

通用、可扩展的图卷积神经网络

——《web 信息处理》课程报告

姓名：王涵之

学号：2019000163

一、背景介绍：

图，作为计算机科学领域中一类重要的数据结构，提供了一种抽象表示事物之间关系的方法。图结构包含两类主要元素——“节点”和“边”，其中，“节点”常被用作表示各种事物，“边”被用作表示事物之间的关系，由此抽象出现实世界真实关系的表达形式，这对我们研究实际生活中复杂的关系网络提供了可能。尽管另一重要的数据结构“树”也含有“节点”和“边”两种元素，但是，图结构比树结构更具灵活性。图结构中节点的平等关系和自由的连边方式，使其可以表示出事物之间的多种关系形式，但树结构的表达能力却会因为自身定义而受到限制。比如，我们无法用树结构表示一个关系闭环，也很难在现实世界中找到一个占有绝对主导地位“根节点”。从这一点中，我们更能看出图结构在关系网络的表达方面具有的天然优势。

由于图结构在关系表达方面的出色特性，我们常将其应用在多种实际场景中。比如，在交通网络中，我们常用节点表示城市，边表示城市之间的道路，边上的权重表示道路的长短，进而延伸出图节点间的最短路径问题等，以此提升路径规划问题的效率。在论文引用网络中，我们常用节点表示论文，边表示论文之间的引用关系，进而帮助人们梳理论文体系、辨识重要论文等。在生命科学中的蛋白质运输路径研究领域，我们可以用节点表示蛋白质分子，用边表示蛋白质之间的物质运输关系，进而帮助研究人员梳理蛋白质的活动规律，识别具有相似功能的蛋白质等。在社交网络中，我们常用节点表示用户，用边表示用户之间的好友关系，由此刻画出用户之间的亲疏关系，进而展开社区发现、相似用户推断、兴趣产品推荐等领域的研究等。

近年来，根据图结构信息和节点和边上携带的特征信息进行表示学习与挖掘的问题吸引了研究者的广泛关注，相关理论、算法和应用系统相继涌现，研究成果日益丰富。但是与此同时，海量数据规模的复杂网络结构对现有的图表示学习研究带来了艰巨的挑战。为了能有效获取属性异质图数据所携带的结构信息和属性、特征信息，在进行属性异质图的表示学习与挖掘中，现有研究工作普遍会在图信息传播阶段，将初始给定的节点、边的属性和特征信息按照图结构沿邻边进

行聚合，进而得到高质量的节点表示向量，再将节点表示向量放入神经网络训练框架中进行训练。为了得到高质量的图信息传播结果，现有方法大多会选择一个适合的图节点邻近度衡量指标，通过计算图节点邻近度间接获得图信息的传播结果。但是，现有工作使用的图节点邻近度衡量指标各不相同，缺乏通用的节点邻近度计算范式来指导图信息的聚合过程，从而难以宏观理解图传播过程的核心，也不易提出通用的图信息传播优化算法，以统一提升所有图表示学习与挖掘过程中的图信息传播效率。

二、算法概述：

图节点邻近度用于衡量图上节点相对某一给定起始节点的相对距离。根据图学习理论，邻近度较高的节点普遍具有较高的相似性。以节点分类任务为例，由于节点邻近度的高低间接指示了图结构上节点间的相似关系，进而包含了各节点的类别信息，故而可以使用节点邻近度计算结果进行神经网络的节点分类训练。目前常用的属性异质图节点邻近度衡量指标主要包括：Personalized PageRank(PPR)、heat kernel PageRank(HKPR)、转移概率(transition probability)、Katz 等。由于不同的节点邻近度衡量指标可能会影响最后的预测结果，因此，目前的研究工作普遍会根据具体的下游任务，选择计算最合适的邻近度衡量指标，以获得图结构、属性信息的集成结果。所以，设计一个通用的节点邻近度计算范式对统一不同的邻近度计算步骤大有裨益。

