Leveraging the sparse data for large power transmission blackouts with generative statistical models

Ian Dobson

IOWA STATE UNIVERSITY

Foundation Models for the Electric Grid Workshop
Argonne National Laboratory, February 2025

Funding in part from USA NSF is gratefully acknowledged

Overview

- Large power system blackouts are complicated and rare but high risk
- **Complicated** ⟹ hard to model and simulate
- **Rare** ⟹ not enough data to train AI models
- Can AI help in unusual or extreme conditions when operator and planning guidance is most needed but data is scarce?
- One solution: use **generative models to create synthetic data**
- We give an example of generative modeling of the fast-time scale spread of line outages on the power transmission system network

The general nature of power transmission systems

- Engineered system designed and operated to “keep the lights on”
- Large, heterogeneous system with continuous, nonlinear, hybrid, uncertain and varying dynamics at time-scales from milliseconds to decades
- Involves a large network, but also generator, load, and other device dynamics and a myriad of control and protection systems
- Follows physical laws & cost, business, regulatory, political constraints
- Modeling is hard: There is no such thing as a power system model; there are many different models at many size and time scales with very different approximations for different purposes

Background

- Transmission network is the bulk electric power system (above 100 kV) that spans states and regions
- Transmission lines = network edges; Substations = network nodes
- The protection system quickly (time scale of cycles = 1/60 second) detects and isolates faults/short circuits by removing lines. It is complex and has back up systems
- There are many other diverse interactions affecting the system at slower time-scales, which are not (yet) addressed in our generative modeling

Fast time-scale line outages on the network

- Transmission line outages spreading on the network at a fast time scale (< 1 minute) due to the protection system contribute in part to blackouts
- The patterns of transmission line outages on the network at the fast time scale can be easily obtained from utility data and statistically modeled
- A generative model can quickly produce patterns of line outages that match key statistics of the observed patterns … this could improve simulations and provide enough data for model training

OBSERVING AND MODELING FAST PROTECTION ACTIONS

- Here we observe and model at the fast protection time scale (<1 minute)
- It is very hard to make a detailed model of protection at transmission system scale including misoperation, stuck breaker, substation effects, sympathetic trips, etc …. but these effects are included in observed outages spreading on the network.
- Each minute of outage data includes all these effects, and it is simple to extract the patterns of all the line outages that start in each minute.
- Our data is public and from
  BPA = Bonneville Power Administration, including WA, OR states
  NYISO = New York Independent System Operator, NY state and surrounding

