Towards a general framework for spatio-temporal transcriptomics

Julie Pinol, Paul Villoutreix, Thierry Artières

Towards a general framework for spatio-temporal transcriptomics

Julie Pinol
LIS, UMR 7020
Aix-Marseille University, France
julie.pinol@agroparistech.fr

Thierry Artières
Laboratoire d'Informatique et Systèmes UMR 7020
Aix-Marseille University, CNRS, Ecole Centrale de Marseille, France
thierry.artieres@centrale-marseille.fr

Paul Villoutreix
LIS, IBDM, Turing Center for Living Systems
Aix Marseille University, France
paul.villoutreix@univ-amu.fr

Abstract

Position and dynamics of cells are essential pieces of information for the study of embryonic development. Unfortunately, this information is lost in many cell gene expression analysis processes, such as single cell RNA sequencing. Being able to predict the physical positions and the temporal dynamics of cells from gene expression data is therefore a major challenge. After motivating our study with data from *C. elegans* development, we first review current methods based on optimal transport that aim at either predicting the spatial position of cells from transcriptomic data or interpolating differentiation trajectories from time series of transcriptomic data. However, they are not designed to capture simple temporal transformations of spatial data such as a rotation, we propose an extension of the framework proposed by Nitzan et al. [8] including a temporal regularization for the inference of the optimal transport plan. This new framework is tested on artificial data using a combination of the Sinkhorn algorithm and gradient descent. We show that we can successfully learn simple dynamic transformations from very high dimensional data.

1 Introduction

To understand how an embryo forms from a single cell or to predict the development of a tumor, we need to understand the relationship between the dynamics of thousands of genes and the dynamics of large assembly of cells such as tissues. High throughput single cell RNA sequencing techniques provide a way to measure the quantity of individual transcripts within each of the cells of a tissue, however, because the tissue needs to be dissociated during the measurement process, this information is obtained at the cost of losing precise spatial and temporal information [8]. On the other hand, live microscopy techniques give access to precise temporal and spatial information about developing tissues, however they are limited in the number of variables that can be measured at the same time [17]. Several studies have

Preprint. Under review.

generated time series of single cell RNASeq such as in the model system *C. elegans* [9]. In this particular system, the invariance of the cell lineage (i.e. the history of cell division) [15], enables one to map the individual vectors obtained from RNASeq data for each of the cells onto their spatial positions, obtained from live imaging [6] . Such constructed dataset would provide a ground for the study of the link between temporal, spatial and transcriptomic dimensions of a developing embryo at a large scale, therefore calling for a spatio-temporal framework for the study of transcriptomic information.

Previous studies have explored the link between transcriptomic data and spatial positioning of cells [8], as well as the link between transcriptomic and temporal information [12]. Both of these studies[8, 12] rely on optimal transport (OT) [1, 10]. In Nitzan *et al.*[8], the authors rely on the principle that the distances between individual cells are preserved between the transcriptomic space and the physical space; stated otherwise, this means that the cells that are nearby in space have similar gene expression, and vice versa. This principle is formalized within a global optimization framework and framed as an optimal transport problem. In Schiebinger *et al.*[12], the problem solved is in that case orthogonal to Nitzan *et al.*[8]; it consists in inferring the coupling between probability distributions at different time points. The problem is stated as an unbalanced optimal transport problem. The use of optimal transport is promising when considering the relationships between multiple spaces in very high dimensions, as obtained from single cell RNA sequencing technique. However, the joint study of spatial and temporal problems with optimal transport hasn't been addressed in the community yet. In this paper, we propose an extension to the framework proposed by Nitzan *et al.*[8] to capture temporal information.

