# 图神经网络导论

## 富信息图神经网络


授课教师：周晟

浙江大学　软件学院

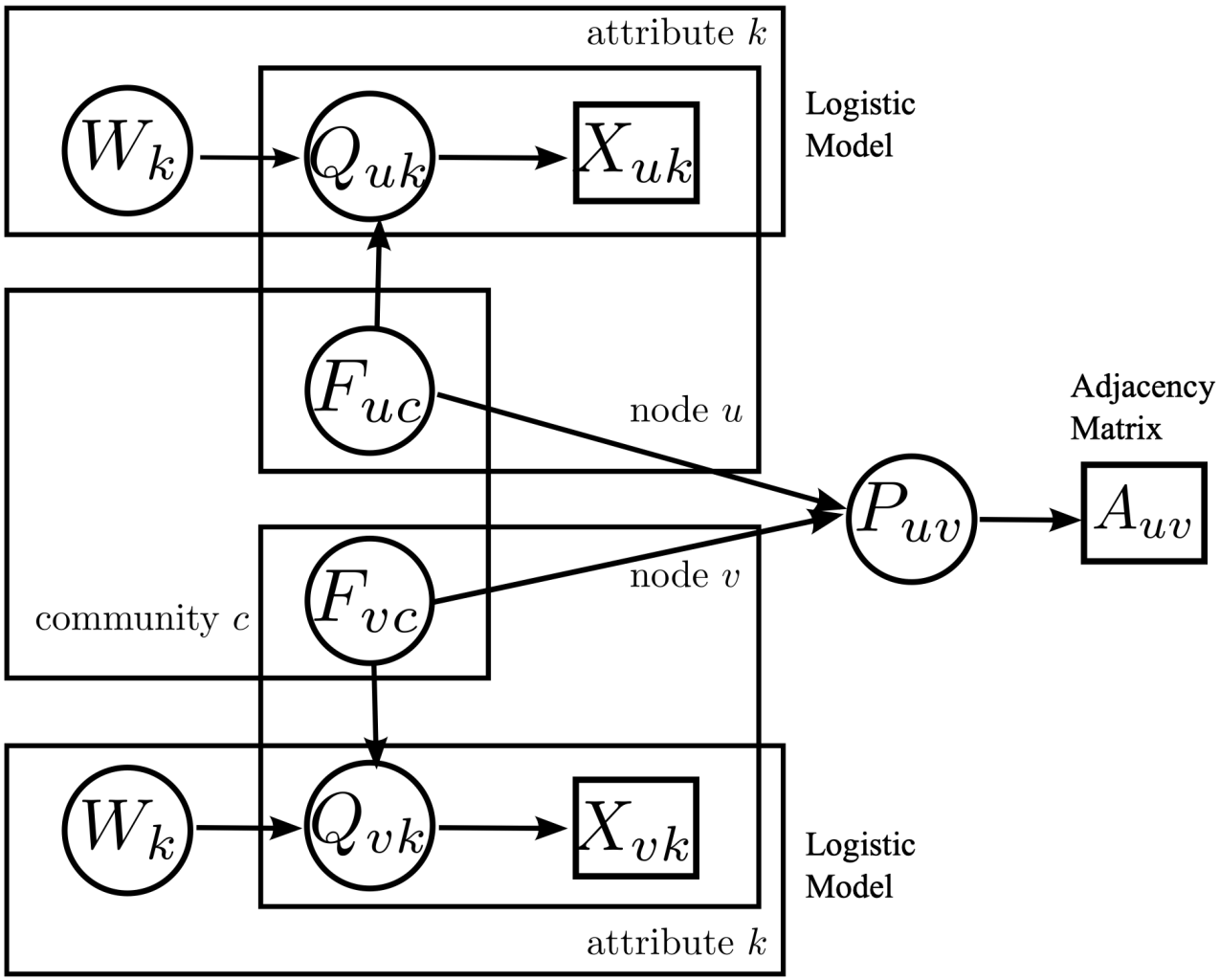

2021.12

# 课程内容



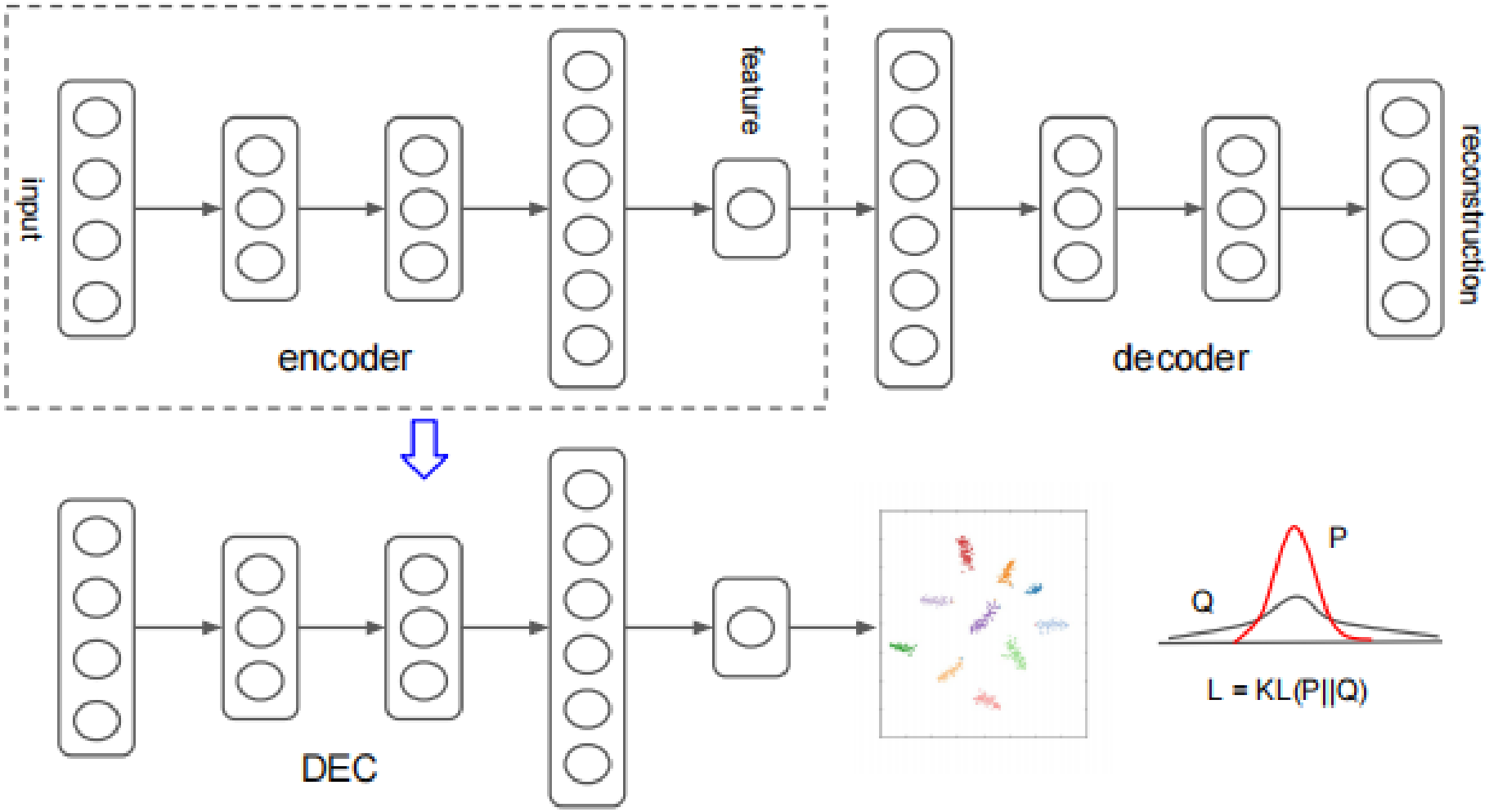
# 上节课回顾

深度社区发现：



# 上节课回顾

基于自训练的社区发现算法：



# 富信息网络

## 图的数学定义

图一般定义为：$G = \{V, E, X\}$，其中 $V$ 是节点的集合，$E$ 是边的集合，$X$ 是节点属性的集合。

## 富信息网络

富信息网络是指网络结构、节点属性、边属性等信息丰富的网络。大多数真实世界的网络都是富信息网络

# 富信息网络

计算机网络

社交网络

交通网络

神经元网络

卫星网络

# 富信息图神经网络

## 困难与挑战

1. 传统图神经网络难以充分建模富信息网络特点
2. 不同类型富信息网络需要设计不同的图神经网络结构

## 常见的富信息图神经网络

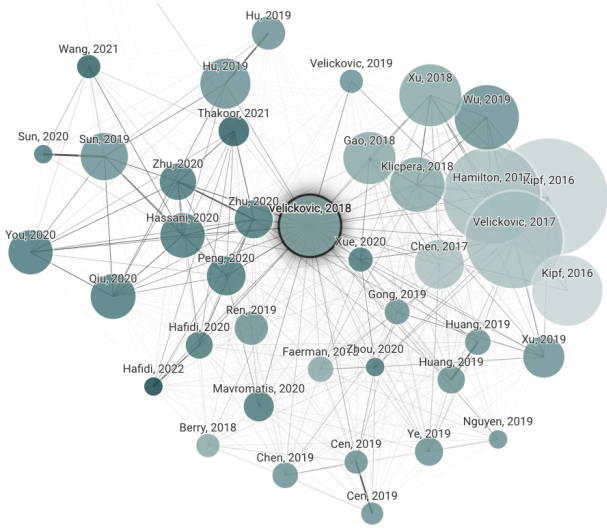
1. 有向图
2. 异构图
3. 动态图

# 常见的有向网络

## 有向网络/图

有向网络（Directed Network/Graph）是指边带有方向的网络。有向网络中的边除了包含临近信息外，往往还包含层次或语义信息。

社交网络




引用网络

# 有向图神经网络的挑战

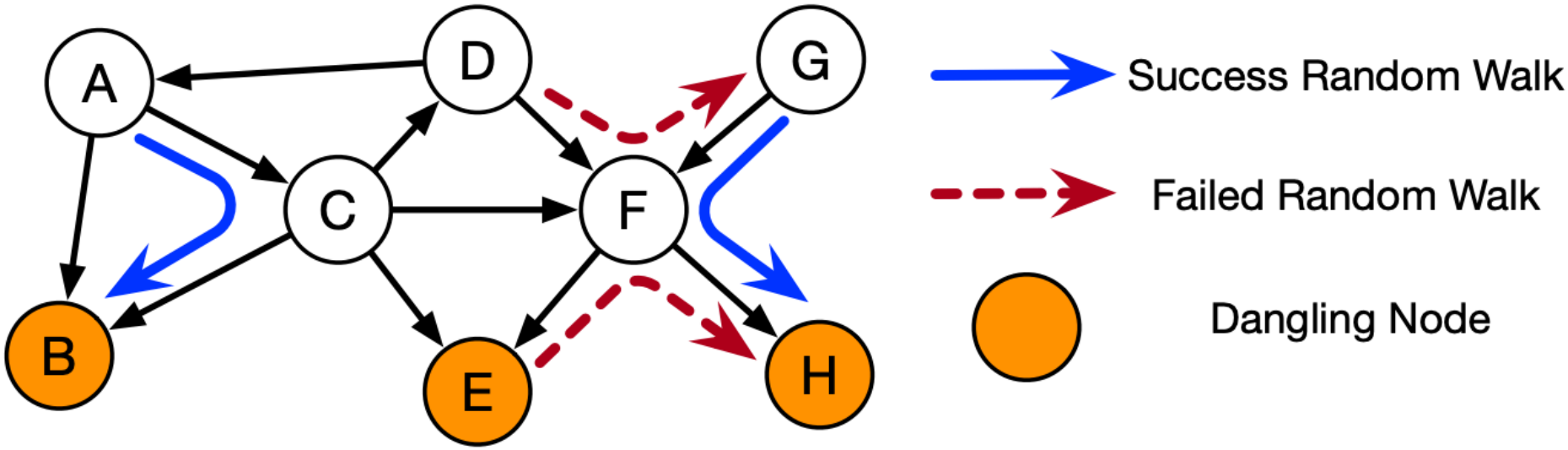
## 有向图的特点

1. 节点间临近关系非对称
2. 网络中存在层次性等特殊结构

## 有向图神经网络的难点

1. 邻接矩阵非对称
2. 邻居关系多样化
3. 有向特征难以捕获

# Direction-Aware User Recommendation Based on Asymmetric Network Embedding[1]

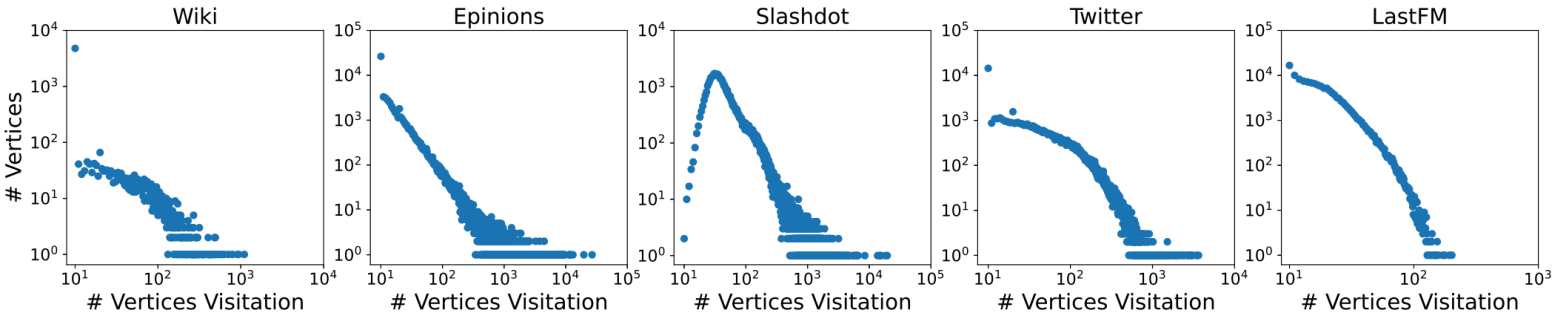



有向网络中的随机游走

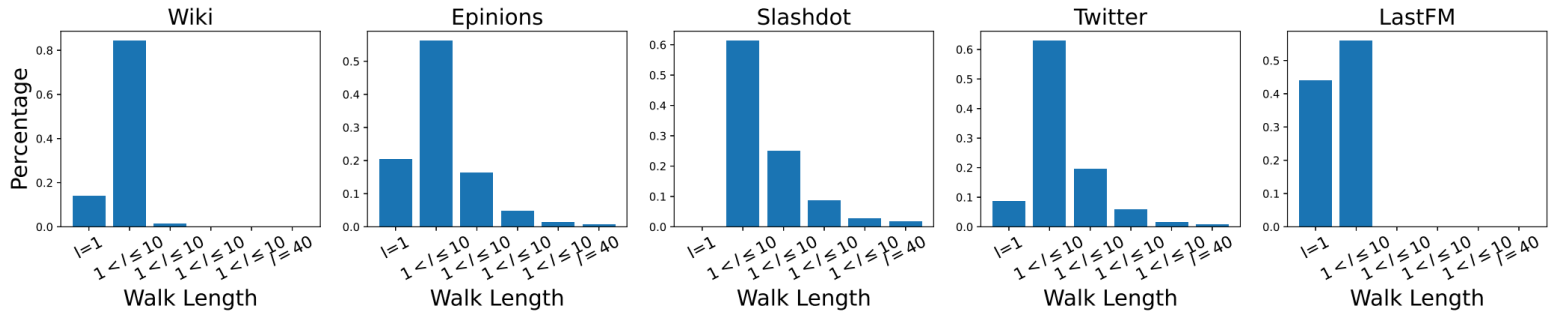
[1]Direction-Aware User Recommendation Based on Asymmetric Network Embedding[Zhou et al., 2021]

# Direction-Aware User Recommendation Based on Asymmetric Network Embedding

在有向网络中使用传统随机游走的问题：




节点被访问次数问题




随机游走长度问题

# InfoWalk 随机游走

- 每一步随机游走时忽略边的方向

$$P\left(\mathcal{R}_{v_i}^{k+1}=b \mid \mathcal{R}_{v_i}^{k}=a\right)=\begin{cases}\frac{1}{d_a^{\text{out}}+d_a^{\text{in}}} & \mathrm{E}_{ab}=1 \text{ or } \mathrm{E}_{ba}=1\\ 0 & \text{otherwise}\end{cases}$$

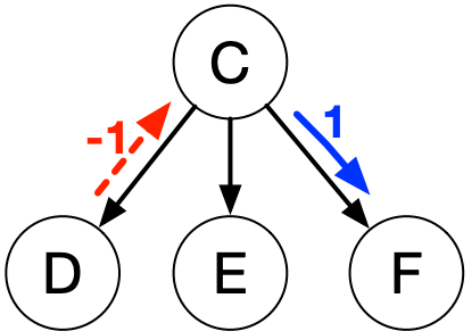
- 每一步随机游走时边的方向使用参数记录信息

$$r_{i,i+1}=\begin{cases}1 & \text{if } \mathrm{E}_{i,i+1}=1 \text{ and } \mathrm{E}_{i+1,i}=0\\ -1 & \text{if } \mathrm{E}_{i,i+1}=0 \text{ and } \mathrm{E}_{i+1,i}=1\\ 0 & \text{if } \mathrm{E}_{i,i+1}=1 \text{ and } \mathrm{E}_{i+1,i}=1\end{cases}$$

- 随机游走得到带参序列 $\mathcal{R}_{v_i}: v_i \xrightarrow{r_{i,j}} v_j \xrightarrow{r_{j,j+1}} \ldots \xrightarrow{r_{k-1,k}} v_k$

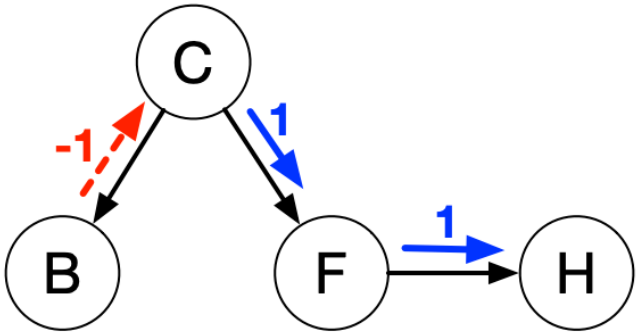
$$s_{i,i+k}=\frac{1}{k}\sum_{j=i}^{i+k-1} r_{j,j+1}$$

