

Toward Graph Foundation Models via Synthetic Graph Priors

Jeongwhan Choi (최정환)

Jang Young-Sil Postdoctoral Fellow (Excellence Track)

Information & Electronics Research Institute, KAIST · jeongwhan.choi@kaist.ac.kr

KCC 2026 · 신진연구자 세션



Who I Am

- **Postdoc & Jang Young-Sil Fellow, KAIST** 2025.09 / 2026.04–
Information & Electronics Research Institute · advisor: Noseong Park
- **Ph.D. in Artificial Intelligence, Yonsei University** 2020–2025
co-advised by Sung-Bae Cho & Noseong Park
- **B.S. in Software Engineering, Jeonbuk Nat'l University** 2016–2020
Presidential Award for Outstanding Undergraduate

Selected recognition

- 10+ first-author papers at **ICLR, NeurIPS, ICML, SIGIR, CIKM, AAAI**
- **2× AAAI Oral** (2022, 2026) · **1× KDD Oral** (2026)
- Qualcomm Innovation Fellowship Korea 2024 Winner
- KAIST–Samsung DS Best Paper Runner-up 2025

Research Roadmap

Graph-based Deep Learning

Signal Processing on Graphs

spectral graph filters

GREAD · GFSA · Fractal Node

Differential-Equation-Based DL

continuous-depth dynamics

STG-NCDE · LT-OCF · NCDENow

Scientific AI & Mesh GNNs

physics-informed simulation

HCMT · PIORF · PANDA

NOW

NodePFN

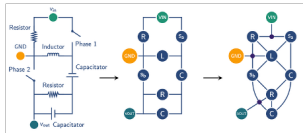
ICLR 2026

in-context learning
over graphs

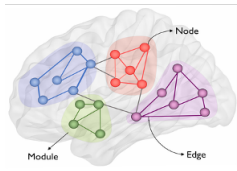
THE GOAL

Graph
Foundation
Models

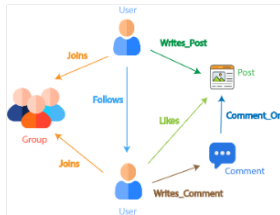
Graphs Are Everywhere



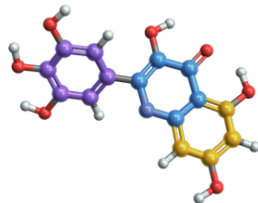
Circuits
devices · wires



Brain
regions · links

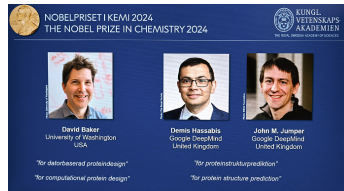


Social / web
users · links



Molecules
atoms · bonds

- Just **nodes** and **edges** — yet they model much of science and engineering.
- **Why now?** The **2024 Nobel Prize in Chemistry**: AlphaFold casts protein structure as *3D graph inference*.



Why Graph Foundation Models Are Hard

- We still **train a new model for every graph**.
- Unlike text/images, there is **no web-scale corpus of labeled graphs** to pre-train on.
- Real graphs also vary by regime (homophily, communities, features), so a single trained model rarely generalizes across graphs.

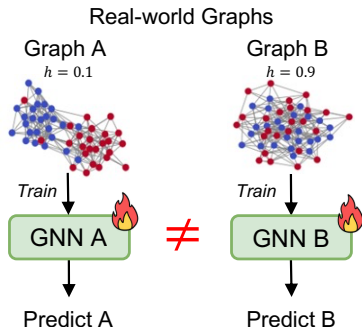
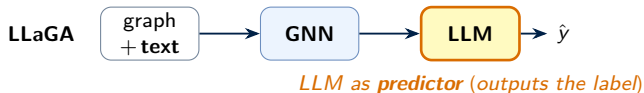


Figure: A new model for every graph

Question: Can a *single* model classify nodes on *any* graph, with *no* graph-specific training?

The Gap: Existing Graph FMs Lean on LLMs

- Most “graph foundation model” attempts need **text-attributed** graphs and an **LLM** in the loop^{[1][2]} — the LLM just plays *different roles*.