2 Problem statement

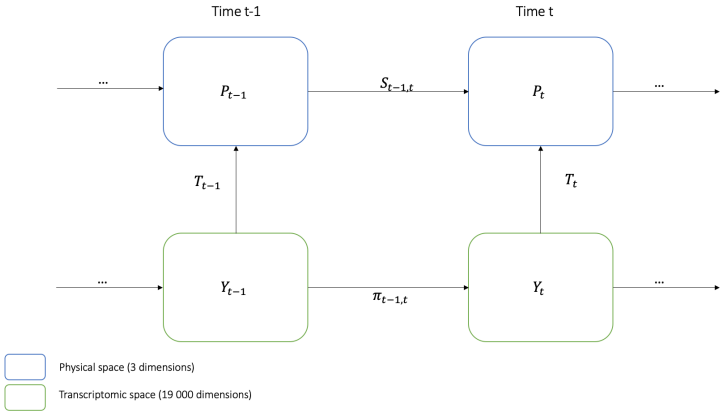

Figure 1: Schematic modeling of the different spaces of our problem, Y_{t-1} and Y_t correspond to transcriptomic matrices at time $t-1$ and t respectively, while P_{t-1} and P_t correspond to positions in the physical space. The matrix T_t is the *transfer* matrix between $Y(t)$ and $P(t)$, and similarly for $t-1$.

Figure 1 illustrates the problem we tackle. We consider the evolution of cells with time, both their spatial positions $P(t)$, their gene expression $Y(t)$ and the links between these variables. We are particularly interested in inferring the sequence of spatial positions of the cells $P(t)$ based on the knowledge of their gene expression in the transcriptomic space $Y(t)$. We give below more details on the considered quantities and variables and we introduce notations. Note that we will use lowercase for vectors (e.g. u), uppercase for matrices (e.g. X) with $u[i]$ denoting the i^{th} element of vector u, $X[i,:]$ the i^{th} row of matrix X, and $X[:,j]$ the j^{th} column of X. Finally we will use superscript $u^{(t)}$ to index quantities on time.

The number of cells varies with time t because of cell division or cell apoptosis (cell death). We note $N(t)$ the number of cells at time t. The cells are characterized by the expression (a real value) for each of the n genes. $Y^{(t)} \in \mathbb{R}^{N(t)\times n}$ stands for the transcriptomic matrix at time t

whose i^{th} row represents expression profile of the i^{th} cell at time t. The links between the set of cells at two successive time steps are usually encoded in a matrix $\pi^{(t-1,t)} \in \mathbb{R}^{N(t-1)\times N(t)}$.

The cells' 3D positions are considered to belong to a set of M predefined locations $X \in \mathbb{R}^{M\times 3}$ which is a regular grid covering the 3D space of interest. The 3D position of a cell is characterized by a distribution over these M predefined position. We note $T^{(t)} \in \mathbb{R}^{N(t)\times M}$ the matrix whose i^{th} row represents the probability distribution of cell i location in the M positions in the grid. Hence the expected 3D positions of the cells at time t, $P(t) \in \mathbb{R}^{N(t)\times 3}$, may be found according to $P(t) = T^{(t)} \times X^{(t)}$. The link between $Y^{(t)}$ and $P(t)$ is then fully determined by $T^{(t)}$.

3 State of the art

Before presenting our method, we report here two recent approaches, which, using optimal transport, aim at inferring some of the variables introduced on figure 1, without, however, tackling the full problem.

3.1 Predicting $T^{(t)}$ from $Y^{(t)}$ at a single time step

Novosparc[8] aims at predicting cell spatial locations from their transcriptional profiles with no or a little number of reporter genes (i.e. genes for which we know the spatial distribution). Since the method is designed for a single time step we drop the time index t everywhere in this section and consider the problem of inferring positions at time t, noted here P (or equivalently T), from transcriptomic vectors at time t, $Y \in \mathbb{R}^{N\times n}$ where n is the number of genes.

The main idea is that, even if the space of hypotheses for cell spatial allocations is very large, a lot of potential allocations do not follow some basic organization which exists in every biological tissues. Consequently, with simple hypotheses, the space of potential solutions can be reduced. The main assumption is that *the distance between two cells in the transcriptomic space and the distance in the spatial space are related, even proportional.* Indeed, the observation on which this approach relies is that cells with similar gene expression profiles tend to be physically close in a developing embryo.

The problem is then the following: one assumes that measurements in each space (transcriptomic and physical) correspond to probabilistic distributions. With this assumption, one can find a transport plan between the two distributions which minimizes the discrepancy between distances in the two spaces.

