浙江大学
ZHEJIANG UNIVERSITY

图基础模型

Graph Foundation Model

周晟

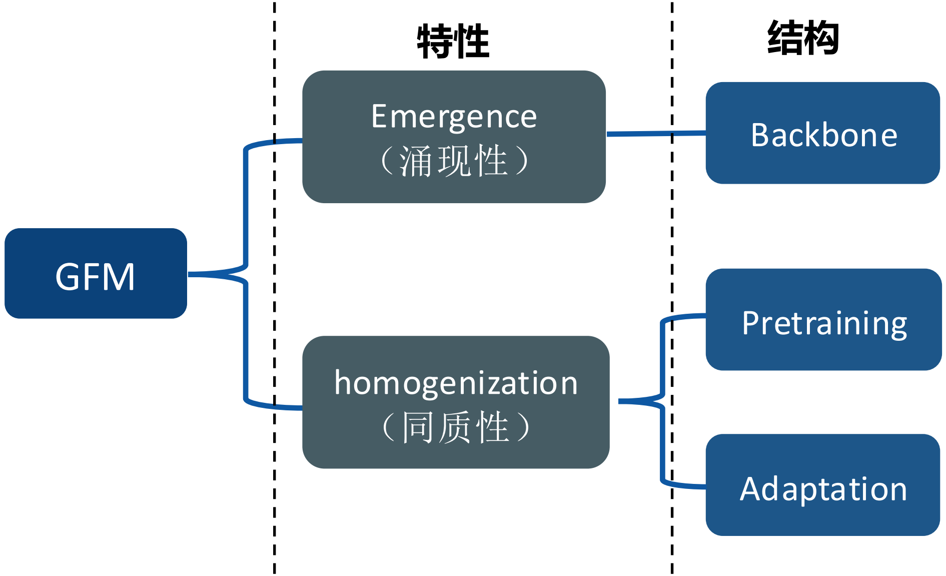
2025年12月30日

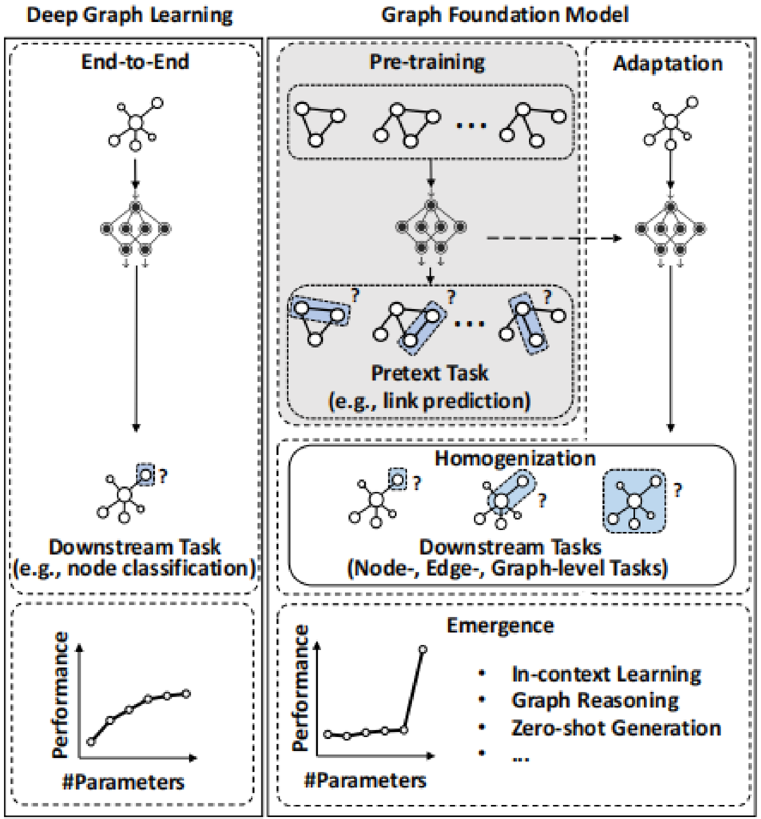
目录

定义

基础模型（Foundation Model）设计

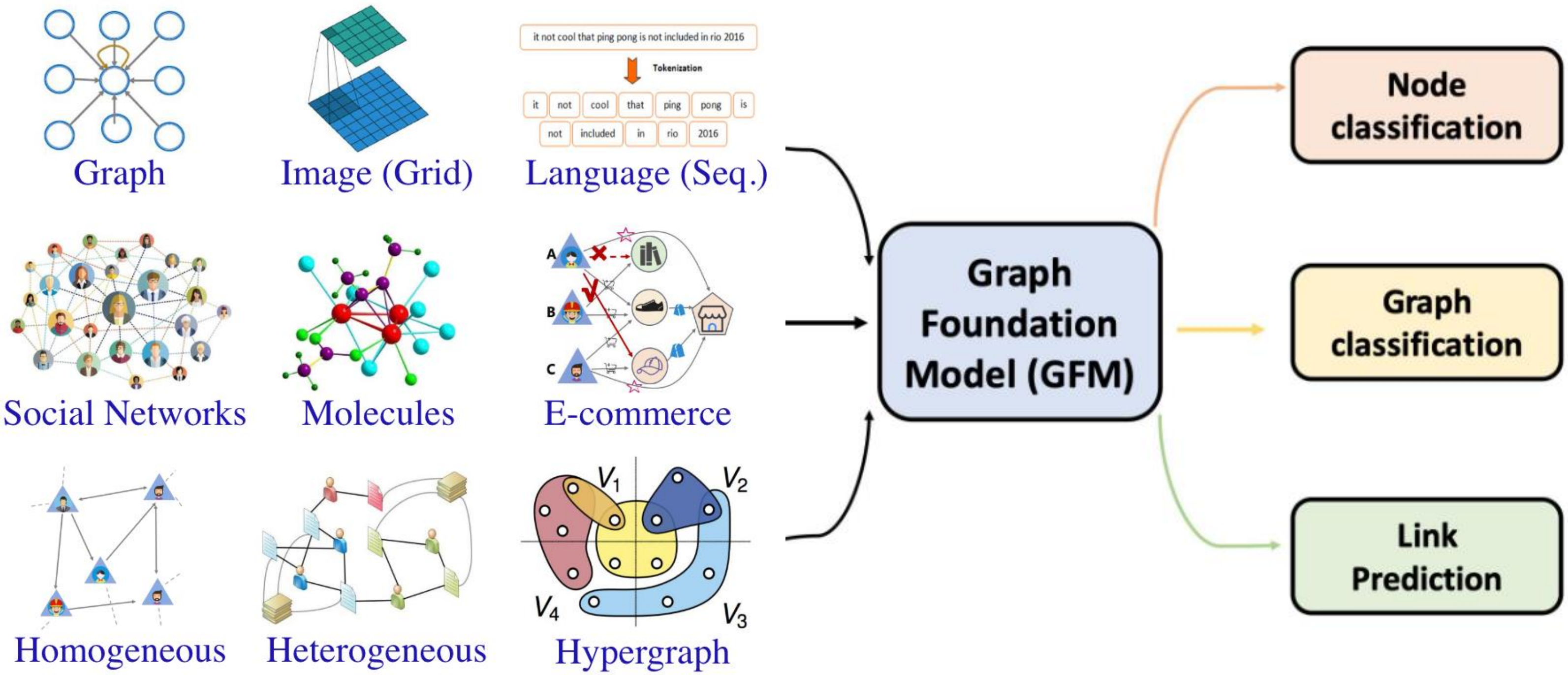
目标：在广泛的数据上进行训练，并能够适应各种下游任务。





[1] Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, Chuan Shi, “Towards Graph Foundation Models: A Survey and Beyond” Preprint 2023.

定义



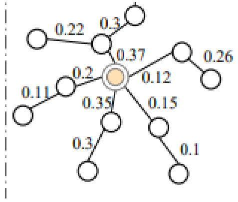
Philip S. Yu, Chuan Shi, Cheng Yang, Yuan Fang, Lichao Sun, “Towards Graph Foundation Models” WWW 2024 Tutorial.

挑战

挑战一：图的类型/任务过多

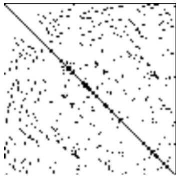
同构图

异构图

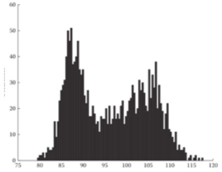


动态图

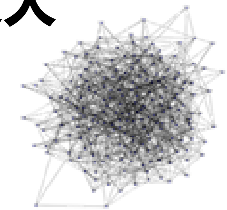
挑战二：图的规模更大



稀疏的边

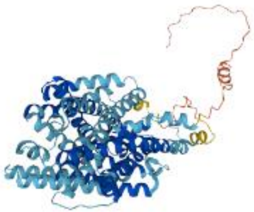


更多的噪声



长距离依赖问题

挑战三：图领域的多样性



蛋白质结构预测

迁移困难

社交网络

GFM特点

GFM和LFM（language foundation model）的对比

		Language Foundation Model	Graph Foundation Model
Similarities	Goal	Enhancing the model's expressive power and its generalization across various tasks	
	Paradigm	Pre-training and Adaptation	
Intrinsic differences	Data	Euclidean data (text)	Non-Euclidean data (graphs) or a mixture of Euclidean (e.g., graph attributes) and non-Euclidean data
	Task	Many tasks, similar formats	Limited number of tasks, diverse formats
Extrinsic differences	Backbone Architectures	Mostly based on Transformer	No unified architecture
	Homogenization	Easy to homogenize	Difficult to homogenize
	Domain Generalization	Strong generalization capability	Weak generalization across datasets
	Emergence	Has demonstrated emergent abilities	No/unclear emergent abilities as of the time of writing

[1] Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, Chuan Shi, "Towards Graph Foundation Models: A Survey and Beyond" Preprint 2023.

