Effective Resistance and Conductance for Probability Measures on Graphs

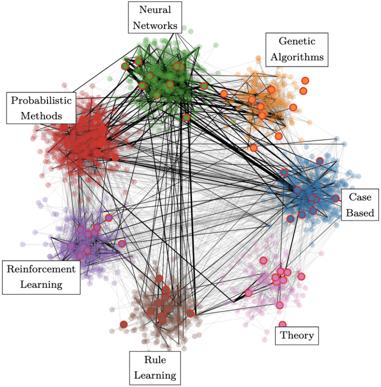
CodEx Seminar 2025

Sawyer Jack Robertson
UC San Diego

July 15, 2025

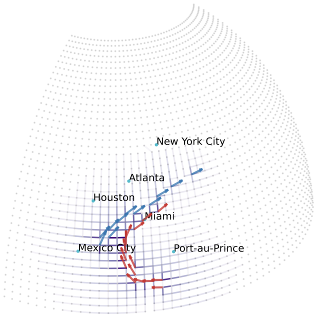
Graphs emerge from a variety of data science problems

Graphs emerge from a variety of data science problems



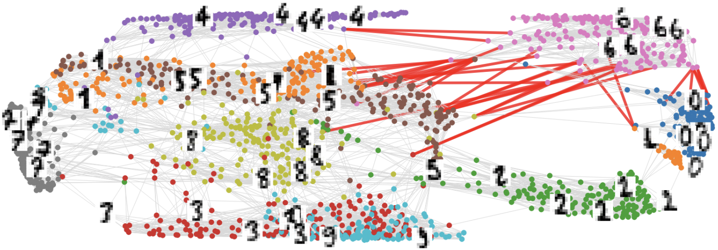
Example 1: An illustration of the citation dataset Cora (nodes are CS research papers, edges if one paper cites another). Nodes colored according to their topic.

Graphs emerge from a variety of data science problems



Example 2: A geometric graph with vector fields corresponding to tropical storms shown. Nodes correspond to grid points in latitude-longitude coordinates, edges based on proximity (not shown).

Graphs emerge from a variety of data science problems



Example 3: An affinity graph drawn on the Sklearn Digits dataset. Each node corresponds to one 8×8 grayscale image of a handwritten digit, and edges are based on Euclidean proximity.

The setting of this talk is two worlds of graph-based methods in data science...

The setting of this talk is two worlds of graph-based methods in data science...

Optimal transportation on graphs, which concerns families of optimization problems on the probability simplex of the vertices of a graph;

The setting of this talk is two worlds of graph-based methods in data science...

Optimal transportation on graphs, which concerns families of optimization problems on the probability simplex of the vertices of a graph;

Graph-based semi-supervised learning, which concerns methods for building data classification models in label-sparse regimes built upon well-designed graphs.

The setting of this talk is two worlds of graph-based methods in data science...

Optimal transportation on graphs, which concerns families of optimization problems on the probability simplex of the vertices of a graph;

Graph-based semi-supervised learning, which concerns methods for building data classification models in label-sparse regimes built upon well-designed graphs.

The goal of this talk is to convince the listener of a surprising duality between these worlds, and cover some new work and results along the way.

Our first topic is optimal transportation on graphs.

Our first topic is optimal transportation on graphs.

- $G = (V, E, w)$ is a weighted, connected, undirected graph on n vertices and m edges,

Our first topic is optimal transportation on graphs.

- $G = (V, E, w)$ is a weighted, connected, undirected graph on n vertices and m edges,
- $\mathcal{P}(V) = \{\mu : V \to \mathbb{R} : \mu \geq 0, \mathbf{1}^\top \mu = 1\}$, and

Our first topic is optimal transportation on graphs.

- $G = (V, E, w)$ is a weighted, connected, undirected graph on n vertices and m edges,
- $\mathcal{P}(V) = \{\mu : V \to \mathbb{R} : \mu \geq 0, \mathbf{1}^\top \mu = 1\}$, and
- $d(i, j)$ weighted shortest path distance between $i, j \in V$.

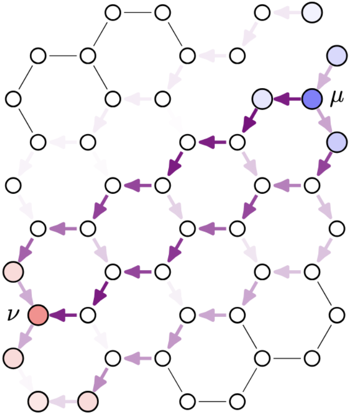
Our first topic is optimal transportation on graphs.

- $G = (V, E, w)$ is a weighted, connected, undirected graph on n vertices and m edges,
- $\mathcal{P}(V) = \{\mu : V \to \mathbb{R} : \mu \geq 0, \mathbf{1}^\top \mu = 1\}$, and
- $d(i, j)$ weighted shortest path distance between $i, j \in V$.

Note: We identify functions $f : V \to \mathbb{R}$ and vectors $f \in \mathbb{R}^n$; moreover, since our graph is finite, probability measures and density vectors can be used interchangeably.

