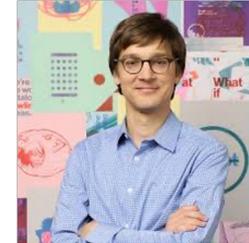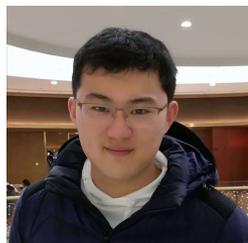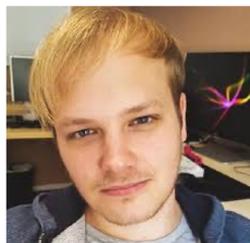# Open Graph Benchmark: Large-Scale Challenge

Weihua Hu

Joint work with
Matthias Fey, Hongyu Ren, Maho Nakata, Yuxiao Dong, Jure Leskovec

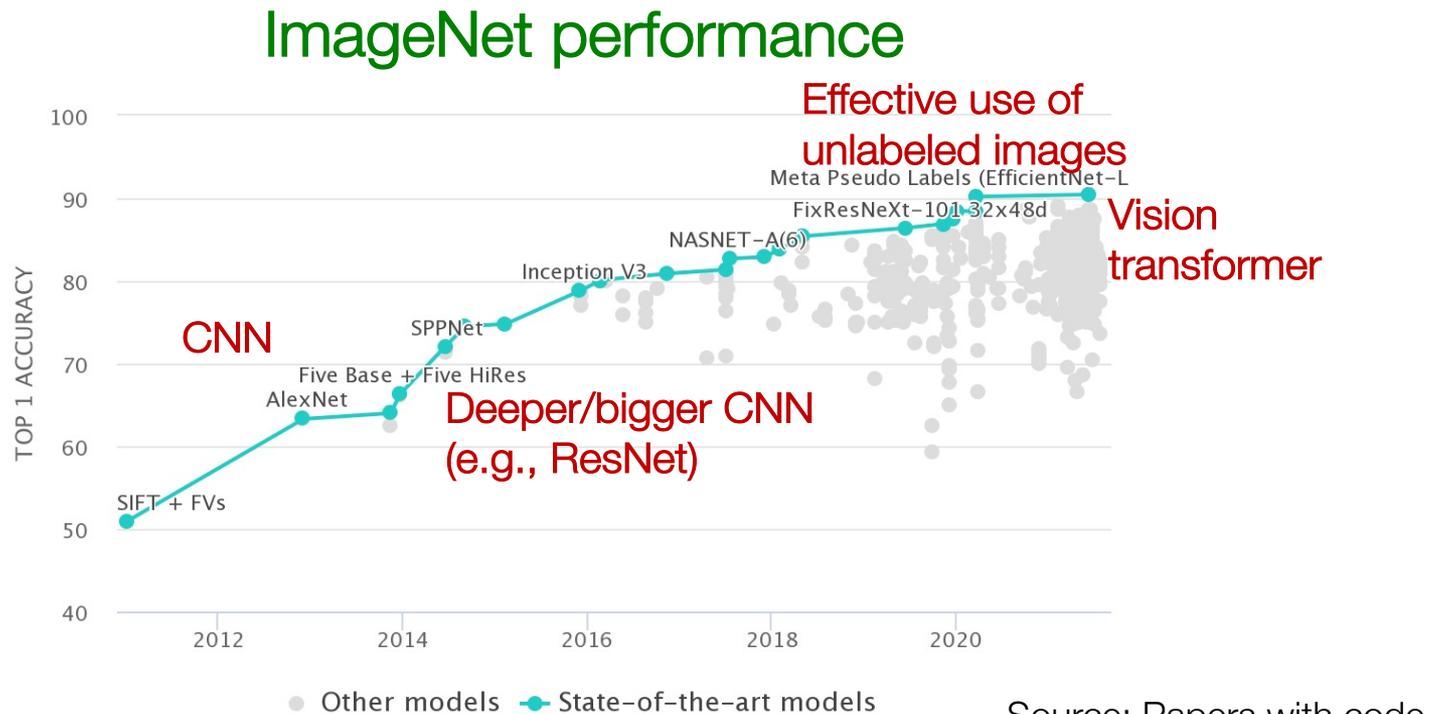# Benchmarks are Important

- Historically…
  - Computer vision
  - Natural Language Processing

Benchmarks are Important

- Challenging and realistic benchmark has driven methodological innovation.

