

AI 前沿讲座向世明老师作业

异质图神经网络

学生： 槐 泽 鹏

导师： X X X

学号:xxxxxxxxxxxxxxxxxxxx

联系方式: xxxxxxxxxxxxxxx

2021 年 11 月 3 日

目 录

第 1 节 背景	1
1.1 异质图及图神经网络	1
第 2 节 同质图神经网络	2
第 3 节 异质图神经网络	3
3.1 使用元路径	4
3.2 不使用元路径	6
3.3 基于子图	9
第 4 节 小结	11
第 5 节 参考文献	12

第 1 节 背景

本报告旨在完成 2020 年人工智能前沿讲座选做大作业中向世明老师布置的作业：“针对深度学习中某一项内容：本人选择图神经网络中的异质图神经网络，写一份综述报告”。

本人博士研究方向为“图数据挖掘”，其中“异质图神经网络”是基础性工作，即使用异质图神经网络来得到节点表示，之后用于下游任务如推荐等，因此本次报告拟对“异质图神经网络”做一个综述，同时也是近期调研工作的一个总结。

1.1 异质图及图神经网络

异质图又称异质信息网络 (Heterogeneous Graph Embedding, HIN)，是指在图中包含不同的节点类型和边类型。目前公认的是，异质图是一类对现实世界复杂问题有较强表征能力的模型，即现实世界中许多场景和应用可以转化为异质图结构化数据，例如商品推荐、基因图谱、社交网络等。由于 HIN 的普遍性，因此异质图的表示学习是基础性关键工作，获取准确和准确表达语义的表示向量有利于下游任务的更好地完成。

图神经网络 (Graph Neural Network, GNN)，起源于 CNN 和图表示学习。1 传统的卷积网络只能处理规则的欧式空间数据，例如图像，文本等，但是现实问题中很多数据是不规则的，是非欧式空间的，这两类情况区别的一个典型示意图如图 1.1 所示。2 图表示 (graph representation learning, GRL) 是图结构化数据的基础性工作，旨在对每个节点得到一个低维的向量来表示其语义或完成相应任务如节点分类。传统的 GRL 方法有 Deepwalk^[10]、node2vec^[4] 等，这些方法都是将图上节点先采样 (采样方法多基于随机游走)，再按照 word2vec^[7,8] 中 SkipGram 思想去训练。综上，GNN 以得到图节点表示为目的，同时借鉴 CNN 中核心要素：局部连接、参数全局共享和多层网络^[21]，模型思想是聚合邻居节点 (不一定是节点，可以是 path、subgraph 等) 信息，最终得到节点表示向量。

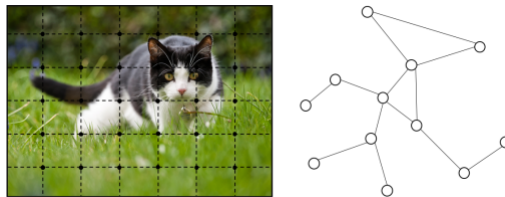


图 1.1 规则和非规则数据示意图

如何将 GNN 应用于异质图？这需要从同质图神经网络说起。同质图是指图中节点和边均为同类，GNN 在同质图上基本采取“采样-聚合-更新-输出”的架构，而在聚合步骤中，同质图聚合的均是同类节点，但异质图周围有很多异质节点，这为异质图神经网络带来了许多困难但也带来了研究空间（比如采用不同的方式处理 or 聚合异质节点，可以将异质图转化为同质图，可以直接聚合，可以分层聚合等等）。

为了帮助理解，下面我们首先简单介绍一些经典同质图神经网络模型，之后再重点介绍异质图神经网络。

第 2 节 同质图神经网络

这里我们介绍 3 个经典的全质图上的 GNN 模型。

GCN^[13] 是 GNN 的开山之作。GCN 是谱方法的图神经网络，这个方法来源是在傅里叶空间里进行卷积操作，因此需要将节点表示通过过滤器进行谱变换，在谱空间卷积操作完之后，再逆傅里叶变换回来得到节点表示；但是这一方法数学公式十分复杂，GCN 之前已有一些工作对这一数学推到过程进行了简化，但是最终，还是 GCN 将其简化到了一个被广为接受的形式，如式 1 所示。其中， X 和 Z 分别是更新前、后的节点表示，GCN 模型示意图如图 2.1 所示。

$$Z = \hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} X \Theta \quad (1)$$

