Active Spectral Clustering via Iterative Uncertainty Reduction

Fabian L. Wauthier, UC Berkeley

with

Nebojsa Jojic, Microsoft Research
Michael I. Jordan, UC Berkeley

15th of August 2012
Spectral Clustering

- Cluster data using only pairwise similarities.
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$. 
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$.
2. Cluster by thresholding at 0.
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$.
2. Cluster by thresholding at 0.

Data
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$.
2. Cluster by thresholding at 0.
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$.
2. Cluster by thresholding at 0.
Spectral Clustering

- Cluster data using only pairwise similarities.
  1. Using similarities, embed datapoints in $\mathbb{R}^1$.
  2. Cluster by thresholding at 0.
Spectral Clustering

- Cluster data using only pairwise similarities.

1. Using similarities, embed datapoints in $\mathbb{R}^1$.
2. Cluster by thresholding at 0.
Spectral Clustering with Missing Data

**Problem:**
- Similarities often expensive or noisy.
Spectral Clustering with Missing Data

Problem:
- Similarities often expensive or noisy.
- E.g., Clustering photos by location.
Spectral Clustering with Missing Data

**Problem:**
- Similarities often expensive or noisy.
- E.g., Clustering photos by location.

- Requires similarities reflecting co-location: $w_{ij} = p(i \sim j) = p(i \text{ and } j \text{ were taken in same room})$
- Need human annotation. Expensive/noisy. Does not scale!
Spectral Clustering with Missing Data

Problem:

- Similarities often expensive or noisy.
- E.g., Clustering photos by location.

- Requires similarities reflecting co-location: e.g.

\[ w_{ij} = p(i \sim j) = p(i \text{ and } j \text{ were taken in same room}) \]
Spectral Clustering with Missing Data

Problem:

▶ Similarities often expensive or noisy.
▶ E.g., Clustering photos by location.

Requires similarities reflecting co-location: e.g.

\[ w_{ij} = p(i \sim j) = p(i \text{ and } j \text{ were taken in same room}) \]

▶ Need human annotation. Expensive/noisy. Does not scale!
Spectral Clustering with Missing Data

Problem:
- Similarities often expensive or noisy.
- E.g., Clustering proteins by binding properties.
Spectral Clustering with Missing Data

**Problem:**

- Similarities often expensive or noisy.
- E.g., Clustering proteins by binding properties.

```latex
\text{Binds to A} \quad \text{Binds to B}
```

Images: Liu et al.
Spectral Clustering with Missing Data

Problem:
- Similarities often expensive or noisy.
- E.g., Clustering proteins by binding properties.

\[ w_{ij} = p(i \sim j) = p(i \text{ and } j \text{ bind to similar targets}) \]
Spectral Clustering with Missing Data

**Problem:**
- Similarities often expensive or noisy.
- E.g., Clustering proteins by binding properties.

\[ p(i \sim j) = p(i \text{ and } j \text{ bind to similar targets}) \]

- Requires similarities reflecting co-binding: e.g.

Images: Liu et al.

- Need experimental data. Expensive/noisy. Does not scale!
Spectral Clustering with Missing Data

- Subsample and impute missing with 0 (e.g. Shamir et al.).
- But performance poor if missing at random.
Spectral Clustering with Missing Data

- Subsample and impute missing with 0 (e.g. Shamir et al.).
- But performance poor if missing at random.
Spectral Clustering with Missing Data

- Subsample and impute missing with 0 (e.g. Shamir et al.).
- But performance poor if missing at random.

**Similarities**

**Embedding**
Spectral Clustering with Missing Data

▶ Subsample and impute missing with 0 (e.g. Shamir et al.).
▶ But performance poor if missing at random.

This talk: Actively measure similarities to minimize measurement cost while achieving good clustering performance.
Overview

Introduction

Spectral Embeddings

Active Clustering

Active Clustering with Noisy Similarities
Overview

Introduction

Spectral Embeddings

Active Clustering

Active Clustering with Noisy Similarities
Spectral Embeddings

\[
\begin{align*}
&\text{Similarities} \quad \Rightarrow \quad \text{Embedding} \\
\end{align*}
\]
Spectral Embeddings

- Given similarity matrix $W$, define $L = \text{diag}(W\mathbf{1}) - W$.

$\Rightarrow$

Similarities

Embedding

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$
Spectral Embeddings

Given similarity matrix $W$, define $L = \text{diag}(W1) - W$.

Embedding $v^*$ solves:

$$v^* = \arg\min v \, v^T L v$$

s.t. $v^T v = 1 \quad v^T 1 = 0$.

