

A NETWORK TOUR OF DATA SCIENCE

LAB SESSIONS

Teachers

Pierre VANDERGHEYNST
Pascal FROSSARD

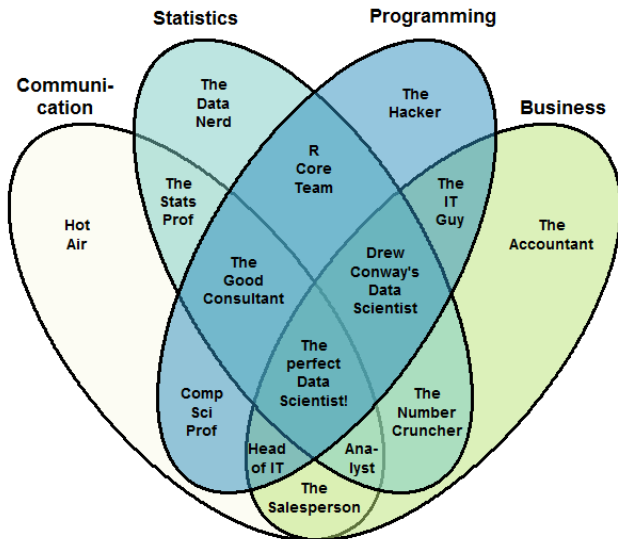
Assistants

Michaël DEFFERRARD
Effrosyni SIMOU
Hermina PETRIC MARETIĆ

EPFL LTS2 & LTS4 laboratories

September 22, 2017

Data Scientist



Goal

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.
- ▶ Demo techniques, e.g. how to collect data from Twitter.
- ▶ Explain the assignments and give directions.
- ▶ Answer questions about the assignments and project.

We expect you to:

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

Schedule (tentative)

Sep 29 Data Science in Python

Oct 2 Lab 1 – Network properties

Oct 13 Lab 1 – Network properties

Oct 23 Lab 2 – Network models

Oct 30 Lab 2 – Network models

Nov 6 Lab 3 – Spectral graph theory

Nov 13 Lab 3 – Spectral graph theory

Nov 24 Project discussion

Nov 27 Lab 4 – Graph signal processing

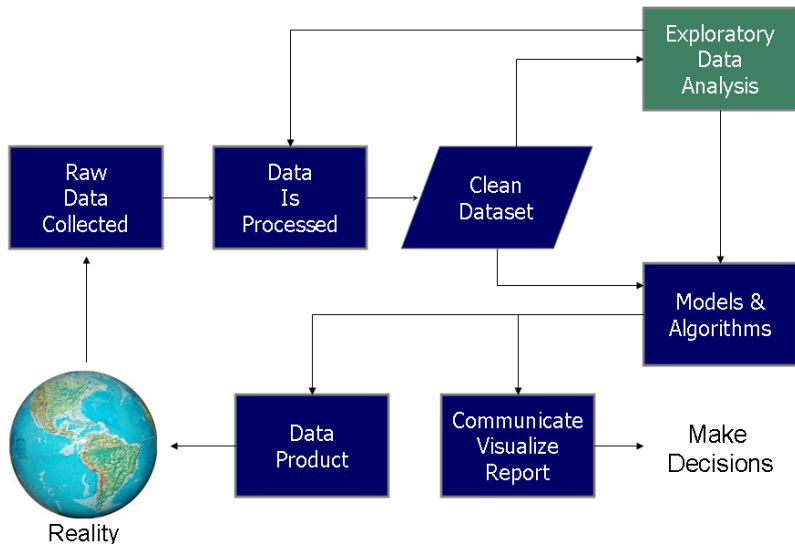
Dec 4 Lab 4 – Graph signal processing

Dec 11 Lab 5 – Machine learning

Dec 18 Lab 5 – Machine learning

Dec 22 Project discussion

Data Science Process



Python scientific stack + git

To be installed with Anaconda:

- ▶ Python: programming language
- ▶ IPython: 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

Grading

1. 50% assignments (reports)
→ the continuous evaluation.
2. 50% project (report & presentation)
→ the final exam!

Assignments

1. Template notebook with instructions given on Github.
2. 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.

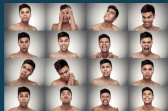
Topics from lectures, with a Data Science taint:

1. Network properties
2. Network models
3. Spectral Graph Theory
4. Graph Signal Processing
5. (Machine Learning)

Project

1. **Proposal:** define a problem you are interested in.
 - ▶ Single page document. Organize yourselves in groups.
 - ▶ Deadline in November (to be confirmed). Upload on Moodle.
 - ▶ Not graded. Discussion with TAs will follow.
2. **Report:** your solution, using the theory seen in class and the practical skills trained during labs.
 - ▶ Jupyter notebook with text, math, code, analyzes and results.
 - ▶ Deadline in January (to be confirmed). Upload on Moodle.
 - ▶ Graded.
3. **Presentation:** impress us!
 - ▶ Presentation of 15 minutes in front of the class.
 - ▶ Held in January (to be confirmed).
 - ▶ Graded.

