Exploring Variational Graph Autoencoders for Distribution Grid Data Generation

Syed Zain Abbas, Ehimare Okoyomon

TECHNISCHE UNIVERSITÄT MÜNCHEN

NEURAL INFORMATION PROCESSING SYSTEMS

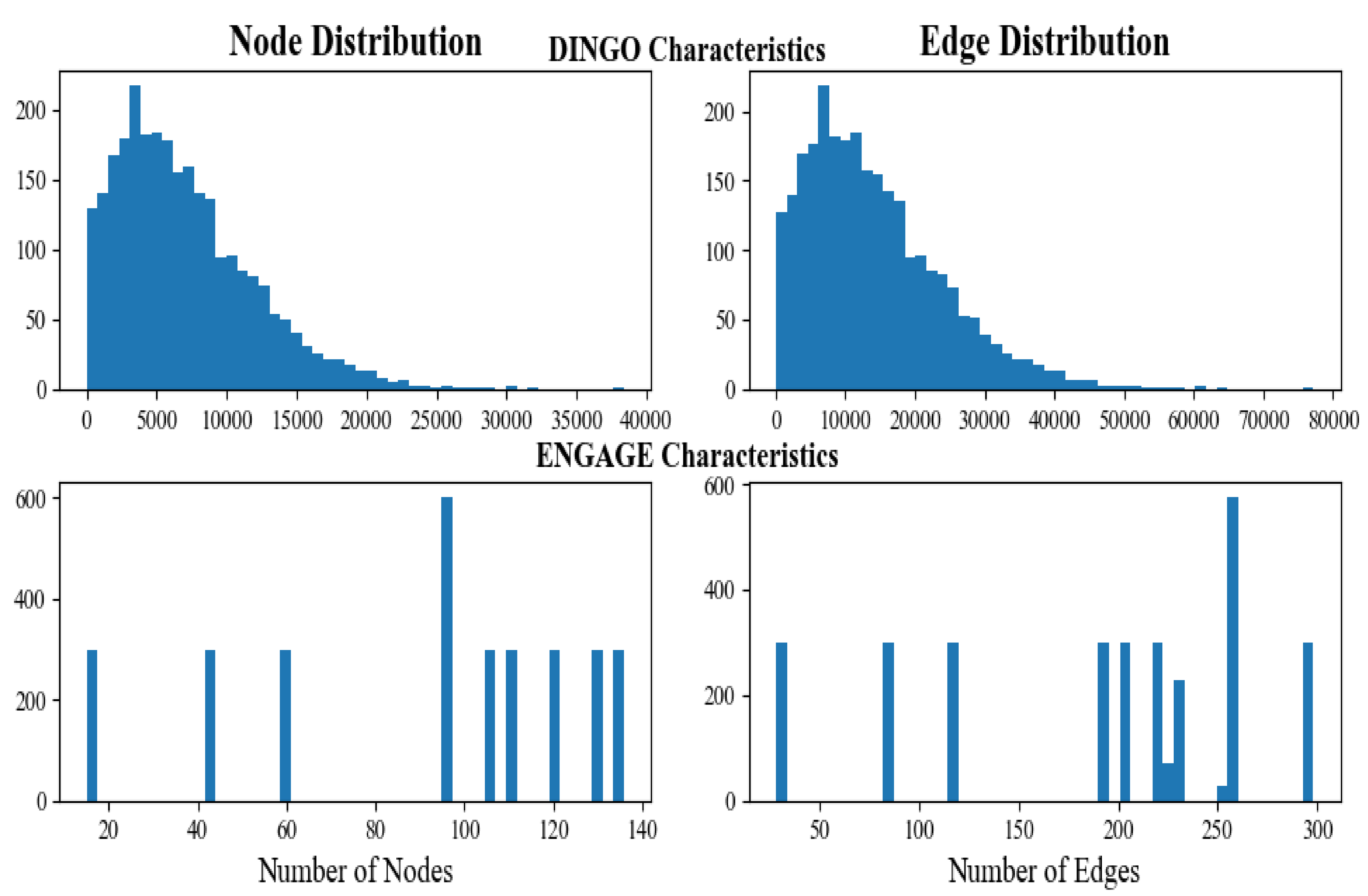
Why synthetic power grid data?

- Distribution grids are critical for renewable integration and flexibility planning.
- Real feeder data is **rarely shared — security & privacy concerns** limit access.
- Synthetic datasets enable algorithm benchmarking without exposing sensitive infrastructure.
- Current methods rely on statistical models or heuristic algorithms, often missing complex topological correlations and producing unrealistic or overly simplified networks.
- Graph generative models offer a new paradigm — already successful in biology & chemistry.
- Few applications exist for power grids, which motivates the exploration of Variational Graph Autoencoders (**VGAEs**).

How to generate realistic grids?

