Data analysis of cascading outages using historical data to mitigate blackouts

Kai Zhou

Program of study committee:

Ian Dobson, major professor

Zhaoyu Wang, major professor

Arka P. Ghosh

Jarad Niemi

Aditya Ramamoorthy

Preliminary oral exam

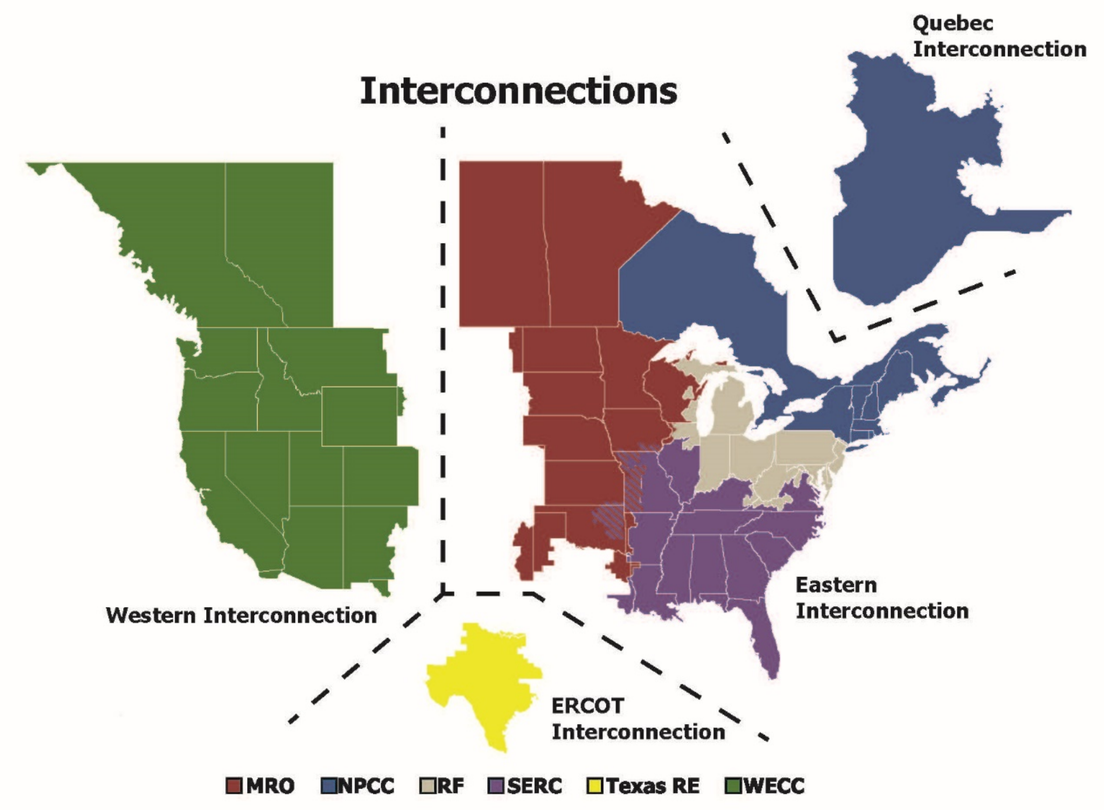
10/12/2020

Outline

- Background and motivation
- Problem statement
- Literature review
- Part 1. Markovian influence graph driven by historical data
- Part 2. Bayesian hierarchical model of individual transmission line outage rates
- Conclusions
- Future work plan

Background and motivation

- A simple power system model is a graph with transmission lines as edges and substations as vertices.
- Power systems are interconnected, so outages can propagate in the network.



Source: North American Electric Reliability Corporation

Background and motivation

- A cascade is a sequence of outages that starts with initial outages and then propagates.
- **Initial** outages: random outages at random times.

Various causes initiate cascading, like weather, earthquakes, human, tree, equipment failures, operational or planning errors, etc.

- **Propagating** outages: arise jointly from the preceding outages and the changing power system state.

Outages propagate because of transmission line overload, hidden failures of protection systems, misoperations or designed operations, etc.

Background and motivation

- Cascading is one of the main causes of blackouts.
- Large cascading outages are rare but have high impact.
- A small set of system components contributes to large blackouts.
- Upgrade critical components to mitigate blackouts and reduce risk.

Table. The total number of power outages recorded in the US from 2008 to 2016

Year	Total number of outages	People affected
2008*	2,169	25.8 million
2009	2,840	13.5 million
2010	3,149	17.5 million
2011	3,071	41.8 million
2012	2,808	25.0 million
2013	3,236	14.0 million
2014	3,634	14.2 million
2015	3,571	13.2 million
2016	3,879	17.9 million

*****Partial-year data. Data collection began on February 16, 2008.*

Source: Eaton, Blackout tracker United States annual report 2016.

Background and motivation

- Transmission Availability Data System(TADS) collects outage data in North America.
- Bonneville Power Administration(BPA) makes its outage data public.

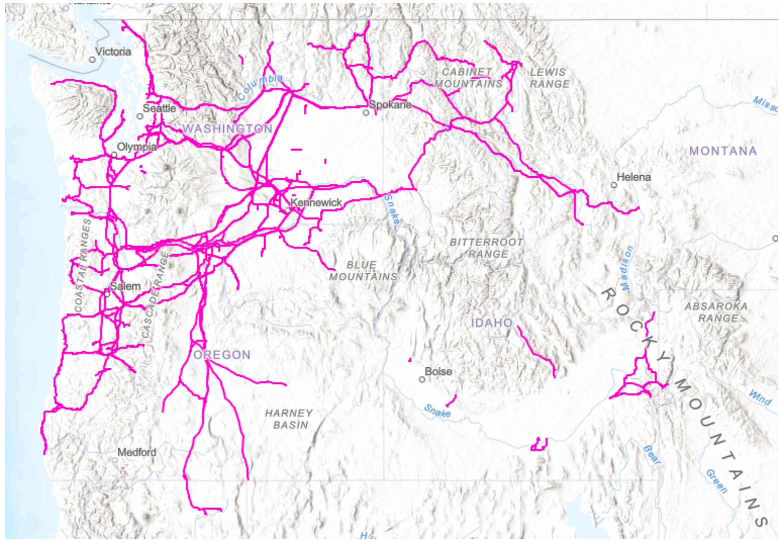


Figure. BPA area

Table. Raw outage data

Out Datetime	In Datetime	Name	Voltage (kV)	Duration (minutes)	Outage Type	Dispatcher Cause	Field Cause	System In Control	MW Intrpt	District	Outage ID
01/29/1999 08:10	01/29/1999 08:10	xxx	115.0	0	Auto	Weather		BPA Trouble	Unknown	LGV	114174
01/29/1999 08:10	01/29/1999 08:10	xxx	34.5	0	Auto	Weather		BPA Trouble	Unknown	LGV	114175
01/29/1999 08:10	01/29/1999 08:10	xxx	115.0	0	Auto	Weather		BPA Trouble	Unknown	LGV	114176

Source: Bonneville Power Administration

Background and motivation

Why use historical outage data?

- Standard outage data is routinely collected; some data is publicly available.
- Studying real outage data does not need to make assumptions about cascading mechanisms.
- Model-based approaches can only approximate a subset of cascading mechanisms.
- Simulation is just starting to be benchmarked and validated.

Problem statement

Given historical outage data,

- how are cascading outages propagating in power systems?
- which lines contribute most to the propagation of outages?
- which lines have high outage rates?
- how to test the mitigation effect after upgrading critical lines?

