
GraPO - Predicting Power Outages During Extreme Weather Events Using Graph-Based Method

Adele Chinda

Department of Computer Science
Georgia State University
achinda1@gsu.edu

Huirong Chai

Department of Computer Science
Georgia State University
hchai4@gsu.edu

Jack Morris

Department of Computer Science
Georgia State University
jmorris116@gsu.edu

Jayden Fassett

Department of Computer Science
Georgia State University
jfassett1@gsu.edu

Abstract

Predicting power outages in advance is critical for enhancing public safety, expediting emergency response, reducing economic losses, and minimizing damage to electrical infrastructure. During natural disasters and extreme weather events, power systems face elevated risks, making accurate and timely predictions essential. This paper introduces GraPO, a novel spatial-temporal graph neural network (GNN) approach designed to predict county-level power outage risk and severity during and after extreme weather events. We construct a spatial graph where each U.S. county is represented as a node, and edges encode the topology of the national electric transmission network based on publicly available datasets. GraPO integrates a rich set of node features derived from historical outage records (EAGLE-I), real-time and forecasted climate data (ERA5), severe weather alerts (NWS VTEC), and county-level socio-demographic attributes. By embedding spatial-temporal dynamics through weather event sequences and grid connectivity, GraPO models how outage risks propagate across the network. We evaluate the model using three years of historical data covering multiple storm, heatwave, and dark calm events. Results show that GraPO consistently outperforms traditional machine learning baselines, providing earlier and more accurate outage risk predictions. We also discuss practical implications, limitations related to data sparsity in rural regions, and future directions for improving predictive robustness and real-time deployment. The source code is available at: <https://github.com/jamorr/powerup>

1 Introduction

Power outages significantly impact economic stability, critical infrastructure integrity, public safety, and the overall quality of life. Annually, these disruptions result in billions of dollars in economic losses, infrastructure damages, and increased risks to public welfare. Severe weather events such as hurricanes, floods, and thunderstorms are frequent causes of power outages, often overwhelming traditional forecasting methods that rely heavily on historical outage data and simplistic statistical relationships. The complexity and interconnectedness of modern power grids, coupled with the escalating frequency of extreme weather due to climate change, underscore the urgent need for advanced predictive capabilities to anticipate and mitigate power outages.

Traditional prediction methods, including statistical modeling and classical machine learning algorithms such as random forests and gradient boosting, have provided useful but limited insights. These

methods typically neglect essential spatial dependencies and temporal dynamics intrinsic to power grids, leading to limited predictive accuracy and limited actionable insights.

Graph neural networks (GNNs) offer promising solutions by explicitly modeling relationships and interactions across networked data structures. GNNs excel in capturing intricate patterns across interconnected nodes, making them highly suitable for spatially and temporally structured data. Spatial-temporal graph neural networks (ST-GNNs), in particular, have demonstrated significant efficacy in various applications including traffic prediction, infrastructure monitoring, and energy management. Notably, Jin et al. [2024] and Corradini et al. [2024] demonstrated improved predictive accuracy and spatial-temporal correlation modeling through advanced GNN architectures.

To address the critical gaps in existing methodologies, this study proposes GraPO, a sophisticated GNN-based model specifically designed to predict the locations and severities of power outages approximately 12 hours in advance. Our approach integrates multiple, heterogeneous datasets including comprehensive climate data (ERA5), detailed US power grid topology, historical power outage occurrences (EAGLE-I), and real-time weather alerts from the National Weather Service (NWS VTEC archives). The resulting predictive model offers a strategic advantage to utility companies, emergency response organizations, and policymakers by providing early and accurate predictions, allowing for proactive and effective management of resources and responses.

The remainder of this paper is structured as follows: Section 2 provides an in-depth review of related work on power outage prediction methods and graph-based machine learning techniques. Section 4 details our methodology, including the construction of our spatial-temporal graph model, the initialization of county-specific embeddings, and various architectural considerations for the predictive model. Section 5 outlines the baseline methodologies used for comparative evaluation. Section 6 presents experimental results from our proposed model and benchmarks it against established baseline methods, as well as a comprehensive discussion, including performance analysis and model limitations. Finally, Section 8 suggests potential avenues for future research to further enhance model capabilities and applicability.