Two open-source datasets are used to evaluate scalability and generalization:

- **ENGAGE**: Based on SimBench feeders, small (≈100 nodes), discrete topologies; 3000 training grids.
- **DINGO**: Large, diverse collection (4,500–7,000 nodes), reflecting real grid variability; 2722 training grids.



VGAE Model Architecture

The VGAE framework consists of a GCN-based encoder and a probabilistic latent space. Four decoder variants were evaluated:

- **Inner Product**
- **MLP**
- **GCN**
- **Iterative GCN** (with refinement loop)

Objective & Metrics

Loss Function:

$$\mathcal{L} = \mathcal{L}_{reconstruction} + \beta \cdot \mathcal{L}_{KL}$$

Evaluation Metrics:

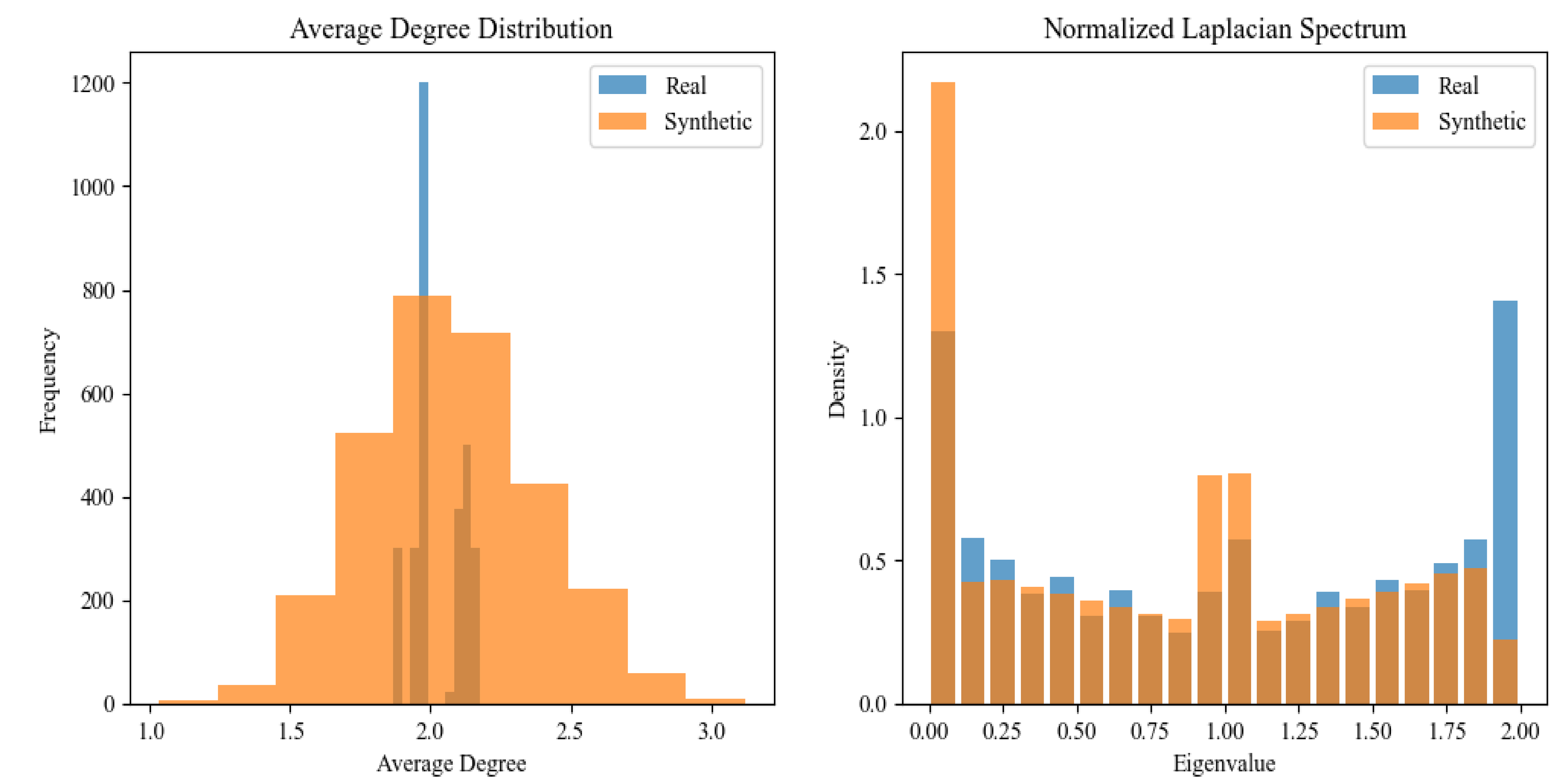
- Average Degree — measures connectivity & radiality.
- Laplacian Spectrum (Wasserstein Distance) — measures global structural similarity.

What did we find?