A challenge is that historical outage data is limited.

Objective

Develop statistical methods applied to real outage data so that we can understand the propagation of cascading outages, identify critical lines in propagating outages and initial outages, and evaluate the mitigation effect.

In particular,

- form a Markovian influence graph that describes probabilities of transitions between line outages. The quasi-stationary distribution of this Markov chain identifies critical lines that are most involved in the propagation of large cascades.
- build a Bayesian hierarchical model leveraging partial transmission line dependencies to estimate individual transmission line outage rates.

Literature review

- Influence graph models
  - Influence is the relation between two components in terms of outages; influence graph model visualizes influences on a graph.
  - are first proposed in [1], and further developed in [2][3].
  - quantify influences between component outages in three ways: (1) conditional probabilities [1-3]; (2) line outage distribution factors [4]; (3) correlations [5].
  - previous influence graphs consider single components and do not exploit Markov structure.
- Our influence graph is formed from real outage data, solves sparse data problem, has multiple outages as states, exploits Markov structure, and computes uncertainty of results.

[1] P. D. H. Hines, et at, "'Dual Graph' and 'Random Chemistry' Methods for Cascading Failure Analysis," in *Hawaii Intl. Conf. System Sciences*, Jan. 2013.

[2] J. Qi, et al, "An Interaction Model for Simulation and Mitigation of Cascading Failures," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 804–819, Mar. 2015.

[3] P. D. H. Hines, et al, "Cascading Power Outages Propagate Locally in an Influence Graph That is Not the Actual Grid Topology," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 958-967. Mar. 2017.

[4] Z. Ma, et al, "Fast screening of vulnerable transmission lines in power grids: A pagerank-based approach," IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 1982–1991, Mar. 2019.

[5] Y. Yang, et al, "Vulnerability and cosusceptibility determine the size of network cascades," Phys. Rev. Lett., vol. 118, no. 4, 2017, Art. no. 048301.

[6] A. Wang, et al, "Vulnerability assessment scheme for power system transmission networks based on the fault chain theory," IEEE Trans. Power Syst., vol. 26, no. 1, pp. 442-450, Feb. 2011.

[7] M. Rahnamay-Naeini, et al "Stochastic analysis of cascading-failure dynamics in power grids," IEEE Trans. Power Syst., vol 29, no. 4, pp. 1767-1779, Jul. 2014.

Literature review

- Estimating transmission line outage rates:
  - Some research uses Bayesian models, estimates outage rates for a group of components: outage counts in a substation district [8], counts in arbitrary areas [9], counts in distribution feeders [10].
  - Many researchers study predicting outage probabilities in a short term according to weather conditions [11][12][13].
- little research is on individual transmission line outage rates.

[8] H. Li, L. A. Treinish, and J. R. M. Hosking, “A statistical model for risk management of electric outage forecasts,” IBM J. Res. Dev., vol. 54, no. 3, pp. 8:1–8:11, May 2010
[9] T. Iesmantas and R. Alzbutas, “Bayesian spatial reliability model for ˇ power transmission network lines,” Electr. Power Syst. Res., vol. 173, pp. 214–219, 2019.
[10] A. Moradkhani et al., “Failure rate estimation of overhead electric distribution lines considering data deficiency and population variability,” Int. Trans. Electr. Energ. Syst., vol. 25, no. 8, Apr. 2015.
[11] Y. Zhou, et al, “Modeling weather-related failures of overhead distribution lines,” IEEE Trans. Power Syst., vol. 21, no. 4, pp. 1683–1690, Nov. 2006
[12] H. Li, L. A. Treinish, and J. R. M. Hosking, “A statistical model for risk management of electric outage forecasts,” IBM J. Res. Dev., vol. 54, no. 3, pp. 8:1–8:11, May 2010.
[13] T. Dokic et al., “Spatially aware ensemble-based learning to predict weather-related outages in transmission,” in Proc. 52th Hawaii Intl. Conf. System Science, Maui, HI, USA, Jan. 2019

Part 1. Markovian influence graph driven by historical data

Collaborators:

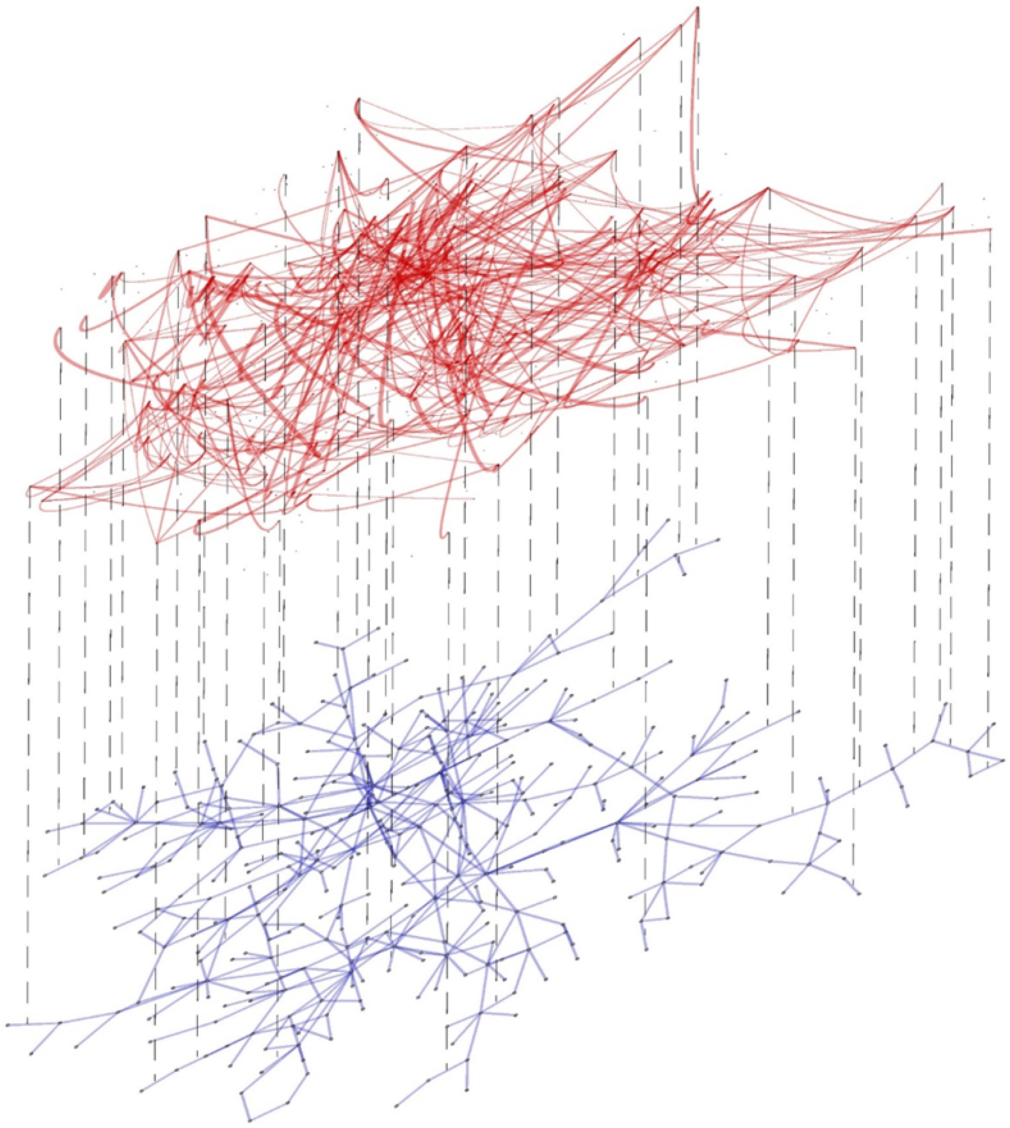
Arka Ghosh, Department of Statistics, Iowa State University