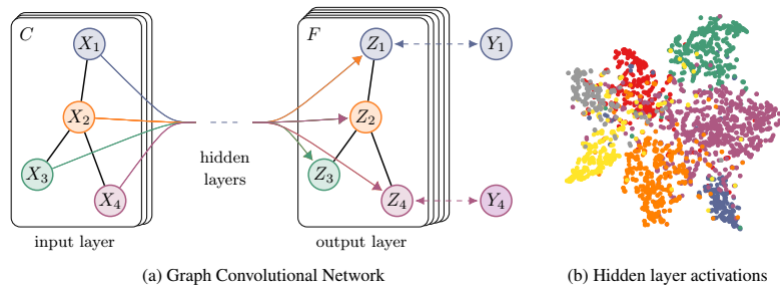


图 2.1 GCN 示意图

GraphSage^[5] 是空域 GNN 的开山之作。相对于频域 GNN，空间域 GNN 具备 inductive(中文不太好翻译这个词，大致意思是不用预先知道图的全部拓扑结构，可以随时随地得生成节点表示) 表示能力，并且空间域简单，可解释性强，目前为止绝大多数 GNN 均是空域方法。GraphSage 经典的诠释了“聚合——更新”这一 GNN 通过聚合邻居节点来生成中心节点的思想。GraphSage 核心公式

如式 2所示。聚合时采用了平均池化 (这也为 attention 留下伏笔)。GraphSage 模型示意图如图 2.2所示。

$$\begin{aligned} h_{N(v)}^k &\leftarrow \text{AGGREGATE}_k(h_u^{k-1}, \forall u \in N(v)) \\ h_v^k &\leftarrow \sigma(W^k \cdot \text{CONCAT}(h_v^{k-1}, h_{N(v)}^k)) \end{aligned} \quad (2)$$

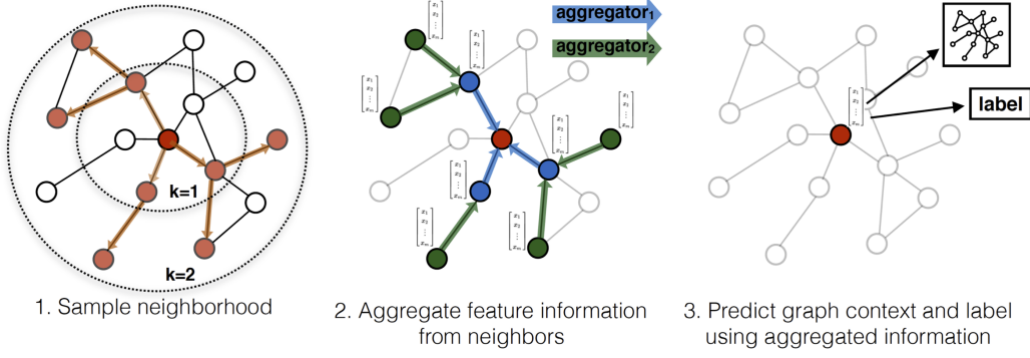


图 2.2 GraphSage 示意图

最后介绍 GAT^[15]。GAT 是同质图巅峰和经典之作，目前异质图神经网络里大量使用了 GAT 的思想。GAT 核心 idea 是在 GNN 中引入了 attention 机制，其 motivation 是不同邻居节点对中心节点的影响是不同的。因此，先对邻居节点计算 attention 系数，最后加权聚合邻居节点，这一过程相较于 mean-pooling 大大改变了聚合方式并提升了节点表示能力。GAT 核心公式就是把 GrapgSage 式 2中的 AGGREGATE 换成 attention，其模型示意图如图 2.3所示。

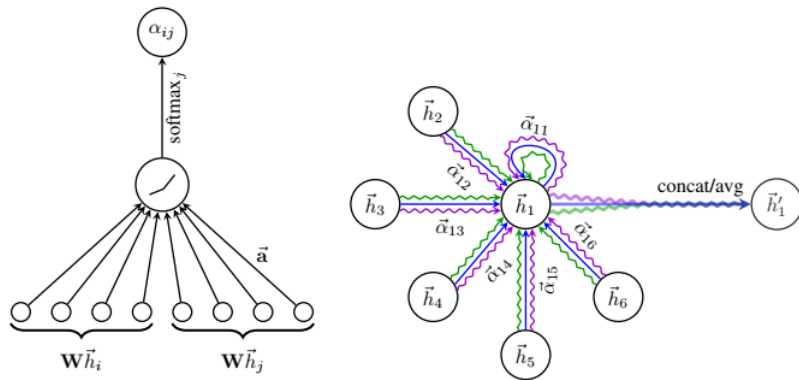


图 2.3 GAT 示意图

第 3 节 异质图神经网络

这里我们介绍三类模型：1 使用元路径 2 不使用元路径 3 基于子图信息。

3.1 使用元路径

元路径是指定义了一条路径的范式，即定义了一条路径上所有节点的类型，在这一范式下存在多个实例。

metapath2vec^[2] 是元路径异质图表示的开山之作。借鉴 word2vec 的 skipgram 思想，这里，metapath2vec 在元路径指导下按照随机游走策略在原图上进行采样，得到许多元路径实例后，对每一条元路径实例上的节点，按照 skipgram 去最大化其周围每类异质节点的共现概率，如式 3所示。其中， $N_t(v)$ 代表节点 v 周围类别为 t 的邻居节点。metapath2vec 的模型示意图和 word2vec 以及 node2vec 也很相似，如图 3.1所示。

$$\arg \max_{\theta} \sum_{v \in V} \sum_{t \in T_v} \sum_{c_t \in N_t(v)} \log p(c_t | v; \theta) \quad (3)$$

