Connecting Graph Spectral Clustering and Partial Differential Equations

Max Daniels, Catherine Huang, Chloe Makdad, and Shubham Makharia

Summer@ICERM Final Presentation

August 4, 2020

Overview

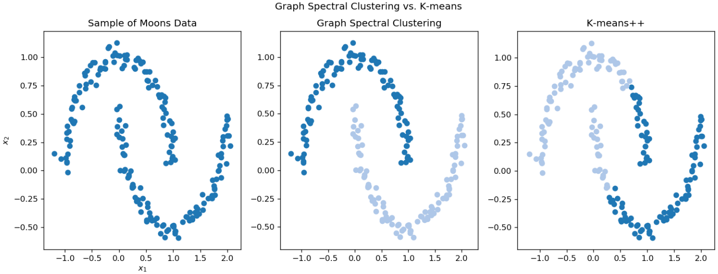
High Dimensional Data

Given a dataset $X \subset \mathbb{R}^l$ containing points $(x_1, \ldots, x_l)$, where l is quite large.

When the dimension of data is high, clustering becomes a challenge.

High Dimensional Data

Global measurements like distance become less informative with increasing dimension. Consider the Two Moons data set:



What's the Deal with Spectral Clustering?

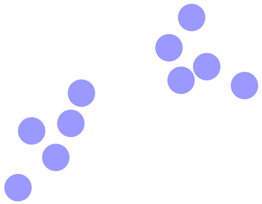
Why does spectral clustering work on the Two Moons data set, while k-means fails? Spectral clustering utilizes a dimension reduction technique that rewards "local linearity."

Roughly, a locally linear embedding is a map that looks linear when considering points sufficiently close to each other.

How do we find this embedding?

What's the Deal with Spectral Clustering?

Say we have some data we want to cluster:



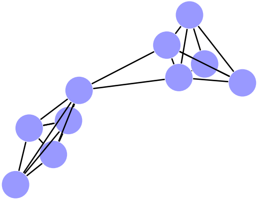
What's the Deal with Spectral Clustering?

Using some similarity metric, we construct a similarity matrix...

What's the Deal with Spectral Clustering?

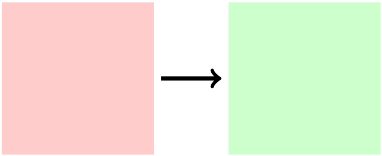
Using some similarity metric, we construct a similarity matrix...

...which allows us to construct a similarity graph

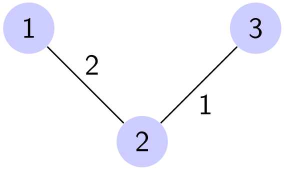


What's the Deal with Spectral Clustering?

Using this similarity matrix, we construct our graph Laplacian.



What's the Deal with Spectral Clustering?



$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad W = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

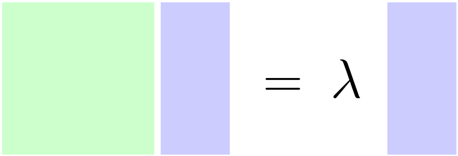
$$L := D - W = \begin{bmatrix} 2 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

In this presentation, we will explain the connection of the graph Laplacian to the broader idea of locally linear embeddings through physical interpretations of clustering.

What's the Deal with Spectral Clustering?

We can refactor the Laplacian with respect to a fixed number k, and then work with a smaller matrix ($n \times k$).

We find the eigenvectors corresponding to the first k eigenvalues of the Lapalcian, where k should correspond to the number of clusters.

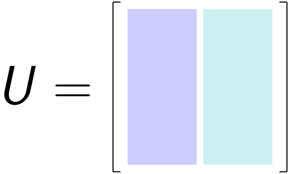


What's the Deal with Spectral Clustering?

We specifically want the k eigenvectors corresponding to the k smallest eigenvalues

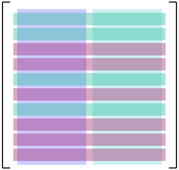
What's the Deal with Spectral Clustering?

