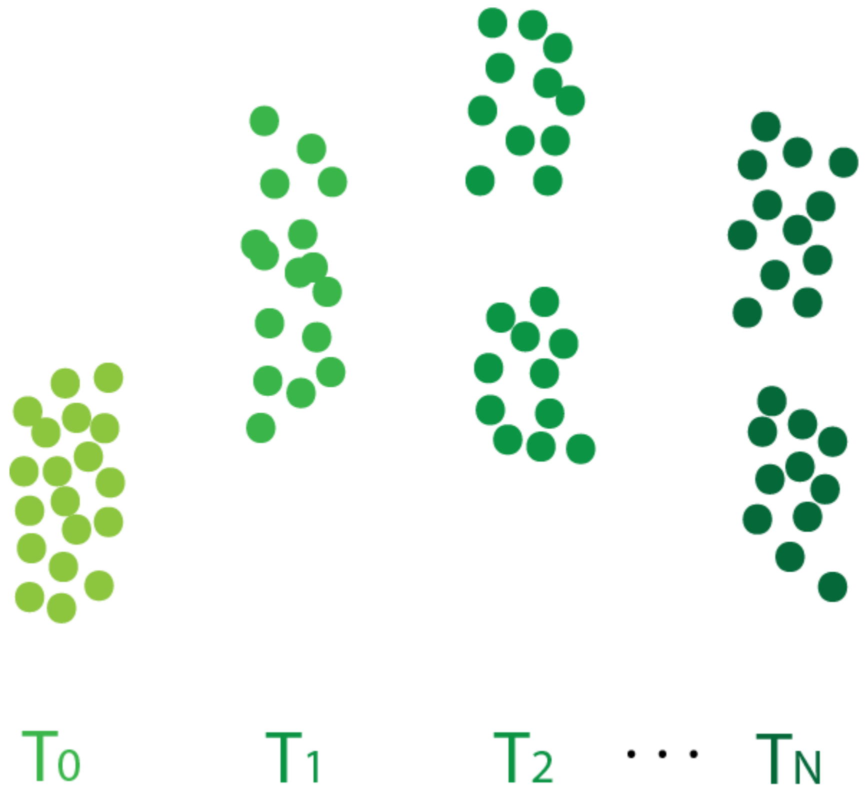
TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, Smita Krishnaswamy

July 2020

Yale

Motivation



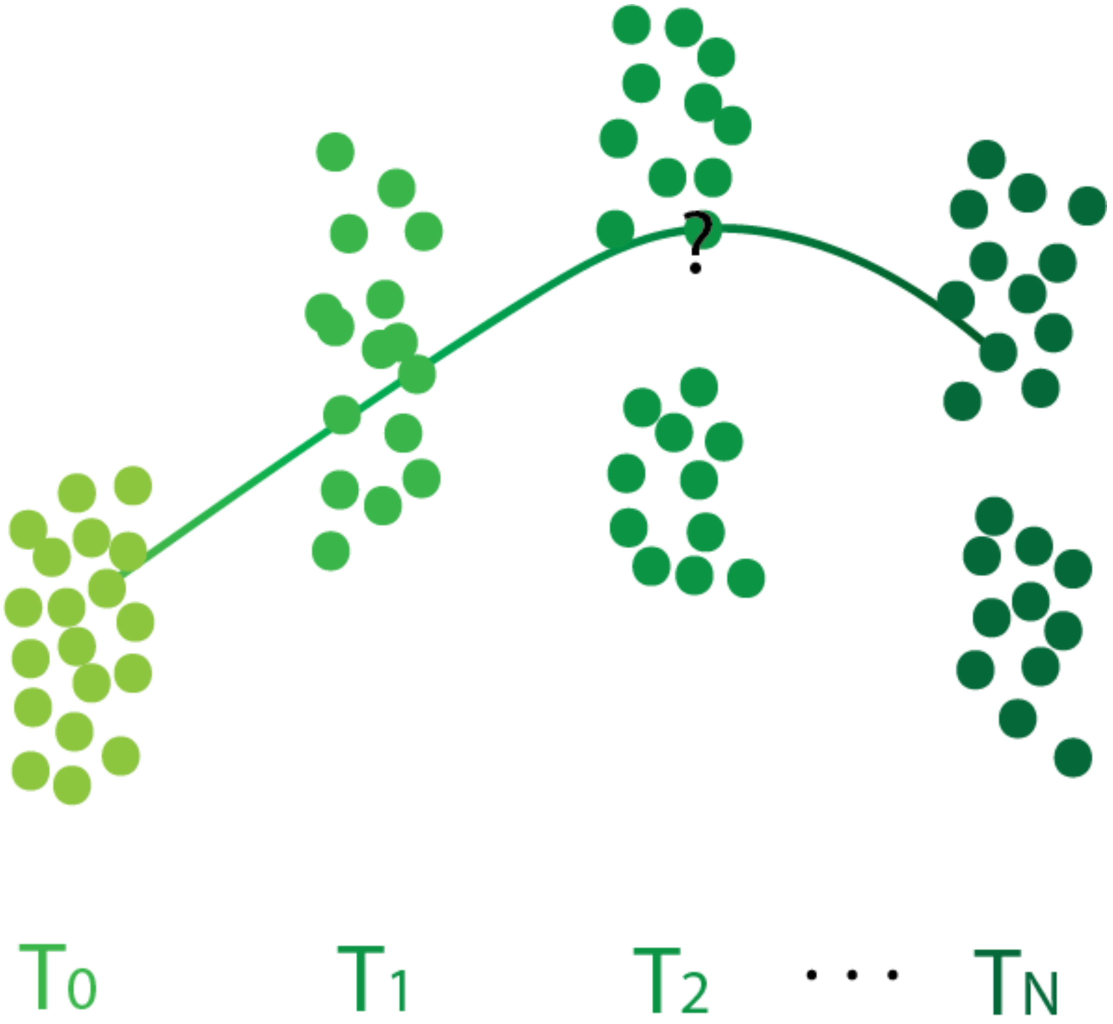
Longitudinal inference from cross sectional snapshot measurements

Motivation

Longitudinal inference from cross sectional measurements

Tasks:

- Predict trajectory of a point

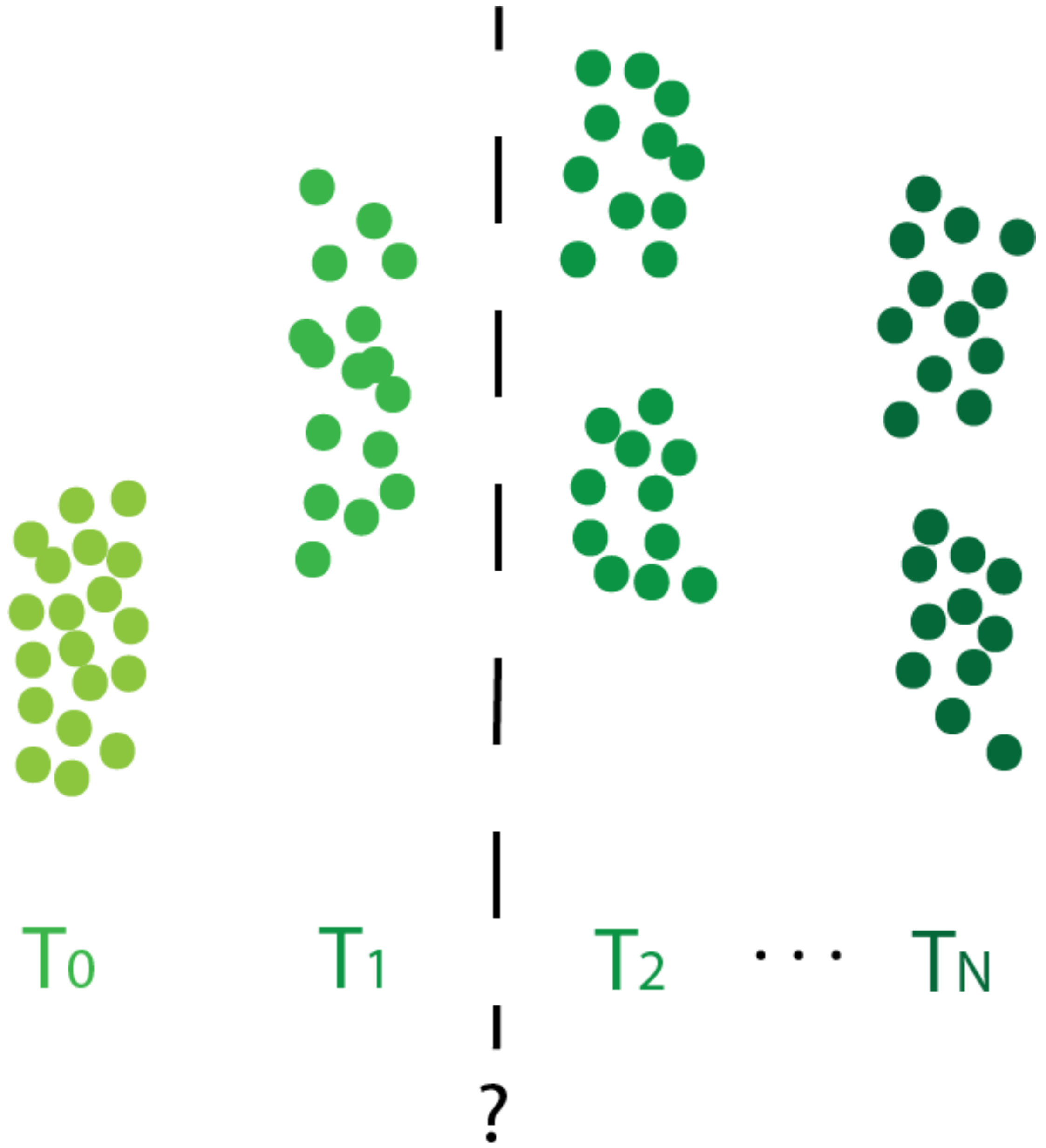


Motivation

Longitudinal inference from cross sectional measurements

Tasks:

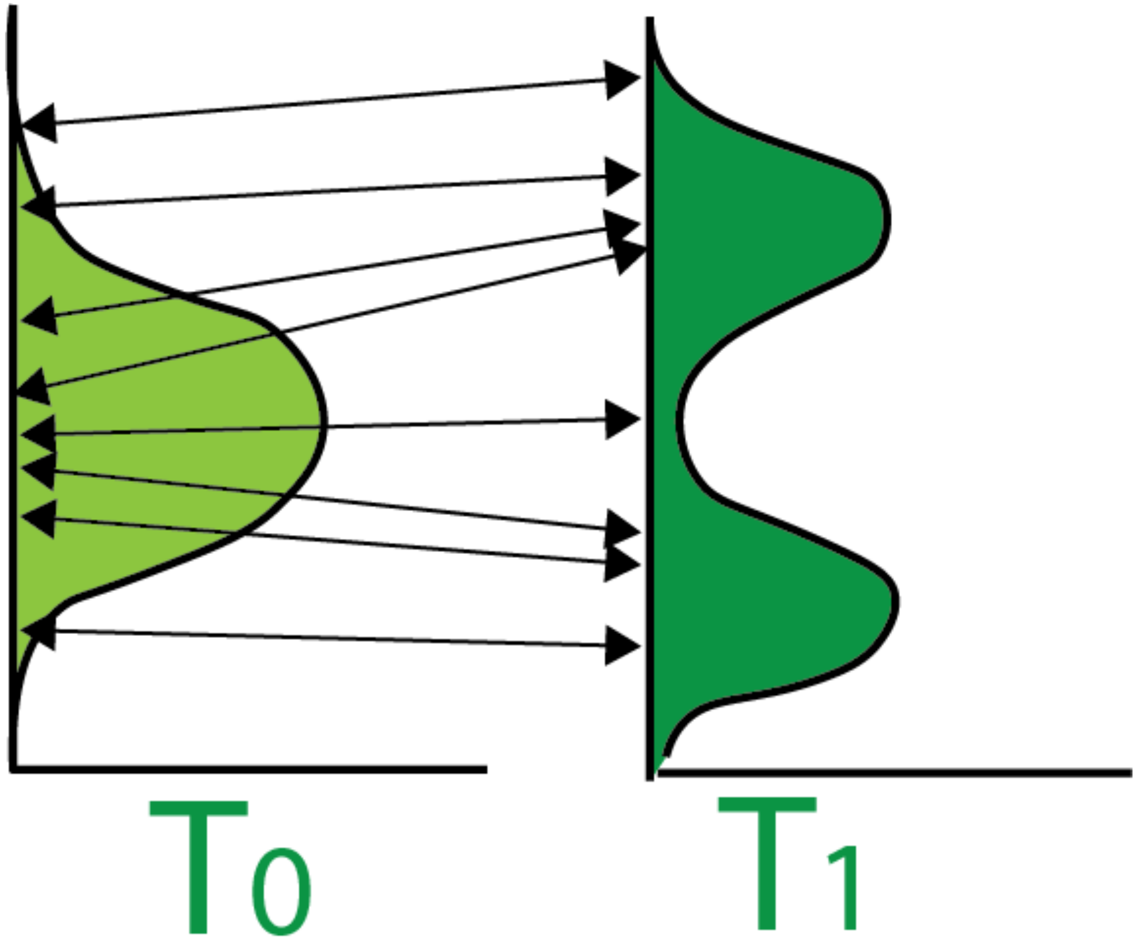
- Predict trajectory of a point
- Predict distribution at test timepoint



Normalizing Flows (NFs)

- Begin with a simple distribution
  $p_{t_0}(x) \sim \mathcal{N}(0,1)$
- Apply an invertible transformation(s)
  $x_{t_1} = f(x_{t_0})$
- Use change of variables to calculate probability

$$\log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \log \det \left| \frac{\partial f}{\partial x_{t_0}} \right|$$



Deep Normalizing Flows (NFs)