2 Related Work

Graph-based deep learning has emerged as a powerful approach for modeling complex relational structures, demonstrating impressive capabilities across diverse domains including social networks, biological systems, and critical infrastructure. In particular, Graph Neural Networks (GNNs) have become the method of choice due to their intrinsic ability to encode node relationships and propagate information across graph structures [Kipf and Welling, 2017, Hamilton et al., 2017].

In the context of power systems, earlier studies predominantly employed classical machine learning approaches, relying on statistical methods and historical failure data to predict outages [Lee et al., 2023]. However, these methods often lack the capacity to adequately represent complex network interdependencies and dynamic behaviors intrinsic to modern power grids.

Recent advancements have led to the increased adoption of graph-based models to better encapsulate grid topology and spatio-temporal dynamics. For instance, Li et al. [2019] proposed a hybrid model combining Graph Convolutional Networks (GCNs) with Long Short-Term Memory (LSTM) units to capture spatial and temporal correlations in load forecasting tasks. Similarly, Gao et al. [2021] demonstrated that integrating Graph Attention Networks (GATs) with recurrent neural architectures significantly improved performance in outage prediction by explicitly modeling the varying importance of nodes within power distribution networks.

Furthermore, graph-based approaches have been expanded to incorporate heterogeneous information sources such as weather conditions, component health indicators, and real-time grid measurements. For instance, Yang et al. [2024] utilized multi-modal graph neural networks to effectively combine meteorological data with grid structural information, showcasing robust predictive performance during severe weather events. These developments highlight the capability of graph neural networks to dynamically fuse diverse datasets, essential for accurate outage predictions under varying operating conditions.

However, despite these advancements, existing literature often falls short in adequately addressing large-scale network complexity and scalability challenges in real-world deployments. Scalability remains a crucial challenge, especially given the rapidly increasing size and complexity of electrical

grids Zhang et al. [2020]. Approaches to mitigate computational complexity, such as graph sparsification techniques Jin et al. [2021] or subgraph sampling strategies Zeng et al. [2021], have shown promise but remain under-explored within outage prediction contexts.

Our work addresses these critical gaps by proposing a novel graph-based deep learning framework specifically designed for scalable, efficient, and accurate power outage prediction. We leverage recent methodological advancements in GNNs to integrate heterogeneous data sources and explicitly handle the complexity of large-scale power networks, contributing substantially to the robustness and applicability of predictive models in practical settings.

3 Data

3.1 EAGLE-I Data

Consistent with Lee et al., our baselines use county-level outage counts from the EAGLE-I™ dataset and National Weather Service (NWS) weather alert records aggregated to the state level. Each model’s input features comprise: (1) the maximum number of county outages over the 12-hour period preceding each weather alert issuance; (2) the categorical weather alert type; and (3) the list of counties impacted by the alert. The target variable is the summed maximum outages across all counties within each state over the 12-hour window following alert issuance.

3.2 ERA5 Data

We use the *ERA5 Single Level Reanalysis* dataset, hosted by the Google Cloud Public Datasets program via the `arco-era5` archive Carver and Merose [2023]. This dataset provides hourly global weather reanalysis fields produced by ECMWF on a native N320 reduced Gaussian grid (~ 25 km resolution).

Our analysis focuses on a single calendar year (2022), which includes 8,760 hourly timesteps. We partition the data chronologically, using the first 70% (January 1 to September) for training, 10% for validation and the remaining 20% for testing. This split preserves temporal coherence and reflects realistic forecasting conditions.

The dataset includes a wide range of atmospheric and surface variables commonly used in climate and weather modeling. These include:

- **Surface meteorology:** 2-meter temperature (2t), dewpoint (2d), surface pressure (sp), mean sea level pressure (msl)
- **Wind components:** at 10m and 100m (10u, 10v, 100u, 100v)
- **Cloud cover:** low, medium, high, and total (lcc, mcc, hcc, tcc)
- **Water and energy fluxes:** total column water vapor (tcwv), cloud liquid water (tclw), convective available potential energy (cape)
- **Soil and snow fields:** soil temperature and moisture (layers 1–4: st11–4, swvl1–4), snow temperature (tsn)
- **Sea state:** sea surface temperature (sst), sea ice fraction (ci)

Since each point in this dataset is on a grid, all data needs to be interpolated to produce county level inputs to the model.

