Workshop Results - Alarm Pattern Recognition

IPSW 2023 - Hitachi Problem 2

2023-08-24

Hitachi Energy

Contributors in the team

Students: Piotr Morawiecki
Judith Tavarez
Benjamin Nguyen
Rohan Singh
Nicholas Goulet

Faculty: Moncef Chioua
Slim Ibrahim

Industry: Antony Hilliard

Introduction – Alarms in Power Systems

Context

- Alarms play an important role in detecting and mitigating abnormal situations.
- Well configured alarms = Facilitate operations
- Poorly configured alarms = Increase operator workload

Problem

- Alarm avalanches:
  - Too many alarms too fast for operator.
  - Many lines of text to read
- Alarm origin:
  - It is difficult to determine which alarm is the originating alarm.

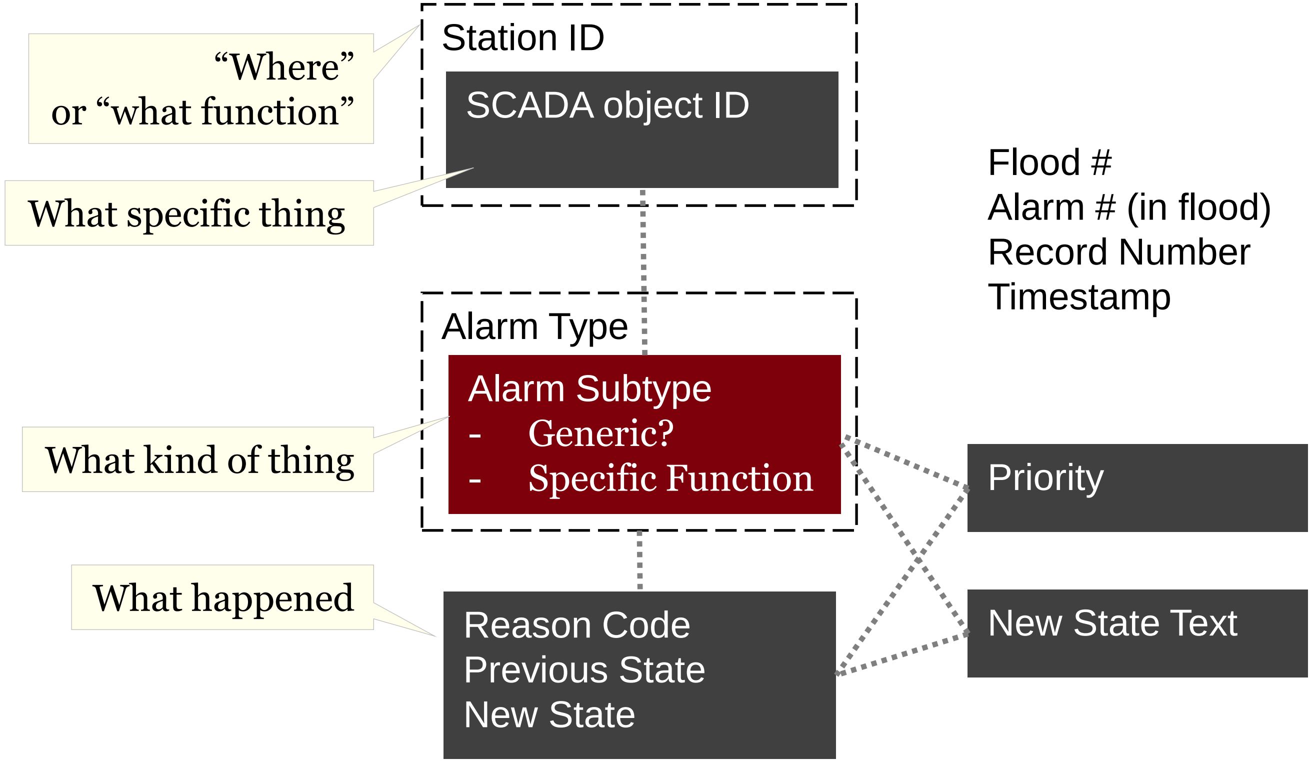
Introduction - Dataset

HITACHI
Inspire the Next

Periods of rapid alarms from a real power system

- 5680 examples of “floods” or “avalanches”
  - January 2022 – August 2022
  - ~2000 with uninformative data
- Uninformative data includes:
  - “Chattering” alarms
  - Communication loss alarms
- 729 floods remaining after pre-processing for Approach 1