- Apply a series of transformations

$$x_{t_1} = f(x_{t_0}) \quad \Rightarrow \quad x_{t_N} = f_N \circ f_{N-1} \circ \cdots \circ f_1(x_{t_0})$$

- Use change of variables to calculate probability

$$\log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \log \det \left| \frac{\partial f}{\partial x_{t_0}} \right| \quad \Rightarrow \quad \log p_{t_N}(x_{t_N}) = \log p_{t_0}(x_{t_0}) - \sum_{n=1}^{N} \log \det \left| \frac{\partial f_n}{\partial x_{t_{n-1}}} \right|$$

Continuous Normalizing Flows

$$x_{t_N} = f_N \circ f_{N-1} \circ \cdots \circ f_1(x_{t_0}) \quad \Rightarrow \quad x_{t_1} = F(x_{t_0}) = \int_{t_0}^{t_1} f(x(t), t)\mathrm{d}t$$

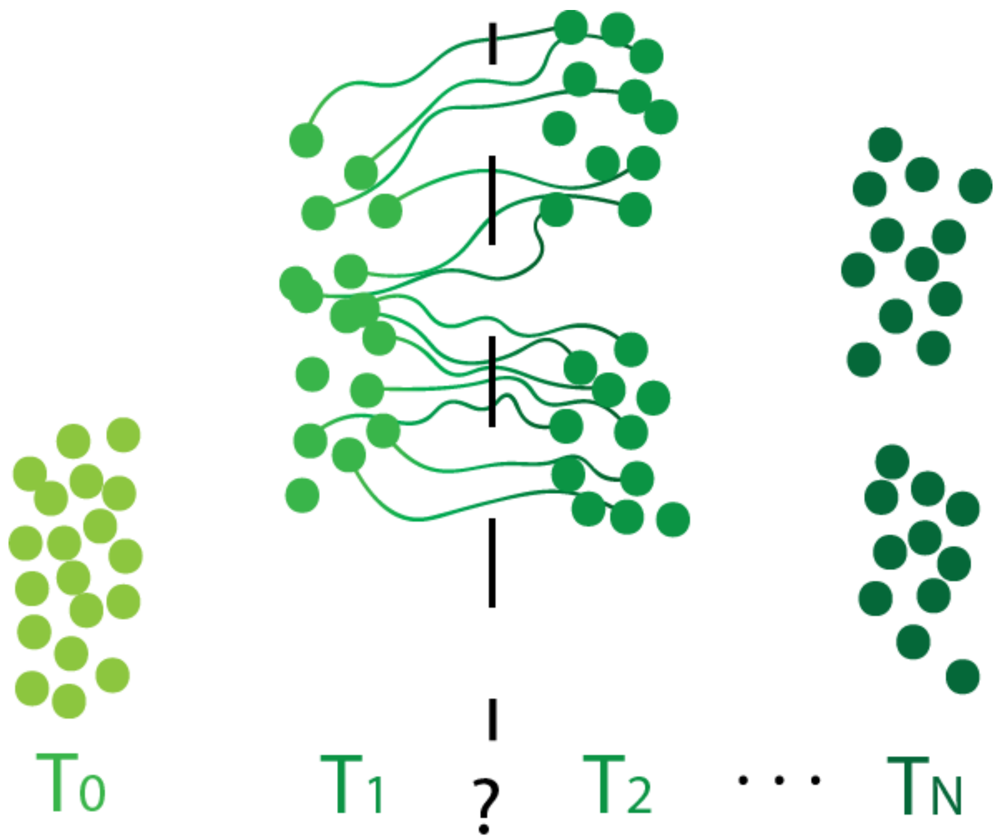
$$\log p_{t_N}(x_{t_N}) = \log p_{t_0}(x_{t_0}) - \sum_{n=1}^{N} \log\det\left|\frac{\partial f_n}{\partial x_{t_{n-1}}}\right| \quad \Rightarrow \quad \log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \int_{t_0}^{t_1} \mathrm{Tr}\left(\frac{\partial f}{\partial x(t)}\right)\mathrm{d}t$$

Cannot model:

$$F(x) = -x$$

[Chen et al. 2018]

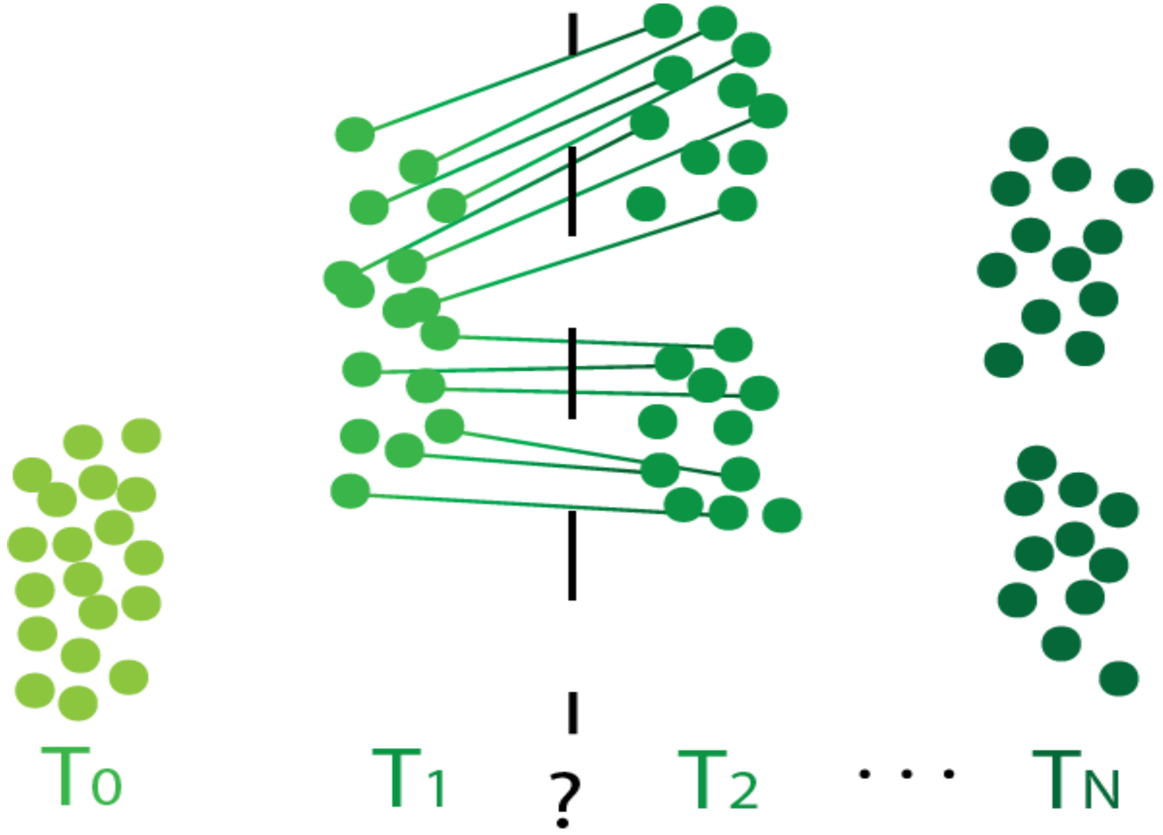
CNFs create continuous paths



Creates continuous paths, but they may not be biologically plausible — no restriction on circuitous paths!

