

Modeling Cellular Dynamics with Continuous Normalizing Flows

Alexander Tong¹, Jessie Huang¹, Guy Wolf², David van Dijk^{1†}, and Smita Krishnaswamy^{1†}

¹ Yale University, ² Université de Montréal, [†] Equal Contribution.

Problem Statement

Single-cell technologies are destructive making it impossible to measure a cell lineage.

The goal is to model individual cells as dynamic and *continuously* evolving entities over gene space from population level data.

This gives a continuous picture of lineage from coarse-grained time measurements.

Main Idea

Constrain continuous normalizing flows to match assumptions on cellular dynamics.

Background

Dynamic Optimal Transport [1] Defined as the flow with the minimum energy from a source distribution to a target distribution.

This relates the kinetic energy of a fluid flow to the Wasserstein distance.

$$\partial_t \rho + \nabla \cdot (\rho f) = 0, \quad \rho(t_0, \cdot) = \rho_0, \quad \rho(t_1, \cdot) = \rho_1$$

$$W(\rho_0, \rho_1)_2^2 = \inf_{\rho, f} (t_1 - t_0) \int_{\mathbb{R}^d} \int_{t_0}^{t_1} \rho(x, t) \|f(x, t)\|^2 dt dx$$

Continuous Normalizing Flows [2] optimize efficiently over paths between distributions.

The network predicts the instantaneous cell state change based on current state and time.

Defines a density over states for a given time.

Trained using maximum likelihood.