Fields that were used in these approaches



Hitachi Energy

Approach 1:

Clustering by Jaccard distance

Objective and Solution Concept

Create a tool that classifies alarm floods for operator decision support

Pipeline of an implementation:

1. Alarm flood data filtering →
2. Clustering to create alarm flood fault classes (Unsupervised) →
3. Training classification model based on clusters (Supervised) →
4. Deploy model online as operator decision support tool

Method: Clustering Steps

Methodology based on Ahmed et al. 2013, “Similarity analysis of industrial alarm flood data”:

1. Define what a unique alarm is to base the clustering

- Equipment sub-type + new state
- Cluster alarm floods that affect same equipment in the same way

Alarm Subtype
- Voltage level
- Switch position

New State Text
- “Into LOW zone”
- “Into HIGH zone”

2. Calculate similarity between every pair of alarm flood

- Jaccard distance

3. Cluster the most similar floods and interpret each cluster

- Hierarchical Clustering

Theory: Determining Similarity

Similarity Metric: Jaccard distance

- Measures the level of **dissimilarity** between two alarm flood sets:

$$J_{S_1, S_2} = \frac{b + c}{a + b + c}$$

$$a: \#\ of\ the\ same\ alarms\ in\ both\ S_1 and\ S_2$$
$$b: \#\ of\ alarms\ that\ are\ in\ S_1\ but\ not\ S_2$$
$$c: \#\ of\ alarms\ that\ are\ in\ S_2 but\ not\ S_1$$

Example:

S_1	S_2	S_3
A_1	A_2	A_5
A_2	A_3	A_6
A_3	A_4	A_7

S_1 vs. S_2

$$a = 2\ (A_2, A_3) \quad b = 1\ (A_1) \quad c = 1\ (A_4)$$

$$J_{i,j} = \frac{1 + 1}{2 + 1 + 1} = 0.5$$

S_1 vs. S_3

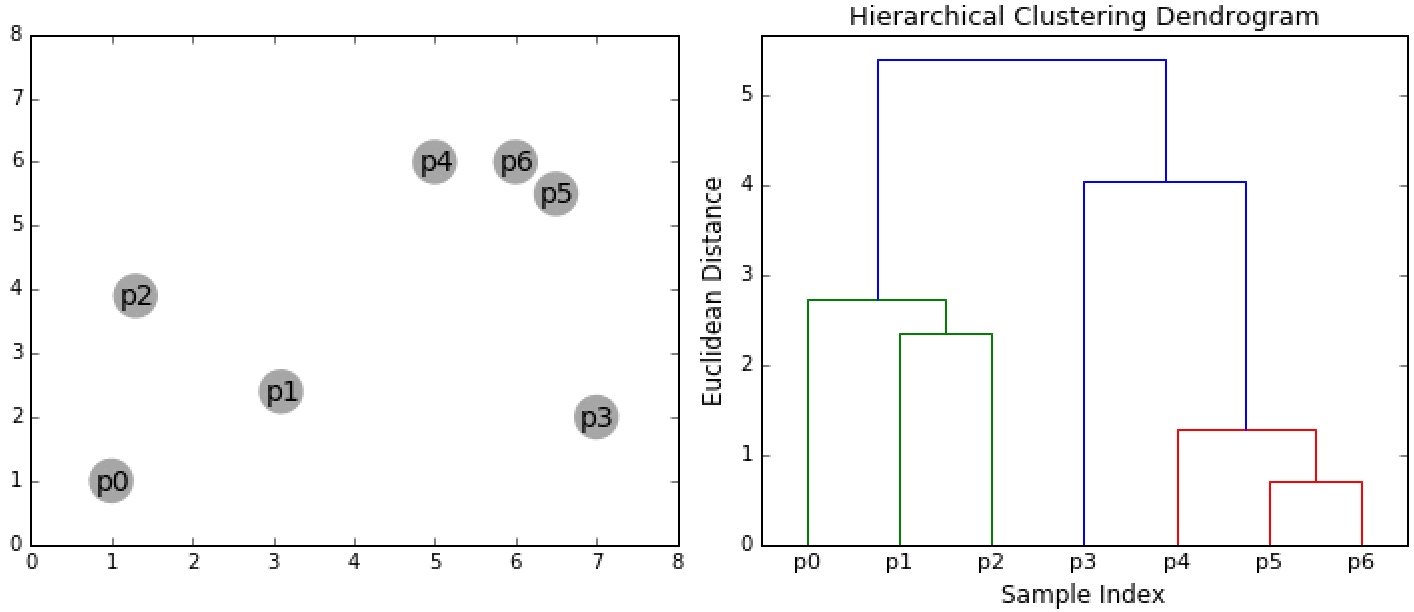
$$a = 0 \quad b = 3 \quad c = 3$$

$$J_{i,k} = \frac{3 + 3}{0 + 3 + 3} = 1$$

Theory: Hierarchical Clustering Algorithm

Clusters from bottom to top (agglomerative)

