

# Graph Representation Learning and Pre-Training

**Yuxiao Dong**

Knowledge Engineering Group (KEG)  
Department of Computer Science and Technology  
Tsinghua University

<https://keg.cs.tsinghua.edu.cn/yuxiao/>  
[yuxiaod@tsinghua.edu.cn](mailto:yuxiaod@tsinghua.edu.cn)

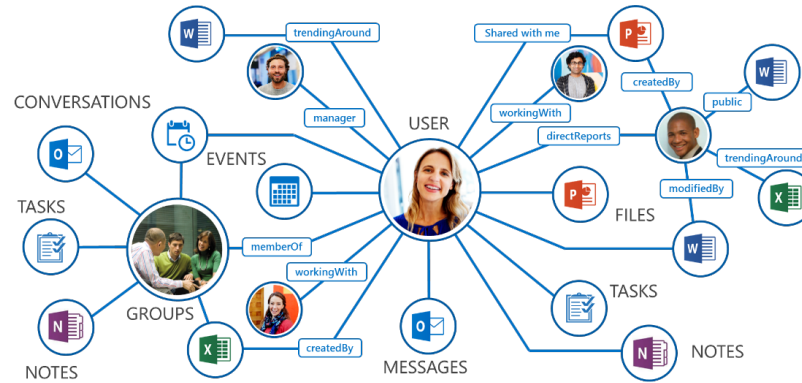
# Joint Work with

*Xiao Liu, Jiezhong Qiu, Ziniu Hu, Wenzheng Feng, Zhenyu Hou  
Yukuo Cen, Weihua Hu, Jie Zhang, Chenhui Zhang, Yuyang Xie  
Hao Ma, Kuansan Wang, Yizhou Sun, Jure Leskovec, Jie Tang*