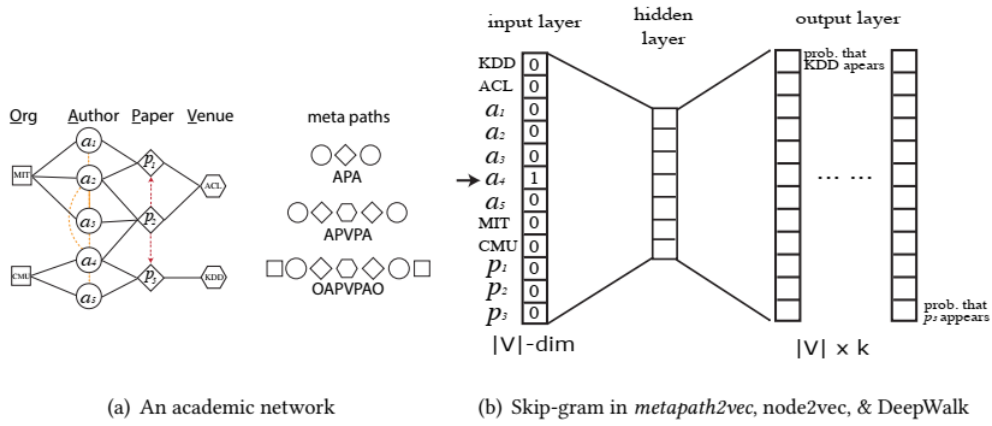


图 3.1 metapath2vec 示意图

Esim^[11] 和 metapath2vec 十分类似，也是在元路径指导下按照随机游走采样，再最大化同一个元路径实例上节点间的共现概率。但是与 metapath2vec 不同的是，Esim 没有采用 skipgram 策略去建模共现概率，而是自定义了两种条件概率，如式 4所示。其中，第一种是序列化条件概率，即认为每个节点只与自己左边 or 右边节点有关；第二种是成对概率，即计算每个节点与路径中所有其他节点的条件概率。实际上，Esim 定义的这两种条件概率与 skipgram 在本质上相同，都是为了计算同一个元路径下节点的共现概率。但是 metapath2vec 名气更大一些。

$$\begin{aligned}
& \arg \max_{\theta} \prod_{k=1}^L Pr(v_k | u_k, M_{k,k}, \theta) \\
& \arg \max_{\theta} \prod_{s=1}^L \prod_{t=s}^L Pr(v_t | u_s, M_{s,t}, \theta)
\end{aligned} \tag{4}$$

HIN2vec^[17] 是另一种计算元路径上节点共现概率的模型。HIN2Vec 将元路径视为“关系”，如文献图谱中 author1-paper-author2(代表同一篇文章的两个作者)，则 A1 和 A2 在“A-P-A”元路径下视为相邻节点。如此，HIN2Vec 定义了很多元路径范式，在此范式下去衡量两个节点是否相连，即每给一对节点进行一个二分类，其核心公式如式 5所示。从另一个角度，可以说 HIN2Vec 考虑的是一条元路径两端节点的共现概率。HIN2Vec 模型示意图如图 3.2所示。

$$o_{x,y,r}(x, y, r) = \begin{cases} P(r|x, y), & \text{if } L(x, y, r) = 1 \\ 1 - P(r|x, y), & \text{if } L(x, y, r) = 0 \end{cases} \tag{5}$$

$$P(r|x, y) = \text{sigmoid}(\sum W_x x \cdot W_y y \cdot W_r r)$$

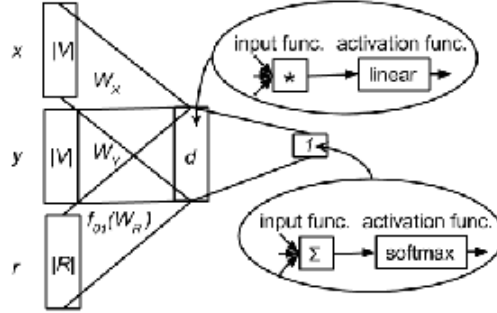


图 3.2 HIN2vec 示意图

可以说，以上模型都是为了最大化一条元路径实例上的节点的共现概率，即“相邻”（不一定是原图上拓扑相连的点，也可以是在自定义规则下相连的点）的同类节点应该具备相近的表示向量，这一思想是正确的，但仍存在局限性，即原图上相近的节点也可以标签差别很大，而更加准确或者完备的思想是：相同标签的节点应该具备相似的邻域（相邻节点因为拓扑接近所以具备相似邻域，因此具备相近的表示。因此前一思想是这一思想的特例。但后一思想可以区分开原图相邻但标签不同的节点，因为这两个节点的邻域信息存在部分相似和部分不相似，而前一思想无法解决部分相似和部分不相似的场景）。以下的模型则是后一思想的体现。

