at least $1 - c_1 n \exp\left(-c_2 n\varepsilon^d\right) - c_3 \exp\left(-c_4 n \varepsilon^d\right) / \varepsilon^d$ the commute-time distance on the ε-neighborhood graph satisfies:

$$\left| \frac{n\varepsilon^d}{\operatorname{vol}(G)} c_{ij} - \left(\frac{n\varepsilon^d}{d_i} + \frac{n\varepsilon^d}{d_j} \right) \right| \le \begin{cases} c_5 / n\varepsilon^d & \text{if } d > 3; \\ c_6 \log(1/\varepsilon) / n\varepsilon^3 & \text{if } d = 3; \\ c_7 / n\varepsilon^3 & \text{if } d = 2. \end{cases}$$

The probability converges to 1 if $n \to \infty$ and $n\varepsilon^d / \log(n) \to \infty$. Under these conditions, if p is continuous and if $\varepsilon \to 0$, then

$$\frac{n\varepsilon^d}{\operatorname{vol}(G)} c_{ij} \xrightarrow{\text{a.s.}} \frac{1}{\eta_d\, p(\boldsymbol{x}_i)} + \frac{1}{\eta_d\, p(\boldsymbol{x}_j)}, \text{ where } \eta_d := \operatorname{vol}(B_1(\mathbf{0})) = \frac{2\pi^{d/2}}{d\,\Gamma(d/2)}.$$

Main Results of Von Luxburg, Radl, & Hein . . .

Theorem (Commute-distance on unweighted k-NN graphs)

Let $\mathscr{X}$ be a valid region with bottleneck h and density bounds $p_{\min}$ and $p_{\max}$. Consider an unweighted k-NN graph (either symmetric or mutual) such that $(k/n)^{1/d}/2p_{\max} \le h$, built from $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ that have been drawn i.i.d. from the pdf p. Fix i and j. Assume that $\operatorname{dist}(\boldsymbol{x}_\ell, \partial\mathscr{X}) \ge h$, $\ell = i, j$ and that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \ge 4(k/n)^{1/d}/p_{\max}$. Then, there exist constants $c_\ell > 0$, $\ell = 1, \ldots, 6$ such that with probability at least $1 - c_1 n \exp(-c_2 k)$ the commute-time distance on the k-NN graph satisfies:

$$\left| \frac{k}{\operatorname{vol}(G)} c_{ij} - \left(\frac{k}{d_i} + \frac{k}{d_j} \right) \right| \le \begin{cases} c_4/k & \text{if } d > 3; \\ c_5 \log(n/k)/k & \text{if } d = 3; \\ c_6 n^{1/2}/k^{3/2} & \text{if } d = 2. \end{cases}$$

The probability converges to 1 if $n \to \infty$ and $k/\log(n) \to \infty$. Under these conditions, if p is continuous and if $k/n \to 0$, then

$$\frac{k}{\operatorname{vol}(G)} c_{ij} \xrightarrow{\text{a.s.}} 2.$$

Main Results of Von Luxburg, Radl, & Hein ...

Theorem (Commute-distance on fully connected weighted graphs)

Consider a fixed, fully connected weighted graph with weight matrix A (not necessarily Gaussian weights). Assume that $0 < a_{\min} \le a_{ij} \le a_{\max}$ for all i, j. Then, uniformly for all $i, j \in N = \{1, \dots, n\}$, $i \neq j$,

$$\left| \frac{n}{\mathrm{vol}(G)} m(j \mid i) - \frac{n}{d_j} \right| \le 4n \frac{a_{\max}}{a_{\min}} \frac{a_{\max}}{d_{\min}^2} \le 4 \frac{a_{\max}^2}{a_{\min}^3} \frac{1}{n}.$$

Theorem (Gaussian graphs with adapted bandwidth)

Let $\mathscr{X} \subset \mathbb{R}^d$ be a compact set and p a continuous, strictly positive pdf on $\mathscr{X}$. Consider a fully connected, weighted similarity graph built from $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ drawn i.i.d. from p. Let the weight function be $k_h(\boldsymbol{x}_i, \boldsymbol{x}_j) := \frac{1}{(2\pi h^2)^{d/2}} \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / 2h^2)$. If $n \to \infty$, $h \to 0$, and $nh^{d+2}/\log(n) \to \infty$, then

$$\frac{n}{\mathrm{vol}(G)} c_{ij} \xrightarrow{a.s.} \frac{1}{p(\boldsymbol{x}_i)} + \frac{1}{p(\boldsymbol{x}_j)}.$$

Main Results of Von Luxburg, Radl, & Hein . . .

