Network analysis and data visualization

Academic Year: (2021 / 2022) **Review date: 21-05-2021**

Department assigned to the subject: Department of Mathematics

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

Link to document

DESCRIPTION OF CONTENTS: PROGRAMME

1. Networks: general concepts and definitions
   1.1 Network introduction
      - Network importance and examples
      - Historical background of network science
      - Network types and attributes
   1.2 Network measures
      - Degree distributions and correlations

- Transitivity and clustering coefficient
- Connectedness and giant component

Workshop 1: Gephi (network visualization)

2. Network communities
   2.1 Centrality measures
   - Distances on networks, radius and diameter
   - Degree, closeness, harmonic and betweenness centralities
   - Eigenvector, Katz and PageRank centralities

   2.2 Network mesoscale analysis
   - Cliques and network motifs
   - Modularity measure
   - Community detection algorithms

   Workshop 2: iGraph and graph visualization (R)

3. Network models
   3.1 Random network models
   - Erdoös-Rényi (ER) random graph
   - Random Geometric Graph (RGG)
   - Configuration network models

   3.2 Simple rule network models
   - Stochastic block model
   - Barabási-Albert (BA) scale-free network model
   - Watts-Strogatz (WS) small-world network model

   Workshop 3: Netlogo, community detection algorithms and network models (R)

4. Social Networks
   4.1 Local and global properties of social networks
   - Examples of social networks and their properties
   - (Generalized) Friendship paradox
   - Six degrees of separation
   - Dunbar¿s numbers

   4.2 Social mechanisms
   - Homophily
   - Triadic closure
   - Strength of relationships

   Workshop 4: Network analysis

5. Network dynamics and applications
   5.1 Link prediction
   - Assortative, relational and proximity algorithms
   - Graph distance methods
   - Common neighbors methods
   - Preferential attachment
   - Katz score and hitting time
   - Community-based heuristics

   5.2 Spreading processes
   - Susceptible-Infected (SI) model
   - Susceptible-Infected-Removed (SIR) model
   - Susceptible-Exposed-Infected-Removed (SEIR) model
   - More advanced models

   Workshop 5: Link prediction and spreading processes

6.1 Data visualization
   6.1 Introduction to visualization
   - Types of visualizations
   - Examples of good visualizations
   - Examples of bad visualizations

   6.2 Introduction to data and charts
   - Types of data
   - Types of charts
   - Visualization tools

   Workshop 6: Data visualization (ggplot)
   Workshop 7: GoogleVis, R shiny app and geolocalised data visualization

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 two midterm assignments to be done individually or in groups: 60%

Final assignment: 40%

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