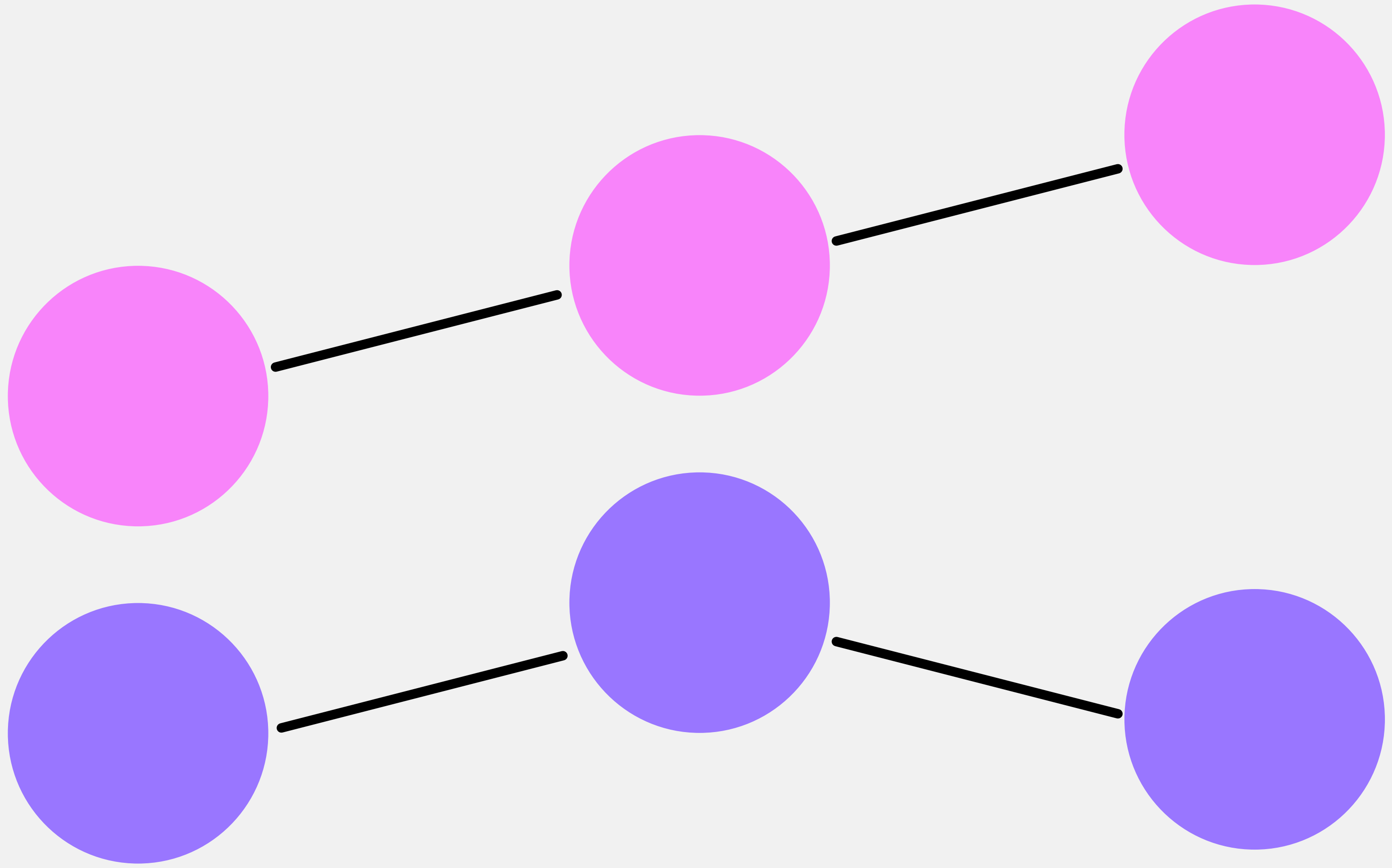
MACHINE LEARNING 4 HEALTHCARE 2021

TAGS: Time Adaptive Global State Model



Sujay Nagaraj, Cait Harrigan, Aslesha Pokhrel

Motivation

Unsupervised state detection is a general challenge in heathcare data

- Don't have ground truth for underlying population structure
- Can be hard to compare across individuals

One approach: define physiological states based on signal correlations (Tozzo et al. 2021)

- Efficiently summarize high-dimensional relationships with sparse graphical structure
- Strong physiological grounding - relationship between features more important than the raw feature values
- Interpretable - state visitation can be directly compared within individuals

Problem Setting - Stress

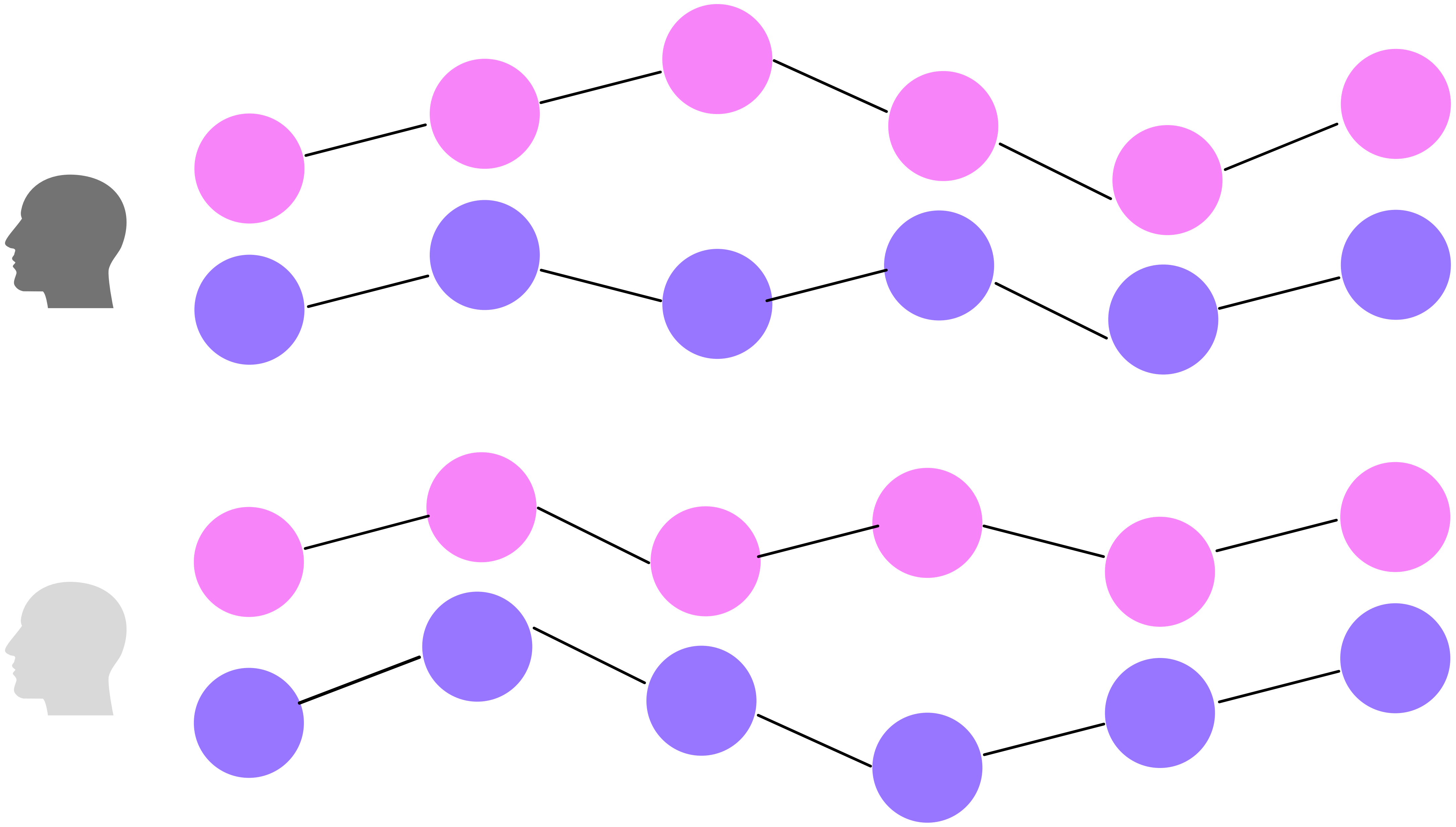
- Wearables collect high-dimensional and high-resolution timeseries data
- Stress is latent and difficult to measure
  - We don't have good labels in the wild
  - We don't fully understand how it manifests

