NEURAL INFORMATION PROCESSING SYSTEMS

Yandex Research

GraphLand: Evaluating Graph Machine Learning Models on Diverse Industrial Data

Gleb Bazhenov* Oleg Platonov* Liudmila Prokhorenkova

Problem

- Popular graph ML benchmarks for node prediction cover a very narrow set of data domains (mostly citation graphs)
- They focus on text-attributed graphs, ignoring heterogeneous tabular features and cross-domain transfer
- And often miss realistic settings with temporal splits and inductive evaluation under distribution shifts
- Strong industrial baselines are underrepresented, and many tasks do not verify whether graph structure actually helps

Contribution

- GraphLand: 14 diverse industrial datasets with rich heterogeneous tabular node features and varied structural properties
- Different data splits which enable controlled random vs temporal and transductive vs inductive comparisons
- Extensive evaluation of GNNs, GFMs, and GBDTs, plus NFA as simple graph-based features for graph-agnostic models
- Key findings are: attention GNNs often lead, GBDTs with NFA are strong baselines, current GFMs fail to compete

Results

top performing models

(a) Results for classification datasets. Accuracy is reported for multiclass classification datasets and Average Precision is reported for binary classification datasets.

	multiclass classification			binary classification			
	hm-categories	pokec-regions	web-topics	tolokers-2	city-reviews	artnet-exp	web-fraud
best const. pred.	19.46 ± 0.00	3.77 ± 0.00	28.36 ± 0.00	21.82 ± 0.00	12.09 ± 0.00	10.00 ± 0.00	0.66 ± 0.00
ResMLP	37.72 ± 0.18	4.88 ± 0.01	42.41 ± 0.02	41.16 ± 1.13	71.32 ± 0.11	35.07 ± 2.34	8.77 ± 0.18
XGBoost	40.04 ± 0.09	4.93 ± 0.01	TLE	45.76 ± 1.00	74.70 ± 0.13	41.92 ± 0.82	11.54 ± 0.04
LightGBM	39.73 ± 0.08	4.89 ± 0.00	TLE	44.60 ± 0.12	74.51 ± 0.04	41.21 ± 0.12	TLE
CatBoost	40.72 ± 0.40	TLE	TLE	46.10 ± 0.35	74.77 ± 0.10	42.50 ± 0.12	TLE
ResMLP-NFA	48.72 ± 0.38	8.05 ± 0.03	MLE	48.14 ± 1.40	76.02 ± 0.14	38.25 ± 0.56	MLE
LightGBM-NFA	56.55 ± 0.15	9.53 ± 0.01	TLE	56.16 ± 0.28	78.33 ± 0.04	45.40 ± 0.13	TLE
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
GraphSAGE	56.75 ± 0.53	37.88 ± 0.41	47.41 ± 0.13	53.73 ± 0.53	77.82 ± 0.13	42.65 ± 0.59	12.11 ± 0.23
GAT	67.96 ± 0.33	46.17 ± 0.32	48.25 ± 0.05	53.78 ± 1.34	77.67 ± 0.13	46.62 ± 0.32	13.32 ± 0.29
GT	69.23 ± 0.50	46.47 ± 0.16	48.00 ± 0.05	54.50 ± 1.20	76.97 ± 0.21	45.16 ± 0.46	12.74 ± 0.42
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
TS-GNN (ICL)	20.09 ± 1.29	MLE	MLE	38.54 ± 0.94	43.46 ± 5.17	20.44 ± 1.05	MLE
GCOPE (FT)	19.51 ± 0.07	TLE	TLE	28.67 ± 1.42	67.38 ± 1.23	16.10 ± 2.79	TLE

(b) Results for regression datasets. R^2 is reported for all datasets.

	hm-prices	avazu-ctr	city-roads-M	city-roads-L	twitch-views	artnet-views	web-traffic
best const. pred.	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
ResMLP	62.66 ± 0.37	24.54 ± 0.36	54.77 ± 0.15	46.47 ± 0.29	13.35 ± 0.02	29.71 ± 0.60	72.42 ± 0.05
XGBoost	65.68 ± 0.16	26.72 ± 0.02	59.14 ± 0.11	53.75 ± 0.07	13.39 ± 0.00	32.74 ± 0.04	TLE
LightGBM	65.44 ± 0.09	25.83 ± 0.04	57.76 ± 0.10	52.65 ± 0.08	13.38 ± 0.01	32.47 ± 0.04	TLE
CatBoost	66.85 ± 0.28	26.10 ± 0.04	57.53 ± 0.18	51.43 ± 0.17	13.20 ± 0.03	32.89 ± 0.05	TLE
ResMLP-NFA	67.19 ± 0.30	31.11 ± 0.30	57.82 ± 0.14	50.85 ± 0.18	51.43 ± 0.60	51.03 ± 0.41	MLE
LightGBM-NFA	70.46 ± 0.09	31.72 ± 0.06	61.00 ± 0.05	55.26 ± 0.04	60.20 ± 0.01	56.55 ± 0.04	TLE
GCN	69.76 ± 0.38	30.47 ± 0.27	59.05 ± 0.16	53.26 ± 0.14	75.55 ± 0.05	55.99 ± 0.26	82.07 ± 0.14
GraphSAGE	70.54 ± 0.21	31.84 ± 0.24	57.51 ± 0.53	52.43 ± 0.25	66.87 ± 0.11	49.79 ± 0.51	83.50 ± 0.11
GAT	73.17 ± 0.50	33.20 ± 0.20	59.11 ± 0.20	53.43 ± 0.20	72.93 ± 0.17	53.36 ± 0.78	84.68 ± 0.06
GT	71.87 ± 0.65	30.87 ± 0.47	58.05 ± 0.58	53.38 ± 0.12	72.19 ± 0.14	54.23 ± 0.22	84.49 ± 0.07

Source code at GitHub

Datasets API at PyG (in progress)

Raw datasets at Zenodo

Follow **Yandex Research** works at **research.yandex.ru**