# Graphs in Society



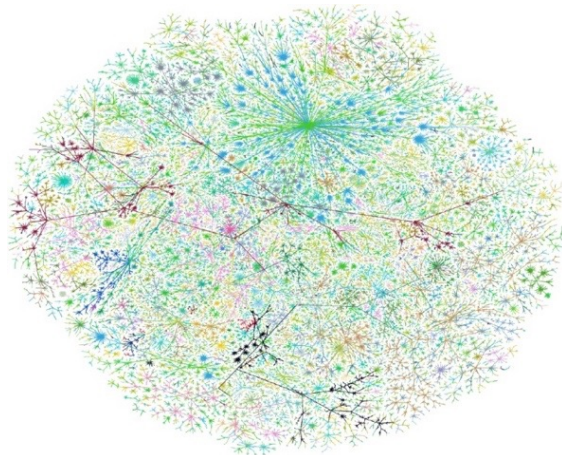
Academic Graph



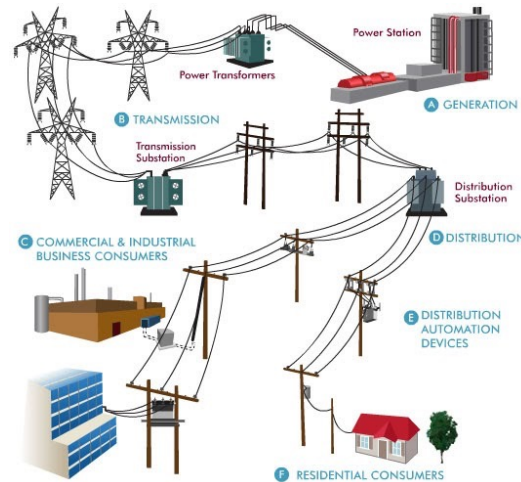
Social & Office Graph



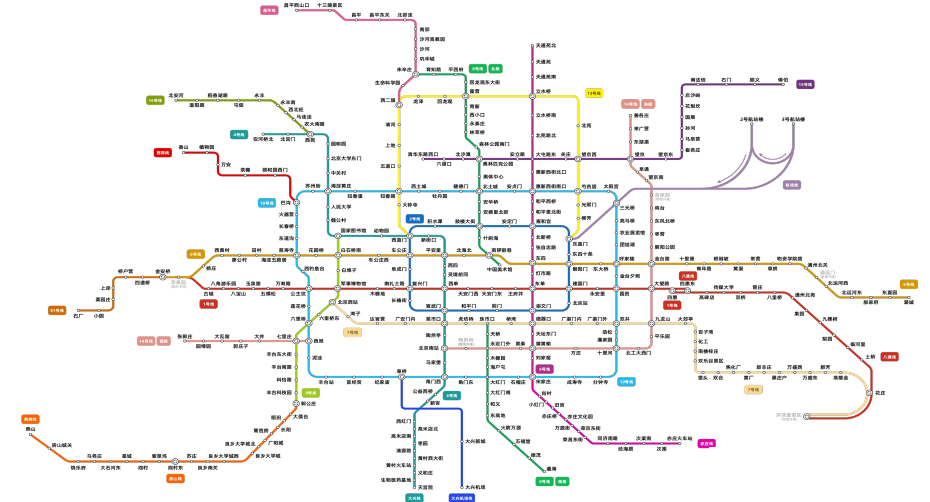
Knowledge Graph



Internet

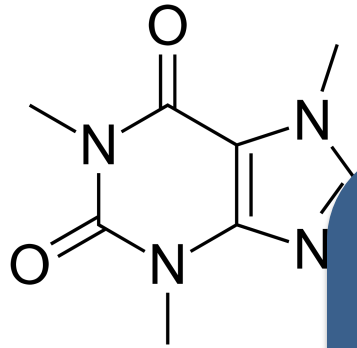


Electrical Grid Network

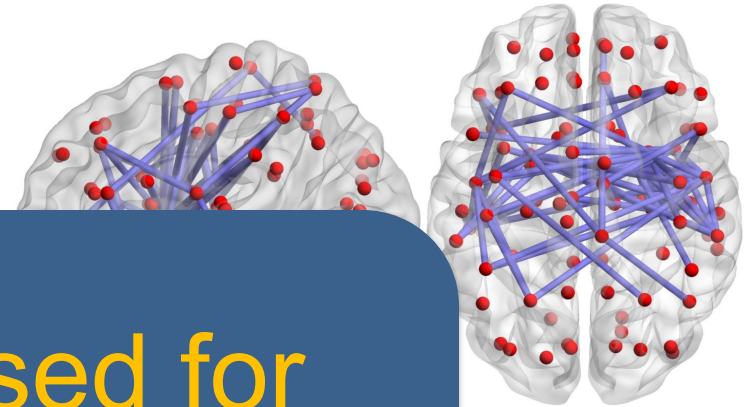


Transportation

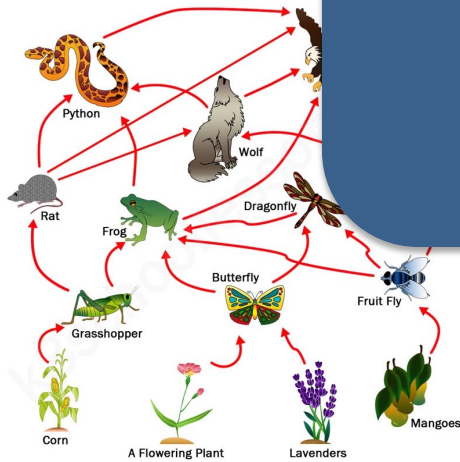
# Graphs in Nature



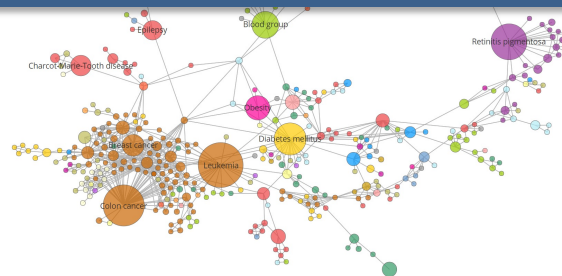
Molecules



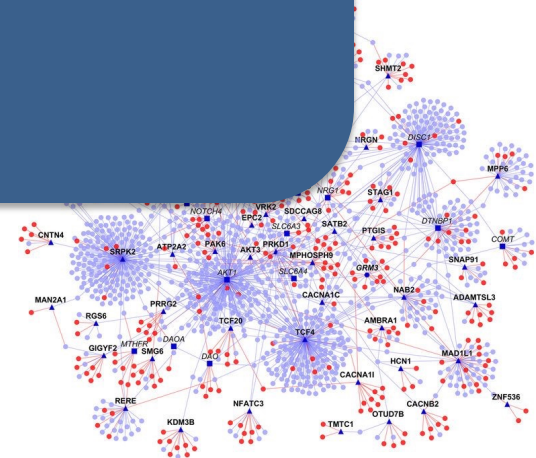
Graphs are widely used for abstracting complex systems of interacting objects!