下面介绍两个非常经典的模型: HAN 和 MAGNN。

HAN^[16] 是发表在 2019 年 WWW 的工作。这篇第一次在异质信息网络中引入了 attention 机制。HAN 的 idea 非常经典：在元路径上寻找与中心节点同类的节点，然后聚合同类节点时采用两层 attention 机制：先聚合一个元路径范式下的所有实例，再聚合不同元路径得到的中心节点表示。可以说，HAN 开创了 HIN 聚合方式的新篇章，即使用多层 attention 机制直接聚合邻域信息，但是 HAN 只聚合同类节点，这也为后续模型留下了提升空间。HAN 模型示意图如图 3.3 所示。

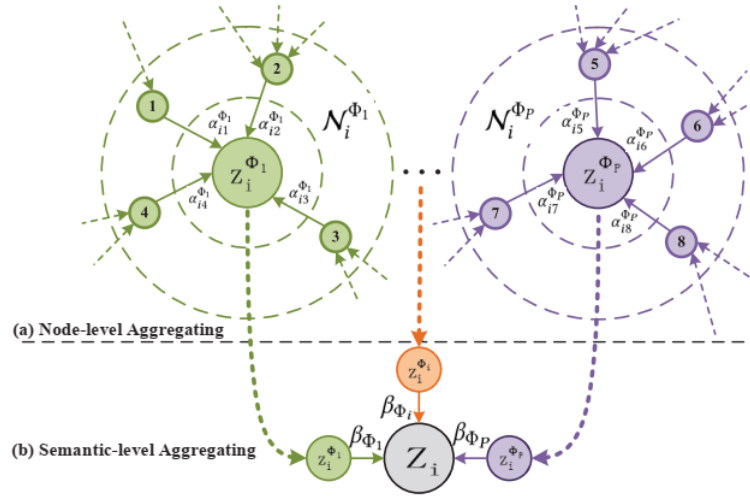


图 3.3 HAN 示意图

MAGNN^[3] 是截止今天 (2021.1.14) 效果最好的异质图神经网络模型。MAGNN 和 HAN 相比，其聚合了一条元路径上所有的节点，包含同类或异质节点。具体地说，就是对每条元路径实例进行编码，将一条元路径上所有异质节点信息综合为一个向量，再按照两层 attention 机制，对中心节点先聚合同一个元路径范式下的实例，再聚合不同元路径范式。MAGNN 可以说将元路径聚合几乎做到了极致，因为一条元路径上能聚合的信息也就这么多，除非出现其他的聚合框架或信息，MAGNN 的实验效果也是目前为止最好的结果。MAGNN 模型示意图如图 3.4 所示。

3.2 不使用元路径

PME^[1] 是不使用元路径而直接对异质节点间拓扑结构直接建模的经典模型。PME 认为每类关系两边的节点，投影至该关系对应的隐藏空间后，这两个点的距离应该接近。其核心公式如式 6 所示。其实类似于元路径路线中最大化共

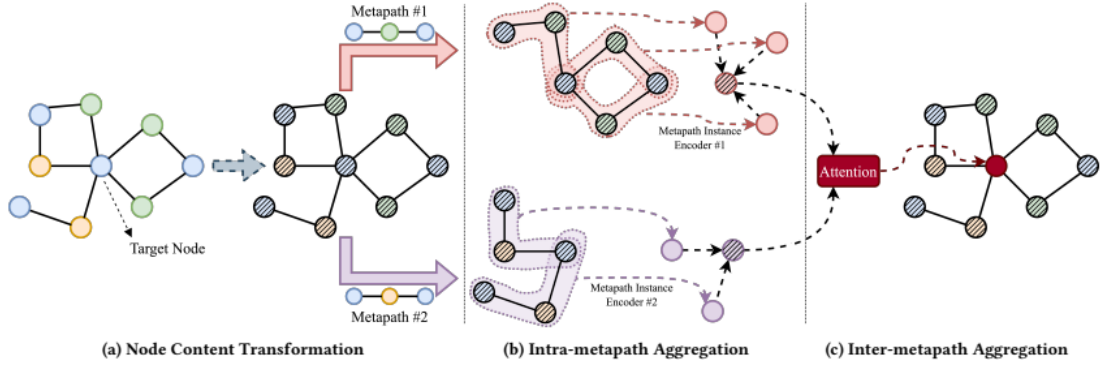


图 3.4 MAGNN 示意图

现概率的思想，这里最大化在原图中直接相连的两个异质节点的共现概率，只不过元路径方法采用元路径去考虑两个点是否相连，PME 直接乘上了一个与异质边类型相关的矩阵，但本质都是最大化共现概率。PME 模型示意图如图 3.5所示。

$$f_r(v_i, v_j) = w_{ij} \|M_r v_i - M_r v_j\|$$

$$\arg \min \sum_{(v_i, v_j) \in D_r} \sum_{(v_i, v_k) \notin D_r} [m + f_r(v_i, v_j)^2 - f_r(v_i, v_k)^2] \quad (6)$$