Problem Setting - Stress

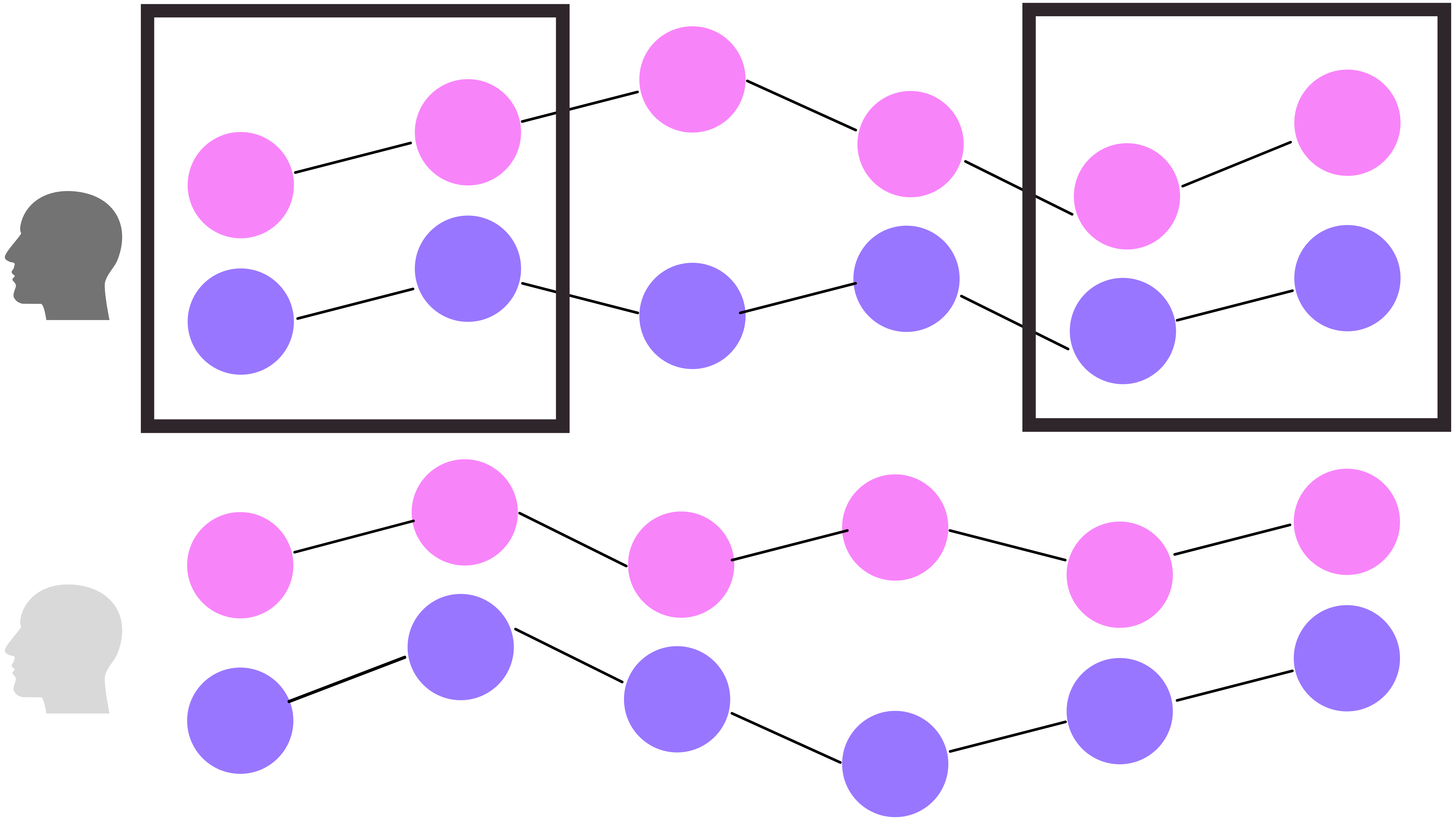
- Wearables collect high-dimensional and high-resolution timeseries data
- Stress is latent and difficult to measure
  - We don't have good labels in the wild
  - We don't fully understand how it manifests
- Stress might look different for different people
  - Overarching question: are there subtypes (or archetypes) of stress
- State change can be volatile
- Stress is it's own disease process - but also may impact other diseases

