

# 大数据分析第一次作业

(最晚提交时间 2025.04.18,18:00 之前提交到微助教上)

1. 假定不采用抽税法的情况下, 计算图 5.7 中每个网页的 PageRank 值(10分)。

(备注: 采用 Matrix Formulation 的形式, 需要计算出初始  $r$  值和第一轮迭代后的  $r$  值,

不要求解出最终的稳态  $r$  值。后续其他关于 PageRank 类题型都是一样的要求!)

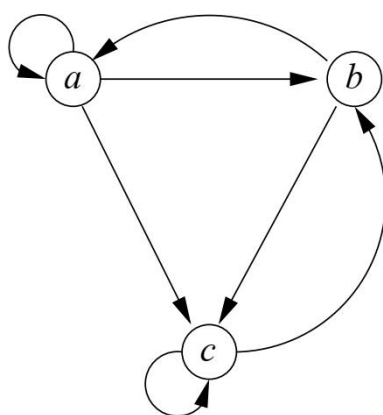


Figure 5.7: An example graph for exercises

2. 假定 $\beta = 0.8$ , 计算图 5.7 中每个网页的 PageRank 值 (20 分)。

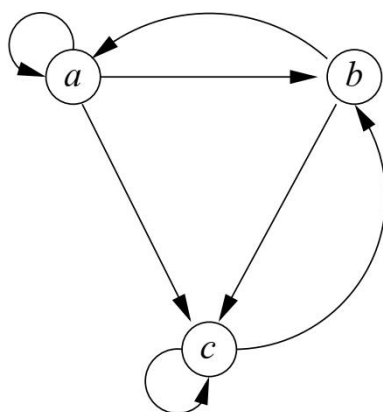


Figure 5.7: An example graph for exercises

3. 假设随机跳转集合为 A 和 C 组成的集合,  $\beta = 0.8$ , 计算图 5.15 所对应的面向主题的 PageRank 值 (20 分)

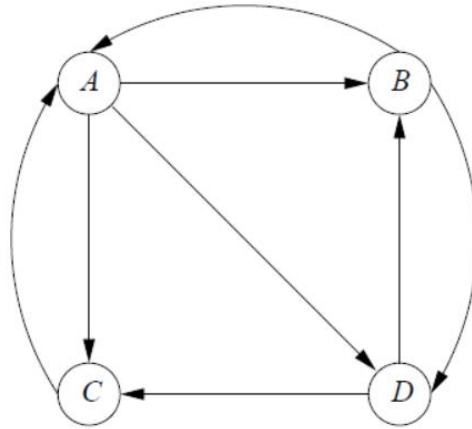


Figure 5.15: Repeat of example Web graph

4. 在垃圾农场的分析章节中 (1-linkanalysis PPT 95-99 页, Section 1.6.2), 我们对图 5.16 进行了分析, 其中每个支持网页都链向目标网页。重复垃圾农场的分析过程: (30 分)

- (a) 每个支持网页只链向自己, 而不是链向目标网页
- (b) 每个支持网页不链向任何网页
- (c) 每个支持网页同时链向自己和目标网页

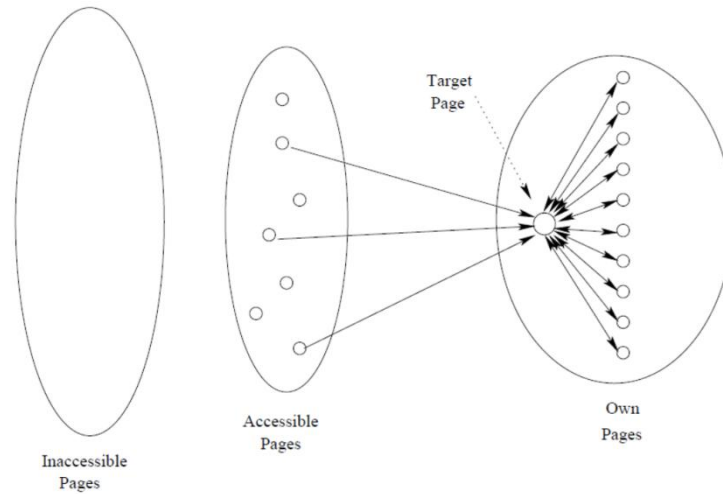


Figure 5.16: The Web from the point of view of the link spammer

备注：Supporting page(支持页), 表示 own pages 里面除了 target page (目标页)t 以外的其他网页.

5. 计算图 5.1 给出的图中所有节点的导航度和权威度 (20 分)。

(类似地, 需要计算出初始  $h, a$  值和第一轮迭代后的  $h, a$  值, 不要求解出最终的稳态  $h, a$  值)

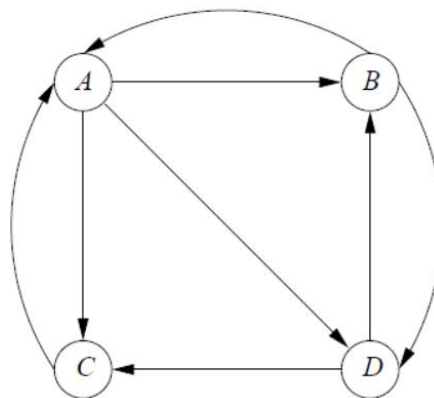


Figure 5.1: A hypothetical example of the Web