ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Practical Informations

Teachers
Pierre Vandergheynst
Pascal Frossard

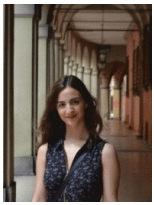
Assistants
Michaël Defferrard
Effrosyni Simou
Hermina Petric Maretić
Rodrigo Pena
Eda Bayram
Benjamin Ricaud

Team

Pierre
Vandergheynst

Pascal
Frossard

Michaël
Defferrard



Effrosyni
Simou

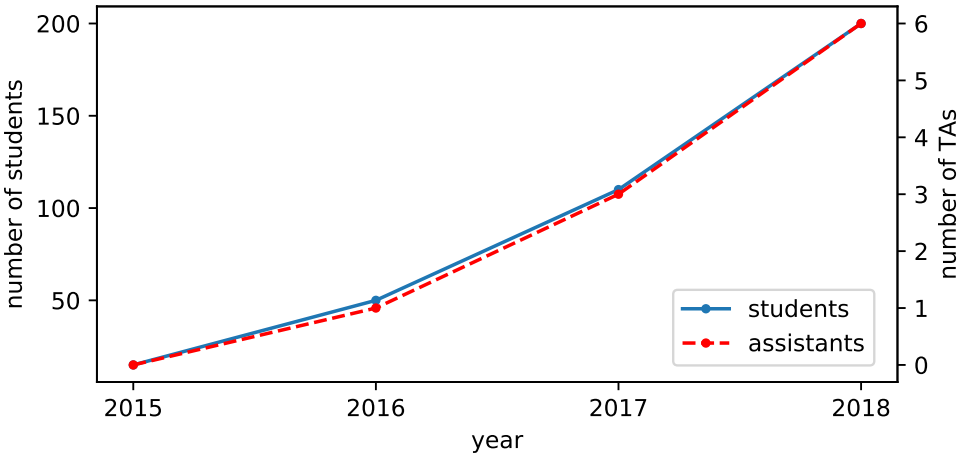
Hermina
Petric Maretić

Rodrigo
Pena

Eda
Bayram

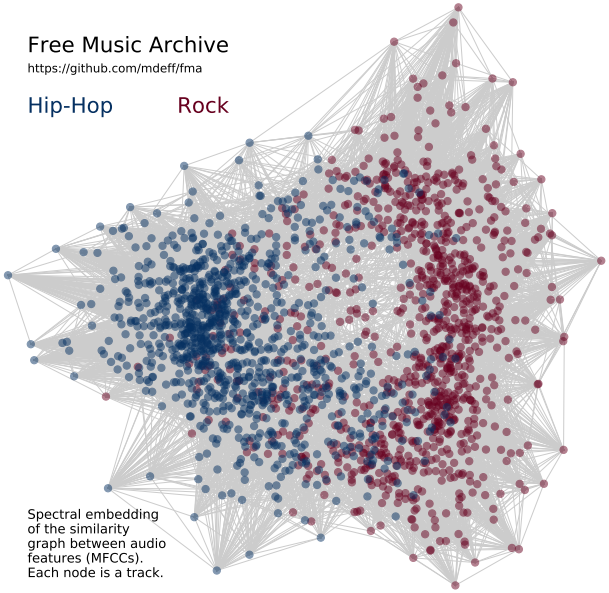
Benjamin
Ricaud

Enrollment


number of students
200
150
100
50
number of TAs
6
5
4
3
2
1
0
2015
2016
2017
2018
year
students
assistants

