南京大学120周年校庆
120th ANNIVERSARY
NANJING UNIVERSITY
1902 - 2022

南京大学120周年校庆

120 anniversary of Nanjing University

2022-10-6

Convolutional Graph Neural Networks

A brief introduction

目录
CONTENT

王世奇 2022/10/6

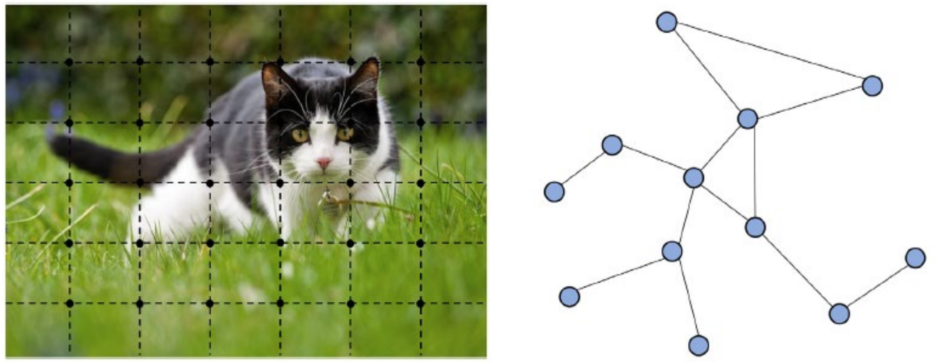
南京大学120周年校庆
120th ANNIVERSARY
NANJING UNIVERSITY
1902 - 2022

01 背景与框架

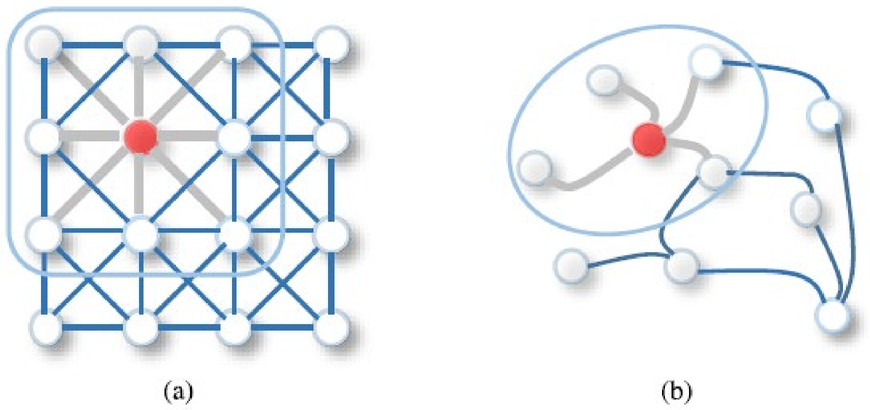
Background and Framework

What is Convolutional Graph Neural Networks?

- **Graph Neural Networks(GNNs):** deep learning based methods that operate on graph domain;
- **Convolutional GNNs(conv-GNNs):** the main idea is to generalize convolutions from other domain to the graph domain, it’s the mostly used kind of GNNs;



Images(left) are in Euclidean space V.S. graphs(right) in non-Euclidean space



2-D convolution(left) V.S. graph convolution(right)

I. Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
II. Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE transactions on neural networks and learning systems, 2020, 32(1): 4-24.

General Design Pipeline of GNNs

I. Find your graph

You may need to build the graph at first ; e.g. 语法树, 金融图谱......

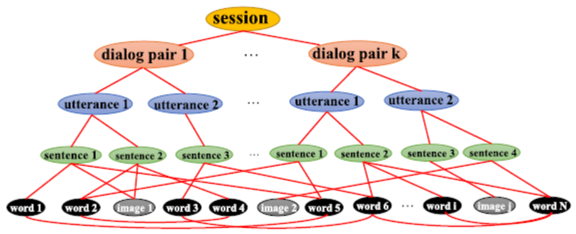


Fig1: 从对话系统构造的图

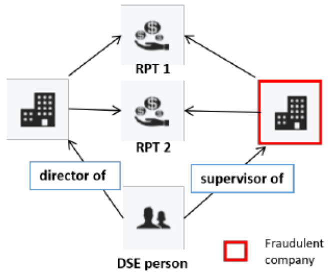


Fig2: 从上市公司构造的图

II. Specify graph type and scale

Graph type: homogeneous or heterogeneous?
Scale：is this big or not? Is sampling required?

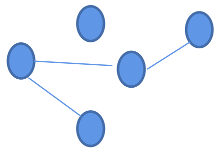


Fig3: homogeneous graph(同构图)

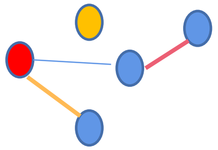


Fig4: heterogeneous graph(异构图)

