INSTITUTE FOR MATHEMATICS
STRUCTURES CLUSTER OF EXCELLENCE
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Michael Bleher
Institute for Mathematics, Heidelberg University

– MATRIX – 14 May 2026 –

Single cells, population dynamics, and Euler characteristic profiles

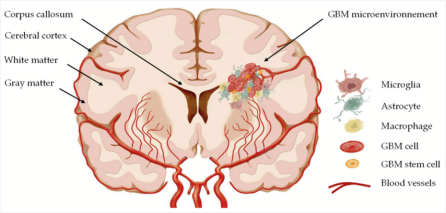
Motivation

Cell population dynamics of Glioblastoma (GBM) (Brain cancer)

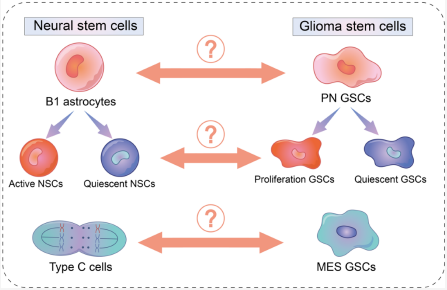
GBM = neural stem cell dynamics 'gone wrong'.

Roughly: dysregulated differentiation process.

w/ Anna Marciniak-Czochra (Heidelberg),
Ana Martin-Vilalba (DKFZ),
Marta Marszweska (Gdansk, Warsaw),
Justyna Signerska-Rynkowska (Gdansk),
Pawel Dlotko (Warsaw)

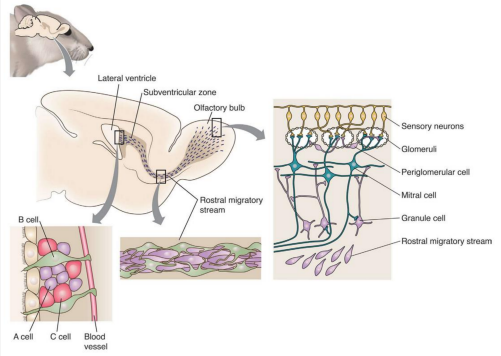


el Kheir et al., 2022



Wang et al., 2021

Model System: Neurogenesis in mature mice



Sanes et al., 2019

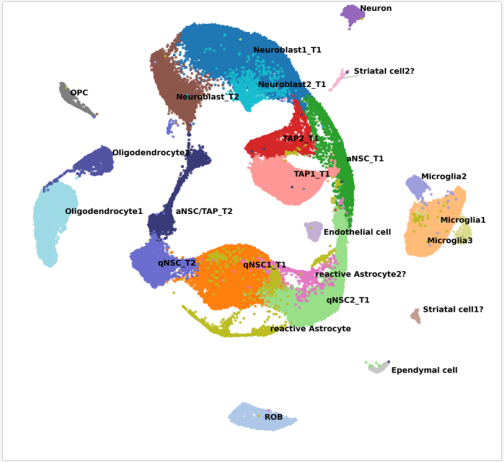
(Main) Cell Types

- Quiescent Neural Stem Cells (Q)
- Active Neural Stem Cells (A)
- Differentiated Cells, e.g. Neurons (D)

Questions

- How do cells transition between
$$Q \leftrightarrow A \leftrightarrow D$$
- How do transitions depend on population size (signalling), time (aging), external factors (e.g., inflammation)?
- How do these dynamics change in disease?

Model System: Neurogenesis in mature mice



Martin-Vilalba et al. (unpublished)

(Main) Cell Types

- Quiescent Neural Stem Cells (Q)
- Active Neural Stem Cells (A)
- Differentiated Cells, e.g. Neurons (D)

Questions

- How do cells transition between
$$Q \leftrightarrow A \leftrightarrow D$$
- How do transitions depend on population size (signalling), time (aging), external factors (e.g., inflammation)?
- How do these dynamics change in disease?

Population Dynamics

Compartmental model:

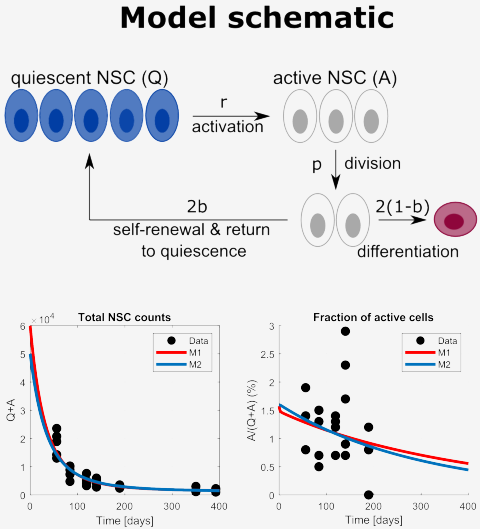
- Graph (compartments + transitions)
- Transition rates between compartments

$$\frac{dQ}{dt} = -rQ + 2bpA$$
$$\frac{dA}{dt} = rQ - pA$$
$$\frac{dD}{dt} = 2(1-p)bA$$

Problem

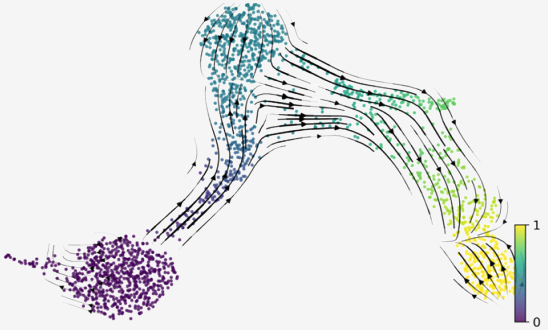
Population dynamics not identifiable from data.