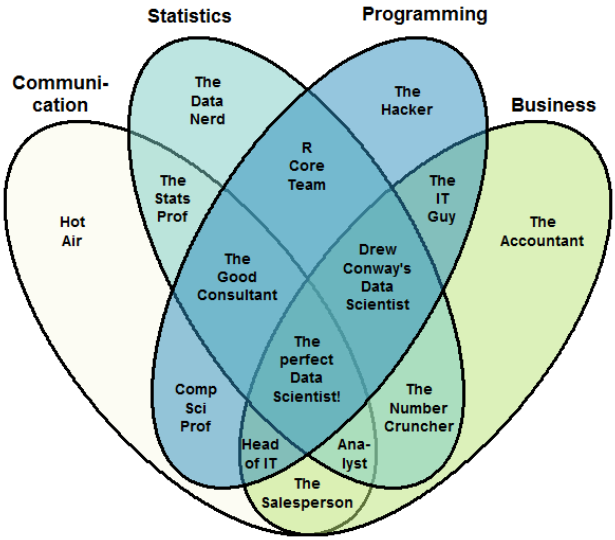
Content: **A Network Tour** of Data Science

1. Network Science
2. Spectral Graph Theory
3. Graph Signal Processing
4. Machine Learning



Spectral embedding of the similarity graph between audio features (MFCCs). Each node is a track.

Content: A Network Tour **of Data Science**



Evaluation

Joint evaluation of theoretical and practical skills through a project.

Two parts:

1. Guided with four milestones that follow the lectures.
2. Open ended extra work (e.g., changing the graph, adding features, exploring other algorithms, gathering complementary data).

Grading:

- 50% for acquiring the course material in a structured way.
- 50% for being creative and able to understand data, i.e., Data Science.

Project flow

1. Form teams of four and choose a project from a list of ten. Register on Moodle.
2. Complete and handle (on Moodle) the four milestones (Jupyter notebooks).
3. Work on the open ended extension.
4. Write and handle (on Moodle) a coherent 5 pages PDF report that tells a data story.
5. Handle all the code produced for the project as a git repository. On GitHub, with a proper readme, license, etc.
6. Impress us in a presentation! Presentation of 15 minutes in front of the class.

Milestones

1. Template notebook with instructions given on GitHub.
2. Around two weeks to complete.
3. At least one lab session to ask questions.
4. Completed notebook to be handled on Moodle.
5. Solutions posted on GitHub.
6. Grades given on Moodle.

Four topics that follow the lectures, with a Data Science taint:

1. Network properties
2. Network models
3. Spectral Graph Theory
4. Graph Signal Processing

Deadlines (tentative)

Oct 2	form groups of four and choose a project
Oct 23	handle milestone 1 (data loading, network properties)
Nov 12	handle milestone 2 (network models)
Nov 26	handle milestone 3 (spectral graph theory)
Dec 11	handle milestone 4 (graph signal processing)
Jan 11	handle project report and GitHub repository
Jan 22	project presentations

Practical sessions

Apply the material learned in class in a Data Science context.

During the labs, we will:

- Demo tools, e.g., how to manipulate a graph in Python.
- Explain the milestones and give directions.
- Answer questions about the milestones and project.

We expect you to:

- Bring your laptop.
- Work outside the hours on the milestones and project.

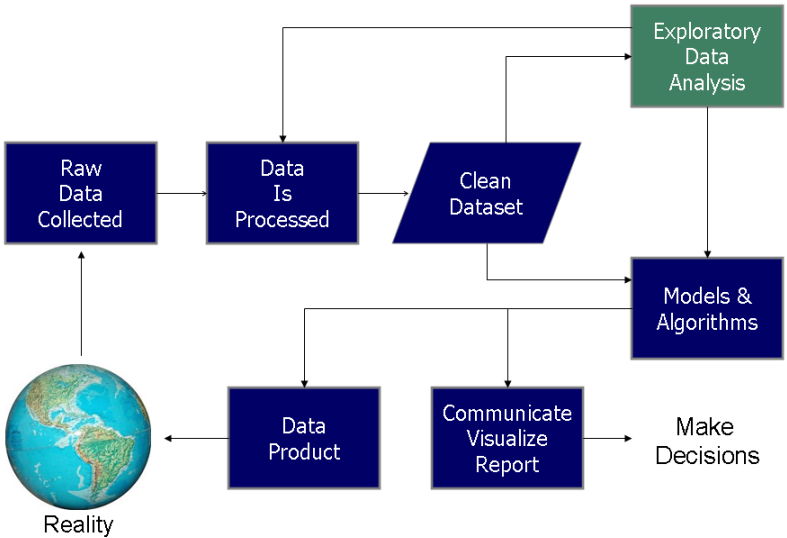
Tools

Python scientific stack & git

To be installed with conda:

- Python: programming language
- Jupyter: interactive computing
- NumPy: *n*-dimensional arrays
- SciPy: scientific computing
- matplotlib: visualization
- pandas: data analysis
- NetworkX: network science
- graph-tool: network science
- PyGSP: graph signal processing

Data science process



Data science process

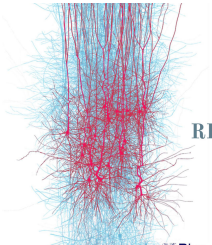
1. Data acquisition: from the web, a database, a flat file, etc. This includes cleaning the data.

2. Data exploration: some exploratory analysis to describe properties of the data and understand the content.

3. Data exploitation: use the data to solve a task, to infer knowledge, to draw conclusions. The concepts or algorithms taught in class must be used.

4. Conclusion: discuss the results and summarize your findings. What did we learn from the data and the project?

Example projects from 2017



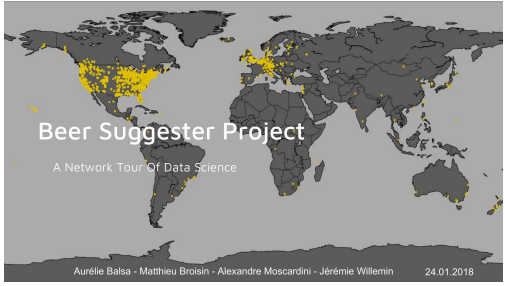
GSP ON THE DIGITAL RECONSTRUCTION OF THE BRAIN

Stefania Ebli - Christopher Elin - Florian Roth

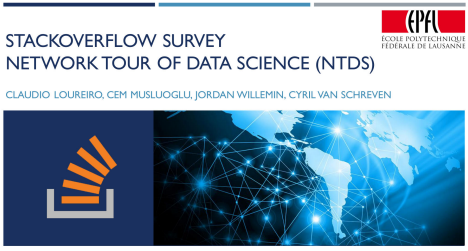
NTDS Project 2017/18

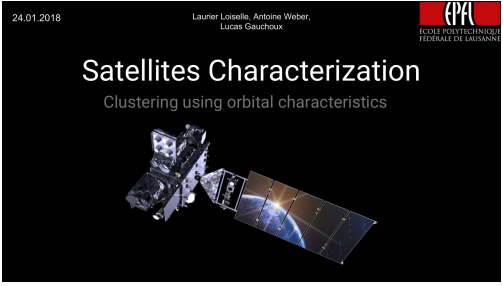
Blue Brain Project

1



Example projects from 2017





Proposed projects

- Free Music Archive
- US Senators
- Wikipedia
- Researchers on Twitter
- Scientific co-Authorship
- Spammers on Social Networks
- Citation Network
- Terrorist Attacks and Relations
- IMDb Films and Crews
- Flight Routes

Rules

- Form groups of 4 students. No less, no more.
  - One member of the group uploads the deliverables.
  - The names of all members should appear clearly.
- Projects should use tools and ideas from the lectures. While the second part is quite free, it should include graph and network data aspects, and more generally fall under the scope of the class.
- The project should follow the data acquisition, exploration and exploitation workflow.
- Each member of a team shall contribute equally to the project.

Online

Moodle: `https://moodle.epfl.ch/course/view.php?id=15299`

- slides
- grades
- official announcements
- discussion forum

GitHub: `https://github.com/mdeff/ntds_2018`

- installation instructions
- tutorials
- milestones
- projects

Questions?