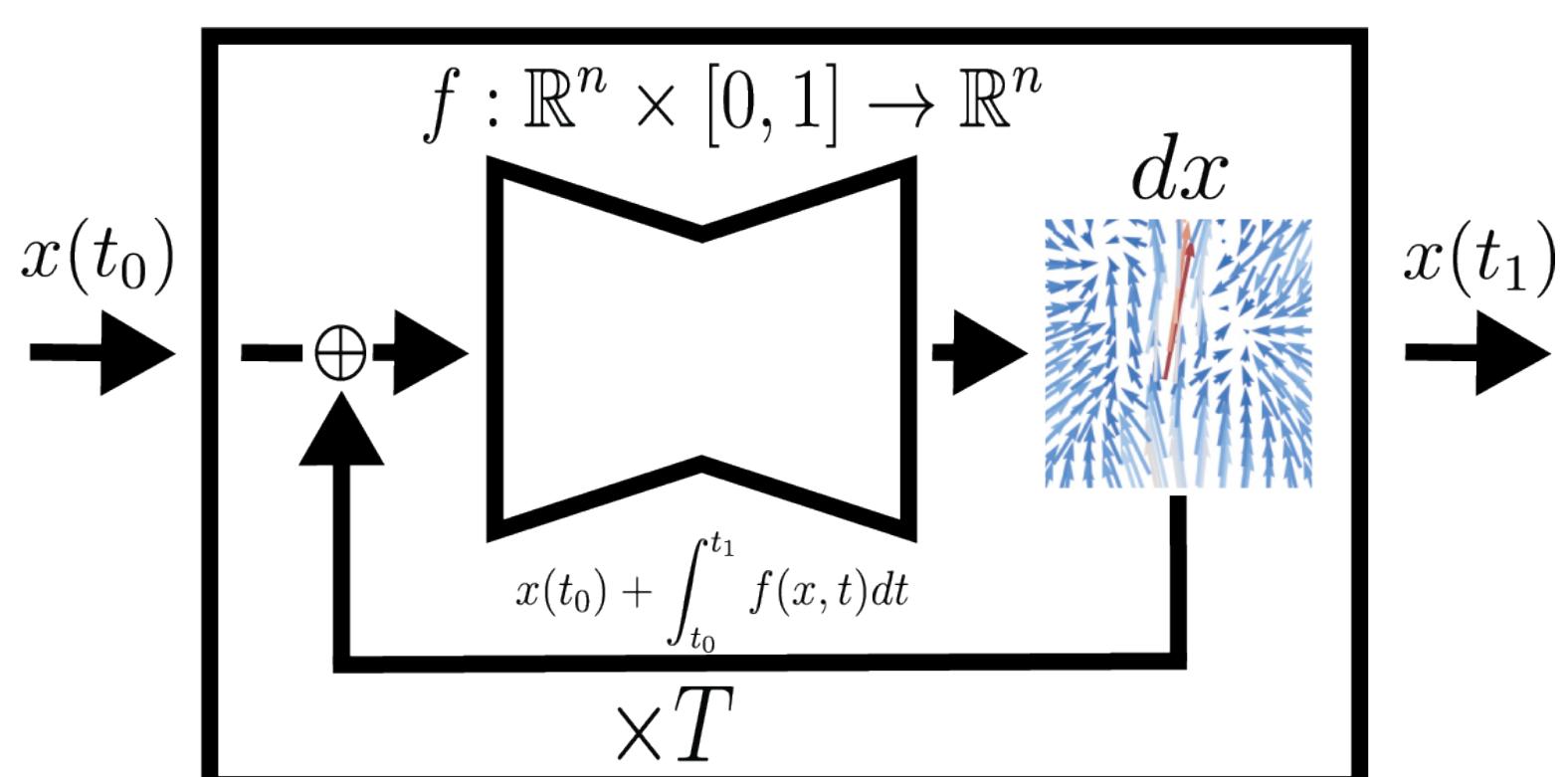


Fig 7: Continuous normalizing flow architecture. A CNF models the instantaneous change at every cell state / time. This defines a flow over time from the initial density to t_1 .