Detailed outage data from BPA on the web

Automatic (Unplanned) Transmission Line Outages: 2009 Complete

CHRONOLOGICAL ORDER

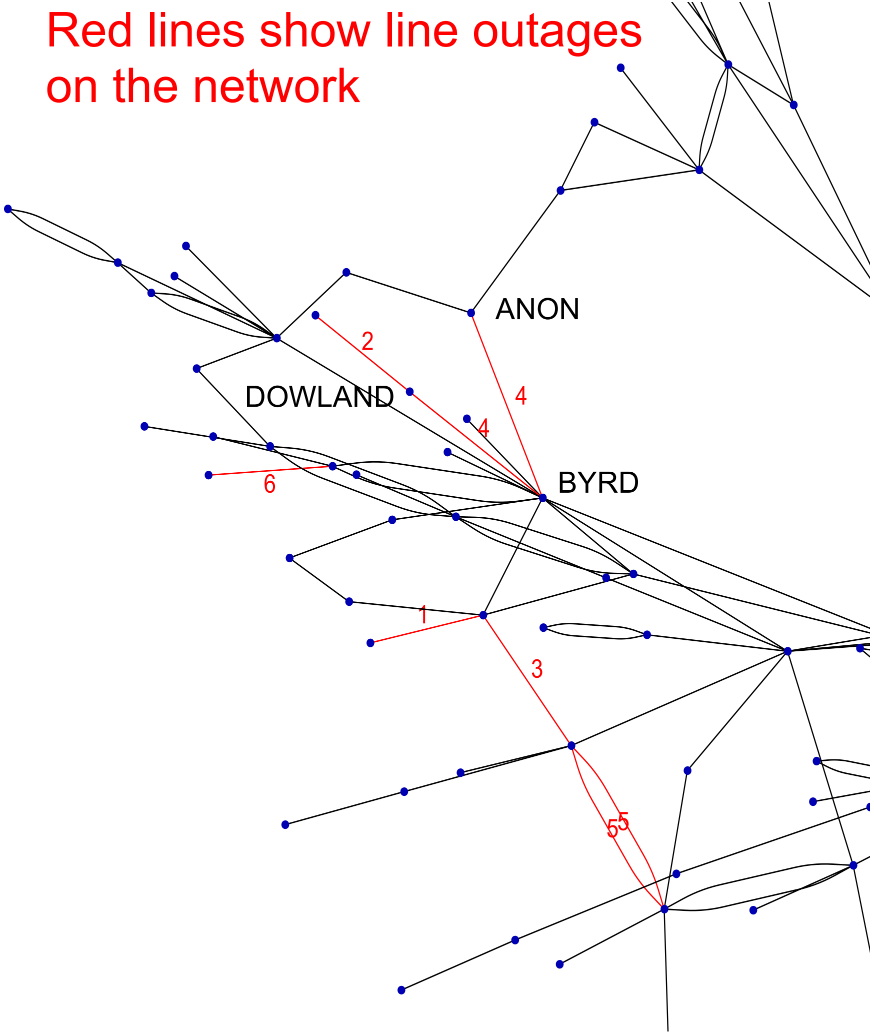
Outage#	Tred ID	Line Name	Gen Flag	kV	District	Own Code	Length (Mi)	Out Date/Time	In Date/Time	Out Mins	Disp Cause
157560	339	xxxx-xxxxxx (230 kV)		230	xxxx	2	0.5	6/18/07 23:48	2/23/09 14:38	886550	81
164651	140	xxxx-xxxxxx (230 kV)	G	230	xxxx	1	61.9	1/2/09 2:35	1/2/09 17:43	908	31
164652	497	xxxx-xxxxxx (115 kV)	G	115	xxxx	1	24.8	1/2/09 3:55	1/2/09 6:59	184	90

Includes 10942 automatic line trip times *to the nearest minute*;
Required data is available to all USA utilities and is reported to regulators

Use 19 years of BPA data. We also use 12 years of NYISO data

Outages that start in the same minute are at the fast protection time scale
Outages that start in the same minute form the **patterns**

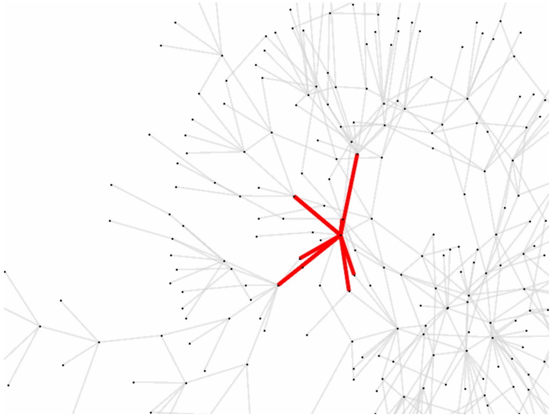
Red lines show line outages on the network