PROJECT NTDS Face Emotion Recognition



Patryk OLENIUK
Carmen GALOTTA



How Do Fake-News Go Viral?

Or why Bernie Sanders could replace
Trump with little-known loophole.



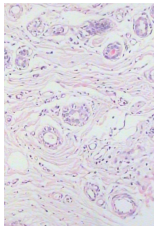
EE-559: A Network Tour of Data Science

William Trouleau & Victor Kristof



Epileptic Seizures Prediction

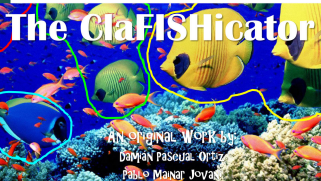
Sophie du Bois



BREAST CANCER CLASSIFICATION

based on histopathological images

by:
Robin Demesmaeker
MA3, EE



Open Source Software Support

A Network Tour of Data Science

Matthaios Olma
Pavlos Nikolopoulos
Stefanos Skalis

Predicting an Election from Tweets

Michaël Juillard, Mikhail Vorobiev, Chiara Ercolani

A Network Tour of Data Science

18th January 2017



Data Acquisition Pipeline



For all posts in 2016:

- Post id
- Time of creation
- Type of post
- Text message of the post
- Number of shares
- Number of comments
- Number of reactions

Add BuzzFeed rating:

- Mostly true
- Mixture of true and false
- Mostly False
- No factual content

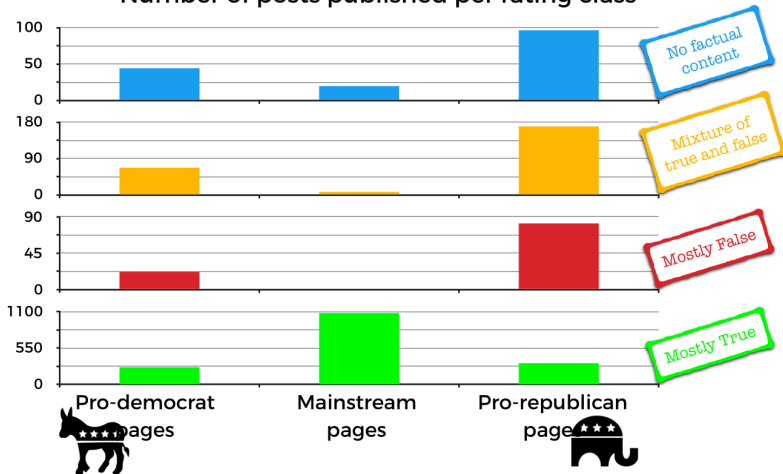
Add political orientation category:

- Pro-democrat page
- Pro-republican page
- Mainstream page

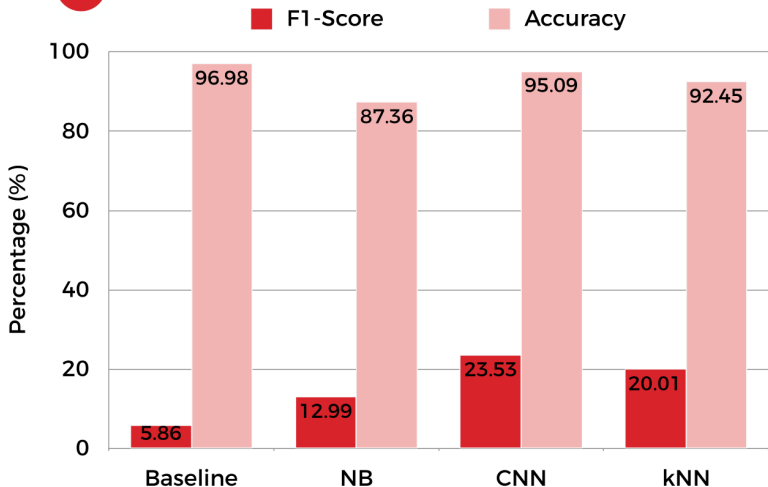


How much can we trust these pages?

Number of posts published per rating class



Evaluation: k-Nearest Neighbors



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

- ▶ Slides.
- ▶ Grades.
- ▶ Official announcements.
- ▶ Discussion forum.

GitHub: https://github.com/mdeff/ntds_2017

- ▶ Installation instructions.
- ▶ Demos.
- ▶ Assignments.
- ▶ Projects.

Questions?