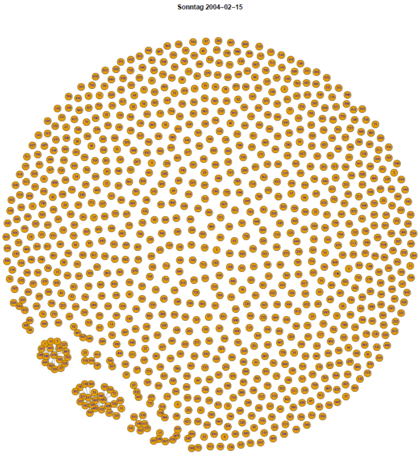
Beyond Nodes: Understanding Static Networks Through Edges and Flows

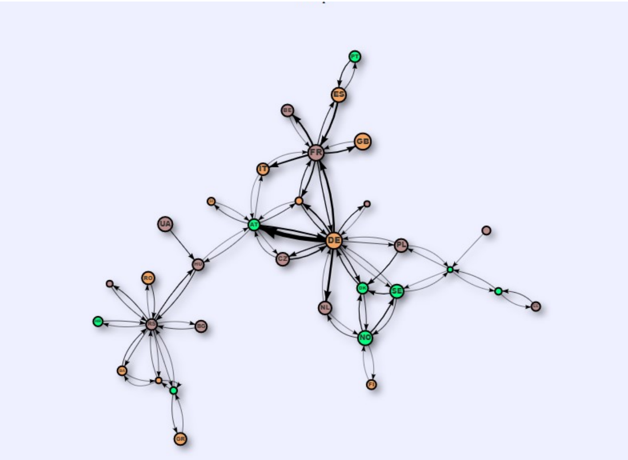
Dr. Anna Malinovskaya
Data Scientist, Lecturer

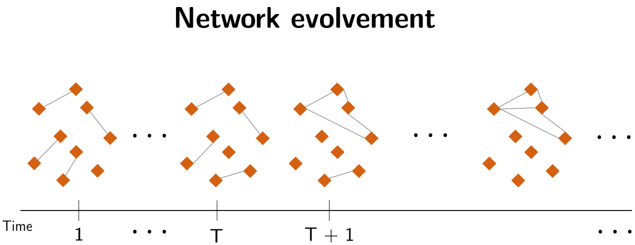
Guest lecture, University of Koblenz

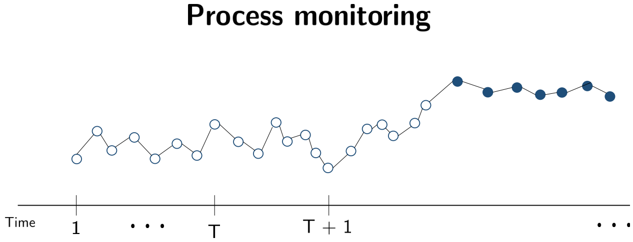
July 2, 2025

What can be anomalous?









- Process monitoring provides the information whether a process evolves optimally.
- Goal: Detection of abnormal behaviour to prevent system failures.
- Network monitoring: The change should be detected based on the changes in network's topology or / and attributed processes.
- **How can we monitor a network which structure doesn't change?**

Questions to answer

1. Why monitoring networks without structural changes is challenging?
2. How to use already developed network models in this novel context?
3. What is a possible use case for temporal edge network processes?

Overview

1. Definition of TEN processes
2. How to model and monitor?
3. Monitoring of cross-border physical electricity flows

Network monitoring

Random network monitoring

- Changing nodes and links;
- Focus of the monitoring: Network itself, *i.e.* network structure;
- Example: Detection of anomalous behaviour in a social network.

Surveillance *of* networks

Fixed network monitoring

- Nodes and links are given;
- Focus of the monitoring: Node- or edge-level variables observed over time;
- Example: Detection of flow deviations in an electricity flow network.

Surveillance *on* networks

- Network = graph

Definition

A graph $G = (V, E)$ consists of a set of vertices V and a set of edges E connecting pairs of vertices.

