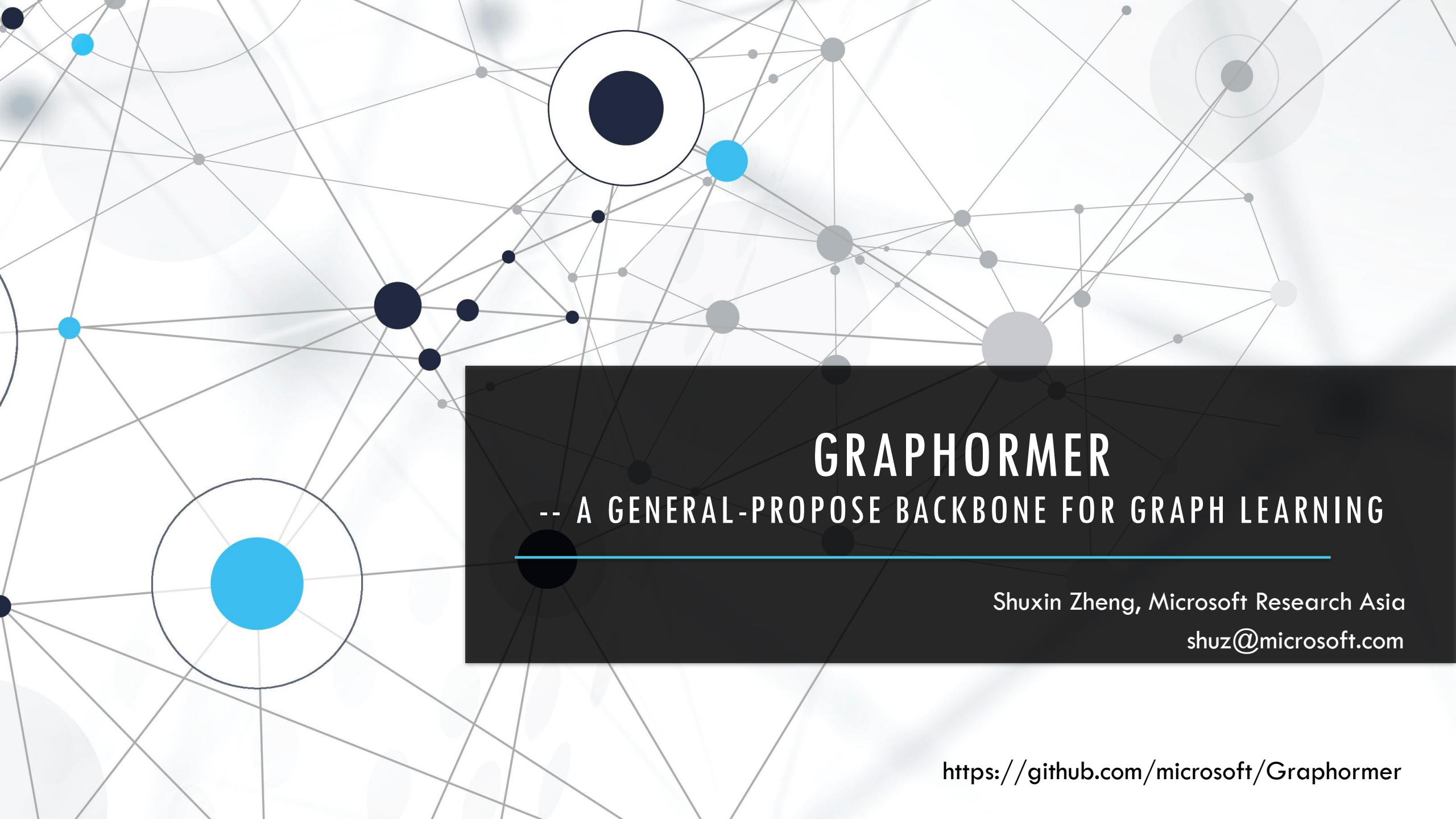


GRAPHORMER

-- A GENERAL-PROPOSE BACKBONE FOR GRAPH LEARNING

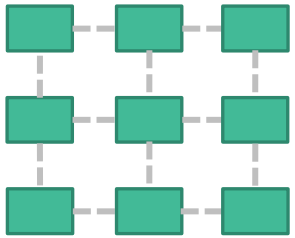
Shuxin Zheng, Microsoft Research Asia
shuz@microsoft.com