Theorem (Commute-distance on fully connected weighted graphs)

Consider a fixed, fully connected weighted graph with weight matrix A *(not necessarily Gaussian weights). Assume that* $0 < a_{\min} \le a_{ij} \le a_{\max}$ *for all* i, j*. Then, uniformly for all* $i, j \in N = \{1, \ldots, n\}$, $i \neq j$,

$$\left| \frac{n}{\mathrm{vol}(G)} m(j \mid i) - \frac{n}{d_j} \right| \le 4n \frac{a_{\max}}{a_{\min}} \frac{a_{\max}}{d_{\min}^2} \le 4 \frac{a_{\max}^2}{a_{\min}^3} \frac{1}{n}.$$

Theorem (Gaussian graphs with adapted bandwidth)

Let $\mathscr{X} \subset \mathbb{R}^d$ *be a compact set and* p *a continuous, strictly positive pdf on* $\mathscr{X}$*. Consider a fully connected, weighted similarity graph built from* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ *drawn i.i.d. from* p*. Let the weight function be* $k_h(\boldsymbol{x}_i, \boldsymbol{x}_j) := \frac{1}{(2\pi h^2)^{d/2}} \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / 2h^2)$*. If* $n \to \infty$, $h \to 0$*, and* $nh^{d+2} / \log(n) \to \infty$*, then*

$$\frac{n}{\mathrm{vol}(G)} c_{ij} \xrightarrow{a.s.} \frac{1}{p(\boldsymbol{x}_i)} + \frac{1}{p(\boldsymbol{x}_j)}.$$

Outline

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \mathrm{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^\delta : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^\delta(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \leq \delta$.

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \operatorname{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j{}^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j{}^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^\delta : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^\delta(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \leq \delta$.

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \operatorname{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j{}^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j{}^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \ge \mu_1 \ge \cdots \ge \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^\delta : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^\delta(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \le \delta$.

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \mathrm{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^{\delta} : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^{\delta}(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \leq \delta$.

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \mathrm{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^\delta : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^\delta(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \leq \delta$.

Diffusion Maps

- Consider the transition matrix $P = D^{-1}A$ of a weighted graph G.
- Due to the nonsymmetry of P, it has the left and right eigenvectors, i.e., $P = \Phi M \Psi^{\mathsf{T}}$ where $M := \mathrm{diag}(\mu_0, \ldots, \mu_{n-1})$, $\boldsymbol{\psi}_j{}^{\mathsf{T}} P = \mu_j \boldsymbol{\psi}_j{}^{\mathsf{T}}$, $P\boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j$, $j = 0, \ldots, n-1$. Note $\Phi^{\mathsf{T}}\Psi = \Psi^{\mathsf{T}}\Phi = I_n$, i.e., $\{\boldsymbol{\phi}_j\}$ and $\{\boldsymbol{\psi}_j\}$ are *biorthogonal bases* of $\mathbb{R}^n$.
- Recall $\lambda_j^{\mathrm{sym}} = 1 - \mu_j$. Hence, for a connected graph, $1 = \mu_0 \geq \mu_1 \geq \cdots \geq \mu_{n-1} > -1$.
- The *diffusion map* $\Phi_t : V = X \to \mathbb{R}^n$ is defined as

$$\Phi_t(\boldsymbol{x}_i) := [\mu_0^t \boldsymbol{\phi}_0(i), \mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_{n-1}^t \boldsymbol{\phi}_{n-1}(i)]^{\mathsf{T}} \quad t > 0.$$

- Often the first coordinate $\mu_0^t \boldsymbol{\phi}_0(i)$ is neglected since its common for all i's ($\mu_0 = 1$, $\boldsymbol{\phi}_0$ is a constant vector), and not providing useful information.
- A truncated version $\Phi_t^\delta : V \to \mathbb{R}^m$, $0 < \delta < 1$, is defined by

$$\Phi_t^\delta(\boldsymbol{x}_i) := [\mu_1^t \boldsymbol{\phi}_1(i), \ldots, \mu_m^t \boldsymbol{\phi}_m(i)]^{\mathsf{T}} \quad t > 0,$$

where $m \ll n$ is chosen by $|\mu_m|^t > \delta$, $|\mu_{m+1}|^t \leq \delta$.