Generically speaking, optimal transportation concerns a variety of optimization problems which model how costly it is to move mass (given by probability distributions) from one "location" μ to another ν in a metric space.

Generically speaking, optimal transportation concerns a variety of optimization problems which model how costly it is to move mass (given by probability distributions) from one “location” μ to another ν in a metric space.

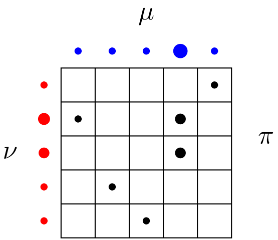


Suppose $\mu, \nu \in \mathcal{P}(V)$. Then we define the set of μ, ν-*couplings*, denoted $\Gamma(\mu, \nu)$ by

$$\Gamma(\mu, \nu) = \left\{ \pi \in \mathbb{R}^{n \times n} \, : \, \pi_{ij} \geq 0, \sum_{i \in V} \pi_{ij} = \mu_j, \sum_{j \in V} \pi_{ij} = \nu_i \right\}.$$

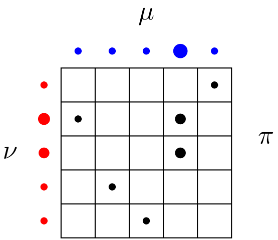
Suppose $\mu, \nu \in \mathcal{P}(V)$. Then we define the set of μ, ν-*couplings*, denoted $\Gamma(\mu, \nu)$ by

$$\Gamma(\mu, \nu) = \left\{ \pi \in \mathbb{R}^{n \times n} \, : \, \pi_{ij} \geq 0, \sum_{i \in V} \pi_{ij} = \mu_j, \sum_{j \in V} \pi_{ij} = \nu_i \right\}.$$



Suppose $\mu, \nu \in \mathcal{P}(V)$. Then we define the set of μ, ν-*couplings*, denoted $\Gamma(\mu, \nu)$ by

$$\Gamma(\mu,\nu) = \left\{ \pi \in \mathbb{R}^{n \times n} \;:\; \pi_{ij} \geq 0, \sum_{i \in V} \pi_{ij} = \mu_j, \sum_{j \in V} \pi_{ij} = \nu_i \right\}.$$



Wasserstein distance is defined as follows, for $\mu, \nu \in \mathcal{P}(V)$ fixed:

$$W_p(\mu,\nu)^p = \inf_{\pi \in \Gamma(\mu,\nu)} \left\{ \sum_{i,j} \pi_{ij} d(i,j)^p \right\},$$

$$1 \leq p < \infty.$$

For $p = 1$, W_1 can be recast as a **min cost flow problem**:

$$W_1(\mu, \nu) = \inf \left\{ \sum_e |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

where $B \in \mathbb{R}^{n \times m}$ is the oriented graph vertex-edge incidence matrix.

For $p = 1$, W_1 can be recast as a **min cost flow problem**:

$$W_1(\mu, \nu) = \inf \left\{ \sum_e |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

where $B \in \mathbb{R}^{n \times m}$ is the oriented graph vertex-edge incidence matrix.

$$B_{ie_j} = \begin{cases} 1 & \text{if } e_j = (i, \cdot) \\ -1 & \text{if } e_j = (\cdot, i) \\ 0 & \text{otherwise.} \end{cases}$$

For $p = 1$, W_1 can be recast as a **min cost flow problem**:

$$W_1(\mu, \nu) = \inf \left\{ \sum_e |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

where $B \in \mathbb{R}^{n \times m}$ is the oriented graph vertex-edge incidence matrix.

$$B_{ie_j} = \begin{cases} 1 & \text{if } e_j = (i, \cdot) \\ -1 & \text{if } e_j = (\cdot, i) \\ 0 & \text{otherwise.} \end{cases}$$

$$BJ(i) = (\text{outgoing mass at } i) - (\text{incoming mass at } i)$$

For $p = 1$, W_1 can be recast as a **min cost flow problem**:

$$W_1(\mu, \nu) = \inf \left\{ \sum_e |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

where $B \in \mathbb{R}^{n \times m}$ is the oriented graph vertex-edge incidence matrix.

$$B_{ie_j} = \begin{cases} 1 & \text{if } e_j = (i, \cdot) \\ -1 & \text{if } e_j = (\cdot, i) \\ 0 & \text{otherwise.} \end{cases} \qquad \begin{aligned} BJ(i) = {} & (\text{outgoing mass at } i) \\ & - (\text{incoming mass at } i) \end{aligned}$$

This is a well-known result and can be shown by, e.g., computing Lagrangian duals repeatedly.

W_1 is known as the "Beckmann problem" on Graphs.

W_1 is known as the "Beckmann problem" on Graphs.

Has been used in computer graphics, notions of Ricci curvature on graphs, weather pattern modeling, ...

W_1 is known as the "Beckmann problem" on Graphs.

Has been used in computer graphics, notions of Ricci curvature on graphs, weather pattern modeling, ...

Challenges: non-uniqueness of J (can be addressed via regularization); expensive at scale

W_1 is known as the "Beckmann problem" on Graphs.

