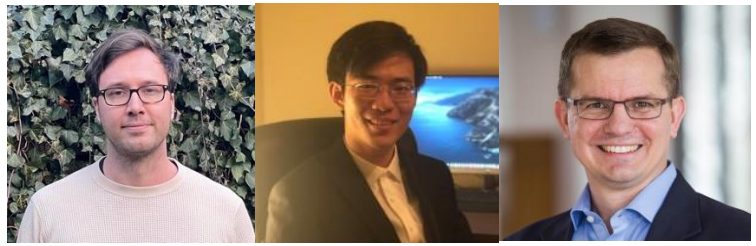
# Optimal Transport Modeling of Cellular Differentiation: From Low-Rank Structure to Temporal Dynamics

**Speaker: Peter Halmos**
Joint work with Julian Gold, Xinhao Liu, and Ben Raphael

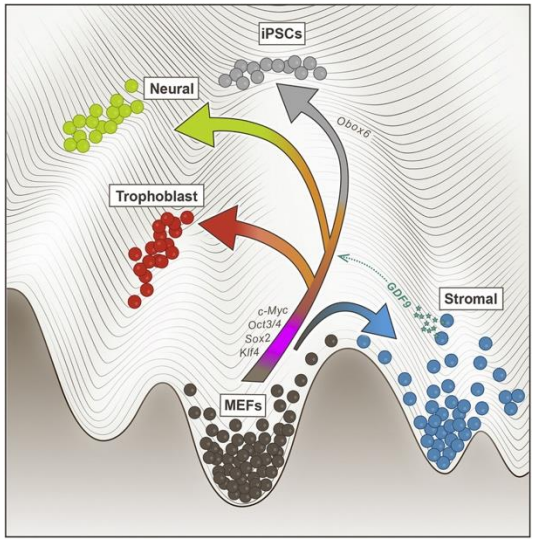


PRINCETON UNIVERSITY

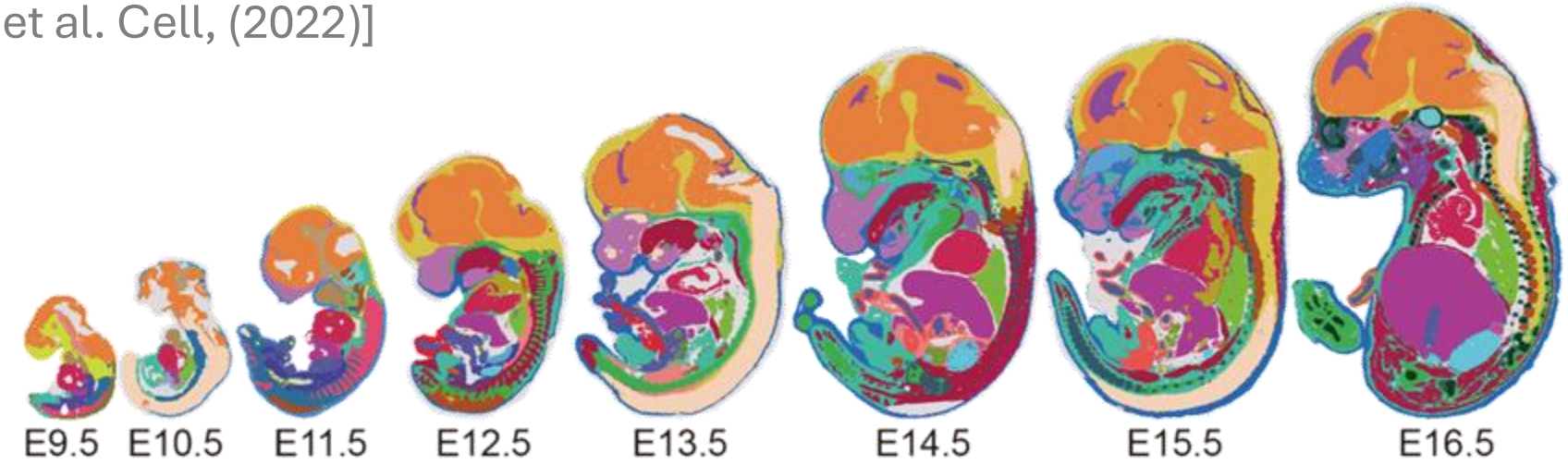
siam

# **Temporal and Spatiotemporal transcriptomics**: Sequencing across multiple time points during developmental and reprogramming processes

Reprogramming of fibroblasts to induced pluripotent stem cells [Schiebinger, et al. Cell, (2019)]




Spatial Transcriptomics of mouse embryos across 8 developmental stages
[Chen, et al. Cell, (2022)]

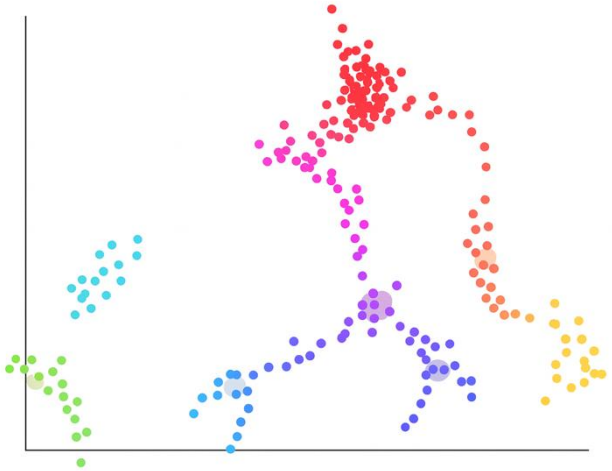


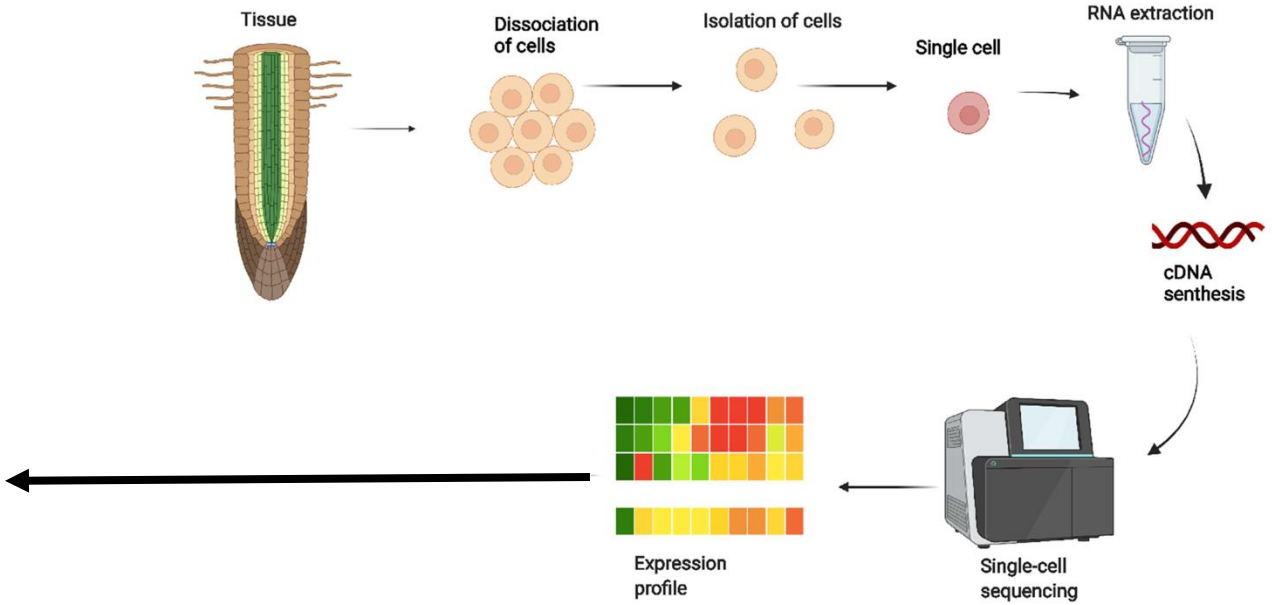

**And others!*
*(Pijuan-Sala et al., Nature, 2019)*
(Liu et al. *Developmental Cell*, 2022)

...

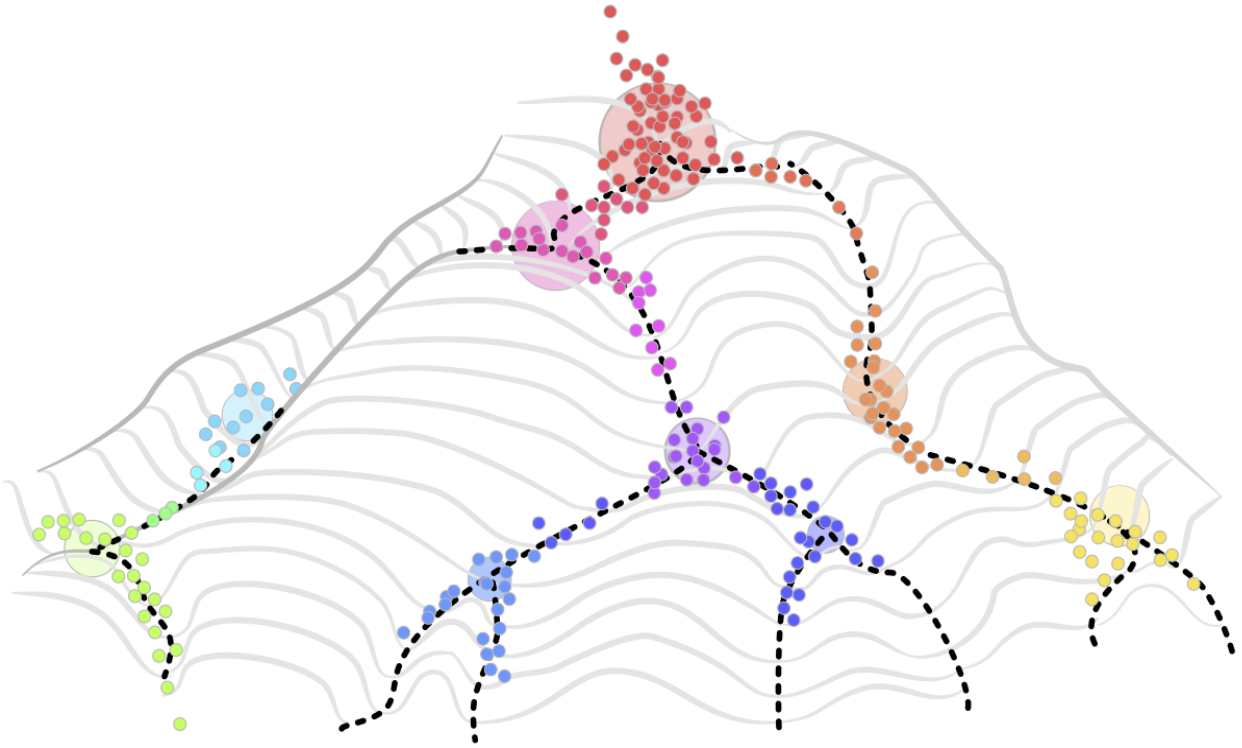
# Defining Cell State Space

- Two key experimental technologies: **Single-cell transcriptomics** and **Spatial transcriptomics**.
  - **Transcriptomics:** D-dimensional "state-space" of vector with scalar expression of the mRNA for each protein in cell (= $\mathbb{R}^{D}$)
  - ***Single cell:*** Cell $i$ is its transcriptional state $X^{(i)} \in \mathbb{R}^{D}$
  - ***Spatial transcriptomics*:** Cell $i$ is its transcriptional state $X^{(i)} \in \mathbb{R}^{D}$ augmented with a spatial position $s^{(i)} \in \mathbb{R}^{2 \text{ or } 3}$





# Temporal and Spatiotemporal transcriptomics: Opens the Analysis of Fundamental Biological Questions!



The "Waddington Landscape"

## Questions:

1. Ancestor-descendant relationships between cells across two timepoints?
2. Cell-states or types which index the temporal process of development?
3. Trajectories between these cell types?