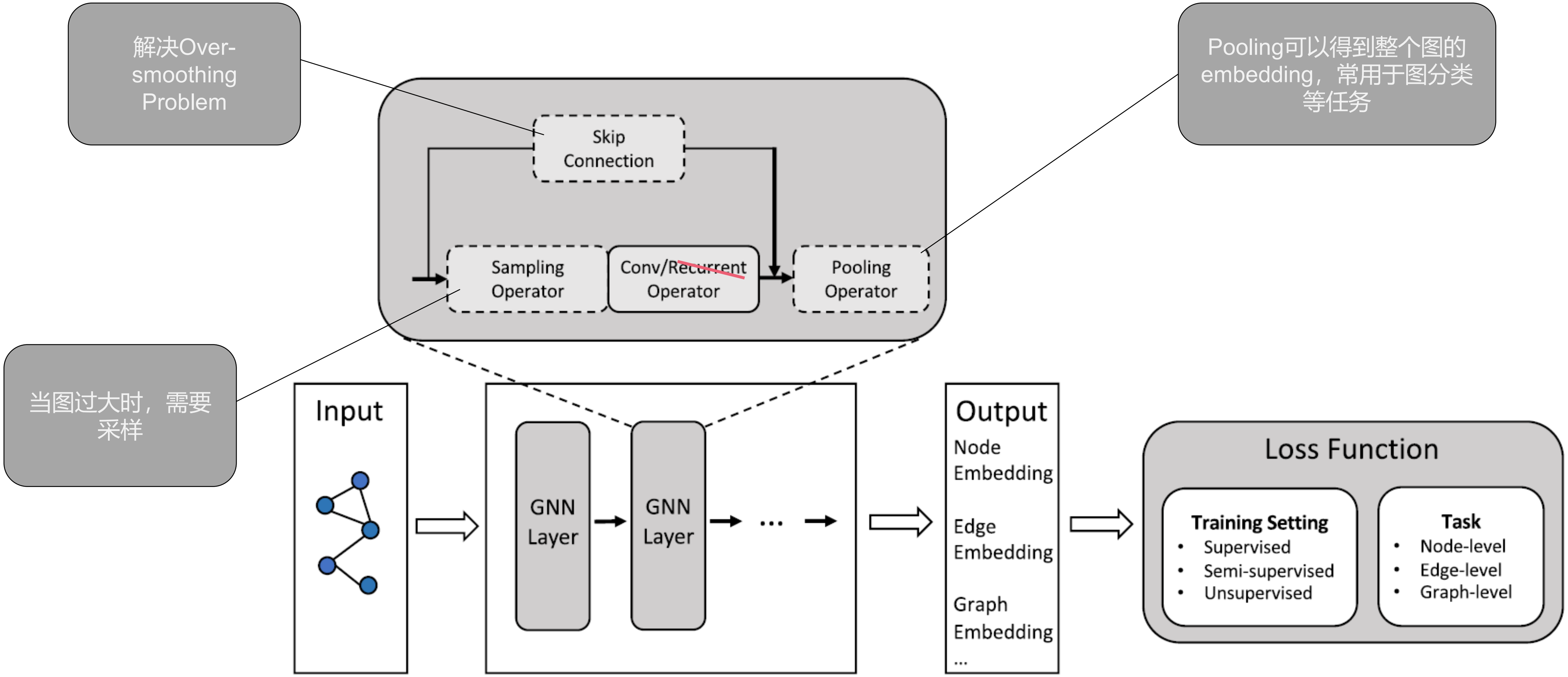
III. Specify your task and define loss function

General Tasks: Node classification; Graph classification; Link prediction.....

IV. Choose or building the model

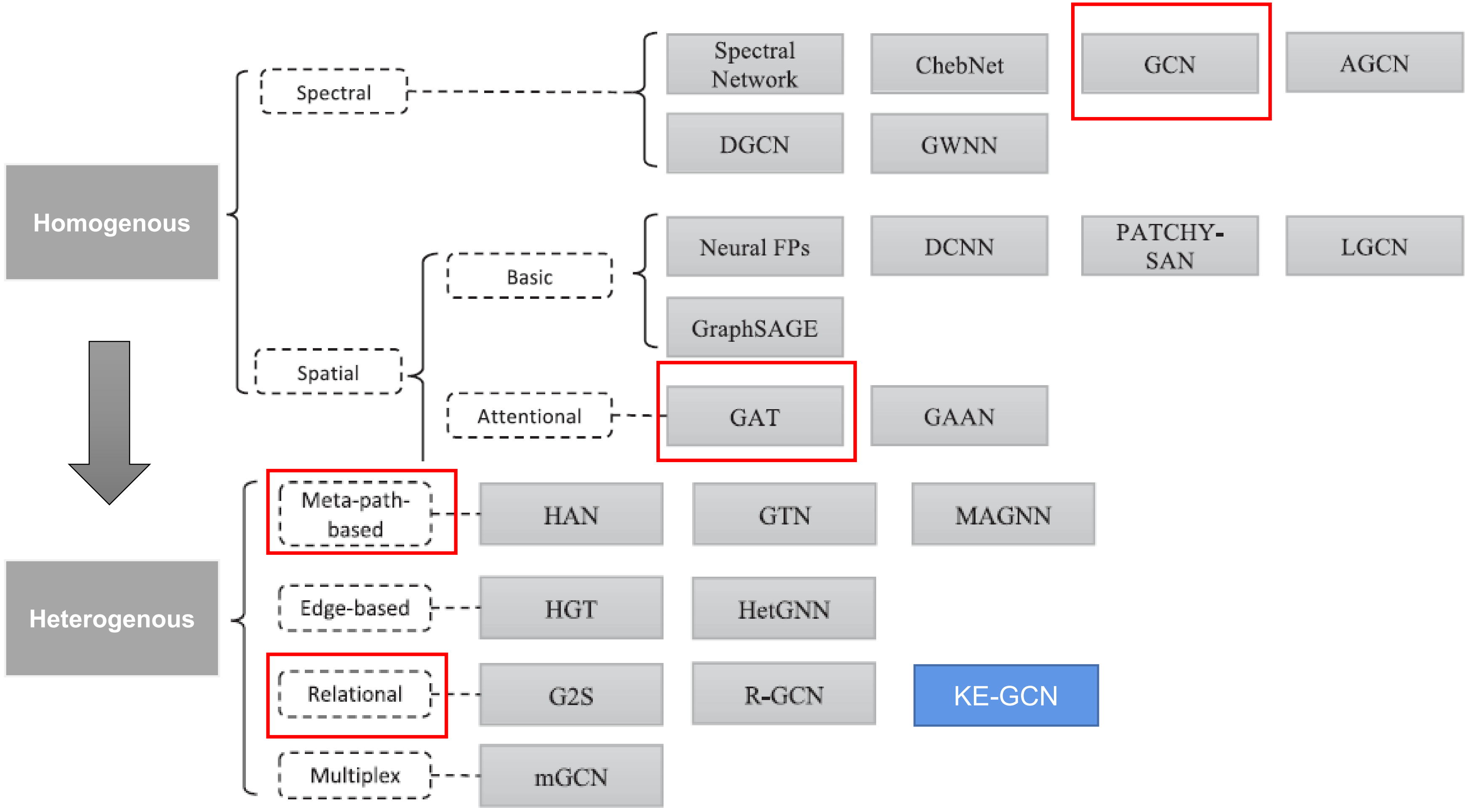
Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
Fig 1. Nie L, Jiao F, Wang W, et al. Conversational Image Search[J]. IEEE Transactions on Image Processing, 2021, 30: 7732-7743.