# InfoWalk 随机游走



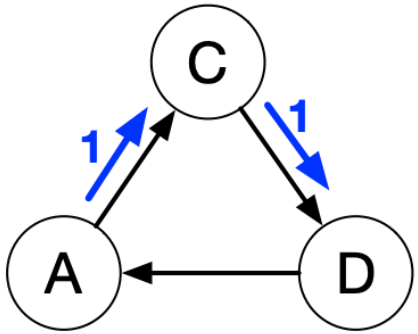

$S_{D\to C\to F}=0$

**(a)**



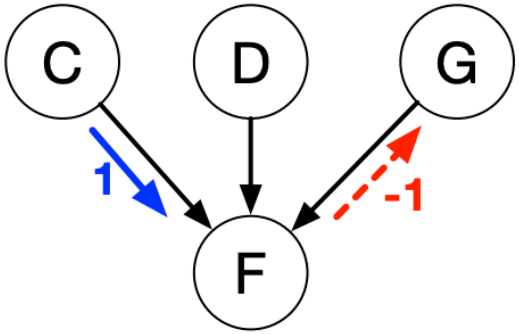

$S_{B\to C\to F\to H}=1$

**(c)**



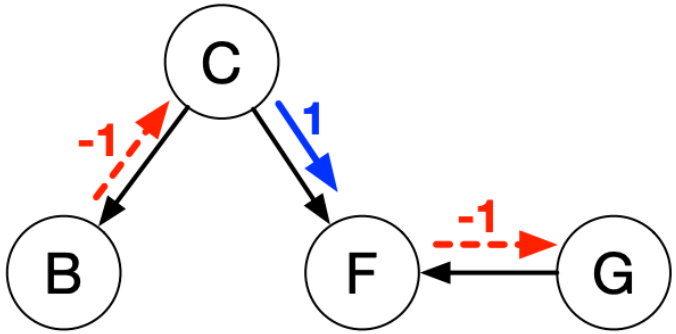

$S_{A\to C\to D}=2$

**(e)**



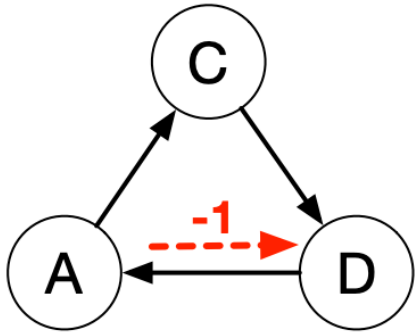

$S_{C\to F\to G}=0$

**(b)**




$S_{B\to C\to F\to G}=-1$

**(d)**

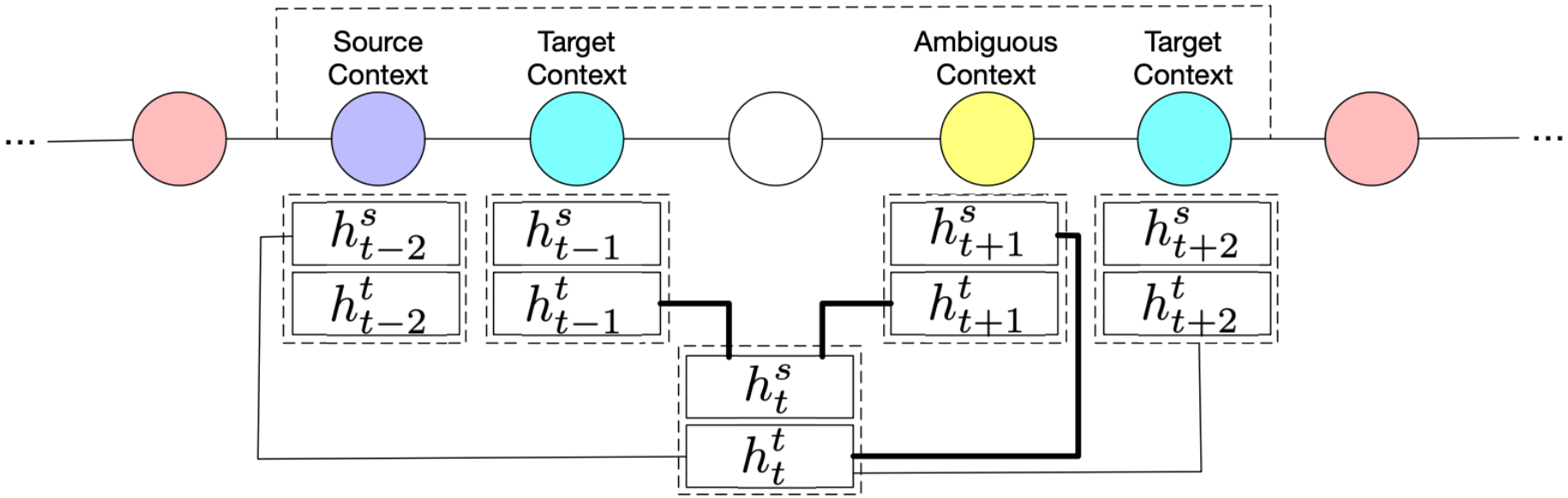



$S_{A\to D}=-1$

**(f)**

InfoWalk 随机游走

# InfoWalk 随机游走




基于 InfoWalk 的节点分类

$$\max_{\mathrm{H}^s,\mathrm{H}^t} \sum_{u\in \mathrm{V}} \sum_{v\in DC_u} \log P\left(v \mid u, s_{u,v}\right)$$

# 基于 InfoWalk 的节点表征学习

- **定性学习**

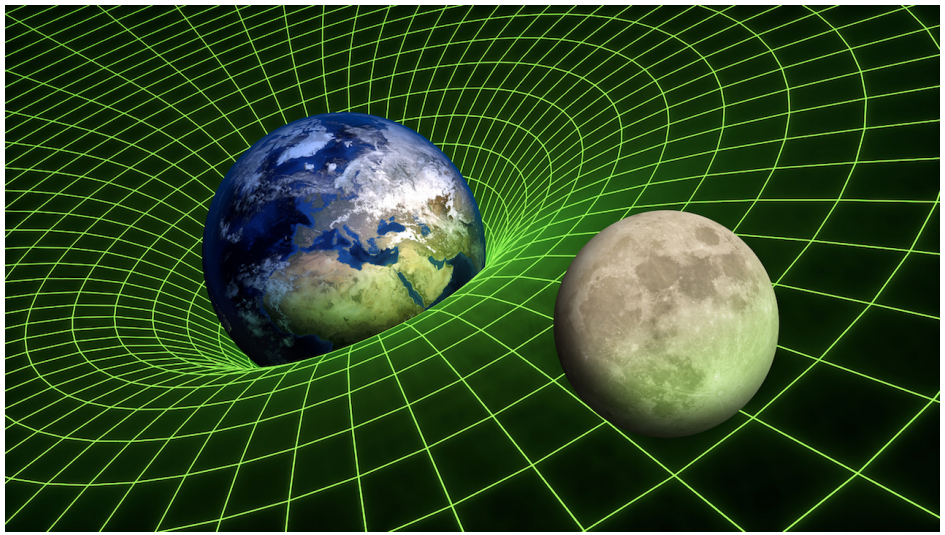
$$P\left(v \mid u, s_{u,v} > 0\right) = \frac{\exp\left(h_u^s \cdot h_v^t\right)}{\sum_{k \in V} \exp(h_u^s \cdot h_v^t)}$$
$$P\left(v \mid u, s_{u,v} < 0\right) = \frac{\exp\left(h_v^s \cdot h_u^t\right)}{\sum_{k \in V} \exp(h_v^s \cdot h_u^t)}$$
$$P\left(v \mid u, s_{u,v} = 0\right) = \frac{\exp\left(h_v^s \cdot h_u^t + h_u^s \cdot h_v^t\right)}{\sum_{k \in V} \exp(h_v^s \cdot h_u^t + h_u^s \cdot h_v^t)}$$

- **定量学习**

$$\pi_{u,v} = \log\left(\frac{s_{u,v}+1}{v-u} + b\right)$$

$$\max_{\mathrm{H}^s, \mathrm{H}^t} \sum_{u \in \mathrm{V}} \sum_{v \in DC_u} \log P\left(v \mid u, \pi_{u,v}\right) = \log \frac{\pi_{u,v} \cdot \exp\left(h_u^s \cdot h_v^t\right)}{\sum_{k \in V} \exp\left(h_u^s \cdot h_v^t\right)}$$

# Gravity-Inspired Graph Autoencoders for Directed Link Prediction[2]



万有引力理论

$$F = \frac{Gm_1m_2}{r^2}, \quad a_{1\to 2} = \frac{F}{m_1} = \frac{Gm_2}{r^2}, \quad a_{2\to 1} = \frac{F}{m_2} = \frac{Gm_1}{r^2}$$

[2]Gravity-Inspired Graph Autoencoders for Directed Link Prediction[Salha et al., 2019]

# Gravity-Inspired Graph Autoencoders for Directed Link Prediction

**基于万有引力的有向边生成模型**（Decoder）：

$$\begin{aligned}\hat{A}_{ij} &= \sigma\left(\log a_{i\to j}\right) \\ &= \sigma(\underbrace{\log G m_j}_{\tilde{m}_j} - \log \|z_i - z_j\|_2^2)\end{aligned}$$

$$p\left(A_{ij}=1 \mid z_i, z_j, \tilde{m}_j\right) = \sigma\left(\tilde{m}_j - \log \|z_i - z_j\|_2^2\right)$$

$$p(A \mid Z, \tilde{M}) = \prod_{i=1}^{n}\prod_{j=1}^{n} p\left(A_{ij} \mid z_i, z_j, \tilde{m}_j\right)$$

# Gravity-Inspired Graph Autoencoders for Directed Link Prediction

基于图神经网络的参数推断（Encoder）：

$$q((Z, \tilde{M}) \mid A, X) = \prod_{i=1}^{n} q\left((z_i, \tilde{m}_i) \mid A, X\right)$$

使用高斯分布建模：

$$q\left((z_i, \tilde{m}_i) \mid A, X\right) = \mathcal{N}\left((z_i, \tilde{m}_i) \mid \mu_i, \mathrm{diag}\left(\sigma_i^2\right)\right)$$

使用图神经网络建模参数：

$$\mu = \mathrm{GCN}_{\mu}(A, X) \text{ and } \log \sigma = \mathrm{GCN}_{\sigma}(A, X)$$

# 有向图神经网络的实验验证

不同于传统的链接预测任务，有向网络表征学习的链接预测大致分为如下三类：

## 标准链接预测

随机删除部分边，并抽取部分不存在边的节点对，进行二分类。

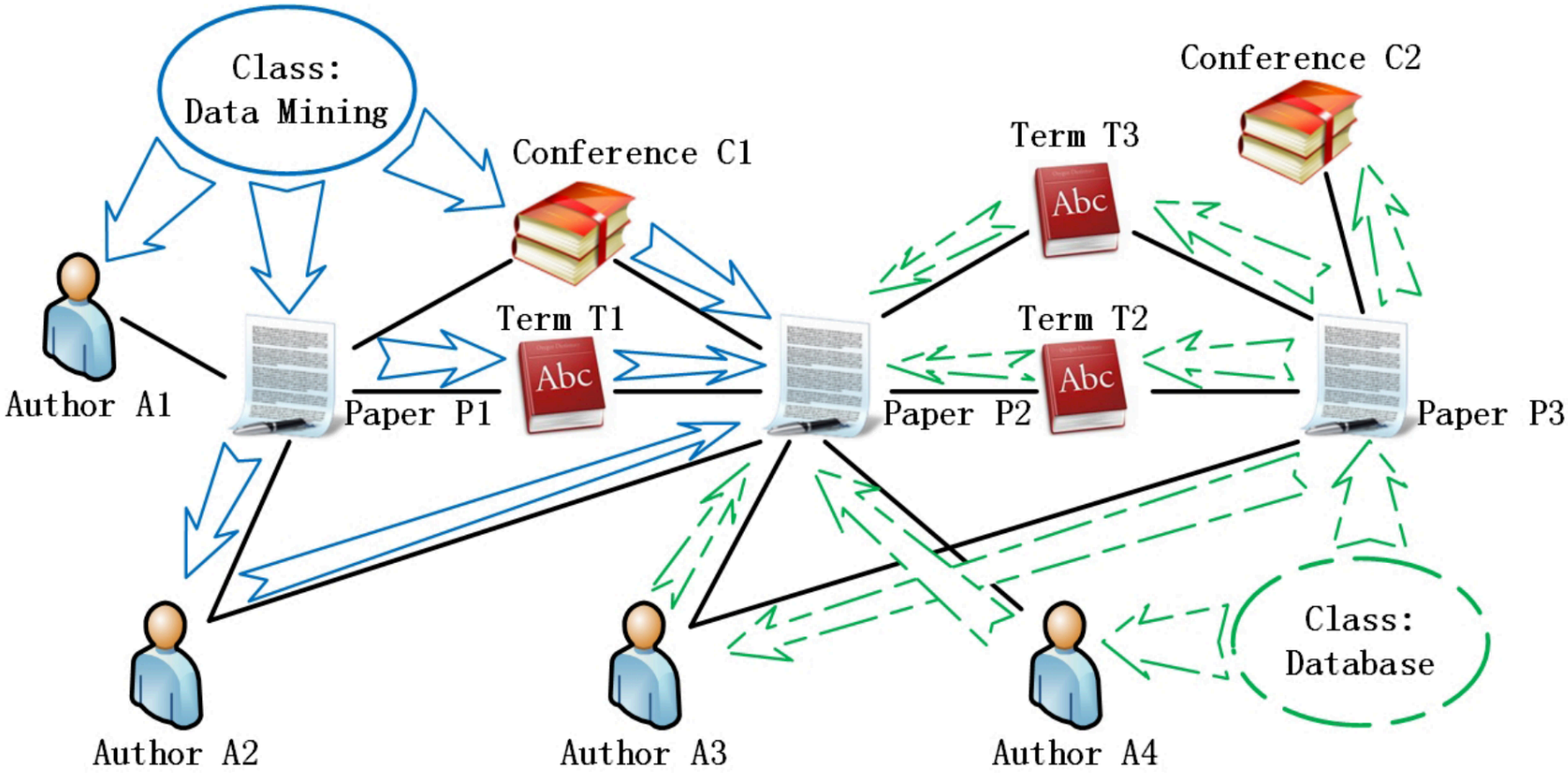
## 有偏链接预测

随机删除部分单向边，并抽取部分不存在边的节点对，进行二分类。

## 双向链接预测

随机删除双向边中的一条，并抽取部分不存在边的节点对，进行二分类。

# 异构信息网络

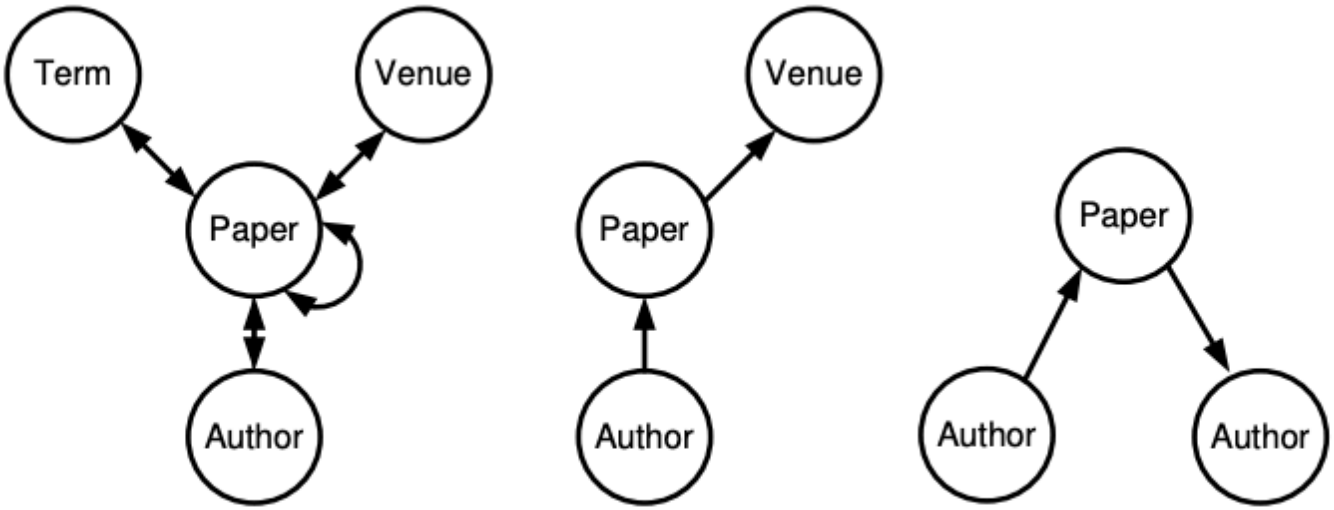



经典异构信息网络

# 异构信息网络

异构图神经网络的核心：

1. 元路径（Meta-path）
2. 语义空间（Semantic Space）

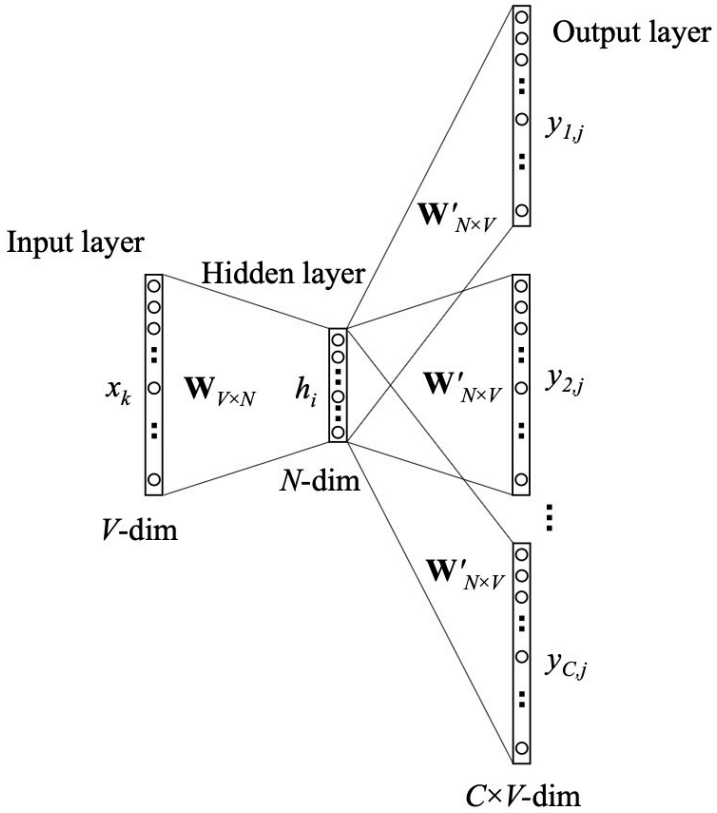



(a) Network Schema (b) Meta Path: APC (c) Meta Path: APA

# Word2Vec

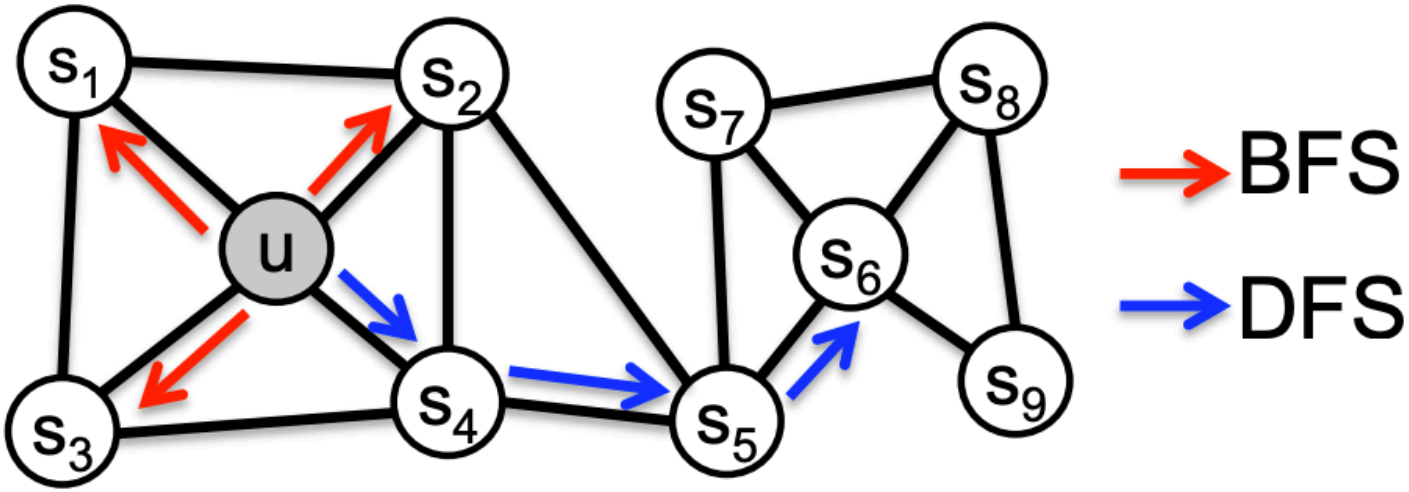
- 对于不同的周围词，隐含层向量 $h_i$ 相同，权重矩阵 $W'$ 共享，那理论上预测出来的词都是相同的，与实际不符。