## Limitations:

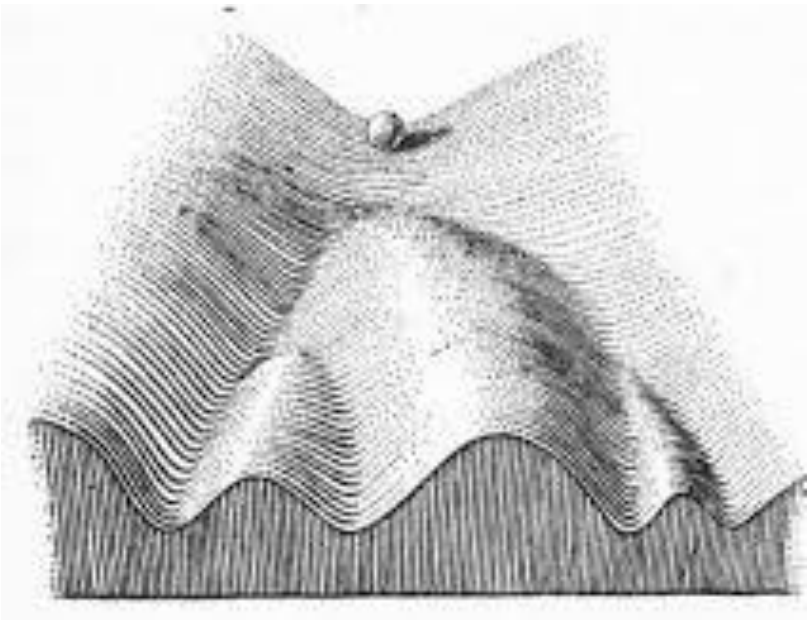
- Technology is destructive – each sample from a different individual
- Do not have ground-truth trajectories!

# Waddington Potential Landscape

- What's a reasonable modeling assumption for how cells evolve?
- In *The Strategy of the Genes '57*, C.H. Waddington Conjectured:
  - Cell differentiation pathways ≈ A gradient-flow on a potential landscape

i.e. minimize $\mathrm{V}(\mathrm{x})$ via dissipative descent:

$$\dot{x} = -\nabla V(x)$$



The "Waddington Landscape"
Photo cred: (Waddington, 1957)

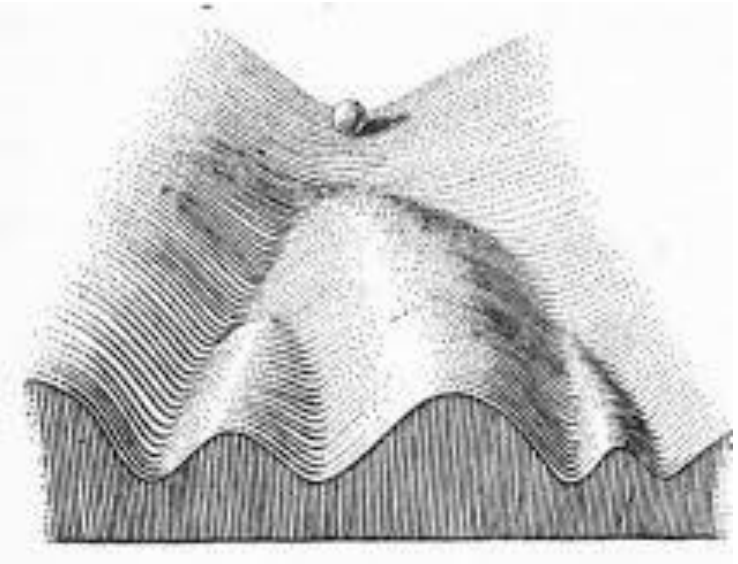
# Waddington Potential+ Landscape

- Since we observe cell distribution marginals at different time-points $\rho_t(x)$ with stochasticity, one needs to augment the particle-flow view!
- This can be viewed as a descent over the particle-distribution $\rho_t(x)$ itself with Langevin dynamics (Biondo '25, Lavenant & Schiebinger '21)

Continuity-equation: probability mass
(rather than points) flows down landscape

$$\partial_t \rho_t = \nabla \cdot (\rho_t \nabla V) + T \Delta \rho_t$$

$$dX_t = -\nabla V(x) dt + \sqrt{2T} dB_t$$

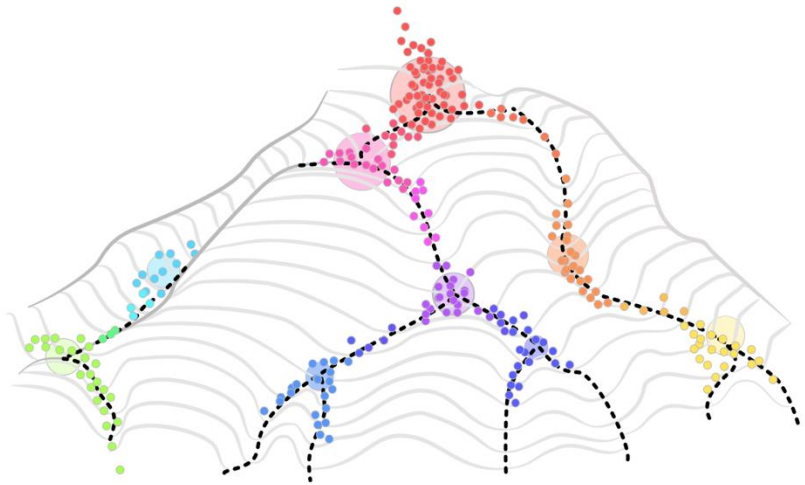


# Resolving Trajectories on the Landscape

- There are a few practical difficulties with applying this model:
  1. We do not know the landscape a priori
  2. We only observe discrete “snapshots”

$$\left(\rho_t = \frac{1}{n_t}\sum_{i=1}^{n_t}\delta_{x_i}\right)_{t=1}^{N}$$

     generated from the unknown landscape.
  3. Without the landscape, we do not know which cells transitioned to which!

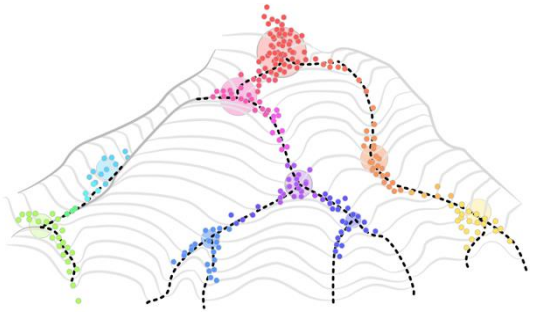


# Resolving Trajectories on the Landscape

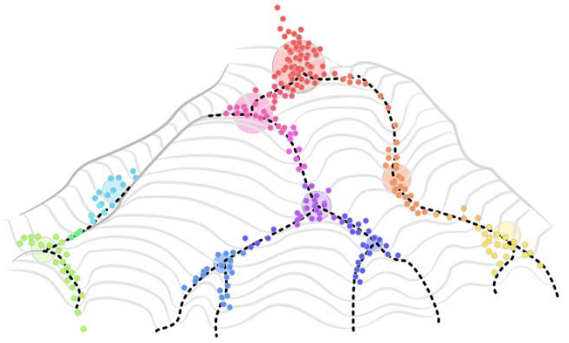
- Let's look at a pair of densities $\rho_0, \rho_1$ and apply Occam's Razor to the velocities which can take $\rho_0 \to \rho_1$:

Q: What is the time-varying distribution $\rho_t$ and velocity field $v_t$ of ***least kinetic energy*** between the two distributions?

$$\inf_{(\rho_t, v_t)} \left\{ \int_0^1 \int_{\mathbb{R}^d} \|v_t\|_2^2 \, \rho_t(dx) \;\middle|\; \partial_t \rho_t = -\nabla \cdot (\rho_t v_t),\ \rho_0 = \rho_{t=0},\ \rho_1 = \rho_{t=1} \right\}$$



# Solution: Optimal Transport!



- **(Benamou & Brenier, 2000):** The minimal kinetic energy value for this probability-flow is the Wasserstein distance of optimal transport (OT)

$$\inf_{(\rho_t, v_t)} \left\{ \int_0^1 \int_{\mathbb{R}^d} \|v_t\|_2^2 \, \rho_t(dx) \;\middle|\; \partial_t \rho_t = -\nabla \cdot (\rho_t v_t),\ \rho_0 = \rho_{t=0},\ \rho_1 = \rho_{t=1} \right\}$$
$$= W_2^2(\rho_0, \rho_1)\,!$$

- The optimal velocity field is given by the transport map! We can *just* solve for OT velocity directly. Coincides with gradient flow velocity in limit of taking the time-interval to zero (**Jordan, Kinderlehrer & Otto, '98**): discrete OT on fine time-intervals can recover any dynamics.

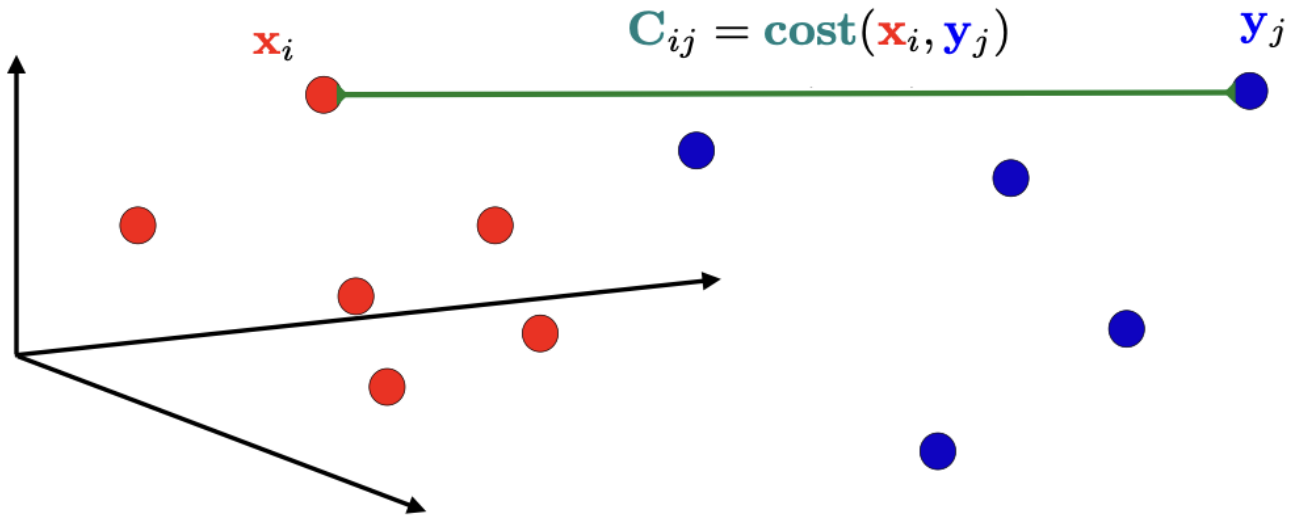
Conclusion: OT is *"most natural"* choice!