General Components of a GNNs model：



Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.

An overview of heterogeneous and homogeneous GNNs models:



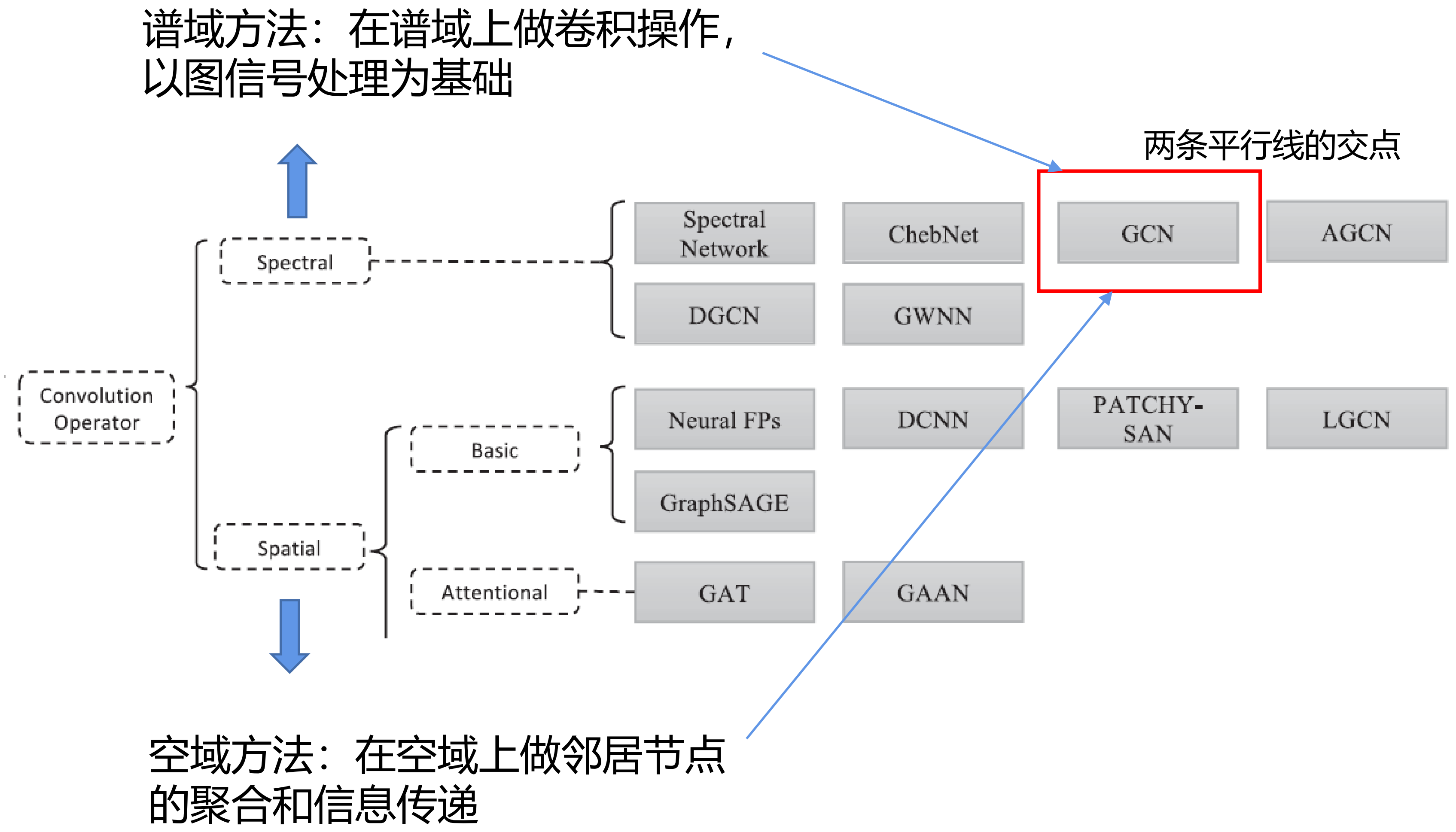
Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.

南京大学120周年校庆
120th ANNIVERSARY
NANJING UNIVERSITY
1902 - 2022

02 同构图卷积神经网络

Homogeneous conv-GNNs

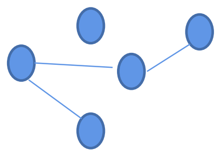
An overview of heterogeneous GNNs models：



Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.

Graph Convolutional Network(GCN[1])

Common Spectral Approach：



graph signal **X**

graph Fourier transform → Spectral Domain → Convolution Operation ★ →

$$\mathbf{g}_w \star \mathbf{x} = \mathbf{U}\mathbf{g}_w\mathbf{U}^T\mathbf{x}.$$

g_w: filter in the spectral domain；

U： the matrix of eigenvectors of the normalized graph Laplacian

Basic ideas： To design and simplify the filter g_w

GCN：

$$\mathbf{g}_w \star \mathbf{x} \approx w\left(\mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{x}.$$

防止D中出现0，并且

renormalization trick

$$\begin{cases} \mathbf{H} = \boxed{\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}}\mathbf{X}\mathbf{W} \\ \tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij} \\ \tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N \end{cases}$$

对邻接矩阵做row normalization

例如：

1	0	1
0	1	0
1	0	1

Row normalization →

0.5	0	0.5
0	1	0
0.5	0	0.5

I_N: 单位矩阵

A: 图的邻接矩阵

D: 度矩阵，对角线元素为每个节点的度，其他为0

1: Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. Proceedings of ICLR.