More formally, one wants to solve:

$$T^* = \underset{T}{\operatorname{argmin}}(1-\alpha)D_1(T) + \alpha D_2(T) - \epsilon H(T) \tag{1}$$

where D_1 is a measure of agreement between pairwise distances in the transcriptomic space and pairwise spatial distances, D_2 encodes prior information about gene expression in space, and H an entropic constraint on T, α and ϵ are positive coefficients that scale the various terms for an optimal trade-off. Actually D_1 reads:

$$D_1(T) = \sum_{i,j,i',j'} L(D^{trans}_{i,i'}, D^{space}_{j,j'}) T_{i,j} T_{i',j'} \tag{2}$$

where L is a mean square loss function and $D^{trans} \in \mathbb{R}^{N\times N}$ are geodesic distances in the transcriptomic space and $D^{space} \in \mathbb{R}^{M\times M}$ are geodesic distances in the grid. This term represents the Gromov-Wasserstein discrepancy as introduced in Peyre *et al.* [11] . It allows to generalize the optimal transport problem in cases where the two spaces cannot be compared directly [11, 10]. When prior information is available it can be added to the model with the second term. If the expression of some genes (usually very few amongst the n genes) is known for the M positions, a term representing the distance (restricted to the marker genes) between the transcriptomic vectors and the known expression on each location may be added. We note G the set of the indices of the marker genes (it is a subset

of $[1...n]$) and $Y[i, G]$ the i^{th} row of of matrix Y restricted to the columns whose index are in G. Then, the prior term D_2 reads:

$$D_2(T) = \sum_{i,j} D_{i,j}^{exp,space} T_{i,j} \tag{3}$$

with $D^{exp,space} \in R^{n \times M}$ a matrix of disagreement. Each term in the sum stands for the distance between the transcriptomic vector (restricted to genes in G) of the i^{th} cell and the predicted transcriptomic vector (restricted to genes in G) at the considered location j ($D^{exp,space}[i, j]$), weighted by the probability of the i^{th} cell being at location j ($T_{i,j}$). Using such a term stabilizes the prediction with known gene positions acting as anchors for the overall prediction.

The final term is an entropic regularization:

$$H(T) = -\sum_{i,j} T_{i,j} log(T_{i,j}). \tag{4}$$

This is the Shannon Boltzmann entropy. Its use in optimal transport was introduced in Wilson *et al.* [19]. As the function is concave, its maximisation forces the solution to tend towards a uniform distribution and makes the whole problem convex.

Once T is inferred with this strategy, one can use it to predict the spatial distribution of the expression of each of the genes by computing $Y^T T$. The accuracy of the prediction on actual biological datasets is described in details in [8].

3.2 Modeling the dynamics in transcriptomic time series $\left(Y^{(t)}\right)_t$

If we now consider the dynamical part of the problem presented on figure 1, a recent work has been successfully using optimal tranport to infer trajectories over time series of transcriptional profiles [12]. Optimal transport is used here to infer the couplings between probability distributions at different time points.

The transcriptional profiles are seen as samples of a time-varying distribution $Y^{(t)}$ on the space of genes. The method looks for a joint distribution between $t-1$ and t, it infers a coupling $\pi_{t-1,t}$ between Y_{t-1} and Y_t, such as :

$$\pi_{t-1,t} Y_{t-1} = Y_t \tag{5}$$

Because of cell proliferation or cell apoptosis, the couplings are found using an unbalanced optimal transport problem.

Another possibility is to consider the whole time sequence as a dynamic optimal transport problem [16]. By adding a time interpolation variable the problem is related to fluids dynamic and facilitate the interpolation task [10, 16]. Using a neural network [16] is computationally efficient in the case of single cells sequences. In brief, the main difference between the two articles lies in the fact that, by considering a succession of static optimal transport rather than a dynamic optimal transport scheme, Schiebinger *et al.* [12] results in a trajectory which is not as smooth as in Tong *et al.* [16].

