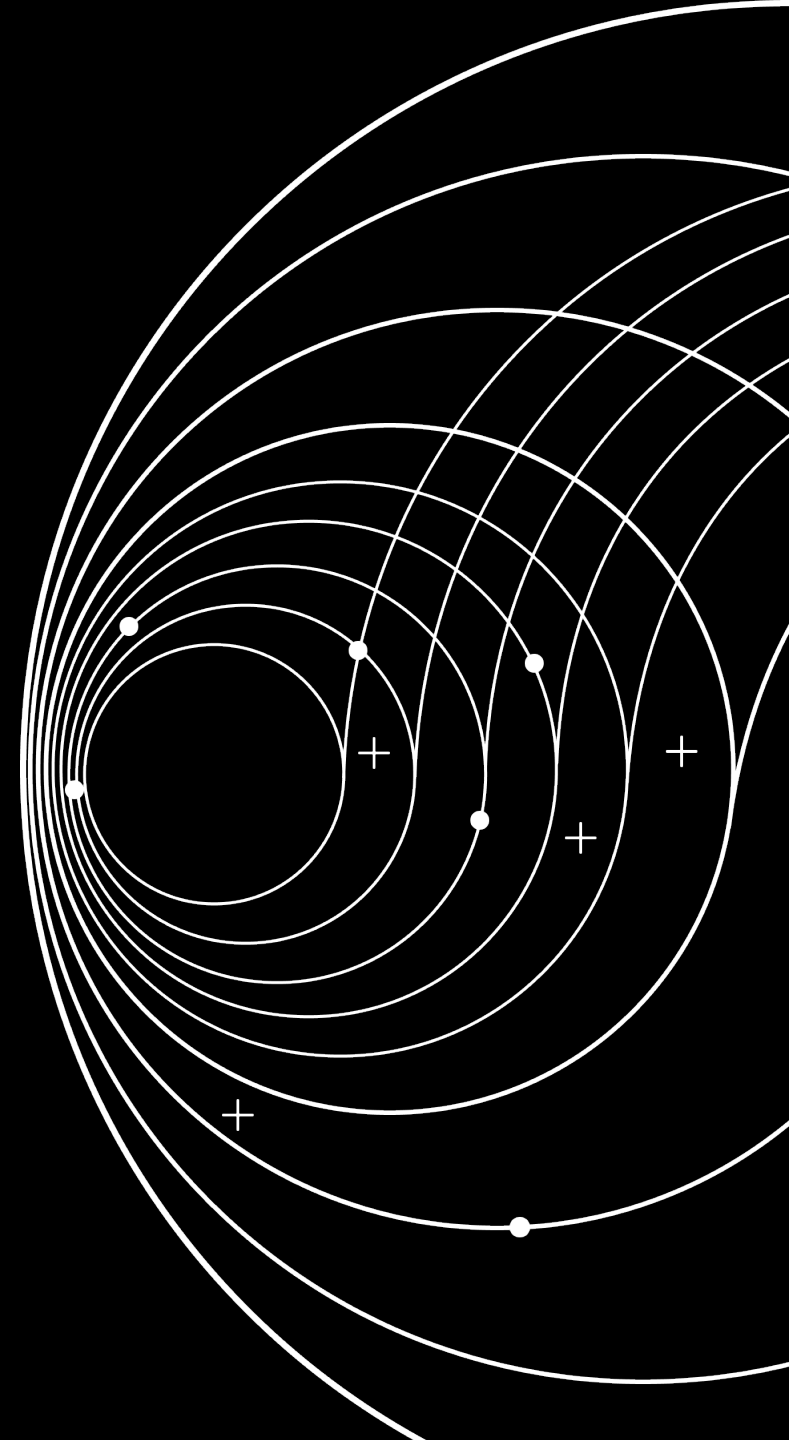


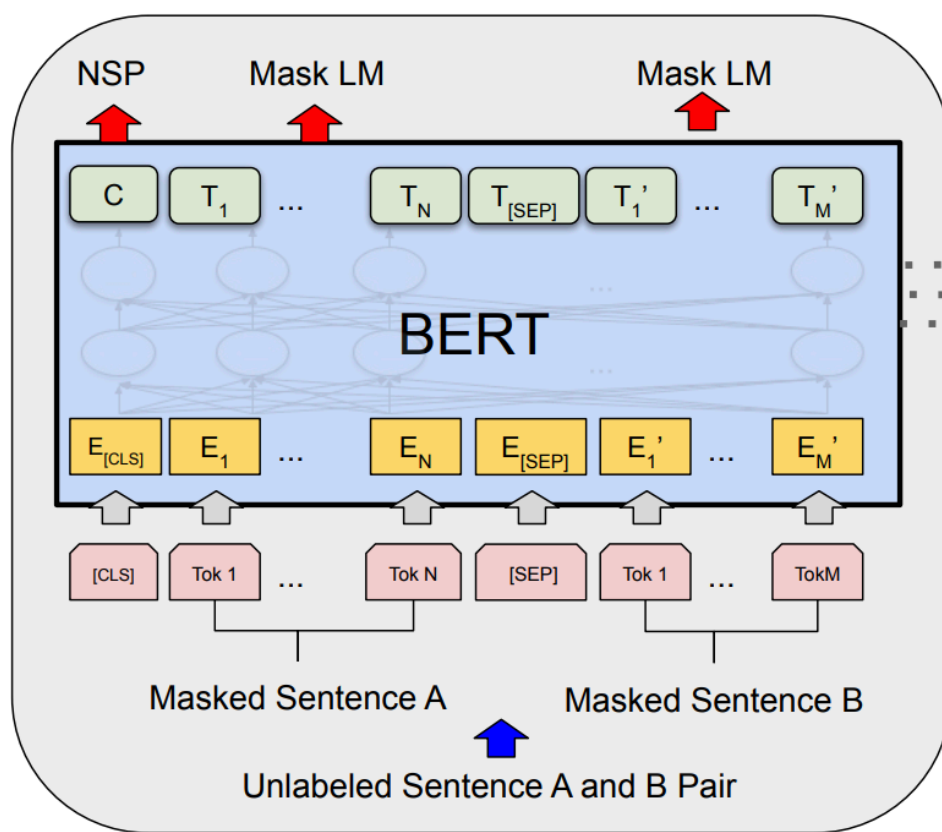
Yandex Research

Graph foundation models

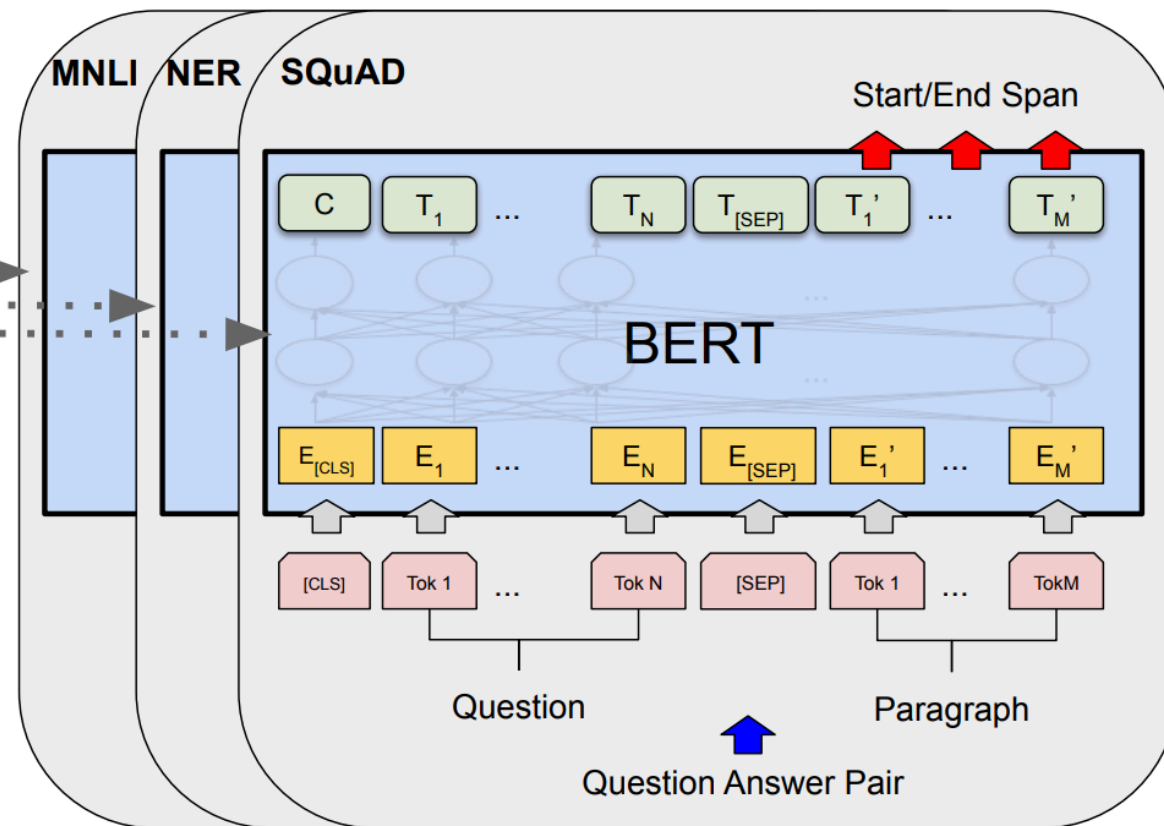
Dmitry Ereemeev



Motivation: FMs in other domains



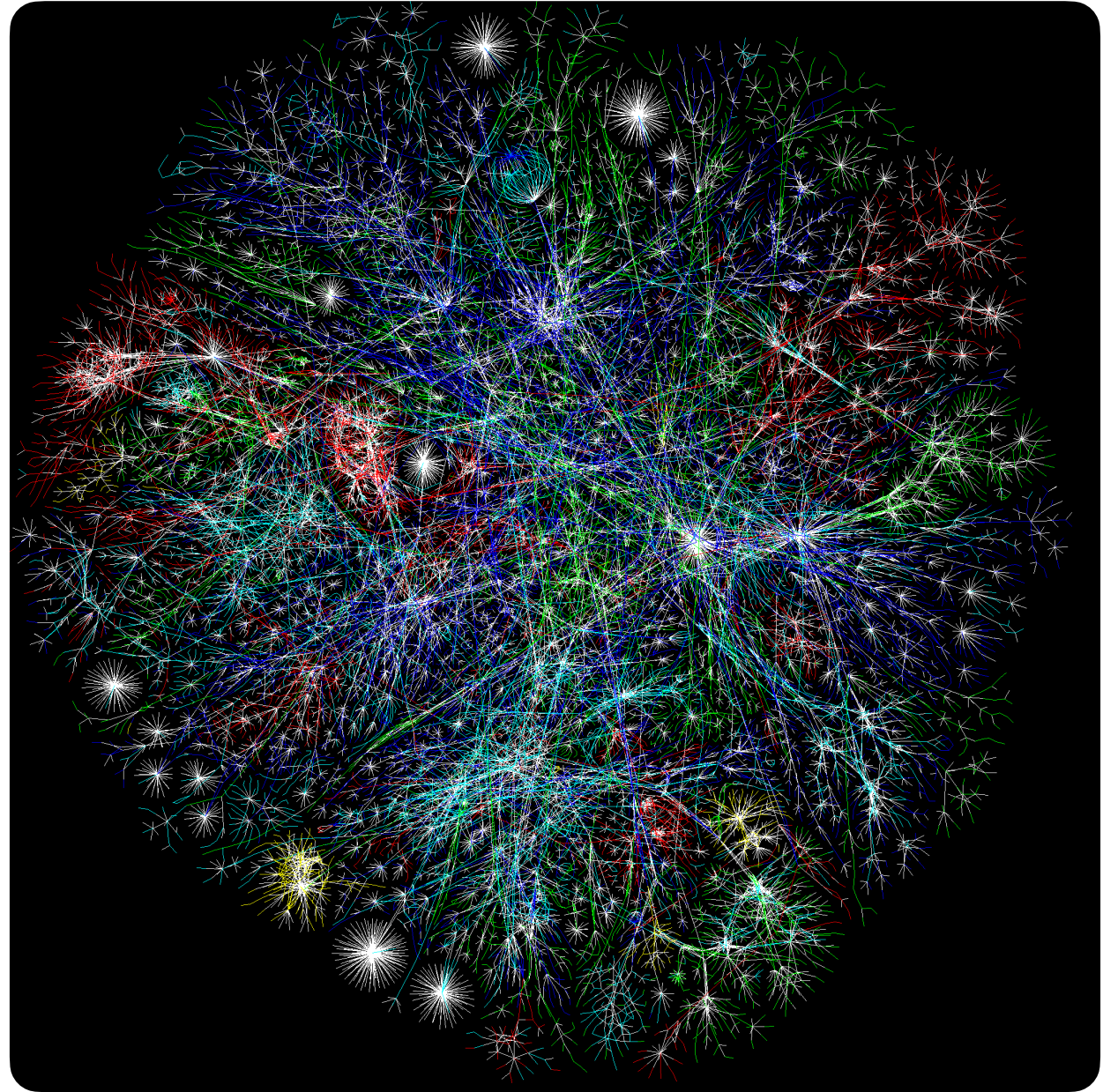
Pre-training



Fine-Tuning

Setup

- Large graph
- All nodes have features
- Some nodes have labels, want to predict labels for remaining nodes
- Non-fewshot splits
- Arbitrary feature/label spaces