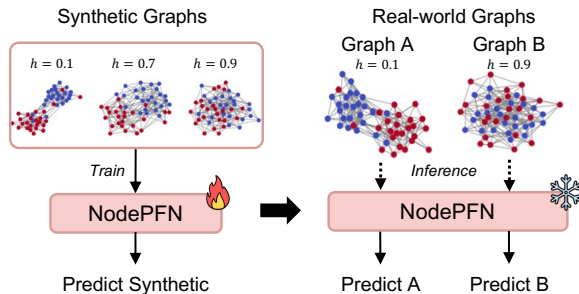
Enhancer or predictor — both *depend on an LLM and text*. **NodePFN** needs neither —
so where does it learn from?

[1] Hao Liu et al. “One For All: Towards Training One Graph Model For All Classification Tasks”. In: *ICLR*. 2024.

[2] Runjin Chen et al. “Llaga: Large language and graph assistant”. In: *arXiv preprint arXiv:2402.08170* (2024).

Our Idea: Learn from Synthetic Graphs

- NodePFN **never trains on real graphs**. We pre-train *once* on **synthetic** graphs.
- At test time, read labeled nodes as **context** and predict the rest in **one forward pass** (in-context).



Why does training on *synthetic* graphs generalize? It learns a **posterior predictive distribution**.

Learning Target: PPD — Approximated by a Prior-Fitted Network

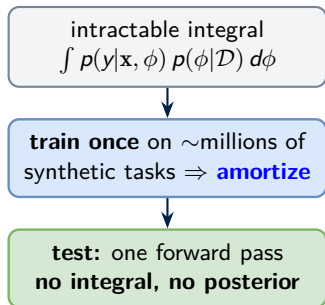
- **Setup.** Predict a **test** node from labeled **train** nodes as context, $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, y_i)\}$.
- **PPD.** Rather than pick one **hypothesis** ϕ , marginalize over *all* hypotheses, weighted by posterior:

$$\underbrace{p(y_{\text{test}} \mid \mathbf{x}_{\text{test}}, \mathcal{D}_{\text{train}})}_{\text{PPD}} = \int \underbrace{p(y_{\text{test}} \mid \mathbf{x}_{\text{test}}, \phi)}_{\text{prediction under } \phi} \underbrace{p(\phi \mid \mathcal{D}_{\text{train}})}_{\text{posterior over } \phi} d\phi$$

- The **best possible** predictor (lowest expected loss), but **intractable**.
- **Müller et al. (2022)**: train a **Transformer** f_θ on **unlimited synthetic datasets** (the prior).

Theorem^[3]: loss minimization $\Rightarrow$ PPD

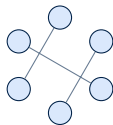
With enough capacity, minimizing the loss makes f_θ *become* the true PPD.



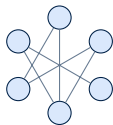
[3] Samuel Müller et al. "Transformers Can Do Bayesian Inference". In: *ICLR*. 2022.

Synthetic Graph Priors: Inheriting Network Science

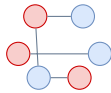
- We don't hand-design a prior — we **inherit** the canonical **random graph models** of network science as **inductive bias**.



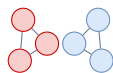
Sparse ER



Dense ER



cSBM, low h



cSBM, high h



BA, hub

- Random (ER): pure random edges. (Erdős–Rényi, 1959)^[4]
- Contextual stochastic block model (cSBM): planted **community** blocks with node features. (Holland 1983^[5]; Deshpande 2018^[6])
- Scale-free (BA): a few high-degree **hubs**. (Barabási–Albert, 1999)^[7]

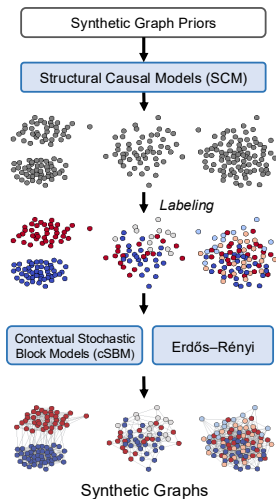
[4] Paul Erdős and Alfréd Rényi. “On random graphs I”. In: *Publ. math. debrecen* 6.290–297 (1959), p. 18.

[5] Paul W Holland et al. “Stochastic blockmodels: First steps”. In: *Social networks* 5.2 (1983), pp. 109–137.

