laSalle ENG
Universitat Ramon Llull

Practical assignment 1
PageRank

Subject: Data Mining
Deadline: 22 October 2021
Year: 2021/2022

PageRank algorithm: analysis and implementation

PART 0: MOTIVATION AND OBJECTIVES

Motivation

PageRank is an algorithm used by Google Search to rank websites in their search engine results.

Objectives

a) Get skills with the Python environment.
b) Implement an algorithm from formal equations.
c) Know the PageRank algorithm and discuss about its parameters

Instructions
Read all instructions in this section thoroughly.

- Try to understand the python script skeleton. To do so, run the script according the code structure.
- Find the comment “Introduce your code here”, this area corresponds to the PageRank’s implementation and has intentionally left blank and expect you to implement.
- Describe in detail how to modify the parameter Beta to obtain the ranks. Discuss about this parameter.

Formatting: Your solution must be implemented in python, following the precise instructions included in Part 2. The programming exercise ask you to add code and analyse the results. Your analysis should be written in a text file "Readme.txt" and upload along with your Python file (a zip file is strongly recommended: SURNAME_NAME.zip). The submission will be evaluated according the execution time and result analysis.

PART I: PROBLEM SET
The file “gr0.California.txt ” gives you a graph with all the webs related to the query "California". There are two differentiated parts:

- Node number (n) with its ID and URL

```
n 31 http://www.assembly.ca.gov/
n 32 http://www.ucop.edu/ucophome/ucservers.html
n 33 http://www.ftb.ca.gov/
n 34 http://www.floydsordeal.com/ring.htm
n 35 http://www.dfg.ca.gov/
n 36 http://www.cpuc.ca.gov/
n 37 http://www.city.net/countries/united_states/california/
```

- An edge number (e) indicating IDFROM to IDTWO

```
e 1841 7902
e 1841 41
e 1841 93
e 1841 196
e 1841 160
e 1841 376
```

PART II: ENVIRONMENT
Before starting the exercise, you will need to make certain that you are working on a computer with particular software:

- Python 3.8.x (https://www.python.org/downloads)
- numpy (https://www.numpy.org/)
- scipy (https://www.scipy.org/)
- Python editor (PyCharm, Atom, Spyder, Notepad++). You can also use notebook such as Jupyter notebook.
- Necessary resources for the assignment can be download from eStudy.