Diffusion Distances

- Now define the *diffusion distance* between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$D_t(\boldsymbol{x}_i, \boldsymbol{x}_j) := \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2.$$

- $D_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a weighted ℓ^2-distance between the probability clouds after t time steps of random walks starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
- From the Markov chain/random walk interpretation, we have

$$(P)_{ij} = p_{ij} = \Pr(s(t+1) = \boldsymbol{x}_j \,|\, s(t) = \boldsymbol{x}_i) \quad \text{for any } t \in \mathbb{N} \cup \{0\}.$$

- Hence,

$$(P^t)_{ij} =: p^t_{ij} = \Pr(s(t) = \boldsymbol{x}_j \,|\, s(0) = \boldsymbol{x}_i),$$

  i.e., the entries of P^t give us the probability to get from one state to another in t *time steps*. t can be viewed as a *scale* parameter.
- Thanks to the biorthogonality, we have

$$P^t = \Phi M^t \Psi^{\top}, \quad p^t_{ij} = \sum_{k=0}^{n-1} \mu_k^t \boldsymbol{\phi}_k(i) \boldsymbol{\psi}_k(j).$$

Diffusion Distances

- Now define the *diffusion distance* between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$D_t(\boldsymbol{x}_i, \boldsymbol{x}_j) := \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2.$$

- $D_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a weighted ℓ^2-distance between the probability clouds after t time steps of random walks starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
- From the Markov chain/random walk interpretation, we have

$$(P)_{ij} = p_{ij} = \Pr(s(t+1) = \boldsymbol{x}_j \mid s(t) = \boldsymbol{x}_i) \quad \text{for any } t \in \mathbb{N} \cup \{0\}.$$

- Hence,

$$(P^t)_{ij} =: p^t_{ij} = \Pr(s(t) = \boldsymbol{x}_j \mid s(0) = \boldsymbol{x}_i),$$

i.e., the entries of P^t give us the probability to get from one state to another in t *time steps*. t can be viewed as a *scale* parameter.
- Thanks to the biorthogonality, we have

$$P^t = \Phi M^t \Psi^{\mathsf{T}}, \quad p^t_{ij} = \sum_{k=0}^{n-1} \mu_k^t \boldsymbol{\phi}_k(i) \boldsymbol{\psi}_k(j).$$

Diffusion Distances

- Now define the *diffusion distance* between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$D_t(\boldsymbol{x}_i, \boldsymbol{x}_j) := \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2.$$

- $D_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a weighted ℓ^2-distance between the probability clouds after t time steps of random walks starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
- From the Markov chain/random walk interpretation, we have

$$(P)_{ij} = p_{ij} = \Pr(s(t+1) = \boldsymbol{x}_j \,|\, s(t) = \boldsymbol{x}_i) \quad \text{for any } t \in \mathbb{N} \cup \{0\}.$$

- Hence,

$$(P^t)_{ij} =: p^t_{ij} = \Pr(s(t) = \boldsymbol{x}_j \,|\, s(0) = \boldsymbol{x}_i),$$

i.e., the entries of P^t give us the probability to get from one state to another in t *time steps*. t can be viewed as a *scale* parameter.
- Thanks to the biorthogonality, we have

$$P^t = \Phi M^t \Psi^{\mathsf{T}}, \quad p^t_{ij} = \sum_{k=0}^{n-1} \mu^t_k \boldsymbol{\phi}_k(i) \boldsymbol{\psi}_k(j).$$

Diffusion Distances

- Now define the *diffusion distance* between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$D_t(\boldsymbol{x}_i, \boldsymbol{x}_j) := \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2.$$

