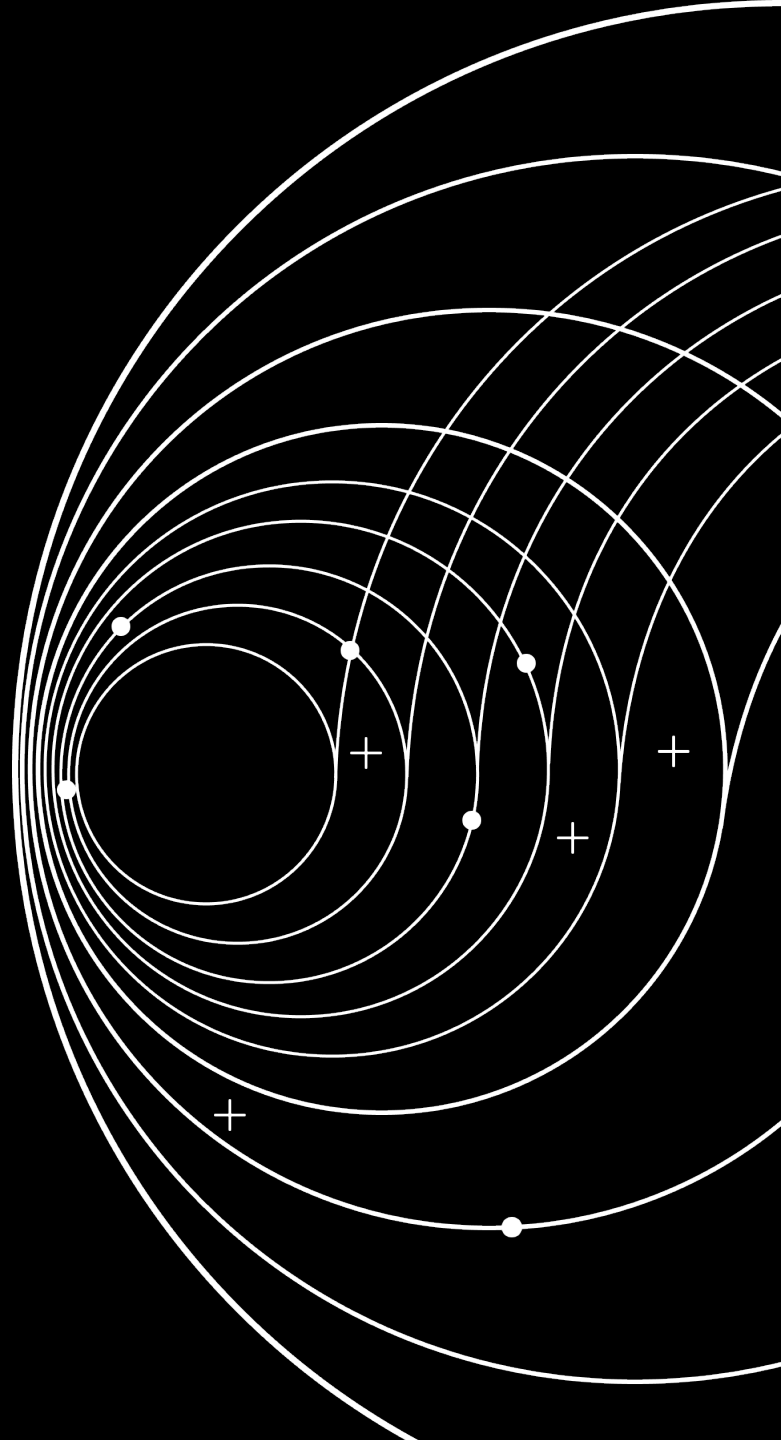
Yandex Research

GraphPFN: A Prior-Data Fitted Graph Foundation Model

Dmitry Eremeev, Oleg Platonov, Gleb Bazhenov,
Artem Babenko, Liudmila Prokhorenkova