浙江大学 ZHEJIANG UNIVERSITY

目录

方法与分类

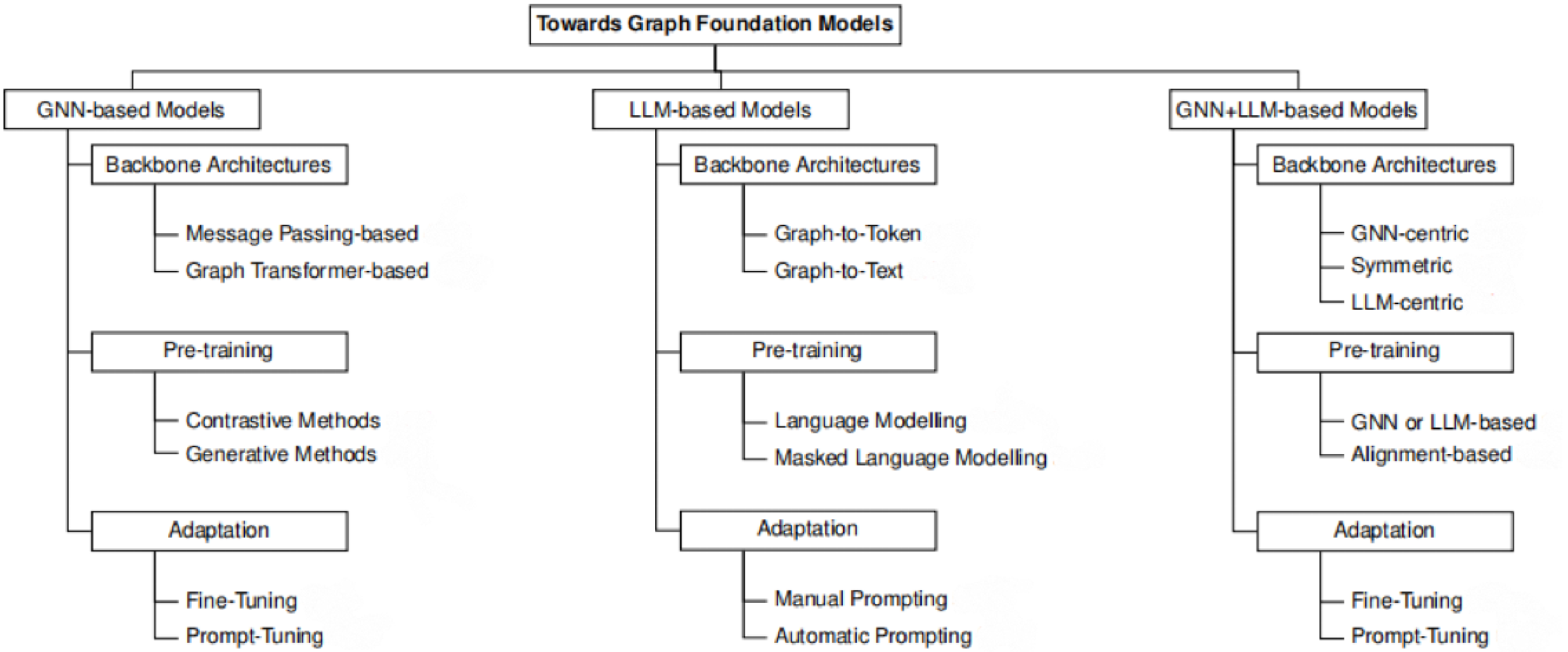


Fig. 2: Taxonomy of existing works towards graph foundation models.

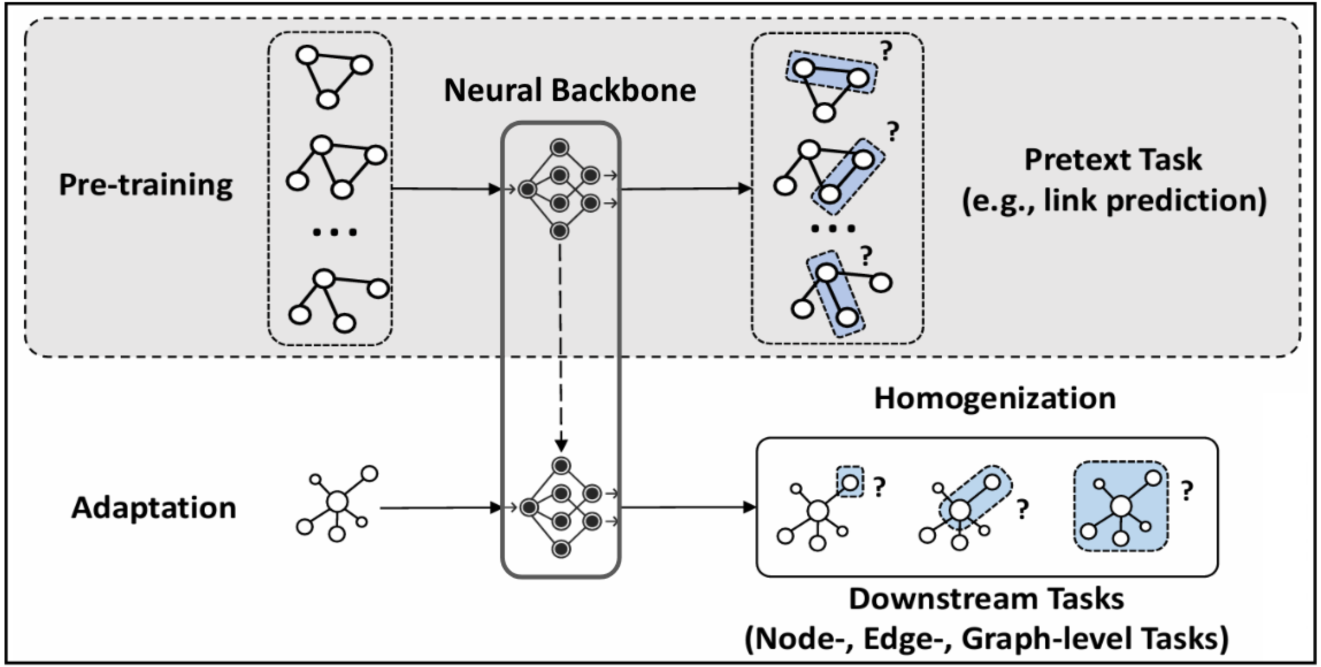
[1] Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, Chuan Shi, "Towards Graph Foundation Models: A Survey and Beyond" Preprint 2023.

GNN-based Methods

Backbone： No unified architecture

（Message Passing/Graph Transformer）

Paradigm： Pre-training + Adaptation

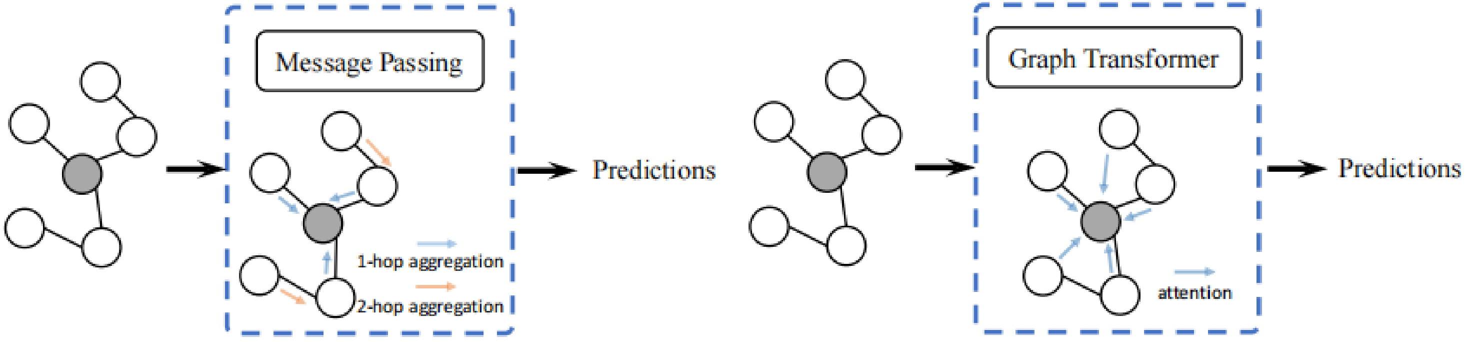


[1] Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, Chuan Shi, “Towards Graph Foundation Models: A Survey and Beyond” Preprint 2023.

GNN-based model — Backbone

GFM的底座选型：

- 基于消息传递（Message Passing-Based）
- 基于图transformer（Graph Transformer-Based）

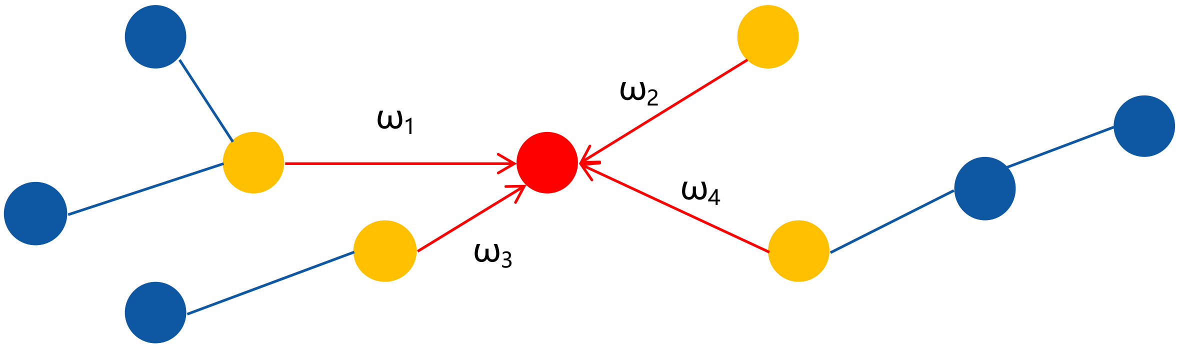


(a) Message Passing.　　(b) Graph Transformer.

Philip S. Yu, Chuan Shi, Cheng Yang, Yuan Fang, Lichao Sun, “Towards Graph Foundation Models” WWW 2024 Tutorial.

GNN-based model — Backbone

Message Passing



GCN
采用谱图卷积的近似，来捕获图结构信息和节点特征

GAT
引入注意力机制，调节聚合的权重

GIN
图同构网络，表达能力相当于1-WL测试

Graphsage
一类inductive学习框架，可以生成未知节点嵌入

HGT
可以针对异构图来提取特征

GNN-based model — Backbone

浙江大学 ZHEJIANG UNIVERSITY

Graph Transformer

- 它将图形视为完全连接，这意味着它会考虑并测量图形中每对节点之间的相似性。
- 长距离建模能力和强大的表达能力。

方法创新方向

- 增强图的空间结构信息：Graph-bert、Graphomer
- 提高模型泛化能力：GROVER
- 过全局化问题：CoBFormer
- 动态图：simplyDyG

Transformer

- **全局注意力机制:**

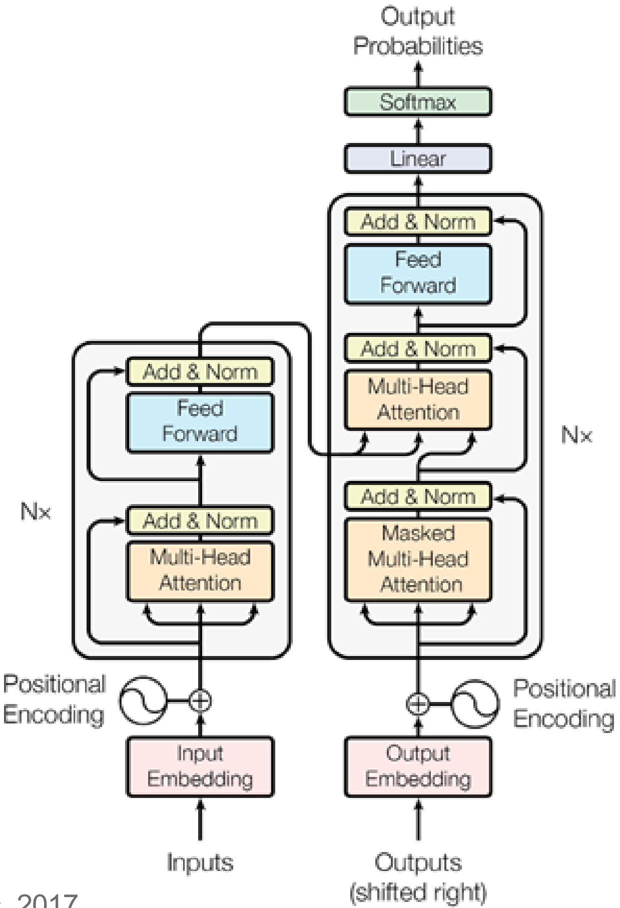
$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$Q=XW^Q, K=XW^K, V=XW^V$$

- **(序列)位置编码:**

$$PE_{(pos,2i)}=sin\left(pos/10000^{2i/d_{\text{model}}}\right)$$
$$PE_{(pos,2i+1)}=cos\left(pos/10000^{2i/d_{\text{model}}}\right)$$

- PE_{pos+k}**可以表示为**PE_{pos}**的线性函数**



Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.