# Discrete Optimal Transport

- Practically, when $\rho_0 = \sum_{i=1}^{n_0} a_i \delta_{x_i}$ and $\rho_1 = \sum_{j=1}^{n_1} b_j \delta_{x_i}$, (usually uniform), OT is solved by the discrete optimization problem:

$$\mathbf{P}^\star = \underset{\mathbf{P} \in \Pi(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \mathbf{C}, \mathbf{P} \rangle$$

$$= \underset{\mathbf{P} \in \Pi(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \sum_{i,j} \mathbf{C}_{ij} \mathbf{P}_{ij}$$

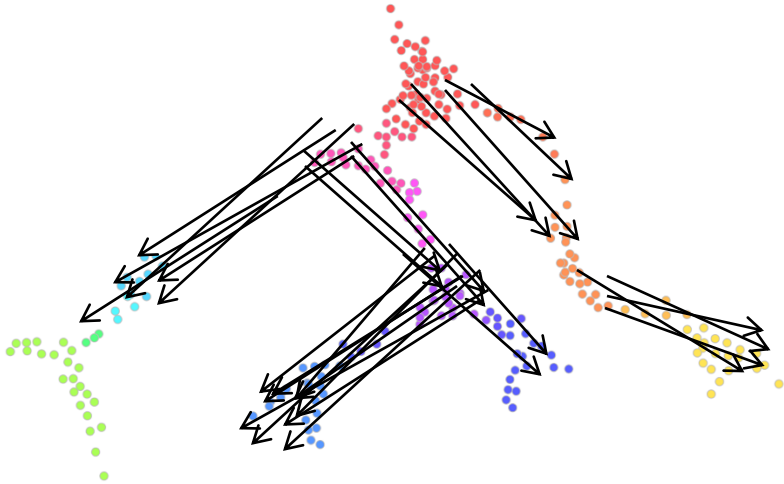



$$\Pi(\boldsymbol{a}, \boldsymbol{b}) = \{\mathbf{P} \in \mathbb{R}_+^{n \times m} : \mathbf{P}\mathbf{1}_m = \boldsymbol{a},\ \mathbf{P}^T \mathbf{1}_n = \boldsymbol{b}\}$$

# Existing Techniques for Optimal Transport

- Many existing methods map the time-dynamics of cells using variations of optimal transport (Schiebinger et al 2019, Zeira et al 2022, Klein et al 2025, Liu & Halmos et al 2025) with great success
- However, they infer full-rank structure through cell-cell couplings, as opposed to *low-rank* structures in the mapping

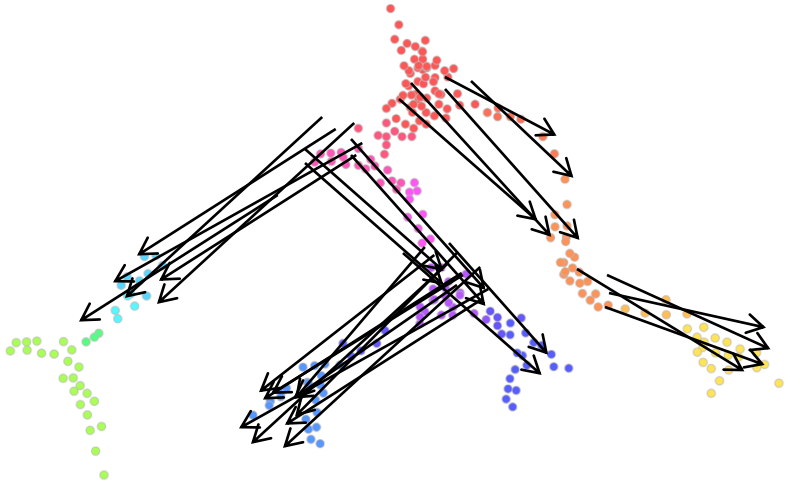
Cell-to-cell Coupling: Full-rank structure



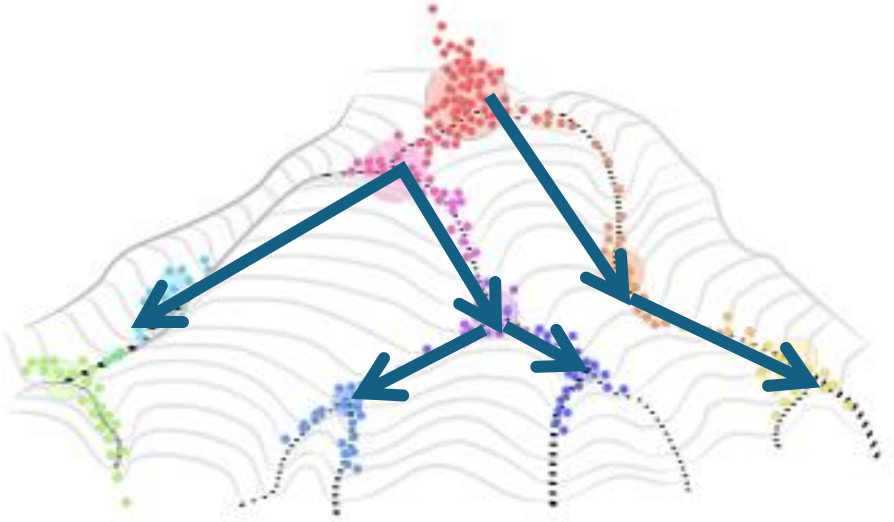
# Low-Rank Structure in the Transport

- OT maps (without entropy regularization) are bijections
- No true low-rank structure and *flat* spectrum of singular values = 1!

Cell-to-cell Coupling: Full-rank structure

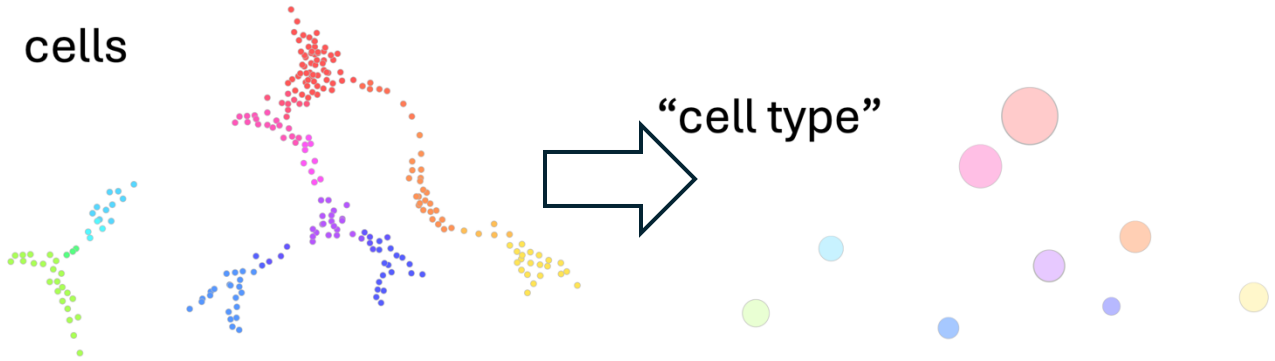


Latent Trajectories: *Low-Rank* Structure

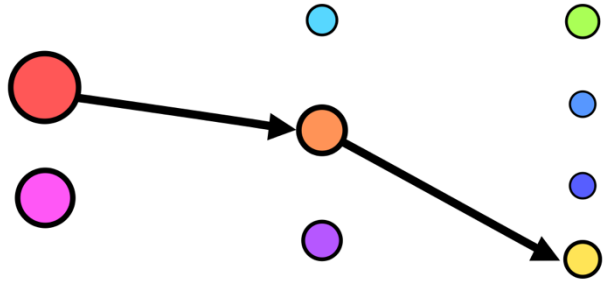


# Low-Rank Structure in the Transport

- Examples of low-rank structure in the transport include latent *cell-state* and the *differentiation map* between these states
  - *The "canals" or "latent-trajectories" of Waddington's landscape are* ***low-rank*** *structures!*
- Existing techniques assume cell-state inference is *distinct* from transport and don't bridge the two.




A *cell-type* is a coarse-graining of cells into clusters.



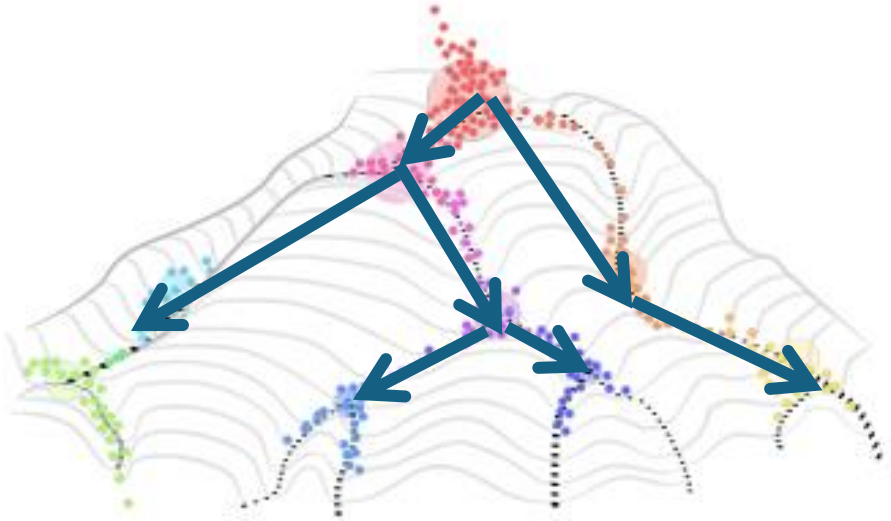
*A differentiation map* is a directed acyclic graph giving the ancestral relationship between cell-types.

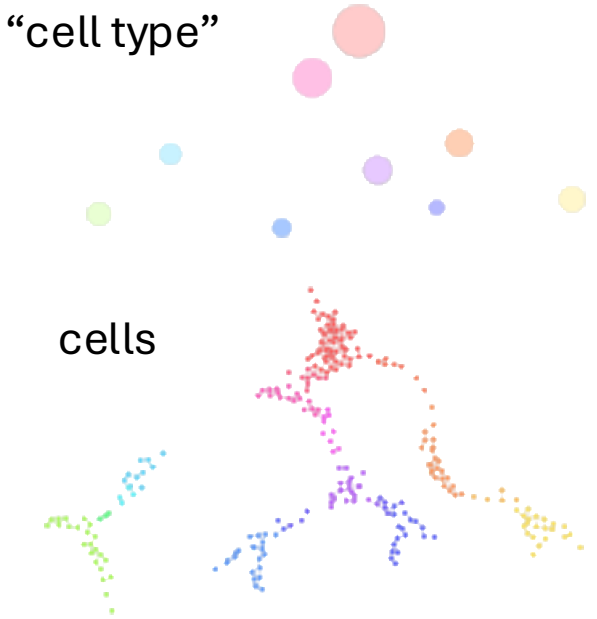
# Latent Trajectories over Cell-State

Our work addresses the following questions:

1. What cell-states "index" the temporal development process?
2. What is the *differentiation map* (DAG) between these cell states?