Has been used in computer graphics, notions of Ricci curvature on graphs, weather pattern modeling, ...

Challenges: non-uniqueness of J (can be addressed via regularization); expensive at scale

What if we focus on the minimum cost flow program to investigate new ways of modeling transport on graphs?

Enter **effective resistance**

Enter **effective resistance** between nodes $i, j \in V$:

$$r_{ij} = (\delta_i - \delta_j)^T L^\dagger (\delta_i - \delta_j)$$

where $L^\dagger$ is the Moore-Penrose pseudoinverse of the graph Laplacian $L = D - A$ and δ_i, δ_j are Dirac masses.

Enter **effective resistance** between nodes $i, j \in V$:

$$r_{ij} = (\delta_i - \delta_j)^T L^\dagger (\delta_i - \delta_j)$$

where $L^\dagger$ is the Moore-Penrose pseudoinverse of the graph Laplacian $L = D - A$ and δ_i, δ_j are Dirac masses.

$$\begin{aligned} r_{ij} &= \|L^{-\dagger/2}\delta_i - L^{-\dagger/2}\delta_j\|^2 \\ &= \inf\left\{\sum_e |J(e)|^2 w_e : J : E \to \mathbb{R}, BJ = \delta_i - \delta_j\right\} \qquad (1) \\ &= \ldots \end{aligned}$$

Enter **effective resistance** between nodes $i, j \in V$:

$$r_{ij} = (\delta_i - \delta_j)^T L^\dagger (\delta_i - \delta_j)$$

where $L^\dagger$ is the Moore-Penrose pseudoinverse of the graph Laplacian $L = D - A$ and δ_i, δ_j are Dirac masses.

$$\begin{aligned} r_{ij} &= \|L^{-\dagger/2}\delta_i - L^{-\dagger/2}\delta_j\|^2 \\ &= \inf \left\{ \sum_e |J(e)|^2 w_e : J : E \to \mathbb{R}, BJ = \delta_i - \delta_j \right\} \qquad (1) \\ &= \ldots \end{aligned}$$

Originated from electrical network models as the resistance in a network between two unit sources.

Enter **effective resistance** between nodes $i, j \in V$:

$$r_{ij} = (\delta_i - \delta_j)^T L^\dagger (\delta_i - \delta_j)$$

where $L^\dagger$ is the Moore-Penrose pseudoinverse of the graph Laplacian $L = D - A$ and δ_i, δ_j are Dirac masses.

$$\begin{aligned} r_{ij} &= \|L^{-\dagger/2}\delta_i - L^{-\dagger/2}\delta_j\|^2 \\ &= \inf\left\{\sum_e |J(e)|^2 w_e : J : E \to \mathbb{R}, BJ = \delta_i - \delta_j\right\} \qquad (1) \\ &= \ldots \end{aligned}$$

Originated from electrical network models as the resistance in a network between two unit sources.

Has deep connections to the simple random walk: $r_{ij} = \frac{1}{\mathrm{vol}(G)} C(i, j)$ where $C(i, j)$ is commute time (expected time to go from i to j and back).

Enter **effective resistance** between nodes $i, j \in V$:

$$r_{ij} = (\delta_i - \delta_j)^T L^\dagger (\delta_i - \delta_j)$$

where $L^\dagger$ is the Moore-Penrose pseudoinverse of the graph Laplacian $L = D - A$ and δ_i, δ_j are Dirac masses.

$$\begin{aligned} r_{ij} &= \| L^{-\dagger/2}\delta_i - L^{-\dagger/2}\delta_j \|^2 \\ &= \inf \left\{ \sum_e |J(e)|^2 w_e : J : E \to \mathbb{R}, BJ = \delta_i - \delta_j \right\} \\ &= \ldots \end{aligned} \tag{1}$$

Originated from electrical network models as the resistance in a network between two unit sources.

Has deep connections to the simple random walk: $r_{ij} = \frac{1}{\mathrm{vol}(G)} C(i, j)$ where $C(i, j)$ is commute time (expected time to go from i to j and back).

Applications in graph sparsification, GNNs.

Why not extend (1) to probability measures?

Why not extend (1) to probability measures? For $\mu, \nu \in \mathcal{P}(V)$, we introduce the p-**Beckmann metric** $B_{w,p}(\mu, \nu)$ via:

Why not extend (1) to probability measures? For $\mu, \nu \in \mathcal{P}(V)$, we introduce the *p*-**Beckmann metric** $B_{w,p}(\mu, \nu)$ via:

For $1 \leq p < \infty$, we set:

$$B_{w,p}(\mu, \nu)^p = \inf \left\{ \sum_{e \in E} |J(e)|^p w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

For $p = \infty$, we set

$$B_{w,\infty}(\mu, \nu) = \inf \left\{ \max_{e \in E} |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

Why not extend (1) to probability measures? For $\mu, \nu \in \mathcal{P}(V)$, we introduce the p-**Beckmann metric** $B_{w,p}(\mu, \nu)$ via:

For $1 \leq p < \infty$, we set:

$$B_{w,p}(\mu,\nu)^p = \inf \left\{ \sum_{e \in E} |J(e)|^p w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

