EnergyREV Research Fellow Presentation

Colin Stephen
Coventry University

Overview

Recent research	Time series feature engineering for machine learning
	Topology of complex networks for machine learning
Connections to EnergyREV Cyber-Physical Advances	Anomaly and fault detection in time series and networks
	Data pipelines for forecasting and prediction
Research related expertise	Architecture and algorithm design Software development lifecycle best practices

Time Series Feature Engineering for ML

How can we differentiate these deterministic signals using their "global" geometric features?

(Standard frequency and correlation statistics are not sufficient to distinguish them.)

Given:

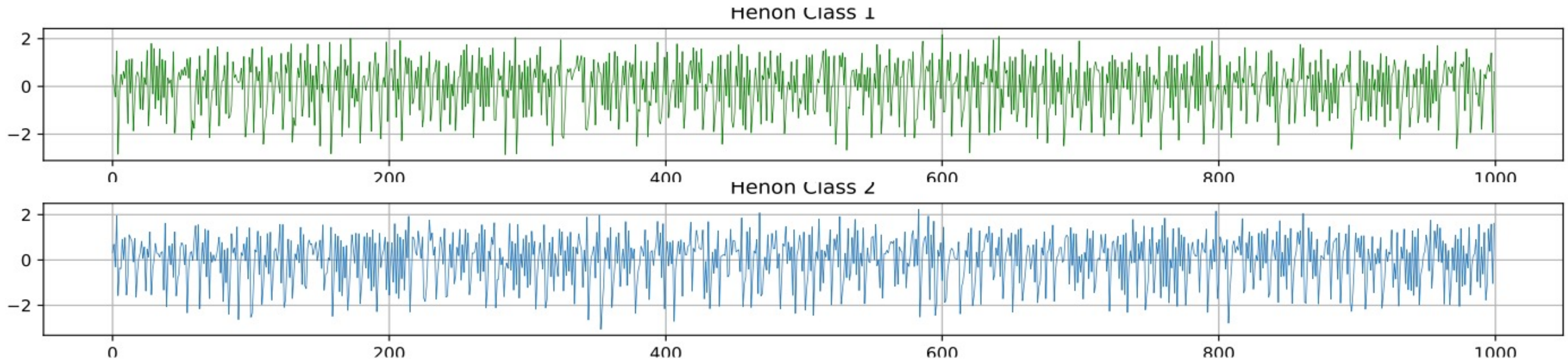
- Two classes of labelled z-normalized time series measured from some chaotic system
- An unlabelled time series from one class

Find:

- A good choice of label for the unlabelled instance

Subject to:

- The underlying dynamic model is unknown
- Signal to noise ratio may be low
- Robust identification needs long time series

