Course guide
270617 - CSN - Complex and Social Networks

Last modified: 13/07/2022

Unit in charge:	Barcelona School of Informatics
Teaching unit:	723 - CS - Department of Computer Science.

Degree:	MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).
	MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).

Academic year: 2022   **ECTS Credits:** 6.0   **Languages:** English

LECTURER

Coordinating lecturer:	RAMON FERRER CANCHO
Others:	Primer quadrimestre:
	ARGIMIRO ALEJANDRO ARRATIA QUESADA - 10
	RAMON FERRER CANCHO - 10

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE3.1. Capability to identify computational barriers and to analyze the complexity of computational problems in different areas of science and technology as well as to represent high complexity problems in mathematical structures which can be treated effectively with algorithmic schemes.
CEE3.2. Capability to use a wide and varied spectrum of algorithmic resources to solve high difficulty algorithmic problems.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.

Generical:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Basic:

CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.

CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

TEACHING METHODOLOGY

The theory sessions will be done primarily by the professor using either the blackboard or projected slides.

The lab work will be done in front of the computer. Students are expected to be working on their assignment, and the professor will explain all that is necessary to follow the class in the beginning of the session. Each lab session will be accompanied by a thorough guide describing the work that the students need to do.

All the material relevant for the course will be available from the course's website.

LEARNING OBJECTIVES OF THE SUBJECT

STUDY LOAD

Type	Hours	Percentage
Hours large group	54,0	36.00
Self study	96,0	64.00

Total learning time: 150 h

CONTENTS

Introduction. What are networks? Simple measures and models for networks

Description:
- Examples of real networks: social networks, information networks, technological networks and biological networks.
- Terminology: points (vertices, nodes, sites, actors) and lines (edges, arcs, links, bonds, ties,...).
- Types of networks: unweighted/weighted networks, undirected/directed networks, ...
- Classic network properties: the small-world phenomenon (distance metrics), heterogeneous degree distribution (power-law degree distribution, high clustering or transitivity.
- Classic network models: the Erdös-Rényi model, the Watts-Strogatz model, the Barabási-Albert model.

The degree distribution of a network and its analysis

Description:
- The empirical and the theoretical degree distribution. Undirected degree, in-degree and out-degree.
- Theoretical distributions: zeta family, binomial distribution.
- Fitting of the degree distribution: visual fitting, linear and non-linear regression and maximum likelihood.
- Introduction to standard model selection: parsimony versus quality of fit, Akaike information criterion.

Measures of networks

Description:
- Distance metrics: geodesic paths, local distance metrics (mean geodesic distance, closeness centrality), global distance metrics (diameter, mean geodesic distance, mean closeness centrality). Algorithms for computing distance.
- Clustering metrics: transitivity, clustering (different metrics). Algorithms for computing clustering.
- Degree correlations: assortative vs disassortative mixing by degree, degree correlation metrics (Pearson correlation vs Spearman rank correlation, mean degree of nearest neighbours.

Statistical testing of network measures

Description:
- Introduction to hypothesis testing: qualitative versus quantitative testing, families of null hypotheses (Erdös-Renyi model, configuration model, switching model,...), p-values.
- Introduction to Monte Carlo testing: general scheme, uniform random number generators, uniformly random permutation, random Erdös-Rényi graph with constant and variable number of edges.
- The configuration or matching model.
- The switching model.

Advanced measures for networks. Centrality.

Description:
Centrality:
- Qualitative notions of centrality of a node.
- Quantitative definitions of centrality: degree centrality, closeness centrality, betweenness centrality, eigenvector centrality and PageRank (google).

Finding community structure in networks

Description:
- Introduction to community structure.
- How to quantify the quality of a community structure: intra-cluster versus inter-cluster density. Other metrics: conductance, expansion, internal density, cut ratio, normalized cut, Flake's out degree fraction .... and modularity.
- Methods for detection of community structure: hierarchical clustering algorithms (agglomerative hierarchical clustering, node similarity measures), Girvan-Newman algorithm, modularity optimization algorithms (the modularity Q, algorithms to maximize modularity, the Louvain method, spectral modularity optimization), graph partitioning algorithms (minimum bisection algorithms, Kernighan-Lin algorithm) and clique percolation method.

