Cliques of single-cell RNA-seq profiles reveal insights into cell ecology during development and differentiation

Baihan Lin

Department of Systems Biology, Columbia University, New York, USA
Department of Applied Mathematics, University of Washington, Seattle, USA
Single-cell data are noisy

(Faure et al., *Cell Systems* 2017)
scTDA

- Nonlinear
- Model-independent
- Unsupervised

- resolve **asynchrony** and **continuity** in cellular identity

(Rizvi et al., *Nature Biotechnology* 2017)
Vertebrate Embryogenesis scRNA-Seq

- 38,731 zebrafish embryo cells
- 25 cell types
- 12 time steps

(Farrell et al., *Science* 2018; Wagner et al., *Science* 2018; Briggs et al., *Science* 2018)
TDA and t-SNE colored by time

- Selected 103 genes
- Mapped to 2D t-SNE
**TDA and t-SNE colored by time**

**Observation:** the 12 time points are not perfectly mapped here.
**Observation**: Unlike the scTDA paper, the temporal structure here is not a skeleton / flare.

(Rizvi et al., 2017)
Are there more **distinguishable** time-dependent features we can extract, other than the relative distance in TDA?

(Rizvi et al., 2017)
Can we characterize time points better?

- Temporal progression → Beyond TDA skeleton
- Geometric invariants → Beyond Betti Number 1
Simplicial Architectures

- **Cech complex** (Nerve)
  - Nonempty spherical intersection
  - Used in most TDA mapping
  - Benefit from Nerve theorem

- **Vietoris-Rips complex**
  - Distance between any pair < $\epsilon$
  - Easier to compute
Simplicial Architectures of scRNA-Seq

- Back to the scTDA paper
- 1,529 cells
- 5 time points

The correlation between sampling time point and cell complexity are not obvious.

Cell complexity was defined as the number of genes whose expression is detected in a cell.

(Rizvi et al., 2017)
Simplicial Architectures of scRNA-Seq

• Back to the scTDA paper
• 1,529 cells
• 5 time points

• Preprocessing:
  • 1,415 cells selected
  • 197 genes selected
  • PCA[0] and PCA[1]

Here we are interested in the **intercellular interaction** within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Architectures of scRNA-Seq

- Back to the scTDA paper
- 1,529 cells
- 5 time points

- Preprocessing:
  - 1,415 cells selected
  - 197 genes selected
  - PCA[0] and PCA[1]

Here we are interested in the **intercellular interaction** within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Architectures of scRNA-Seq

- Back to the scTDA paper
- 1,529 cells
- 5 time points

- Preprocessing:
  - 1,415 cells selected
  - 197 genes selected
  - PCA[0] and PCA[1]

Here we are interested in the intercellular interaction within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Architectures of scRNA-Seq

- Back to the scTDA paper
- 1,529 cells
- 5 time points

- Preprocessing:
  - 1,415 cells selected
  - 197 genes selected
  - PCA[0] and PCA[1]

Here we are interested in the **intercellular interaction** within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Architectures of scRNA-Seq

• Back to the scTDA paper
• 1,529 cells
• 5 time points

• Preprocessing:
  • 1,415 cells selected
  • 197 genes selected
  • PCA[0] and PCA[1]

Here we are interested in the intercellular interaction within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Architectures of scRNA-Seq

• Back to the scTDA paper
• 1,529 cells
• 5 time points

• Preprocessing:
  • 1,415 cells selected
  • 197 genes selected
  • PCA[0] and PCA[1]

Here we are interested in the intercellular interaction within the same type of cells, in this case, within each time point, rather than their relationships, as in scTDA.
Simplicial Filtration

- Control Models:
  - Renyi-Edros graph
  - Shuffled pairwise distances
Simplicial Architectures of scRNA-Seq

As shown, **Betti numbers** only offer limited information on cellular complexity.
As shown, simplicial complexes are much more informative than Betti numbers.
As shown, simplicial complexes are much more informative than Betti numbers.
It shows that **gastulation stage** is a very critical stage in vertebrate development.
Simplicial Architectures of scRNA-Seq

- **Process:**
  - the embryo begins differentiation to establish distinct cell lineages

- **Before Gastrulation:**
  - the embryo is a continuous epithelial sheet of cells

- **After Gastrulation:**
  - **Organogensis**: individual organs develop within the newly formed germ layers

It shows that **gastulation stage** is a very critical stage in vertebrate development.
Simplicial Architectures of scRNA-Seq

- Interesting questions in biology:
  - Can we determine developmental stages without physiological features?
  - Can we generate pseudo-time series based on scRNA-Seq?

It shows that gastrulation stage is a very critical stage in vertebrate development.
Normalized simplicial complexity
Simplicial Dynamics Mapping
As shown, simplicial complexes can also facilitate **lineage tracing and analysis**.
Applications for simplicial architectures

• Low-dimensional representation
  • Multi-scale dynamics
  • Cell lineage analysis
  • Developmental analysis
  • Critical stage identification

• Directed vs. non-directed graphs
  • Flexible setup
  • Temporal dependency
Back to **TDA** colored by time…

Is there a better way to capture the **information flow** across time?
Back to scRNA-seq TDA colored by time...

The temporal filtration has a better capture of the information flow
Back to scRNA-seq TDA colored by time…

The temporal filtration has a better capture of the information flow.
Back to scRNA-seq TDA colored by time…

The temporal filtration has a better capture of the information flow

Track 1

Track 2

gastrulation

Track 2
Better characterize single-cell states?

Cell states arise transiently during time-dependent processes.

Idea:

Can we compare the traditional TDA with time-filtrated TDA to get this type of information?

(Wagner et al., Nature Biotechnology 2016)
References


Acknowledgements

• Raul Rabada (Columbia)
• Ioan Filip (Columbia)
• Nikolaus Kriegeskorte (Columbia)
• Guillermo Cecchi (IBM Research)
Thanks! Questions?