- $D_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a weighted ℓ^2-distance between the probability clouds after t time steps of random walks starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
- From the Markov chain/random walk interpretation, we have

$$(P)_{ij} = p_{ij} = \Pr(s(t+1) = \boldsymbol{x}_j \,|\, s(t) = \boldsymbol{x}_i) \quad \text{for any } t \in \mathbb{N} \cup \{0\}.$$

- Hence,

$$(P^t)_{ij} =: p^t_{ij} = \Pr(s(t) = \boldsymbol{x}_j \,|\, s(0) = \boldsymbol{x}_i),$$

i.e., the entries of P^t give us the probability to get from one state to another in *t time steps*. t can be viewed as a *scale* parameter.
- Thanks to the biorthogonality, we have

$$P^t = \Phi M^t \Psi^\top, \quad p^t_{ij} = \sum_{k=0}^{n-1} \mu_k^t \boldsymbol{\phi}_k(i) \boldsymbol{\psi}_k(j).$$

Diffusion Distances

- Now define the *diffusion distance* between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$D_t(\boldsymbol{x}_i, \boldsymbol{x}_j) := \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2.$$

- $D_t(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a weighted ℓ^2-distance between the probability clouds after t time steps of random walks starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.
- From the Markov chain/random walk interpretation, we have

$$(P)_{ij} = p_{ij} = \Pr(s(t+1) = \boldsymbol{x}_j \,|\, s(t) = \boldsymbol{x}_i) \quad \text{for any } t \in \mathbb{N} \cup \{0\}.$$

- Hence,

$$(P^t)_{ij} =: p^t_{ij} = \Pr(s(t) = \boldsymbol{x}_j \,|\, s(0) = \boldsymbol{x}_i),$$

  i.e., the entries of P^t give us the probability to get from one state to another in *t time steps*. t can be viewed as a *scale* parameter.
- Thanks to the biorthogonality, we have

$$P^t = \Phi M^t \Psi^\top, \quad p^t_{ij} = \sum_{k=0}^{n-1} \mu_k^t \boldsymbol{\phi}_k(i) \boldsymbol{\psi}_k(j).$$

Diffusion Distances . . .

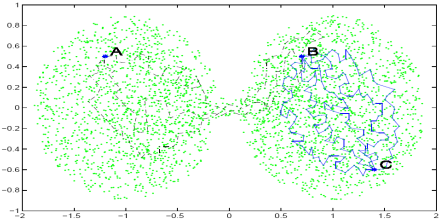

Figure: Courtesy: R. R. Coifman & S. Lafon

- Diffusions between A and B have to go through the bottleneck while C is easily reachable from B.
- The Markov matrix defining a diffusion could be given by a kernel or by inference between neighboring nodes.
- The diffusion distance accounts for *preponderance of inference links*. The shortest path (i.e., the geodesic distance) between A and C is roughly the same as that between B and C.
- The diffusion distance between A and B, however, is larger than that between B and C since diffusion occurs through a bottleneck.

Diffusion Distances . . .

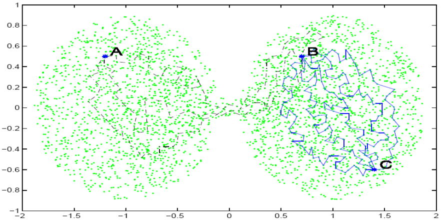

Figure: Courtesy: R. R. Coifman & S. Lafon

- Diffusions between A and B have to go through the bottleneck while C is easily reachable from B.
- The Markov matrix defining a diffusion could be given by a kernel or by inference between neighboring nodes.
- The diffusion distance accounts for *preponderance of inference links*. The shortest path (i.e., the geodesic distance) between A and C is roughly the same as that between B and C.
- The diffusion distance between A and B, however, is larger than that between B and C since diffusion occurs through a bottleneck.

Diffusion Distances . . .

