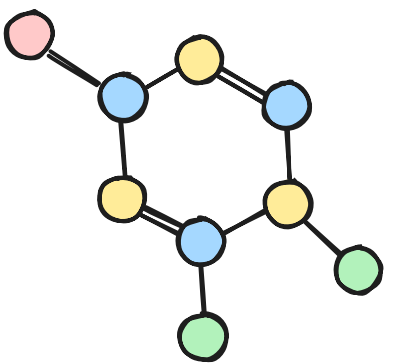
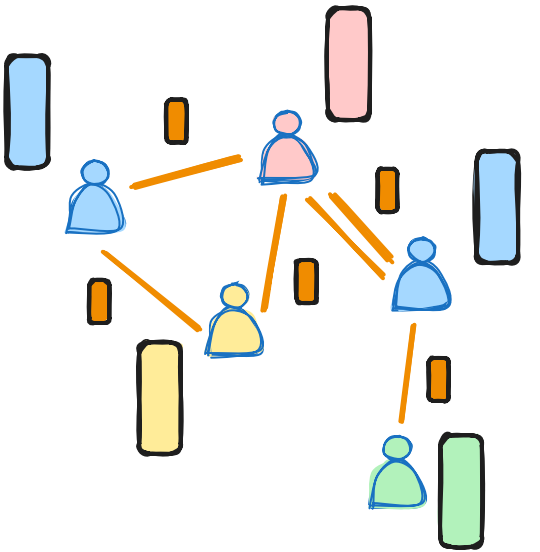
Generative Graph Pattern Machine

Jonathan Maillefaud
Adviser: Saku Peltonen

What is graph learning?

$$G = (V, E) \text{ with } E \subseteq V \times V$$

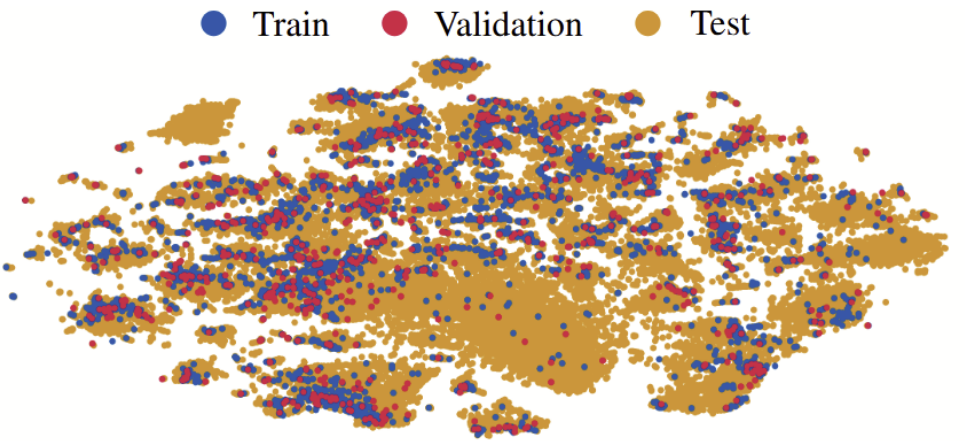




Homophily/Heterophily

Node Classification

Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., & Leskovec, J. (2021). *Open Graph Benchmark: Datasets for Machine Learning on Graphs*. http://arxiv.org/abs/2005.00687



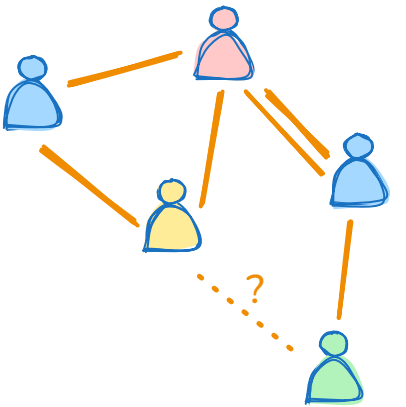
obgn-products

Which category is this amazon product?

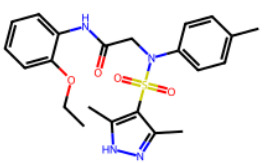
Edge Prediction

obgn-collab

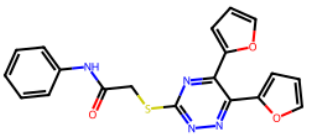
Will these researchers work together?