For $p = \infty$, we set

$$B_{w,\infty}(\mu,\nu) = \inf \left\{ \max_{e \in E} |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

Theorem R., Z. Wan, A. Cloninger (2024)

1. When $p = 1$: $B_{w^{-1},1}(\mu,\nu) = W_{\widetilde{d},1}(\mu,\nu)$ for $\widetilde{d}$ shortest path distance on (V, E, w^{-1})
2. When $p = 2$: $B_{w^{-1},2}(\mu,\nu)^2 = (\mu - \nu)^T L^\dagger (\mu - \nu)$

Why not extend (1) to probability measures? For $\mu, \nu \in \mathcal{P}(V)$, we introduce the p-**Beckmann metric** $B_{w,p}(\mu, \nu)$ via:

For $1 \leq p < \infty$, we set:

$$B_{w,p}(\mu, \nu)^p = \inf \left\{ \sum_{e \in E} |J(e)|^p w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

For $p = \infty$, we set

$$B_{w,\infty}(\mu, \nu) = \inf \left\{ \max_{e \in E} |J(e)| w_e : J : E \to \mathbb{R}, BJ = \mu - \nu \right\}$$

Theorem R., Z. Wan, A. Cloninger (2024)

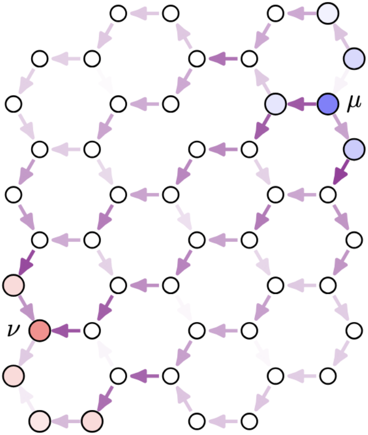
1. When $p = 1$: $B_{w^{-1},1}(\mu, \nu) = W_{\widetilde{d},1}(\mu, \nu)$ for $\widetilde{d}$ shortest path distance on (V, E, w^{-1})
2. When $p = 2$: $B_{w^{-1},2}(\mu, \nu)^2 = (\mu - \nu)^T L^\dagger (\mu - \nu)$

Important to think of the weights as *affinities* in this setting

$\widetilde{d}$ is a more natural distance than d

Beckmann-2 distance, or **measure effective resistance** can be thought of as a “quadratic” min cost flow problem for probability measures

Beckmann-2 distance, or **measure effective resistance** can be thought of as a "quadratic" min cost flow problem for probability measures



This notion seems to have some nice theory, for example, if we define the **graph Sobolev seminorm**

$$\|f\|^2_{\dot{H}^1(V)} = \sum_{\{i,j\} \in E} w_{ij} |f(i) - f(j)|^2$$

This notion seems to have some nice theory, for example, if we define the **graph Sobolev seminorm**

$$\|f\|^2_{\dot{H}^1(V)} = \sum_{\{i,j\}\in E} w_{ij}|f(i) - f(j)|^2$$

and its dual seminorm

$$\|f\|_{\dot{H}^{-1}(V)} = \sup\left\{f^\top g : \|f\|_{\dot{H}^1(V)} \leq 1\right\},$$

This notion seems to have some nice theory, for example, if we define the **graph Sobolev seminorm**

$$\|f\|_{\dot{H}^1(V)}^2 = \sum_{\{i,j\} \in E} w_{ij} |f(i) - f(j)|^2$$

and its dual seminorm

$$\|f\|_{\dot{H}^{-1}(V)} = \sup \left\{ f^\top g : \|f\|_{\dot{H}^1(V)} \leq 1 \right\},$$

then we can obtain a graph version of the Benamou-Brenier formula:

Theorem R., Z. Wan, A. Cloninger (2024)

$$B_2(\mu, \nu)^2 = \inf \left\{ \int_0^1 \|d\mu_t\|_{\dot{H}^{-1}(V)}^2 dt : \mu_t \in C^1([0,1]), \mu_0 = \mu, \mu_1 = \nu \right\}$$

This notion seems to have some nice theory, for example, if we define the **graph Sobolev seminorm**

$$\|f\|^2_{\dot{H}^1(V)} = \sum_{\{i,j\}\in E} w_{ij}|f(i) - f(j)|^2$$

and its dual seminorm

$$\|f\|_{\dot{H}^{-1}(V)} = \sup\left\{f^\top g : \|f\|_{\dot{H}^1(V)} \leq 1\right\},$$

then we can obtain a graph version of the Benamou-Brenier formula:

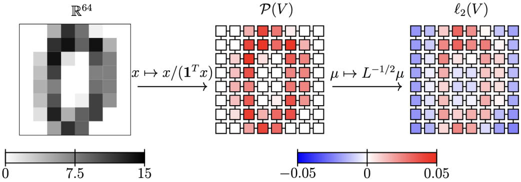
Theorem R., Z. Wan, A. Cloninger (2024)

$$B_2(\mu,\nu)^2 = \inf\left\{\int_0^1 \|d\mu_t\|^2_{\dot{H}^{-1}(V)} dt : \mu_t \in C^1([0,1]), \mu_0 = \mu, \mu_1 = \nu\right\}$$

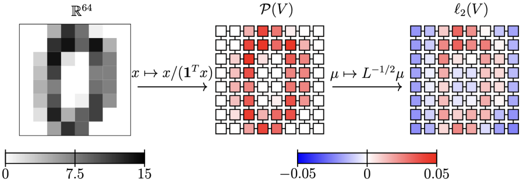
There are also connections to random walks on the graph, but these require some additional machinery to state.