Latent Trajectories in Transport



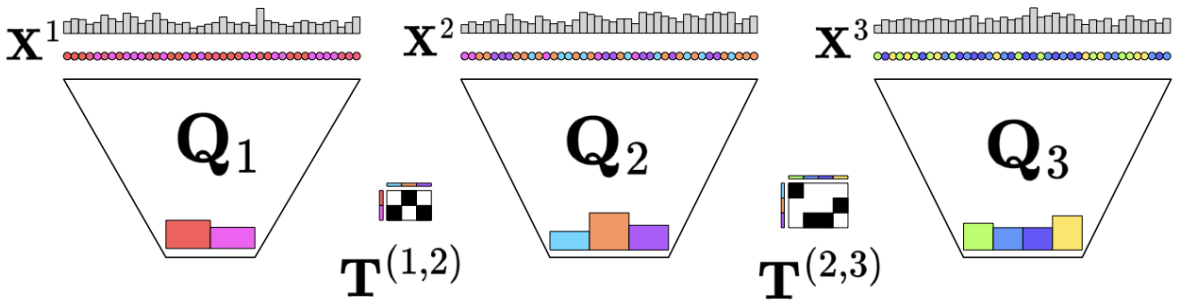


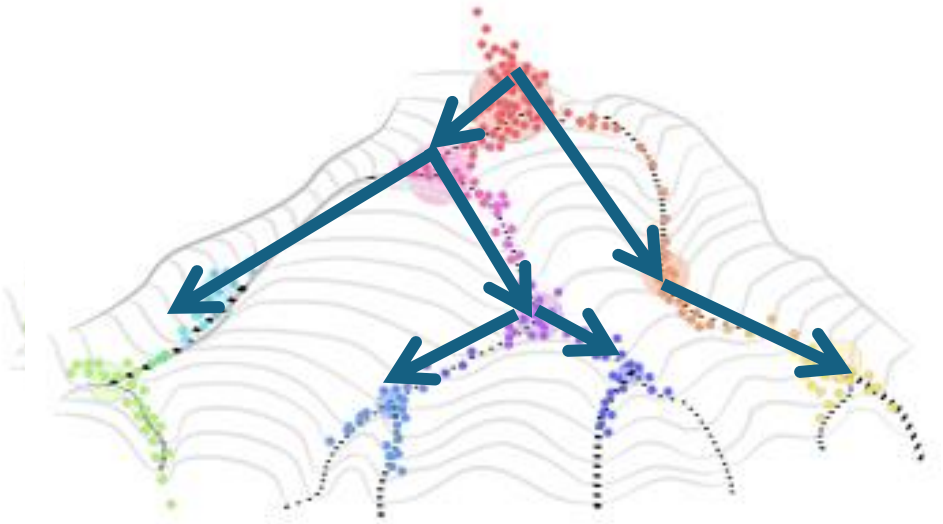
# Hidden Markov Optimal Transport (HM-OT)

(1) Discovers latent cell types and aligns individual cells to them.

(2) Maps between the cell types while minimizing an optimal transport cost.

(3) Uses *low-rank optimal transport* (Forrow et al '19, Scetbon et al '20, Lin et al '21, Halmos et al '24) to do (1) and (2) simultaneously across multiple timepoints.

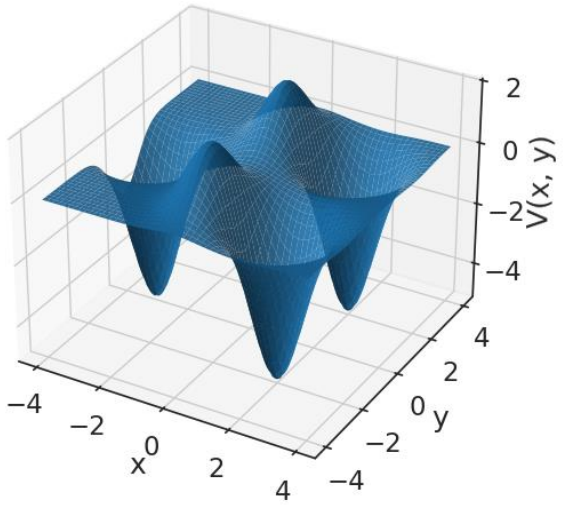


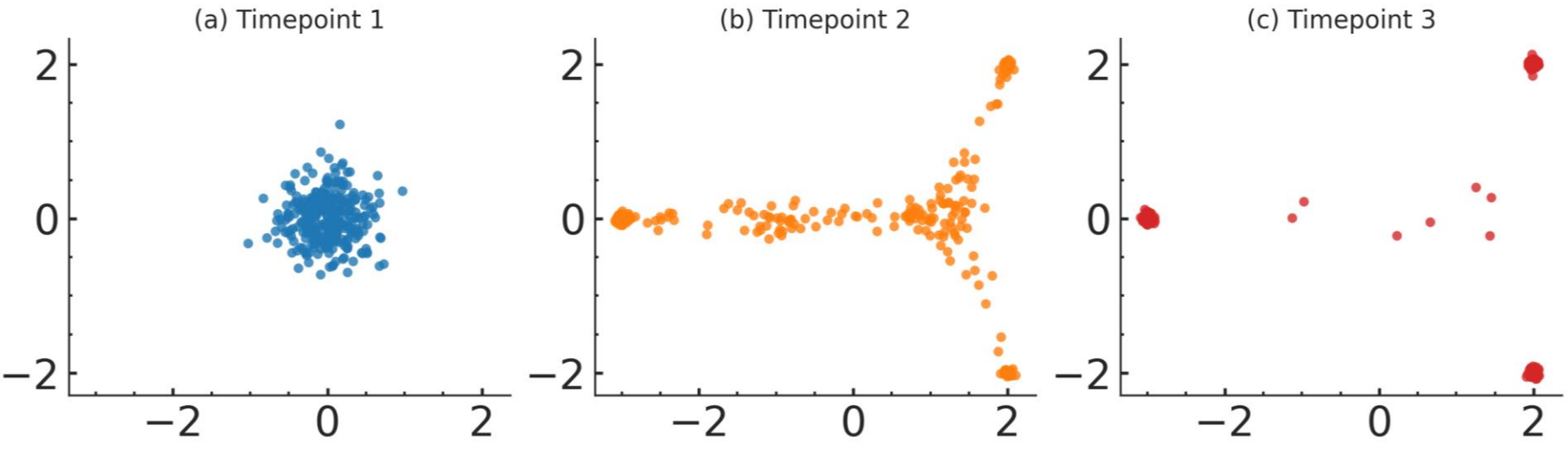



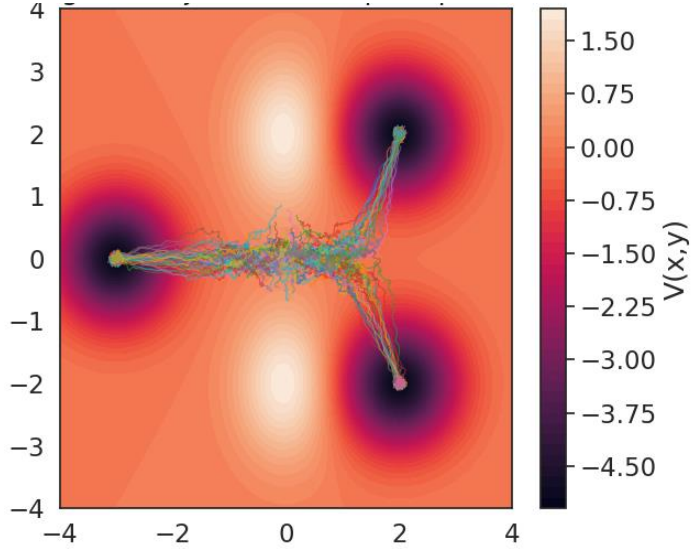
# Example: Langevin on Tristable Potential

- Existing techniques assume cell-state inference is *distinct* from transport. *Where does this go wrong?*

- Suppose cells follow Langevin on a tristable landscape (**Bhattacharya '11**) and we measure trajectories at 3 timepoints



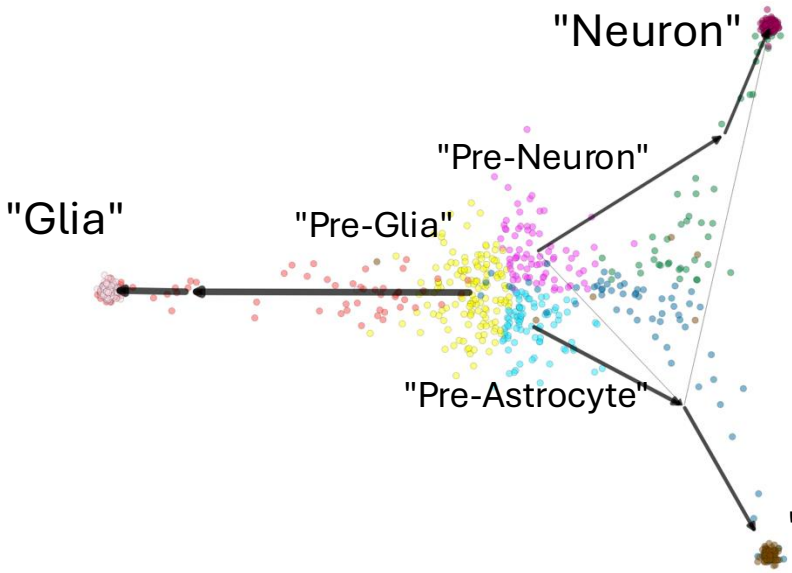



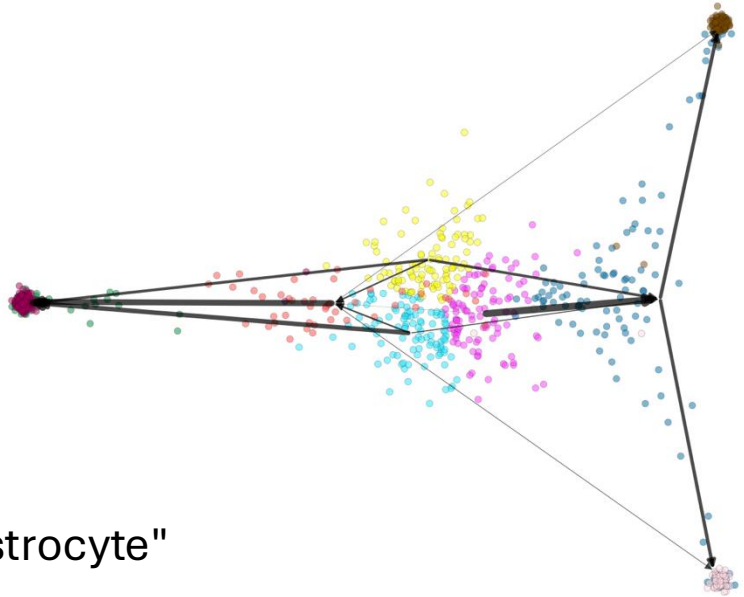
# Example: Langevin on Tristable Potential

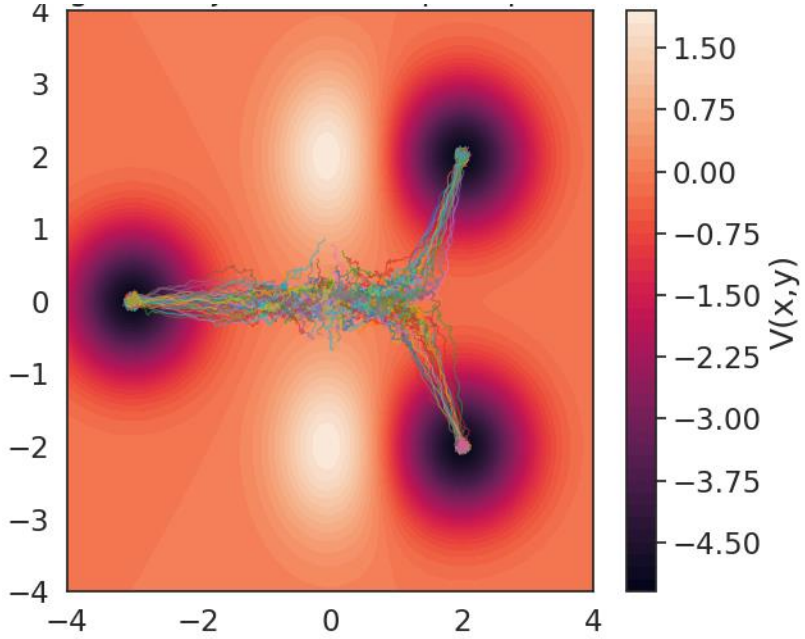
- Clustering at each timepoint independently + full-rank OT fails to identify the three true latent trajectories on the tri-stable landscape.
- Latent trajectories must be optimized *jointly* with the transport!