Key challenges for GFMs

Transfer across datasets

- Different semantics of nodes/edges
- Different feature/label spaces
- Diverse domains

Data scarcity

- Amount of publicly available diverse data is significantly lower compared to CV/NLP
- Even if we collect large graph (e.g., web graph), diversity remains an issue

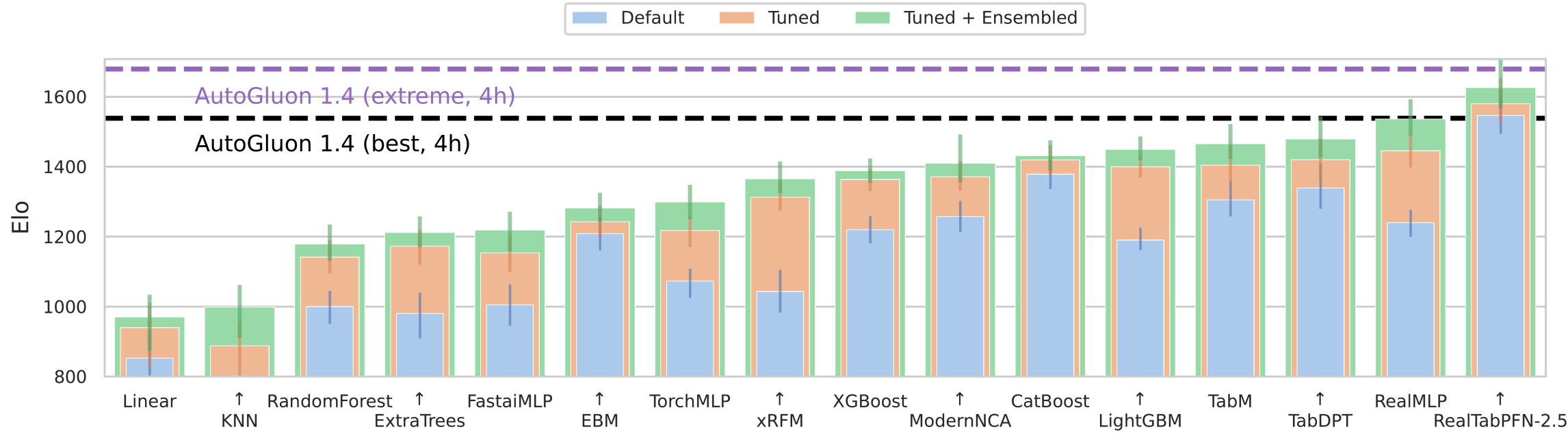
GFM on GraphLand

Performance on GraphLand under the 10/10/80 train/val/test split in the transductive setting.
Some publicly available GFMs published before Aug 2025.

Model	Multiclass classification (accuracy)			Binary classification (average precision)			
	hm-categories	pokec-regions	web-topics	tolokers-2	city-reviews	artnet-exp	web-fraud
GCN	61.70 ± 0.35	34.96 ± 0.38	46.45 ± 0.10	51.32 ± 0.96	77.15 ± 0.28	43.09 ± 0.38	10.02 ± 0.18
OpenGraph (ICL)	9.49 ± 0.93	1.73 ± 0.31	RTE	40.49 ± 0.31	58.44 ± 1.08	15.65 ± 1.23	RTE
AnyGraph (ICL)	15.47 ± 2.36	24.65 ± 1.51	6.67 ± 3.88	31.33 ± 2.89	64.37 ± 1.29	13.14 ± 1.15	0.68 ± 0.03
GCOPE (FT)	19.51 ± 0.07	TLE	TLE	28.67 ± 1.42	67.38 ± 1.23	16.10 ± 2.79	TLE

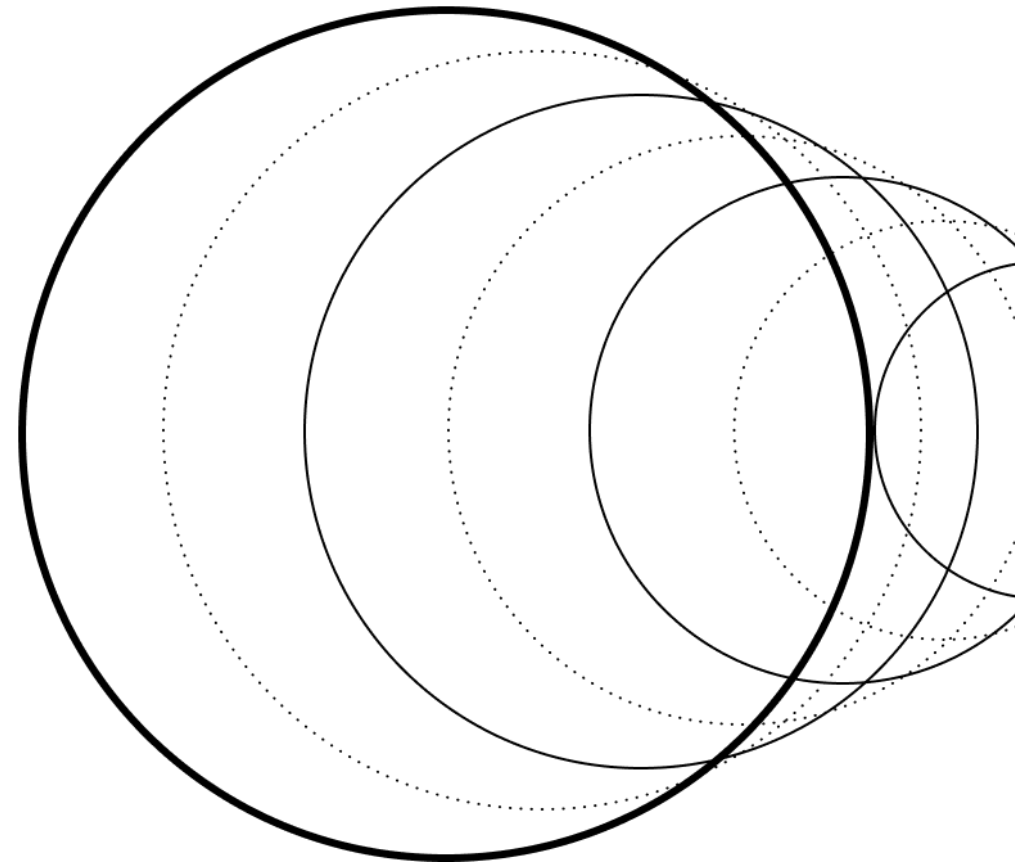
The End?

Tabular Foundation Models

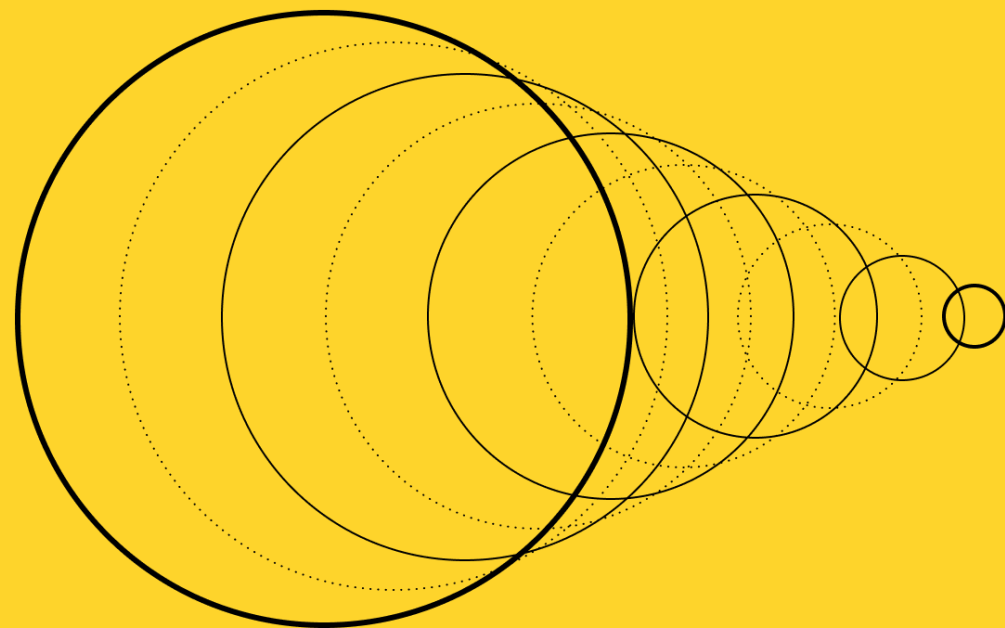


Caveats

- 01 There could be working GFM's for text-attributed graphs and few-shot settings
- 02 There could be working GFM's that are not publicly available (e.g., [KumoRFM](#))