We then concatenate our eigenvectors to form the matrix U



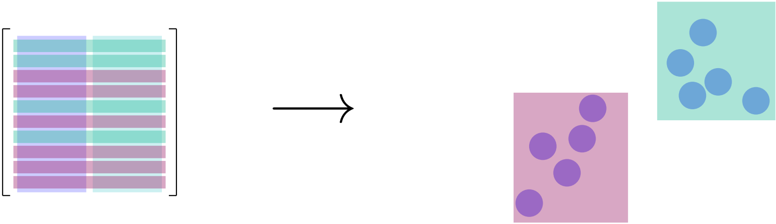
What's the Deal with Spectral Clustering?

Running k-means on the rows of U, we form our clusters.

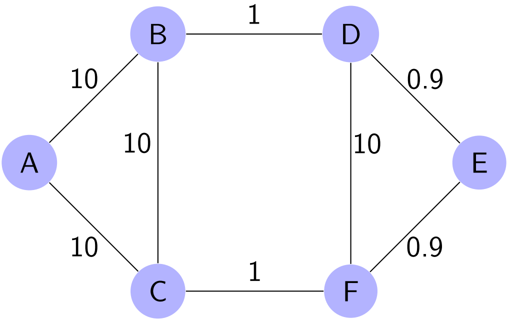
$U =$ 

What's the Deal with Spectral Clustering?

Since each row of the matrix U corresponds to a data point, we get our clusters!



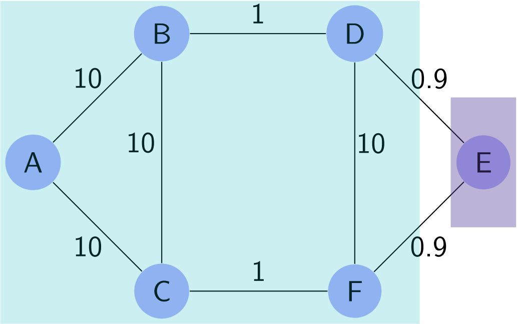
Ratio Cut vs. Min Cut



Ratio Cut vs. Min Cut

Definition (Min Cut)

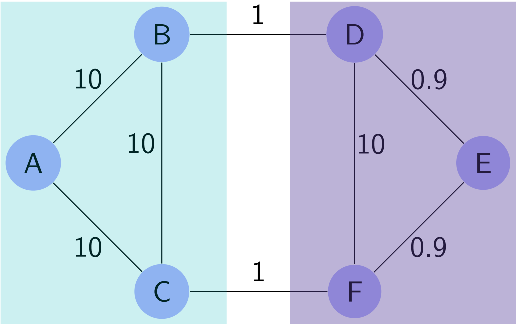
A partition of the vertices of a graph into disjoint subsets such that the sum of the weights going between the two sets is the smallest possible.



Ratio Cut vs. Min Cut

Definition (Ratio Cut)

A partition of the vertices of a graph into disjoint subsets such that the size of each cluster is taken into consideration.



Posing Ratio Cut as an Optimization Problem

- A, B are sets that contain the nodes in each respective cluster.
- The ratio cut problem can be expressed as an optimization problem.

$$\text{Ratio Cut } = \min_{A,B} \left(\sum_{i \in A, j \in B} w_{ij} \right) \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

- When w_{ij} is large, we want nodes i and j in the same cluster. This formulation also prioritizes clusters with roughly equal sizes.

How does this relate to the Laplacian?

- We'll show that $\min_{\boldsymbol{f}} \boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \text{RatioCut}$
- We define an indicator vector $\boldsymbol{f}$ where

$$f_i = \begin{cases} \sqrt{\frac{|A|}{|B|}}, & \text{node i} \in \text{A} \\ -\sqrt{\frac{|B|}{|A|}}, & \text{node i} \in \text{B} \end{cases}$$

- Notice one is positive and the other is negative. Normalization makes the math work out but don't worry too much about it!

