TOWARDS PREDICTING DYNAMIC STABILITY OF POWER GRIDS WITH GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

To mitigate climate change, the share of renewable energies in power production needs to be increased. Renewables introduce new challenges to power grids regarding the dynamic stability due to decentralization, reduced inertia and volatility in production. However, dynamic stability simulations are intractable and exceedingly expensive for large grids. Graph Neural Networks (GNNs) are a promising method to reduce the computational effort of analyzing dynamic stability of power grids. We provide new datasets of dynamic stability of synthetic power grids and find that GNNs are surprisingly effective at predicting highly non-linear targets from topological information only. We show that large GNNs outperform GNNs from previous work as well as as handcrafted graph features and semi-analytic approximations. Further, we demonstrate GNNs can accurately identify *trouble maker*-nodes in the power grids. Lastly, we show that GNNs trained on small grids can perform accurately on a large synthetic Texan power grid model, which illustrates the potential of our approach.

1 INTRODUCTION

Adaption to and mitigation of climate change jointly influence the future of power grids: 1) Mitigation of climate change requires power grids to be carbon-neutral, with the bulk of power supplied by solar and wind generators. These are more decentralized and have less inertia than traditional power generators and their production is more volatile. Hence, sustainable power grids operate in different states and the frequency dynamics need to explored in more detail. Renewable energies will have to start contributing to the dynamical stability of the system (Milano et al., 2018; Christensen et al., 2020) in the future, requiring a new understanding of complex dynamics. 2) A higher global mean temperature increases the likelihood as well as the intensity of extreme weather events such as hurricanes or heatwaves (Field et al., 2012; Pörtner et al., 2022) which result in great challenges to power grids. Building a sustainable grid as well as increasing the resilience of existing power grids towards novel threats are challenging tasks on their own. Tackling climate change in the power grid sector calls for a solution to both at the same time and requires for new methods to investigate the dynamic stability.

Power grids are complex networks, consisting of nodes that represent different actors such as producers, consumers or transformers and edges that represent power lines. In contrast to many other networks, the interaction of nodes and edges can be described by underlying physical properties. Power grids may behave unintuitively. For example, the Braess paradoxon describes the phenomenon that adding lines to a power grid may actually reduce its stability ((Witthaut & Timme, 2012; Schäfer et al., 2022)). Those phenomenons can be non-local, e.g. the less stable parts might be far away from the added line, which sets another challenge and requires appropriate modeling to adequately consider all relevant interactions. Our work deals with the important issue of predicting the dynamic stability of the grid after non-linear perturbations.

When simulating frequency dynamics after perturbations, grid operators are currently limited to analyze individual contingencies in the current state of the transmission grid only. Transmission grids transport electricity at high voltages across large distances. Distributed renewable generation, however, is connected at lower grid levels. These are currently only assessed statically, i.e. by studying flow patterns induced in the networks, which can not reveal dynamic vulnerabilities. Conducting high-fidelity simulations of the whole dynamic hierarchy of the power grid and exploring

all possible states is not feasible (Liemann et al., 2021). For future power grids an understanding of how to design robust dynamics is required. This has led to a renewed interdisciplinary interest in understanding the collective dynamics of power grids (Brummitt et al., 2013), with a particular focus on the robustness of the self-organized synchronization mechanism underpinning the stable power flow (Rohden et al., 2012; Motter et al., 2013; Dörfler et al., 2013) by physicists and control mathematicians Witthaut et al. (2022).

To understand which structural features impact the self-organized synchronization mechanism, it has proven fruitful to take a probabilistic view. Probabilistic approaches are well established in the context of static power flow analysis (Borkowska, 1974). In the dynamic context, considering the expected likelihood of failure given a non-linear, random perturbation effectively averages over the various contingencies. Such probabilities are thus well suited to reveal structural features that enhance the system robustness or vulnerability. This approach has been highly successful in identifying particularly vulnerable grid regions (Menck et al., 2014; Schultz et al., 2014a; Nitzbon et al., 2017) and revealing general mechanisms of desynchronization (Hellmann et al., 2020). Probabilistic stability assessments recently gained more attention in the engineering community as well (Liu & Zhang, 2017; Liu et al., 2019; Liemann et al., 2021).

If graph neural networks (GNNs) were able to reliably predict probabilistic dynamic stability, they could be used to select promising candidate configurations for which a more detailed assessment should be carried out. Moreover, the analysis of the decision process of ML-models might lead to new unknown relations between dynamical properties and structural aspects, such as the grids topology or the distribution of loads and generators. Such insights may ultimately inform the design and development of power grids. Since datasets of probabilistic stability in power grids of sufficient size do not exist yet, we introduce new datasets, that consist of synthetic models of power grids and statistical results of dynamical simulations. We simulated datasets of increasing complexity to get closer to reality step by step. There are 10,000 small grids, 10,000 medium-sized grids and for evaluation purposes one large grid based on a synthetic Texan power grid model.

Related work on power grid property prediction Since power grids have an underlying graph structure, the recent development of graph representation learning (Bronstein et al., 2021; Hamilton, 2020) introduces promising methods to use machine learning in the context of power grids. There is a number of applications using GNNs for different power flow-related tasks (Donon et al., 2019; Kim et al., 2019; Bolz et al., 2019; Retiére et al., 2020; Wang et al., 2020; Owerko et al., 2020; Gama et al., 2020; Misyris et al., 2020; Liu et al., 2021; Bush et al., 2021; Liu et al., 2020; Jhun et al., 2022) and to predict transient dynamics in microgrids (Yu et al., 2022).

In (Nauck et al., 2022) small GNNs are used to predict the dynamic stability on small datasets. The authors demonstrate the general feasibility of the approach, but do not compare to conventional baselines. We add such baselines and introduce datasets that have ten times as many grids. This allows us to train GNNs with much higher capacity to achieve higher predictive power and sets a foundation for developing new methods on the path to practical applications.

Our main contributions are: We introduce new datasets of probabilistic dynamic stability of synthetic power grids. The new datasets have 10 times the size of previously published ones and and include a Texan power grid model to map the path towards real-world applications. We also introduce a new method to identify critical nodes the at which perturbations are strongly amplified. Such cases may be dangerous to hardware and the overall grid stability. We call such nodes: *trouble makers* and their identification leads to an additional classification task. We train strong baselines and benchmark models to evaluate the difficulty of all tasks. Our results demonstrate i) that the larger dataset allows to train more powerful GNNs, (ii) which outperform the baselines, and (iii) transfer from the new datasets to a real-sized power grid. The general approach is visualized in Figure 1.

2 Generation of the datasets

2.1 Structure and modeling of power grids

Mathematically speaking, power grids are characterized by distinctive topological properties. Power grids have more edges with respect to the number of nodes than a tree network, but are still sparsely

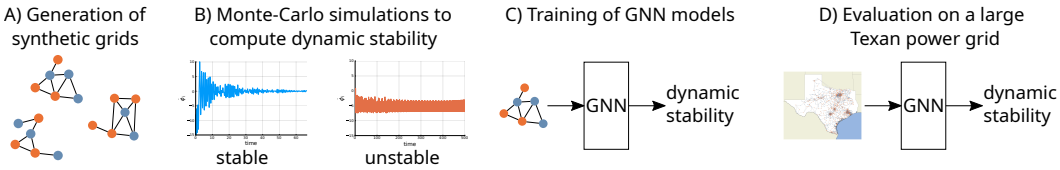

Figure 1: We generate new, large datasets of the dynamic stability of power grids, based on synthetic power grids (A) and statistics of dynamical simulations (B). Then, GNN models are trained to predict the dynamic stability of the synthetic grids (C) and, additionally, evaluated on a Texan power grid model (D).

connected. The degree distribution has a local maximum at very small degrees (e.g. ≈ 2.3), an exponentially decaying tail and the mean of the degree is ≈ 2.8 (Schultz et al., 2014b). Hence, most of the nodes are only connected to a small number of other nodes, whereas there are very few nodes connected to a larger number of nodes. The redundancy of lines leads to distinct distributions of local clustering coefficient and algebraic connectivity.