Results

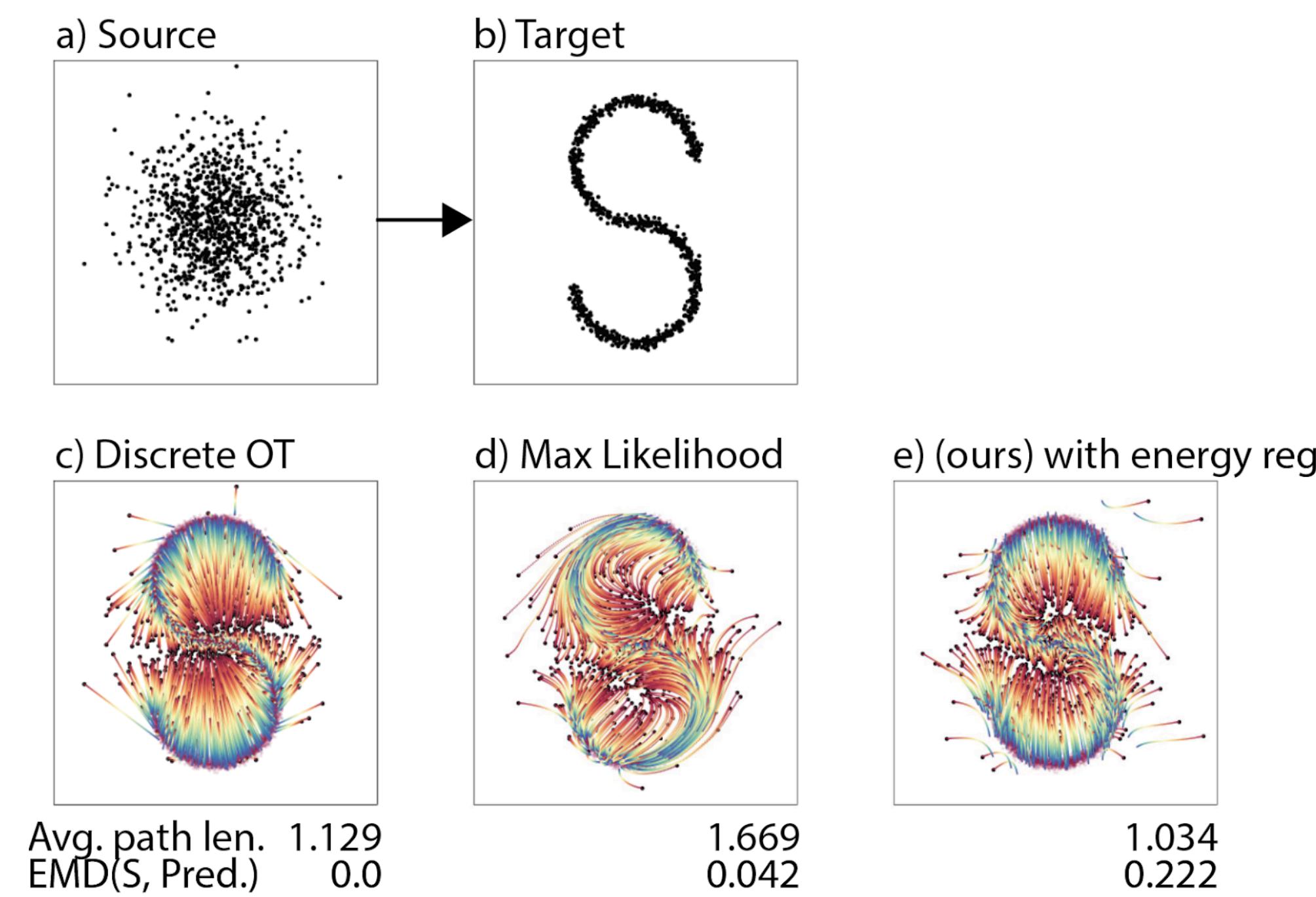


Fig. 1: Paths learned from a gaussian to S distribution (a-b) for three methods: (c) a discrete OT optimization (d) an unregularized continuous normalizing flow (CNF) and (e) a regularized CNF using R_{energy} . Adding R_{energy} straightens paths by minimizing average kinetic energy over the paths.