[Chen et al. 2018]

Obtaining straight paths via regularization



Penalize path energy: the squared L2-norm of the derivatives

Regularized CNF approximates dynamic optimal transport

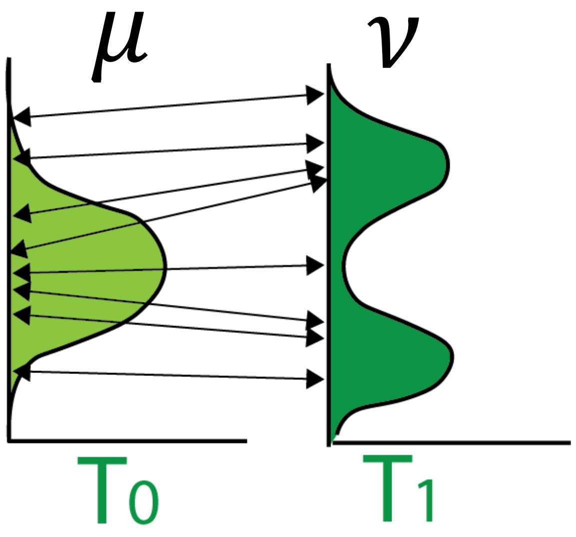
Dynamic OT:

$$W(\mu,\nu)_2^2 = \underbrace{\inf_{p,f} \int_{\mathbb{R}^d} \int_0^1 p(x,t)\, \|f(x,t)\|\, \mathrm{d}t\mathrm{d}x}_{\text{Eulerian frame}} = \underbrace{\inf_{p,f} \mathop{\mathbb{E}}_{x_0 \sim \mu} \int_0^1 \|f(x(t),t)\|\, \mathrm{d}t}_{\text{Lagrangian frame}}$$

Subject to:

$$\underbrace{\partial_t p + \nabla \cdot (pf) = 0}_{\text{Mass is preserved}}$$

$$p(\cdot, 0) = \mu \qquad p(\cdot, 1) = \nu$$



Benamou and Brenier 2000

Regularized CNF approximates dynamic optimal transport

~~Dynamic OT:~~ regularized CNF

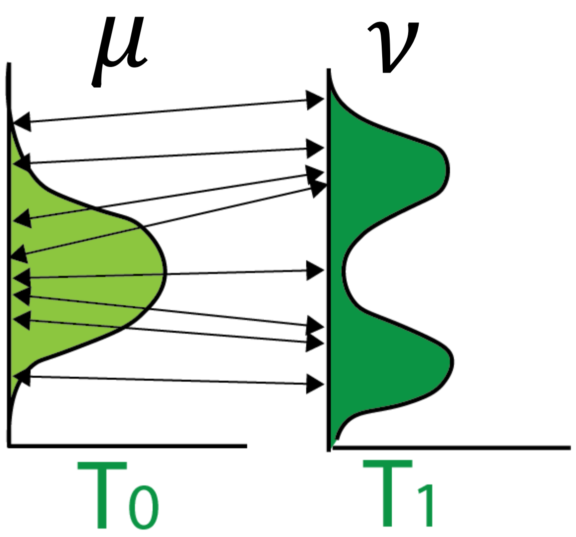
$$W(\mu,\nu)_2^2 = \underbrace{\inf_{p,f} \int_{\mathbb{R}^d} \int_0^1 p(x,t)\, \|f(x,t)\|\, \mathrm{d}t\mathrm{d}x}_{\text{Eulerian frame}} = \underbrace{\inf_{p,f} \mathop{\mathbb{E}}_{x_0\sim\mu} \int_0^1 \|f(x(t),t)\|\, \mathrm{d}t}_{\text{Lagrangian frame}}$$

$$+\ \mathrm{KL}(p(\cdot,1)\|\nu)$$

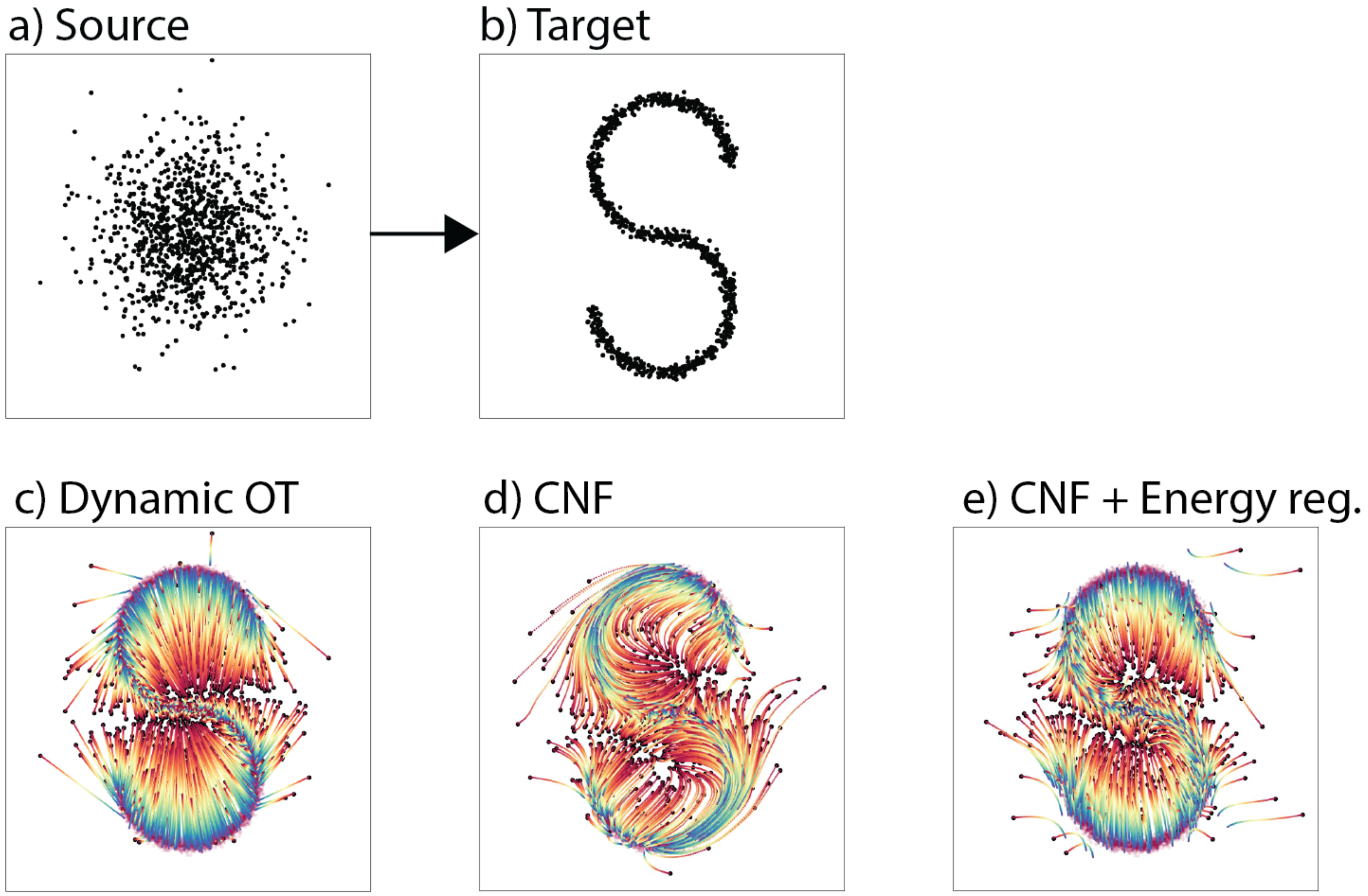
Subject to:

$$\underbrace{\partial_t p + \nabla\cdot(pf) = 0}_{\text{Mass is preserved}}$$

$p(\cdot,0) = \mu$ ~~$p(\cdot,1) = \nu$~~



CNFs model Dynamic Optimal Transport

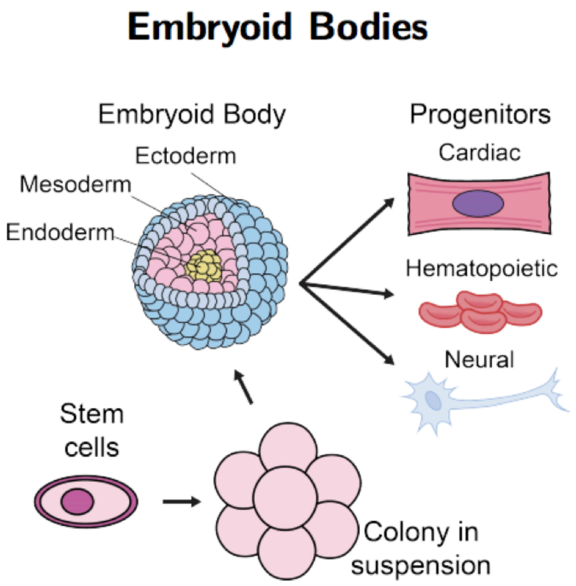


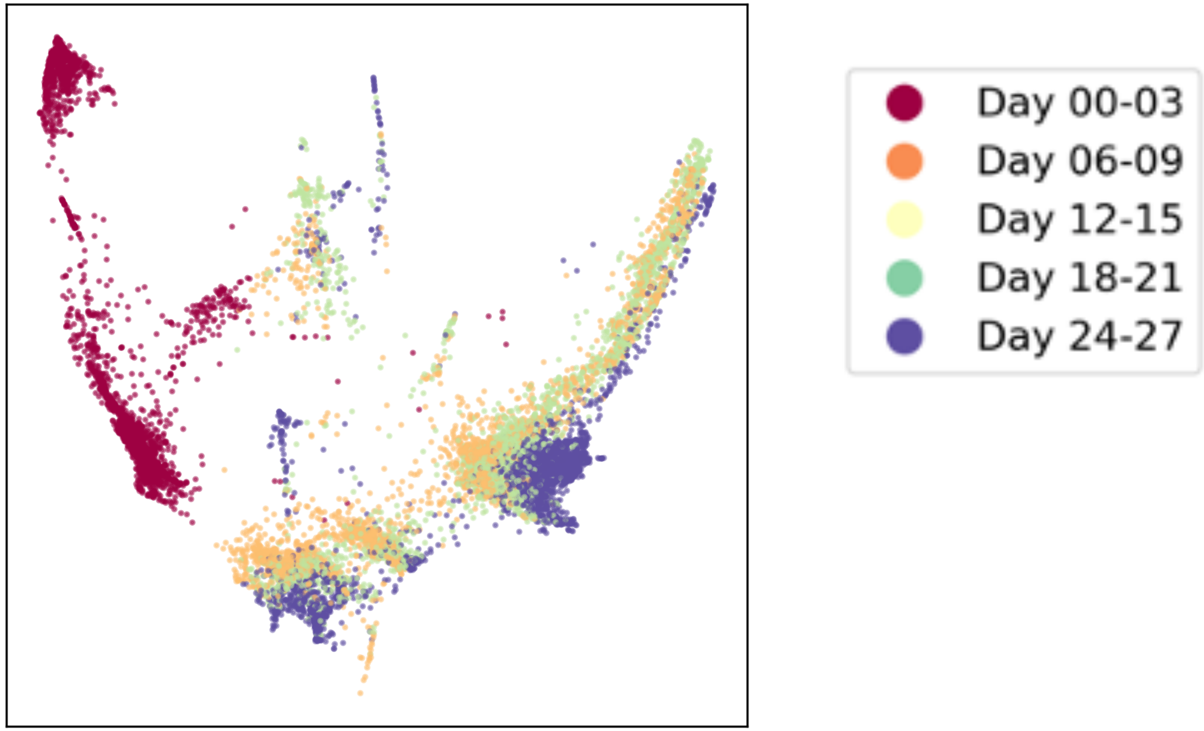
Dynamic OT via TrajectoryNet

- Dynamic OT via TrajectoryNet can be utilized to infer continuous trajectories of any populations adhering to energy or transport constraints
  - Population migration
  - Disease spread

- However, cellular systems are more constrained, and other domain specific priors apply