Ratio Cut in terms of the Laplacian matrix

$$\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \sum_{(i,j) \in E} w_{ij} (f_i - f_j)^2$$

Ratio Cut in terms of the Laplacian matrix

$$\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \sum_{(i,j)\in E} w_{ij}(f_i - f_j)^2$$

$$= \sum_{i\in A, j\in B} w_{ij} \left(\sqrt{\frac{|A|}{|B|}} - \left(-\sqrt{\frac{|B|}{|A|}} \right) \right)^2$$

Ratio Cut in terms of the Laplacian matrix

$$\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \sum_{(i,j)\in E} w_{ij}(f_i - f_j)^2$$

$$= \sum_{i\in A, j\in B} w_{ij} \left(\sqrt{\frac{|A|}{|B|}} - \left(-\sqrt{\frac{|B|}{|A|}} \right) \right)^2$$

$$= \left(\sum_{i\in A, j\in B} w_{ij} \right) \left(\frac{|A|}{|B|} + \frac{|B|}{|A|} + 2 \right)$$

Ratio Cut in terms of the Laplacian matrix

$$\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \sum_{(i,j) \in E} w_{ij} (f_i - f_j)^2$$

$$= \sum_{i \in A, j \in B} w_{ij} \left(\sqrt{\frac{|A|}{|B|}} - \left(-\sqrt{\frac{|B|}{|A|}} \right) \right)^2$$

$$= \left(\sum_{i \in A, j \in B} w_{ij} \right) \left(\frac{|A|}{|B|} + \frac{|B|}{|A|} + 2 \right)$$

$$= \left(\sum_{i \in A, j \in B} w_{ij} \right) \left(\frac{|A| + |B|}{|A|} + \frac{|A| + |B|}{|B|} \right)$$

This is the Ratio Cut!

Eigenvector Problem!

- We said $\boldsymbol{f}$ is an indicator vector with 2 options.
- Solving this minimization problem is computationally intractable for large matrices L.
- *Relax:* Remove restrictions on $\boldsymbol{f}$

$$\min_{\boldsymbol{f}} \boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f}, \text{an eigenvector problem!}$$

- This is why we want the eigenvectors corresponding to the smallest eigenvalues.

PDEs involving the Laplace Operator

Linear PDEs

$$\text{Laplace Equation: } \Delta u = 0$$

$$\text{Poisson Equation: } \Delta u = f$$

$$\text{Wave Equation: } \frac{\partial^2 u}{\partial t^2} = c^2 \Delta u$$

$$\text{Heat Equation: } \frac{\partial u}{\partial t} = c \Delta u$$

$$\text{Schrodinger Equation: } \frac{\partial u}{\partial t} = ic \Delta u$$

Nonlinear PDEs

$$\text{Klein-Gordon equation: } \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} + c' u = \Delta u$$

$$\text{Calabi-Flow: } \frac{\partial g_{ij}}{\partial t} = (\Delta R) g_{ij}$$

The Heat and Wave Equations

Spectral clustering has been investigated through the lens of both the heat and wave equations:

Definition (1D Heat Equation)

The one-dimensional heat equation is defined as

$$\frac{\partial}{\partial t} u(x, t) = \frac{\partial^2}{\partial x^2} u(x, t)$$

where $u(x, t)$ outputs the temperature at position x and time t.

Definition (1D Wave Equation)

The one-dimensional wave equation is defined as

$$\frac{\partial^2}{\partial t^2} u(x, t) = \frac{\partial^2}{\partial x^2} u(x, t)$$

where $u(x, t)$ outputs the displacement at position x and time t.

The Heat and Wave Equations

Recall that in one dimension, the Laplacian $\Delta = \frac{\partial^2}{\partial x^2}$, so we can rewrite the heat equation as

$$\frac{\partial}{\partial t} u(x,t) = \Delta u(x,t)$$

and the wave equation as

$$\frac{\partial^2}{\partial t^2} u(x,t) = \Delta u(x,t).$$

Look at these Laplacians!

We must be able to make a connection to graph spectral clustering...

Focus: The Heat Equation

Let's further investigate the heat equation.

Focus: The Heat Equation

Let's further investigate the heat equation.

We'll focus on these questions:

1. How can we think about the heat equation in physical terms?
2. How do we solve the heat equation?
3. How do we discretize the heat equation?

