Luca Pasquali
Fachbereich Mathematik und Informatik, Freie Universität Berlin

Freie Universität Berlin

VAMPnets

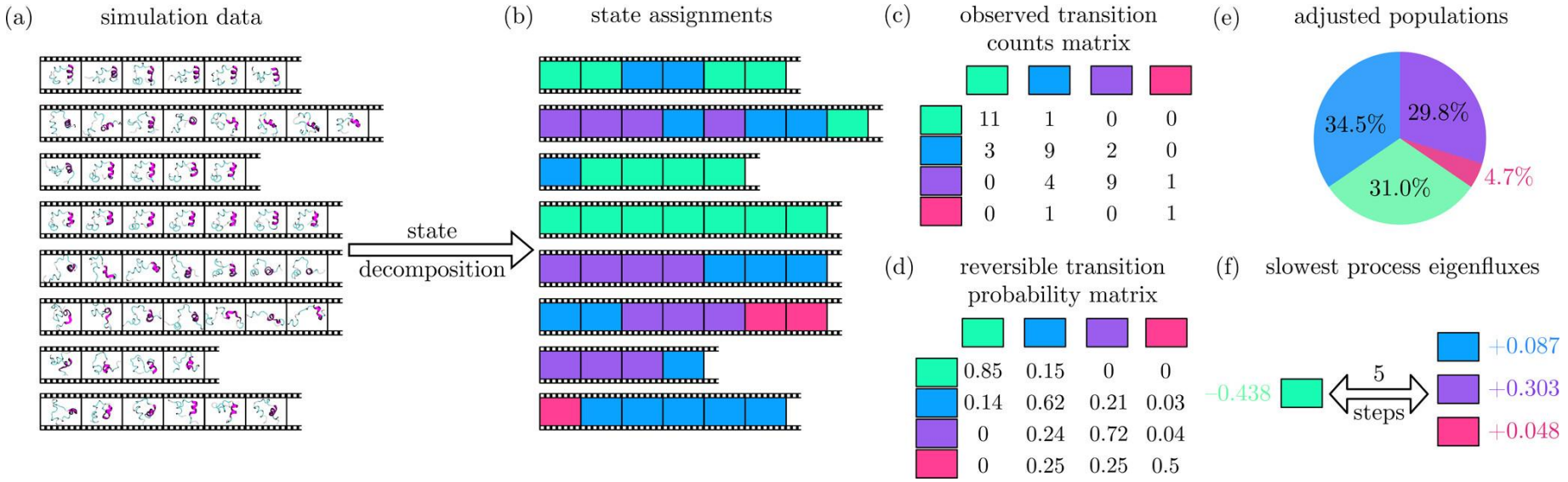
Deep Learning of Molecular Dynamics

Luca Pasquali
Frank Noé, Freie Universität Berlin
Pyemma Workshop, 19/02/2020

Introduction

High-dimensional raw data → Interpretable transition-based model

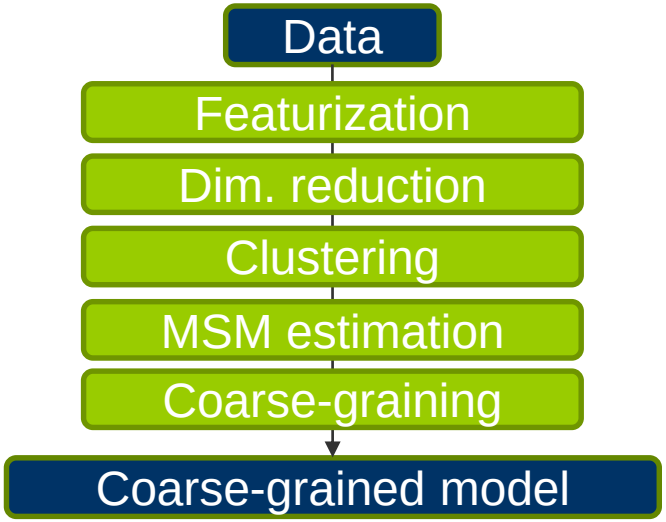
MSMs



Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. Journal of the American Chemical Society 2018, 140, 2386–2396.