https://github.com/microsoft/Graphormer

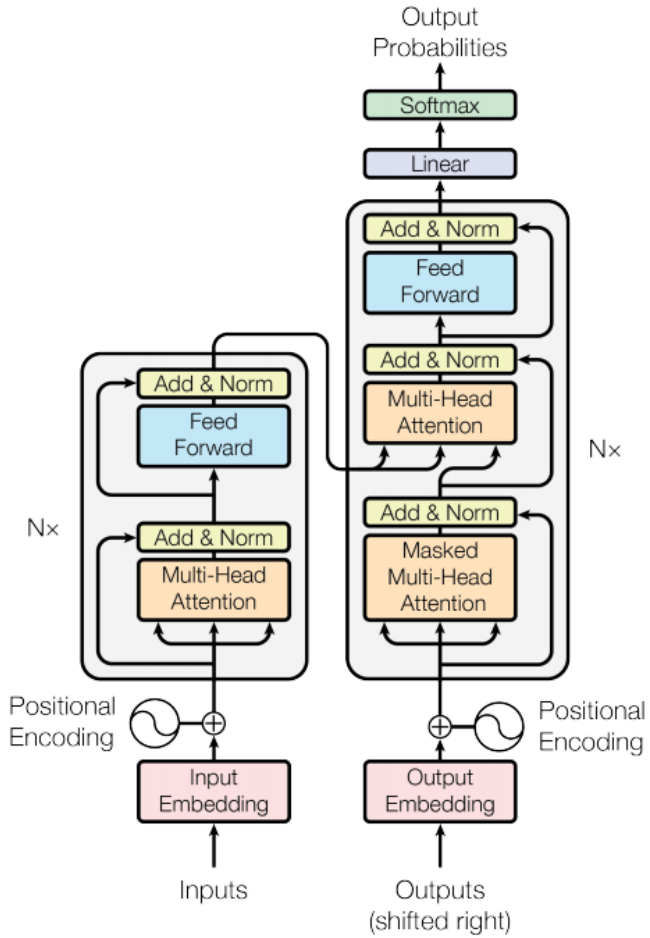
TRANSFORMER BECOMES DOMINANT ON SEQUENCE DATA

Sequence Data (1D): 

- Speech
- Language
- Protein

Grid Data (2D) : 

- Image
- Video

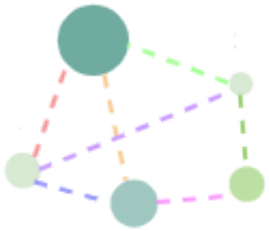


Today's Transformer Model:

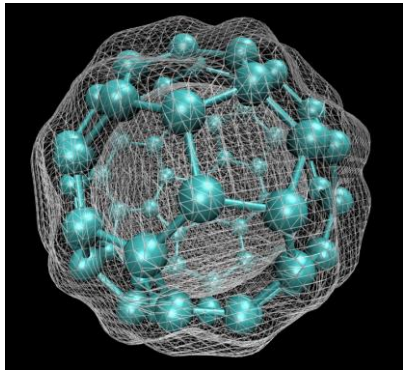
Become Bigger, Deeper, Wider

GNN IS STILL THE FIRST CHOICE FOR GRAPH DATA

Graph Data:

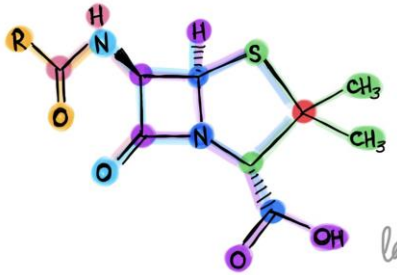


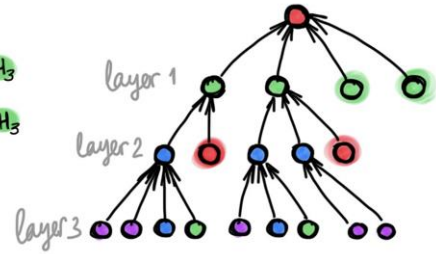
- **Molecule**
- Social Network
- **Combinatorial Optimization**



Today's Graph Neural Network:

Keep Slim, Shallow,
and Simple Operations





GIN: 3-5 layers
Operations: Sum + 2-layer FFN

EXPRESSIVENESS VS. CAPABILITY OF MODELING GRAPH

Attempts:

1. Graft Existing Modules to GNN

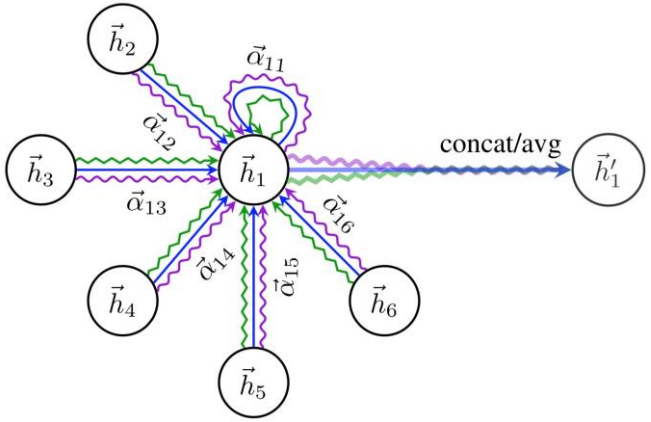
   Graph Attention Network, etc.