The study of non-linear self-organizing processes in power grids, specifically the interplay of synchrony and active power flows has been an active topic in the research of complex power grids (Witthaut et al., 2022). Depending on the task, different modeling strategies and timescales are used. When analyzing seasonal influence for example, larger timescales are used than for short-term market analyses. For our new dataset, we consider timescales of seconds and analyze the capability of the grid to remain in stable operation after different perturbations. In this context, the second-order Kuramoto model (Kuramoto, 2005; Rodrigues et al., 2016) is state-of-the-art, as it strikes a careful balance of capturing the key processes for the self-organized power flow and the synchronization of generators, while remaining computationally - and to some degree analytically - tractable. It is useful for understanding important structural and conceptual relationships that are expected to occur in real power grids, but it should not be taken to be a fully adequate model for all situations. The model includes inertia, which is important to properly simulate the reaction of power grids when using short timescales. To investigate the dynamical stability of power grids, basin stability Menck et al. (2014) is an important measure, because it allows to statistically analyze the dynamical behaviour of power grids.

Power grids as dynamical systems From a dynamical systems point of view, the central challenge is characterizing the interplay of network topology with the dynamics of synchronizing oscillators, which we address here. The dynamics of real-world power grids feature synchronizing oscillators coupled on complex networks. However, the full dynamics certainly contain many more aspects. A full scale analysis that can treat high-fidelity models of real systems is currently out of reach for several reasons. These include that real world data does not exist or is not accessible, synthetically generating large numbers of realistic dynamical models is challenging, and that models can not be simulated fast enough with current software Liemann et al. (2021). These problems force trade-offs on us, most notably reducing the details of the intrinsic behaviors of oscillators to the inertial Kuramoto model Kuramoto (1975); Bergen & Hill (1981). As a consequence, our results are not directly transferable to real-world applications. Nevertheless, any future treatment that moves towards more accurate dynamical models will also have to solve the challenging subproblem of the impact of topology on synchrony that we consider here.

Quantifying dynamic stability of power grids We quantify probabilistic dynamic stability with the widely used measures single-node basin stability (SNBS)(Menck et al., 2013) and survivability (SURV) (Hellmann et al., 2016). These are probabilistic measures that capture non-linear, dynamical effects after nodal perturbations. Crucially, they are not the result of a single simulation but describe a statistical behaviour, i.e. the expected value of a Bernoulli experiment. They takes values in $[0, 1]$, with higher values indicate more stability. Even though the perturbations are applied at a single node, SNBS and SURV capture the reaction of the entire grid and hence are not purely local properties. SNBS is an asymptotic measure, that considers if the system returns to its stable operational state after an initial perturbation at a single node. If this is the case for any perturbation applied then SNBS takes the value 1. On the other hand, SURV does not only consider the final state, but observes the transient dynamics and evaluates the likelihood that a trajectory of the perturbed system does not leave a region of desirable states. We set up a classification task based on a new interpretation of SURV by focusing on the identification of *trouble makers*. The idea is presented in more detail in the next subsection.

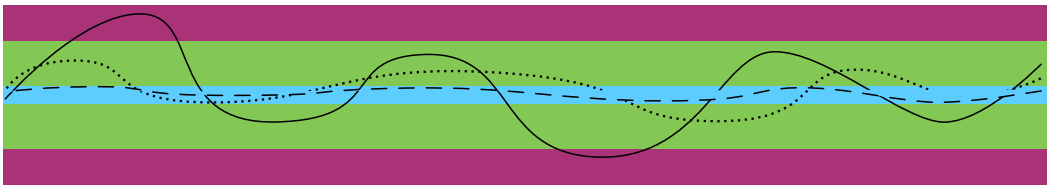

Figure 2: Identification of *trouble makers* based on trajectories. There are three spaces: blue for the space of perturbations, green for the *safe*-space and purple for the *trouble*-space. The dashed trajectory stays within the initial space of the perturbations, the dotted line leaves that space, but stays below the *trouble*-threshold and the solid line represents a trouble-maker-trajectory.

2.2 New method to identify trouble makers

The measure SURV characterizes the transient stability after non-linear perturbations. We use it to study nodes for which perturbations are significantly amplified along the trajectory. Operating power grids always involves dealing with small frequency changes and most of the times they are damped and do not result in critical outcomes. However, there are circumstances, when relatively small perturbations lead to problems and we want to investigate if we can introduce a new method to identify potential *trouble makers*. We consider trajectories to be critical, if an initially small perturbation is amplified by at least a factor of 6 throughout the simulations (see Figure 2). The factor was chosen for the following two reasons: 1) For smaller factors a strong correlation between the degree and the stability exists, so the prediction becomes less challenging. 2) For larger factors less and less nodes are classified as trouble makers. We consider the factor of 6 a good compromise between a challenging task and still a considerable number of nodes that are trouble makers (approx. 4 %). On the contrary to the SNBS task, the identification of trouble makers is an imbalanced classification task. More information can be found in Appendix A.5

2.3 Procedure of generating dataset20 and dataset100

We closely follow the methods in Nauck et al. (2022) and extend their work by computing 10 times as many grids. To investigate different topological properties of differently sized grids, we generate two datasets with either 20 or 100 nodes per grid, referred to as dataset20 and dataset100. To enable the training of complex models, both datasets consist of 10,000 graphs. The simulations take more than 550,000 CPU hours. Details regarding the modeling strategy are provided in Appendix A.4.

2.4 Modeling of the Texan power grid

To take a further step towards real-world applications, we evaluate the performance of our GNN models by analyzing the dynamic stability of a real-sized power grid. Real power grid data are not available due to security reasons and calculating an entire SNBS assessment of the fully parameterized synthetic model by Birchfield appears not to be feasible due to the computational effort (Liemann et al., 2021). We chose a synthetic model derived from the Texan power grid topology, introduced by Birchfield et al. (2017). The synthetic Texan power grid model consists of 1,910 nodes after removing 90 nodes that are not relevant for the dispatching.

We use the same modelling approach as for the other grids, i.e we use only the topological properties. As a consequence, we only investigate the potential applicability of GNNs to real-sized grids and can not make any statements about the real-world Texan power grid. Even after applying the simplifications, the simulations are already very expensive due to the large number of nodes. To manage the computational cost of simulating dynamic stability of such a large grid, we reduce the number of simulated perturbations from 10,000 to 1,000. Nonetheless, the simulation of that grid takes 127,000 CPU hours. Computing less perturbations results in an increased standard error of our estimates of approximately ± 0.031 for the 95 % confidence interval.

2.5 Properties of the datasets

Examples of the grids of dataset20, dataset100 and the Texan power grid as well as the distributions of SNBS characterized by multiple modes and SURV (imbalanced classification task) are given in Figure 3. Interestingly, the SNBS distribution of Texan power grid has a third mode which is challenging for prediction tasks. Overall, the power grid datasets consist of the adjacency matrix en-

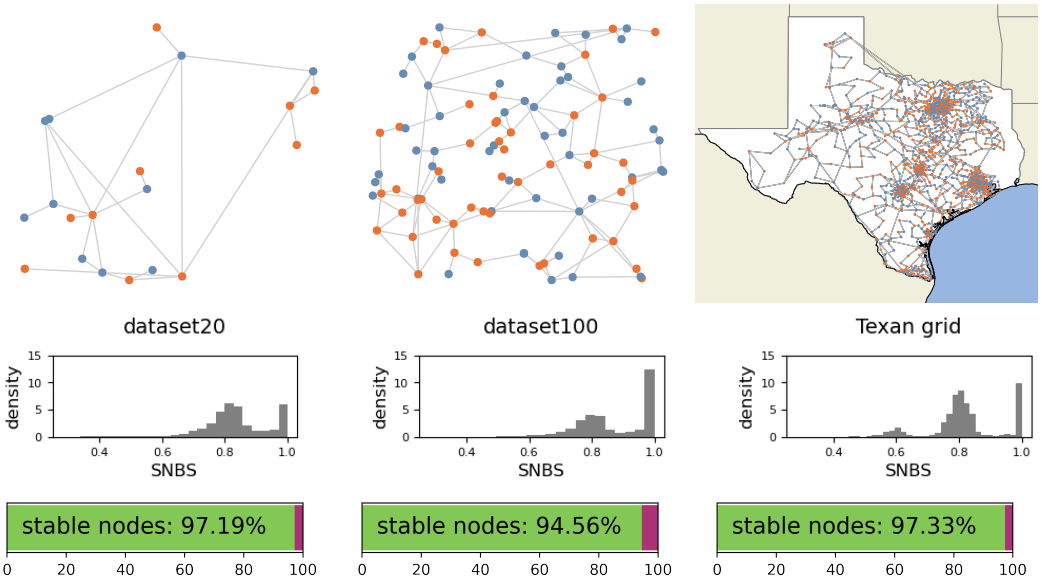

Figure 3: At the top, there are examples of the power grids of the datasets with 20 nodes (left) and 100 nodes (center) and the Texan power grid model (right). Blue color denotes sources and the orange sinks. For SURV we show the imbalance of the classifcation task where green represent stable nodes and purple show the trouble makers.

coding the topology, the binary injected power P per node as input features, and nodal SNBS/SURV as target values.