Alexander Roitershtein, Texas A&M

Markovian influence graph (red) from real data on real grid (blue)

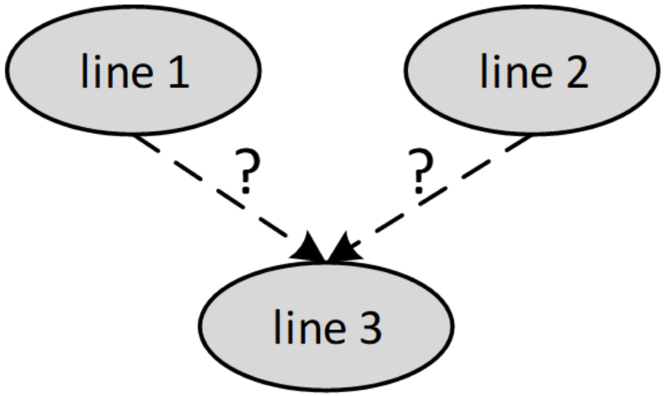
Physical network and influence graph

- Components interconnected with each other form a **physical network**.
- **Influence graph** describes influences between components probabilistically in cascading outages.



Introduction

- Why a new influence graph?
  - Form a influence graph from historical data.
  - Simultaneous outages are common. Previous influence graphs consider single components [2][3][9].



---

[2] J. Qi, et al, “An Interaction Model for Simulation and Mitigation of Cascading Failures,” IEEE Trans. Power Syst., vol. 30, no. 2, pp. 804–819, Mar. 2015.
[3] P. D. H. Hines, et al, “Cascading Power Outages Propagate Locally in an Influence Graph That is Not the Actual Grid Topology,” IEEE Trans. Power Syst., vol. 32, no. 2, pp. Mar. 2017.

Introduction

- Group line outages into cascades, then into generations within each cascade according to outage times.

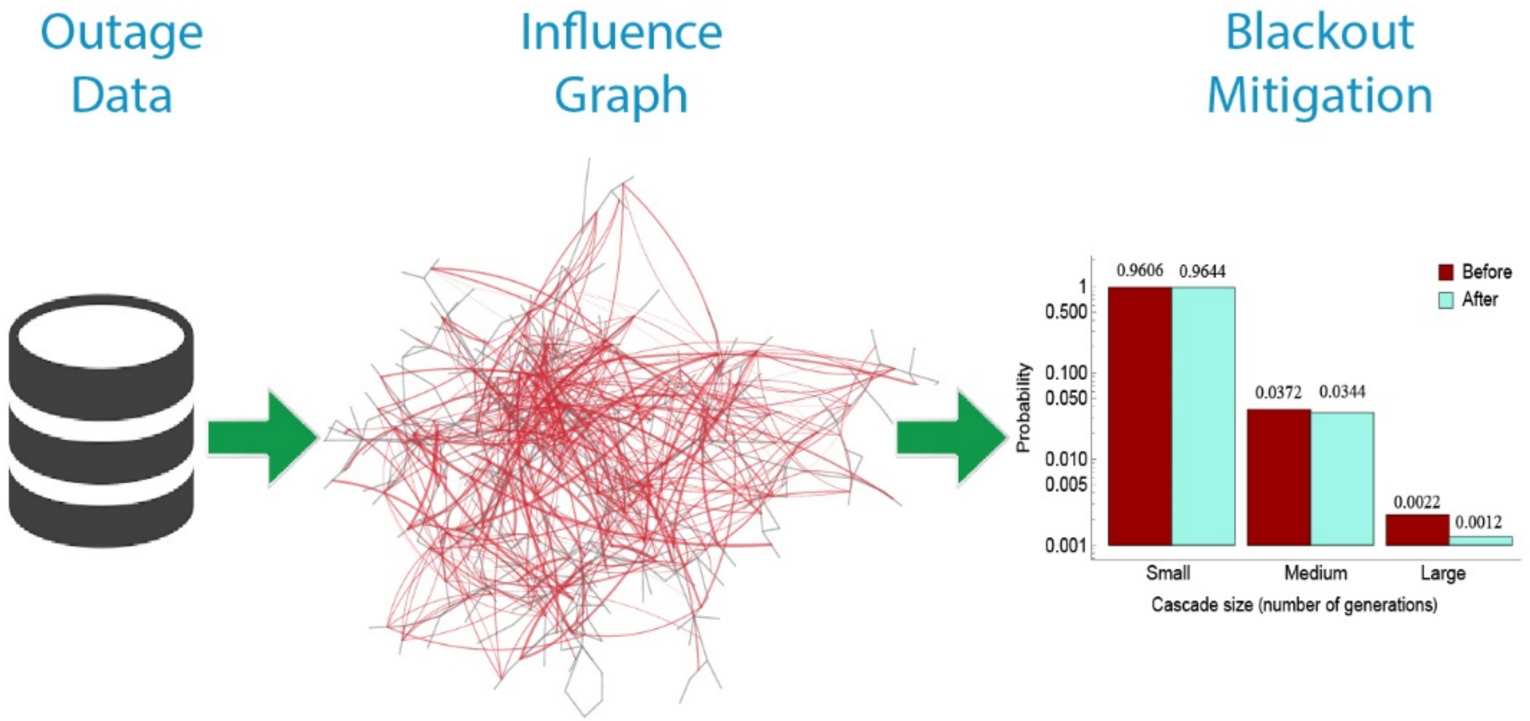
Out Datetime	In Datetime	Name	Voltage (kV)	Duration (minutes)	Outage Type	Dispatcher Cause	Field Cause	System In Control	MW Intrpt	District	Outage ID
01/29/1999 08:10	01/29/1999 08:10	xxx	115.0	0	Auto	Weather		BPA Trouble	Unknown	LGV	114174
01/29/1999 08:10	01/29/1999 08:10	xxx	34.5	0	Auto	Weather		BPA Trouble	Unknown	LGV	114175
01/29/1999 08:10	01/29/1999 08:10	xxx	115.0	0	Auto	Weather		BPA Trouble	Unknown	LGV	114176

cascade number	generation 0 X_0	generation 1 X_1	generation 2 X_2	generation 3 X_3
1	{line 1}	{line 3}	{line 2}	{}
2	{line 2}	{line 1, line 3}	{}	{}
3	{line 3}	{line 1}	{}	{}
4	{line 1}	{}	{}	{}

Introduction

Objective:

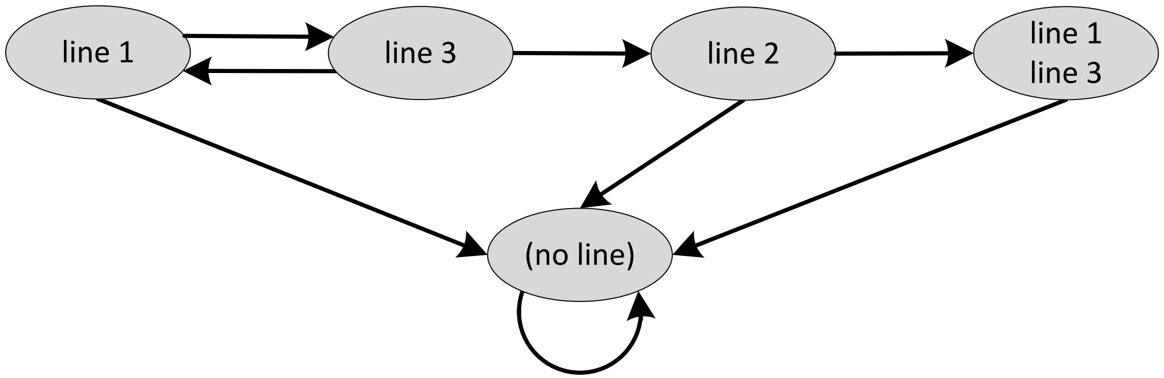
Form a rigorous Markovian influence graph using historical outage data to identify critical lines and test mitigation of large cascades.