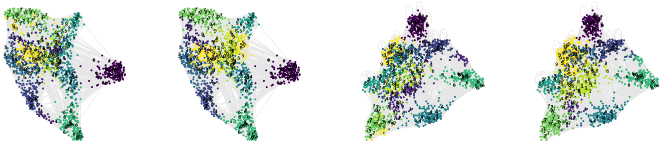
This setup leads to a natural pipeline for scalable graph-based unsupervised learning.

This setup leads to a natural pipeline for scalable graph-based unsupervised learning.



This setup leads to a natural pipeline for scalable graph-based unsupervised learning.





Nodes colored by ground truth clusters | Nodes colored by spectral clustering | Nodes colored by ground truth clusters | Nodes colored by spectral clustering

Beckmann-2 Distance | Wasserstein-2 Distance

I buried the lede a few slides ago in stating this result:

I buried the lede a few slides ago in stating this result:

Theorem SJR, Z. Wan, A. Cloninger (2024)

1. When $p = 1$: $B_{w^{-1},1}(\mu, \nu) = W_{\widetilde{d},1}(\mu, \nu)$ for $\widetilde{d}$ shortest path distance on (V, E, w^{-1})
2. When $p = 2$: $B_{w^{-1},2}(\mu, \nu)^2 = (\mu - \nu)^T L^\dagger (\mu - \nu)$

I buried the lede a few slides ago in stating this result:

Theorem SJR, Z. Wan, A. Cloninger (2024)

1. When $p = 1$: $B_{w^{-1},1}(\mu, \nu) = W_{\widetilde{d},1}(\mu, \nu)$ for $\widetilde{d}$ shortest path distance on (V, E, w^{-1})
2. When $p = 2$: $B_{w^{-1},2}(\mu, \nu)^2 = (\mu - \nu)^T L^\dagger (\mu - \nu)$

What about $p = \infty$?

Usually when someone mentions duality in an optimization context they are referring to the *Lagrangian dual* problem:

Usually when someone mentions duality in an optimization context they are referring to the *Lagrangian dual* problem:

- One forms a Lagrangian function with the objective and its constraints by introducing multipliers; then,

Usually when someone mentions duality in an optimization context they are referring to the *Lagrangian dual* problem:

- One forms a Lagrangian function with the objective and its constraints by introducing multipliers; then,
- by minimizing this Lagrangian over the primal variables and maximizing over the multipliers, one solves the *dual problem*.

Usually when someone mentions duality in an optimization context they are referring to the *Lagrangian dual* problem:

- One forms a Lagrangian function with the objective and its constraints by introducing multipliers; then,
- by minimizing this Lagrangian over the primal variables and maximizing over the multipliers, one solves the *dual problem*.

Strong duality refers to the primal and dual problems achieving the same optimal value.

Gauge optimization refers to programs whose objectives are gauge functions (nonnegative, positively homogeneous, and vanishing at $\vec{0}$):

$$\min_{x \in \mathcal{C}} \kappa(x), \qquad \mathcal{C} \subseteq \mathbb{R}^n \text{ closed, convex.}$$

Gauge optimization refers to programs whose objectives are gauge functions (nonnegative, positively homogeneous, and vanishing at $\vec{0}$):

$$\min_{x \in \mathcal{C}} \kappa(x), \qquad \mathcal{C} \subseteq \mathbb{R}^n \text{ closed, convex.}$$

Gauge duality transforms a primal problem by replacing the objective with its polar and the constraint set with its antipolar:

Gauge optimization refers to programs whose objectives are gauge functions (nonnegative, positively homogeneous, and vanishing at $\vec{0}$):

$$\min_{x\in\mathcal{C}} \kappa(x), \qquad \mathcal{C} \subseteq \mathbb{R}^n \text{ closed, convex.}$$

Gauge duality transforms a primal problem by replacing the objective with its polar and the constraint set with its antipolar:

$$\begin{cases} \min_{y\in\mathcal{C}'} \kappa^\circ(y) & \kappa^\circ(y) = \inf\{\mu > 0 : \langle x, y\rangle \le \mu\kappa(x) \;\; \forall\, x \in \mathbb{R}^n\} \\ & \mathcal{C}' = \{y \in \mathbb{R}^n : \langle y, x\rangle \ge 1 \;\; \forall\, x \in \mathcal{C}\}. \end{cases}$$

Gauge optimization refers to programs whose objectives are gauge functions (nonnegative, positively homogeneous, and vanishing at $\vec{0}$):

$$\min_{x \in \mathcal{C}} \kappa(x), \qquad \mathcal{C} \subseteq \mathbb{R}^n \text{ closed, convex.}$$

Gauge duality transforms a primal problem by replacing the objective with its polar and the constraint set with its antipolar:

$$\begin{cases} \min_{y \in \mathcal{C}'} \kappa^\circ(y) & \kappa^\circ(y) = \inf\{\mu > 0 : \langle x, y \rangle \leq \mu\kappa(x) \ \ \forall\, x \in \mathbb{R}^n\} \\ & \mathcal{C}' = \{y \in \mathbb{R}^n : \langle y, x \rangle \geq 1 \ \ \forall\, x \in \mathcal{C}\}. \end{cases}$$

Strong duality in this setting refers to a *reciprocal* relationship between the optimal values of the primal and gauge dual programs.

