**读书报告**

61518424 王贵涛

1. **问题与解答**

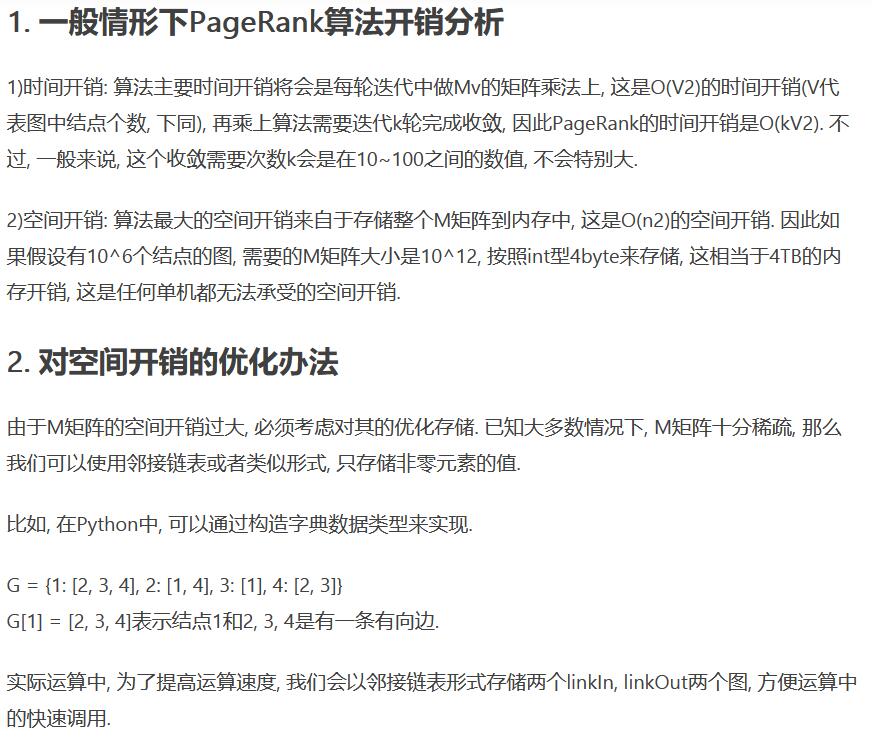
我提出问题：

1. 从一个结点到其余每一个结点的概率为什么都相等？

讨论结果：这个应该可以理解为PageRank提出的时候所依托的假设，而实际情况可能有所区别，所以假设的话就比较自由了。

1. 现实中相互链接的网页众多，这样转移矩阵阶数会很高，计算起来会不会过于复杂？

讨论结果：如果直接使用书本原始算法，确实开销较大，时间和空间几乎都是难以承受的，因此可以采用一些优化算法。





别人提出的问题：

1. 阻尼因子d在应用中应该如何选择，是否可以理解为阻尼因子越大，PageRank越不稳定？

我的解答：阻尼因子d取值由经验决定。当d接近1时，随机游走主要依照转移矩阵M进行，当d接近0时， 随机游走主要以等概率随机访问各个结点。

1. 阻尼因子的作用是什么？其过大或者过小造成的影响是什么？

我的解答：在任意一个网页上，浏览者或者以概率d决定按照超链接随机跳转，这时以等概率从连接出去的超链接跳转到下一个网页或者以概率（1-d）决定完全随机跳转，这时以等概率1/n跳转到任意一个网页。第二个机制保证从没有连接出去的超链接的网页也可以跳转出。这样可以保证平稳分布，即一般PageRank的存在，因而一般PageRank适用于任何结构的网络。