Motivation

- Data processing pipeline for dynamical analysis is prone to errors
- We propose a neural network (NN) framework to substitute the whole pipeline
- Entire mapping from molecular coordinates to coarse-grained model encoded in a NN



Koopman Theory

$$\chi_0, \chi_1 : \mathbb{R}^n \to \mathbb{R}^m, \qquad \mathbb{E}_t[\chi_1(x_{t+\tau})] \approx K^T \mathbb{E}_t[\chi_0(x_t)]$$

Given χ_0, χ_1, the K that minimizes the prediction error is:

$$\mathrm{K} = C_{00}^{-1} C_{01}, \qquad \begin{aligned} C_{00} &= \mathbb{E}_t[\chi_0(x_t)\chi_0(x_t)^T] \\ C_{01} &= \mathbb{E}_t[\chi_0(x_t)\chi_1(x_{t+\tau})^T] \\ C_{11} &= \mathbb{E}_t[\chi_1(x_{t+\tau})\chi_1(x_{t+\tau})^T] \end{aligned}$$

This leaves the choice of χ_0 and χ_1

Wu H., Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. *J Nonlinear Sci* **30**, 23-66 (2020). https://doi.org/10.1007/s00332-019-09567-y

VAMP score

We define the VAMP-2 score of two functions χ_0, χ_1:

$$\widehat{R_2}[\chi_0, \chi_1] = \left\| C_{00}^{-1/2} C_{01} C_{11}^{-1/2} \right\|_F^2$$

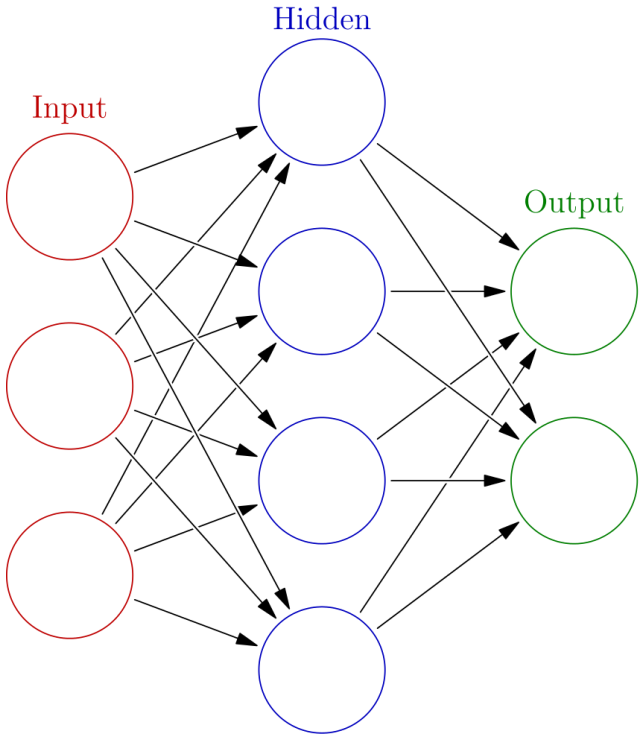
- VAMP variational principle:

  Maximum value of $\widehat{R_2}$ is achieved when the m dominant left and right Koopman singular functions are a linear combination of $(\chi_{01}, \dots, \chi_{0m})$ and $(\chi_{11}, \dots, \chi_{1m})$ respectively

- We can identify the χ_0 and χ_1 which can preserve the dominant processes in the Koopman operator via maximization of the VAMP-2 score

Wu H., Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. *J Nonlinear Sci* **30**, 23-66 (2020). https://doi.org/10.1007/s00332-019-09567-y

Neural Networks



$$\hat{y} = f_2(f_1(x * W_1)W_2)$$

- Layers consist of nodes, which are connected to the nodes of the subsequent layer
- Connections are weighted through trainable parameters W_i
- Nodes apply nonlinear functions f_i to the sum of inputs
- Universal approximation theorem: 3-layered NNs can approximate any function