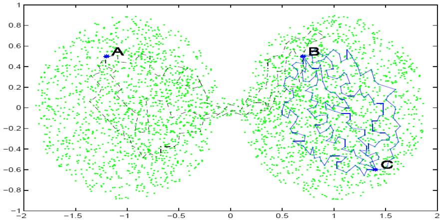

Figure: Courtesy: R. R. Coifman & S. Lafon

- Diffusions between A and B have to go through the bottleneck while C is easily reachable from B.
- The Markov matrix defining a diffusion could be given by a kernel or by inference between neighboring nodes.
- The diffusion distance accounts for *preponderance of inference links*. The shortest path (i.e., the geodesic distance) between A and C is roughly the same as that between B and C.
- The diffusion distance between A and B, however, is larger than that between B and C since diffusion occurs through a bottleneck.

Diffusion Distances . . .

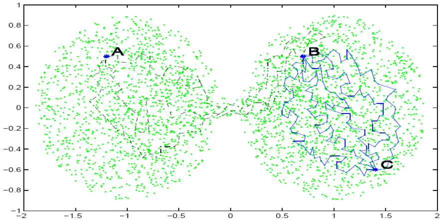

Figure: Courtesy: R. R. Coifman & S. Lafon

- Diffusions between A and B have to go through the bottleneck while C is easily reachable from B.
- The Markov matrix defining a diffusion could be given by a kernel or by inference between neighboring nodes.
- The diffusion distance accounts for *preponderance of inference links*. The shortest path (i.e., the geodesic distance) between A and C is roughly the same as that between B and C.
- The diffusion distance between A and B, however, is larger than that between B and C since diffusion occurs through a bottleneck.

Diffusion Distances . . .

- Now let's compute the weighted ℓ^2-distance between the probability clouds $P^t(i,:)$ and $P^t(j,:)$, i.e., the probability distribution of the random walks after t steps starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively.
- Let's choose the weights as $D^{-1} = \mathrm{diag}(1/d_1, \ldots, 1/d_n)$ in the ℓ^2-distance (i.e., the higher the degree of a node, the smaller its influence on the distance).

$$\begin{aligned}
\|P^t(i,:) - P^t(j,:)\|^2_{2,D^{-1}} &= \left((\boldsymbol{e}_i^{\mathsf{T}} - \boldsymbol{e}_j^{\mathsf{T}})\Phi M^t \Psi^{\mathsf{T}}\right) D^{-1} \left((\boldsymbol{e}_i^{\mathsf{T}} - \boldsymbol{e}_j^{\mathsf{T}})\Phi M^t \Psi^{\mathsf{T}}\right)^{\mathsf{T}} \\
&= (\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathsf{T}} \Phi M^t \Psi^{\mathsf{T}} D^{-1} \Psi M^t \Phi^{\mathsf{T}} (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&\overset{(*)}{=} (\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathsf{T}} \Phi M^{2t} \Phi^{\mathsf{T}} (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&= \sum_{k=0}^{n-1} \mu_k^{2t} (\phi_k(i) - \phi_k(j))^2 \\
&= D_t^2(\boldsymbol{x}_i, \boldsymbol{x}_j).
\end{aligned}$$

where $(*)$ is true since $D^{-1/2}\Psi$ is unitary (recall the properties of L_{rw} in Lecture 7).

Diffusion Distances . . .

- Now let's compute the weighted ℓ^2-distance between the probability clouds $P^t(i,:)$ and $P^t(j,:)$, i.e., the probability distribution of the random walks after t steps starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively.
- Let's choose the weights as $D^{-1} = \mathrm{diag}(1/d_1, \ldots, 1/d_n)$ in the ℓ^2-distance (i.e., the higher the degree of a node, the smaller its influence on the distance).

$$
\begin{aligned}
\|P^t(i,:) - P^t(j,:)\|^2_{2,D^{-1}} &= \left((\boldsymbol{e}_i^\top - \boldsymbol{e}_j^\top)\Phi M^t \Psi^\top\right) D^{-1} \left((\boldsymbol{e}_i^\top - \boldsymbol{e}_j^\top)\Phi M^t \Psi^\top\right)^\top \\
&= (\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \Phi M^t \Psi^\top D^{-1} \Psi M^t \Phi^\top (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&\overset{(*)}{=} (\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \Phi M^{2t} \Phi^\top (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&= \sum_{k=0}^{n-1} \mu_k^{2t} (\boldsymbol{\phi}_k(i) - \boldsymbol{\phi}_k(j))^2 \\
&= D_t^2(\boldsymbol{x}_i, \boldsymbol{x}_j).
\end{aligned}
$$

where $(*)$ is true since $D^{-1/2}\Psi$ is unitary (recall the properties of L_{rw} in Lecture 7).