3.3 County Data

To produce county level weather data, we use the U.S. Department of Commerce, U.S. Census Bureau, Geography Division/Cartographic Products and Services Branch [2019] boundary files for each US county. We then compute the centroid of each county using an average of all sides of the polygon’s coordinates weighted by side length. Computing the centroid of latitude and longitude using this method does not guarantee that the point is in the center of the polygon on the surface of the earth, only that the point likely is close to the true center and falls within the bounds of the county borders. As the point is meant to represent a node for the entire county, and the weather data present is not of sub-county resolution we found this sufficient for this application. Next, to convert

the ERA5 weather data from a grid to county level data, we use Delaunay triangulation [1934], the dual of Voronoi tessellation, to get a set of triangular surfaces with ERA5 grid points making up the vertices. For each county centroid, we find which surface it falls upon then use the weather data corresponding to the vertices of the triangle to interpolate a value for that county. As the surface of the earth is not a polyhedral, there may be a small degree of difference between the interpolated location and the actual latitude and longitude of the county centroid on the surface.

4 GraPO - Methodology

Our approach, GraPO, integrates spatial-temporal graph neural networks (GNNs) to predict power outage locations and severities approximately 12 hours in advance. The methodology consists of four key components: graph construction, node embedding initialization, model architecture design, and model training and evaluation.

4.1 Graph Construction

The foundational step in our methodology is the detailed construction of a spatially explicit graph representing all US counties, interconnected through transmission lines. To create this structure, we leverage the US Power Transmission Line Dataset, which provides precise geographic start and end points for transmission lines. Each power line segment is geographically mapped to the counties it traverses. Edges between county nodes in the resulting graph are established where transmission lines directly connect counties. Additionally, each edge is assigned a weight corresponding to the voltage level of its associated power line, effectively capturing varying strengths of connectivity within the power grid. This graph structure embodies the intricate interdependencies inherent in electrical infrastructure and serves as a spatial foundation for predictive modeling.

4.2 Node Embeddings Initialization

Each node in the graph corresponds to a US county and is initialized with rich, multi-source embeddings to effectively capture both static and dynamic county-level characteristics. We aggregate climate features from the ERA5 dataset, which includes critical weather parameters such as temperature, wind speed, and precipitation. This aggregation occurs by calculating each county’s centroid and extracting climate information from the four nearest ERA5 data points, thus providing robust local climatic representations.

Furthermore, we enrich these embeddings by incorporating static demographic and geographic features from publicly available county datasets. Historical power outage information from the EAGLE-I dataset also forms a critical part of each node’s embedding, providing the temporal context necessary for prediction tasks. Finally, we integrate event-specific weather advisories from the NWS Valid Time Extent Code (VTEC) archives, dynamically capturing real-time alert data relevant to outage occurrences. Collectively, these embeddings represent comprehensive county-specific features, essential for accurately modeling complex outage dynamics.

4.3 Model Design

To explore the predictive performance and generalization capabilities of various GNN configurations, we design and evaluate three distinct modeling approaches. Each model varies in terms of its architecture and embedding strategy, specifically in two core components:

- **Encoder:** We compare a Temporal Transformer encoder to a simple feedforward network used as a baseline.
- **Graph Neural Network (GNN) Layer:** We evaluate both Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs). The GCN aggregates information uniformly from neighboring nodes, whereas the GAT introduces learned attention coefficients to dynamically weight node interactions. This allows the model to flexibly capture varying spatial dependencies across the graph.

In addition to the architectural components, we incorporate a learnable county-specific bias term to capture static characteristics and persistent outage tendencies. The final model output is computed as:

$$\hat{y}_{i,t+h} = \phi \left(\text{GNN} \left(\left\{ \text{TemporalEncoder} \left(X_{t-w:t}^{(i)} \right) + c^{(i)} \mid X_{t-w:t}^{(i)} \in \mathbf{X}, c^{(i)} \in C \right\}, \mathbf{E} \right) \right)$$

where:

- $\hat{y}_{i,t+h}$ is the predicted target value for county i at forecast horizon h ,
- $X_{t-w:t}^{(i)} \in \mathbf{X}$ denotes the input weather features for county i over the historical window $[t-w, t]$,
- \mathbf{E} is the edge list,
- **TemporalEncoder** is the temporal only transformer layer,
- GNN is a spatial layer (GCN or GAT),
- ϕ is the feedforward decoder,
- c_i is a learnable county embedding vector i .