Can use formulation to guide active learning.

If $v^*$ unstructured, can maximize "overall" change to $v^*$ (e.g. Shamir and Tishby).

But data clusters, so $v^*$ structured. Can do better!
Spectral Embeddings

Given similarity matrix \( W \), define \( L = \text{diag}(W1) - W \).

Embedding \( v^* \) solves:

\[
  v^* = \arg\min_v v^T Lv = \arg\min_v \sum_{ij} w_{ij} (v(i) - v(j))^2
\]

s.t. \( v^T v = 1 \) \( v^T 1 = 0 \).
Spectral Embeddings

Given similarity matrix $W$, define $L = \text{diag}(W\mathbf{1}) - W$.

Embedding $v^*$ solves:

$$v^* = \arg\min_v v^T L v = \arg\min_v \sum_{ij} w_{ij} (v(i) - v(j))^2$$

s.t. $v^T v = 1 \quad v^T \mathbf{1} = 0$.

Can use formulation to guide active learning.
Given similarity matrix $W$, define $L = \text{diag}(W \mathbf{1}) - W$.

Embedding $v^*$ solves:

$$v^* = \arg\min_v v^T L v = \arg\min_v \sum_{ij} w_{ij} (v(i) - v(j))^2$$

s.t. $v^T v = 1$  $v^T \mathbf{1} = 0$.

Can use formulation to guide active learning.

If $v^*$ unstructured, can maximize “overall” change to $v^*$ (e.g. Shamir and Tishby).
Spectral Embeddings

Given similarity matrix $W$, define $L = \text{diag}(W\mathbf{1}) - W$.

Embedding $v^*$ solves:

$$v^* = \arg\min_v v^T L v = \arg\min_v \sum_{ij} w_{ij} (v(i) - v(j))^2$$

s.t. $v^T v = 1$, $v^T \mathbf{1} = 0$.

- Can use formulation to guide active learning.
- If $v^*$ unstructured, can maximize “overall” change to $v^*$ (e.g. Shamir and Tishby).
- But data clusters, so $v^*$ structured. Can do better!
Spectral Embeddings

- We threshold $\nu^*$ at 0.

Clustering is "uncertain" for points near threshold.

Idea:
- Actively measure similarities to reduce uncertainty.
Spectral Embeddings

- We threshold $v^*$ at 0.
- If for some $i$, $\forall j, w_{ij} \approx c > c_0$, then $v^*(i) \approx \frac{\sum_{j \neq i} v^*(j)}{n-1} \approx 0$
Spectral Embeddings

- We threshold $v^*$ at 0.
- If for some $i$, $\forall j, w_{ij} \approx c > c_0$, then $v^*(i) \approx \frac{\sum_{j \neq i} v^*(j)}{n-1} \approx 0$

Similarities

$\Rightarrow$

Embedding
Spectral Embeddings

- We threshold $v^*$ at 0.
- If for some $i$, $\forall j, w_{ij} \approx c > c_0$, then $v^*(i) \approx \frac{\sum_{j \neq i} v^*(j)}{n-1} \approx 0$

![Diagram showing similarities and embedding]
We threshold \( v^* \) at 0.

If for some \( i \), \( \forall j, w_{ij} \approx c > c_0 \), then
\[
v^*(i) \approx \frac{\sum_{j \neq i} v^*(j)}{n-1} \approx 0
\]

Clustering is “uncertain” for points near threshold.
Spectral Embeddings

- We threshold $v^*$ at 0.
- If for some $i$, $\forall j, w_{ij} \approx c > c_0$, then $v^*(i) \approx \frac{\sum_{j \neq i} v^*(j)}{n-1} \approx 0$

Clustering is “uncertain” for points near threshold.

Idea:
- **Actively** measure similarities to reduce uncertainty.
Overview

Introduction

Spectral Embeddings

Active Clustering

Active Clustering with Noisy Similarities
Active Clustering

- Iteratively measure similarities which can most change the embedding near the threshold.
Active Clustering

- Iteratively measure similarities which can most change the embedding near the threshold.

![Similarities](image1)

![Embedding](image2)
Active Clustering

Iteratively measure similarities which can most change the embedding near the threshold.

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
\( v_2 \)

Similarities

Embedding
Active Clustering

Iteratively measure similarities which can most change the embedding near the threshold.

Similarities

\[
\left| \frac{d \star}{dw_{1,4}} \right|_0 = \text{small}
\]

Embedding

SMALL change

Perturb similarity \( w_{1,4} \)
Active Clustering

- Iteratively measure similarities which can most change the embedding near the threshold.

\[
\frac{d \star}{dw_{1,4}} \bigg|_0 = \text{small}
\]
Active Clustering

Iteratively measure similarities which can most change the embedding near the threshold.

