INFORMATION RETRIEVAL

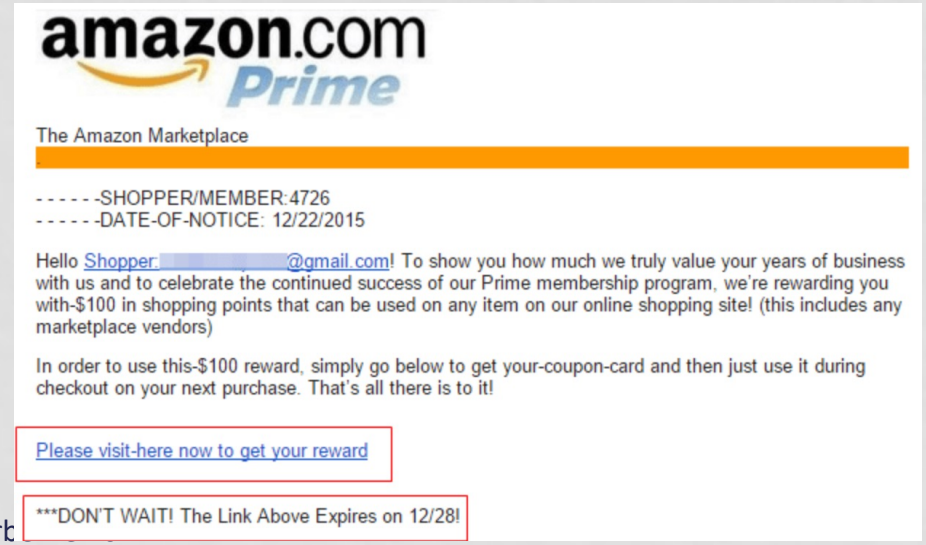
HOMEWORK EXERCISES L10. LINK ANALYSIS

SUZAN VERBERNE 2022

Universiteit Leiden

EXERCISE 1 - SOLUTION

- Anchor texts can also be misleading on purpose. Give an example of misleading anchor text, and describe why it is misleading
  - Examples of phishing: “visit here now to get your reward!” / “click here to reset your password”
  - Examples of click bait:
    “the reason why will shock you”


amazon.com Prime

The Amazon Marketplace

------SHOPPER/MEMBER:4726
------DATE-OF-NOTICE: 12/22/2015

Hello Shopper: @gmail.com! To show you how much we truly value your years of business with us and to celebrate the continued success of our Prime membership program, we're rewarding you with-$100 in shopping points that can be used on any item on our online shopping site! (this includes any marketplace vendors)

In order to use this-$100 reward, simply go below to get your-coupon-card and then just use it during checkout on your next purchase. That's all there is to it!

Please visit-here now to get your reward

***DON'T WAIT! The Link Above Expires on 12/28!

EXERCISE 2 - SOLUTION

- (Exercise 21.3): Given the collection of anchor texts for a web page x, suggest a heuristic for choosing one anchor text from this collection that is most descriptive of x. If possible, give the equation.

  - 'anchor frequency – inverse page frequency' (af-ipf)
    - af: the number of times this anchor text is used for page x
    - ipf: the inverse of the number of pages the anchor text is used for
    - Same log-weighting as for tf-idf:

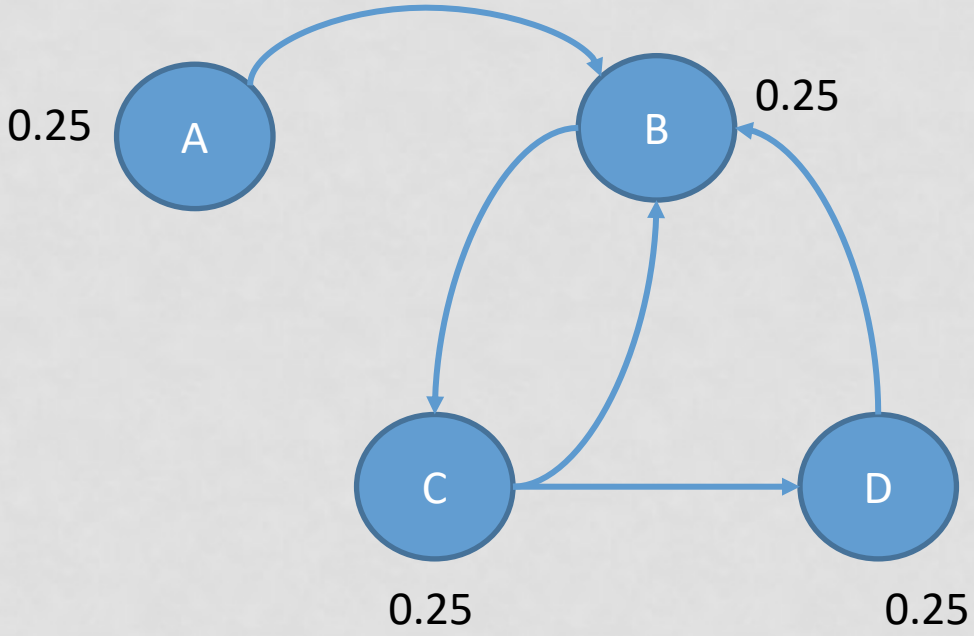
$$1 + \log(c_{a,x}) * \log\left(\frac{N}{n_a}\right)$$

    - $c_{a,x}$: count of a used as anchor text for x
    - n_a: number of pages a refers to
    - N: number of pages in the collection

EXERCISE 3 - SOLUTION

- Iteratively compute the PageRank score for nodes A, B, C, D with $\alpha = 0$. Initialize with equal probabilities for all nodes in iteration 0. Stop after iteration 3.