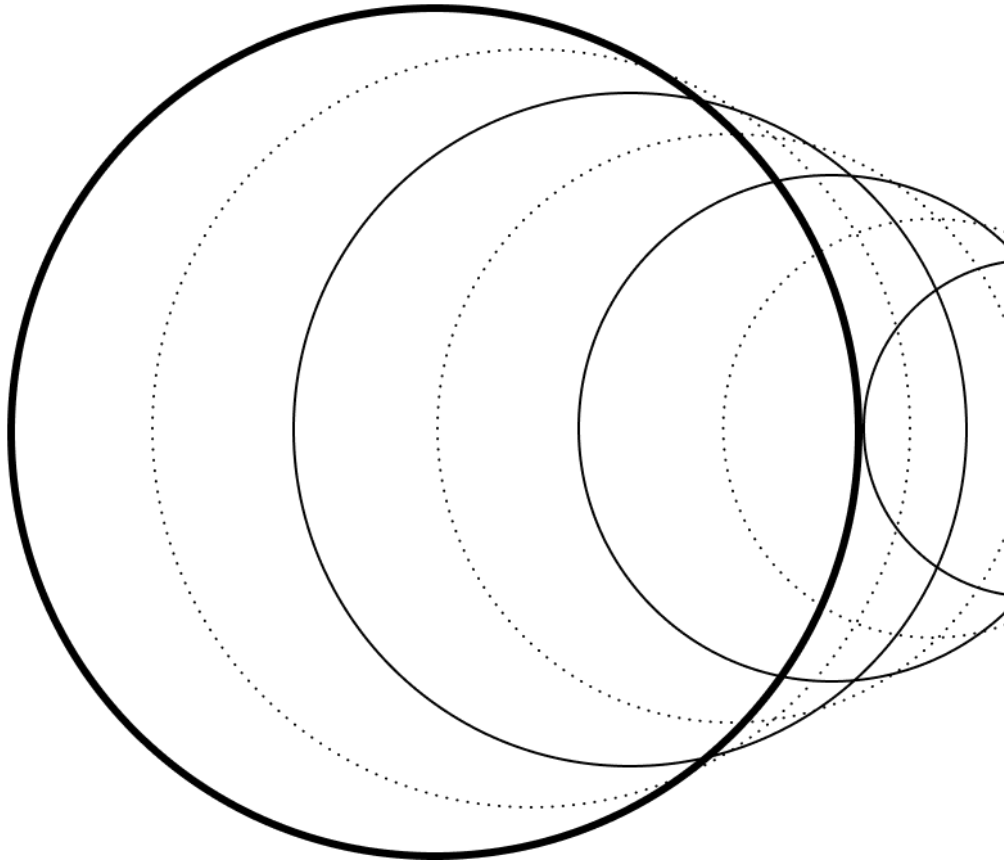


GraphPFN: Overview

01 Based on the PFNs framework popularized for tabular data by TabPFN and extends it to graph-structured data via message-passing adapters