Relevance of the datasets Despite several simplifications, we see our work as a necessary step towards datasets of even more realistic power grid models. Those datasets will be more and more expensive to compute. We intend to sequentially increase the complexity of power grid datasets incrementally and develop appropriate ML methods on the way. For future synthetic power grids, several aspects could be integrated from Xu et al. (2017), who use more realistic modelling. Crucially, the current datasets are already challenging for GNN architectures, and thus provide an excellent starting point. Furthermore, the datasets connect well to existing approaches in network science which facilitates the comparison of GNNs and network science based approaches. This enables the comparison of known relations of topological properties and the dynamic stability to new results obtained from analyzing the explainability of the GNN models.

3 Predicting dynamic stability of power grids using GNNs

In this section, we analyze the dynamic stability on the new datasets using GNNs. We start by investigating the prediction of SNBS on the two grid sizes as well as out-of-distribution capabilities from small to large grids. Subsequently, we establish the benefit of our large dataset in comparison to previous work. Afterwards, we evaluate the GNNs trained on our datasets on the Texan power grid as a larger, more realistic test-case. Lastly we identify trouble makers based on SURV.

3.1 Experimental setup

On both datasets, we train GNNs on a nodal tasks using a regression setup in case of SNBS and a classification task for SURV (visualized in Figure 4). As input the GNNs are given an adjacency matrix representing the topology of power grids and a binary feature vector representing sources and sinks. We split the datasets in training, validation and testing sets (70:15:15). The validation set is used for the hyperparameter optimization, we report the performance on the test set. To minimize the effect of initializations we use 5 different initializations per model and consider the three best to compute average performances.

To evaluate the robustness of the GNNs, we analyze the performance of different models based on several GNN-architectures: GNNs with ARMA filters by Bianchi et al. (2021), Graph Convolutional Networks (GCN) by Kipf & Welling (2017), SAmple and aggreGatE (SAGE) by Hamilton et al. (2017) and Topology Adaptive Graph Convolution (TAG) by Du et al. (2017). We refer to the models by ArmaNet, GCNNet, SAGENet and TAGNet.

We conduct hyperparameter studies to optimize the model structure regarding number of layers, number of channels and layer-specific parameters using dataset20. The resulting models have the following properties: ArmaNet has 3 layers and 189,048 parameters. GCNNet has 7 layers and 523,020 parameters. SAGENet has 8 layers and 728,869 parameters. TAGNet has 13 layers and 415,320 parameters. Afterwards we optimize learning rate, batch size and scheduler of the best models for dataset20 and dataset100 separately. Details on the hyperparameter study are given in Appendix A.7.

Baseline models To better assess the GNN performance, we set up several baseline models. Schultz et al. (2014a) were the first to attempt predicting nodal dynamic stability of power grid models using logistic regression based on hand-crafted features. The hand-crafted features are used as inputs for the regression and they are based on different network measures. For the first baselines, we set up similar regression models and use a linear regression for SNBS and logistic regression for SURV. The following input features are used: degree, average-neigbor-degree, clustering-coefficient, current-flow-betweenness-centrality and closeness-centrality, as well as the nodal feature P as inputs. Additionally, we use more complex Multilayer perceptrons (MLPs) trained on the same features as baselines. We investigate the performance of two differently sized MLPs. MLP1 has one hidden layer with 35 units per hidden layer, resulting in 1,541 parameters and MLP2 consists of 6 hidden layers and 500 hidden units per layer leading to 1,507,001 parameters. We conducted hyperparameter studies to optimize the batch sizes and learning rates. Finally, for the SURV task, the 'semi-analytic lower bound for survivability´ (Hellmann et al., 2016) is used as a further baseline. This bound can be directly computed from the input data and hence requires no training.

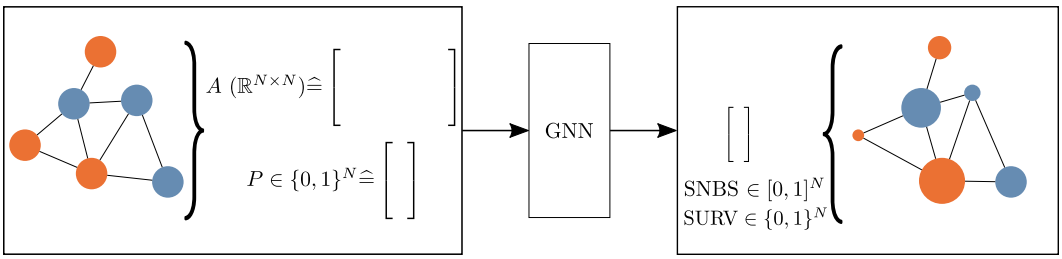

Figure 4: Prediction of SNBS using GNN. The GNN takes the adjacency matrix A and the injected power P as input to obtain the nodal SNBS as output. The different sizes of the nodes in the plot represent more or less stable nodes. The prediction of SNBS does not consider individual faults or any other variables, but operates only on topological properties of the power grid.

3.2 Metrics for evaluation

To analyze the performance, we use the coefficient of determination (R^2-score) for regression and F2-score for classification. F2-score is a modified F-score giving more weight to recall and less to the precision in the calculation of the score. In case of identifying vulnerabilities of power grids it is more important to identify all critical states, even if this increases the number of false positives. Hence, we also use recall to compare the performance. The details of computing the metrics are provided in Appendix A.6.

3.3 Prediction of SNBS

GNNs can accurately predict SNBS In our first set of benchmark experiments, we establish the predictive power of GNN models for SNBS. The key result is the surprisingly high performance across all datasets, see the first two columns in Table 1. R^2 reaches values above 82 % for dataset20 and above 88 % for dataset100. SNBS is a highly nonlinear property and the obtained performance exceeds expectations. The predictive performance captures not only the general trend but the

modalities in the data as well (cf. Figure 5). Interestingly the previously published Ar-bench (Nauck et al., 2022) performs worse than the MLPs, but the larger GNNs outperform all baselines.

Table 1: Results of predicting SNBS represented by R^2 score in %. Each column represents the evaluation on a different test set, e.g. *tr20ev20* denotes that the models are trained and evaluated using dataset20. Additionally, we analyze the out-of-distribution capabilities by evaluating the models on different datasets without retraining, e.g. we train a model on dataset20 and evaluate it on dataset100 and refer to this by *tr20ev100*. Besides the performance of the GNNs, we show the performance of a linear regression and Multilayer perceptrons using hand-crafted features as baselines.

model	tr20ev20	tr100ev100	tr20ev100	tr20evTexas	tr100evTexas
Ar-bench	51.82 $\pm$ 2.388	60.34 $\pm$ 0.299	38.80 $\pm$ 1.327	38.55 $\pm$ 5.897	55.22 $\pm$ 6.349
ArmaNet	80.63 $\pm$ 0.848	87.47 $\pm$ 0.073	**66.75** $\pm$ 1.500	60.54 $\pm$ 5.622	73.47 $\pm$ 2.786
GCNNet	70.64 $\pm$ 0.262	75.49 $\pm$ 0.276	59.46 $\pm$ 0.450	2.27 $\pm$ 6.210	-46.02 $\pm$ 1.575
SAGENet	65.46 $\pm$ 0.208	75.57 $\pm$ 0.228	52.27 $\pm$ 0.784	33.47 $\pm$ 0.509	55.38 $\pm$ 1.390
TAGNet	**82.49** $\pm$ 0.455	**88.22** $\pm$ 0.135	66.10 $\pm$ 0.508	**62.69** $\pm$ 2.029	**84.41** $\pm$ 0.947
linreg	41.75	36.29	5.98	-11.39	-22.62
MLP1	58.47 $\pm$ 0.149	63.59 $\pm$ 0.051	28.49 $\pm$ 1.493	-34.52 $\pm$ 17.934	19.79 $\pm$ 8.659
MLP2	58.20 $\pm$ 0.0422	65.52 $\pm$ 0.038	19.65 $\pm$ 2.109	5.81 $\pm$10.58	58.46 $\pm$ 0.480