ImageNet performance

Effective use of unlabeled images

Vision transformer

Deeper/bigger CNN (e.g., ResNet)

CNN

Meta Pseudo Labels (EfficientNet-L

FixResNeXt-101-32x48d

NASNET-A(6)

Inception V3

SPPNet

Five Base + Five HiRes

AlexNet

SIFT + FVs

Other models — State-of-the-art models

Source: Papers with code

Evaluating Graph ML

To advance research in graph ML, it is critical for our community to

develop diverse, challenging, and realistic benchmark datasets for machine learning on graphs

Issues with Existing Graph Benchmark

1. Datasets are small

2. Experimental protocol is not unified

3. Dataset splits follow conventional random splits

Hard to rigorously evaluate graph ML models.

Open Graph Benchmark

- In May 2020, we introduced **OGB**: Realistic and diverse benchmark datasets for graph ML
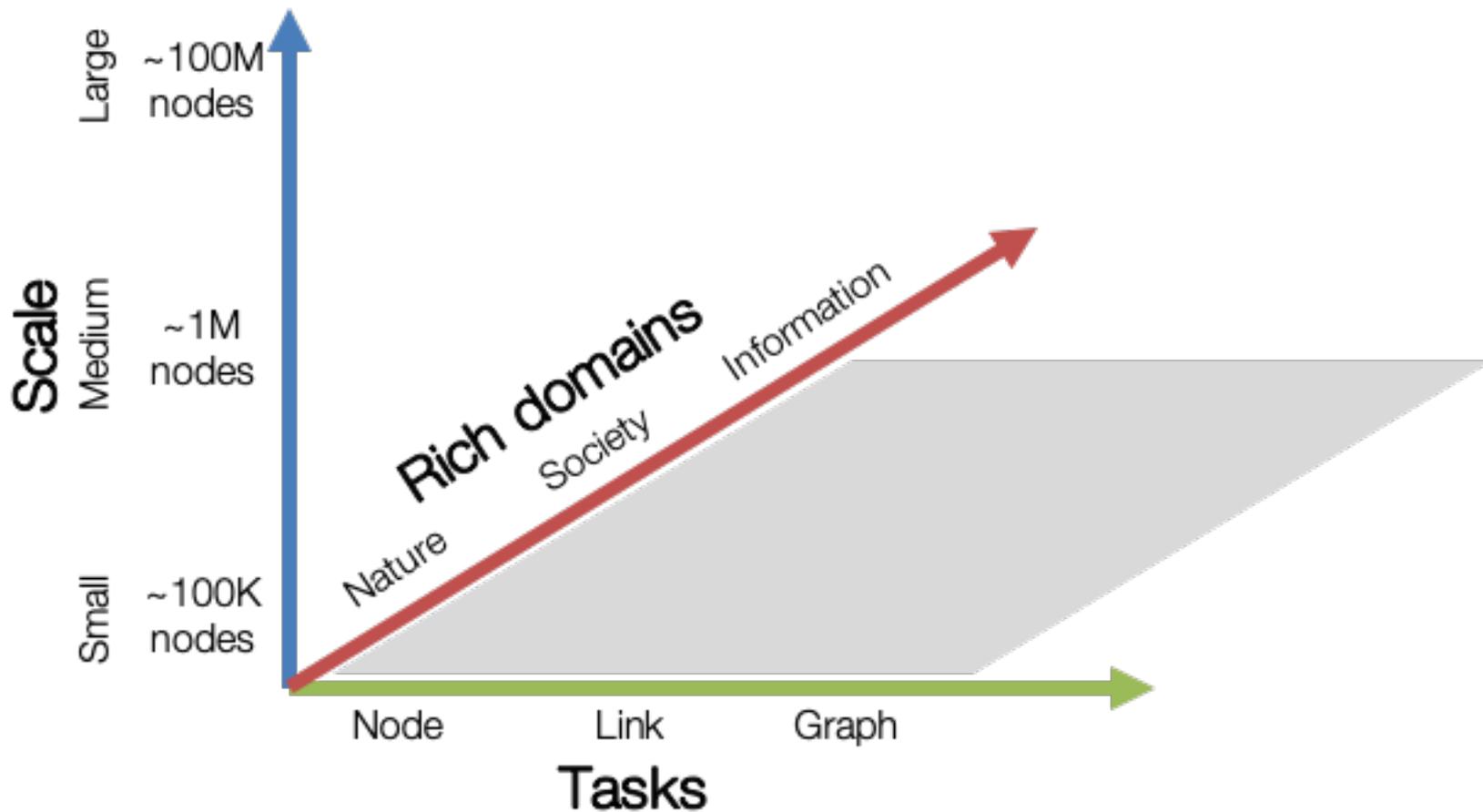

