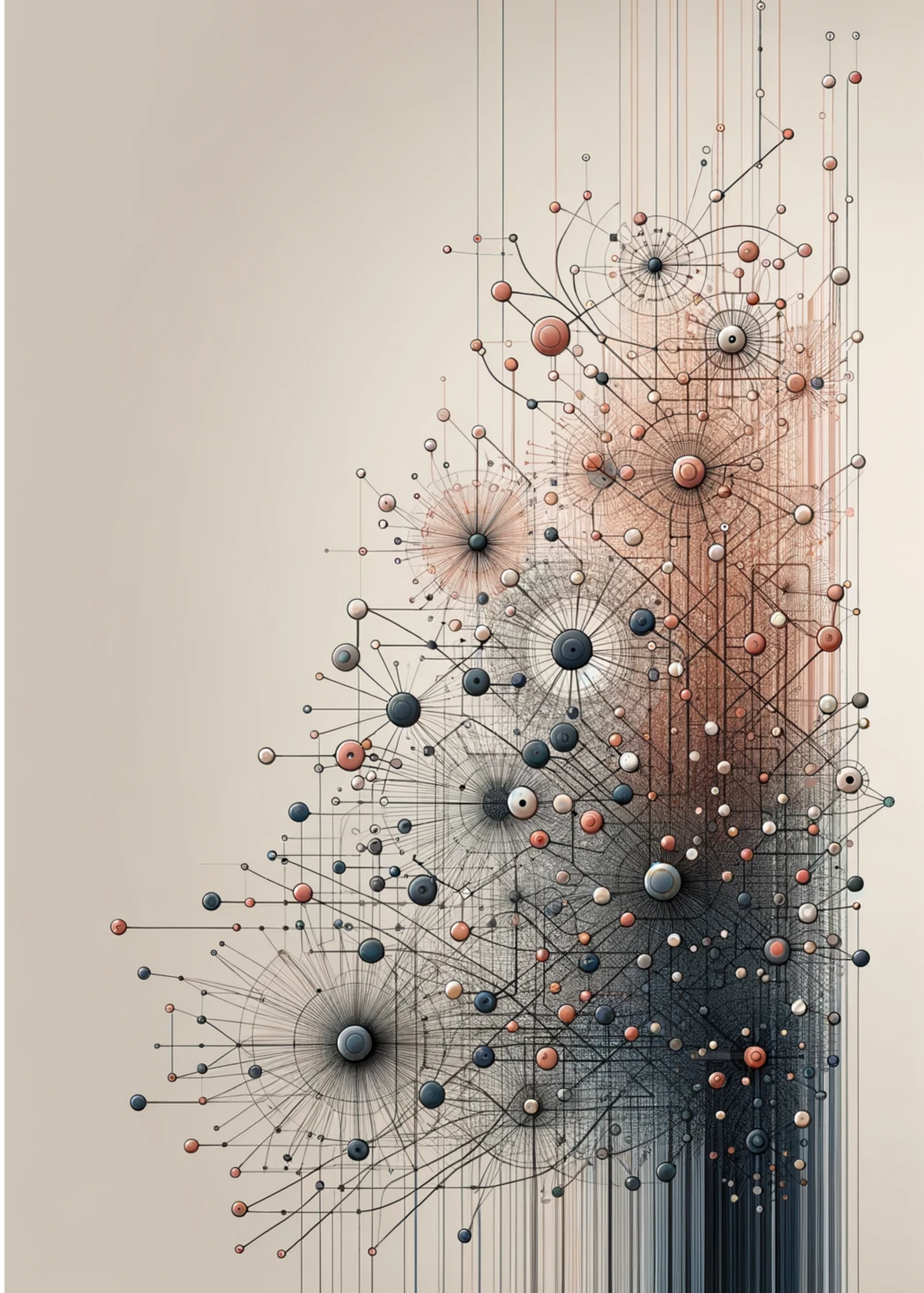
Analisi e Visualizzazione di Reti Complesse

NS01 - Introduction to the course

Prof. Rossano Schifanella



UNIVERSITÀ DI TORINO

Instructor

Prof. Rossano Schifanella

- Associate Professor at Computer Science Department at UNITO
- Researcher at ISI Foundation

Contacts:

- Email: rossano.schifanella@unito.it
- Homepage: http://www.di.unito.it/~schifane
- Office: 032_D_P03_0030 (3rd floor)