Graph Convolutional Network(GCN)

Understand GCN in a Spatial way:

$$H = f(\bar{A}XW)$$

等价于 ↓

$$\mathbf{h}_v = f\left(\mathbf{W}^T\left(\boxed{\sum_{u \in N(v) \cup v} \bar{A}_{v,u}\mathbf{x}_u}\right)\right) \quad \forall v \in V.$$

邻居节点信息聚合

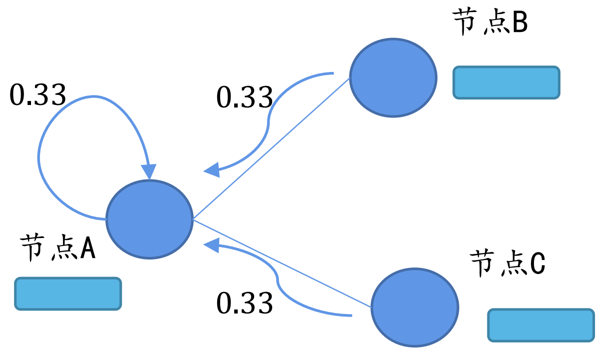
卷积操作的filter，共享参数

- N(v) 代表节点v 的所有邻居节点；
- V是所有节点的集合

An examples of GCN message passing:

A	B	C
1	1	1
1	1	0
1	0	1

邻接矩阵



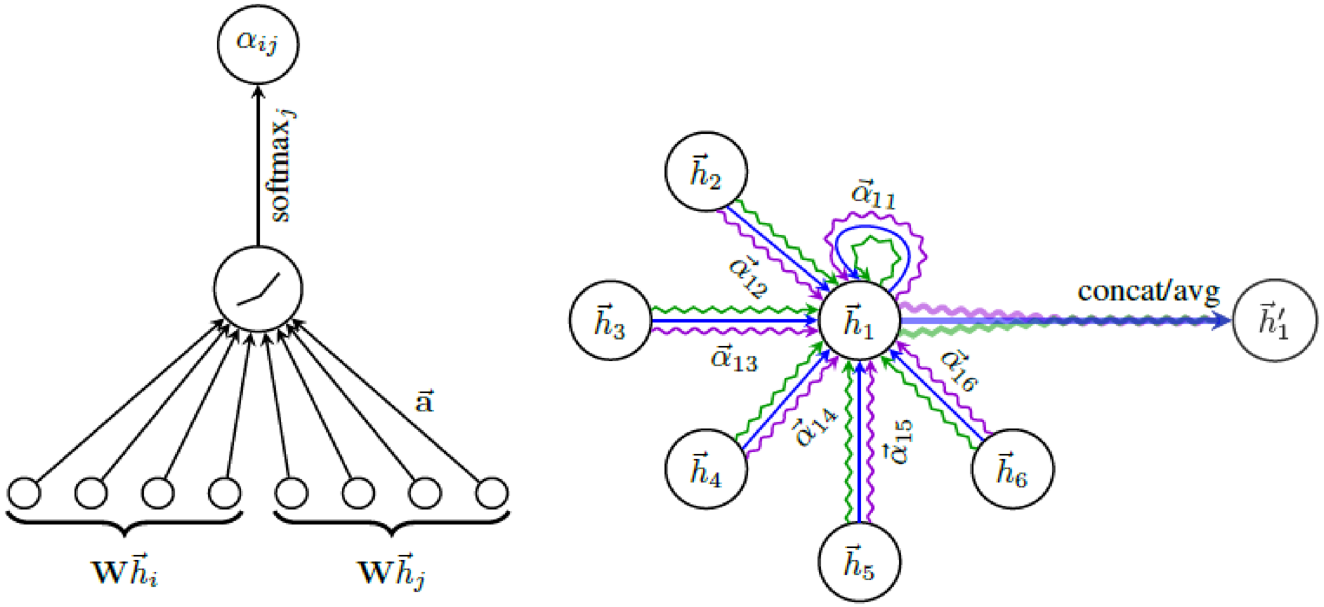
节点A的更新过程

Predefined weight!

1: Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. Proceedings of ICLR.2017.

Graph Attention Networks(GAT[1])

GCN: Predefined weight → GAT: Use Attention to learn weight



Attention calculation in GAT(left); Multi-head attention mechanism in GAT(right)

不再使用全局信息(邻接矩阵),因此适用于 inductive learning tasks

按照attention权重邻居节点信息聚合

$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\boxed{\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k\mathbf{W}^k\vec{h}_j}\right)$$

N_i stands for all neighbors of node I,
K is the number of head,
α stands for attention value

1: Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[C] Proceedings of ICLR. 2018.

Others

Skip Connection —— 残差连接

- Highway GCN. [I]
- Jump Knowledge Network. [II]
- DeepGCNs. [III]

Sampling Modules —— 采样

- GraphSAGE (node sampling) [IV]
- FASTGCN (layer sampling) [V]
- GraphSAINT (subgraph sampling) [VI]

Pooling Modules —— 池化

Direct pooling
Hierarchical pooling

I. Rahimi, A., Cohn, T., Baldwin, T., 2018. Semi-supervised user geolocation via graph convolutional networks. In: Proceeding of ACL
II. Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., Jegelka, S., 2018. Representation learning on graphs with jumping knowledge networks. In: Proceeding of ICML,
III. Li, G., Muller, M., Thabet, A., Ghanem, B., 2019a. Deepgcns: can gcns go as deep as cnns?.In: Proceedings of ICCV, pp. 9267–9276.
IV. Hamilton, W.L., Ying, Z., Leskovec, J., 2017a. Inductive representation learning on large graphs. In: Proceedings of NIPS, pp. 1024–1034.
V. Chen, J., Ma, T., Xiao, C., 2018b. Fastgcn: fast learning with graph convolutional networks via importance sampling. Proceedings of ICLR.
VI. Zeng, H., Zhou, H., Srivastava, A., Kannan, R., Prasanna, V.K., 2020. Graphsaint: graph sampling based inductive learning method. In: Proceedings of ICLR.

