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

Michael Bleher
Institute for Mathematics, Heidelberg University

– PEPS Seminar – January 19, 2026 –

Population dynamics, scRNA-seq, and Euler Characteristic Profiles

Motivation

Cell population dynamics is crucial in:

- Stem cell biology
- Cancer treatment
- Regenerative medicine

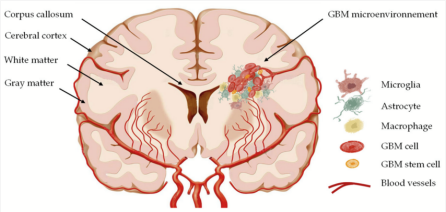
Example: Glioblastoma

Glioblastoma tumours (GBM) arise from neural stem cell dynamics 'gone wrong'.

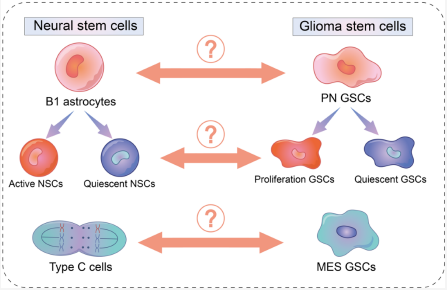
Roughly: dysregulated differentiation process.

High 'stemness' of tumour cells
$\rightarrow$ therapy resistance, relapse, poor prognosis.

w/ Ana Martin-Vilalba (DKFZ),
Anna Marciniak-Czochra (Heidelberg)

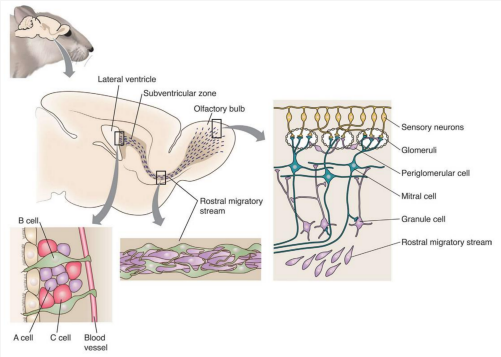


el Kheir et al., 2022



Wang et al., 2021

Model System: Neurogenesis in maturating 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?

The Population-Level Problem

Problem

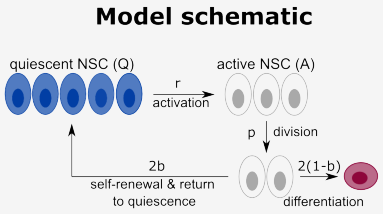
Population dynamics not identifiable from data.

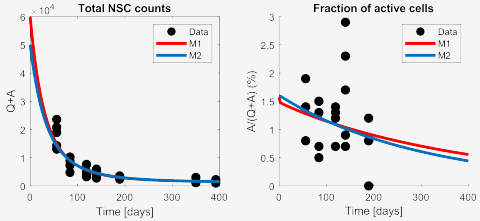
Why?

Compartmental models are determined by:

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

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





The Population-Level Problem

Problem

Population dynamics not identifiable from data.

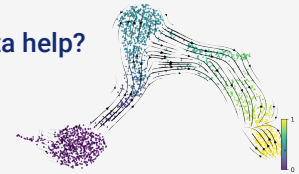
Why?

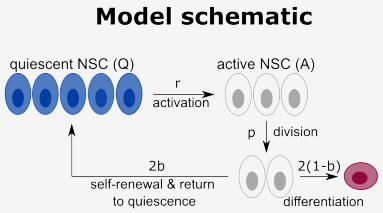
Compartmental models are determined by:

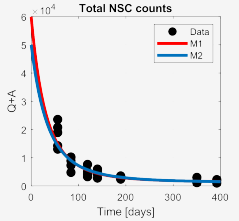
- Graph (compartments + transitions)
- Transition rates between compartments

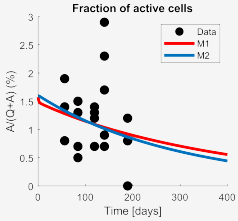
Many models fit the same population data.
(different graphs, rates, non-linearities, ...)

Q: Can single-cell data help?

