Dynamic Pseudo-time Warping of Complex Single-Cell Trajectories

Van Hoan Do[4], Mislav Blažević[1], Pablo Monteagudo[4], Luka Borozan[1], Khaled Elbassioni[2], Sören Laue[3], Francisca Rojas Ringeling[4], Domagoj Matijević[1], and Stefan Canzar[4(✉)]

[1] Department of Mathematics, University of Osijek, Osijek, Croatia
[2] Khalifa University of Science and Technology, Abu Dhabi, UAE
[3] Friedrich-Schiller-Universität Jena, Jena, Germany
[4] Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany
canzar@genzentrum.lmu.de

1 Introduction

Single-cell RNA sequencing enables the construction of trajectories [1] describing the dynamic changes in gene expression underlying biological processes such as cell differentiation and development. The comparison of single-cell trajectories under two distinct conditions can illuminate the differences and similarities between the two and can thus be a powerful tool for analysis [2]. Recently developed methods for the comparison of trajectories [2, 3] rely on the concept of dynamic time warping (dtw), originally proposed for the comparison of two time series and consequently restricted to simple, linear trajectories. Here, we adopt and theoretically link arboreal matchings to dtw and implement a suite of exact and heuristic algorithms suitable for the comparison of complex trajectories of different characteristics in our tool Trajan (Fig. 1). Trajan's alignment enables the meaningful comparison of gene expression dynamics along a common pseudo-time scale. Trajan is available at https://github.com/canzarlab/Trajan.

2 Methods

Dynamic time warping (dtw) is the algorithmic workhorse underlying current methods that compare linear single-cell trajectories. We develop Trajan, the first method to compare and align complex trajectories (trees) with multiple branch points. Trajan aligns each path in one tree to at most one path in the second tree and vice versa and, similar to dtw, preserves the order of nodes along the paths. In [4] we have introduced *arboreal matchings* that formalize such a consistent path-by-path alignment of trees.

We devise scoring schemes for arboreal matchings that yield (guaranteed) similar distance measures between linear trajectories as dtw, but naturally

The full version of this paper is available as preprint at bioRxiv 522672.
Van Hoan Do and Mislav Blažević—equal contribution.

L. J. Cowen (Ed.): RECOMB 2019, LNBI 11467, pp. 294–296, 2019.
https://doi.org/10.1007/978-3-030-17083-7

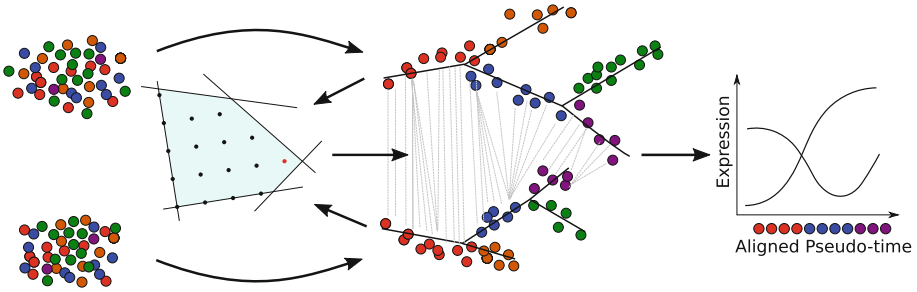

Fig. 1. Complex trajectories, reconstructed from single-cell RNA measurements using, e.g., Monocle 2, are aligned by Trajan based on arboreal matchings. The matching warps individual pseudo-time scales into a shared one along which expression kinetics can be compared.

extend to complex trajectories. Trajan implements a thoroughly engineered branch-and-cut algorithm that allows to practically compare complex single-cell trajectories. It repeatedly determines cutting planes that strengthen the LP relaxation in [4] in polynomial-time and uses an in-house developed, non-commercial, non-linear solver for all continuous optimization problems. For trajectories with a small number of cell fates k we employ a fixed-parameter tractable algorithm, parameterized by k, that applies a dynamic program similar to [5] to align them optimally.

3 Results

Adopting a strategy similar to [2], we re-analyzed two public single-cell datasets: human skeletal muscle myoblast (HSMM) differentiation and human fibroblasts undergoing MYOD-mediated myogenic reprogramming (hFib-MyoD). Trajan is able to align the core paths of each complex trajectory, without any previous knowledge of myoblast differentiation markers. From Trajan's alignment, we construct gene expression kinetics for a set of genes that were assessed in [2] and are able to reproduce their key findings, including the molecular barriers identified in [2] that hinder the efficient reprogramming of fibroblasts to myotubes.

In a perturbation experiment we demonstrate the benefits in terms of robustness and accuracy of our model which compares entire trajectories at once, as opposed to a pairwise application of dtw.

Acknowledgments. Sören Laue has been funded by Deutsche Forschungsgemeinschaft (DFG) under grant LA 2971/1-1. Mislav Blažević was supported in part by BAYHOST. Francisca Rojas Ringeling was supported by the Bavarian Gender Equality Grant (BGF).

References

1. Qiu, X.: Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods **14**(10), 979–982 (2017)
2. Cacchiarelli, D.: Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome. Cell Syst. 1–18 (2016)
3. Alpert, A.: Alignment of single-cell trajectories to compare cellular expression dynamics. Nat. Methods **15**(4), 267–270 (2018)
4. Böcker, S., Canzar, S., Klau, G.W.: The generalized robinson-foulds metric. In: Darling, A., Stoye, J. (eds.) WABI 2013. LNCS, vol. 8126, pp. 156–169. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40453-5_13
5. Zhang, K.: Simple fast algorithms for the editing distance between trees and related problems. SIAM J. Comput. **18**, 1245–1262 (1989)