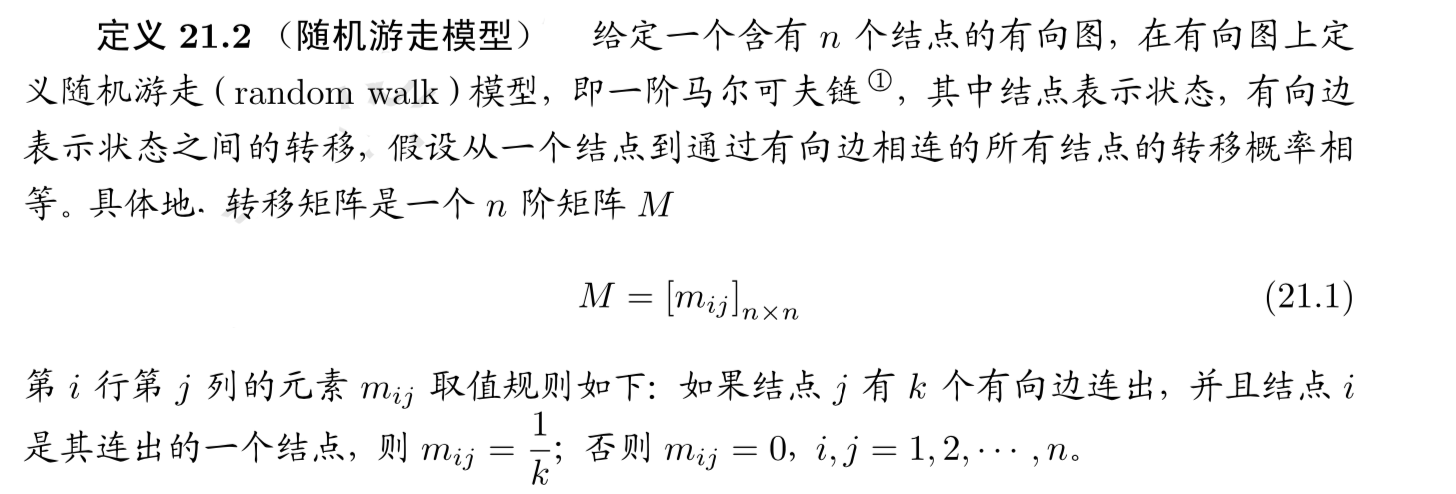
1. **下周计划安排**

看《Speech and Language Processing》第四章。

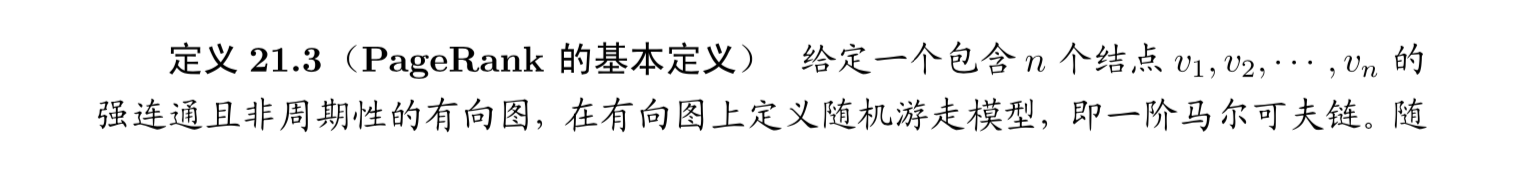
1. **读书收获**

