1. D矩阵的意义：

为每个没有链出的页面，再多增加一个可以等概率跳出到任何一个页面的可能。

1. HITS和PageRank的假设异同点：

同：

二者都关注“某一网页的链接情况”，同时关注链接的“数量”与“质量”，并以此评估网页重要度。

异：

1. PageRank只关注链入的链接，而HITS同时关注链入和链出
2. PageRank将所有网站一视同仁，而HITS将网站视作“hub和Authority”的结合
3. HITS在query得到以后进行在线计算，而PageRank离线计算，与query无关。
4. 公式各项的含义：
   1. 0.01：防止co\_degree为0时f为0.
   2. -1：降低c与w偶然同现的情况对结果造成的影响
   3. En的计算：在w和c出现彼此独立的情况下，c和w同时出现的文章数。（区别于n\_{cw},指的是实际同时出现的情况；二者差值越大，说明c和w关系越密切，越不独立）
   4. idf的含义：考察query中的每个w的重要性，出现越少，越重要idf越高。