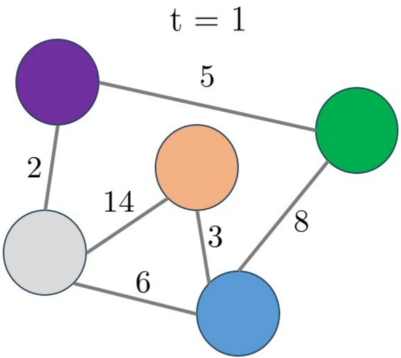
- Two vertices $i, j \in V$ are **adjacent** if they are connected by an edge $e \in E$
- The adjacency matrix $\mathbf{Y} = (Y_{ij})_{i,j=1,\dots,|V|}$ encodes the structure of G

$$Y_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E,\ i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

- $\mathbf{Y}$ is symmetric if G is undirected.

- **Temporal Egde Network (TEN) processes: each edge $e \in E$ has a time series $\{x_{e,t}\}_{t=1}^{T}$, forming a process X on G.**

Illustration of a TEN process for two time points



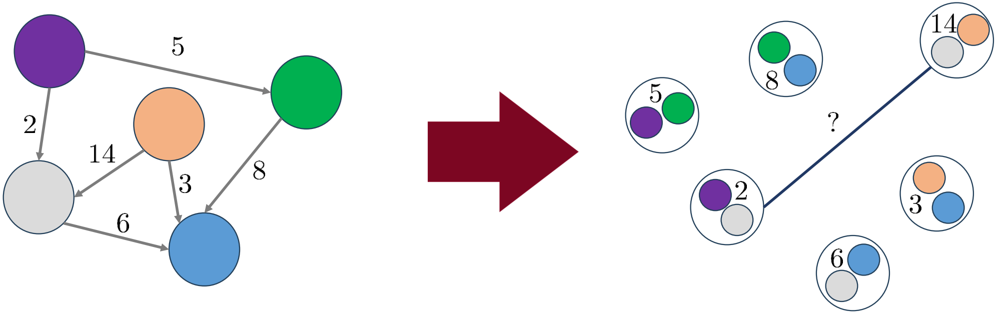
Why is monitoring fixed networks challenging?

- **Popularity of graph augmentation perspective:** Nodal or edge attributes provide contextual information, acting as an extra layer to understand edge likelihood;
- **Field focus:** The majority of existing network models are designed to analyse graph structure, not evolving processes on networks;
- **Inter-flow dependencies:** Flows may be dependent, involving multiple nodes and mechanisms governing exchange dynamics;
- **From nodes to pairs:** The unit of analysis shifts from single nodes to node pairs;
- **Representation shift:** A traditional adjacency matrix is insufficient to model such dependencies accurately.

We need (monitoring) methods that focus on temporal and contextual changes *on* the network, rather than changes *to* the network.

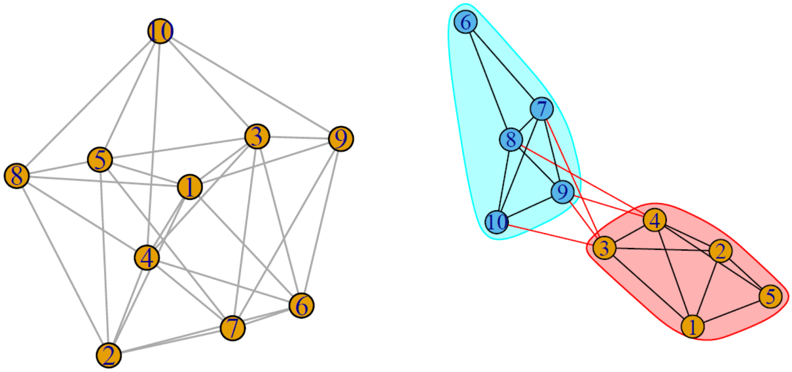
Transforming from G to G^F

...and what about the structure of the connections?



Which options exist to construct $\boldsymbol{Y}$ for G^F?

1. **Random construction**
   - **Erdős–Rényi Model**
   - Each graph with a fixed number of nodes and exactly d edges is equally probable.
   - Suitable when no structural knowledge is available.
   - **Stochastic Block Model (SBM)**
   - Generates community structures (e.g. intra-community edges more likely).
   - Effective when communities are known or nodes can be meaningfully clustered.
2. **Deterministic construction**
   - **Line Graph** $L(G)$
   - Nodes represent edges of G; edges connect if the original edges share a node.
   - Provides a logical structure when random models are unsuitable.
   - Preserves local connectivity constraints without prior domain knowledge.

