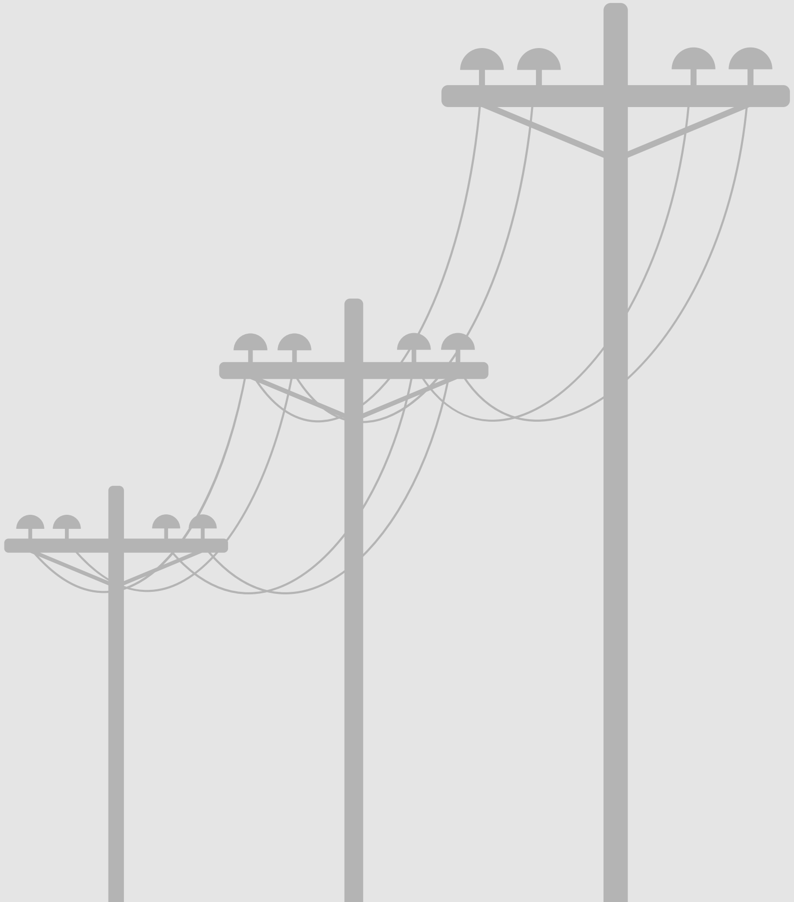
PFΔ: A Benchmark Dataset for Power Flow



Anvita Bhagavathula*,
Alvaro Carbonero*,
Ana K. Rivera*,
Priya L. Donti

Massachusetts Institute of Technology

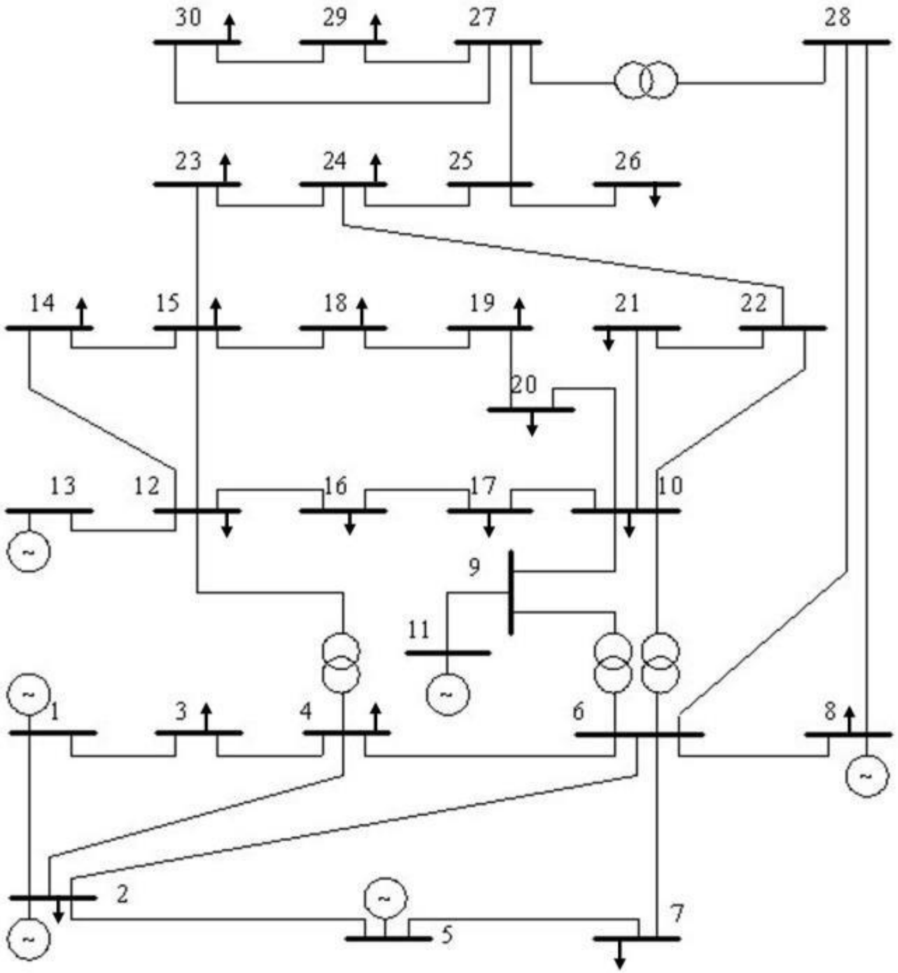
Climate change affects electric grids

Keeping the lights on

Grid operators constantly analyze several different scenarios to guarantee a stable grid.

Climate change’s impact

More weather anomalies and large amounts of renewable energy greatly increase the number of scenarios to be analyzed.



Power Flow

- Power into and out of each bus must be balanced:

$$\mathrm{diag}(v)\bar{Y}\bar{v} = (p_g - p_d) + (q_g - q_d)j$$

- We only know a fraction of the voltages and power values. What are the remaining values?

Machine learning can help

We need more speed

Power flow is usually solved using Newton Raphson (NR). While accurate, NR is not fast enough in very large grids.

ML for power flow

To create a machine learning proxy to NR, we need a benchmark that ranks proposed architectures with realistic scenarios and field-relevant metrics.

Related Work: Power flow

Model	Dataset
Graph Neural Solver	Varying load/line characteristics/grid sizes. Largest grid 6k. No N-1 perturbations.