A toy example

- Markovian influence graph is a rigorous Markov chain:
  - Discrete;
  - States: generations of cascades, multiple outages as a single state;

cascade number	generation 0 X_0	generation 1 X_1	generation 2 X_2	generation 3 X_3
1	{line 1}	{line 3}	{line 2}	{}
2	{line 2}	{line 1, line 3}	{}	{}
3	{line 3}	{line 1}	{}	{}
4	{line 1}	{}	{}	{}

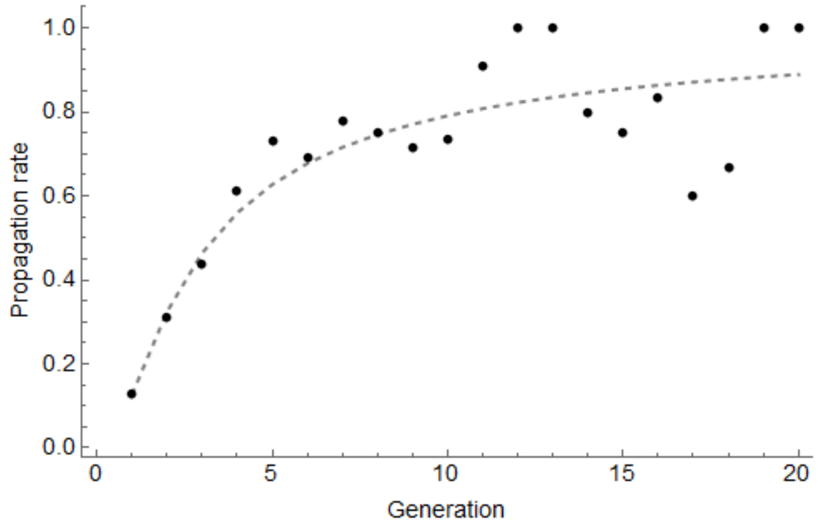


What type of Markov chain

- This Markov chain has an absorbing state: {} or {no line}.

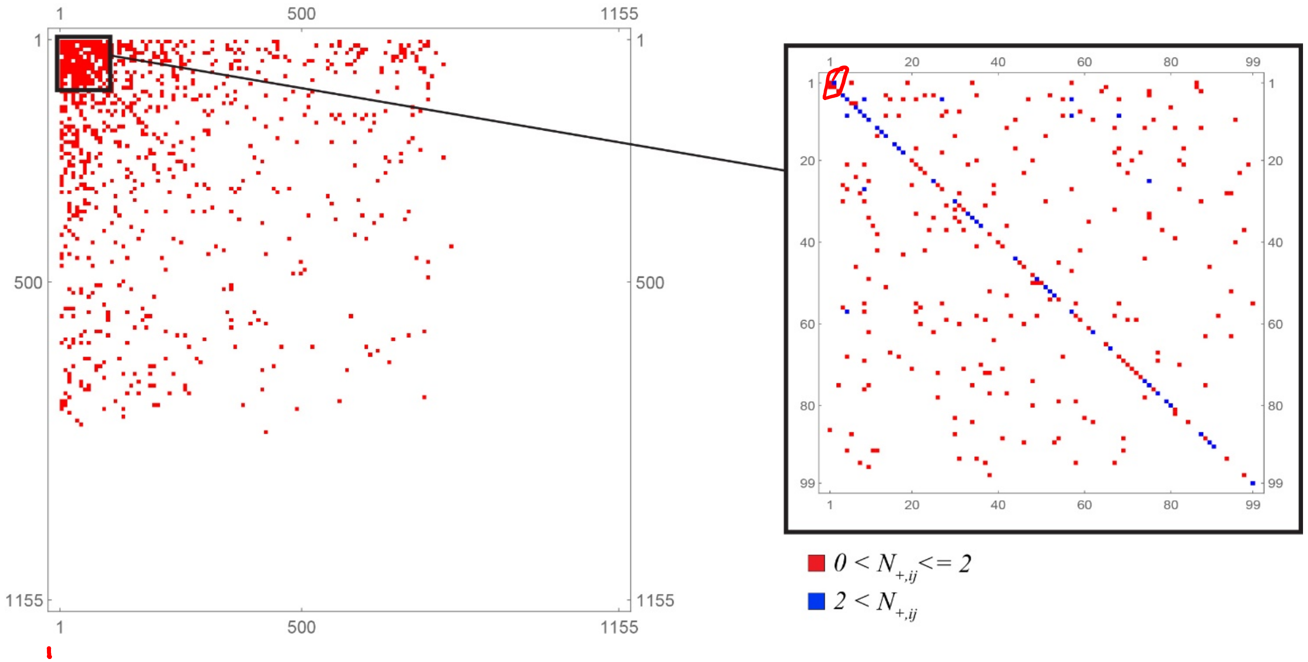
$$
P_k = \begin{array}{c} \\ \{\} \\ \{1\} \\ \{2\} \\ \{3\} \\ \{1,3\} \end{array}
\begin{array}{c}
\begin{array}{ccccc} \{\} & \{1\} & \{2\} & \{3\} & \{1,3\} \end{array} \\
\left(\begin{array}{c|cccc}
1 & 0 & 0 & 0 & 0 \\ \hline
u_1 & Q_{11} & Q_{12} & Q_{13} & Q_{14} \\
u_2 & Q_{21} & Q_{22} & Q_{23} & Q_{24} \\
u_3 & Q_{31} & Q_{32} & Q_{33} & Q_{34} \\
u_4 & Q_{41} & Q_{42} & Q_{43} & Q_{44}
\end{array}\right)
\end{array}
$$

- Transition matrices P_k is variant with time step k to capture increasing propagation rates.