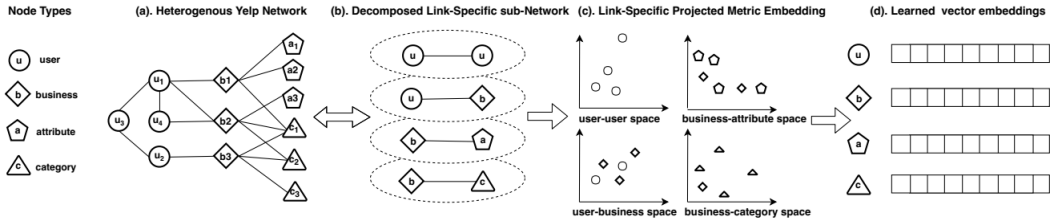


图 3.5 PME 示意图

HEER^[12] 是一篇与 PME 类似的文章，核心也是最大化共现概率，同样考虑在同一类型边两边的节点应满足：在“处理”后应“距离”接近。HEER 和 PME 不同的点在于处理方式和距离定义。PME 是采取矩阵投影，距离是欧式距离。HEER 是采用神经网络来计算两节点之间的关系向量，再使该类型边的所有两边节点均尽可能靠近同一个向量，如式 7所示。HEER 模型示意图如图 3.6所示。

$$g_{uv} = f(u, v)$$

$$\arg \max \sum_{R(u,v) \in r} \mu_r^T g_{uv} \quad (7)$$

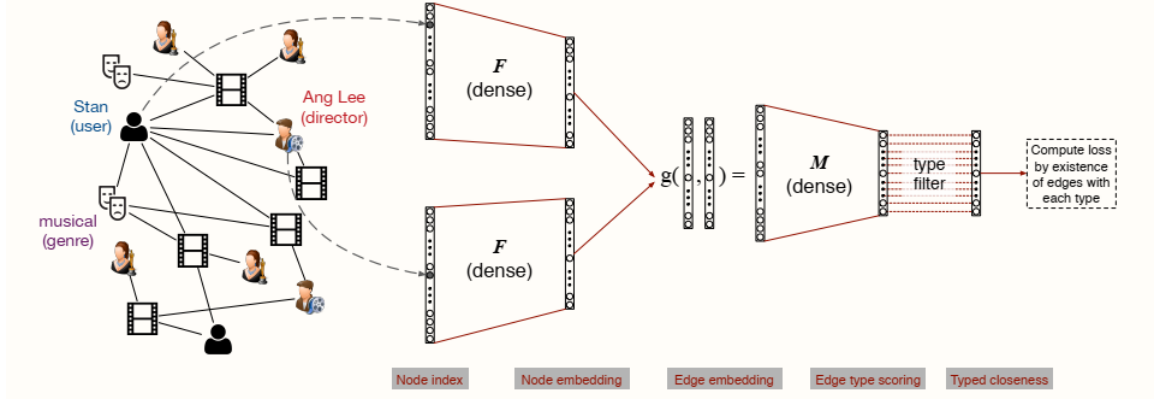


图 3.6 HEER 示意图

JUST^[6] 是另一篇以最大化共现概率为 loss 设计准则的文章。但是 JUST 采用和 metapath 同样的计算共现概率的准则：skipgram，但是 JUST 没有采用元路径采样，采用了自定义的采样准则：JUp and STay。对比的说，1JUST 和 PME、HEER 相比，相同的是：均没有采用元路径；不同的是：PME 和 HEER 直接处理原图一条边相连的两个顶点，JUST 基于 JUST 采样策略得到节点序列，同时 PME、HEER 贡献概率采用同类边最大化思想，而 JUST 采用 skipgram。2JUST 和 metapath2vec 相比，相同的是：均采用 skipgram 计算共现概率；不同的是：JUST 采用自定义策略得到节点序列，metapath2vec 采用元路径采样。可以说，JUST 开启了自定义采样规则的篇章。JUST 采样策略如图 3.7所示。