The Minimum Linear Arrangement problem

Description:
- Linear arrangements of networks: linear/Euclidean distance, metrics (mean dependency length).
- Mean dependency length in trees: expectation in random linear arrangements, lower bounds, upper bounds (non-crossing trees),...
- The minimum linear arrangement problem: cost function and complexity.
- The minimum linear arrangement of trees: algorithms, particular cases.
- Introduction to crossing theory for trees. Number of crossings: upper bound, expectation in random linear arrangements (the particular case of uniformly random trees).

Network dynamics

Description:
Introduction to network dynamics:
- Classic models that generate networks: the Barabáasi-Albert model (growth and preferential attachment), copying models, fitness based models and optimization models.
- The Barabási-Albert model. The statistical properties: vertex degree as a function of time, the power-law degree distribution.
- The copying model. The statistical properties: vertex degree as a function of time and degree distribution.
- Fitness models. Fitness functions and degree distribution.
- Optimization models. Trade-off between geodesic distance and link density. Cost function.

Advanced network models:
- Modifications of the Barabási-Albert model
- Liu et al's hybrid model.
- Bianconi-Barabási hybrid model.
- Dorogovtsev-Mendes model (accelerated growth).
- Other models.

Sampling in networks

Description:
- Introduction to sampling: motivation and importance and goals.
- Sampling strategies: random node selection, random edge selection, crawling-based.
- Biases of sampling strategies.
- How to compensate for biases or how to reduce them.

Epidemic spreading in networks

Description:
- Introduction to the dynamics of epidemics: relevant questions and the full mixing assumption.
- Classic epidemic models (full mixing): the SI model (the logistic growth equation), the SIR model and the SIS model. Thresholding phenomena.
- Epidemic models over networks: homogeneous network models, scale-free network model for SIS and a general network model for SIS. Thresholding phenomena.

Percolation and network resilience

Description:
- Introduction to percolation: percolation in lattices, giant cluster.
- Network resilience upon node removal:
- Uniform node removal: the configuration model (thresholding phenomena), Erdös-Rényi networs, scale-free networks.
- Non-uniform node removal: random versus targeted attack, exponential versus scale-free network. Size of the giant cluster.

Other dynamic processes over networks: random walks and diffusion processes; search on networks

Description:
Other dynamic processes over networks: random walks and diffusion processes; search on networks

GRADING SYSTEM

There will be no exam in this course. Grading is done entirely through reports on various tasks throughout the course.

Students are expected to hand in 7 lab work reports about two weeks after its corresponding lab session, which count toward 50% of the final grade. We will also have a course project that the students will hand in towards the end of the course that accounts for 50% of the final grade.

The formula to compute the final grade is therefore:

$0.1 * (L1 + L2 + L3 + L4 + L5) + 0.5 * CP - (0.25 * P)$

where Li, for i=1..5, stands for the grade for the best 5 lab reports, CP stands for grade of the course project, and P is the nr of lab reports not handed in.

For each lab report not handed in, there is a penalty of 0.25 points towards the final grade. So, if a student does not hand in 2 reports, then 0.5 points will be deducted from the final grade. Grades are over 10.

BIBLIOGRAPHY

Basic:
- Newman, M.E.J. Networks. 2nd ed. Oxford University Press, 2018. ISBN 0198805098.
- Kolaczyk, E.D. Statistical analysis of network data: methods and models. Springer, 2009. ISBN 9780387881454.
- Trends in Cognitive Sciences.
- Barrat, A.; Barthélemy, M.; Vespignani, A. Dynamical processes on complex networks. Cambridge University Press, 2008. ISBN 9780521879507.

Complementary:
- Pastor, R.; Vespignani, A. Evolution and structure of the Internet: a statistical physics approach. Cambridge Univeristy Press, 2004. ISBN 0521826985.
- Caldarelli, G. Scale-free networks: complex webs in nature and technology. Oxford University Press, 2007. ISBN 9780199211517.

RESOURCES

Hyperlink:
- http://www.cs.upc.edu/~CSN/