

Inference of trajectory presence by tree dimension

Lovemore Tenha¹ and Mingzhou Song^{1,2}

¹Department of Computer Science, ²Molecular Biology and Interdisciplinary Life Sciences Graduate Program, New Mexico State University

Contact email: ltenha@nmsu.edu, joemsong@nmsu.edu

Modern biology now exploits massive availability of data to study the nature and dynamics of complex biological processes. The complexity of biological processes such as cell differentiation is reflected in dynamic transitions between cellular states. Trajectory inference arranges the states from an early state to a developed state using gene expression data generated from single-cell biology. However, current trajectory inference methods, all returning a best trajectory, do not address the problem of determining the existence of a trajectory in observed data before inference, leading to uncertainty in inferred trajectories. We introduce the tree dimension statistical test for presence of trajectory in single-cell data. The tree dimension test comprises of dimension measure of Euclidean minimum spanning tree, a test statistic, and a null distribution. The tree dimension measure captures the degree of branching in the minimum spanning tree more effectively than globally insensitive number of leaves or tree diameter indifferent to secondary branches. The test statistic quantifies presence of trajectory, and its null distribution is estimated under the null hypothesis of no trajectory in data. On simulated and real single-cell datasets, the test outperformed the intuitive number of leaves and tree diameter statistics. Tree dimension measure is computable in linear time to tree size and has no user parameter to tune. Our work opens a window to prioritize cellular dynamics in development and other multivariate dynamical systems.