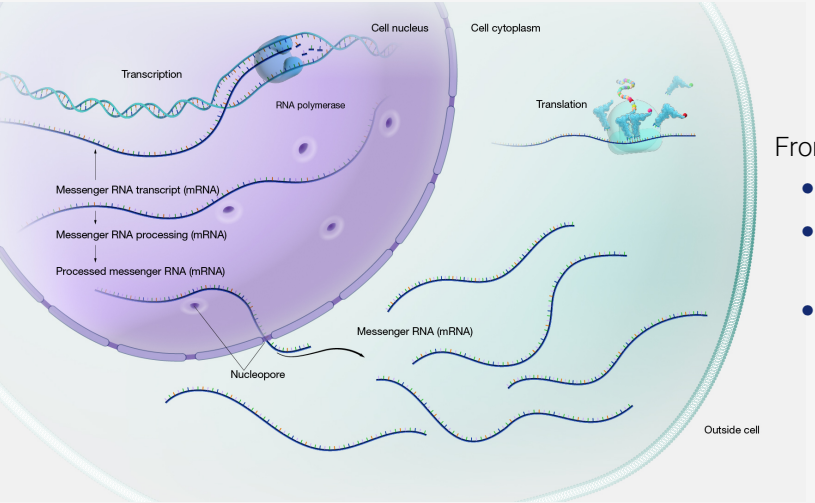








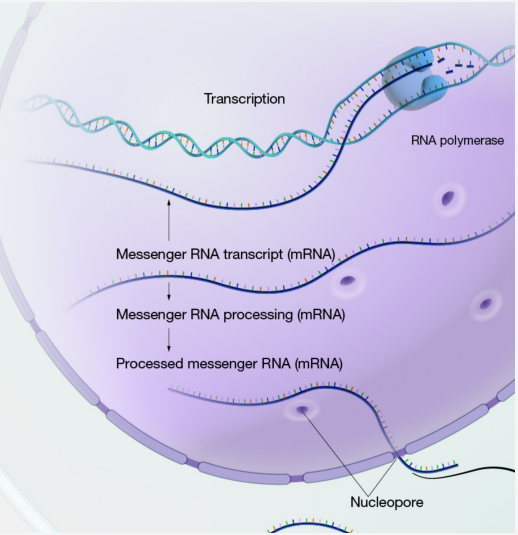
Single-cell Gene Expression



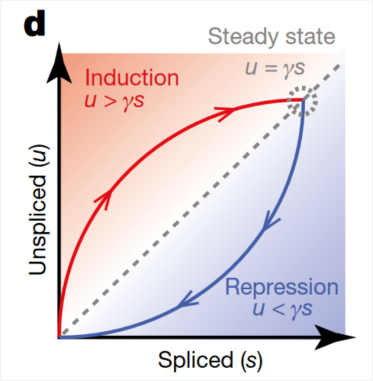
From code to function

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

RNA velocity

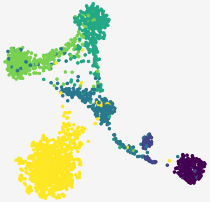


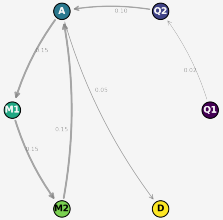
La Manno, G. et al. (2018) 'RNA velocity of single cells', Nature, 560(7719), pp. 494–498. Fig 1.



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

Why single-cell data might help





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 a **hierarchy of problems**.

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: Graph structure = Topology

Can we differentiate between graphs using single cell information?

Approach: Euler Characteristic Curves.

with Marta Marszweska (Gdansk, Warsaw), Justyna Signerska-Rynkowska (Gdansk), Pawel Dlotko (Warsaw)

Next 20 mins: intro to relevant topological concepts and application to dynamical systems.

Directed Graphs

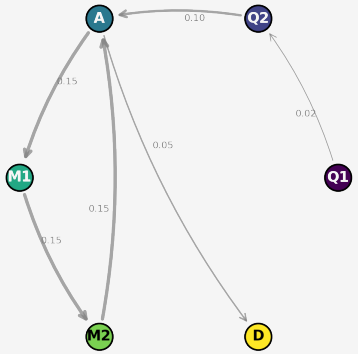
A **directed graph** is a collection of **vertices** connected by **directed edges**.

The Challenge

- For N vertices $\approx 2^{N^2}/N!$ directed graphs.
- For 3 compartments (Q, A, D) have **85 possibilities!**

Need mathematical tool that **distinguishes graph topologies** and that:

- is built from single cell data (many points!),
- captures the underlying Markov chain (few states!),
- is invariant as we go from the first to the second.