Transformer ⇒ Graph Transformer

浙江大学 ZHEJIANG UNIVERSITY

■ 全局注意力机制：

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$Q=XW^Q, K=XW^K, V=XW^V$$

■ （序列）位置编码：

$$PE_{(pos,2i)}=sin\left(pos/10000^{2i/d_{\text{model}}}\right)$$

$$PE_{(pos,2i+1)}=cos\left(pos/10000^{2i/d_{\text{model}}}\right)$$

- PE_{pos+k}**可以表示为**PE_{pos}**的线性函数**

- **把基于序列的位置编码，推广到 图的拓扑结构 上。**
- **设计不同的全局/局部的位置/结构编码。**

Transformer ⇒ Graph Transformer

■ 全局注意力机制:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$Q=XW^Q, K=XW^K, V=XW^V$$

■ (序列)位置编码:

$$PE_{(pos,2i)}=sin\left(pos/10000^{2i/d_{\text{model}}}\right)$$
$$PE_{(pos,2i+1)}=cos\left(pos/10000^{2i/d_{\text{model}}}\right)$$

- PE_{pos+k}可以表示为PE_{pos}的线性函数

- 时空复杂度 $O(n^2d)$，无法应对大型图
- 常用两种解决方案
  - 局部注意力机制:
    - 采样/保留低阶邻居
  - 全局注意力机制:
    - 对Transformer进行 **时间和空间上的优化**

GCN vs. GAT vs. Transformer

- **传统的图卷积网络：**

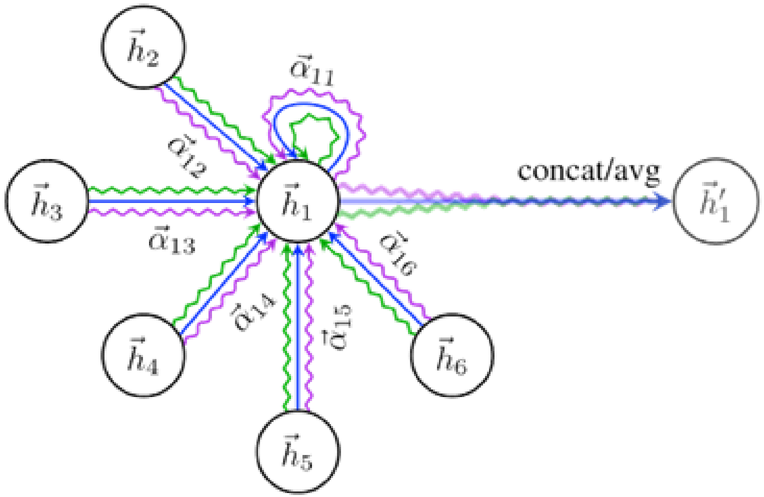
$$\mathbf{z}_i = \sum_{j \in \mathcal{N}_i} \mathbf{z}_j$$

- **图注意力网络：**

$$\mathbf{z}_i = \sum_{j \in \mathcal{N}_i} \text{similarity}(i, j) \cdot \mathbf{z}_j$$

- **Transformer：**

$$\mathbf{z}_i = \sum_{j} \text{similarity}(i, j) \cdot \mathbf{z}_j$$



Edge Channel: Graph Transformer

- **Graph Transformer**

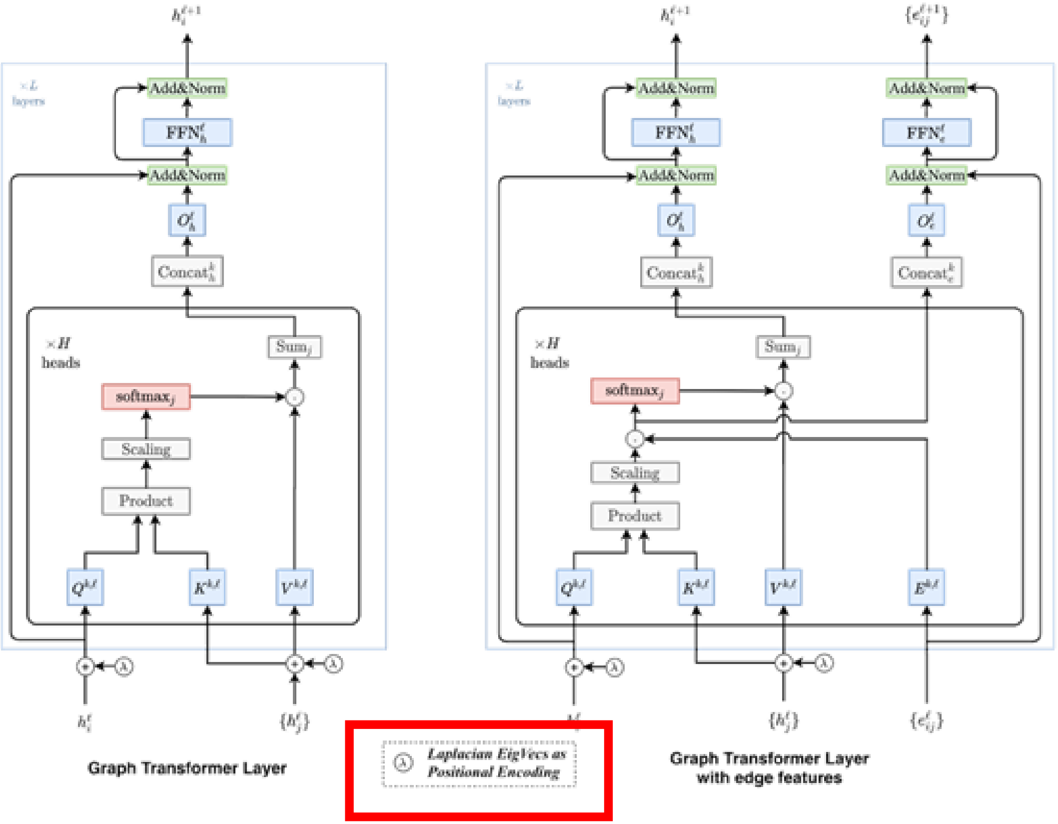
$$\hat{h}_i^{l+1} = O_h^l \mathop{\Big\Vert}_{k=1}^{H} \left(\sum_{j \in \mathcal{N}_i} w_{ij}^{k,l} V^{k,l} h_j^l \right)$$

$$w_{ij}^{k,l} = softmax_j \left(\frac{Q^{k,l} h_i^l \cdot K^{k,l} h_j^l}{\sqrt{d_k}} \right)$$

- **Graph Transformer with edge features**

$$\hat{e}_{ij}^{l+1} = O_e^l \mathop{\Big\Vert}_{k=1}^{H} \left(\frac{Q^{k,l} h_i^l \cdot K^{k,l} h_j^l}{\sqrt{d_k}} \cdot E^{k,l} e_{ij}^l \right)$$

$$w_{ij}^{k,l} = softmax_j \left(\frac{Q^{k,l} h_i^l \cdot K^{k,l} h_j^l}{\sqrt{d_k}} \cdot E^{k,l} e_{ij}^l \right)$$



Dwivedi V P, Bresson X. A Generalization of Transformer Networks to Graphs[J]. AAAI Workshop on Deep Learning on Graphs: Methods and Applications, 2021.

Encoding：Graphormer

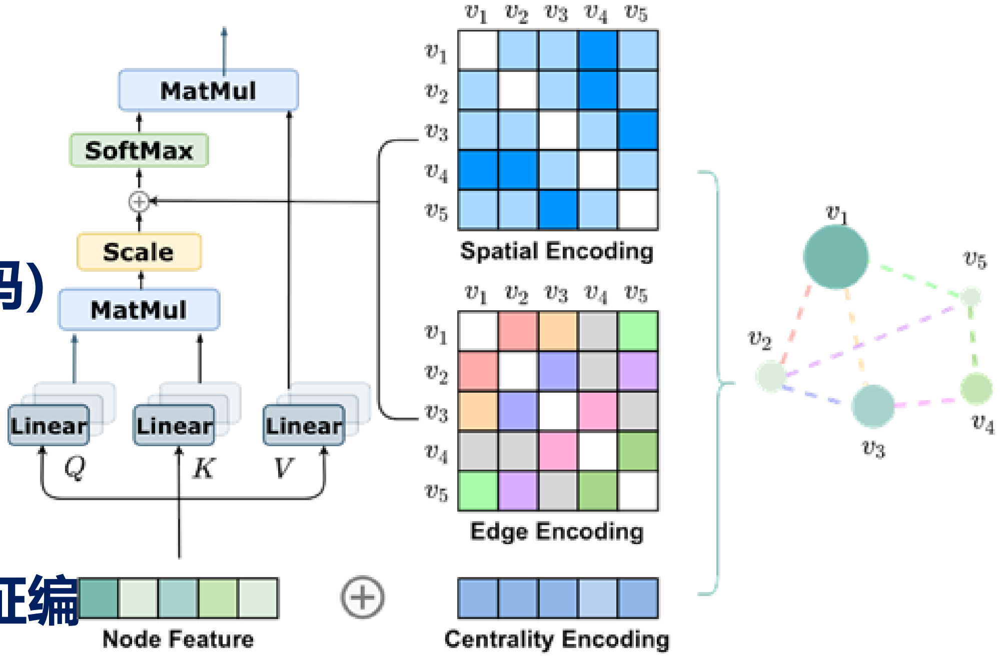
- **Centrality Encoding（度编码）**

$$h_i^{(0)} = x_i + z^-_{\deg^-(v_i)} + z^+_{\deg^+(v_i)}$$

- **Spatial Encoding（最短路径编码）**

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i,v_j)}$$

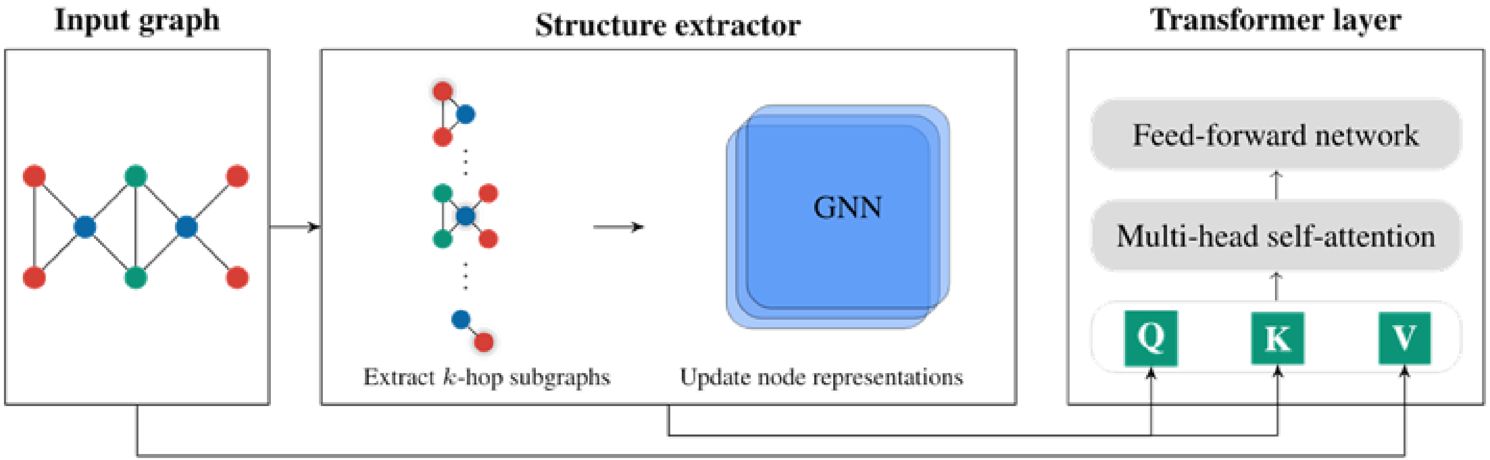
- **Edge Encoding（最短路径边特征编码）**



$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i,v_j)} + c_{ij}, \qquad \text{where } c_{ij} = \frac{1}{N}\sum_{n=1}^{N} x_{e_n}(w_n^E)^T$$

Ying C, Cai T, Luo S, et al. Do transformers really perform badly for graph representation?[J]. Advances in neural information processing systems, 2021, 34: 28877-28888.

