# Semi-supervised manifold alignment with Diffusion Transport

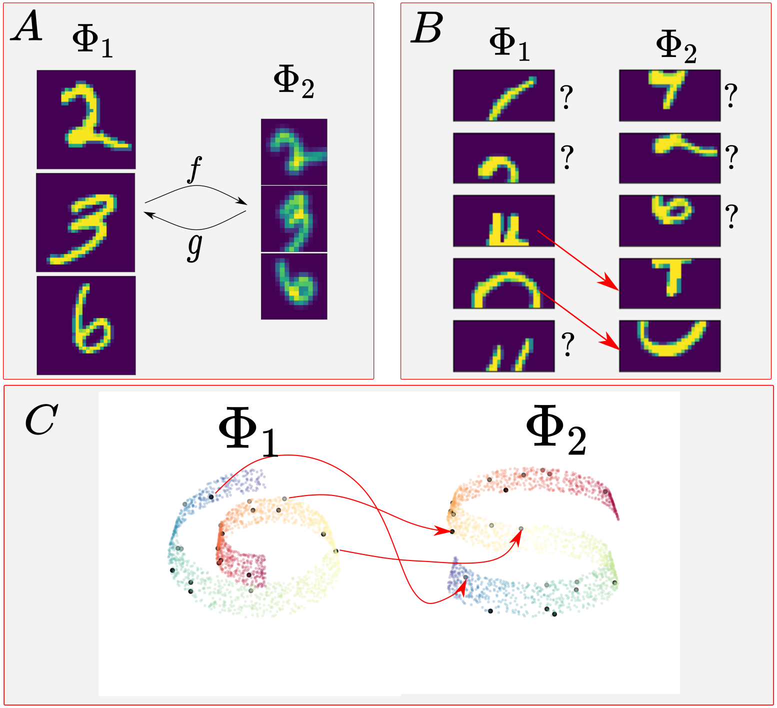
Andres Duque
Advisor: Dr. Kevin moon

Utah State University
Math & Stats department

April 8, 2022

# What are we looking for?

The integration of multimodal data presents a challenge in cases when the study of a given phenomena by different instruments or conditions generates distinct but related domains ($\Phi_1$, $\Phi_2$).



# State of the art

## Semi-supervised Manifold Alignment (SSMA) Ham et al. 2005

1. Find the mappings $f : \Phi_1 \rightarrow \mathcal{S}$ and $g : \Phi_2 \rightarrow \mathcal{S}$, ($\mathcal{S}$ is a common embedding space) by minimizing:

$$C(f, g) = u \sum_{i \in \mathcal{C}} ||f_i - g_i||^2 + f^T \mathcal{L}_{\Phi_1} f + g^T \mathcal{L}_{\Phi_2} g \qquad (1)$$

# State of the art

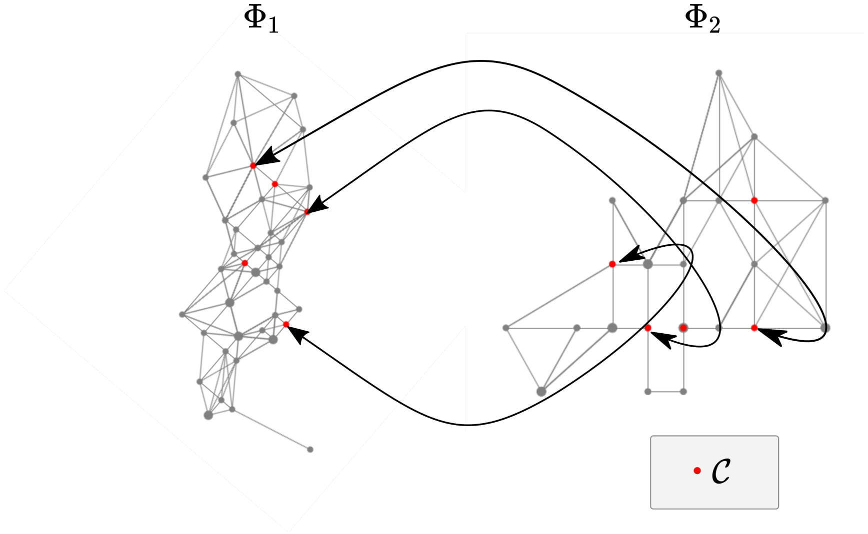
## Manifold alignment Procrustes approach - C. Wang et al. 2008

1. Compute $\mathcal{L}_{\Phi_1}$ and $\mathcal{L}_{\Phi_2}$.
2. Find their eigenvectors as embedding coordinates.
3. Use Procrustes analysis on the known correspondences to find a mapping from the embedding of one domain to the other.

# Diffusion Transport Alignment (DTA)

We propose to use an inter-domain diffusion process that connects both domains.

