A Network Tour of Data Science

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Projects

Teachers
Pierre Vandergheynst
Pascal Frossard

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

EPFL LTS2 & LTS4 laboratories

December 5, 2018

Goal

The milestones gave you

- an understanding of your data,
- practical experience with some tools.

Use that knowledge to

tell us a data story.

Process

1. Define a problem.
2. Use your dataset to solve it.
3. Use the tools seen during the course.
4. Present your conclusions.

Beware:

- Set a goal that is attainable given your understanding of the data and tools.
- It's not a presentation about what you did during the milestones!
- Projects should use tools and ideas from the lectures. They should include graph and network data aspects, and more generally fall under the scope of the class.
- Building a good graph is as important as analyzing it. Your graph should be built towards the goal of solving your problem.[1] That includes sub-sampling.

[1] We didn't care much for the milestones as there was no real problem to solve.

Two directions

- Use the data to answer a research question you are curious about.

- Use the data to build a “product”, e.g., predict some variables given others.

Examples

IMDb How do actors choose a film to play with? Do they form communities? What are the characteristics of those communities?

FMA Can an online music platform recommend annotations (such as tags or genres) to artists regarding their newly uploaded songs?

Senators Are senators truly divided in republicans and democrats? Can I predict the votes of a senator knowing the votes of "similar" senators?

Good examples from last year:[2]

- A Network Tour of StackOverflow
- GSP on the Digital Reconstruction of the Brain
- Graph-based Recommendation for lastFM
- Graph-based Nutrition Guide
- Buda + Pest = Budapest
- GraphLang

[2] available at `https://github.com/mdeff/ntds_2017`

Deliverables

report Describe your motivations, explain what you did and why (e.g., how you built the graph and why), and state your conclusions. The report contains an exploration part (e.g., some important or interesting facts about your data and graph), and an exploitation part (how you used the data to solve your problem). The report is a 5 pages PDF.

code All the code you developed for the project must be stored in a GitHub repository. It should contain a useful readme and a license. Code should be organized and clean.

presentation Impress us! Presentations are 12 minutes long, followed by 3 minutes of questions. Each group member must talk.

Deadlines

Jan 18 upload the project report on Moodle

Jan 18 send a link to your GitHub

Jan 22-23 give an oral presentation

Jan 25 upload the presentation slides on Moodle

Have fun! Questions?