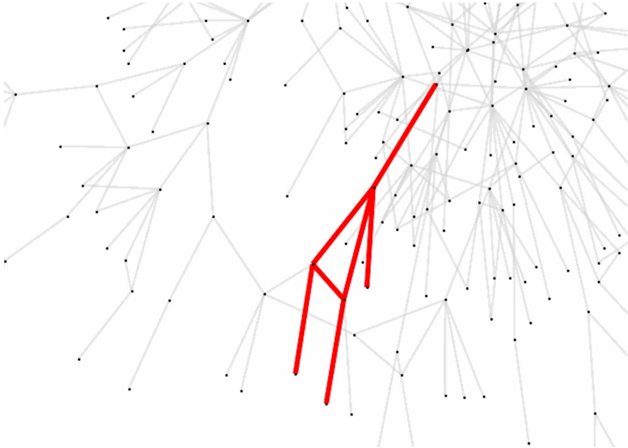


- Mostly protection removes one line to make the pattern
- Dowland-Byrd and Anon-Byrd start in the same minute and make the pattern

transmission line	outage start time hour:minute
JOSQUIN – ISAAC	15:22
GIBBONS – DOWLAND	15:25
ISAAC – OCKEGHEM	15:27
DOWLAND – BYRD	15:37
ANON – BYRD	15:37
OCKEGHEM – DUFAY No 1	15:49
TYE – TALLIS	15:57

Larger patterns also occur but more rarely:

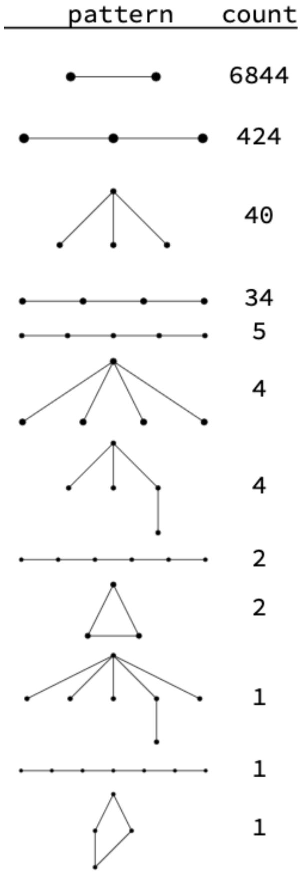




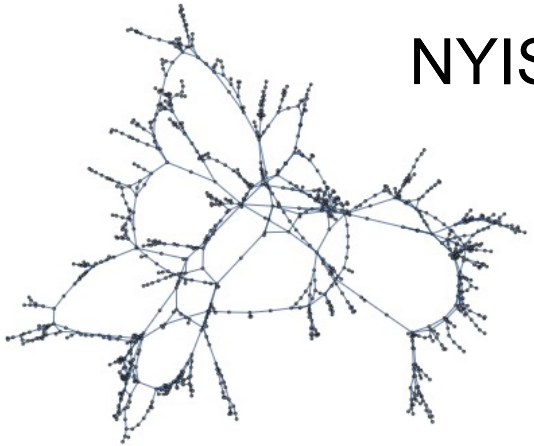
Observed NYISO patterns of line outages starting at the same minute

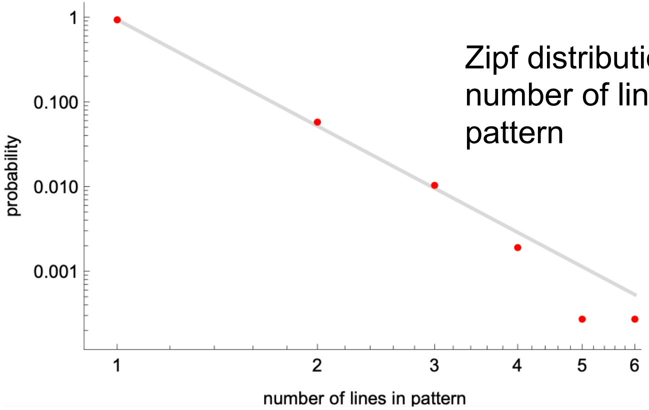
- Patterns are mainly small connected trees.
- Large patterns are rare, but occur with heavy tail.

