ANALYZING SOCIAL MEDIA

Bachelor in Data and Business Analytics BDBA SEP-2023
ASM-DBA.3.M.A

Area Others

Number of sessions: 20

Academic year: 23-24

Degree course: THIRD

Number of credits: 3.0

Semester: 1º

Category: COMPULSORY

Language: English

Professor: **JUAN JOSE MANJARIN COLON**

E-mail: jjmanjarin@faculty.ie.edu

Juanjo Manjarín received the highest grade for his Ph.D. in Theoretical Physics in String Theory and M-Theory from the Universdad Autónoma de Madrid (UAM), after earning an Advanced Studies Diploma (DEA) in Theoretical Physics from the Universdad Autónoma de Madrid (UAM). His teaching experience includes different international universities where he lectured on Complex Variable Analysis, Theoretical Mechanics or Classical Electrodynamics and in the IE University Mathematics, Statistics and Econometrics. He also teaches the course of Math and Stats for Data Analysis in the Bootcamp for Data Science in the IE Exponential Learning. He has published a number of papers on international journals on mathematics and theoretical physics. He was reviewer for Mathematical Reviews from 2003 to 2005 and during 2016 he also worked in El Pais in the realization of some divulgative science materials.

He has also corporate experience on different TV and cinema production companies: Gestmusic Endemol, 7 y Acción S.L., Hill Valley S.L., 100 Balas S.L. or Zebra Producciones in TV shows such as “El Hormiguero” or “Esto es vida!”, receiving prizes Ondas and Rose d'Or in 2008 and 2009. He has director, producer and post-producer of different short films and now as director of E8 Producciones is recording a documentary film.

jjmanjarin@faculty.ie.edu

SUBJECT DESCRIPTION

We can see networks everywhere, from the www or the electric web to how viruses or innovations spread. These are a special type of situation in which not only the individuals are relevant but most of all, the type of relationship linking them: the last allow us to determine or create different patterns among the first ones.

The focus of this course is to provide an introduction to theory and methods used in network science covering basic topics as the small—world problem or the identification of relevant individuals in a network and an introduction to more advanced ones as the spreading phenomena or similarity and influence in networks.

This will be done both analytically and with the help of computers: coding in R and using Gephi.

LEARNING OBJECTIVES

The course provides an introduction to network analysis. Students will learn a wide range of social network methods and how to understand the theoretical background allows them to choose the best approach. The main objectives of the course are:

Unit 1: Fundamentals to Networks

- Know the types of networks.
- Understand what a network dataset is and their use as explanatory or outcome.
- Know the Local, Dyadic, and Global level analysis in a network study.
- Learn how to describe a network mathematically.
- Understand the main measures to describe a network.
- Understand the Correlations in Networks.
- Learn the difference between Influence, Homophily and Environment effects.
- Study the subgroups and communities in a network.
- Understand the limits of the classification algorithms.

Unit 2: Models of Networks

- Learn the Statistical Description of a Network.
- Understand Hubs and Universality.
- Learn the laws governing real networks and their evolution.
- Development of ERGM models and their interpretation.

Unit 3: Dynamical Networks

- Understand the basic classical models to describe the spreading phenomena.
- Study how the network properties affect these phenomena.
- Understand Herd Behaviour, and ICM models for Bayesian Influence.
- Understand French's Theory of Social Power and the DeGroot's generalization.

TEACHING METHODOLOGY

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	33.33 %	25.0 hours
Discussions	6.67 %	5.0 hours
Exercises in class, Asynchronous sessions, Field Work	13.33 %	10.0 hours
Group work	20.0 %	15.0 hours
Individual studying	26.67 %	20.0 hours
TOTAL	100.0 %	75.0 hours

PROGRAM

UNIT I: FUNDAMENTALS OF NETWORKS

SESSION 1 (LIVE IN-PERSON)

Networks and their Data Sets

About the course; Syllabus objectives; Introduction to Networks

- Why networks
- Types of Networks
- Types of Relations
- Levels of Analysis
- Network data: compositional and structural
- Networks as outcome and as explanatory variables

Book Chapters: Networks; Chapter 6.1 (See Bibliography

Book Chapters: Analyzing Social Networks; Chapter 1 (See Bibliography)

