ANALYZING SOCIAL MEDIA

Dual Degree in Business Administration & Data and Business Analytics BBADBA SEP-2024 ASM-NBDA.3.M.A

Area Data Science

Number of sessions: 20

Academic year: 24-25

Degree course: THIRD

Number of credits: 3.0

Semester: 2º

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 where he has taught Mathematics, Statistics, Econometrics, Social Media Analytics and Programming in R and Python. 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 and he was reviewer for Mathematical Reviews from 2003 to 2005. His research interests are Quantum Information and Computing and Network Science together with their applications in Data Science.

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 was director, producer and post-producer of different short films and now as director of E8 Producciones is recording a documentary film. He also worked in El Pais in the realization of some divulgative science materials.

Office Hours

Office hours will be on request. Please contact at:

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 mostly analytically, but you will the student will also use specific software as **Gephi**, or programming in **Python** and/or in **R**

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: Basics of Networks

- Know the types of networks
- Understand what a network dataset is and its use as explanatory or outcome.
- Learn how to describe a network mathematically
- Understand de Node, Dyadic, and Network level of analysis
- Understand the main metrics to describe a network at the different levels
- Understand the Correlations in Networks
- Learn the difference between Influence, Homophily and Environment effects

Unit 2: Models of Networks

- Learn the Statistical Description of a Network: Random Networks, Preferential attachment
- Understand Hubs and Universality
- Learn the laws governing real networks and their evolution
- Perform statistical tests to determine the nature of the Degree Distributions
- Perform statistical tests in the different levels of analysis

Unit 3: Applications

- Study the subgroups and communities in a network
- Understand the basic classical models to describe the spreading phenomena
- Understand Herd Behaviour, and ICM models for Bayesian Influence
- Study Schelling model of seggregation
- Study the Social influence Models (French's and DeGroot's)

TEACHING METHODOLOGY

IE University teaching method is defined by its collaborative, active, and applied nature. Students actively participate in the whole process to build their knowledge and sharpen their skills. Professor's main role is to lead and guide students to achieve the learning objectives of the course. This is done by engaging in a diverse range of teaching techniques and different types of learning activities such as the following:

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	33.3 %	25.0 hours
Discussions	6.7 %	5.0 hours
Exercises in class, Asynchronous sessions, Field Work	13.3 %	10.0 hours
Group work	20.0 %	15.0 hours
Individual studying	26.7 %	20.0 hours
TOTAL	100.0 %	75.0 hours

AI POLICY

In today's world, generative artificial intelligence (GenAI) is changing how we work, study and, in general, how we get things done. However, in the context of this course, the use of GenAI is not permitted, unless it is otherwise stated by the instructor. The use of GenAI tools would jeopardize the students' ability to acquire fundamental knowledge or skills of this course.

If a student is found to have used AI-generated content for any form of assessment, it will be considered academic misconduct, and the student might fail the respective assignment or the course.

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)

- Mandatory Readings:
  - [FCNS] Ch. 0
  - [SNA] Ch. 1.3, Ch. 2 sections 1.1, 3 and 4

- Additional Readings
  - Course Slides
  - [NEW] Ch. 6.1

SESSION 2 (LIVE IN-PERSON)

The Mathematical Description of Networks

- Representation of networks: the Adjacency Matrix;
- Simple, Weighted and Directed Networks
- Connectivity: Trees, Walks, Paths, and Cycles;
- Components; Bridges and Cut-points
- Complete Networks
- Degree and Degree Distribution
- Affiliation Networks: The Incidence Matrix
- The Graph Laplacian

Book Chapters: Networks; Chapter 6 (See Bibliography

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

- Mandatory Readings:
    - [SEN] Chapter 2, Section 1
    - [SNA] Chapter 4, Sections 2 and 3, except 4.2.9, 4.2.10, 4.3.8

- Additional Readings and documentation
    - [NEW] Chapter 6
    - Course Slides
    - Course Video: "The Mathematics of Networks”

SESSION 3 (LIVE IN-PERSON)

Local Level of Analysis

- Local Clustering Coefficient
- Centrality
    - 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)

- Madatory Readings:
    - [FCNS] Chapter 3, Sections 1, 2, and 3
    - [SNA] Chapter 5, Sections 1, 2, and 3, except 5.2.4

- Additional Readings:
    - [NEW] Chapter 7, Sections 1-4
    - [SEN] Chapter 2, Section 2
    - Course Slides

SESSION 4 (LIVE IN-PERSON)

Lab Session: Local Level

- Solution to selected exercises from the problem set

For this session it is important that the students have read the chapters below and are a bit familiar with the Python package NetworkX

- Mandatory readings:
    - [FCNS] Chapter 1
    - [CNAP] Chapter 3

SESSION 5 (LIVE IN-PERSON)