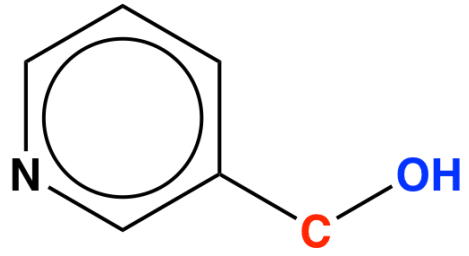
2. Transform Graph to Sequence

Flow Graph to Text, Molecule to SMILES, etc.

3. Modify Transformer by Heuristic on Graph

   Still not appear on leaderboards.

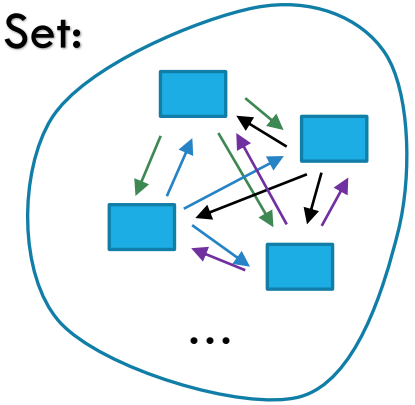




SMILES: C(c1cnccc1)O

TRANSFORMER ON GRAPH

Set:

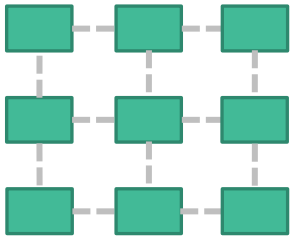


Self-Attention: Calculate Correlation Between Tokens/Patches…

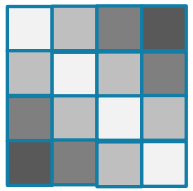
$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) V$$

Sequence Data:

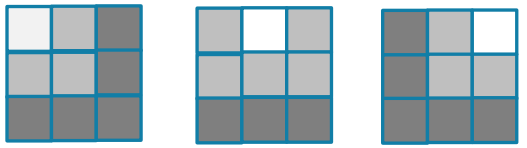
Grid Data:



Relative Positional Encoding[1]:

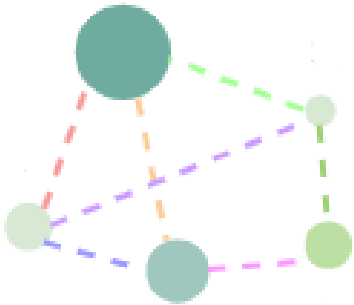


2D Relative Positional Encoding:

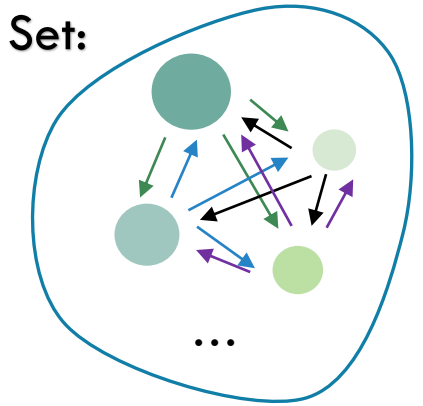


[1] Ke, Guolin, Di He, and Tie-Yan Liu. "Rethinking the Positional Encoding in Language Pre-training." *ICLR*(2021)

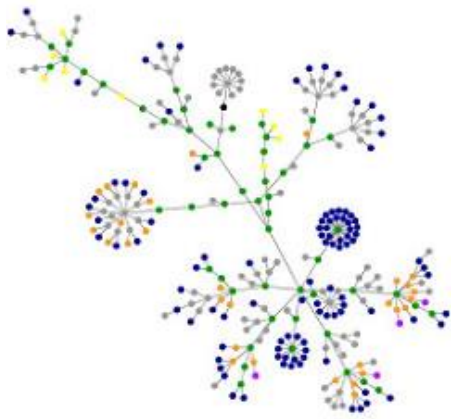
KEY INSIGHT: STRUCTURAL ENCODINGS



Self-Attention: Calculate **Correlation** Between Nodes…

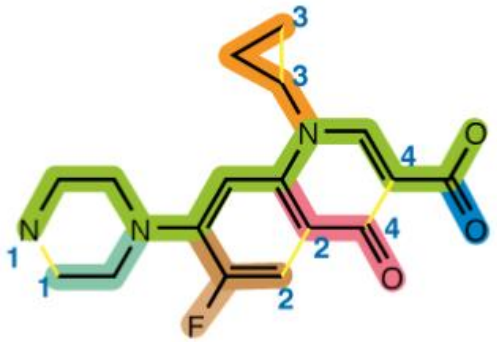


What affects the Correlation between Nodes:



Spatial Position

Centrality



Edge Feature

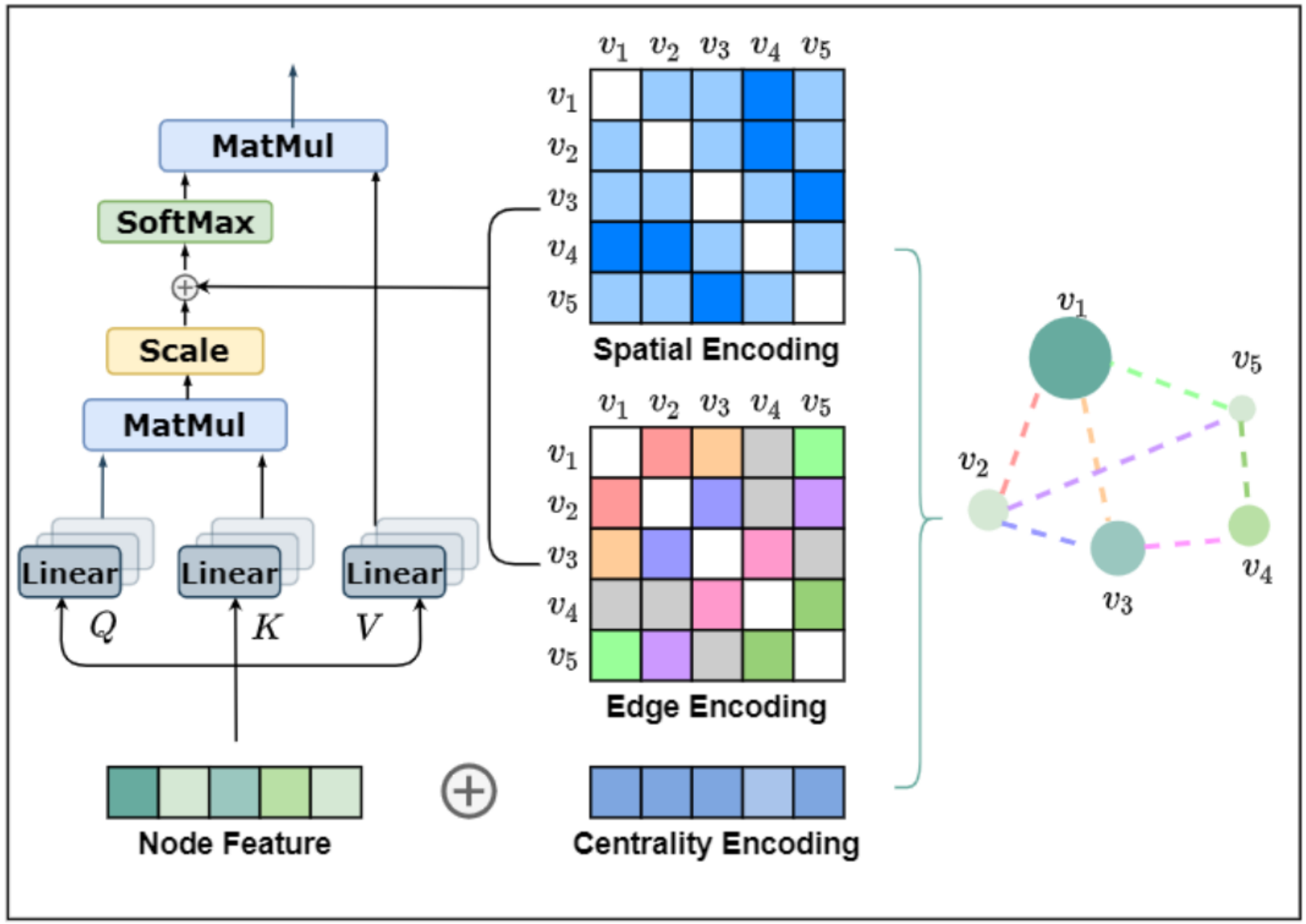
GRAPHORMER

= Pure Transformer

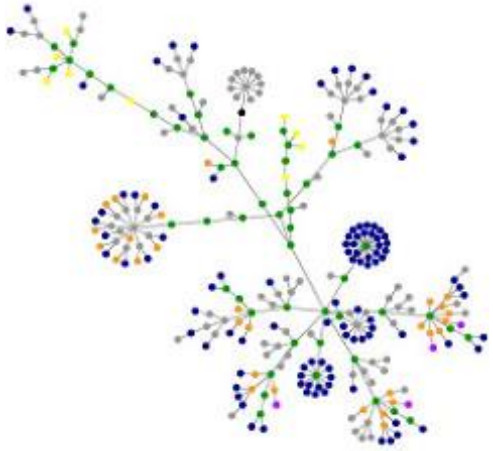
+ Spatial Encoding

+ Centrality Encoding

+ Edge Encoding



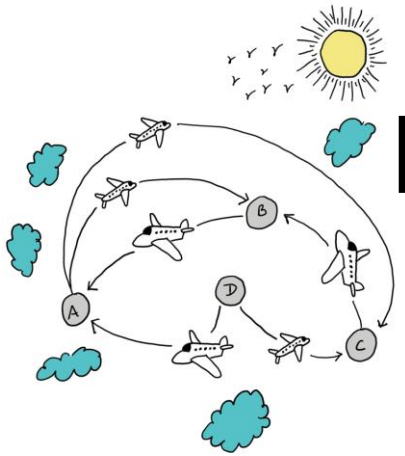
GRAPHORMER: SPATIAL ENCODING



$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}} + b_{\phi(v_i,v_j)}\right) V$$

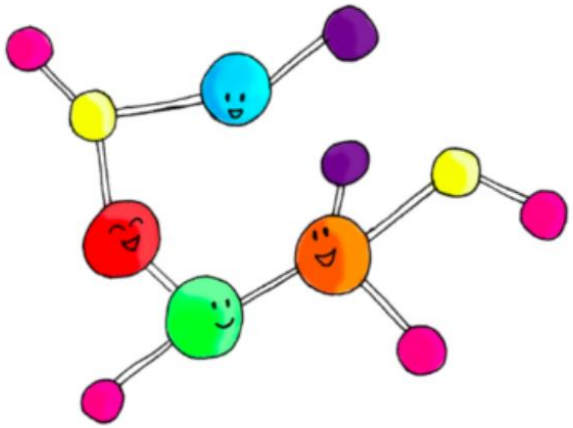
Spatial Position

$\phi(v_i, v_j)$: Any Metric that Measures the Distance Between vi & vj.