- As a prerequisite, the Unit I of [New] should have been read
- Lectures
    - Course Slides
    - [New] Ch. 6.1
    - [BEJ] Ch. 1

SESSION 2 (LIVE IN-PERSON)

The Mathematical Description of Networks

- Matrix representation of networks:
    - The Adjacency Matrix;
    - The Incidence Matrix
    - The Graph Laplacian
- Simple, Weighted and Directed Networks
- Degree and Degree Distribution
- Connectivity
    - Trees, Walks and Paths;
    - Components

Book Chapters: Networks; Chapter 6 (See Bibliography

Book Chapters: Social and Economic Networks; Chapter 2, Section 1 (See Bibliography)

- Lectures
    - [New] Chapter 6
    - [Jac] Chapter 2, Section 1

- Course Video: "The Mathematics of Networks”

SESSION 3 (LIVE IN-PERSON)

Local Level of Analysis

- Degree derived centralities
    - Degree Centrality
    - Eigenvector Centrality
    - Katz Centrality
    - PageRank Centrality
    - Diffusion Centrality
- Betweenness Centrality;
- Closeness Centrality;
- Valued Networks:
    - PN-Centrality

Article: Centrality in Social Networks, a conceptual clarification (Social Networks, 1 (1978/79) 215-239) (CED)

Book Chapters: Networks; Chapter 7, Sections 1-4 (See Bibliography

Book Chapters: Social and Economic Networks; Chapter 2, Section 2 (See Bibliography)

Book Chapters: Analyzing Social Networks; Chapter 10 (See Bibliography)

- Lectures:
    - [New] Chapter 7, Sections 1-4
    - [Jac] Chapter 2, Section 2
    - [BEJ] Ch. 10

- Course Video:
    - “Centralities in R”

SESSION 4 (LIVE IN-PERSON)

Lab Session: Local Level

- Selected Exercises from [New] and [Jac]
- Computer problems in R

SESSION 5 (LIVE IN-PERSON)

Dyadic Level of Analysis

- Assortativity and Disassortativity;
- Forces of assortativity
- Measuring Degree Correlations;
- Correlations in Real Networks;
- Impact of Correlations

Book Chapters: Networks; Chapter 7, Sections 6-7 (See Bibliography

Book Chapters: Social and Economic Networks; Chapter 3, Section 2 (See Bibliography)

Book Chapters: Network Science; Chapter 7, Sections 2, 3, 5 & 7 (See Bibliography)

- Lectures
    - [New] Chapter 7, Sections 6-7

- [Jac] Chapter 3, Section 2
- [ALB] Chapter 7, Sections 2, 3, 5 & 7

- Course Videos:
  - “Assortativity”

SESSION 6 (LIVE IN-PERSON)

Dyadic Level of Analysis

- Assortative Mixing: Influence, Homophily and Environment in Networks;
- Statistical Tests to Distinguishing between Homophily and Influence

Book Chapters: Networks; Chapter 7, Section 7; Chapter 10, Section 7 (See Bibliography
Book Chapters: Social Media Mining: An Introduction; Chapter 8 (See Bibliography)
Book Chapters: Statistical Analysis of Network Data with R; Chapter 4, Section 5 (See Bibliography)

- Lectures
  - [New] Chapter 7, Section 7; Chapter 10, Section 7
  - [KC] Chapter 4, Section 5
  - [SMM] Chapter 8

SESSION 7 (LIVE IN-PERSON)

Lab Session: Dyadic Level

- Solutions to selected exercises from [New] and [Jac]
- Computer problems in R

SESSION 8 (LIVE IN-PERSON)

Global Level of Analysis

- Group Cohesion
- Reciprocity
- Transitivity:
  - Global and Local Clustering Coefficient
  - Triad Census
- Centralization,

Book Chapters: Analyzing Social Networks; Chapter 9 (See Bibliography)
Lectures:

- [BEJ] Ch. 9

SESSION 9 (LIVE IN-PERSON)

Global Level of Analysis

- Cliques:
  - N-cliques
  - N-clans
  - N-clubs
- Core-Periphery Strucutre

- K-cores
- K-plexes

- Definition of Communities

Book Chapters: Networks; Chapter 14 (See Bibliography

Book Chapters: Analyzing Social Networks; Ch. 11.1, 11.2, 11.4 and 11.7 (See Bibliography)