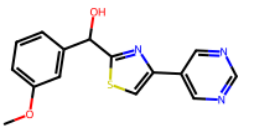
Graph Regression



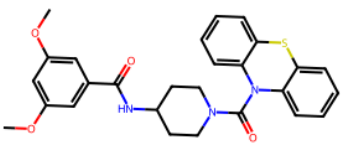
idx=257520



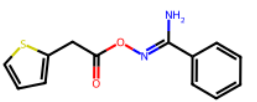
idx=280006



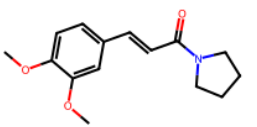
idx=27634



idx=176607



idx=337825

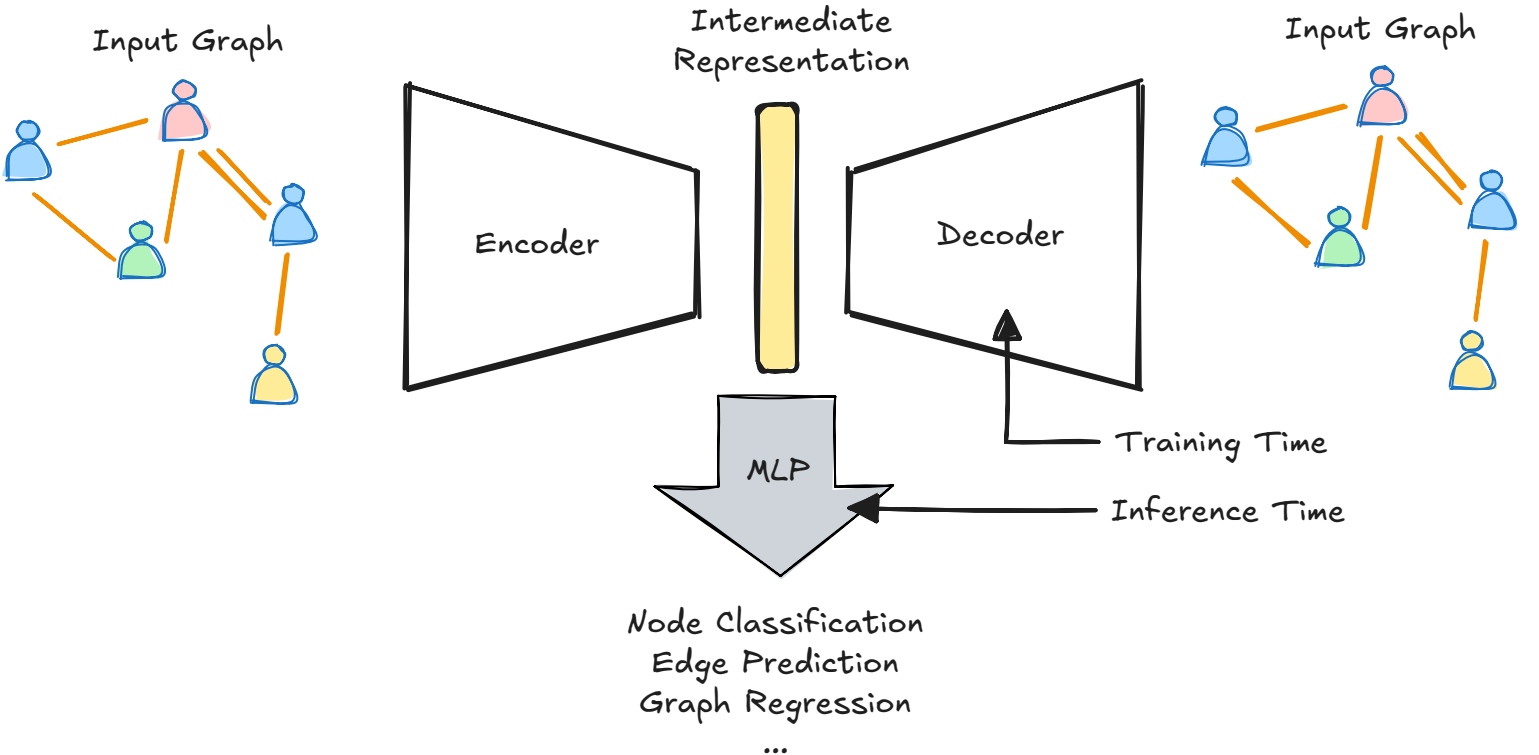


idx=321724

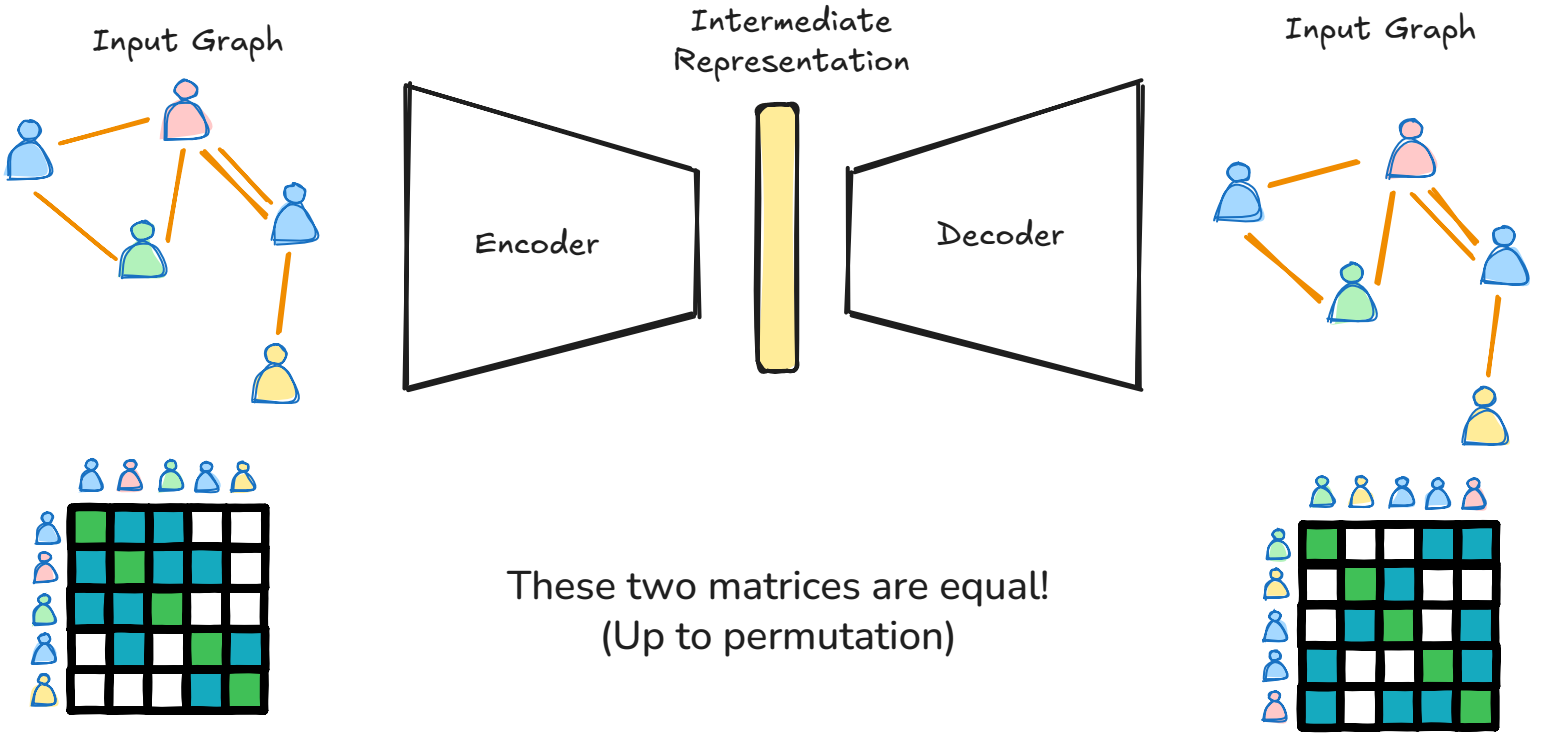
ogbg-molpcba

What is the homo-lumo gap of this molecule?

Graph Autoencoders

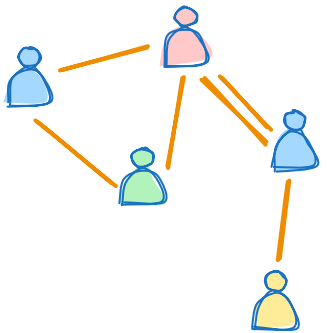


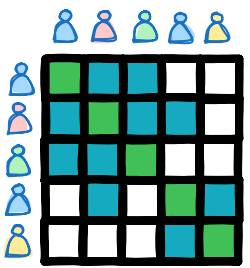
The problem


Input Graph
Intermediate Representation
Input Graph
Encoder
Decoder
These two matrices are equal!
(Up to permutation)

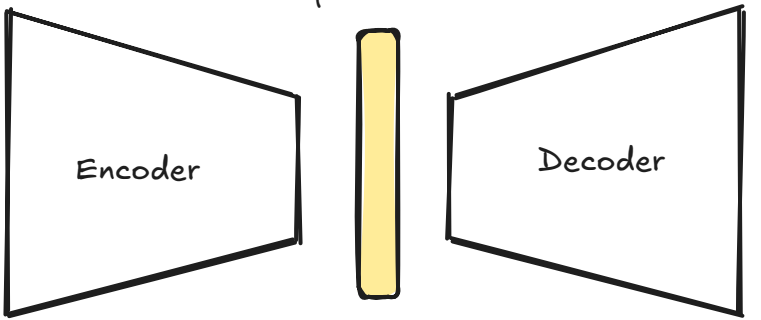
The problem

Input Graph

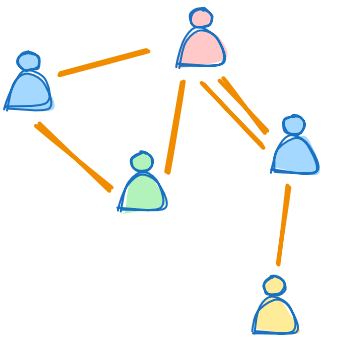


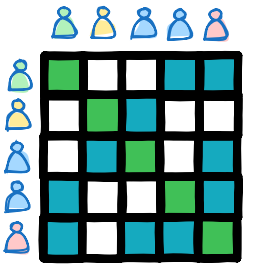


Intermediate Representation



Input Graph



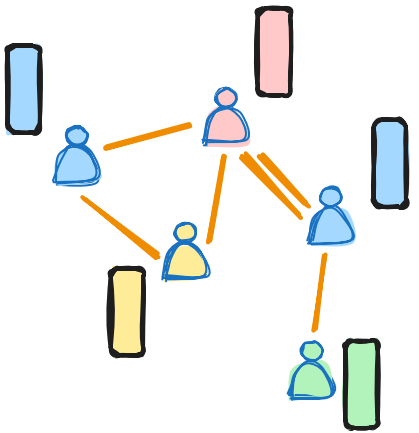


Our loss is:

$$\min_{\text{Permutation P}} \|A - P^T \tilde{A} P\|^2$$

(n! permutations, intractable)