Diffusion Distances . . .

- Now let's compute the weighted ℓ^2-distance between the probability clouds $P^t(i,:)$ and $P^t(j,:)$, i.e., the probability distribution of the random walks after t steps starting at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively.
- Let's choose the weights as $D^{-1} = \mathrm{diag}(1/d_1, \ldots, 1/d_n)$ in the ℓ^2-distance (i.e., the higher the degree of a node, the smaller its influence on the distance).

$$
\begin{aligned}
\|P^t(i,:) - P^t(j,:)\|^2_{2,D^{-1}} &= \left((\boldsymbol{e}_i^\top - \boldsymbol{e}_j^\top)\Phi M^t \Psi^\top\right) D^{-1} \left((\boldsymbol{e}_i^\top - \boldsymbol{e}_j^\top)\Phi M^t \Psi^\top\right)^\top \\
&= (\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \Phi M^t \Psi^\top D^{-1} \Psi M^t \Phi^\top (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&\overset{(*)}{=} (\boldsymbol{e}_i - \boldsymbol{e}_j)^\top \Phi M^{2t} \Phi^\top (\boldsymbol{e}_i - \boldsymbol{e}_j) \\
&= \sum_{k=0}^{n-1} \mu_k^{2t} (\boldsymbol{\phi}_k(i) - \boldsymbol{\phi}_k(j))^2 \\
&= D_t^2(\boldsymbol{x}_i, \boldsymbol{x}_j).
\end{aligned}
$$

where $(*)$ is true since $D^{-1/2}\Psi$ is unitary (recall the properties of L_{rw} in Lecture 7).

Diffusion Distances . . .

In practice, we use the truncated version Φ_t^δ instead of Φ_t.

Proposition (A. Singer (2011?))

$$\|\Phi_t(\boldsymbol{x}_i)-\Phi_t(\boldsymbol{x}_j)\|_2^2-\frac{2\delta^2}{d_{\min}}(1-\delta_{ij}) \le \|\Phi_t^\delta(\boldsymbol{x}_i)-\Phi_t^\delta(\boldsymbol{x}_j)\|_2^2 \le \|\Phi_t(\boldsymbol{x}_i)-\Phi_t(\boldsymbol{x}_j)\|_2^2,$$

where δ_{ij} is Kronecker's delta.

Proof. Recall $D^{+1/2}\Phi$ is unitary. Hence,

$$\begin{aligned}\|\Phi(i,:)-\Phi(j,:)\|_2^2 &= (\boldsymbol{e}_i-\boldsymbol{e}_j)^\top\Phi\Phi^\top(\boldsymbol{e}_i-\boldsymbol{e}_j)\\ &= (\boldsymbol{e}_i-\boldsymbol{e}_j)^\top D^{-1}(\boldsymbol{e}_i-\boldsymbol{e}_j)\\ &= \frac{1}{d_i}+\frac{1}{d_j}-\frac{2}{d_i}\delta_{ij}\\ &\le \frac{2}{d_{\min}}(1-\delta_{ij}).\end{aligned}$$

Diffusion Distances . . .

In practice, we use the truncated version Φ_t^δ instead of Φ_t.

Proposition (A. Singer (2011?))

$$\|\Phi_t(\boldsymbol{x}_i)-\Phi_t(\boldsymbol{x}_j)\|_2^2-\frac{2\delta^2}{d_{\min}}(1-\delta_{ij})\le\|\Phi_t^\delta(\boldsymbol{x}_i)-\Phi_t^\delta(\boldsymbol{x}_j)\|_2^2\le\|\Phi_t(\boldsymbol{x}_i)-\Phi_t(\boldsymbol{x}_j)\|_2^2,$$

where δ_{ij} is Kronecker's delta.