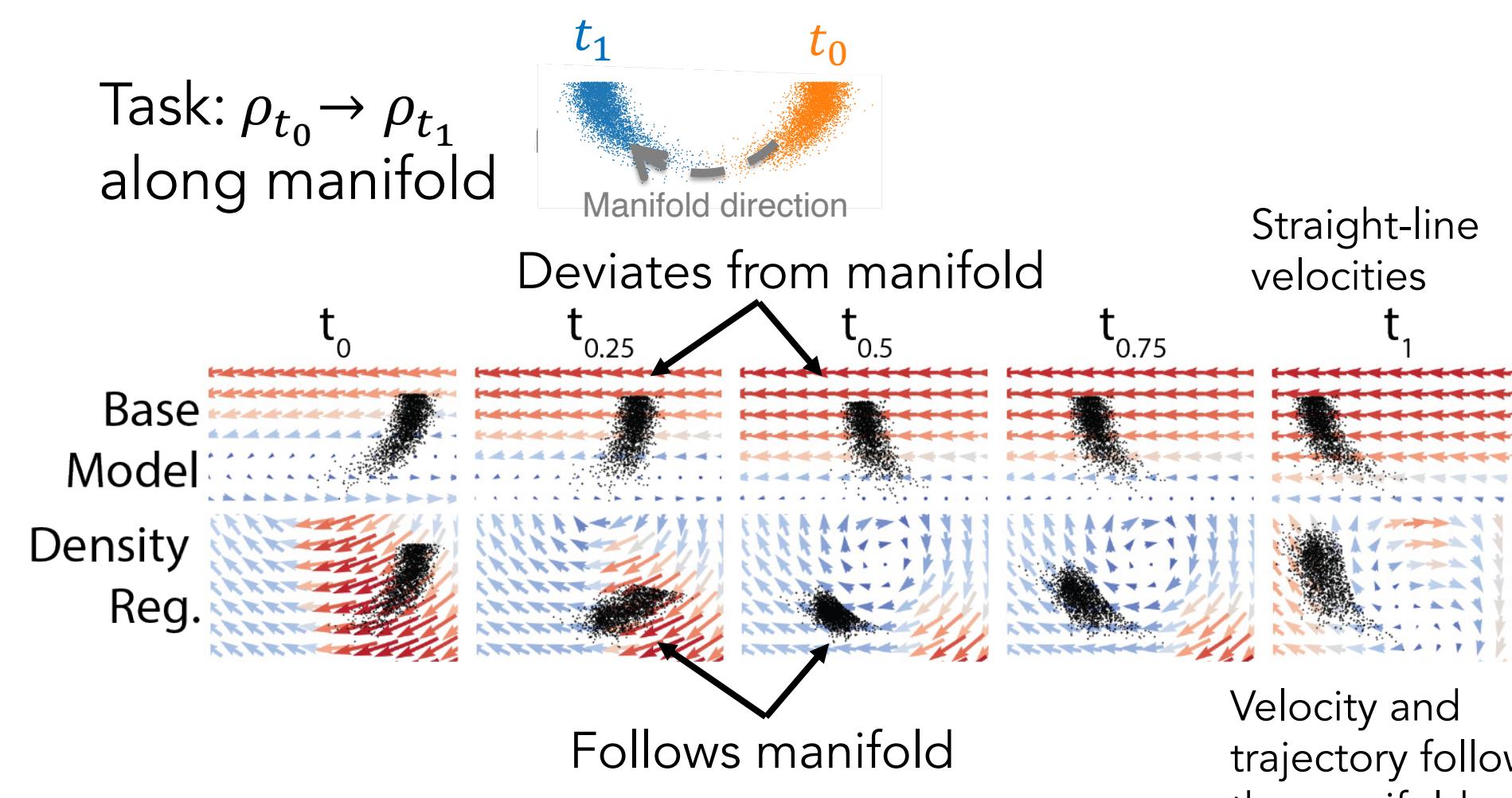


Fig. 2: For a simple 1d manifold embedded in 2d adding R_{density} encourages following the manifold over time.

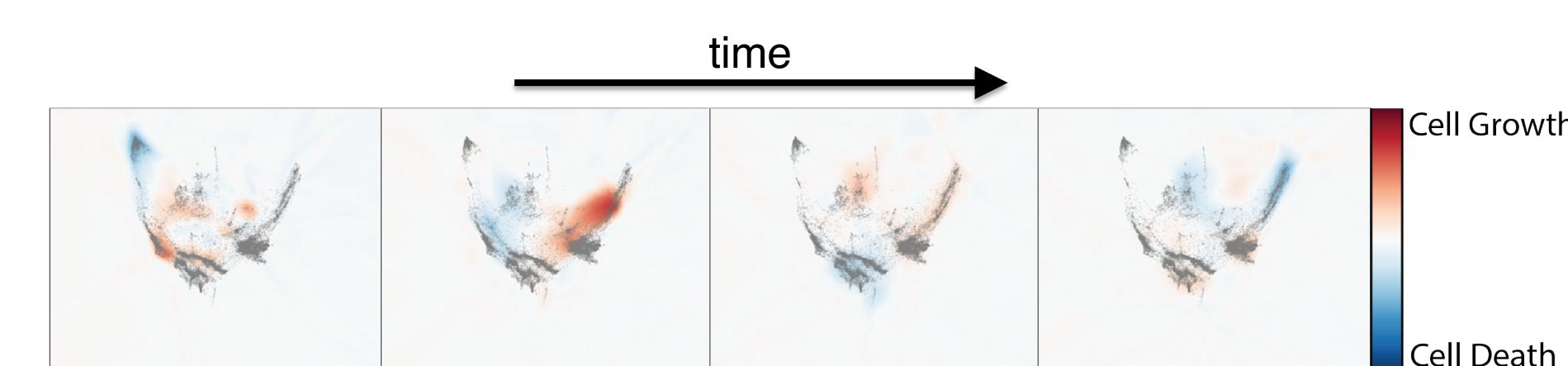


Fig. 3: To model cell growth / death in a continuous system we learn a growth function $g : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$ from unbalanced OT.