Many models fit the same population data.
(different graphs, rates, non-linearities, ...)

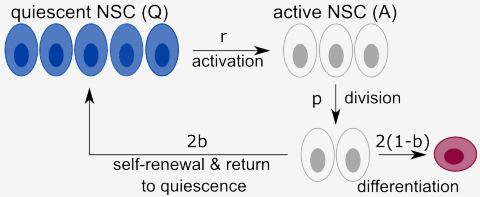


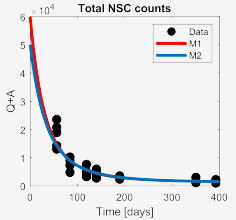
Population Dynamics

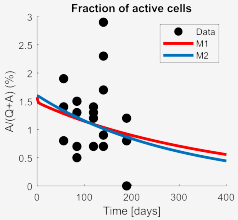
Q: Can single-cell data help?



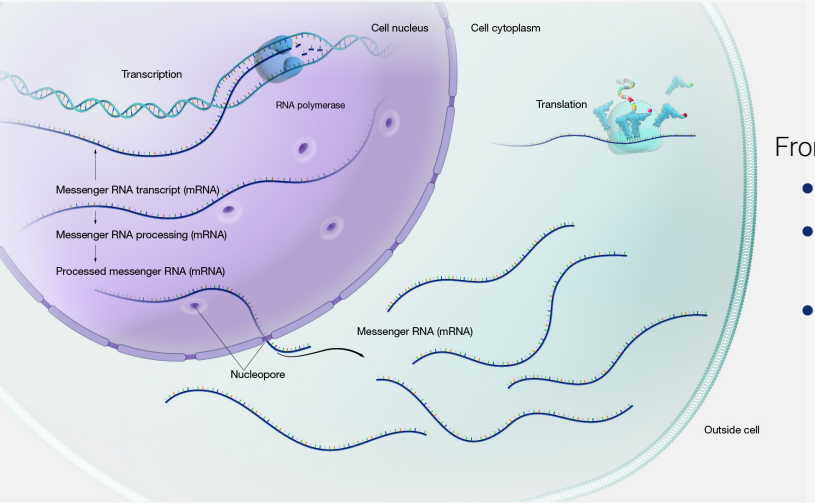
Model schematic







Single-cell Gene Expression



From code to function

- DNA $\rightarrow$ mRNA $\rightarrow$ proteins
- gene expression
  $\simeq$ # mRNA snippets
- proxy for cell's current biological state $\mathbf{x} \in \mathbb{R}^N$

Single-cell dynamics: a minimal picture

1. Cell state dynamics

Each cell can be in one of several cell states

$$z \in \{Q, A, D, \ldots\}$$

determines transcription rate $\alpha_{z,g}$ for each gene g

State probability follows a continuous-time Markov chain:

$$\frac{d}{dt} p_z(t) = \sum_{z'} H_{zz'} p_{z'}(t)$$

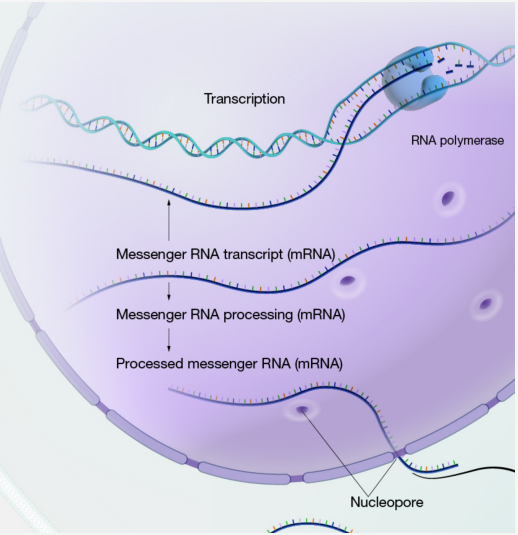
2. Gene transcription dynamics

Gene expression follows a transcription-degradation process:

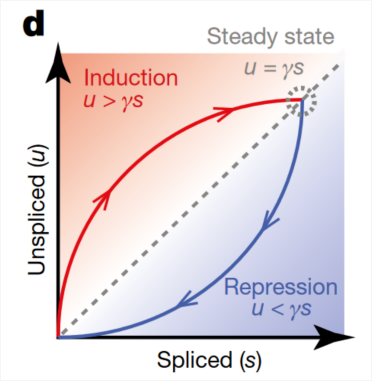
$$\frac{dx_g}{dt} = \alpha_{z(t),g} - \gamma x_g$$

- x_g: expression of gene g
- α_{g,z_t}: transcription rate
- γ: degradation rate

RNA velocity



La Manno, G. et al. (2018) 'RNA velocity of single cells', Nature, 560(7719), pp. 494–498. Fig 1.