- Lectures
  - [BEJ] Ch. 11.1, 11.2, 11.4 and 11.7
  - [New] Chapter 14

SESSION 10 (LIVE IN-PERSON)

Lab Session: Global Level

- Solution to selected exercises
- Computer problems in R

SESSION 11 (LIVE IN-PERSON)

MIDTERM EXAM

UNIT II: MODELS OF NETWORKS

SESSION 12 (LIVE IN-PERSON)

Random Networks

- Static Random-Graph Models:
  - Binomial and Poisson Networks: the Erdös-Renyi Model
  - Small-Worlds: the Watts-Strogatz Model
  - Markov graphs
  - The Configuration Model
- Properties of Ranom Networks
  - Degree Distribution
  - Thresholds and Phase transitions
  - Connectedness
  - Giant Components

Book Chapters: Networks; Chapter 10, Sections 1-2; Chapter 11 (See Bibliography

Book Chapters: Social and Economic Networks; Chapter 4, Sections 1-4 (See Bibliography)

Book Chapters: Statistical Analysis of Network Data with R; Chapter 5, Sections 2-3, 4.1, 5.2 (See Bibliography)

- Lectures
  - [New] Chapter 10, Sections 1-2; Chapter 11
  - [Jac] Chapter 4, Sections 1-4
  - [KC] Chapter 5, Sections 2-3, 4.1, 5.2

SESSION 13 (LIVE IN-PERSON)

Growing Networks

- Scale-Free networks
    - Preferential Attachment
    - Growth
    - Universality
    - Degree Exponent and Regimes
- Hybrid Models
- Small Worlds, Clustering, and Assortativity

Book Chapters: Networks; Chapter 10 (See Bibliography
Book Chapters: Social and Economic Networks; Chapter 5, sections 2 - 4 (See Bibliography)
Book Chapters: Network Science; Chapter 4, Sections 1 - 7 (See Bibliography)

- Lectures
    - [Jac] Ch- 5, sections 2 - 4
    - [New] Chapter 10
    - [ALB] Chapter 4, Sections 1 - 7

- Course Videos: "Scale Free Networks"

SESSION 14 (LIVE IN-PERSON)

Statistical Models of Networks

- Exponential Random Graph Models (ERGM);
    - Definition
    - Understanding the terms of the model
    - Estimation and interpretation of the coefficients
    - Goodness of fit

Book Chapters: Statistical Analysis of Network Data with R; Chapter 6, Sections 1-3 (See Bibliography)
Book Chapters: Exponential Random Graph Models in Social Networks; Chapters 2 - 5 (See Bibliography)

- Lectures
    - [KC] Chapter 6, Sections 1-3
    - [LKR] Chapters 2 - 5

SESSION 15 (LIVE IN-PERSON)

Lab Session:

- Computer problems in R

UNIT III: APPLICATIONS OF NETWORKS

SESSION 16 (LIVE IN-PERSON)

Groups and Communities

- Graph Partitioning;
- Communities;
    - Working Hypotheses
    - Detection Algorithms;
    - Balanced, Modular and Hierarchical Communities;
    - Limitations of the communities algorithms

Book Chapters: Networks; Chapter 14 (See Bibliography

Book Chapters: Statistical Analysis of Network Data with R; Chapter 4, Sections 3-4 (See Bibliography)

- Lectures
    - [New] Chapter 14
    - [KC] Chapter 4, Sections 3-4

SESSION 17 (LIVE IN-PERSON)

Spreading Phenomena

- Herd Behaviour;
- ICM;
- The Role of the Network Structure
- Modelling the Diffusion of Innovations

Book Chapters: Social and Economic Networks; Chapter 7, Sections 1 - 2 (See Bibliography)

Book Chapters: Social Media Mining: An Introduction; Chapter 7, Sections 1 - 3 (See Bibliography)

- Lectures
    - [Jac] Chapter 7, Sections 1 - 2
    - [SMM] Chapter 7, Sections 1 - 3

SESSION 18 (LIVE IN-PERSON)

Social Influence and Learning in Networks

- Opinion Leaders;
- French's Theory of Social Power;
- Learning in Networks:
    - DeGroot's Model