在本文中，我们提出了如下的图节点邻近度计算范式，以概括所有主流的图信息传播框架：

$$Y = \sum_{i=0}^{\infty} w_i \cdot (D^{-a} A D^{-b})^i \cdot X.$$

其中， A 、 D 分别为图邻接矩阵和度数矩阵， X 为初始特征矩阵， w_i 是一个权重系数， a 、 b 是图传播的拉普拉斯系数。目前绝大多数图表示学习所依赖的图信息传播框架，均可归纳为上述计算范式：

- **Personalized PageRank(PPR)**：作为 PageRank 的特殊变形，PPR 在每步传播时，都以 α 的概率停止在当前节点，以 $1 - \alpha$ 的概率随机移动到当前节点的任一出邻居。PPR 的传播框架可以抽象为： $Y = \sum_{i=0}^{\infty} \alpha(1 - \alpha)^i (A D^{-1})^i \cdot X$ 。其中， A 、 D 分别为图邻接矩阵和度数矩阵， α 是转移概率参数。在图表示学习框架中，我们常将转移矩阵替换为对称归一化的邻接矩阵，即 $D^{-1/2} A D^{-1/2}$ 。因此，上式变为： $Y = \sum_{i=0}^{\infty} \alpha(1 - \alpha)^i (D^{-1/2} A D^{-1/2})^i \cdot X$ 。对比我们提出的通用范式，当设置 $a = b = -1/2$ ， $w_i = \alpha(1 - \alpha)^i$ ，则通用范式可以转化为基于 PPR 的传播框架。

- **热核 PageRank(HKPR):** 热核 PageRank 同样属于 PageRank 的变形,但与 PPR 不同的是, HKPR 传播过程中, 每一步的停止概率不一定相同。具体地, 给定热核常数 t , HKPR 传播过程中的游程长度 t 服从泊松分布 $P(t) = e^{-t} \frac{t^i}{i!}$, 其传播式可以抽象为: $Y = \sum_{i=0}^{\infty} e^{-t} \frac{t^i}{i!} (D^{-1/2} A D^{-1/2})^i \cdot X$ 。这里我们直接使用了对称归一化的邻接矩阵 $D^{-1/2} A D^{-1/2}$ 。当设置通用范式中的 $a = b = \frac{-1}{2}$, $w_i = e^{-t} \frac{t^i}{i!}$, 则可将通用范式转化为基于 HKPR 的传播框架。
- **转移概率(transition probability):** 区别于 PPR 和 HKPR 在每步转移时都有一定的停止概率, 转移概率通过指定具体的传播层数, 保证整个传播过程只在最后一部停止。如果指定传播层数为 k , 则对应的传播式可以抽象为: $Y = (D^{-1/2} A D^{-1/2})^k \cdot X$ 。当设置通用范式中的 $a = b = \frac{-1}{2}$, $w_k = 1$, $w_i = 0$ if $i \neq k$, 则可将通用范式转化为基于转移概率的传播框架。
- **Katz:** 作为一种常用的邻近度衡量指标, Katz 使用图结构上两点间的可达路经数来进行节点邻近度的衡量手段。其传播框架可以抽象为: $Y = \sum_{i=0}^{\infty} \beta^i (A)^i \cdot X$ 。其中, A 是图结构的邻接矩阵, β 是收敛参数, 为了保证上式的收敛性, 我们常设置 $\beta < \frac{1}{\lambda_1}$, λ_1 是矩阵 A 的最大特征值。为了将上式统一到通用传播范式中, 我们可以设置 $a = b = 0$, $w_i = \beta^i$ 。指的注意的是, 基于 Katz 的信息传播框架不满足 $a + b = -1$ 的要求。因此, 现有通用范式均无法满足这一情形。

基于该通用计算范式, 本文提出了一种能统一绝大多数图节点邻近度衡量指标的计算范式, 并给出通用的计算方法。具体而言, 本文使用逐层传播的方法, 确定性计算各层的传播结果 $(D^{-a} A D^{-b})^i \cdot X$, 并在各层传播完成之后, 与设置的权重参数 w_i 相结合, 汇总得到最终的传播结果。同时, 通过引入随机性可以进一步加速上述过程的计算效率。