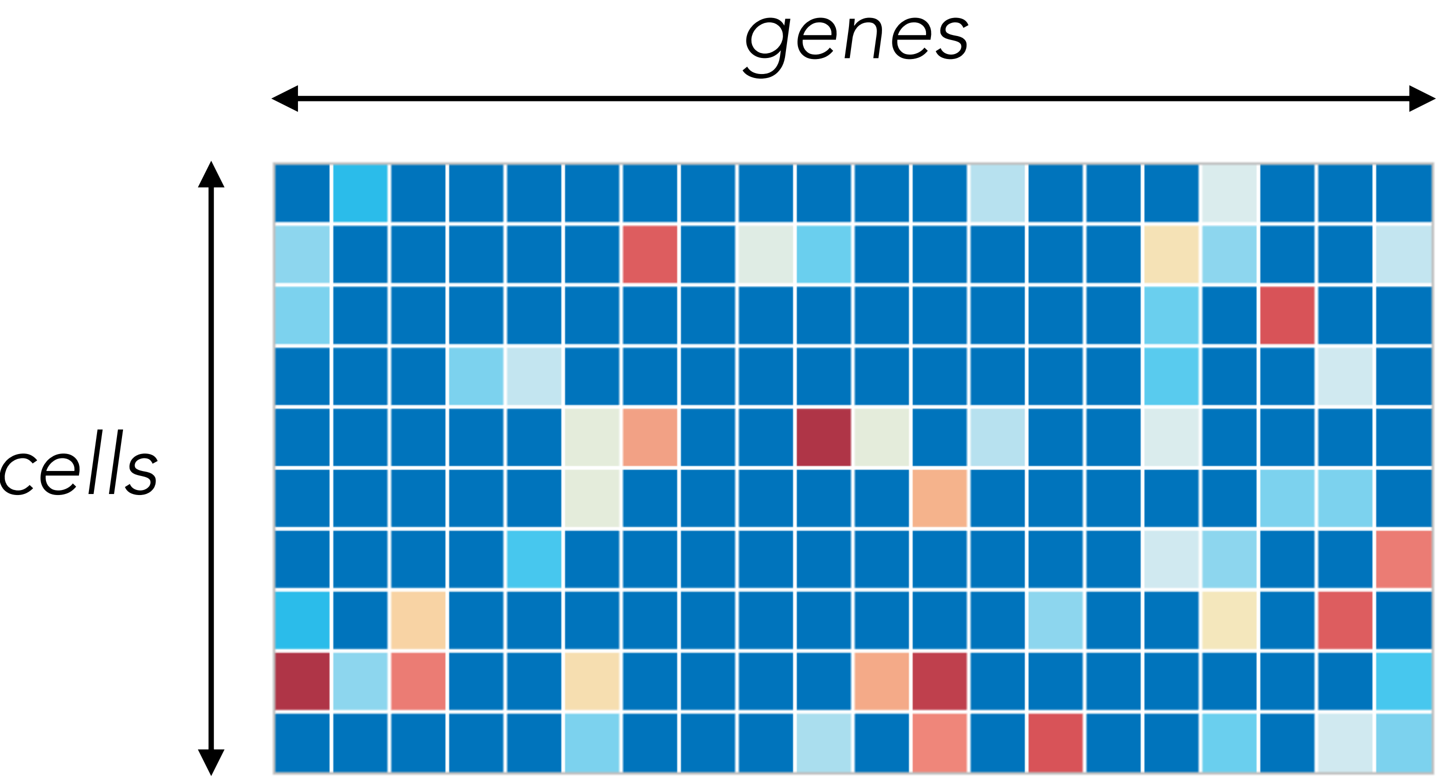
Single Cell Embryonic Stem Cell Data

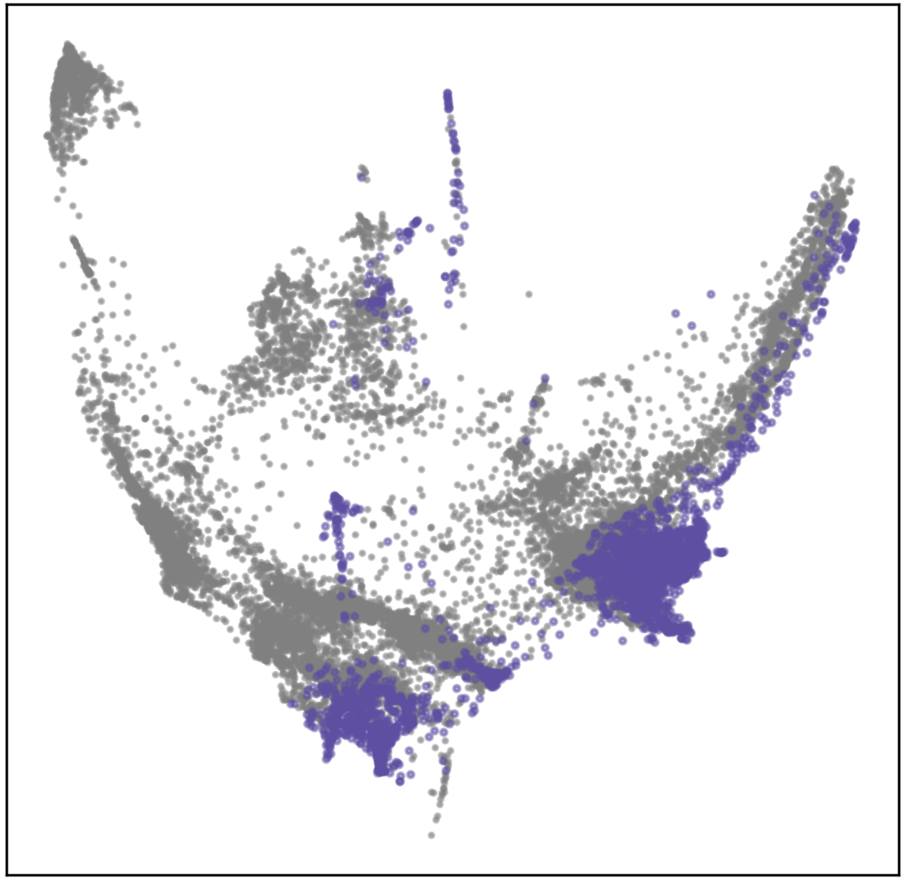




27 day timecourse collected at 5 timepoints, measurements destroy cells at each timepoint (same cell cannot be measured at more than one timepoint)

Inferring Continuous Flow in Static Snapshots



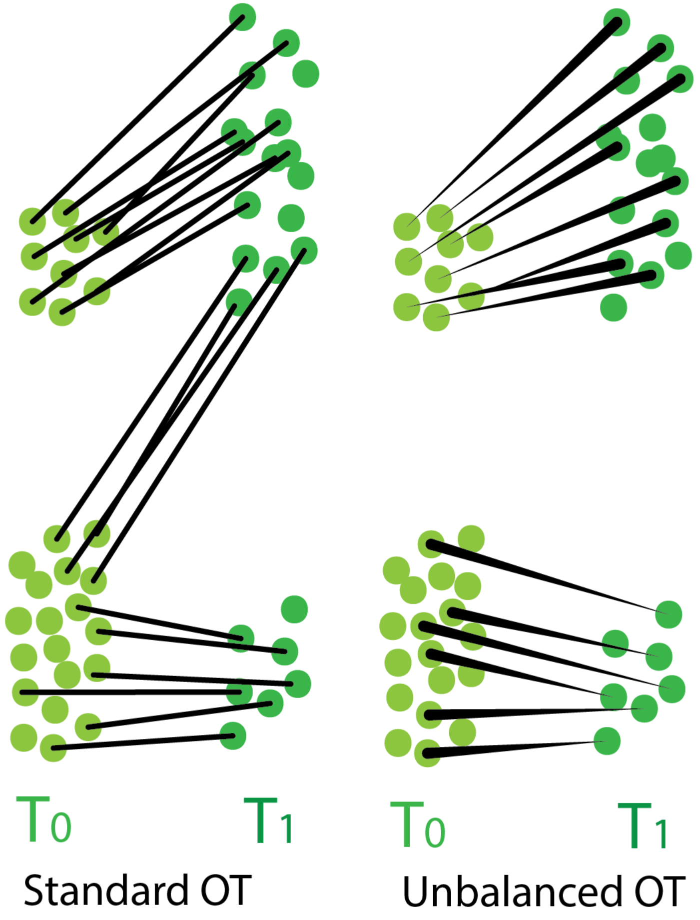


Additional Properties of cells

1. Cells are not simply transported from one timepoint to another, ***they cells divide and die.***
2. Cells cannot travel in straight paths through Euclidean space in terms of measured dimensions, ***cells only travel along a cellular manifold.***
3. Though cells are destroyed when measured, we can estimate their direction of transition-based ***RNA velocity***