State Space Models for clustering are needed

Motivating Example

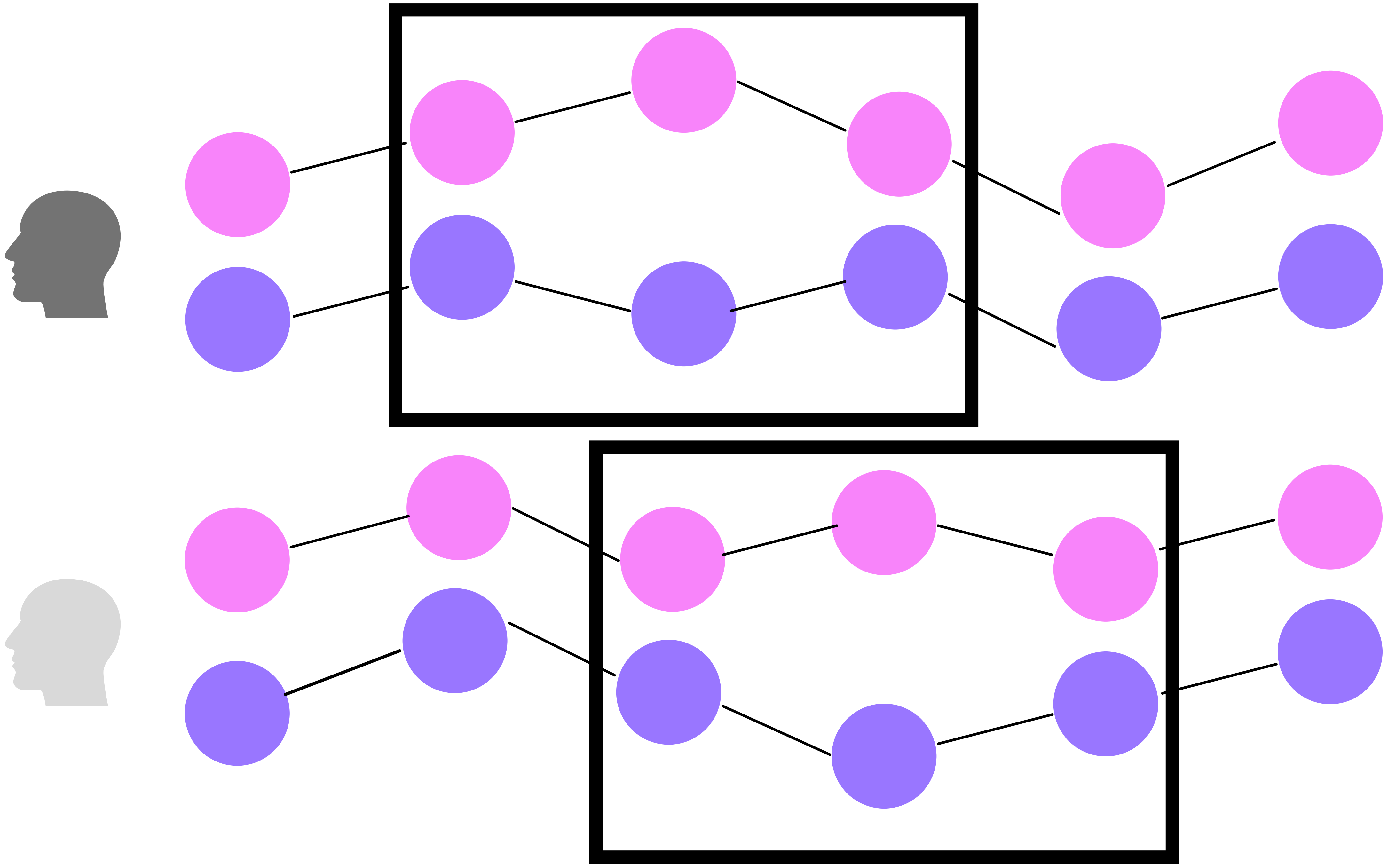


Motivating Example



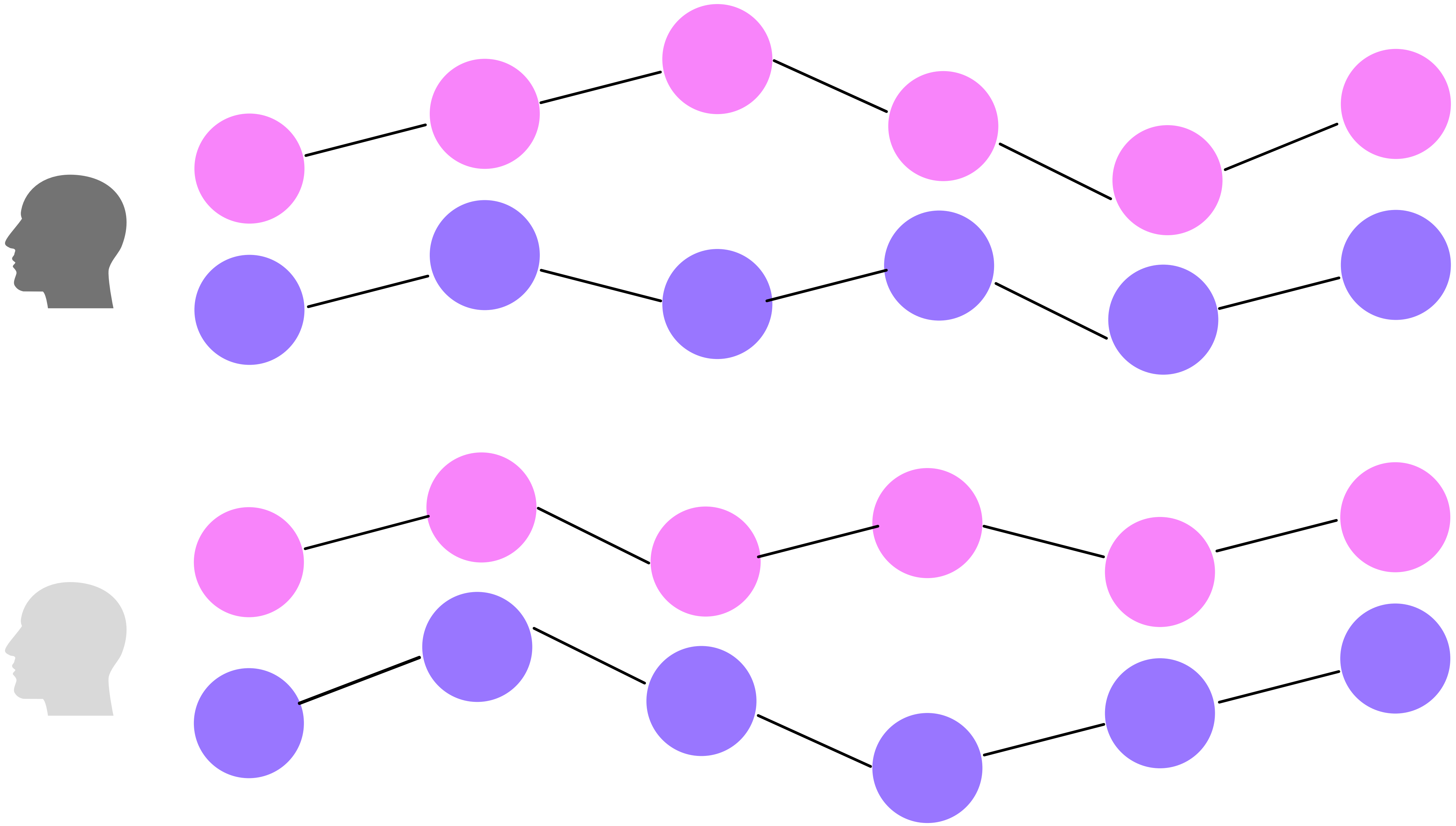
Local States

Motivating Example



Global States

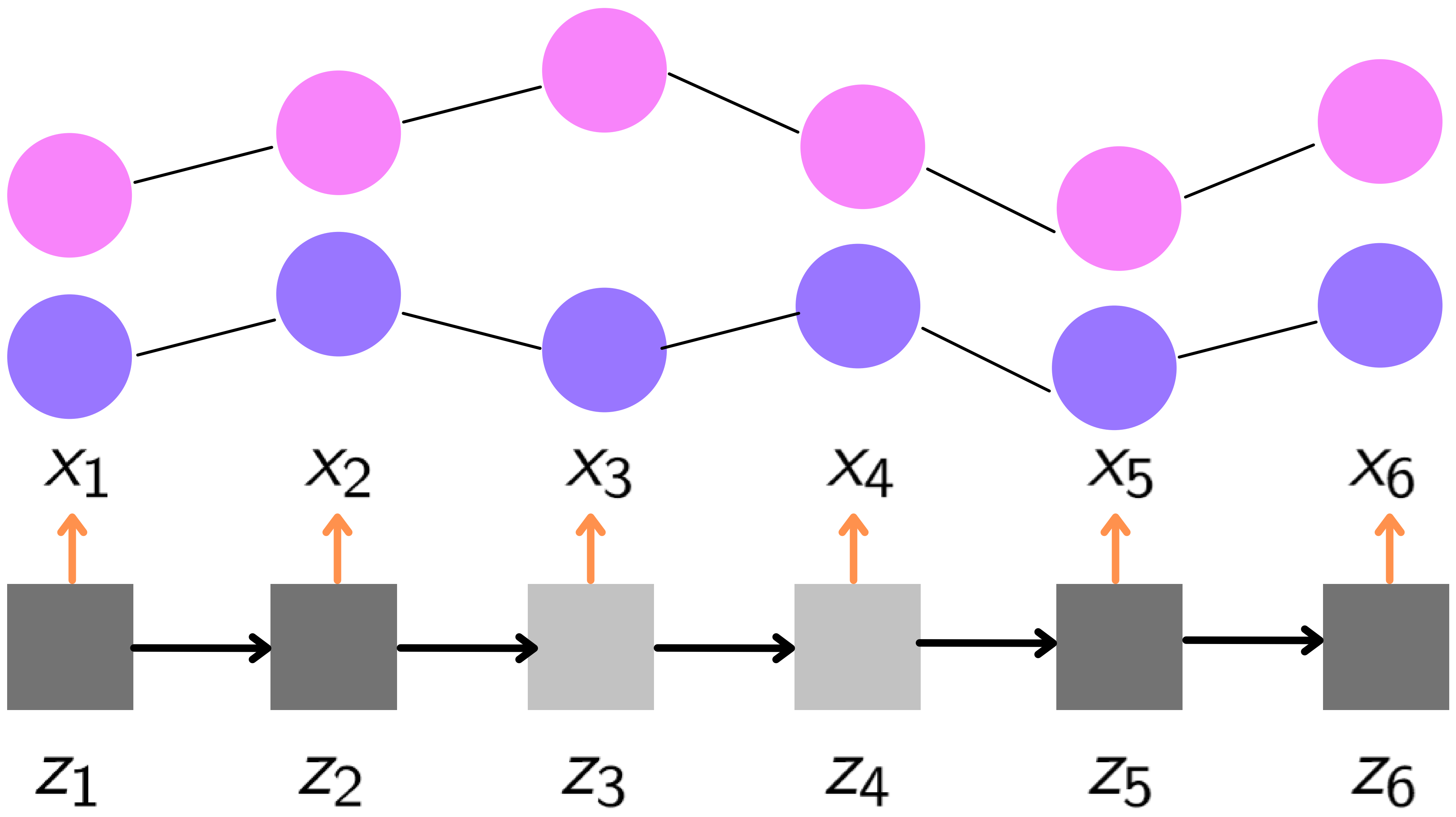
Motivating Example



Sharing **global information** about states **between individuals** may give us insight into population structure

- What states some or all individuals pass through?
- Are there sub-populations within characteristic state visitation behaviours?
- Is there heterogeneity within a given state across individuals?

Timeseries is modelled as HMM



$\longrightarrow$ Transition probability $p(z_{t+1}|z_t)$

$\uparrow$ Emission probabilities $p(x_t|z_t)$

$x_1, x_2, x_5, x_6 \sim N(\blacksquare)$

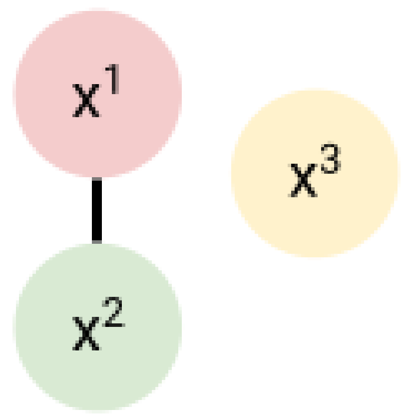
$x_3, x_4 \sim N(\square)$

Features are modeled as multivariate normal, parameterized according to the state assignment

State Definitions

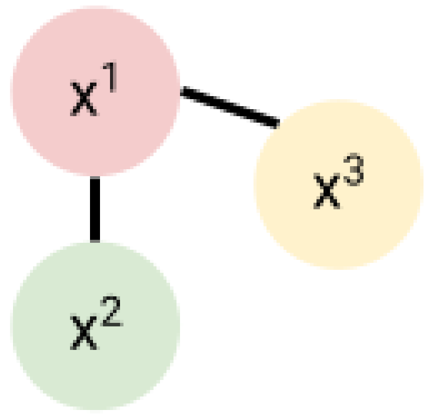
- Each state is defined by conditional dependencies between features
- Edge in graph corresponds to >0 covariance
- Physiological grounding - each patient may have their own baseline feature value