Dyadic Level of Analysis

- Notions of Equivalence:
    - Structural Similarity
    - Regular Similarity
- Reciprocity
- Structural Balance
    - Triad Census
    - Biplots
    - Distribution of a triad census
    - Testing Structural hypothesis

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

- Mandatory Readings
    - [SNA] Chapter 14
    - Course Slides

- Additional Readings:
    - [NEW] Chapter 7, Section 6
    - Course Slides

SESSION 6 (LIVE IN-PERSON)

Dyadic Level of Analysis

- Contexts
- Assortative Mixing: Assortativity and Disassortativity;
    - Measuring Degree Correlations;
    - Correlations in Real Networks;
    - Impact of Correlations

- Forces of assortativity
- 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)

- Mandatory Readings
  - [FCSN] Chapter 2, Section 1
  - [SEN] Chapter 3, Section 2

- Additional Readings
  - [NEW] Chapter 7, Section 7; Chapter 10, Section 7
  - [SMM] Chapter 8, Section 4
  - [CNAP] Chapter 8, Section 5
  - Course Slides

SESSION 7 (LIVE IN-PERSON)

Lab Session: Dyadic Level

- Solutions to selected exercises of the Dyadic Level from the problem set

SESSION 8 (LIVE IN-PERSON)

Global Level of Analysis

- Global Metrics:
  - Diameter,
  - Density,
  - Global Clustering Coefficient
- Centralization
- Core-Periphery Structure

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

- Mandatory Readings:
  - [CNAP] Chapter 8
  - [SNA] Chapter 5.1.3, 5.2.1, 5.2.2, 5.2.3
  - Course Slides
- Additional Readings:
  - [NEW] Chapter 14.7.3

SESSION 9 (LIVE IN-PERSON)

Global Level of Analysis

- Group Cohesion
- Groups and Subgroups
- Groups based on Nodal degree:
  - Cliques
  - K-cores

        - K-plexes
    - Groups Based on Reachability
        - N-cliques
        - N-clans
        - N-clubs
    - Mandatory Readings
        - [SNA] Chapter 7, Sections 1 to 4
        - [CNAP] Chapter 11, Sections 1 to 5
        - Course Slides

SESSION 10 (LIVE IN-PERSON)

Groups and Communities

- Definition of Community
    - Main Hypotheses
- Types of Communities and Algorithms
    - Node Based Communities
    - Group Based Communities
        - Balanced Communities
            - Graph Bisection
            - Fiedler Vectors
        - Hierarchical Communities
            - Girvan-Newman and Ravasz algorithms
        - Modular Communities;
            - Greedy and Louvain Algorithms
            - Limitations of the communities algorithms
- Evolution of Communities
    - Linear Threshold Models
    - Schelling's Model of Seggregation

Article: Modularity and Community Structure in Networks (Proc. Natl. Acad. Sci. USA 103, 8577-8582 (2006)) (CED)

Book Chapters: Networks; Chapter 14 (See Bibliography

- Mandatory Readings:
    - [FCNA] Chapter 6
    - [CNAP] Chapter 11, Section 6

- Additional Readings:
    - [NEW] Chapter 14
    - [SEN] Chapter 13, Section 2
    - [SMM] Chapter 6, and Chapter 8, Section 2
    - Course Slides

SESSION 11 (LIVE IN-PERSON)

Lab Session: Global Level

- Solution to selected exercises for the Global Level of Analysis from the problem set
- Group Graded Activity: Schelling model on Python

Article: Scale-Free Networks (Scientific American , Vol. 288, No. 5 (MAY 2003), pp. 60-69) (CED)

SESSION 12 (LIVE IN-PERSON)

MIDTERM EXAM

UNIT II: MODELS OF NETWORKS

SESSION 13 (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 Random 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)

- Mandatory Readings
  - [FCNS] Chapter 5, Sections 1 to 3
- Additional Readings
  - [NEW] Chapter 10, Sections 1-2; Chapter 11
  - [SEN] Chapter 4, Sections 1-4
  - Course Slides

SESSION 14 (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)

- Mandatory Readings
    - [FCSN] Chapter 5, Sections 4 and 5

- Additional Lectures
    - [SEN] Chapter 5, sections 2 to 4
    - [NEW] Chapter 10
    - Course Slides

SESSION 15 (LIVE IN-PERSON)

Statistical Inference on Networks

- Inference on the node, dyadic, and network levels
- Inference on network models
    - Random Networks: Goodness of fit with Chi-square test and G-test
- Inference on Scale-Free network models
    - Linear models for scale free networks
    - Maximum likelihood and KS-test
- Mandatory Readings
    - Course Slides

SESSION 16 (LIVE IN-PERSON)

Lab Session:

- Solution to selected exercises from the problem set
- Graded Group Activity: Testing Statistical Models on Python

UNIT III: APPLICATIONS OF NETWORKS

SESSION 17 (LIVE IN-PERSON)

Spreading Phenomena

- Information in Networks
- Diffusion processes
    - Herd Behaviour and Bayesian Herd Behaviour
    - Information Cascades:
        - The Independent Cascade Model (ICM)
    - Diffusion of Innovations
    - Epidemics

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)

- Mandatory Readings

- [FCSN] Chapter 7, Sections 1 to 3
- Course Slides

- Lectures
  - [SEN] Chapter 7, Sections 1 and 2
  - [SMM] Chapter 7, Sections 1 to 4

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
- Learning Biases

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)

- Mandatory Readings
  - [SIN] Chapter 1, Sections 1.1, and 1.3, Chapter 2, Sections 1.1, and 1.2

- Additional Readings
  - [SEN] Chapter. 8, Sections 1 - 3
  - [SIN] Chapters 5 and 6
  - Course Slides

SESSION 19 (LIVE IN-PERSON)

Lab on Communities and Diffusion

- Solution to selected exercises of Diffusion processes and Social Influence processes
- Graded Group activity: De Groot model on Python

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	30 %		
Activities	15 %		
Class Participation	20 %		

RE-SIT / RE-TAKE POLICY

Activities

This part of the grade accounts for the 15% of the final grade and will be completed and submitted in groups. For the last 30 minutes of some sessions you will receive a computational task to be solved in Python as an application of some of the unit concepts to a real network:

1. Session 11: Schelling Model
2. Session 16: Inference on Networks and Network Models
3. Session 19: DeGroot's model

Each activity has a weight of the 5% (0.5 points) in the final grade. Note that everyone in the group will receive the same grade, and that if you are absent in any of them you will be marked with a 0 without the possibility to do it individually other day.

Midterm Exam

The midterm exam accounts for the 30% of the final grade.It will be a closed books, individual exam which will contain analytical and computational questions. The analytical part will be split into multiple choice, mathematical questions and a case analysis. The computational questions willb e mostly based, but not only, on the analysis of pseudocode. Note that everything that is said in any session may be a question of the exam, which will cover the contents from session 1 to 11.

The students may bring a formula sheet and a calculator to this exam. The formula sheet will consist of two handwritten A4 pages (4 sides), at most, with anything that they consider as important.

The date of this exam cannot be changed under any circumstance, and individual exams will not be placed for any student. Consider this before making any plans that may affect your attendance. Failing to attend this exam implies a 0 on this part of the grade.

Final Exam

The Final exam accounts for the 35% of the final grade. It is a closed books, individual exam, with a similar structure to the midterm exam. It will contain analytical and computational questions covering the whole course but mostly focused in the second and third units If the student does not achive a minimum of 3.5 points (out of 10) in this exam, the course will be failed and none of the other grades will be evaluated.

The students may bring a formula sheet and a calculator to this exam. The formula sheet will consist of three handwrtitten A4 pages (6 sides), at most, with anything that they consider as important.

The date of this exam cannot be changed under any circumstance, and individual exams will not be placed for any student. Consider this before making any plans that may affect your attendance. Failing to attend this exam implies that the student fails the course.

Class Participation

Class participation is evaluated as:

- **In-** and **Out- class** participation: answering or making questions during the sessions, or requesting office hours, for example. This will give you a maximum of 3 points (0.6 points in the overall final grade)
- Theory session's **mini-quizzes**: As you can see in the section describing the Program in this document, each session has some mandatory readings for that session. Then, at the beginning of the sessions, the students will receive a 5 minutes with two random questions about the readings. Overall there will be 12 of those mini-quizzes (not considering the first session of the course, the exams and the labs). The average of these mini-quizzes will give your 7 points in participation (1.4 points in the overall final 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 80% 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

- Filippo Menczer, Santo Fortunato, and Clayton A. Davis. (2023). *A First Course in Network Science.* First Edition (2020), version 3. Cambridge University Press. ISBN 9781108471138 (Printed)

  FCNS in the sessions.

- Stanley Wasserman and Katherine Faust. (2019). *Social Network Analysis.* First, 6th printing. Cambridge University Press. ISBN 9780521387071 (Printed)

  SNA in the sessions

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

  SEN 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

- Dmitry Zinoviev. (2018). *Complex Network Analysis in Python.* First. The Pragmatic Programmers LLC. ISBN 9781680502695 (Printed)

  CNAP in the sessions

Recommended

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

  NEW in the session

- 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 sessions

- Edward L. Platt. (2019). *Network Science with Python and NetworkX Quick Start Guide.* First. Packt Publishing Ltd.. ISBN 9781789955316 (Printed)

NSPN in the sessions

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.