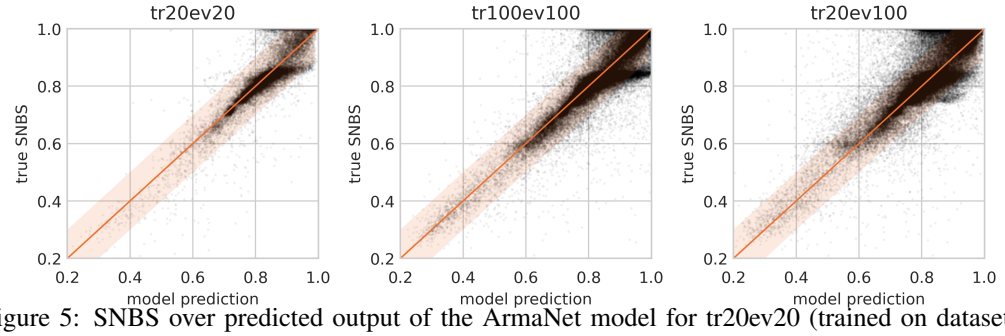

Figure 5: SNBS over predicted output of the ArmaNet model for tr20ev20 (trained on dataset20 and evaluated on dataset20), tr100ev100 and tr20ev100. The diagonal represents a perfect model ($R^2 = 1$), the band indicates a region for accurate predictions when using a threshold of 0.1.

GNNs generalize from dataset20 to dataset100 Training on small grids without loss of generalization and predictive power would be a huge advantage to scale to real power grids. To evaluate the potential of our datasets and GNN models to that end, we apply an out-of distribution task by training the models on dataset20 and evaluating the performance without any further training on dataset100. The third column in Table 1 (tr20ev100) show that all GNNs generalize well and are able to predict SNBS with R^2 exceeding 66 %. We would like to emphasize the significance of that finding. Given sufficient size and complexity in the source dataset, GNNs can robustly predict highly nonlinear stability metrics for grids several times larger than the source. We did not expect grids of size 20 to be large enough to contain enough relevant structures to generalize to larger grids. However, we expect grids of size 100 to generalize well to even larger grids and are encouraged by the out-of-distribution and the Texan power grid (cf. Section 3.3) results. Generalizing from small, numerically solvable grids to large grids is key for real world application. The computational cost of the dynamic simulations grows faster than linearly with the size of the grids, so computational time can be saved when training models on smaller networks or sections of real-sized grids. In comparison to the baselines, the generalization capabilities of the new GNN models are much better

Training on more data increases the performance of all models General Machine Learning convention assumes that larger dataset size allows to train larger models to higher performance. In this section, we investigate the influence of the size of the training set to show the relevance of the larger datasets. We train the models on the smaller dataset introduced in Nauck et al. (2022) after specifically optimizing the learning rate for the analysis of the smaller training set. Our experiments show that training on less data results in lower performance, see Figure 6. Instead of peak values of R^2 of 82.49 %, we only obtain 74.77 % for dataset20 and only 83.92 % instead of 88.22 % for dataset100. The results of all models are given in Appendix A.9. Comparing the performance

differences on dataset20 and dataset100, the improvements are larger for dataset20. A reasonable explanation is the total number of nodes used for the training. For dataset20, the training set based on 800 grids contains 16,000 nodes in total, the dataset100 contains 800,000 nodes.

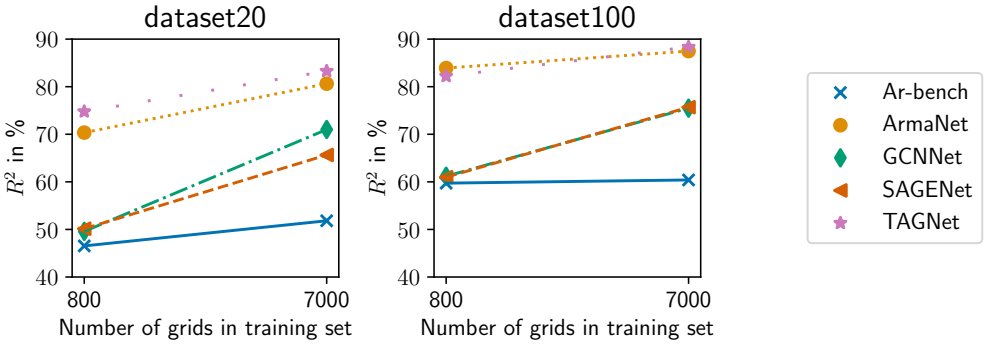

Figure 6: Comparison of the performance based on the size of the training set using 800 or 7,000 grids. Training on our larger dataset improves performance on all models. The 800 grids used for the training follow Nauck et al. (2022) and all models are evaluated on the newly introduced test set.

Predicting SNBS on a Texan power grid model using the previously trained models Using GNNs for SNBS prediction becomes feasible, if they can be trained on relatively simple datasets and still perform well on large, complex grids. As an example of a large and complex grid, we use a Texan power grid model and evaluate the models previously trained on dataset20 and dataset100. The benchmark models achieve surprisingly high performance with R^2 values (above 84 %) and GCN as a sole exception, see the colums tr20evTexas and tr100evTexas in Table 1. Details on the poor performance of the GCN is given in Appendix A.10. Hence, the approach of training models on grids which are smaller by more than one order of magnitude is feasible. We want to emphasize that one successful attempt of a real-sized power grid should illustrate the general potential of this approach, but we still consider the hard evidence to be the generalization from 20 to 100.

The performance is significantly better for the models trained on dataset100. We hypothesize that the repetition of geometrical structures more prevalent in dataset100 is useful for even larger grids. Grids of size 20 might still be to small to generalize to large grids, but the size 100 might actually be sufficient for many applications.

3.4 Identification of trouble makers

In this section we analyze the capabilities of GNNs predicting trouble makers according to the idea introduced in Section 2.2. Nodes are considered trouble makers, if small disturbances are amplified by a factor of at least 6. From the perspective of grid operators, it is more important to identify potential troublemakers than to identify stable nodes. Hence the recall of a classifier is more important than its precision and accordingly the F2-score is used as evaluation metric. The results are given in Table 2 using the F2-score and in Table 3 using the recall. The GNNs outperform the baselines and especially TAGNet achieves a good performance. TAGNet achieves recalls of more than 93 % throughout all tasks, including the out-of-distribution generalizations, while still keeping F2-score above 82 %.

3.5 Benefits of using GNNs to predict dynamic stability

As the experiments above show, GNNs are suitable for the analysis of the dynamic stability, both in terms of computational effort and predictive power. The performance at the regression task exceeded our expectations. In case of TAGNet, which is the best model, roughly 95 % of the estimations differ only by 0.1 from the target values for tr100ev100 and tr100evTexas. This corresponds to the banded region shown in Figure 5. The threshold of 0.1 can be motivated by considering the distributions in Figure 3, because the modes are separated by 0.1. To achieve a similar accuracy by conducting dynamical simulations with Monte-Carlo sampling, 100 perturbations per node are needed, which is a significant reduction to 10,000 perturbations we used for the generation of the datasets. With 100 perturbations, the computation for one grid of size 100 takes roughly 30 minutes and for the Texan grid 530 days using one CPU. Evaluating the GNNs takes less than one second per grid. Hence

Table 2: SURV prediction: F2-score in %. In case of MLP2, the training on dataset100 was successful with only one of the five seeds. We only report that result as over-estimation of a strong baseline. (*) Since the semi-analytic baseline does not require training, its performance is directly evaluated on the test set.

	tr20			tr100	
model	ev20	ev100	evTexas	ev100	evTexas
Ar-bench	74.31 $\pm$ 0.230	77.08 $\pm$ 0.989	78.49 $\pm$ 0.790	88.29 $\pm$ 0.331	83.3 $\pm$ 1.856
ArmaNet	86.85 $\pm$ 0.503	84.94 $\pm$ 4.050	82.43 $\pm$ 1.747	95.75 $\pm$ 0.049	85.14 $\pm$ 1.335
TAGNet	**89.78** $\pm$ 0.079	**88.77** $\pm$ 0.309	**82.78** $\pm$ 2.122	**96.36** $\pm$ 0.063	**96.51** $\pm$ 0.028
logreg	44.74	55.92	21.66	73.47	55.81
MLP1	58.06 $\pm$ 1.451	72.68 $\pm$ 6.142	49.52 $\pm$ 9.990	79.61 $\pm$ 0.001	86.02 $\pm$ 0.000
MLP2	69.43 $\pm$ 0.724	41.11 $\pm$ 7.200	32.16 $\pm$ 12.34	91.60	73.22
semi-analytic	22.86*	38.53*	15.73*	38.53*	15.73*