Encoding：SAT



$$\mathrm{SA}-\mathrm{attn}(v)=\sum_{u\in V}\frac{\kappa_{\mathrm{graph}}(S_G(v),S_G(u))}{\sum_{w\in V}\kappa_{\mathrm{graph}}(S_G(v),S_G(u))}f(x_u)$$

K-subtree：提取以v为根的深度为k的子树做GNN
K-subgraph：把节点邻域的K-subtree的结果再聚合一次

$$\kappa_{\mathrm{graph}}(S_G(v),S_G(u))=\kappa_{\exp}(\varphi(v,G),\varphi(u,G))$$

$\varphi(v,G)$：k-subtree GNN extractor / k-subgraph GNN extractor

Chen D, O'Bray L, Borgwardt K. Structure-aware transformer for graph representation learning[C]//International Conference on Machine Learning. PMLR, 2022: 3469-3489.

Encoding Summary: GPS

Encoding type	Description	Examples
Local PE *node features*	Allow a node to know its position and role within a local cluster of nodes. *Within a **cluster**, the closer two nodes are to each other, the closer their local PE will be, such as the position of a word in a sentence (not in the text).*	• Sum each column of non-diagonal elements of the m-steps random walk matrix. • Distance between a node and the centroid of a cluster containing the node.
Global PE *node features*	Allow a node to know its global position within the graph. *Within a **graph**, the closer two nodes are, the closer their global PE will be, such as the position of a word in a text.*	• Eigenvectors of the Adjacency, Laplacian [15, 36] or distance matrices. • SignNet [39] (includes aspects of relative PE and local SE). • Distance from the graph's centroid. • Unique identifier for each connected component of the graph.
Relative PE *edge features*	Allow two nodes to understand their distances or directional relationships. *Edge embedding that is correlated to the distance given by any global or local PE, such as the distance between two words.*	• Pair-wise node distances [38, 3, 36, 63, 44] based on shortest-paths, heat kernels, random-walks, Green's function, graph geodesic, or any local/global PE. • Gradient of eigenvectors [3, 36] or any local/global PE. • PEG layer [57] with specific node-wise distances. • Boolean indicating if two nodes are in the same cluster.
Local SE *node features*	Allow a node to understand what substructures it is a part of. *Given an SE of radius m, the more similar the m-hop subgraphs around two nodes are, the closer their local SE will be.*	• Degree of a node [63]. • Diagonal of the m-steps random-walk matrix [16]. • Time-derivative of the heat-kernel diagonal (gives the degree at $t = 0$). • Enumerate or count predefined structures such as triangles, rings, etc. [6, 68]. • Ricci curvature [54].
Global SE *graph features*	Provide the network with information about the global structure of the graph. *The more similar two graphs are, the closer their global SE will be.*	• Eigenvalues of the Adjacency or Laplacian matrices [36]. • Graph properties: diameter, girth, number of connected components, # of nodes, # of edges, nodes-to-edges ratio.
Relative SE *edge features*	Allow two nodes to understand how much their structures differ. *Edge embedding that is correlated to the difference between any local SE.*	• Pair-wise distance, encoding, or gradient of any local SE. • Boolean indicating if two nodes are in the same substructure [5] (similar to the gradient of sub-structure enumeration).

- **MPNN Layer:**
  PE (Positional Encoding)
  +SE (Structural Encoding)
  +Graph features

Rampášek L, Galkin M, Dwivedi V P, et al. Recipe for a general, powerful, scalable graph transformer[J]. Advances in Neural Information Processing Systems, 2022, 35: 14501-14515.

Encoding Summary: GPS

- **Graphormer**
  - **Centrality Encoding**
  - **Spatial Encoding**
  - **Edge Encoding**
- **SAT**
  - **Structure Extractor**
- **SAN (Laplacian)**
  - **Node-wise LPE**
  - **Edge-wise LPE**

Encoding type	Description	Examples
Local PE *node features*	Allow a node to know its position and role within a local cluster of nodes. *Within a **cluster**, the closer two nodes are to each other, the closer their local PE will be, such as the position of a word in a sentence (not in the text).*	• Sum each column of non-diagonal elements of the m-steps random walk matrix. • Distance between a node and the centroid of a cluster containing the node.
Global PE *node features*	Allow a node to know its global position within the graph. *Within a **graph**, the closer two nodes are, the closer their global PE will be, such as the position of a word in a text.*	• Eigenvectors of the Adjacency, Laplacian [15, 36] or distance matrices. • SignNet [39] (includes aspects of relative PE and local SE). • Distance from the graph's centroid. • Unique identifier for each connected component of the graph.
Relative PE *edge features*	Allow two nodes to understand their distances or directional relationships. *Edge embedding that is correlated to the distance given by any global or local PE, such as the distance between two words.*	• Pair-wise node distances [38, 3, 36, 63, 44] based on shortest-paths, heat kernels, random-walks, Green's function, graph geodesic, or any local/global PE. • Gradient of eigenvectors [3, 36] or any local/global PE. • PEG layer [57] with specific node-wise distances. • Boolean indicating if two nodes are in the same cluster.
Local SE *node features*	Allow a node to understand what sub-structures it is a part of. *Given an SE of radius m, the more similar the m-hop subgraphs around two nodes are, the closer their local SE will be.*	• Degree of a node [63]. • Diagonal of the m-steps random-walk matrix [16]. • Time-derivative of the heat-kernel diagonal (gives the degree at $t = 0$). • Enumerate or count predefined structures such as triangles, rings, etc. [6, 68]. • Ricci curvature [54].
Global SE *graph features*	Provide the network with information about the global structure of the graph. *The more similar two graphs are, the closer their global SE will be.*	• Eigenvalues of the Adjacency or Laplacian matrices [36]. • Graph properties: diameter, girth, number of connected components, # of nodes, # of edges, nodes-to-edges ratio.
Relative SE *edge features*	Allow two nodes to understand how much their structures differ. *Edge embedding that is correlated to the difference between any local SE.*	• Pair-wise distance, encoding, or gradient of any local SE. • Boolean indicating if two nodes are in the same sub-structure [5] (similar to the gradient of sub-structure enumeration).

Rampášek L, Galkin M, Dwivedi V P, et al. Recipe for a general, powerful, scalable graph transformer[J]. Advances in Neural Information Processing Systems, 2022, 35: 14501-14515.

Complexity: Nodeformer

- **Traditional Transformer:**

$$\mathbf{z}_u^{(l+1)} = \sum_{v=1}^{N} \frac{\exp(\mathbf{q}_u^T \boldsymbol{k}_{\mathrm{v}})}{\sum_{w=1}^{N} \exp(\mathbf{q}_u^T \boldsymbol{k}_w)} \cdot \mathbf{v}_v$$

- **Nodeformer:**

$$\mathbf{z}_u^{(l+1)} = \sum_{v=1}^{N} \frac{\phi(\mathbf{q}_u)^T \phi(\mathbf{k}_v)}{\sum_{w=1}^{N} \phi(\mathbf{q}_u)^T \phi(\mathbf{k}_w)} \cdot \mathbf{v}_v = \frac{\phi(\mathbf{q}_u)^T \sum_{v=1}^{N} \phi(\mathbf{k}_v) \cdot \mathbf{v}_v^T}{\phi(\mathbf{q}_u)^T \sum_{w=1}^{N} \phi(\mathbf{k}_w)}$$

- **where $\phi(\mathrm{x})$ using Positive Random Features (PRF):**

$$\phi(x) = \frac{\exp\left(\frac{-\|\mathbf{x}\|_2^2}{2}\right)}{\sqrt{m}} [\exp(\mathbf{w}_1^T \mathbf{x}), \cdots, \exp(\mathbf{w}_m^T \mathbf{x})]$$

Wu Q, Zhao W, Li Z, et al. Nodeformer: A scalable graph structure learning transformer for node classification[J]. Advances in Neural Information Processing Systems, 2022, 35: 27387-27401.

Complexity：Nodeformer

- **Latent Graph**
  - **The categorical distribution of edge $< u, v >$ in latent graph:**

$$\pi_{uv}^{(l)} = p(v|u) = \mathrm{softmax}(\mathbf{q}_v \mathbf{k}_u)$$

  - **Gumble-Softmax（离散变量建模神器） with Kernel function**

$$\mathbf{z}_u^{(l+1)} = \sum_{v=1}^{N} \frac{\phi(\mathbf{q}_u/\sqrt{\tau})^T \phi(\mathbf{k}_v/\sqrt{\tau}) e^{g_w/\tau}}{\sum_{w=1}^{N} \phi(\mathbf{q}_u/\sqrt{\tau})^T \phi(\mathbf{k}_w/\sqrt{\tau}) e^{g_w/\tau}} \cdot \mathbf{v}_v$$
$$= \frac{\phi(\mathbf{q}_u/\sqrt{\tau})^T \sum_{v=1}^{N} \phi(\mathbf{k}_v/\sqrt{\tau}) \cdot \mathbf{v}_v^T}{\phi(\mathbf{q}_u/\sqrt{\tau})^T \sum_{w=1}^{N} \phi(\mathbf{k}_w/\sqrt{\tau})}$$

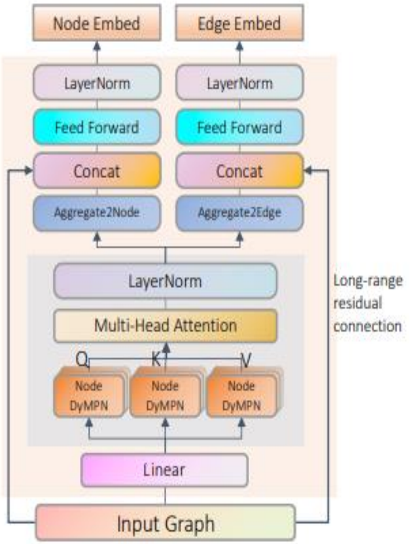
- **Adding Graph Topology Info**

$$\mathbf{z}_u^{(l+1)} \leftarrow \mathbf{z}_u^{(l+1)} + \sum_{v, a_{uv}=1} \sigma\left(b^{(l)}\right) \cdot \mathbf{v}_v$$