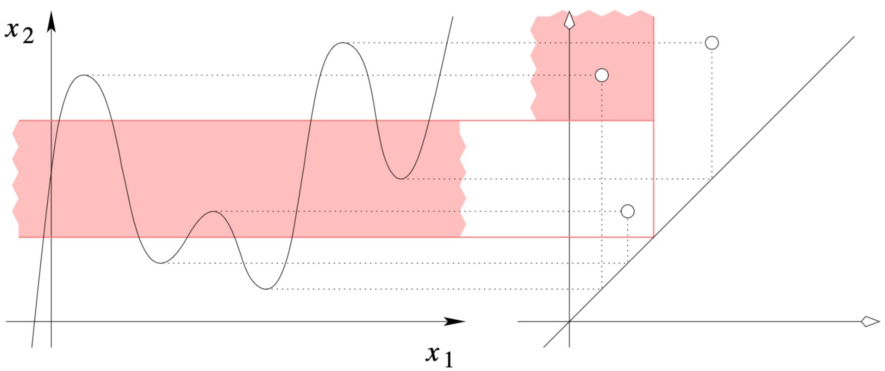


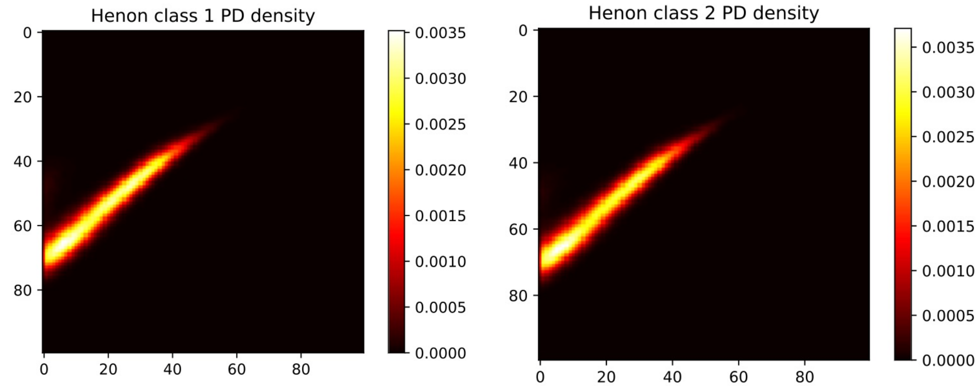
Time Series Feature Engineering for ML

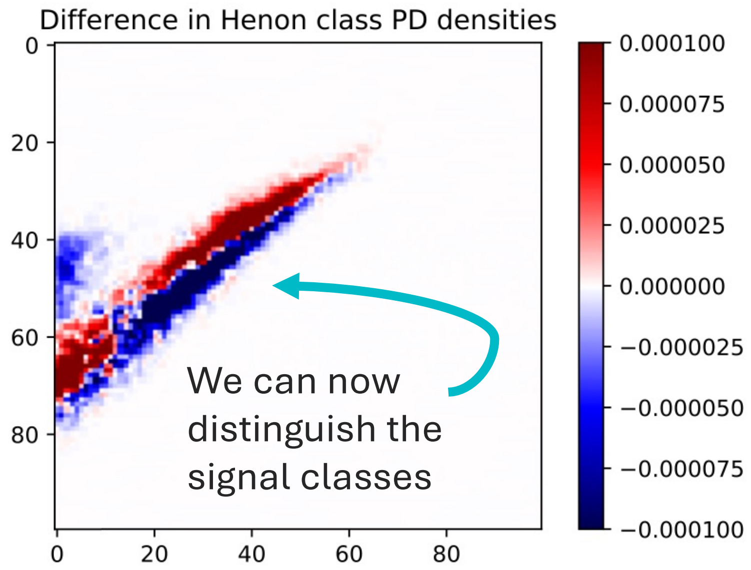
Idea: translate each signal to a multiset of points in the plane, based on relative ordering of its critical points. This approach works in high dimensions (TDA). We show it can be used in 1D (time series).

Benefits: the map can be made invariant to warping and the map can be made stable wrt perturbations

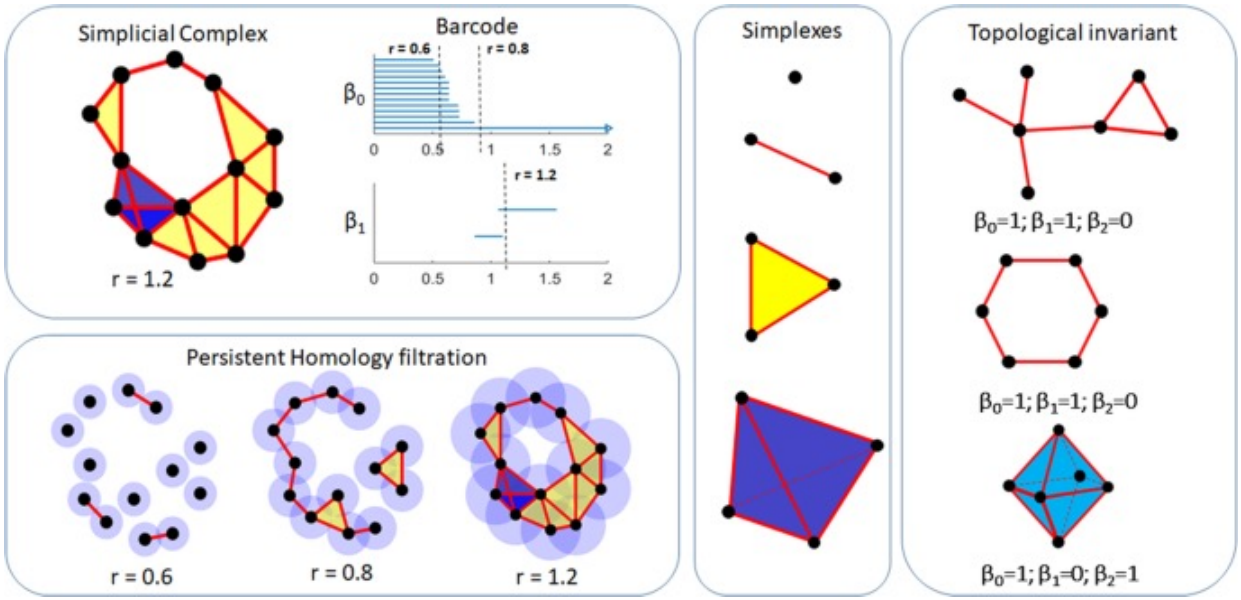
Output: kernel density estimators (KDEs) of in-class persistence densities