02 Pretrains on 1’600’000 synthetic datasets with more than 1 billion nodes in total

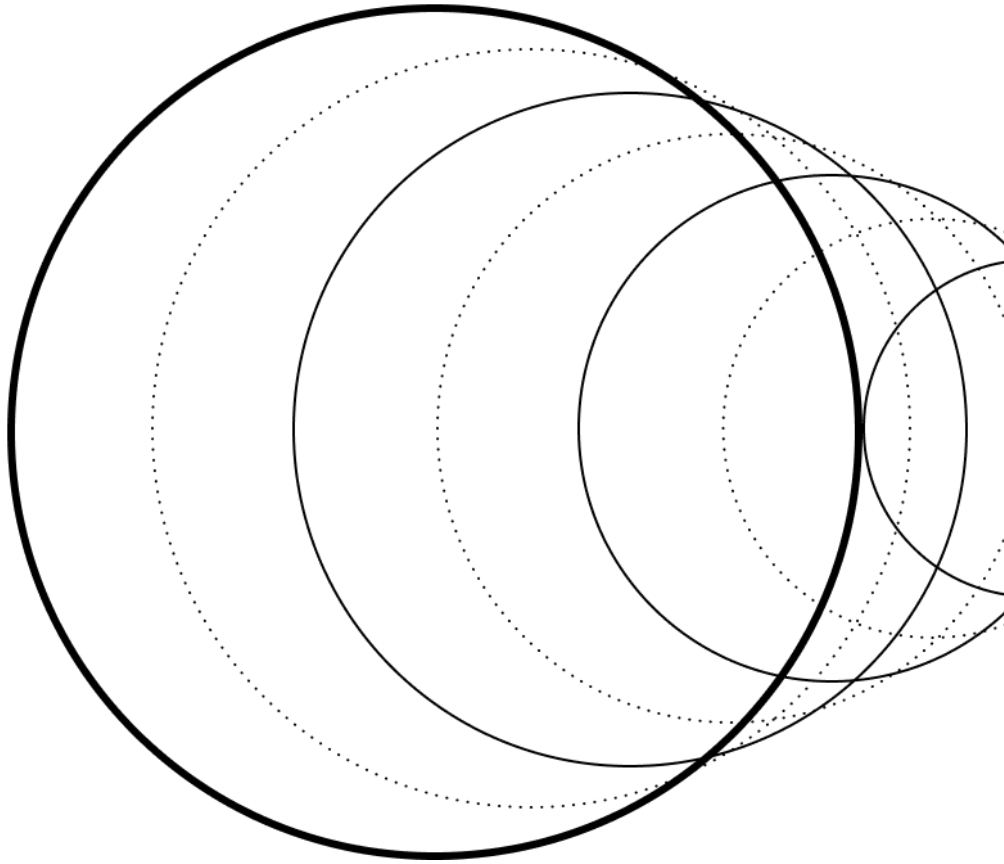
03 Achieves state-of-the art results in both in-context learning and finetuning regimes



Key Challenges for GFMs

01 **Transferability:** Graphs vary widely in features, labels, and semantics, making learned representations hard to transfer across datasets.

02 **Data scarcity:** Graph datasets lag behind CV and NLP in scale and diversity; massive single-domain graphs are feasible, but do not help with domain coverage.