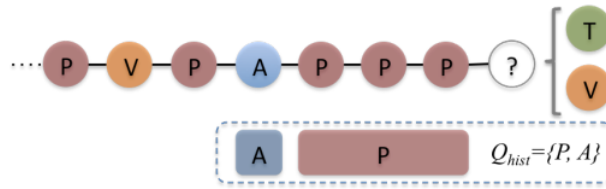


图 3.7 JUST 采样策略示意图

HetGNN^[18] 是不使用元路径的典范，这篇旨在解决一个新问题：异质节点具备多模态特征，如图像，文本、属性等信息。HetGNN 是将多层聚合机制发挥到了极致：第一层聚合是对节点的多模态特征采用双向 LSTM 进行特征提取；

第二层是对中心节点的每一类邻居节点通过平均池化进行聚合；第三层是对中心节点的不同类邻居节点通过 attention 进行聚合。这三层聚合成功的解决了具备多模态特征信息的异质图表示问题。此篇是多层聚合的典范，和推荐系统中 DHGAT^[9] 在多层聚合中有相似思想。同时，HetGNN 采用随机游走采样，并没有采用元路径。HetGNN 模型如图 3.8所示。

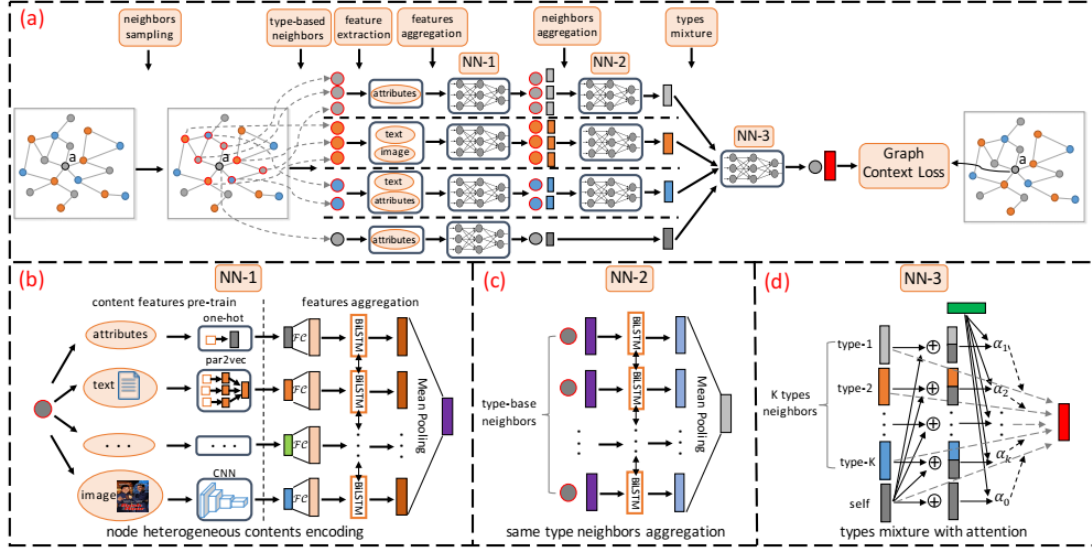


图 3.8 HetGNN 示意图

3.3 基于子图

基于子图是指，在异质图上基于局部子图满足的规律，来学习节点表示。以上其实有两个规律：原图拓扑相邻的节点具有相近的表示向量，相似的节点应该具有相似的邻域。这两个思想都是节点级的，都还没有去保留/还原/编解码局部子图的信息。可以说，如何利用好子图信息，不仅仅是异质图表示的未来趋势，更是整个图数据挖掘的重要方向，“如何利用子图信息”这一工作仍处于探索和百花齐放的阶段。

NSHE^[20] 处理子图的方式是：子图补全。NSHE 定义了原图范式 (类似于元路径，但是这个范式不是单一的串联路径形式)，在这个范式下，会有许多实例，每一个实例每次掩码掉一个节点，用其余节点去推测被掩码掉节点的概率，这一概率应该尽可能大。NSHE 利用子图的方式是：局部子图里的节点具有语义相关性，可用局部子图里的部分节点还原/推断出其余节点，这一思想也符合现实情境：如我们可以基于一个人在不同事情的表现去推测其在另一件事情上的选择和表现。但是 NSHE 的效果其实并没有 MAGNN 好，可以说子图思想的模型还