Webpage: https://ogb.stanford.edu/
Paper: https://arxiv.org/abs/2005.00687
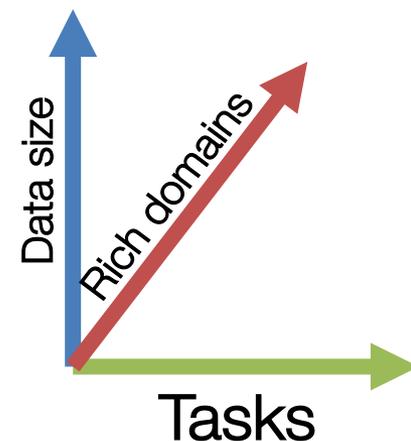Github: https://github.com/snap-stanford/ogb

OGB Datasets are Diverse

OGB Datasets are Diverse

Core tasks:

- Node property prediction
  - Node label classification

- Link property prediction
  - Link existence prediction
  - KG completion

- Graph property prediction
  - Molecule property prediction
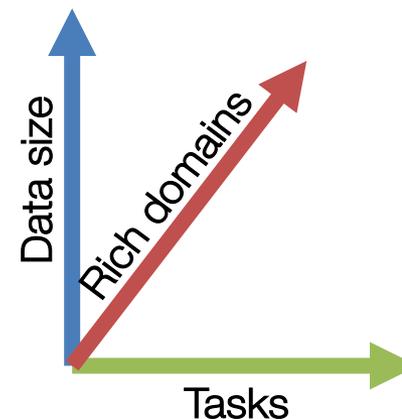 - Method naming of code snippets.

OGB Datasets are Diverse

- ## Domains:
  - Natural sciences (chemistry, biology)
  - Social and information networks
  - Knowledge graphs, code
  - **Richness of node/edge features**

- ## Dataset size:
  - Small rich graphs with 100K nodes.
  - Medium graphs with 1M to 3M nodes.
  - Large graphs with more than 100M nodes.

Data size

Rich domains

Tasks

Open Graph Benchmark

- OGB includes 15 datasets from diverse domains and tasks.

Task	Node property prediction ogbn-		
Domain	Nature	Society	Information
Small		arxiv	
Medium	proteins	products	mag
Large		papers100M	

Task	Link property prediction ogbl-		
Domain	Nature	Society	Information
Small	ddi	collab	biokg
Medium	ppa	citation2	wikikg2
Large			

Task	Graph property prediction ogbg-		
Domain	Nature	Society	Information
Small	molhiv		
Medium	molpcba / ppa		code2
Large			

Open Graph Benchmark

- Many methods have been developed.
  - Over 300 leaderboard submissions
  - Drastic accuracy improvement on many datasets

ogbg-molpcba (molecule classification)

+4% AP improvement over our best baseline

ogbn-products (product classification)

+5% accuracy improvement over our best baseline
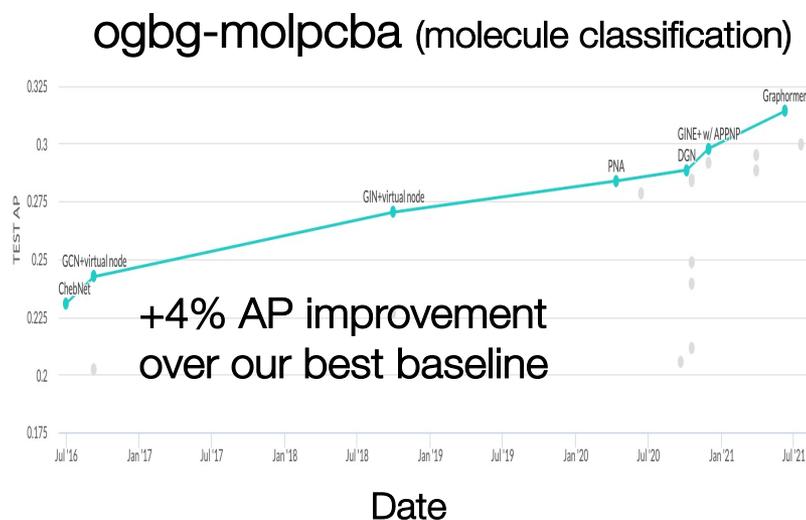
Source: Papers with code