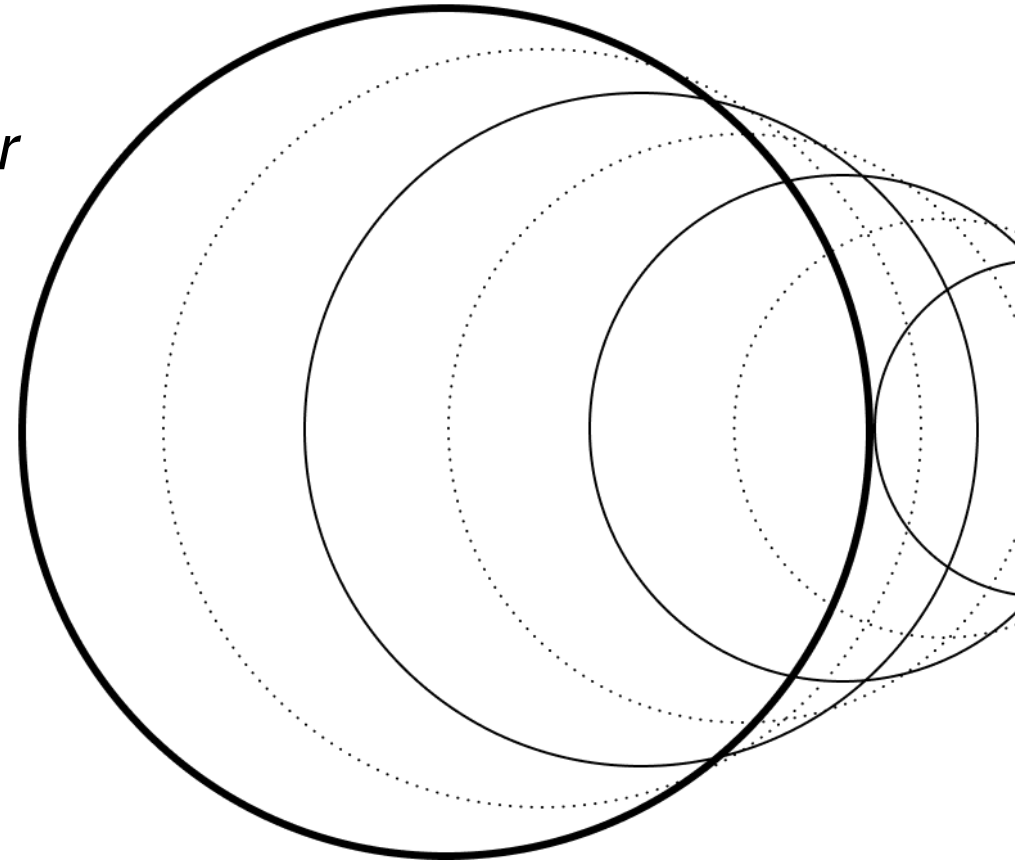
Yandex Research



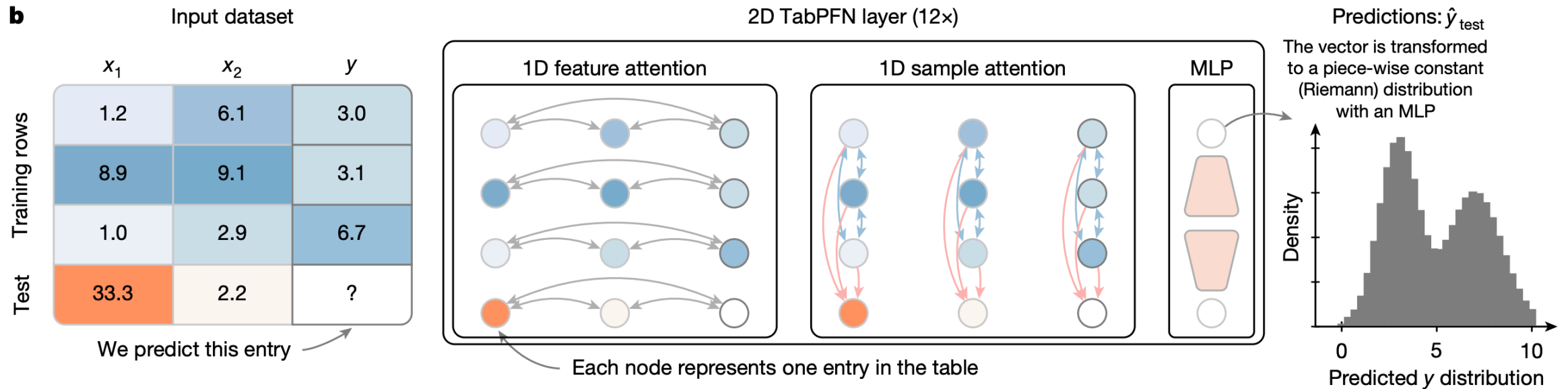
Prior-data fitted networks

PFN Framework

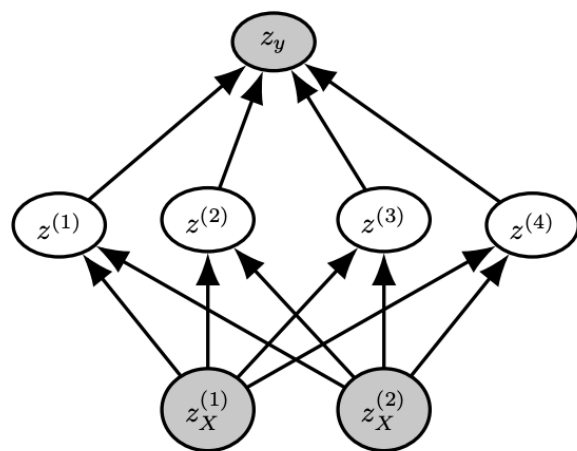
- 01 Predictions in an in-context learning regime
- 02 Pretrained on synthetic datasets from a *prior*



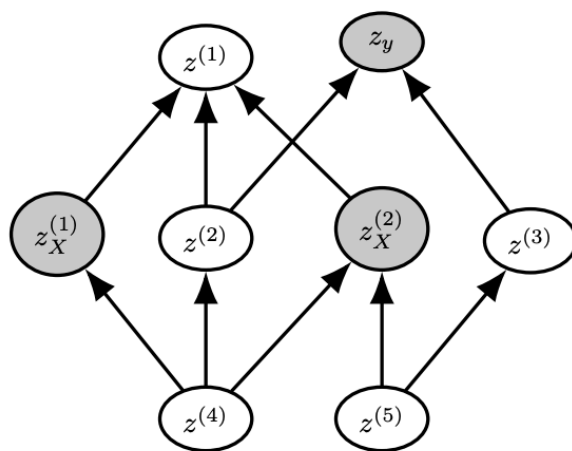
In-context learning



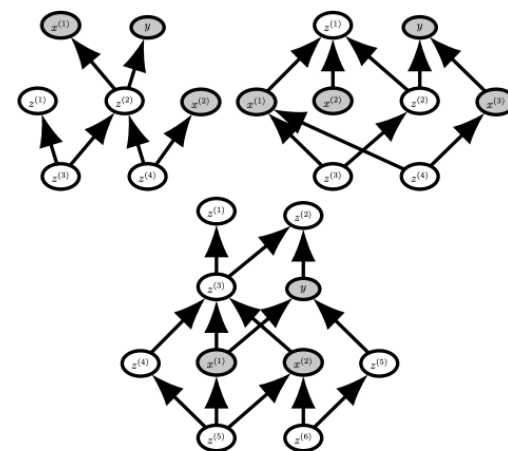
Prior for synthetic dataset generation



(a) A BNN



(b) An SCM



(c) SCMs sampled from the prior

Figure 2: Overview of graphs generating data in our prior. Inputs x are mapped to the output y through unobserved nodes z . Plots based on Müller et al. (2022).

Theoretical motivation

Insight 1. *The proposed objective ℓ_θ is equal to the expectation of the cross-entropy between the PPD and its approximation q_θ : $\ell_\theta = \mathbb{E}_{x,D \sim p(\mathcal{D})} [H(p(\cdot|x, D), q_\theta(\cdot|x, D))]$*

Proof. The above can be shown with a simple derivation. We mark transformations for easy reading.

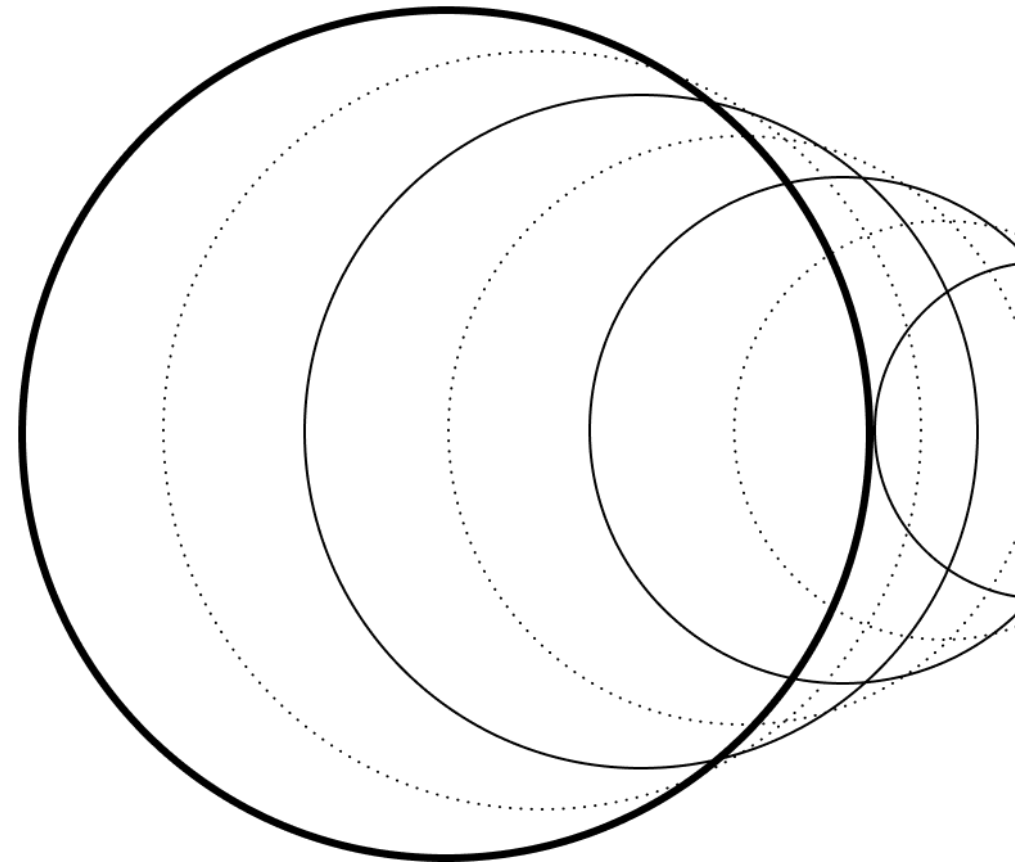
$$\ell_\theta = - \int_{D,x,y} p(x, y, D) \log q_\theta(y|x, D) = - \int_{D,x} p(x, D) \int_y p(y|x, D) \log q_\theta(y|x, D) \quad (3)$$

$$= \int_{D,x} p(x, D) \mathbf{H}(p(\cdot|x, D), q_\theta(\cdot|x, D)) = \mathbb{E}_{x,D \sim p(\mathcal{D})} [\mathbf{H}(p(\cdot|x, D), q_\theta(\cdot|x, D))] \quad (4)$$

□

Hypothesis about PFNs success

- 01 Synthetics is a good option for tabular tasks
- 02 ICL allows to capture dataset-level patterns



ICL allows to capture dataset-level patterns?

Imagine you are ResNet-like TFM, and you need to produce embedding for each row.

ICL allows to capture dataset-level patterns?

Imagine you are ResNet-like TFM, and you need to produce embedding for each row.

What information will you store for this row?

27.0,12.0,1.0,28.0,68.0,0.0,0.3971,0.1765,0.0147,0.4118,0.0,23.0,0,1,1,3

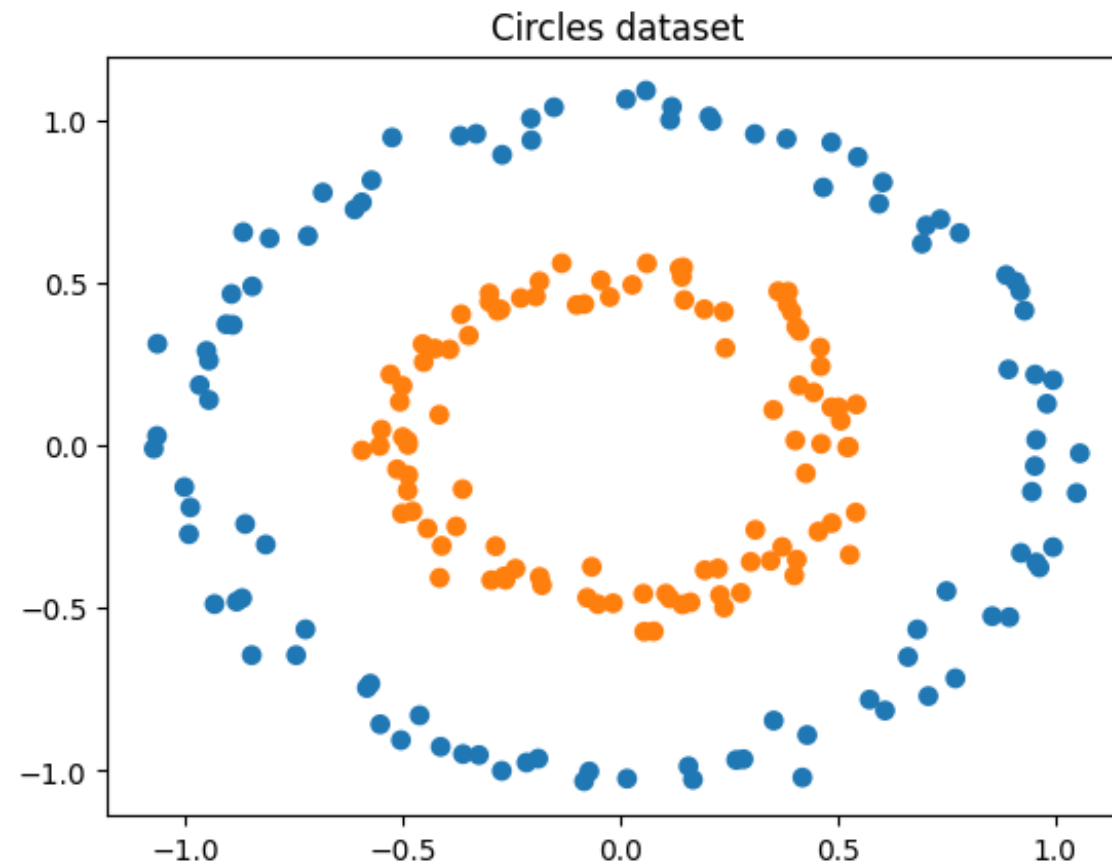
ICL allows to capture dataset-level patterns?