Hidden-Markov OT (HM-OT)




Clustering + Full-rank OT





# *Low-rank* OT: Latent Structure in Transport

- A rank-r coupling P may always represented (Scetbon ’21, Cohen & Rothblum) by the factorization

$$\mathbf{P} = \mathbf{Q}\mathrm{diag}(\mathbf{g}^{-1})\mathbf{R}^\top$$

- *Low-rank* optimal transport (Forrow ‘19, Scetbon ‘21) solves primal OT with this factorization

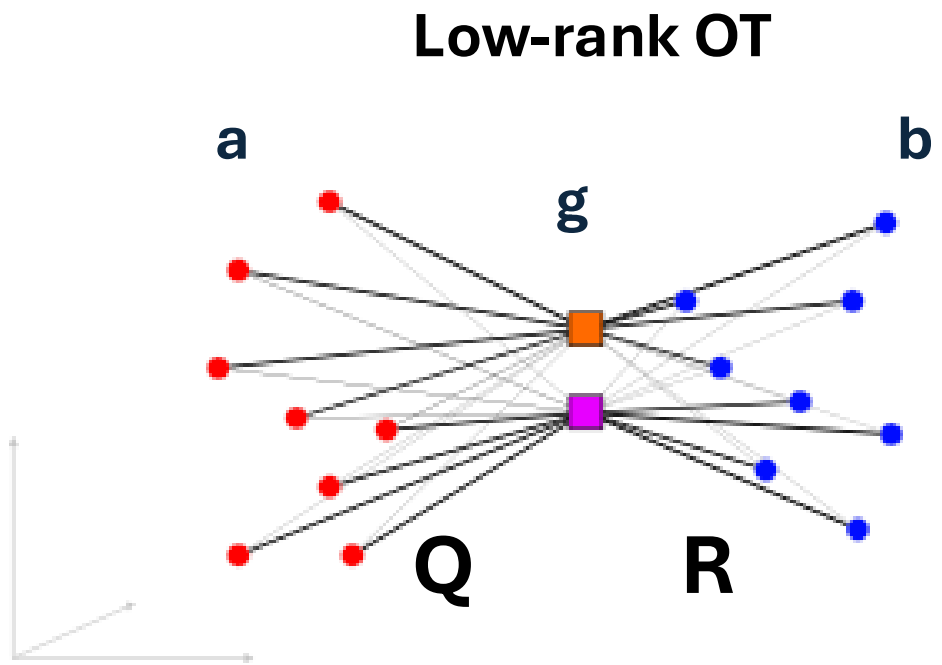
(Parametrization) $$\boldsymbol{P} = \boldsymbol{Q}\mathrm{diag}(1/\boldsymbol{g})\boldsymbol{R}^{\mathrm{T}}$$

(Loss) $$\mathcal{L}_{\mathrm{LOT}} := \langle \boldsymbol{C}, \boldsymbol{Q}\mathrm{diag}(1/\boldsymbol{g})\boldsymbol{R}^{\mathrm{T}} \rangle$$

(Constraints) $$\mathsf{FC}_{\boldsymbol{a},\boldsymbol{b}}(r) := \{(\boldsymbol{Q},\boldsymbol{R},\boldsymbol{g}) \in \mathbb{R}_+^{n\times r} \times \mathbb{R}_+^{m\times r} \times (\mathbb{R}_+^*)^r : \boldsymbol{Q} \in \Pi_{\boldsymbol{a},\boldsymbol{g}}, \boldsymbol{R} \in \Pi_{\boldsymbol{b},\boldsymbol{g}}\}$$

# *Low-rank* OT: Latent Structure in Transport

$$\mathbf{P} = \mathbf{Q}\mathrm{diag}(\mathbf{g}^{-1})\mathbf{R}^\top$$




"Outer marginal:" the points

$$\mathbf{Q}\mathbf{1}_r = \mathbf{a},\ \mathbf{R}\mathbf{1}_r = \mathbf{b}$$

"Inner marginal:" the latent co-clusters

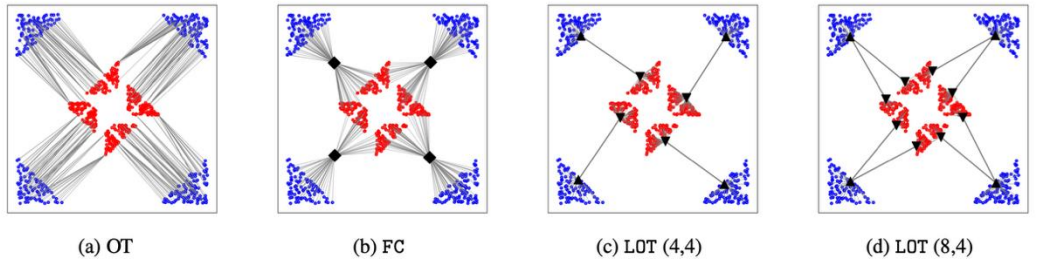
$$\boldsymbol{g} = \mathbf{Q}^\top \mathbf{1}_n = \mathbf{R}^\top \mathbf{1}_n$$

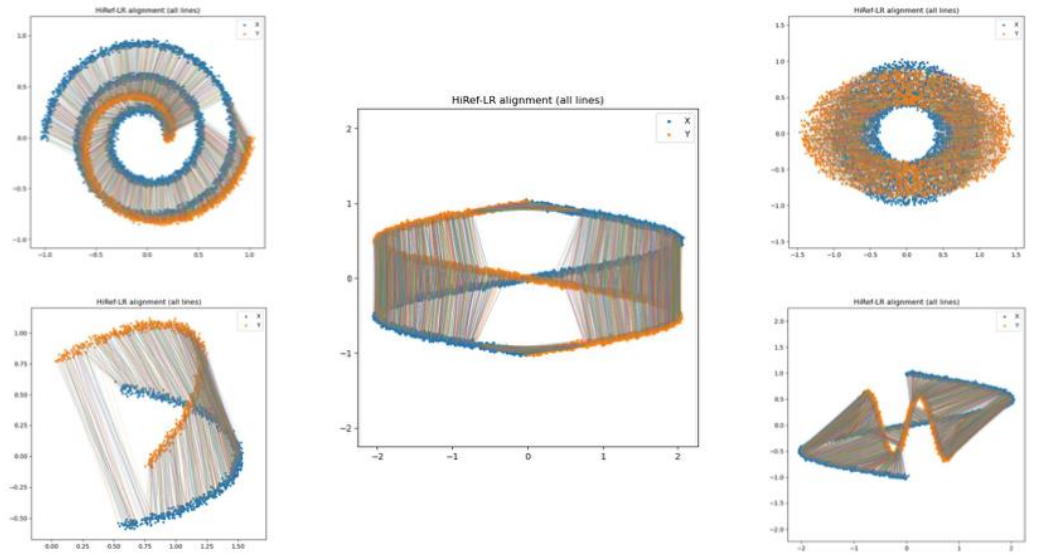
# *Low-rank* OT: key benefits

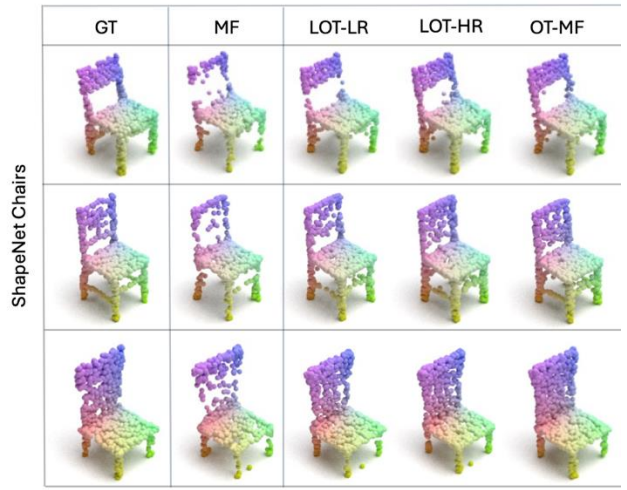
1. It captures **latent, interpretable** structure in the transport (Forrow ‘19, Lin ’21) with *linear* complexity.

2. It offers a framework for ***co-clustering*** and generalizes K-means to a pair of datasets (Scetbon ‘22).

3. Full-rank OT ***can reduce to*** low-rank OT with linear space and **O(n log n) time*** with Hierarchical Refinement (Halmos&Gold, ICML ’25).



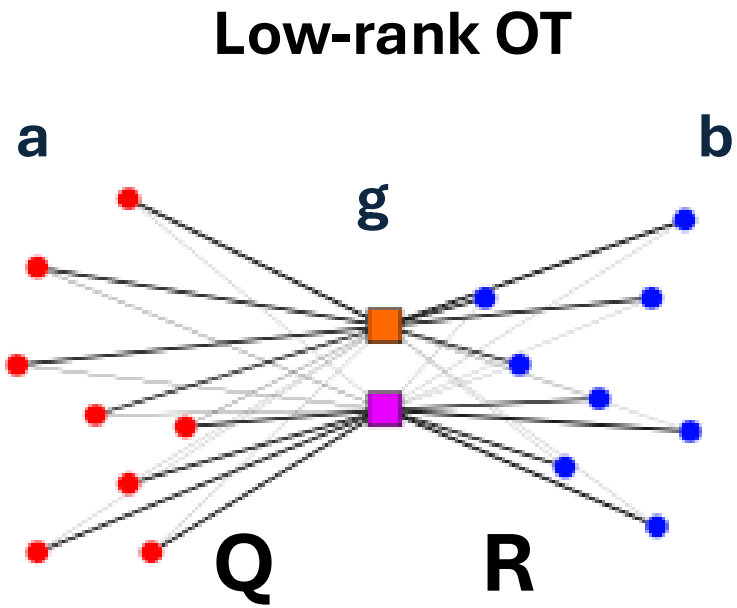



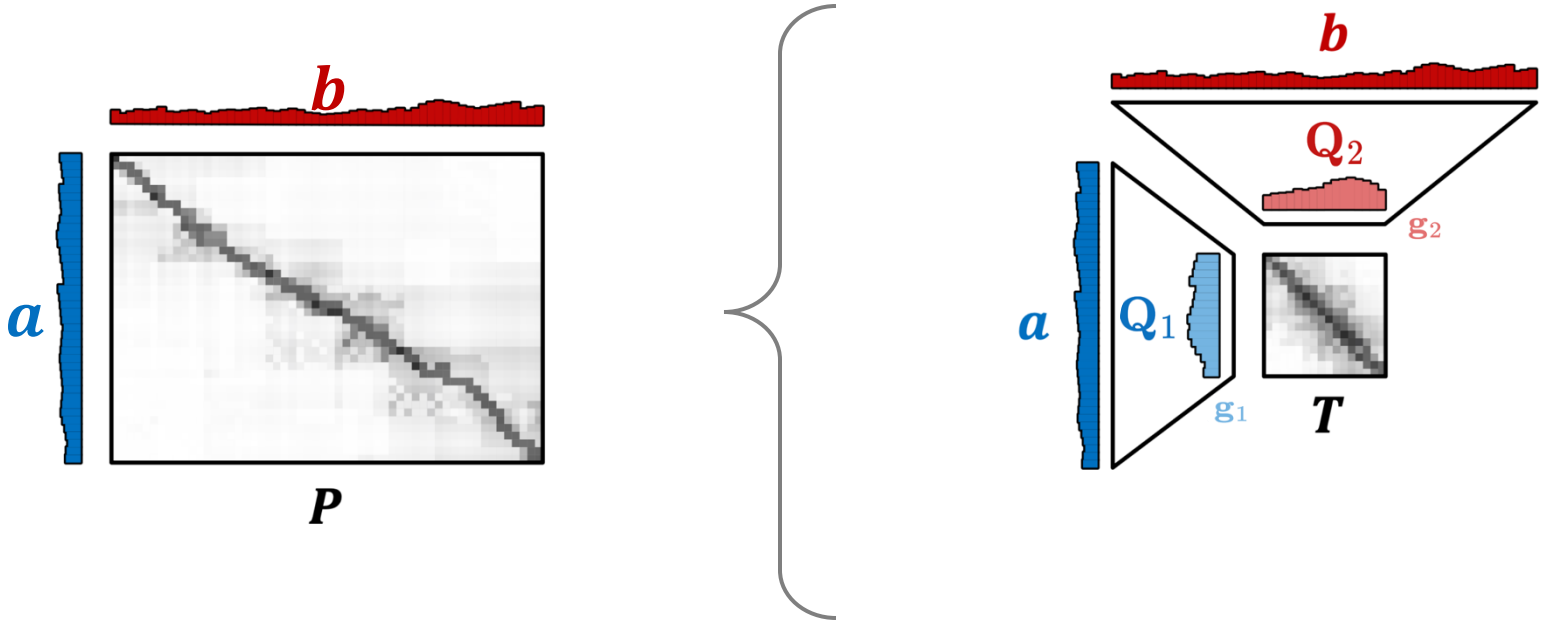
# *Low-rank* Optimal Transport

- One issue with this factorization: It does not account for hidden transition or DAG structure!