Similarities

\[
\left| \frac{d \star}{dw_{1,4}} \right|_0 = \text{small} \quad \left| \frac{d \star}{dw_{2,8}} \right|_0 = \text{medium}
\]
Active Clustering

Iteratively measure similarities which can most change the embedding near the threshold.

\[ \frac{d \star}{dw_{1,4}} |_0 = \text{small} \]
\[ \frac{d \star}{dw_{2,8}} |_0 = \text{medium} \]
Active Clustering

Iteratively measure similarities which can most change the embedding near the threshold.

\[
\begin{align*}
|\frac{d \star}{dw_{1,4}}|_0 &= \text{small} \\
|\frac{d \star}{dw_{2,8}}|_0 &= \text{medium} \\
|\frac{d \star}{dw_{8,11}}|_0 &= \text{large}
\end{align*}
\]
Active Clustering

- Iteratively measure similarities which can most change the embedding near the threshold.

\[ \text{Similarities} \]

\[
\begin{align*}
\left| \frac{d\star}{dw_{1,4}} \right|_0 &= \text{small} & \left| \frac{d\star}{dw_{2,8}} \right|_0 &= \text{medium} & \left| \frac{d\star}{dw_{8,11}} \right|_0 &= \text{large}
\end{align*}
\]

- \( \Rightarrow \) Similarity \( w_{8,11} \) is most influential. Measure it next.
Matrix Perturbation Theory

- Recall spectral embedding:

\[ \nu^* = \arg\min_{\nu} \nu^T L \nu \]

\[ \text{s.t. } \nu^T \nu = 1 \quad \nu^T \mathbf{1} = 0. \]
Matrix Perturbation Theory

▶ Recall spectral embedding:

\[ v^* = \arg\min_v v^\top L v \]

\[ \text{s.t. } v^\top v = 1, \quad v^\top 1 = 0. \]

▶ Embedding \( v^* \) is eigenvector of \( L \).
Matrix Perturbation Theory

▶ Recall spectral embedding:

\[ \mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbf{v}^\top \mathbf{L} \mathbf{v} \]

\[ \text{s.t. } \mathbf{v}^\top \mathbf{v} = 1 \quad \mathbf{v}^\top \mathbf{1} = 0. \]

▶ Embedding \( \mathbf{v}^* \) is eigenvector of \( \mathbf{L} \).

▶ If \( k_{\min} = \arg\min_i |\mathbf{v}^*(i)| \), then

\[ \star = \mathbf{v}^*(k_{\min}) \]

\[ \frac{d \star}{dw_{ij}} \bigg|_0 = \frac{d \mathbf{v}^*(k_{\min})}{dw_{ij}} \bigg|_0. \]
Matrix Perturbation Theory

- Recall spectral embedding:

\[
\nu^* = \arg\min_{\nu} \nu^T L \nu \\
\text{s.t. } \nu^T \nu = 1, \quad \nu^T 1 = 0.
\]

- Embedding \(\nu^*\) is eigenvector of \(L\).

- If \(k_{\min} = \arg\min_i |\nu^*(i)|\), then

\[
\star = \nu^*(k_{\min})
\]

\[
\frac{d \star}{d w_{ij}} \bigg|_0 = \frac{d \nu^*(k_{\min})}{d w_{ij}} \bigg|_0.
\]

- Matrix Perturbation Theory:

\[
\frac{d \nu^*(k_{\min})}{d w_{ij}} \bigg|_0 = \sum_{p>2}^n \nu_2^\top \left[ \frac{\partial L}{\partial w_{ij}} \right] \nu_p \frac{\nu_p(k_{\min})}{\lambda_2 - \lambda_p}.
\]
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} | \frac{dv^*(k_{\text{min}})}{dw_{ij}} |_0 $.
3. Measure $w_{i^*, j^*}$ and add to $W$. 
Algorithm Sketch

Impute unobserved similarities in \( W \) with 0.

Iterate:

1. Compute embedding \( v^* \).
2. \((i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0 \).
3. Measure \( w_{i^*, j^*} \) and add to \( W \).

00 measurements out of 66

\[ \begin{align*}
\text{Similarities} & \quad \implies \quad \text{Embedding}
\end{align*} \]
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 

Simulations out of 66
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

**Iterate:**

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 

![Diagram showing similarities and embedding](image-url)
Algorithm Sketch

Impute unobserved similarities in $W$ with $0$.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$. 