Table 3: SURV prediction: recall in %. In case of MLP2, the training on dataset100 was successful with only one of the five seeds. We only report that result as over-estimation of a strong baseline.

model	tr20			tr100	
	ev20	ev100	evTexas	ev100	evTexas
Ar-bench	86.30 $\pm$ 0.727	83.33 $\pm$ 0.983	98.69 $\pm$ 0.654	92.56 $\pm$ 0.798	98.04 $\pm$ 0.000
ArmaNet	88.85 $\pm$ 1.019	86.74 $\pm$ 5.455	90.20 $\pm$ 2.264	96.85 $\pm$ 0.272	**98.69** $\pm$ 0.654
TAGNet	**93.84** $\pm$ 0.039	**96.10** $\pm$ 0.590	**99.35** $\pm$ 0.654	**97.16** $\pm$ 0.120	97.15 $\pm$ 0.100
logreg	86.69	95.32	98.04	91.82	94.12
MLP1	76.44 $\pm$ 0.312	68.98 $\pm$ 6.874	44.44 $\pm$ 9.628	90.70 $\pm$ 0.000	94.12 $\pm$ 0.000
MLP2	75.85 $\pm$ 0.513	36.08 $\pm$ 6.902	28.10 $\pm$ 11.563	89.86	68.63
semi-analytic	92.70*	**97.77***	98.04*	**97.77***	98.04*

using GNNs is more than 1,800 times faster for grids of size 100 and 4,6 $\times 10^7$ times faster for the synthetic Texan power grid. This demonstrates the potential of analyzing a large number of different configurations using GNNs.

4 Conclusion and Outlook

In this work, we analyze the probabilistic dynamic stability of synthetic power grids using GNNs. We generate new datasets that are 10 times larger than previous ones to enable the training of high capacity GNNs. Our benchmark results significantly improve over previous work and show that highly nonlinear SNBS can be predicted at surprisingly high accuracy by using more complex models and training on our larger dataset. We show that the models trained on our datasets can be used for prediction on a real-world-sized Texan power grid model. The results indicate the potential benefits of machine learning in this important domain. We expect the datasets to attract attention by research groups working on complexity science and non-linear dynamics as well. Furthermore, we introduce a new method to identify trouble makers in power grids and and show that GNNs are accurate classifiers for this task. Besides further improving the performance, future work might try to explain the decision process of GNNs to generate new insights on the relation of topology and stability. Encouraged by our results, we will continue to extend our datasets with increasingly more complex and realistic grids, aiming at real power grids. All data and code will be published upon publication (see Appendix A.1). For reviewing, the SNBS training is provided in a zip container at: `https://drive.google.com/drive/folders/19hHqgup12xypYnS0ceEDaZRhqdqUG7gi?usp=sharing`. The open access enables the community to develop new methods to analyze future renewable power grids.

REFERENCES

A.R. Bergen and D.J. Hill. A Structure Preserving Model for Power System Stability Analysis. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(1):25–35, January 1981. ISSN 0018-9510. doi: 10.1109/TPAS.1981.316883. Conference Name: IEEE Transactions on Power Apparatus and Systems.

Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 2017. ISSN 00361445. doi: 10.1137/141000671. _eprint: 1411.1607.

Filippo Maria Bianchi, Daniele Grattarola, Lorenzo Livi, and Cesare Alippi. Graph Neural Networks with Convolutional ARMA Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021. ISSN 1939-3539. doi: 10.1109/TPAMI.2021.3054830. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Adam Birchfield. ACTIVSg2000: 2000-bus synthetic grid on footprint of Texas, visited on Nov. 1, 2021. URL https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg2000/.

Adam B. Birchfield, Ti Xu, Kathleen M. Gegner, Komal S. Shetye, and Thomas J. Overbye. Grid Structural Characteristics as Validation Criteria for Synthetic Networks. *IEEE Transactions on Power Systems*, 32(4):3258–3265, July 2017. ISSN 1558-0679. doi: 10.1109/TPWRS.2016.2616385. Conference Name: IEEE Transactions on Power Systems.

Valentin Bolz, Johannes Rueß, and Andreas Zell. Power Flow Approximation Based on Graph Convolutional Networks. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 1679–1686, December 2019. doi: 10.1109/ICMLA.2019.00274.

Barbara Borkowska. Probabilistic Load Flow. *IEEE Transactions on Power Apparatus and Systems*, PAS-93(3):752–759, May 1974. ISSN 0018-9510. doi: 10.1109/TPAS.1974.293973. Conference Name: IEEE Transactions on Power Apparatus and Systems.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv:2104.13478 [cs, stat]*, May 2021. URL http://arxiv.org/abs/2104.13478. arXiv: 2104.13478.

Charles D. Brummitt, Paul D. H. Hines, Ian Dobson, Cristopher Moore, and Raissa M. D'Souza. Transdisciplinary electric power grid science. *Proceedings of the National Academy of Sciences*, 110(30):12159–12159, July 2013. doi: 10.1073/pnas.1309151110. URL https://www.pnas.org/doi/full/10.1073/pnas.1309151110. Publisher: Proceedings of the National Academy of Sciences.

Brian Bush, Yuzhou Chen, Dorcas Ofori-Boateng, and Yulia R. Gel. Topological Machine Learning Methods for Power System Responses to Contingencies. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17):15262–15269, May 2021. ISSN 2374-3468. URL https://ojs.aaai.org/index.php/AAAI/article/view/17791. Number: 17.

Peter Christensen, Gert Karmisholt Andersen, Matthias Seidel, Sigrid Bolik, Sönke Engelken, Thyge Knueppel, Athanasios Krontiris, Klaus Wuerflinger, Thorsten Bülo, Jörg Jahn, Mario Ndreko, Saeed Salehi, Ioannis Theologitis, Bernd Weise, Helge Urdal, Agusti Egea Alvarez, Andrew J Roscoe, and Jens Fortmann. *High Penetration of Power Electronic Interfaced Power Sources and the Potential Contribution of Grid Forming Converters*. Brussels, January 2020. URL https://www.entsoe.eu/events/2020/01/30/workshop-on-high-penetration-of-power-electronic-interfaced-power-sources-and-the-potential-contribution-of-grid-forming-converters/.

Balthazar Donon, Benjamin Donnot, Isabelle Guyon, and Antoine Marot. Graph Neural Solver for Power Systems. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2019-July. Institute of Electrical and Electronics Engineers Inc., July 2019. ISBN 978-1-72811-985-4. doi: 10.1109/IJCNN.2019.8851855.

Jian Du, Shanghang Zhang, Guanhang Wu, Jose M. F. Moura, and Soummya Kar. Topology Adaptive Graph Convolutional Networks. October 2017. URL http://arxiv.org/abs/1710.10370. _eprint: 1710.10370.

Florian Dörfler, Michael Chertkov, and Francesco Bullo. Synchronization in complex oscillator networks and smart grids. *Proceedings of the National Academy of Sciences of the United States of America*, 2013. ISSN 00278424. doi: 10.1073/pnas.1212134110.

Matthias Fey and Jan E Lenssen. FAST GRAPH REPRESENTATION LEARNING WITH PYTORCH GEOMETRIC. pp. 9, 2019. URL http://arxiv.org/abs/1903.02428.

C.B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Chan, 2012.

Fernando Gama, Ekaterina Tolstaya, and Alejandro Ribeiro. Graph Neural Networks for Decentralized Controllers. March 2020. URL http://arxiv.org/abs/2003.10280. _eprint: 2003.10280.

William L. Hamilton. Graph Representation Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159, September 2020. ISSN 1939-4608, 1939-4616. doi: 10.2200/S01045ED1V01Y202009AIM046. URL https://www.morganclaypool.com/doi/10.2200/S01045ED1V01Y202009AIM046.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, 2017. ISSN: 10495258 _eprint: 1706.02216.

Frank Hellmann, Paul Schultz, Carsten Grabow, Jobst Heitzig, and Jürgen Kurths. Survivability of Deterministic Dynamical Systems. *Scientific Reports*, 6(1):29654, July 2016. ISSN 2045-2322. doi: 10.1038/srep29654. URL https://www.nature.com/articles/srep29654. Number: 1 Publisher: Nature Publishing Group.

