A145 · GeoSinkhorn Flow: Geometry-Aware Flow Matching for Conditional Dynamics in Single cell Data Phenoscapes

Presenter: Ke Xu — Yale University, Department of Computer Science

Authors: Ke Xu, Shuang Ni, Alistair Wilkinson, Guillaume Huguet, Christopher J. Tape, Guy Wolf, Mark B. Gerstein, Smita Krishnaswamy

Session: Day 1 · Thu Oct 1, 2026 · 9:45–10:45 AM

Single-cell technologies increasingly enable profiling of biological systems across many perturbations, but understanding how a system reorganizes across a space of conditions is difficult to read off single-cell embeddings alone. Phenoscapes address this by embedding entire cell-state distributions—rather than individual cells—as points in a landscape, such that one dot represents one experimental condition. Phenoscapes remain a new and underexplored representation, and existing static phenoscapes position conditions relative to one another without specifying how populations move between them. Flow- and trajectory-based models are numerous, but existing approaches operate on individual cells, are not built to respect phenoscape geometry, and typically cannot generalize predicted flows to unseen conditions. Here we introduce , a flow-matching framework specialized to convert a static phenoscape into a dynamic one, in which condition-level points are connected by geodesic, condition-aware transport trajectories consistent with the underlying geometry. uses heat-diffusion geometry to define manifold-aware transport costs and trains a geometric autoencoder whose latent Euclidean distances remain aligned with phenoscape geometry while extending it to unseen cells. By conditioning transport on perturbation identity, generalizes trajectories to new conditions and produces intermediate progression states. An unbalanced optimal transport variant further models changes in relative cell-state abundance together with state movement. Across synthetic manifolds, patient-derived organoid drug responses, Perturb-seq knockdowns, and time-resolved CyTOF signaling, recovered intrinsic geometry, inferred biologically coherent trajectories, and generalized to held-out drug combinations, perturbation genes, and time points. Together, these results establish as a phenoscape-consistent dynamical framework that generalizes across experimental conditions.

GeoSinkhorn Flow: Geometry-Aware Flow Matching for Conditional Dynamics in Single-Cell Data Phenoscapes

Ke Xu[5†], Shuang Ni[1,3,†], Alistair Wilkinson[4,†], Guillaume Huguet[1,2], Christopher J. Tape[4], Guy Wolf[1,2], Mark B. Gerstein[5,7,8,9,10], and Smita Krishnaswamy[5,6,*]

[1]Mila – Quebec AI Institute, Montréal, Quebec, Canada; [2]Department of Mathematics and Statistics, Université de Montréal, Montréal, Quebec, Canada; [3]Department of Computer Science and Operations Research, Université de Montréal, Montréal, Quebec, Canada; [4]UCL Cancer Institute, University College London, London, United Kingdom; [5]Department of Computer Science, Yale University, New Haven, CT, USA; [6]Department of Genetics, Yale School of Medicine, New Haven, CT, USA; [7]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; [8]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; [9]Department of Statistics and Data Science, Yale University, New Haven, CT, USA; [10]Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA.

[†]These authors contributed equally to this work. [*]Correspondence: smita.krishnaswamy@yale.edu.

Abstract.

Single-cell technologies increasingly enable profiling of biological systems across many perturbations, but understanding how a system reorganizes across a space of conditions is difficult to read off single-cell embeddings alone. Phenoscapes address this by embedding entire cell-state distributions—rather than individual cells—as points in a landscape, such that one dot represents one experimental condition. Phenoscapes remain a new and underexplored representation, and existing static phenoscapes position conditions relative to one another without specifying how populations move between them. Flow- and trajectory-based models are numerous, but existing approaches operate on individual cells, are not built to respect phenoscape geometry, and typically cannot generalize predicted flows to unseen conditions.

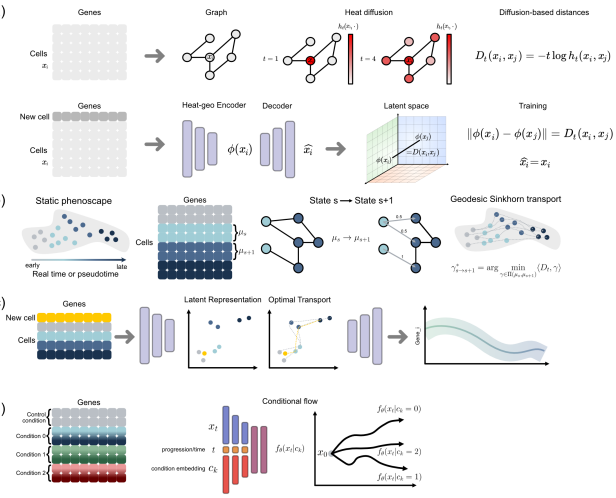

Figure 1: Overview of the GeoSinkFlow pipeline. (a) Heat diffusion defines graph-based cell geometry and supervises a geometry-preserving autoencoder. (b) Geodesic Sinkhorn couples neighboring states along real time or pseudotime. (c) Latent Euclidean transport extends the learned geometry to unseen cells. (d) Conditional flow matching learns perturbation-specific trajectories across progression.

Here we introduce GeoSinkFlow, a flow-matching framework specialized to convert a static phenoscape into a dynamic one, in which condition-level points are connected by geodesic, condition-aware transport trajectories consistent with the underlying geometry. GeoSinkFlowuses heat-diffusion geometry to define manifold-aware transport costs and trains a geometric autoencoder whose latent Euclidean distances remain aligned with phenoscape geometry while extending it to unseen cells. By conditioning transport on perturbation identity, GeoSinkFlowgeneralizes trajectories to new conditions and produces intermediate progression states. An unbalanced optimal transport variant further models changes in relative cell-state abundance together with state movement.

Across synthetic manifolds, patient-derived organoid drug responses, Perturb-seq knockdowns, and time-resolved CyTOF signaling, GeoSinkFlowrecovered intrinsic geometry, inferred biologically coherent trajectories, and generalized to held-out drug combinations, perturbation genes, and time points. Together, these results establish GeoSinkFlowas a phenoscape-consistent dynamical framework that generalizes across experimental conditions.