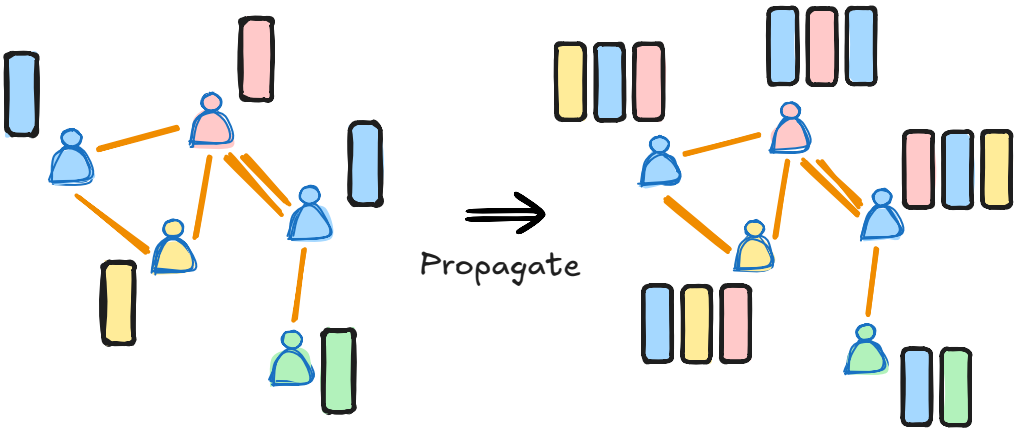
Message passing



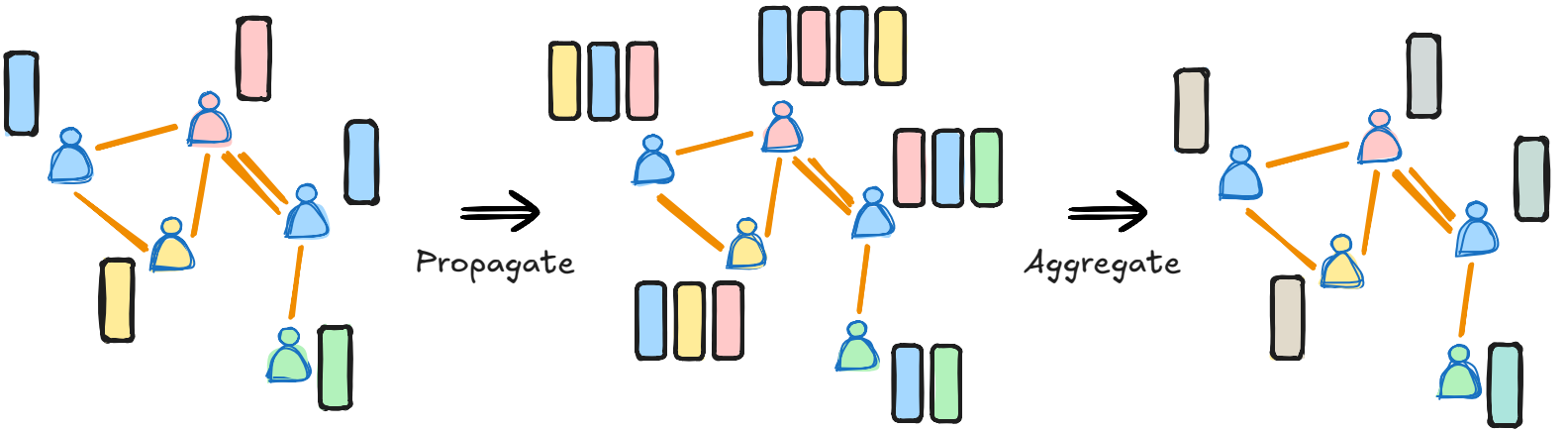
GCN: Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