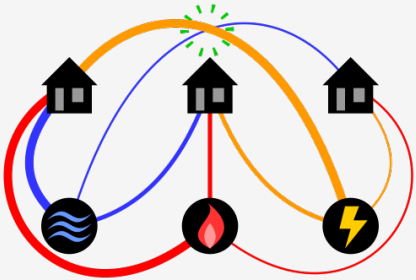


The 3-Utilities Puzzle

The Puzzle

Can you connect the 3 houses to the 3 utilities (water, gas, electricity) without crossing lines?

The 3-Utilities Puzzle



The Puzzle

Can you connect the 3 houses to the 3 utilities (water, gas, electricity) without crossing lines?

Answer: No.

For any graph without crossings in the plane,

$$\chi = V - E + F = 2 \quad \text{Euler's formula } \textit{(not obvious)}$$
$$\implies F = 5$$

But (1) no *house-house* or *utilities-utilities* connections $\implies$ every face boundary has length ≥ 4.
And (2) Each edge is in the boundary of exactly two faces.

$$2E \geq 4F \implies 18 \geq 20 \quad \text{contradiction.}$$

The 3-Utilities Puzzle

The Puzzle (revisited)

Can you connect the 3 houses to the 3 utilities (water, gas, electricity) without crossing lines on a mug?

Answer: Your Turn.

The Euler Characteristic

For graphs $\chi = V - E$

For polygons $\chi = V - E + F$

In general

For a **simplicial complex** K:

$$\chi(K) = \sum_{i=0}^{n} (-1)^i |K_i| = V - E + F - \dots$$

where K_i is the set of i-simplices in K.

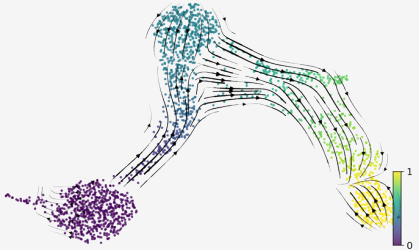
Why is χ useful?

- **Captures connectivity**
  Connected components, cycles, holes.
- **Distinguishes structures**
  Different $\chi \Rightarrow$ different topologies.
- **Generalizes**
  Extends to higher-dimensions.
- **Invariant**
  Preserved under continuous deformations, e.g. stretching, bending, or contracting connected components.
- **Computable**

From Point Clouds to Graphs, and beyond

What We Have (in principle)

- scRNA-seq: high-dimensional point cloud data
- RNA velocity: estimates of where cells are moving



What We Want

- The underlying graph structure of the Markov chain on cell states

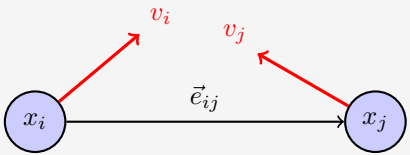
The Key Question

How do we extract graph topology from point cloud data (with velocity information)?

TDA Approach: Connect points at different scales/for different relevant parameter thresholds.

Bifiltered Complexes from RNA-Velocity

1. Start with kNN-graph of the scRNA-seq data.
2. For each edge (x_i, x_j), compute edge vector: $\vec{e}_{ij} = x_j - x_i$.
3. **Filtration 1** (ϵ_1): Cosine distance between $\vec{e}_{ij}$ and velocity v_i at x_i.
4. **Filtration 2** (ϵ_2): Cosine distance between $-\vec{e}_{ij}$ and velocity v_j at x_j.
5. Add higher order simplices by standard clique construction.



What this does

If x_i is moving toward x_j, then v_i aligns with $\vec{e}_{ij}$ (small ϵ_1).
$\Rightarrow$ Observe connection between x_i and x_j at small ϵ_1 values.

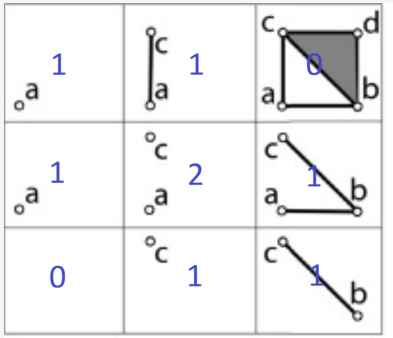
The Euler Characteristic Profile (ECP)

Given a point cloud:

e.g., scRNA-seq data with velocity information

1. For each choice of parameters $(\epsilon_1, \epsilon_2, \ldots, \epsilon_k)$,
2. build a simplicial complex[1] $K_{\epsilon_1,\ldots,\epsilon_k}$,
3. and compute $\chi(K_{\epsilon_1,\ldots,\epsilon_k}) = |V| - |E| + |F| - \ldots$

The **ECP** is the function $\chi : \mathbb{R}^k \to \mathbb{Z}$ obtained from this.



