Exercise 1: Google

◻ Consider a set of 4 web pages a, b, c and d. a hyperlinks to b, b hyperlinks to a, c and d, and c hyperlinks to d, and d hyperlinks to a. **Estimate** the (simplified) PageRank score of these pages and **rank** the pages according to the score.

Use the definition of the score, i.e., the score of a page z, s(z), is equal to the sum of the scores of the pages that hyperlinks to it (e.g., x, y, w) divided by the number of their outbound links (s(z) = s(x)/|O(x)| + s(y)/|O(y)| + s(w)/|O(w)|). Moreover the sum of the scores of all the pages must be 1.

Exercise 2: Google

- A user of a browser can, in addition to clicking a hyperlink on the page *x* he is currently browsing, use the *back button* to go back to the page from which he arrived at *x*.

- Can such a use of back button be modeled as a Markov chain? How would we model repeated invocations of the back button?

Exercise 3: google

- Download from the course web site the file PageRank.zip
- Unzip the file and brose the content of the class directory
- Transform the input graphs tiny.txt and medium.txt into the corresponding transition matrices using the java program Transition
    - `> java Transition < tiny.txt`
- Use the generated transition matrices to perform:
    - the random surf simulation
        - `> java Transition < tiny.txt | java RandomSurfer 1000`
    - the random surfer visualization histogram
        - `> java Transition < tiny.txt | java RandomSurferHistogram 1000`
    - the power method
        - `> java Transition < tiny.txt | java Markov 20`

Exercise 3 cont.

- Observe the PageRank values computed by the random surf simulation and the power method for increasing values of the iteration parameter

  - random surf simulation: 10, 100, 1000, 10000, …

  - power method: 0, 1, 2, 3, 4, …

- Check when you have convergence of the PageRank values with both approaches on the two available graphs

Exercise 4: part 1

- Make 3 small groups
- Look slide 31 (Exploratory Search) and read the paper by Marchionini (pages 42-43) – it is on the course web site
- Each group selects a domain among: music, video, books
- For each type of search (lookup, learn, investigate) select one/two tasks among those listed in fig. 1
- Consider one of the following tools and check if it supports your selected tasks (maybe you can divide the group and analyze more than one system)
  - Books: shelfari, goodreads, amazon, librarything.com yournextread.com
  - Movies/Video: youtube, ted.com, rottentomatoes, movielens, imdb
  - Music: last.fm, musicovery, allmusic.com, spotify
- Think about new tools/functions that can better support your selected tasks.

Exercise 5: part 1

◻ Think about the **similarities** and **differences** between collaborative filtering and google page-rank in scoring and ranking items for users

◻ Evaluate them with respect to the following dimensions:

  ▪ How do they exploit User-Generated-Content, i.e., any kind of content that is produced by the users (e.g., ratings, reviews, pictures, links, etc.)

  ▪ Are they personalized?

  ▪ Are they context-dependent, i.e., do they behave differently if the context of the user changes (e.g., what the user knows about a topic, or the user activity on the tool in the 30 mins before the current time, or if its is night or morning)

  ▪ When the heaviest part of the computation is performed (before or during the query reply)?