GAT: Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

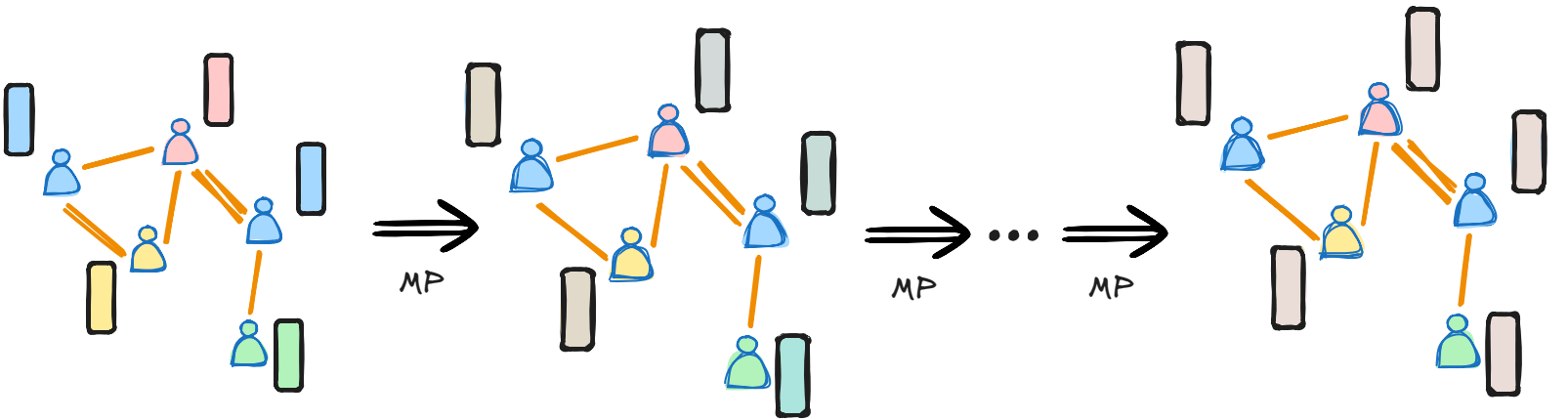
Message passing



Message passing

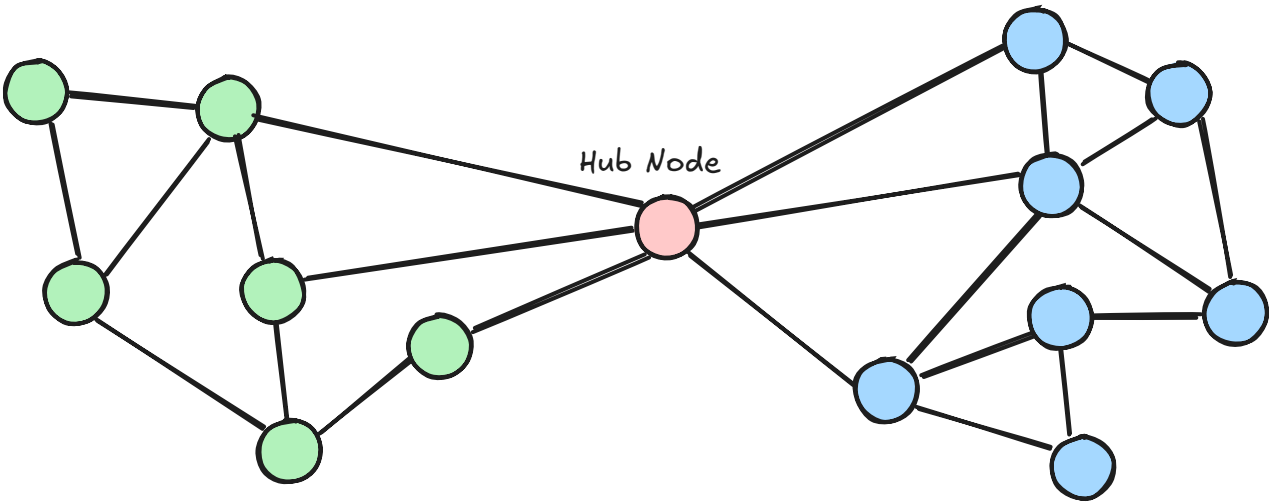


Problem 1: Over-smoothing

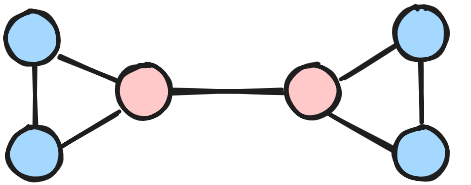


Message Passing acts as a low-pass filter!

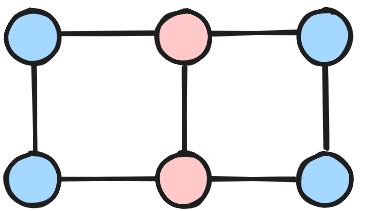
Problem 2: Over-squashing



Problem 3: Expressiveness

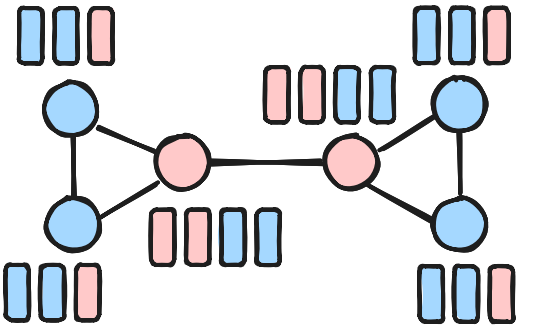


Message passing is as expressive as the Weisfeiler-Leman Graph Isomorphism Test!

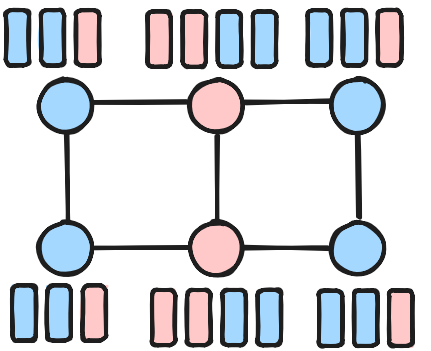


Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). *How Powerful are Graph Neural Networks?* https://arxiv.org/abs/1810.00826

Problem 3: Expressiveness

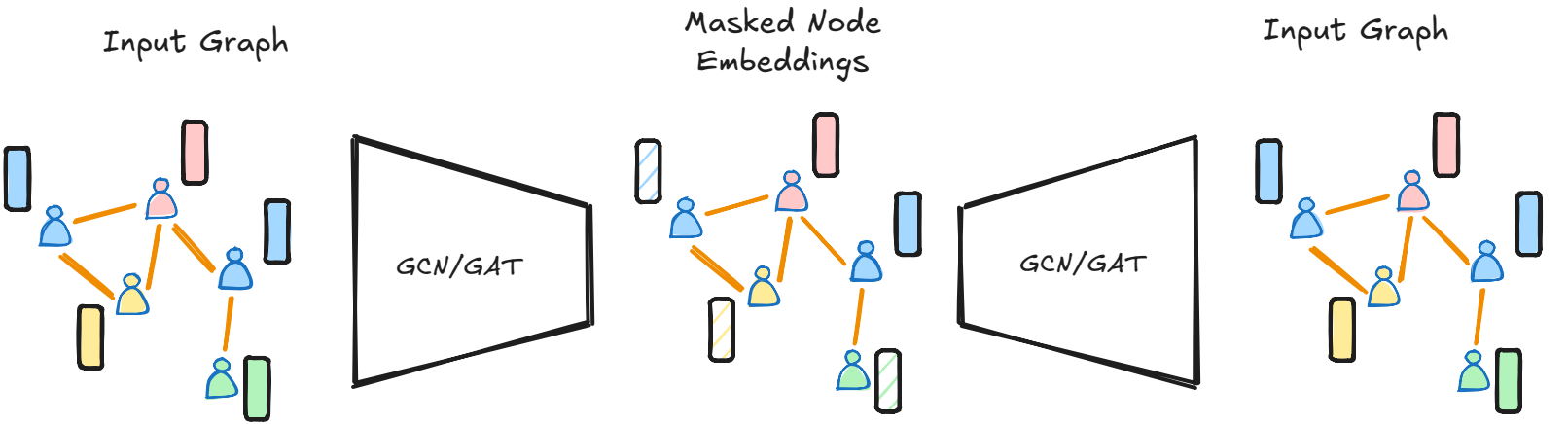


Message passing is as expressive as the Weisfeiler-Leman Graph Isomorphism Test!



Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). *How Powerful are Graph Neural Networks?* https://arxiv.org/abs/1810.00826

Graph Masked Autoencoders



Hou, Z., Liu, X., Cen, Y., Dong, Y., Yang, H., Wang, C., & Tang, J. (2022). GraphMAE: Self-Supervised Masked Graph Autoencoders. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 594–604. https://doi.org/10.1145/3534678.3539321

Hou, Z., He, Y., Cen, Y., Liu, X., Dong, Y., Kharlamov, E., & Tang, J. (2023). GraphMAE2: A Decoding-Enhanced Masked Self-Supervised Graph Learner. *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, 737–746. https://doi.org/10.1145/3543507.3583379

Graph Transformers

Graph Transformer	Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., & Liu, T.-Y. (2021). *Do Transformers Really Perform Bad for Graph Representation?* http://arxiv.org/abs/2106.05234
Graphormer	Dwivedi, V. P., & Bresson, X. (2021). *A Generalization of Transformer Networks to Graphs*. http://arxiv.org/abs/2012.09699
NAGphormer	Chen, J., Gao, K., Li, G., & He, K. (2023). *NAGphormer: A Tokenized Graph Transformer for Node Classification in Large Graphs*. http://arxiv.org/abs/2206.04910

There are many more...

Real graphs are huge

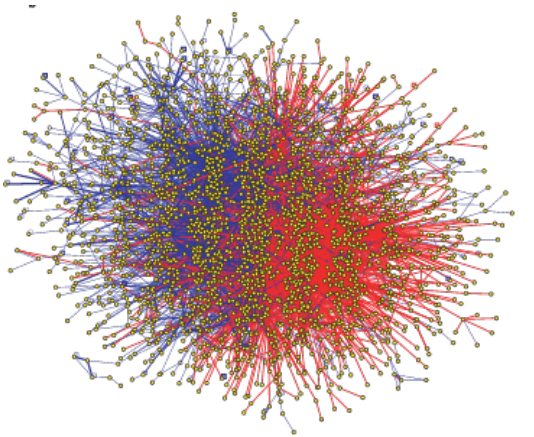
Document networks

Grandjean, Martin (2014). "La connaissance est un réseau". *Les Cahiers du Numérique* **10** (3): 37-54. DOI:10.3166/LCN.10.3.37-54.