有待挖掘。NSHE 的模型如图 3.9所示。

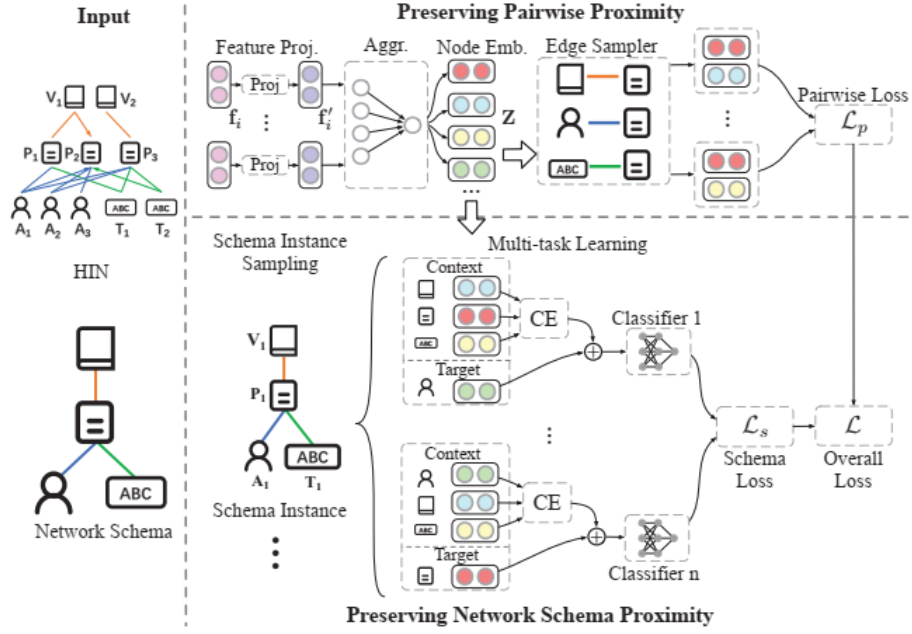


图 3.9 NSHE 示意图

mg2vec^[19] 处理子图的方式是：保持子图的拓扑结构。具体地说，mg2vec 定义了子图范式，同样的一个子图范式下存在很多实例。这里，mg2vec 不对实例编码，而是直接对范式编码（每一个范式对应一个嵌入）。之后，mg2vec 从两个方面去保持子图结构：一是每个节点处于不同的子图实例上，那么最大化该节点与其所处的子图范式的共现概率（此概率包含一个子图下实例的个数，个数越多代表该节点与该范式的关系越深）；二是选取一个节点对，考虑这对节点所处的所有子图实例，之后最大化这对节点与所处子图范式的共现概率。上述两个方面在原文中称为 1-阶和 2-阶信息。实际上，以上两个方面目的是尽可能保持子图的结构信息，一阶信息重点在保持节点周围子图结构，二阶信息重点在保持一对节点间的子图结构。可以说，mg2vec 是 metapath2vec、Esim、HIN2Vec、JUST 的升级版，前者旨在保持子图结构，后者旨在保持路径结构；而 PME 和 HEER 旨在保持节点连接的拓扑结构，因此这三类是：边-路径-子图的逐步升级过程，因此模型的效果肯定也是越来越好。值得一提的是，mg2vec 是自监督的训练方式。mg2vec 模型示意图如图 3.10所示。

DHNE^[14] 处理子图的方式是：保持超图 (hyper-networks) 的一阶二阶结构。具体地说，超图的一条边相当于一元路径，一阶结构是指不同类型的节点出现在同一类超边 (hyperedge) 上，因此一阶结构的 loss 是最大化一条超边上节点的

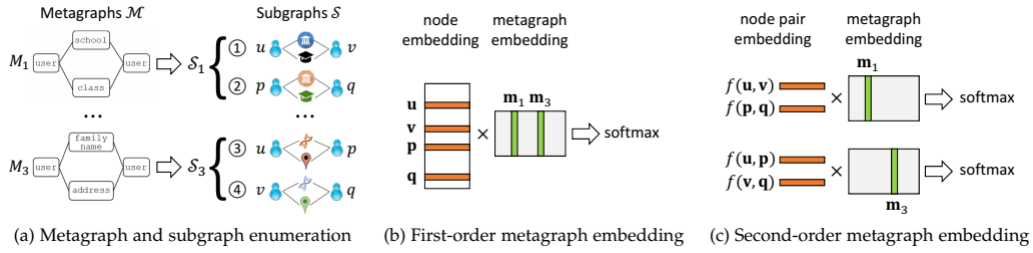


图 3.10 mg2vec 示意图

共现概率；二阶结构是指两个节点具有相同的超边邻居，这一关系采用重构连接矩阵 A 的 loss 表征。DHNE 一共三层模型：第一层通过编码器得到节点表示向量，第二层得到超边的向量，第三层是输出层判断一条超边是否存在。其中，第一层输出后跟了一个解码器用以还原二阶相邻信息的连接矩阵 A ，第三层输出代表一阶结构，最后两个 loss 线性加权得到最终 loss。DHNE 模型示意图如图 3.11 所示。