TypedGNN	Varying load/line characteristics/N-1 perturb./generator set point. Largest grid: 118.
PowerFlowNet	Varying load/generator set point/grid size. Largest grid: 6k. No N-1 perturbations.

All load variations are drawn from a uniform distribution around a base case.

Related Work: Optimal Power Flow

Dataset	Description
OPFData	Varying load/N-1 perturbations. Largest grid: 13k. No line characteristic perturbations. **Load drawn from a uniform distribution.**
OPFLearn	Varying load perturbations. Largest grid: 118. **Grid fixed**. **Space of feasible loads is approximated and drawn from.**

Space of feasible loads approx. -> high diversity in values -> realistic.

Power Flow Δ: A dataset

Dataset	Load Profile	Generator Profile	Grid Sizes	N-1	> 1000 Buses
OPFData	✓	×	✓	✓	✓
OPFLearn	✓	×	✓	×	✓
GraphNeuralSolver	✓	✓	✓	×	×
PowerFlowNet	✓	✓	✓	×	✓
TypedGNN	✓	✓	✓	✓	×
PFΔ	✓	✓	✓	✓	✓

We also provide grid perturbations and achieve high feature diversity.

Power Flow Δ: A benchmark

Standardized tasks

We provide multiple tasks in which models can be ranked and compared, including in data-efficient and out-of-distribution settings.

Model comparison

We implement multiple architectures and compare them on metrics capturing physical feasibility and speed.