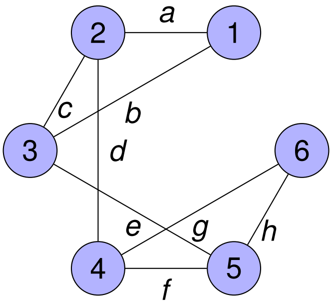


Representation of connectivity types using graph random models.

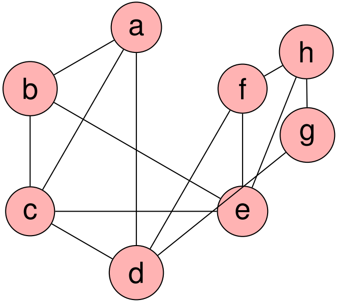
Exploring line graph to construct $\boldsymbol{Y}$ for G^F

Line graph or edge-to-vertex dual graph

A line graph $L(G)$ of a simple graph G is obtained by associating a vertex with each edge of the graph and connecting two vertices with an edge iff the corresponding edges of G have a vertex in common (Gross and Yellen 2006, p. 20).

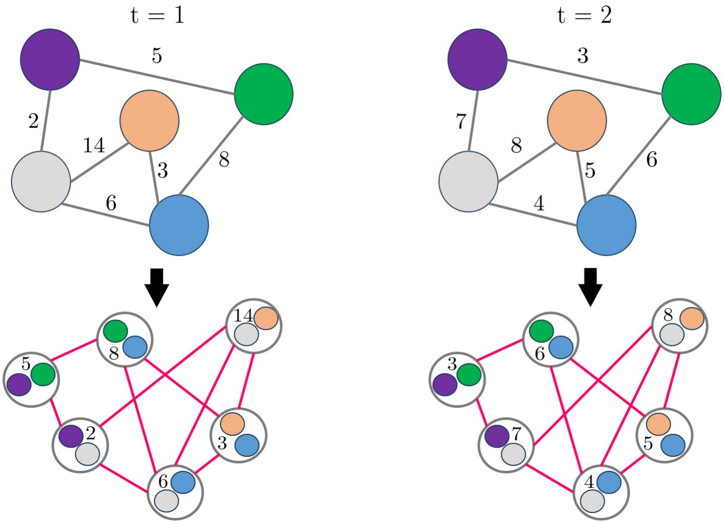


(a) Original Graph G



(b) Line Graph (Vertex to Edge) G^F

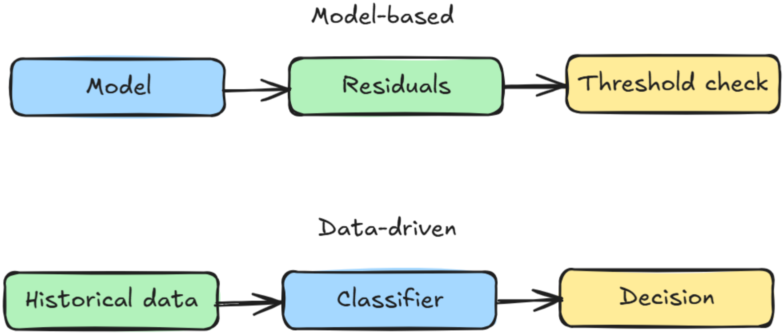
TEN process for two time points and with the new representation



- **Statistical network monitoring** = (1) Network modelling + (2) Process Monitoring
- We decide for the **model-based** monitoring approach

1. Select a suitable model for the novel representation of TEN processes
    - Models for processes on (stochastic temporal) networks
    - Random walks on stochastic temporal networks
    - Non-Poissonian processes on networks
    - Volatility models on networks
    - Generalized Network Autoregressive (GNAR) models
2. Define what and how to monitor
    - Compare results obtained by using a model to the actual observations
    - Analyse deviance residuals
    - Suitable technique – control charts

Model-based vs. data-driven process monitoring



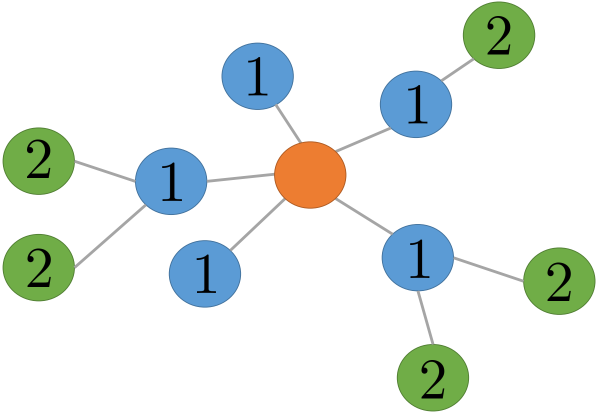
How to model?