Need for Large-Scale Benchmark

- To yield breakthroughs in graph ML, we need a large-scale graph benchmark

- Bigger and advanced models benefit from larger data

  - Deeper and more expressive GNNs can yield performance breakthrough

- ML on large-scale graphs is challenging and requires innovations:

  - Training GNNs on large graphs requires non-trivial mini-batch sampling of nodes/edges

OGB Large-Scale Challenge

- For the ACM KDD Cup 2021, we provide a set of three challenging large-scale graph datasets

Webpage: https://ogb.stanford.edu/kddcup2021/
Paper: https://arxiv.org/abs/2103.09430
Github: https://github.com/snap-stanford/ogb

OGB-LSC Stats

- Attracted huge attention from the community
  - 500+ registrations across the globe.
  - 143 teams submitted to the initial test submission.
  - 123 teams submitted to the final test submission.
- Institutions:
 - Academia: 60%, Industry: 40%

OGB-LSC Datasets

- LSC datasets are orders-of-magnitude larger than any exiting datasets

Task type	Dataset	Statistics	
Node-level	MAG240M–LSC	#nodes:	244,160,499
		#edges:	1,728,364,232
Link-level	WikiKG90M–LSC	#nodes:	87,143,637
		#edges:	504,220,369
Graph-level	PCQM4M–LSC	#graphs:	3,803,453
		#edges (total):	55,399,880