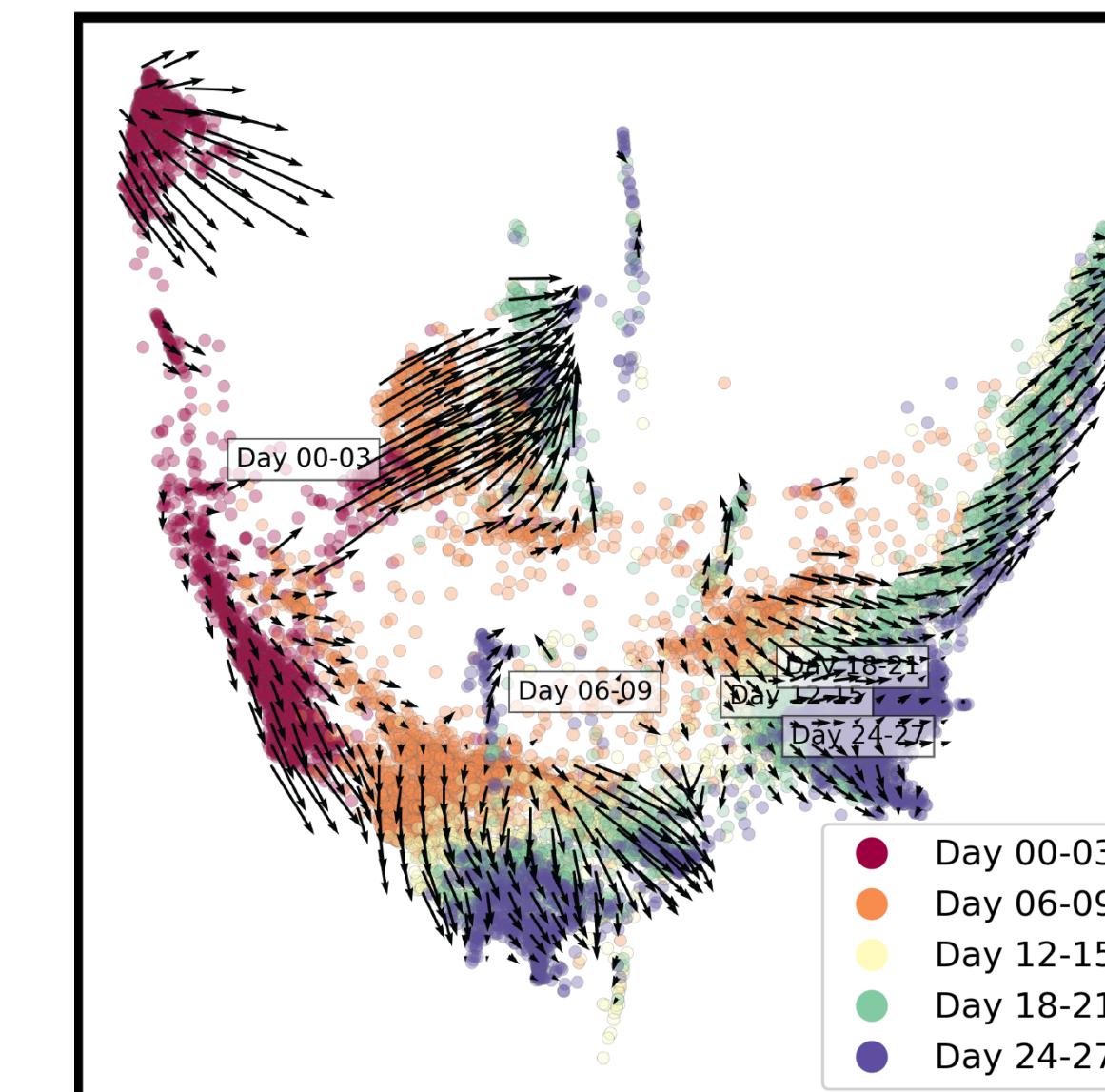


Fig. 4: Embryoid body data [3] with RNA velocity provides an instantaneous time direction of change at each cell. We regularize the cosine angle between RNA velocity and our model at all observed cells / timepoints.