Frank Hellmann, Paul Schultz, Patrycja Jaros, Roman Levchenko, Tomasz Kapitaniak, Jürgen Kurths, and Yuri Maistrenko. Network-induced multistability through lossy coupling and exotic solitary states. *Nature Communications*, 11(1):592, January 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14417-7. URL https://www.nature.com/articles/s41467-020-14417-7. Number: 1 Publisher: Nature Publishing Group.

Bukyoung Jhun, Hoyun Choi, Yongsun Lee, Jongshin Lee, Cook Hyun Kim, and B. Kahng. Prediction and mitigation of nonlocal cascading failures using graph neural networks, July 2022. URL http://arxiv.org/abs/2208.00133. arXiv:2208.00133 [physics].

Cheolmin Kim, Kibaek Kim, Prasanna Balaprakash, and Mihai Anitescu. Graph Convolutional Neural Networks for Optimal Load Shedding under Line Contingency. In *2019 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, August 2019. doi: 10.1109/PESGM40551.2019.8973468. ISSN: 1944-9933.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907 [cs, stat]*, February 2017. URL http://arxiv.org/abs/1609.02907. arXiv: 1609.02907.

Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. *Mathematical Problems in Theoretical Physics*, 39:420–422, January 1975. doi: 10.1007/BFb0013365. URL https://ui.adsabs.harvard.edu/abs/1975LNP....39..420K. ADS Bibcode: 1975LNP....39..420K.

Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics*. 2005. doi: 10.1007/bfb0013365.

Sebastian Liemann, Lia Strenge, Paul Schultz, Holm Hinners, Johannis Porst, Marcel Sarstedt, and Frank Hellmann. Probabilistic Stability Assessment for Active Distribution Grids. In *2021 IEEE Madrid PowerTech*, pp. 1–6, June 2021. doi: 10.1109/PowerTech46648.2021.9494855.

Michael Lindner, Lucas Lincoln, Fenja Drauschke, Julia M. Koulen, Hans Würfel, Anton Plietzsch, and Frank Hellmann. NetworkDynamics.jl—Composing and simulating complex networks in Julia. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(6):063133, June 2021. ISSN 1054-1500. doi: 10.1063/5.0051387. URL https://aip.scitation.org/doi/10.1063/5.0051387. Publisher: American Institute of Physics.

Yuxiao Liu, Ning Zhang, Dan Wu, Audun Botterud, Rui Yao, and Chongqing Kang. Guiding Cascading Failure Search with Interpretable Graph Convolutional Network, January 2020. URL http://arxiv.org/abs/2001.11553. arXiv:2001.11553 [cs, eess].

Yuxiao Liu, Ning Zhang, Dan Wu, Audun Botterud, Rui Yao, and Chongqing Kang. Searching for Critical Power System Cascading Failures with Graph Convolutional Network. *IEEE Transactions on Control of Network Systems*, pp. 1–1, 2021. ISSN 2325-5870. doi: 10.1109/TCNS.2021.3063333. Conference Name: IEEE Transactions on Control of Network Systems.

Zhao Liu and Ziang Zhang. Quantifying transient stability of generators by basin stability and Kuramoto-like models. In *2017 North American Power Symposium (NAPS)*, pp. 1–6, September 2017. doi: 10.1109/NAPS.2017.8107260.

Zhao Liu, Xi He, Zhenhuan Ding, and Ziang Zhang. A Basin Stability Based Metric for Ranking the Transient Stability of Generators. *IEEE Transactions on Industrial Informatics*, 15(3):1450–1459, March 2019. ISSN 1941-0050. doi: 10.1109/TII.2018.2846700. Conference Name: IEEE Transactions on Industrial Informatics.

Peter J. Menck, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. How basin stability complements the linear-stability paradigm. *Nature Physics*, 9(2):89–92, February 2013. ISSN 1745-2481. doi: 10.1038/nphys2516. URL https://www.nature.com/articles/nphys2516. Number: 2 Publisher: Nature Publishing Group.

Peter J. Menck, Jobst Heitzig, Jürgen Kurths, and Hans Joachim Schellnhuber. How dead ends undermine power grid stability. *Nature Communications*, 5(1):3969, June 2014. ISSN 2041-1723. doi: 10.1038/ncomms4969. URL https://www.nature.com/articles/ncomms4969. Number: 1 Publisher: Nature Publishing Group.

Federico Milano, Florian Dörfler, Gabriela Hug, David J. Hill, and Gregor Verbič. Foundations and Challenges of Low-Inertia Systems (Invited Paper). In *2018 Power Systems Computation Conference (PSCC)*, pp. 1–25, June 2018. doi: 10.23919/PSCC.2018.8450880.

George S. Misyris, Andreas Venzke, and Spyros Chatzivasileiadis. Physics-Informed Neural Networks for Power Systems. In *2020 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, August 2020. doi: 10.1109/PESGM41954.2020.9282004. ISSN: 1944-9933.

Philipp Moritz, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol, Zongheng Yang, William Paul, Michael I. Jordan, and Ion Stoica. Ray: A Distributed Framework for Emerging AI Applications. *arXiv:1712.05889 [cs, stat]*, September 2018. URL http://arxiv.org/abs/1712.05889. arXiv: 1712.05889.

Adilson E. Motter, Seth A. Myers, Marian Anghel, and Takashi Nishikawa. Spontaneous synchrony in power-grid networks. *Nature Physics*, 9(3):191–197, March 2013. ISSN 1745-2481. doi: 10.1038/nphys2535. URL https://www.nature.com/articles/nphys2535. Number: 3 Publisher: Nature Publishing Group.

Christian Nauck, Michael Lindner, Konstantin Schürholt, Haoming Zhang, Paul Schultz, Jürgen Kurths, Ingrid Isenhardt, and Frank Hellmann. Predicting basin stability of power grids using graph neural networks. *New Journal of Physics*, 2022. ISSN 1367-2630. doi: 10.1088/1367-2630/ac54c9. URL http://iopscience.iop.org/article/10.1088/1367-2630/ac54c9.

J. Nitzbon, P. Schultz, J. Heitzig, J. Kurths, and F. Hellmann. Deciphering the imprint of topology on nonlinear dynamical network stability. *New Journal of Physics*, 19(3):033029, March 2017. ISSN 1367-2630. doi: 10.1088/1367-2630/aa6321. URL https://doi.org/10.1088/1367-2630/aa6321. Publisher: IOP Publishing.

Damian Owerko, Fernando Gama, and Alejandro Ribeiro. Optimal Power Flow Using Graph Neural Networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5930–5934, May 2020. doi: 10.1109/ICASSP40776.2020.9053140. ISSN: 2379-190X.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Anton Plietzsch, Raphael Kogler, Sabine Auer, Julia Merino, Asier Gil-de Muro, Jan Liße, Christina Vogel, and Frank Hellmann. PowerDynamics.jl – An experimentally validated open-source package for the dynamical analysis of power grids. *arXiv:2101.02103 [cs, eess]*, January 2021. URL http://arxiv.org/abs/2101.02103. arXiv: 2101.02103.

Hans-Otto Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2022.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(1): 15, May 2017. ISSN 2049-9647. doi: 10.5334/jors.151. URL http://openresearchsoftware.metajnl.com/articles/10.5334/jors.151/. Number: 1 Publisher: Ubiquity Press.

Nicolas Retiére, Dinh Truc Ha, and Jean-Guy Caputo. Spectral Graph Analysis of the Geometry of Power Flows in Transmission Networks. *IEEE Systems Journal*, 14(2):2736–2747, June 2020. ISSN 1937-9234. doi: 10.1109/JSYST.2019.2928852. Conference Name: IEEE Systems Journal.

Francisco A. Rodrigues, Thomas K. DM. Peron, Peng Ji, and Jürgen Kurths. The Kuramoto model in complex networks. *Physics Reports*, 610:1–98, January 2016. ISSN 0370-1573. doi: 10.1016/j.physrep.2015.10.008. URL https://www.sciencedirect.com/science/article/pii/S0370157315004408.

Martin Rohden, Andreas Sorge, Marc Timme, and Dirk Witthaut. Self-Organized Synchronization in Decentralized Power Grids. *Physical Review Letters*, 109(6):064101, August 2012. doi: 10.1103/PhysRevLett.109.064101. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.064101. Publisher: American Physical Society.

Paul Schultz. *https://github.com/PIK-ICoNe/SyntheticNetworks.jl*. 2020. URL https://github.com/PIK-ICoNe/SyntheticNetworks.jl.