Social Networks

Al-Molhem, N.R., Rahal, Y. & Dakkak, M. Social network analysis in Telecom data. *J Big Data* **6**, 99 (2019). https://doi.org/10.1186/s40537-019-0264-6

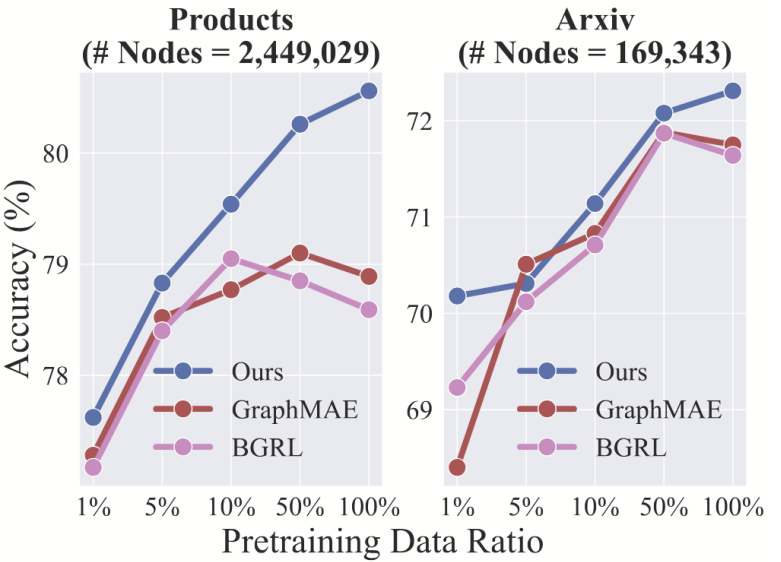
Protein Networks



Rual, JF., Venkatesan, K., Hao, T. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005). https://doi.org/10.1038/nature04209

Graph complexity scales in $O(n^2)$!

Current methods don't scale...



Transformer Masked Autoencoders Scale

GPT-1, iGPT, Jukebox, VideoGPT

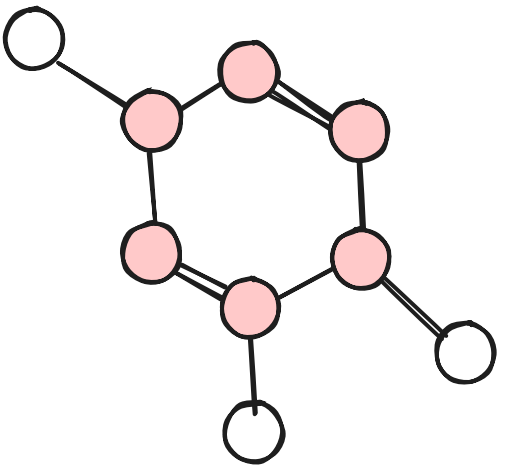
Radford, Alec, et al. “Improving Language Understanding by Generative Pre-Training.” https://openai.com/index/language-unsupervised/.

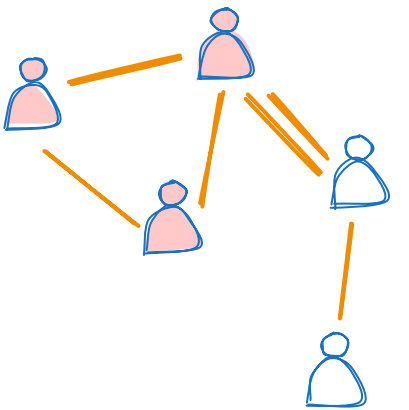
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929

Generative pretraining on high-level semantic tokens

What if we applied Generative Pre-training to Graphs?

Substructures as tokens





The bridge between NLP and Graph learning

DeepWalk: Online Learning of Social Representations

Bryan Perozzi
Stony Brook University
Department of Computer Science

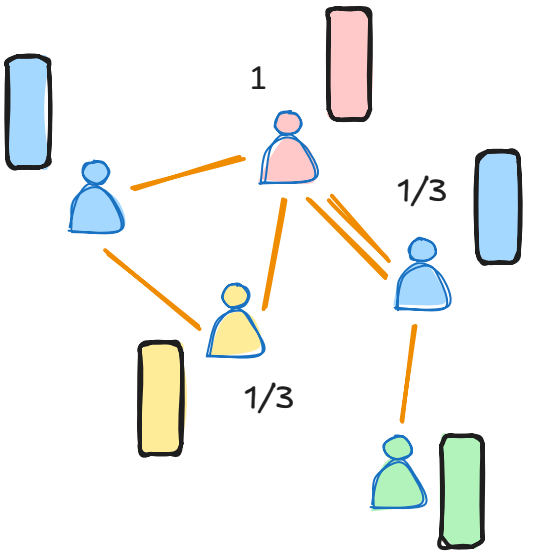
Rami Al-Rfou
Stony Brook University
Department of Computer Science

Steven Skiena
Stony Brook University
Department of Computer Science

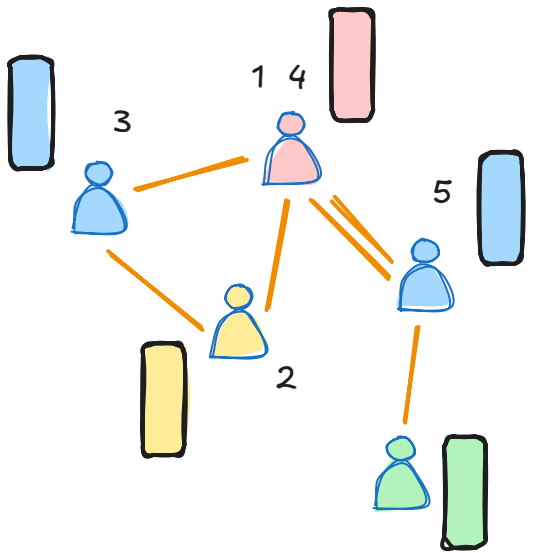
{bperozzi, ralrfou, skiena}@cs.stonybrook.edu

Random walks represent local substructure as a sentence!

