

ANALYSIS OF LARGE SCALE SOCIAL NETWORKS

Introduction

Bart Thijs

The objective of this course is to provide you with knowledge and insight in social network analysis on large scale data.

It builds on SNA from Social Sciences and on Graph Theory from Mathematics.

After this course you should be able to design, implement and finalize a SNA project on large graph datasets within the requirements of the requested application.

COURSE OBJECTIVE

- ▶ 10 Lectures on Monday (Starting from February 15th)
 - ▶ First: February 14th
 - ▶ Last: May 16th
 - ▶ No lecture: April 4th, 11th & 18th
- ▶ 5 Exercises on Fridays
 - ▶ March 11th, 25th, April 1st, 29nd May 13th
- ▶ 1 Project

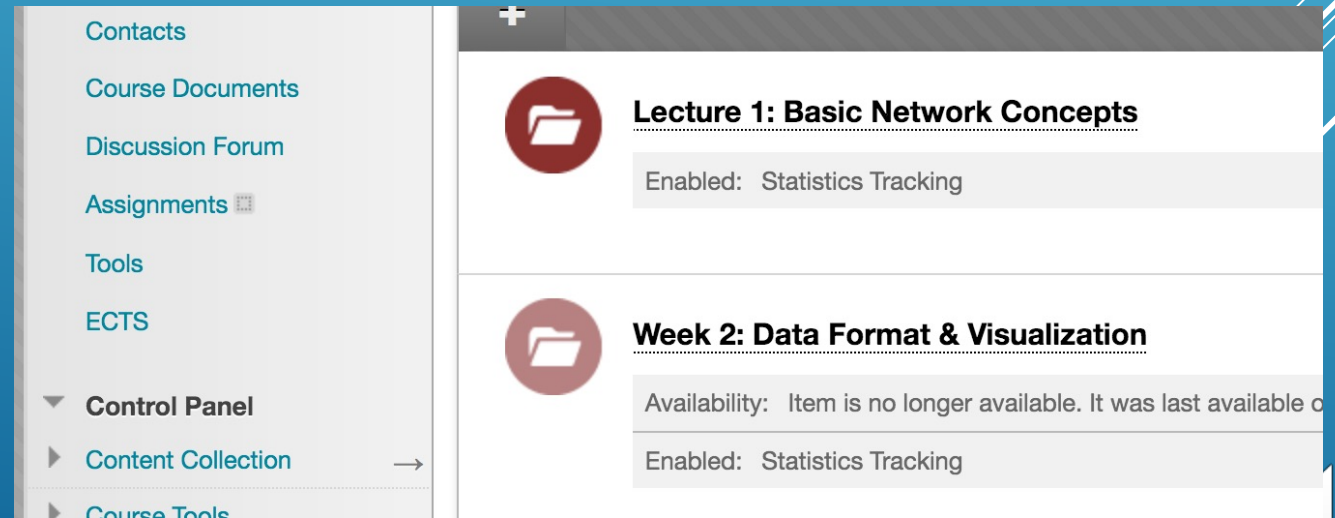
THREE LEARNING ACTIVITIES

Flipped Classroom Format

- ▶ Content (Text, Slides or Media) is given in the week prior to the lecture
- ▶ Material is to be read and processed at home. Small assignments can be given.
- ▶ During lectures we discuss the material using additional scientific publications in order to achieve a better understanding

Material can be found on Toledo

LECTURES



1. Basic Concepts & Centrality Measures
2. Data formats & Visualizations
3. Graph Traversals
4. Graph Databases vs RDBMS
5. Distributed Graph Processing: GraphX, Pregel
6. Random Graph Models & Scale Free Networks
7. Clustering & Partitioning 1: Modularity
8. Clustering & Partitioning 2: Other approaches
9. Graph Embeddings 1: Supervised Learning
10. Graph Embeddings 2: Unsupervised Learning

LECTURES

5 sessions: Online

Software: Python (iGraph, Stellargraph, Spark, GraphX)

1. Graph Processing & Visualizations(igraph)
2. Graph Databases (Neo4J)
3. Distributed Graph Processing (Spark; GraphX)
4. Clustering, Partitioning (igraph)
5. Embeddings (stellar graph)

EXERCISES

Topics:

1. Analytics Project: Community Detection, Graph Embedding
2. Graph Database (Neo4J)
3. Graph Programming (Python, Spark – GraphX, StellarGraph)

Team work: groups of 3 – 4 students

Grading: 25% of final grades
based on final report

PROJECT

2 Sections

1. Application section = open ended questions about a scientific publication covering topics discussed in the lectures (45min)
2. Exercise section = processing and analysing a small network combined with multiple choice questions on t

3 hours in a computer class

Grading: 65% of final grades

EXAM

1. Participation in the online discussion prior, during and after the lecture
2. Participation in class room discussions of scientific papers

Grading: 10% of final grades

PARTICIPATION