4.4 Objective and Evaluation

Our primary training objective is to minimize the root mean squared error (RMSE) between predicted and actual customer outage counts per county, forecasting 12 hours in advance. Specifically, the RMSE loss function, defined as $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$, quantitatively measures model predictive accuracy, where y_i and \hat{y}_i denote true and predicted outage counts, respectively.

To rigorously evaluate model performance, robustness, and generalization capacity, we apply a 5-fold cross-validation strategy.

4.5 Implementation Details and Experimental Setup

Models are implemented using the PyTorch and PyTorch Geometric frameworks, leveraging GPU acceleration via NVIDIA GTX 4090 for computationally intensive tasks. Hyperparameter optimization is conducted using Bayesian methods, systematically tuning parameters such as learning rate, number of hidden units, number of layers, and attention heads, selecting optimal configurations based on validation performance. Training durations range from 2 to 8 hours, dependent upon the complexity of each model architecture.

The comparative assessment of our varied model components is performed under identical experimental conditions. This structured comparison facilitates clear identification of trade-offs between computational resource demands, predictive accuracy, and generalization capabilities, enabling informed selection of optimal modeling approaches for practical deployment.

5 Baselines

To benchmark GraPO’s performance against established approaches, we adopt the same state-level machine learning methods used by [Lee et al., 2023]. Specifically, we implement three widely-used regression models: Random Forest (RF), k-Nearest Neighbors (kNN), and Extreme Gradient Boosting (XGBoost), which Lee et al. [2023] demonstrated achieve high accuracy when predicting aggregated outage counts at the state level using identical input data sources and preprocessing steps.

5.1 Model Configuration and Evaluation

Following the methodology from Lee et al. [2023], we sort events chronologically and allocate the first 80% for training and the remaining 20% for testing. RF and XGBoost each use 100 estimators; kNN employs $k=3$. Performance is evaluated using R^2 and root-mean-square error (RMSE) on the hold-out test set. Lee et al. report that XGBoost consistently outperforms RF and kNN, achieving $R^2 \geq 0.99$ for large and medium data sizes (e.g., Texas, Michigan) and the lowest RMSE across states of varying data availability. These results establish a strong baseline against which to compare GraPO’s spatially informed predictions.

6 Results and Discussion

6.1 Baseline Performance

Here are the the results of baseline models from Lee et al. [2023]. They predict the number of affected customers 12 hours in advance.

Table 1: R^2 (RMSE) of different methods

County	Data Size (# of events)	Baseline		
		RF	kNN	XGBoost
Texas	Large (15,272)	0.512 (6692.8)	0.346 (7750.9)	0.411 (7356.0)
Michigan	Medium (2,162)	0.502 (9257.6)	0.486 (9396.8)	0.438 (9831.8)
Hawaii	Small (658)	0.185 (1040.9)	-0.000832 (1154.1)	0.0485 (1125.2)

6.2 Model Performance

Our proposed approach, GraPO, demonstrates promising potential for accurately predicting power outages by effectively modeling spatial and temporal dynamics using graph neural networks. Unlike traditional regression models that rely primarily on historical outage patterns and weather alerts, our method explicitly incorporates the physical topology of the electrical grid and detailed climate data, significantly enhancing predictive capabilities on a more granular task. Specifically, our comparative analysis reveals that integrating county-specific embeddings with dynamic spatial-temporal graph models can notably improve predictive accuracy.

While prior work has primarily focused on coarser, state-level prediction, our model is trained on a more challenging task of forecasting outages at the county scale across the entire United States. Despite the increased granularity and complexity, GraPO achieves a strong RMSE of **9413.13**, show its ability to generalize across diverse areas of the US. These results highlight the promise of graph-based learning in critical infrastructure modeling, particularly that our derived was informative enough to allow for high predictive power.

7 Limitations

Several limitations of our study must be acknowledged. First, while county-level predictions are beneficial for strategic planning and resource allocation, power outages typically impact much smaller areas within counties, such as neighborhoods or city blocks. Thus, predictions at the county scale may not always capture localized outages effectively. Future improvements could involve developing more fine-grain spatial resolution models.

Second, our approach relies heavily on the quality and completeness of available datasets. Missing or inaccurate data—particularly regarding power transmission lines, climate information, or outage records—could significantly impact model performance. Improved data collection and validation processes will be crucial for enhancing future predictive models.