$\rightsquigarrow \mathbf{v} \in \mathbb{R}^N$ **RNA velocity**

Single-cell dynamics: a slightly extended picture

1. Cell state dynamics

Each cell can be in one of several cell states

$$z \in \{Q, A, D, \ldots\}$$

determines transcription rate $\alpha_{z,g}$ for each gene g

State probability follows a continuous-time Markov chain:

$$\frac{d}{dt}p_z(t) = \sum_{z'} H_{zz'}p_{z'}(t)$$

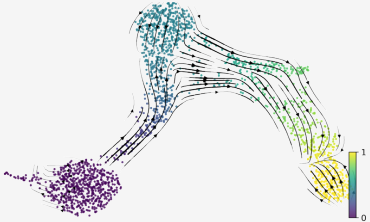
2. Gene transcription dynamics

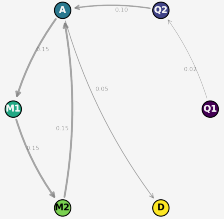
Gene expression follows a transcription-splicing-degradation process:

$$\frac{du_g}{dt} = \alpha_{z(t),g} - \beta u_g$$
$$\frac{ds_g}{dt} = \beta u_g - \gamma s_g$$

- u_g/s_g: unspliced/spliced expression
- α_{g,z_t}: transcription rate
- β: splicing rate
- γ: degradation rate

From single cells to populations





Single-cell level

Continuous-time Markov Chain (CTMC)

(interacting particles / mean field)

$$\frac{d}{dt}\begin{pmatrix} p_Q \\ p_A \\ p_D \end{pmatrix} = \begin{pmatrix} -\lambda_{QA} & \lambda_{AQ} & 0 \\ \lambda_{QA} & -(\lambda_{AQ}+\lambda_{AD}) & \lambda_{DA} \\ 0 & \lambda_{AD} & -\lambda_{DA} \end{pmatrix}\begin{pmatrix} p_Q \\ p_A \\ p_D \end{pmatrix}$$

$\xrightarrow{LLN}$

Population level

Occupation number ODEs

(non-linear in population sizes)

$$\frac{d}{dt}\begin{pmatrix} Q \\ A \\ D \end{pmatrix} = \begin{pmatrix} -f_{QA} & f_{AQ} & 0 \\ f_{QA} & -(f_{AQ}+f_{AD}) & f_{DA} \\ 0 & f_{AD} & -f_{DA} \end{pmatrix}\begin{pmatrix} Q \\ A \\ D \end{pmatrix}$$

A Hierarchy of Problems

Linking scRNA-seq data to population dynamics requires answering

1. What **compartments/states** can individual cells be in?
2. What **transitions** occur between these states?
   For example: can cells move back into (deep) quiescence or do they remain active?
3. What are the **rates** of these transitions?
4. How do rates **depend on** population size, time, external factors?

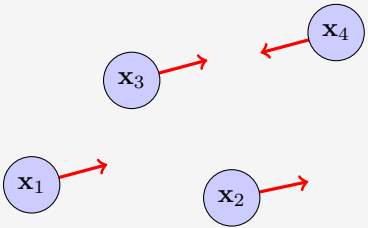
Problems 1 & 2: Topology

Can we differentiate between candidate graphs using scRNA-seq data?

Approach: Euler Characteristic Profiles.

based on work by Marta Marszweska, Justyna Signerska-Rynkowska, Pawel Dlotko

From labeled points to filtered graphs



What We Have (in principle)

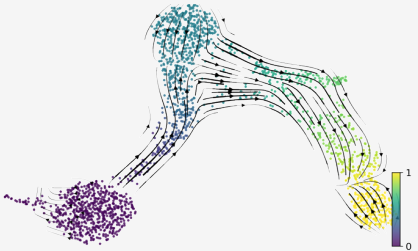
For each cell

- current gene expression $\mathbf{x} \in \mathbb{R}^N$
- RNA velocity $\mathbf{v} \in T_{\mathbf{x}}\mathbb{R}^N$

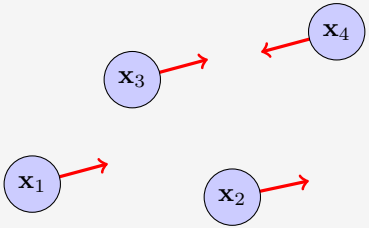
$$X = \{(\mathbf{x}, \mathbf{v})\} \subset T\mathbb{R}^N$$