UNIVERSITÀ DI TORINO

Course

- **INF0007 Analisi e Visualizzazione delle Reti Complesse** (9 credits)
    - Magistrale in Informatica
- Borrowed by:
    - **MFN0954 Reti Complesse** (6 credits)
        - Magistrale in Informatica
    - **FIS0127 Analisi e Visualizzazione delle Reti Complesse** (9 credits)
        - Magistrale in Fisica dei Sistemi Complessi
- The course is divided into two sub-modules:
    - **Network Science** (NS) (48h)
    - **Data Vizualization** (DV) (24h)

Schedule

	Tuesday	Wednesday	Friday
Module	**NetSci**	DataViz	**NetSci**
When	**4pm-6pm**	11am-1pm	**2pm-4pm**
Where	**Aula F**	Aula E	**Aula E**

The schedule could be subject to variations that will be communicated via the proper channels (e.g., Moodle forum, email).

This is why it is **crucial** that all students are registered to the Moodle module to be up-to-date.

Material

- Course on Moodle:
  - https://informatica.i-learn.unito.it/course/view.php?id=2870
- GitHub code repo:
  - https://github.com/rschifan/avrc-2324
- External references listed in Moodle
- The material will be updated throughout the course, so make sure to **download the latest version** when available.

Textbooks

> Filippo **Menczer**, Santo **Fortunato**, Clayton A. **Davis**, *A First Course in Network Science*, Cambridge University Press, 2020

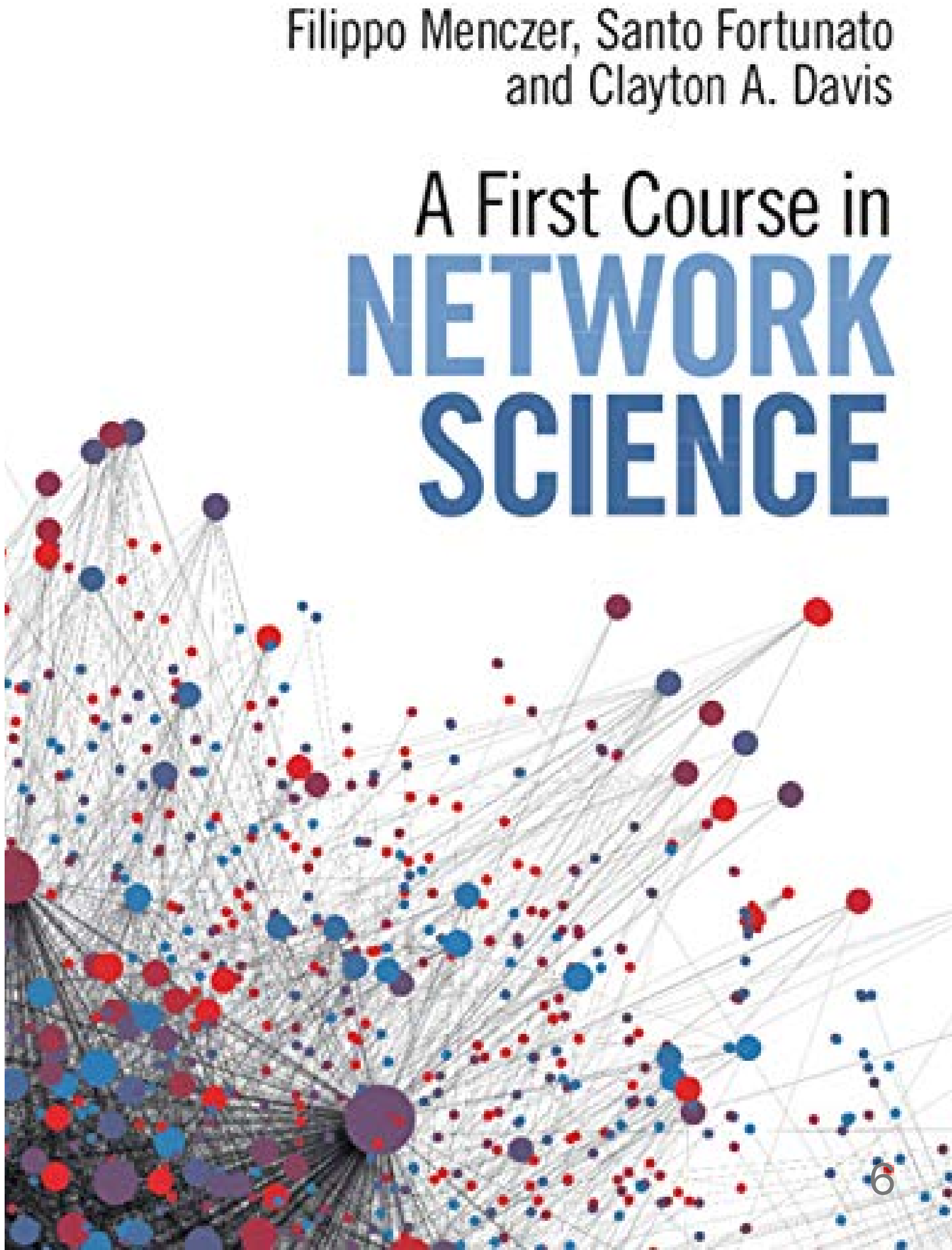
Fresh new textbook!