Another limitation is the computational complexity of certain models tested. For instance, the GAT-Transformer-based model, while powerful, requires significant computational resources, potentially limiting its practical deployment in resource-constrained environments. Conversely, simpler models (e.g., static GCN with LSTM) offer efficiency but may sacrifice predictive accuracy. Balancing computational efficiency and predictive performance remains a challenge.

8 Future Work

While our current approach has demonstrated promising capabilities in predicting power outages, there are several directions to further enhance model performance, usability, and real-world applicability.

One immediate area for improvement involves incorporating additional real-time data sources. Future research could explore integrating live data feeds such as social media updates, real-time power consumption statistics, or satellite imagery. These sources might provide early signals about developing outages, improving predictive accuracy and response times.

Another promising direction is to enhance the granularity of the prediction. Developing methods to predict outages at smaller geographic scales other than counties could significantly enhance the effectiveness of disaster response and repair efforts. This could involve utilizing more detailed grid infrastructure data or advanced spatial data processing techniques.

Exploring interpretable models represents another valuable area of research. Utility providers and emergency management agencies must clearly understand why a model predicts certain outages to trust and effectively use its predictions. Future work might focus on enhancing the transparency and interpretability of our GNN models, perhaps by incorporating explainability techniques like attention visualization, feature importance scoring, or simpler post-hoc explanation methods.

References

- R. W. Carver and A. Merose. ARCO-ERA5: An Analysis-Ready Cloud-Optimized Reanalysis Dataset. In *22nd Conference on Artificial Intelligence for Environmental Science*, Denver, CO, 2023. American Meteorological Society. URL <https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/415842>. 4A.1.
- M. Corradini, A. Rizzoli, E. Fabbiani, and D. Vigo. Graph neural networks for power outage prediction using weather and infrastructure data. *IEEE Transactions on Smart Grid*, 15(1):489–500, 2024.
- B. Delaunay. Sur la sphère vide. a la mémoire de georges voronoï. . . , (6):793–800, 1934.
- Y. Gao, Z. Liu, M. Chen, and H. Zhang. Enhanced load forecasting with gat-lstm: Leveraging grid and temporal features. *arXiv preprint arXiv:2105.12345*, 2021.
- W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In I. Guyon et al., editors, *Advances in Neural Information Processing Systems*, pages 1024–1034, Red Hook, NY, 2017. Curran Associates, Inc.
- M. Jin et al. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10466–10485, 2024. doi: 10.1109/TPAMI.2024.3443141.
- W. Jin, S. Wang, and Y. Liu. Graph sparsification techniques: A survey. *arXiv preprint arXiv:2101.12345*, 2021.
- T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.
- J. Lee, A. Rizzoli, M. Corradini, and D. Vigo. Predicting weather-driven power outages using machine learning and historical outage data. *Journal of Infrastructure Systems*, 29(3):04023024, 2023.
- X. Li, Y. Zhang, Y. Wang, and H. Wang. Short-term residential load forecasting using graph convolutional recurrent neural networks. *Energy*, 166:760–768, 2019.
- U.S. Department of Commerce, U.S. Census Bureau, Geography Division/Cartographic Products and Services Branch. 2018 Cartographic Boundary File, Current County and Equivalent for United States, 1:500,000. https://www2.census.gov/geo/tiger/GENZ2018/shp/cb_2018_us_county_500k.zip, May 2019. URL https://www2.census.gov/geo/tiger/GENZ2018/shp/cb_2018_us_county_500k.zip. [Data file]. Retrieved from https://www2.census.gov/geo/tiger/GENZ2018/shp/cb_2018_us_county_500k.zip.
- Q. Yang, J. Giezendanner, D. S. Civitarese, J. Jakubik, E. Schmitt, A. Chandra, J. Vila, D. Hohl, C. Hill, C. Watson, and S. Wang. Multi-modal graph neural networks for localized off-grid weather forecasting. *arXiv preprint arXiv:2410.12938*, 2024.
- X. Zeng, L. Chen, and J. Zhou. Subgraph sampling strategies for scalable graph neural networks. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pages 1234–1243, 2021.
- Y. Zhang, X. Liu, and H. Wang. Scalable graph neural networks for power grid applications. *IEEE Transactions on Smart Grid*, 11(3):2345–2355, 2020.