Proof. Recall $D^{+1/2}\Phi$ is unitary. Hence,

$$\begin{aligned}\|\Phi(i,:)-\Phi(j,:)\|_2^2 &= (\boldsymbol{e}_i-\boldsymbol{e}_j)^\mathsf{T}\Phi\Phi^\mathsf{T}(\boldsymbol{e}_i-\boldsymbol{e}_j)\\ &= (\boldsymbol{e}_i-\boldsymbol{e}_j)^\mathsf{T}D^{-1}(\boldsymbol{e}_i-\boldsymbol{e}_j)\\ &= \frac{1}{d_i}+\frac{1}{d_j}-\frac{2}{d_i}\delta_{ij}\\ &\le \frac{2}{d_{\min}}(1-\delta_{ij}).\end{aligned}$$

Finally,

$$
\begin{aligned}
\|\Phi_t^\delta(\boldsymbol{x}_i) - \Phi_t^\delta(\boldsymbol{x}_j)\|_2^2 &= \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2^2 - \sum_{k:|\mu_k|^t<\delta} \mu_k^{2t}(\boldsymbol{\phi}_k(i) - \boldsymbol{\phi}_k(j))^2 \\
&\geq \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2^2 - \delta^2 \sum_{k:|\mu_k|^t<\delta} (\boldsymbol{\phi}_k(i) - \boldsymbol{\phi}_k(j))^2 \\
&\geq \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2^2 - \delta^2 \sum_{k=0}^{n-1} (\boldsymbol{\phi}_k(i) - \boldsymbol{\phi}_k(j))^2 \\
&= \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2^2 - \delta^2 \|\Phi(i,:) - \Phi(j,:)\|_2^2 \\
&\geq \|\Phi_t(\boldsymbol{x}_i) - \Phi_t(\boldsymbol{x}_j)\|_2^2 - \delta^2 \frac{2}{d_{\min}}(1 - \delta_{ij}).
\end{aligned}
$$

On the other hand, the inequality of the other direction is obvious. □

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x}, \boldsymbol{y}) = D_t(\boldsymbol{y}, \boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x}, \boldsymbol{y}) \leq D_t(\boldsymbol{x}, \boldsymbol{y}) + D_t(\boldsymbol{x}, \boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x}, \boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x}, \boldsymbol{y}) = D_t(\boldsymbol{y}, \boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x}, \boldsymbol{y}) \leq D_t(\boldsymbol{x}, \boldsymbol{y}) + D_t(\boldsymbol{x}, \boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x}, \boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x}, \boldsymbol{y}) = D_t(\boldsymbol{y}, \boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x}, \boldsymbol{y}) \leq D_t(\boldsymbol{x}, \boldsymbol{y}) + D_t(\boldsymbol{x}, \boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x}, \boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x},\boldsymbol{y}) = D_t(\boldsymbol{y},\boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x},\boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x},\boldsymbol{y}) \leq D_t(\boldsymbol{x},\boldsymbol{y}) + D_t(\boldsymbol{x},\boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x},\boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x}, \boldsymbol{y}) = D_t(\boldsymbol{y}, \boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x}, \boldsymbol{y}) \leq D_t(\boldsymbol{x}, \boldsymbol{y}) + D_t(\boldsymbol{x}, \boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x}, \boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #

Is the Diffusion Distance a Metric?

Unfortunately, the answer is NO:

- Symmetry: $D_t(\boldsymbol{x}, \boldsymbol{y}) = D_t(\boldsymbol{y}, \boldsymbol{x})$ ✓
- Nonnegativity: $D_t(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ ✓
- Triangle inequality: $D_t(\boldsymbol{x}, \boldsymbol{y}) \leq D_t(\boldsymbol{x}, \boldsymbol{y}) + D_t(\boldsymbol{x}, \boldsymbol{z})$ ✓
- Identity of indiscernibles: $D_t(\boldsymbol{x}, \boldsymbol{y}) = 0 \overset{\not\Rightarrow}{\Leftarrow} \boldsymbol{x} = \boldsymbol{y}$.

Consider the case when $A(i,:) = \alpha A(j,:)$, $\exists \alpha > 0$, e.g., a path with 3 vertices, v_1, v_2, v_3 with the uniform edge weights 1. Then, $D_t(v_1, v_3) = 0$ although $v_1 \neq v_3$. #