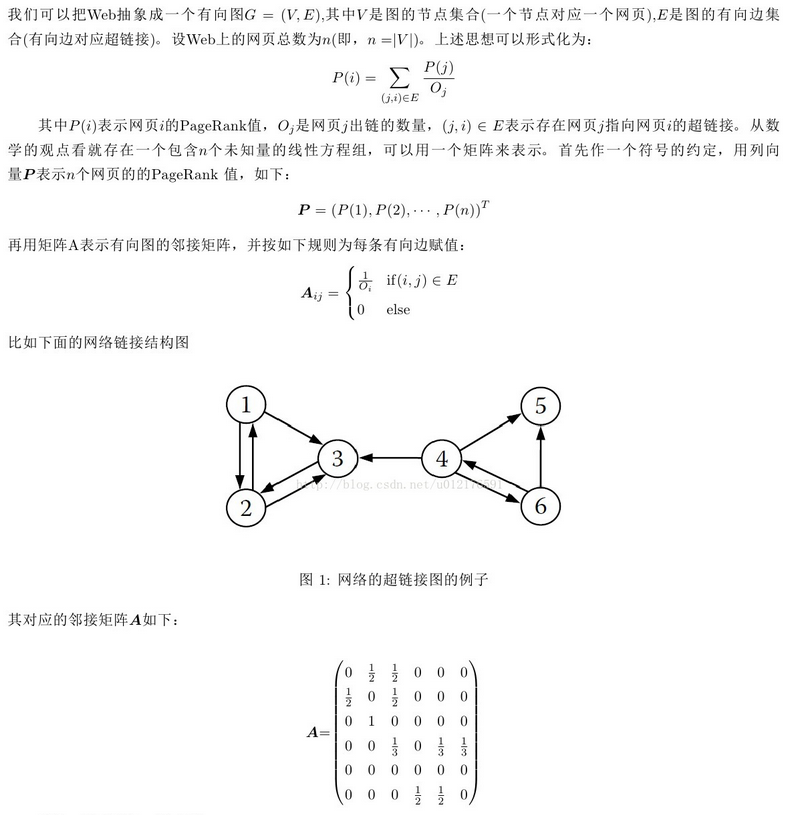
**PageRank Algorithm**

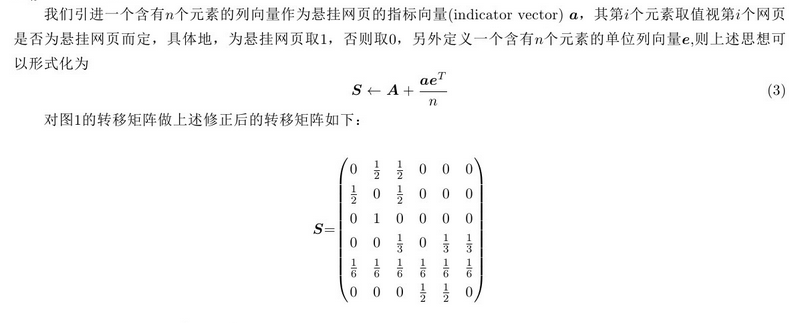
谷歌的两位创始人，当时还是美国斯坦福大学 (Stanford University) 研究生的佩奇 (Larry Page) 和布林 (Sergey Brin) 开始了对网页排序问题的研究。后来发表一篇文章《[The PageRank Citation Ranking: Bringing Order to the Web](http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf)》（ODU,Norfolk,1998）



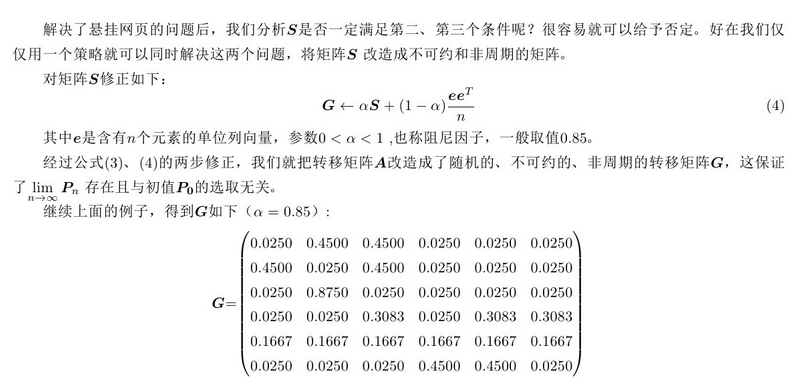
主要思想：原来网页搜索得出的结果是根据分类返回，随着海量数据，手工分类太累，因此必须通过某种方法自动分类。按照合理解释，每个网页的排序不应该靠自己控制，例如：一个垃圾网页不论把关键词重复多少次，它仍然是垃圾网页，用户在搜索该关键词该网页的排序仍然是排名靠后，相反好的网页同理。那么排名靠前靠后不通过人为分类，必须使用其他指标表示，比如链接数。即：一个网页被其他网页链接的越多，它的排名就应该越靠前，而且，一个网页越是被排序靠前的网页所链接的话，他的排名也应该越靠前。比如，一篇论文被大牛所引用，显然这个文章也差不到哪去。



A中每一行表示对其他节点的出度分子一，列表示对该行的入度。



转移矩阵S其实就是在原来A基础之上加上只有入度没有出度的节点分子一，其他无变化。a表示第几个节点为悬挂网页，其值为1，其他全为0；e单位矩阵貌似为全为1的列向量才对。



e\*e^T表示n\*n全为1的矩阵。

