Emergent Abilities of Graph Neural Networks for Large-scale Power System Analysis

Yuhong Zhu\textsuperscript{1}, Yongzhi Zhou\textsuperscript{1}, Lei Yan\textsuperscript{1}, Zuyi Li\textsuperscript{1}, Huanhai Xin\textsuperscript{1}, and Wei Wei\textsuperscript{1}

\textsuperscript{1}Affiliation not available

March 30, 2024

Abstract

The scale-up of AI models for analyzing largescale power systems necessitates a thorough understanding of their scaling properties. Existing studies on these properties provide only partial insights, showing that loss function decreases predictably with increased model scales; yet no scaling law for power system AI models has been established, and model performance remains unpredictable due to “emergent abilities”. This study pioneers the discussion on the emergent abilities of graph neural network (GNN) for analyzing large-scale power systems, revealing that model performance improves dramatically once model scale exceeds a threshold. Furthermore, we introduce an empirical power-law formula to quantify the relationship between this threshold and the power system size. Our theory accurately predicts the threshold for the appearance of emergent ability in large-scale power systems, including a synthetic 10,000- bus and a real-world 19,402-bus systems.
Emergent Abilities of Graph Neural Networks for Large-scale Power System Analysis

Yuhong Zhu, Yongzhi Zhou, Lei Yan, Zuyi Li, Huanhai Xin, and Wei Wei

Abstract—The scale-up of AI models for analyzing large-scale power systems necessitates a thorough understanding of their scaling properties. Existing studies on these properties provide only partial insights, showing that loss function decreases predictably with increased model scales; yet no scaling law for power system AI models has been established, and model performance remains unpredictable due to “emergent abilities”. This study pioneers the discussion on the emergent abilities of graph neural network (GNN) for analyzing large-scale power systems, revealing that model performance improves dramatically once model scale exceeds a threshold. Furthermore, we introduce an empirical power-law formula to quantify the relationship between this threshold and the power system size. Our theory accurately predicts the threshold for the appearance of emergent ability in large-scale power systems, including a synthetic 10,000-bus and a real-world 19,402-bus systems.

Index Terms—Graph neural networks, graph transformers, emergent abilities, large-scale power system analysis, scaling law.

I. INTRODUCTION

Graph neural network and its variant, graph attention network (GAT), have achieved state-of-the-art performance in multiple power system analysis tasks by effectively recognizing and processing power system topologies [1]–[3]. In the context of deepening interconnection of modern power grids, the abilities of AI models to scale for large-scale power systems are vital for their adoption in the power industry [4].

Understanding the scaling properties thoroughly is crucial for the success of large-scale AI models like Chatgpt 4 [5]. Research has suggested that model scale enlargement correlates with performance improvement. Such a relationship is quantified by the well-known scaling law [6], which empirically shows that reducible loss decreases linearly with the logarithmic increase of the model scale, i.e., characterized by the number of parameters contained in a model. Another noteworthy empirical observation is that models below a specific scale exhibit negligible performance, characterized by a loss nearly equivalent to that of a randomly initialized model. But once the scale is beyond a certain threshold, performance dramatically improves, a phenomenon termed “emergent abilities” by Google and Stanford researchers [7]. These insights have transformed the scaling of large AI models from a blind process into a methodology underpinned by empirical assurance [8]. Yet, existing scaling laws overlook the size characteristics of power systems, leading to a gap in scaling AI models to adapt to large-scale power systems. Consequently, one has to resort to speculative methods like grid search to determine model scale, leading to unnecessary computational resource expenditure.

This work bridges the gap by proposing an empirical yet quantitative formula from the perspective of emergent abilities. Drawing on a vast number of carefully designed experiments across power systems of varying sizes and GNNs of different types, we postulate that the scale threshold for a GNN’s emergent ability follows a power-law relationship with the power system’s size. This hypothesis has been corroborated when we forecast GNNs’ emergent abilities in large-scale power systems, including a synthetic 10,000-bus and a real-world 19,402-bus systems. Code is available at https://github.com/zyh1996saa/Emergent-Abilities-GNN-for-PS.

II. EMERGENT ABILITIES

A. A Quick Review of Emergent Abilities

Philip Anderson, the 1977 Nobel Prize-winning physicist, defines emergence in nature as a phenomenon where quantitative changes in a system lead to qualitative behavioral changes. In deep learning, such phenomenon is marked by increased computational resources and model parameters, directly influencing task-specific model performance. This concept underpins the definition that an ability is emergent if it is not present in smaller models but is present in larger models [7].

Evidence strongly indicates that large AI models inherently possess emergent abilities. Research from Google highlights that large models like GPT-3, along with LaMDA, Gopher, Chinchilla, and PaLM, display dramatic improvements in complex tasks such as multi-step arithmetic, college-level exams, and contextual word meaning identification, at certain model scales [9]. These observations underscore the critical need for exploring emergent abilities to inform the development and scaling of AI models designed for large-scale power system applications.