Consider the following program: For distributions $\mu, \nu \in \mathcal{P}(V)$, consider the program, for $1 \leq p < \infty$:

$$C_p(\mu, \nu) = \min_{\varphi \in \mathbb{R}^n,\ \varphi^\top(\mu - \nu) = 1} \left(\sum_{\{i,j\} \in E} w_{ij} |\varphi_i - \varphi_j|^p \right)^{1/p},$$

Consider the following program: For distributions $\mu, \nu \in \mathcal{P}(V)$, consider the program, for $1 \leq p < \infty$:

$$C_p(\mu,\nu) = \min_{\varphi \in \mathbb{R}^n,\ \varphi^\top(\mu-\nu)=1} \left(\sum_{\{i,j\}\in E} w_{ij} |\varphi_i - \varphi_j|^p \right)^{1/p},$$

and with the usual modification when $p = \infty$.

Consider the following program: For distributions $\mu, \nu \in \mathcal{P}(V)$, consider the program, for $1 \leq p < \infty$:

$$C_p(\mu, \nu) = \min_{\varphi \in \mathbb{R}^n,\ \varphi^\top(\mu - \nu) = 1} \left(\sum_{\{i,j\} \in E} w_{ij} |\varphi_i - \varphi_j|^p \right)^{1/p},$$

and with the usual modification when $p = \infty$.

This defines the p-**conductance** between μ and ν.

Consider the following program: For distributions $\mu, \nu \in \mathcal{P}(V)$, consider the program, for $1 \le p < \infty$:

$$C_p(\mu, \nu) = \min_{\varphi \in \mathbb{R}^n,\ \varphi^\top(\mu-\nu)=1} \left(\sum_{\{i,j\} \in E} w_{ij} |\varphi_i - \varphi_j|^p \right)^{1/p},$$

and with the usual modification when $p = \infty$.

This defines the p-**conductance** between μ and ν.

"p-conductance" is used since...

$C_p(\mu, \nu)$ is gauge dual to the **Beckmann formulation** of optimal transport:

[1] Alamgir, V. Luxburg. *Phase transition in the family of p-resistances*, 2011.

$C_p(\mu, \nu)$ is gauge dual to the **Beckmann formulation** of optimal transport:

Theorem R., C. Holtz, Z. Wan, G. Mishne, A. Cloninger (2025)

We have the primal-dual relationship:

$$1/\mathcal{C}_p(\mu, \nu) = \begin{cases} \mathcal{B}_{\infty, w^{-1}}(\mu, \nu) & \text{if } p = 1, \\ \mathcal{B}_{q, w^{1-q}}(\mu, \nu) & \text{if } p \in (1, \infty) \text{ and } 1/p + 1/q = 1, \\ \mathcal{B}_{1, w^{-1}}(\mu, \nu) & \text{if } p = \infty. \end{cases}$$

[1] Alamgir, V. Luxburg. *Phase transition in the family of p-resistances*, 2011.

$C_p(\mu, \nu)$ is gauge dual to the **Beckmann formulation** of optimal transport:

Theorem R., C. Holtz, Z. Wan, G. Mishne, A. Cloninger (2025)

We have the primal-dual relationship:

$$1/\mathcal{C}_p(\mu, \nu) = \begin{cases} \mathcal{B}_{\infty, w^{-1}}(\mu, \nu) & \text{if } p = 1, \\ \mathcal{B}_{q, w^{1-q}}(\mu, \nu) & \text{if } p \in (1, \infty) \text{ and } 1/p + 1/q = 1, \\ \mathcal{B}_{1, w^{-1}}(\mu, \nu) & \text{if } p = \infty. \end{cases}$$

Interpretation: p-conductance solves a dual norm problem to transporting mass across the graph with minimum cost.

[1] Alamgir, V. Luxburg. *Phase transition in the family of p-resistances*, 2011.

$\mathcal{C}_p(\mu,\nu)$ is gauge dual to the **Beckmann formulation** of optimal transport:

Theorem R., C. Holtz, Z. Wan, G. Mishne, A. Cloninger (2025)

We have the primal-dual relationship:

$$1/\mathcal{C}_p(\mu,\nu) = \begin{cases} \mathcal{B}_{\infty,w^{-1}}(\mu,\nu) & \text{if } p = 1, \\ \mathcal{B}_{q,w^{1-q}}(\mu,\nu) & \text{if } p \in (1,\infty) \text{ and } 1/p + 1/q = 1, \\ \mathcal{B}_{1,w^{-1}}(\mu,\nu) & \text{if } p = \infty. \end{cases}$$

Interpretation: p-conductance solves a dual norm problem to transporting mass across the graph with minimum cost.