南京大学120周年校庆
120th ANNIVERSARY
NANJING UNIVERSITY
1902 - 2022

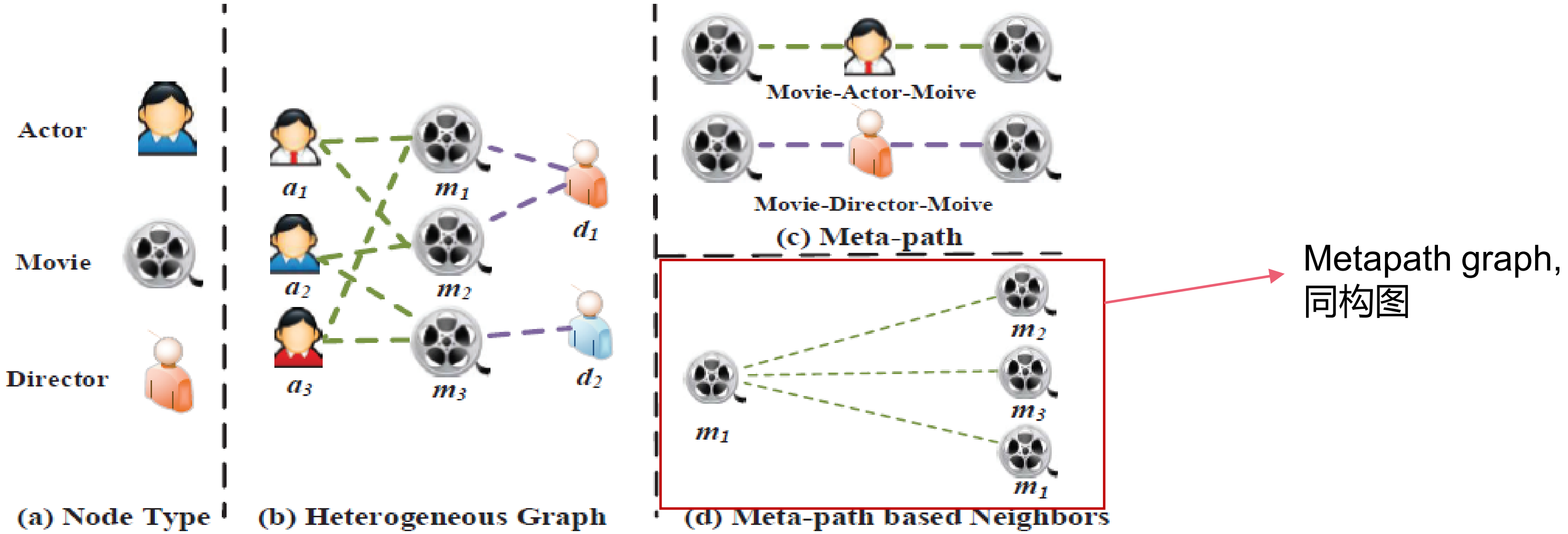
03 异构图卷积神经网络

Heterogeneous conv-GNNs

Heterogeneous Graph Attention Network(HAN[1])

GAT: Homogeneous graph → HAN: Heterogenous Graph

meta-paths: 一个预先设定好的路径，代表了异构图上的一种可能对问题有帮助的关系
例如：

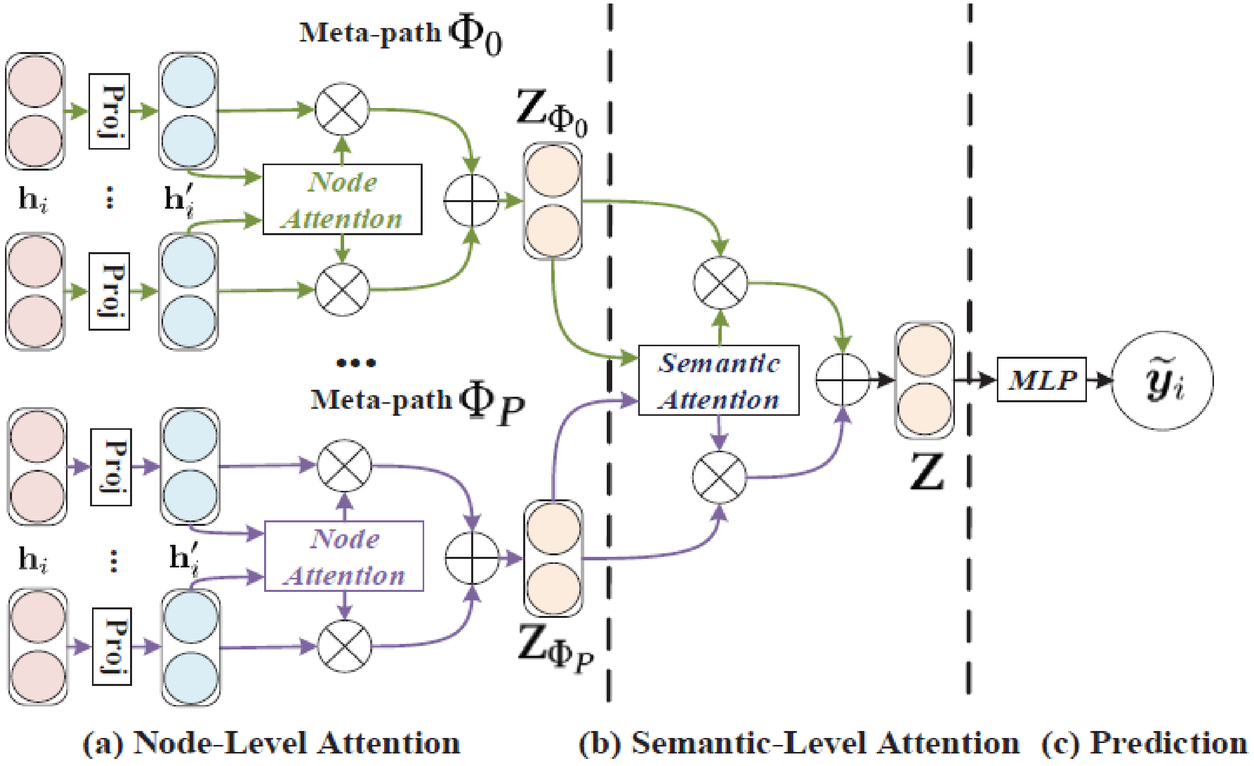


电影图谱中可能的meta-paths

1. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S., 2019a. Heterogeneous graph attention network. In: Proceedings of WWW, pp. 2022–2032.