Food Web

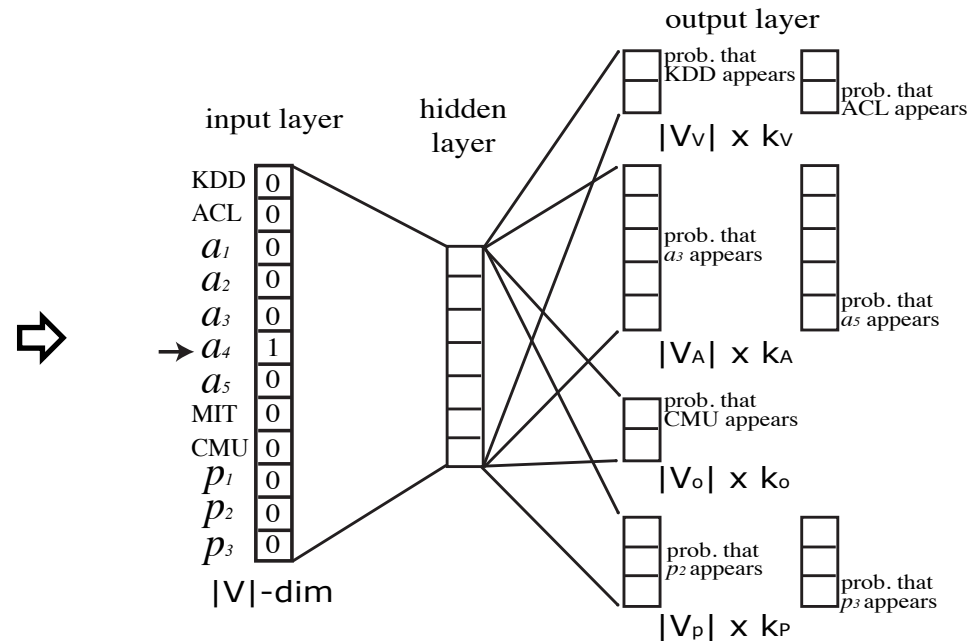


Human Disease Networks

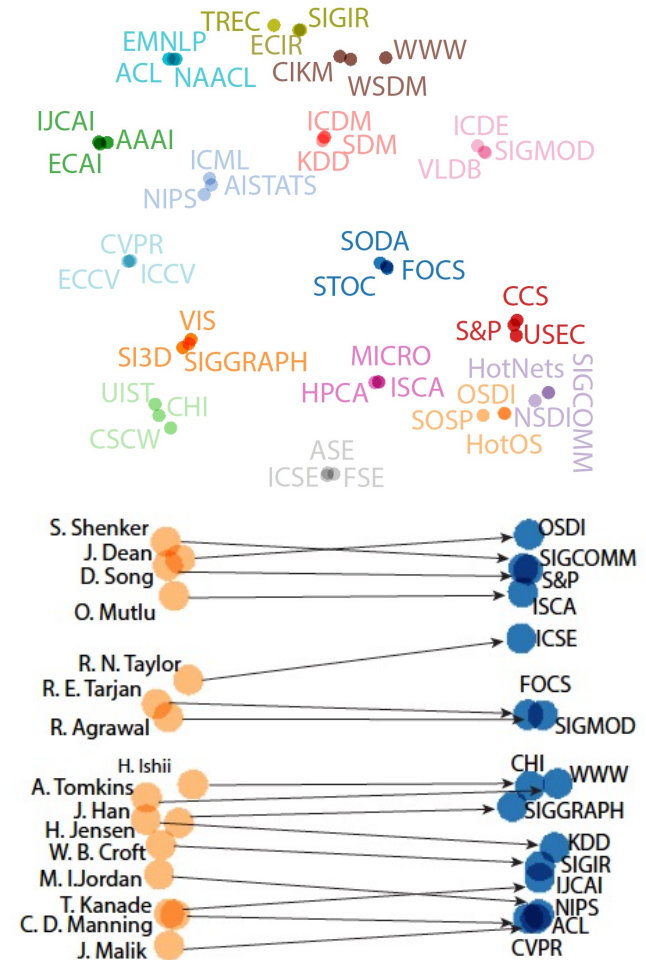


Protein-Protein Interactions

# Graph Representation Learning

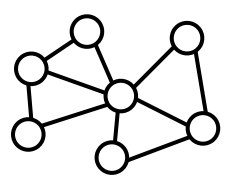
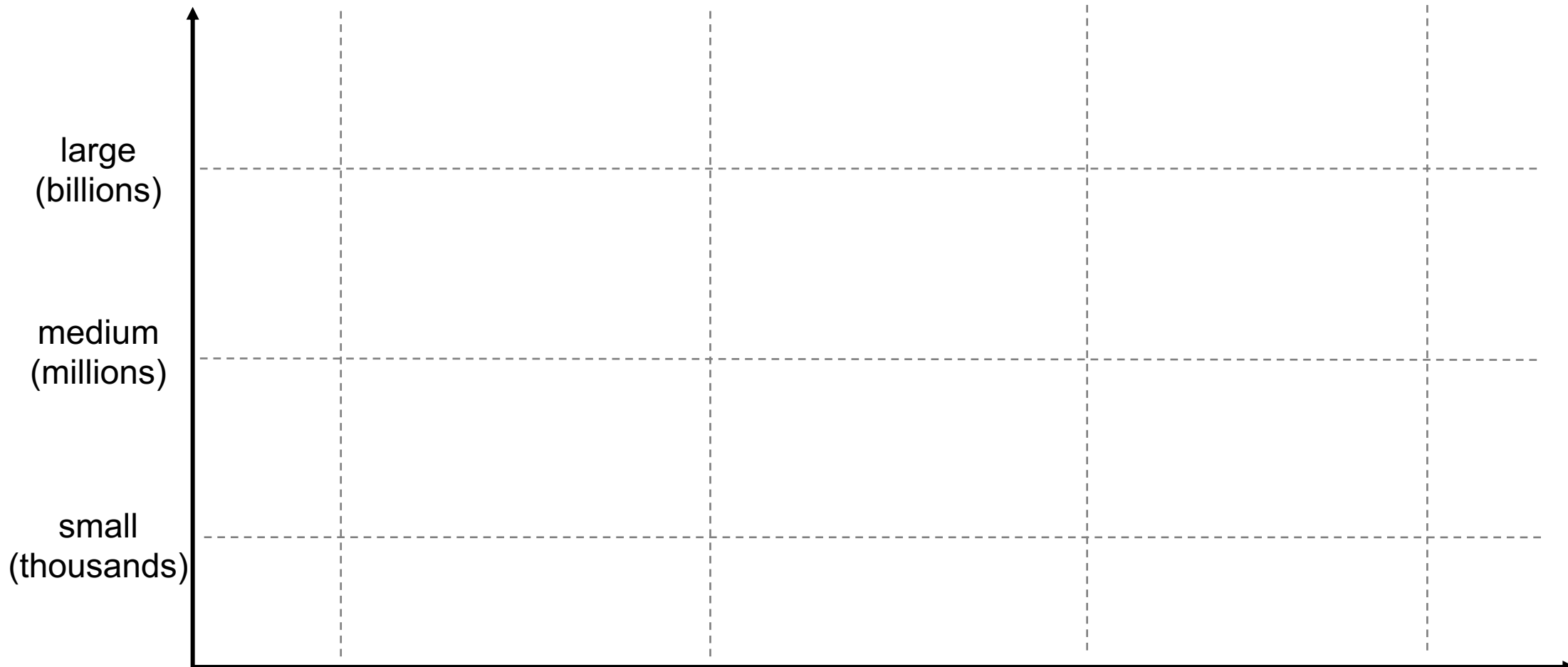


- Input: a graph  $G = (V, E)$
- Output:  $Z \in R^{|V| \times k}$ ,  $k \ll |V|$ ,  $k$ -dim vector  $Z_v$  for each node  $v$ .

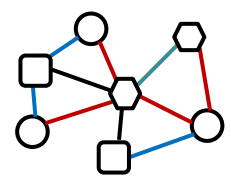




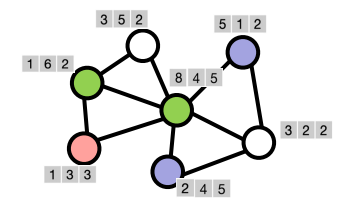




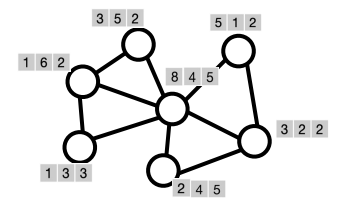
structure



heterogeneous structure / knowledge graph

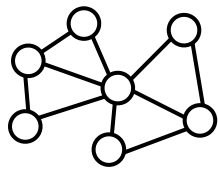


structure with features

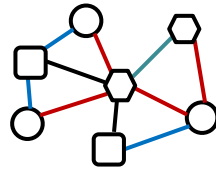


no label (pre-training)