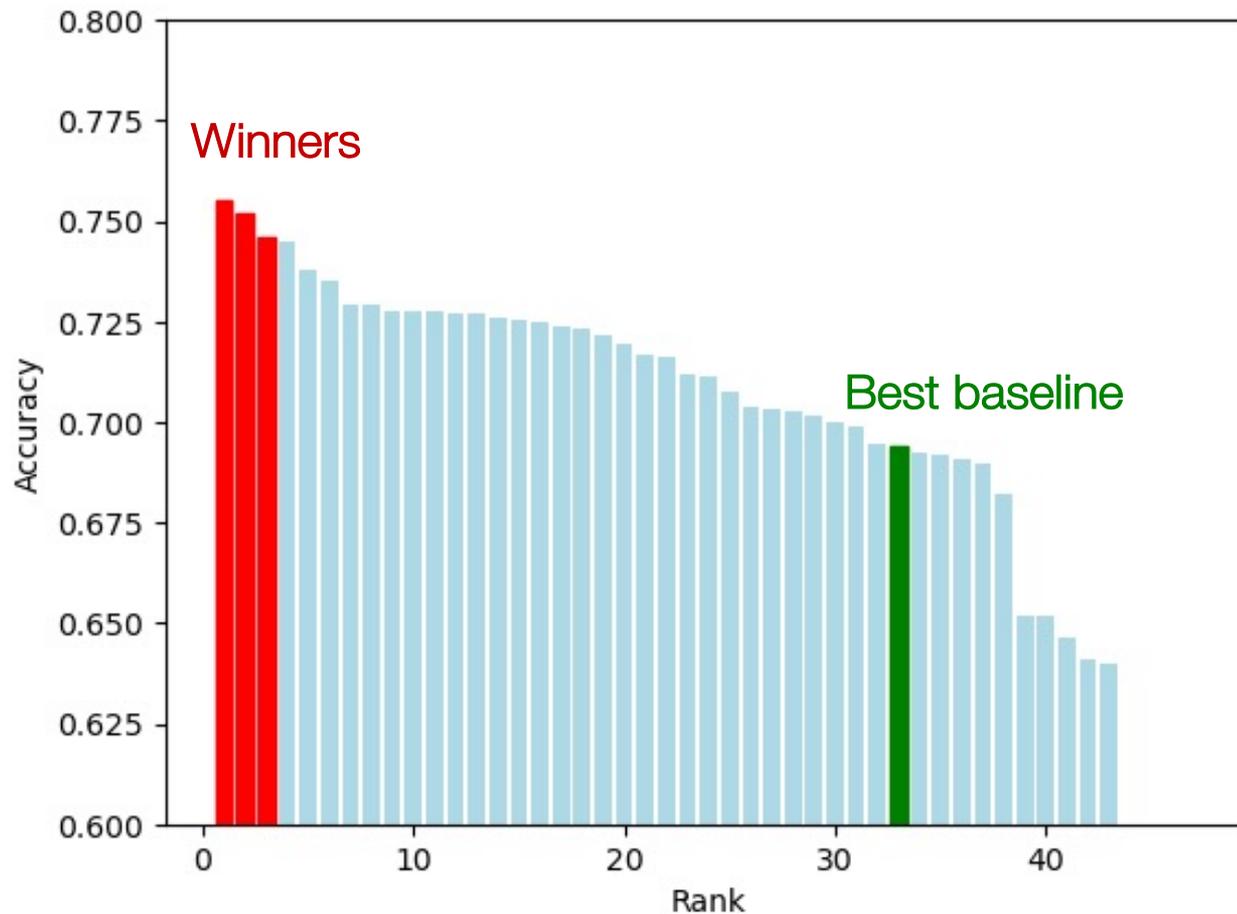- Each dataset is practically relevant

Node-Level: MAG240M

- Heterogeneous academic graph
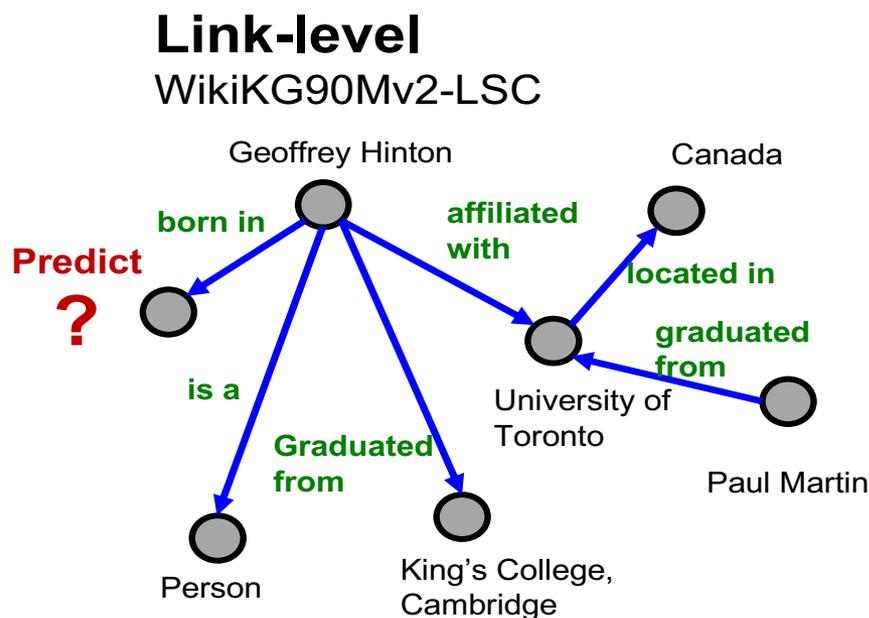- Task: Predict the subject areas