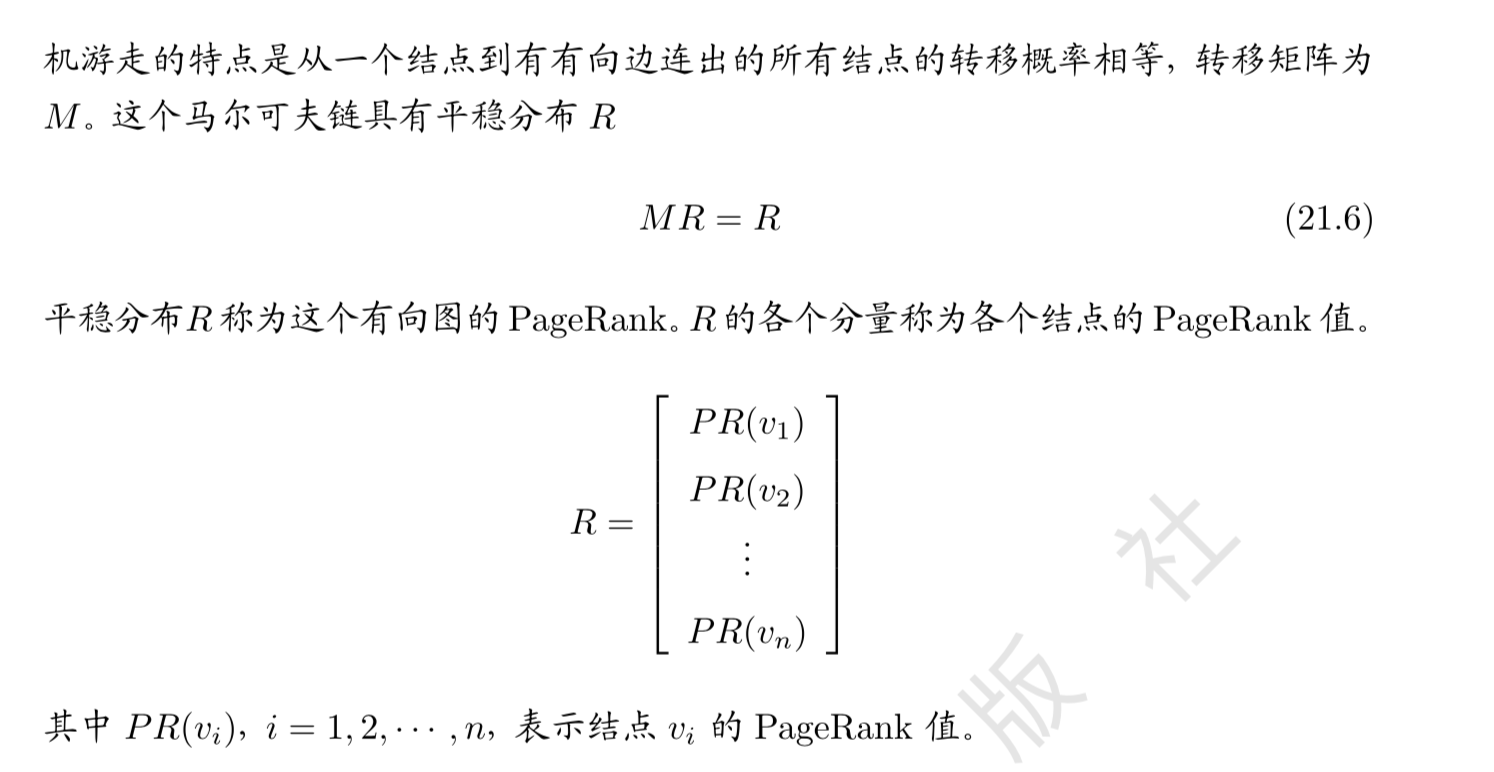
PageRank是定义在网页集合上的一个函数，它对每个网页给出一个正实数，表示网页的重要程度，整体构成一个向量。PageRank值越高，网页就越重要，在互联网搜索的排序中可能就被排在前面。