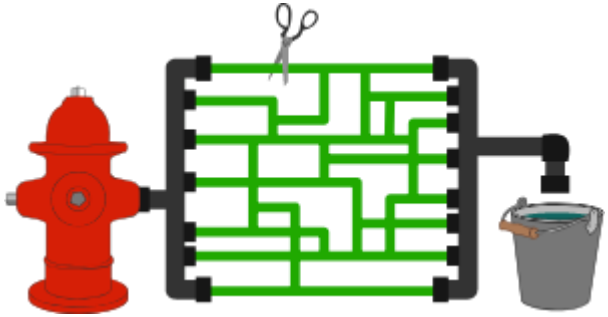
☑ Unweighted Shortest Path

Weighted Shortest Path



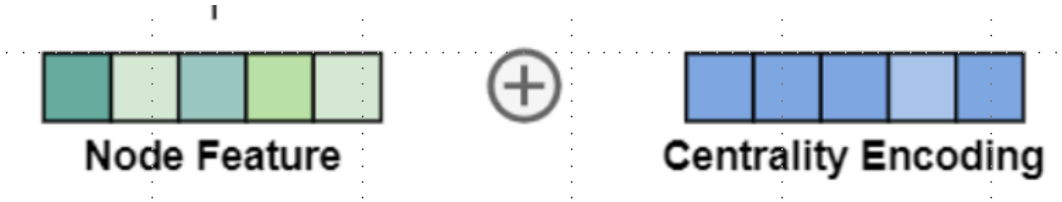
3D Euclidean Distance

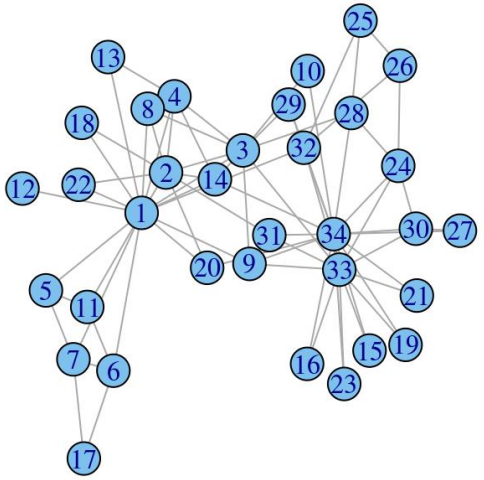
Max Flow

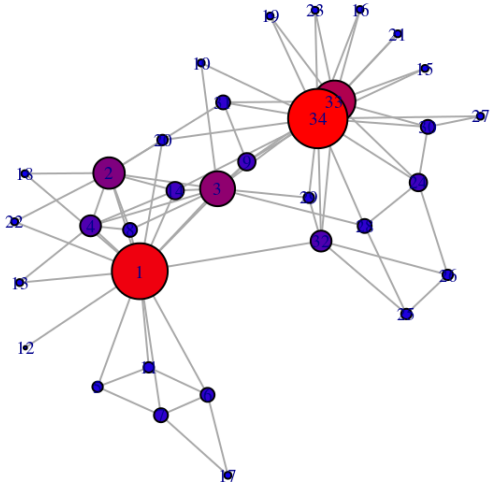


GRAPHORMER: CENTRALITY & EDGE ENCODINGS

Node Centrality: **Degree** ☑

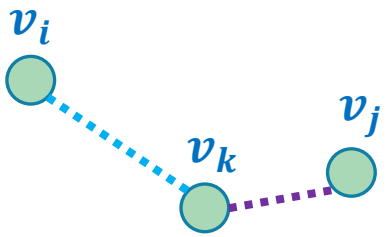






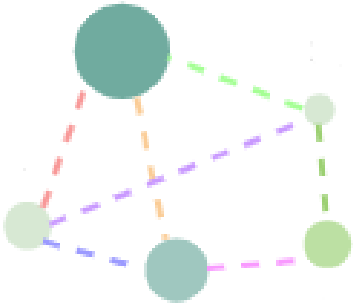
Edge Encoding:

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}} + b_{\phi(v_i,v_j)} + c_{ij}\right) V$$

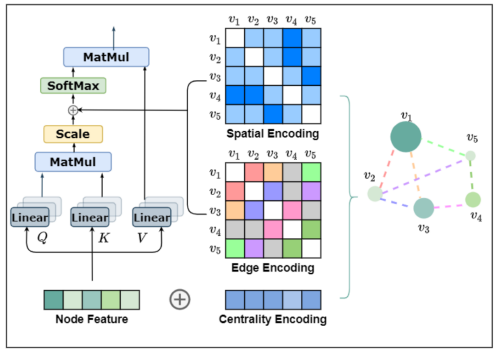