Methods of estimating transition matrices

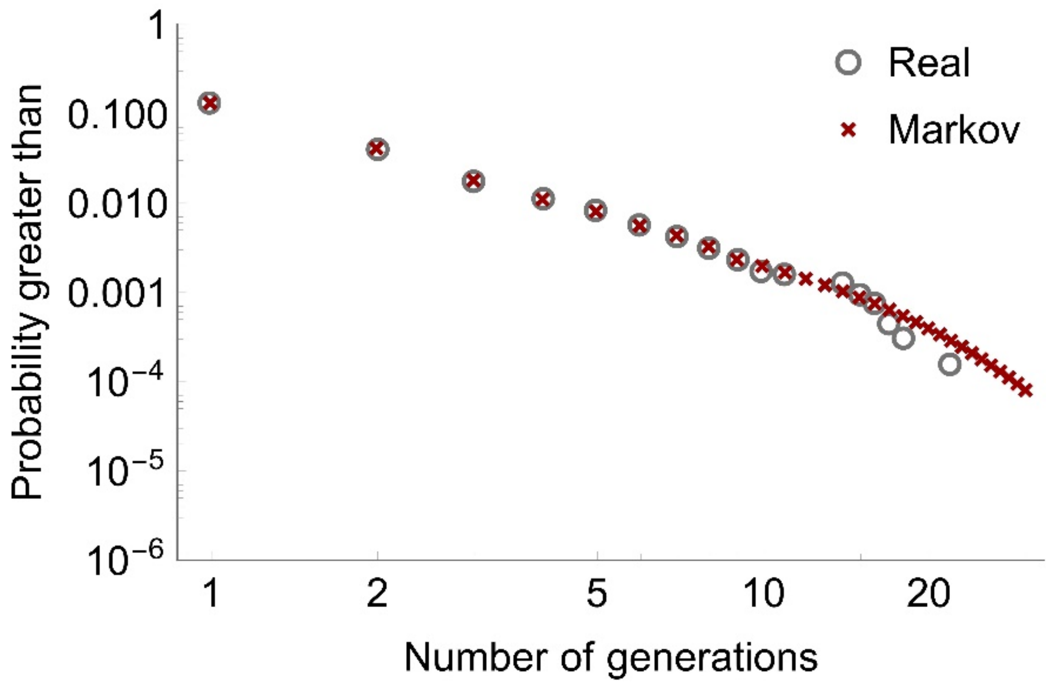
- Three steps to mitigate the problem of limited data:
  - group data of high generations together;
  - Bayesian methods estimate stopping probabilities;
  - adjust transition matrices to match propagation rates.

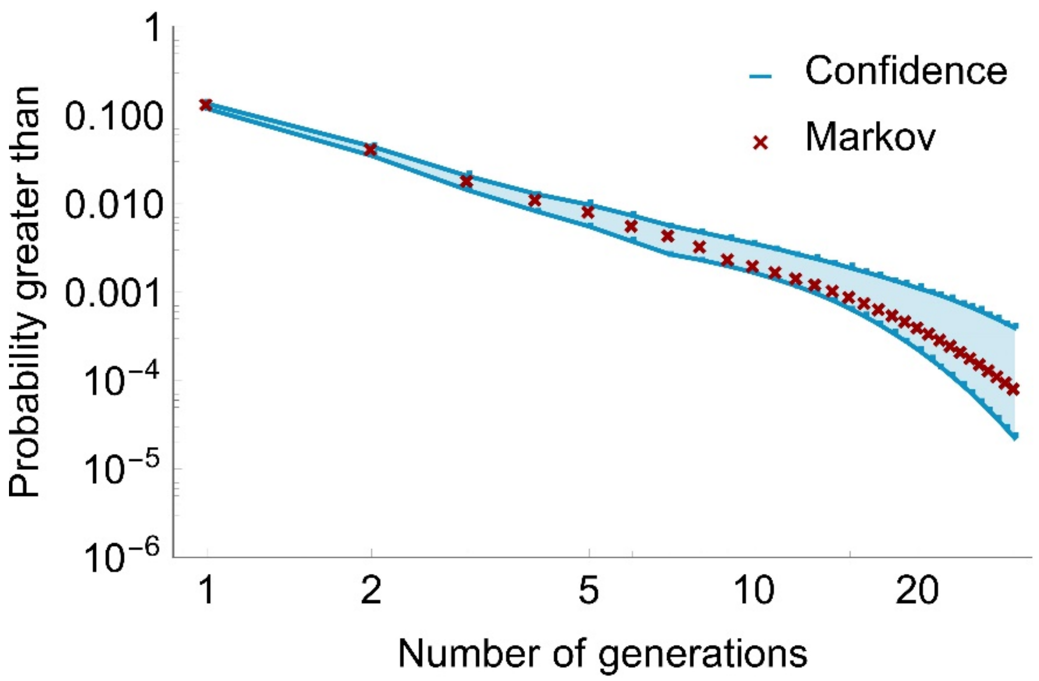


Transition count matrix

Compute cascade size distribution and its uncertainty

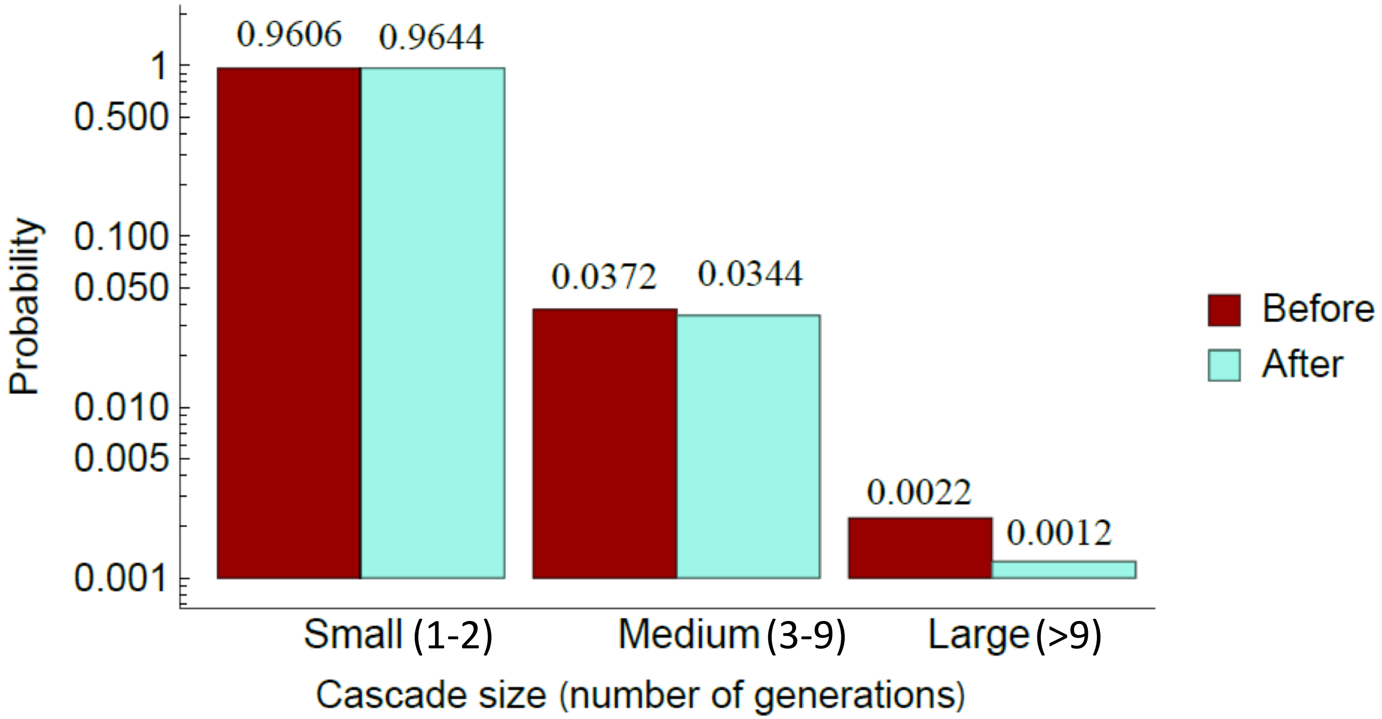
- Verify the model by comparing empirical cascade size (number of generations) distribution and distribution produced by the influence graph.
- Use bootstrap to estimate 95% confidence intervals.
- Probability of large cascades (> 9 generations) is within a factor of 1.5.