Physical Interpretation

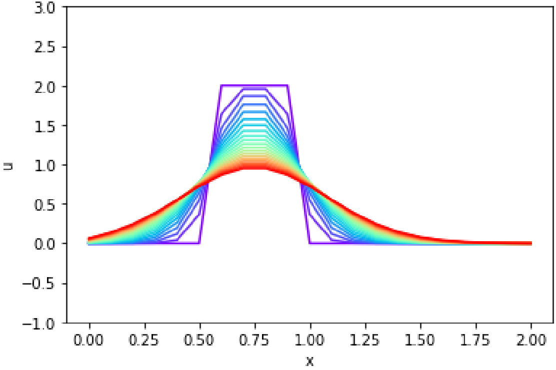
The quintessential interpretation of the one-dimensional heat equation is to think of a heated rod.

Consider what happens if we dip the ends of the rod in cold water.

Physical Interpretation

The quintessential interpretation of the one-dimensional heat equation is to think of a heated rod.

Consider what happens if we dip the ends of the rod in cold water.



Solving the Heat Equation

Now that we have some physical intuition, let's work on solving the heat equation.

Solving the Heat Equation

Now that we have some physical intuition, let's work on solving the heat equation. Setup:

$$\begin{aligned} \text{Find } u(x,t) : [0,1] \to \mathbb{R} \text{ such that:} \quad & \frac{\partial u}{\partial t} = \Delta u(x,t) \\ \text{Boundary Condition:} \quad & u(0,t) = \alpha \qquad u(1,t) = \beta \\ \text{Initial Condition:} \quad & u(x,0) = f(x) \end{aligned}$$

Solving the Heat Equation

Now that we have some physical intuition, let's work on solving the heat equation. Setup:

$$\text{Find } u(x,t) : [0,1] \to \mathbb{R} \text{ such that:} \quad \frac{\partial u}{\partial t} = \Delta u(x,t)$$

$$\text{Boundary Condition:} \quad u(0,t) = \alpha \qquad u(1,t) = \beta$$

$$\text{Initial Condition:} \quad u(x,0) = f(x)$$

Definition (Boundary Condition)

Describes behavior in space at the boundary of the domain.

Definition (Initial Condition)

Describes the initial behavior in space at time $t = 0$.

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x, t)$

We can solve the heat equation by separation of variables.

Assume **Dirichlet Boundary Conditions**:

$$u(0, t) = u(1, t) = 0$$

With initial conditions $u(x, 0) = f(x)$. Then, assume

$$u(x, t) = X(x)T(t)$$

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x,t)$

We can solve the heat equation by separation of variables.

Assume **Dirichlet Boundary Conditions**:

$$u(0,t) = u(1,t) = 0$$

With initial conditions $u(x,0) = f(x)$. Then, assume

$$u(x,t) = X(x)T(t)$$

Plugging this into the heat equation, we get

$$XT' - X''T = 0$$

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x, t)$

By way of algebra, we can write

$$XT' - X''T = 0$$

as

$$\frac{T'}{T} = \frac{X''}{X}$$

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x, t)$

By way of algebra, we can write

$$XT' - X''T = 0$$

as

$$\frac{T'}{T} = \frac{X''}{X}$$

and define

$$-\lambda := \frac{X''}{X}$$

for some constant λ.

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x,t)$

This gives us a system of ODEs,

$$T' = -\lambda T$$
$$X'' = -\lambda X$$

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x,t)$

Under Dirichlet Boundary conditions, these equations have special solutions:

$$X''(x) = -\lambda X(x) \implies X(x) = e^{-i\omega_n x} \qquad \omega_n = \sqrt{\lambda} = n\pi$$

$$T'(t) = -\lambda T(t) \implies T(t) = e^{-\omega_n^2 t}$$

Solving the Heat Equation $\frac{\partial u}{\partial t} = \Delta u(x,t)$

Under Dirichlet Boundary conditions, these equations have special solutions:

$$X''(x) = -\lambda X(x) \implies X(x) = e^{-i\omega_n x} \qquad \omega_n = \sqrt{\lambda} = n\pi$$
$$T'(t) = -\lambda T(t) \implies T(t) = e^{-\omega_n^2 t}$$

Moreover, any linear combination is also solution. Given initial conditions, this implies:

$$f(x) = u(x,0) = \sum_{n\in\mathbb{Z}} a_n e^{-i\omega_n x}$$
$$\implies u(x,t) = \sum_{n\in\mathbb{Z}} a_n e^{-\omega_n^2 t} e^{-i\omega_n x}$$

Discretization of the Heat Equation

"But Chloe, graph spectral clustering is a discrete problem."