A recently developed extension of GNAR models called GNARX model incorporates exogenous time-dependent node-specific regressors.

- Consider multivariate time series $\{X_{i,t}\}$ be observations collected on variable $i \in V = \{1, \ldots, N\}$ at time $t = 1, \ldots, T \in \mathbb{N}$, $N \in \mathbb{N}$ is the number of nodes.
- Let $\{Z_{h,i,t}\}$ be the h-th stationary exogenous node-specific regressor series.
- The GNARX $(p, \boldsymbol{s}, \boldsymbol{p}')$ model, where $(p, \boldsymbol{s}, \boldsymbol{p}') \in \mathbb{N} \times \mathbb{N}_0^p \times \mathbb{N}_0^H$ is specified as

$$X_{i,t} = \sum_{l=1}^{p} \left(\alpha_{i,l} X_{i,t-l} + \sum_{r=1}^{s_l} \beta_{l,r} \sum_{q \in \mathcal{N}^{(r)}(i)} \omega_{i,q} X_{q,t-l} \right) + \sum_{h=1}^{H} \sum_{l'=0}^{p'_h} \lambda_{h,l'} Z_{h,i,t-l'} + \epsilon_{i,t},$$

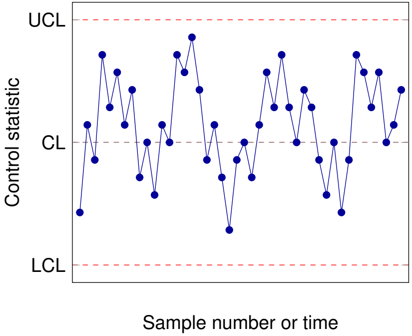
- the parameters $\alpha_{i,l},\ \beta_{l,r},\ \lambda_{h,l'} \in \mathbb{R}$ for all $i,\ l,\ r,\ h,\ l'$,
- p is the autoregressive order of the model and maximum order of neighbour time lags,
- p'_h is the maximum lag of the hth exogenous regressor,
- s_l the maximum stage of neighbour dependence for time lag l.



Stage-1 (blue) and stage-2 (green) neighbourhood of the orange node.

How to monitor?

- Statistical Process Monitoring (SPM) – form of online surveillance to detect a change point when a process starts deviating from its target state.
- SPM tools are applied for achievement of the process stability. Key technique – control chart.
- Task of the control chart: to detect as quickly as possible the occurrence of unusual variation in the process.
- **Which control chart to choose?**



$H_{0,t}$: The observed TEN process coincides with the fitted GNARX model.

$H_{1,t}$: The observed TEN process does not coincide with the fitted GNARX model.

Requirements and solutions

- Changes may affect the mean and/or variance $\rightarrow$ joint monitoring is possible when using residuals.
- In realistic monitoring settings, we might have hundreds of flows $\rightarrow$ employ parallel univariate control charts to detect *local changes* in individual flows.
- We are still interested in detecting anomalies on a network level $\rightarrow$ local changes per flow can be aggregated to identify *global changes*.

Step 1: Local change detection with CUSUM

- We use a generalised Page's CUSUM based on centred squared deviance residuals $u_{\iota,t}$, where ι indicates the novel representation.
- The test statistic from Grundy (2025) is