$\theta_1 =$



$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

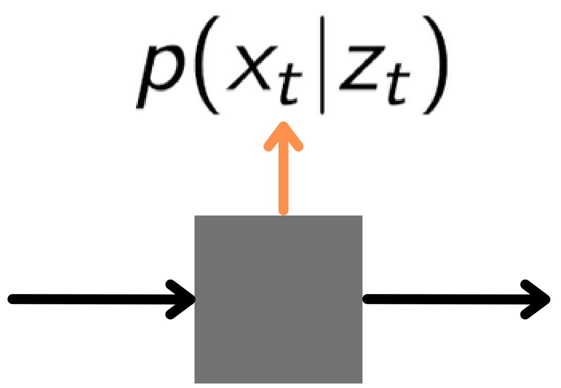
$\theta_2 =$



$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

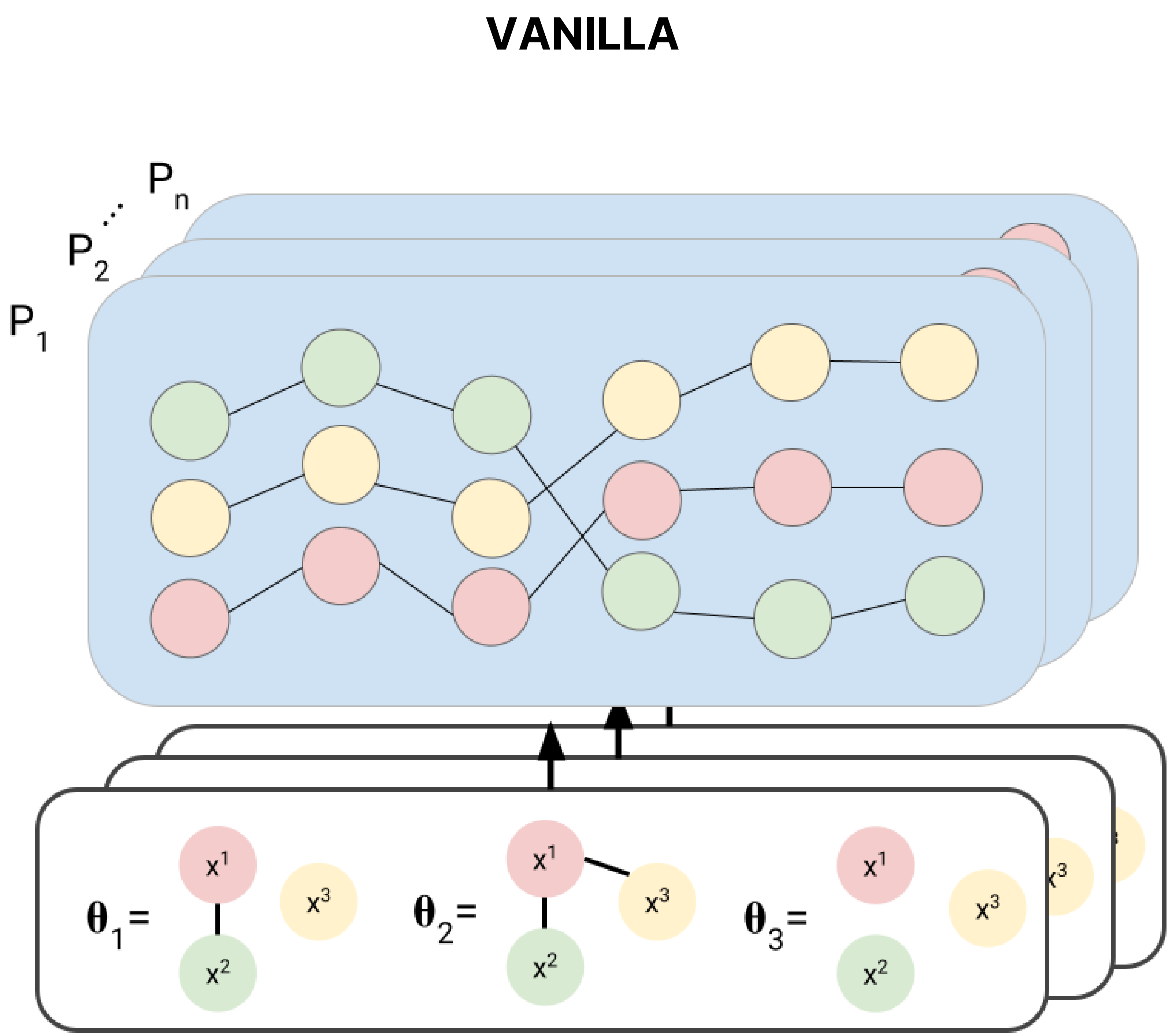
$$x_1, x_2, x_5, x_6 \sim N(\mu_p, \theta_1^{-1})$$

$$x_3, x_4 \sim N(\mu_p, \theta_2^{-1})$$



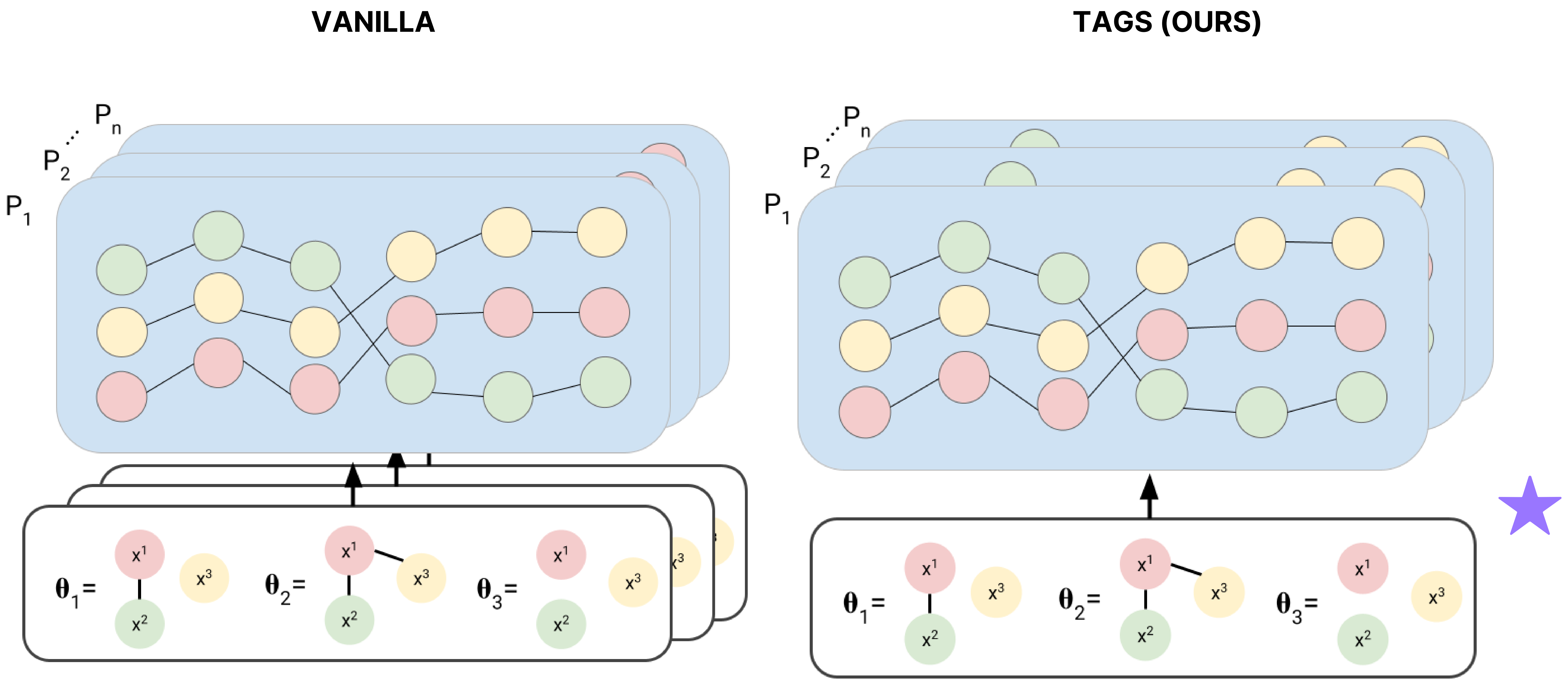
Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3):432–441, 2008

Model

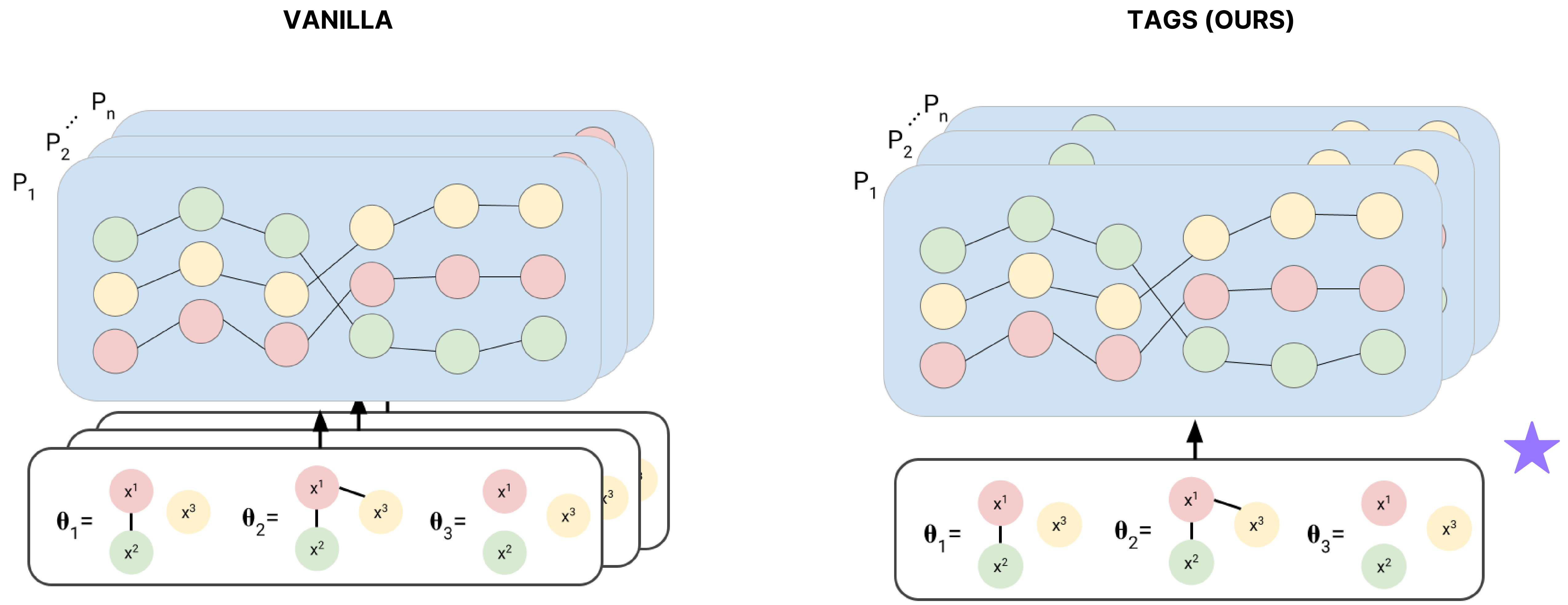


Veronica Tozzo, Frederico Ciech, Davide Garbarino, Alessandro . Verri Statistical Modela Coupling Allows for Complex Local Multivariate Time Series Analysis, in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, New York, NY, USA, Aug. 2021

