

关于论文《Attributed Network Embedding with Micro-Meso Structure》的学习报告

这篇论文是针对network embedding问题的研究，network embedding主要是用来把网络中的节点在低维空间中表示出来，并且要求这一过程能够保存下尽量多关于原本网络的信息。通篇来看，这篇论文的研究方向和研究方法与论文《Community Preserving Network Embedding》很相似。

当前存在的network embedding方法中，一部分方法是仅仅考虑了网络的基本架构而没有考虑网络节点的属性，一部分方法虽然在network embedding中整合了节点的属性，但是只考虑了网络的microscopic proximity structure，而忽略了网络的mesoscopic community structure。而本篇论文的最大亮点就是在考虑network embedding任务的时候同时考虑了网络节点的属性、网络的microscopic proximity structure，网络的mesoscopic community structure三方面。

一、网络的microscopic proximity structure

microscopic proximity structure体现的是基本的网络架构中的连通信息，在论文中作者只考虑了一阶连通和二阶连通，所谓一阶连通就是原本网络中节点之间的直接连通情况，当两个节点之间有边连接到一起的时候，表明这两个节点存在一种关系（如社交网络中的朋友关系），因此这两个节点就具有较大的相似度；但是，在现实的网络中，以社交网络为例，只有互为朋友关系的用户之间才存在直接连通的边，因此直接连通的节点只占了很小的比例，从而导致了相似矩阵的稀疏性问题。因此论文中在考虑microscopic proximity structure的时候也考虑了二阶连通，二阶连通是通过余弦相似度计算得到的，余弦相似度计算的是两个向量之间的正交性情况，数值越大表示二者越相似。

在考虑了网络微观架构的一阶与二阶连通以后，就得到了一个网络的相似度矩阵， $S = S^1 + \eta S^2$ ，这个相似度矩阵有一阶相似度与二阶相似度组成。

对S矩阵进行NMF分解就得到以下的目标函数：

$$\min_{M, U} \|S - MU^T\|_F^2 \quad \text{s.t.} \quad M \geq 0, U \geq 0. \quad (3)$$

这个函数的意义是让分解后的两个矩阵尽量地去拟合原始矩阵，使相似度矩阵分解成的两个矩阵尽可能多地包含原来矩阵的信息，达到分解降维的效果的情况下又不损失太多信息的目的。

二、网络的mesoscopic community structure

与论文《Community Preserving Network Embedding》（简称M-NMF）相比，本篇论文的一大亮点就是关于网络的mesoscopic community structure的衡量。在M-NMF中，这部分的衡量采用的是社区的modularity作为一个惩罚项，并且使用的是社区声明矩阵H声明每一个节点分别属于哪个社区，该矩阵中每一行只有一个entry为1，其他为0，从而导致了矩阵的极大稀疏性。在这篇论文中，作者使用的是BigCLAM生成模型，得到的是社区成员强度矩阵H，与M-NMF方法不一样的是，这里的H矩阵声明的是每一个节点属于每一个社区的

可能性。至于如何评估得到的H矩阵的质量呢？论文中采用了以下的目标函数：

$$\max_{\mathbf{H}} \sum_{(u,v) \in \mathcal{E}} \log(1 - \exp(-\mathbf{H}_u \mathbf{H}_v^T)) - \sum_{(u,v) \notin \mathcal{E}} \mathbf{H}_u \mathbf{H}_v^T \quad \text{s.t.} \quad \mathbf{H} \geq 0. \quad (4)$$

我对这个目标函数的理解是：考虑这样一个前提：假如网络中的某两个点之间存在一条边，表明这两个节点存在一定的关系，那么这两个节点属于同一个社区的可能性就远大于不存在边连接的两个节点。同时，对于两个节点，考虑两种情况，1) 两个节点属于同一个社区，不妨假设H矩阵为[0.9,0.1;0.9,0.1]，即两个节点同属于第一个社区的可能性都很大，得到的 $\mathbf{H}_u \mathbf{H}_v^T$ 就为0.82；2) 两个节点属于两个社区，假设H矩阵为[0.9,0.1;0.1,0.9]，得到的 $\mathbf{H}_u \mathbf{H}_v^T$ 仅为0.18。这是因为节点属于某个社区的可能性是用一个小数来表示的，当两个节点不属于同一个社区， $\mathbf{H}_u \mathbf{H}_v^T$ 相乘的时候必然存在大数值与小数值相乘或小数值与小数值相乘（这里的大小是相对的），如0.9*0.1，而不会存在两个大数值相乘的情况（假如是两个大数值相乘就表明这两个节点是属于同一个社区了）。因而导致对于同一个社区的两个节点的 $\mathbf{H}_u \mathbf{H}_v^T$ 值大于不同社区的两个节点的 $\mathbf{H}_u \mathbf{H}_v^T$ 值。因此，不难理解H的衡量标准就是使得同一个社区的两个点的 $\mathbf{H}_u \mathbf{H}_v^T$ 尽量大而不同社区的两个点的 $\mathbf{H}_u \mathbf{H}_v^T$ 尽量小。基于这一衡量标准与前面提出的前提，就得到了目标函数（4），目标函数中的两项分别对应同一社区与不同社区中的两个点。

基于上述目标，就得到了网络的mesoscopic community structure信息，表示成矩阵H。对H进行NMF矩阵分解就能把网络的mesoscopic community structure信息进行降维表示。因此得到以下的目标函数：

$$\min_{\mathbf{U}, \mathbf{C}} \|\mathbf{H} - \mathbf{UC}^T\|_F^2 \quad \text{s.t.} \quad \mathbf{U} \geq 0, \mathbf{C} \geq 0.$$

三、网络节点的属性

在考虑network embedding时网络节点的属性的时候，采用的方法比较简单，直接从数据集中得到节点的属性矩阵D，对属性矩阵进行NMF分解得到以下的目标函数：

$$\min_{\mathbf{N}, \mathbf{U}} \|\mathbf{D}^T - \mathbf{NU}^T\|_F^2 \quad \text{s.t.} \quad \mathbf{N} \geq 0, \mathbf{U} \geq 0. \quad (6)$$

为了结合以上提到的三个方面来考虑network embedding的任务，论文中作者直接组合了以上的三方面的目标函数，通过优化得到同一个满足上述条件的U得到网络的最终降维表示，因此得到最终的目标函数为：

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{U}, \mathbf{H}, \mathbf{C}, \mathbf{N}} L = & \|\mathbf{S} - \mathbf{MU}^T\|_F^2 + \alpha \|\mathbf{H} - \mathbf{UC}^T\|_F^2 \\ & - \beta \left(\sum_{(u,v) \in \mathcal{E}} \log(1 - \exp(-\mathbf{H}_u \mathbf{H}_v^T)) - \sum_{(u,v) \notin \mathcal{E}} \mathbf{H}_u \mathbf{H}_v^T \right) + \gamma \|\mathbf{D}^T - \mathbf{NU}^T\|_F^2 \\ \text{s.t.} \quad & \mathbf{M} \geq 0, \mathbf{U} \geq 0, \mathbf{H} \geq 0, \mathbf{C} \geq 0, \mathbf{N} \geq 0 \end{aligned} \quad (7)$$