12 years of NYISO data



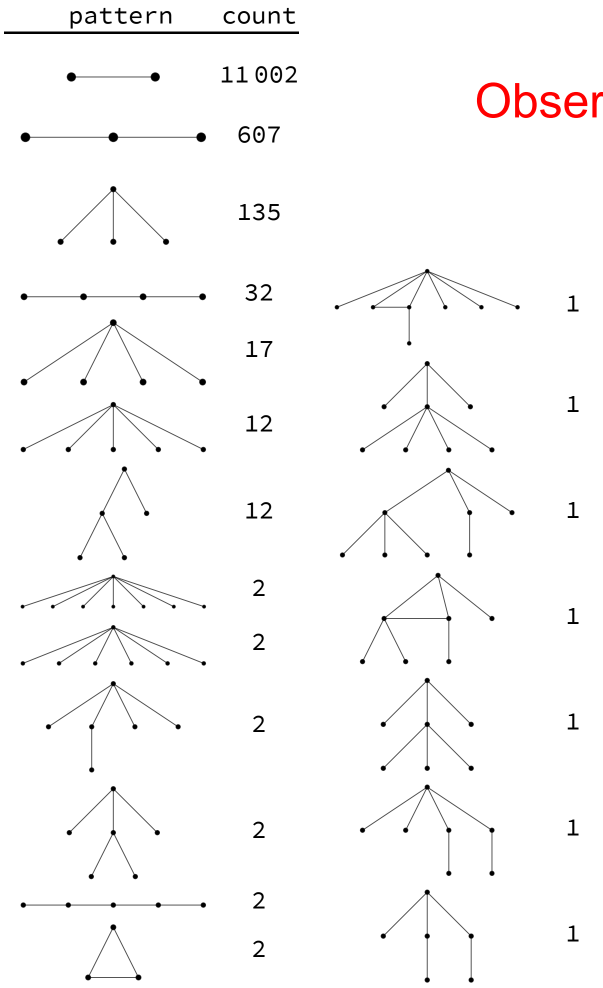
NYISO

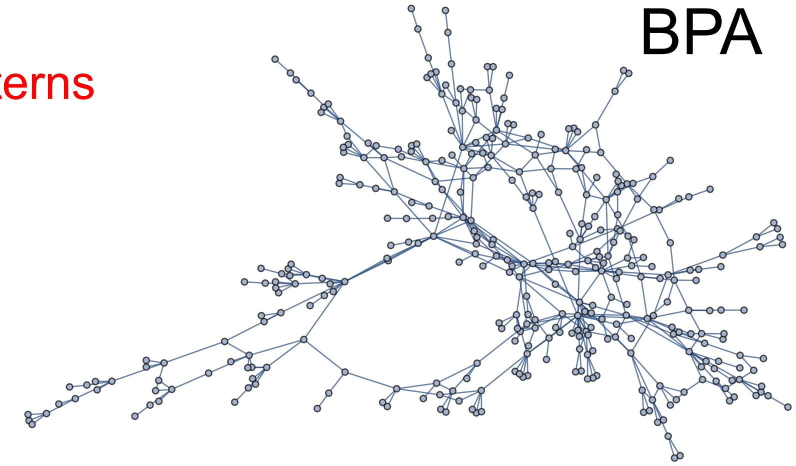


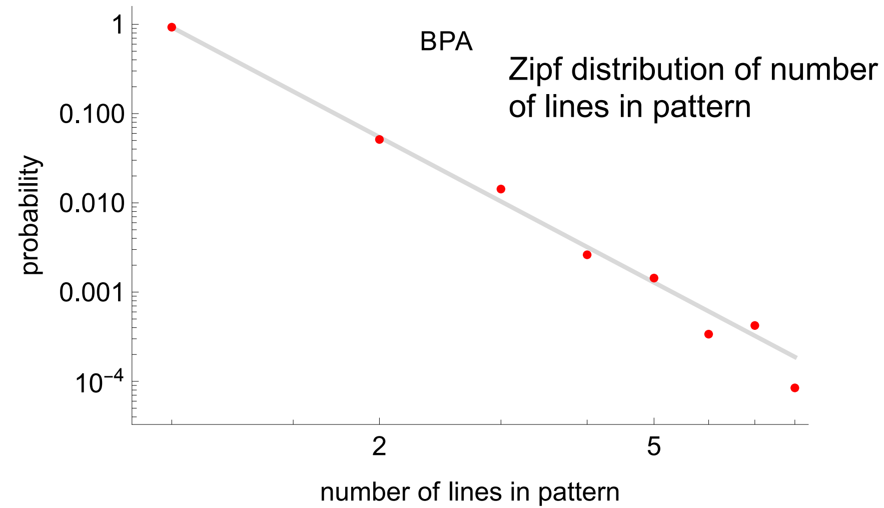


Zipf distribution of number of lines in pattern

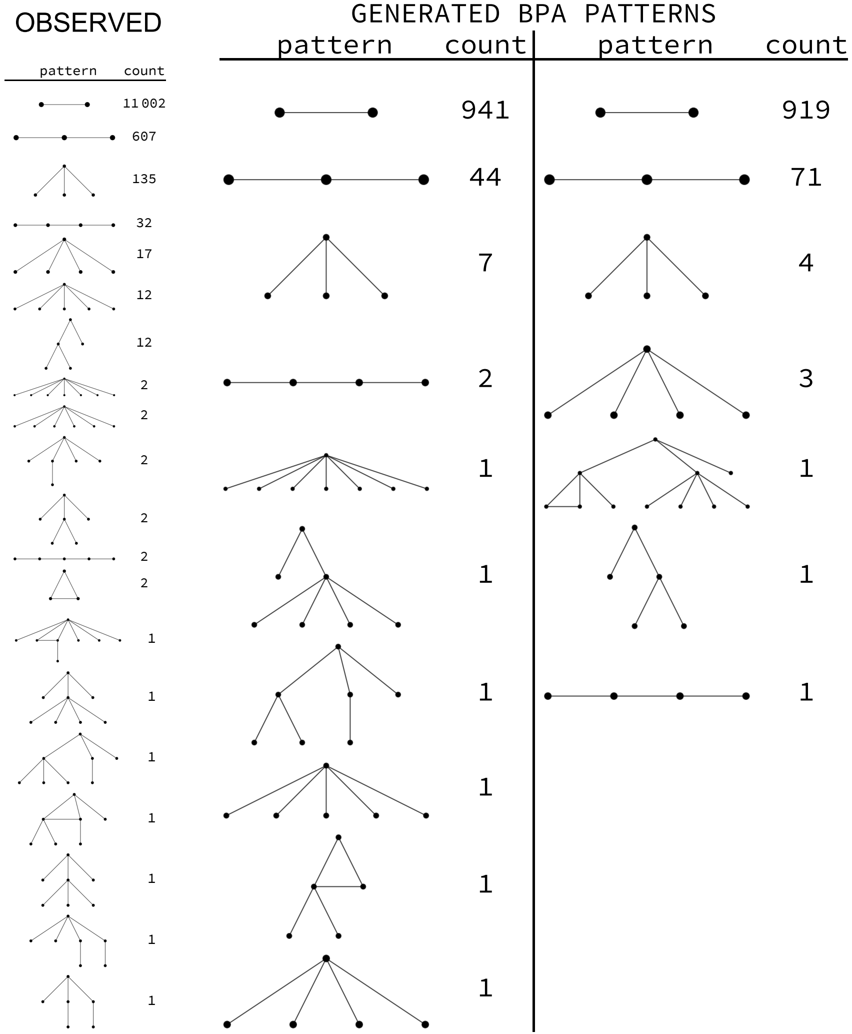
Observed BPA patterns







19 years of data

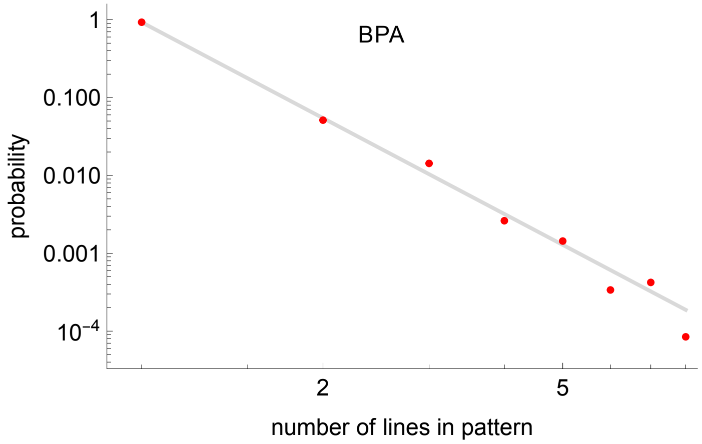


Generative Modeling of Protection at transmission system scale -- a new capability