4 Infering spatio-temporal trajectories from transcriptomic data

We are now interested in using dynamics when inferring the mapping that relates transcriptomic profiles to spatial positions. We assume that we have a time series of transcriptomic profiles and aim at predicting the spatial position of cells. We are showing how we can take advantage of the dynamic undergone by cells to learn small spatial transformations of the cell positions.

4.1 Problem considered

While only subproblems of the full original problem, described in section 2, have been addressed up to now, we are concerned here with a more global approach to solve the full

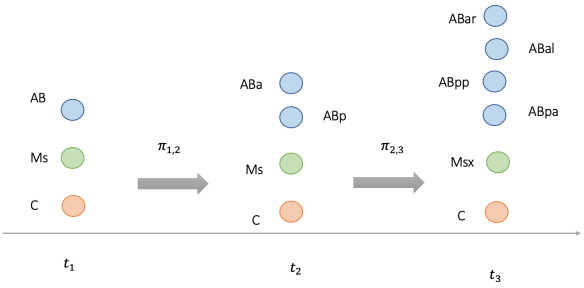

Figure 2: Lineage allows to link cells between time steps. In the classical convention used for describing *C. elegans* development, at every division, a letter is added to the name of a cell, depending on the orientation of the division the letter is either a for anterior, p for posterior, r for right, l for left, d for dorsal and v for ventral. For example the cells ABpa is two divisions away from the cell AB and has ABpp for sister cell.

problem of predicting the sequence of positions $\left(T^{(t)}\right)_t$ of the cells from the sequence of their transcriptomic representations $\left(Y^{(t)}\right)_t$.

Yet, we consider a specific case where the transfer matrices π are known and do not have to be inferred. Indeed in the case of a developing embryo, such as *C. elegans*, where the whole cell lineage is known, i.e. where the history of cell division is known, $\pi^{(t-1,t)}$ can be identified deterministically as presented in figure 2. Daughter cells tend to have gene expression close to their mother cell [9]. This information can be used as a constraint in our model. We show an example of cell lineage representation on figure 2.

4.2 Mathematical framework

When considering the generalization of the problem of inferring the transport map T in a dynamical context as presented on figure 1, one could wonder why not just solve the optimization problem in equation (1) for each time step. As of now, there is no theoretical results that establish in general the stability of the transport plan of optimal transport when varying the probability distributions to be matched [7]. In addition, in its current formulation, when no *a priori* information is given, the solution of equation (1) at a given time step is invariant under rotation or translation, therefore, any simple transformation of the data over time wouldn't be captured by successive application of this inference method.

Building on previous considerations we generalize equation 1 by including constraints that penalize high rate changes in T. The notations are the same as the one defined in equation 1. We are looking for $\mathcal{T} = (T^{(1)}, T^{(2)}, ..., T^{(L)})$ which minimizes the OT-like problem presented in equation (6) where the length of the sequence (the number of time steps) is L and times steps range from 1 to L.

$$\boxed{\mathcal{T}^* = \underset{\mathcal{T}}{\operatorname{argmin}}(1-\alpha-\beta)\sum_{t=1}^{L} D_1(T^{(t)}) + \alpha\sum_{t=1}^{L} D_2(T^{(t)}) + \beta\sum_{t=1}^{L}\Omega(T^{(t)}, \pi^{(t-1,t)}, T^{(t-1)}) - \epsilon\sum_{t=1}^{L} H(T^{(t)})} \tag{6}$$

where α, β and ϵ are positive coefficients that scale the various terms. The definition of D_1 and D_2 have been detailed above. The term $\Omega(T^{(t)}, \pi^{(t-1,t)}, T^{(t-1)})$ is stated in general and denotes constraints on the smoothness of the spatial trajectories of the cells.

In our settings, and in the following sections we will consider the simpler case when there is no death or division of cells so that $\pi^{(t-1,t)} = I$, and $\forall t, N(t) = N$.