![Similarities](image1)

![Embedding](image2)
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \frac{|d v^*(k_{\min})|}{|d w_{ij}|} 0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 

04 measurements out of 66

![Similarities](image)

⇒

![Embedding](image)
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 

![Diagram of similarities and embedding]
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \text{argmax}_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$.
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{d\nu^*(k_{min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

**Iterate:**

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

![Diagram of similarities and embedding]

08 measurements out of 66
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$. 
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$. 
3. Measure $w_{i^*, j^*}$ and add to $W$. 

09 measurements out of 66

![Similarities](image1)

![Embedding](image2)
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:
1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 

10 measurements out of 66

Similarities

Embedding
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \operatorname{argmax}_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^* j^*}$ and add to $W$. 

11 measurements out of 66

Similarities

Embedding
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.

2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.

3. Measure $w_{i^* \cdot j^*}$ and add to $W$.
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.

2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.

3. Measure $w_{i^*j^*}$ and add to $W$.

13 measurements out of 66
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$. 
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

16 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$.

17 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

18 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$.

19 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

**Iterate:**

1. Compute embedding $v^*$.  
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.  
3. Measure $w_{i^*, j^*}$ and add to $W$.

20 measurements out of 66  

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\mathbf{v}^*$.
2. $(i^*, j^*) = \underset{(i, j) \in \text{Unobserved}}{\text{argmax}} \left| \frac{d\mathbf{v}^*(k_{\text{min}})}{dW_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

21 measurements out of 66

Similarities

Embedding

Separates well w/ 23% of similarities
**Algorithm Sketch**

Impute unobserved similarities in $W$ with 0.

**Iterate:**

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$.

22 measurements out of 66

Similarities

Embedding

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

23 measurements out of 66

Separates well w/ 23% of similarities
Active Clustering

Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.

2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\min})}{dw_{ij}} \right|_0$.

3. Measure $w_{i^*, j^*}$ and add to $W$.

24 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

**Iterate:**

1. Compute embedding $v^*$. 
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^* (k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

25 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

26 measurements out of 66

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $v^*$.
2. $(i^*, j^*) = \arg\max_{(i, j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

27 measurements out of 66

Similarities

Embedding

Separates well w/ 23% of similarities
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.
2. $(i^*, j^*) = \text{argmax}_{(i,j) \in \text{Unobserved}} |\frac{dv^*(k_{\text{min}})}{dw_{ij}}|_0$.
3. Measure $w_{i^*, j^*}$ and add to $W$.

28 measurements out of 66

Separates well w/ 23% of similarities

Similarities

Embedding
Algorithm Sketch

Impute unobserved similarities in $W$ with 0.

Iterate:

1. Compute embedding $\nu^*$.
2. $(i^*, j^*) = \operatorname{argmax}_{(i,j) \in \text{Unobserved}} \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0$.
3. Measure $w_{i^*j^*}$ and add to $W$.

29 measurements out of 66

Similarities  \[\Rightarrow\]  Embedding

Separates well w/ 23% of similarities
Active Clustering

Application: Clustering Photos by Location

Cluster 100 photos into kitchen/living room.

- Similarities $w_{ij} = p(i \sim j)$ are HITs on Mechanical Turk: How likely is it the photos were taken in the same room?
- Similarities are median of three noisy HIT responses.
- Can cluster with all similarities. Expensive (US$222)!
Application: Clustering Photos by Location

- Cluster 100 photos into kitchen/living room.
- Similarities $w_{ij} = p(i \sim j)$ are HITs on Mechanical Turk:
  How likely is it the photos were taken in the same room?
Application: Clustering Photos by Location

Cluster 100 photos into kitchen/living room.

Similarities $w_{ij} = p(i \sim j)$ are HITs on Mechanical Turk:
**How likely is it the photos were taken in the same room?**

Similarities are median of three noisy HIT responses.
Active Clustering

Application: Clustering Photos by Location

- Cluster 100 photos into kitchen/living room.
- Similarities $w_{ij} = p(i \sim j)$ are HITs on Mechanical Turk: How likely is it the photos were taken in the same room?
- Similarities are median of three noisy HIT responses.
- Can cluster with all similarities. Expensive (US$222)!
Active Clustering

Application: Clustering Photos by Location

- Proportion $p$ means $p \binom{n}{2}$ of $3 \binom{n}{2}$ HIT responses used.