$$\mathbf{P} = \mathbf{Q}\mathrm{diag}(\mathbf{g}^{-1})\mathbf{R}^{\top}$$



# Low-Rank Optimal Transport with Latent Coupling: A Special Parametrization

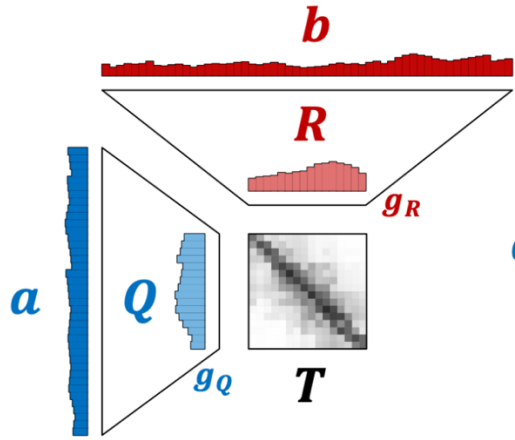



**Factor relaxation with latent coupling (FRLC)**

Low rank approximation of optimal transport

Halmos*, Liu*, Gold*, R. NeurIPS (2024)

# Factor-Relaxation and Latent Coupling



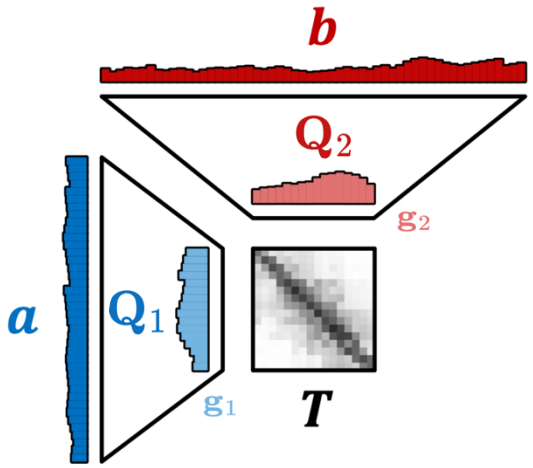

- LC (latent coupling) parametrization decouples problem using coordinate mirror-descent, with the introduction of a *latent coupling* **T** between two inner marginals $\mathbf{g_Q}$ and $\mathbf{g_R}$

(Parametrization) $$\boldsymbol{Q}\mathrm{diag}(1/\boldsymbol{g}_Q)\boldsymbol{T}\mathrm{diag}(1/\boldsymbol{g}_R)\boldsymbol{R}^{\mathrm{T}} =: \boldsymbol{P}_{(\boldsymbol{Q},\boldsymbol{R},\boldsymbol{T})}$$

(Loss) $$\mathcal{L}_{\mathrm{LC}}(\boldsymbol{Q},\boldsymbol{R},\boldsymbol{T}) := \langle \boldsymbol{C}, \boldsymbol{P}_{(\boldsymbol{Q},\boldsymbol{R},\boldsymbol{T})}\rangle_F,$$

(Constraints) $$\mathrm{LC}_{\boldsymbol{a},\boldsymbol{b}}(r) := \{(\boldsymbol{Q},\boldsymbol{R},\boldsymbol{T}) \in \mathbb{R}_+^{n\times r} \times \mathbb{R}_+^{m\times r} \times \mathbb{R}_+^{r\times r} : \boldsymbol{Q} \in \Pi_{\boldsymbol{a},.}, \boldsymbol{R} \in \Pi_{\boldsymbol{b},.}, \boldsymbol{T} \in \Pi_{\boldsymbol{g}_Q,\boldsymbol{g}_R}\}$$

# Factor-Relaxation and Latent Coupling (Algorithm)




$$\boldsymbol{Q}_k \leftarrow \tilde{\Pi}_{\boldsymbol{a},\boldsymbol{g}_Q}(\boldsymbol{Q}_k \odot \exp(-\gamma_k \nabla_{\boldsymbol{Q}}))$$
$$\boldsymbol{R}_k \leftarrow \tilde{\Pi}_{\boldsymbol{b},\boldsymbol{g}_R}(\boldsymbol{R}_k \odot \exp(-\gamma_k \nabla_{\boldsymbol{R}}))$$
$$\boldsymbol{g}_Q, \boldsymbol{g}_R = \boldsymbol{Q}_k^{\mathrm{T}} \mathbf{1}_n, \boldsymbol{R}_k^{\mathrm{T}} \mathbf{1}_m$$
$$\boldsymbol{T}_k \leftarrow \Pi_{\boldsymbol{g}_R,\boldsymbol{g}_Q}(\boldsymbol{T}_k \odot \exp(-\gamma_{\boldsymbol{T}} \nabla_{\boldsymbol{T}}))$$

**Factor relaxation with latent coupling (FRLC)**

Low rank approximation of optimal transport

Halmos*, Liu*, Gold*, R. NeurIPS (2024)

Efficiently solved with Mirror-Descent: alternating Sinkhorn projections!

# Interpretation of Latent Coupling

**LC (latent coupling) factorization** of P imposes *rank constraint* and decomposes P into 3 factors while keeping it a feasible coupling

$$\implies \mathbf{P} = \mathbf{Q}_1 \mathrm{diag}(1/\mathbf{g}_1)\mathbf{T}\mathrm{diag}(1/\mathbf{g}_2)\mathbf{Q}_2^{\mathrm{T}}$$

$\mathbf{Q}_1 \in \Pi(\mathbf{a}, \mathbf{g}_1)$

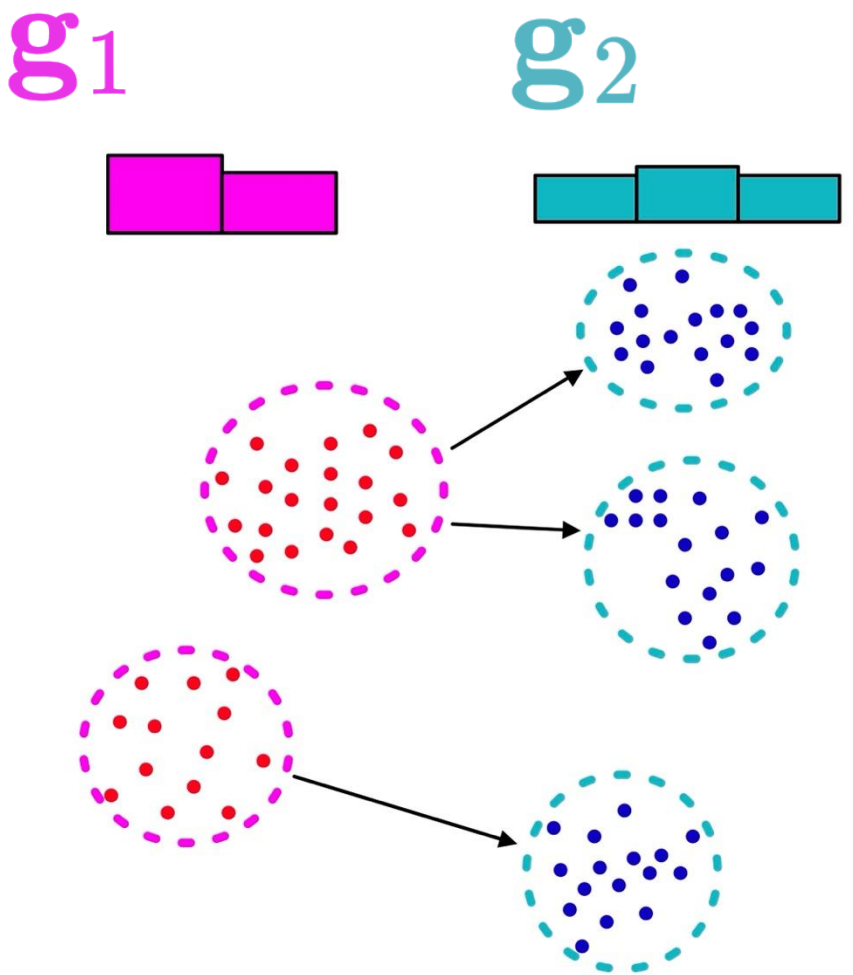
Couples point distribution at time 1 to cell-state distribution at time 1

$\mathbf{T} \in \Pi(\mathbf{g}_1, \mathbf{g}_2)$

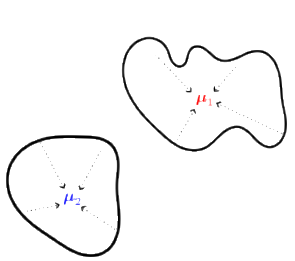
Couples cell-type distribution at time 1 to cell-state distribution at time 2

$\mathbf{Q}_2^{\mathrm{T}} \in \Pi(\mathbf{g}_2, \mathbf{b})$

Couples cell-state distribution at time 2 to point distribution at time 2

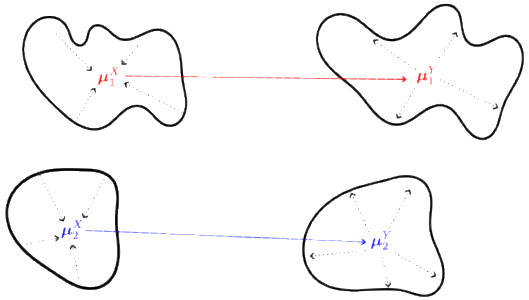


# Hidden-Markov OT:
# Linking Latent Trajectories *Through Time*



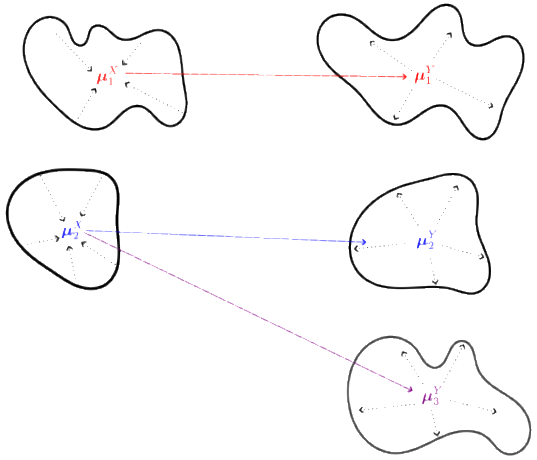
$$\langle \mathbf{C}, \mathbf{Q}^{(0)} \mathrm{diag}(1/\boldsymbol{g}) \mathbf{Q}^{(0),\top} \rangle_F$$

