

GCN 长什么样

GCN 的公式看起来还是有点吓人的，论文里的公式更是吓破了我的胆儿。但后来才发现，其实 90% 的内容根本不必理会，只是为了从数学上严谨地把事情给讲清楚，但是完全不影响我们的理解，尤其对于我这种“追求直觉，不求甚解”之人。

下面进入正题，我们直接看看 GCN 的核心部分是什么亚子：

假设我们手头有一批图数据，其中有 N 个节点 (node)，每个节点都有自己的特征，我们设这些节点的特征组成一个 $N \times D$ 维的矩阵 X ，然后各个节点之间的关系也会形成一个 $N \times N$ 维的矩阵 A ，也称为邻接矩阵 (adjacency matrix)。 X 和 A 便是我们模型的输入。

GCN 也是一个神经网络层，它的层与层之间的传播方式是：

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) .$$

这个公式中：

\tilde{A} 波浪 = $A + I$ ， I 是单位矩阵

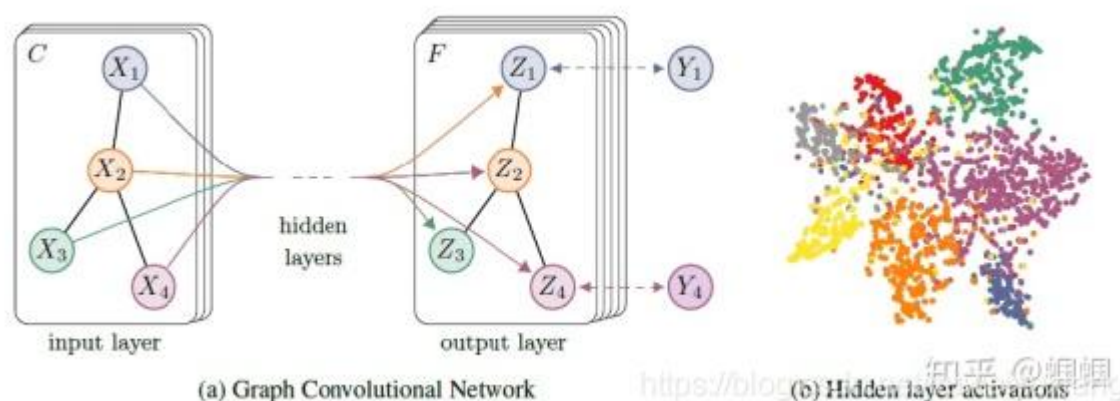
\tilde{D} 波浪是 \tilde{A} 波浪的度矩阵 (degree matrix)，公式为

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$$

H 是每一层的特征，对于输入层的话， H 就是 X

σ 是非线性激活函数我们先不用考虑为什么要这样去设计一个公式。

我们现在只用知道：这个部分，是可以事先算好的，因为 D 波浪由 A 计算而来，而 A 是我们的输入之一。所以对于不需要去了解数学原理、只想应用 GCN 来解决实际问题的人来说，你只用知道：哦，这个 GCN 设计了一个牛逼的公式，用这个公式就可以很好地提取图的特征。这就够了，毕竟不是什么事情都需要知道内部原理，这是根据需求决定的。为了直观理解，我们用论文中的一幅图：



上图中的 GCN 输入一个图，通过若干层 GCN 每个 node 的特征从 X 变成了 Z，但是，无论中间有多少层，node 之间的连接关系，即 A，都是共享的。

假设我们构造一个两层的 GCN，激活函数分别采用 ReLU 和 Softmax，则整体的正向传播的公式为：

$$Z = f(X, A) = \text{softmax}\left(\hat{A} \text{ReLU}\left(\hat{A}XW^{(0)}\right)W^{(1)}\right).$$