Model



Veronica Tozzo, Frederico Ciech, Davide Garbarino, Alessandro . Verri Statistical Modela Coupling Allows for Complex Local Multivariate Time Series Analysis, in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, New York, NY, USA, Aug. 2021



	Vanilla	TAGS
	Fits one single time series	Fits several time series with state information shared between patients
E-step M-step	Update state assignments for all timepoints Update individual $\mu_p, \pi_p, A_p, \theta_p$	Update state assignments for all timepoints Update individual μ_p, π_p, A_p and global θ

EM for HMMs

Goal: Learn parameters $\underset{\text{Shared parameters}}{\theta}, \underbrace{\overset{\text{baseline mean}}{\mu_p}, \overset{\text{initial state probability}}{\pi_p}, \overset{\text{state transition matrix}}{A_p}, \overset{\text{state assignment of timepoint n (one-hot)}}{z_{pn}}}_{\text{Inidividual parameters}}$

state definitions

baseline mean

initial state probability

state transition matrix

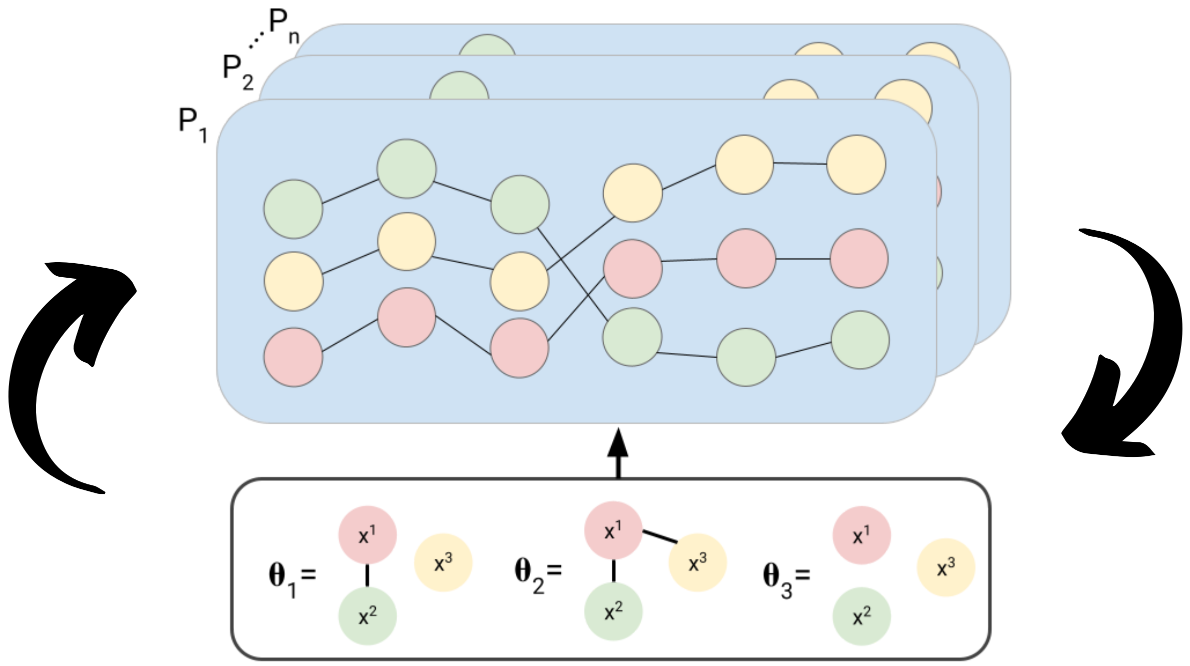
state assignment of timepoint n (one-hot)

Shared parameters

Inidividual parameters

Model posterior: $$\prod_{p=1}^{P}\left(p(z_{p,1}|\pi_p)\prod_{n=2}^{N_p}[p(z_{pn}|z_{p,n-1}, A_p)]\prod_{n=1}^{N_p}\prod_{k=1}^{K}\mathcal{N}(x_{pn}|\mu_{pk}, \theta_k^{-1})^{z_{pnk}}\right)$$

EM for HMMs



E-step: Assign each timepoint to a state (via maximum likelihood)

$$\prod_{p=1}^{P}\left(p(z_{p,1}|\pi_p)\prod_{n=2}^{N_p}[p(z_{pn}|z_{p,n-1},A_p)]\prod_{n=1}^{N_p}\prod_{k=1}^{K}\mathcal{N}(x_{pn}|\mu_{pk},\theta_k^{-1})^{z_{pnk}}\right)$$

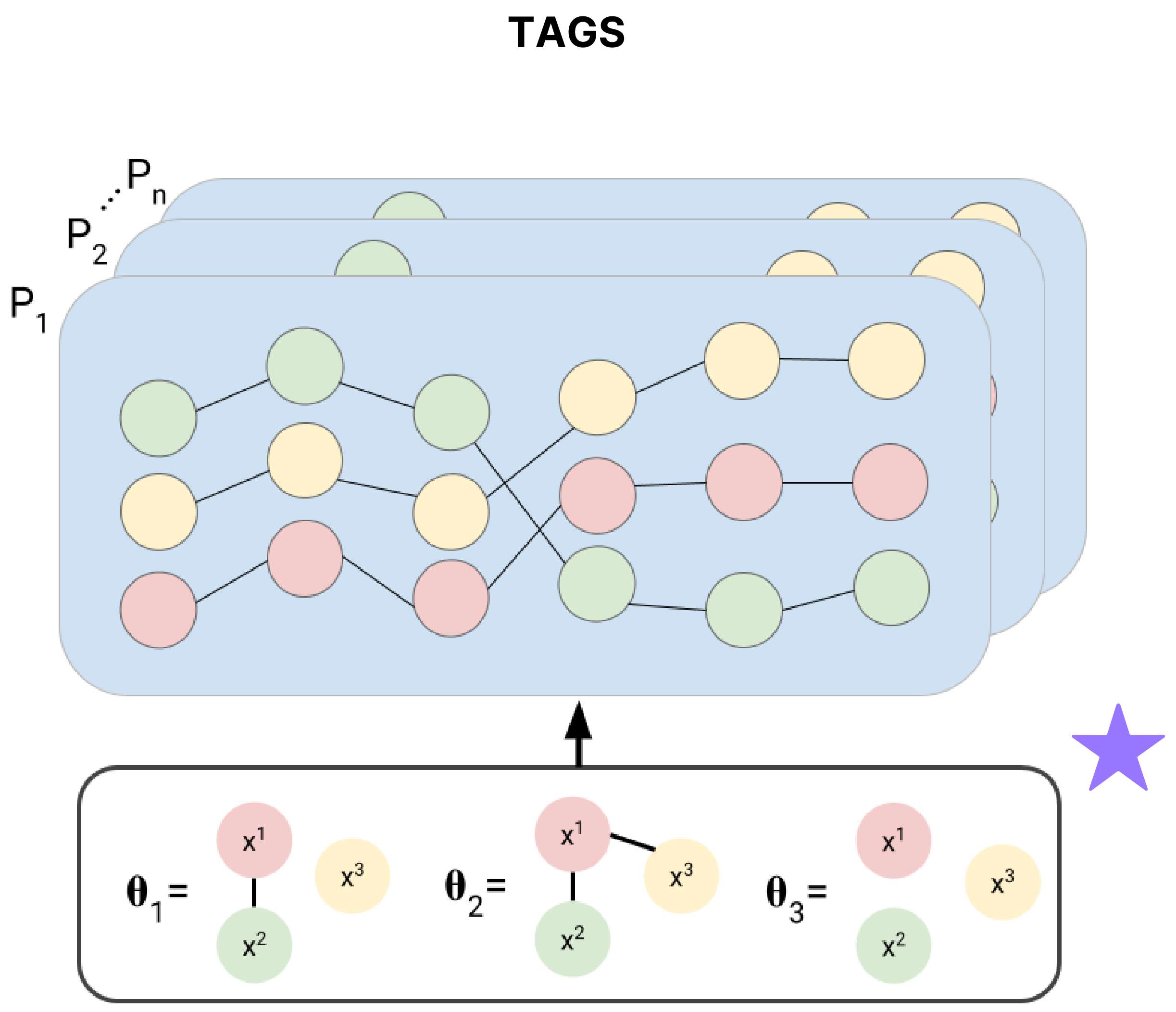
M-step: Update how each state is defined (based on all the timepoints assigned to them)

$$\prod_{p=1}^{P}\left(p(z_{p,1}|\pi_p)\prod_{n=2}^{N_p}[p(z_{pn}|z_{p,n-1},A_p)]\prod_{n=1}^{N_p}\prod_{k=1}^{K}\mathcal{N}(x_{pn}|\mu_{pk},\theta_k^{-1})^{z_{pnk}}\right)$$