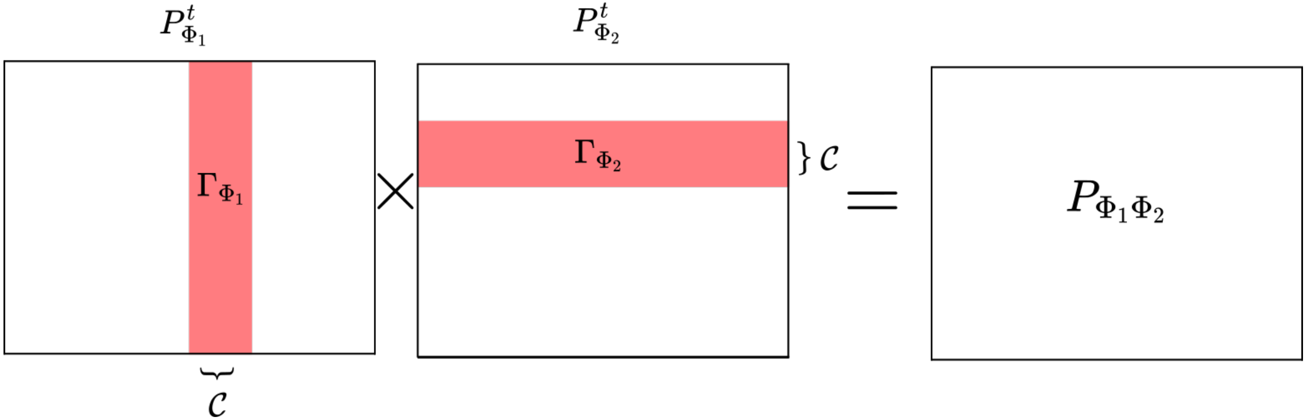
1. DTA starts off computing a diffusion operator in each domain by row normalizing an $\alpha-$decay Kernel (Moon et al. 2019).



# Diffusion Transport Alignment (DTA)

We propose to use an inter-domain diffusion process that connects both domains.

2. Extract the probabilities of transition from every observation in $\Phi_1$ to every $c \in \mathcal{C}$. And from every $c \in \mathcal{C}$ to $\Phi_2$.
3. Compute $P_{\Phi_1\Phi_2}$ as follows:



# Diffusion Transport Alignment (DTA)

We propose to use an inter-domain diffusion process that connects both domains.

4. Similarly, in the opposite direction we compute $P_{\Phi_2\Phi_1}$. These two quantities represent the transition probabilities from observations of one domain to the other. Thus, we can extract a distance as follows:

$$D_{ij} = ||P_{\Phi_1\Phi_2}(i,:) - P^t_{\Phi_2}(j,:)||_2 + ||P_{\Phi_2\Phi_1}(j,:) - P^t_{\Phi_1}(i,:)||_2 \tag{1}$$

or

$$D_{ij} = ||\log P_{\Phi_1\Phi_2}(i,:) - \log P^t_{\Phi_2}(j,:)||_2 + ||\log P_{\Phi_2\Phi_1}(j,:) - \log P^t_{\Phi_1}(i,:)||_2 \tag{2}$$

or

$$D_{ij} = 1 - \frac{\langle P_{\Phi_1\Phi_2}(i,:), P^t_{\Phi_2}(j,:)\rangle}{||P_{\Phi_1\Phi_2}(i,:)||||P^t_{\Phi_2}(j,:)||} + 1 - \frac{\langle P_{\Phi_2\Phi_1}(j,:), P^t_{\Phi_1}(i,:)\rangle}{||P_{\Phi_2\Phi_1}(j,:)||||P^t_{\Phi_1}(i,:)||} \tag{3}$$

# Diffusion Transport Alignment (DTA)

We propose to use an inter-domain diffusion process that connects both domains.

5. Now, how do we assign observations from one domain to the other?.

# Diffusion Transport Alignment (DTA)

We propose to use an inter-domain diffusion process that connects both domains.

5. Now, how do we assign observations from one domain to the other?. $D_{ij}$ contains a “cost” to move from one point to another. Thus, we can solve an optimal transpor problem:

$$
\begin{aligned}
\min_{T} \quad & \sum_{i \in \Phi_1, j \in \Phi_2} D_{ij} T_{ij} \\
\text{s.t.} \quad & \sum_{i \in \Phi_1} T_{ij} = q_j, \quad \forall j \in \Phi_2 \\
& \sum_{j \in \Phi_2} T_{ij} = v_i, \quad \forall i \in \Phi_1 \\
& T_{ij} \geq 0, \quad \forall i \in \Phi_1, \forall j \in \Phi_2.
\end{aligned}
\tag{1}
$$