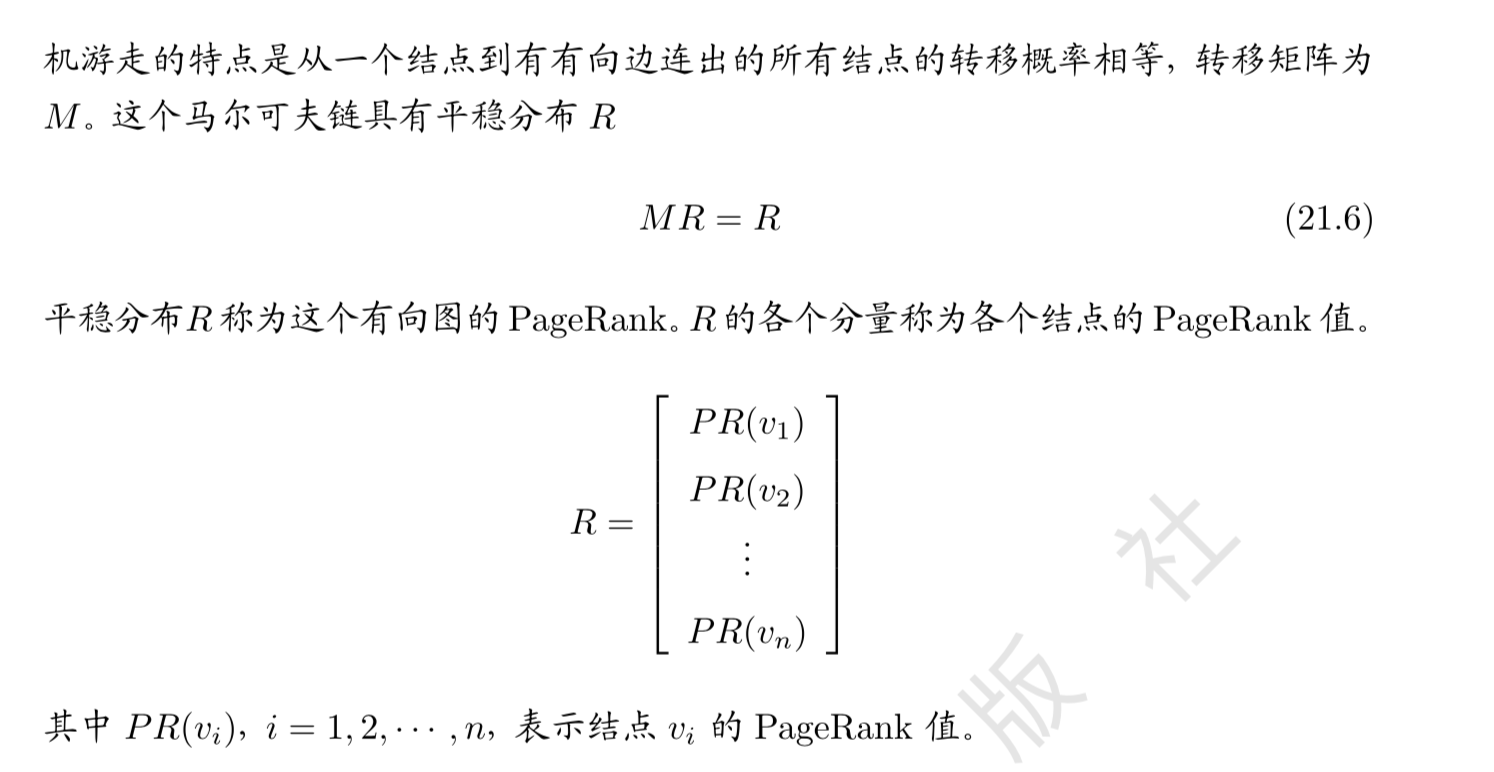
假设互联网是一个有向图，在其基础上定义随机游走模型，即一阶马尔可夫链，表示网页浏览者在互联网上随机浏览网页的过程。假设浏览者在每个网页依照连接出去的超链接以等概率跳转到下一个网页，并在网上持续不断进行这样的随机跳转，这个过程形成一阶马尔可夫链。PageRank表示这个马尔可夫链的平稳分布。每个网页的PageRank值就是平稳概率。