Intuition

- Multi-parameter filtrations capture topological information for all parameter choices simultaneously.
- Different network topologies produce different ECPs.

[1] Need to be *included* in each other as we increase parameter values.

Challenges and Considerations

Technical Challenges

- **Curse of Dimensionality:** Noisy vectors are essentially orthogonal.
- **Computational cost:** 2D parameter space for bifiltration
- **Sampling:** How many cells are sufficient for reliable ECP estimates?

Fundamental Questions

- Which graph features are topologically distinguishable?
- How do transition rates affect ECP signatures?
- How does noise affect ECP signatures?

Challenges and Considerations: Can ECP distinguish between different graphs?

Goal

Systematically understand which topological features ECP can capture.

Create synthetic benchmark

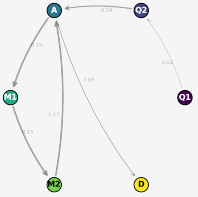
1. Select candidate graph topologies
   (e.g., linear chains, cycles, branching structures, graphs with/without specific edges)
2. For each graph: simulate data with known ground truth
3. Build bifiltered complexes (kNN + velocity alignment)
4. Compute ECP for each dataset
5. Use distances between ECPs to run clustering / statistical tests.
6. Assess: which features are distinguishable, which aren't?

Synthetic Data: A Markov-Modulated Splicing Model

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

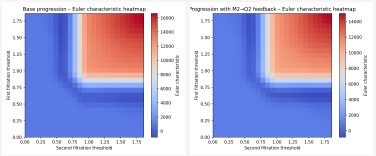
- **Latent state process:**
  Continuous-time Markov chain on a state graph
  - States $z_1, \ldots, z_n$ with probability p_i
  - $\frac{d}{dt} p_i = \sum_j Q_{ij} p_j$
  - Asymmetric transition rates $Q_{ij} \neq Q_{ji}$, encode directionality
- **Gene expression:**
  Standard transcription-splicing-degradation 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)

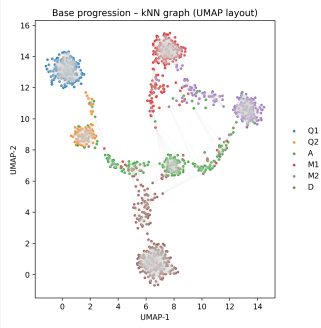
Output: u/s-counts + **current state** for each cell at time t.

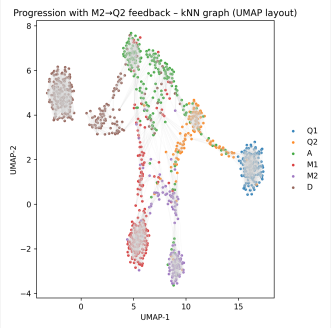


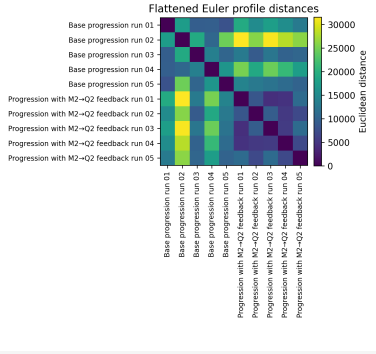
Preliminary Results

- 5 simulation runs each from two graphs: one with $M2 \to Q2$ transition, one without.
- Computed ECP for each dataset.
- Computed pairwise euclidean distances between flattened ECPs.









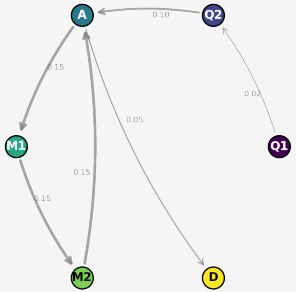
Application Goal: Uncovering Cell State Transitions

The Biological Question

After benchmarking validates the method, we can apply it to answer:

Consider the (Q, A, D) system:

- Does an $A \to Q2$ transition exist? How about an $A \to Q1$ transition? Or an $M2 \to Q1$ transition?
- That is: can cells land in (deep) quiescence after division?



Current Status and Next Steps

What We Have

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

Immediate Next Steps

1. Develop benchmarks (standard datasets + scoring)
2. Validate on benchmark: ECP provides meaningful classification/clustering
3. Move beyond classification/clustering; can we recover the ground truth?

Longer-Term Goal

Apply to scRNA-seq data to test biological hypotheses about cell fate transitions and constrain population-level models.