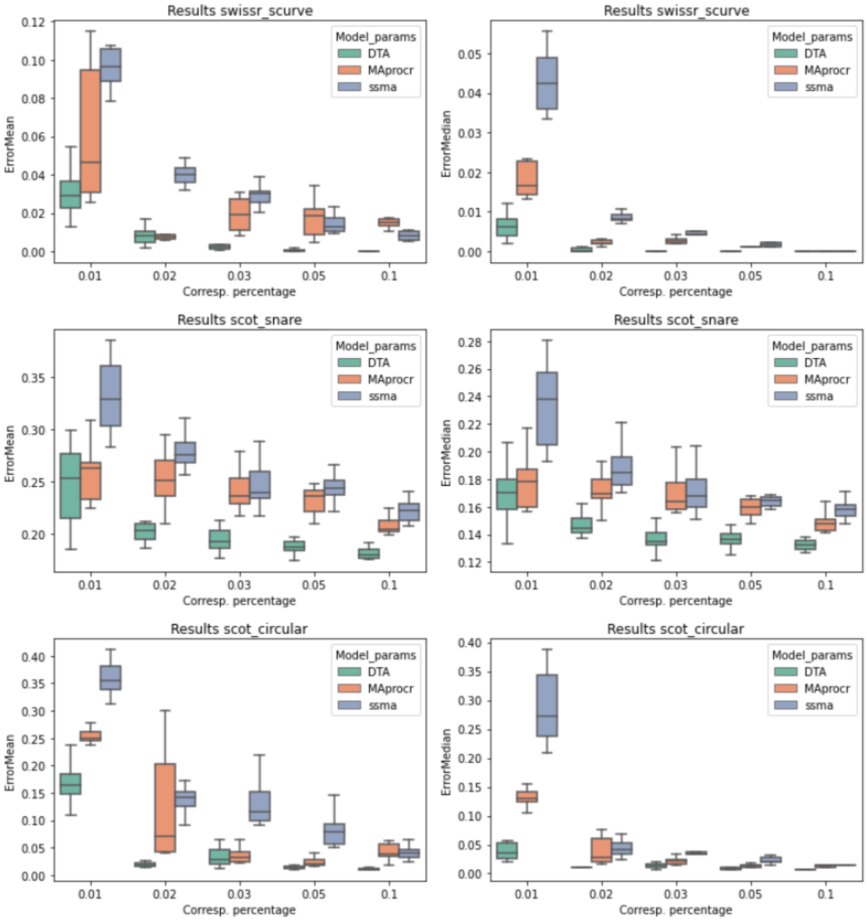
# Results

In order to quantify the alignment performance, we are interested in finding how close to the true match, a given observation is assigned in the other domain.

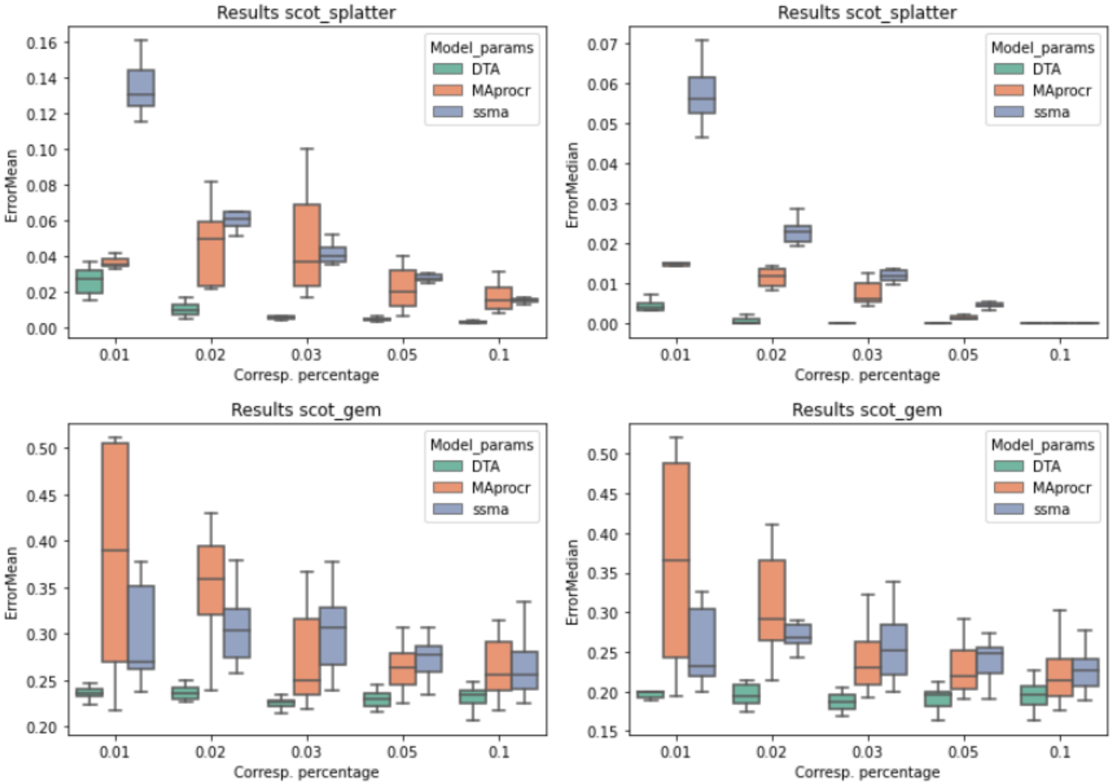
For this purpose, we define the error in the alignment of a given point as the rank of its assigned counterpart in the nearest neighbors of its true match.