Heterogeneous Graph Attention Network(HAN[1])

HAN: Do node-level attention between meta-path neighbors and then do semantic level attention to aggregate different meta-path info



The overall framework of HAN

However:

- Metapath needs experts knowledge;
- This two-stage process makes the results easily effected by different meta-paths
- It did NOT make use of Intra-path information

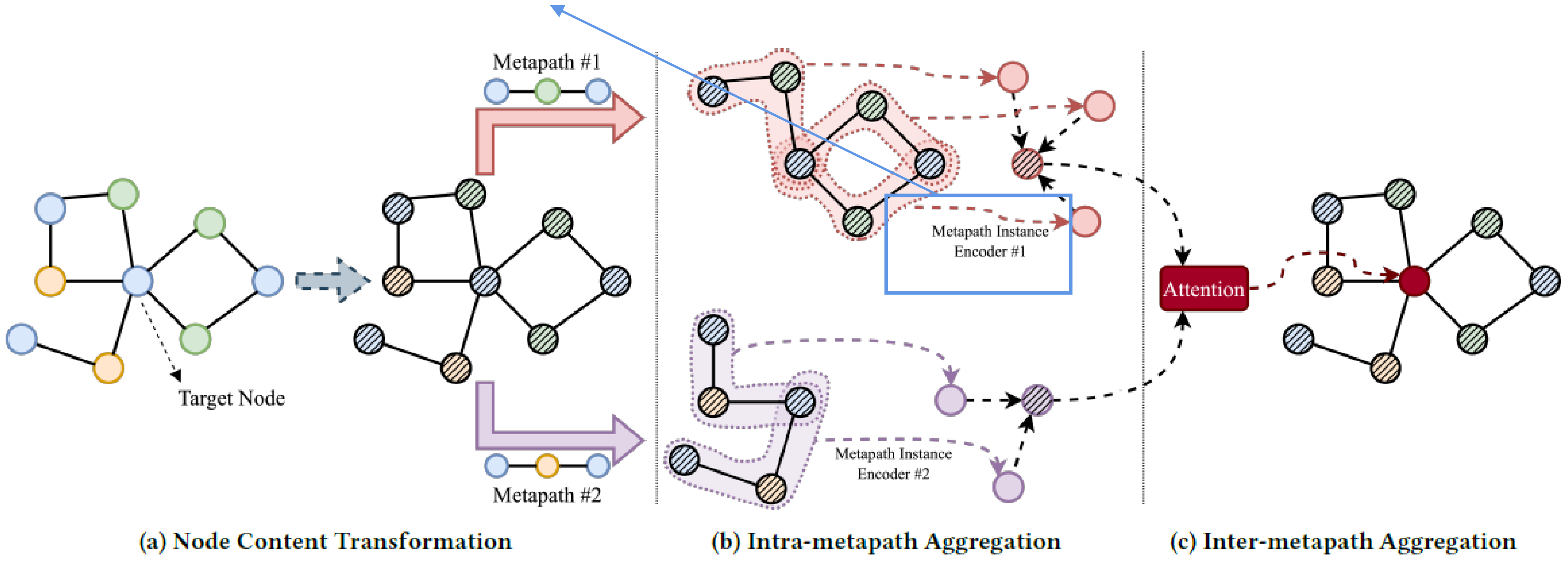
1. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S., 2019a. Heterogeneous graph attention network. In: Proceedings of WWW, pp. 2022–2032.

Metapath Aggregated Graph Neural Network(MAGNN[1])

HAN : inter-path aggregation → MAGNN: intra and inter-path aggregation

Issues：To use intra-metapath node representations

MAGNN：Metapath Instance Encoder（Mean / W*Mean / Relational Rotation）



The overall architecture of MAGNN

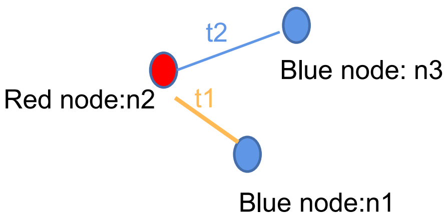
1. Fu, X., Zhang, J., Meng, Z., King, I., Magnn, 2020. Metapath aggregated graph neural network for heterogeneous graph embedding. Proceedings of WWW 2331–2341.

Graph Transformer Network(GTN[1])

HAN : predefined meta-path → GTN: learn meta-paths by the model

一个重要的观察：Given two adjacency matrices Q1 and Q2, the meta-path adjacency matrix is computed by **matrix multiplication**