Imagine you are ResNet-like TFM, and you need to produce embedding for each row.
What information will you store for this row? And for this one?

0.0, 1.0

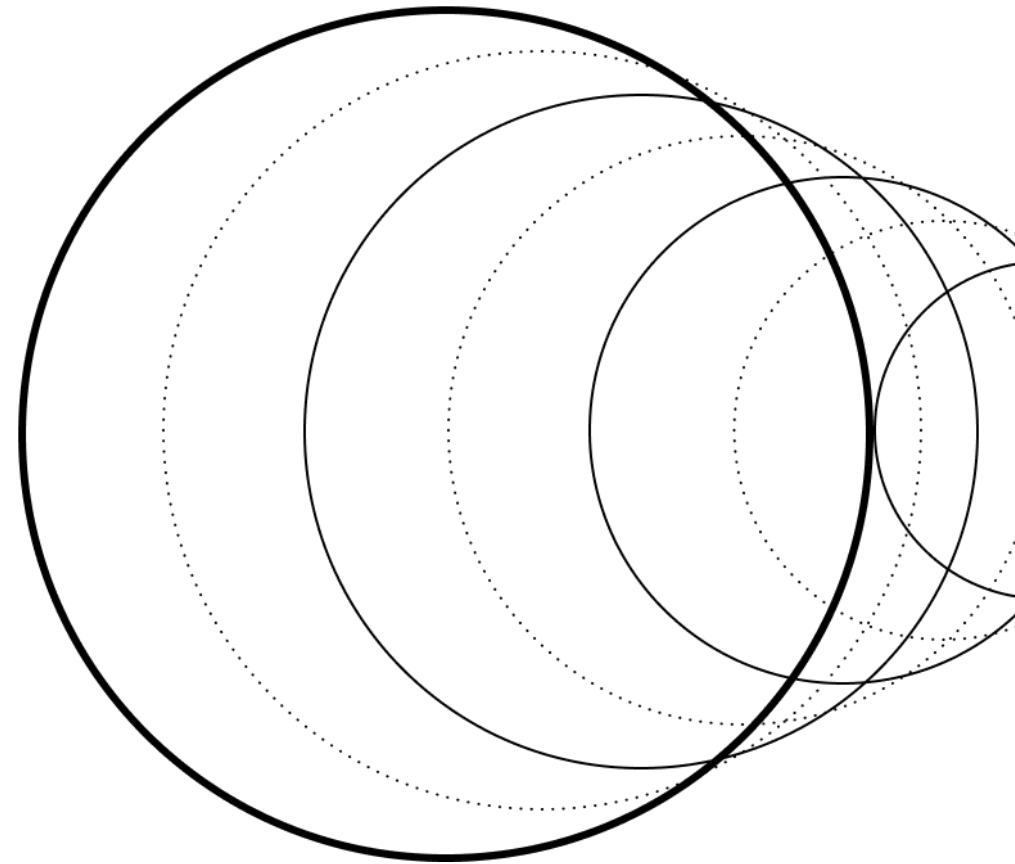
ICL allows to capture dataset-level patterns?

And if you are able to see the whole dataset?



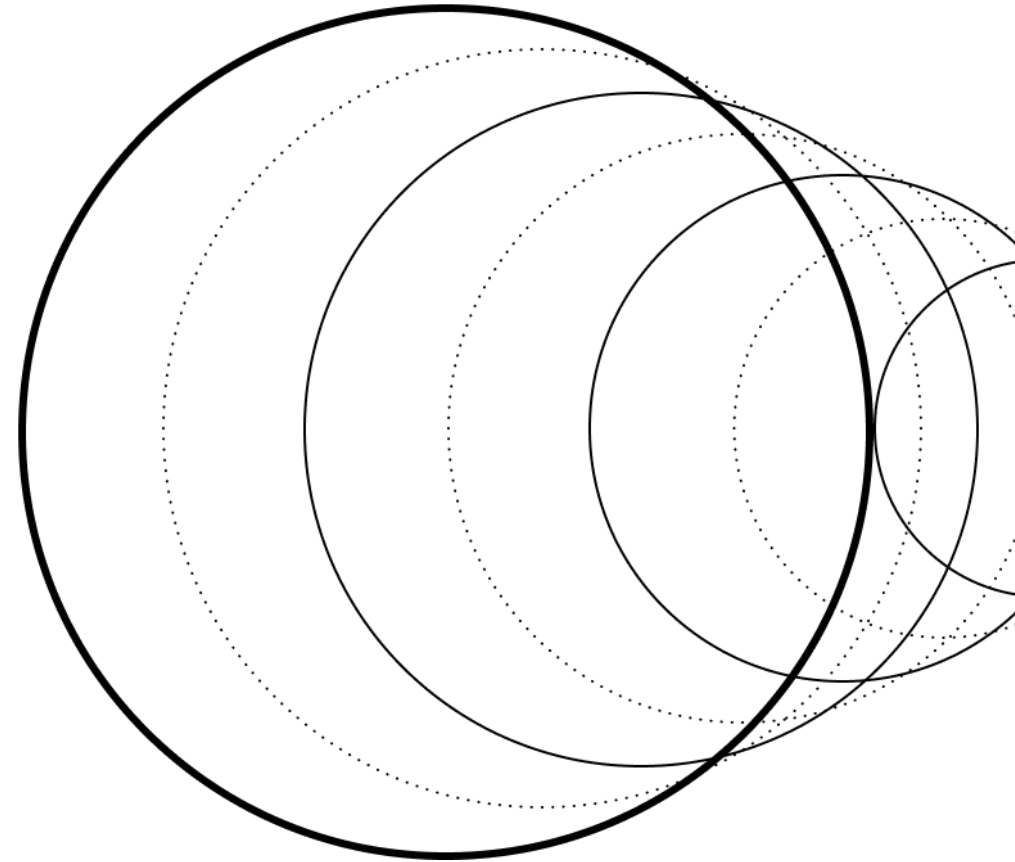
See also

- 01 PFN-based TFM: [\[TabPFNv2\]](#), [\[TabPFN-2.5\]](#), [\[TabICLv2\]](#), [\[LimiX\]](#), [\[Mitra\]](#)
- 02 Finetuning on collection of real-world datasets helps PFNs: [\[Real-TabPFN\]](#)
- 03 Making PFNs more robust to distribution shifts: [\[Drift-Resilient\]](#)

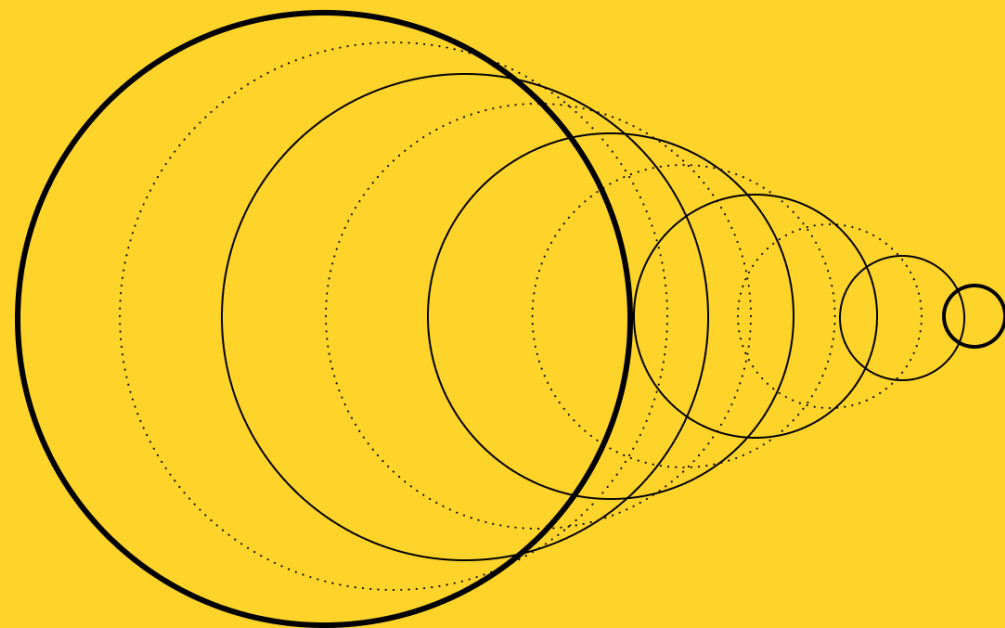


Caveats

- 01 Cross-table pretraining without in-context learning: [\[XTab\]](#)
- 02 TFM pretrained solely on real-world data: [\[TabDPT\]](#)
- 03 Pretraining TFM on a single dataset: [\[SingleDataset\]](#)



Yandex Research



TFMs for graphs

TFMs for node-level tasks

01

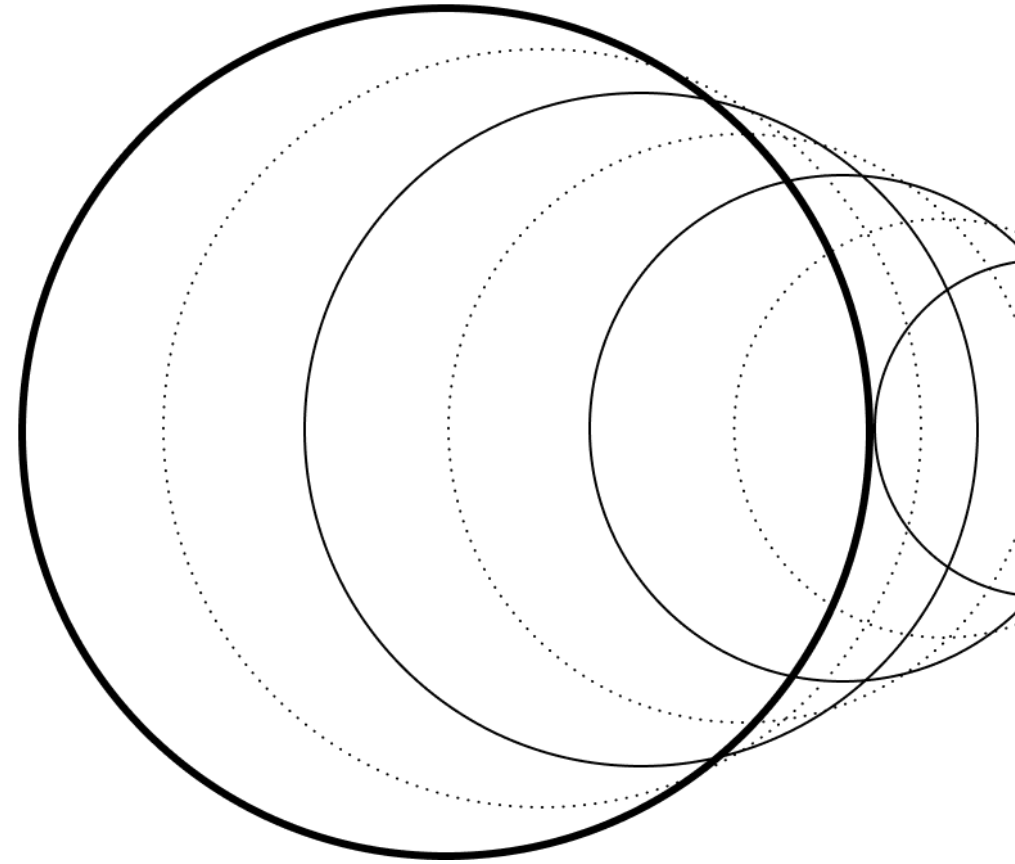
[G2T-FM]