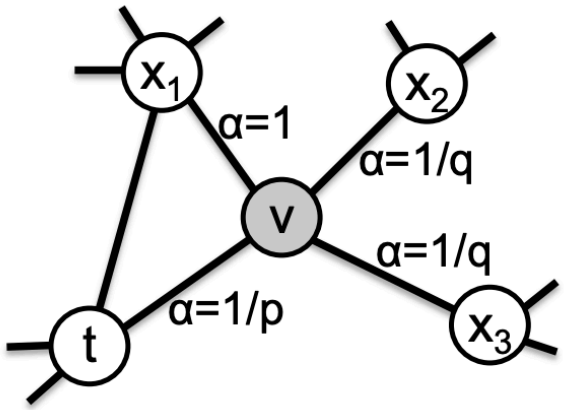
- Skip-gram 预测周围词的准确率并不是我们真正的目标——我们希望的是学到语义尽可能准确的词向量。因此，尽管预测上可能不准确，但从语义上理解是可行的。



# Node2Vec

改进的 RandomWalk 策略：



# Metapath2Vec

## 异构 Skip-gram

$$\arg\max_{\theta} \sum_{v \in V} \sum_{t \in T_V} \sum_{c_t \in N_t(v)} \log p\left(c_t \mid v; \theta\right)$$

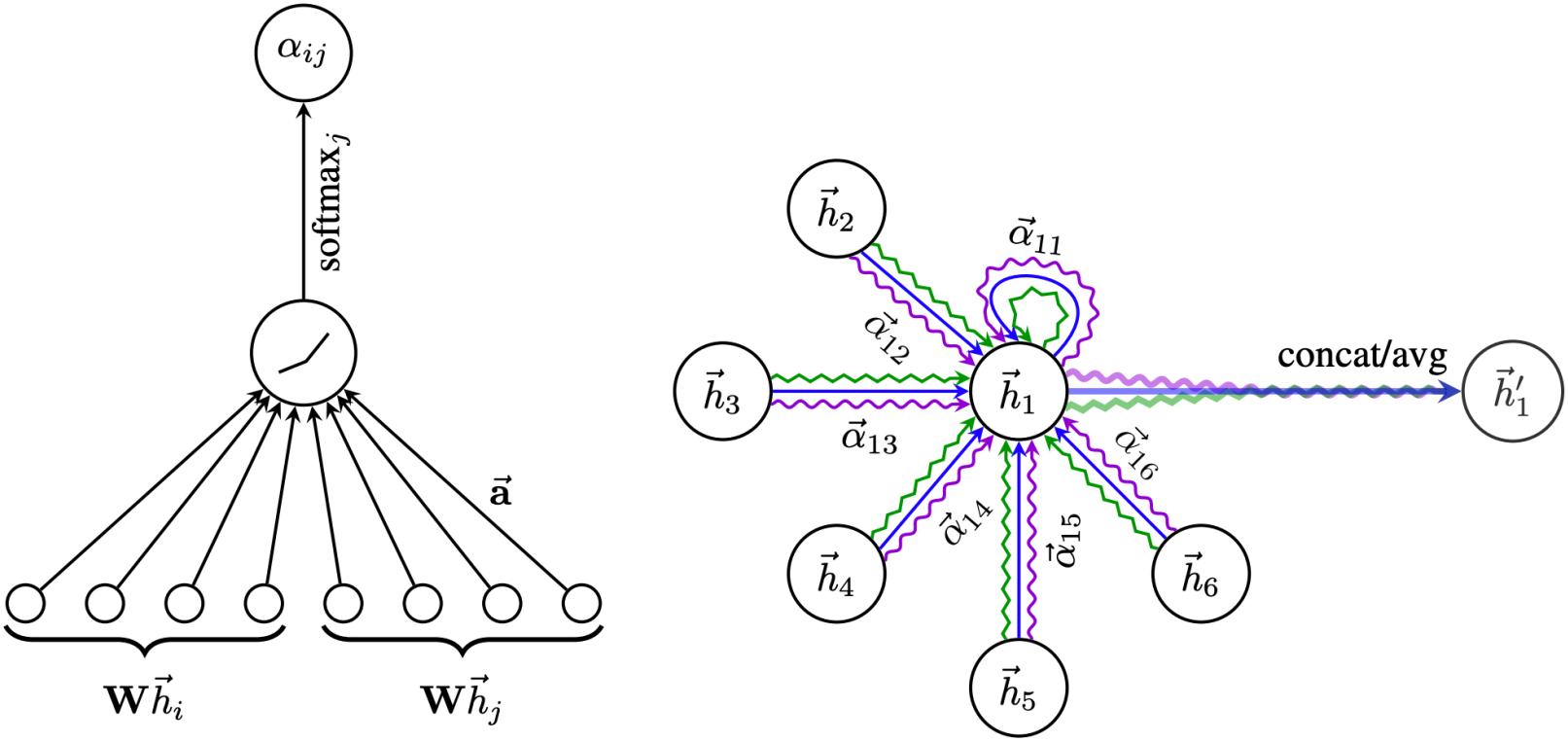
$$p\left(c_t \mid v; \theta\right) = \frac{e^{X_{c_t} \cdot X_v}}{\sum_{u \in V} e^{X_u \cdot X_v}}$$

## 异构随机游走

给定元路径，随机游走的转移概率描述为：

$$p\left(v^{i+1} \mid v_t^i, \mathcal{P}\right) = \begin{cases} \frac{1}{\left|N_{t+1}\left(v_t^i\right)\right|} & \left(v^{i+1}, v_t^i\right) \in E, \phi\left(v^{i+1}\right) = t+1 \\ 0 & \left(v^{i+1}, v_t^i\right) \in E, \phi\left(v^{i+1}\right) \neq t+1 \\ 0 & \left(v^{i+1}, v_t^i\right) \notin E \end{cases}$$

# Graph Attention Network




Graph Attention Network

**如何在异构网络上整合邻居信息？**

# Heterogenous Graph Attention Network

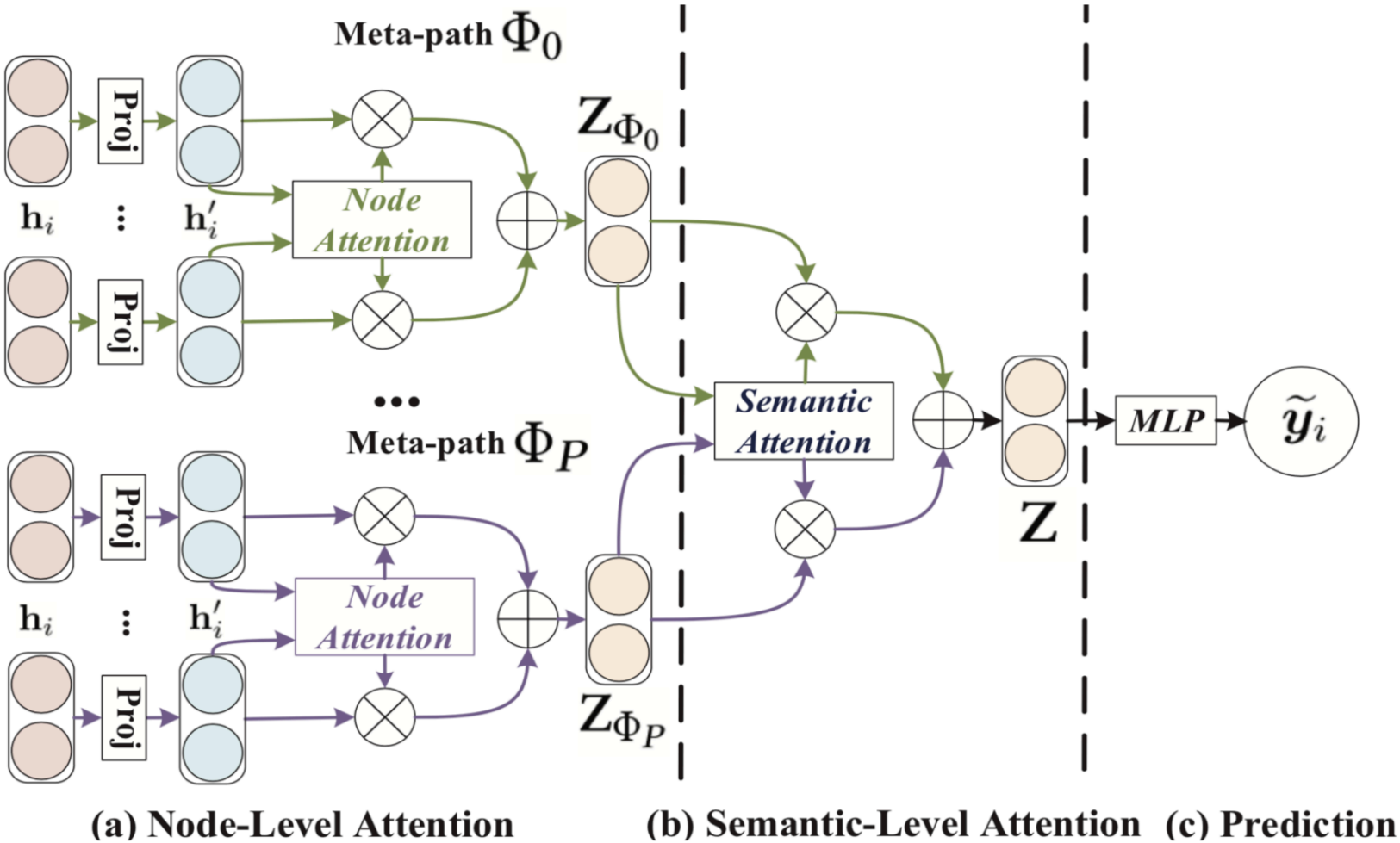
## 异构图神经网络的特点

1. 不同 meta-path 对应不同语义空间
2. 每个语义空间内节点的邻居和信息均不同
3. 节点的最终表征与每个语义空间均有关

## 异构图注意力网络

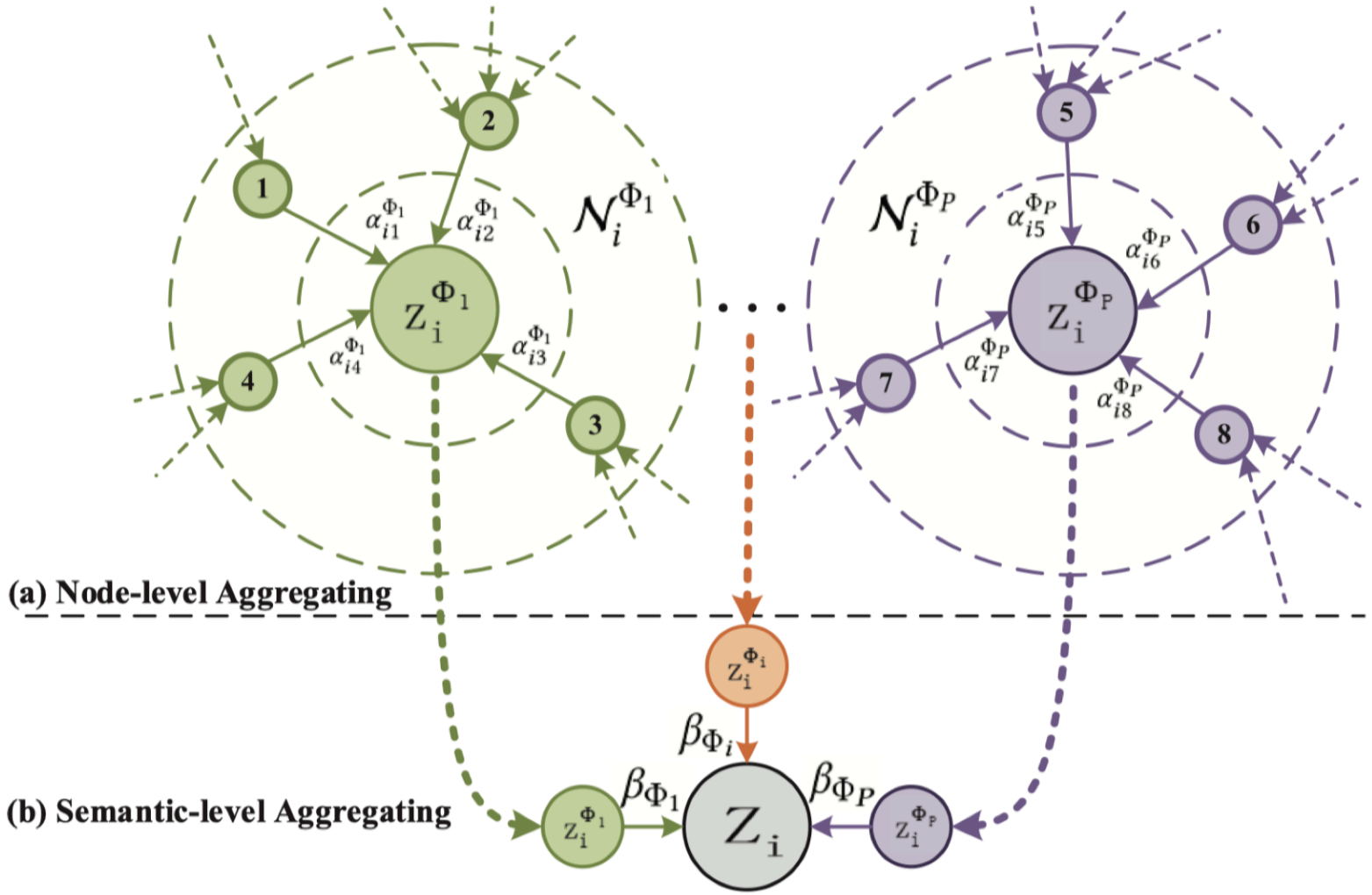
1. 使用 meta-path 异构信息网络投影到多个同构图
2. 在每个同构图内使用注意力网络整合邻居信息
3. 对不同 meta-path 对应的多个同构图使用全局注意力机制

# Heterogenous Graph Attention Network




HAN 模型架构

# Heterogenous Graph Attention Network



# Heterogenous Graph Attention Network

1. Node-level Attention

$$\alpha_{ij}^{\Phi} = \mathrm{softmax}_j\left(e_{ij}^{\Phi}\right) = \frac{\exp\left(\sigma\left(\mathbf{a}_{\Phi}^{\mathrm{T}} \cdot \left[\mathbf{h}_i' \| \mathbf{h}_j'\right]\right)\right)}{\sum_{k \in \mathcal{N}_i^{\Phi}} \exp\left(\sigma\left(\mathbf{a}_{\Phi}^{\mathrm{T}} \cdot \left[\mathbf{h}_i' \| \mathbf{h}_k'\right]\right)\right)}$$

$$\mathbf{z}_i^{\Phi} = \sigma\left(\sum_{j \in \mathcal{N}_i^{\Phi}} \alpha_{ij}^{\Phi} \cdot \mathbf{h}_j'\right)$$

2. Semantic-level Attention

$$w_{\Phi_p} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathbf{q}^{\mathrm{T}} \cdot \tanh\left(\mathbf{W} \cdot \mathbf{z}_i^{\Phi_p} + \mathbf{b}\right)$$

$$\beta_{\Phi_p} = \frac{\exp\left(w_{\Phi_p}\right)}{\sum_{p=1}^{P} \exp\left(w_{\Phi_p}\right)}$$

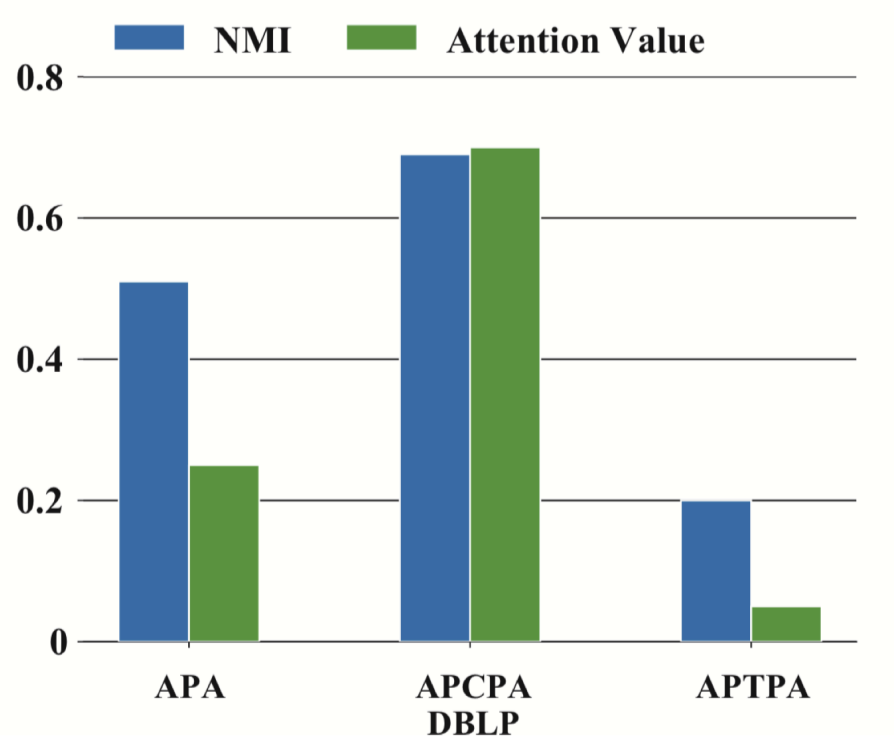
$$\mathbf{Z} = \sum^{P} \beta_{\Phi_p} \cdot \mathbf{Z}_{\Phi_p}$$