4.3 Constraining spatial dynamics

As already stated, we assume closeness between spatial distributions at two close time steps and introduce corresponding constraints that rely on implicit assumptions on the spatial dynamics. In our particular setting constraints terms can be rewritten as:

$$\Omega(T^{(t)}, \pi^{(t-1,t)}, T^{(t-1)}) = \sum_{i=1}^{i=N} l\left(T^{(t)}[i,:], T^{(t-1)}[i,:]\right) \tag{7}$$

where $l\left(T^{(t)}[i,:], T^{(t-1)}[i,:]\right)$ stands for the constraint on the position of the i^{th} cell. Note that we actually put the constraints on the $T(t)$ matrices rather than on the $P(t)$ matrices but this is actually equivalent since $P(t) = T(t) \times X$ where X is the fixed matrix of the grid positions.

The simplest constraint which we call *static constraint* states that the successive positions of a cell should be similar which translates in $\|P(t) - P(t-1)\|^2$ (or $\|T(t) - T(t-1)\|^2$) should be small. A finer constraint states that the spatial trajectory of a cell should be smooth. Noting $\Delta(t) = P(t) - P(t-1)$ the approximate speed vector of $P(t)$, we implemented this idea by constraining $\|\Delta(t) - \Delta(t-1)\|^2$ to be small, we call such a constraint a *dynamic constraint*.

4.4 Alternate optimization scheme

To solve the optimisation problem stated by Eq. (6), we use an alternating optimization scheme whose sketch is illustrated in algorithm 1. The algorithm starts by computing solutions (inferring $T^{(t)}$ from $Y(t)$) at time step 1 and time step 2 independently using the Sinkhorn-Knopp algorithm [13, 14, 8], which we note $T^{(t)} = \text{sinkhorn}\ (D^{trans}, D^{space}, Y^{(t)})$ in the Algorithm, where $D^{trans} \in \mathbb{R}^{N \times N}$ and $D^{space} \in \mathbb{R}^{M \times M}$ are the distance matrices defined in section 3.1.

Then one successively considers the problems for time step t from 3 to L; when solving the problem at time t (inferring $T^{(t)}$ from $Y(t)$), the solutions obtained for previous time steps (e.g. $T^{(t-2)}, T^{(t-1)}$) are used to constrain the solution at time t. More formally, at time step t, we first solve the problem with Sinkhorn without smoothing constraint as in Nitzan *et al.* [8] (equation 1). Then we use the solution obtained as the initialization of the following optimization problem where we introduce an additional smoothing constraint. This constrained objective that we want to minimize for $T^{(t)}$ reads:

$$(1-\beta)D_1(T^{(t)}) + \beta(\Omega(T^{(t)}, \pi^{(t-1,t)}, T^{(t-1)})) \tag{8}$$

This type of alternative resolution on OT problem has already been used on close problems, for example by Xu *et al.* [20] . We present a gradient descent based optimization but note that our attempts to instantiate this optimization scheme with the Sinkhorn-Knopp algorithm[13, 14] instead, while adding smoothing constraints through appropriately designed D_2 term as in Eq. 1 did not allow to reach interesting results and are not presented here. Finally, the smoothing constraint introduced here involves a norm, whose convexity should help the minimization task, as the entropy does.

We used the Sinkhorn algorithm as implemented in the POT library [2] on Python and taken up by [8], creating the Novosparc package. The theory formalized in Nitzan *et al.* [8] is directly generalizable to make prediction in 3D, however the code was only made to predict 2D data, we thus edited it to handle 3D data.

4.5 Experiments

The problem we are tackling is motivated by a real dataset, constructed by combining transcriptomic data from published *C. elegans* single cell RNASeq [9] and the corresponding time series of cell spatial positions obtained from live microscopy [6]. However, the dataset was not robust enough to provide a ground truth to compare various algorithms as there