$$c_{ij} = \frac{1}{2}(e_{ik}w_1^T + e_{kj}w_2^T)$$

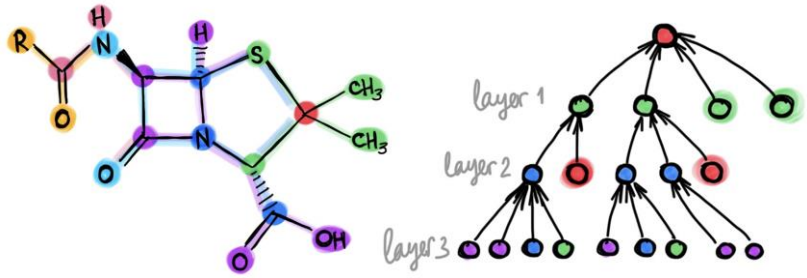
HOW POWERFUL IS GRAPHORMER?



Why Graphormer is Better? Theoretical Facts:



Special Cases



Graphormer

GIN, GCN, GraphSage …

Example: Mean Aggregation

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}} + b_{\phi(v_i,v_j)} + c_{ij}\right) V$$

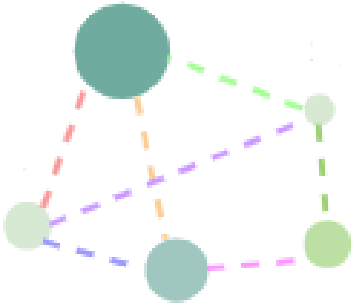
Let:

$W_Q = W_K = 0, W_v = I,$

$b_{\phi(v_i,v_j)} = 0$, if v_i and v_j are neighbor, else $b_{\phi(v_i,v_j)} = -\infty$,

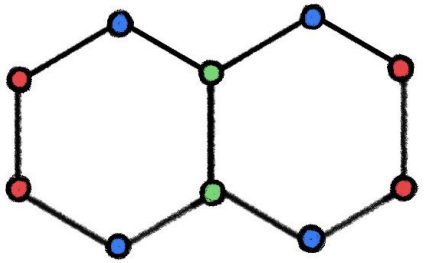
$c_{ij} = 0.$

HOW POWERFUL IS GRAPHORMER?