Data Generation Process

Load perturbation

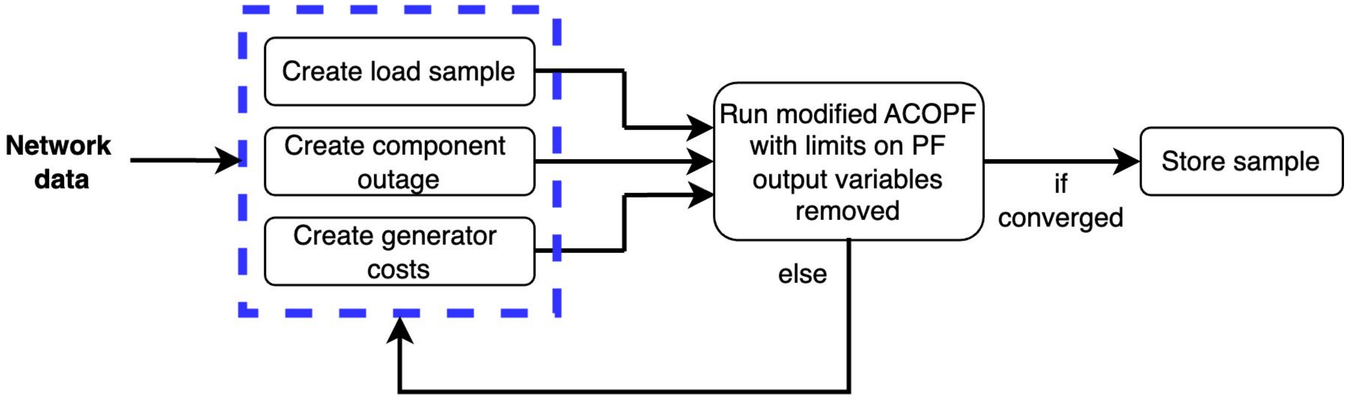
±20% of the nominal load

Topological perturbation

N-1 outages (branches or generators)