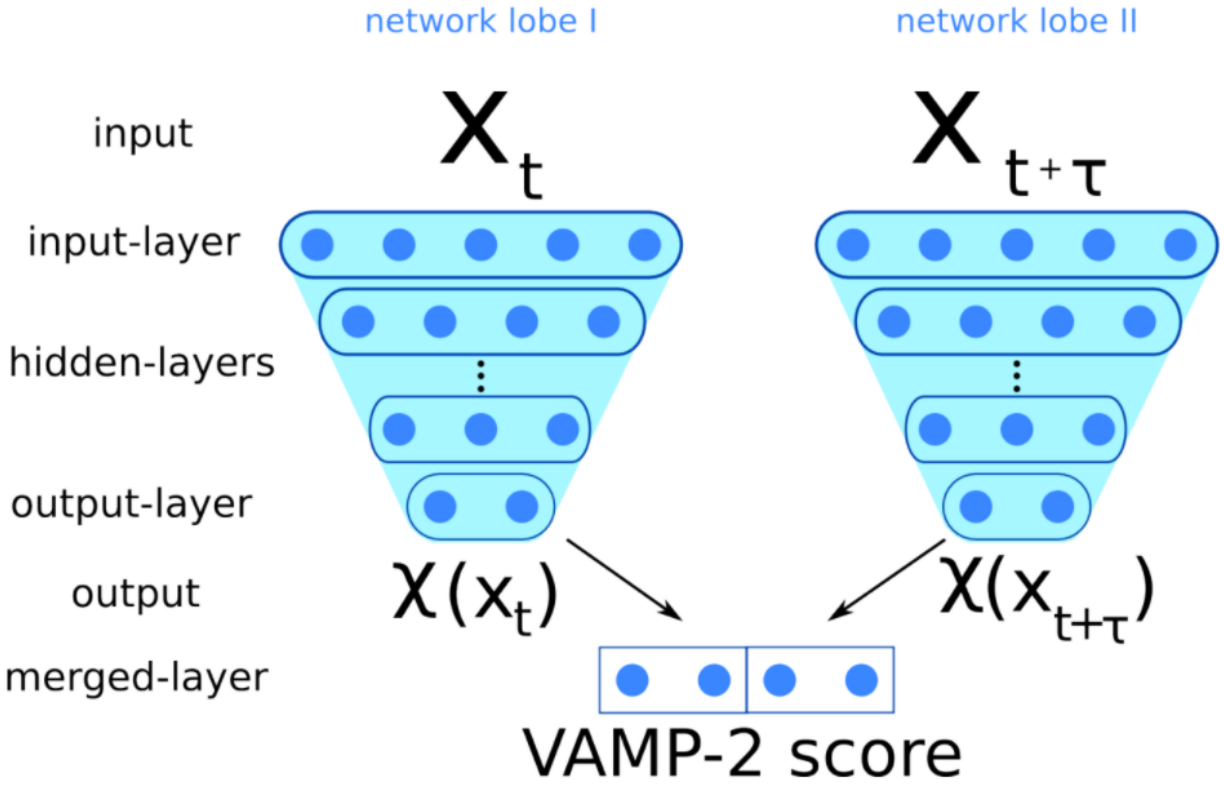
NN training

- Process of updating the parameters of the network
- Target of the update is the minimization of a given loss function:

1. Initialize $\boldsymbol{w}$
2. Until stopping criterion is reached:
    1. For every training sample $\boldsymbol{x}^{(i)}$:
        1. Compute output value $\hat{\boldsymbol{y}}^{(i)}$
        2. Compute value of loss function L($\hat{\boldsymbol{y}}^{(i)}$)
        3. Compute $\nabla_{\boldsymbol{W}} L$
        4. Update $\boldsymbol{w}$
    2. Evaluate $F_{\boldsymbol{w}}$ on validation set
3. Evaluate $F_{\boldsymbol{w}}$ on test set

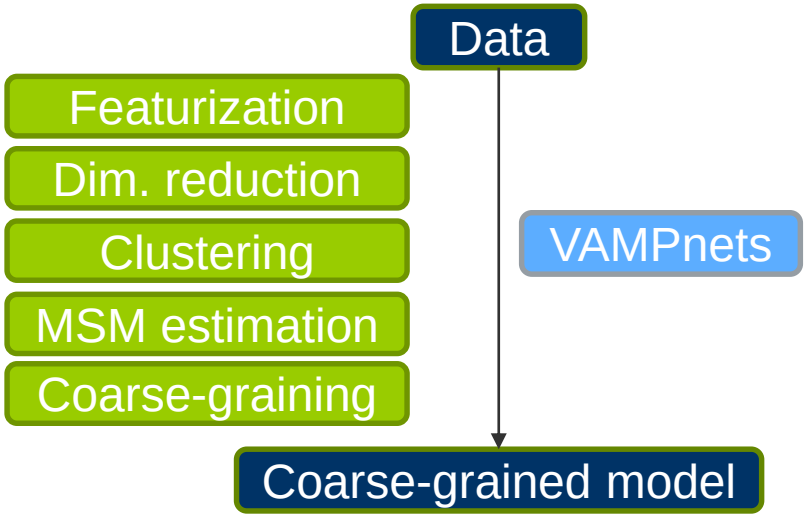
VAMPnets

- NNs are used to implement $\chi_0 = \chi_1 = \chi$
- Trained by maximizing VAMP-2 score