"A First course": very **basic** , but wide and updated overview

Few mathematical formalisms, some **illustrative code** (in Python)

Some copies are available in the computer science library.

We will cover **every chapter**



> David **Easley** and Jon **Kleinberg** , *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* , Cambridge University Press, 2010

Very **multidisciplinary** !

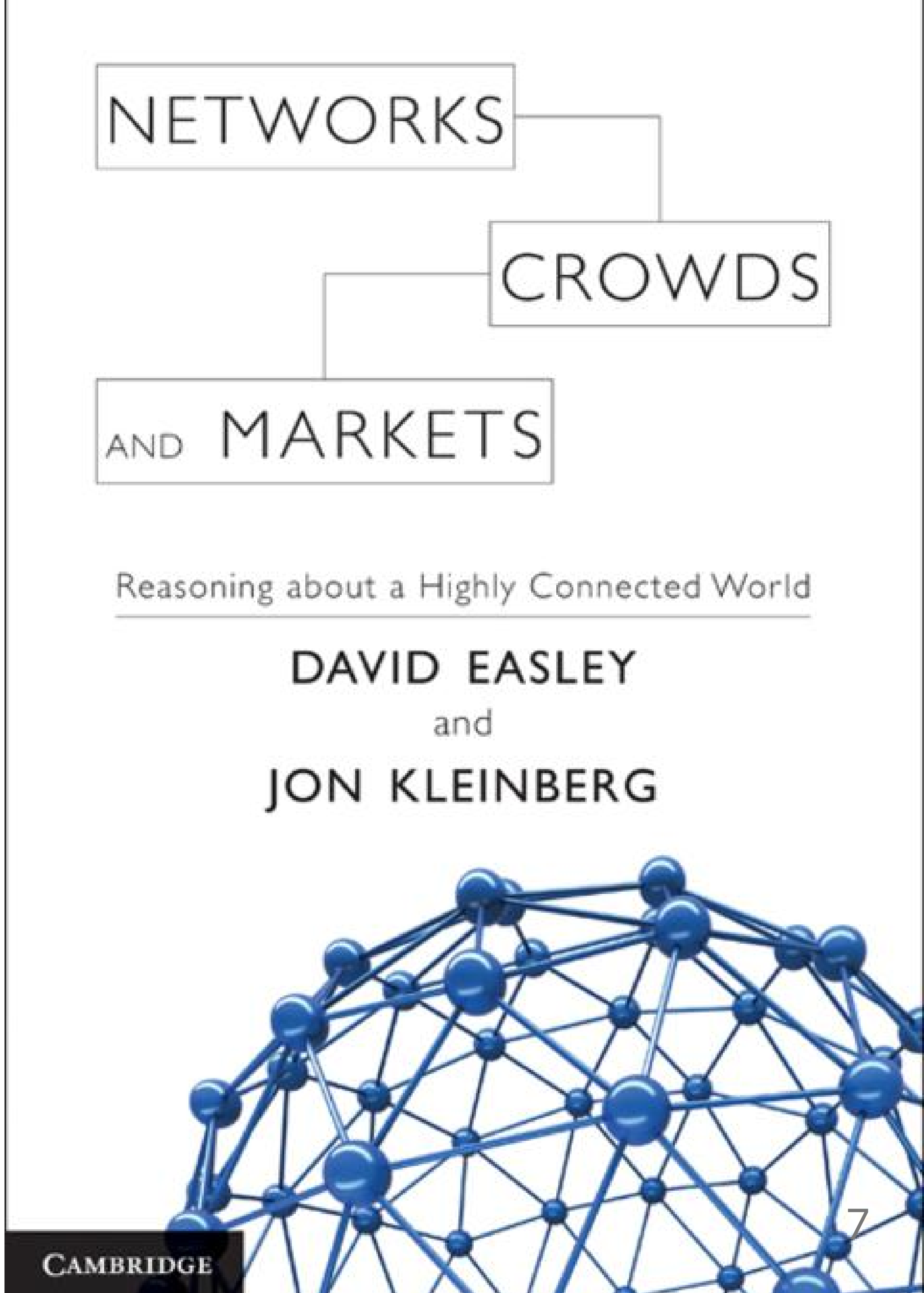
Richer overview (much beyond 'networks'); updated up to 2010

Few mathematical formalisms, **a lot of examples in many different applicative domains**, no illustrative code

Available in our libraries but also online (pre-print free!)