Paul Schultz, Jobst Heitzig, and Jürgen Kurths. Detours around basin stability in power networks. *New Journal of Physics*, 16(12):125001, December 2014a. ISSN 1367-2630. doi: 10.1088/1367-2630/16/12/125001. URL https://iopscience.iop.org/article/10.1088/1367-2630/16/12/125001.

Paul Schultz, Jobst Heitzig, and Jürgen Kurths. A random growth model for power grids and other spatially embedded infrastructure networks. *The European Physical Journal Special Topics*, 223(12):2593–2610, October 2014b. ISSN 1951-6401. doi: 10.1140/epjst/e2014-02279-6. URL https://doi.org/10.1140/epjst/e2014-02279-6.

Benjamin Schäfer, Thiemo Pesch, Debsankha Manik, Julian Gollenstede, Guosong Lin, Hans-Peter Beck, Dirk Witthaut, and Marc Timme. Understanding Braess' Paradox in power grids. *Nature Communications*, 13(1):5396, September 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32917-6. URL https://www.nature.com/articles/s41467-022-32917-6. Number: 1 Publisher: Nature Publishing Group.

Dawei Wang, Kedi Zheng, Qixin Chen, Gang Luo, and Xuan Zhang. Probabilistic Power Flow Solution with Graph Convolutional Network. In *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 650–654, October 2020. doi: 10.1109/ISGT-Europe47291.2020.9248786.

Dirk Witthaut and Marc Timme. Braess's paradox in oscillator networks, desynchronization and power outage. *New Journal of Physics*, 2012. ISSN 13672630. doi: 10.1088/1367-2630/14/8/083036.

Dirk Witthaut, Frank Hellmann, Jürgen Kurths, Stefan Kettemann, Hildegard Meyer-Ortmanns, and Marc Timme. Collective nonlinear dynamics and self-organization in decentralized power grids. *Reviews of Modern Physics*, 94(1):015005, February 2022. doi: 10.1103/RevModPhys.94.015005. URL https://link.aps.org/doi/10.1103/RevModPhys.94.015005. Publisher: American Physical Society.

Ti Xu, Adam B Birchfield, Komal S Shetye, and Thomas J Overbye. Creation of Synthetic Electric Grid Models for Transient Stability Studies. pp. 7, 2017.

Yin Yu, Xinyuan Jiang, Daning Huang, and Yan Li. PIDGeuN: Graph Neural Network-Enabled Transient Dynamics Prediction of Networked Microgrids Through Full-Field Measurement. *arXiv:2204.08557 [cs, eess]*, April 2022. URL http://arxiv.org/abs/2204.08557. arXiv: 2204.08557.

A APPENDIX

This section includes additional information to generate the datasets, reproduce the presented results and additional results that are not already shown in the main section.

We start by providing information on the availability of the datasets and the used software, before providing more details regarding the modeling for the dataset generations. Subsequently, details regarding the training, evaluating as well as the hyperparameter study are provided. Afterwards additional results are shown.

A.1 AVAILABILITY OF THE DATASETS

The new datasets and full code for the training, evaluation and generation of the figures will be published upon publication on Zenodo and GitHub and it will be licensed under CC-BY 4.0 to enable the community to contribute to this challenge.

A.2 SOFTWARE FOR GENERATING THE DATASETS

Julia is used for the simulations Bezanson et al. (2017) and the dynamic simulations rely on the package DifferentialEquations.jl Rackauckas & Nie (2017). For simulating more realistic power grids in future work we recommend the additional use of NetworkDynamics.jl Lindner et al. (2021) and PowerDynamics.jl Plietzsch et al. (2021).

A.3 SOFTWARE FOR TRAINING

The training is implemented in Pytorch (Paszke et al., 2019). For the graph handling and graph convolutional layers we rely on the additional library PyTorch Geometric (Fey & Lenssen, 2019). We use the SGD-optimizer and as loss function we use the mean squared error [1]. Furthermore `ray` (Moritz et al., 2018) is used for parallelizing the hyperparameter study.

A.4 MODELLING DETAILS OF GENERATING THE DATASETES

In the detail of the modeling and the size of the dataset we attempt to strike a balance between relevance to real-world applications, computational tractability and conceptual simplicity. Therefore we employ the following criteria: (i) generate synthetic network topologies that mimic real world power grids; (ii) model the main dynamics of self-organised power flow and synchronization; (iii) minimize statistical and numerical errors with highly accurate simulations; (iv) to study out-of distribution tasks and scale effects, consider grid sizes of different orders of magnitude.

The most important simplifications in comparison to real-world power grids are homogeneous edges, fixed magnitudes of sources/sinks and modeling all nodes by the swing equation. In contrast to our modeling, real power grid lines have different properties and there are more complex models for generators and loads. However, previous studies have shown that many interesting observations are still possible under these assumptions Nitzbon et al. (2017).

To investigate different topological properties of differently sized grids, we generate two datasets with either 20 or 100 nodes per grid, referred to as dataset20 and dataset100. To enable the training of complex models, both datasets consist of 10,000 graphs. Additionally, probabilistic dynamic stability values of a synthetic model of the Texan power grids are provided for evaluation purposes.

Modeling of the power grids This section covers the precise modeling of power grids used for the dataset generation and may be skipped. To generate realistic topologies we use the package `SyntheticNetworks`(Schultz et al., 2014b; Schultz, 2020). The sources and sinks are assigned randomly. For the dynamic simulations all nodes are represented by the 2nd-order-Kuramoto model (Kuramoto, 2005; Rodrigues et al., 2016), which is also called swing equation:

$$\ddot{\phi}_i = P_i - \alpha\dot{\phi}_i - \sum_j^n K_{ij}\sin(\phi_i - \phi_j), \tag{1}$$

[1] corresponds to MSELoss in Pytorch

where $\phi, \dot{\phi}, \ddot{\phi}$ denotes the voltage angle and its time derivatives. The parametrization is as follows: the injected power $P_i \in \{-1, 1\}$ where $\sum_i P_i = 0$ to guarantee power balance, the damping coefficient $\alpha = 0.1$ the coupling matrix K is based on the graphs' adjacency matrix A which encodes the graph topology and we use a uniform coupling strength, that is $K_{ij} = 9A_{ij}$. Using homogeneous coupling strength can be interpreted as considering power grids that only have one type of power line and comparable distances between all nodes. This assumption is justifiable, because in real power grids longer lines are built stronger, so the actual coupling does not scale as much with the length.

To estimate SNBS we rely on the approach presented in Nauck et al. (2022): "[F]or every node in a graph, $M = 10,000$ samples of perturbations per node are constructed by sampling a phase and frequency deviation from a uniform distribution with $(\phi, \dot{\phi}) \in [-\pi, \pi] \times [-15, 15]$ and adding them to the synchronized state. Each such single-node perturbation serves as an initial condition of a dynamic simulation of our power grid model, [cf. Equation (1)]. At $t = 500$ the integration is terminated and the outcome of the Bernoulli trial is derived from the final state. A simulation outcome is referred to as *stable* if at all nodes $\dot{\phi}_i < 0.1$. Otherwise it is referred to as *unstable*. The classification threshold of 0.1 is chosen accounting for minor deviations due to numerical noise and slow convergence rates within a finite time-horizon."

To ensure the reliability of our results we try to minimize numerical and statistical errors: For the dynamical simulations a higher order Runge-Kutta methods with adaptive time stepping and low error tolerances is used. For the Monte-Carlo sampling, 10,000 simulations per node result in standard errors for our probabilistic measure SNBS of less than ± 0.01.

Furthermore, we want to provide some numbers regarding the simulation time. The computation of a single perturbation in case of dataset20 takes .056 seconds, in case of dataset100 0.189 seconds and in case of the Texan power grid 239.97 seconds.

A.5 Identification of trouble makers using SURV

For the survivability we consider the maximum magnitude of $\dot{\phi}$ throughout the entire time region and consider a configuration stable if $\dot{\phi}(t) <$ threshold for all t, and the threshold is set to the maximum magnitude of our applied perturbations. Hence, the survivability is also a probability, where large values mean that more trajectories survived and low values mean that the node is more vulnerable to the applied perturbations. For our analysis, we reformulate this problem to a classification task and we focus on the identification of *trouble makers*, hence we consider trajectories as critical if the applied perturbations grows by a factor of 6 during the transient simulations. Similar to SNBS, we count the number of critical trajectories to compute their likelihood.