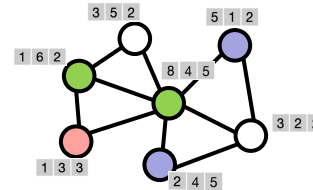
# Network Embedding



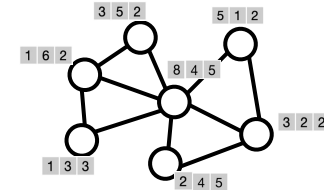
structure



heterogeneous structure / knowledge graph



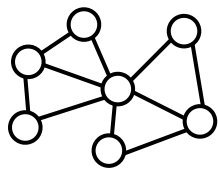
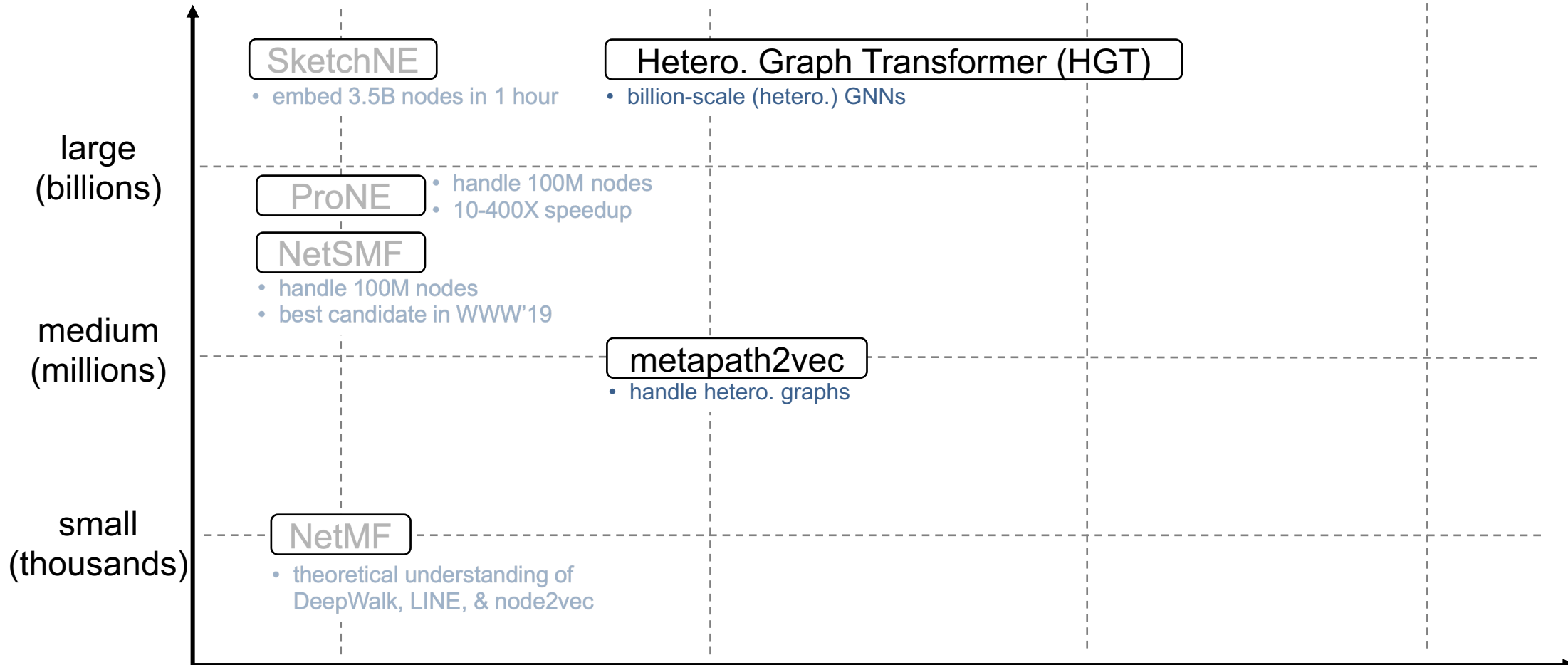
structure with features



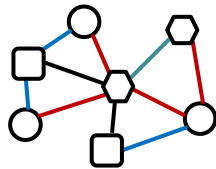
no label (pre-training)



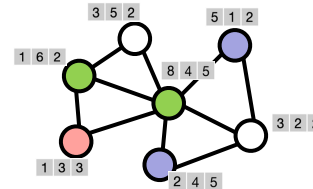
# Heterogeneous Graphs



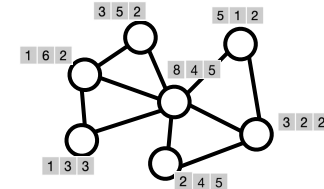
structure



heterogeneous structure / knowledge graph

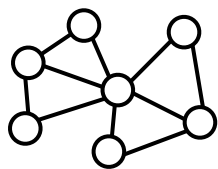
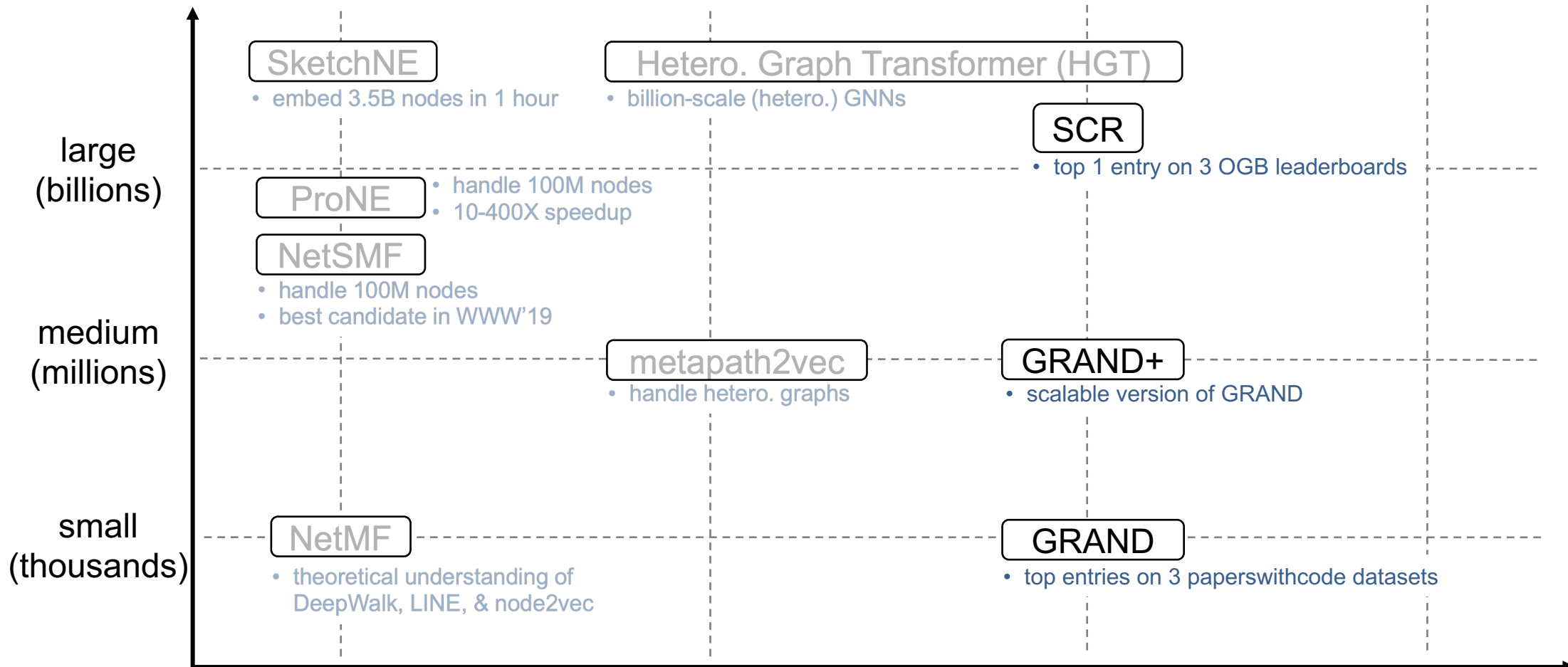


structure with features

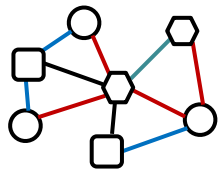


no label (pre-training)