- S&T need US$32 to reach error rate 0.05. We need US$17!
Overview

Introduction

Spectral Embeddings

Active Clustering

Active Clustering with Noisy Similarities
Noisy Similarities

- Can use median of repeat measurements to reduce noise.
- Influence of noise on embedding captured by \[ \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0 \].
  \[ \Rightarrow \text{Only need to know influential similarities accurately.} \]
Noisy Similarities

- Can use median of repeat measurements to reduce noise.
- Influence of noise on embedding captured by \( \left| \frac{dv^*(k_{\text{min}})}{dw_{ij}} \right|_0 \).
  \[ \Rightarrow \text{Only need to know influential similarities accurately.} \]

At each measurement have choice. Can reduce

- noise by re-measuring similarities.
- cluster uncertainty by measuring most influential similarity.
Noisy Similarities

- Can use median of repeat measurements to reduce noise.
- Influence of noise on embedding captured by \( \left| \frac{d v^*(k_{\text{min}})}{d w_{ij}} \right|_0 \).
  - Only need to know influential similarities accurately.

At each measurement have choice. Can reduce
- noise by re-measuring similarities.
- cluster uncertainty by measuring most influential similarity.

Idea:

- Measure the most uncertain and influential similarity.
- Augment algorithm to estimate:
  - running medians \( \bar{w}_{ij} \) of similarities
  - standard deviations \( \sigma_{ij} \) of medians (frequentist or Bayesian)

- Measure similarity \( w_{ij} \) where \( \left| \frac{d v^*(k_{\text{min}})}{d w_{ij}} \right|_{\bar{w}_{ij}} \cdot \sigma_{ij} \) is largest
Active Clustering with Noisy Similarities

- Measure the most uncertain and influential similarity.
Active Clustering with Noisy Similarities

- Measure the most **uncertain** and **influential** similarity.

Similarities

Embedding
Active Clustering with Noisy Similarities

- Measure the most **uncertain** and **influential** similarity.

Similarities

Embedding
Active Clustering with Noisy Similarities

- Measure the most **uncertain** and **influential** similarity.

\[
\begin{align*}
\frac{d \sigma}{dw_{1,4}} |_{\bar{w}_{1,4}} = \text{small} \\
\sigma_{1,4} = \text{small}
\end{align*}
\]
Active Clustering with Noisy Similarities

- Measure the most **uncertain** and **influential** similarity.

\[ \frac{d \star \bar{w}_{1,4}}{d w_{1,4}} \mid \sigma_{1,4} = \text{small} \]
Active Clustering with Noisy Similarities

Measure the most **uncertain** and **influential** similarity.

Perturb $w_{8,11}$

$\sigma_{8,11} = \text{med.}$

$\sigma_{1,4} = \text{small}$

$\sigma_{2,8} = \text{large}$
Active Clustering with Noisy Similarities

Measure the most **uncertain** and **influential** similarity.

![Similarities](image1)

![Embedding](image2)

\[
\begin{align*}
\frac{d \star}{d w_{1,4}} \quad & \quad \bar{w}_{1,4} \\
\sigma_{1,4} = \text{small} \\
\frac{d \star}{d w_{8,11}} \quad & \quad \bar{w}_{8,11} \\
\sigma_{8,11} = \text{med.}
\end{align*}
\]
Active Clustering with Noisy Similarities

▶ Measure the most **uncertain** and **influential** similarity.

Perturb $w_{2,8}$

LARGE uncertainty

$\sigma_{1,4} = \text{small}$

$\sigma_{8,11} = \text{med.}$

$\sigma_{2,8} = \text{large}$

$\frac{d \sigma_{1,4}}{d w_{1,4}} \approx 0.2, \quad \frac{d \sigma_{8,11}}{d w_{8,11}} \approx 0.1$
Active Clustering with Noisy Similarities

- Measure the most **uncertain** and **influential** similarity.

\[ \text{Perturb } w_{2,8} \]

\[ \text{LARGE uncertainty} \]

\[ \begin{align*}
\frac{d}{dw_{1,4}} \bar{w}_{1,4} & \quad \sigma_{1,4} = \text{small} \\
\frac{d}{dw_{8,11}} \bar{w}_{8,11} & \quad \sigma_{8,11} = \text{med.} \\
\frac{d}{dw_{2,8}} & \quad \sigma_{2,8} = \text{large}
\end{align*} \]

- \[ \Rightarrow w_{2,8} \text{ most uncertain and influential. Measure it next.} \]
Application: Clustering Photos by Location

Proportion $p$ means $p \binom{n}{2}$ of $3 \binom{n}{2}$ HIT responses used.
Application: Clustering Photos by Location

- Proportion $p$ means $p \binom{n}{2}$ of $3 \binom{n}{2}$ HIT responses used.

![Graph showing error rate vs. proportion of HITs]
Questions?