- https://www.cs.cornell.edu/home/kleinber/networks-book/

We will cover **many chapters**



UNIVERSITÀ DI TORINO

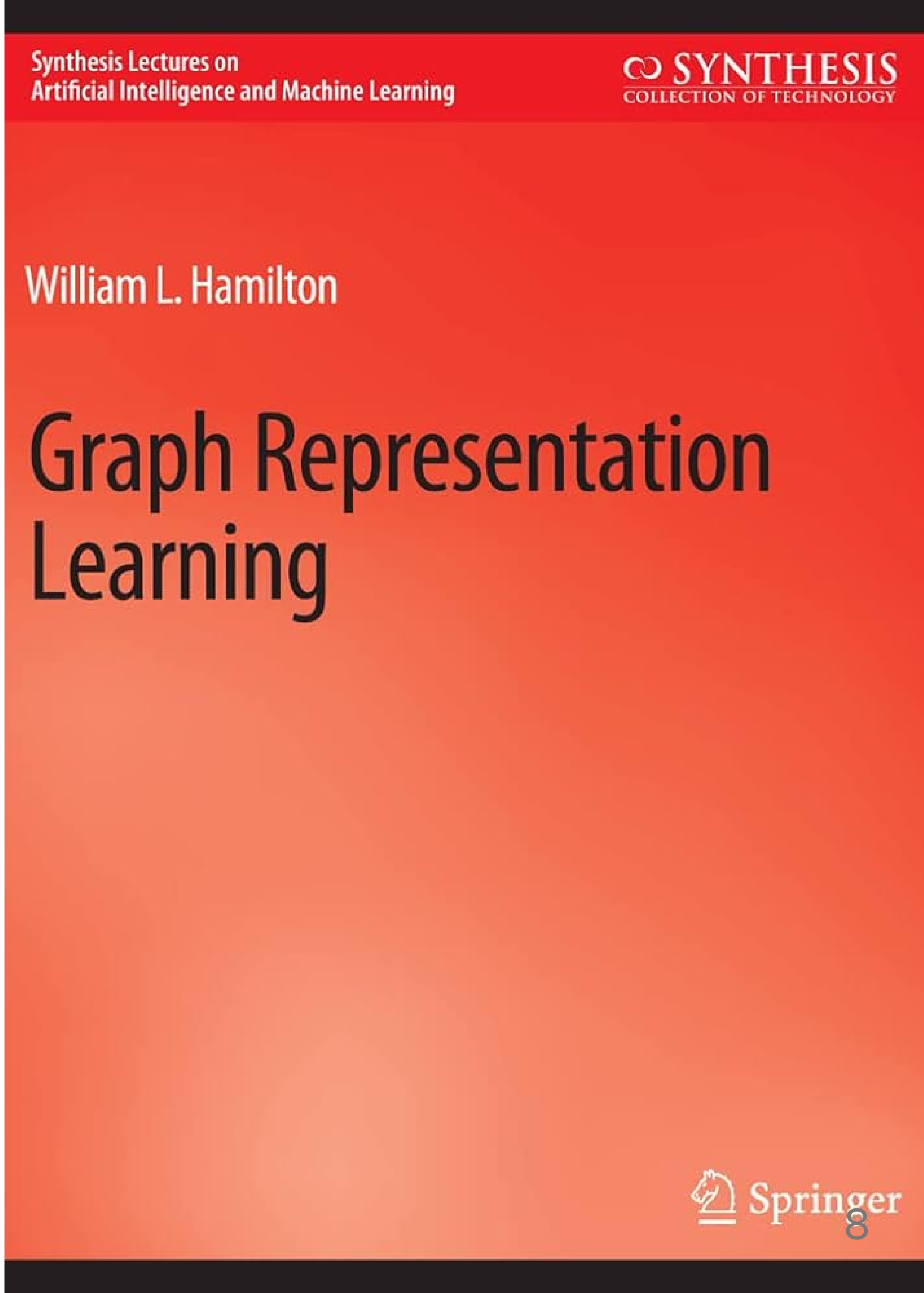
> Hamilton, William L.. Graph Representation Learning. Switzerland: Springer International Publishing, 2022.