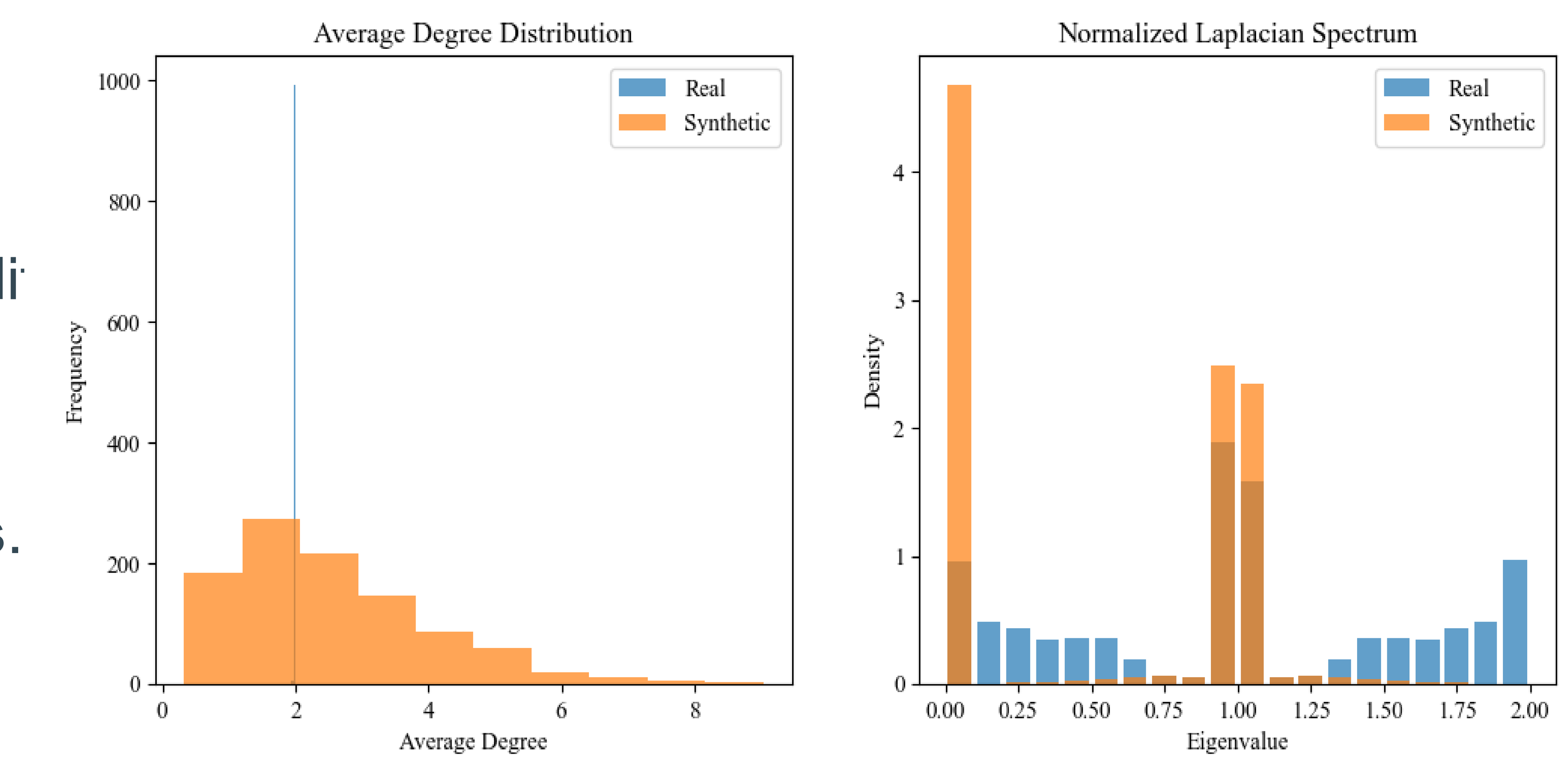
- Iterative-GCN consistently achieved the best structural fideli across both datasets.
- **ENGAGE**:
  1. Synthetic grids are **aligned closely with real topologies**.
  2. Strong agreement in both average degree and spectral metrics.
- **DINGO**:
  1. All models **struggled with scale and diversity**.
  2. Common artifacts: over-connected components, repeated motifs.
  3. Reveals scalability limitations of current VGAE architectures.
- **Takeaway**:
  1. Simple VGAE models reproduce small, uniform feeders well.
  2. Large, heterogeneous grids demand more expressive and physics-aware generative methods.

Dataset	Avg. Degree (Real)	Avg. Degree (Synthetic)	Wasserstein Distance
ENGAGE	2.0521	2.0697	0.1039
DINGO	1.9986	2.5300	0.5072

ENGAGE Results



DINGO Results



Future Work

Richer decoders: Explore attention-based or diffusion architectures to improve expressiveness.

Physics-aware learning: Incorporate power-flow feasibility and operational constraints during generation.

Scalability: Extend models and datasets to handle larger, more diverse grid topologies.