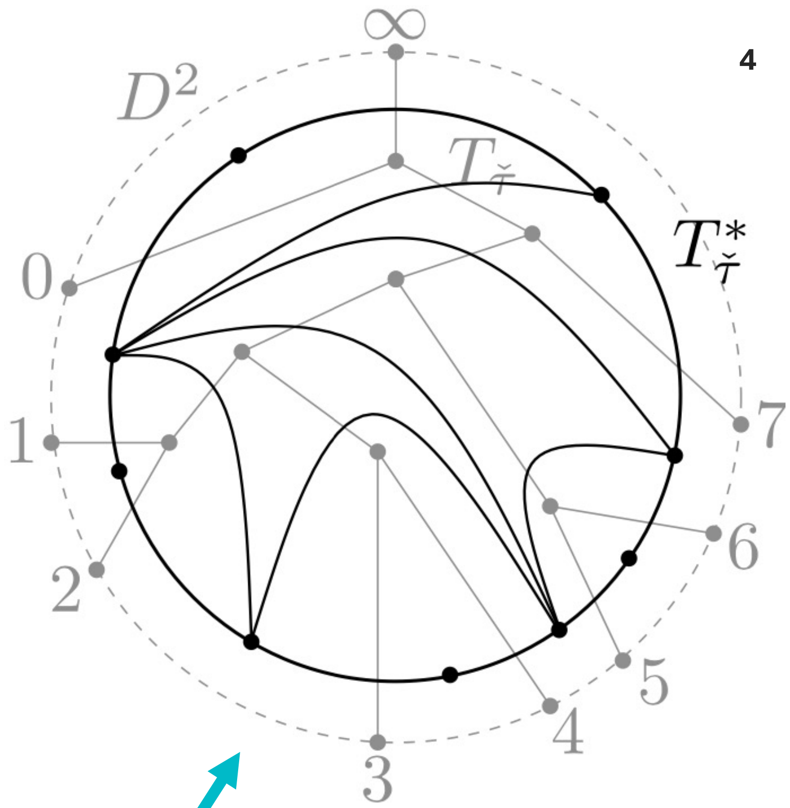






Topology of Complex Networks for ML





2. A persistent homology data pipeline extracts features representing **qualitative** aspects of the network topology *at all scales*

1. Network can represent a **time series** directly or a set of interacting nodes in some **weighted graph / simplicial complex**

