

# Back Pagerank

Chunwei Yan

2012 年 10 月 26 日

## 1 原始模型

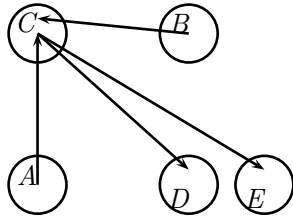


Figure1

如部分图 Figure1, 只考虑 C 节点的情况  
之前课上想到的模型由三部分组成:

1. 原始的按照链接指向的漫步到 C 的概率  
 $\frac{P_A}{L_A} + \frac{P_B}{L_B}$
2. 某种情况下从 C 退回的概率  $\frac{P_A}{L_A} + \frac{P_B}{L_B}$
3. 随机漫步到任何一个节点的概率  $\frac{1}{N}$

进行组合, 得到最初的模型, 也就是课上演示的模型:

$$P_C = k[\frac{P_A}{L_A} + \frac{P_B}{L_B}] + (1-k)\{\alpha[\frac{P_A}{L_A} + \frac{P_B}{L_B}] + (1-\alpha)\frac{1}{N}\}$$

其中,  $\alpha$  是 back 操作的概率

但是, 这个模型很明显是错误的, 主要有如下问题:

1. 从 C 退回的概率并不能算入 C 的权重中, 而应该是从 C 指向的其他节点退回的概率参与 C 权重的计算

2. 如果把 Walker 沿着链接指向访问各节点作为正常的访问方式, 那么退回操作 (back) 是否应该看做和 Walker 随机选取一个节点访问同类型的操作

## 2 新模型

1. 将 Back 操作作为一种权重的回赠, 如在 Figure1 中, C 指向 D, 则在 D 中会有一定的几率通过 Back 操作回赠一部分权重给 C
2. Back 作为在图中访问的一种普通的操作, 即 Walker 在沿着链接漫游和随机选取一个节点漫游时, 均可能有 Back 操作, 假设
3. 在沿着链接漫游时, back 操作的概率为  $\alpha$

如此, 同样将模型分为三个部分:

1. 原始的按照链接指向的漫步到 C 的概率  
 $\frac{P_A}{L_A} + \frac{P_B}{L_B}$
2. 在其他节点, 通过 back 操作返回到 C 的概率,  
 $\frac{P_C}{L_C} + \frac{P_C}{L_C}$
3. Random Walker 随机选择一个节点访问, 选到 C 的概率  $\frac{1}{N}$

最终的模型是

$$P_C = k[\frac{P_A}{L_A} + \frac{P_B}{L_B}] + (1-k)\{\alpha[\frac{P_C}{L_C} + \frac{P_C}{L_C}] + (1-\alpha)\frac{1}{N}\}$$