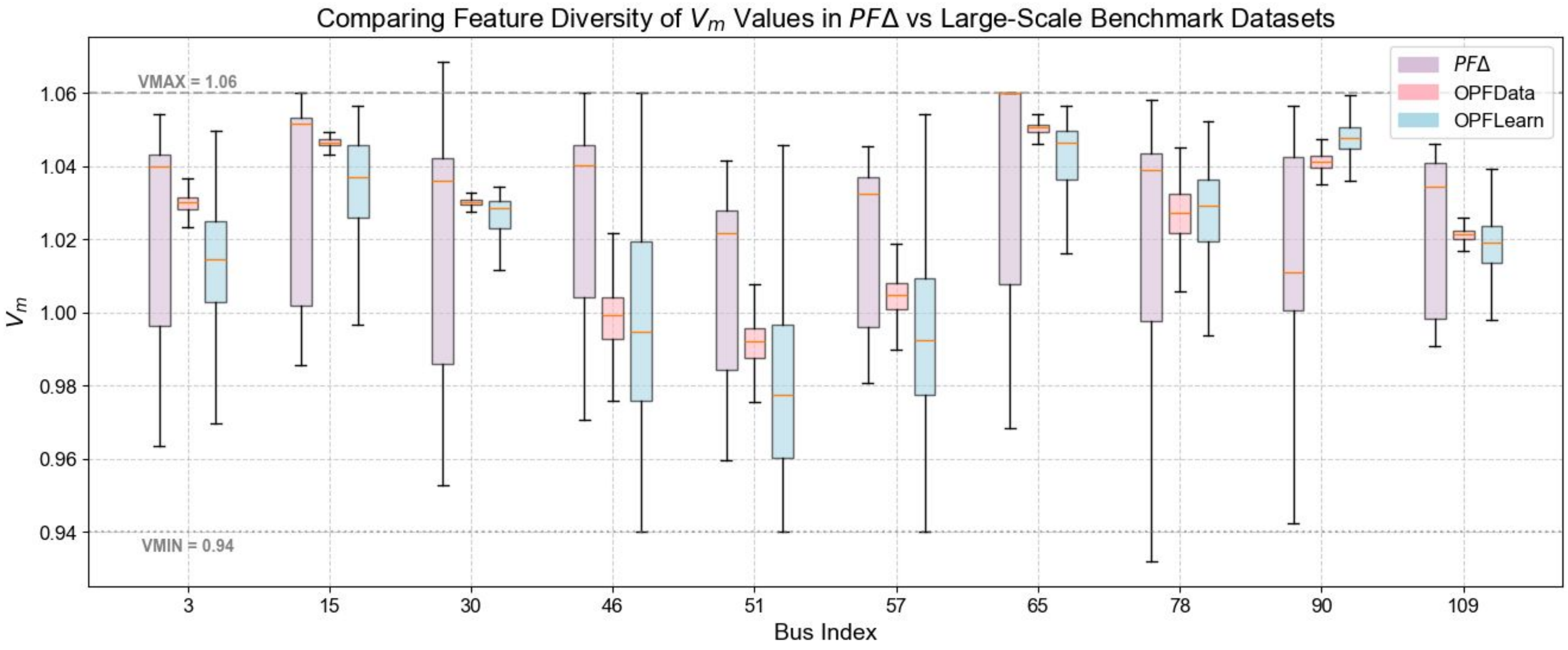
Generator cost perturbation

Permute cost parameters



Data generation process for a single sample within PFΔ

Feature Diversity



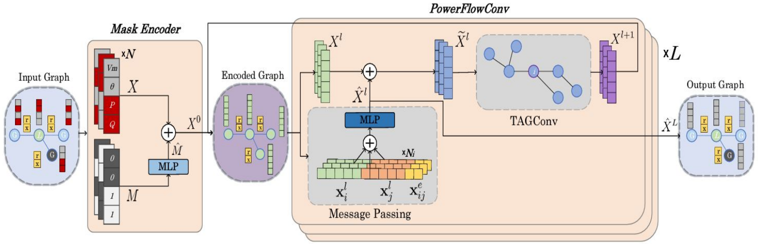
IEEE 118-bus system

Tested Models

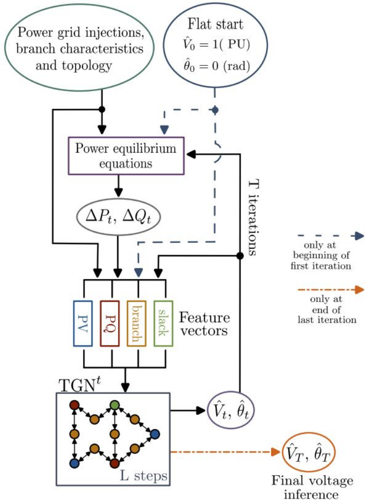
Newton-Raphson[1]

MATPOWER

PowerFlowNet[2]



TypedGNN[3]



[1] Ray D Zimmerman and Carlos E Murillo-S´ anchez. MATPOWER, May 2024.
[2] Nan Lin, Stavros Orfanoudakis, Nathan Ordonez Cardenas, Juan S. Giraldo, and Pedro P. Vergara. Powerflownet: Power flow approximation using message passing graph neural networks. International Journal of Electrical Power Energy Systems, 160:110112, 2024. ISSN 0142-0615. doi: https://doi.org/10.1016/j.ijepes.2024.110112.
[3] Tania B. Lopez-Garcia and Jos´ e A. Dom´ ınguez-Navarro. Power flow analysis via typed graph neural networks. Engineering Applications of Artificial Intelligence, 117:105567, 2023. ISSN 0952- 1976. doi: https://doi.org/10.1016/j.engappai.2022.105567.

Model Performance Metrics

Power Balance Loss

$$\mathrm{Mean}_{\mathrm{sample}}(\Delta S) = \frac{1}{N}\sum_{i=1}^{N} |\Delta S_i|$$

$$\mathrm{L2}_{\mathrm{sample}}(\Delta S) = \sqrt{\sum_{i=1}^{N} (\Delta S_i)^2}$$

$$\mathrm{Max}_{\mathrm{sample}}(\Delta S) = \max_{i=1}^{N} |\Delta S_i|$$

ΔS_i is the power mismatch for the i-th bus

Runtime

$$\frac{\sum_{\mathrm{batches}}(\text{runtime on batch} \times b)}{\text{total samples}}$$

where b is the batch size

GNNs are run in batches whereas MATPOWER is run sequentially (b=1)

Standardized Evaluation Tasks

Data Efficiency

Echo scenarios with limited historical data or distribution shifts due to evolving grid conditions.

OOD Generalization

Capture cases where models are deployed on unfamiliar grid sizes, such when a grid expands.