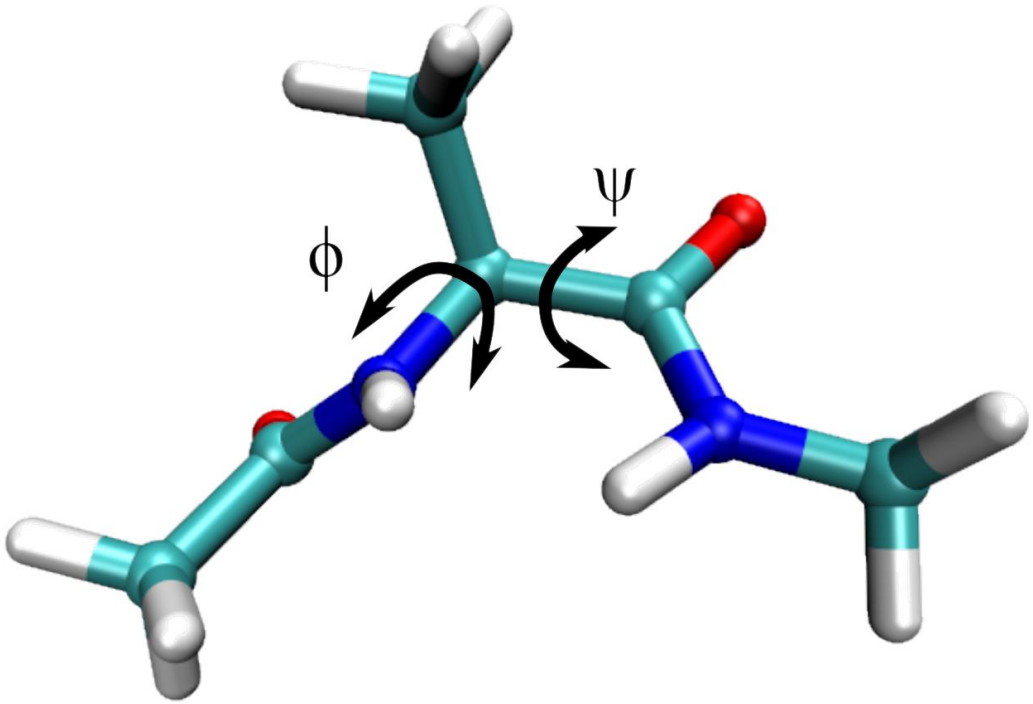


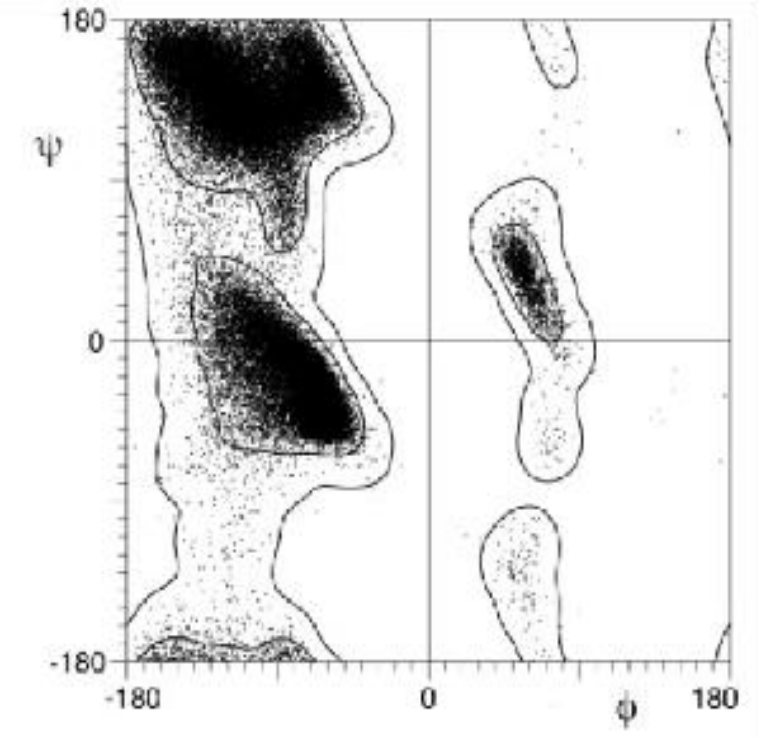
VAMPnets

- Dimensionality reduction obtained through lowering number of output states
- Coarse-graining is implemented through softmax output layers
- Output nodes = number of states, and the output value represents probability of being in that state
- We can calculate the K matrix on the transformed trajectory and test for dynamical processes