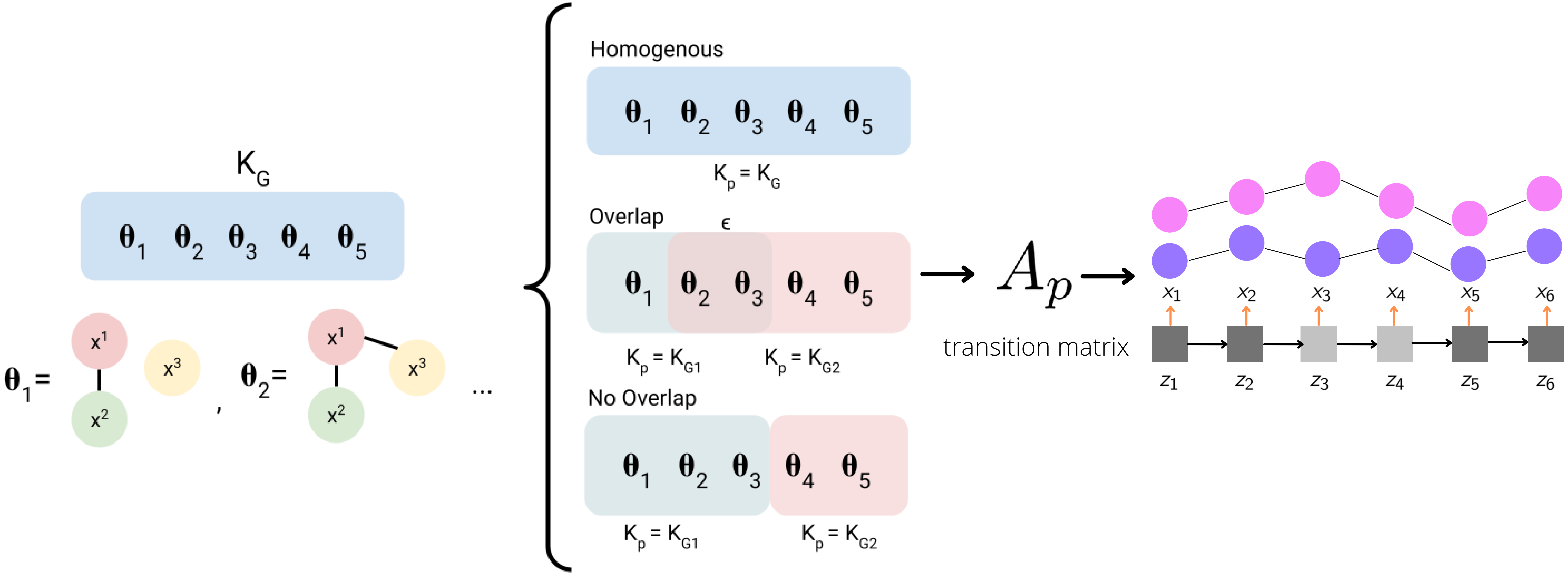
States are clustered during training

Our Contributions

- Shared set of global states for all time-series that is jointly optimized
- States are interpretable with respect to underlying biological processes
- State visitation can be comparable across patients



Dataset generation

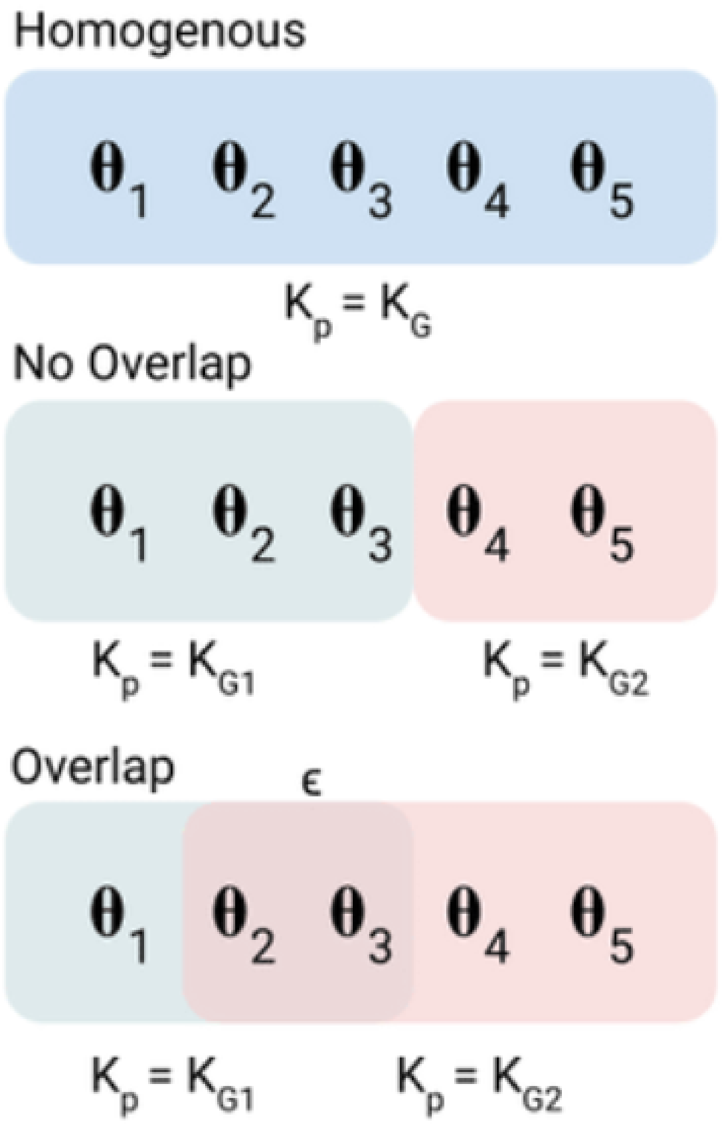


Overlap parameter epsilon

0 ← → 1

No overlap
Distinct sub-populations

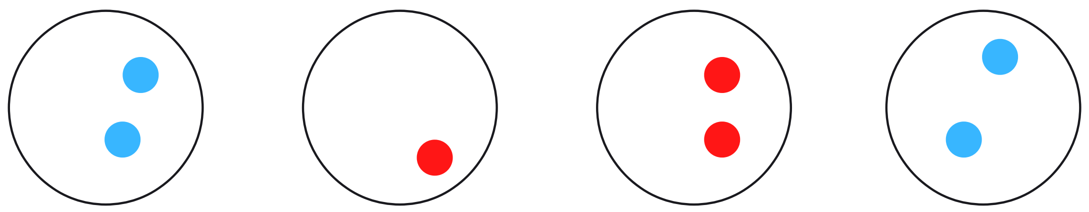
Homogenous
No sub-population structure



We tested our model's performance by simulating a variety of population structure settings: epsilon = [0, 0.2, 0.4, 0.6, 0.8, 1]