Results

Table 1: Comparison of MATPOWER's Newton-Raphson (NR) solver, PowerFlowNet (PFNet), and TypedGNN (TGNN) on 3000 test samples, across various bus system sizes. GNN model performance is evaluated in low, medium, and high data efficiency regimes (LDE, MDE, HDE).

Model	**PBL (118)**		**PBL (30)**		**PBL (2000)**		**Runtime(s)**		
	Mean	*Max*	*Mean*	*Max*	*Mean*	*Max*	*118*	*30*	*2000*
NR (Warm-Start)	3.3e-6	9.5e-5	1.1e-6	1.6e-5	4.7e-4	8.6e-1	1.3e-2	1.2e-2	3.0e-2
NR (Flat-Start)	3.3e-6	9.5e-5	1.1e-6	1.6e-5	Did not converge		1.5e-2	1.3e-2	N/A
PFNet (LDE)	1.2e0	1.2e2	2.2e0	1.1e2	1.3e3	9.2e5	7.7e-3	7.6e-3	8.4e-3
PFNet (MDE)	1.3e0	1.2e2	5.6e0	6.0e1	1.7e3	5.9e5	7.6e-3	7.5e-3	7.7e-3
PFNet (HDE)	1.3e0	1.5e2	1.0e1	1.4e2	1.1e3	7.3e5	7.4e-3	7.6e-3	7.9e-3
TGNN (LDE)	**2.2e-1**	1.5e1	7.1e-1	4.6e0	1.1e1	1.3e2	6.8e-3	7.1e-3	7.4e-3
TGNN (MDE)	2.5e-1	**1.4e1**	7.9e-1	5.8e0	1.2e1	**1.2e2**	**6.4e-3**	**6.4e-3**	**6.6e-3**
TGNN (HDE)	5.4e-1	2.1e1	**5.0e-1**	**2.3e0**	**9.9e0**	1.7e2	7.0e-3	7.2e-3	7.4e-3

Key Insights & Open Problems

ML models outperform traditional solvers in terms of runtime.

Producing feasible solutions remains an issue.

Self-supervised methods outperform supervised methods.

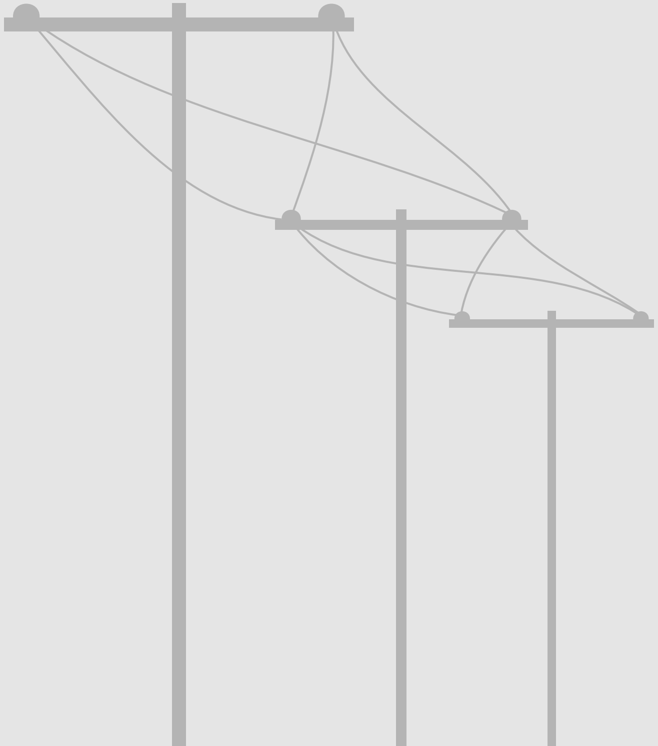
Models show better out-of-distribution generalization to smaller grids than larger ones.

Improving Grid Scenario Diversity

- Include broader range of feasible load profiles
- Including “hard” power flow cases
- Topological perturbations to include *N-k* contingencies

• • • •

Pathway to Impact



- Standardized evaluation framework to support grid operators and planners
- Facilitate integration of high-performing solutions into open-source and commercial power system software