

HW4 Report

计11 周韧平 2021010699

题目1

如果G是有向图，则平均节点度数为 $\frac{m}{n}$

如果G是无向图，则平均节点度数为 $\frac{2m}{n}$

题目 2

根据讲义中的公式可以计算得到 v 的clustering coefficient

$$cc(v) = \frac{2 \times 6}{7 \times (7-1)} \approx 0.286$$

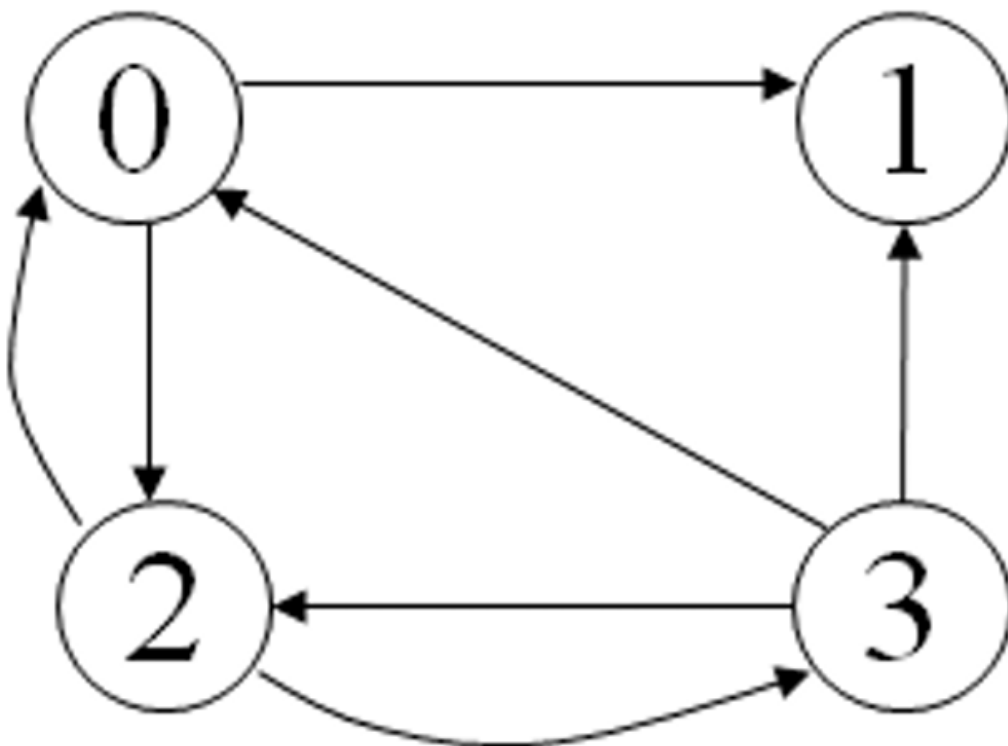
题目 3

D, 一般来说 k 的值都要小于 $|V|$, 一般来说, 隐空间的维度应该小于原始的维度

题目 4

算法介绍

PageRank 算法由 Google 创始人 Larry Page 在斯坦福读大学时提出, 该算法基于两个基本假设: 一个网页接收到指向它的入链越多, 这个网站越重要 (数量假设), 一个质量高的网页指向另一个网页, 说明那个网页也比较重要 (质量假设)。具体来说, PageRank算法作用在的模型可以抽象成一个有向图结构 (如下图), 对于这个图, 我们可以抽象出一个邻接矩阵 A_{ij} 来表示节点之间的关系, $a_{ij} = 1$ 表示从节点 j 到节点 i 有一条边。而PageRank则是通过迭代的方式, 去给这些节点一个重要性排名



实现思路

我的PageRank算法实现可以分为以下几步

1. 初始化：一开始，初始化每个节点的重要性为1，即 $x_i^{(0)} = 1, i = 0, 1, \dots, n$ ，其中 $|V| = n$ 表示图中的节点个数
2. 第 t 步到第 $t+1$ 步，第 i 个节点重要性值的更新公式为 $x_i^{(t+1)} = \beta \sum_{j=1}^n \frac{a_{ij}}{d_j^{out}} x_j^{(t)} + (1 - \beta) \frac{1}{n}$ ，其中 $\beta = 0.85$ 为阻尼系数， d_j^{out} 表示节点 j 的出度，其实际含义为让算法以 β 的概率按照图中的结构迭代重要性，否则随机游走跳跃到任意节点。这里我发现课件中提供的矩阵形式公式存在问题，因此在代码实现时，我采用的两层循环实现迭代。此外，为了应对 dead end 的情况，即某一个节点出度为 0，此时原有的公式会因为分母为 0 而失效，因此我调整为将 a_{ij}/d_j^{out} 设为 $1/n$ ，该操作等价于邻接矩阵对应的第 j 行设为相同的值，即节点 j 以相等的权重指向其它节点
3. 判停准则，当重要性向量两次迭代的差值最大分量小于 $\epsilon = 10^{-3}$ 时，停止迭代，该操作等价于 $|\Delta x|_{\infty} < \epsilon$

运行结果和分析

迭代收敛后，节点0到3的重要性为 0.26979373 0.26979373 0.26979373 0.21007725，其实际含义可以理解为0，1，2都相同重要，而节点3相对不那么重要，这一结果也可以通过PageRank前两条假设来解释：由于前三个节点都有两个节点指向他们，且其中一个节点是3，而3号只有2号节点指向它，因此从数量假设来看，3号节点相对不那么重要