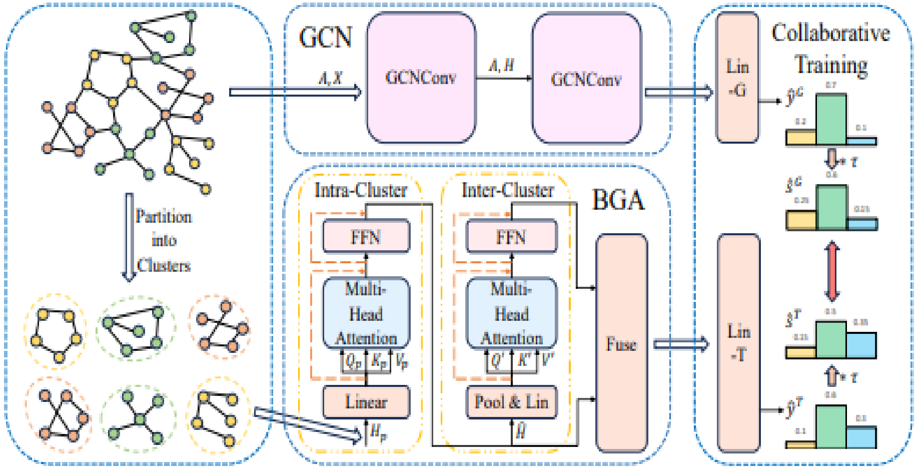
Wu Q, Zhao W, Li Z, et al. Nodeformer: A scalable graph structure learning transformer for node classification[J]. Advances in Neural Information Processing Systems, 2022, 35: 27387-27401.

GNN-based model — Backbone

提高模型泛化能力



过全局化问题（舍近求远）



GROVER[1]:使用Dynamic MPN动态提取不同阶邻居节点特征（heterophily? 🤔）

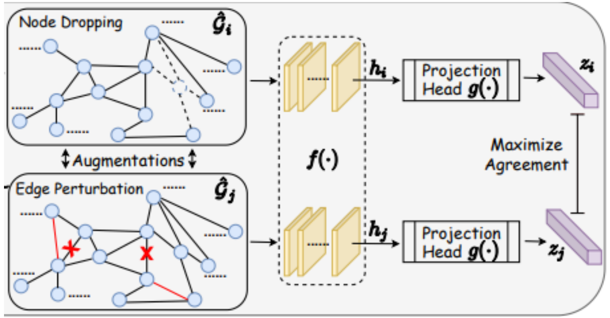
CoBFormer[2]:由两个GT组成，一个得到簇内特征，一个得到簇间特征（GCN+Transformer）

[1] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, “Self-supervised graph transformer on large-scale molecular data,” Proc. of NeurIPS, 2020.
[2] Y. Xing, X. Wang, Y. Li, H. Huang, and C. Shi, “Less is more: on the over-globalizing problem in graph transform,” in Forty-first International Conference on Machine Learning, 2024.

GNN-based model — Pre-training

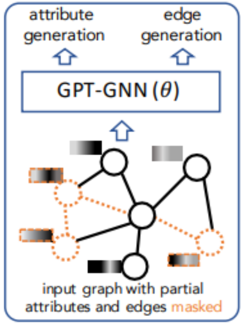
对比式（Contrastive Methods）

- 同尺度对比学习
- 跨尺度对比学习

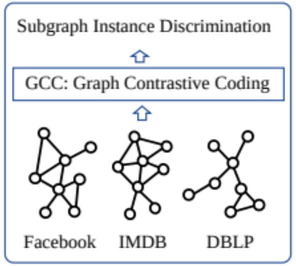


生成式（Generative Methods）

- 图重构
- 属性预测



Pre-Training



Pre-Training

Contrastive Methods

同尺度对比

- **GCC[1]**

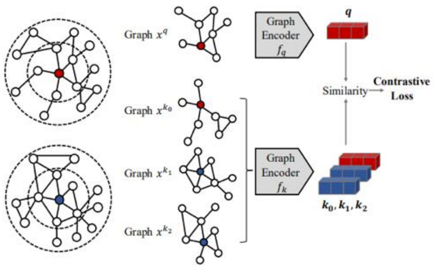
节点的子图的嵌入视为节点嵌入,不同子图进行对比。

- **GraphCL[2]**

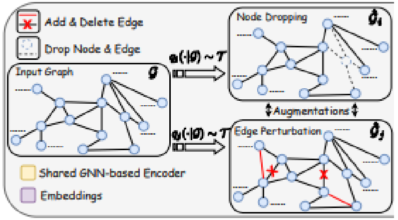
使用图增强方法来获得扰动后的view，再通过对比学习来对图的拓扑结构进行学习。

- **MA-GCL[3]**

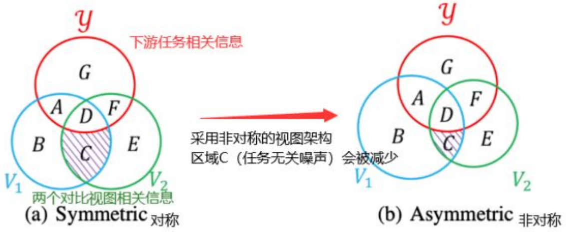
在模型结构上进行扰动，比如使用两个共享参数但是传播层数不同的encoder，将两个结果进行对比，去除噪声。



GCC



GraphCL



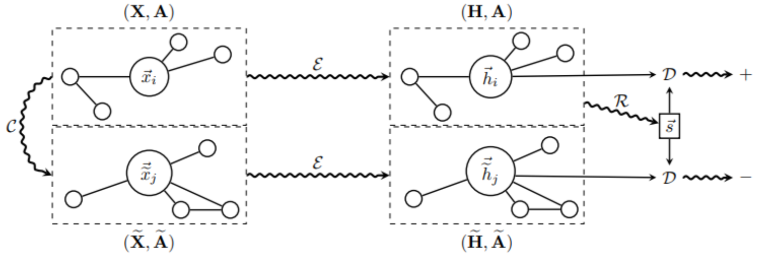
MA-GCL

[1] J. Qiu, Q. Chen, Y. Dong, J. Zhang, H. Yang, M. Ding, K. Wang, and J. Tang, “Gcc: Graph contrastive coding for graph neural network pre-training,” in Proc. of KDD, 2020.
[2] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, “Graph contrastive learning with augmentations,” Proc. of NeurIPS, 2020.
[3] X. Gong, C. Yang, and C. Shi, “Ma-gcl: Model augmenta_x0002_tion tricks for graph contrastive learning,” in Proc. of AAAI, 2023.

GNN-based model — Pre-training

Contrastive Methods

DGI跨尺度对比:将高维节点信息嵌入聚合为图嵌入，然后使用图嵌入和节点嵌入进行对比，最大化节点嵌入和图嵌入之间互信息。



1. Sample a negative example by using the corruption function: $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{A}}) \sim \mathcal{C}(\mathbf{X}, \mathbf{A})$.
2. Obtain patch representations, $\vec{h}_i$ for the input graph by passing it through the encoder: $\mathbf{H} = \mathcal{E}(\mathbf{X}, \mathbf{A}) = \{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_N\}$.
3. Obtain patch representations, $\vec{\widetilde{h}}_j$ for the negative example by passing it through the encoder: $\widetilde{\mathbf{H}} = \mathcal{E}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{A}}) = \{\vec{\widetilde{h}}_1, \vec{\widetilde{h}}_2, \ldots, \vec{\widetilde{h}}_M\}$.
4. Summarize the input graph by passing its patch representations through the readout function: $\vec{s} = \mathcal{R}(\mathbf{H})$.
5. Update parameters of $\mathcal{E}$, $\mathcal{R}$ and $\mathcal{D}$ by applying gradient descent to maximize Equation].

[1]] P. Velickovi ˇ c, W. Fedus, W. L. Hamilton, P. Li ´ o, Y. Bengio, `and R. D. Hjelm, “Deep graph infomax,” in Proc. of ICLR, 2019.

Graph Reconstruction

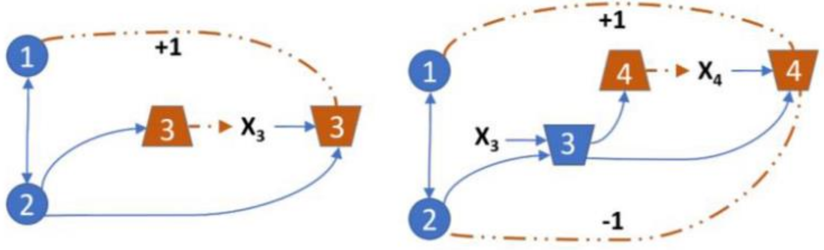
VGAE[1]

对高阶结构和节点属性不敏感

GPT-GNN[2]

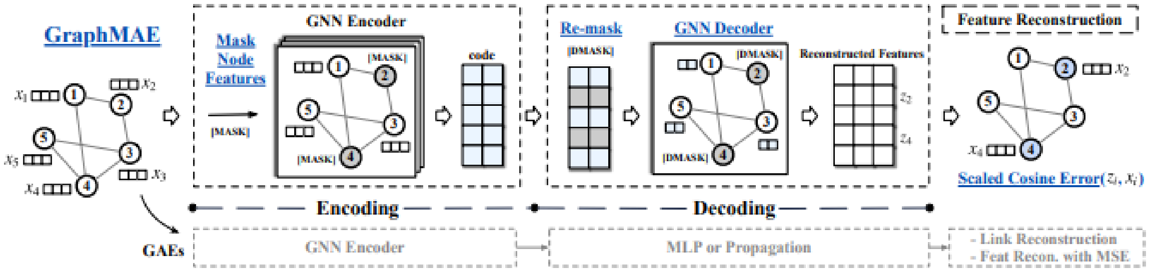
过于重视图结构信息

GraphMAE[3]



(d) Generate attribute and masked edges for node 3.

(e) Generate attribute and masked edges for node 4.



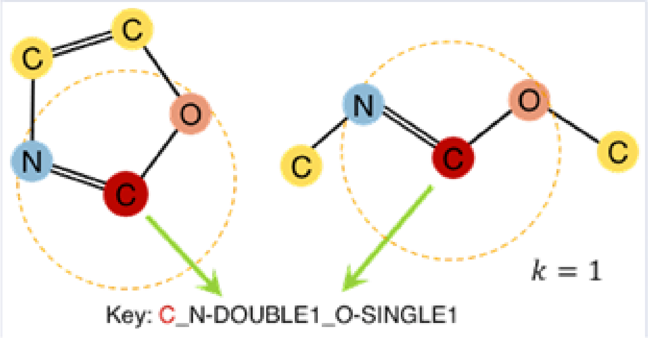
[1] T. N. Kipf and M. Welling, “Variational graph auto-encoders,” NIPS Workshop on Bayesian Deep Learning, 2016.
[2] Z. Hu, Y. Dong, K. Wang, K.-W. Chang, and Y. Sun, “Gpt-gnn: Generative pre-training of graph neural networks,” in Proc. of KDD, 2020.
[3] Z. Hou, X. Liu, Y. Cen, Y. Dong, H. Yang, C. Wang, and J. Tang, “Graphmae: Self-supervised masked graph autoencoders,” in Proc. of KDD, 2022.

