Universitat Autònoma de Barcelona

2022/2023

Graph Analysis and Information Search

Code: 104363
ECTS Credits: 6

Degree	Type	Year	Semester
2503758 Data Engineering	OB	2	1

Contact

Name: Josep Lladós Canet
Email: josep.llados@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Other comments on languages

Many distributed materials are written in English.

Teachers

Josep Lladós Canet
Cristina Perez Sola
Jordi Herrera Joancomarti

Prerequisites

For a better follow-up of the subject, it is necessary to have passed the subject of second semester of first year "Graphs, topology and discrete geometry" which explains the basic theory of graphs, optimization of paths, algorithms on graphs and complexity of algorithms and problems. During the semester, the Data Structures course provides additional knowledge about the data structures for the storage and management of graphs.

Objectives and Contextualisation

Graph structures are very useful in representing data from their relationships. They are very useful in the representation and navigation of social networks, modeling of molecules, geographical representation for navigation software, biometrics (fingerprint recognition), knowledge representation in Artificial Intelligence, etc. In addition, graph theory offers a robust and mathematically grounded methodology for characterizing, traversing or analyzing this type of structure. This subject covers the following objectives: Acquire knowledge in data mining for graphs, dynamic graphs, dissemination and propagation of information in networks, internet search, analysis of social networks, communities and profiles of public users, recommendation systems.

Competences

- Analyse data efficiently for the development of smart systems with the capacity for autonomous learning and/or data mining.

- Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
- Prevent and solve problems, adapt to unforeseen situations and take decisions.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.

Learning Outcomes

1. Apply specific data mining techniques for graphs, carefully choosing the algorithms on the basis of the objective of the analysis, the volume of the data to be processed, and the available processing capacities.
2. Choose the algorithms that allow data collection in graph form on the basis of their impact on the characteristics of the data captured.
3. Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
4. Prevent and solve problems, adapt to unforeseen situations and take decisions.
5. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.

Content

Topic 1. Introduction. Definitions, concepts, representations.

Topic 2. Metrics in graphs. Algorithms of centrality and prestige (Degree centrality, Closeness centrality, betweenness centrality).

Topic 3. Generation of graphs. Random graphs. Synthesis of graphs.

Topic 4. Graph visualization. Gephi platform.

Topic 5. Search and routes. Graph traversal (Breadth First Search, Depth First Search), Shortest Path, A *, Minimum Spanning Tree,Random Walk.

Topic 6. Search the internet. Ranking and relevance for web models. PageRank Algorithm. Crawling.

Topic 7. Algorithms for the detection of influential communities and nodes. Graph clustering. Common subgroups. Reputation.

Topic 8. Propagation of labels and diffusion in graphs. Diffusion models. Epidemics. Belief propagation. Message passing.

Topic 9. Recognition based on graphs. Matching, kernels and embeddings.

Topic 10. Case study: graphs and social networks.

Methodology

The subject consists of 4 classroom hours per week (two sessions of 2 hours each). Classes will be held in a classroom with computers (or it will be recommended that the student has a laptop in the classroom). There is no distinction between schedules of theory, problems and laboratory practices in different hours and classroom, but will be alternating as appropriate to the same classroom. In general, a week will be devoted to each topic (in some subjects it will be two weeks). For each subject the theoretical concepts will be introduced, starting from materials that will be recommended to have looked previously, alternating with more applied activities (problem solving or seminars). Active participation in solving problems / exercises will be encouraged, working in the classroom, and encouraging the analytical presentation and discussion of the results. Some sessions, mainly the last ones of the course, will consist of seminars of more practical work, where a problem to solve of more general character will be raised, that requires the design and implementation of a solution

from the combination of the theoretical concepts. To facilitate follow-up, a detailed calendar of sessions will be distributed on the first day of the course.

Theory sessions. It consists of master classes with multimedia material available on the UAB Virtual Campus. The main objective of these classes is to introduce the basic notions that should allow the student to take a real vision of graph theory and its application in information search systems, in particular on the internet. To encourage more interactive sessions, and validate that knowledge is acquired, materials will be previously distributed through the virtual campus, recommending that they have looked at the session beforehand. During the theory session, practical activities will be introduced (resolution of exercises or more extensive case studies), and other follow-up activities (interactive questionnaires).

Problem sessions. During the course will alternate sessions of lectures with the resolution of exercises in front of the computer, with support tools such as the environment *NetworkX*. They will be exercises that will allow to monitor the understanding of the theoretical contents. A collection of exercises will be provided as a basis for work. Although it is not a strict division, of the two weekly sessions, one will be devoted more to theoretical content, and another to working problems of the corresponding topic. There may be evaluation of the delivery of some exercises (previously announcing which are the ones that the students must deliver).

Seminars / practical cases. Some sessions will be devoted to solving a larger and more general problem, which requires the combination of several tools seen in the corresponding topics. In the seminars, the ability to analyze and synthesize is fundamentally promoted, as well as the critical reasoning and decision-making of the student in order to solve the problem. There will be a practical case that corresponds to the last topic of the course on graphs and social networks. At mid-course there may be a case study plus course that covers the first topics.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Problems sessions	15	0.6	1, 3, 2, 4, 5
Seminars / practical cases	15	0.6	1, 3, 2, 4, 5
Theory lectures	30	1.2	1, 2, 5
Type: Supervised			
Problems/seminars preparation	15	0.6	1, 3, 2, 4, 5
Tutoring	15	0.6	1, 3, 2, 4, 5
Type: Autonomous			
Personal work	30	1.2	1, 3, 2, 4, 5
Study / exam preparation	22.5	0.9	1, 3, 2, 4, 5

Assessment

The evaluation of the subject (over 10 points) will consist of the following evaluative tests:

Partial exams (N1). Two partial exams will be done during the course (E1 and E2). The exams will contain questions of theoretical concepts as well as exercises of style that will have been done in class. The minimum mark to pass each exam is 5. In case that one of the two partials is not exceeded, a revision exam will be available at the end of the part that has not been passed. Corresponds to 50% of the final mark.