# Heterogenous Graph Attention Network

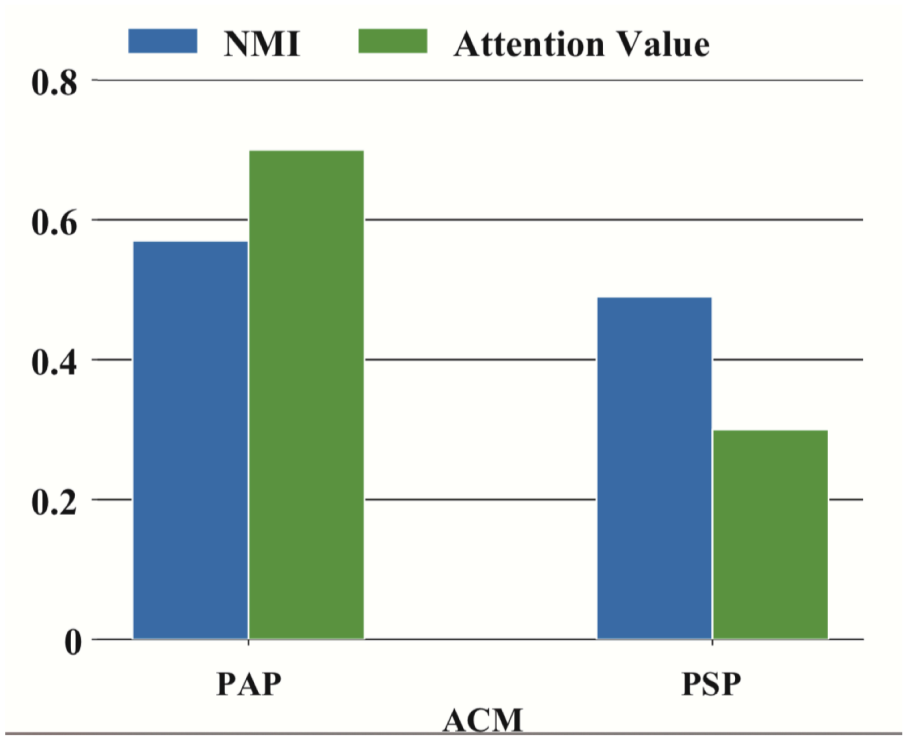
**Table 3: Qantitative results (%) on the node classification task.**

| Datasets | Metrics | Training | DeepWalk | ESim | metapath2vec | HERec | GCN | GAT | $HAN_{nd}$ | $HAN_{sem}$ | HAN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ACM | Macro-F1 | 20% | 77.25 | 77.32 | 65.09 | 66.17 | 86.81 | 86.23 | 88.15 | 89.04 | **89.40** |
| | | 40% | 80.47 | 80.12 | 69.93 | 70.89 | 87.68 | 87.04 | 88.41 | 89.41 | **89.79** |
| | | 60% | 82.55 | 82.44 | 71.47 | 72.38 | 88.10 | 87.56 | 87.91 | **90.00** | 89.51 |
| | | 80% | 84.17 | 83.00 | 73.81 | 73.92 | 88.29 | 87.33 | 88.48 | 90.17 | **90.63** |
| | Micro-F1 | 20% | 76.92 | 76.89 | 65.00 | 66.03 | 86.77 | 86.01 | 87.99 | 88.85 | **89.22** |
| | | 40% | 79.99 | 79.70 | 69.75 | 70.73 | 87.64 | 86.79 | 88.31 | 89.27 | **89.64** |
| | | 60% | 82.11 | 82.02 | 71.29 | 72.24 | 88.12 | 87.40 | 87.68 | **89.85** | 89.33 |
| | | 80% | 83.88 | 82.89 | 73.69 | 73.84 | 88.35 | 87.11 | 88.26 | 89.95 | **90.54** |
| DBLP | Macro-F1 | 20% | 77.43 | 91.64 | 90.16 | 91.68 | 90.79 | 90.97 | 91.17 | 92.03 | **92.24** |
| | | 40% | 81.02 | 92.04 | 90.82 | 92.16 | 91.48 | 91.20 | 91.46 | 92.08 | **92.40** |
| | | 60% | 83.67 | 92.44 | 91.32 | 92.80 | 91.89 | 90.80 | 91.78 | 92.38 | **92.80** |
| | | 80% | 84.81 | 92.53 | 91.89 | 92.34 | 92.38 | 91.73 | 91.80 | 92.53 | **93.08** |
| | Micro-F1 | 20% | 79.37 | 92.73 | 91.53 | 92.69 | 91.71 | 91.96 | 92.05 | 92.99 | **93.11** |
| | | 40% | 82.73 | 93.07 | 92.03 | 93.18 | 92.31 | 92.16 | 92.38 | 93.00 | **93.30** |
| | | 60% | 85.27 | 93.39 | 92.48 | 93.70 | 92.62 | 91.84 | 92.69 | 93.31 | **93.70** |
| | | 80% | 86.26 | 93.44 | 92.80 | 93.27 | 93.09 | 92.55 | 92.69 | 93.29 | **93.99** |
| IMDB | Macro-F1 | 20% | 40.72 | 32.10 | 41.16 | 41.65 | 45.73 | 49.44 | 49.78 | **50.87** | 50.00 |
| | | 40% | 45.19 | 31.94 | 44.22 | 43.86 | 48.01 | 50.64 | 52.11 | 50.85 | **52.71** |
| | | 60% | 48.13 | 31.68 | 45.11 | 46.27 | 49.15 | 51.90 | 51.73 | 52.09 | **54.24** |
| | | 80% | 50.35 | 32.06 | 45.15 | 47.64 | 51.81 | 52.99 | 52.66 | 51.60 | **54.38** |
| | Micro-F1 | 20% | 46.38 | 35.28 | 45.65 | 45.81 | 49.78 | 55.28 | 54.17 | 55.01 | **55.73** |
| | | 40% | 49.99 | 35.47 | 48.24 | 47.59 | 51.71 | 55.91 | 56.39 | 55.15 | **57.97** |
| | | 60% | 52.21 | 35.64 | 49.09 | 49.88 | 52.29 | 56.44 | 56.09 | 56.66 | **58.32** |
| | | 80% | 54.33 | 35.59 | 48.81 | 50.99 | 54.61 | 56.97 | 56.38 | 56.49 | **58.51** |

# Heterogenous Graph Attention Network




(a) NMI values on DBLP




(b) NMI values on ACM

Attention 权重与 meta-path 质量对比

# Heterogenous Graph Attention Network

HAN 模型总结：

## 优点

1. 模型结构清晰易懂
2. 使用 Attention 模型实现多重融合

## 缺点

1. 需要半监督指导

# 对比学习

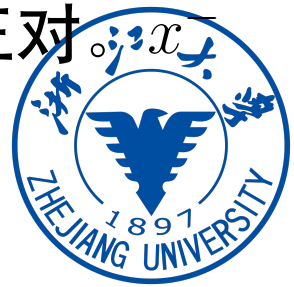
## 对比学习

对比学习（Contrastive Learning）是无监督表征学习中的一类，它的核心思想是在特征空间中将正样本对尽量靠近，负样本对尽量拉远，并以对比的形式进行训练。

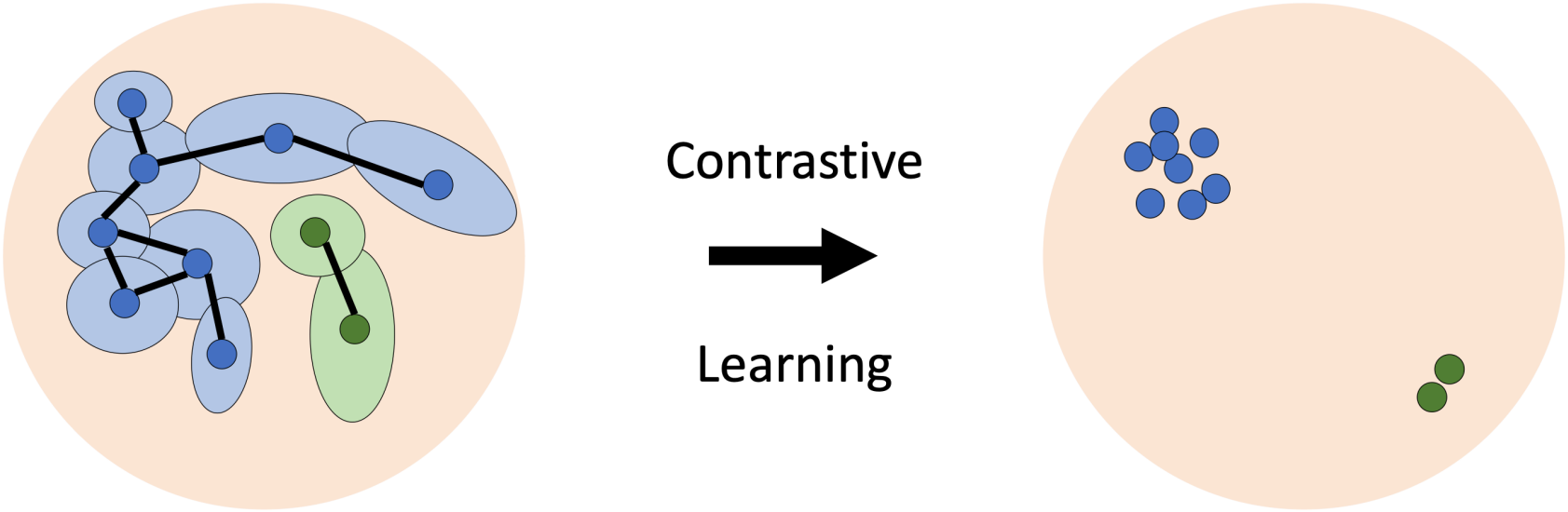
对任意样本数据 $x$，对比学习的目标是学习一个编码器 $f$ 使得：

$$score(f(x), f(x^+)) >> score(f(x), f(x^-))$$

其中，$x^+$ 为与 $x$ 相似的样本，称为正样本，两者组成正对。$x^-$ 为与 $x$ 不相似的样本，称为负样本，两者组成负对。

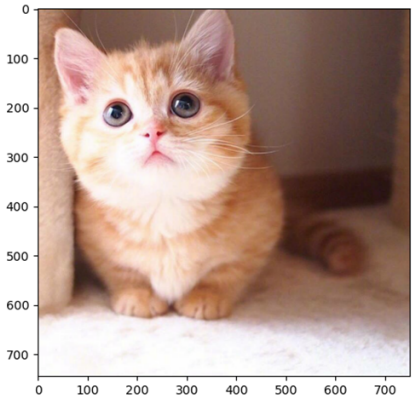
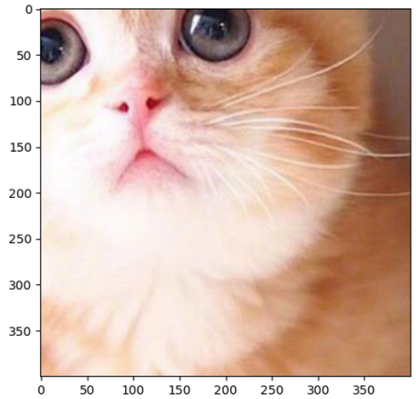


# 对比学习

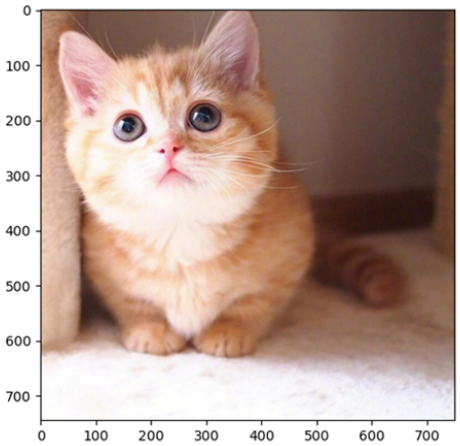
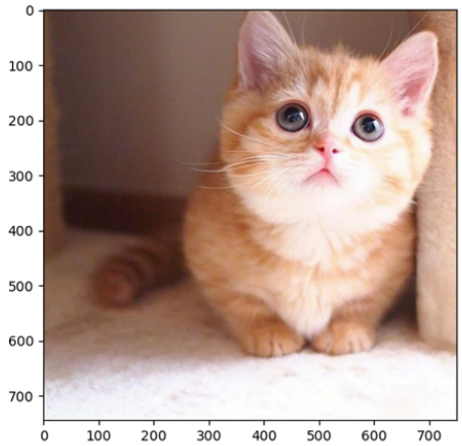
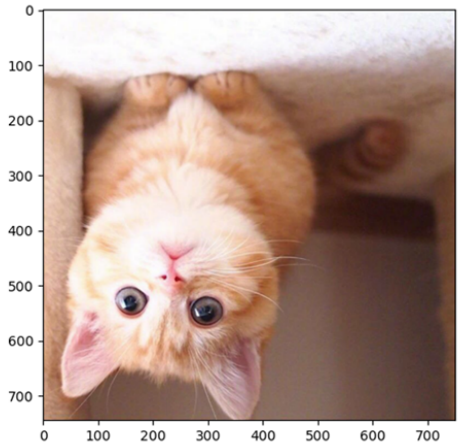



对比学习原理 [Anonymous, 2022]