of papers situated in the heterogeneous graph (node classification)
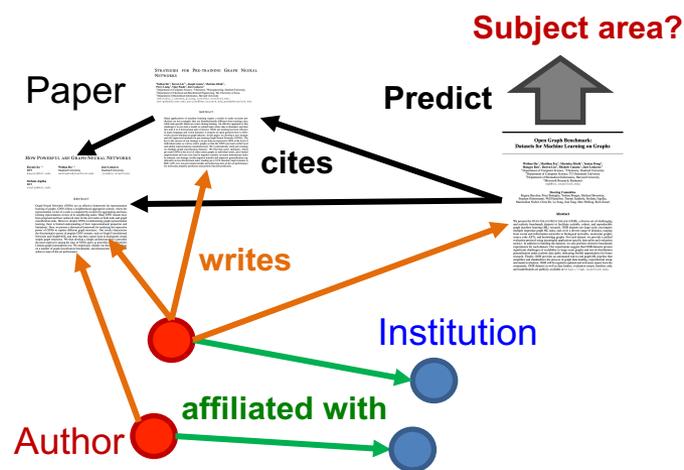
Performance Improvement
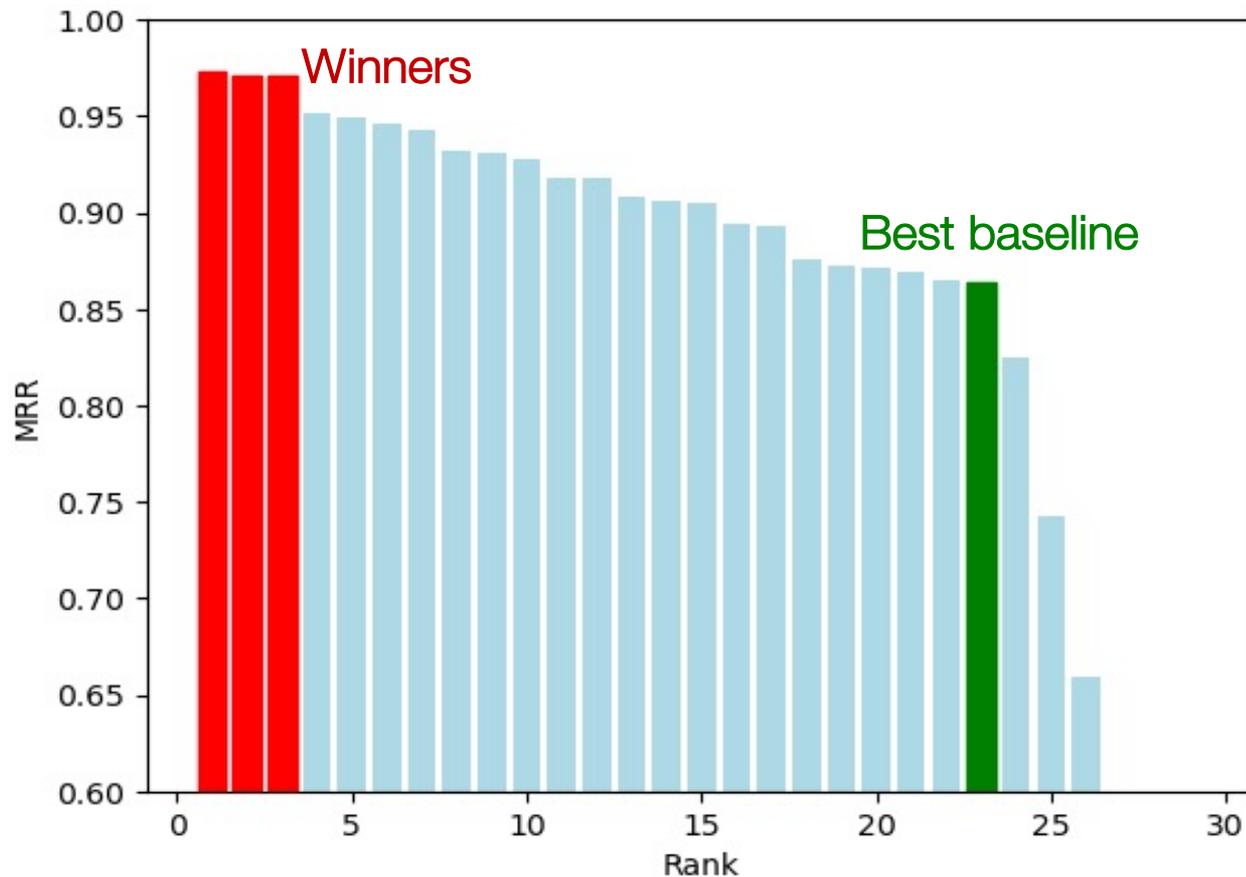
Accuracy: The higher, the better.