Updated to 2020 (the research area is changing very rapidly and constantly)

We will cover **a few chapters**

There are several equivalent books and online resources that you can adopt instead

- https://www.cs.mcgill.ca/~wlh/grl_book/



Textbooks - Optional

> Albert-László **Barabási**, *Network Science*,
> Cambridge University Press

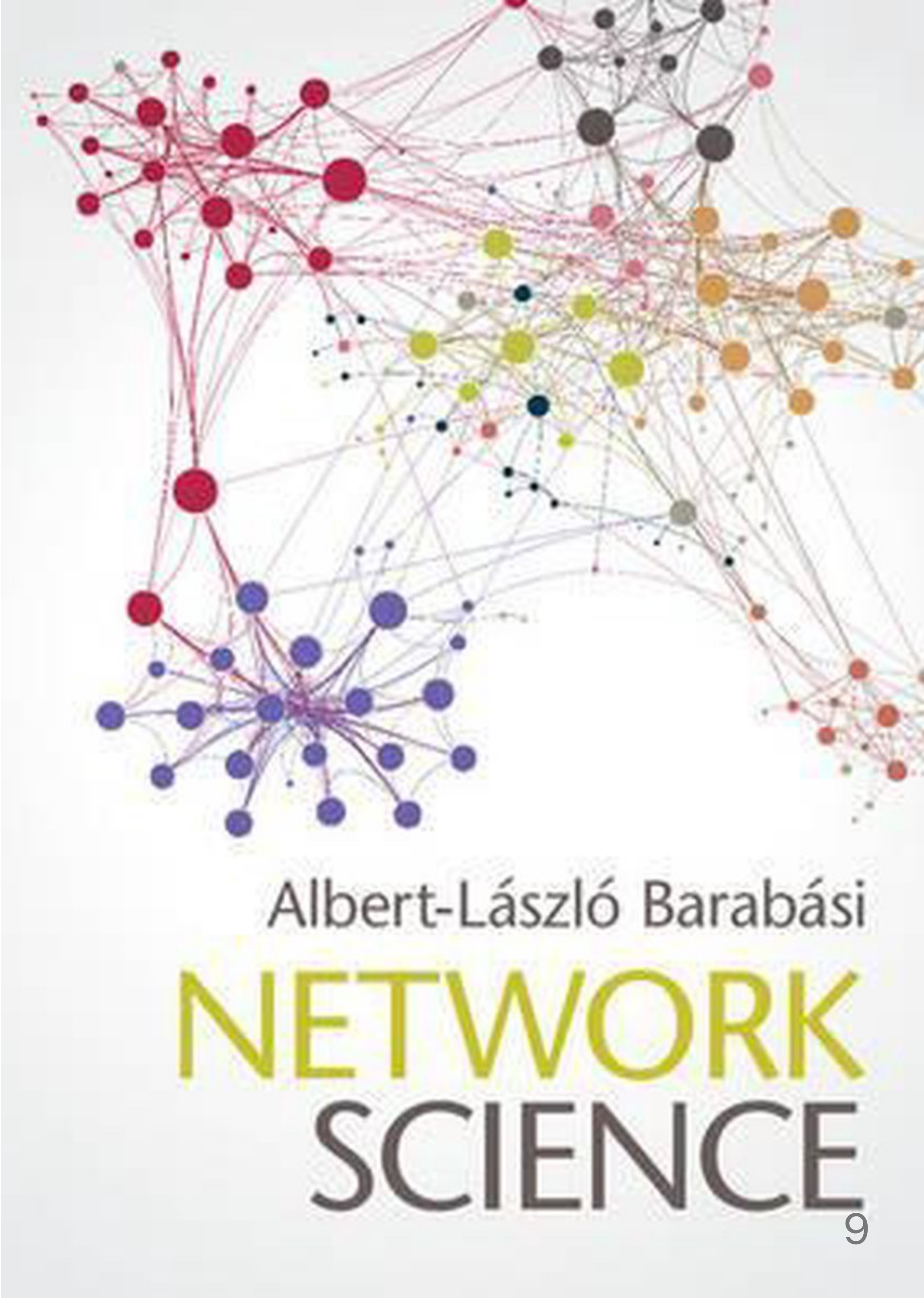
Read it if you are interested in **models** more than in data science.

Plenty of mathematical formalisms (even if simplified), many plots, some applicative examples, no illustrative code

Available in our libraries but also on-line (free!)

- http://networksciencebook.com

This is mainly **optional,** even if we may adopt some parts.