Book Chapters: Social and Economic Networks; Chapter. 8, Sections 1 - 3 (See Bibliography)

Book Chapters: Social Influence Network Theory; Chapter 2, Section 1 (See Bibliography)

- Lectures
    - [Jac] Chapter. 8, Sections 1 - 3
    - [SIN] Chapter 2, Section 1

SESSION 19 (LIVE IN-PERSON)

Lab on Communities and Diffusion

- Selected Exercises from [New] and [Jac]
- Computer problems in R

SESSION 20 (LIVE IN-PERSON)

FINAL EXAM

EVALUATION CRITERIA

The evaluation will involve the following weights

criteria	percentage	Learning Objectives	Comments
Final Exam	35 %		
Midterm Exam	25 %		
Cases Analysis	20 %		
Class Participation	20 %		

RE-SIT / RE-TAKE POLICY

Case Analysis

By the end of the course you will have to submit an analysis of a real data set. You will be given the activity at the begnining of the course and will have to complete the following milestones:

- EDA of the network: metrics, local, dyadic and global analysis
- Model: Identification of the degree distribution and ERGM simulation
- Simulation of diffusion processes and Community detection

It wil be completed in groups. These groups will be chosen by the students at the beginning of the course, however, if after the first week these have not been created, the professor will create them randomly. This analysis will account for the 20% of the final grade.

Midterm Exam

The midterm exam accounts for the 30% of the final grade.It will be an open books, individual exam which will contain both, analytical and computational questions, and the interpretation of the computations will have the largest impact in the grade (at least the 60%). Note that everything that is said in the sessions may be a question of the exam, which will cover the contents from session 1 to 10.

The date of this exam cannot be changed under any circumstance: it will be on the session number 11, as established in this syllabus. Consider this before making any plans that may affect your attendance.

Final Exam

The Final exam accounts for the 35% of the final grade. It is an open books, individual exam, which is mostly computer based in term of computations, but the largest impact of the grade will be on the interpretations of you results (at least the 60%). The material covered in this exam will be the whole course since a proper interpretation of your simulations will require knowledge from the 10 first sessions, however, the main questions will be about finding communities, simluation diffusion processes and understanding the role of the network structure in them.

The date of this exam cannot be changed under any circumstance: it will be on the session number 20, as established in this syllabus. Consider this before making any plans that may affect your attendance.

Class Participation

Class participation is evaluated taking into account both: in- and out- class participation. The participation in the lab days solving exercises is an important part of the grade.

Apart from this, general participation will be evaluated based on the following criteria:

Quality (not quantity) of your participation in class discussion: The most important dimension of participation concerns what it is that you are saying. A high-quality comment reveals depth of insight, rigorous use of case evidence, consistency of argument, and realism. Frequency refers to the attainment of a threshold quantity of contributions that is sufficient for making a reliable assessment of comment quality.

The logic is simple: if contributions are too few, one cannot reliably assess the quality of your remarks. However, once threshold quantity has been achieved, simply increasing the number of times you talk does not automatically improve your evaluation. Beyond the threshold, it is the quality of your comments that must improve. In particular, one must be especially careful that in claiming more than a fair share of "airtime", quality is not sacrificed for quantity. Finally, your attempts at participation should not be such that the instructor has to "go looking for you". You should be attempting to get into the debate on a regular basis.

You might want to avoid being classified as one of the following types of students:

- **Repeaters**, i.e., students that, consciously or unconsciously, make comments that are just repeats/rephrasing of what has already been said (by other students, or you). This wastes time and adds nothing to learning.
- **Ramblers**, i.e., students that take a lot of time to say simple things or they may tell long personal/professional stories, or they roam into topics that are not relevant, or simply make low-quality comments just to participate. They waste valuable time and prevent other students from being able to participate.
- Students that have been distracted (by Facebook, etc.) or who have stopped paying attention and then, later, when they realized they have missed a term or concept, they ask you about it.

As for the evaluation: 60% of your grade corresponds to the lab days, and the 40% to the continuous evaluation.

RE-SIT / RE-TAKE POLICY

Each student has four (4) chances to pass any given course distributed over two (2) consecutive academic years. Each academic year consists of two calls: one (1) ordinary call (during the semester when the course is taking place); and one (1) extraordinary call (or "re-sit") in June/July.