例如：



	n1	n2	n3
	0	1	0
	1	0	0
	0	0	0

Adjacency matrix for t1

X

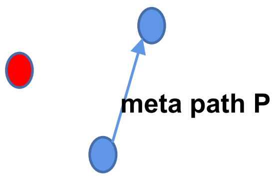
	n1	n2	n3
	0	0	0
	0	0	1
	0	1	0

Adjacency matrix for t2

→

	n1	n2	n3
n1	0	0	1
n2	0	0	0
n3	0	0	0

Adjacency matrix for meta path P:
bluenode-t1-rednode-t2-bluenode



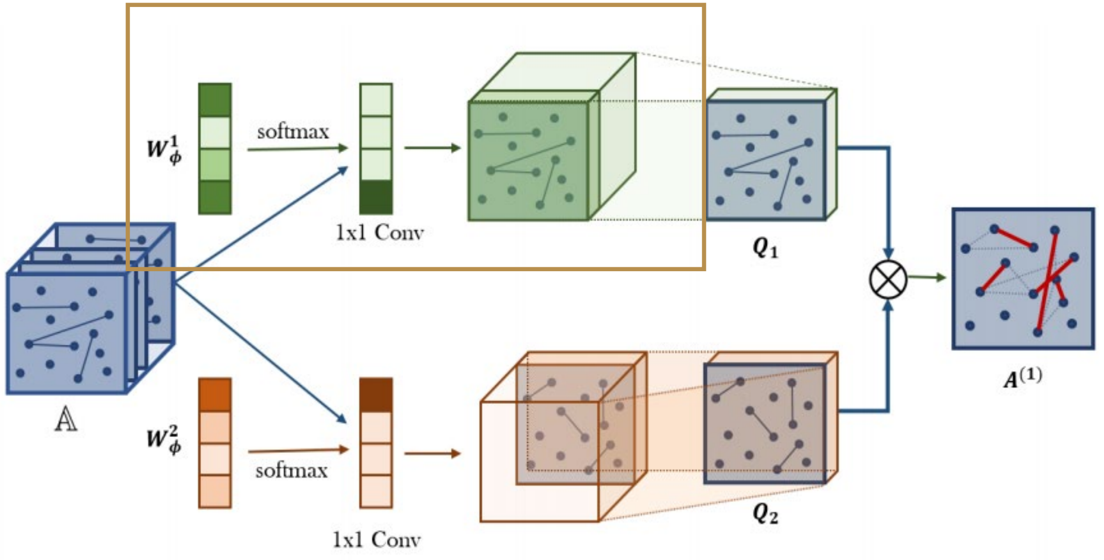
So, the adjacency matrix of arbitrary length ℓ meta-paths can be calculated by:

$$A_P = \left(\sum_{t_1 \in \mathcal{T}^e} \alpha_{t_1}^{(1)} A_{t_1}\right)\left(\sum_{t_2 \in \mathcal{T}^e} \alpha_{t_2}^{(2)} A_{t_2}\right) \ldots \left(\sum_{t_l \in \mathcal{T}^e} \alpha_{t_l}^{(l)} A_{t_l}\right)$$

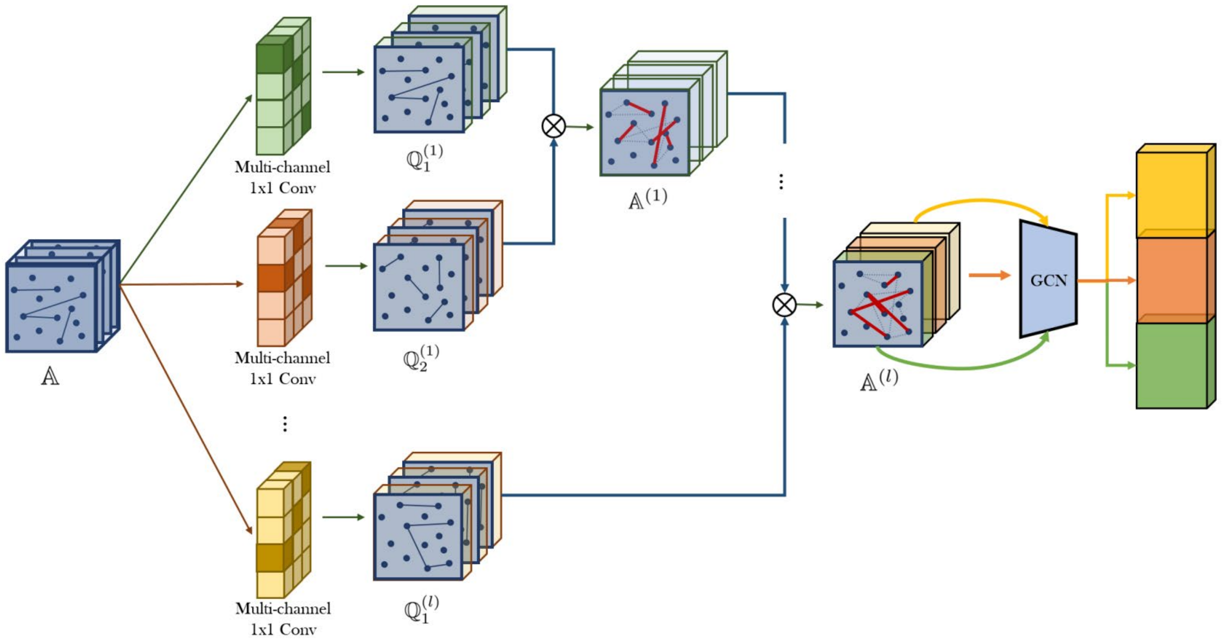
1. Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J., 2019. Graph transformer networks. In: Proceedings of NeurIPS, pp. 11983–11993

Graph Transformer Network(GTN[1])

GTN: **Softly select** adjacency matrix and then do multiplication



Graph transformer layer



Graph Transformer Networks (GTNs)

1. Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J., 2019. Graph transformer networks. In: Proceedings of NeurIPS, pp. 11983–11993

Relation based heterogeneous GNNs —— RGCN

- edges and semantic information might be more important
- and their type numbers might be too large for meta-path based methods.

(e.g. **Knowledge Graph**)

Relational Graph Convolutional Networks (RGCN[1]):

Neighbor aggregation of particular relation

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right)$$

Normalization weight

Sum up aggregated information from different relations

R : relation set
N_i^r :Neighbors of node i with relation r
σ: activation function

1. Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M., 2018. Modeling relational data with graph convolutional networks. In: Proceedings of ESWC. Springer, pp. 593–607.