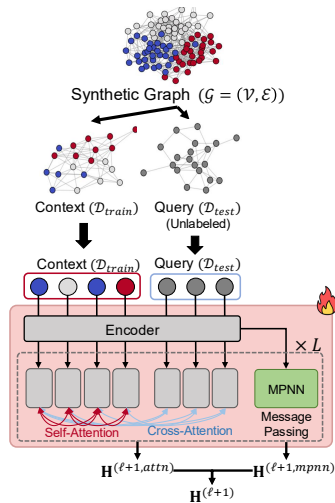
[6] Yash Deshpande et al. “Contextual stochastic block models”. In: *Advances in neural information processing systems* 31 (2018).

[7] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.

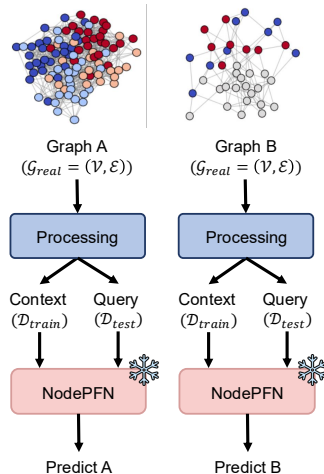
Method: Three Phases & In-Context Mechanism



(a) Generate priors



(b) Pretrain (once)



(c) Inference (one pass)

Results: One Model on 23 Benchmarks

- One pre-trained NodePFN on **23 real benchmarks** (13 homophily + 10 heterophily) — **no per-graph tuning**.

Table: Average over 23 node-classification benchmarks (a single pre-trained NodePFN).

Method	Homo.(13)	Hetero.(10)	Avg.	Rank↓
MLP	56.43	58.17	57.30	7.41
GCN	73.05	58.84	66.63	5.86
GAT	74.39	59.11	67.67	5.57
GraphAny ^[8]	71.45	61.62	66.24	4.38
NodePFN	77.39	65.14	71.27	1.70

[8] Jianan Zhao et al. "Fully-inductive Node Classification on Arbitrary Graphs". In: *ICLR*. 2025.

Results: vs. Training-free & LLM-based GFM

Table: Training-free models vs. NodePFN.

Method	Cora	Pubmed	Wisc.	Texas
Linear	52.80	59.30	80.00	32.35
SGC ^[9]	78.20	72.98	57.64	46.03
HGC	22.50	46.32	64.32	57.54
LabelProp ^[10]	60.30	63.44	16.08	23.53
TFGNN ^[11]	60.03	40.04	14.51	19.91
NodePFN	82.06	78.00	81.18	76.22

Table: LLM-based GFMs under GLBench^[12]

Method	Cora	Cite.	Pubmed	WikiCS
InstructGLM ^[13]	69.10	51.87	71.26	45.73
GraphText ^[14]	<u>76.21</u>	59.43	<u>75.11</u>	67.35
LLaGA ^[15]	74.42	55.73	68.82	73.88
OFA ^[16]	75.24	73.04	75.61	<u>77.34</u>
NodePFN	76.38	<u>63.08</u>	68.18	76.29

[9] Felix Wu et al. "Simplifying graph convolutional networks". In: *ICML*. 2019.

[10] Xiaojin Zhu and Zoubin Ghahramani. "Learning from labeled and unlabeled data with label propagation". In: *Technical Report CMU-CALD-02-107*. 2002.

[11] Ryoma Sato. *Training-free Graph Neural Networks and the Power of Labels as Features*. 2024.

[12] Yuhan Li et al. "Glbentch: A comprehensive benchmark for graph with large language models". In: *NeurIPS* (2024).

[13] Ruosong Ye et al. "Language is all a graph needs". In: *arXiv preprint arXiv:2308.07134* (2023).

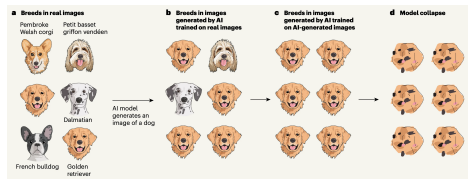
[14] Jianan Zhao et al. "Graphtext: Graph reasoning in text space". In: *arXiv preprint arXiv:2310.01089* (2023).

[15] Runjin Chen et al. "Llaga: Large language and graph assistant". In: *arXiv preprint arXiv:2402.08170* (2024).