02

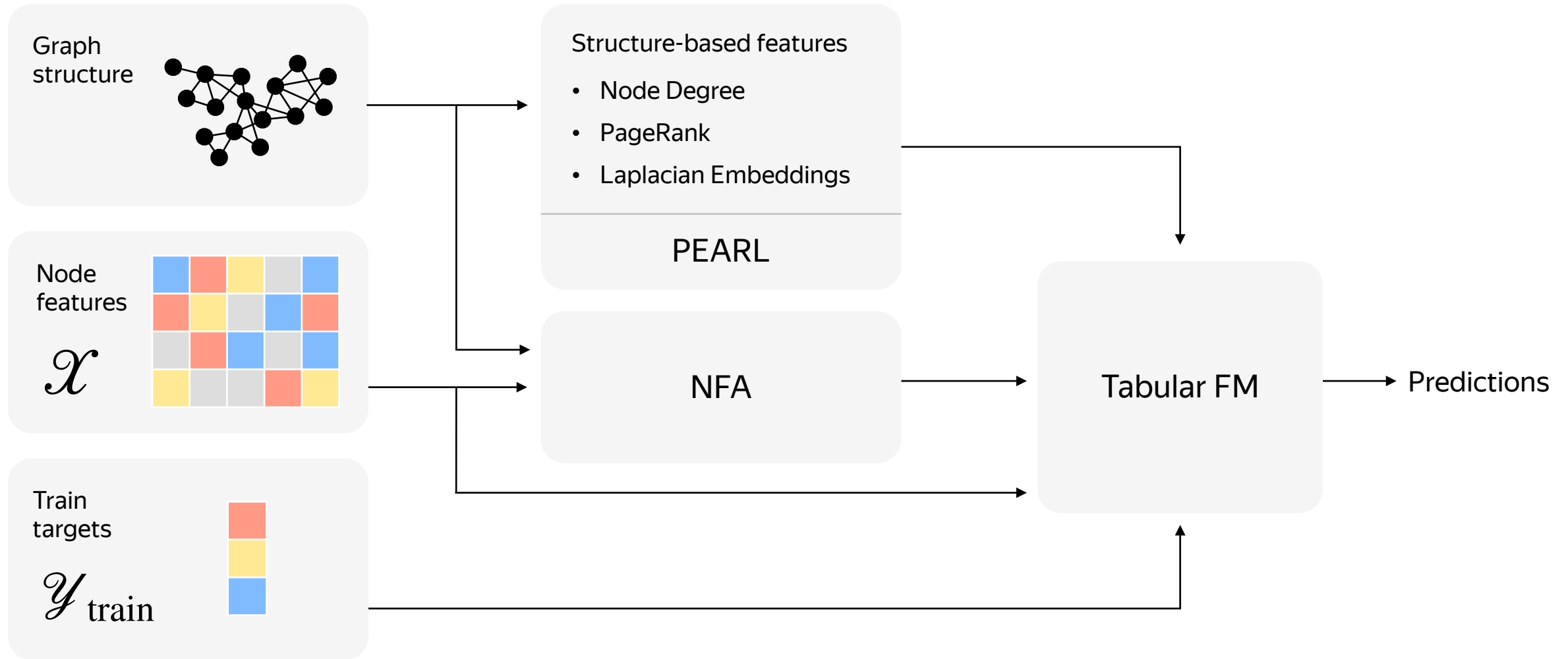
[TAG]

03

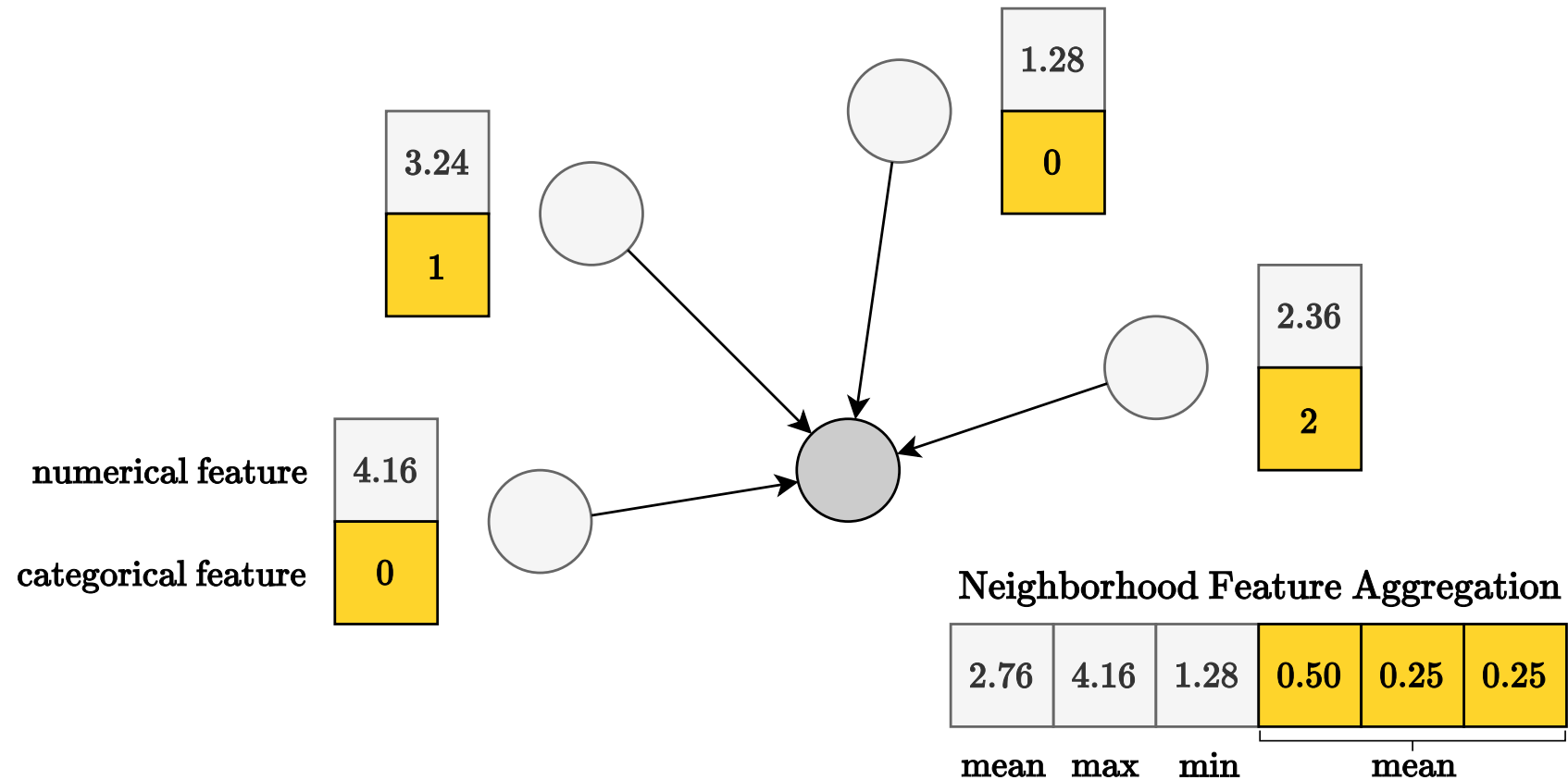
[TabPFN-GN]



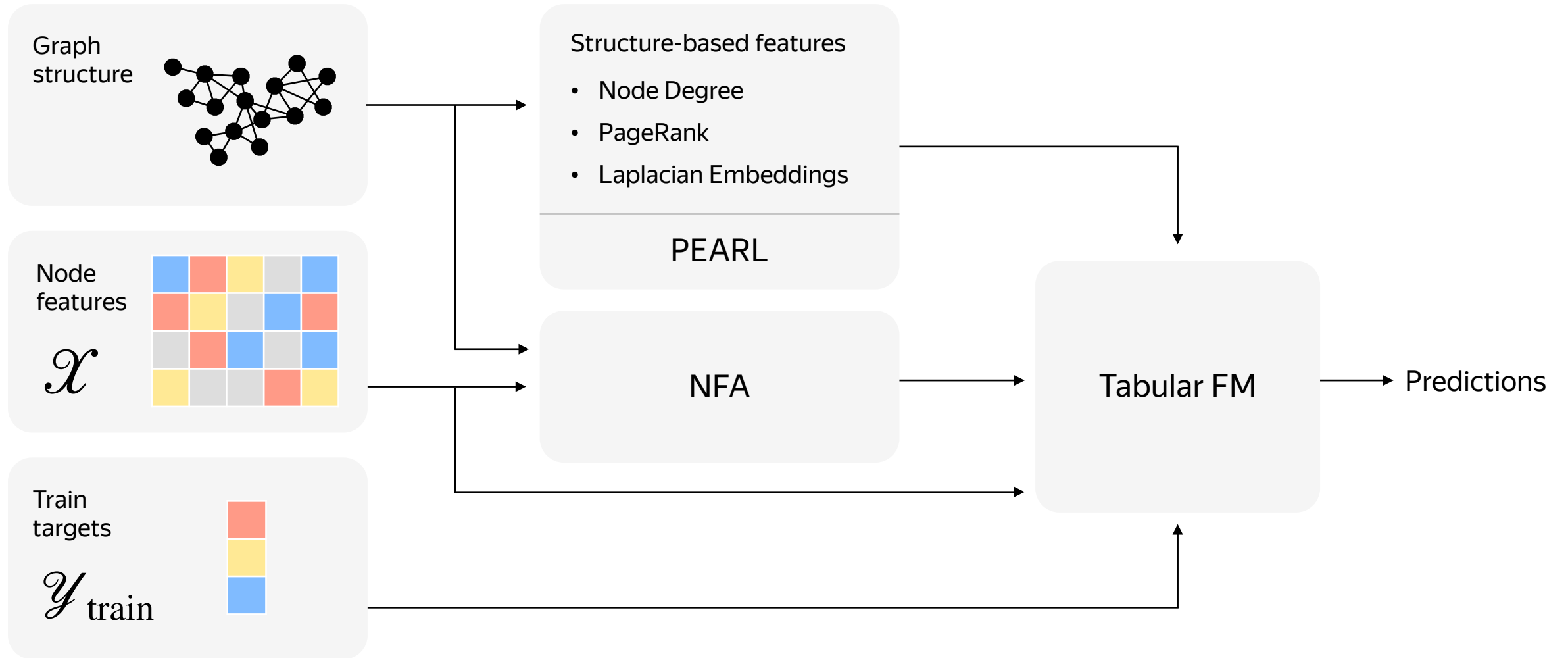
G2T-FM (Graph-to-Table Foundation Model)



Neighborhood feature aggregation



G2T-FM (Graph-to-Table Foundation Model)