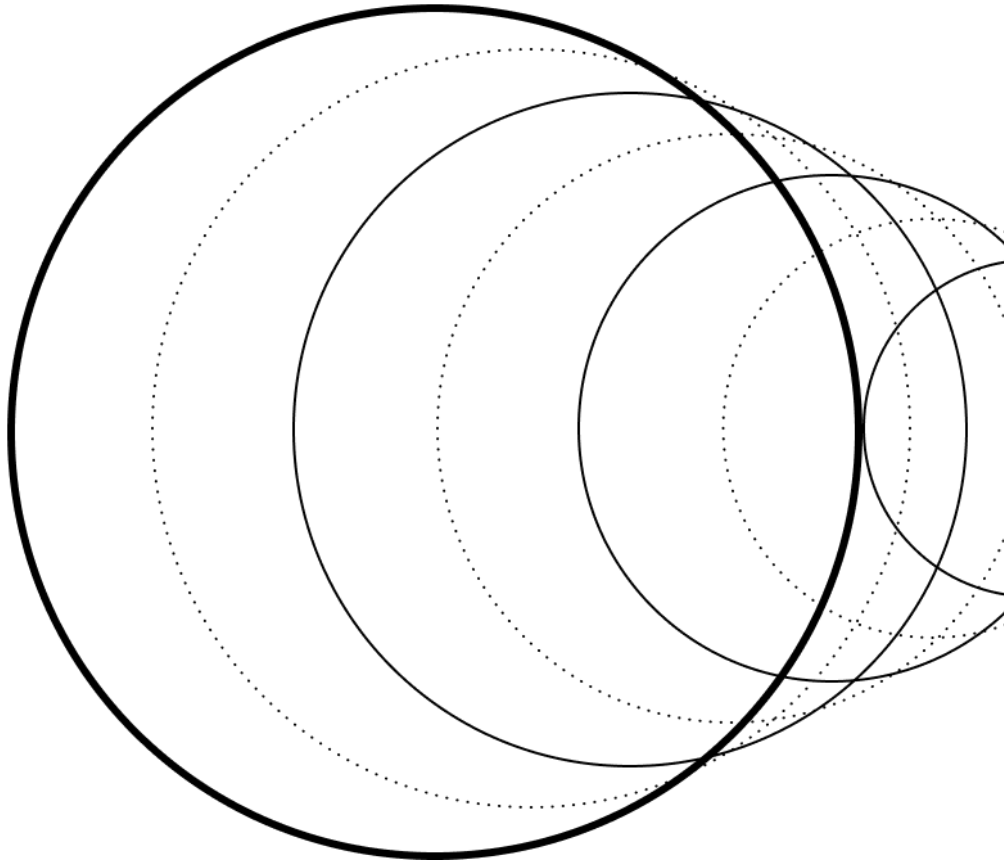


Prior-Data Fitted Networks

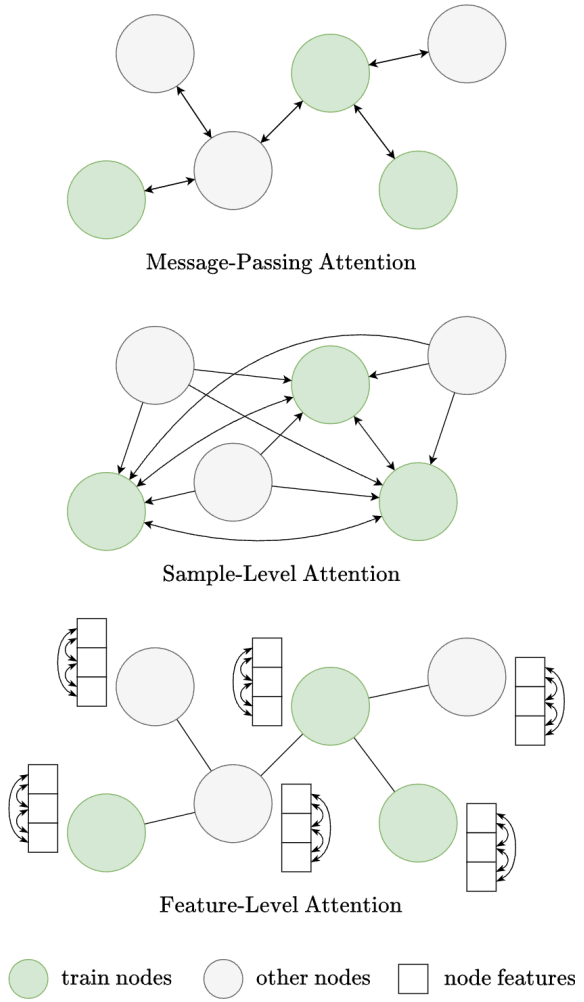
TFMs face similar challenges, yet successful PFN-based TFMs emerged in last years

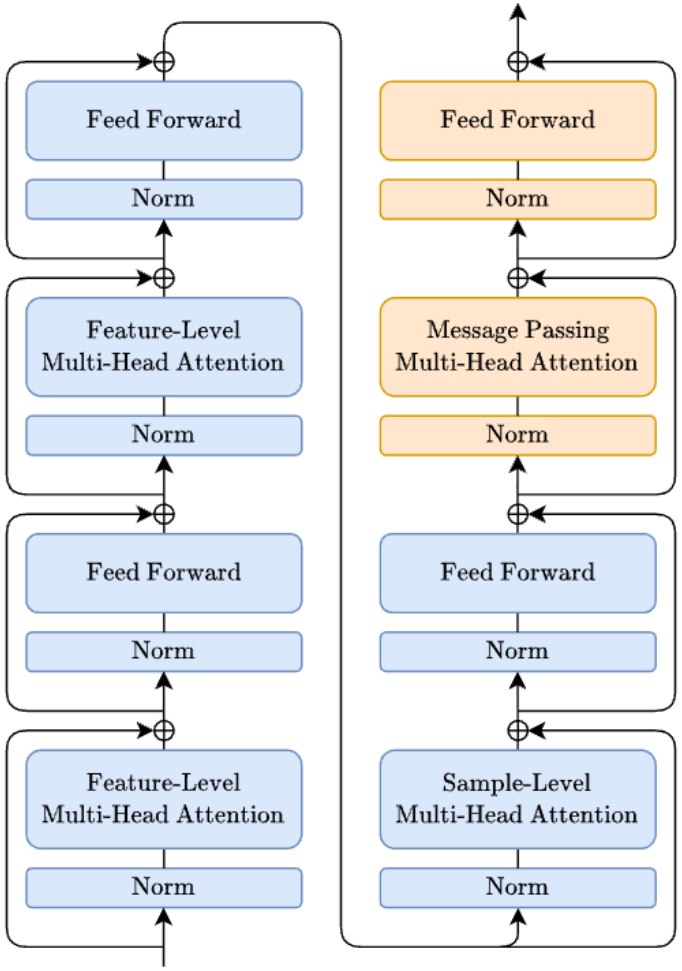
01 Predictions in an in-context learning regime