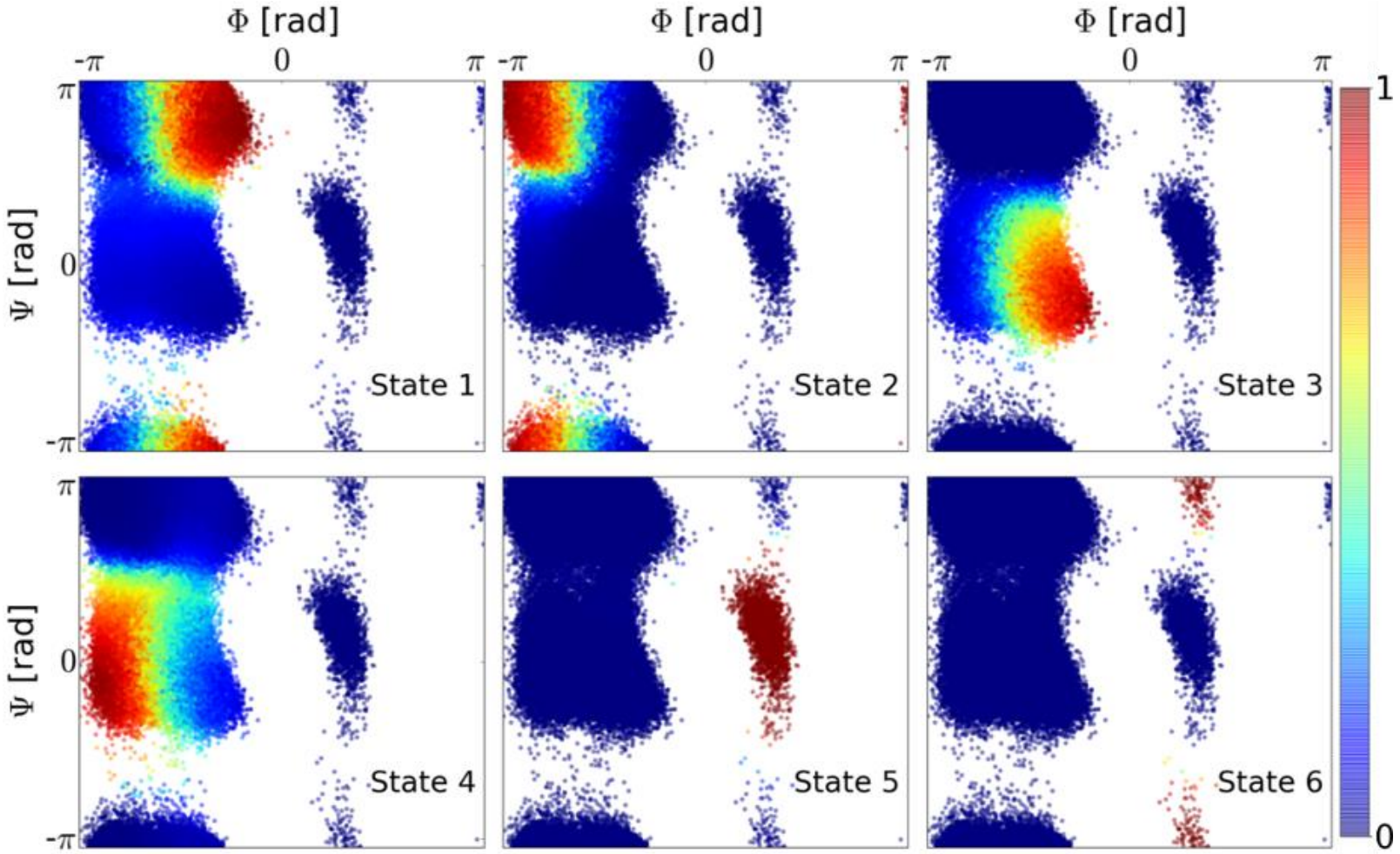


Examples: Alanine Dipeptide

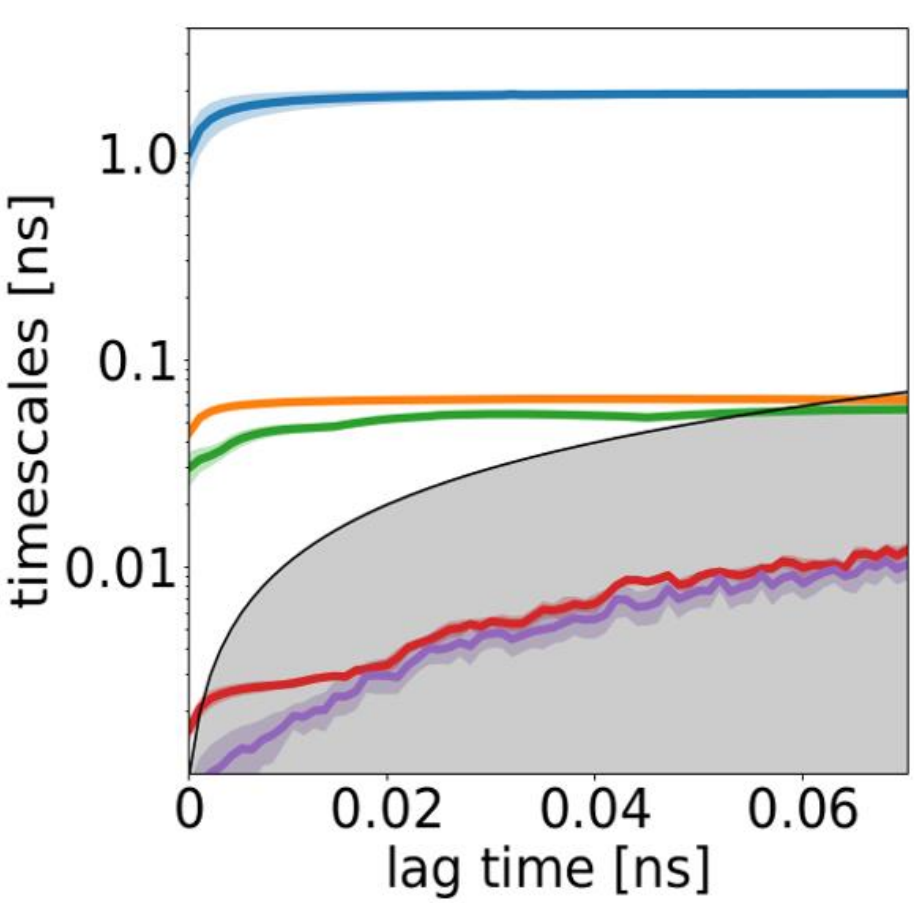


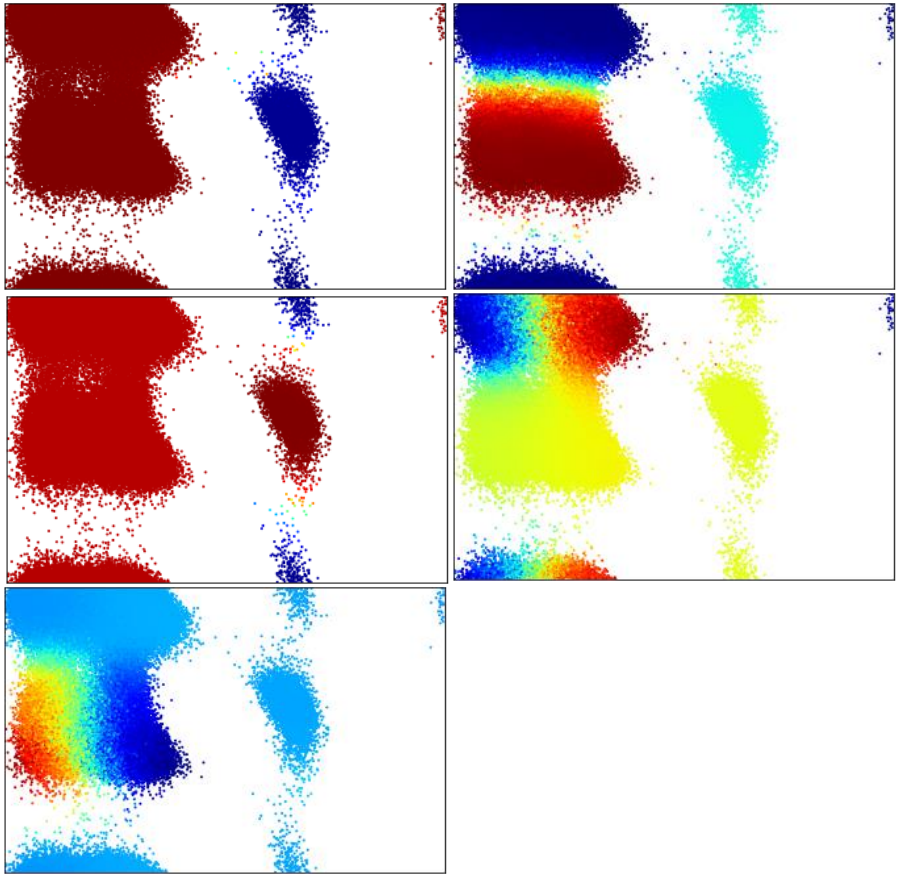


Results: Alanine Dipeptide