[16] Hao Liu et al. "One For All: Towards Training One Graph Model For All Classification Tasks". In: *ICLR*. 2024.

Vision & Future Work — The Bigger Picture

- In many domains real data is **locked**: medical (privacy), semiconductor (trade secrets), finance (compliance).
- Worse: public data may be **exhausted this decade**^[17], and training on AI-generated data risks **model collapse**^[18].
- **Synthetic-prior pretraining** (NodePFN's recipe) is a path forward for **data-scarce / data-locked** domains.



“With the right mathematical prior, we can build a working foundation model for a domain.”

[17] [Katrina Krämer](#). “AI & robotics briefing: The year machines will run out of data.”. In: *Nature* (2024).

[18] [Iliia Shumailov et al.](#) “AI models collapse when trained on recursively generated data”. In: *Nature* 631.8022 (2024), pp. 755–759.

Vision & Future Work — Toward Graph Foundation Models

- Broaden the paradigm(*ongoing*)
 - Multi-task PFN: link prediction, graph classification.
 - **SRPFN** for sequential recommendation — **KDD 2026**.
- Scale to real graphs
 - Linear / efficient attention for **100k+ node** graphs.
- Richer, principled priors
 - My scientific-AI / dynamics line recast as **PDE-/physics-based priors**.
- Multimodal & trustworthy
 - **PFN + LLM** (structured priors + language).
 - **Privacy-preserving** via synthetic-only training.

My goal: **graph foundation models that generalize across domains without large-scale real-data pretraining.**

Thank You!

Questions?

Jeongwhan Choi (최정환) · KAIST

Homepage: jeongwhanchoi.com

Email: jeongwhan.choi@kaist.ac.kr



Reference I

- [1] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.
- [2] Runjin Chen et al. “Llaga: Large language and graph assistant”. In: *arXiv preprint arXiv:2402.08170* (2024).
- [3] Yash Deshpande et al. “Contextual stochastic block models”. In: *Advances in neural information processing systems* 31 (2018).
- [4] Paul Erdős and Alfréd Rényi. “On random graphs I”. In: *Publ. math. debrecen* 6.290-297 (1959), p. 18.
- [5] Aditya Grover and Jure Leskovec. “Node2vec: Scalable Feature Learning for Networks”. In: *KDD*. 2016, pp. 855–864.
- [6] Paul W Holland et al. “Stochastic blockmodels: First steps”. In: *Social networks* 5.2 (1983), pp. 109–137.

Reference II

- [7] Noah Hollmann et al. “TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second”. In: *ICLR*. 2023.
- [8] Katrina Krämer. “AI & robotics briefing: The year machines will run out of data.”. In: *Nature* (2024).
- [9] Yuhan Li et al. “Glbench: A comprehensive benchmark for graph with large language models”. In: *NeurIPS* (2024).
- [10] Hao Liu et al. “One For All: Towards Training One Graph Model For All Classification Tasks”. In: *ICLR*. 2024.
- [11] Samuel Müller et al. “Transformers Can Do Bayesian Inference”. In: *ICLR*. 2022.
- [12] Ryoma Sato. *Training-free Graph Neural Networks and the Power of Labels as Features*. 2024.

Reference III

- [13] Ilia Shumailov et al. “AI models collapse when trained on recursively generated data”. In: *Nature* 631.8022 (2024), pp. 755–759.
- [14] Jian Tang et al. “LINE: Large-Scale Information Network Embedding”. In: *TheWebConf (former WWW)*. 2015, pp. 1067–1077.
- [15] Felix Wu et al. “Simplifying graph convolutional networks”. In: *ICML*. 2019.
- [16] Ruosong Ye et al. “Language is all a graph needs”. In: *arXiv preprint arXiv:2308.07134* (2023).
- [17] Jianan Zhao et al. “Fully-inductive Node Classification on Arbitrary Graphs”. In: *ICLR*. 2025.
- [18] Jianan Zhao et al. “Graphtext: Graph reasoning in text space”. In: *arXiv preprint arXiv:2310.01089* (2023).
- [19] Xiaojin Zhu and Zoubin Ghahramani. “Learning from labeled and unlabeled data with label propagation”. In: *Technical Report CMU-CALD-02-107*. 2002.