Network analysis and data visualization

Academic Year: (2019 / 2020) **Review date: 22-04-2020**

Department assigned to the subject: Mathematics Department

Coordinating teacher: ANTONIONI , ALBERTO

Type: Electives ECTS Credits : 3.0

Year : 1 Semester : 2

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

It is recommended to have completed the Mathematics, Statistics subjects and a good level in programming in R or Python

OBJECTIVES

Basic Skills

- Acquisition of knowledge and skills that provide with a background of creativity in the development and application of ideas, often within a research context.
- Ability to apply acquired knowledge and to solve problems under novel or almost novel situations or within broader (multidisciplinary) contexts related with big data.
- Acquisition of skills for learning in an autonomous and continuous manner.

General Skills

- Ability to apply the theoretical foundation of collect, storage, processing and presentation of information, especially for big data volumes.
- Ability to identify the most suitable data analysis technique in each problem, and to apply it for obtaining the most appropriate solution to each one.
- Ability to obtain practical and efficient solution for processing of big data volumes.
- Skill to synthesize data analysis conclusions, and to communicate it clearly and convincingly in a bilingual environment.
- Ability to generate new ideas and to anticipate new situations, within the context of data analysis and decision making.
- Skill to working collaboratively and to collaborate with others autonomously.

Specific Skills

- Skill to design data processing systems, from the data gathering to statistical analysis and presentation of final results.
- Ability to apply the basic principles of network science and apply them to the study of different data to model and forecast their behavior using features extracted from network science.
- Ability to design effective visualizations of large data sets that can lead to the discover, interpretation and access to those datasets..
- Ability to identify the opportunity to apply network science and visualization techniques for solving real problems.

Learning outcomes

- Basic knowledge about network science techniques.
- Understanding of basic network science techniques.
- Making practical use of network science techniques in real problems
- Basic knowledge of data visualization techniques
- Ability to use visualization techniques to explain and solve real problems

DESCRIPTION OF CONTENTS: PROGRAMME

1. Graphs and general concepts
   1.1 Mathematical definition and examples
   1.2 Graph Theory, historical introduction
   1.3 Weighted and directed graphs
   1.4 Bipartite graphs
   1.5 Adjacency matrix
   1.6 Degree, average degree, degree distributions
   1.7 Topological concepts in graphs (distance, shortest path, diameter)

1.8 Practical example
1.9 Centrality metrics
1.10 Cliques, motifs, clustering and communities
1.11 Types of networks: random networks, small world, scale-free
1.12 Hubs and preferential attachment

2. Social Networks
2.1 Definition and context
2.2 Local and global properties of social networks
2.3 Difference between social networks and other networks
2.4 Social mechanisms
2.5 Applications of social networks: fraud detection, recommendation systems, product adoption, churn, etc.

3. Graph analysis / Social Network Analysis
3.1 Overview of software/libraries for SNA
3.2 Introduction to the igraph library
3.3 Introduction to the networkX library in Python
3.4 Practical example
3.5 Create a graph
3.6 Analyze a graph
3.7 Simulate a graph
3.8 Test a graph

4. Practical examples of graph analysis
4.1 Link prediction: application to friend recommendation
4.2 Epidemic models in networks
4.3 Build, analyze and visualize information networks: the case of Twitter and its API
4.4 Analysis and visualization of dynamic networks

5. Introduction to data visualization
5.1 Data types and sources
5.2 Main tools to visualize data. Introduction to Tableau, ggplot and D3
5.3 Data reduction techniques
5.4 Static visualization of data
5.5 Visualization of one-dimensional data
5.6 Visualization of multi-dimensional data
5.7 Geo-spatial data
5.8 Content (text) visualization
5.9 Time-series and predictive model visualization
5.10 Graph visualization
5.11 Dynamic data visualization
5.12 Visualizaton of transport data (world-wide flights)
5.13 Visualization of large social networks from Twitter
5.14 Visualization of movie ratings

LEARNING ACTIVITIES AND METHODOLOGY

The course is imparted in specific rooms and laboratories for the Master Program. It will include:
- Lectures for the presentation, development and analysis of the contents of the course.
- Practical sessions for the resolution of individual problems and practical projects in the laboratory
- Seminars for discussion with reduced groups of students

ASSESSMENT SYSTEM

Continuous evaluation: Participation during the classes and three small homework assignments to be done individually or in groups:40%

Final exam: 60%

Extra test. Those students who have not passed the subject in the regular period will have the opportunity to do an extra final test. Its percentage in the grade will be 100%, although for those students who did the partial tests in the regular period, I will apply the same rules as in the regular period, whenever this improves the grade of the extra test.

% end-of-term-examination:	40
% of continuous assessment (assigments, laboratory, practicals…):	60

BASIC BIBLIOGRAPHY

- A-L Barabasi Network science, http://barabasi.com/book/network-science#network-science, 2018
- E. Tufte The Visual Display of Quantitative Information (2nd Edition)., Graphic Press, 2001
- M.E.J. Newman Networks: An Introduction , Oxford University Press, 2010
- Rafa Donahue Fundamental Statistical Concepts in Presenting Data, http://biostat.mc.vanderbilt.edu/wiki/Main/RafeDonahue, 2018

ADDITIONAL BIBLIOGRAPHY

- Alberto Cairo The Truthful Art: Data, Charts, and Maps for Communication, New Riders, 2016
- Nathan Yau Visualize This, John Wiley & Sons, 2011