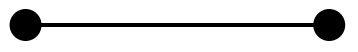
- Can make model to generate patterns *statistically similar* to those observed… since data-driven it captures combined effect of many mechanisms.
- Simplest version simply attaches another line to any outage with probability 0.07
- Can also easily generate the larger and rarer patterns:
  (1) given any single line out, sample total number of lines out from Zipf distribution
  (2) attach these lines to form the pattern according to probability of attaching at edges or middle of pattern.

Generative modeling to produce patterns

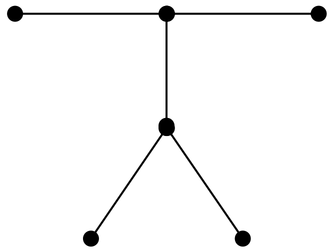
Sample number of lines in pattern



Start pattern with one line chosen at random



Add the remaining lines at edges with probability p_{1+} or middle of pattern with probability 1 - p_{1+}



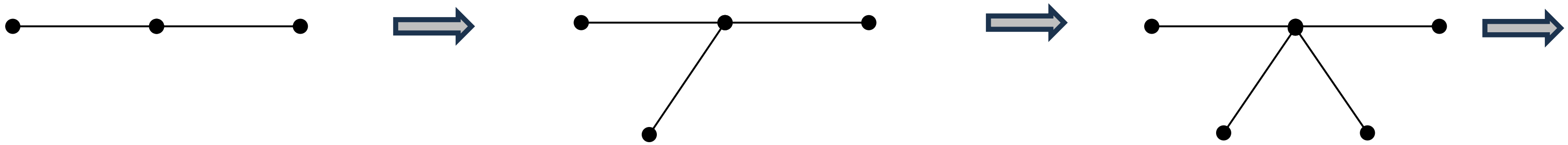
pattern edge = degree 1
middle of pattern = degree ≥ 2