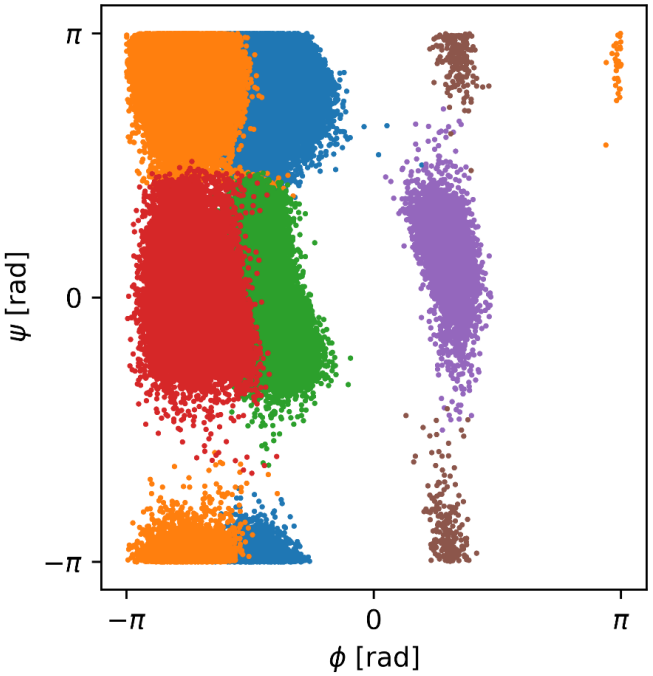
Results: Alanine Dipeptide

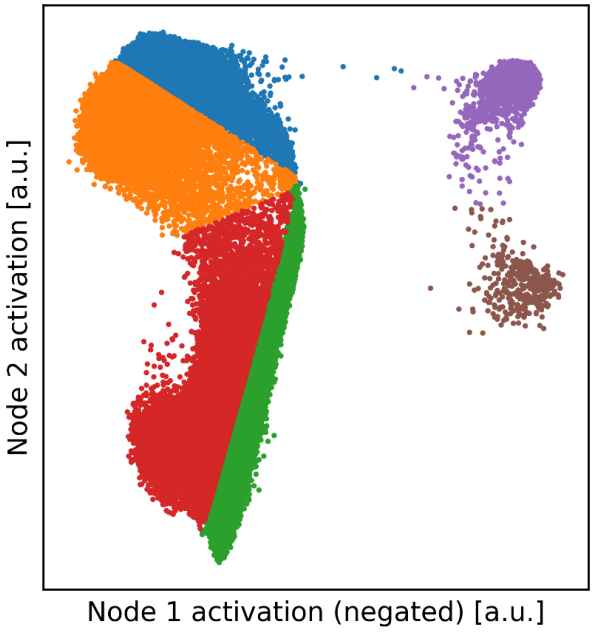




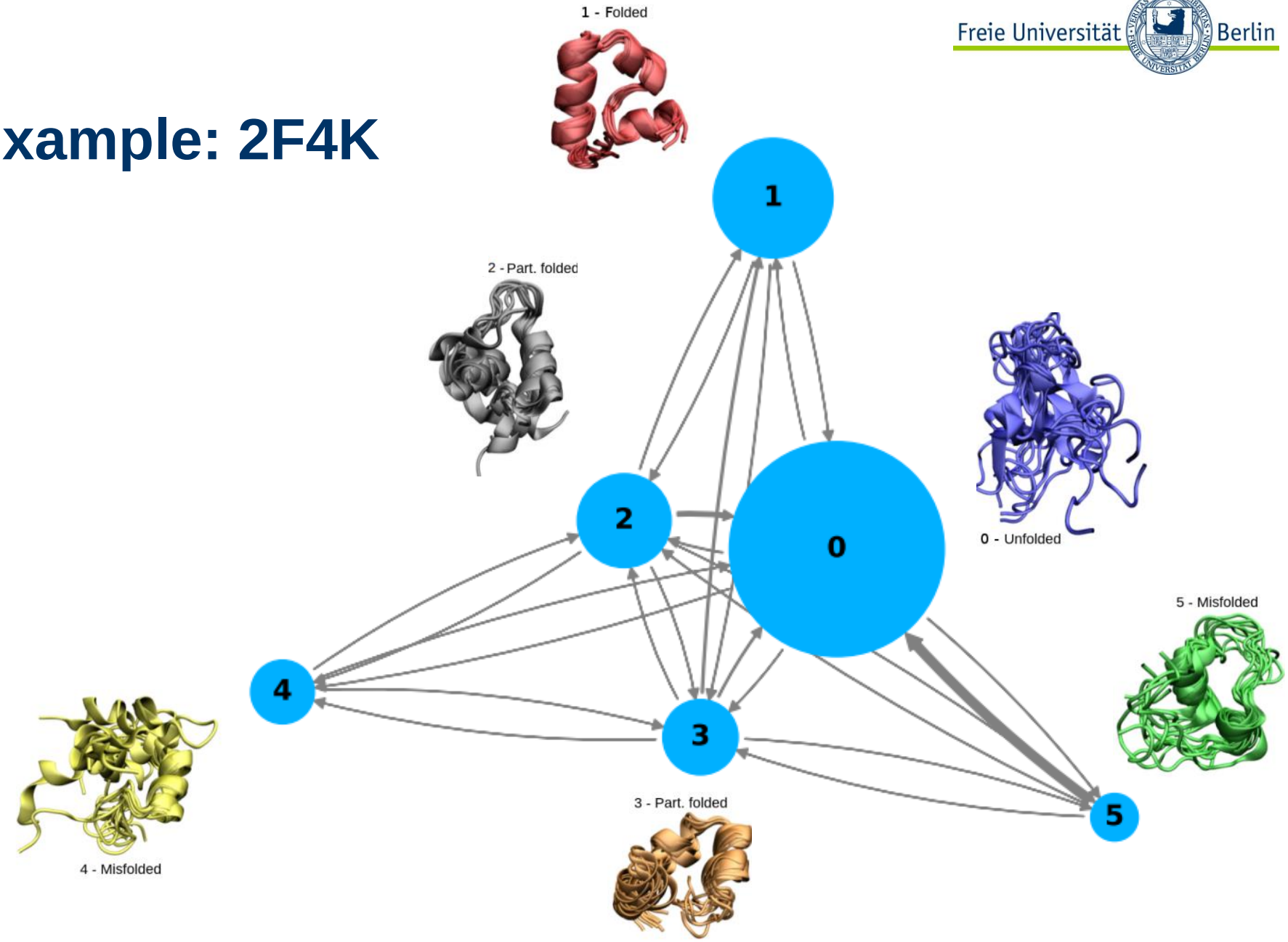
What is the Network Learning?

- VAMPnets with a 2-node bottleneck before the output layer
- Network is forced to learn a 2-dimensional representation of the data





Example: 2F4K



Conclusions

- VAMPnets provides a nonlinear transformation to a space where linear propagation is possible
- Low-dimensional model is easier to interpret
- Unsupervised learning method → low knowledge of system required