Substructure Sampling

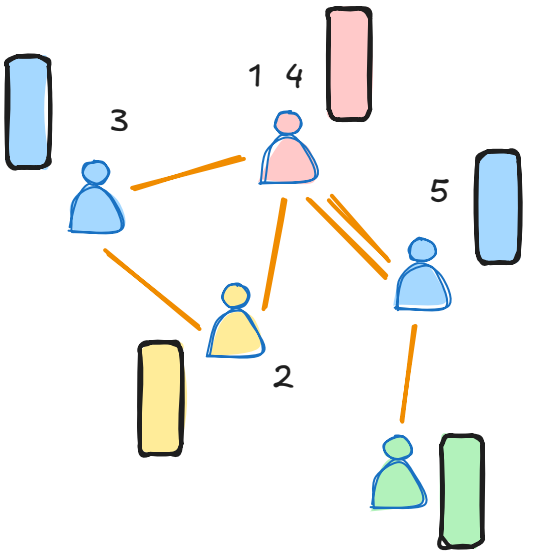


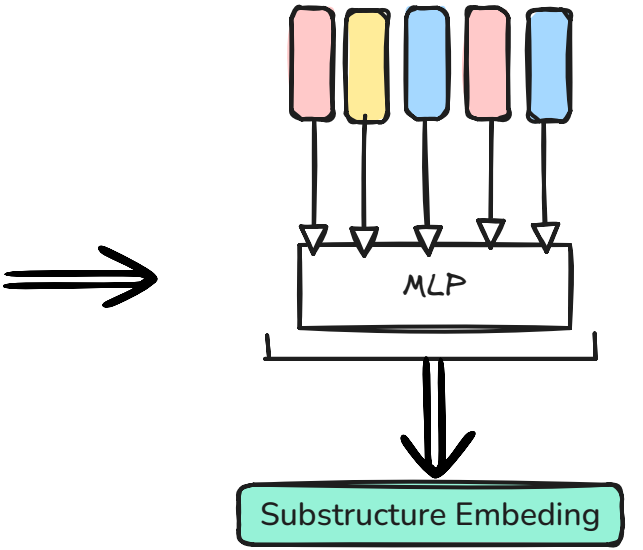
Substructure Sampling



(Path length is a parameter of the model)

Substructure Sampling

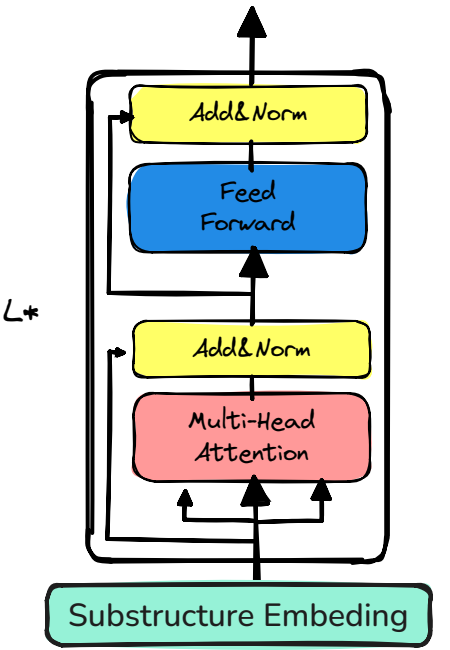




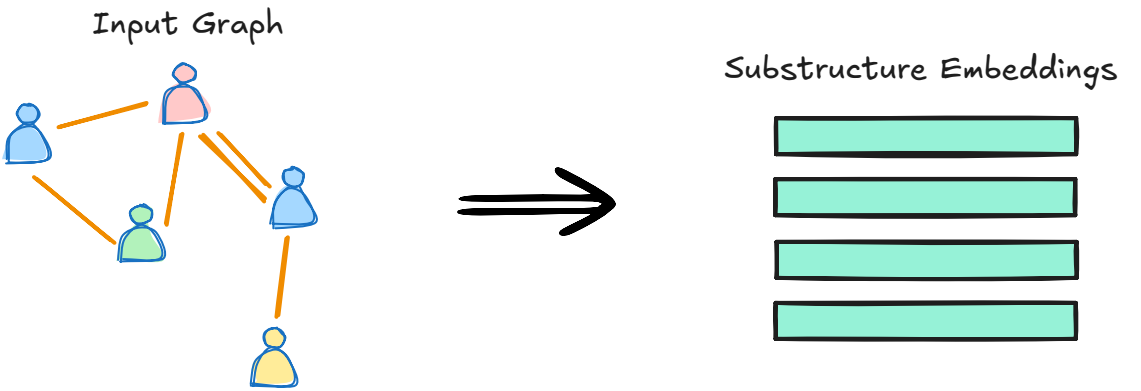
(No need to do attention over all nodes)

Encoder in detail

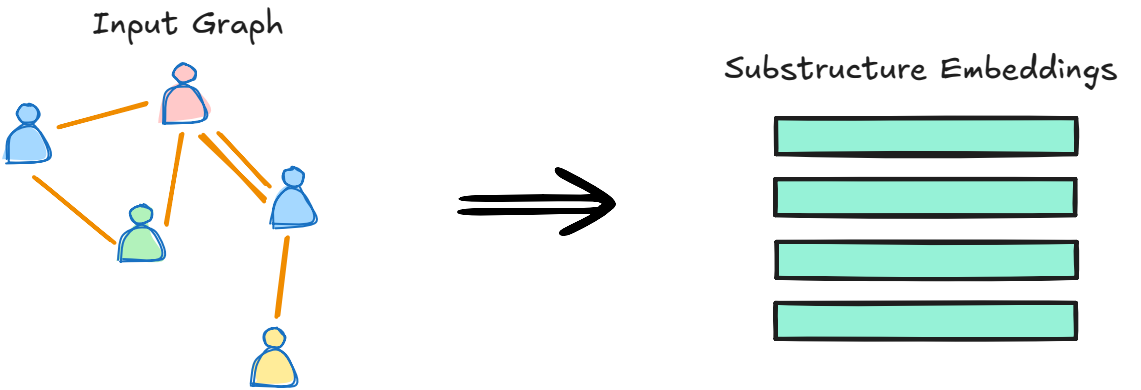
The encoder selects the substructures it cares about!