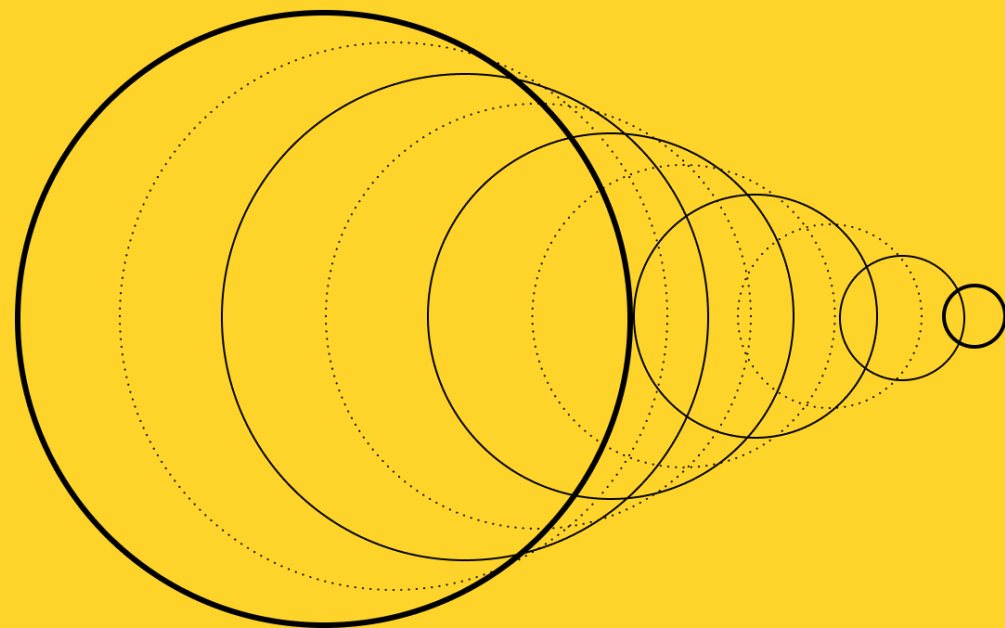


Evaluation results

Table 2: Evaluation results on datasets with tabular features (datasets from the GraphLand benchmark under the RL (random low) data split). For each column, we highlight **first**, **second**, and **third** best results with a color.

	tolokers-2	city-reviews	artnet-exp	hm-prices	avazu-ctr	city-roads-M	twitch-views	artnet-views	AR
LightGBM+NFA	56.34 $\pm$ 0.06	78.53 $\pm$ 0.01	46.13 $\pm$ 0.03	70.84 $\pm$ 0.04	31.71 $\pm$ 0.01	61.18 $\pm$ 0.03	60.14 $\pm$ 0.01	56.10 $\pm$ 0.02	5.38
GCN	56.27 $\pm$ 0.29	77.81 $\pm$ 0.14	44.86 $\pm$ 0.34	68.02 $\pm$ 0.40	32.00 $\pm$ 0.15	58.82 $\pm$ 0.24	75.51 $\pm$ 0.05	56.03 $\pm$ 0.24	6.25
GraphSAGE	54.43 $\pm$ 0.32	78.17 $\pm$ 0.09	45.14 $\pm$ 0.34	70.00 $\pm$ 0.70	31.44 $\pm$ 0.15	59.44 $\pm$ 0.26	66.29 $\pm$ 0.31	49.32 $\pm$ 0.86	7.12
GAT	57.41 $\pm$ 0.80	77.74 $\pm$ 0.20	45.06 $\pm$ 0.49	72.07 $\pm$ 1.16	32.63 $\pm$ 0.16	59.86 $\pm$ 0.19	72.89 $\pm$ 0.25	53.60 $\pm$ 0.23	5.12
GT	56.98 $\pm$ 0.53	77.34 $\pm$ 0.20	46.41 $\pm$ 0.68	69.44 $\pm$ 0.89	31.11 $\pm$ 0.47	59.55 $\pm$ 0.27	72.13 $\pm$ 0.13	53.37 $\pm$ 0.43	6.62
OpenGraph (ICL)	40.38 $\pm$ 1.13	59.09 $\pm$ 0.72	15.16 $\pm$ 0.83	N/A	N/A	N/A	N/A	N/A	11.00
AnyGraph (ICL)	28.75 $\pm$ 3.56	63.71 $\pm$ 1.45	12.84 $\pm$ 0.93	N/A	N/A	N/A	N/A	N/A	12.33
TS-GNN (ICL)	38.54 $\pm$ 0.94	43.46 $\pm$ 5.17	20.44 $\pm$ 1.05	N/A	N/A	N/A	N/A	N/A	11.33
GCOPE (FT)	28.81 $\pm$ 1.28	67.16 $\pm$ 0.98	14.92 $\pm$ 1.56	N/A	N/A	N/A	N/A	N/A	11.33
G2T-TabPFNv2 (ICL)	60.42 $\pm$ 0.27	77.46 $\pm$ 0.10	45.84 $\pm$ 0.03	66.68 $\pm$ 0.09	26.38 $\pm$ 0.07	60.47 $\pm$ 0.04	70.00 $\pm$ 0.06	58.75 $\pm$ 0.15	6.38
G2T-TabPFNv2 (FT)	57.65 $\pm$ 1.92	79.12 $\pm$ 0.21	47.31 $\pm$ 0.59	71.05 $\pm$ 0.91	28.52 $\pm$ 0.43	63.08 $\pm$ 0.28	74.06 $\pm$ 0.16	60.29 $\pm$ 0.13	3.62
G2T-LimiX (ICL)	61.48 $\pm$ 0.30	77.72 $\pm$ 0.54	48.43 $\pm$ 0.18	74.96 $\pm$ 0.06	32.39 $\pm$ 0.14	64.53 $\pm$ 0.07	71.08 $\pm$ 0.07	60.95 $\pm$ 0.10	3.12
G2T-LimiX (FT)	61.17 $\pm$ 0.49	80.13 $\pm$ 0.05	49.88 $\pm$ 0.13	76.32 $\pm$ 0.17	33.94 $\pm$ 0.34	65.87 $\pm$ 0.10	73.16 $\pm$ 0.40	62.12 $\pm$ 0.10	1.38

Yandex Research



PFNs for graphs

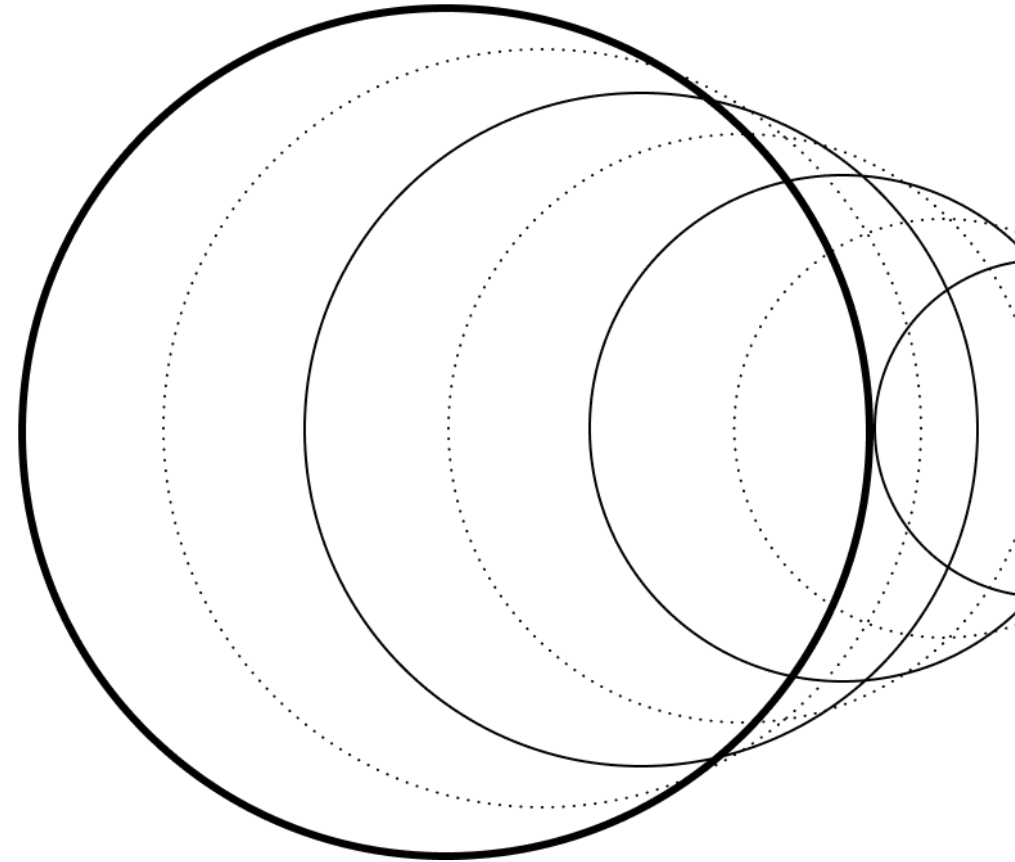
Can we pretrain PFNs specifically for graphs?

01

[GraphPFN]

02

[NodePFN]