This extends a result of Alamgir and von Luxburg which considers the case of vertices, i.e, Dirac measures.[1]

[1] Alamgir, V. Luxburg. *Phase transition in the family of p-resistances*, 2011.

The program $C_p(\mu, \nu)$ is more than just an artifact of a proof.

The program $C_p(\mu, \nu)$ is more than just an artifact of a proof.

The case of $p = 1$ in particular can be connected to minimum cuts between probability measures (and, in the case of Dirac measures, the standard vertex mincut program).

The program $C_p(\mu, \nu)$ is more than just an artifact of a proof.

The case of $p = 1$ in particular can be connected to minimum cuts between probability measures (and, in the case of Dirac measures, the standard vertex mincut program).

On the applied side, $C_p(\mu, \nu)$ led to a novel graph-based semi-supervised learning method.

In semi-supervised learning, we are tasked with building a classification model from a large dataset of examples, of which only a *small fraction* are labeled.

In semi-supervised learning, we are tasked with building a classification model from a large dataset of examples, of which only a *small fraction* are labeled.

Graph-based methods seek to leverage the unlabeled examples by building a graph on the dataset as a whole and utilizing the underlying structure of the dataset to inform the classification model.

In semi-supervised learning, we are tasked with building a classification model from a large dataset of examples, of which only a *small fraction* are labeled.

Graph-based methods seek to leverage the unlabeled examples by building a graph on the dataset as a whole and utilizing the underlying structure of the dataset to inform the classification model.

Training and unlabeled datapoints are identified as *vertices* in a graph, and are connected with *affinity-weighted edges* when datapoints are close in some metric.

Laplace learning was the earliest such method: Assuming binary class labels, let $f : V \to \mathbb{R}$ denote a $\{-1, 1\}$ indicator of the training node labels on $T \subseteq V$, and the "extend f to the rest of the graph" via

$$\widetilde{f} \in \operatorname{argmin}_{\varphi} \left\{ \sum_{\{i,j\} \in E} |\varphi(i) - \varphi(j)|^2 : \varphi|_T = f|_T \right\}.$$

[2] Zhu, Ghahramani, Lafferty. *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, 2003.

Laplace learning was the earliest such method: Assuming binary class labels, let $f : V \to \mathbb{R}$ denote a $\{-1, 1\}$ indicator of the training node labels on $T \subseteq V$, and the "extend f to the rest of the graph" via

$$\widetilde{f} \in \operatorname{argmin}_{\varphi} \left\{ \sum_{\{i,j\} \in E} |\varphi(i) - \varphi(j)|^2 : \varphi|_T = f|_T \right\}.$$

Class predictions: $\operatorname{sign}(\widetilde{f})$.[2]

[2] Zhu, Ghahramani, Lafferty. *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, 2003.

Laplace learning was the earliest such method: Assuming binary class labels, let $f : V \to \mathbb{R}$ denote a $\{-1, 1\}$ indicator of the training node labels on $T \subseteq V$, and the "extend f to the rest of the graph" via

$$\widetilde{f} \in \operatorname{argmin}_\varphi \left\{ \sum_{\{i,j\} \in E} |\varphi(i) - \varphi(j)|^2 : \varphi|_T = f|_T \right\}.$$

Class predictions: $\operatorname{sign}(\widetilde{f})$.[2]

The p-conductance program defines an SSL method as follows: Model the training labels of one class using a measure μ, the other ν, and then find:

$$\widetilde{f} \in \operatorname{argmin}_\varphi \left\{ \sum_{\{i,j\} \in E} w_{ij} |\varphi_i - \varphi_j|^p : \varphi^\top (\mu - \nu) = 1 \right\}$$

[2] Zhu, Ghahramani, Lafferty. *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, 2003.

Laplace learning was the earliest such method: Assuming binary class labels, let $f : V \to \mathbb{R}$ denote a $\{-1, 1\}$ indicator of the training node labels on $T \subseteq V$, and the "extend f to the rest of the graph" via

$$\widetilde{f} \in \operatorname{argmin}_{\varphi} \left\{ \sum_{\{i,j\} \in E} |\varphi(i) - \varphi(j)|^2 : \varphi|_T = f|_T \right\}.$$

Class predictions: $\operatorname{sign}(\widetilde{f})$.[2]

The p-conductance program defines an SSL method as follows: Model the training labels of one class using a measure μ, the other ν, and then find:

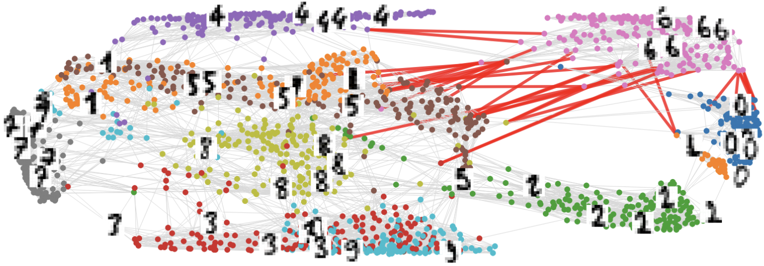
$$\widetilde{f} \in \operatorname{argmin}_{\varphi} \left\{ \sum_{\{i,j\} \in E} w_{ij} |\varphi_i - \varphi_j|^p : \varphi^\top (\mu - \nu) = 1 \right\}$$

Class predictions: $\operatorname{sign}(\widetilde{f} - \overline{\widetilde{f}})$.