What We Want

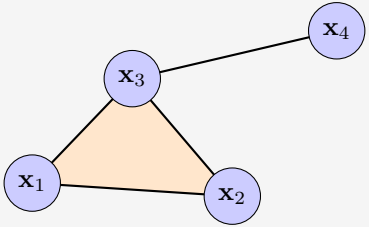
- Graphs of underlying Markov processes
- Or at least to tell them apart



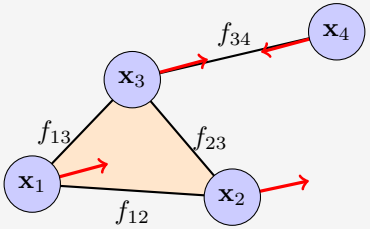
From labeled points to filtered graphs to filtered complexes



From labeled points to filtered graphs to filtered complexes



From labeled points to filtered graphs to filtered complexes



$$f(e_{ij}) = \big(d_{\cos}(\mathbf{v}_i, \mathbf{x}_j - \mathbf{x}_i), d_{\cos}(\mathbf{v}_j, \mathbf{x}_j - \mathbf{x}_i)\big)$$

$$f(\triangle) = \max(f_{12}, f_{13}, f_{23})$$

The Euler Characteristic Profile (ECP)

$$K_\bullet(\Gamma) : (P, \leq) \to (\mathrm{SimpComp}, \subseteq)$$

Applying homology $\implies$ multi-parameter persistence module.

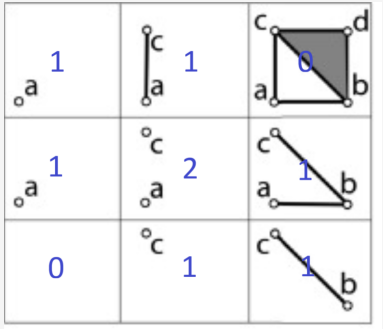
Hard to interpret, hard to compute.

Euler Characteristic Profile

$$ECP: P \to \mathbb{Z}$$

$$(a,b) \mapsto \chi\big(K_{(a,b)}\big) = \sum_{i=0}^{n} (-1)^i \#\{\sigma \in K_{(a,b)} \mid \dim \sigma = i\}$$

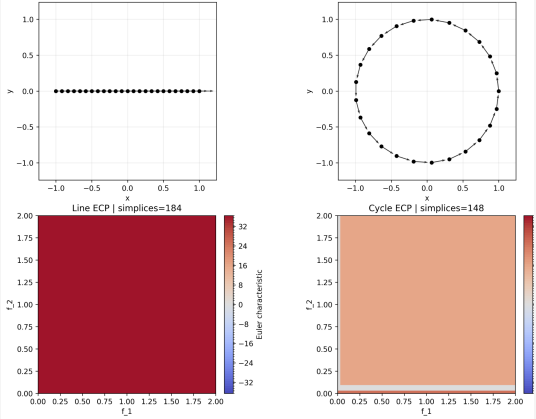
$$= \sum_{i=0}^{n} (-1)^i \operatorname{rk} H_i\big(K_{(a,b)}\big)$$



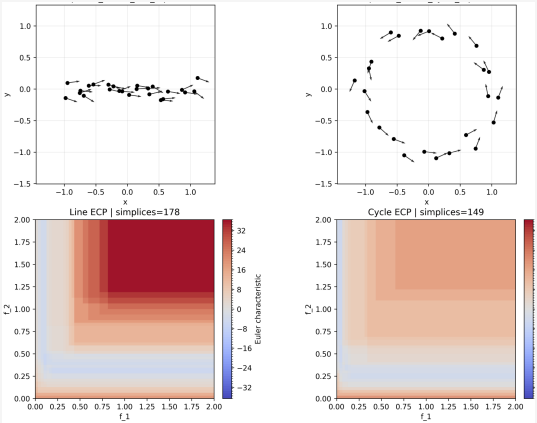
Hard to interpret, embarrassingly easy to compute.