Discretization of the Heat Equation

Right! Let's think about discretizing the heat equation. Once again consider our heated rod:

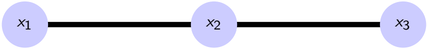
Discretization of the Heat Equation

Right! Let's think about discretizing the heat equation. Once again consider our heated rod:

What if instead of solving the heat equation analytically, we approximated solutions using Euler's method for the time operator and a finite difference method for space.

Discretization of the Heat Equation

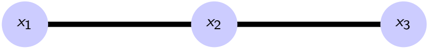
Right! Let's think about discretizing the heat equation. Once again consider our heated rod:



What if instead of solving the heat equation analytically, we approximated solutions using Euler's method for the time operator and a finite difference method for space.

This allows us to discretize the domain of u!

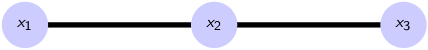
Discretization of the Heat Equation



We can write

$$u(x_2, t+1) = u(x_2, t) + (u(x_3, t) - 2u(x_2, t) + u(x_1, t))$$

Discretization of the Heat Equation

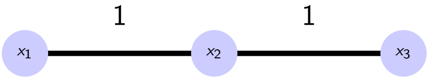


We can write

$$u(x_2, t+1) = u(x_2, t) + \boxed{(\mathrm{u}(\mathrm{x}_3, t) - 2u(x_2, t) + u(x_1, t))}$$

The part of the equation that is boxed corresponds to the negative Laplacian matrix:

Discretization of the Heat Equation



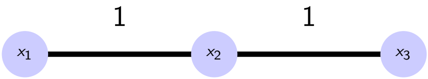
We can write

$$u(x_2, t+1) = u(x_2, t) + \boxed{(u(x_3, t) - 2u(x_2, t) + u(x_1, t))}$$

The part of the equation that is boxed corresponds to the negative Laplacian matrix:

$$-L = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

Discretization of the Heat Equation



We can write

$$u(x_2, t+1) = u(x_2, t) + \boxed{(\mathrm{u}(x_3, t) - 2u(x_2, t) + u(x_1, t))}$$

The part of the equation that is boxed corresponds to the negative Laplacian matrix:

So,

$$\begin{aligned} u(x_2, t+1) &= u(x_2, t) - Lu(x_2, t) \\ &= (I - L)u(x_2, t) \end{aligned}$$

How can we use diffusion for clustering?

Solution of heat equation looks ugly, but has a nice interpretation:

$$u(x,t) = \sum_{n \in \mathbb{Z}} a_n e^{-\omega_n^2 t} e^{-i\omega_n x}$$

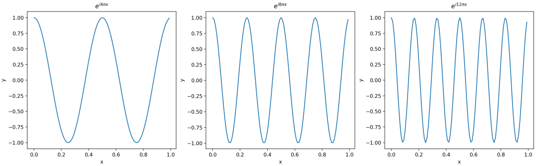


Figure: Real part of complex exponentials

Takeaway: $u(x,t)$ has a decomposition in *locally dissimilar* components. The spikier each component, the faster that component diffuses.

How can we use diffusion for clustering?

More precisely, high frequency components of f are attenuated exponentially faster than low frequency parts of f.

Animation

How can we use diffusion for clustering?

More precisely, high frequency components of u are attenuated exponentially faster than low frequency parts of u.

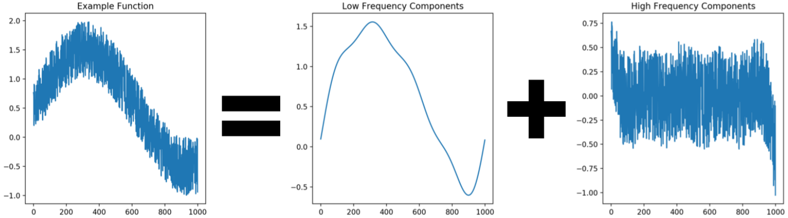


Figure: High and low frequency components of a function.

How do low frequencies help in clustering?

