C Access to PowerGraph Dataset

C.1 Dataset documentation and intended uses

PowerGraph is the collection of the following GNN datasets: UK, IEEE24, IEEE39, IEEE118 power grids. We use `InMemoryDataset` [27] class of Pytorch Geometric, which processes the raw data obtained from the Cascades [61] simulation. For each dataset UK, IEEE24, IEEE39, and IEEE118, we provide a folder containing the raw data organized in the following files for node-level tasks, i.e., power flow and optimal power flow analyses:

- `edge_attr.mat`: edge feature matrix for the power flow problem
- `edge_attr_opf.mat`: edge feature matrix for the optimal power flow problem
- `edge_index.mat`: branch list for the power flow problem
- `edge_index_opf.mat`: branch list for the optimal power flow problem
- `X.mat`: node feature matrix for the power flow problem
- `Xopf.mat`:node feature matrix for the optimal power flow problem
- `Y_polar.mat`: node output matrix for the power flow problem
- `Y_polar_opf.mat`: node output matrix for the optimal power flow problem

For graph-level tasks, i.e., cascading failure analysis:

- `blist.mat`: branch list also called edge order or edge index
- `of_bi.mat`: binary classification
- `of_reg.mat`: regression labels
- `of_mc.mat`: multi-class labels
- `Bf.mat`: node feature matrix
- `Ef.mat`: edge feature matrix
- `exp.mat`: ground-truth explanation

C.2 Download Dataset

The dataset can be viewed and downloaded by the reviewers from https://figshare.com/articles/dataset/PowerGraph/22820534 (node-level ∼1.08GB and graph-level ∼2.7GB, when uncompressed):

Node-level data:

```
#!/bin/bash
wget -O data.tar.gz "https://figshare.com/ndownloader/files/46619152"
tar -xf data.tar.gz
```

Graph-level data:

```
#!/bin/bash
wget -O data.tar.gz "https://figshare.com/ndownloader/files/46619158"
tar -xf data.tar.gz
```

C.3 Author statement

The authors state here that they bear all responsibility in case of violation of rights, etc., and confirm that this work is licensed under the CC BY 4.0 license.

C.4 Hosting, Licensing, and Maintenance Plan

The code to obtain the PowerGraph dataset in the `InMemoryDataset` [27] format and to benchmark GNN and explainability methods is available as a public GitHub organization at https://github.com/PowerGraph-Datasets/. The authors are responsible for updating the code in case issues are raised and maintaining the datasets. We aim to extend the PowerGraph with new datasets and include additional power grid analyses, including solutions to the unit commitment problem. Over time, we plan to release new versions of the datasets and provide updates to the results for both the GNN accuracy and the explainability analysis. In addition,

the code will be updated if new pytorch/torch-geometric versions are released or crucial python packages are updated. The data is hosted on figshare at https://figshare.com/articles/dataset/PowerGraph/22820534. The authors give public free access to the PowerGraph dataset. The datasets are identified with the DOI:10.6084/m9.figshare.22820534. The work in this paper (code, data) is licensed under the CC BY 4.0 license.

C.5 Code Implementation

We run a hyper-parameters grid search over different GNN models, using torch-geometric 2.3.0 [27] and Torch 2.0.0 with CUDA version 11.8 [62, 63]. The benchmark node and graph classification and regression models experiments are performed on the GPU nodes of the ETH Euler clusters [64]. For the explainability analysis, experiments are conducted on 8 AMD EPYC 7742 CPUs with a memory of 5GB each on the ETH Euler clusters [64].