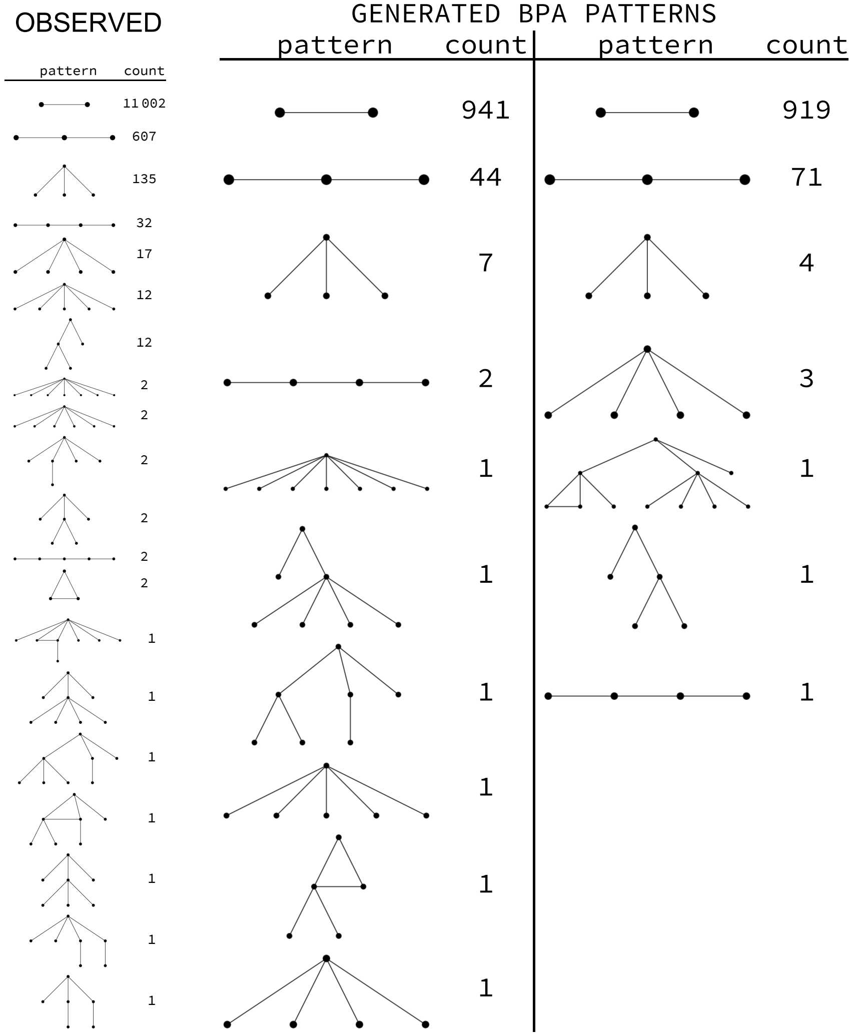
p1+ is obtained from the observed patterns

Generative modeling to produce patterns

Only add at edges = degree 1 gives linear patterns:

Only add at middle = degree ≥ 2 gives star patterns:





Generative Modeling of Protection at transmission system scale -- a new capability

- Can make model to generate patterns *statistically similar* to those observed… since data-driven it captures combined effect of many mechanisms.
- Simplest version simply attaches another line to any outage with probability 0.07
- Can also easily generate the larger and rarer patterns:
  (1) given any single line out, sample total number of lines out from Zipf distribution
  (2) attach these lines to form the pattern according to probability of attaching at edges or middle of pattern.

Testing the generated patterns

How close are sets of generated patterns to observed patterns?
We set up a distance between patterns and sets of patterns and produce 1000 sets of generated patterns. Then

- Mean Wasserstein distance between observed and generated patterns corresponds to changing 107 lines out of 11836 lines (BPA) or changing 99 lines out of 7632 lines (NYISO)

- A permutation test indicates that is not possible to distinguish between generated and observed data most of the time

Conclusions for observing and modeling fast protection actions

- Easy to observe propagating outages in protection system from outage data
- Simple model additionally outages neighboring line with probability 7%
- More elaborate generative model generates larger patterns that match some key statistics of the observed patterns
- Similar generative models in spread of blog posts and discussion threads
- Potential applications of generative models are
  (1) Better prediction of outage spread in simulations assessing blackout risk
  (2) Generate enough synthetic data for rare protection events to train AI models
- Future work: similar methods for statistics and generative modeling for weather impacts and cascading at slower time scales.

New description and generative modeling of propagating outages across a large transmission system at the fast protection time scale

References

Reference for fast protection system statistics and generative models

I Dobson, DA Maldonado, M Anitescu, The statistical spread of transmission outages on a fast protection time scale based on utility data, Arxiv preprint July 2024, https://arxiv.org/abs/2407.15059#

Reference towards generative modeling for resilience

MR Kelly-Gorham, PDH Hines, I Dobson, Ranking the impact of interdependencies on power system resilience using stratified sampling of utility data, IEEE Trans. Power Systems, 39(1), 2024 https://iandobson.ece.iastate.edu/PAPERS/kellyPS24preprintpdfA.pdf

CONTACT
Ian Dobson dobson@iastate.edu