Selection of perturbations for the identification of trouble makers The idea is visualized in fig. 2. If perturbations are amplified by more than a factor of 6, the desired space is left and we consider the trajectories as troublesome. If that occurs for at least one perturbation applied at a node, we consider that node a trouble maker, because such a strong amplification is a danger for safe grid operation. For the analysis of SURV, we only consider perturbations within $(\dot{\phi}) \in [-2.5, 2.5]$. This reduces the minimum number of analyzed perturbations per node to 1,595 for dataset20, 1,569 for dataset100 and 129 for the Texan power grid and increases the error bounds of the 95 % confidence interval to .1.

A.6 Evaluation of the performance using different metrics

For the regression, we use the coefficient of determination: $R^2 = 1 - \frac{mse(y,t)}{mse(t_{mean},t)}$, where mse denotes the mean squared error, y the output of the model, t the target value and t_{mean} the mean of all considered targets of the test dataset. R^2 captures the mean square error relative to a null model that predicts the mean of the test-dataset for all points. The R^2-score is used to measure the portion of explained variance in a dataset. By design, a model that predicts the mean of SNBS per grids has $R^2 = 0$.

For the classification, we use F2-score which is based on recall and precision. The recall is defined by: Recall $= \frac{TP}{TP+FN}$, where TP denotes true positives and FN false negatives. The precision

is defined by: precision $= \frac{TP}{TP+FP}$. This leads to the F2-score which is defined by: $F-2 = \frac{5 \times precision \times recall}{4 \times precision + recall}$.

A.7 Hyperparameter optimization

We conduct hyperparameter studies in two steps. First, we optimize model properties such as the number of layers and channels as well as layer-specific parameters e.g. the number of stacks and internal layers in case of ArmaNets. For this optimization we use dataset20 and the SNBS task only. For all models we investigated the influence of different numbers of layers and the numbers of channel between multiple layers. We limit the model size to just above four million parameters, so we did not investigate the full presented space, but limited for example the number of channels when adding more layers.

Afterwards we optimize the learning rate, batch size and scheduler of the best models for dataset20 and dataset100 and the tasks SNBS/SURV separately. Hence, our models are not optimized to perform well at the out-of distribution task. The best model from (Nauck et al., 2022) referred to as Ar-bench, is used as another baseline. It is a GNN model consisting of 1,050 parameters and based on 2 Arma-layers. The only adjustment to that model is the removal of the fully connected layer after the second Arma-Convolution and before applying the Sigmoid-layer, which improves the training.

A.8 Details of the training of the benchmark models

To reproduce the obtained results, more information regarding the training is provided in this section. Detailed information on the training as well as the computation time is shown in Table 4. In case of dataset20, a scheduler is not applied, in case of dataset100, schedulers are used for Ar-bench (stepLR), GCNNet (ReduceLROnPlateau). The default validation and test set batch size is 150. The validation and test batchsize for Ar-bench and ArmaNet3 is 500 in case of dataset20 and 100 for dataset100. The number of trained epochs differs, because the training is terminated in case of significant overfitting. Furthermore, different batch sizes have significant impact on the simulation time. Most of the training success occurs within the first 100 epochs, afterwards the improvements are relatively small.

Table 4: Properties of training models and regarding the training time, we train 5 seeds in parallel using one nVidia V100.

name	number of epochs		training time		train batch size		learning rate	
dataset	20	100	20 (hours)	100 (days)	20	100	20	100
Ar-bench	1,000	800	26	4	200	12	0.914	.300
ArmaNet	1,500	1,000	46	6	228	27	3.00	3.00
GCNNet	1,000	1000	29	5	19	79	.307	.286
SAGENet	300	1000	9	5	19	16	1.10	1.23
TAGNet	400	800	11	4	52	52	0.193	.483

A.9 Detailed results of training on a smaller dataset

To investigate the influence of available training data and to connect with previous work, we train all models on only 800 grids, from Nauck et al. (2022). The results are shown in Table 5.

A.10 Poor performance of GCN when applying to the Texan power grid

The GCN model is not able to predict the dynamic stability for the Texan power grid. To understand this behaviour, we compare it to the ArmaNet model at the out-of-distribution task from dataset20 and dataset100 to predict SNBS of the Texan power grid. The scatter plots are shown in Figure 7. We can clearly see that the model is not able to predict lower values of SNBS correctly. The limited

Table 5: Performance after training on smaller training set. All models are trained on the same 800 grids as in Nauck et al. (2022), but evaluated on the newly introduced test set. The results are represented by R^2 score in %.

model	tr20ev20	tr100ev100	tr20ev100
Ar-bench	46.54 $\pm$ 2.378	59.73 $\pm$ 0.886	31.75 $\pm$ 1.204
ArmaNet	70.35 $\pm$ 1.226	**83.92** $\pm$ 0.263	55.84 $\pm$ 0.598
GCNNet	49.59 $\pm$ 0.513	61.18 $\pm$ 1.663	36.08 $\pm$ 0.625
SAGENet	50.15 $\pm$ 0.255	60.98 $\pm$ 0.279	39.89 $\pm$ 0.089
TAGNet	**74.77** $\pm$ 0.370	82.21 $\pm$ 0.017	**60.31** $\pm$ 0.732

output of the GCNNet results in a bad performance in case of the distribution of the Texan power grid that has three modes. As a consequence, a model that predicts the mean of the distribution would achieve better performance. Furthermore, we provide the scatter plots of the GCNNet for the three tasks dataset20, dataset100 and tr20ev100 in Figure 8, that can be compared to Figure 5.

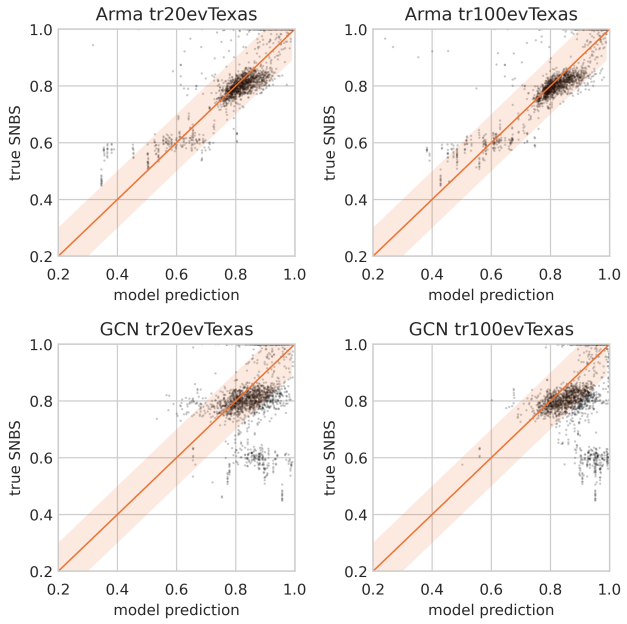

Figure 7: SNBS over predicted output of the Arma and GCN models for the out-of-distribution task to predict SNBS of the Texan power grid. The diagonal represents a perfect model ($R^2 = 1$), the banded region indicates prediction errors ≤ 0.1. To account for the small number of nodes, a lower transparency is used in comparison to Figures 5 and 8.

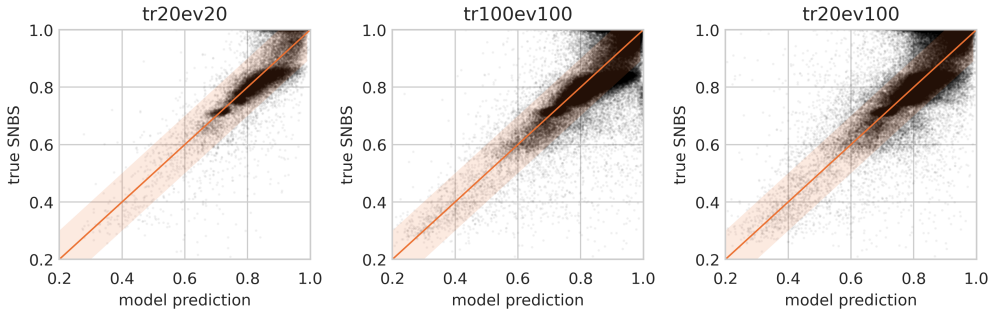

Figure 8: SNBS over predicted output of the GCNNet model for tr20ev20, tr100ev100 and tr20ev100. The diagonal represents a perfect model ($R^2 = 1$), the band indicates the region for for accurate predictions based on a threshold of 0.1.