Identify critical lines and test mitigation

- identify 10 critical lines (1.6% of total lines) in propagation by calculating the quasi-stationary distribution, which is the dominant left eigenvector of submatrix of transition matrix;
- test mitigation: upgrade critical lines by reducing the corresponding columns of transition matrices by 80%, and recompute the cascade size distribution.



Summary

The new influence graph generalizes and improves previous work in several ways:

- uses real outage data routinely collected by utilities.
- mitigates limited data problem with several new methods.
- obtains a clearly defined Markov chain by including multiple outages as states.
- exploits Markov structure to identify critical lines for mitigation.
- uses bootstrap to estimate uncertainty.

Part 2. Bayesian hierarchical model of individual transmission line outage rates

Collaborators:
Louis Wehenkel, University of Liege Belgium
James R. Cruise, Riverlane Research England
Chris J. Dent, Amy L. Wilson, University of Edinburgh Scotland

Introduction

Problem:

- Line outage rates are fundamental to reliability, but there is limited data.

Objective:

- Get better estimates of individual line outage rates by exploiting partial dependencies between transmission lines.

Introduction

Two existing ways of estimating line outage rates.

- A straightforward way assumes independent lines:

$$\text{annual outage rate of a line} = \frac{\text{outage counts}}{\text{number of years}}$$

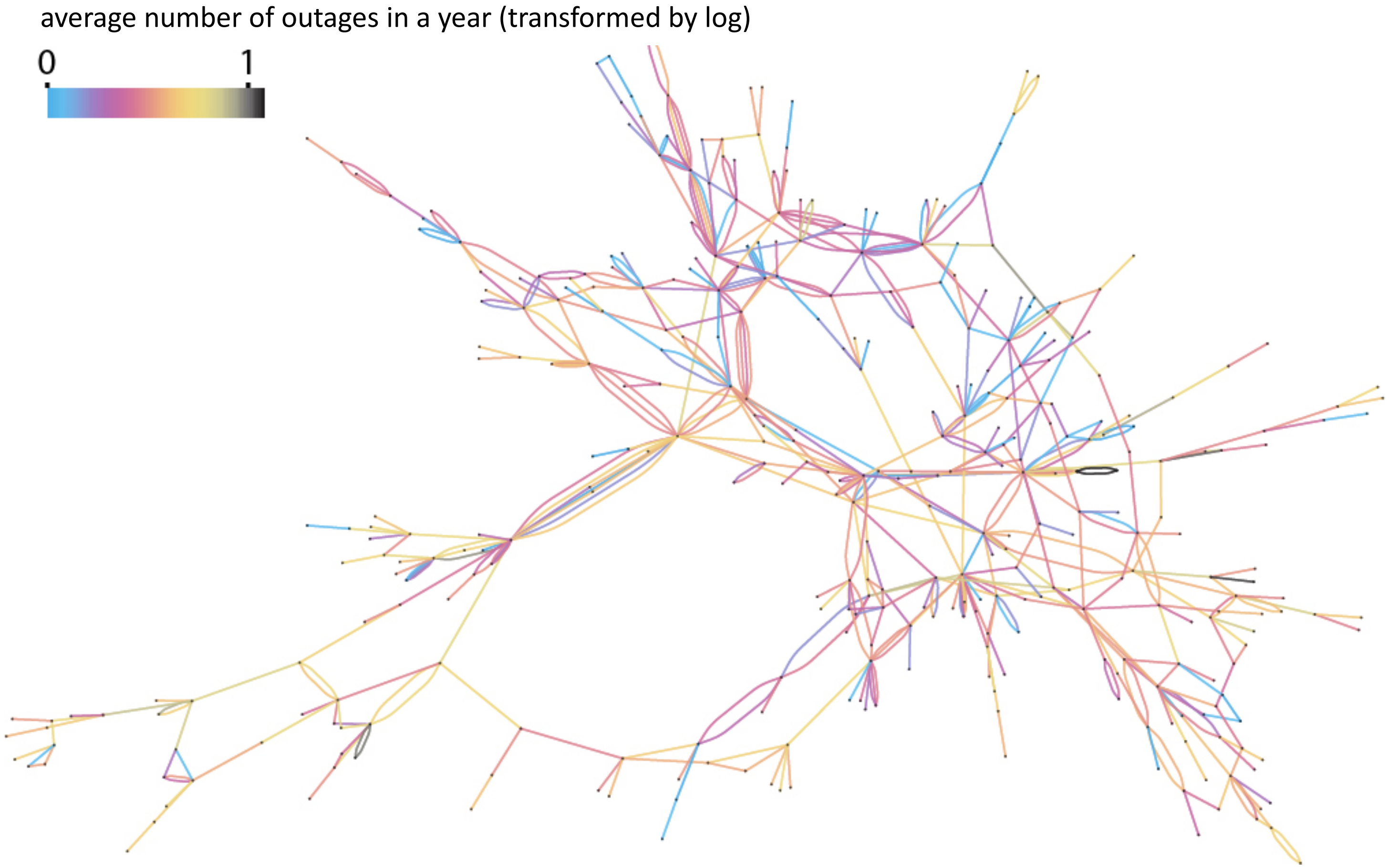
  the estimate is highly variable due to limited data.

- Group similar lines assuming perfect dependencies within a group, such as estimate a single outage rate for all 230kV lines or all lines in an area.
- The is a middle ground between the two methods.

Introduction

Line outage rates are correlated in several ways

- Length, voltage rating, geographical location, network proximity
- Modeling dependencies: adjacency matrix, Gamma field.
- Our solution is the Bayesian hierarchical model.



Introduction

Utility transmission line outage data

- Outages of 549 lines over 14 years.
- Neglect scheduled and momentary outages.

Table 1: Annual Outage Counts and Line attributes

Line	Outage counts in different years														Line attributes		
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Voltage(kV)	Length(mile)	District
29	0	0	0	0	0	0	0	0	3	2	0	0	0	0	230	8.3	P
11	0	0	1	0	0	1	0	0	1	0	0	0	0	1	500	22.65	N
2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	230	7.62	A
8	1	2	4	2	1	2	2	2	2	1	3	8	6	2	500	148.86	E

Source: Bonneville Power Administration transmission services operations & reliability website." [Online]. Available: https://transmission.bpa.gov/Business/Operations/Outages

Bayesian hierarchical model

$$N_i \sim \text{Poisson}(\lambda_i t_i), \quad i = 1, \ldots, n$$

$$\lambda_i \sim \text{Gamma}(\alpha, \alpha/\mu_i), \quad i = 1, \ldots, n$$

$$\ln \boldsymbol{\mu} = \boldsymbol{\beta}_0 + \beta_L \boldsymbol{x}_L + \beta_V \boldsymbol{x}_V$$

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(m\mathbf{1}, \sigma^2(w\boldsymbol{\Sigma}_1 + (1-w)\boldsymbol{\Sigma}_2))$$

- Considering line similarities between length x_L and voltage rating x_V.
- Covariance matrix $\sigma^2(w\boldsymbol{\Sigma}_{\mathbf{1}} + (1-w)\boldsymbol{\Sigma}_{\mathbf{2}})$ models line spatial dependencies
  - Σ_1 : districts.
  - Σ_2 : network distance.