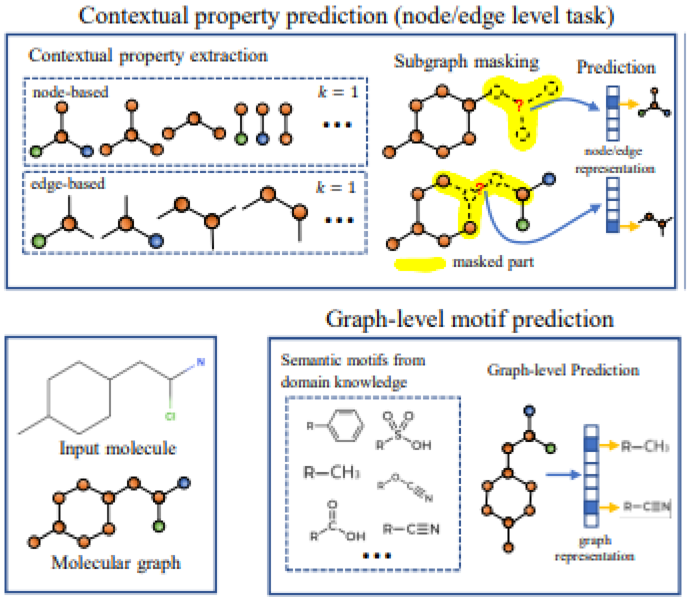
GNN-based model — Pre-training

Property Prediction

- **Grover**

使用的是节点和边的任务，预测子图中的上下文感知属性contextual property prediction，并且使用motif（graph-level motif prediction）进行预测训练，增强模型的表达能力。





[1] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, “Self-supervised graph transformer on large-scale molecular data,” Proc. of NeurIPS, 2020.

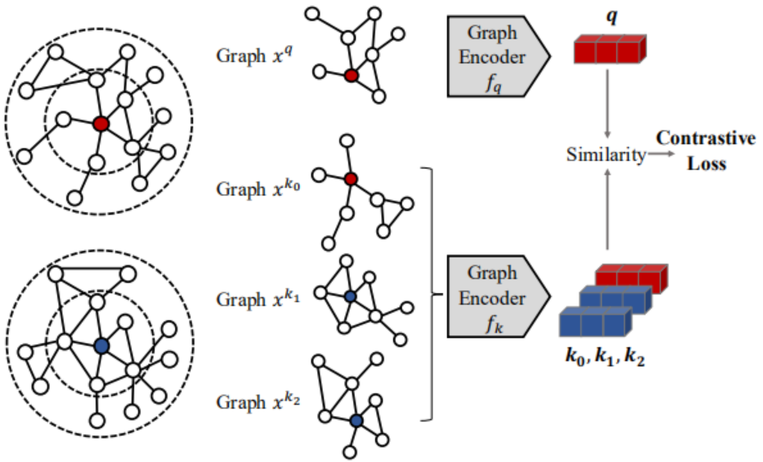
GNN-based model — Pre-training

浙江大学 ZHEJIANG UNIVERSITY

预训练具体步骤

关键：没有节点特征和节点标签，所以只用一个点作为输入是无效的

<u>需要样本点连接边的拓扑信息</u>



r-ego网络产生：
1、带有重启的随机行走：以点z开始进行采样
2、产生子图
3、将采样的子图匿名化（Node Index）

来源于同一点z产生的子图组成一个similar pair
来源于不同点z产生的子图组成一个dissimilar pair

将样本点的子图通过GNN网络进行编码，通过infoNCE得到对比损失。目的是让similar pair中的距离接近，dissimilar pair中的距离远离。

GNN-based model — Adaptation

Fine-tuning

利用预训练模型生成节点嵌入或图形嵌入，然后微调外部任务特定层，将预训练模型推广到下游任务。

DGI/GRACE:

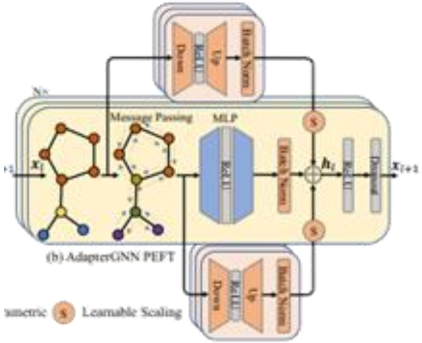
同领域数据进行预训练+微调

GPT-GNN：
进行了数据的迁移训练
论文中将迁移情况分为三类:
不同时间，不同领域，不同时间且不同领域

parameter-efficient fine-tuning

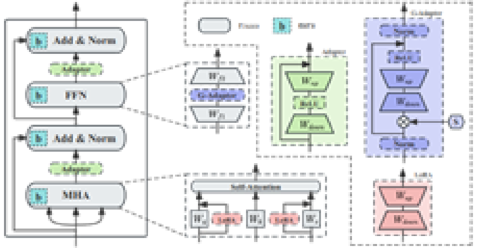
AdapterGNN[1]

- 在message passing 前后都有添加bottleneck结构
- 专门针对GNN模型



G-Adapter[2]

- 利用图卷积运算引入图结构作为归纳偏置来引导更新过程
- 专门针对Transformer模型



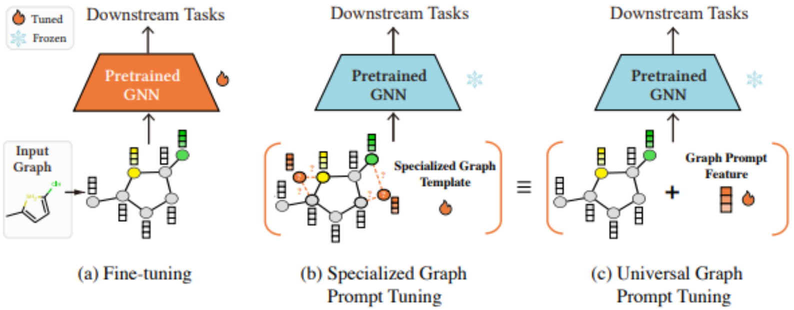
[1] S. Li, X. Han, and J. Bai, “Adaptergnn: Parameter-efficient fine-tuning improves generalization in gnns,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 13600–13608, 2024.
[2] A. Gui, J. Ye, and H. Xiao, “G-adapter: Towards structure aware parameter-efficient transfer learning for graph transformer networks,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 12226–12234, 2024.

GNN-based model — Adaptation

prompt tuning

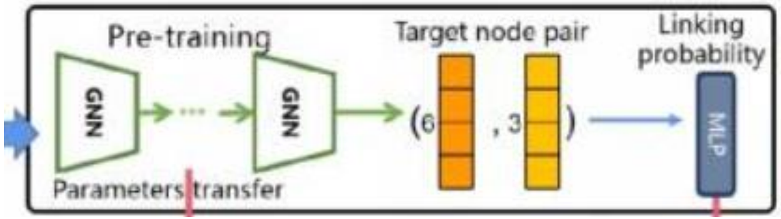
Pre-Prompt

GPF[1]：
维护一个特定的prompt vector，来对所有节点进行更新



Post-Prompt

GPPT[2]：
为每个类设定一个虚拟节点，这样就把分类任务转换成了连接预测任务



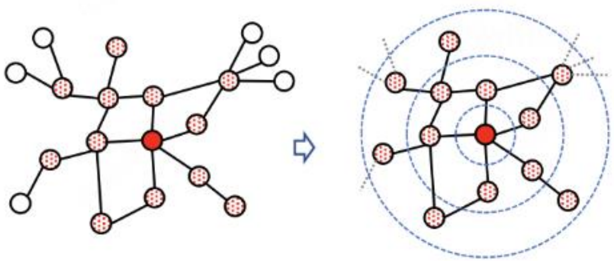
[1] T. Fang, Y. M. Zhang, Y. Yang, C. Wang, and L. CHEN, “Universal prompt tuning for graph neural networks,” in Thirty-seventh Conference on Neural Information Processing Systems, 2023.
[2] M. Sun, K. Zhou, X. He, Y. Wang, and X. Wang, “Gppt: Graph pre-training and prompt tuning to generalize graph neural networks,” in Proc. of KDD, 2022.

GNN-based model — Adaptation

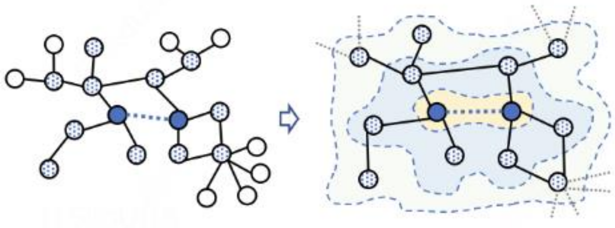
将节点级别和边级别的任务转为图级别的任务，构建induced-graph。

抽取中心节点附近的子图，节点分类转换为子图分类

两个节点之间的边预测/分类，变成两个节点附近的子图连接之后的子图的分类



(a) Induced graphs for nodes



(b) Induced graphs for edges

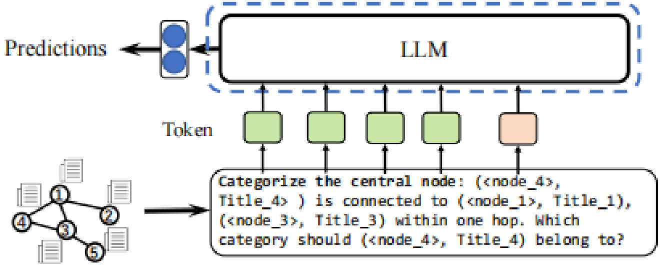
[1] Sun, X., Cheng, H., Li, J., Liu, B., & Guan, J. All in one: Multi-task prompting for graph neural networks. KDD2023.

LLM-based model — Backbone

Backbone Architecture

Graph-to-token

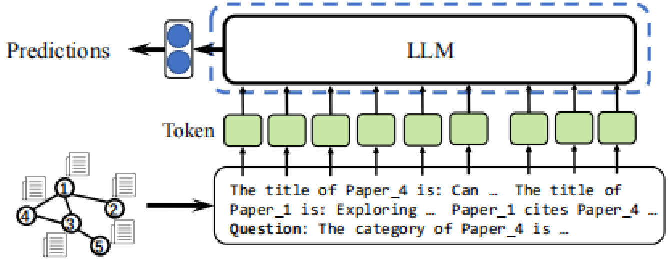
将图信息进行标记化，模仿基于 Transformer 模型的标准输入格式，最后喂给LLM



(a) Graph-to-token.

Graph-to-text

通过使用自然语言描述图信息，使图数据与自然语言对齐，直接将文本喂给LLM



(b) Graph-to-text.