Relation based heterogeneous GNNs —— WGCN

Weighted Graph Convolutional Networks (WGCN[1]):

Take relation directions into consideration

$$\mathbf{m}_v^{l+1} = \boxed{\sum_{(u,r)\in\mathcal{N}_{\text{in}}(v)} W^l(\alpha_r^l \mathbf{h}_u^l) + \sum_{(u,r)\in\mathcal{N}_{\text{out}}(v)} W^l(\alpha_r^l \mathbf{h}_u^l)}$$

$$\mathbf{h}_v^{l+1} = \sigma(\mathbf{m}_v^{l+1} + W^l \mathbf{h}_v^l)$$

relation-specific learnable parameter

$N_{in}(v)$:Neighbors of node i with an in link

σ： activation function

$N_{\text{out}}(v)$:Neighbors of node i with an out link

1. Shang C, Tang Y, Huang J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 3060-3067.

Knowledge Embedding Based Graph Convolutional Networks (KE-GCN[1]):

Vanilla GCN (spatial, omit the normalization):

$$\mathbf{m}_v^{l+1} = \sum_{u \in \mathcal{N}(v)} \mathbf{h}_u^l \quad (1)$$

$$\mathbf{h}_v^{l+1} = \sigma(W^l \boxed{(\mathbf{m}_v^{l+1} + \mathbf{h}_v^l}))) \quad (2)$$

Def score function

$$f(\mathbf{h}_u, \mathbf{h}_v) = \mathbf{h}_u^T \mathbf{h}_v$$

$$\boxed{\mathbf{m}_v^{l+1} = \sum_{u \in \mathcal{N}(v)} \frac{\partial f(\mathbf{h}_u^l, \mathbf{h}_v^l)}{\partial \mathbf{h}_v^l} = \frac{\partial(\sum_{u \in \mathcal{N}(v)} f(\mathbf{h}_u^l, \mathbf{h}_v^l))}{\partial \mathbf{h}_v^l}} \quad (3)$$

$m_v^{l+1} + h_v^l$ can be seen as one step gradient ascent to maximize the sum of scoring function f

We can use **score function in Knowledge graph** to give GCN the ability for heterogeneous graph !!!

N(V) : the set of immediate neighbors of node v

h_v^l: the embedding of node v at layer ℓ

m_v^{l+1}: the aggregated representation of neighbors

σ： activation function

1. Yu D, Yang Y, Zhang R, et al. Knowledge embedding based graph convolutional network[C]//Proceedings of the WWW 2021. 2021: 1619-1628.

Knowledge Embedding Based Graph Convolutional Networks (KE-GCN[1]):

KE-GCN : Use score function (f) in Knowledge embedding to help GNN

$$\mathbf{m}_v^{l+1} = \sum_{(u,r)\in\mathcal{N}_{\text{in}}(v)} W_r^l \frac{\partial f_{\text{in}}(\mathbf{h}_u^l, \mathbf{h}_r^l, \mathbf{h}_v^l)}{\partial \mathbf{h}_v^l} \quad (4)$$

$$+ \sum_{(u,r)\in\mathcal{N}_{\text{out}}(v)} W_r^l \frac{\partial f_{\text{out}}(\mathbf{h}_v^l, \mathbf{h}_r^l, \mathbf{h}_u^l)}{\partial \mathbf{h}_v^l} \quad (5)$$

$$\mathbf{h}_v^{l+1} = \sigma_{\text{ent}}(\mathbf{m}_v^{l+1} + W_0^l \mathbf{h}_v^l) \quad (6)$$

(a).Node representation updating

$$\mathbf{m}_r^{l+1} = \sum_{(u,v)\in\mathcal{N}(r)} \frac{\partial f_r(\mathbf{h}_u^l, \mathbf{h}_r^l, \mathbf{h}_v^l)}{\partial \mathbf{h}_r^l} \quad (7)$$

$$\mathbf{h}_r^{l+1} = \sigma_{\text{rel}}(W_{\text{rel}}^l(\mathbf{m}_r^{l+1} + \mathbf{h}_r^l)) \quad (8)$$

(b).Relation representation updating

N(V) : the set of immediate neighbors of node v

h_v^l: the embedding of node v at layer ℓ

h_r^l: the embedding of relation r at layer ℓ

m_v^{l+1}: the aggregated representation of neighbors

1. Yu D, Yang Y, Zhang R, et al. Knowledge embedding based graph convolutional network[C]//Proceedings of the WWW 2021. 2021: 1619-1628.

Knowledge Embedding Based Graph Convolutional Networks (KE-GCN[1]):

KE-GCN is powerful:

- *R-GCN can be fully recovered by KE-GCN when*

1) $f_{in}(h_u^l, h_r^l, h_v^l) = f_{out}(h_v^l, h_r^l, h_u^l)$ $=(h_u^l)^T\ h_v^l$ *; and 2)* $h_l^r\ = 0$*(no relation embedding)*

- *W-GCN can be fully recovered by KE-GCN when*

1) $f_{in}(h_u^l, h_r^l, h_v^l) = f_{out}(h_v^l, h_r^l, h_u^l)$ $=(h_u^l)^T\ h_v^l$ *; and 2)* $W_r^l\ = W^l \alpha_l^r$*; and 3)* $h_l^r\ = 0$*(no relation embedding)*

- *CompGCN can be fully recovered by KE-GCN when*

  ……

KE-GCN is powerful yet more complex and hard to train:

You have to maintain **a full triples set** (like in KG but can **NOT** do sampling)

1. Yu D, Yang Y, Zhang R, et al. Knowledge embedding based graph convolutional network[C]//Proceedings of the WWW 2021. 2021: 1619-1628.

Others

Edge based

- HetGNN. [I] (KDD2019)

Use Random walk to generate neighbors rather than meta-path

Zhang, C., Song, D., Huang, C., Swami, A., Chawla, N.V., 2019b. Heterogeneous graph neural network. In: Proceedings of KDD, pp. 793–803.

- HGT. [II] (WWW 2020)

Transformer => Graph domain

Hu, Z., Dong, Y., Wang, K., Sun, Y., 2020a. Heterogeneous graph transformer. In: Proceedings of WWW, pp. 2704–2710.

……

完结，撒花

Question & Answer

Feel free to ask any questions, I will try my best to help you!

谢谢观看

诚耀百廿 雄创一流