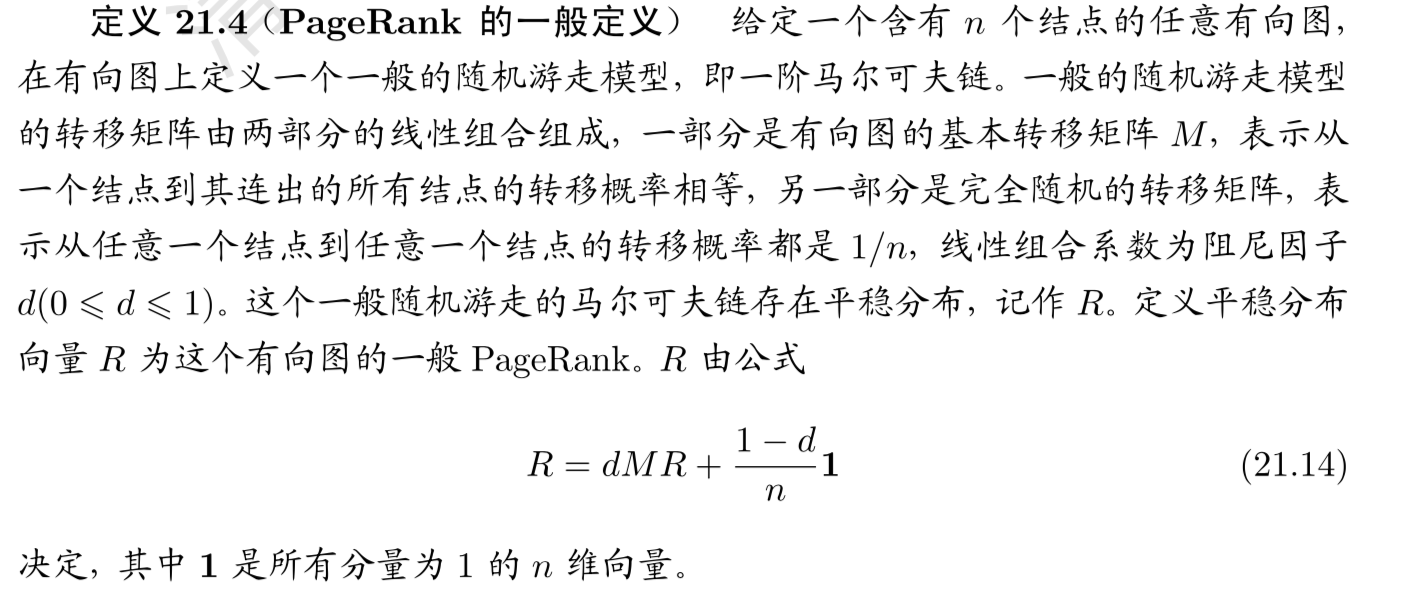
**PageRank的基本定义**





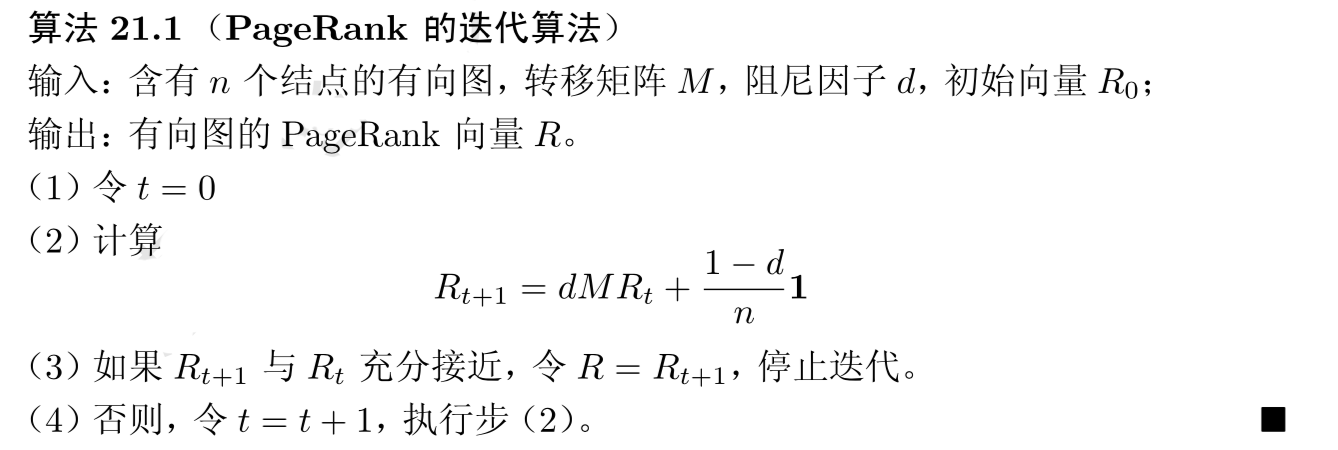


**PageRank的一般定义**

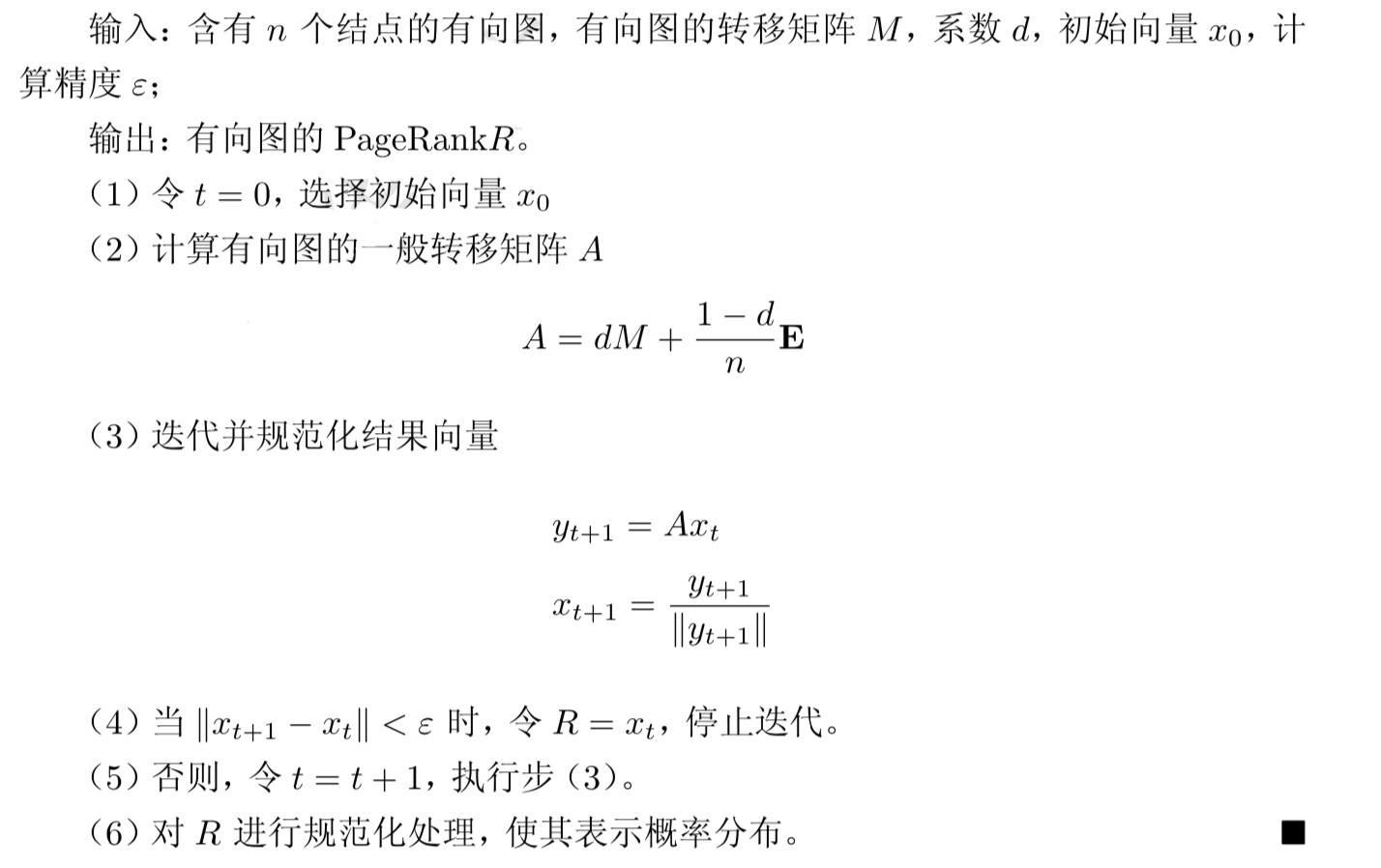


一般PageRank的定义意味着互联网浏览者，按照以下方法在网上随机游走：在任意一个网页上，浏览者或者以概率d决定按照超链接随机跳转，这时以等概率从连接出去的超链接跳转到下一个网页。或者以概率（1-d)决定完全随机跳转，这时以等概率1/n跳转到任意一个网页。第二个机制保证从没有连接出去的超链接的网页也可以跳转出。这样可以保证平稳分布，即一般PageRank的存在，因而一般PageRank适用于任何结构的网络。

**PageRank的计算**



计算一般PageRank的幂法：



代数算法：