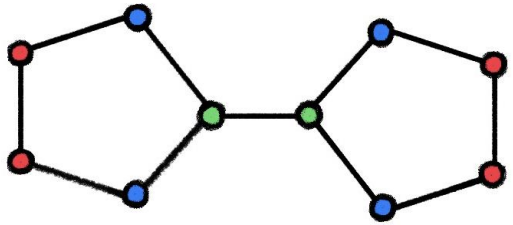


Why Graphormer is Better? Theoretical Facts:

Expressiveness: Graphormer > 1-WL Test ≥ Graph Neural Network



(a)



(b)

Spatial Encoding

KDD CUP 2021 – 1ST PLACE AWARD

Awardees of PCQM4M-LSC Track (Leaderboard)

Winners

1st place: MachineLearning (contact)

- **Team members**: ChengxuanYing (Dalian University of Technology), Mingqi Yang (Dalian University of Technology), Shengjie Luo (Peking University), Tianle Cai (Princeton University), Guolin Ke (MSRA), Di He (MSRA), Shuxin Zheng (MSRA), Chenglin Wu (Xiamen University), Yuxin Wang (Dalian University of Technology), Yanming Shen (Dalian University of Technology)
- **Method**: Graphormer (10 ensemble) + ExpC (8 ensemble)
- **Short summary**: We adopt Graphormer and ExpC as our basic models. We train each model by 8-fold cross-validation, and additionally train two Graphormer models on the union of training and validation sets with different random seeds. For final submission, we use a naive ensemble for these 18 models by taking average of their outputs.
- **Learn more:** Technical report, code
- **Test MAE**: 0.1200

2nd place: SuperHelix (contact)

- **Team members**: Zhang Shanzhuo (Baidu), Liu Lihang (Baidu), Gao Sheng (Baidu), He Donglong (Baidu), Li Weibin (Baidu), Huang Zhengjie (Baidu), Su Weiyue (Baidu), Wang Wenjin (Baidu)
- **Method**: LiteGEM
- **Short summary**: Deep graph neural network with self-supervised tasks on topology and geometry information. 73 models with different tasks and hyper-parameters are ensembled.
- **Learn more:** Technical report, code
- **Test MAE**: 0.1204

3rd place: Quantum (contact)

- **Team members**: Petar Velickovic (DeepMind), Peter Battaglia (DeepMind), Jonathan Godwin (DeepMind), Alvaro Sanchez (DeepMind), David Budden (DeepMind), Shantanu Thakoor (DeepMind), Jacklynn Stott (DeepMind), Ravichandra Addanki (DeepMind), Sibon Li (DeepMind), Andreea Deac (DeepMind)
- **Method**: Very Deep GN Ensemble + Conformers + Noisy Nodes
- **Short summary**: A combination of a 32-layer deep Graph Network over RDKit conformer features, and 50-layer deep Graph Network for molecules for which conformers cannot be computed. Denoising regularisation with Noisy Nodes was applied. 20 models with different initialisation and validation splits are ensembled.
- **Learn more:** Technical report, code
- **Test MAE**: 0.1205

1st Place: MSRA

BIOASSAY

OGB

Get Started | Updates | Large-Scale Challenge | Datasets | Leaderboards | Paper | Team | Github

Leaderboard for ogbg-molpcba

The Average Precision (AP) score on the test and validation sets. The higher, the better.

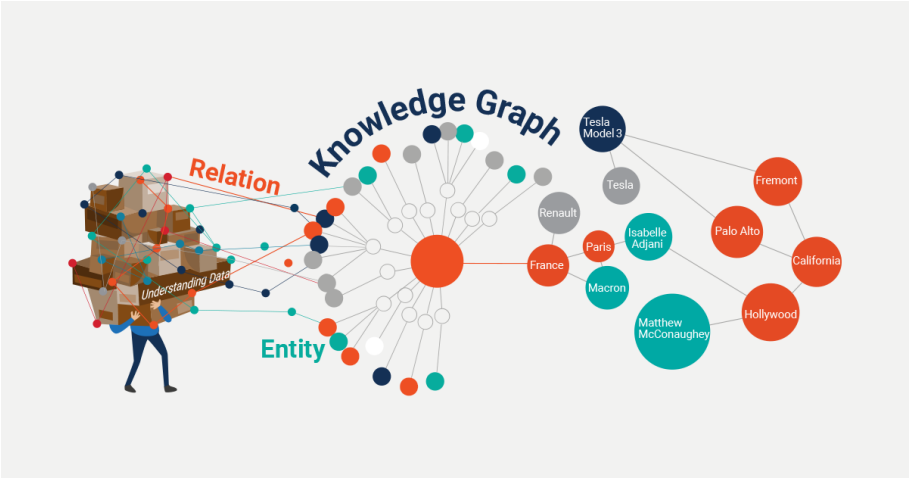
Note: The evaluation metric has been changed from PRC-AUC (Aug 11, 2020).