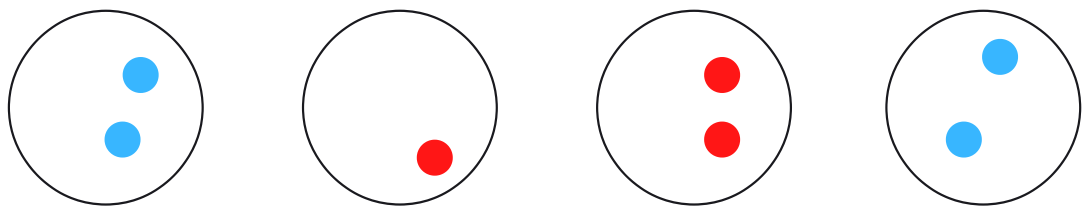
V-Measure

V-Measure

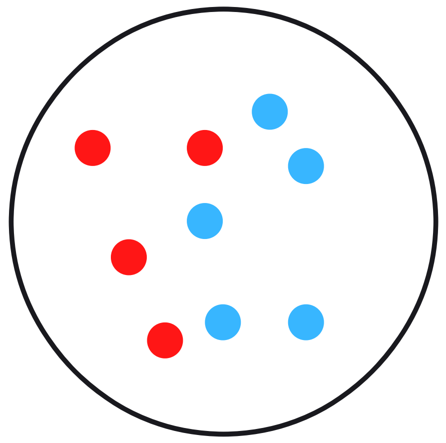


High Homogeneity, Low Completeness

V-Measure

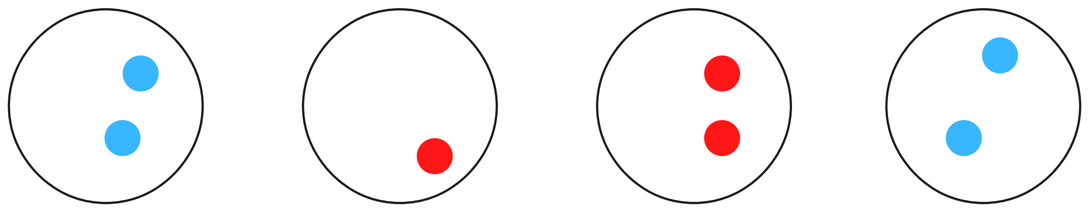


High Homogeneity, Low Completeness

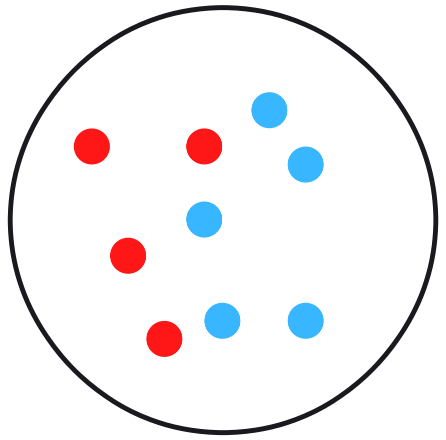


High Completeness, Low Homogeneity

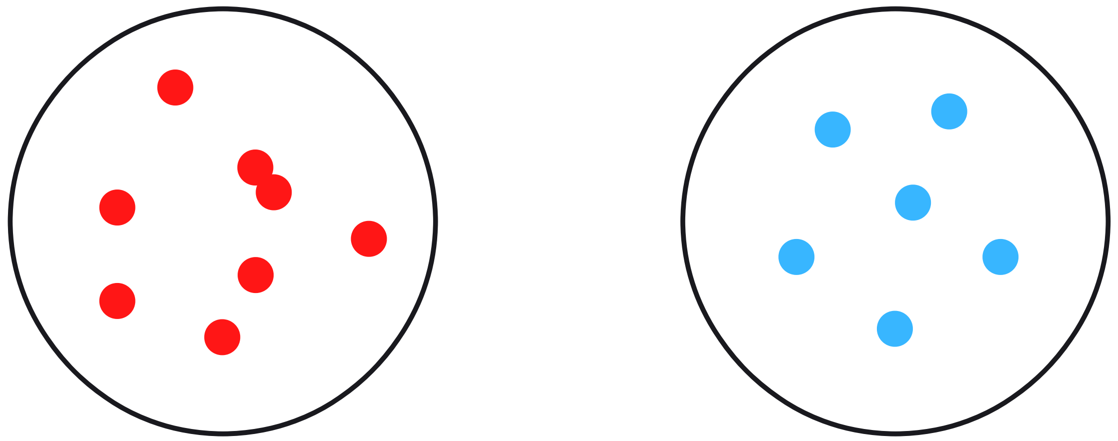
V-Measure



High Homogeneity, Low Completeness

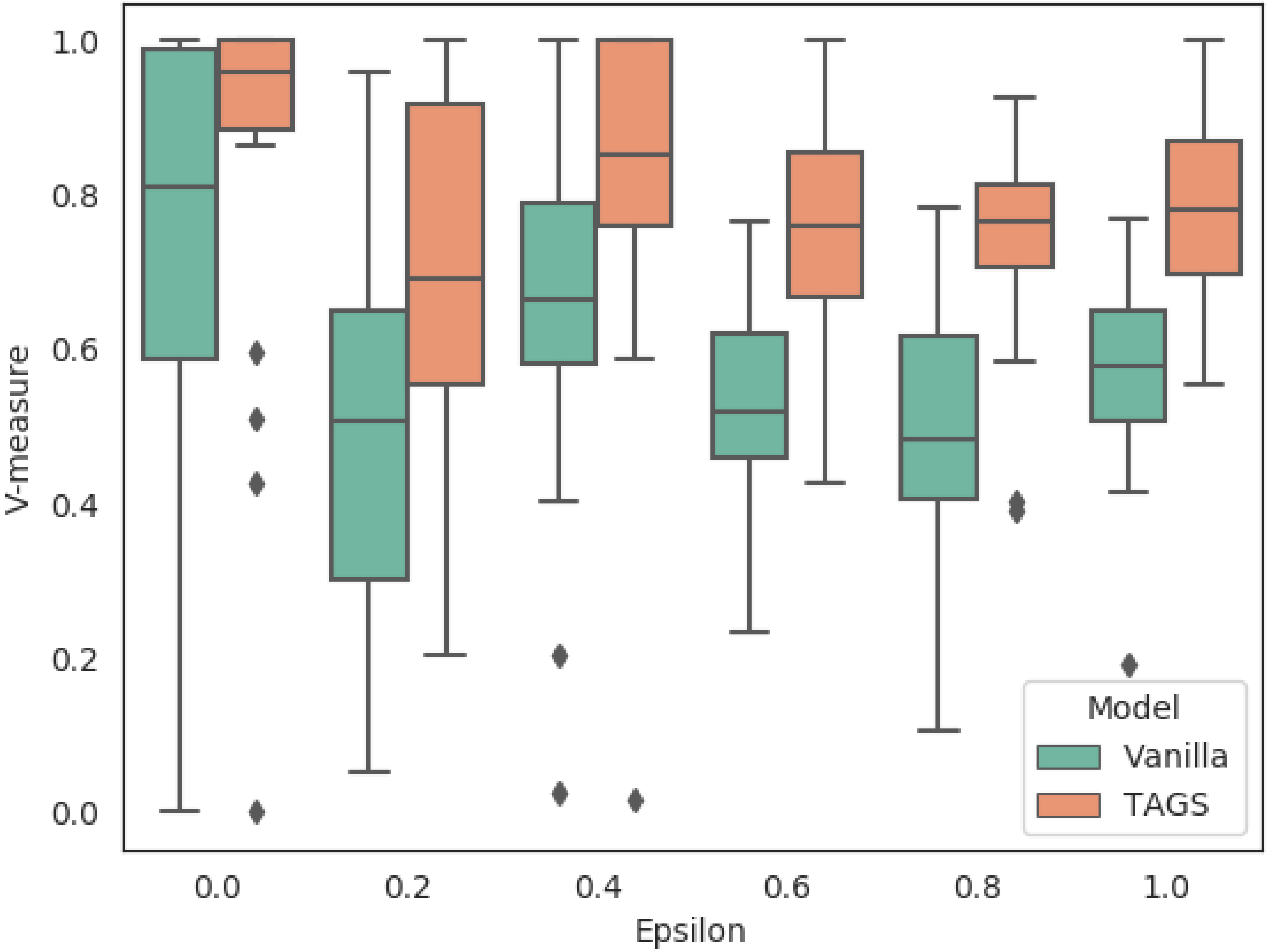


High Completeness, Low Homogeneity



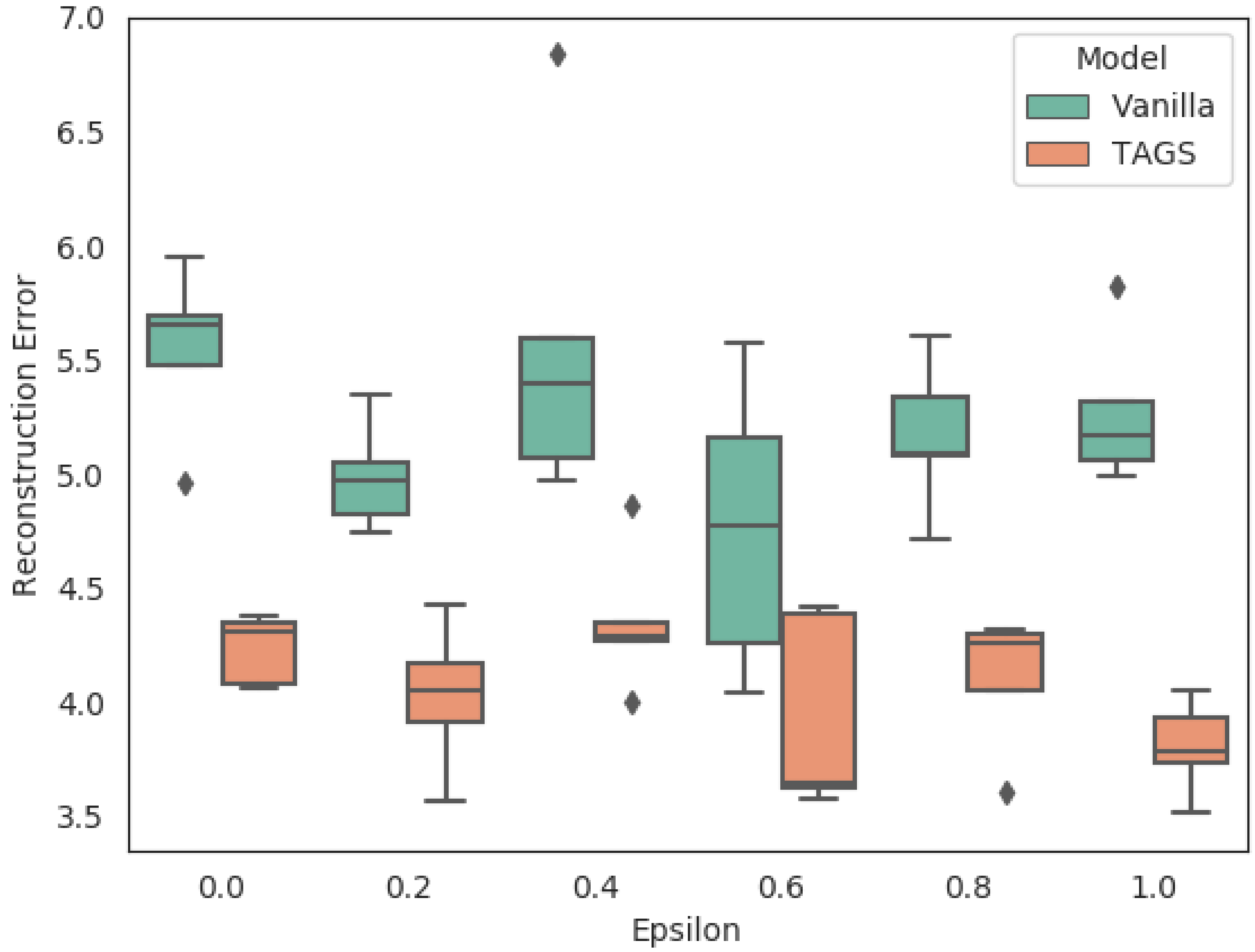
V-Measure: harmonic mean between completeness and homogeneity