K-Means



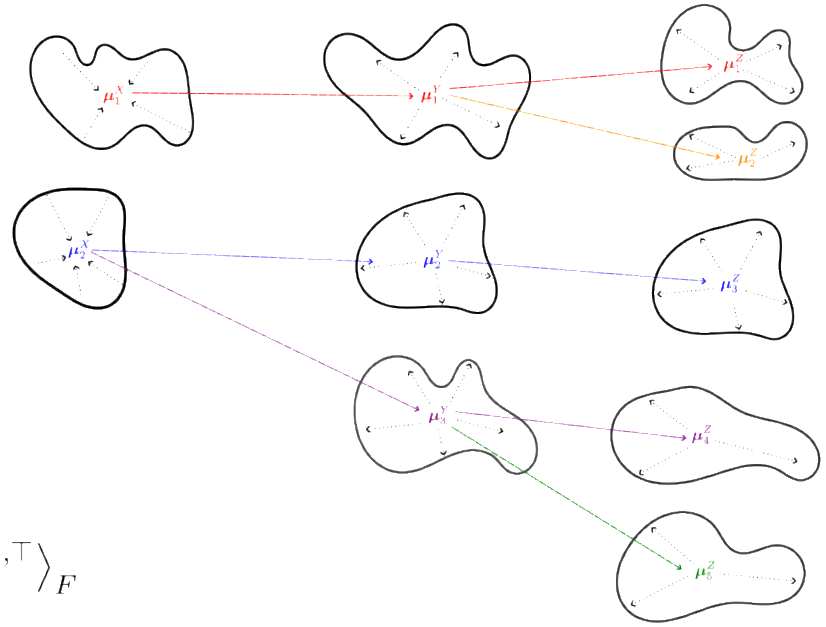
$$\langle \mathbf{C}, \mathbf{Q}^{(0)} \mathrm{diag}(1/\boldsymbol{g}) \mathbf{Q}^{(1),\top} \rangle_F$$

Classical Low-Rank
Optimal Transport



$$\langle \mathbf{C}, \mathbf{Q}^{(0)} \mathrm{diag}(1/\boldsymbol{g}_0) \mathbf{T}^{(0,1)} \mathrm{diag}(1/\boldsymbol{g}_1) \mathbf{Q}^{(1),\top} \rangle_F$$

Latent Coupling Low-Rank
Optimal Transport



$$\sum_{t=1}^{N-1} \langle \mathbf{C}^{(t,t+1)}, \mathbf{Q}^{(t)} \mathrm{diag}(1/\boldsymbol{g}_t) \mathbf{T}^{(t,t+1)} \mathrm{diag}(1/\boldsymbol{g}_{t+1}) \mathbf{Q}^{(t+1),\top} \rangle_F$$

Hidden-Markov Optimal Transport

# Hidden-Markov Optimal Transport

- The key idea of HM-OT is to (1) use the latent-coupling parametrization to capture cell differentiation and transitions, and (2) extend the loss function of low-rank OT to account for many timepoints
- Introduce cells encoded as distributions $\boldsymbol{a}_t$, cell-type vectors $\boldsymbol{g}_t$, and a coupling (latent representation) between these cells their types $\boldsymbol{Q}_t$

Parametrization and constraints are same as FRLC, but with time index!

(Parametrization) $\mathbf{P}^{(t,t+1)} := \mathbf{Q}_t \mathrm{diag}(1/\mathbf{g}_t) \mathbf{T}^{(t,t+1)} \mathrm{diag}(1/\mathbf{g}_{t+1}) \mathbf{Q}_{t+1}^T$

(Constraints) $\mathsf{LC}_{\mathbf{a}_t, \mathbf{a}_{t+1}}(r_t, r_{t+1})$

# Hidden-Markov Optimal Transport

- The factors need to be *linked* across time for cell-state to be defined consistently: multi-marginal extension of low-rank OT
- Cost minimizes the global distance of latent trajectories across time

Different loss! Generalized and linked across many timepoints.

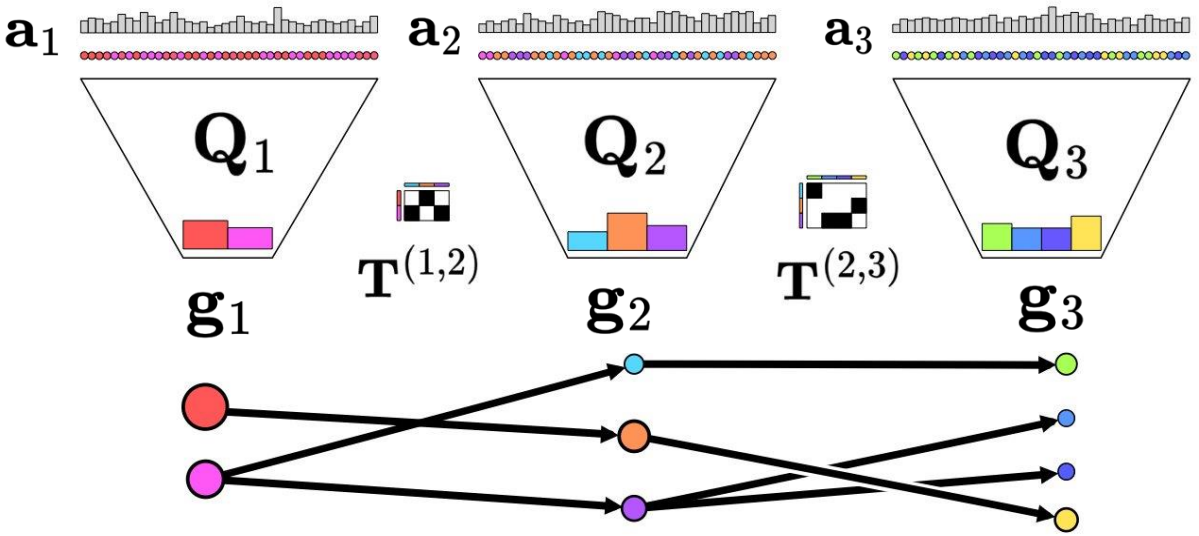
(Loss) $$\min_{\mathbf{Q},\mathbf{T}:(\mathbf{Q}_t,\mathbf{Q}_{t+1},\mathbf{T}^{(t,t+1)})\in \mathsf{LC}_{\mathbf{a}_t,\mathbf{a}_{t+1}}(r_t,r_{t+1})} \sum_{t=1}^{N-1} \langle \mathbf{C}^{(t,t+1)}, \mathbf{P}^{(t,t+1)} \rangle_F$$

# Hidden-Markov Optimal Transport

**Problem:** Given empirical distributions $(\boldsymbol{a}_t)_{t=1,..,N}$ find the latent factors $(\boldsymbol{Q}_t)_{t=1,..,N}$ and differentiation maps $(T^{(t,t+1)})_{t=1,..,N-1}$ that minimize the Wasserstein cost of the latent states through time.

$$\min_{\mathbf{Q},\mathbf{T}:(\mathbf{Q}_t,\mathbf{Q}_{t+1},\mathbf{T}^{(t,t+1)})\in \mathsf{LC}_{\mathbf{a}_t,\mathbf{a}_{t+1}}(r_t,r_{t+1})} \sum_{t=1}^{N-1} \langle \mathbf{C}^{(t,t+1)}, \mathbf{P}^{(t,t+1)} \rangle_F$$

$$\mathbf{P}^{(t,t+1)} := \mathbf{Q}_t \mathrm{diag}(1/\mathbf{g}_t) \mathbf{T}^{(t,t+1)} \mathrm{diag}(1/\mathbf{g}_{t+1}) \mathbf{Q}_{t+1}^T$$



# Hidden-Markov OT (Algorithm)

- Decouples sequential problem in time into greedy Forward-Backward estimates of ($Q_t$ , $T^{(t,t+1)}$, $g_t$) which are solved with FRLC.

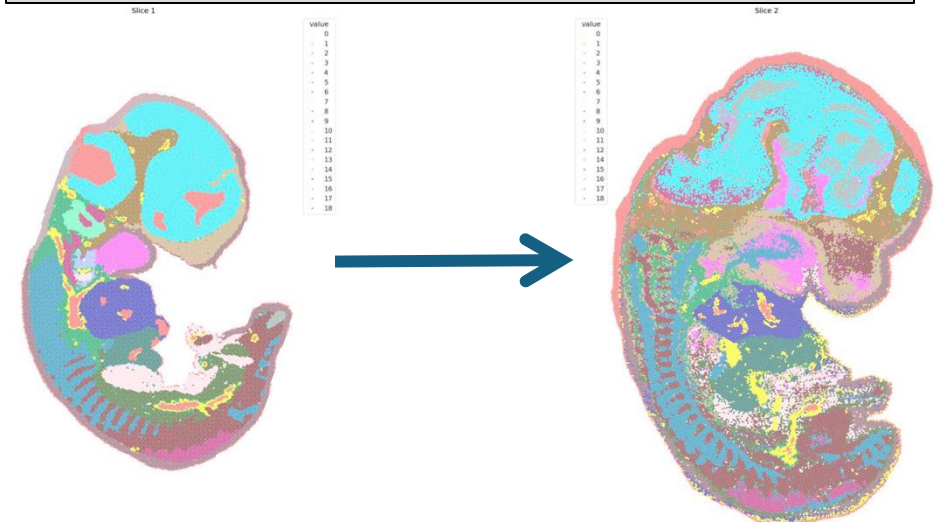
- Decoupling is like the forward-backward algorithm for Hidden Markov Models (HMM)

# Hidden-Markov OT (Flexibility)

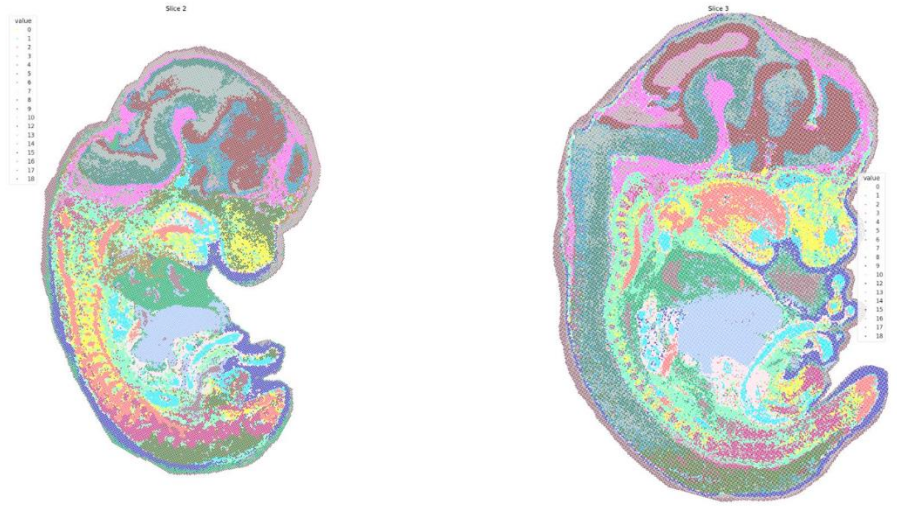
- Highly flexible in terms of input information! One can either run fully unsupervised or constrain any subset of the variables:
  - Cell-type proportions ($\boldsymbol{g_t}$)
  - Cell to cell-type assignments ($\boldsymbol{Q_t}$)
  - Cell-type to cell-type (e.g. lineage) transition structure ($\boldsymbol{T}^{(t,t+1)}$)