GraphPFN

01. Architecture

02. Graph prior

03. Evaluation results

GraphPFN

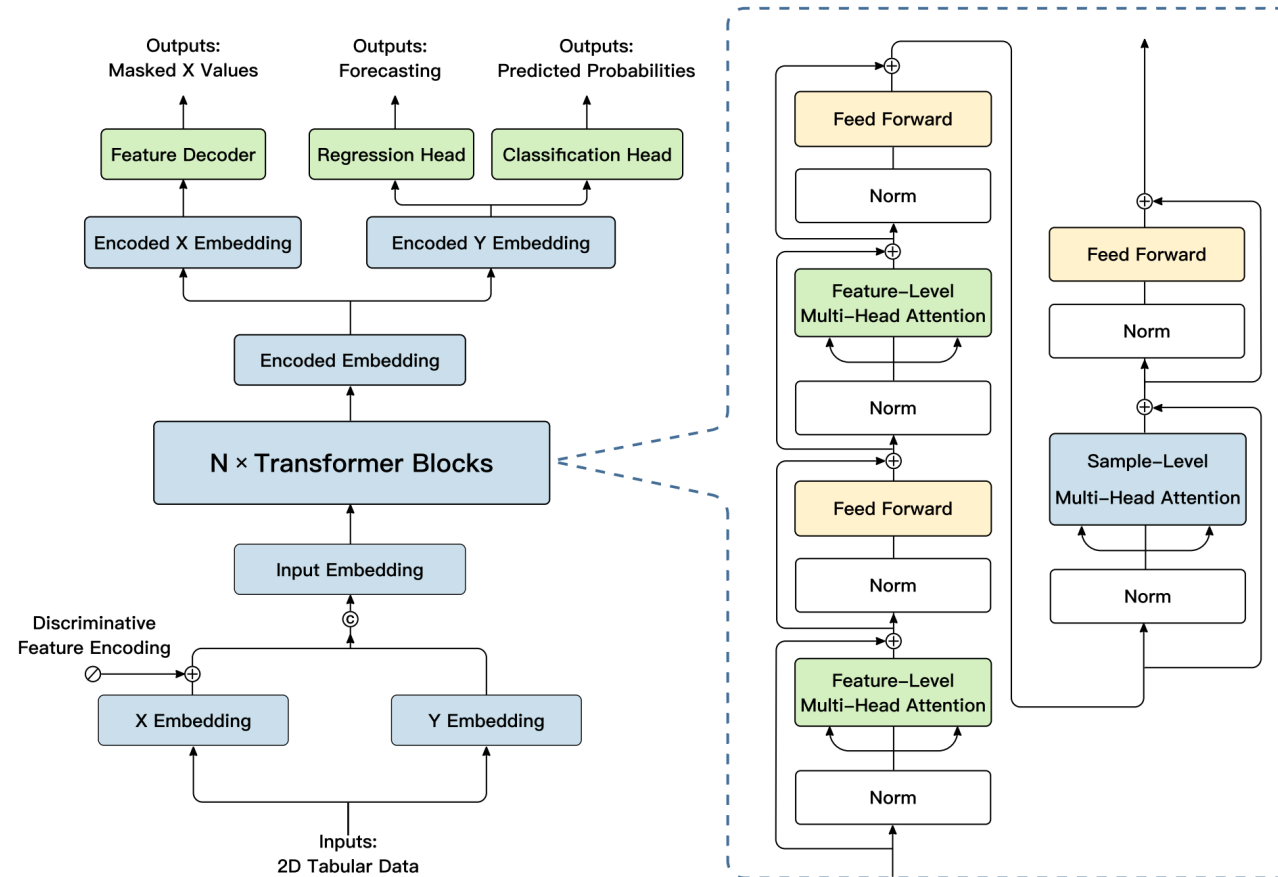
01. Architecture

02. Graph prior

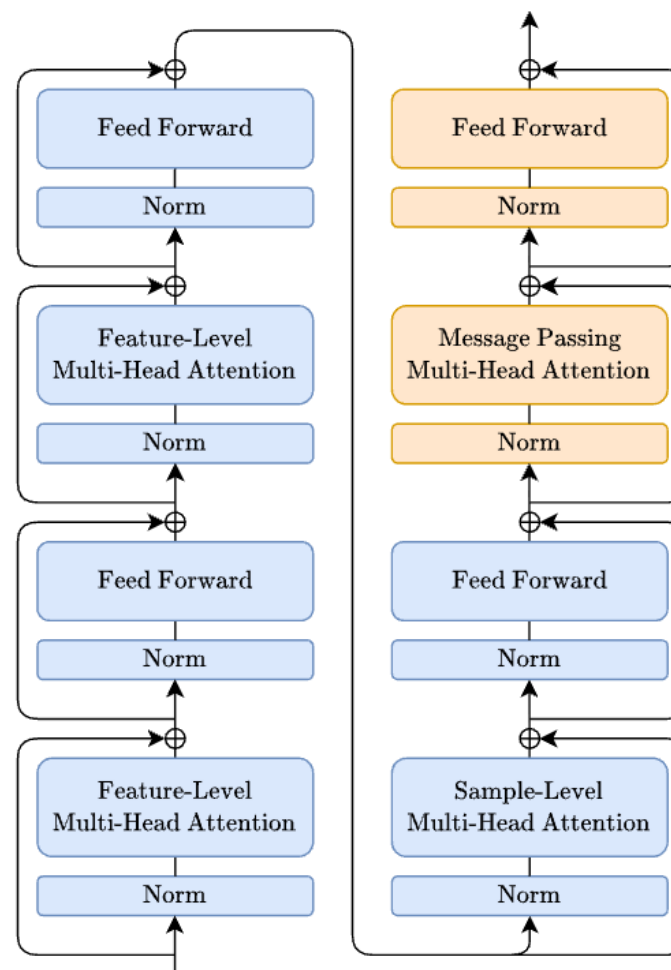
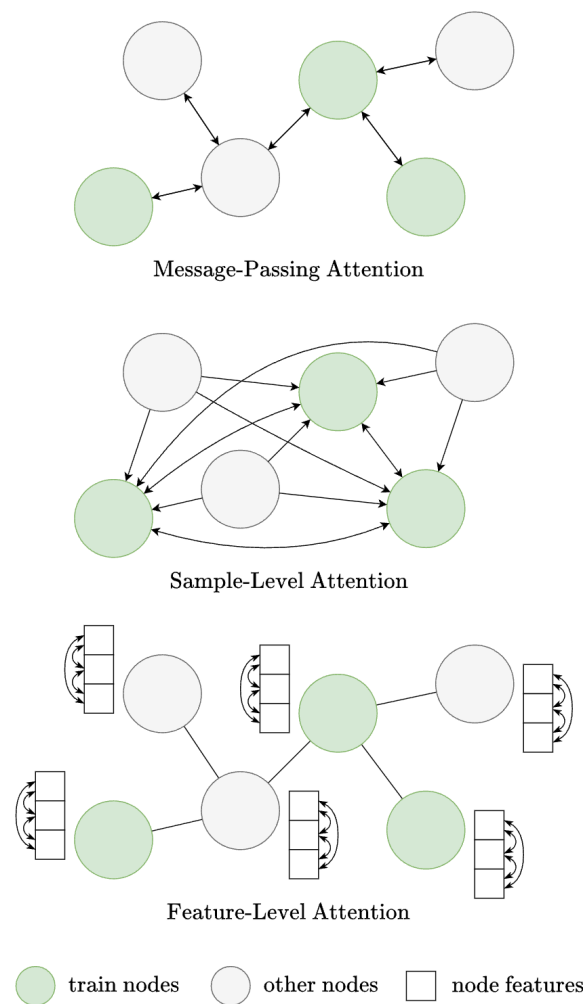
03. Evaluation results

Architecture: LimiX

TFMs can handle heterogeneous features and ICL, so let's take TFM as a backbone



Architecture: GraphPFN



GraphPFN

01. Architecture

02. Graph prior

03. Evaluation results

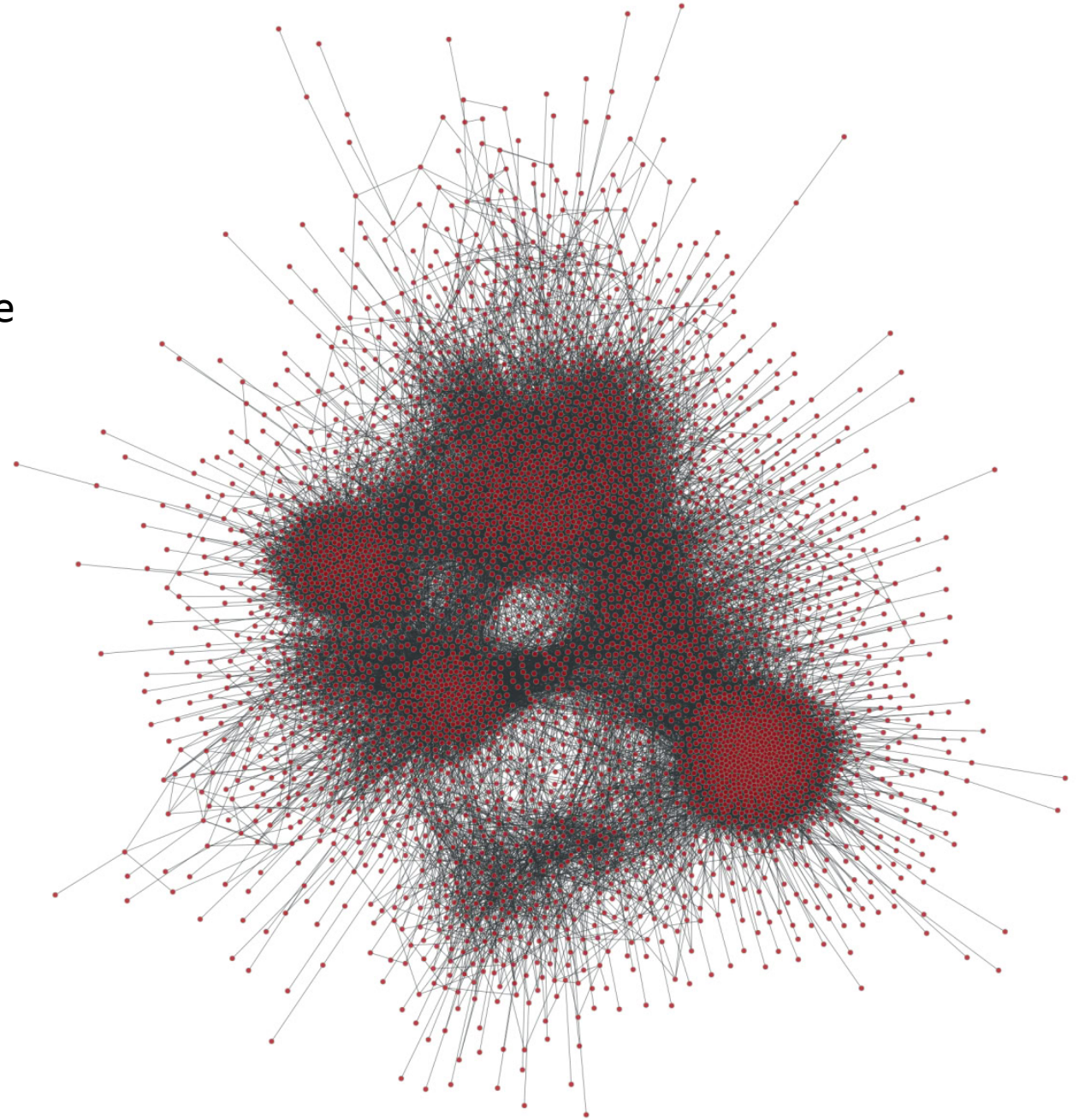
Prior: graphs

01

Multi-level degree-corrected SBM for community structure and desired degree distribution

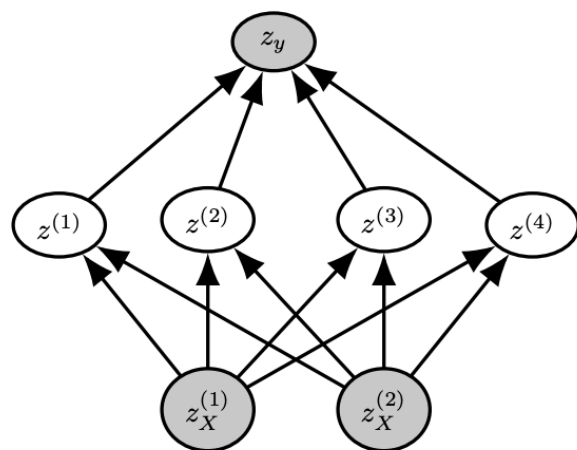
02

Preferential attachment for core-periphery structure

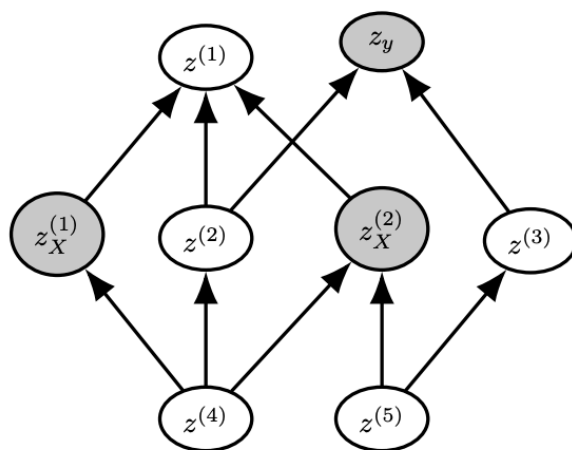


Prior: attributes

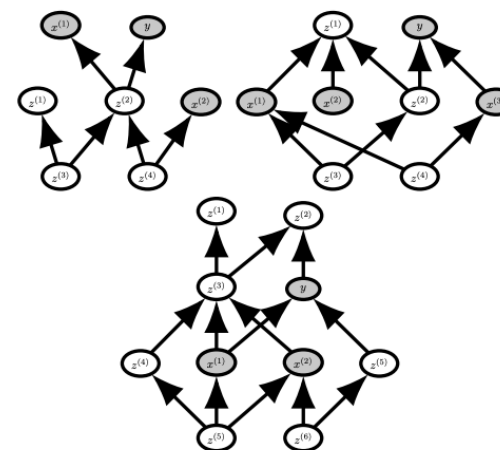
Take tabular SCMs and randomly replace some neurons with GNN neurons



(a) A BNN



(b) An SCM



(c) SCMs sampled from the prior

Figure 2: Overview of graphs generating data in our prior. Inputs x are mapped to the output y through unobserved nodes z . Plots based on Müller et al. (2022).