Link-Level: WikiKG90M

- Knowledge graph
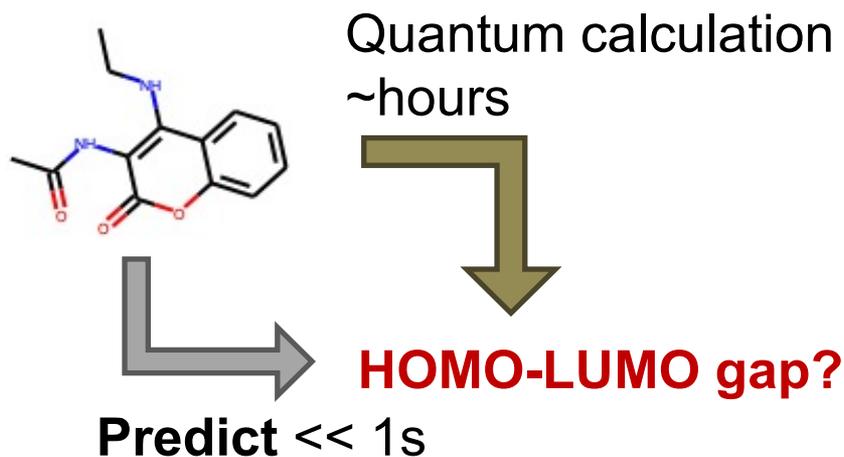
- Task: Impute missing triplets (link prediction)

Link-level
WikiKG90Mv2-LSC

Performance Improvement

MRR: The higher, the better

Winners

Best baseline

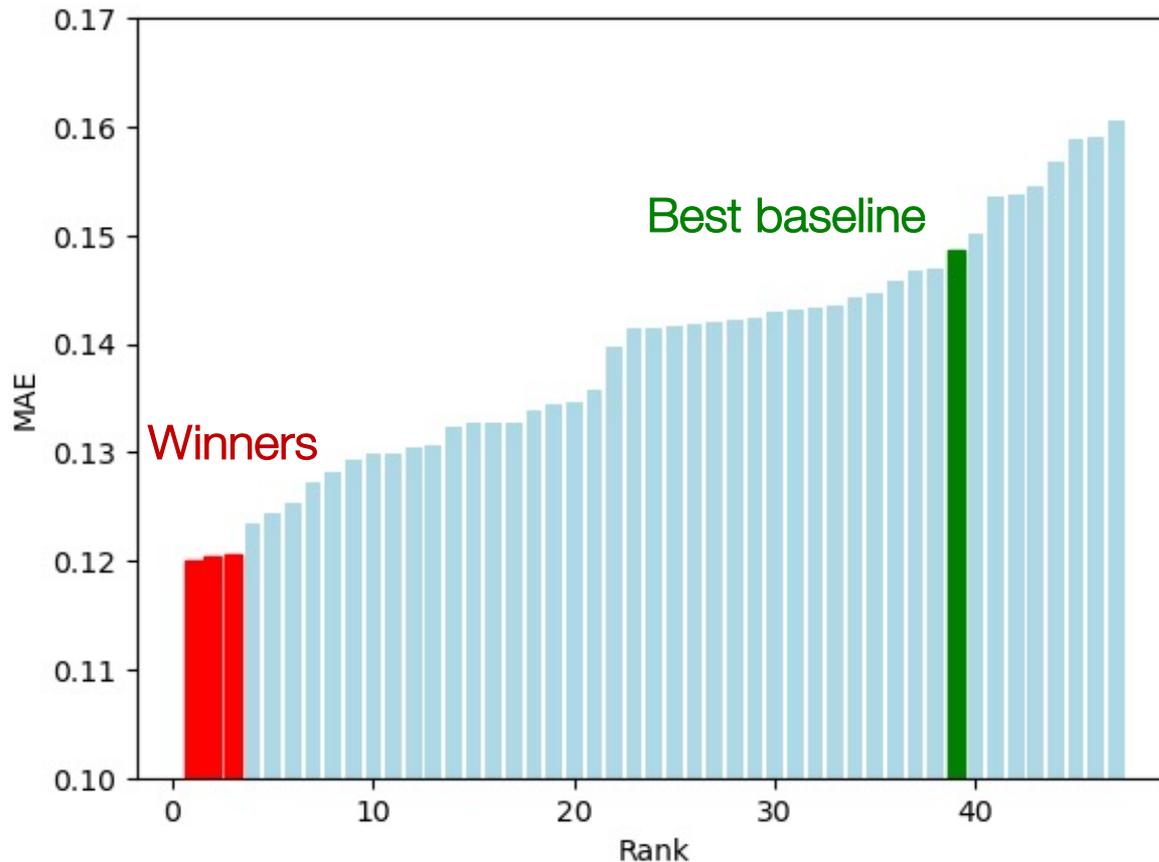# Graph-Level: PCQM4M

- Molecular graphs

- Task: Predict an important quantum chemistry property, the HOMO-LUMO gap, of a given molecule (graph regression).