- Each flood is a cluster → calculate pairwise similarity → cluster pair with highest similarity → re-calculate similarities → repeat until one cluster exists
- Used “sklearn.cluster.AgglomerativeClustering”



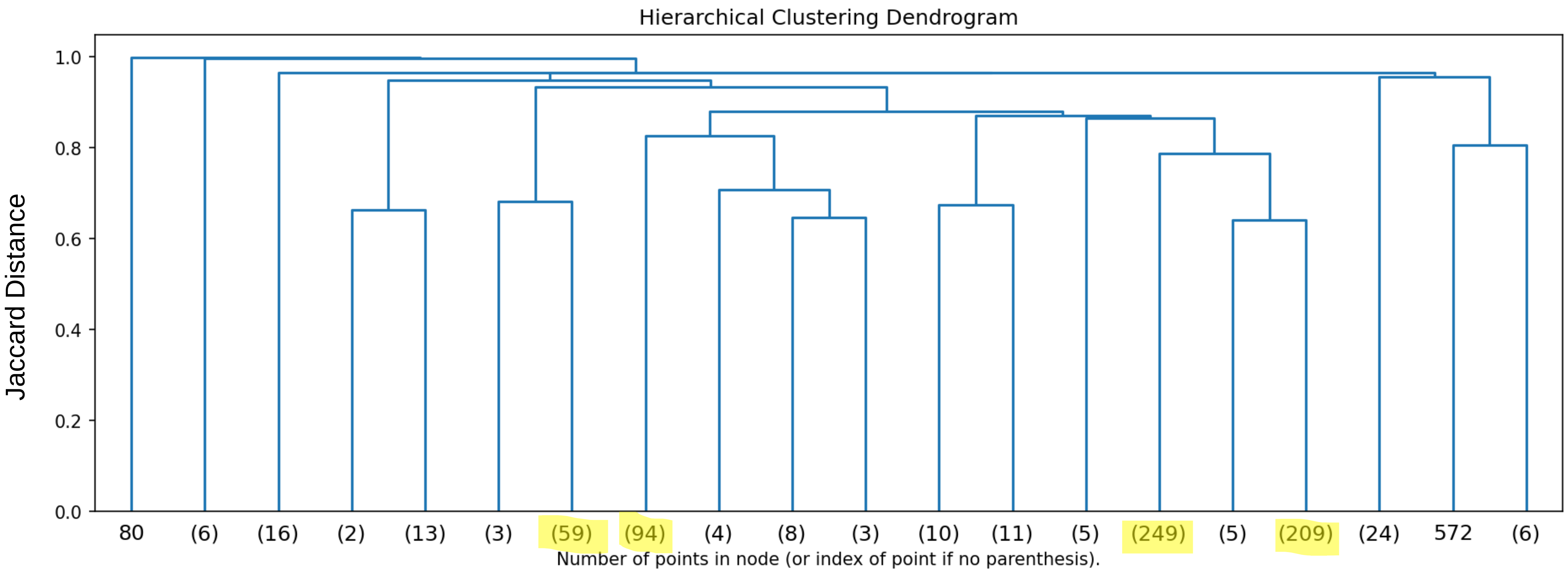
Taken from Sheehan 2017:
https://dashee87.github.io/data%20science/general/Clustering-with-Scikit-with-GIFs/

Hitachi Energy

Results: Dendrogram

HITACHI
Inspire the Next

Can identify 4 major clusters + a few smaller clusters
Dendrogram shows which clusters are more related (part of same branch)



Cluster number: 12 17 9 18 3 7 **1** **0** 14 8 4 16 5 10 **2** 19 **6** 13 15 11