State Clustering Performance



*Higher is better
1= perfect clustering

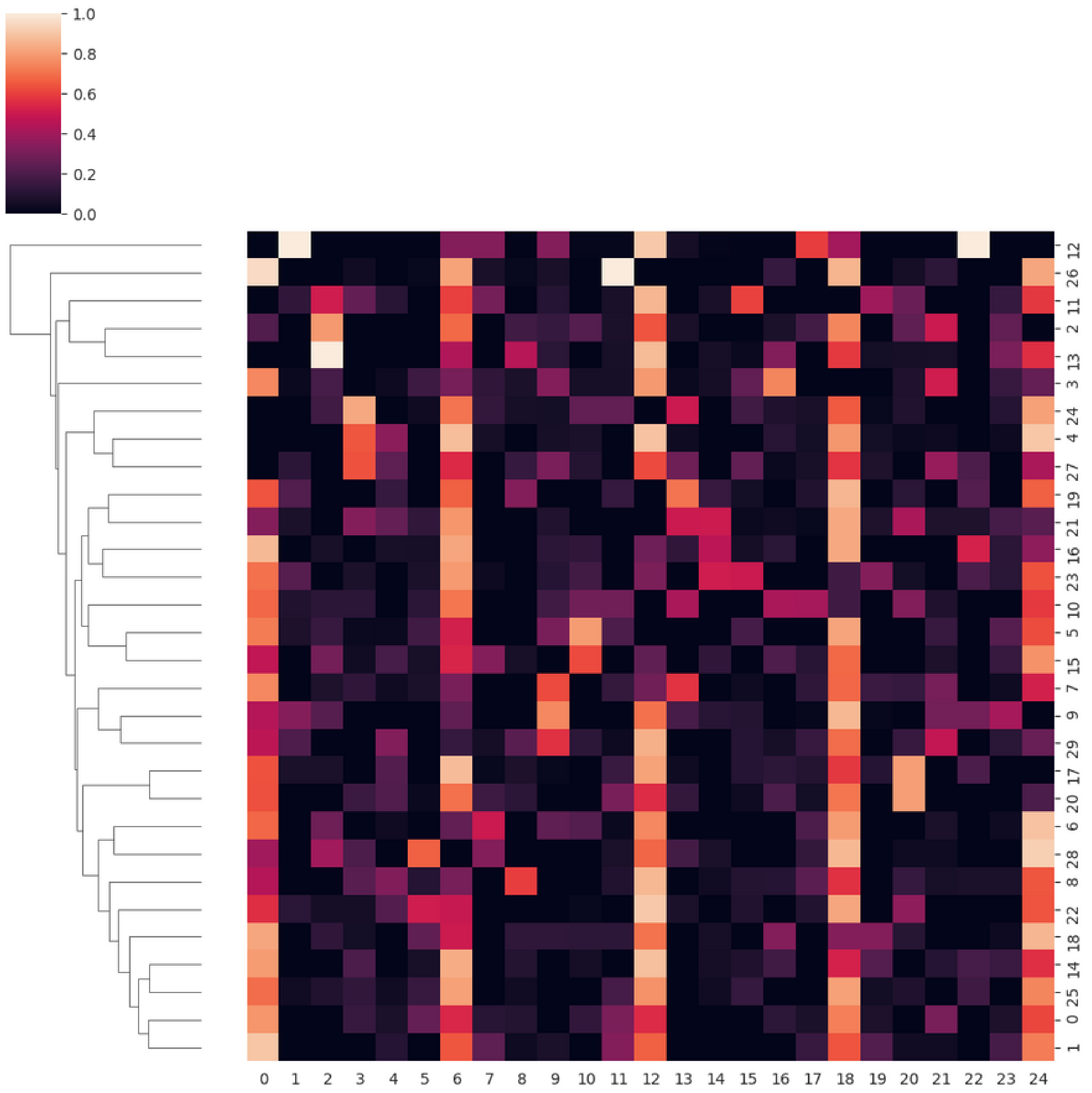
State Reconstruction



Reconstruction Error: Euclidean distance between predicted graph and true graph

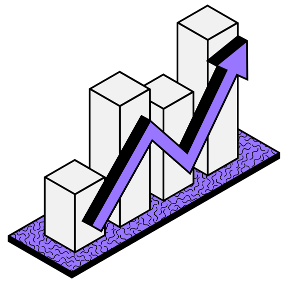
*Lower is better

Transition Probability clustering



Our model makes it possible to compare state visitation across individuals (not possible without global state information)

Key Findings

- TAGS has better clustering performance
- Improved state reconstructions 
- Interpretable!
- Faster!

Limitations

- Feature missingness is not handled.
- We still have to specify the number of states locally and globally (hyperparameter).
  - Did not test robustness against a misspecified number of states.
  - Not flexible to unseen states (fixed K limits predictive power).
- Assumes that sharing state information across patients is useful. This may not be true in all settings.
  - The performance gain is probably higher in a large-population regime.

Next Steps

- Finish running on real-world data ★
    - State detection from wearables - human activity recognition
- Recreating the global subgroups given the state assignments
- More complex graph clustering techniques
- Hierarchical state definitions (baseline global + individual variation)

Do you have any questions?

Modified Baum-Welch (Variant of EM algorithm for HMMs)

Goal: Learn parameters θ μ_p, π_p, A_p

Posterior: $$\prod_{p=1}^{P} \left(p(z_{p,1}|\pi_p) \prod_{n=2}^{N_p} [p(z_{pn}|z_{p,n-1}, A_p)] \prod_{n=1}^{N_p} \prod_{k=1}^{K} \mathcal{N}(x_{pn}|\mu_{pk}, \theta_k^{-1})^{z_{pnk}} \right)$$

E-Step: update state assignments given observed data

M-step (i): Update all individual's private parameters

M-step (ii): Update all globally shared population parameters via Graphical Lasso

Modified Baum-Welch (Variant of EM algorithm for HMMs)

Goal: Learn parameters θ μ_p, π_p, A_p

E-Step: update state assignments given observed data

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n | \mathbf{X}, \theta_{\text{old}})$$
$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n | \mathbf{X}, \theta_{\text{old}})$$

M-step (i): Update all individual's private parameters

M-step (ii): Update all globally shared population parameters via Graphical Lasso

Modified Baum-Welch (Variant of EM algorithm for HMMs)

Goal: Learn parameters θ μ_p, π_p, A_p

E-Step: update state assignments given observed data

M-step (i): Update all individual's private parameters

$$A_{j,k} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{n,k})}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{n,l})}$$

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{n,k}) \mathbf{x}_n}{\sum_{n=1}^{N} \gamma(z_{n,k})}$$

M-step (ii): Update all globally shared population parameters via Graphical Lasso

Modified Baum-Welch (Variant of EM algorithm for HMMs)

Goal: Learn parameters θ μ_p, π_p, A_p

E-Step: update state assignments given observed data

M-step (i): Update all individual's private parameters

M-step (ii): Update all globally shared population parameters via Graphical Lasso

$$\Theta_k = \underset{\Theta > 0}{\operatorname{argmin}}\, tr(\Theta S_k) - \log\det(\Theta) + \lambda||\Theta||_{1,od}$$

$$S_k = \frac{1}{\sum_{p=1}^{P}\sum_{n=1}^{N_p} z_{pnk}} \sum_{p=1}^{P}\sum_{n=1}^{N_p} z_{pnk}(x_{pn} - \hat{\mu}_k)(x_{pn} - \hat{\mu}_{pk})^T$$

$$\hat{\mu}_k = \frac{1}{\sum_{p=1}^{P}\sum_{n=1}^{N_p} z_{pnk}} \sum_{p=1}^{P}\sum_{n=1}^{N_p} z_{pnk} x_{pn}$$

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3):432–441, 2008