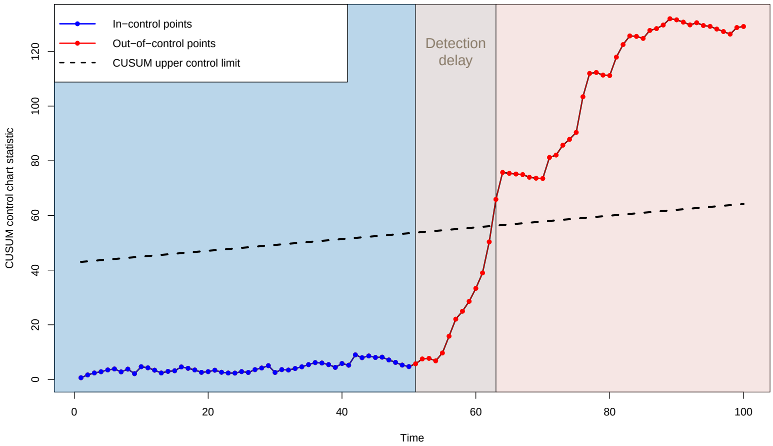
$$Q_\iota(m,k) = \sum_{t=m+1}^{m+k} (u_{\iota,t} - \hat{b})^2 - \frac{k}{m} \sum_{t=1}^{m} (u_{\iota,t} - \hat{b})$$

  - m = length of Phase I
  - k = the current time point in Phase II
  - $\hat{b}$ = the mean estimate of the deviance residuals computed in Phase I
- The control chart statistic is

$$D_\iota(m,k) = \max_{0 \le a \le k} |Q_\iota(m,k) - Q_\iota(m,a)|$$

- A local change is detected when $D_\iota(m,k) > UCL$.

Change in $\alpha + 0.3$ (Original $\alpha = 0.2$)



Simulation study: example of monitoring one specific flow from a TEN process.

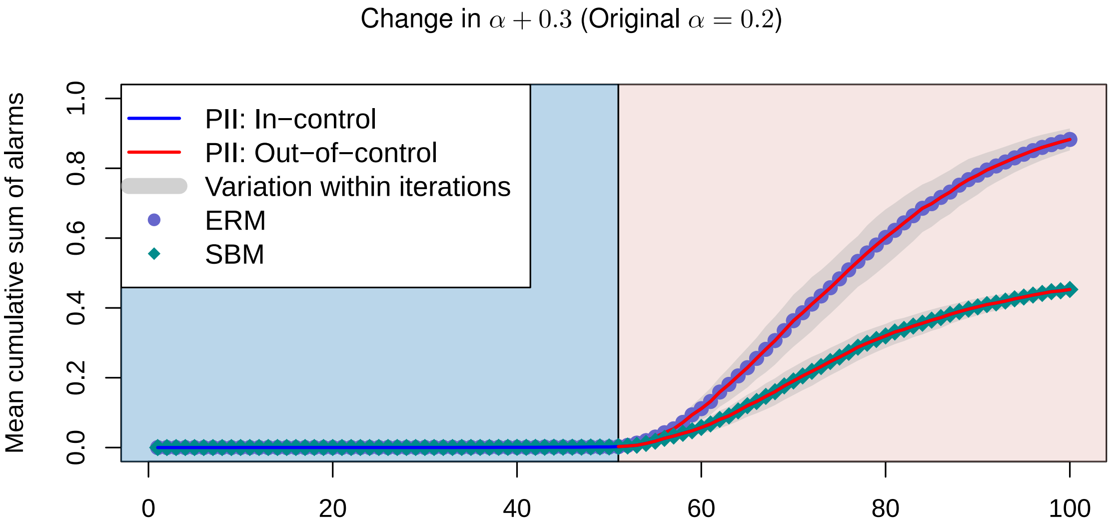
Step 2: Aggregating local signals into global change

- Create a **threshold-based rule**, setting a minimum share of flows that must signal a change.
- Implement $n = |V^F|$ control charts.
- Define a cumulative change intensity function – percentage of edges with alarms as

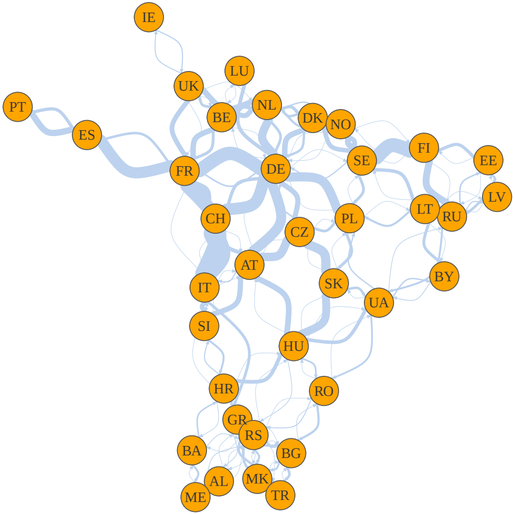
$$I_{\mathbf{X}}(T) = \sum_{k=1}^{T} \frac{\sum_{\iota=1}^{n} \mathbf{1}_{[UCL,\infty)}[D_\iota(m,k)]}{n}$$

- If in one flow ι a change was detected, the monitoring of this flow stops, resulting in $\max(I) = 1$.
- To determine when a global alarm should occur, define threshold W.