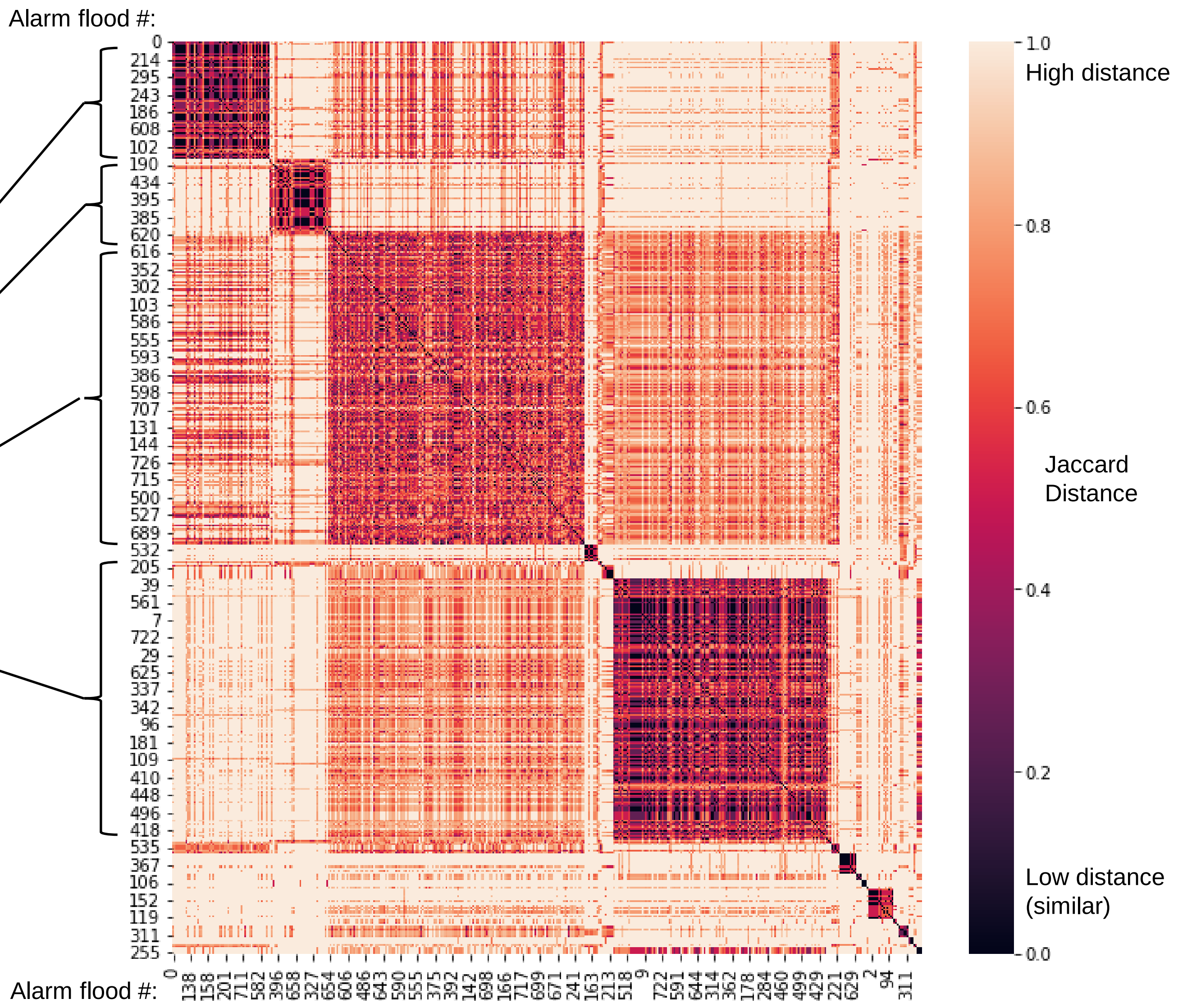
Hitachi Energy

Results: Clustered distance matrix

Physical Meaning	Cluster Number
Low voltage in multiple stations	**0**
High voltage in multiple stations	**1**
Manual grid configuration	**2**
Automatic grid configuration	**6**