最后，我们针对所有带标签的节点计算 cross entropy 损失函数：

$$\mathcal{L} = - \sum_{l \in \mathcal{Y}_L} \sum_{f=1}^F Y_{lf} \ln Z_{lf}$$

就可以训练一个 node classification 的模型了。由于即使只有很少的

node 有标签也能训练，作者称他们的方法为半监督分类。当然，你也可以用这个方法去做 graph classification、link prediction，只是把损失函数给变化一下即可。

3 GCN 为什么是这个样子

我前后翻看了很多人的解读，但是读了一圈，最让我清楚明白为什么 GCN 的公式是这样子的居然是作者 Kipf 自己的博客：

<http://tkipf.github.io/graph-convolutional-networks/> 推荐大家一读。

作者给出了一个由简入繁的过程来解释：我们的每一层 GCN 的输入都是邻接矩阵 A 和 node 的特征 H ，那么我们直接做一个内积，再乘一个参数矩阵 W ，然后激活一下，就相当于一个简单的神经网络层嘛，是不是也可以呢？

$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)})$$

实验证明，即使就这么简单的神经网络层，就已经很强大了。这个简单模型应该大家都能理解吧，这就是正常的神经网络操作。但是这个简单模型有几个局限性：

- 只使用 A 的话，由于 A 的对角线上都是 0，所以在和特征矩阵 H 相乘的时候，只会计算一个 node 的所有邻居的特征的加权和，该

node 自己的特征却被忽略了。因此，我们可以做一个小小的改动，给 A 加上一个单位矩阵 I ，这样就让对角线元素变成 1 了。

- A 是没有经过归一化的矩阵，这样与特征矩阵相乘会改变特征原本的分布，产生一些不可预测的问题。所以我们对 A 做一个标准化处理。首先让 A 的每一行加起来为 1，我们可以乘以一个 D^{-1} ， D 就是度矩阵。我们可以进一步把 $D^{-1} A D^{-1}$ 拆开与 A 相乘，得到一个对称且归一化的矩阵： $\frac{1}{2} D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ 。

通过对上面两个局限的改进，我们便得到了最终的层特征传播公式：

$$f(H^{(l)}, A) = \sigma \left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

其中 $\hat{A} = A + I$ ， \hat{D} 为 \hat{A} 的 degree matrix。

公式中的 $D^{-1/2} A D^{-1/2}$ 与对称归一化拉普拉斯矩阵十分类似，而在谱图卷积的核心就是使用对称归一化拉普拉斯矩阵，这也是 GCN 的卷积叫法的来历。原论文中给出了完整的从谱卷积到 GCN 的步步推导，我是看不下去的，大家有兴趣可以自行阅读。

4 GCN 有多牛

在看了上面的公式以及训练方法之后，我并没有觉得 GCN 有多么特别，无非就是一个设计巧妙的公式嘛，也许我不用这么复杂的公式，多加一点训练数据或者把模型做深，也可能达到媲美的效果呢。

但是一直到我读到了论文的附录部分，我才顿时发现：GCN 原来这么

牛啊！

为啥呢？因为即使不训练，完全使用随机初始化的参数 W ，GCN 提取出来的特征就以及十分优秀了！这跟 CNN 不训练是完全不一样的，后者不训练是根本得不到什么有效特征的。