Quantum calculation ~hours

HOMO-LUMO gap?

Predict << 1s

Performance Improvement

MAE: The lower, the better.

Overall Observations

- Many novel techniques developed for the OGB-LSC large graphs
    - New mini-batch sampling for hetero-generous graphs
    - New label propagation methods using GNNs.
    - New knowledge graph embedding models
    - New self-supervised learning methods for GNNs
    - Deeper, bigger, and more expressive GNNs

Deeper and Bigger GNNs

- On the molecule dataset:
    - Number of GNN layers
        - Our baseline: 5 layers
        - Winners: 11—50 layers
    - Number of parameters (single model)
        - Our baseline: 6.7M
        - Winners: 50M—450M
    - Bigger models indeed perform better!

More Expressive GNNs

■ The winner uses Transformer-style fully-connected message passing.

Input graph

Fully-connected graph

Attention aggregation from all the nodes
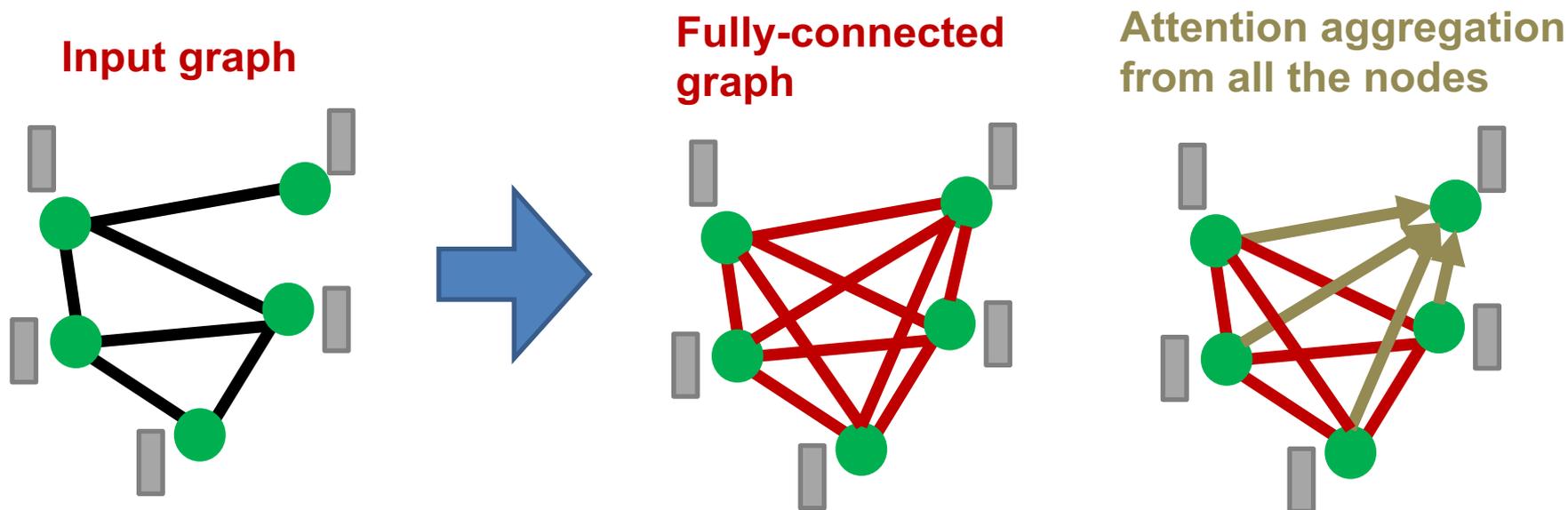
Outlook for the Future

- We hope the community will continue to use the OGB-LSC datasets to develop new methods.

- **Based on the lessons from the KDD Cup,** OGB-LSC datasets will be updated to further facilitate research advances.

- We also hope to facilitate the development of efficient ML systems for large-scale graphs.

Conclusions

- **Large-scale graphs are ubiquitous** in real-world applications but **are challenging to handle.**

- OGB-LSC provides a set of **three unprecedently large graph datasets.**

- At the ACM KDD Cup 2021, **many innovative methods have been developed.**