我们首先回顾一下现有的特征逐层传播方法的运行过程, 之后再具体介绍引入随机性加速特征传播的优化细节。给定特征向量 \vec{x} , 从其各非零项出发进行逐层特征传播。更具体地, 对于特征传播的起始节点 t , 其对图上每个节点 $\forall u \in V$ 都维护两个变量: $\text{residue } r^b(u, t)$ 和 $\text{reserve } \pi^b(u, t)$ 。其中, $\text{residue } r^b(u, t)$ 记录从节点 u 出发的 α -衰减随机游走走到节点 t 的概率, $\text{reserve } \pi^b(u, t)$ 记录从节点 u 出发的 α -衰减随机游走停止在节点 t 的概率。每次选取当前 $\text{residue } r^b(u, t)$ 最大的节点 u 进行 Push, 对节点 u 所有入邻居的 residue 和 reserve 进行更新, 最后用

各节点的 reserve 作为该节点 PPR 的估计值。由于每次 Push 都需要遍历当前节点的所有入邻居，因此整个过程的时间消耗为 $\tilde{O}\left(\frac{\bar{d}}{\delta}\right)$ ，其中 \bar{d} 为节点的平均度。

在上述逐层传播方法的基础上，本文进一步引入随机性，灵活选取各层需要 push 的节点，减少时间消耗。该方法的核心思想为：每次 Push 时，不需要更新所有入邻居的 reserve ，只需更新那些 reserve 增长超过阈值的入邻居，从而避免时间复杂度中 $O(\bar{d})$ 系数的出现。具体而言：对于当前需要进行 Push 操作的节点 u ，每个入邻居 v 的 reserve 期望增长值为 $\frac{(1-\alpha)\pi_l(u,t)}{d_{out}(v)}$ 。可以注意到，该增长值和入邻居 v 的出度成反比，对于出度较大的入邻居，这次 push 对其 reserve 的改变不大。因此，我们只在入邻居 v 的出度满足 $d_{out}(v) \leq \frac{(1-\alpha)\pi_l(u,t)}{\alpha\delta}$ （即 reserve 的增长幅度 $\frac{(1-\alpha)\pi_l(u,t)}{d_{out}(v)} \geq \alpha\delta$ ）时，才确定进行 push 操作。对于其他出度较大的节点，我们使用随机抽样的方法，产生一个 0、1 之间的随机数 r ，如果某一入邻居 v 的出度满足 $d_{out}(v) \leq \frac{(1-\alpha)\pi_l(u,t)}{r \cdot \alpha\delta}$ （即 reserve 增长幅度 $\frac{(1-\alpha)\pi_l(u,t)}{d_{out}(v)} \geq r \cdot \alpha\delta$ ），则增加 $\alpha\delta$ 到入邻居 v 的 reserve 中；否则，不进行本次 push。这样，对于每个要进行 Push 操作的节点 u ，我们只确定地更新了一部分入邻居、采样了一部分入邻居进行更新以保证结果的无偏，而对于剩下出度较大的入邻居，即使进行本次 Push，其 reserve 的变化幅度也低于误差要求，因此我们放弃对该入邻居的 push 以节省时间。

值得注意的是，在上述过程中，我们不能逐个查看每个入邻居，以判断其出度是否符合要求，因为这样也同样会消耗 $O(\bar{d})$ 的时间。我们需要将每个节点的入邻居都按照出度递增的顺序排序，在判断具体要更新当前节点 u 的哪些入邻居时，我们按照出度递增的顺序扫描节点 u 的入邻居，先判断当前入邻居是否满足确定进行 push 操作的条件要求，因为确定进行 push 操作对出度的要求最严格。如果不满足，再判断是否满足借助随机数 r 放宽后的出度要求。如果仍然不满足，则其他入邻居节点也可以直接放弃查看，因为入邻居的出度按照扫描顺序递增。我们可以在读图时完成对入邻居的排序，对于相对误差约束阈值 δ 和随机选取的宿点，上述过程可以在 $\tilde{O}\left(\frac{1}{\delta}\right)$ 的时间内一维特征的传播。