# HM-OT: Flexible Toolbox for (Co) Clustering

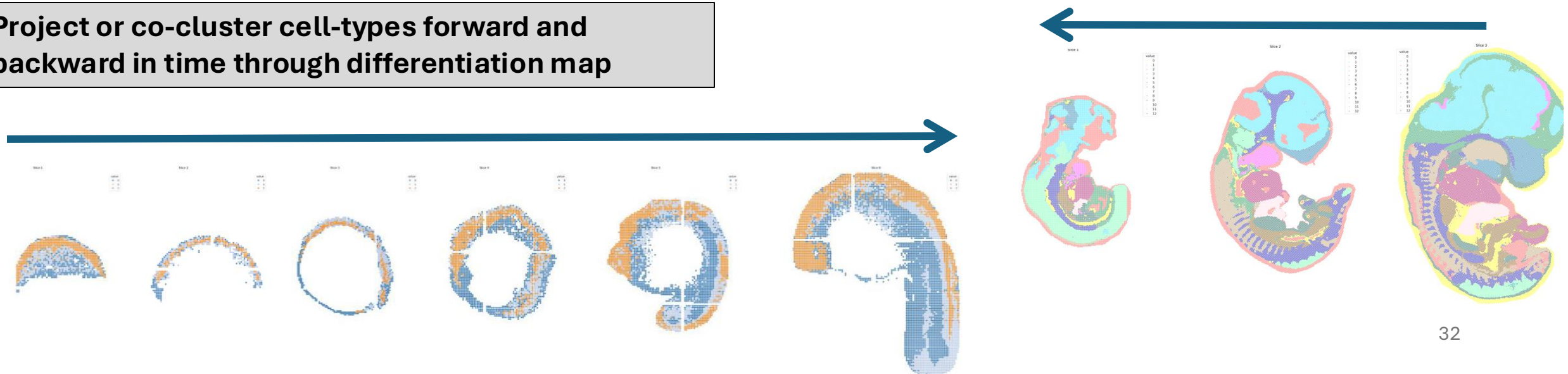
**Transfer known clusters forwards or backwards in time to other data**



**Learn cell state/type from scratch to minimize HM-OT objective**



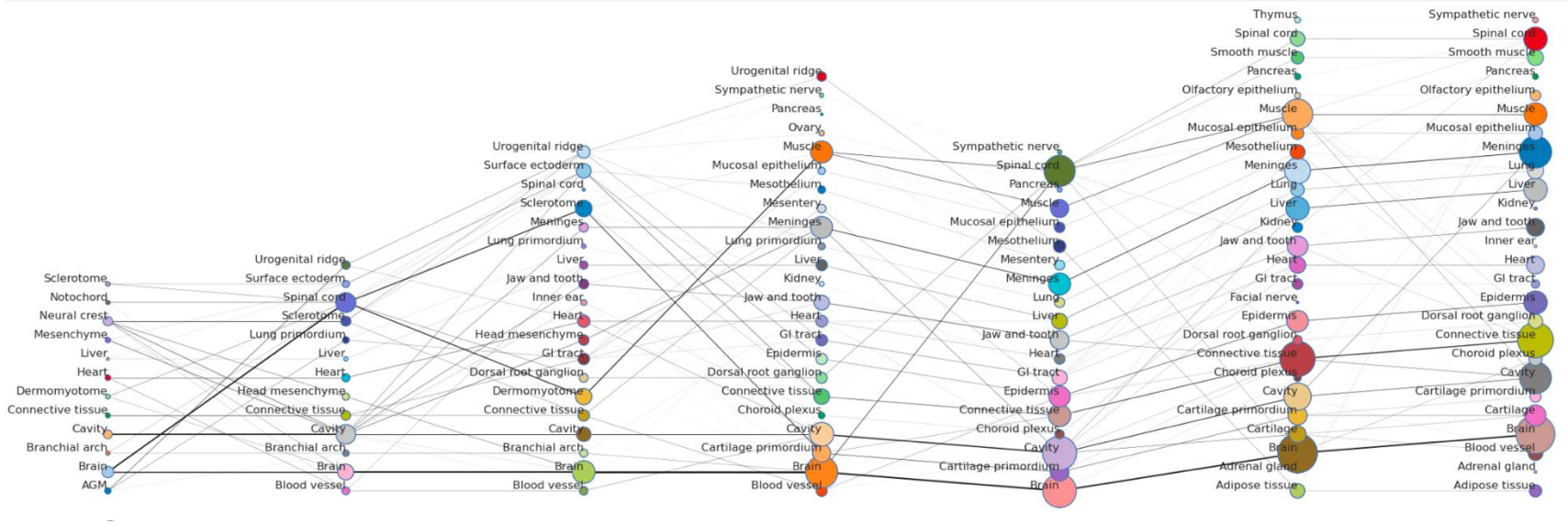
**Project or co-cluster cell-types forward and backward in time through differentiation map**



# Large-Scale Inference of Differentiation Maps

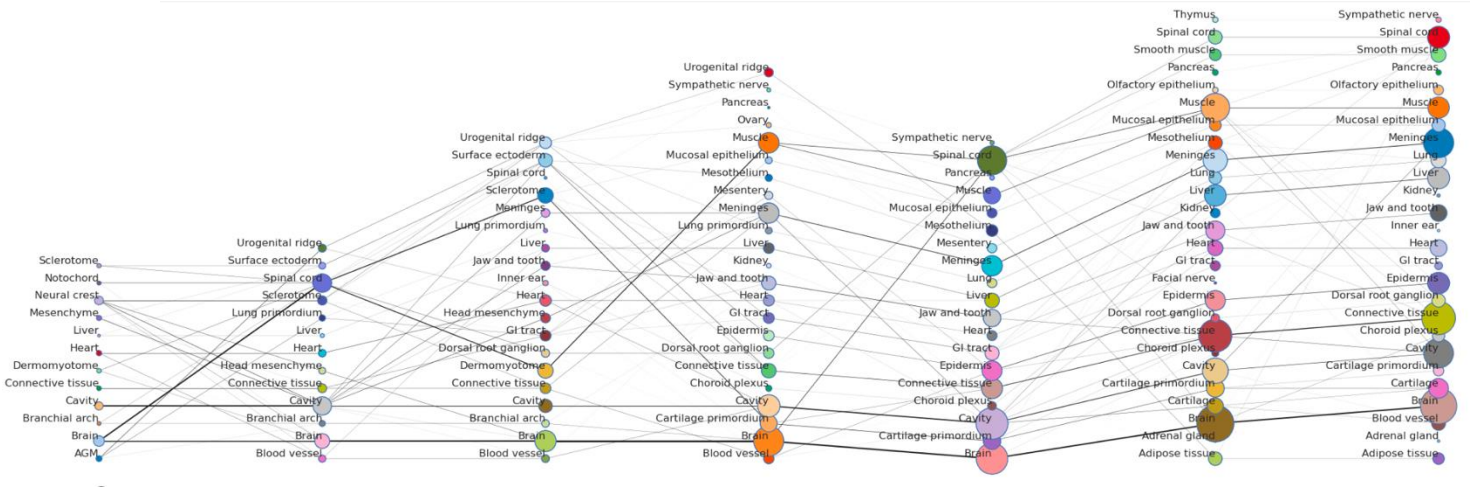
*Lightning fast* and space-efficient; can scale maps to millions of points!

**Spatial (Stereo-Seq) Mouse Development (Chen et al '22)**

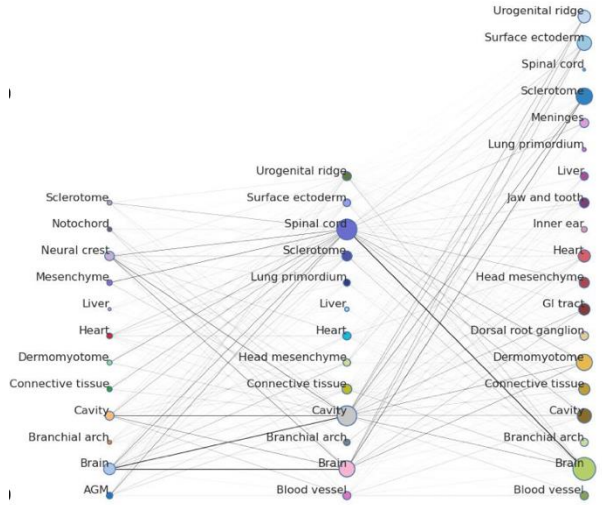



**Temporal (Single-Cell) Mouse Embryogenesis (Qiu et al '24)**

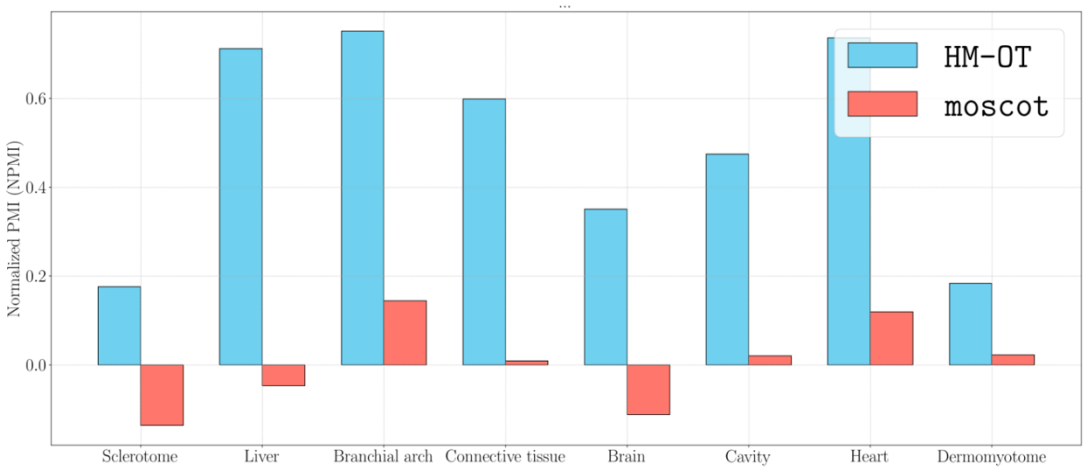
# Large-Scale Inference of Differentiation Maps




**HM-OT**




**moscot spatiotemporal**




Other true transitions identified:

- urogenital ridge transitioning to ovary (NPMI = **0.560**, E11.5-12.5)
- lung primordium transitioning to lung (NPMI=**0.927**, E12.5-13.5)
- dermomyotome transitioning to muscle (NPMI=**0.968**, E11.5-12.5)
- sclerotome transitioning to cartilage primordium (NPMI=**0.866**, E11.5-12.5)
- surface ectoderm transitioning to the epidermis (NPMI=**0.705**, E11.5-12.5)

Spatial (Stereo-Seq) Mouse Development (Chen et al '22)

# Summary

**HM-OT:** a scalable algorithm to infer differentiation maps, discover temporal co-clusters, and track cell-types through time and space.

- HM-OT introduces a novel multi-marginal optimal transport formulation to map cell-type differentiation
- Optimizes this factorization across a full time-series of temporal transcriptomics (or other!) data

Thank you!

https://github.com/raphael-group/HM-OT/

# Acknowledgments


## Raphael Group

**Prof. Ben Raphael**
Dr. Brian Arnold
Dr. Metin Balaban
Dr. Uthsav Chitra
**Dr. Julian Gold**
Dr. Cong Ma
Dr. Uyen Mai
Dr. Hirak Sarkar
Dr. Palash Sashittal
Dr. Yihang Shen
Dr. Alexander Strzalkowski
Dr. Hongyu Zheng
Viola Chen
Gillian Chu
**Peter Halmos**
William Howard-Snyder
Gary Hu
Akhil Jakatdar
**Xinhao Liu**
Sereno Lopez-Darwin
Henri Schmidt
Ahmed Shuaibi
Richard Zhang
Clover Zheng


PRINCETON UNIVERSITY

NIH NATIONAL CANCER INSTITUTE Informatics Technology for Cancer Research

SCHMIDT FUTURES