Introduction to data science and network science

Artificial Intelligence, urbanistics and data science
ITMO, 2024/2025

Liubov Tupikina liubov.tupikina@cri-paris.org
ITMO, Russia, LPI, Bell labs France

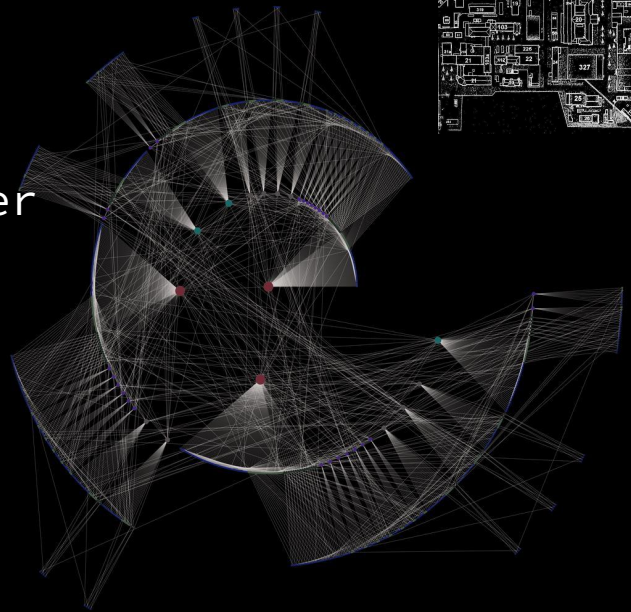
Contacts

Affiliate researcher at LPI,
Researcher, Nokia Bell Labs, France

Main topics:

Explainable AI (geometry of embeddings, analysis of higher-order data)

Geometry of graphs and hypergraphs (processes on graphs, transport applications)



Students contact:
Zahara Farook, Hritika Kathuria

LPI and embodied network science

Outline and connection to other courses

1. Reversed classrooms
2. Project-based learning
3. Teaching and learning by doing (fablab access, art itmo projects, other external collaborations, e.g. lpi cri-paris.org collaboration)
4. Open science (github, X-posts on the results)
5. Open data policy if possible and needed
6. Course co-desing (office hours by requirements), collective learning and self-learning (other platforms)
7. Sustainability oriented courses
8. Mathematics and other deep courses (Skoltech, Spbgu)
9. Network seminar subscription (youtube links)

Codesing of the course

Orientation every class

Outline of Introduction to data science

Outline of Introduction to data science

What are your main projects?

Outline of Introduction to network science

1. Introduction to networks
2. Practical part: notebooks

Assignment for the next course

Resources and libraries for the course

Standard libraries (Python): numpy, matplotlib, scikit learn, seaborn
Network libraries: networkx, osmnx (open street data)

Resources and libraries for the course

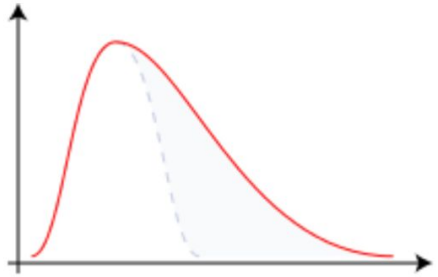
Standard libraries (Python): numpy, matplotlib, scikit learn, seaborn
Network libraries: networkx, osmnx (open street data)

Support materials

- Big data course Marc and Liubov https://github.com/Big-data-course-CRI/
- Correlaid, Complex system conference CSS 2023 and TidyTuesday https://github.com/rfordatascience/tidytuesday
- Network science book http://networksciencebook.com/
- Network repository https://networks.skewed.de/
- Visualisation tools https://gephi.org/users/download/
- Network datasets https://www.complex-networks.net/datasets.html#chap8

Statistics course

Two types of distributions we will work with very often



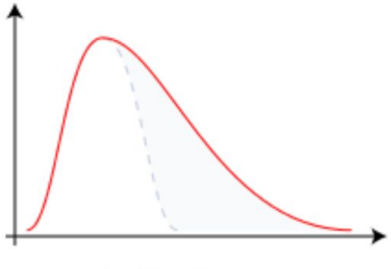
Positive Skew

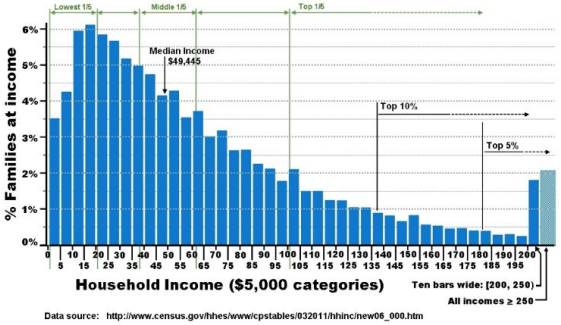
Right skew

Long tail to the right

Statistics course

Types of distributions we will work with very often
(following the recommendations from Michael Szell on his introduction to single variable analysis)
https://github.com/mszell/introdatasci/tree/main/unit12_singlevariableanalysis
Follow the notebook





Central limit theorem and its intuition