B. Observed Emergent Abilities of GAT models Across Power Systems of Varying Sizes

This study marks the initial exploration into the GNNs’ emergent abilities within diverse sizes of power systems, focusing on model generalization across two tasks: power flow approximation regression [3] and $P - 1$ static security assessment classification [10]. These tasks are selected for
their foundational role in enabling more complex analyses, including optimal power flow, contingency screening, and cascading failure prediction.

The ability to perform such tasks is emergent when a model has random performance (close to that of a randomly initialized model) until a certain scale, after which performance increases to well-above random. Fig.1 shows emergent abilities of GAT models spanning five power systems of varying sizes. We observe that GAT models exhibit emergent abilities when their parameter counts exceed certain thresholds—9910, 52819, 164988, 218930, and 412922—for power systems of increasing sizes.

The main parts of the experiment for Fig.1 are stated below.

1) Dataset: All GAT models are trained using an open-source dataset, which encompasses synthetic data from IEEE 300-bus system and synthetic power systems ranging from 1354 to 10000 buses (code available at https://github.com/zyh1996saa/Emergent-Abilities-GNN-for-PS). The dataset is partitioned into training and test sets, with the former for model training and the latter for performance evaluation. To assess model generalization, training and test data are designed to follow similar yet distinct distributions (see Fig.2), reflecting the challenge in real-world applications where training data cannot encompass all actual scenarios.

2) Model Configuration: The model architecture remains consistent with [3], using a 2-layer cascaded GAT layer connected to a 3-layer multi-layer perceptron (MLP). Each GAT layer contains 8 heads and is activated by LeakyReLU function, and dense connections are used between GATs. We gradually enlarge the model scale by increasing the number of neural units in the GAT and MLP layers until emergent ability is observed.

3) Training Strategy: All models are trained using the Adam optimizer, with a learning rate of 0.001 and epsilon set to 1e-7. An early stopping mechanism is implemented to prevent unnecessary training, activating when there is less than 0.0003 improvement in the model’s loss over 30 consecutive epochs. The loss functions used are root mean square error (MSE) for regression tasks and categorical cross-entropy for classification tasks.

4) Software and Hardware Environment: The models are implemented and trained using Tensorflow 2.11.0, leveraging its comprehensive API and optimized computational graph. Training and evaluation of the models are performed on Dell R760xd2 server, equipped with Nvidia A100 80G GPUs.

C. The Proposed Empirical Formula

In this section, we further investigate the quantitative relationship between the scale threshold of GNNs and the size of power systems. We document the model parameter counts of various GNNs at the onset of their emergence abilities across power systems of varying sizes, as illustrated in Fig.3. Experimental results show that the scale threshold of GNNs when emergent abilities appear and the size of the power system obey a power-law, that is:

\[ P = a \cdot (N - D)^b + c \]  

where \( N \) is the number of buses in the selected power system, \( P \) is the number of
parameters beyond which the GNN model exhibits emergent abilities. $a$, $b$, and $c$ are positive coefficients, which are fitted through the nonlinear least squares method. $D$ denotes a disturbance quantity related to the power system itself.

III. APPLICATION AND VERIFICATION

Formula (1) enables the prediction of the minimum parameter count required for a GNN to exhibit emergent abilities in large-scale power systems, bypassing the need for computationally intensive approaches like grid search.

We validate formula (1) through experiments that include: 1) incrementally enlarging the model scale by increasing contained parameters for systems of 300-bus, 3012-bus, 4601-bus, and 8387-bus until emergent abilities are observed; 2) using the nonlinear least squares method to fit coefficients in formula (1) based on the parameter threshold identified in step 1, and predicting the parameter thresholds for emergent abilities in systems of 10,000-bus and 19,402-bus; 3) maintaining the model structure constant, we further increase model parameters for the 10,000-bus and 19,402-bus systems to observe the emergence abilities. Comparative results of these experiments are depicted in Fig.4. We observe that the prediction error is less than 5%, indicating the effectiveness of the proposed empirical formula.

IV. CONCLUSION

This letter introduces a pioneering approach to scale AI models for large-scale power system analysis, emphasizing the critical role of emergent abilities in GNNs. By developing an empirical power-law formula, we bridge the gap in the predictive scaling of GNNs, correlating model scale with power system size. The empirical validation across various power systems underscores the formula’s efficacy and sets a new precedent for AI model scaling in the power sector.

Future work will aim to refine this formula, assess its applicability to various AI architectures, and broaden its utility for different analysis tasks. We contend that this work will guide the development towards a compute-optimal AI model for large-scale power system analysis.

REFERENCES