02 Pretrained on synthetic datasets from a *prior*



Architecture



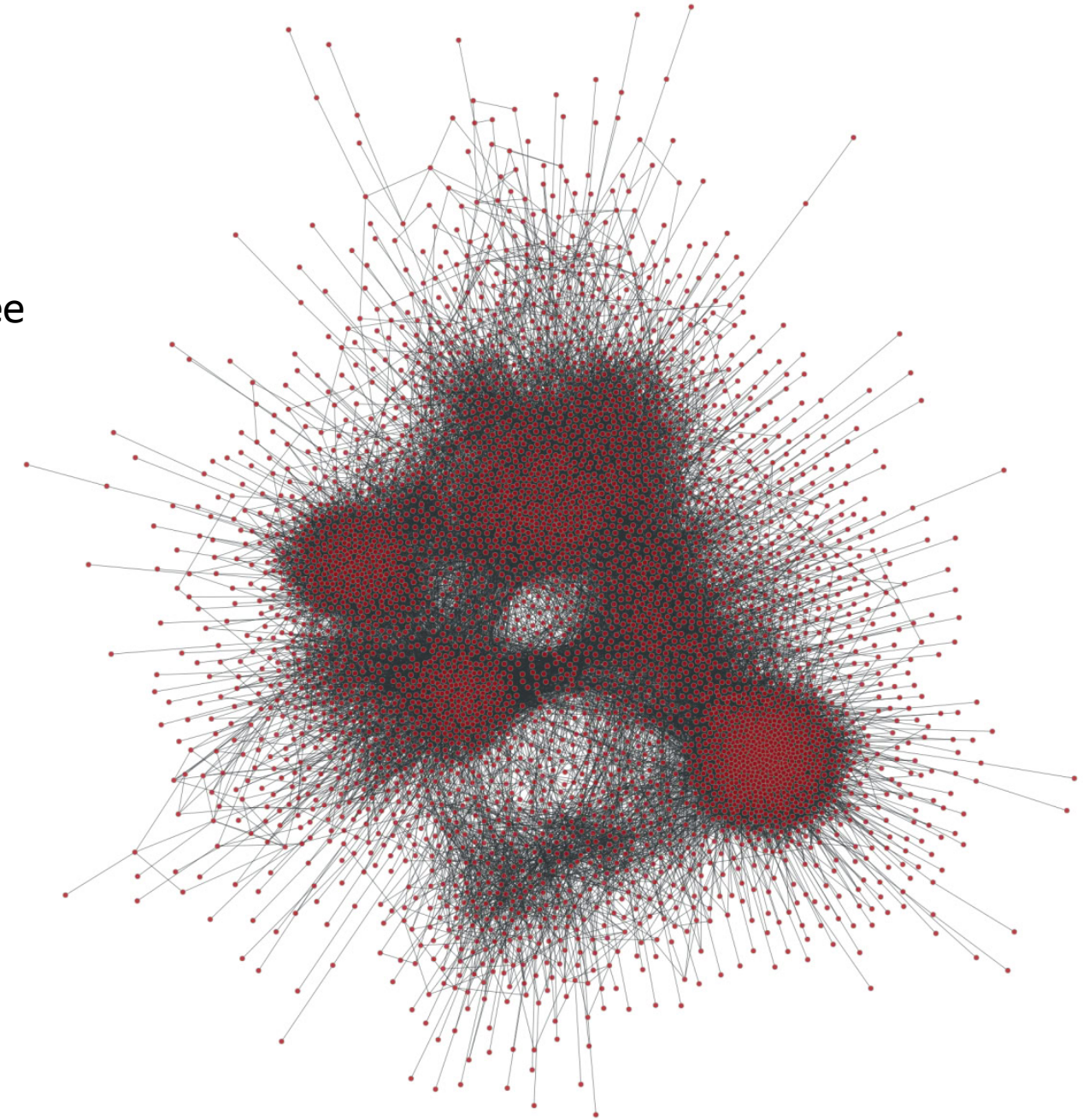


Graph Prior

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

02 Preferential attachment for core-periphery structure

03 Apply random GNN to generate node features and targets*



* – procedure is more complex, see the paper for details

Evaluation Results

	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 ± 0.04	78.53 ± 0.01	56.34 ± 0.06	56.10 ± 0.02	31.71 ± 0.01	61.18 ± 0.03	70.84 ± 0.04	60.14 ± 0.01	**7.33**	**6.00**
GCN	44.86 ± 0.36	77.81 ± 0.15	56.27 ± 0.31	56.03 ± 0.25	32.00 ± 0.16	58.82 ± 0.25	68.02 ± 0.42	75.51 ± 0.05	**9.67**	**6.88**
GraphSAGE	45.14 ± 0.36	78.17 ± 0.10	54.43 ± 0.34	49.32 ± 0.91	31.44 ± 0.16	59.44 ± 0.27	70.00 ± 0.74	66.29 ± 0.32	**8.67**	**7.62**
GAT	45.06 ± 0.52	77.74 ± 0.21	57.41 ± 0.85	53.60 ± 0.24	32.63 ± 0.17	59.86 ± 0.20	72.07 ± 1.22	72.89 ± 0.27	**8.67**	**5.88**
GT	46.41 ± 0.71	77.34 ± 0.21	56.98 ± 0.55	53.37 ± 0.46	31.11 ± 0.49	59.55 ± 0.28	69.44 ± 0.94	72.13 ± 0.13	**8.67**	**7.25**
OpenGraph (ICL)	15.16 ± 0.83	59.09 ± 0.72	40.38 ± 1.13	–	–	–	–	–	**13.67**	–
