Graph as Matrix: PageRank, Random Walks & Embeddings

PageRank (Google Algo)

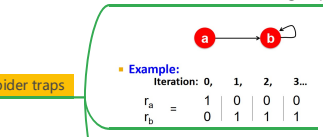
how to solve

- Power iteration
- Repeat until convergence
- $r_j^{(t+1)} = \sum_{i \to j} \frac{r_i^{(t)}}{d_i}$ or equivalently $r = Mr$
- qs
  - Does this converge?
  - Does it converge to what we want?
  - Are results reasonable?

2 problems

Spider traps

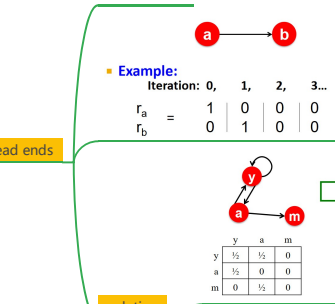
- all out-links are within the group



- solution
  - teleport — out of spider trap within a few time steps
  - With prob β, follow a link at random
  - With prob. 1-β, jump to a random page

dead ends

- no out-links



- solution
  - teleport to somewhere with equal prob

PageRank equation

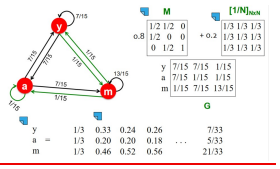
$$r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1-\beta)\frac{1}{N}$$

- formula
  - teleport - 1/N prob i.e. same for all nodes
  - assumes that M has no **dead ends**
  - preprocess M or random teleport
- Matrix form

$$G = \beta M + (1-\beta)\left[\frac{1}{N}\right]_{N \times N}$$

- a recursive problem
  - r = G * r
  - Power method
  - β=0.8,0.9 in practice

toy example



- Random Teleports (β = 0.8)
- (transition matrix * teleport matrix)* r