Alarm Subtype
- Voltage level
- Switch position

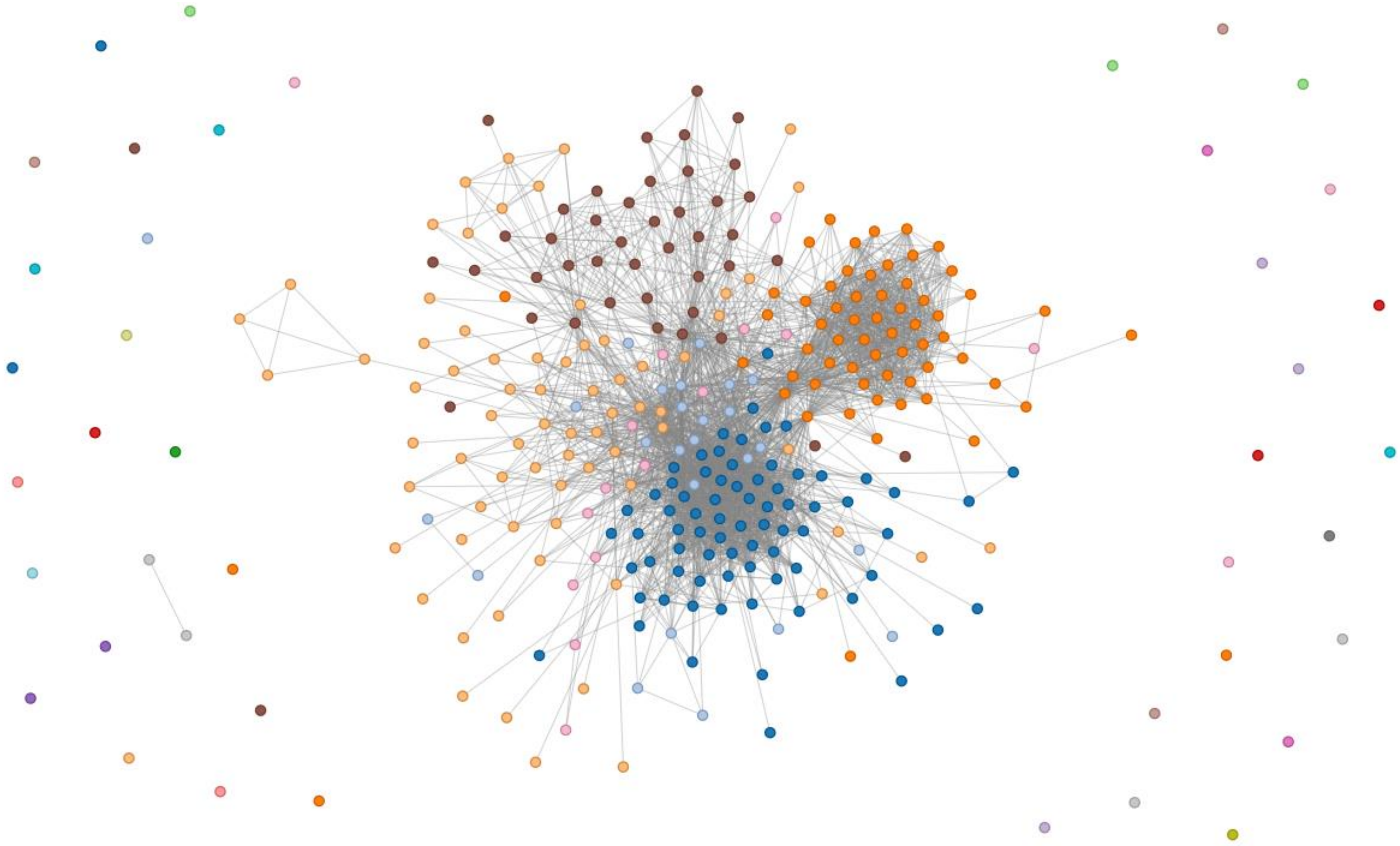
New State Text
- "Into LOW zone"
- "Into HIGH zone"



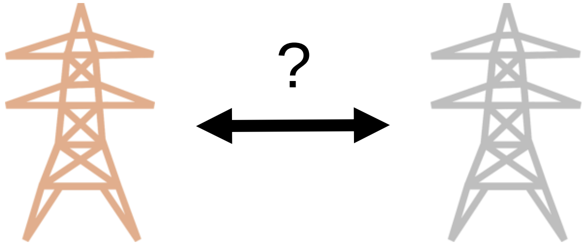
Approach 2:

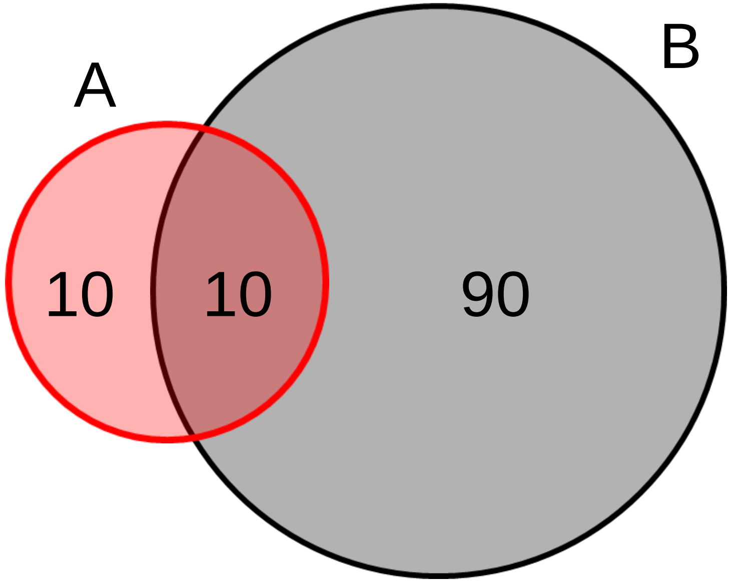
Maximum likelihood flood root identification

Correlation of alarms by network location (stations)



Towards causality: a (de)motivating toy example





We have conditional probabilities:

$$\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)} = 0.5 \qquad \Pr(A|B) = \frac{\Pr(A \wedge B)}{\Pr(B)} = 0.1$$

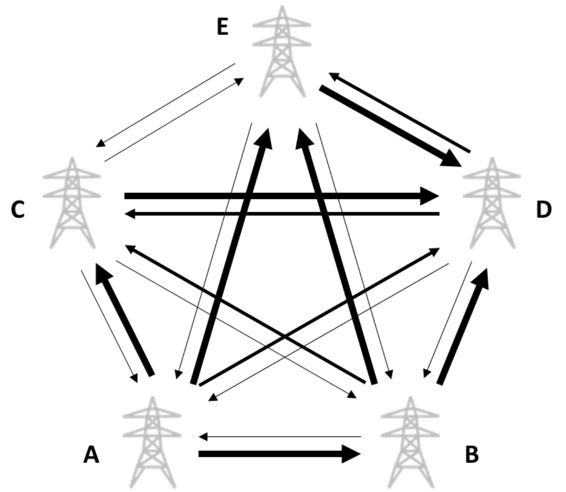
There are two possible interpretations:
- Alarm A triggers alarm B with probability 50%,
- Alarm B triggers alarm A with probability 10%.

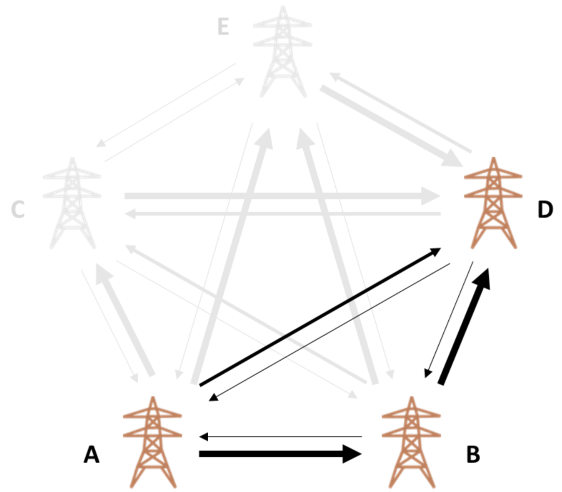
...or anything in between these two scenarios.

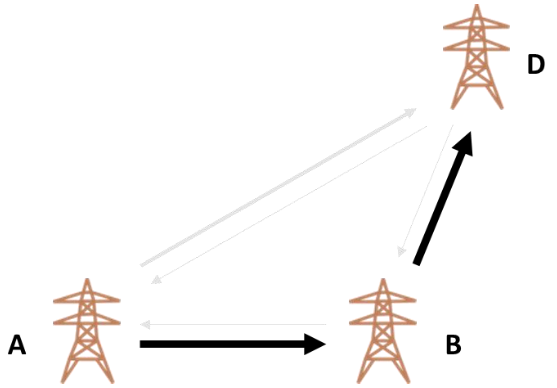
Possible solution: maximum likelihood approach

How about more complex networks?