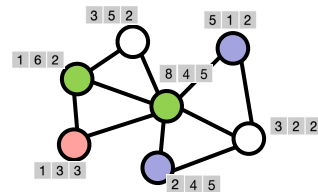
# Graph Neural Networks



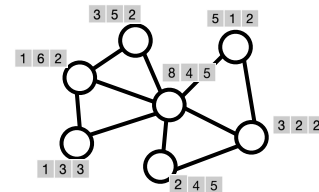
structure



heterogeneous structure / knowledge graph

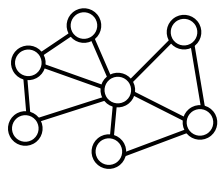
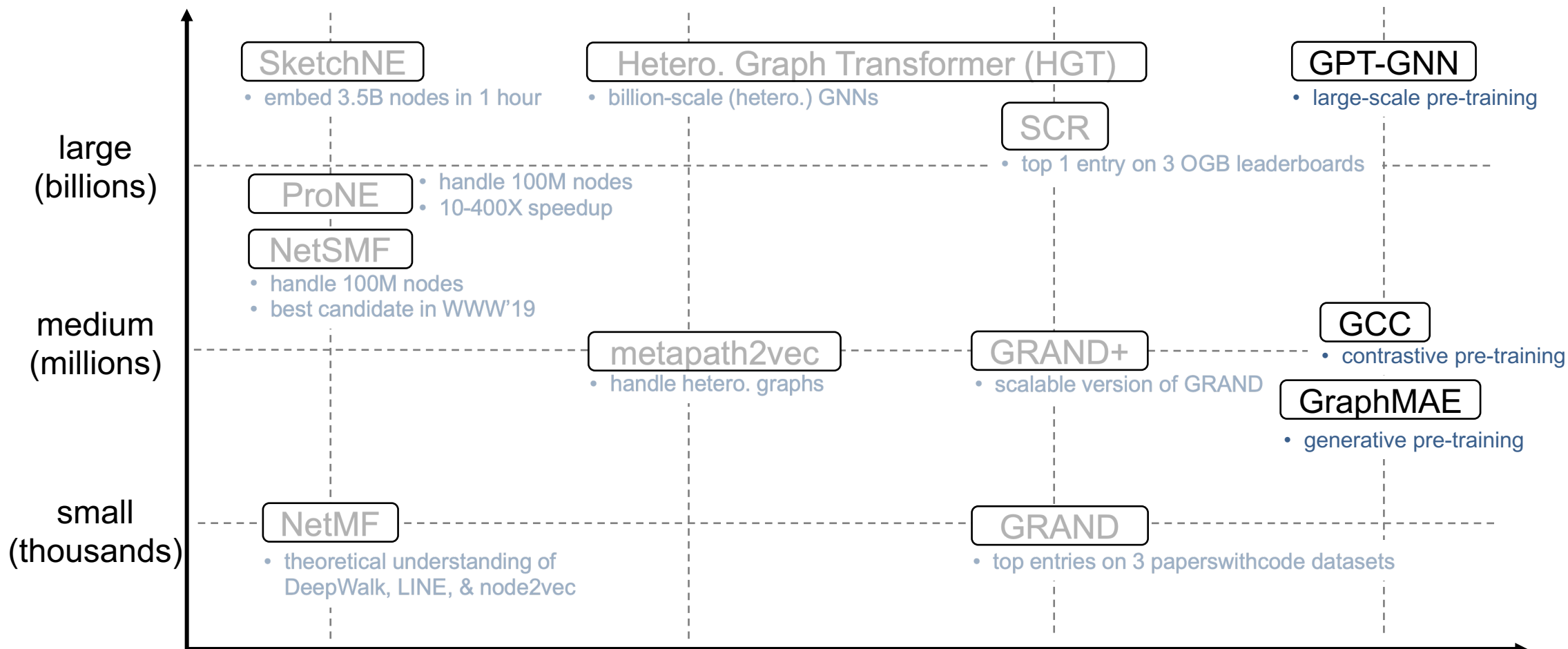


structure with features

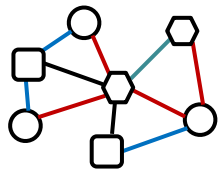


no label (pre-training)

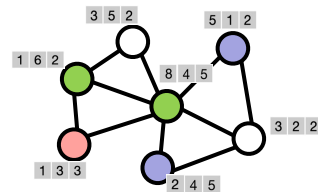
# Graph Pre-Training and Self-Supervised Learning