# 常用的图像变换

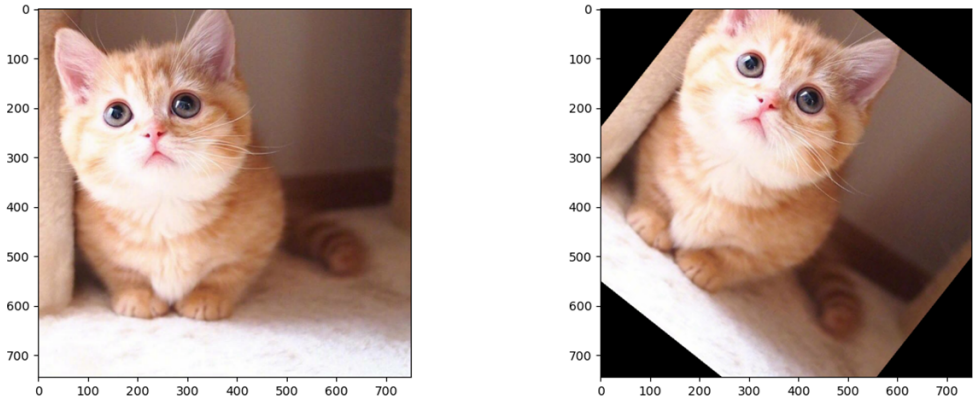



随机裁剪

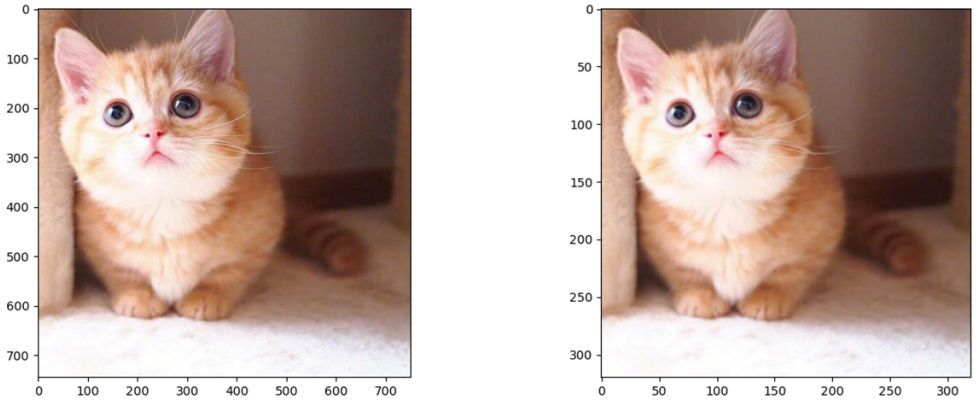




水平和竖直翻转

# 常用的图像变换

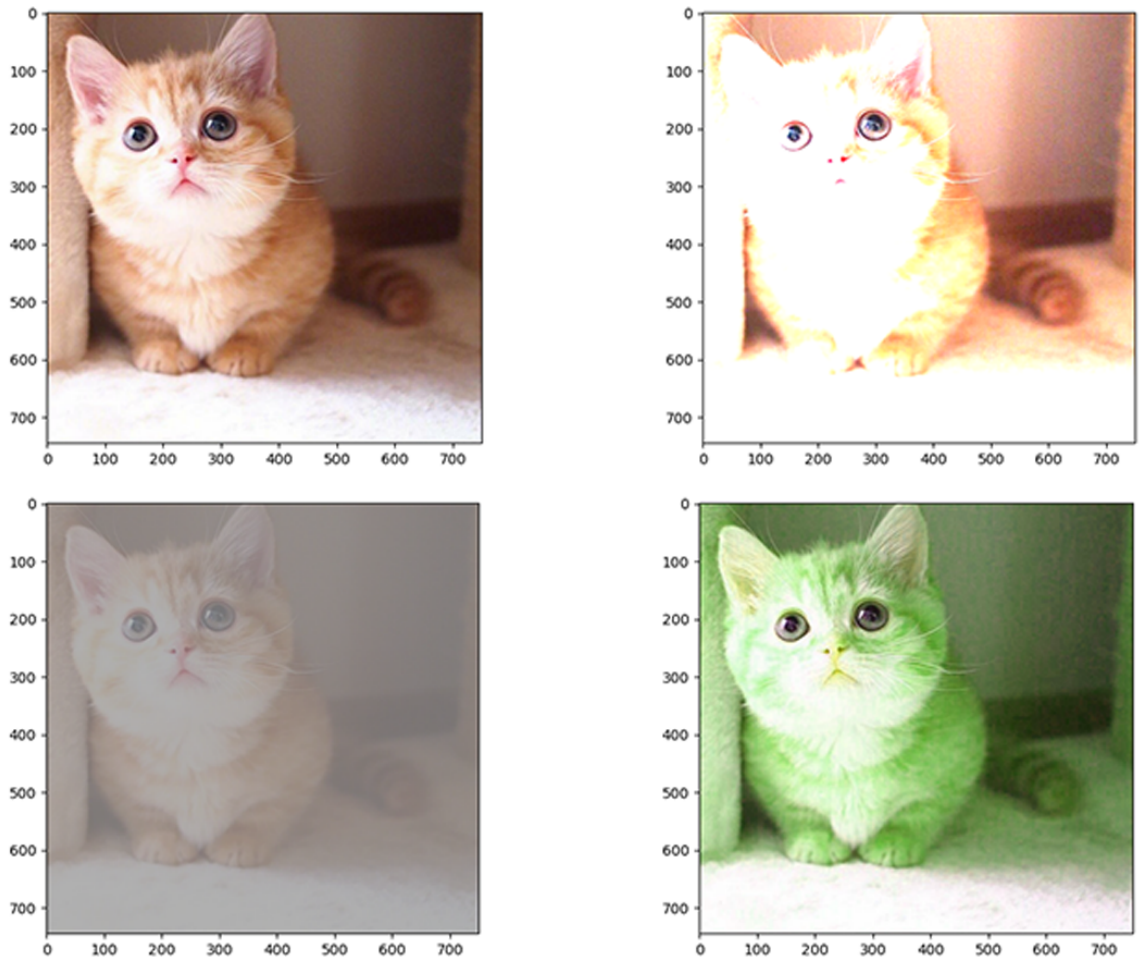


随机旋转



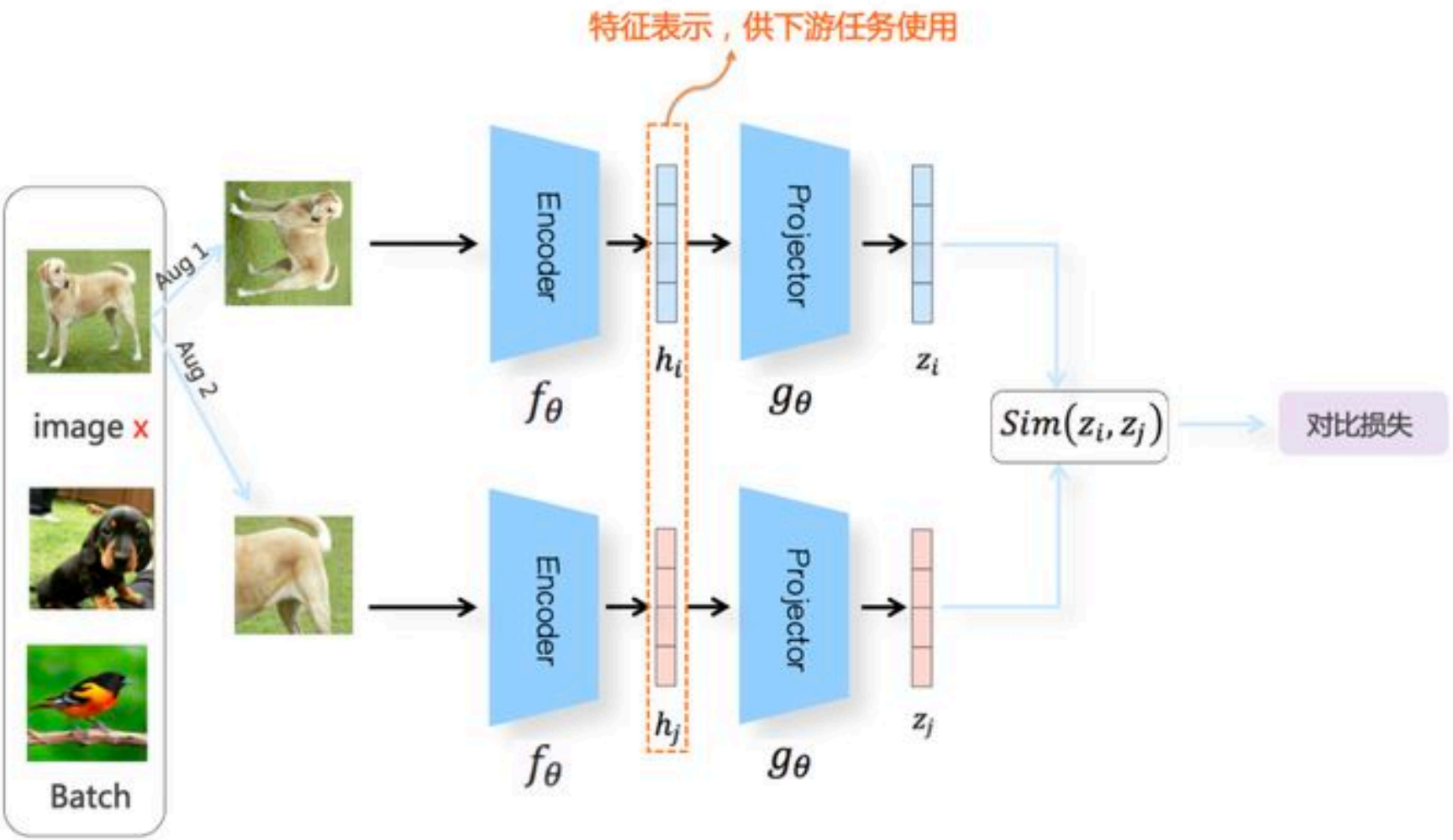
图像缩放

# 常用的图像变换



亮度、对比度和颜色的随机变换

# SimCLR




SimCLR 框架

# SimCLR

SimCLR 使用余弦距离来衡量样本之间的对相似度:

$$S(z_i, z_j) = \frac{z_i^T z_j}{\|z_i\|_2 \|z_j\|_2}$$

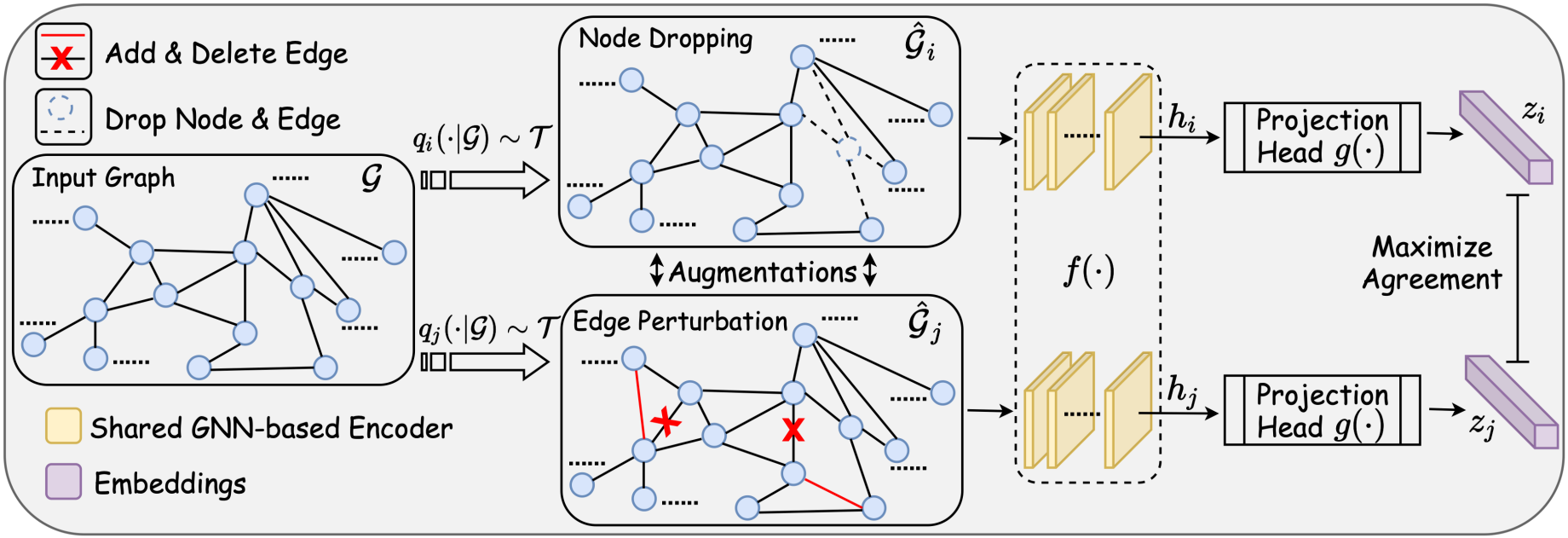
对于样本 $i$，相对应的 InfoNCE Loss 为:

$$L_i = -\log \frac{\exp(S(z_i, z_i^+)/\tau)}{\sum_{j=1}^{N} \exp(S(z_i, z_j)/\tau)}$$

其中 $\tau$ 为温度参数，用于控制整体分布的均匀程度。

分子部分鼓励正样本对间相似度越大越好，分母部分鼓励负样本对之间相似度越小越好。

# 对比图神经网络




对比图神经网络框架

# 图数据增强

常见的图数据增强形式包括：

1. 节点丢弃：随机丢弃部分节点以及对应的边
2. 边扰动：随机增加或者删除部分边
3. 节点属性掩码：随机掩盖部分属性
4. 子图采样：从原始图中随机采样子图

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning

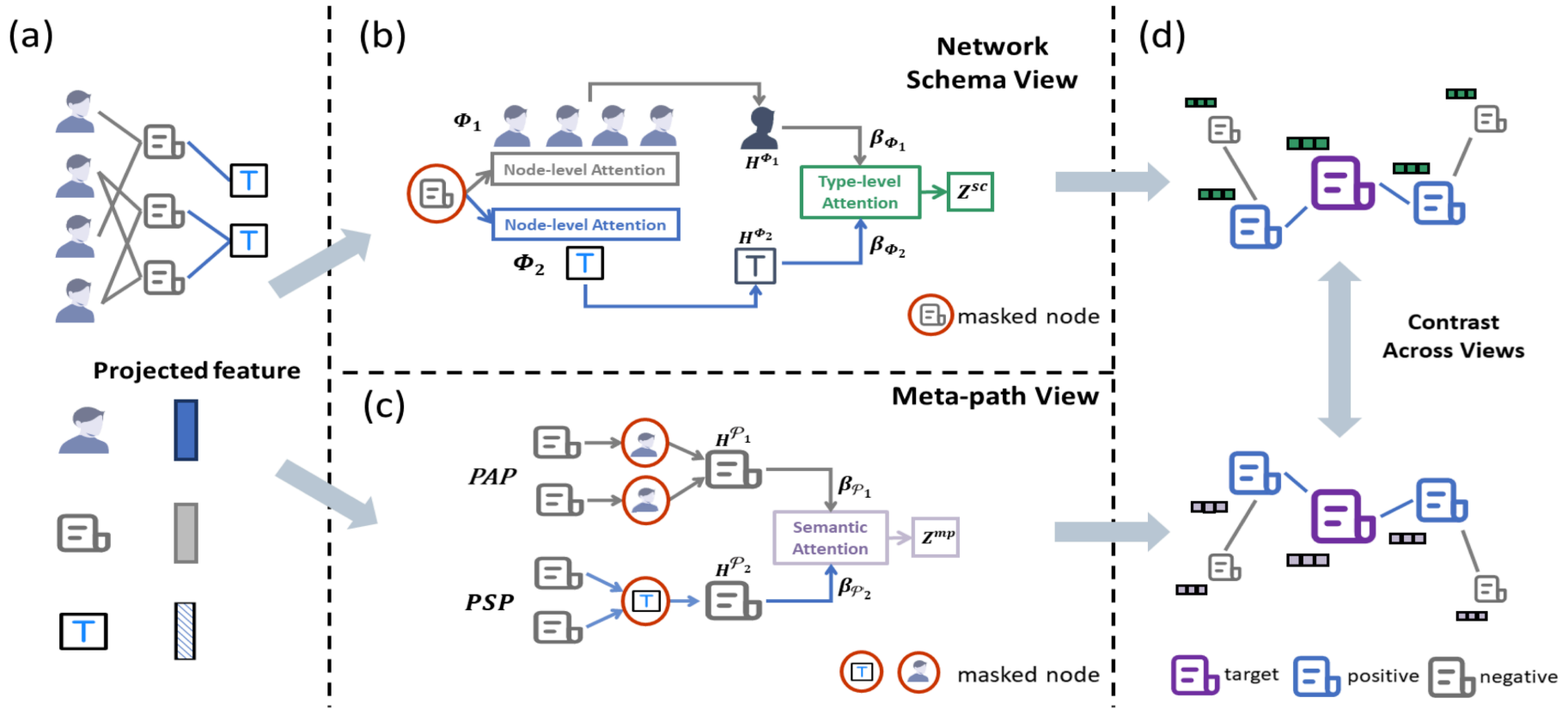
## 异构图上的对比学习难点

1. 异构图上对比的对象是什么？
2. 异构图的特点如何被保留？
3. 如何进行对比学习和训练？

## HeCo 解决方案：

1. 从 network schema 角度和 meta-path 角度构造不同视角
2. 在不同视角中同时捕获异构网络的局部特征和全局特征
3. View Mask 对比学习

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning




HeCo 模型架构

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning

- Network Schema View Encoder

$$h_i^{\Phi_m} = \sigma\Big( \sum_{j \in N_i^{\Phi m}} \alpha_{i,j}^{\Phi_m} \cdot h_j \Big)$$

$$\alpha_{i,j}^{\Phi_m} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}_{\Phi_m}^{\top} \cdot [h_i \| h_j]\right)\right)}{\sum_{l \in N_i^{\Phi m}} \exp\left(\text{LeakyReLU}\left(\mathbf{a}_{\Phi_m}^{\top} \cdot [h_i \| h_l]\right)\right)},$$

$$w_{\Phi_m} = \frac{1}{|V|} \sum_{i \in V} \mathbf{a}_{sc}^{\top} \cdot \tanh\left(\mathbf{W}_{sc} h_i^{\Phi_m} + \mathbf{b}_{sc}\right)$$

$$\beta_{\Phi_m} = \frac{\exp\left(w_{\Phi_m}\right)}{\sum_{i=1}^{S} \exp\left(w_{\Phi_i}\right)}$$

$$z_i^{sc} = \sum_{m=1}^{S} \beta_{\Phi_m} \cdot h_i^{\Phi_m}$$

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning

- Meta-path View Encoder

$$h_i^{\mathcal{P}_n} = \frac{1}{d_i + 1} h_i + \sum_{j \in N_i^{\mathcal{P}_n}} \frac{1}{\sqrt{(d_i + 1)(d_j + 1)}} h_j$$

$$z_i^{mp} = \sum_{n=1}^{M} \beta_{\mathcal{P}_n} \cdot h_i^{\mathcal{P}_n}$$

$$w_{\mathcal{P}_n} = \frac{1}{|V|} \sum_{i \in V} \mathbf{a}_{mp}^{\top} \cdot \tanh\left(\mathbf{W}_{mp} h_i^{\mathcal{P}_n} + \mathbf{b}_{mp}\right)$$

$$\beta_{\mathcal{P}_n} = \frac{\exp\left(w\mathcal{P}_n\right)}{\sum_{i=1}^{M} \exp\left(w_{\mathcal{P}_i}\right)}$$

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning

- View Mask **数据增强**
- **正样本增广**
- **协同对比学习**

$$\mathcal{L}_i^{sc} = -\log \frac{\sum_{j \in \mathbb{P}_i} \exp\left(\text{sim}\left(z_i^{sc} - proj, z_j^{mp} - proj\right)/\tau\right)}{\sum_{k \in \{\mathbb{P}_i \cup \mathbb{N}_i\}} \exp\left(\text{sim}\left(z_i^{sc} - proj, z_k^{mp} - proj\right)/\tau\right)},$$

# Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning

| Datasets | Metric | Split | GraphSAGE | GAE | Mp2vec | HERec | HetGNN | HAN | DGI | DMGI | HeCo |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ACM | Ma-F1 | 20 | 47.13±4.7 | 62.72±3.1 | 51.91±0.9 | 55.13±1.5 | 72.11±0.9 | 85.66±2.1 | 79.27±3.8 | 87.86±0.2 | **88.56±0.8** |
| | | 40 | 55.96±6.8 | 61.61±3.2 | 62.41±0.6 | 61.21±0.8 | 72.02±0.4 | 87.47±1.1 | 80.23±3.3 | 86.23±0.8 | **87.61±0.5** |
| | | 60 | 56.59±5.7 | 61.67±2.9 | 61.13±0.4 | 64.35±0.8 | 74.33±0.6 | 88.41±1.1 | 80.03±3.3 | 87.97±0.4 | **89.04±0.5** |
| | Mi-F1 | 20 | 49.72±5.5 | 68.02±1.9 | 53.13±0.9 | 57.47±1.5 | 71.89±1.1 | 85.11±2.2 | 79.63±3.5 | 87.60±0.8 | **88.13±0.8** |
| | | 40 | 60.98±3.5 | 66.38±1.9 | 64.43±0.6 | 62.62±0.9 | 74.46±0.8 | 87.21±1.2 | 80.41±3.0 | 86.02±0.9 | **87.45±0.5** |
| | | 60 | 60.72±4.3 | 65.71±2.2 | 62.72±0.3 | 65.15±0.9 | 76.08±0.7 | 88.10±1.2 | 80.15±3.2 | 87.82±0.5 | **88.71±0.5** |
| | AUC | 20 | 65.88±3.7 | 79.50±2.4 | 71.66±0.7 | 75.44±1.3 | 84.36±1.0 | 93.47±1.5 | 91.47±2.3 | **96.72±0.3** | 96.49±0.3 |
| | | 40 | 71.06±5.2 | 79.14±2.5 | 80.48±0.4 | 79.84±0.5 | 85.01±0.6 | 94.84±0.9 | 91.52±2.3 | 96.35±0.3 | **96.40±0.4** |
| | | 60 | 70.45±6.2 | 77.90±2.8 | 79.33±0.4 | 81.64±0.7 | 87.64±0.7 | 94.68±1.4 | 91.41±1.9 | **96.79±0.2** | 96.55±0.3 |
| DBLP | Ma-F1 | 20 | 71.97±8.4 | 90.90±0.1 | 88.98±0.2 | 89.57±0.4 | 89.51±1.1 | 89.31±0.9 | 87.93±2.4 | 89.94±0.4 | **91.28±0.2** |
| | | 40 | 73.69±8.4 | 89.60±0.3 | 88.68±0.2 | 89.73±0.4 | 88.61±0.8 | 88.87±1.0 | 88.62±0.6 | 89.25±0.4 | **90.34±0.3** |
| | | 60 | 73.86±8.1 | 90.08±0.2 | 90.25±0.1 | 90.18±0.3 | 89.56±0.5 | 89.20±0.8 | 89.19±0.9 | 89.46±0.6 | **90.64±0.3** |
| | Mi-F1 | 20 | 71.44±8.7 | 91.55±0.1 | 89.67±0.1 | 90.24±0.4 | 90.11±1.0 | 90.16±0.9 | 88.72±2.6 | 90.78±0.3 | **91.97±0.2** |
| | | 40 | 73.61±8.6 | 90.00±0.3 | 89.14±0.2 | 90.15±0.4 | 89.03±0.7 | 89.47±0.9 | 89.22±0.5 | 89.92±0.4 | **90.76±0.3** |
| | | 60 | 74.05±8.3 | 90.95±0.2 | 91.17±0.1 | 91.01±0.3 | 90.43±0.6 | 90.34±0.8 | 90.35±0.8 | 90.66±0.5 | **91.59±0.2** |
| | AUC | 20 | 90.59±4.3 | 98.15±0.1 | 97.69±0.0 | 98.21±0.2 | 97.96±0.4 | 98.07±0.6 | 96.99±1.4 | 97.75±0.3 | **98.32±0.1** |
| | | 40 | 91.42±4.0 | 97.85±0.1 | 97.08±0.0 | 97.93±0.1 | 97.70±0.3 | 97.48±0.6 | 97.12±0.4 | 97.23±0.2 | **98.06±0.1** |
| | | 60 | 91.73±3.8 | 98.37±0.1 | 98.00±0.0 | 98.49±0.1 | 97.97±0.2 | 97.96±0.5 | 97.76±0.5 | 97.72±0.4 | **98.59±0.1** |
| Freebase | Ma-F1 | 20 | 45.14±4.5 | 53.81±0.6 | 53.96±0.7 | 55.78±0.5 | 52.72±1.0 | 53.16±2.8 | 54.90±0.7 | 55.79±0.9 | **59.23±0.7** |
| | | 40 | 44.88±4.1 | 52.44±2.3 | 57.80±1.1 | 59.28±0.6 | 48.57±0.5 | 59.63±2.3 | 53.40±1.4 | 49.88±1.9 | **61.19±0.6** |
| | | 60 | 45.16±3.1 | 50.65±0.4 | 55.94±0.7 | 56.50±0.4 | 52.37±0.8 | 56.77±1.7 | 53.81±1.1 | 52.10±0.7 | **60.13±1.3** |
| | Mi-F1 | 20 | 54.83±3.0 | 55.20±0.7 | 56.23±0.8 | 57.92±0.5 | 56.85±0.9 | 57.24±3.2 | 58.16±0.9 | 58.26±0.9 | **61.72±0.6** |
| | | 40 | 57.08±3.2 | 56.05±2.0 | 61.01±1.3 | 62.71±0.7 | 53.96±1.1 | 63.74±2.7 | 57.82±0.8 | 54.28±1.6 | **64.03±0.7** |
| | | 60 | 55.92±3.2 | 53.85±0.4 | 58.74±0.8 | 58.57±0.5 | 56.84±0.7 | 61.06±2.0 | 57.96±0.7 | 56.69±1.2 | **63.61±1.6** |
| | AUC | 20 | 67.63±5.0 | 73.03±0.7 | 71.78±0.7 | 73.89±0.4 | 70.84±0.7 | 73.26±2.1 | 72.80±0.6 | 73.19±1.2 | **76.22±0.8** |
| | | 40 | 66.42±4.7 | 74.05±0.9 | 75.51±0.8 | 76.08±0.4 | 69.48±0.2 | 77.74±1.2 | 72.97±1.1 | 70.77±1.6 | **78.44±0.5** |
| | | 60 | 66.78±3.5 | 71.75±0.4 | 74.78±0.4 | 74.89±0.4 | 71.01±0.5 | 75.69±1.5 | 73.32±0.9 | 73.17±1.4 | **78.04±0.4** |
| AMiner | Ma-F1 | 20 | 42.46±2.5 | 60.22±2.0 | 54.78±0.5 | 58.32±1.1 | 50.06±0.9 | 56.07±3.2 | 51.61±3.2 | 59.50±2.1 | **71.38±1.1** |
| | | 40 | 45.77±1.5 | 65.66±1.5 | 64.77±0.5 | 64.50±0.7 | 58.97±0.9 | 63.85±1.5 | 54.72±2.6 | 61.92±2.1 | **73.75±0.5** |
| | | 60 | 44.91±2.0 | 63.74±1.6 | 60.65±0.3 | 65.53±0.7 | 57.34±1.4 | 62.02±1.2 | 55.45±2.4 | 61.15±2.5 | **75.80±1.8** |
| | Mi-F1 | 20 | 49.68±3.1 | 65.78±2.9 | 60.82±0.4 | 63.64±1.1 | 61.49±2.5 | 68.86±4.6 | 62.39±3.9 | 63.93±3.3 | **78.81±1.3** |
| | | 40 | 52.10±2.2 | 71.34±1.8 | 69.66±0.6 | 71.57±0.7 | 68.47±2.2 | 76.89±1.6 | 63.87±2.9 | 63.60±2.5 | **80.53±0.7** |
| | | 60 | 51.36±2.2 | 67.70±1.9 | 63.92±0.5 | 69.76±0.8 | 65.61±2.2 | 74.73±1.4 | 63.10±3.0 | 62.51±2.6 | **82.46±1.4** |
| | AUC | 20 | 70.86±2.5 | 85.39±1.0 | 81.22±0.3 | 83.35±0.5 | 77.96±1.4 | 78.92±2.3 | 75.89±2.2 | 85.34±0.9 | **90.82±0.6** |
| | | 40 | 74.44±1.3 | 88.29±1.0 | 88.82±0.2 | 88.70±0.4 | 83.14±1.6 | 80.72±2.1 | 77.86±2.1 | 88.02±1.3 | **92.11±0.6** |
| | | 60 | 74.16±1.3 | 86.92±0.8 | 85.57±0.2 | 87.74±0.5 | 84.77±0.9 | 80.39±1.5 | 77.21±1.4 | 86.20±1.7 | **92.40±0.7** |