HITACHI
Inspire the Next







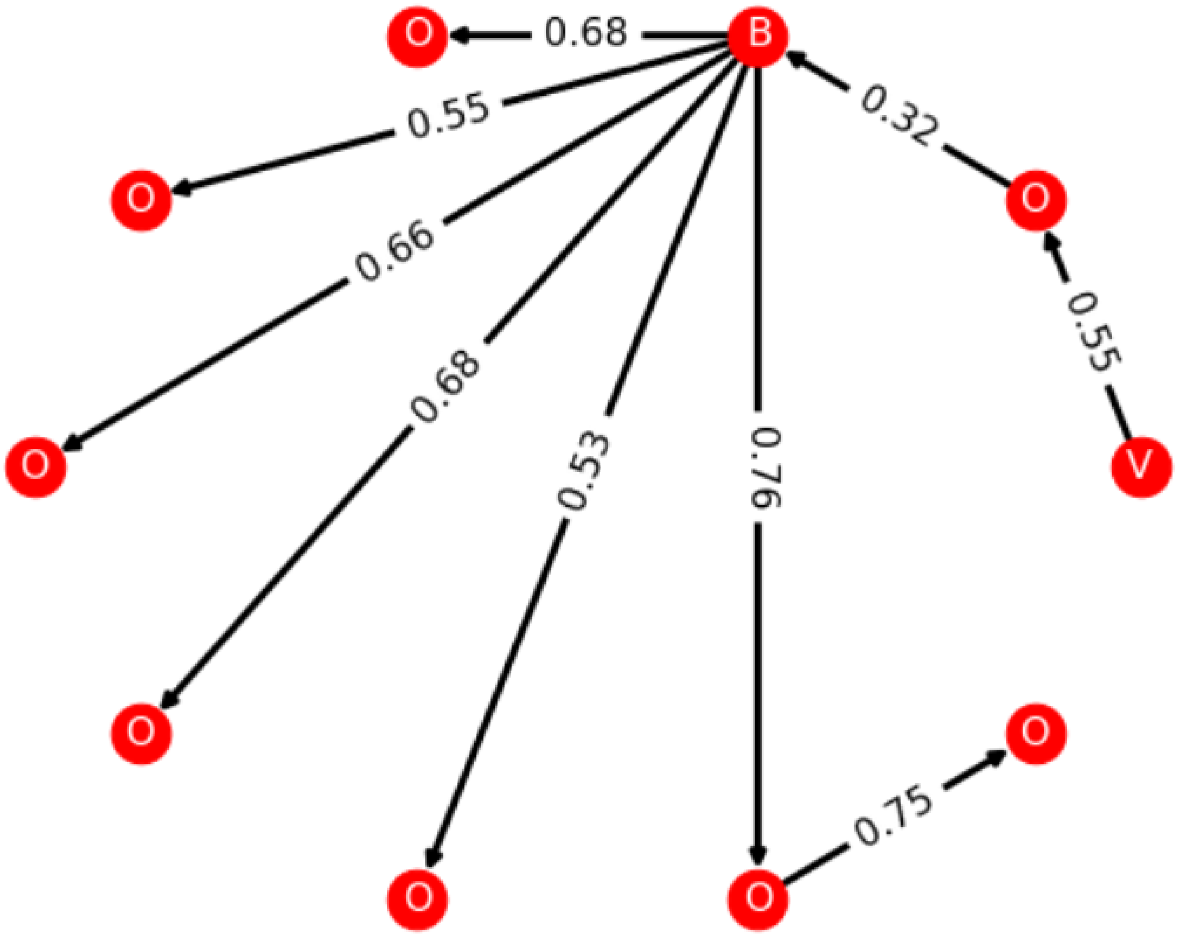
STEP 1: Use historical data to find conditional probabilities between every two alarms using *Bayesian inference* approach.

STEP 2: For the given alarm flood isolate nodes represented in the given alarm flood.

STEP 3: Find a directed tree (spanning arborescence) with highest likelihood using *Edmonds' algorithm*.