UNIVERSITÀ DI TORINO

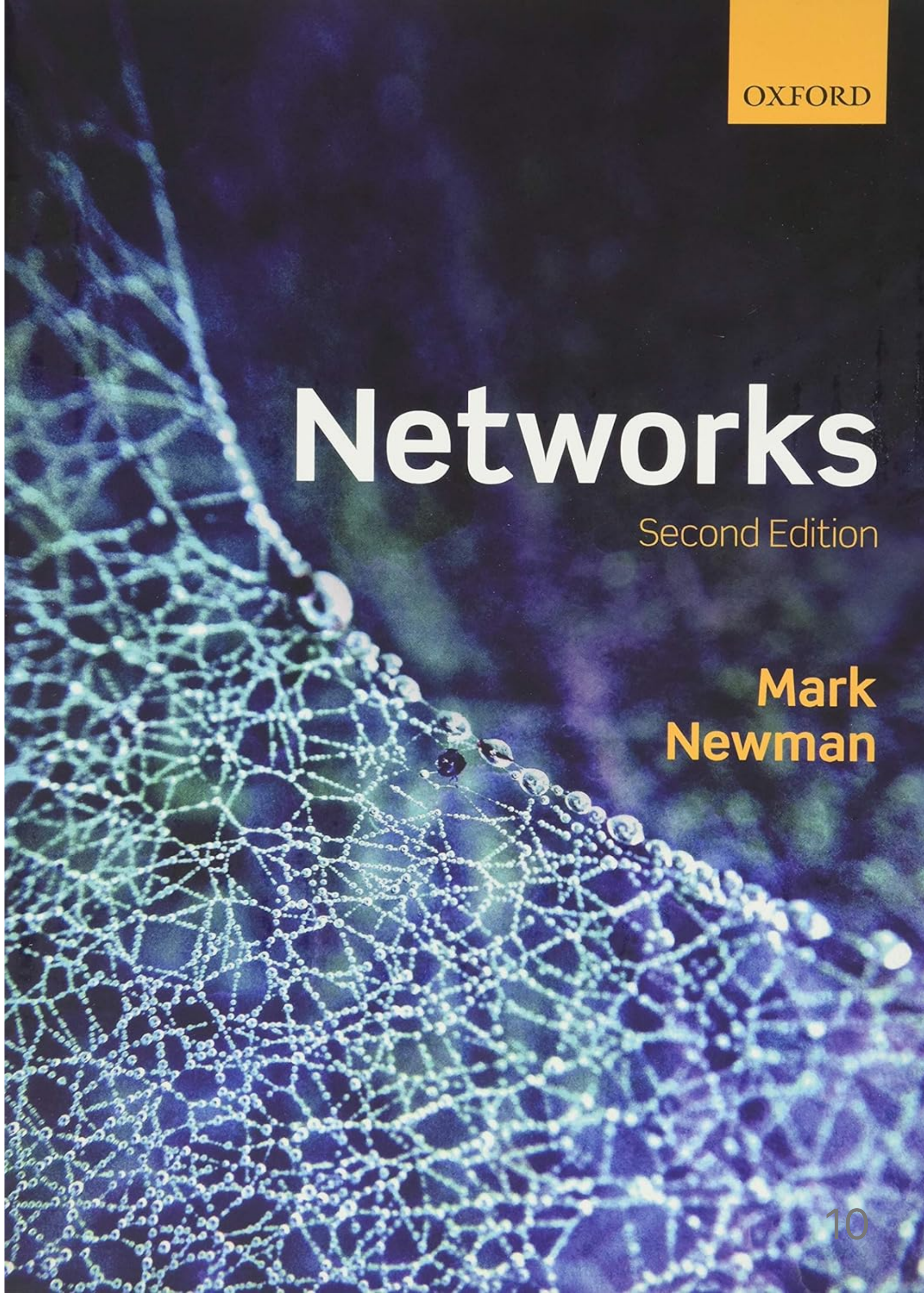
> Mark **Newman**, *Networks*, Oxford University Press

The **Bible** for network scientists: it is a must if you undergo graduate studies or have a master's degree thesis on such topics

Read it if interested to: **models**, deep understanding of **measures**, and **algorithms**