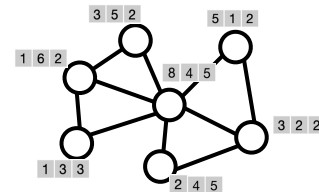
structure



heterogeneous structure / knowledge graph

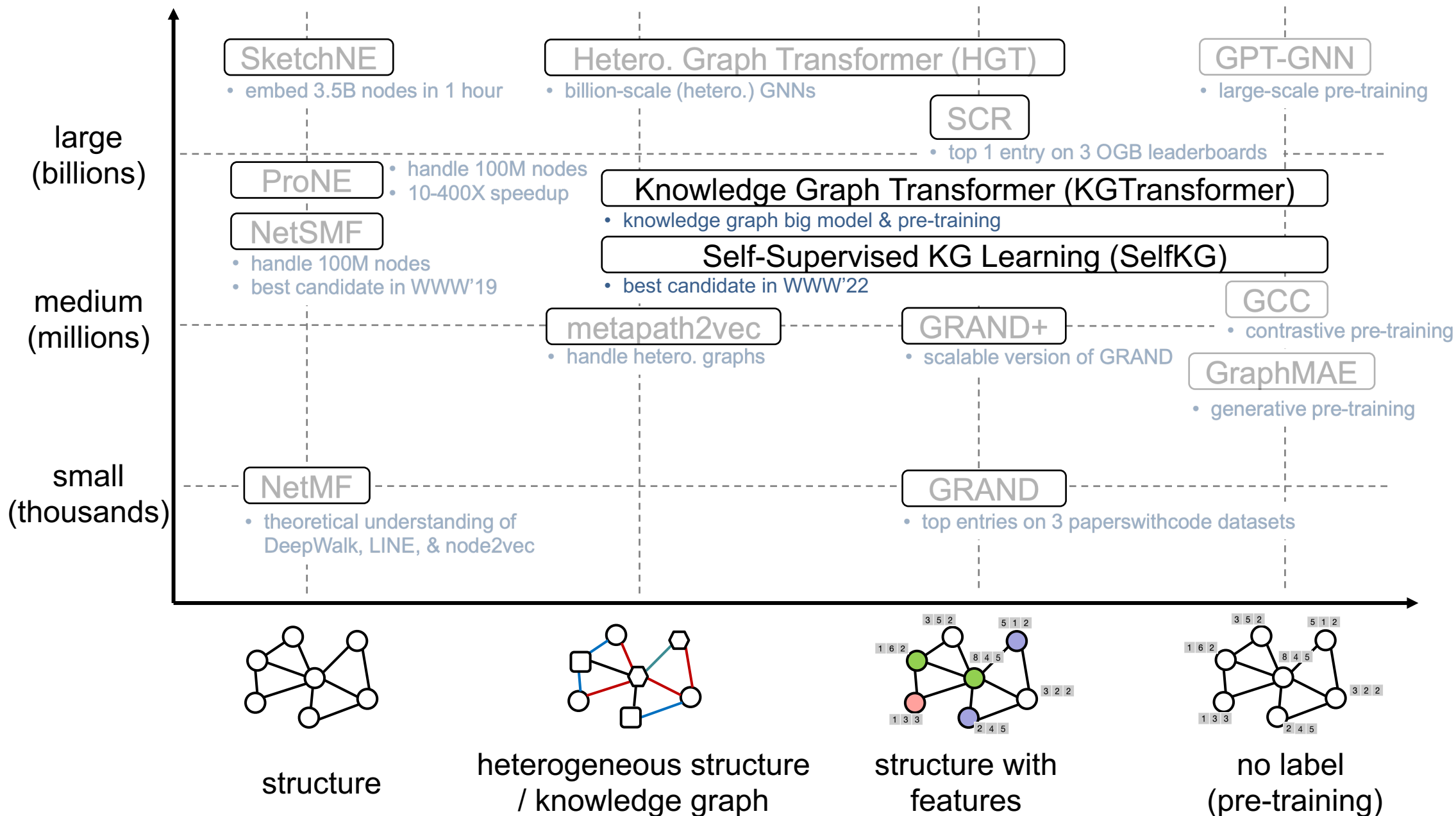


structure with features

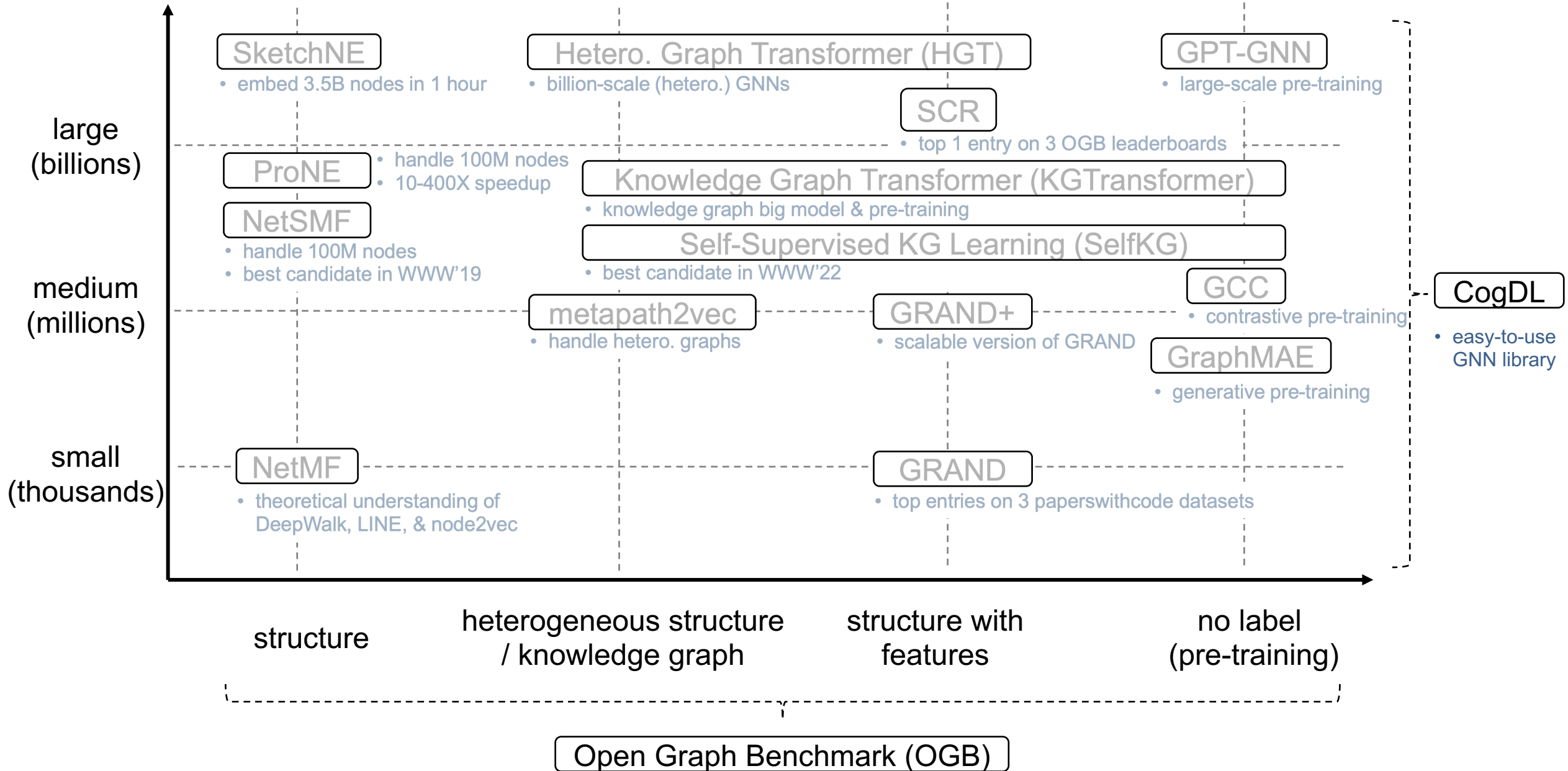


no label (pre-training)