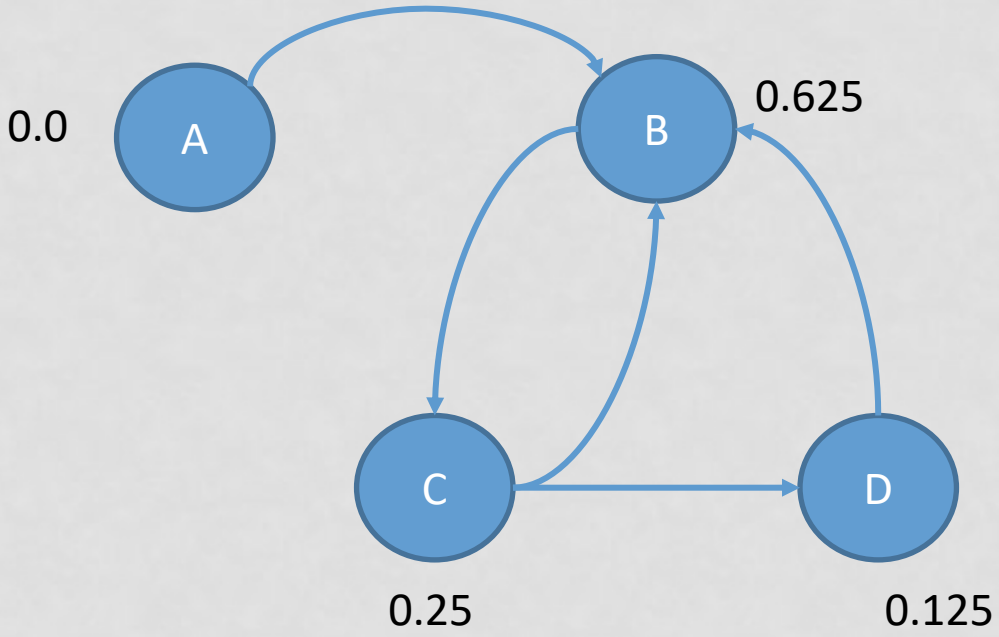
Iteration 0



EXERCISE 3 - SOLUTION

- Iteratively compute the PageRank score for nodes A, B, C, D with $\alpha = 0$. Initialize with equal probabilities for all nodes in iteration 0. Stop after iteration 3.

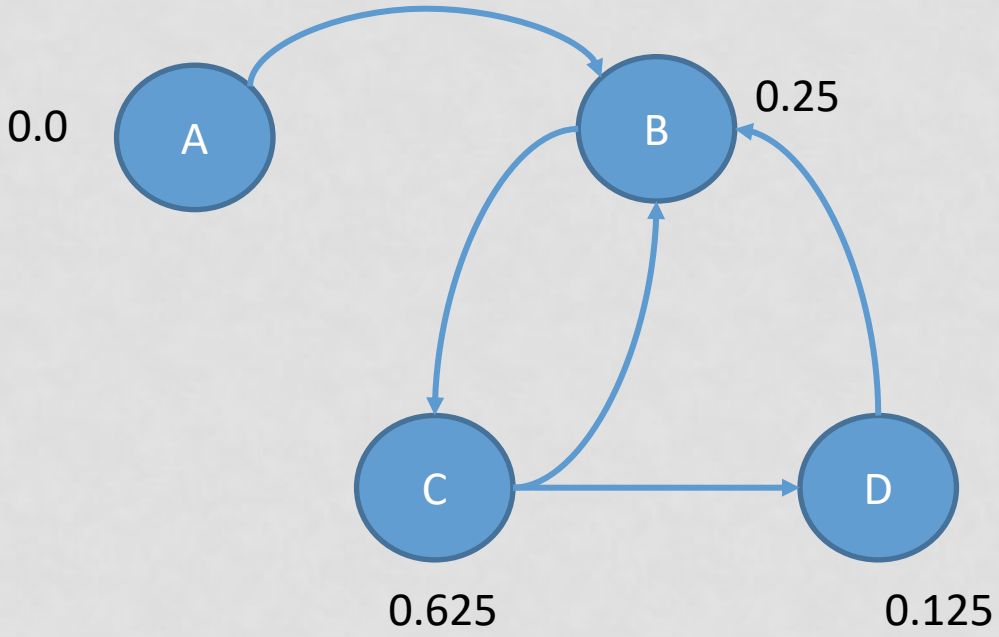
Iteration 1



EXERCISE 3 - SOLUTION

- Iteratively compute the PageRank score for nodes A, B, C, D with $\alpha = 0$. Initialize with equal probabilities for all nodes in iteration 0. Stop after iteration 3.

Iteration 2



EXERCISE 3 - SOLUTION

- Iteratively compute the PageRank score for nodes A, B, C, D with $\alpha = 0$. Initialize with equal probabilities for all nodes in iteration 0. Stop after iteration 3.

Iteration 3