Simulation study with 500 iterations: Effect of increasing α



Monitoring of cross-border physical electricity flows



- Physical flow is defined as the measured real flow of electricity between neighbouring countries on the cross borders.
- Weekly aggregation of observations.
- Phase I: Years 01.01.2018 – 29.12.2019; Phase II: 30.12.2019 – 27.11.2022.
- Covariates: Amounts of electricity generated using fossils and renewable sources.
- Proof of concept: Observe changes during the periods 15.03.2020 – 30.05.2021 and since 01.02.2022.

New representation of cross-border physical electricity flows

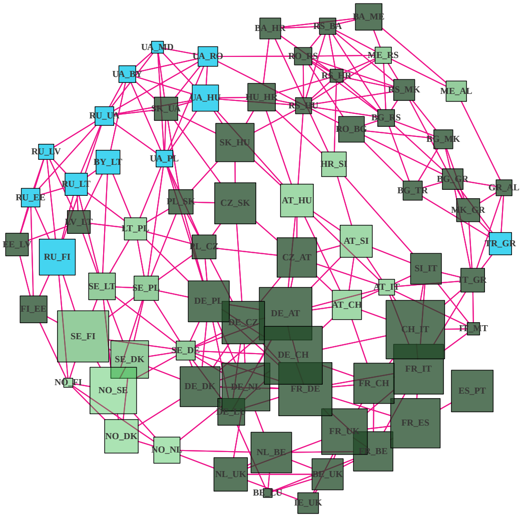
- Aggregate parallel flows f_1 and f_2 to avoid network expansion (three strategies):

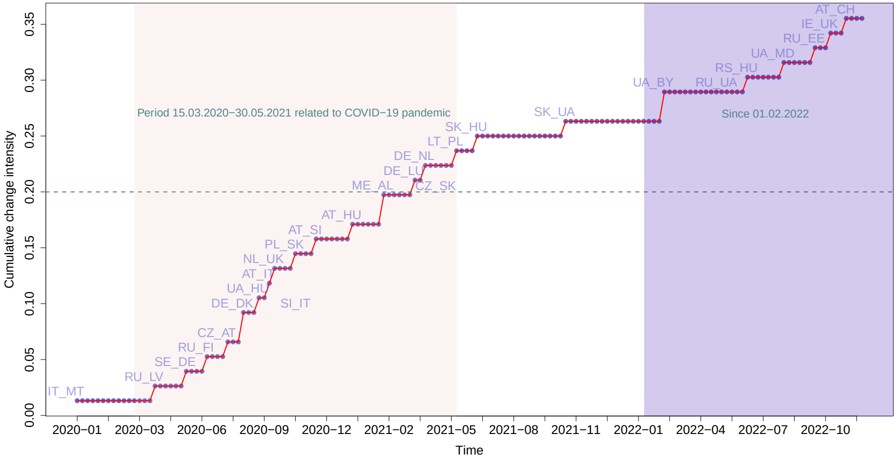
  1. $$\mathscr{M}_1 = \ln(f_1 + f_2 + 1)$$