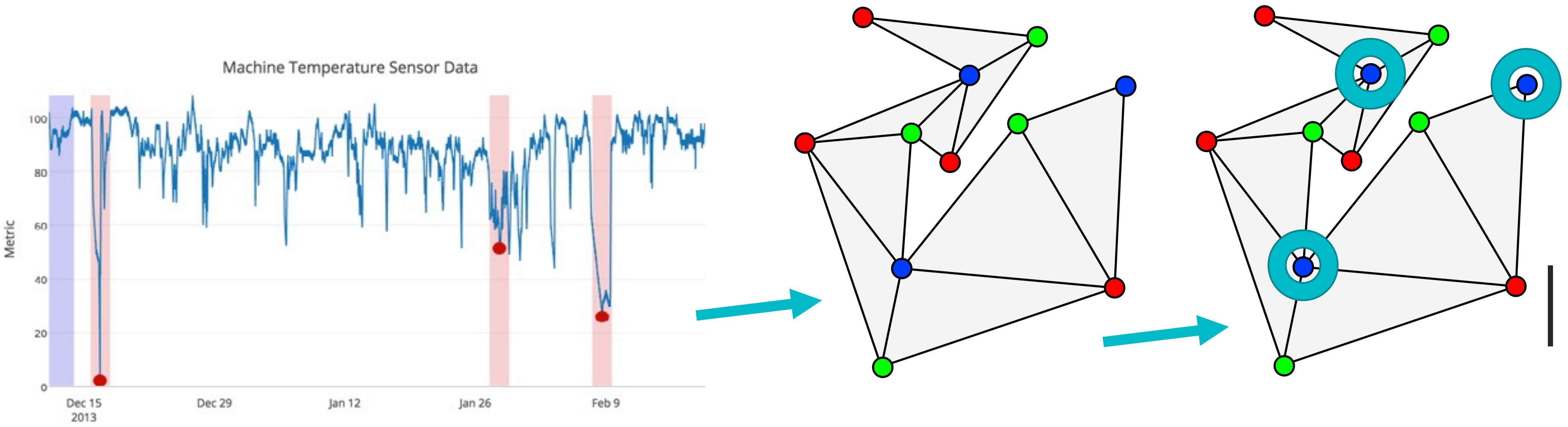
3. The graph topology can thus be **learned** and used in **classifications**/predictions

Anomaly and Fault Detection in Time Series

Time series map to networks whose nodes can be classified using standard graph properties, GNNs, or the previously mentioned filtrations.

This gives new methods of anomaly detection.

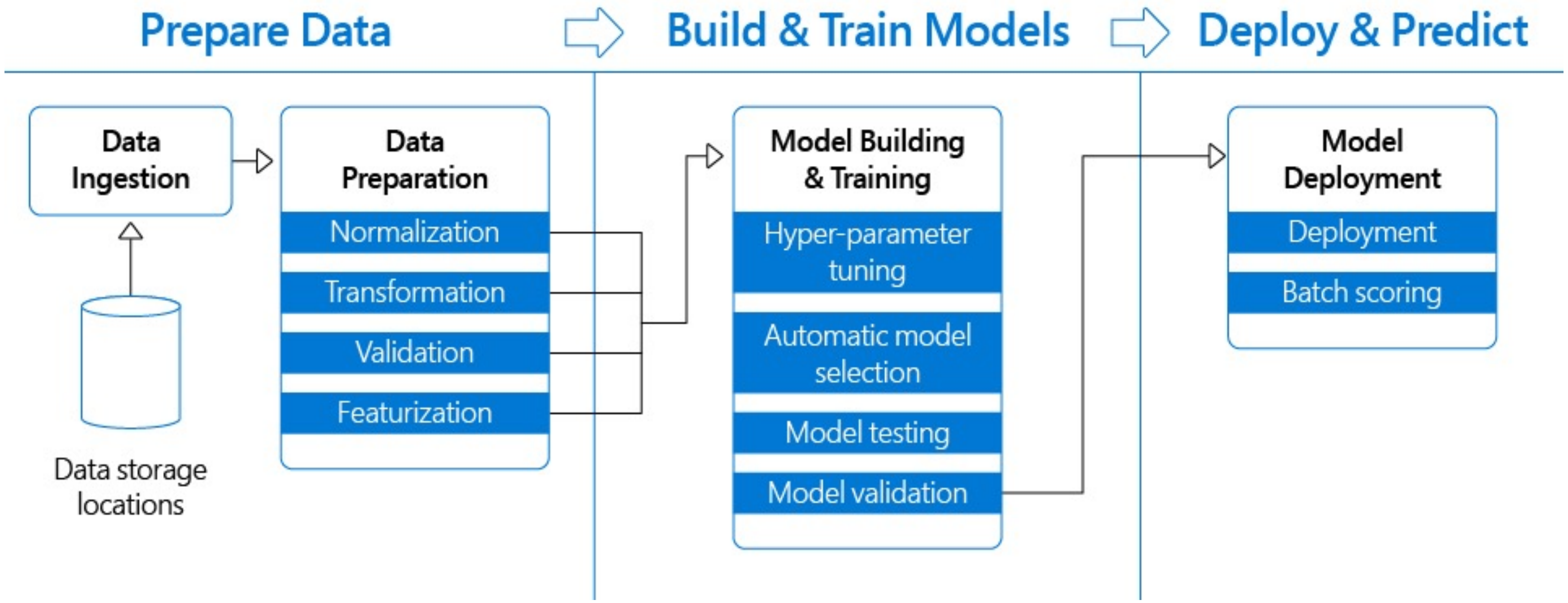
If the TS → Graph representation is a good one, then localized properties of nodes will vary across node classes. Anomalies can be found using these.