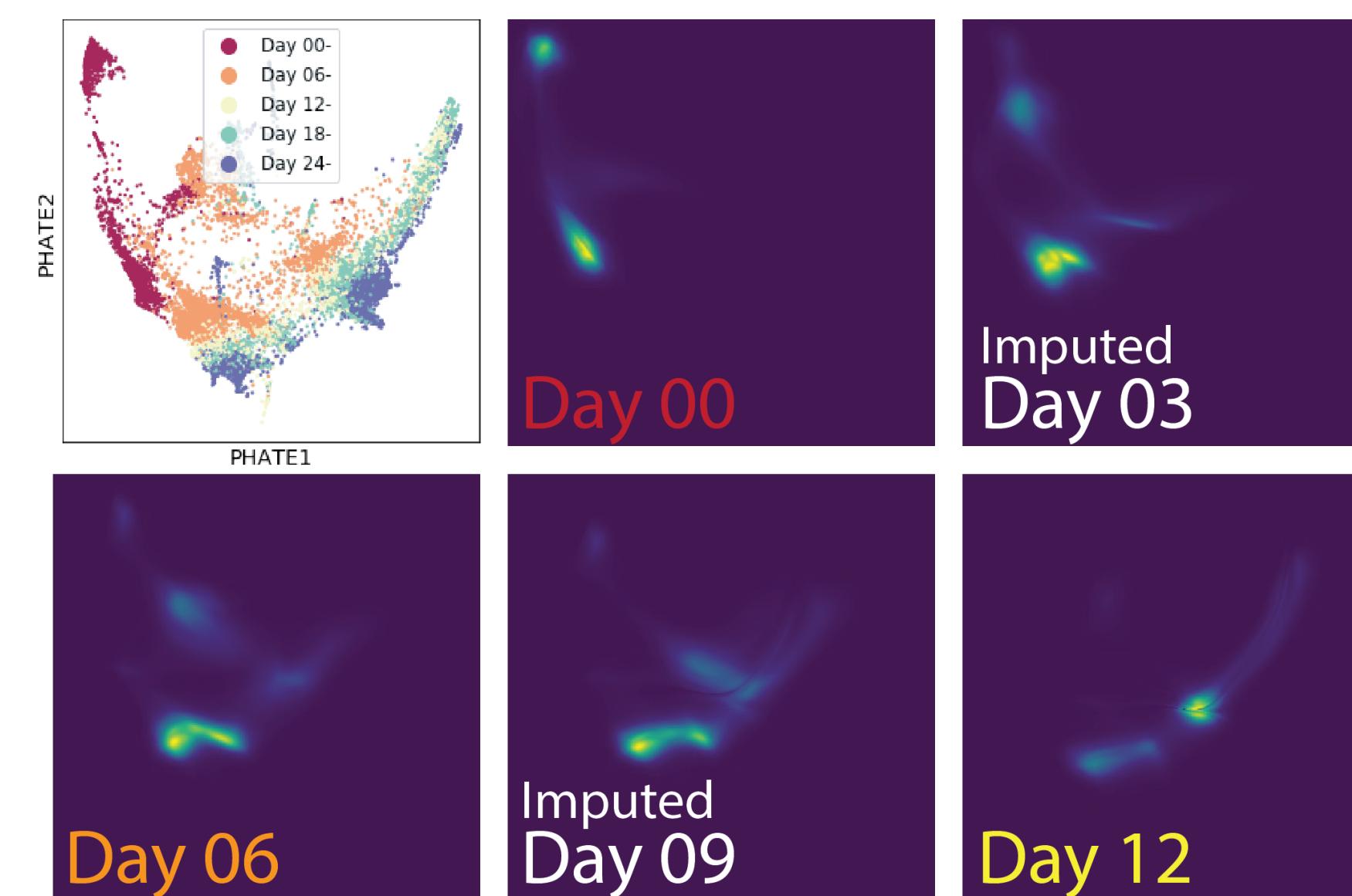


Fig. 5: TrajectoryNet models a density over time. At any time that density is tractable and at measured timepoints matches the measured distribution.