[2] Zhu, Ghahramani, Lafferty. *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, 2003.

The method appears performant in many settings, but there are still many open questions on the theory side!

The method appears performant in many settings, but there are still many open questions on the theory side!

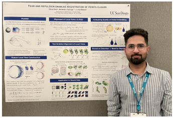


An illustration of a graph built on the Sklearn digits dataset. The red edges highlight where $|\widetilde{f}(i) - \widetilde{f}(j)|$ is large ($p = 1$), illustrating how the predictions of the class of images of the digit six are formed.

I want to acknowledge support from HDSI, the Department of Mathematics, the School of Physical Sciences, and my collaborators, who include but are not limited to:

Alex Cloninger

Gal Mishne



Dhruv Kohli

Zhengchao Wan

Chester Holtz

Questions?

Cora # Labels per class	1	3	5	10	100
Laplace/LP Zhu et al. (2003)	21.8 (14.3)	37.6 (12.3)	51.3 (11.9)	66.9 (6.8)	81.8 (1.1)
Sparse LP Jung et al. (2016)	16.0 (1.8)	19.4 (1.8)	23.1 (2.3)	28.7 (2.2)	47.0 (2.2)
p-laplace Flores et al. (2022)	41.9 (8.9)	57.6 (7.1)	61.9 (6.2)	67.9 (3.6)	79.2 (1.3)
p-eikonal Calder & Ettehad (2022)	40.4 (8.6)	51.8 (7.2)	56.9 (6.8)	64.5 (4.2)	80.2 (1.1)
Poisson Calder et al. (2020)	57.4 (9.2)	67.0 (4.9)	69.3 (3.6)	71.6 (3.0)	76.0 (1.0)
p-conductance ($p = 1, \epsilon = n$)	22.9 (6.8)	22.7 (6.8)	21.4 (6.6)	21.3 (6.6)	35.6 (13.2)
p-conductance ($p = 2, \epsilon = n$)	58.9 (7.2)	67.9 (3.7)	70.2 (2.3)	72.2 (1.9)	75.2 (1.3)
p-conductance ($p = \infty, \epsilon = n$)	44.3 (6.8)	53.7 (5.0)	58.9 (3.6)	63.7 (3.1)	73.7 (1.4)
PoissonMBO Calder et al. (2020)	58.5 (9.4)	68.5 (4.1)	70.7 (3.0)	73.3 (2.3)	80.1 (0.9)
p-conductance ($p = 2, \epsilon = 0$)	**63.1 (8.0)**	**72.9 (3.5)**	**75.5 (1.8)**	**77.9 (1.1)**	**82.9 (0.9)**

Pubmed # Labels per class	1	3	5	10	100
Laplace/LP Zhu et al. (2003)	34.6 (8.8)	35.7 (8.2)	36.9 (8.1)	39.6 (9.1)	74.9 (3.6)
Sparse LP Jung et al. (2016)	32.4 (4.7)	33.0 (4.8)	33.6 (4.8)	33.9 (4.8)	43.2 (4.1)
p-laplace Flores et al. (2022)	44.8 (11.2)	58.3 (9.1)	61.6 (7.7)	66.2 (4.7)	74.3 (1.1)
p-eikonal Calder & Ettehad (2022)	44.3 (11.8)	55.6 (10.0)	58.4 (9.1)	65.1 (5.8)	74.9 (1.5)
Poisson Calder et al. (2020)	55.1 (11.3)	66.6 (7.4)	68.8 (5.6)	71.3 (2.2)	75.7 (0.8)
p-conductance ($p = 1, \epsilon = n$)	39.6 (0.3)	39.6 (0.3)	39.6 (0.3)	40.3 (0.3)	41.2 (0.3)
p-conductance ($p = 2, \epsilon = n$)	58.0 (12.1)	67.5 (7.5)	70.8 (4.9)	**72.4 (2.5)**	77.6 (0.6)
p-conductance ($p = \infty, \epsilon = n$)	48.0 (8.1)	56.5 (5.2)	57.0 (8.2)	62.6 (3.0)	72.9 (1.3)
PoissonMBO Calder et al. (2020)	54.9 (11.4)	65.3 (7.8)	68.2 (5.3)	69.9 (3.0)	74.8 (1.0)
p-conductance ($p = 2, \epsilon = 0$)	**58.7 (11.5)**	**67.5 (7.5)**	**70.8 (4.8)**	**72.4 (2.6)**	**77.8 (0.6)**