Very elegant mathematical formalisms, many examples, no illustrative code

Available in our libraries



UNIVERSITÀ DI TORINO

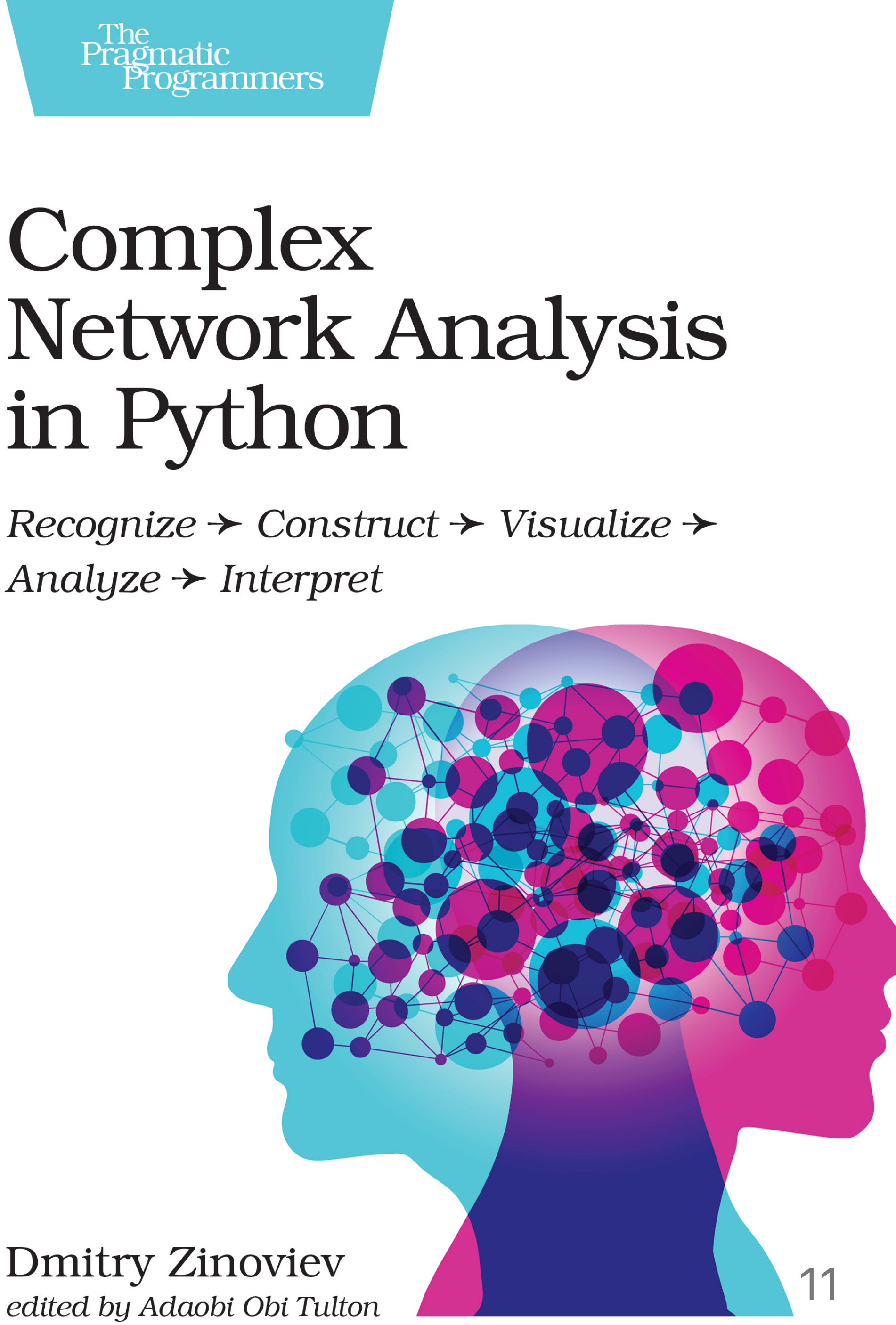
> Dmitry **Zinoviev** , *Complex Network Analysis in Python, Recognize → Construct → Visualize → Analyze → Interpret* , The Pragmatic Bookshelf

Read it if interested to: **practical data science**, **network analysis**;

Very **practical:** not much space for ideas, concepts, and limitations behind network measures

Many illustrative examples, A LOT of illustrative code (Python) with many different datasets