# 异构图神经网络论文写作

1. 符号定义清楚
2. Meta-path 等概念定义清楚
3. 画图清晰
4. 如何捕获异构网络的特性？
5. 对比方法公平公正透明
6. 多做 Case Study

# 动态网络

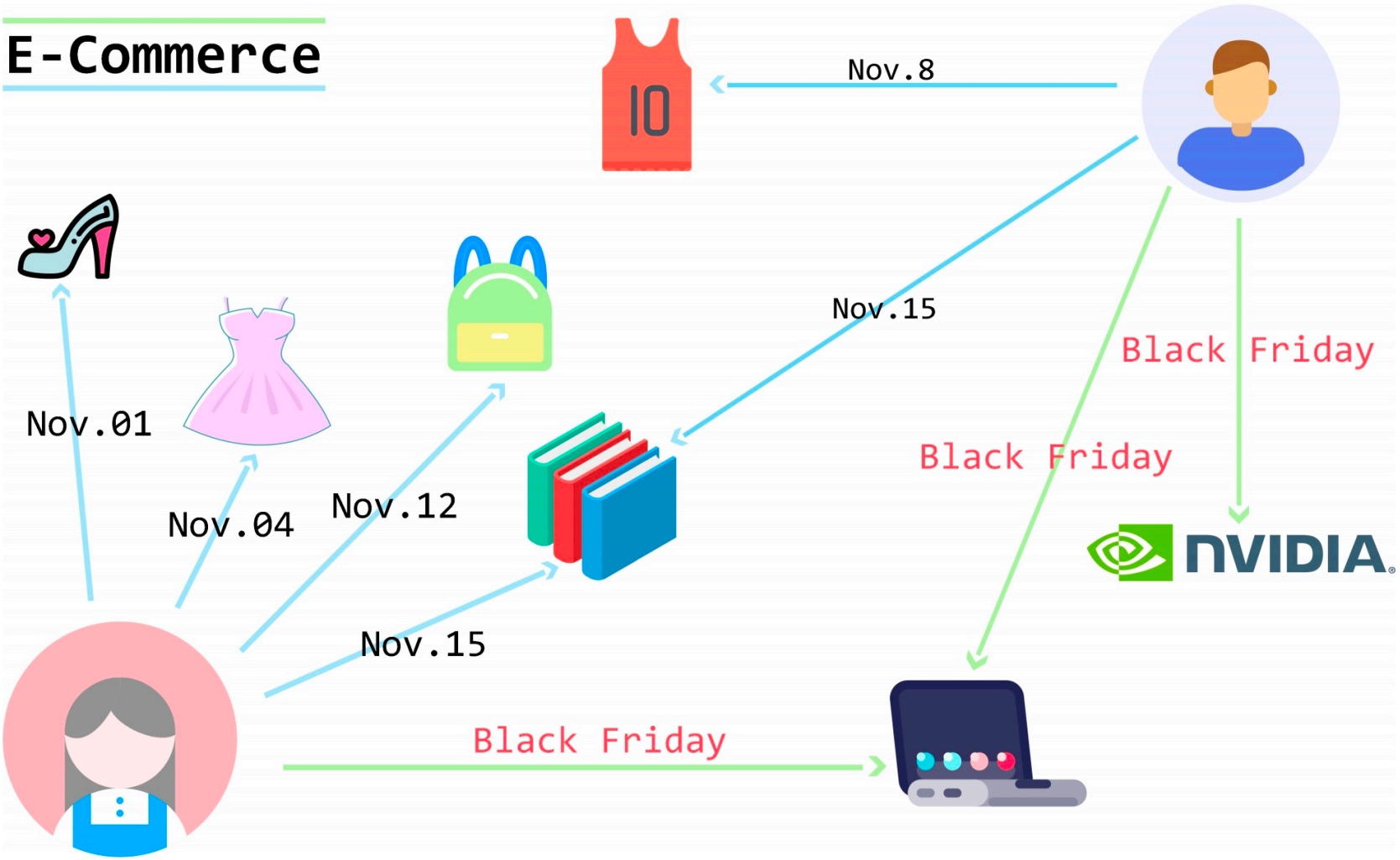
## 动态网络

- 现实世界中的网络往往都是由不断变化的实体组成的。
- 现有的大部分方法都是将其看作静态的网络，没有考虑动态网络的演化趋势。

常见的动态网络：

- 用户-物品网络
- 货币交易网络
- 社交网络
- ……

# 用户-物品网络


E-Commerce
Nov.8
Nov.15
Black Friday
Black Friday
Nov.01
Nov.04
Nov.12
Nov.15
Black Friday
NVIDIA

# 动态网络带来的挑战

## 捕获时间信息

- 边和节点是随时间演化的，邻域聚合会受到时间的约束。
- 学习的节点表示中需要对时间信息进行编码。

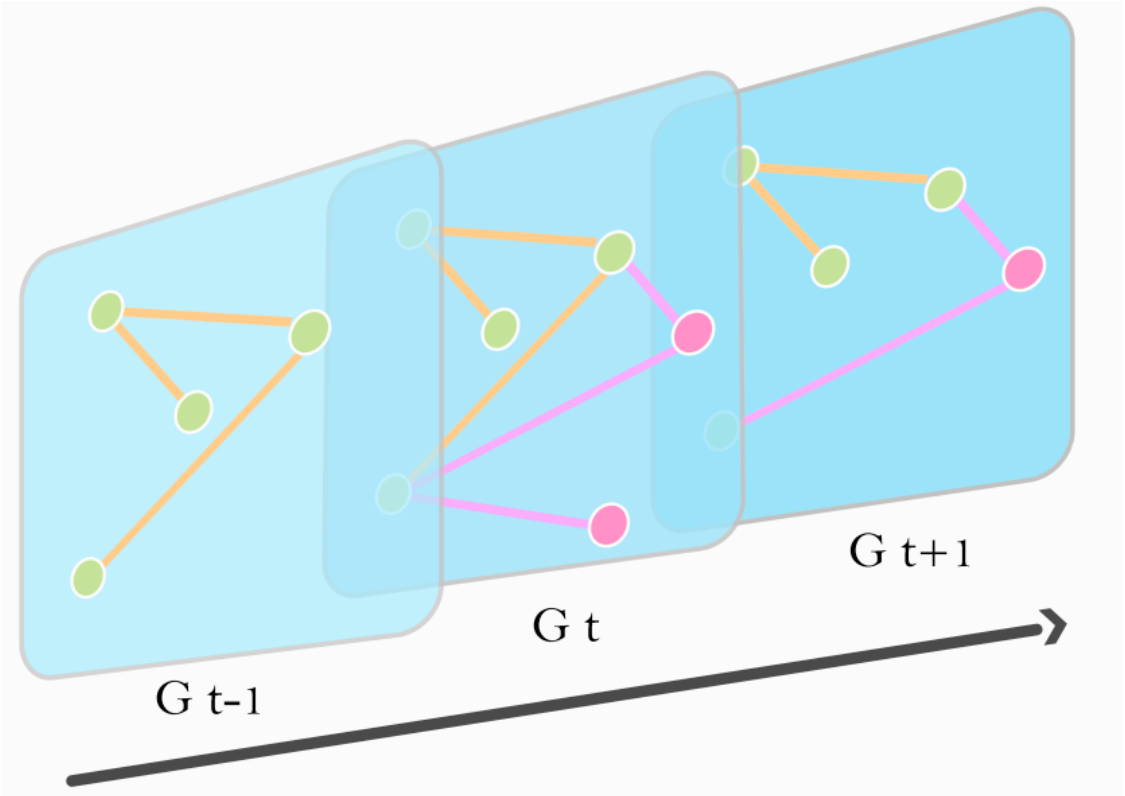
## 高效的表示更新方法

- 一种最简单的更新方法是在每个时间步都应用一次静态网络嵌入方法。然而，这样会消耗大量的时间和计算资源，特别是对于大型的网络。

# 离散模型

## 离散模型

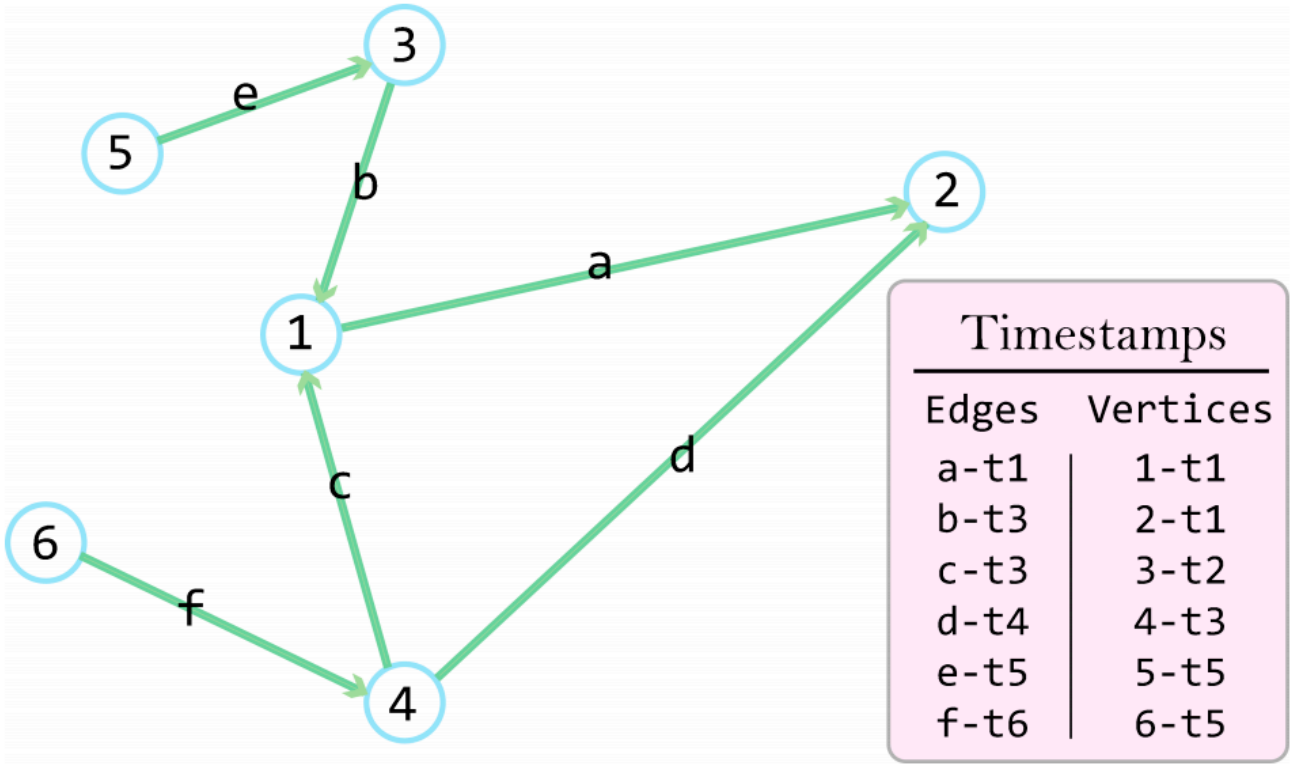
- 将动态网络看作一系列的静态网络，连续两个静态网络之间的区别就是发生的变化。



# 连续模型

## 连续模型

- 为每个边和节点指定特定的时间戳，以此来表示发生的变化。




Timestamps

| Edges | Vertices |
|---|---|
| a-t1 | 1-t1 |
| b-t3 | 2-t1 |
| c-t3 | 3-t2 |
| d-t4 | 4-t3 |
| e-t5 | 5-t5 |
| f-t6 | 6-t5 |

# 动态网络嵌入方法

根据不同的策略，我们可以将动态网络的嵌入方法分为以下几类：

- 基于矩阵分解的嵌入方法
- 基于Skip-Gram的嵌入方法
- 基于自动编码器的嵌入方法
- 基于神经网络的嵌入方法

# 基于矩阵分解的嵌入方法

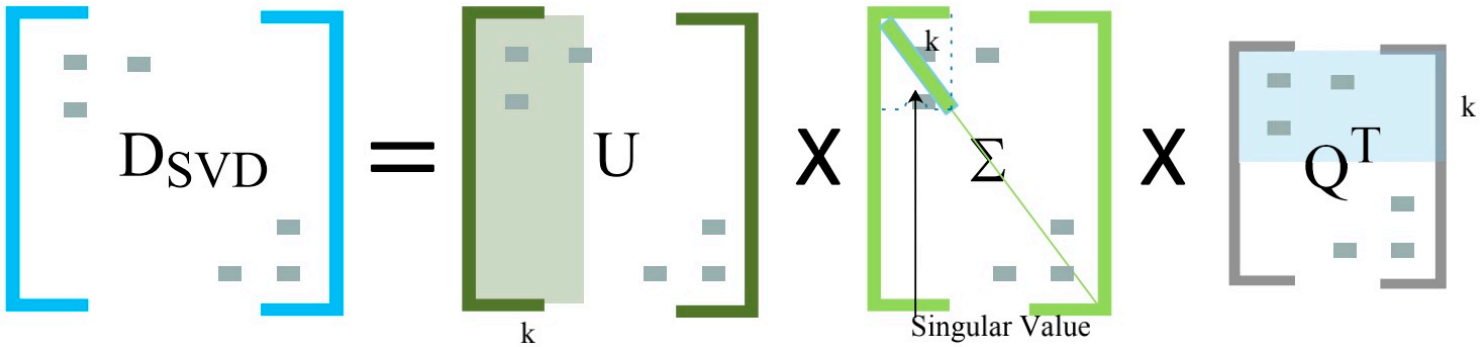
矩阵分解是在网络嵌入领域最普遍的降维方法。其中最具代表性的是奇异值分解（Singular value decomposition，SVD）。

## SVD

- 矩阵 $D_{\mathcal{SVD}} \in R^{m \times n}$ 的 SVD 定义为：

$$D_{\mathcal{SVD}} = U \Sigma Q^T$$

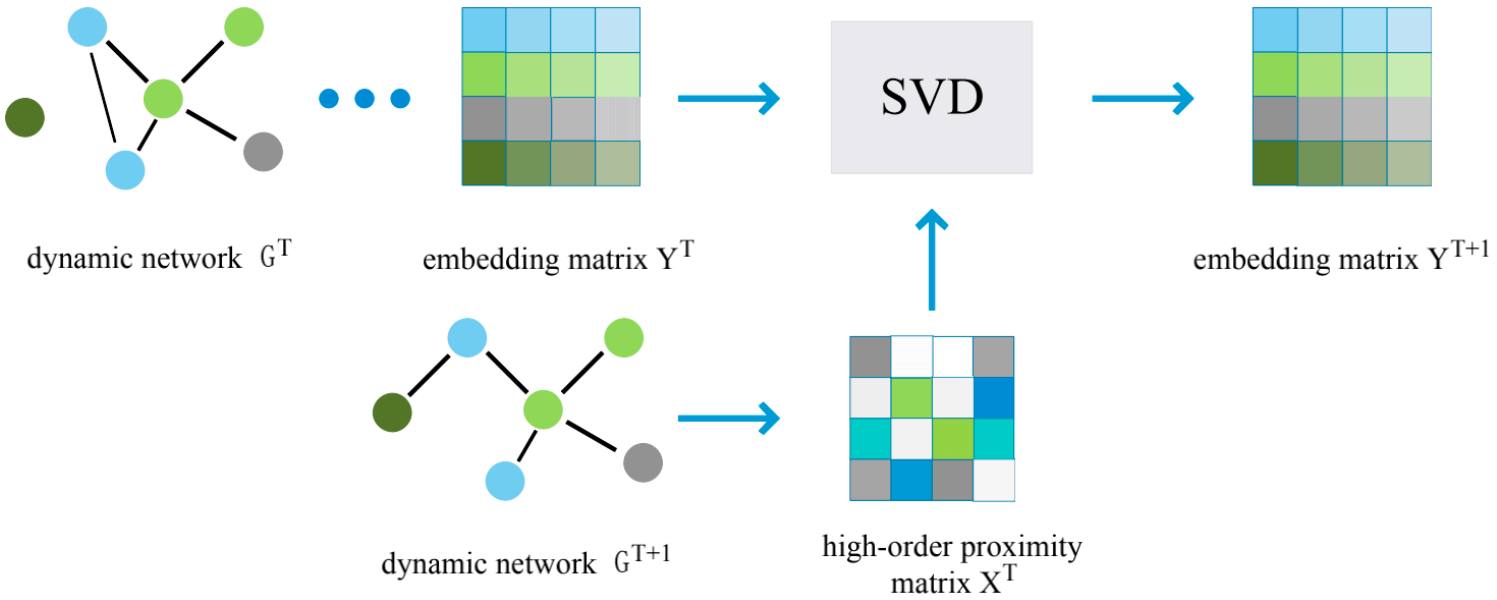
其中 $U \in R^{m \times m}, Q \in R^{n \times n}, \Sigma \in R^{m \times n}$，$\Sigma$ 是一个对角矩阵，主对角线上的每个元素都是一个奇异值。