Hitachi Energy

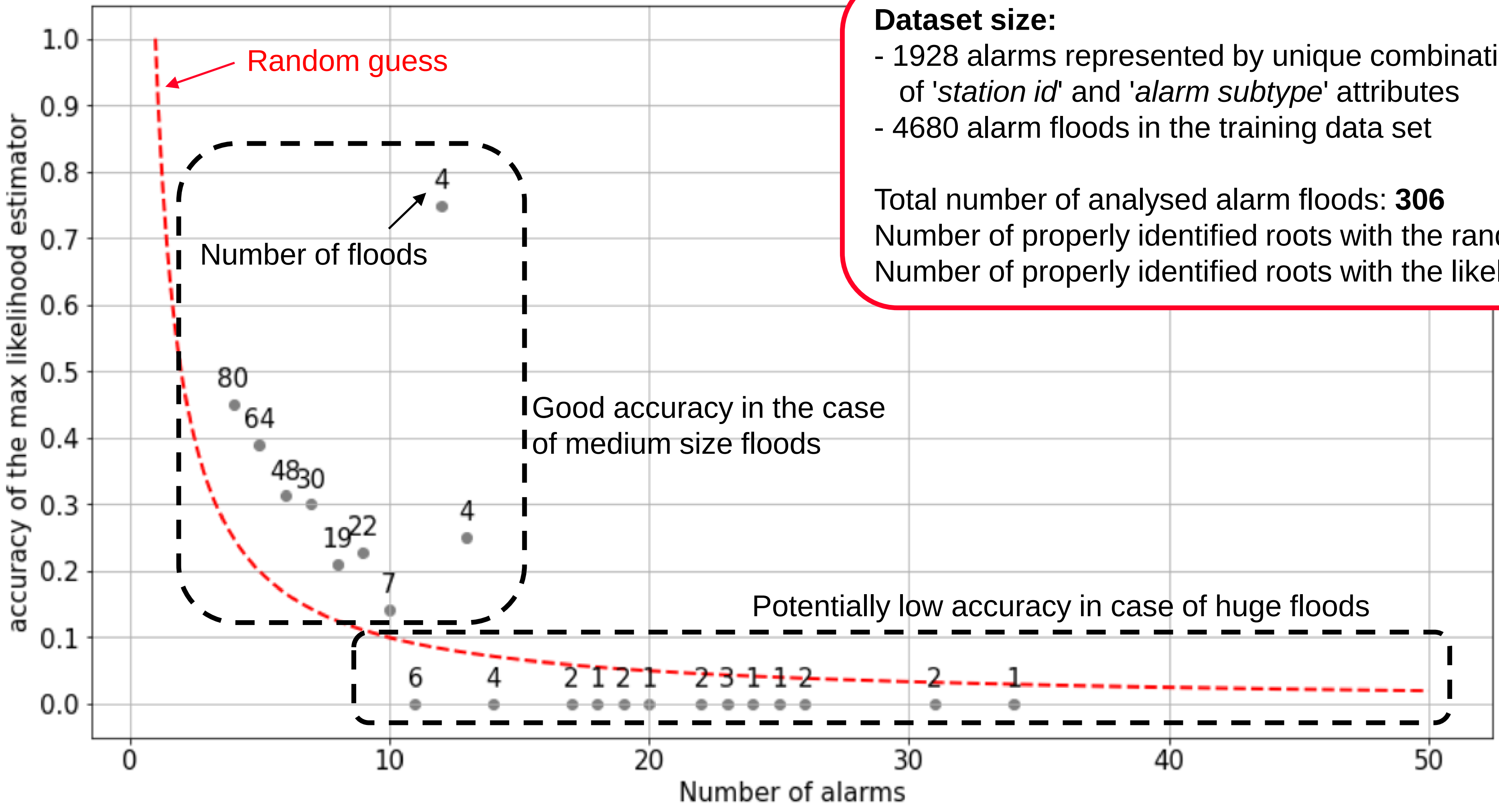
Example of a generated causality tree



Root node

But is this really the earliest alarm activated in the flood?

Numerical results for Hitachi dataset



Dataset size:
- 1928 alarms represented by unique combinations of '*station id*' and '*alarm subtype*' attributes
- 4680 alarm floods in the training data set

Total number of analysed alarm floods: **306**
Number of properly identified roots with the random selection: **53**
Number of properly identified roots with the likelihood estimator: **99**

Conclusions and further work from Approaches 1 & 2

HITACHI
Inspire the Next

Conclusions

1. Data preprocessing is critical to filter alarms which obstruct further classification process.

2. Alarm floods were classified into four key clusters with a clear physical meaning.

3. Causality is impossible to uniquely extract from the available data, but still maximum likelihood approach was used to identify potentially plausible root alarms.

Further work

1. Further refinement of available data, *e.g.* by including information about the electric grid structure and labelled past flood alarm with a known origin.

2. Developing methods for real time classification of new floods to one of the precomputed cluster for the operator decision support.

3. Causality identification can be improved both by refining the maximum likelihood approach and combining it with expert knowledge.

Hitachi Energy