Then, we extract two comparison metrics, one as the average of all the individual errors, and the other as the median.

# Results)

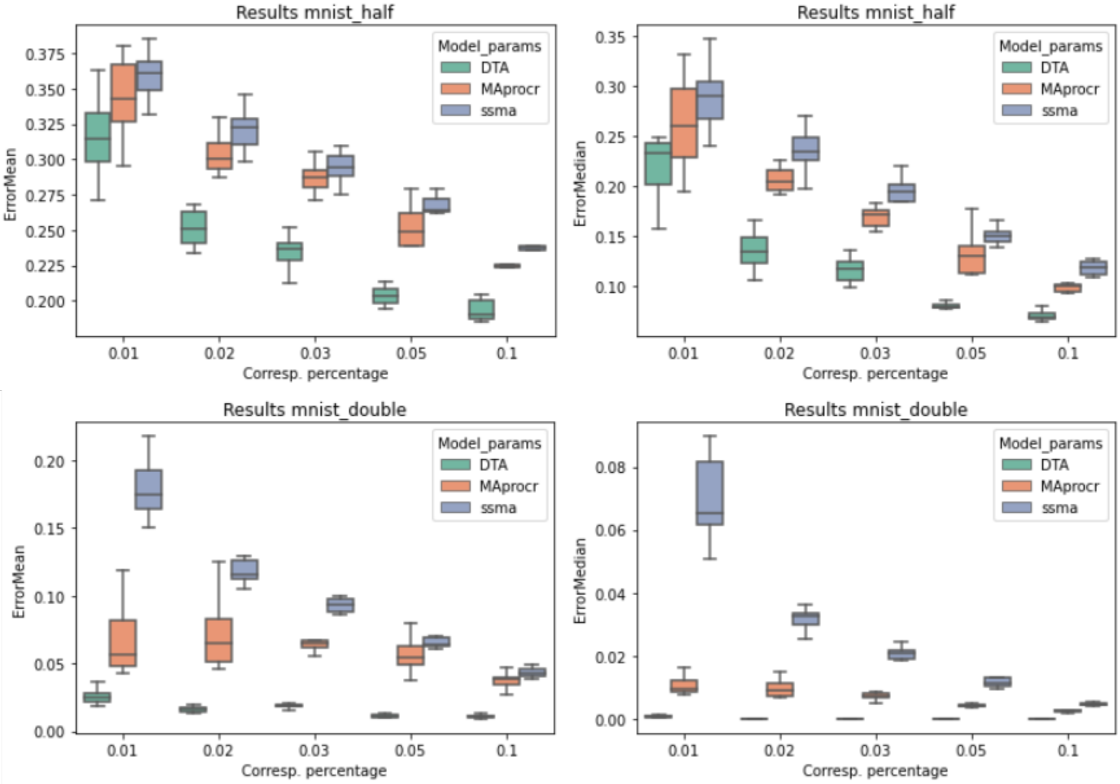

Results swissr_scurve
ErrorMean
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Model_params
DTA
MAprocr
ssma
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Results swissr_scurve
ErrorMedian
0.05
0.04
0.03
0.02
0.01
0.00
Results scot_snare
ErrorMean
0.35
0.30
0.25
0.20
Results scot_snare
ErrorMedian
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
Results scot_circular
ErrorMean
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Results scot_circular
ErrorMedian
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

# Results



Results scot_splatter
ErrorMean
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results scot_splatter
ErrorMedian
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results scot_gem
ErrorMean
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results scot_gem
ErrorMedian
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma

# Results


Results mnist_half
ErrorMean
0.375
0.350
0.325
0.300
0.275
0.250
0.225
0.200
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results mnist_half
ErrorMedian
0.35
0.30
0.25
0.20
0.15
0.10
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results mnist_double
ErrorMean
0.20
0.15
0.10
0.05
0.00
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma
Results mnist_double
ErrorMedian
0.08
0.06
0.04
0.02
0.00
0.01
0.02
0.03
0.05
0.1
Corresp. percentage
Model_params
DTA
MAprocr
ssma

THANKS!!!