Data Pipelines for Forecasting and Prediction

Most of my theoretical research subsequently has been applied using Scikit-Learn and related Python ML libraries and frameworks.

→ Highly familiar with most off-the-shelf models and pipelines for model selection and hyperparameter optimization



Research Related Expertise

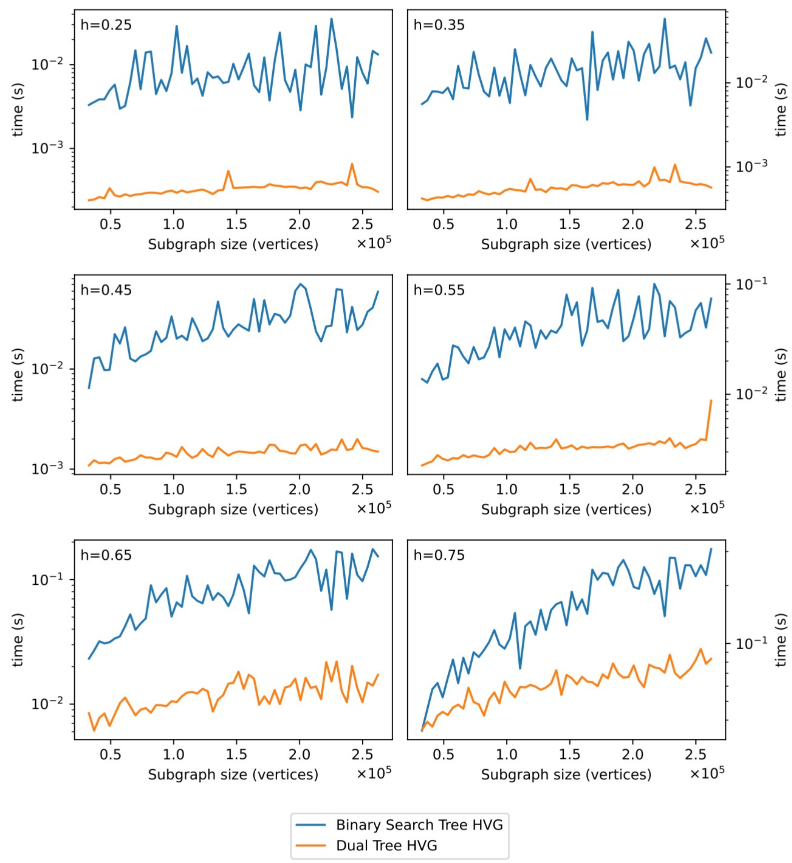
Algorithm design and refinement

- Data processing algorithms need to be well thought-out / chosen for real world use

Software development lifecycle best practices

- documentation, testing
- limit boilerplate, leverage abstractions

Mean merge times, for subgraphs of HVGs containing 1M vertices generated by fractional Brownian motions with varying Hurst index *h*



Thank You

Figure sources:

Stephen, C. (In preparation for J.Phys.A: Mathematical and Theoretical) *The Horizontal Visibility Complex: A Simplicial Complex Representation for Time Series Analysis and Classification*

Stephen, C. (2021, IEEE Big Data - Accepted) *A scalable linear-time algorithm for horizontal visibility graph construction based on a duality with rooted trees*.

Stephen, C. (2019, preprint) *Horizon visibility graphs and time series merge trees are dual*. arXiv:1906.08825 [nlin, physics:physics].

Stephen, C. (2018) *Sinkhorn divergence of topological signature estimates for time series classification*. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 714—721, Orlando, FL, USA.