Pre-training



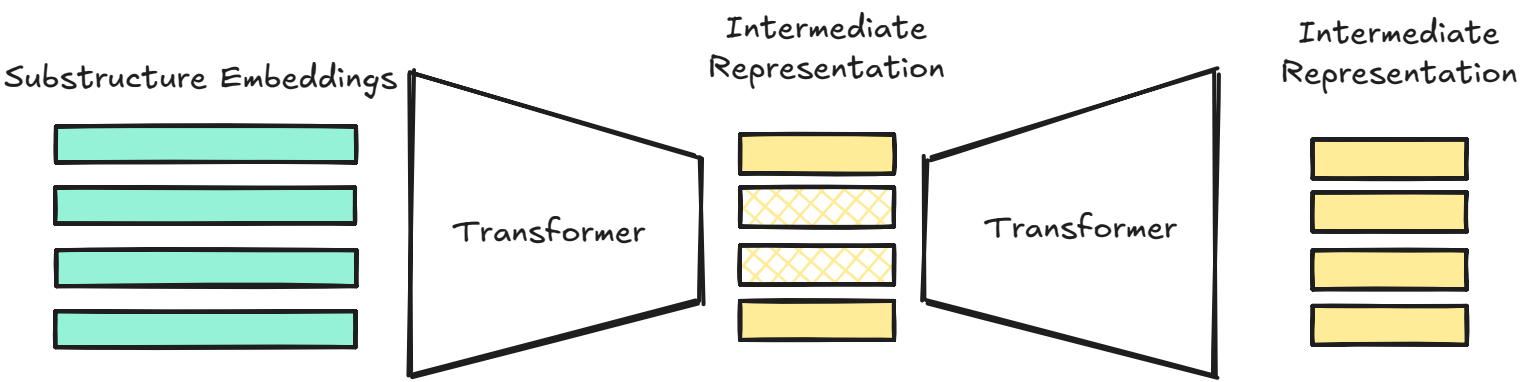
Pre-training



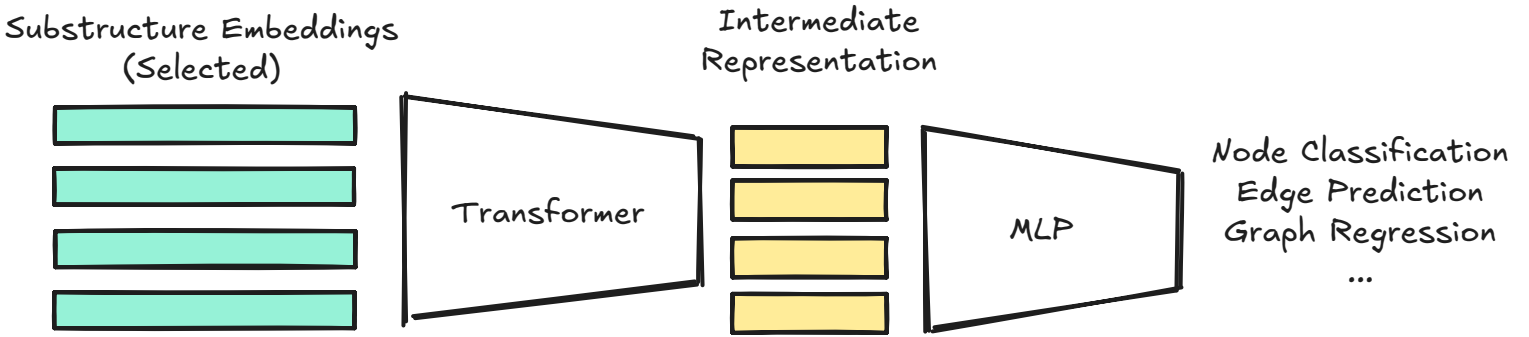
Pre-training

How do we obtain target embeddings?

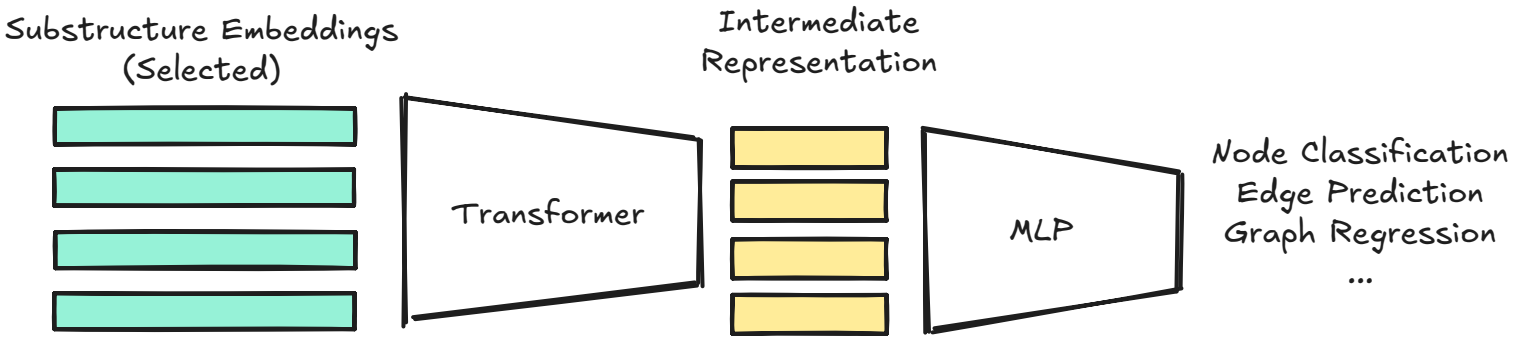
-> Use a EMA teacher encoder



Finetuning


Substructure Embeddings
(Selected)
Transformer
Intermediate
Representation
MLP
Node Classification
Edge Prediction
Graph Regression
...