GraphPFN

01. Architecture

02. Graph prior

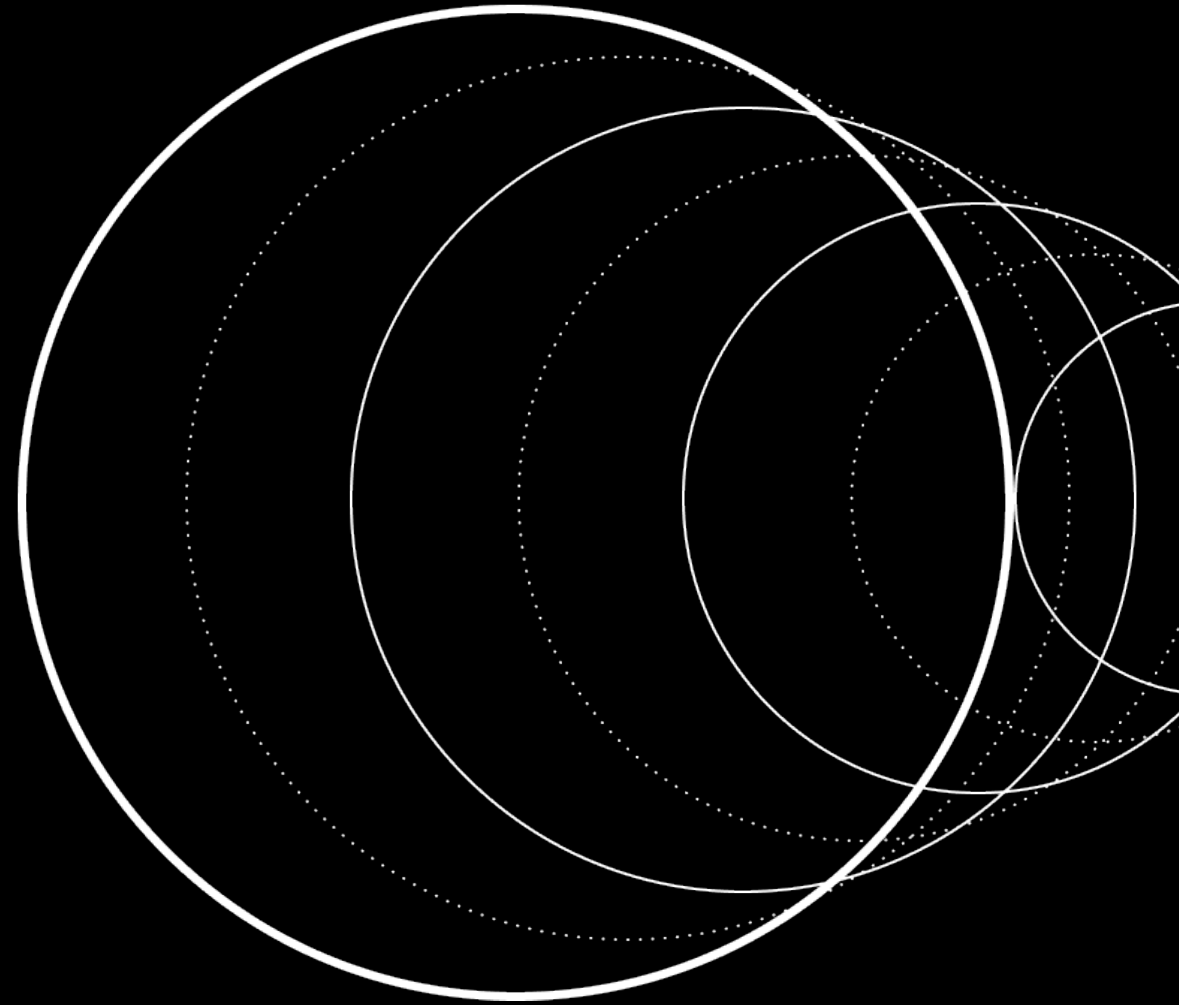
03. Evaluation results

Evaluation

	artnet-exp	city-reviews	tolokers-2	artnet-views	avazu-ctr	city-roads-M	hm-prices	twitch-views	AR (cls)	AR (all)
LightGBM + NFA	46.13 $\pm$ 0.04	78.53 $\pm$ 0.01	56.34 $\pm$ 0.06	56.10 $\pm$ 0.02	31.71 $\pm$ 0.01	61.18 $\pm$ 0.03	70.84 $\pm$ 0.04	60.14 $\pm$ 0.01	11.0	9.75
GCN	44.86 $\pm$ 0.36	77.81 $\pm$ 0.15	56.27 $\pm$ 0.31	56.03 $\pm$ 0.25	32.00 $\pm$ 0.16	58.82 $\pm$ 0.25	68.02 $\pm$ 0.42	75.51 $\pm$ 0.05	13.33	10.25
GraphSAGE	45.14 $\pm$ 0.36	78.17 $\pm$ 0.10	54.43 $\pm$ 0.34	49.32 $\pm$ 0.91	31.44 $\pm$ 0.16	59.44 $\pm$ 0.27	70.00 $\pm$ 0.74	66.29 $\pm$ 0.32	12.33	11.25
GAT	45.06 $\pm$ 0.52	77.74 $\pm$ 0.21	57.41 $\pm$ 0.85	53.60 $\pm$ 0.24	32.63 $\pm$ 0.17	59.86 $\pm$ 0.20	72.07 $\pm$ 1.22	72.89 $\pm$ 0.27	12.67	9.5
GT	46.41 $\pm$ 0.71	77.34 $\pm$ 0.21	56.98 $\pm$ 0.55	53.37 $\pm$ 0.46	31.11 $\pm$ 0.49	59.55 $\pm$ 0.28	69.44 $\pm$ 0.94	72.13 $\pm$ 0.13	12.67	10.88
OpenGraph (ICL)	15.16 $\pm$ 0.83	59.09 $\pm$ 0.72	40.38 $\pm$ 1.13	–	–	–	–	–	17.0	–
AnyGraph (ICL)	12.84 $\pm$ 0.93	63.71 $\pm$ 1.45	28.75 $\pm$ 3.56	–	–	–	–	–	18.33	–
TS-GNN (ICL)	20.44 $\pm$ 1.05	43.46 $\pm$ 5.17	38.54 $\pm$ 0.94	–	–	–	–	–	17.33	–
G2T-LimiX (ICL)	48.44 $\pm$ 0.23	77.81 $\pm$ 0.49	61.45 $\pm$ 0.33	60.94 $\pm$ 0.11	32.39 $\pm$ 0.13	64.53 $\pm$ 0.09	74.96 $\pm$ 0.07	71.08 $\pm$ 0.07	7.33	7.5
GraphPFN (ICL)	50.85 $\pm$ 0.31	79.89 $\pm$ 0.10	60.33 $\pm$ 0.60	61.77 $\pm$ 0.23	30.41 $\pm$ 0.18	63.55 $\pm$ 0.29	75.38 $\pm$ 0.26	73.27 $\pm$ 0.13	5.67	7.12
G2T-LimiX (ICL, E)	48.52 $\pm$ 0.29	77.97 $\pm$ 0.53	61.58 $\pm$ 0.32	60.96 $\pm$ 0.10	32.41 $\pm$ 0.14	64.69 $\pm$ 0.05	74.97 $\pm$ 0.06	71.30 $\pm$ 0.07	6.33	6.5
TAG-TabPFNV2 (ICL, E)	47.87 $\pm$ 0.39	77.38 $\pm$ 0.29	59.33 $\pm$ 1.00	–	–	–	–	–	11.33	–
TAG-LimiX (ICL, E)	50.19 $\pm$ 0.46	78.87 $\pm$ 0.13	59.46 $\pm$ 1.05	–	–	–	–	–	7.0	–
GraphPFN (ICL, E)	51.49 $\pm$ 0.10	80.22 $\pm$ 0.04	60.98 $\pm$ 0.25	62.07 $\pm$ 0.04	30.46 $\pm$ 0.14	64.29 $\pm$ 0.12	75.82 $\pm$ 0.06	73.81 $\pm$ 0.04	3.0	5.25
GCOPE (FT)	14.92 $\pm$ 1.56	67.16 $\pm$ 0.98	28.81 $\pm$ 1.28	–	–	–	–	–	17.33	–
G2T-LimiX (FT)	49.89 $\pm$ 0.18	80.14 $\pm$ 0.06	61.28 $\pm$ 0.70	62.08 $\pm$ 0.12	34.02 $\pm$ 0.27	65.82 $\pm$ 0.18	76.27 $\pm$ 0.24	73.69 $\pm$ 0.35	5.0	4.25
GraphPFN (FT)	50.64 $\pm$ 1.27	80.49 $\pm$ 0.16	60.90 $\pm$ 0.98	63.90 $\pm$ 0.12	34.34 $\pm$ 0.50	66.04 $\pm$ 0.41	79.26 $\pm$ 0.43	78.17 $\pm$ 0.13	4.0	2.75
G2T-LimiX (FT, E)	49.96 $\pm$ 0.09	80.15 $\pm$ 0.07	60.71 $\pm$ 0.57	62.11 $\pm$ 0.06	33.86 $\pm$ 0.33	65.96 $\pm$ 0.11	76.30 $\pm$ 0.19	73.62 $\pm$ 0.49	5.67	4.38
GraphPFN (FT, E)	51.97 $\pm$ 0.40	80.89 $\pm$ 0.05	60.92 $\pm$ 0.56	65.12 $\pm$ 0.11	35.05 $\pm$ 0.23	67.27 $\pm$ 0.14	80.13 $\pm$ 0.37	78.50 $\pm$ 0.36	2.33	1.5

Summary

- GFMs for node-level tasks are feasible on small datasets (~10k train nodes, ~100k total nodes)
- Scalability remains an issue
- PFNs offer a promising path toward generalizable GFMs



References

[[GraphLand](#)] Bazhenov et al. “GraphLand: Evaluating Graph Machine Learning Models on Diverse Industrial Data,” NeurIPS, 2025.

[[GraphFM](#)] Lachi et al. “GraphFM: A Scalable Framework for Multi-graph Pretraining,” arXiv:2407.11907 2024.

[[OpenGraph](#)] Xia et al. “OpenGraph: Towards Open Graph Foundation Models,” Findings of EMNLP 2024.

[[TS-GNN](#)] Finkelshtein et al. “Equivariance Everywhere All At Once: A Recipe for Graph Foundation Models,” arXiv:2506.14291, 2025.

[[GCOPE](#)] Zhao et al. “All in One and One for All: A Simple yet Effective Method Towards Cross-domain Graph Pretraining,” KDD 2024.

[[MDGFM](#)] Wang et al. “Multi-Domain Graph Foundation Models: Robust Knowledge Transfer via Topology Alignment,” ICML 2025.

[[AnyGraph](#)] Xia et al. “AnyGraph: Graph Foundation Model in the Wild,” arXiv:2408.10700, 2024.

[[GFT](#)] Wang et al. “GFT: Graph Foundation Model with Transferable Tree Vocabulary,” NeurIPS 2024.

[[OFA](#)] Liu et al. “One for All: Towards Training One Graph Model for All Classification Tasks,” ICLR 2024.

References

- [[TabPFNV2](#)] Hollmann et al. "Accurate predictions on small data with a tabular foundation model", Nature, 2025.
- [[TabPFN-2.5](#)] Grinsztajn et al. "TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models", arXiv:2511.08667, 2025.
- [[TabICLv2](#)] Qu et al. "TabICLv2: A better, faster, scalable, and open tabular foundation model", 2026.
- [[LimiX](#)] Zhang et al. "LimiX: Unleashing Structured-Data Modeling Capability for Generalist Intelligence", arXiv:2509.03505, 2025.
- [[Mitra](#)] Zhang et al. "Mitra: Mixed Synthetic Priors for Enhancing Tabular Foundation Models", NeurIPS, 2025.
- [[Real-TabPFN](#)] Garg et al. "Real-TabPFN: Improving Tabular Foundation Models via Continued Pre-training With Real-World Data", arXiv:2507.03971, 2025.
- [[DriftResilient](#)] Helli et al. "Drift-Resilient TabPFN: In-Context Learning Temporal Distribution Shifts on Tabular Data", NeurIPS, 2025.
- [[XTab](#)] Zhu et al. "XTab: Cross-table Pretraining for Tabular Transformers", ICML, 2023.
- [[TabDPT](#)] Ma et al. "TabDPT: Scaling Tabular Foundation Models on Real Data", arXiv:2410.18164, 2024.
- [[SingleDataset](#)] Ma et al. "Generalization Can Emerge in Tabular Foundation Models From a Single Table", arXiv:2511.09665, 2025.

References

[G2T-FM] Ereemeev et al. “Turning Tabular Foundation Models into Graph Foundation Models,” arXiv:2508.20906, 2025.

[TAG] Hayler et al. “Bringing Graphs to the Table: Zero-shot Node Classification via Tabular Foundation Models”, arXiv:2509.07143, 2025.

[TabPFN-GN] Choi et al. “Can TabPFN Compete with GNNs for Node Classification via Graph Tabularization?”, arXiv:2512.08798, 2025.

[GraphPFN] Ereemeev et al. “GraphPFN: A Prior-Data Fitted Graph Foundation Model”, arXiv:2509.21489, 2025.

[NodePFN] Choi et al. "Learning Posterior Predictive Distributions for Node Classification from Synthetic Graph Priors", ICLR, 2026.