Bayesian hierarchical model

- Weak informative priors have typical values as their means and have large variances.

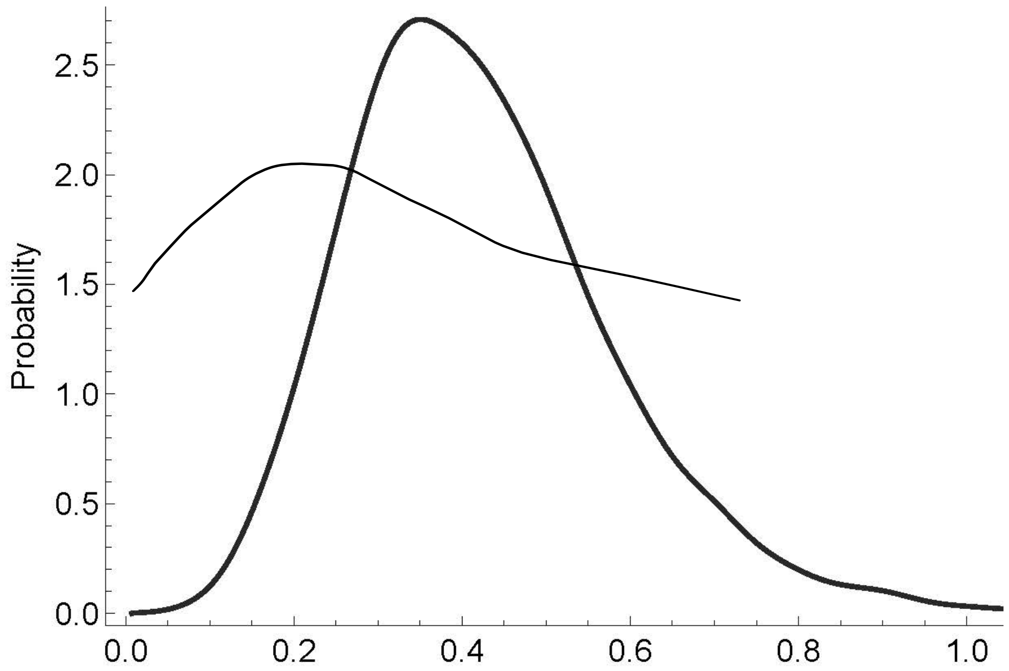
$$\alpha \sim \text{Half Normal}(0.7, 8^2) \quad \beta_L \sim \text{Normal}(0.13, 5^2)$$
$$m \sim \text{Normal}(-1.5, 5^2) \quad \beta_V \sim \text{Normal}(0.12, 5^2)$$
$$\sigma^2 \sim \text{Half Normal}(0, 0.5^2) \quad w \sim \text{Beta}(1, 1)$$

- We use Hamiltonian Monte Carlo to draw posterior distribution samples.

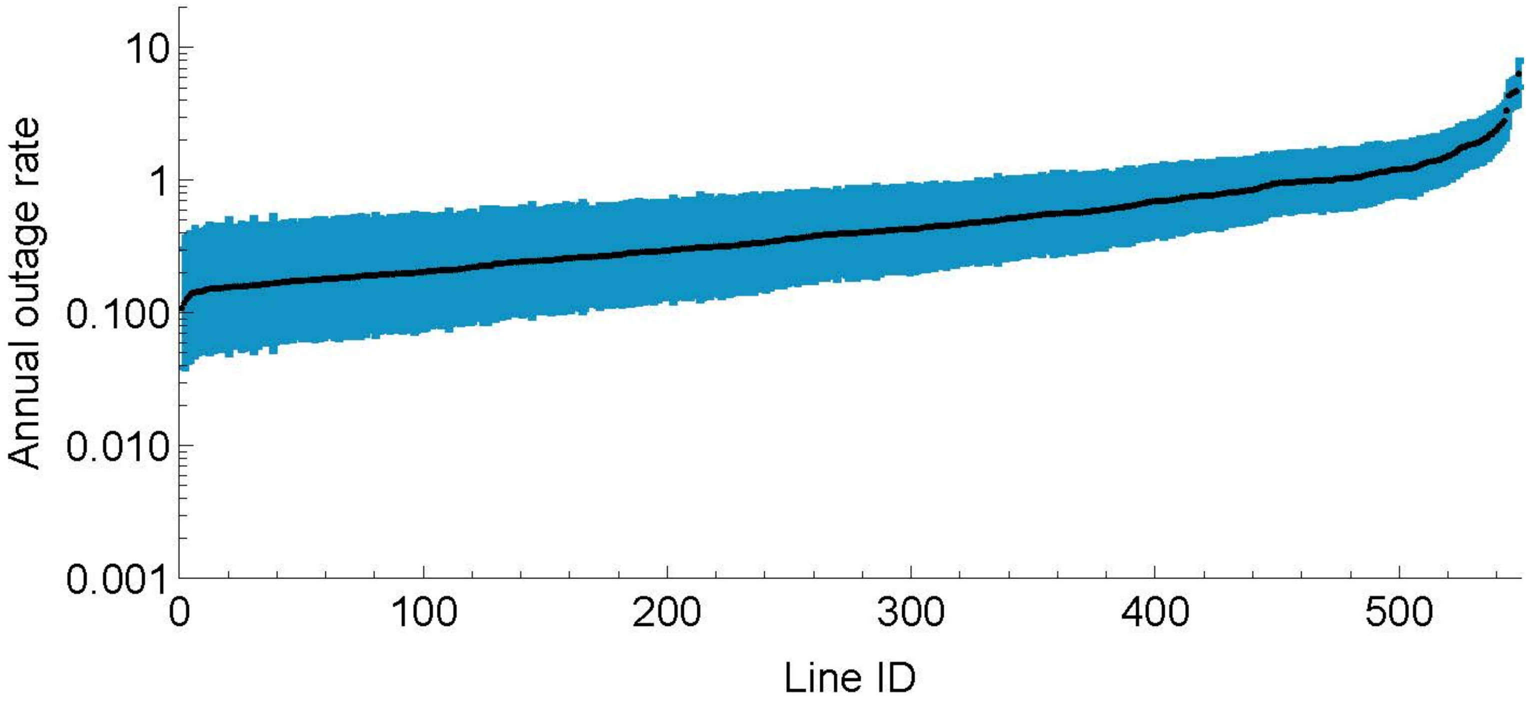
$$p(\boldsymbol{\lambda}|\boldsymbol{N}) = \frac{p(\boldsymbol{N}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})}{p(\boldsymbol{N})}$$

Bayesian model produces uncertainty of outage rates

Probability distribution of the outage rate for line number 12.



95% credible intervals (blue bars) of outage rates and posterior means (black dots).