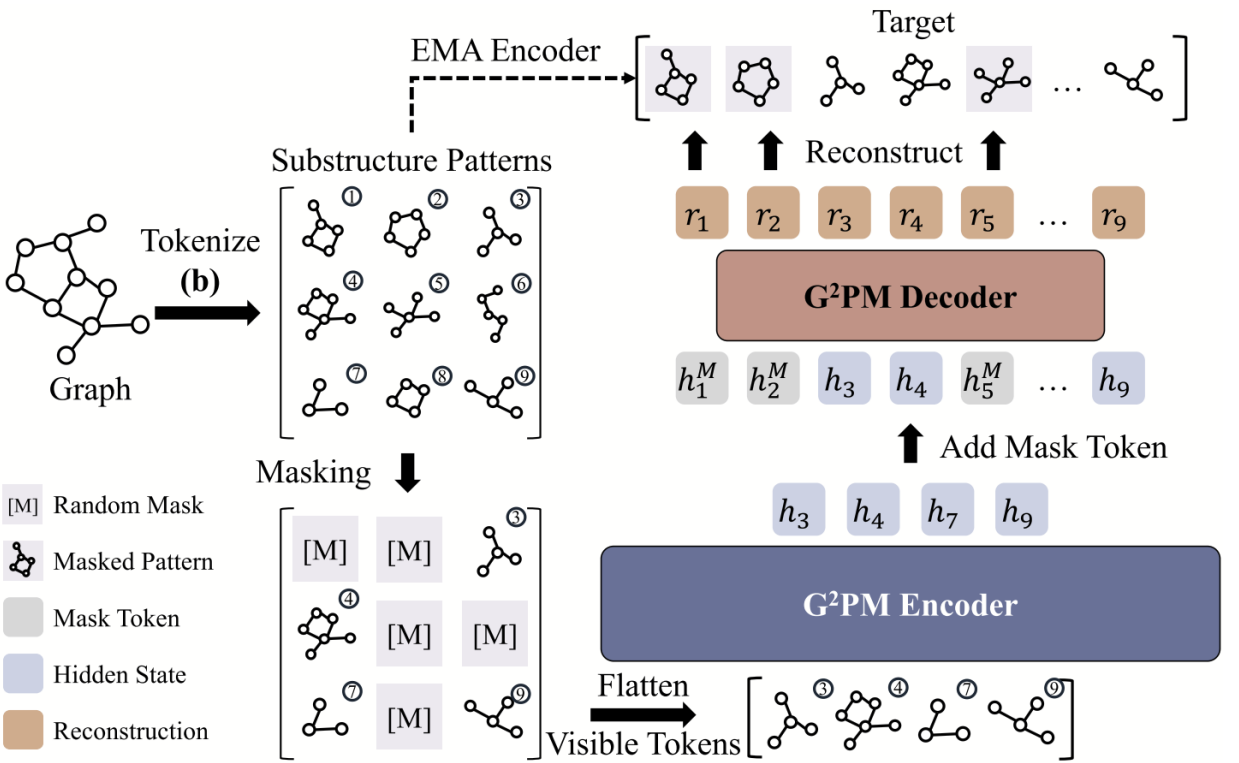
Finetuning



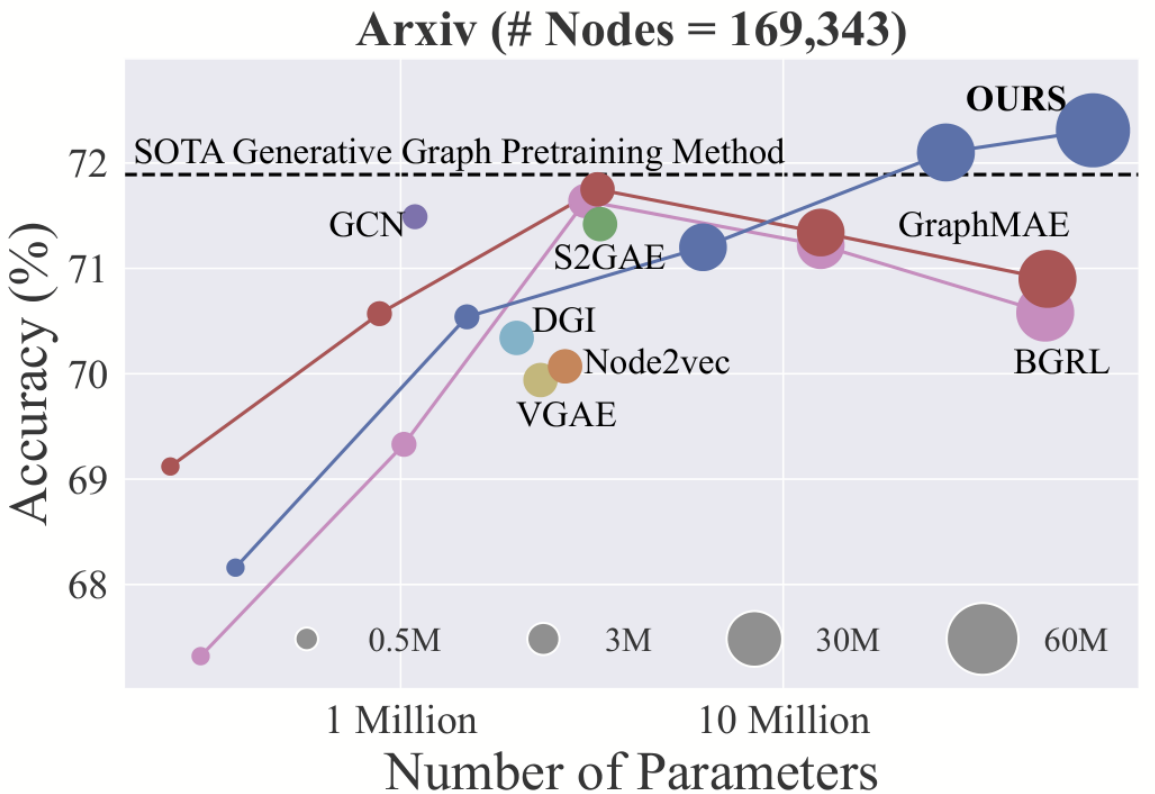
Node prediction	Random walks from the node
Link Prediction	Random walks from the nodes of the link
Graph Prediction	Random walks from any node from any graph

The full model

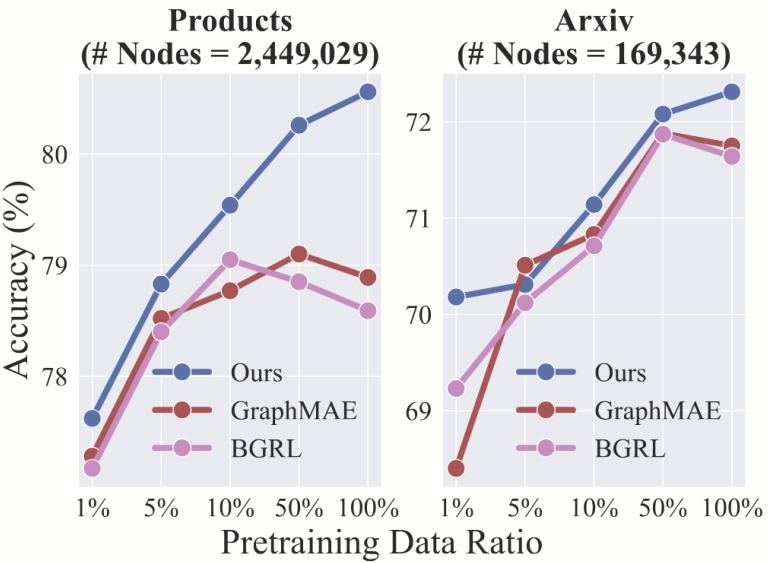
Wang, Z., Zhang, Z., Ma, T., Zhang, C., & Ye, Y. (2025). *Generative Graph Pattern Machine*. http://arxiv.org/abs/2505.16130



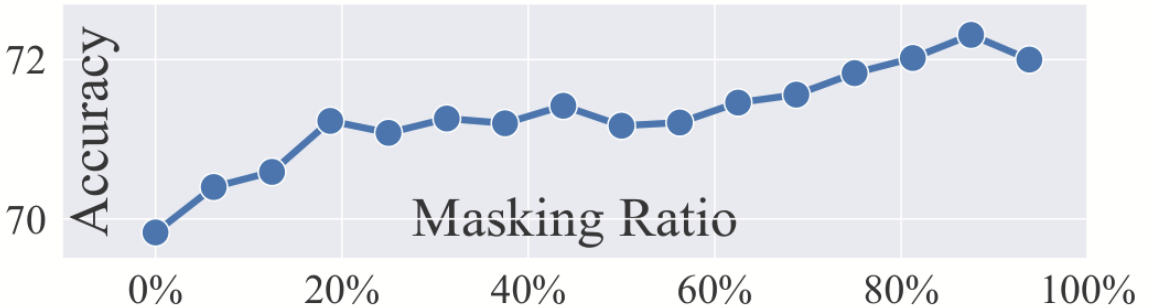
Results


Arxiv (# Nodes = 169,343)
OURS
SOTA Generative Graph Pretraining Method
72
71
70
69
68
Accuracy (%)
GCN
S2GAE
GraphMAE
DGI
Node2vec
BGRL
VGAE
0.5M
3M
30M
60M
1 Million
10 Million
Number of Parameters

Results


Products
(# Nodes = 2,449,029)
Arxiv
(# Nodes = 169,343)
Accuracy (%)
80
79
78
72
71
70
69
Ours
GraphMAE
BGRL
1%
5%
10%
50%
100%
Pretraining Data Ratio

Results



α	**Update Every**	**Arxiv**	**HIV**
0.9	10	70.8	74.3
0.99	10	**72.3**	**78.7**
0.999	10	72.1	76.6
0.99	5	72.2	78.1
0.99	20	**72.3**	76.6
w/o EMA Update		25.2	69.1

Results

Source	**Arxiv**		**HIV**	
Target	**Products**	**HIV**	**Arxiv**	**PCBA**
GNN [60, 73]	78.3 (1.2 ↓)	70.1 (5.7 ↓)	71.1 (1.0 ↓)	71.9 (**1.6** ↑)
GPM [68]	82.0 (0.6 ↓)	74.3 (2.7 ↓)	71.4 (1.5 ↓)	76.4 (**1.3** ↑)
BGRL [53]	78.8 (**0.2** ↑)	72.5 (3.8 ↓)	68.6 (1.9 ↓)	72.9 (0.6 ↓)
GraphMAE [22]	77.5 (1.4 ↓)	74.7 (3.1 ↓)	69.9 (1.9 ↓)	73.4 (**0.2** ↑)
G^2PM	**81.3** (**0.7** ↑)	**76.8** (1.9 ↓)	**72.6** (**0.3** ↑)	**77.9** (**2.3** ↑)

My grain of salt...

Pro:
- Simplicity
- Random walks + attention is complimentary
- Transfer learning

Cons:
- Paper is a bit handwavy
- Some architectural limitations

Q & A