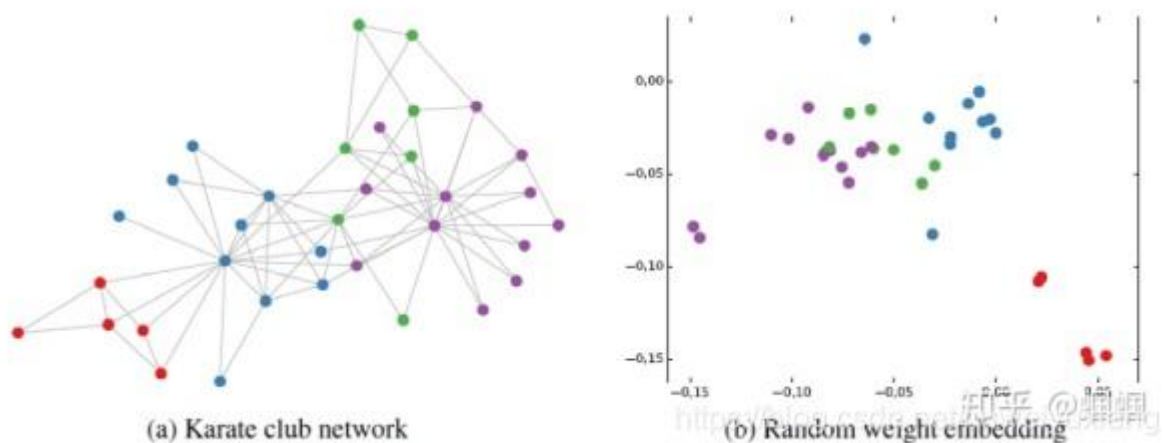
我们看论文原文：

From the analogy with the Weisfeiler-Lehman algorithm, we can understand that even an untrained GCN model with random weights can serve as a powerful feature extractor for nodes in a graph. As an example, consider the following 3-layer GCN model:

$$Z = \tanh\left(\hat{A} \tanh\left(\hat{A} \tanh\left(\hat{A} X W^{(0)}\right) W^{(1)}\right) W^{(2)}\right), \quad (13)$$

with weight matrices $W^{(l)}$ initialized at random using the initialization described in [Glorot & Bengio \(2010\)](#). \hat{A} , X and Z are defined as in Section 3.1.

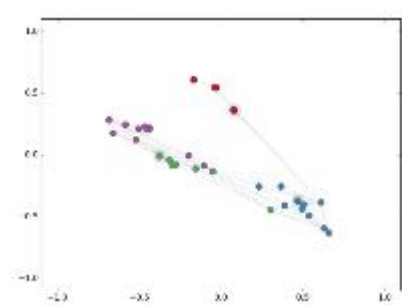
然后作者做了一个实验，使用一个俱乐部会员的关系网络，使用随机初始化的 GCN 进行特征提取，得到各个 node 的 embedding，然后可视化：



可以发现，在原数据中同类别的 node，经过 GCN 的提取出的 embedding，已经在空间上自动聚类了。

而这种聚类结果，可以和 DeepWalk、node2vec 这种经过复杂训练得到的 node embedding 的效果媲美了。

说的夸张一点，比赛还没开始，GCN 就已经在终点了。看到这里我不禁猛拍大腿打呼：“NB！”还没训练就已经效果这么好，那给少量的标注信息，GCN 的效果就会更加出色。作者接着给每一类的 node，提供仅仅一个标注样本，然后去训练，得到的可视化效果如下：



这是整片论文让我印象最深刻的地方。

看到这里，我觉得，以后有机会，确实得详细地吧 GCN 背后的数学琢磨琢磨，其中的玄妙之处究竟为何，其物理本质为何。这个时候，回忆起在知乎上看到的各路大神从各种角度解读 GCN，例如从热量传播的角度，从一个群体中每个人的工资的角度，生动形象地解释。这一刻，历来痛恨数学的我，我感受到了一丝数学之美，于是凌晨两点的我，打开了天猫，下单了一本正版《数学之美》。哦，数学啊，你真如一朵美丽的玫瑰，每次被你的美所吸引，都要深深受到刺痛，我何时才能懂得你、拥有你？

5 其他关于 GCN 的点滴

1. 对于很多网络，我们可能没有节点的特征，这个时候可以使用 GCN 吗？答案是可以的，如论文中作者对那个俱乐部网络，采用的方法就是用单位矩阵 I 替换特征矩阵 X 。

2. 我没有任何的节点类别的标注，或者什么其他的标注信息，可以使用 GCN 吗？当然，就如前面讲的，不训练的 GCN，也可以用来提取 graph embedding，而且效果还不错。
3. GCN 网络的层数多少比较好？论文的作者做过 GCN 网络深度的对比研究，在他们的实验中发现，GCN 层数不宜多，2-3 层的效果就很好了。