

Mining Massive Datasets

1. **Course number and name:** 020MMDES4 Mining Massive Datasets
2. **Credits and contact hours:** 4 ECTS credits, 2x1:15 contact hours
3. **Name(s) of instructor(s) or course coordinator(s):** Dany Mezher
4. **Instructional materials:** Course handouts; slides; in-class problems, E-learning support using Moodle, MOOC on coursera.org

References:

- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey Ullman

5. Specific course information

a. Catalog description:

Introduction to massive data challenges, High Performance File System and MapReduce, Link Analysis in Graphs, Similar Sets, Similar Item sets, Community detection in graphs, mining data streams, recommender systems, Clustering and classifiers.

b. Prerequisites: None

c. Selected Elective for CCE students

6. Educational objectives for the course

a. Specific outcomes of instruction:

- Analyze big data problems and propose adequate solutions.
- Apply the appropriate techniques and skills to solve big data problems.

b. PI addressed by the course:

PI	1.1	1.2	1.3	6.3	6.4	7.1
Covered	x	x	x	x	x	x
Assessed	x	x	x	x	x	

7. Brief list of topics to be covered

- Introduction: Massive Data Challenges (1 lecture)
- HPFS and MapReduce (3 lectures)
- Lab MapReduce (1 lab session)
- PageRank and link analysis in graphs (2 lectures)
- Lab PageRank (1 lab session)
- Similar sets & similar item sets (3 lectures)
- Lab Similar sets (1 lab session)
- Lab Similar item sets (1 lab session)

- Community detection in graphs (Clustering and BigClam) (3 lectures)
- Lab Community detection (clustering) (1 lab session)
- Lab BigClam (1 lab session)
- Mining data streams (2 lectures)
- Lab Mining data streams (1 lab session)
- Recommender systems (3 lectures)
- Lab Recommender systems (1 lab session)
- Clustering (2 lab sessions)
- Lab Clustering (1 lab session)