Algorithme 1 : Alternated optimization scheme

Result : $T = [T^{(1)}, ..., T^{(L)}]$
Input :$\left(Y^{(t)}\right)_t$, X,K;
Initialization ;
D^{trans} = geodesic_distance($Y^{(1)}$);
D^{space} = geodesic_distance(X);
$T^{(1)}$ = sinkhorn (D^{trans},D^{space},$Y^{(1)}$);
D^{trans} = geodesic_distance($Y^{(2)}$);
$T^{(2)}$ = sinkhorn (D^{trans},D^{space},$Y^{(2)}$);
for i *in* $1...K$ **do**
  for t *in* $3...L$ **do**
    D^{trans} = geodesic_distance($Y^{(t)}$);
    $\hat{T}_i^{(t)}$ = sinkhorn (D^{trans},D^{space},$Y^{(t)}$);
    $T_i^{(t)}$ = Minimizer of Loss from Eq. (8) (optimized with gradient descent starting from initial solution $\hat{T}_i^{(t)}$);
  end
end

are many missing data points. In order to make a proof of concept of our method we thus generated an artificial dataset that mimic a real single cell RNAseq dataset with a temporal transformation and a corresponding realistic embedding in physical space. For simplicity, we used an actual time point obtained from Li *et al.* [6] providing a spatial position for each of 333 cells. The random matrix is the same for the various time points, the only change in the time series comes from the rotation applied to the spatial data, which should be projected in the high dimensional space. At every time step we apply a small rotation of 0.3 radians as shown on figure 3 (left panel). Afterwards, each position matrix is projected in a space of 19000 dimensions in order to simulate single cell transcriptomic data. We choose to use random projection [18]. This projection preserves euclidean distances with high probability according to the Jonhson-Lindenstrauss lemma when reducing dimensions [3]. We supposed that the inverse was true and checked if the k-neighbors graphs were approximately the same in the two spaces. This hypothesis was true in 95.6% of the cases, which is consistent with the state of the art. We conclude that the inverse lemma was true enough for our experiments.

Metrics : The first way of studying results is to look at the general appearance of the point clouds. It can be an uneasy task as the different clouds are in 3D, making them difficult to apprehend. Nevertheless, it is still possible to detect some trends by eyes. After this visual inspection step, we defined two metrics at the cellular scale. Firstly, as one of the aim of the model is to smooth trajectories, we chose to measure the series of angles $(\theta_t)_t(i)$ along the trajectory of a cell i as the angle between $\Delta(t-1)[i,:]$ and $\Delta(t)[i,:]$ following notations in section 4.3, where $\Delta(t)[i,:]$ stands for the difference between the spatial position of cell i at t and at $t-1$. A similar method has been used to measure the smoothness of membrane cells in Lavalou *et al.* [5]. Secondly, we aimed at capturing if the inferred trajectory would be chaotic or not. For each cell, we computed the difference between two successive angles. If the difference is low, it means that the trajectory tends to repeat the same patterns between time steps, on the contrary, if the difference is high, the movement does not follow any pattern. For these two measures, we compute the average over all the time points and all the cells. We obtain:

$$\text{Average angle} = \frac{1}{N}\frac{1}{L-1}\sum_{i=1}^{N}\sum_{k=2}^{L}\theta_t(i)$$

Figure 3: Reconstruction of the spatial locations with a dynamic constraint by gradient descent: ground truth trajectories of all cells (left) and reconstructed trajectories for 8 cells with our method with $\beta = 0.99$ (right).

Table 1: Cell trajectories metrics - $\beta = 0.99$

Constraint	Averaged angle (SD)	$D_{\text{successive\_angles}}$(SD)
Ground truth	0.3 (1.85e-16)	3.91e-16 (1.47e-16)
Novosparc	1.65 (0.18)	0.91 (0.26)
Dynamic	0.52 (0.12)	0.36 (0.10)

Table 2: Smoothness with different value of β

β	Averaged angle	$D_{\text{successive\_angles}}$
0.5	0.98	1.03
0.8	0.61	0.58
0.93	0.46	0.35
0.95	0.51	0.40
0.99	0.52	0.36

$$D_{\text{successive\_angles}} = \frac{1}{N}\frac{1}{L-2}\sum_{i=1}^{N}\sum_{k=2}^{L}|\theta_t(i) - \theta_{t-1}(i)|$$

Results : We applied our method on 333 vectors for 10 time steps using the *dynamic constraint*. The results are presented on figure 3 (right panel) which shows a subset of 8 cells over time for $\beta = 0.99$, a rotation seems indeed to have been learned. The tables 1 and 2 show the results obtained and their comparison to Novosparc. We successfully decreased the averaged angle between two time steps and made it close to the 0.3 radians of the actual rotation. We show a significant improvement compared to Novosparc. Moreover, the difference between successive angles suggests that the algorithm did learn a coherent movement and not just random deviations.