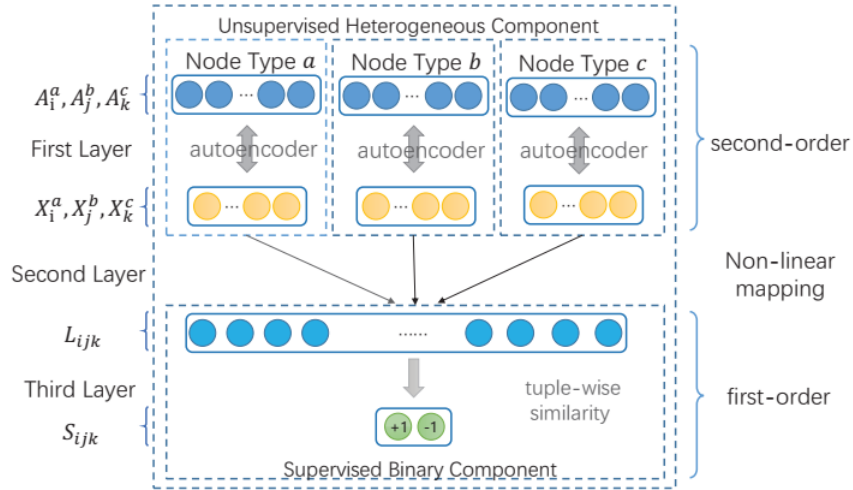


图 3.11 DHNE 示意图

其实，NSHE、mg2vec 和 DHNE 严格意义都不是 GNN 的思路，因为他们都对节点嵌入进行处理，通过处理子图信息 (如保持子图拓扑结构信息、补全子图等) 而不是聚合来更新嵌入。因此，正如此小节一开始提到的那样，如何利用子图信息还是一个值得深挖的点。

第 4 节 小结

本次大作业旨在对最近本人对异质图神经网络的调研做一个总结，同时完成 2020 年人工智能前沿讲座选做大作业中向世明老师布置的作业。图数据挖掘

中，图表示是核心的基础性工作，得到准确和高质量的节点表示是下游任务如推荐的必要前置环节。在图表示中，异质图神经网络是近些年最火的研究方向，从2017年开始，发表在KDD、WWW等顶会上的文章成指数级增长。因此，有必要做好基于异质图神经网络的图表示工作。这篇综述首先简单介绍了图表示和同质图神经网络，之后根据使用的信息介绍了三个类型的模型：使用元路径、不使用元路径和基于子图，较为系统地总结了近些年最新的研究成果，同时给出了个人心得体会。本人研究方向为“图数据挖掘和基于图的推荐系统”，欢迎交流讨论。(huaizepeng2020@ia.ac.cn)

第5节 参考文献

- [1] Hongxu Chen, Hongzhi Yin, Weiqing Wang, Hao Wang, and Xue Li. Pme: Projected metric embedding on heterogeneous networks for link prediction. In *the 24th ACM SIGKDD International Conference*, 2018.
- [2] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. *KDD*, 2017.
- [3] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *The World Wide Web Conference*, 2020.
- [4] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *the 22nd ACM SIGKDD International Conference*, 2016.
- [5] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. *NIPS*, 2017.
- [6] Rana Hussein, Dingqi Yang, and Philippe Cudré-Mauroux. Are meta-paths necessary?: Revisiting heterogeneous graph embeddings. In *the 27th ACM International Conference*, 2018.
- [7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *Computer Science*, 2013.
- [8] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26, 2013.

- [9] Xichuan Niu, Bofang Li, Chenliang Li, and Rong Xiao. A dual heterogeneous graph attention network to improve long-tail performance for shop search in e-commerce. *KDD*, 2020.
- [10] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. *KDD*, 2014.
- [11] Jingbo Shang, Meng Qu, Jialu Liu, Lance M. Kaplan, and Jian Peng. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. *KDD*, 2017.
- [12] Yu Shi, Qi Zhu, Fang Guo, and Chao Zhang. Easing embedding learning by comprehensive transcription of heterogeneous information networks. In *the 24th ACM SIGKDD International Conference*, 2018.
- [13] Kipf T. and Welling M. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.
- [14] Ke Tu, Peng Cui, Xiao Wang, and Fei Wang. Structural deep embedding for hyper-networks. In *AAAI*, 2018.
- [15] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *ICLR*, 2018.
- [16] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Peng Cui, P Yu, and Yanfang Ye. Heterogeneous graph attention network. In *The World Wide Web Conference*, 2019.
- [17] Tao yang Fu, Wang-Chien Lee, and Zhen Lei. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. *CIKM*, 2017.
- [18] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh Vijay Chawla. Heterogeneous graph neural network. In *the 25th ACM SIGKDD International Conference*, 2019.

- [19] Wentao Zhang, Yuan Fang, Zemin Liu, Min Wu, and Xinming Zhang. mg2vec: Learning relationship-preserving heterogeneous graph representations via metagraph embedding. *IEEE Transactions on Knowledge and Data Engineering*, PP(99):1–1, 2020.
- [20] Jianan Zhao, Xiao Wang, Chuan Shi, Zekuan Liu, and Yanfang Ye. Network schema preserving heterogeneous information network embedding. In *Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20*, 2020.
- [21] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. 2018.