LLM-based model — Backbone

Graph-to-token

GIMLET[1]

- 每个节点设置为一个token
- 在注意力层中额外进行编码，目的是获取图的拓扑结构信息

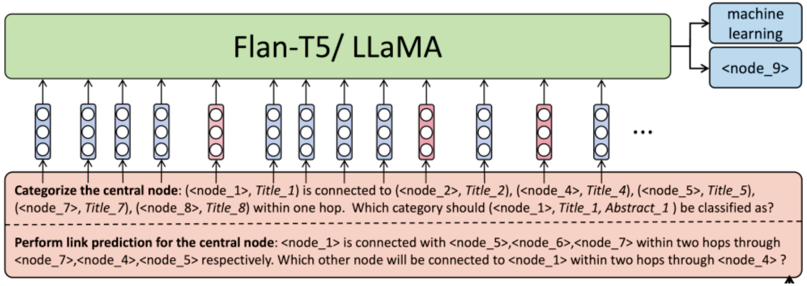
$$\hat{A}_{ij} = \frac{(h_i W^Q)(h_j W^K)^T}{\sqrt{d_k}} + b(i,j)$$

$$b(i,j) = b^D_{\mathrm{POS}(i,j)} + b^M_{i,j} + \underset{k \in \mathrm{SP}(i,j)}{\mathrm{Mean}}\ b^E_{e_k},$$

- **b_{POS}**：相对位置编码，图最短路径（Graph）或序列长度（Text）
- **$b_{i,j}$**:区分字符和图节点（0或正无穷）
- **Mean b_{ek}**：最短路径上边特征平均池化

instructGLM[2]

- 扩展了LLM的vocabulary，基于固有节点特征向量为每个节点生成嵌入
- 下图中红色向量是节点嵌入，蓝色向量是自然语言token



[1] H. Zhao, S. Liu, M. Chang, H. Xu, J. Fu, Z. Deng, L. Kong, and Q. Liu, "Gimlet: A unified graph-text model for instruction-based molecule zero-shot learning," Advances in Neural Information Processing Systems, vol. 36, 2023.
[2] R. Ye, C. Zhang, R. Wang, S. Xu, and Y. Zhang, "Language is all a graph needs," EACL, 2024.

LLM-based model — Backbone

Graph-to-text

LLMtoGraph[1]

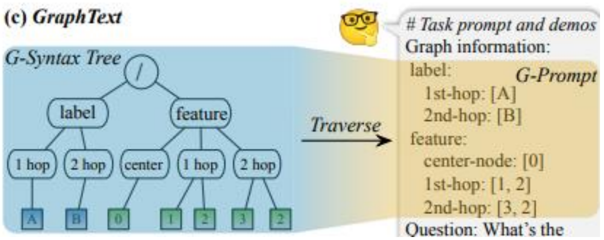
首次使用自然语言来对图数据进行表示，例如节点，边

graph data	given the undirected graph with the specified nodes and edges, nodes: [0, 1, 2, 3, 4, 5, 6, 7], edges: [(0, 5), (1, 6), (3, 6), (4, 5), (5, 6), (6, 7)].
query	find a single path from node 7 to node 6 connected by edges in the given graph.
tail	list the answer after "Ans:" in the format of [0-1-2].

形式上的创新

GraphText[2]

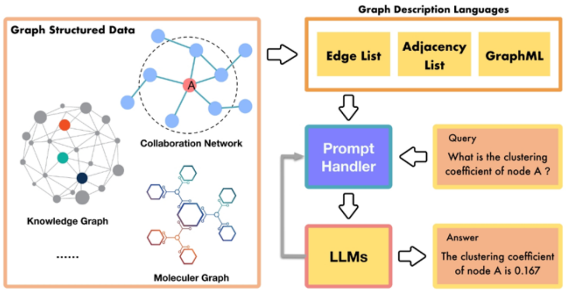
使用语法树来表示图数据



形式上的扩展

GPT4Graph[3]

使用特定的图描述语言表示



[1] C. Liu and B. Wu, “,” arXiv preprint arXiv:2308.11224, 2023. Evaluating large language models on graphs: Performance insights and comparative analysis

[2] J. Zhao, L. Zhuo, Y. Shen, M. Qu, K. Liu, M. Bronstein,Z. Zhu, and J. Tang, “Graphtext: Graph reasoning in text space,” 2023.

[3] J. Guo, L. Du, and H. Liu, “Gpt4graph: Can largelanguage models understand graph structured data? anempirical evaluation and benchmarking,” arXiv preprint arXiv:2305.15066, 2023.

LLM-based model — Pre-training

语言建模（ML）

- 一般只表示单向语言模型

$$p\left(s_{1:L}\right)=\prod_{l=1}^{L} p\left(s_{l} \mid s_{0:l-1}\right).$$

$$p\left(s_{l} \mid s_{0:l-1}\right)=f_{lm}\left(f_{nenc}\left(s_{0:l-1}\right)\right).$$

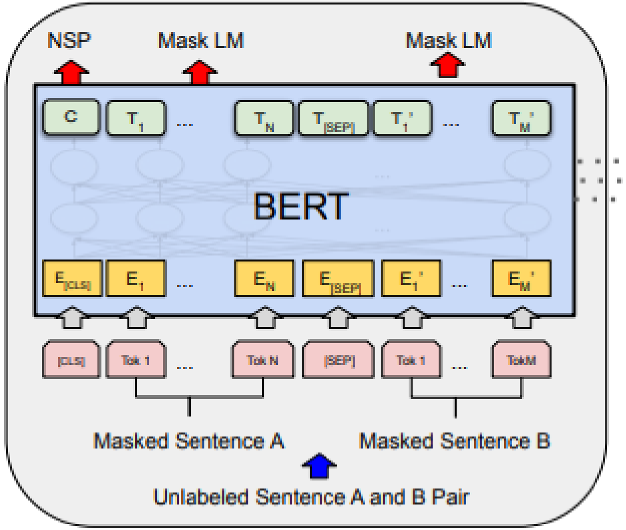
掩码语言建模（MLM）

- 序列到序列掩码语言建模（Seq2Seq MLM）

  mask序列经过encoder后输出嵌入，再通过softmax预测标记

- 增强型掩码语言建模（E-MLM）

  将掩码预测任务扩展到各种类型的语言建模任务



Pre-training

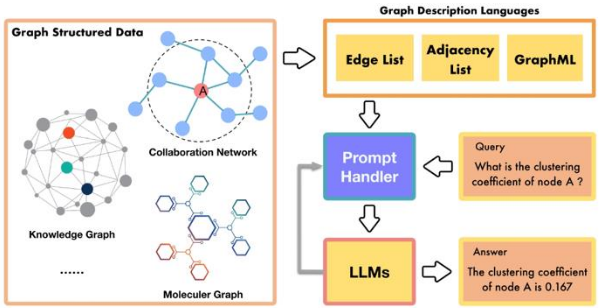
案例—— GPT4Graph

模型：GPT4Graph

类型：基于LLM的模型

Backbone：Graph-to-text

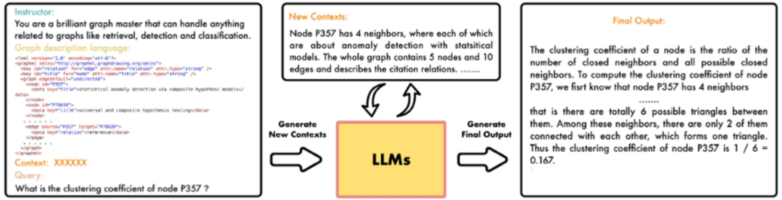
流程



生成图描述语言（GDL）

↓

使用图提示处理程序将用户查询和GDL结合形成LLM的输入 → LLM执行推理并为用户生成答案



提示

· 手动提示：手动增加格式解释。

· 自提示：提取关键特征来生成给定图的摘要。生成的摘要可以作为后续图形相关任务的提示。

图理解基准

· **结构理解**

（1）图大小检测

（2）度检测

（3）边检测

（4）直径计算

（5）聚类系数计算

· **语义理解**

（1）知识图谱问题分类

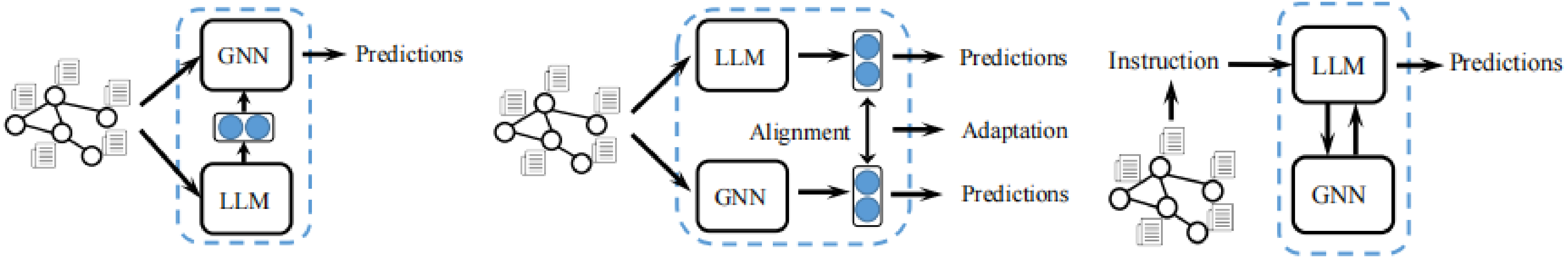
（2）图查询语言生成

（3）结点分类

（4）图分类

GNN+LLM-based model — Backbone

- GNN-centric：用LLM抽取节点特征，用GNN做预测
- Symmetric：将LLM和GNN的表征对齐
- LLM-centric：使用GNN来增强LLM的能力



(a) GNN-centric methods.

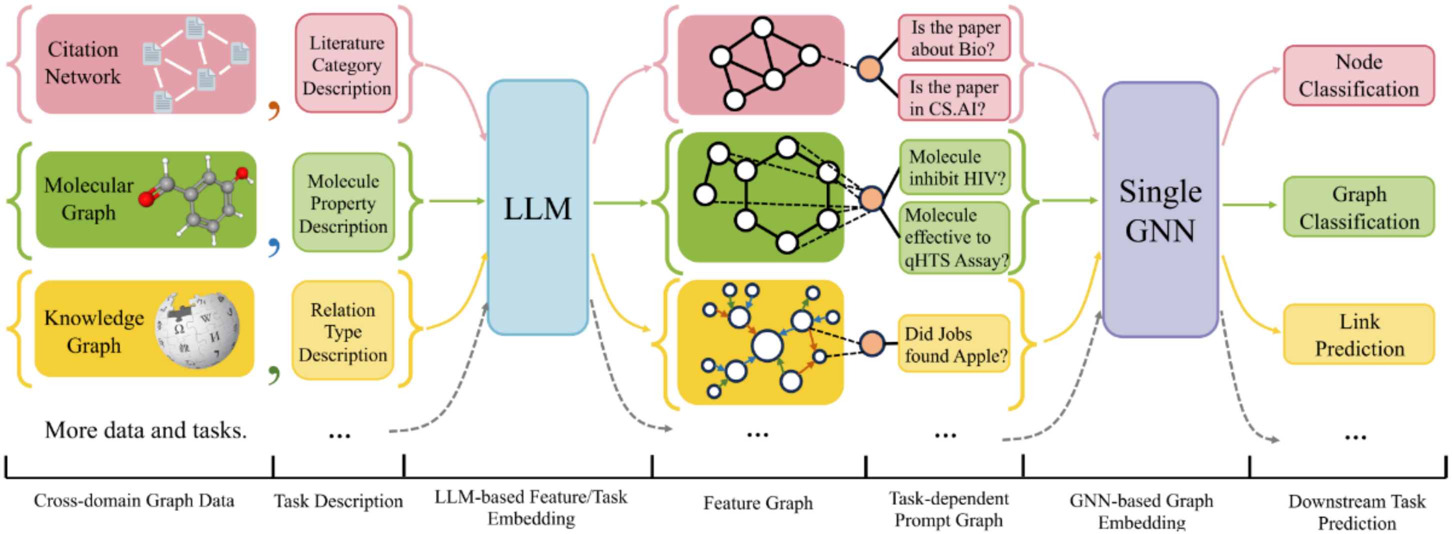
(b) Symmetric methods, where the aligned embeddings can be further utilized for downstream tasks.