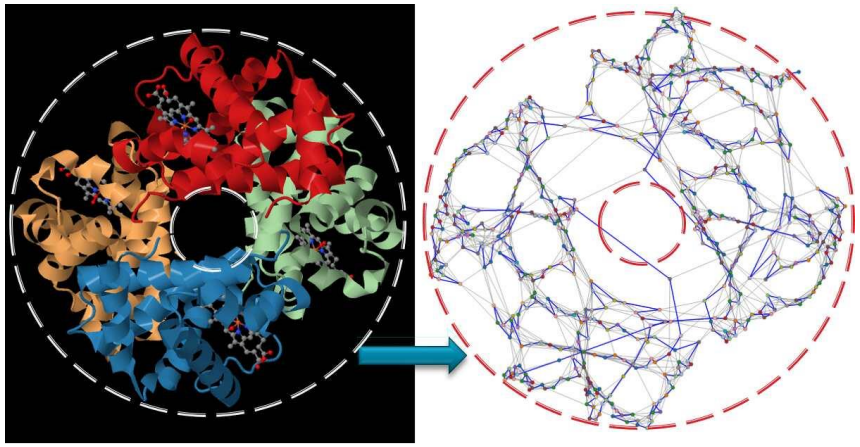
Package: >=1.2.2

Rank	Method	Test AP	Validation AP	Contact	References	#Params	Hardware	Date
1	**Graphormer (pre-trained on PCQM4M)**	0.3140 ± 0.0032	0.3227 ± 0.0024	Shuxin Zheng (Microsoft)	Paper, Code	119,529,664	NVIDIA Tesla V100 (16GB GPU)	Aug 2, 2021
2	GINE+bot	0.2994 ± 0.0019	0.3094 ± 0.0023	Hao Zhang	Paper, Code	5,511,680	Tesla V100(32GB GPU)	Jul 21, 2021
3	GINE+ w/ APPNP	0.2979 ± 0.0030	0.3126 ± 0.0023	Weibin Li (PaddleHelix & PGL)	Paper, Code	6,147,029	Tesla V100 (32GB)	Mar 15, 2021
4	**PHC-GNN**	0.2947 ± 0.0026	0.3068 ± 0.0025	Tuan Le	Paper, Code	1,690,328	Tesla V100 (32GB)	Apr 14, 2021
5	**GINE+ w/ virtual nodes**	0.2917 ± 0.0015	0.3065 ± 0.0030	Rémy Brossard	Paper, Code	6,147,029	GeForce GTX 1080 Ti	Dec 1, 2020
6	**DGN**	0.2885 ± 0.0030	0.2970 ± 0.0021	Dominique Beaini	Paper, Code	6,732,696	NVIDIA T4 GPU (16 GB)	Mar 4, 2021
7	**RandomGIN-vn+FLAG**	0.2881 ± 0.0028	0.3035 ± 0.0047	Giulia Fracastoro (Polito)	Paper, Code	5,572,026	Titan Xp (12GB GPU)	Jul 29, 2021
8	**DeeperGCN+virtual node+FLAG**	0.2842 ± 0.0043	0.2952 ± 0.0029	Kezhi Kong	Paper, Code	5,550,208	NVIDIA Tesla V100 (32GB GPU)	Oct 21, 2020
9	**PNA**	0.2838 ± 0.0035	0.2926 ± 0.0026	Dominique Beaini	Paper, Code	6,550,839	NVIDIA T4 GPU (16 GB)	Mar 4, 2021
10	**GIN+virtual node+FLAG**	0.2834 ± 0.0038	0.2912 ± 0.0026	Kezhi Kong	Paper, Code	3,374,533	GeForce RTX 2080 Ti (11GB GPU)	Oct 21, 2020

https://github.com/microsoft/Graphormer

FUTURE APPLICATIONS… AND THANKS!





Protein 3D structure

Protein graph