AnyGraph (ICL)	12.84 ± 0.93	63.71 ± 1.45	28.75 ± 3.56	–	–	–	–	–	**15.67**	–
TS-GNN (ICL)	20.44 ± 1.05	43.46 ± 5.17	38.54 ± 0.94	–	–	–	–	–	**13.67**	–
SAMGPT (FT)	16.20 ± 0.93	45.91 ± 2.68	36.93 ± 1.07	–	–	–	–	–	**14.33**	–
MDGFM (FT)	16.99 ± 1.75	31.24 ± 8.50	31.97 ± 1.34	–	–	–	–	–	**15.00**	–
GCOPE (FT)	14.92 ± 1.56	67.16 ± 0.98	28.81 ± 1.28	–	–	–	–	–	**14.67**	–
G2T-LimiX (ICL)	48.52 ± 0.29	77.97 ± 0.53	61.58 ± 0.32	60.96 ± 0.10	32.41 ± 0.14	64.69 ± 0.05	74.97 ± 0.06	71.30 ± 0.07	**4.67**	**4.38**
TAG-TabPFNv2 (ICL)	47.87 ± 0.39	77.38 ± 0.29	59.33 ± 1.00	–	–	–	–	–	**7.33**	–
TAG-LimiX (ICL)	50.19 ± 0.46	78.87 ± 0.13	59.46 ± 1.05	–	–	–	–	–	**4.00**	–
GraphPFN (ICL)	51.79 ± 0.11	80.25 ± 0.05	61.29 ± 0.12	62.79 ± 0.08	31.63 ± 0.06	64.85 ± 0.13	77.88 ± 0.09	73.20 ± 0.08	**2.33**	**3.12**
G2T-LimiX (FT)	49.96 ± 0.09	80.15 ± 0.07	60.71 ± 0.57	62.11 ± 0.06	33.86 ± 0.33	65.96 ± 0.11	76.30 ± 0.19	73.62 ± 0.49	**3.67**	**2.88**
GraphPFN (FT)	53.49 ± 0.81	80.90 ± 0.03	62.80 ± 0.39	65.35 ± 0.06	35.07 ± 0.34	67.30 ± 0.21	81.06 ± 0.24	79.00 ± 0.14	**1.00**	**1.00**

Yandex Research

Thank you!