(c) LLM-centric methods, which take an instruction as input and output an answer.

GNN+LLM-based model — Backbone

Text-attributed graph
Task description

LLMs → Prompted graph → GNN → Downstream tasks

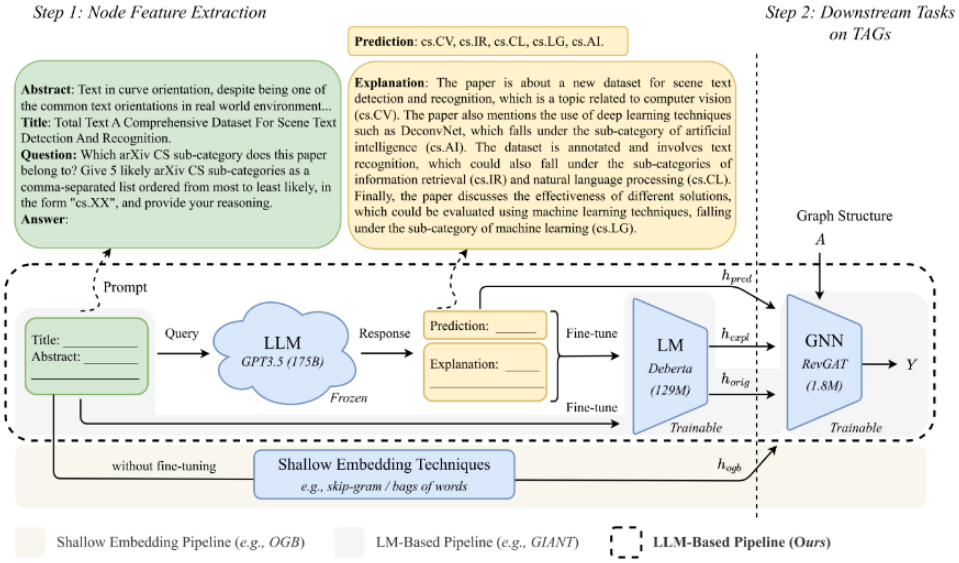


Liu, et al. "One for all: Towards training one graph model for all classification tasks." ICLR'24

GNN+LLM-based model — Backbone

GNN-centric

Textual attributes → LLM → Prediction & Explanation → Fine-tune LM → Node features → GNN



X. He, X. Bresson, T. Laurent, A. Perold, Y. LeCun, and B. Hooi, “Harnessing explanations: Llm-to-lm interpreter for enhanced text-attributed graph representation learning,” in The Twelfth International Conference on Learning Representations, 2024.

GNN+LLM-based model — Backbone

Symmtric

GraphFormer[1]

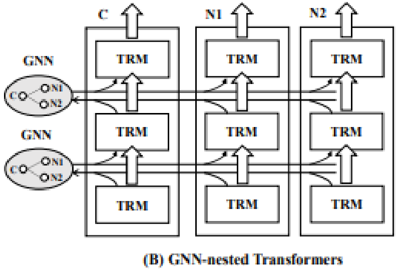
文本嵌入和图聚合融合为迭代流程但有可扩展性问题

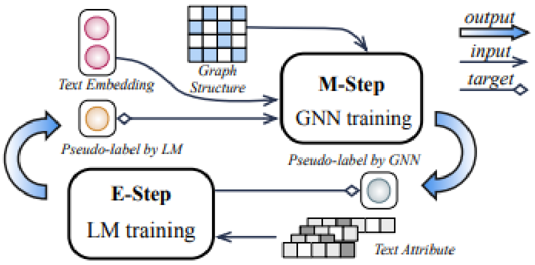
GALM[2]

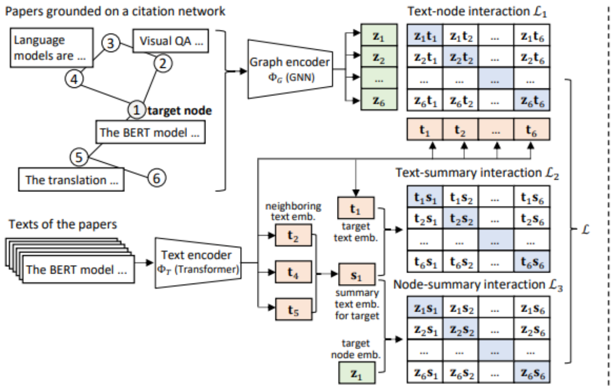
采用变分期望最大化框架交替更新LLM 和 GNN 缓解可扩展性问题

G2P2[3]

用不同编码器在共享潜在空间对齐图节点和文本表示







[1] J. Yang, Z. Liu, S. Xiao, C. Li, D. Lian, S. Agrawal, A. Singh, G. Sun, and X. Xie, “Graphformers: Gnn-nested transformers for representation learning on textual graph,” Proc. of NeurIPS, 2021.

[2] X. He, X. Bresson, T. Laurent, A. Perold, Y. LeCun, and B. Hooi, “Harnessing explanations: Llm-to-lm interpreter for enhanced text-attributed graph representation learning,” in The Twelfth International Conference on Learning Representations, 2024.

[3] Z. Wen and Y. Fang, “Augmenting low-resource text clas_x0002_sification with graph-grounded pre-training and prompting,” in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 506–516, 2023.

GNN+LLM-based model — Backbone

LLM-centric

GraphPrompter[1]

1、使用GNN编码图信息

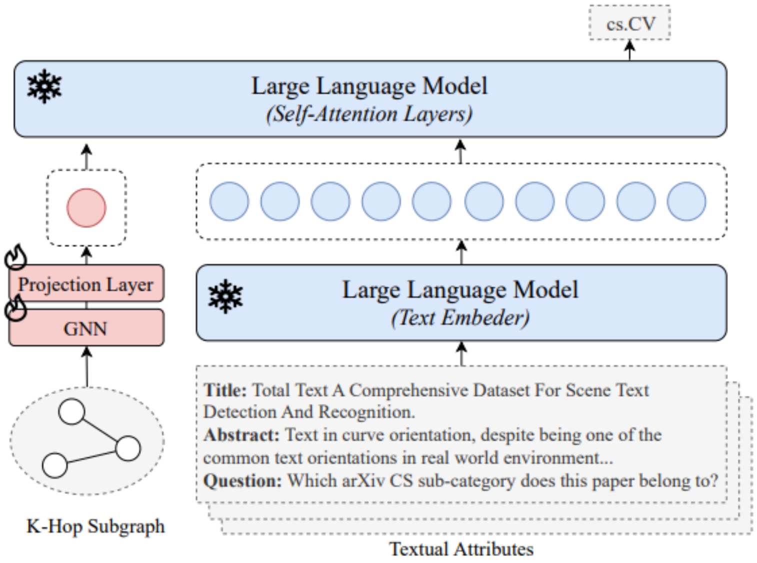
2、使用LLM获得文本嵌入

3、两者编码融合，由最后的LLM给出结果

优势:

1、无需大量微调

2、有效的对齐了图拓扑信息和文本信息



[1] Z. Liu, X. Yu, Y. Fang, and X. Zhang, "Graphprompt: Unifying pre-training and downstream tasks for graph neural networks," in Proceedings of the ACM Web Conference 2023, 2023.

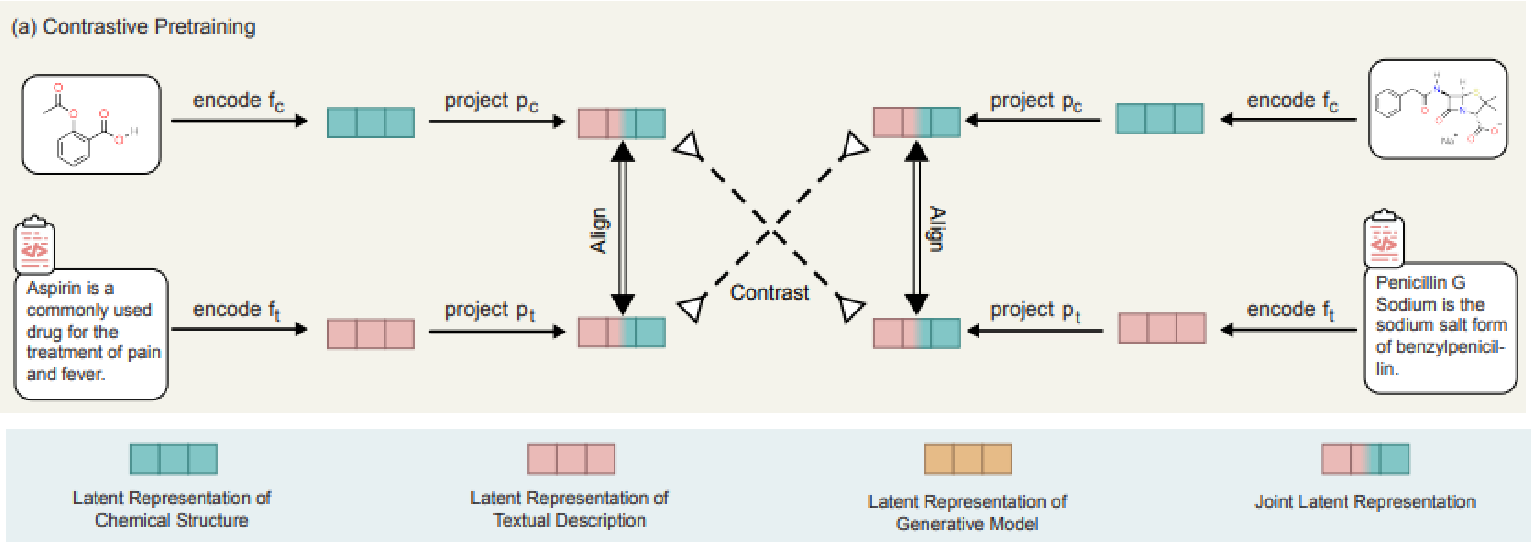
GNN+LLM-based model — Pre-training

- GNN or LLM-based

适用于GNN-centric或LLM-centric模型

- Alignment-based

适用于symmetric模型



目录

前景与挑战

数据和评估方面：
1、缺乏跨领域统一大规模数据集
2、开放式任务缺乏标签

模型方面：
1、骨干架构的设计非常关键
2、预训练时的pretext任务选取

应用方面：
1、是否能有类似大语言模型的突破
2、隐私泄露问题

未来可能的挑战


谢 谢！
浙江大学
ZHEJIANG UNIVERSITY