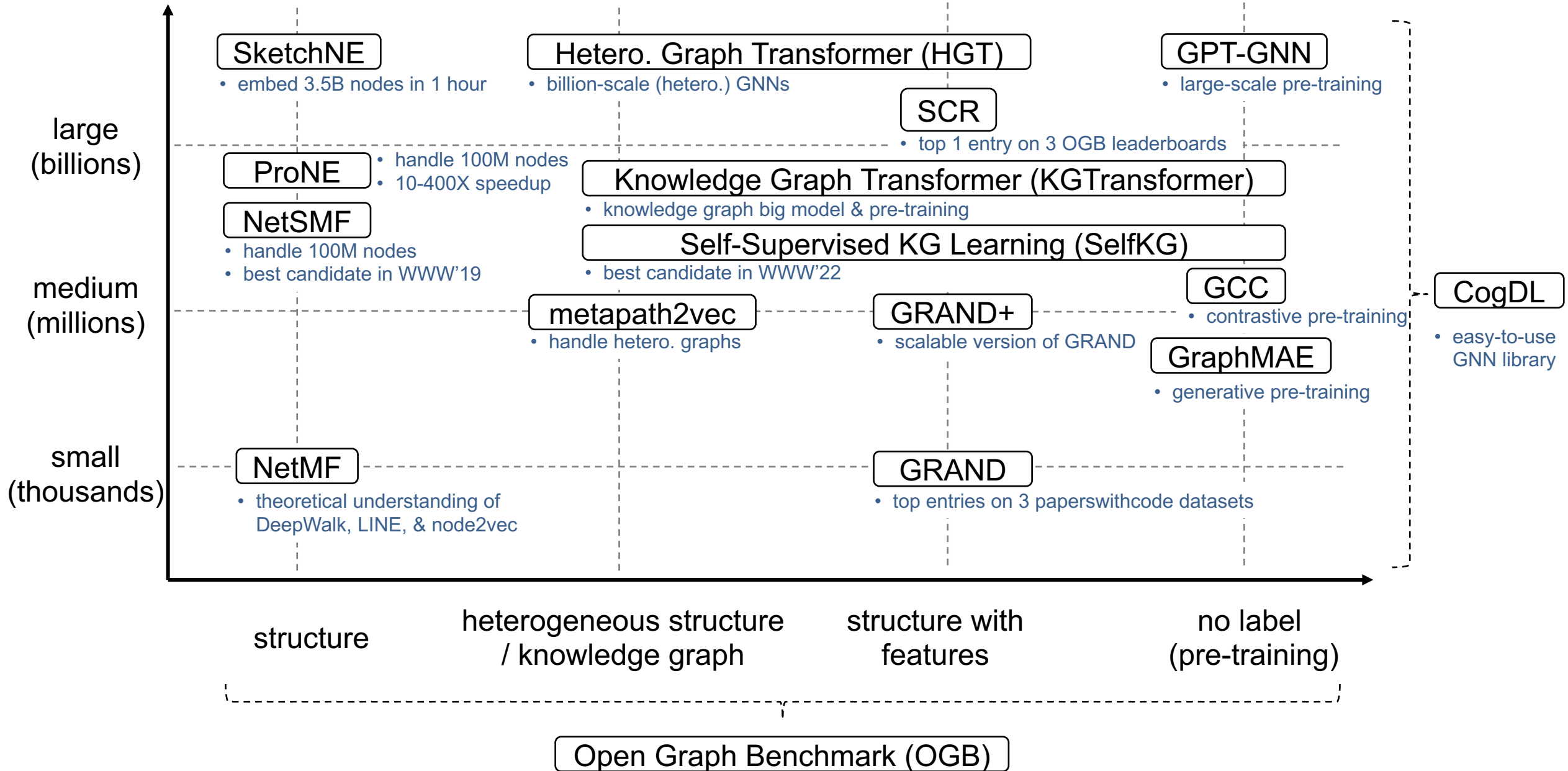
# Graph Pre-Training and Self-Supervised Learning