# 基于矩阵分解的嵌入方法

- 一种传统但有效的方法是对复杂网络中的邻接矩阵和特征矩阵进行分解，从而构造每个节点的嵌入表示。

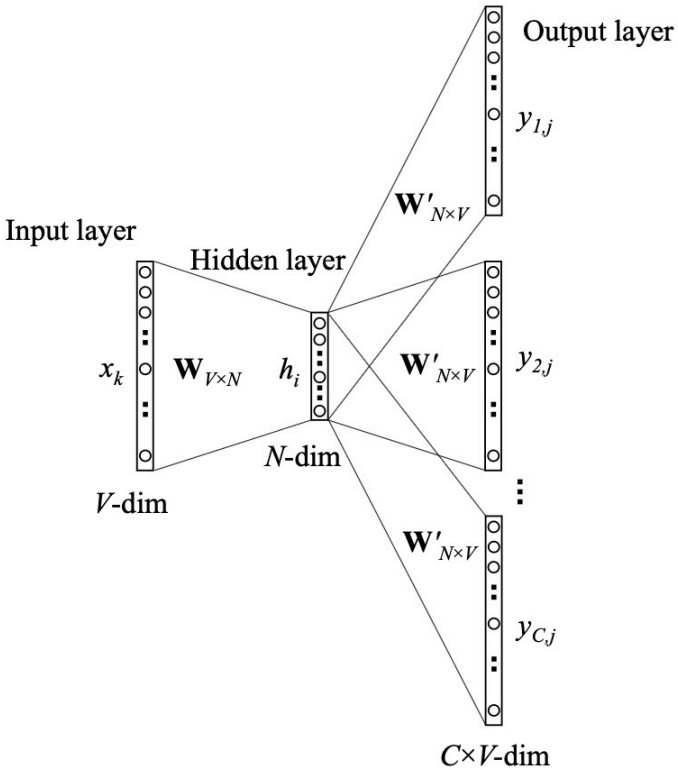
- 根据矩阵分解的角度看，网络的动态演化相当于原始矩阵的不断变化，网络的嵌入根据矩阵扰动理论更新。



# 基于 Skip-Gram 的嵌入方法

## Skip-Gram

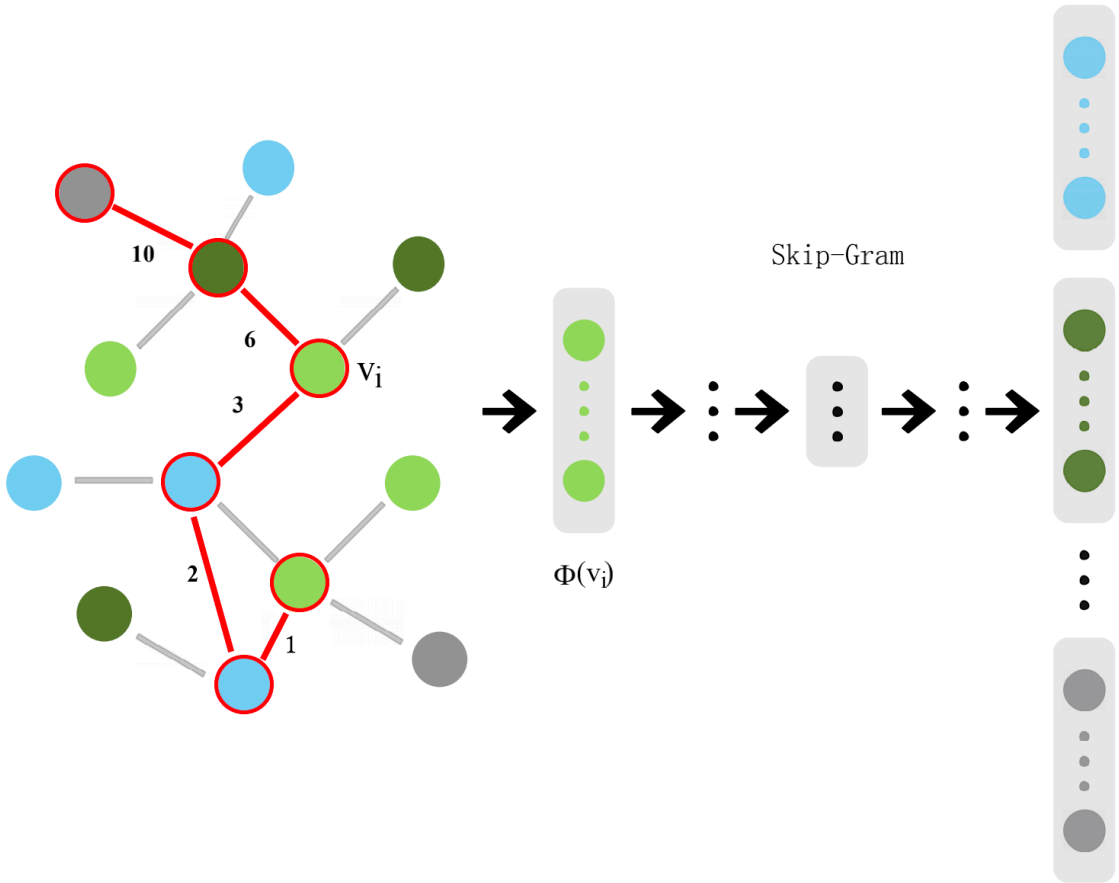
- Word2Vec 的一种经典模型，可以通过输入词来预测上下文。



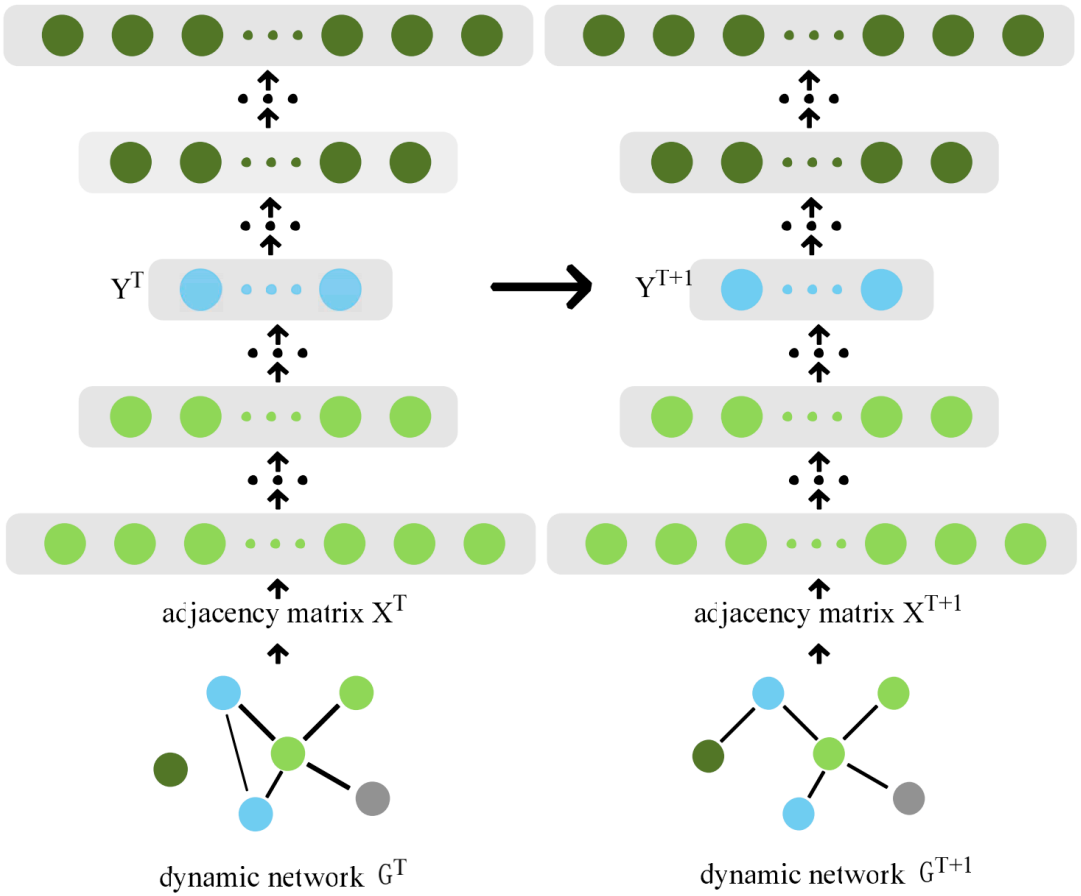

- DeepWalk 提出可以将节点看作为单词，游走序列则对应着句子。
- 由此，大量基于该模型的静态网络嵌入方法被提出。

# 基于 Skip-Gram 的嵌入方法

- 将动态网络看作连续模型，每条边有相对应的时间戳。
- 此时随机游走中的节点通过一系列时间戳递增的边连接。



# 基于自动编码器的嵌入方法

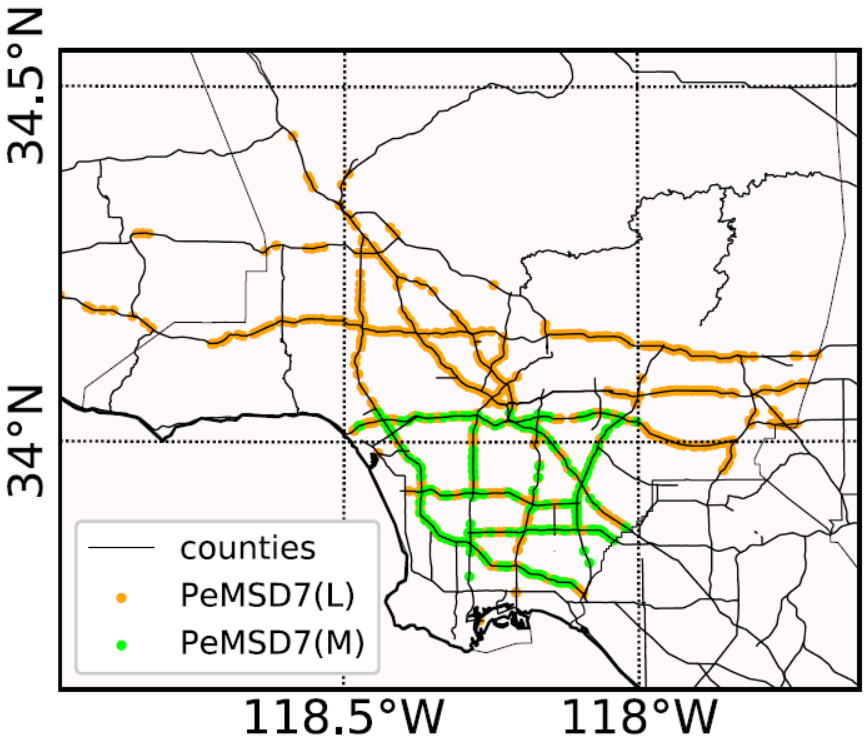



- 离散模型
- 将上一时刻 AE 的权重参数作为下一时刻网络的初始化

# STGCN-动态交通流量图预测

## 交通流量预测

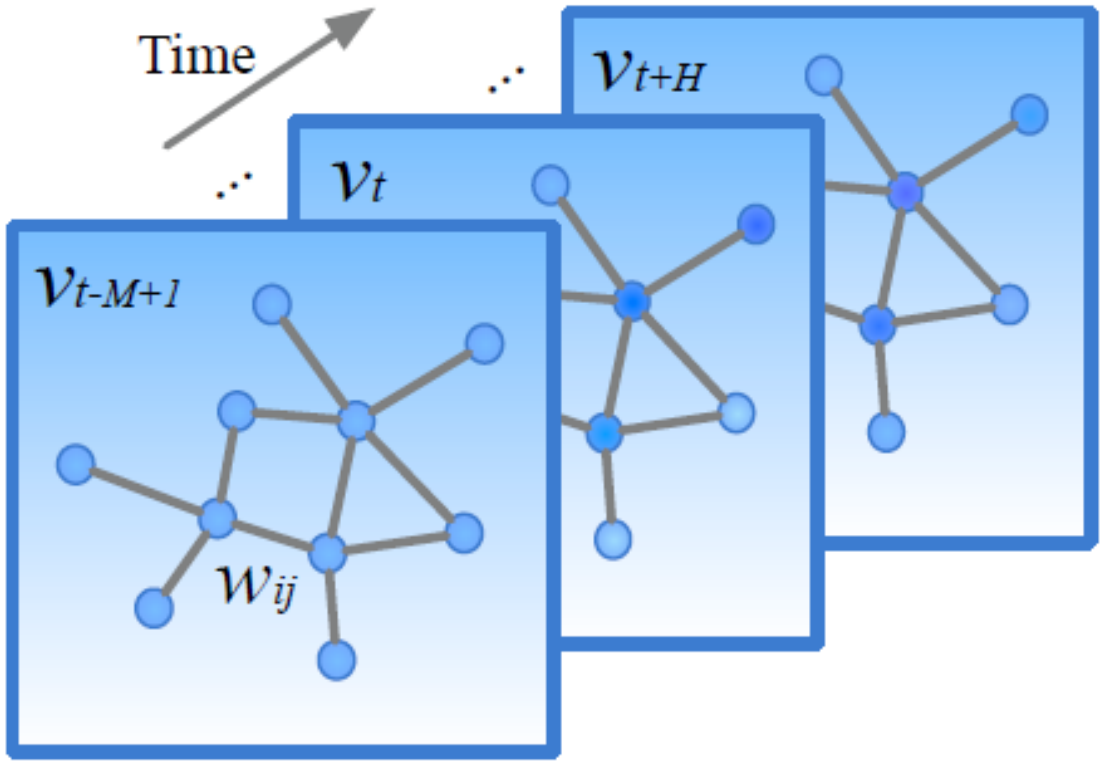
把路口的车流量看作图中节点的属性，道路看作连接节点的边，不同时刻的交通流量网就构成了经典的离散动态图。



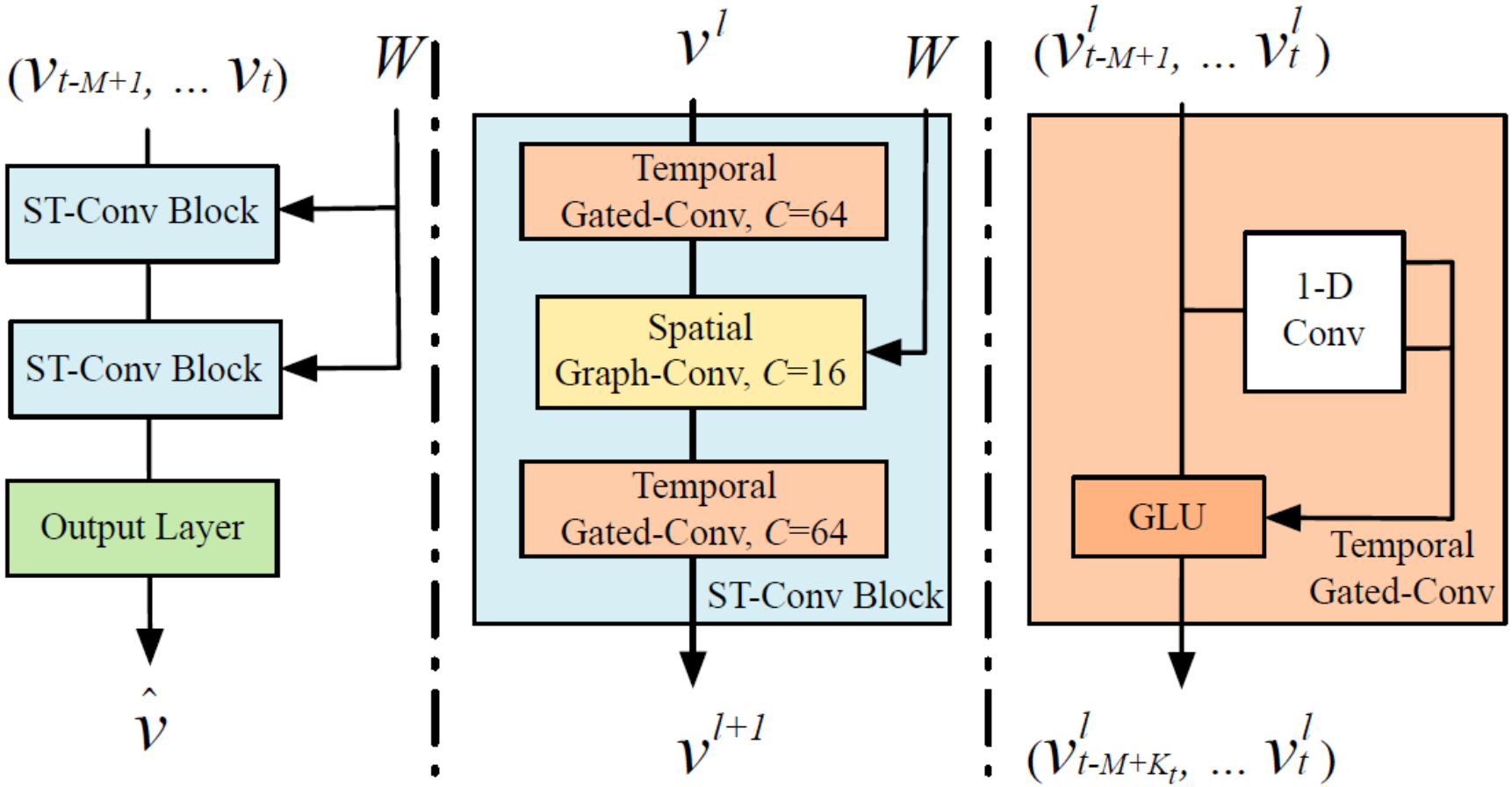
# STGCN-动态交通流量图预测

## 交通流量动态图预测

给定 $M$ 个有序的图 $\mathcal{G}_t = \left(W, X \in R^{M \times n \times C_i}\right)$，预测 t+1 时刻的图 $\mathcal{G}_{t+1} = \left(W, X \in R^{n \times C_i}\right)$ 个时间后节点的特征。



# STGCN-动态交通流量图预测



# STGCN-动态交通流量图预测

## 时域卷积块

- 输入 $X \in R^{M \times C_i}$，沿着时间维度进行一维卷积，卷积核 $\Gamma \in R^{K_t \times C_i}$，个数为 $2C_o$，从而得到 $[PQ] \in R^{(M-K_t+1) \times 2C_o}$，其中 $PQ$ 指向量在 channel 维度的的前后两部分。
- 然后进行 $GLU$(gated linear units) 激活: $\Gamma *_\tau X = P \odot \sigma(Q) \in R^{(M-K_t+1) \times C_o}$，$\odot$ 指两个矩阵对应元素相乘的运算。

对于一张完整的时空图: 输入 $X \in R^{M \times n \times C_i}$，输出 $Y \in R^{(M-K_t+1) \times n \times C_o}$。