Students who do not comply with the 70% attendance requirement in each subject during the semester will automatically fail both calls (ordinary and extraordinary) for that Academic Year and have to re-take the course (i.e., re-enroll) during the next Academic Year.

Regarding to the newly implemented ´liquid learning´ model, all students must still abide by the same IEU attendance policy, including those students who are connecting remotely to class sessions and not physically in the classroom because they are unable to be physically in Spain, on campus. During the sessions, students connecting remotely are required to fully connect their camera and microphone at all times, and must actively participate during the sessions (using all necessary audiovisual equipment), just as their fellow peers who are physically present in the classroom on campus.

The Extraordinary Call Evaluation criteria will be subject to the following rules:

- Students failing the course in the ordinary call (during the semester) will have to re-sit evaluation for the course in June / July (except those students who do not comply with the attendance rule, and therefore will not have that opportunity, since they will fail both calls and must directly re-enroll in the course during the next Academic Year).
- It is not permitted to change the format nor the date of the extraordinary call exams or deadlines under any circumstance. All extraordinary call evaluation dates will be announced in advance and must be taken into consideration before planning the summer (e.g. internships, trips, holidays, etc.)
- The June/July re-sit will consist of a comprehensive evaluation of the course. Your final grade

for the course will depend on the performance in this exam or evaluation only. I.e., continuous evaluation over the semester (e.g. participation, quizzes, projects and/or other grade components over the semester) will not be taken into consideration on the extraordinary call. Students will have to achieve the minimum passing grade of 5 and the maximum grade will be capped at 8.0 (out of 10.0) – i.e., “notable” in the extraordinary call.

- Re-takers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will need to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as re-takers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained as a retaker during the ordinary call (i.e., the 3rd call) is 10.0 (out of 10.0).

After exams and other assessments are graded by the professor (on either the ordinary or extraordinary call), students will have a possibility to attend a review session (whether it be a final exam, a final project, or the final overall grade in a given course). Please be available to attend the session in order to clarify any concerns you might have regarding your grade. Your professor will inform you about the time and place of the review session.

BIBLIOGRAPHY

Compulsory

- Mark Newman. (2018). *Networks.* 2nd Edition. Oxford University Press. ISBN 9780198805090 (Printed)

  This book is denoted as [NEW] in the program section

- Matthew O. Jackson. (2008). *Social and Economic Networks.* 1st. Princeton University Press. ISBN 9780691148205 (Printed)

  [Jac] in the sessions

- Noah E. Friedkin and Eugene C. Johnsen. (2013). *Social Influence Network Theory.* 1st. Cambridge University Press. ISBN 9781107617674 (Printed)

  [SIN] in the sessions

Recommended

- A.L. Barabási. (2016). *“Network Science”.* Cambridge University Press. ISBN 9781107076266 (Printed)

  You can dowloand the book and other resources from the web of the author: http://networksciencebook.com. This book is denoted as [ALB] in the program section

- R.Zafarani, M.A. Abbasi and H. Liu. (2014). *"Social Media Mining: An Introduction".* Cambridge University Press. ISBN 9781107018853 (Printed)

  You can download the book and other resources from the web of the authors: http://dmml.asu.edu/smm. This book is denoted as [SMM] in the program section

- E.D. Kolaczyk and G.Csárdi. (2020). *"Statistical Analysis of Network Data with R".* 2nd. Springer. ISBN 9783030441289 (Printed)

  This book is denoted as [KC] in the program section

- D. Lusher, J. Koskinen and G. Robins. (2013). *Exponential Random Graph Models in Social Networks.* 1st. Cambridge University Press. ISBN 9780521141383 (Printed)

[LKR] in the sessions

- S.P. Borgatti, M.G. Everett and J.C. Johnson.. (2008). *Analyzing Social Networks.* 2nd. Sage. ISBN 9781526404107 (Printed)

[BEJ] in the sessions. Take a look at:

https://sites.google.com/site/analyzingsocialnetworks/home?authuser=0

https://study.sagepub.com/borgatti2e

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

ETHICAL POLICY

Please, check the University's Ethics Code here. The Program Director may provide further indications.