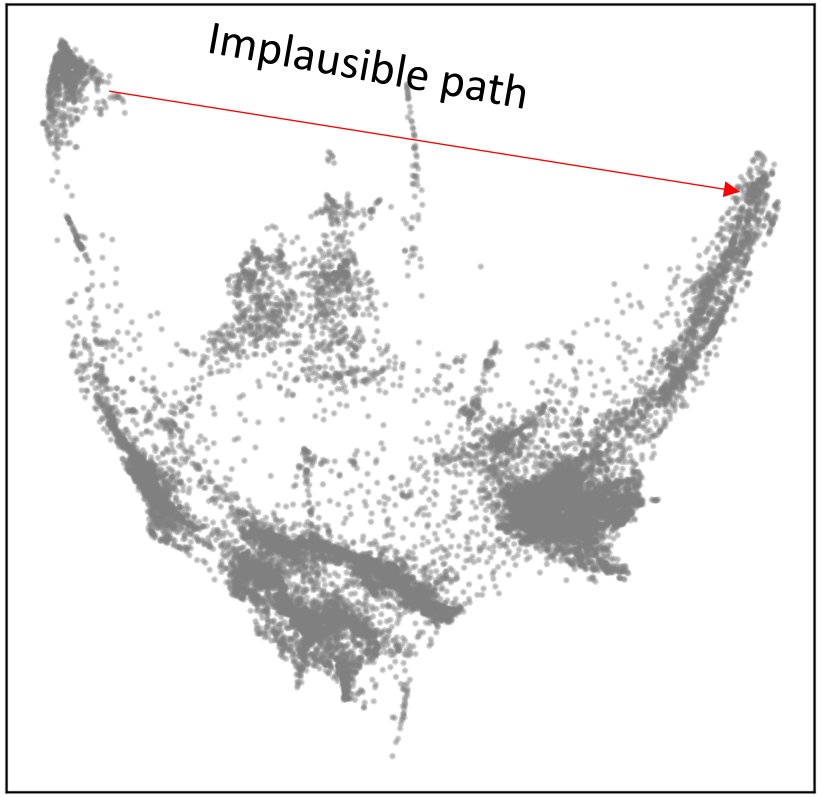
Cell Death and Growth

- Allowing unbalanced transport can let cells “die” instead of moving them to implausible locations
- Unbalanced transport hard to achieve dynamically
- We use discrete optimal transport to assign growth and death rates



Liero et al. 2018

Cellular Manifolds



Cells have to transition through allowable parts of the state space

Enforce this with a density penalty. Based on a knn density estimate

$$L_{density}(x,t) = \sum_k \max(0, \text{min-k}\big(\{\|x(t) - z\| : z \in \mathcal{X}\}\big) - h)$$