Tests based on exercises in problem classes (N2). Questionnaires will be solved, which can be on-line, and / or resolution of additional problems to those seen in class. It is part of the continuous evaluation, and therefore it is not recoverable. Corresponds to 10% of the final mark.

Delivery of a small project based on the work of the seminars. In the seminars a case study will be addressed in a tutorial way. The result of the work must be delivered at the end. This work will be done as a group. It is part of the continuous evaluation and has no recovery, but will be considered recovery or compensation mechanisms in certain cases. Corresponds to 40% of the final mark.

The final grade of the subject will be calculated as follows:

NF = 0.5 * N1 + 0.2 * N2 + 0.3 * N3

To pass the subject it is necessary to have achieved a minimum score of 5 in all marks. At the discretion of the teachers, however, it will be possible to establish compensations between marks N1, N2 and N3.

The subject will be evaluated as Not Evaluable only in the case that the student has not submitted any of the evaluation exams nor has totally or partially delivered the works.

In case of not passing the subject because any of the evaluation activities does not reach the minimum mark required, the numerical score of the student's record will be the lowest value between 4.5 and the weighted average of the marks. With the exceptions that students who do not participate in any of the evaluation activities will be awarded a "non-evaluable" mark, and that the numerical grade of the record will be the lowest value between 3.0 and the weighted average of the notes in case the student has committed irregularities in an evaluation act (and therefore the approved by compensation will not be possible).

will be awarded Matricula de Honor within the maximum allowed by UAB regulations (depending on the number of students enrolled) to higher grades equal to or greater than 9.

For each evaluation activity, a site, date and time of revision will be indicated. in which the student can review the activity with the teacher. In this context, claims may be made on the activity mark, which will be evaluated by the faculty responsible for the subject. If the student does not appear in this review, this activity will not be reviewed later.

See section "PLAGIARISM" on measures in cases of irregularities due to plagiarism in the evaluation activities.

REVISON:

Partial exams (N1). Two partial examinations of liberatory theory will be done during school hours. Students who do not pass this test (with a grade equal to or greater than 5), will have a recovery test on the final evaluation date scheduled for the degree.

Tests based on exercises and in practical case (N2 and N3). The work of exercises and practical evaluates in the form of continuous evaluation during the follow-up in class. Therefore there will be no recovery activity at the end of the course.

EVALUATION DATES:

The dates for evaluation and submission of papers will be published on the virtual campus and may be subject to programming changes for reasons of adaptation to possible incidents. Always be informed in the virtual campus about these changes as it is understood that it is the usual mechanism of exchange of information between teachers and students.

REPEATING STUDENTS:

Partial notes (theory or practices) are not kept from one course to another. However, at the discretion of the teacher and depending on the evaluations of previous courses, compensation may be established. This information will be announced on the day of the presentation of the subject, and the virtual campus.

PLAGIARISM:

Without prejudice to other disciplinary measures deemed appropriate, and in accordance with current academic regulations, irregularities committed by a student that may lead to a variation of the grade will be scored with a zero (0). The evaluation activities qualified in this way and by this procedure will not be recoverable. If it is necessary to pass any of these evaluation activities to pass the subject, this subject will be suspended directly, without the opportunity to recover it in the same course. These irregularities include, among others:

- the total or partial copy of a practice, report, or any other evaluation activity;
- let others copy;
- present a group work not done entirely by the members of the group;
- present as own materials prepared by a third party, even if they are translations or adaptations, and in general works with non original and exclusive elements of the student;
- have communication devices (such as mobile phones, smart watches, etc.) accessible during the theoretical-practical individual assessment tests (exams).

In summary: copying, copying or plagiarizing in any of the evaluation activities is equivalent to a NON PASS with a grade lower than 3.0

FINAL CLARIFICATION:

For any doubt or discrepancy, the most up-to-date information that will be communicated on the day of the presentation of the subject and that will be published in the virtual campus will prevail.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Evaluations based on the exercisis in the problems lectures	20%	1.5	0.06	1, 3, 2, 4, 5
Evaluations based on the practical cases (seminars)	30%	3	0.12	1, 3, 2, 4, 5
Two partial exams	50%	3	0.12	1, 2, 5

Bibliography

The contents of the subject explained in class are extracted from different sources. Many materials are online, extracted from multimedia materials, videos, etc. During the course, the necessary materials for the follow-up of the subject will be provided through the virtual campus. Links to online documentation, software for practical exercises, etc. will also be provided in open access format.
Online books of interest:

- Albert-Laszlo Barabási. Network Science. http://networksciencebook.com/
- David Easley and Jon Kleinberg. Networks, Crowds, and Markets. https://www.cs.cornell.edu/home/kleinber/networks-book/

Software

Open source software such as Gephi and the NetworkX library in Python are used to solve problems and case studies.Some class exercises will be distributed with the Google Colab platform.