# STGCN-动态交通流量图预测

## 空域卷积块

- 空域卷积在每个时间步的图上进行 (不在时间步之间进行)。
  输入

$$X \in R^{n \times C_i}$$

  按照 ChebGCN，输出

$$Y = \sum_{i=0}^{K-1} \theta_i T_i(\tilde{L}) X$$

- 其中 $T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x)$ $\tilde{L} = \frac{2L}{\lambda_{\max}} - I_n$ ，$L = I_n - D^{-1/2} A D^{-1/2}$ ，$A$ 是邻接矩阵，论文中 $K = 3$ 。卷积核 $\Theta \in R^{K \times C_i}$ ，个数为 $C_o$ 。对于一张完整的时空图:
- 输入 $X \in R^{M \times n \times C_i}$ ，输出 $Y \in R^{M \times n \times C_o}$ 。

# STGCN-动态交通流量图预测

- 经过了时域卷积块和空域卷积块后，原本的节点属性维度 $\mathcal{V}$ 已经不再是简单的流量特征，而是通过卷积聚合了时空信息的全新特征。

## 时空卷积块的输出层

- 根据时域卷积块的一维卷积，每经过一个时空卷积块，数据在时间维度的长度减小 $2\left(K_t-1\right)$ 。所以经过两个时空卷积块后，得到 $Y \in R^{(M-4(K_t-1))\times n\times C_o}$ 。
- 输出层包括一个时域卷积层和一个全连接层，时域卷积层的卷积核大小 $\Gamma \in R^{(M-4(K_t-1))\times C_o}$ ，个数为 $C_o$ ，将输出映射到 $Z \in R^{n\times C_o}$ 。全连接层 $\hat{v}=Zw+b$ ，其中 $w \in R^{C_o}$ ，于是输出 $\hat{v} \in R^n$ 。

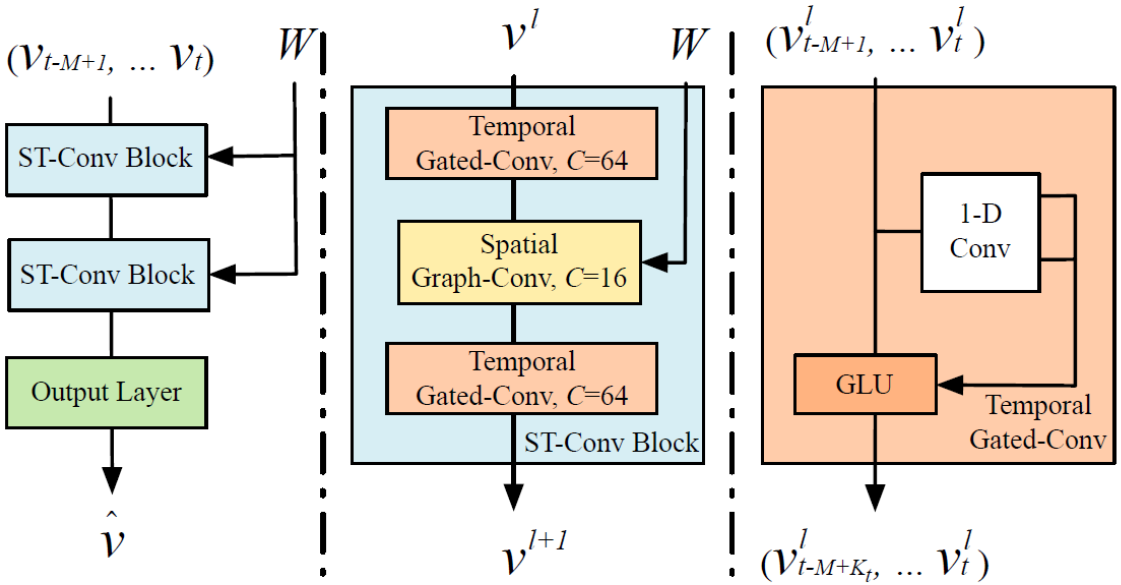
# STGCN-动态交通流量图预测

## 训练

- 损失函数是预测值和真实值的距离度量:

$$L\left(\hat{v}; W_\theta\right) = \left\|\hat{v}\left(v_{t-M+1}, \cdots, v_t, W_\theta\right) - v_{t+1}\right\|^2$$

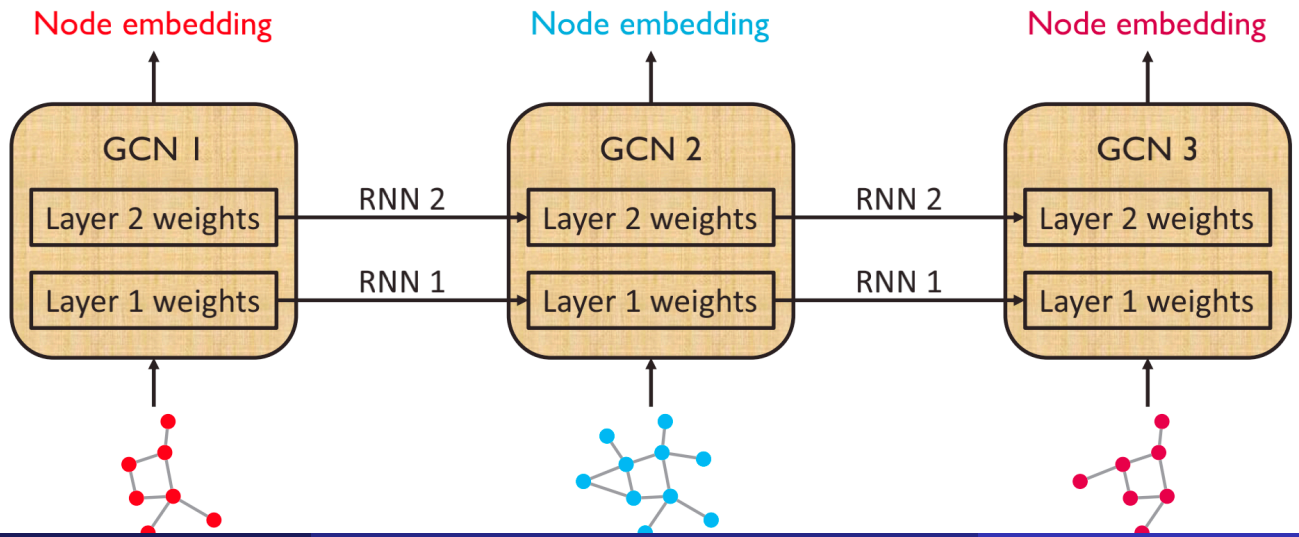
其中 $W_\theta$ 是所有可训练参数，$\hat{v}$ 是预测值，$v_{t+1}$ 是真实值。



# Evolvegcn: Evolving graph convolutional networks for dynamic graphs[3]

## GNN 与 RNN 的结合

- 常见的方式是使用 GNN 作为特征提取器，使用 RNN 从提取的节点特征中学习动态。
- 而 EvolveGCN 则着眼于参数，使用 RNN 来演化 GNN 的参数，从演化的网络参数中捕获动态。



# EvolveGCN

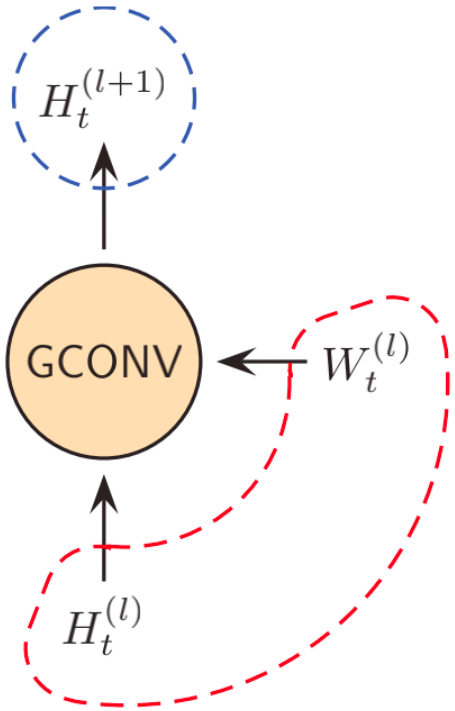
## 核心思想

- 在离散模型中，对于每个静态网络都需要一个 GCN 模型来训练，从而学到每一层的权重 $W$。
- 对于动态的场景，当前时刻 GCN 模型的权重与上一时刻 GCN 模型权重有关。
- 因此，如果我们将各个时刻 GCN 每层的模型参数看做一个序列，就可以从 RNN 来学习动态信息。

# EvolveGCN

- 在时刻 $t$，第 $l$ 层以邻接矩阵 $A_t$ 和节点表征矩阵 $H_t^l$ 作为输入，通过权重矩阵 $W_t^l$ 更新节点表征矩阵 $H_t^{(l+1)}$，将其作为输出:

$$H_t^{(t+1)} = \mathbf{GCONV}(A_t, H_t^l, W_t)$$



# EvolveGCN

如何将每个时刻 $t$ 模型第 $l$ 的参数 $W_t^l$ 应用于 RNN，EvolveGCN 给出了两个方案：

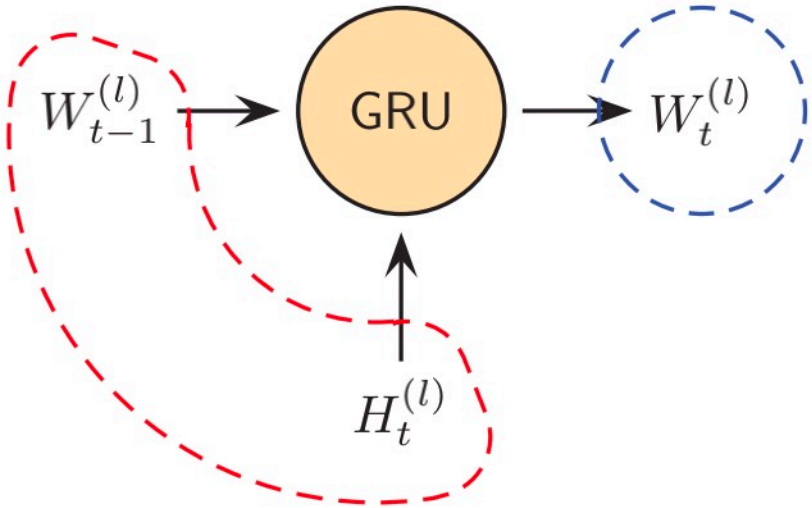
- EvolveGCN-H：将 $W_t^l$ 作为序列数据的隐藏状态，使用 GRU 来更新 $W_t^l$。
- EvolveGCN-O：将 $W_t^l$ 作为序列数据的当前输出，也作为下一时刻的输入，通过 LSTM 来建模。

# EvolveGCN-H

## EvolveGCN-H

- 将 $W_t^l$ 作为序列数据的隐藏状态，使用 GRU 来更新 $W_t^l$。

$$\overbrace{W_t^l}^{\text{GCN weights}} = \underbrace{W_t^l}_{\text{hidden state}} = \mathbf{GRU}\Big(\overbrace{H_t^l}^{\text{node embeddings}}, \underbrace{W_{t-1}^l}_{\text{hidden state}}\Big)$$

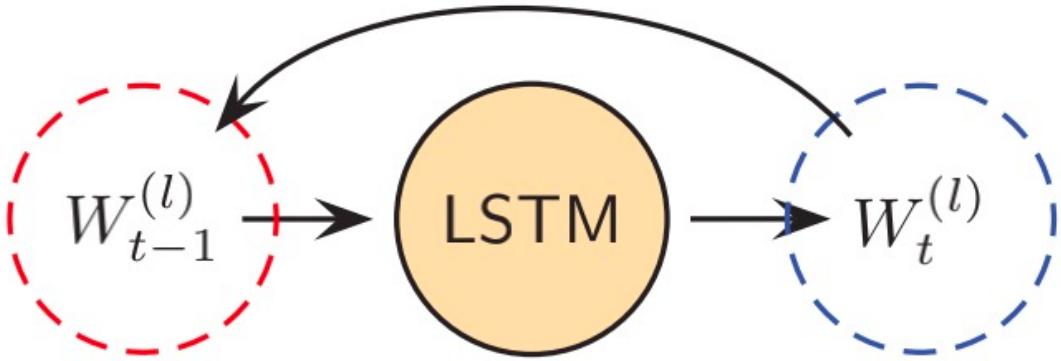


# EvolveGCN-O

## EvolveGCN-O

- 将 $W_t^l$ 作为序列数据的当前输出，也作为下一时刻的输入，通过 LSTM 来建模。

$$\overbrace{W_t^l}^{\text{GCN weights}}_{\text{output}} = \mathbf{LSTM}\Big( \overbrace{W_{t-1}^l}^{\text{GCN weights}}_{\text{input}} \Big)$$



## EGCU

将上述 GCN 单元和 RNN 单元结合，可以组合为新的演化图卷积单元（Evolving Graph Convolution Unit，EGCU）。

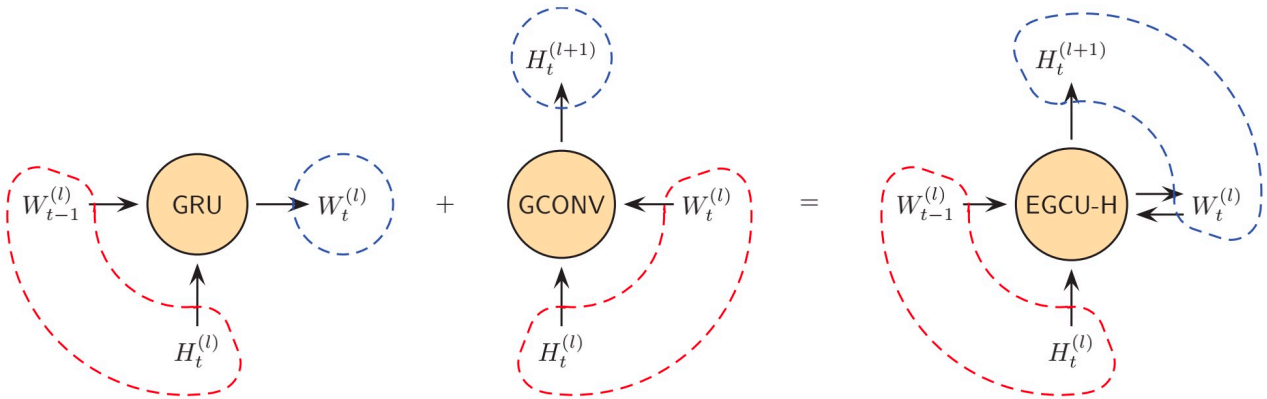
- EGCU-H：

**function** $[H_t^{(l+1)}, W_t^{(l)}] = \text{EGCU-H}(A_t, H_t^{(l)}, W_{t-1}^{(l)})$

$W_t^{(l)} = \text{GRU}(H_t^{(l)}, W_{t-1}^{(l)})$

$H_t^{(l+1)} = \text{GCONV}(A_t, H_t^{(l)}, W_t^{(l)})$

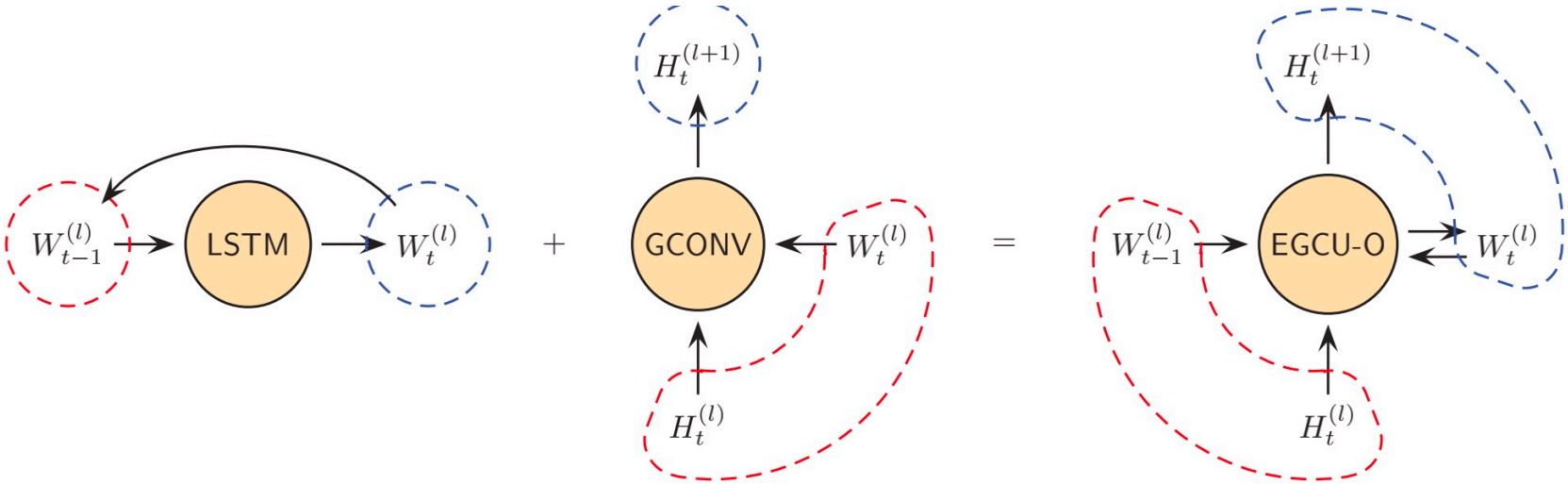
**end function**



# EGCU

- EGCU-O:

1: **function** $[H_t^{(l+1)}, W_t^{(l)}] = \text{EGCU-O}(A_t, H_t^{(l)}, W_{t-1}^{(l)})$
2: $\quad W_t^{(l)} = \text{LSTM}(W_{t-1}^{(l)})$
3: $\quad H_t^{(l+1)} = \text{GCONV}(A_t, H_t^{(l)}, W_t^{(l)})$
4: **end function**



# 不同方案的抉择

## 两种方案

EvolveGCN-H:

$$\overbrace{W_t^l}^{\text{GCN weights}} = \underbrace{\mathbf{GRU}}_{\text{hidden state}}\Big(\overbrace{H_t^l}^{\text{node embeddings}}, \overbrace{W_{t-1}^l}^{\text{GCN weights}}\Big)$$

EvolveGCN-O:

$$\overbrace{W_t^l}^{\text{GCN weights}} = \underbrace{\mathbf{LSTM}}_{\text{output}}\Big(\overbrace{W_{t-1}^l}^{\text{GCN weights}}\Big)$$

- 当节点属性比较丰富时，H 方案会更有效，因为它的 RNN 中包含了额外的节点表征输入。
- 当网络的结构更加重要时，O 方案更加关注结构的变化，效果相对较好。

# 课程总结

1. 研究背景
2. 有向图神经网络
3. 异构图神经网络
4. 动态图神经网络

# References I


Anonymous (2022).
Chaos is a ladder: A new understanding of contrastive learning.
In *Submitted to The Tenth International Conference on Learning Representations*.
under review.

Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T., and Leiserson, C. (2020).
Evolvegcn: Evolving graph convolutional networks for dynamic graphs.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5363–5370.


# References II


Salha, G., Limnios, S., Hennequin, R., Tran, V.-A., and Vazirgiannis, M. (2019).
Gravity-inspired graph autoencoders for directed link prediction.
In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 589–598.

Zhou, S., Wang, X., Ester, M., Li, B., Ye, C., Zhang, Z., Wang, C., and Bu, J. (2021).
Direction-aware user recommendation based on asymmetric network embedding.
*ACM Transactions on Information Systems (TOIS)*, 40(2):1–23.