Bayesian estimates have better individual outage rates

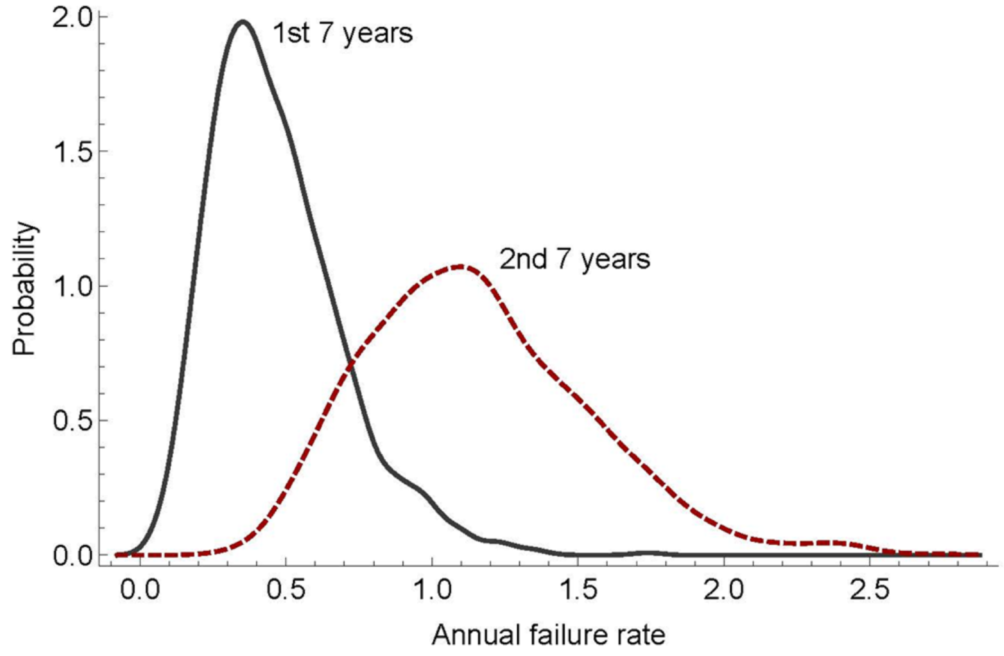
- Bayesian estimates are less variable: they effectively double the data in terms of standard deviation.
  - That is, standard deviation of Bayesian estimates using 1-year data is equivalent to that of straightforward estimates using 2-year data.

- Bayesian estimates are reasonable for rare counts.
  - For example, both Line 29 and Line 11 have 0 count in the first year, but the estimated rates are 0.32 and 0.36 because line 29 has no outages, while line 11 has 2 outages in the next 6 years.

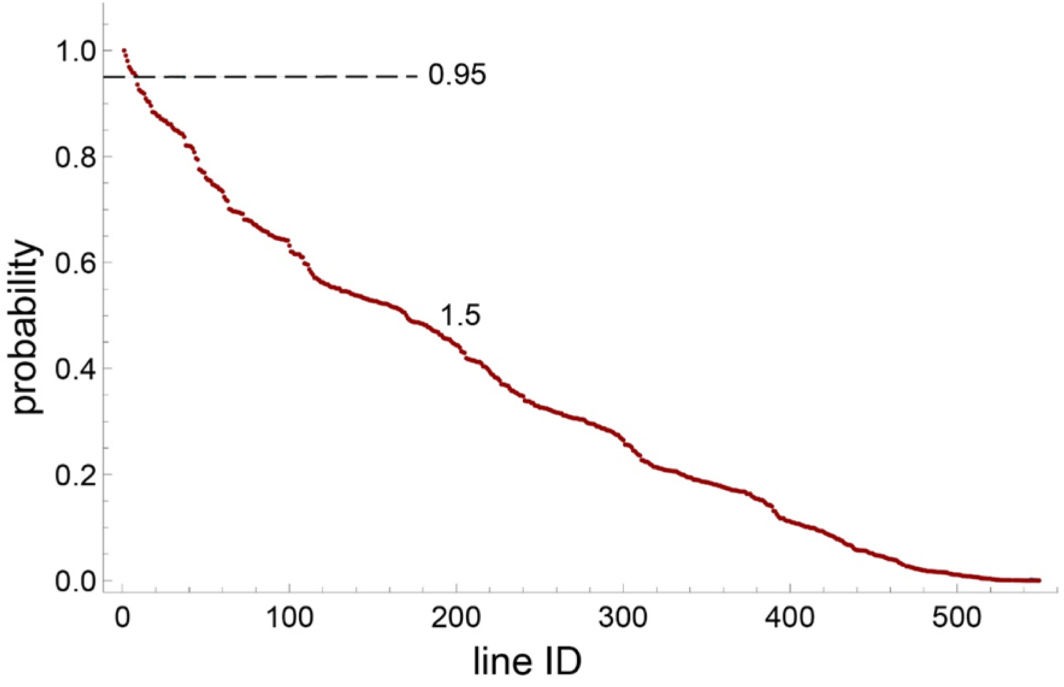
Applying Bayesian estimates of individual outage rates

Detect lines with reduced reliability

- Does the outage rate in the 2nd 7-year increase significantly?



- Check probability $p_k = \mathrm{P}[\lambda_k^{(2)} > 1.5\lambda_k^{(1)}]$
- 8 lines are identified.
- Only 1 line is identified using the straightforward method.



Conclusions

The Markovian influence graph

- uses real data observed and routinely collected by utilities.
- mitigates limited data problem by several new methods.
- obtains a clearly defined Markov chain by including multiple outages as states.
- exploits asymptotic property to identify lines most involved in large cascades.
- estimates uncertainty of cascade size distribution.

The Bayesian hierarchical model

- estimates individual transmission line outage rates.
- leverages partial similarities between lines, including proximity, length, and rated voltage.
- has estimates with a lower SD for given data, or the same SD for less data.
- benefits reliability evaluation.

Publications

Journal

[1] K. Zhou, I. Dobson, Z. Wang, A. Roitershtein, A. P. Ghosh, "A Markovian influence graph formed from utility line outage data to mitigate large cascades, " IEEE Transactions on Power Systems, vol. 35, no. 4, pp. 3224-3235, Jul. 2020.

[2] K. Zhou, J. R. Cruise, C. J. Dent, I. Dobson, L. Wehenkel, Z. Wang, A. Wilson, "Bayesian estimates of transmission line outage rates that consider line dependencies," to appear in IEEE Transactions on Power Systems, 2020. also preprint arXiv:2001.08681 [stat.AP].

Conference

[3] K. Zhou, I. Dobson, P. D. H. Hines, Z. Wang, "Can an infuence graph driven by outage data determine transmission line upgrades that mitigate cascading blackouts?," in IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Boise, ID, USA, Jun. 2018.

[4] K. Zhou, I. Dobson, Z. Wang, "Can the Markovian infuence graph simulate cascading resilience from historical outage data?" in IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Liege, Belgium, Aug. 2020.

[5] K. Zhou, J. R. Cruise, C. J. Dent, I. Dobson, L. Wehenkel, Z. Wang, A. Wilson, "Applying Bayesian estimates of individual transmission line outage rates," in IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Liege, Belgium, Aug.2020.

Future work plan

- Combining historical and simulated data
  - The Markovian influence graph driven by historical data and the model-based simulation are complementary.
  - We can combine the two approached through the influence graph by forming the Markovian influence graph from historical data and simulation data for the same system, then taking the weighted sum of the two transition matrices.
- Testing the Markovian influence graph in terms of load shed
  - The Markovian influence graph in this work measures the cascade size in number of generations as a result of limitations of real outage data.
  - We can develop and test the Markovian influence graph in terms of load shed on three simulation models.
- Exploring the spatial characteristics of cascading propagation
  - The spatial characteristics of cascading outages are not well studied. We can study the spatial patterns of cascading outages using some methods from the graph theory, such as minimum cuts and motifs.
  - Finding these patterns in cascading is very challenging, and no clear patterns may exist, but we have the data to be able to explore this.

Thank you!