Stability Experiments

no noise

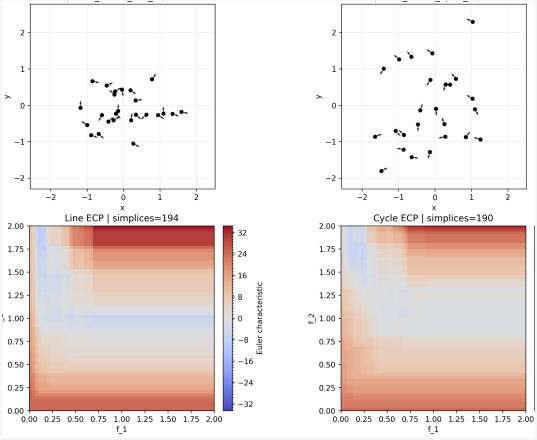


noise ∼ 10% of signal

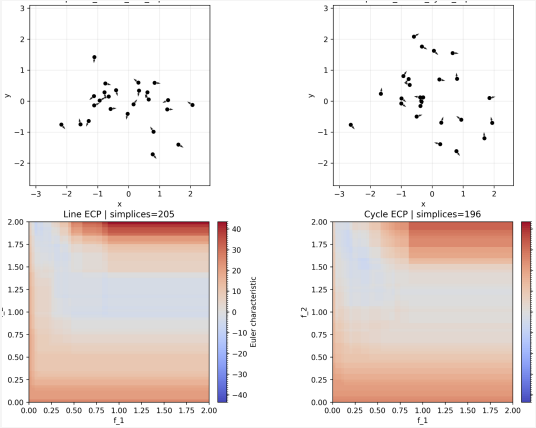


Stability Experiments

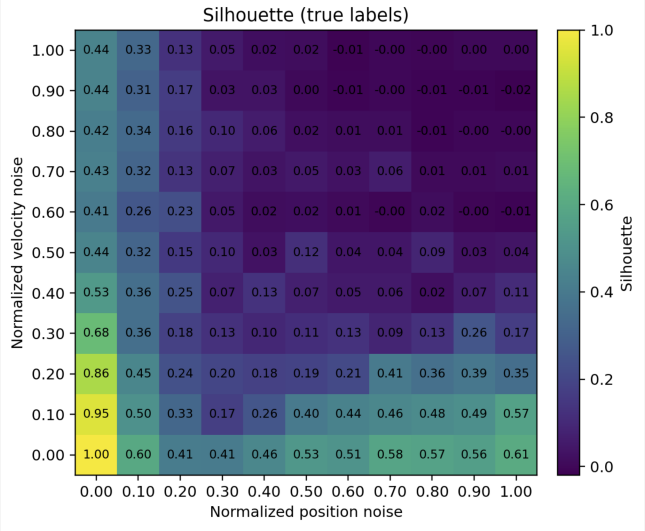
noise ∼ 50% of signal



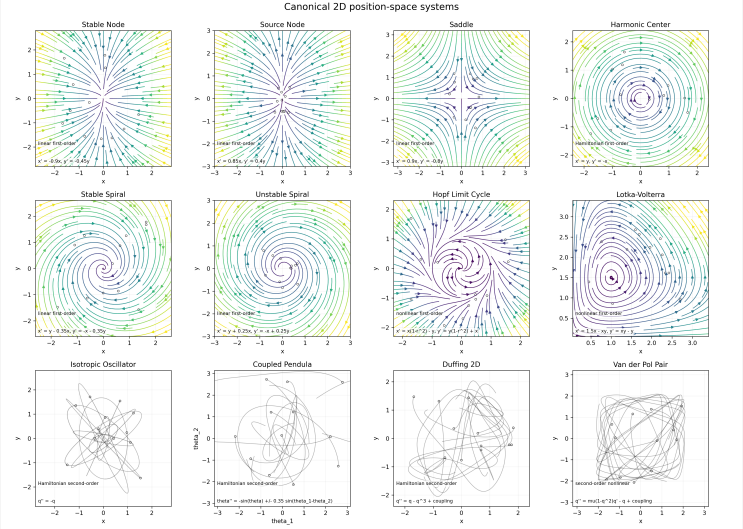
noise ∼ 100% of signal



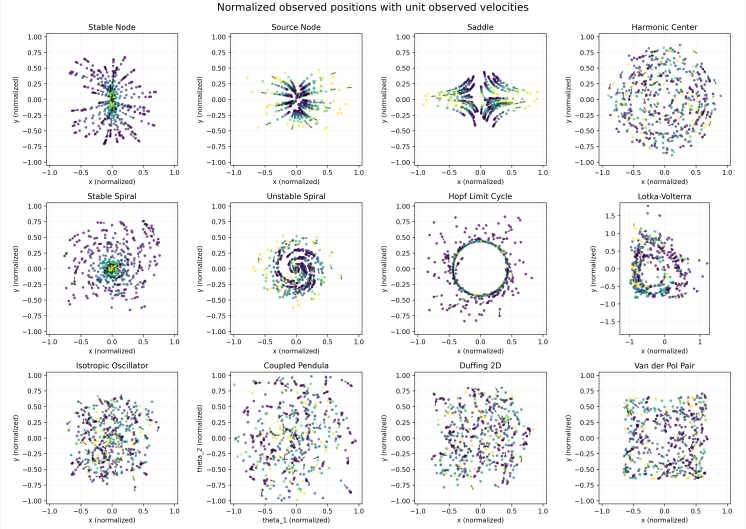
Stability Experiments



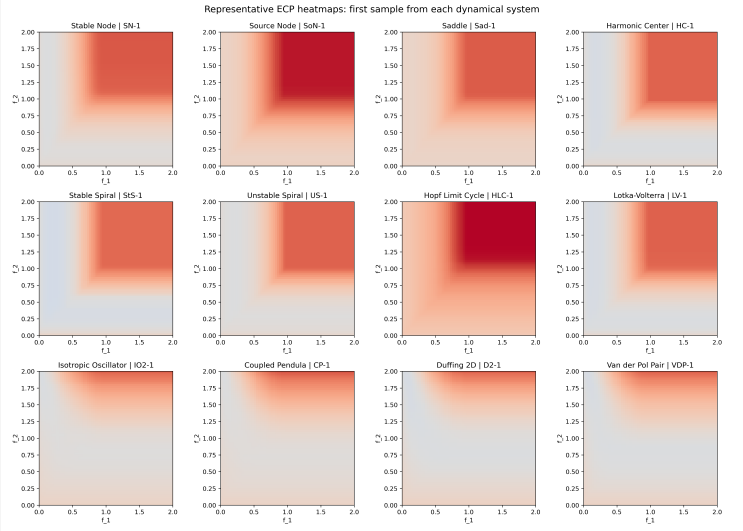
Low-dimensional dynamical systems



