**1、前言**

PageRank 算法计算每一个网页的 PageRank 值，然后根据这个值的大小对网页的重要性进行排序。它的思想是模拟一个悠闲的上网者，上网者首先随机选择一个网页打开，然后在这个网页上呆了几分钟，跳转到该网页所指向的链接，这样无所事事、漫无目的地在网页上跳来跳去，PageRank 就是估计这个悠闲的上网者**分布在各个网页上的概率**。

PageRank 背后的两个基本假设：

**数量假设**：更重要的网页可能被更多的网页链接到。

**质量假设**：有更高的 PageRank 的网页将会传递更高的权重。

**2、PageRank的定义**

**随机游走模型：**

给定一个含有n个结点的有向图，在有向图上定义随机游走模型，即**一阶马尔可夫链**，其中结点表述状态，有向边表示状态之间的转移，假设从一个节点到通过有向边相连的所有结点的转移概率相等，具体地，转移矩阵是一个**n阶矩阵M。**第i行第j列的元素mij取值规则如下:如果节点j有k个有向边连出，并且节点i是其连出的一个结点，则mij=1/k，否则mij=0。随机游走在某时刻t访问各个结点的概率分布就是马尔可夫链在时刻t的状态分布，可以用一个n维向量Rt表示

**PageRank**:

给定一个含有n个结点的任意有向图，在有向图定义一个一般的随机游走模型，即一阶马尔可夫链。一般的随机游走模型的**转移矩阵由两部分的线性组合组成**，一部分是有向图的**基本转移矩阵M**,表示从从一个结点到其连出的所有结点的转移概率相等，另一部分是**完全随机的转移矩阵**，表示从任意一个结点到任意一个结点的转移概率为1/n,线性组合系数为**阻尼因子d(1≤d≤1)** 这个一般随机游走的马尔可夫链存在平稳分布，记作R 定义平稳分布向量R 为这个有向图的一般PageRank.

1. **PageRank的计算: 迭代算法 & 代数算法**