三、实验结果：

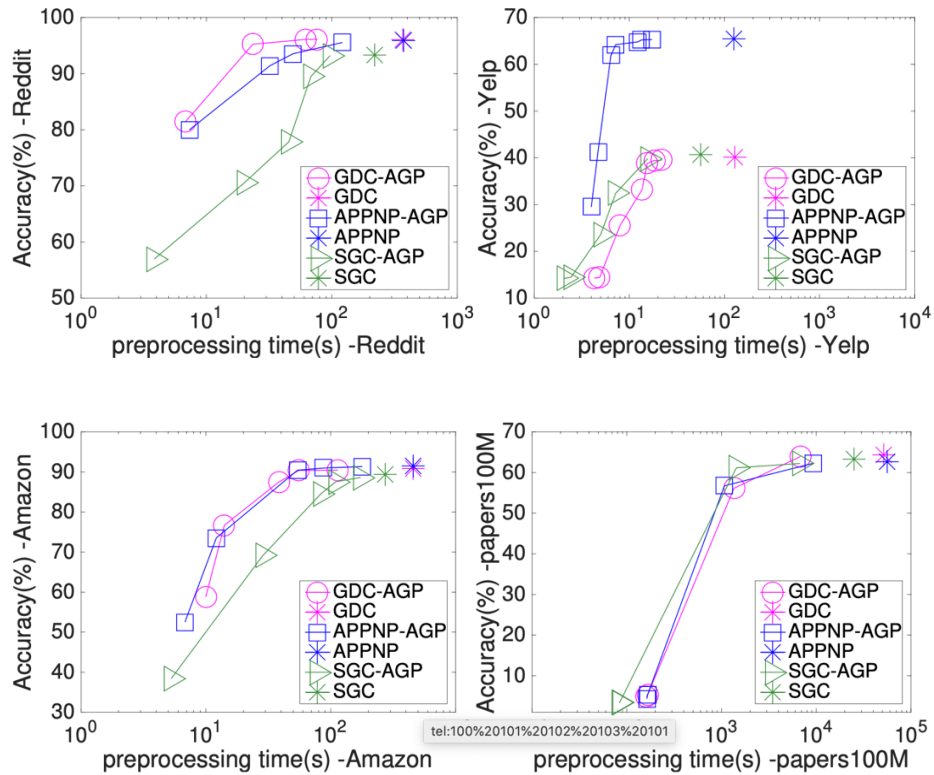
在实验环节，我们选择了 4 个真实图数据，其具体信息如下表所示：

Data Set	n	m	d	Classes	Label %
Reddit	232,965	114,615,892	602	41	0.0035
Yelp	716,847	6,977,410	300	100	0.7500
Amazon	2,449,029	61,859,140	100	47	0.7000
Papers100M	111,059,956	1,615,685,872	128	172	0.0109

在这 4 个数据集上，我们分别运行了现有算法 SGC（按照转移概率框架进行特征传播）、APPNP（按照 PPR 框架进行特征传播）、GDC（按照 HKPR 框架进行特征传播），以及使用我们方法加速后的方法 SGC-AGP、APPNP-AGP 和 GDC-AGP。具体的参数设置方式如下表所示：

Data set	Learning rate	Dropout	Hidden dimension	Batch size	GDC-AGP	APPNP-AGP	SGC-AGP
					t	α	L
Yelp	0.01	0.1	2048	$3 \cdot 10^4$	4	0.9	10
Amazon	0.01	0.1	1024	10^5	4	0.2	10
Reddit	0.0001	0.3	2048	10^4	3	0.1	10
Papers100M	0.0001	0.3	2048	10^4	4	0.2	10

实验效果如下：



Accuracy 为节点分类的 F1-score。可以看到，对于各种传播框架，我们均可以使用通用方法完成特征传播过程加速，上述 AGP 方法可以在更少的时间内达到相同 accuracy 的节点分类结果。