5 Discussion

We showed in this article a new method to infer spatial positions from time series of cell expression profiles. Our framework builds on gene expression cartography [8], a previous method that infer a transport map between transcriptomic data and cell spatial positions a given time step. We add a way to capture a major aspect of biological systems, which is their dependency to time. We proposed to take advantage of the assumption that systems vary continuously through time in order to establish a framework which, in addition to being able to predict cell spatial positions at one step, is able to learn simple dynamical transformation within the data. We showed its efficiency on a simple rotation of the data. This framework is different from previous attempts that aim at inferring temporal relationships within transcriptomic data only [4, 12], as we are concerned here with the mapping that connect

transcriptomic data with spatial data. However, even if we stated the problem in general terms in equation 6, we provided a solution in a particular case, which is when the number of cells is constant. If the system studied was more complicated, with a changing number of cells, or a non-deterministic, or even unknown, cell lineage, we would have to add an inference step for π which can be estimated with optimal transport, e.g. in [12]. We propose here an initial step towards a general framework for the study of spatio-temporal transcriptomics which is a central problem of biology. The study of T as a function of time will likely be very useful to understand extremely complex processes such as embryonic development.

References

[1] V. C. *Optimal Transport, old and new*. Springer, 2008.

[2] R. Flamary and N. Courty. Pot python optimal transport library, 2017.

[3] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

[4] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

[5] J. Lavalou, Q. Mao, S. Harmansa, S. Kerridge, A. C. Lellouch, J.-M. Philippe, S. Audebert, L. Camoin, and T. Lecuit. Formation of mechanical interfaces by self-organized toll-8/cirl gpcr asymmetry. *bioRxiv*, 2020.

[6] X. Li, Z. Zhao, W. Xu, R. Fan, L. Xiao, X. Ma, and Z. Du. Systems properties and spatiotemporal regulation of cell position variability during embryogenesis. *Cell reports*, 26(2):313–321, 2019.

[7] Q. Mérigot, A. Delalande, and F. Chazal. Quantitative stability of optimal transport maps and linearization of the 2-wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3186–3196. PMLR, 2020.

[8] M. Nitzan, N. Karaiskos, N. Friedman, and N. Rajewsky. Gene expression cartography. *Nature*, 576(7785):132–137, Dec. 2019.

[9] J. S. Packer, Q. Zhu, C. Huynh, P. Sivaramakrishnan, E. Preston, H. Dueck, D. Stefanik, K. Tan, C. Trapnell, J. Kim, et al. A lineage-resolved molecular atlas of c. elegans embryogenesis at single-cell resolution. *Science*, 365(6459):eaax1971, 2019.

[10] G. Peyré, M. Cuturi, et al. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[11] G. Peyré, M. Cuturi, and J. Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672, 2016.

[12] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

[13] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.

[14] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[15] J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, et al. The embryonic cell lineage of the nematode caenorhabditis elegans. *Developmental biology*, 100(1):64–119, 1983.

[16] A. Tong, J. Huang, G. Wolf, D. van Dijk, and S. Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics, 2020.

[17] P. Villoutreix, J. Andén, B. Lim, H. Lu, I. G. Kevrekidis, A. Singer, and S. Y. Shvartsman. Synthesizing developmental trajectories. *PLoS computational biology*, 13(9):e1005742, 2017.

[18] R. Ward. Dimension reduction via random projections, 2014. URL: `https://web.ma.utexas.edu/users/rward/madison14.pdf`. Last visited on 2020-09-10.

[19] A. G. Wilson. The use of entropy maximising models, in the theory of trip distribution, mode split and route split. *Journal of transport economics and policy*, pages 108–126, 1969.

[20] H. Xu, D. Luo, H. Zha, and L. Carin. Gromov-wasserstein learning for graph matching and node embedding. *arXiv preprint arXiv:1901.06003*, 2019.