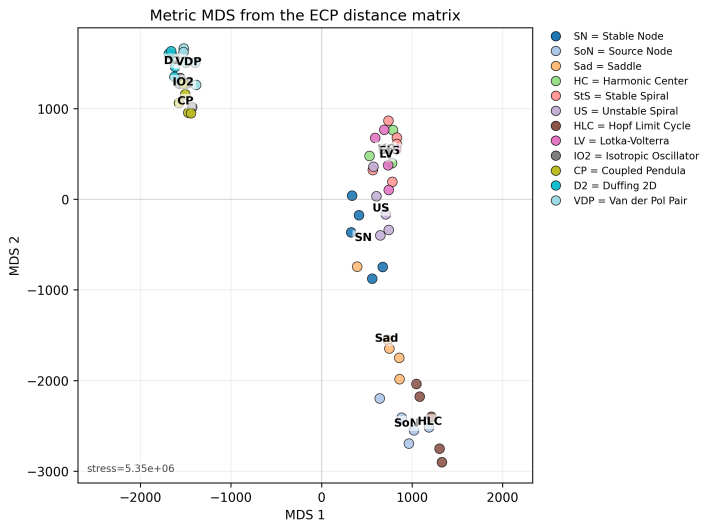
Low-dimensional dynamical systems



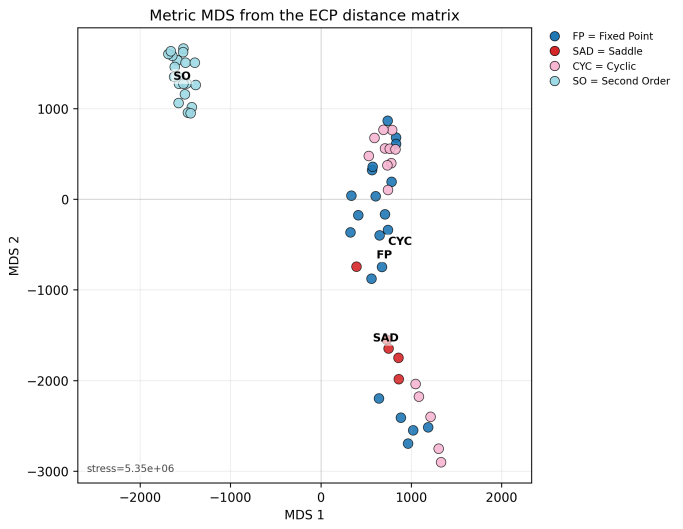
Low-dimensional dynamical systems



Low-dimensional dynamical systems



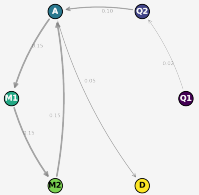
Low-dimensional dynamical systems



Synthetic Data: A Markov-Modulated Splicing Model

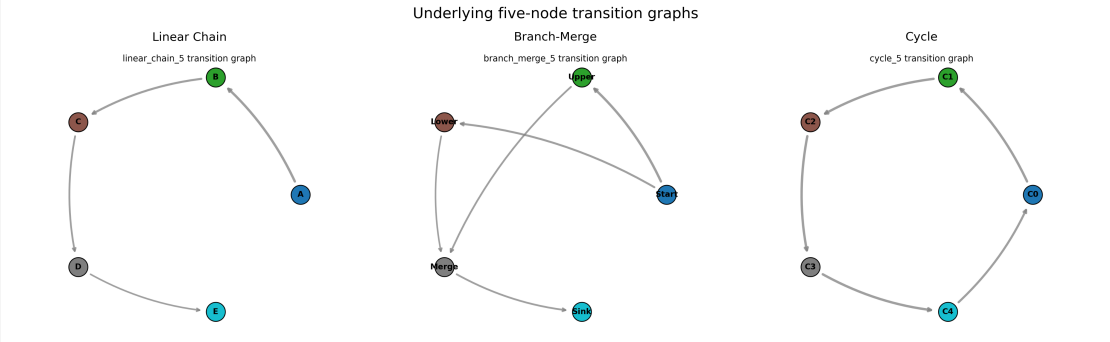
Generate synthetic u/s-counts of single cells using two-level dynamics:

- **Latent state process:**
  Continuous-time Markov chain on a state graph
  - States $z_1, \ldots, z_n$ with probability p_i
  - $\frac{d}{dt} p_i = \sum_j H_{ij} p_j$
  - Asymmetric transition rates $H_{ij} \neq H_{ji}$, encode directionality
- **Gene expression:**
  Standard transcription-splicing-degradation process for cell in state z with state-dependent transcription rate α_z
  - $\frac{du}{dt} = \alpha_z - \beta \cdot u$ (transcription + splicing)
  - $\frac{ds}{dt} = \beta \cdot u - \gamma \cdot s$ (splicing + degradation)

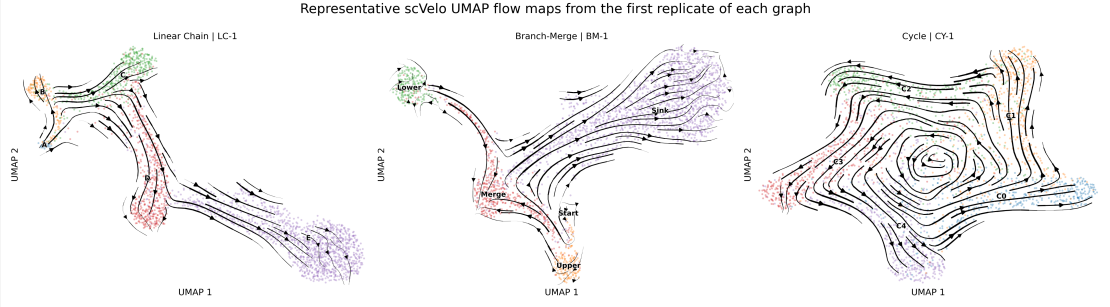


Output: u/s-counts + **current state** for each cell at time t.

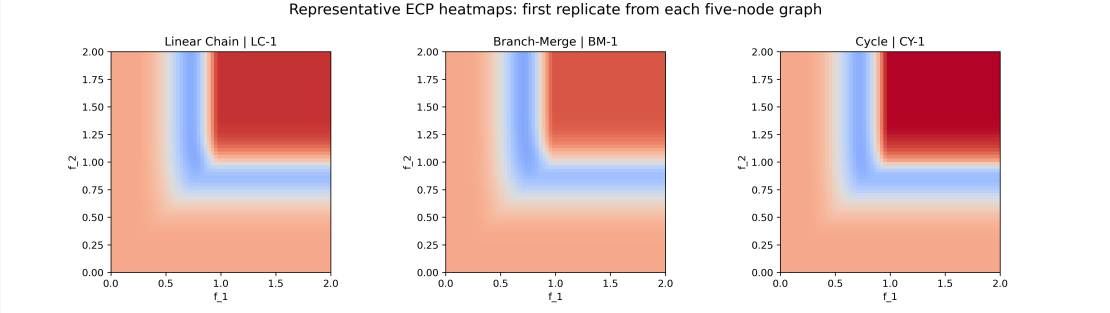
Synthetic scRNA-seq data



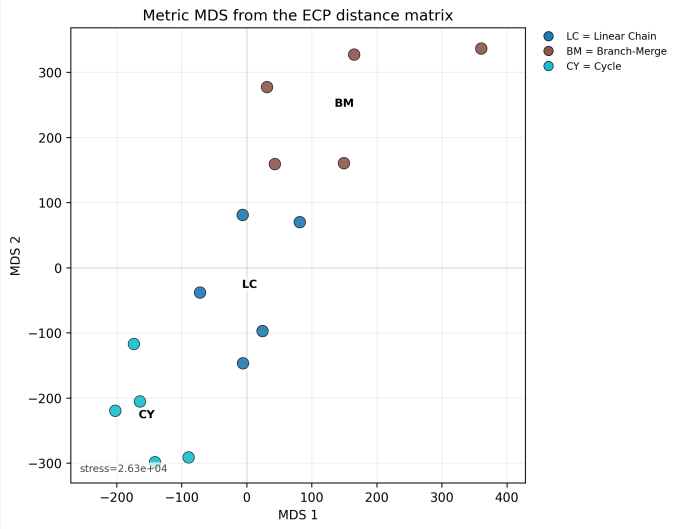
Synthetic scRNA-seq data



Synthetic scRNA-seq data



Synthetic scRNA-seq data



Current Status and Next Steps

What We Have

- Bifiltered complex construction (kNN + velocity alignment)
- ECP computation for bifiltrations
- synthetic ”single cell expression” data for arbitrary graphs

Next Steps

1. Proper benchmarking
2. Systematic noise sensitivity analysis
3. high-dimensional data is famously noisy; too noisy for ECP to be useful?
4. RNA velocity is famously unreliable; too unreliable for ECP to be useful?
5. Move beyond classification/clustering; can we recover the ground truth?

Thanks for your attention

From labeled points to filtered graphs

Construct symmetric kNN graph $\Gamma_{kNN}(\{\mathbf{x}\})$.