Visual Example

We can watch the role of low frequencies in imaging:

Figure: "no no Shubham you gotta pay *us*" - Rishi

Visual Example

Here we zero out high frequencies, while the heat equation attenuates them gradually.

Figure: Low frequency components of Rishi.jpg

Visual Example

Here we zero out high frequencies, while the heat equation attenuates them gradually.

Figure: Low frequency components of Rishi.jpg

White areas are hot, black areas are cold.
Diffusion from the foreground to the background takes a long time.

Clustering via Low Frequencies

How can we interpret clustering as a diffusion process?

- A good clustering puts nodes in groups that *retain their heat* for a long time.
- Equivalently: heat from inside a cluster diffuses slowly outside the cluster.

Why should we care about small eigenvectors of the graph Laplacian?

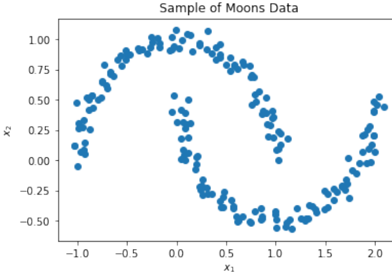
- Small eigenvectors represent heat distributions that diffuse as slowly as possible.

Summary

- Different interpretations of graph spectral clustering
  - Local linear embedding
  - Ratio cut minimization problem
  - Heat diffusion on a graph

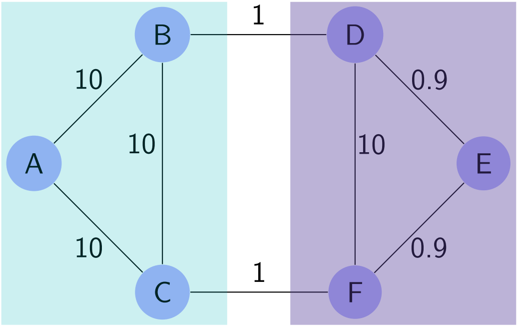
Summary

- Different interpretations of graph spectral clustering
  - Local linear embedding
  - Ratio cut minimization problem
  - Heat diffusion on a graph



Summary

- Different interpretations of graph spectral clustering
  - Local linear embedding
  - Ratio cut minimization problem
  - Heat diffusion on a graph



Summary

- Different interpretations of graph spectral clustering
  - Local linear embedding
  - Ratio cut minimization problem
  - Heat diffusion on a graph

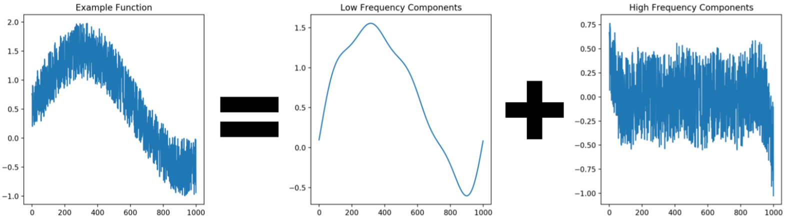


Figure: Heat diffusion preserves low frequencies.

Thank You!

Thank you to all the organizers and TAs: Akil Narayan, Minah Oh, Alexandru Mihai, Liu Yang, all others at ICERM, and our fellow student researchers. It has been a great (virtual) summer!

Thank You!

Thank you to all the organizers and TAs: Akil Narayan, Minah Oh, Alexandru Mihai, Liu Yang, all others at ICERM, and our fellow student researchers. It has been a great (virtual) summer!

Any questions?

References

- Mikhail Belkin and Partha Niyogi. "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation". In: *Neural Computation* 15.6 (2003), pp. 1373–1396. DOI: 10.1162/089976603321780317. URL: http://www.mitpressjournals.org/doi/10.1162/089976603321780317.
- Matthew J Hancock. "18.303 Linear Partial Differential Equations". en. In: (), p. 44.
- Ulrike von Luxburg. "A tutorial on spectral clustering". In: *Statistics and Computing* 17.4 (Dec. 2007). arXiv: 0711.0189 [cs.DS], pp. 395–416. DOI: 10.1007/s11222-007-9033-z. URL: https://doi.org/10.1007/s11222-007-9033-z.