  2. $$\mathscr{M}_2 = \ln(f_1 + 1) - \ln(f_2 + 1)$$

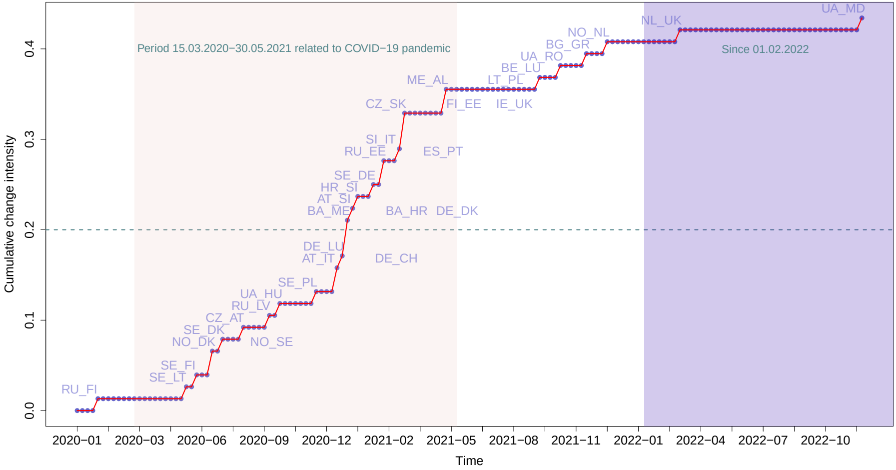
  3. $$\mathscr{M}_3 = \frac{f_1 - f_2}{f_1 + f_2}$$

- The size of the nodes indicates the strength of the electricity exchange and the colour its proportion of renewable energy sources.

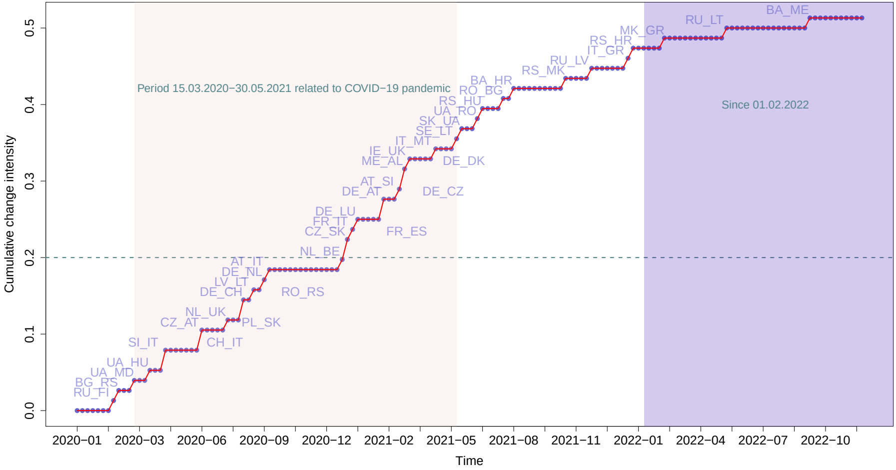




Results of monitoring $\mathscr{M}_1$, obtaining in total 27 change points.



Results of monitoring $\mathscr{M}_2$, obtaining in total 33 change points.



Results of monitoring $\mathscr{M}_3$, obtaining in total 39 change points.

Summary / open questions

- Concept for monitoring and detecting anomalies in TENs by combining the GNARX model and the CUSUM control chart based on deviance residuals.
- Current assumption – TENs observed at discrete times $\rightarrow$ How and when to extend the monitoring to continuous time stamps?
- Another assumption – fixed structure. When different adjacency matrices should be introduced (e.g. when some nodes or flows disappear)?
- We assume GNARX model is suitable for considered empirical data – any alternatives?
- Is there a suitable multivariate monitoring procedure that can omit intensity function?

GNAR(X) Model

Knight, Marina, Kathryn Leeming, Guy Nason, and Matthew Nunes. (2020). *Generalized Network Autoregressive Processes and the GNAR Package.* Journal of Statistical Software 96 (5):1-36.

Nason, Guy P., and James L. Wei. (2022). *Quantifying the economic response to COVID-19 mitigations and death rates via forecasting purchasing managers' indices using generalised network autoregressive models with exogenous variables.* Journal of the Royal Statistical Society Series A: Statistics in Society 185 (4):1778-1792.

Network analysis and statistical process monitoring

Grundy, T., Killick, R., and Svetunkov, I. (2025). Online detection of forecast model inadequacies using forecast errors. *arXiv preprint* arXiv:2502.14173.

Malinovskaya A, Otto P, Peters T (2022) Statistical learning for change point and anomaly detection in graphs. In Artificial Intelligence, Big Data and Data Science in Statistics: Challenges and Solutions in Environmetrics, the Natural Sciences and Technology

Malinovskaya A, Otto P (2021) Online network monitoring. Statistical Methods & Applications 30(5)

Thank you for your attention!