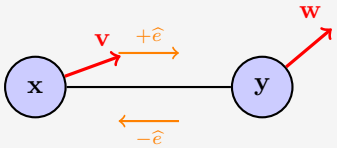
For each undirected edge $\{\mathbf{x}, \mathbf{y}\}$, represent it by unit direction $\widehat{e}_{ij} = (\mathbf{y} - \mathbf{x})/\|\mathbf{y} - \mathbf{x}\|$ normalized velocities $\widehat{\mathbf{v}}, \widehat{\mathbf{w}}$.

Orientation-invariant edge bifiltration

$$q_i^{\pm} = \frac{1 \mp \langle s\widehat{e}_{ij}, \widehat{\mathbf{v}} \rangle}{2}, \qquad q_j^{\pm} = \frac{1 \mp \langle s\widehat{e}_{ij}, \widehat{\mathbf{w}} \rangle}{2},$$

$$p_{ij}^s = \mathrm{sort}_{\uparrow}(q_i^s, q_j^s),$$

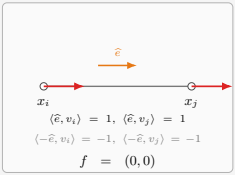
$$f(\{\mathbf{x}, \mathbf{y}\}) = \min\nolimits_{\mathrm{lex}}\{p_{ij}^+, p_{ij}^-\} \in [0,1]^2.$$

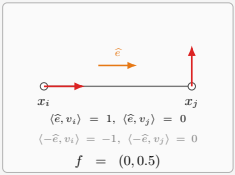


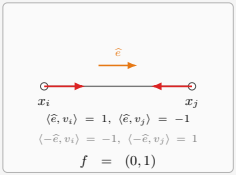
Sort for endpoint order; take both signs for segment orientation.

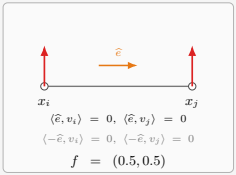
Small means coflow along a local segment.

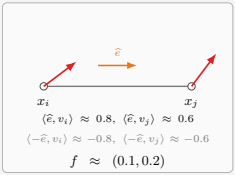
From labeled points to filtered graphs

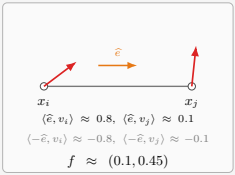


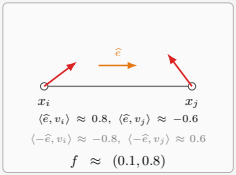


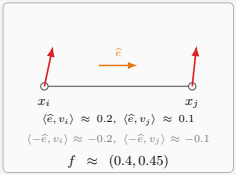


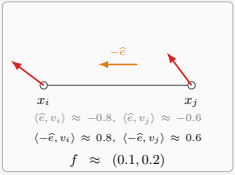


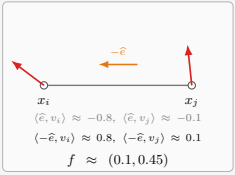


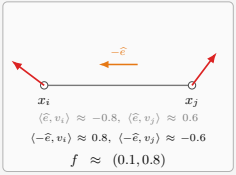


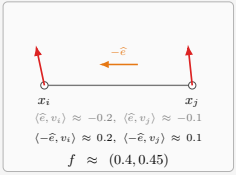










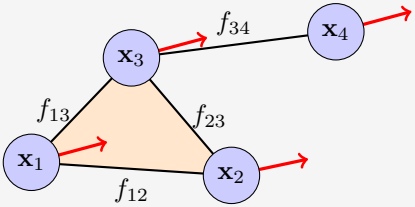


From filtered graphs to filtered complexes

For a graph $\Gamma = (V, E)$, its **clique** / **flag complex** is

$$K(\Gamma) = \big\{\sigma \subseteq V \mid \{u, v\} \in E \text{ for all } u, v \in \sigma\big\}.$$

$$(p+1)\text{-clique in } \Gamma \iff p\text{-simplex in } K(\Gamma)$$



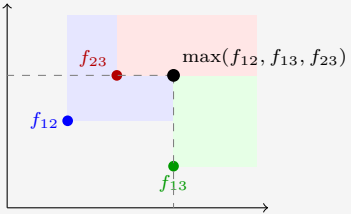
$$f(\triangle) = \max(f_{12}, f_{13}, f_{23})$$

Filtration

- Vertices: $f(\mathbf{x}) = (0, 0)$
- Edges: $f(\{\mathbf{x}, \mathbf{y}\}) = \min_{\text{lex}}\{p_{ij}^{+}, p_{ij}^{-}\}$
- Higher simplices when all faces are present:

$$f(\sigma) = \max_{\{x,x'\}\subseteq\sigma} f(\{x, x'\})$$

$$K_{\boldsymbol{\epsilon}}(\Gamma) \coloneqq \{\sigma \in K(\Gamma) \mid f(\sigma) \leq \boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)\}$$



coordinatewise maximum in $(\mathbb{R}^2, \leq)$