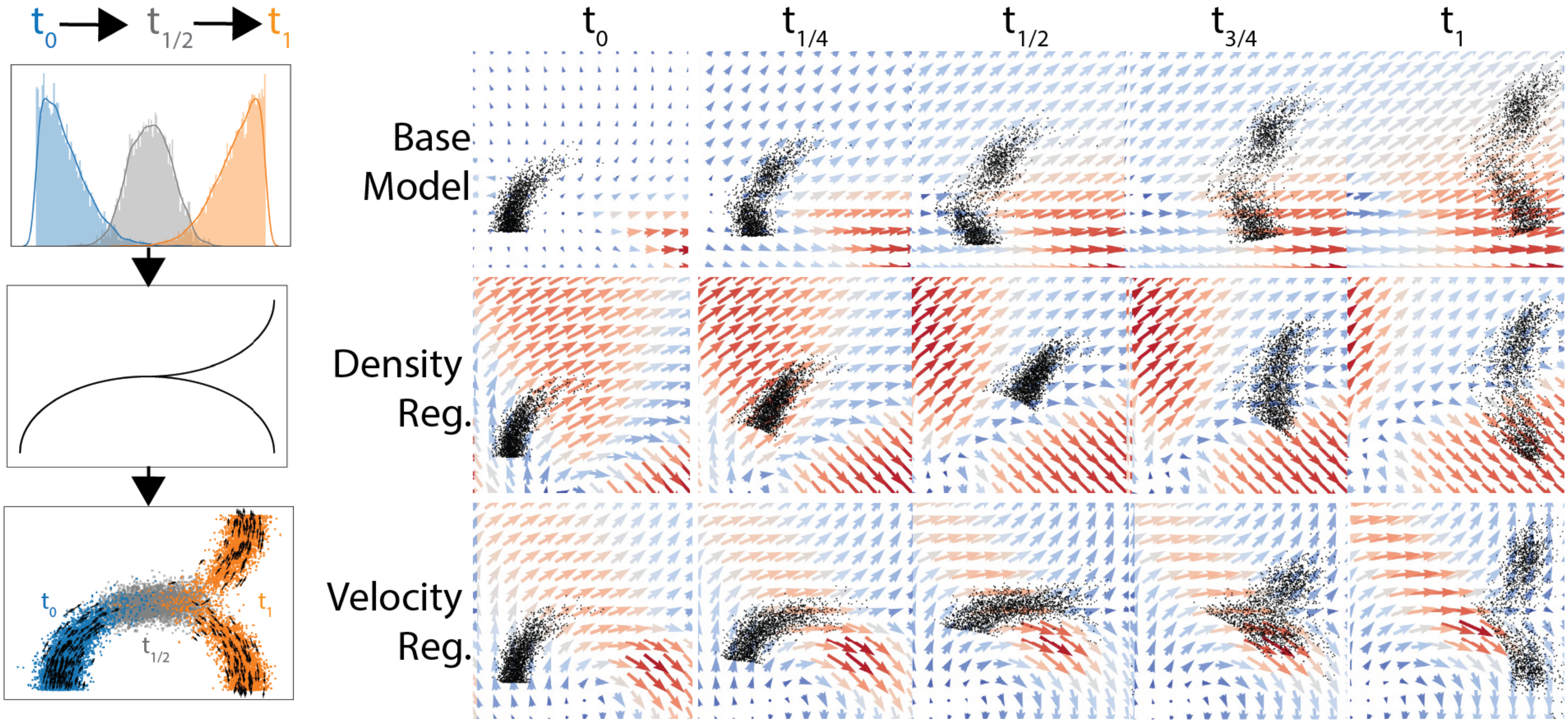
Velocity Regularization

RNA Velocity, estimate of direction of change

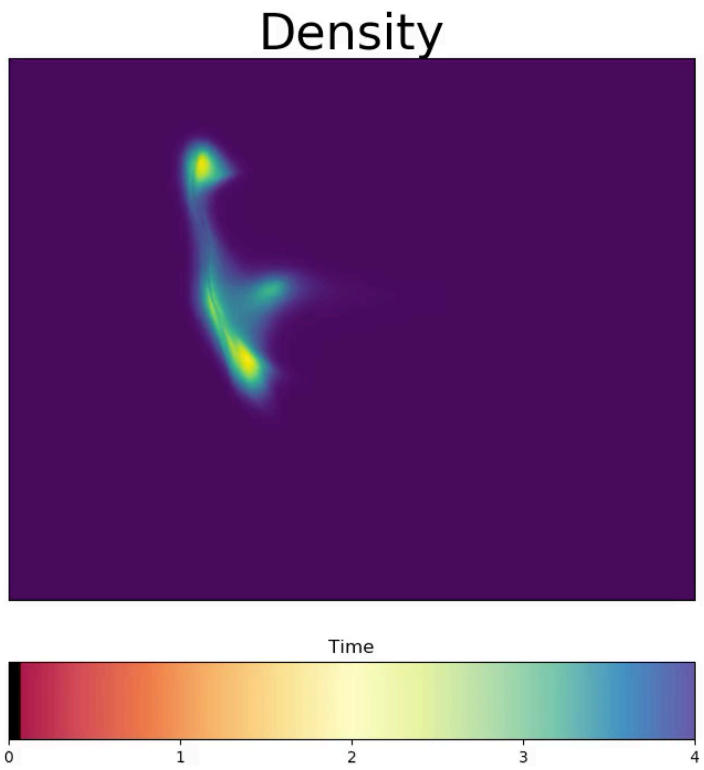
[La Manno et al. 2018 Velocyto; Volker et al. ScVelo]

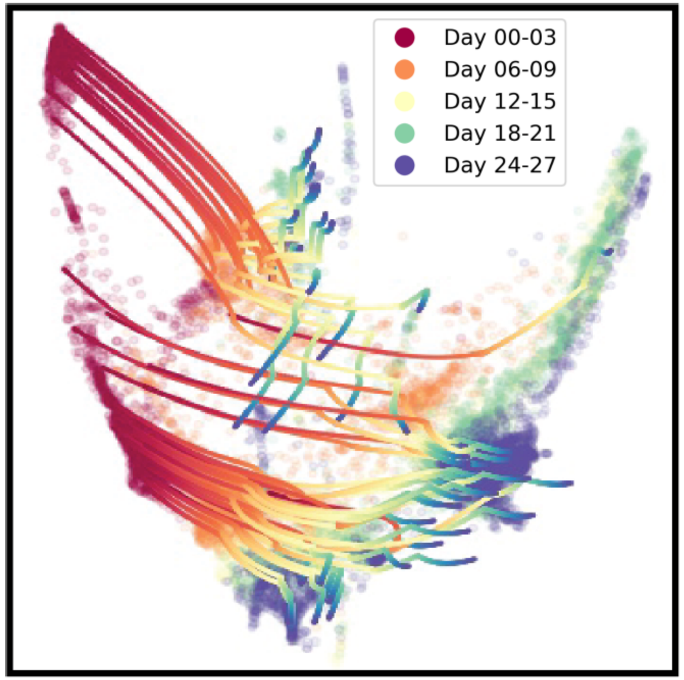
$$(x, t, \widehat{dx/dt}) = \frac{f(x,t) \cdot \widehat{dx/dt}}{\|f(x,t)\| \left\| \widehat{dx/dt} \right\|}$$

Toy Example


t0 → t1/2 → t1
t0
t1
t1/2
t0
t1/4
t1/2
t3/4
t1
Base Model
Density Reg.
Velocity Reg.

Continuous Trajectories in Single Cell Data





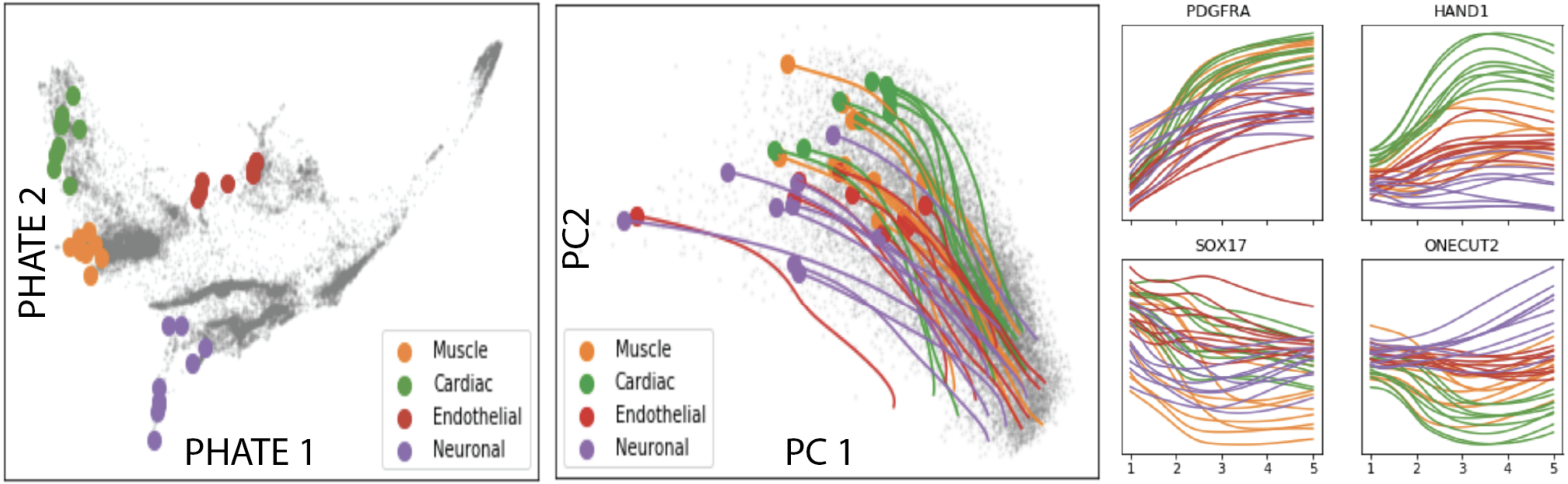
Single Cell Trajectories

Results — Embryoid body dataset

	t=1	t=2	t=3	mean
Base	0.764	0.811	0.863	0.813
Base + D	0.759	**0.783**	0.811	**0.784**
Base + V	0.816	0.839	0.865	0.840
Base + D + V	0.930	0.806	**0.810**	0.848
Base + E	. 0.737	0.896	0.842	0.825
Base + G	**0.700**	0.913	0.829	0.814
OT	0.791	0.831	0.841	0.821
prev	1.715	1.400	0.814	1.309
next	1.400	0.814	1.694	1.302
rand	0.872	1.036	0.998	0.969

- Wasserstein distance between predicted and true distributions for different left out timepoints
- Different regularizations have different assumptions and tradeoffs

Tracing Ancestry



Summary

- Energy regularized CNF performs dynamic optimal transport to find flows between cross-sectional populations

- TrajectoryNet includes additional regularizations that allow for optimal transport on a manifold, with growth and death of individuals over time, and respecting individual velocity data

- Trajectories of individual cells, and gene expression activity can be inferred

Thanks!

Code: https://github.com/krishnaswamylab/TrajectoryNet

Paper: https://arxiv.org/abs/2002.04461

Lab Website: https://www.krishnaswamylab.org