# Open Data & Toolkit



# Graph Representation Learning and Pre-Training





# Pre-Train Graphs with *Language/Image/Knowledge*



Academic Graph



Microsoft Office Graph

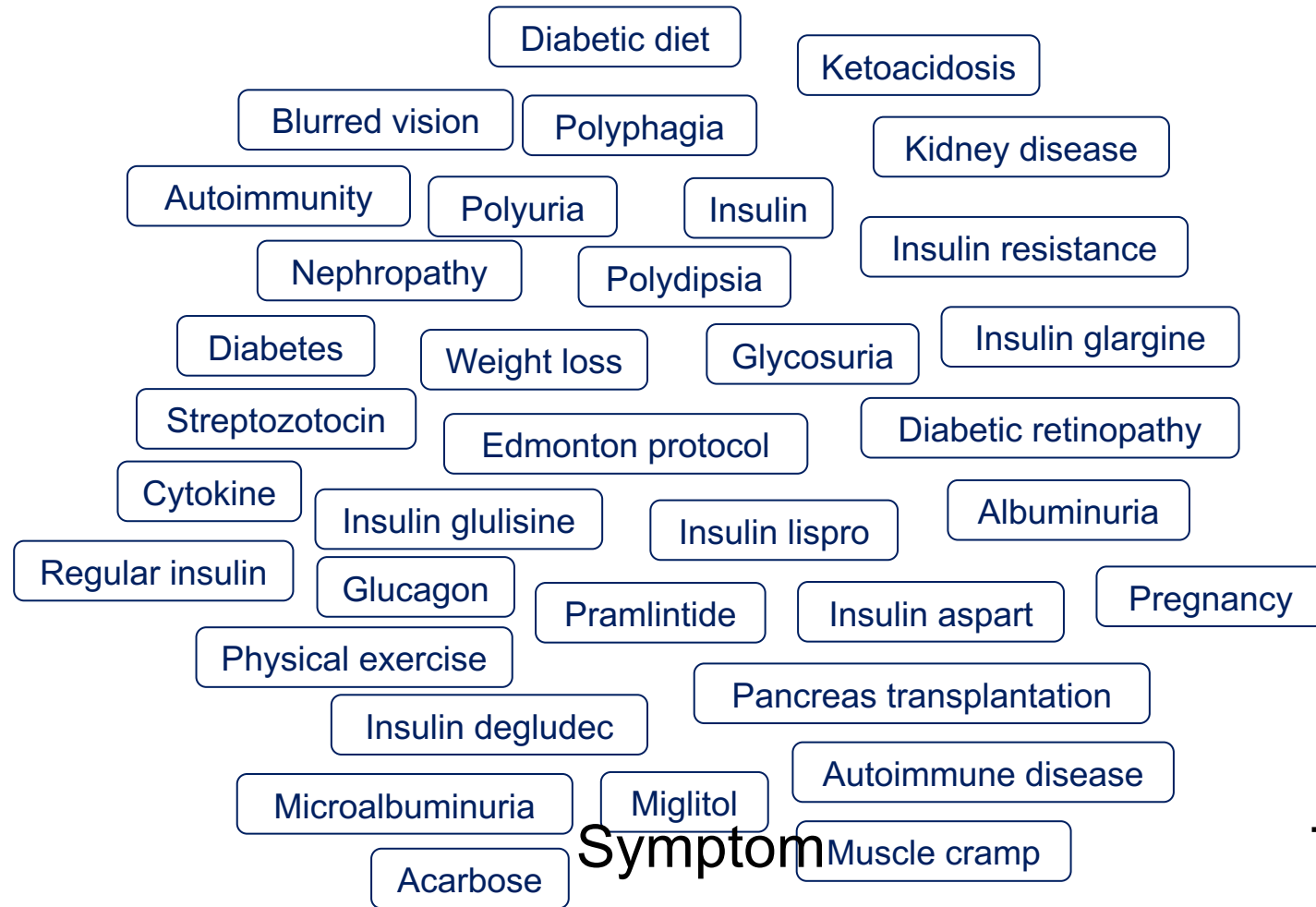


LinkedIn Economic Graph



Facebook Entity Graph

# Pre-Train for Cognitive Reasoning

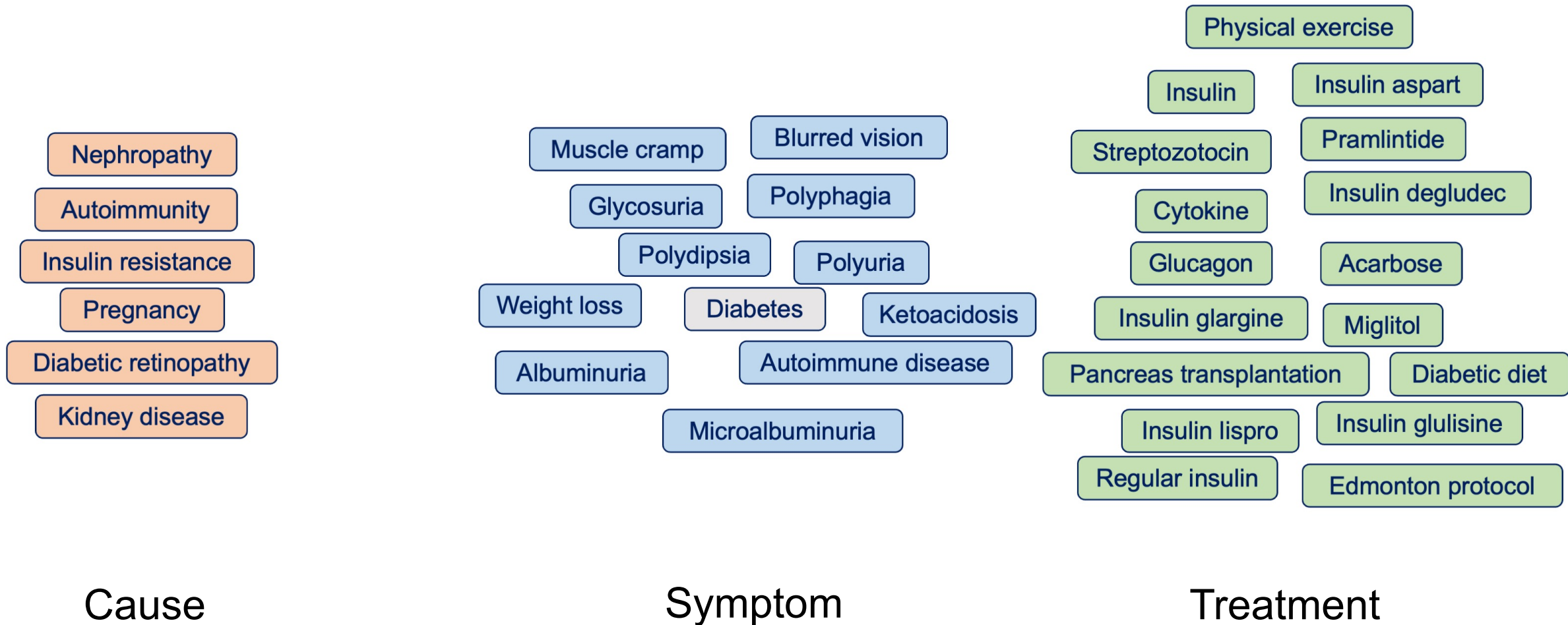


Cause

Symptom

Treatment

# Pre-Train for Cognitive Reasoning



# References

1. Zhenyu Hou, et al. *GraphMAE: Self-Supervised Masked Graph Autoencoders*. KDD 2022.
2. Xiao Liu, et al. *Mask and Reason: Pre-Training Knowledge Graph Transformers for Complex Logical Queries*. KDD 2022.
3. Xiao Liu, et al. *SelfKG: Self-Supervised Entity Alignment in Knowledge Graphs*. WWW 2022. **Best Paper Candidate**.
4. Wenzheng Feng, et al. *GRAND+: Scalable Graph Random Neural Networks*. WWW 2022.
5. Yukuo Cen, et al. *CogDL: A Unified Library for Graph Neural Networks*. <https://cogdl.ai/>.
6. Chenhui Zhang, et al. *SCR: Training Graph Neural Networks with Consistency Regularization*. arXiv.
7. Tinglin Huang, et al. *MixGCF: An Improved Training Method for Graph Neural Network-based Recommender Systems*. KDD 2021.
8. Xu Zou, et al. *TDGIA: Effective Injection Attacks on Graph Neural Networks*. KDD 2021.
9. Wenzheng Feng, et al. *Graph Random Neural Networks for Semi-Supervised Learning on Graphs*. NeurIPS 2020.
10. Weihua Hu et al. *Open Graph Benchmark: Datasets for Machine Learning on Graphs*. NeurIPS 2020.
11. Ziniu Hu et al. *GPT-GNN: Generative Pre-Training of Graph Neural Networks*. KDD 2020. *Top cited in KDD'20*
12. Jiezhong Qiu et al. *GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training*. KDD 2020. *Most cited in KDD'20*
13. Ziniu Hu et al. *Heterogeneous Graph Transformer*. WWW 2020. *Most cited in WWW'20*
14. Yuxiao Dong et al. *Heterogeneous Network Representation Learning*. IJCAI 2020.
15. Jie Zhang et al. *ProNE: Fast and Scalable Network Representation Learning*. IJCAI 2019.
16. Jiezhong Qiu et al. *NetSMF: Large-Scale Network Embedding as Sparse Matrix Factorization*. WWW 2019. **Best Paper Candidate**
17. Jiezhong Qiu et al. *Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec*. WSDM 2018. *Most cited in WSDM'18*
18. Yuxiao Dong et al. *metapath2vec: Scalable Representation Learning for Heterogeneous Networks*. KDD 2017. *Most cited in KDD'17*

Thank you!

Papers & data & code available at <https://keg.cs.tsinghua.edu.cn/yuxiao/yuxiaod@tsinghua.edu.cn>