一、问题讨论：

1.PageRank算法有什么优缺点？

优点：是一个与查询无关的静态算法，所有网页的PageRank值通过离线计算获得；有效减少在线查询时的计算量，极大降低了查询响应时间。

缺点：

1）人们的查询具有主题特征，PageRank忽略了主题相关性，导致结果的[相关性](http://baike.baidu.com/view/1326213.htm)和主题性降低

2）旧的页面等级会比新页面高。因为即使是非常好的新页面也不会有很多上游链接，除非它是某个站点的子站点。

2.PageRank算法必然收敛吗？

PageRank算法的转移矩阵A 满足马尔可夫过程收敛的条件

A为随机矩阵(A矩阵所有元素都大于等于0，并且每一列的元素和都为1)

A为不可约的(当图是强连通时，A为不可约，我们之前定义各个网页都是可相互转跳的)

A为非周期的

以上条件均满足，所以pagerank是收敛的，且与初始值无关。

3.PageRank算法有什么应用？

PageRank算法最初作为互联网网页重要度的计算方法，1996 年由Page和Brin提出，并用于谷歌搜索引擎的网页排序。事实上，PageRank 可以定义在任意有向图上，后来被应用到社会影响力分析、文本摘要等多个问题。

4.PageRank算法对于孤立节点是怎么处理的？

互联网这样的图里存在很多孤立节点，即不被其他节点引用的网页。PageRank增加了一个策略，就是为所有的节点设置一个最小得分，使得搜索用户有一定几率检索到这些网页。具体做法是为PR值的计算公式增加一个阻尼系数。

二、读书小结

PageRanks 是一种链接分析算法，它为超链接文档集（例如万维网）的每个元素分配数字权重，目的是“衡量”其在集合中的相对重要性。该算法可以应用于具有相互引用和引用的任何实体集合。它分配给任何给定元素E的数字权重称为E的PageRank。

PageRank 是基于数学算法的结果，由所有万维网页面创建为节点，超链接为边。排名值表示特定页面的重要性。指向页面的超链接算作支持投票。页面的 PageRank 是递归定义的，取决于链接到它的所有页面的数量和 PageRank 指标。由许多具有高 PageRank 的页面链接到的页面本身会获得较高的排名。

实际上，PageRank 概念可能容易受到操纵。已进行研究以识别受到错误影响的 PageRank 排名。目标是找到一种有效的方法来忽略具有错误影响的 PageRank 的文档中的链接。

PageRank 算法输出一个[概率分布，](https://en.wikipedia.org/wiki/Probability_distribution)用于表示一个人随机点击链接到达任何特定页面的可能性。可以为任意大小的文档集合计算 PageRank。在几篇研究论文中，假设在计算过程开始时，分布在集合中的所有文档中平均分配。PageRank 计算需要多次传递，称为“迭代”，通过集合来调整近似的 PageRank 值以更接近地反映理论真实值。

概率表示为介于 0 和 1 之间的数值。0.5 的概率通常表示为某事发生的“50% 几率”。因此，PageRank 为 0.5 的文档意味着点击随机链接的人有 50% 的机会被定向到该文档。