Università della Svizzera italiana

Facoltà di scienze informatiche

Data Analytics

Academic Year 2022-23

Course Assignment N. 26: Autonomous Systems

Prof. Fabio Crestani

For this assignment you will work individually to carry out simple tasks of data analysis given a specific dataset. The goal of this assignment is to use Python and complementary libraries on a given dataset in order to *explore* and *analyze* the given data and *draw conclusions*.

whats the maximum number of computers in one server?
whats the maximum number of servers that a computer can connect with?
is there any computers that have no servers?
the most computers with the connection has no connection with which network?

Description

Computer networks consist of numerous devices that are connected to one another. Link analysis is a powerful tool that is used in numerous applications and for a number of different tasks. Link analysis on networks and visualization help understanding the relationships (connections) between the computers and the servers. You are given a dataset with computers and a number of connected servers and you are asked to:

- Explore and describe the data (preprocess the data, visualize the variables with different graphs, distribution of the variables).
    - While exploring the data, formulate questions and provide an answer to them, such as which server has the most connected computers? Which two servers are less connected? etc.
- Plot the graph that shows the links between the computers, *i.e.*, how the servers are connected.
    - Use graphics to enlarge the servers that have most centrality, etc.

pagerank

Submission procedure and evaluation

You should produce a report of your work and its evaluation along with the source code. It will be a concise explanation of how you tackled the different tasks, the reasons of your choices, successive conclusions, graphs you produced, results of the decisions and their accuracy *etc*.

Use Jupyter Notebook to produce results of the commands in a single .ipynb file. For more information check: https://jupyter.org/documentation

The report (max 5 pages) and the code of the project need to be submitted via iCorsi.

Please, upload all the required items in a single file and name it following the structure: **noProject_FirstnameLastname.[zip|tar.gz|7z]**. For instance, 05_ NameSurname.tar.gz The dataset regarding this project can be downloaded from the link provided on iCorsi.

- centrality
- directed and undirected with page ranking
- degree and indegree
- check how to do the page rank for the graph adjacency matrix for page rank using undirected to directed graph
-correlation between the four centralities e.g degree centrality etc

which computer has the most connection?
which computer has the least connection?
which connected server has more connection?
is there no server that is not connected?
most important?
what is the most important node(page rank result)?
what if we remove the most connected server?