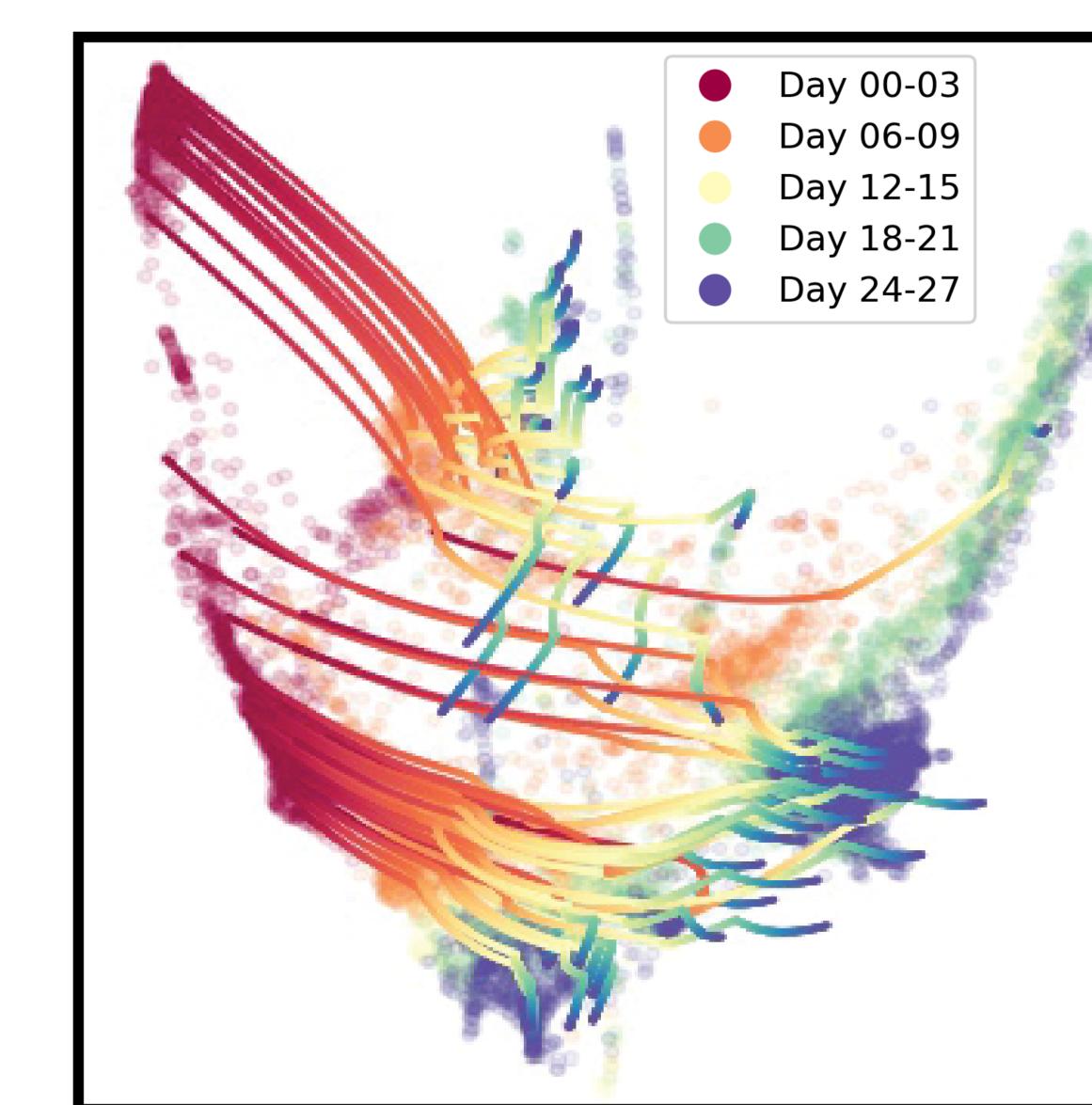


Fig. 6: Learned cell trajectories from TrajectoryNet. Each trajectory is continuous in time and gene space. Collectively paths match density at measured times.

Method

We approximate dynamic optimal transport over gene space with continuous normalizing flows.

We add the following regularizations to build in priors on cellular dynamics.

R_{energy} – Minimize kinetic energy and particle time derivative terms encouraging more Euclidean optimal and energy efficient paths.

R_{density} – At random time interpolations minimize distance to K nearest neighbors for each point preferring trajectories close to the manifold of observed cells.

R_{velocity} – Minimize angle between RNA velocity data and flow at every measured cell building in knowledge of RNA splicing data.

Finally, we learn a growth function g for modeling the cell growth rate using unbalanced optimal transport (see Fig. 3).

Conclusions

Our results relate continuous normalizing flows to dynamic optimal transport and create a flexible model for single cell population modeling.

Our model allows for interpolation of trajectories to unobserved cell states inferring future and past states of individuals from population data.

Further exploration is needed in efficient learning of stochastic and unbalanced models of cell populations.

References

- [1] Benamou, J.-D. & Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* 84, (2000).
- [2] Grathwohl, W. et al. FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models. in *ICLR* (2019).
- [3] Moon, K. R. et al. Visualizing Structure and Transitions for Biological Data Exploration. *Nature Biotech.* (2019).

Further information

Email: alexander.tong@yale.edu

Code: github.com/KrishnaswamyLab

Supported by IVADO (l'institut de valorisation des données) and the Chan-Zuckerberg Initiative.