Available in our libraries



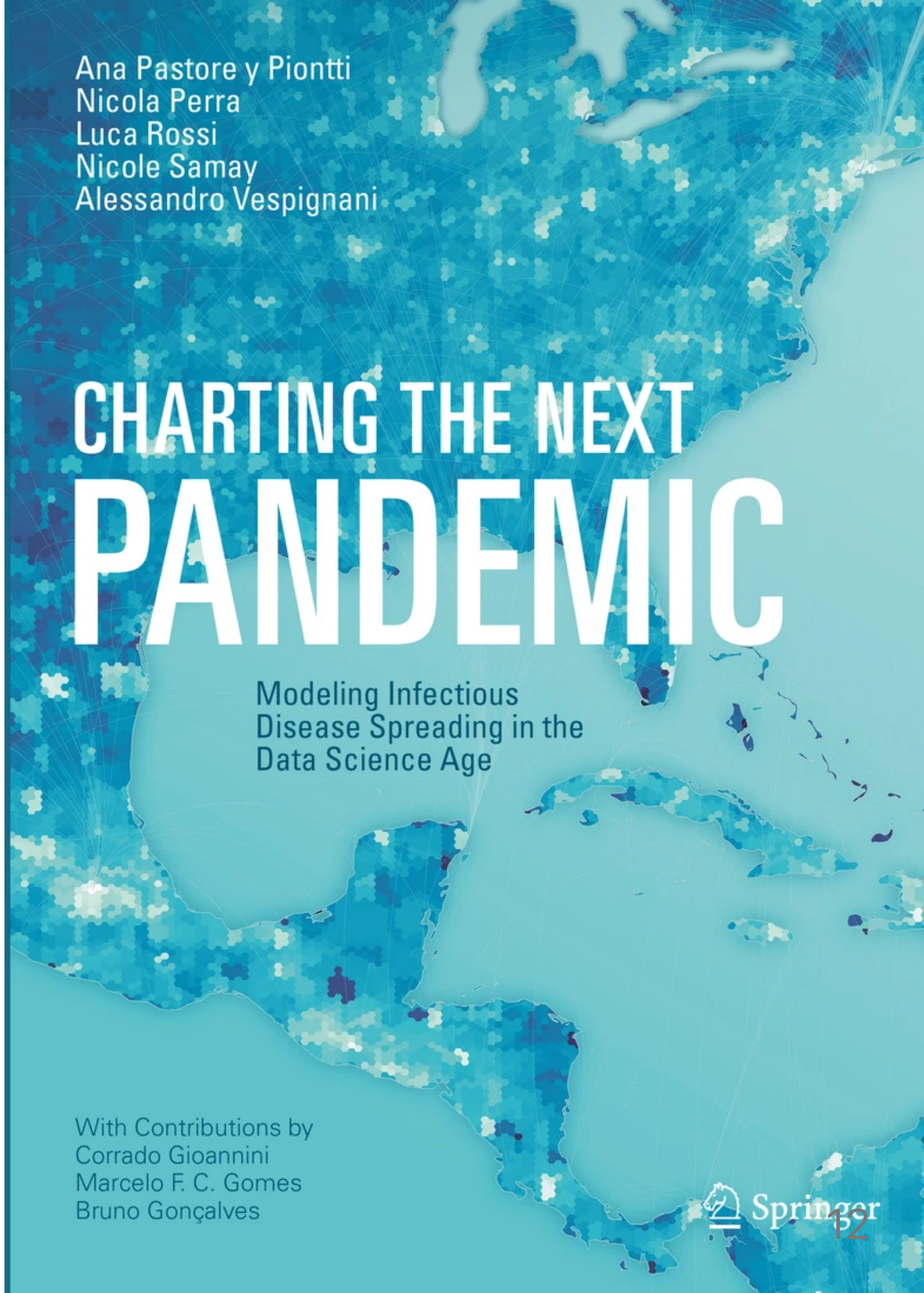
Pastore y Piontti , A., Perra , N., Rossi , L., Samay , N., Vespignani , A., Charting the Next Pandemic: Modeling Infectious Disease Spreading in the Data Science Age, Springer, 2019

Read it if interested in: **dynamical processes in networks**, **epidemics**, **visualization**, **data science**;

Focused on how to model infectious diseases in networks

Many illustrative examples, a lot of plots, and a great link between ideas and reality

This is mainly **optional,** BUT we will use part of it for our dataviz module and when introducing epidemics.



Exam

- **Network Science**:
  - Report: network analysis with real data and interpretation of the results
    - E.g., in the form of a commented Jupyter Notebook
  - Oral Examination
- **Data Visualization**:
  - Project: design and implementation of a data visualization platform to explore some real dataset(s)
    - It could take different forms: e.g., a d3.js project, a Jupyter Notebook using Python visualization libraries
- **Active class participation**:
  - up to +10%
  - You can get the maximum evaluation even if you do not attend classes actively. However, students who participate in Q&A sessions will be more likely to get some extra points.

Network Science module (NetSci): Description and Objectives

- **Networks are pervasive in our lives**
    - friendships, communication, computers, the Web, transportation, brain, protein interaction, and many others.
- We introduce the fundamental concepts, principles, and methods in the interdisciplinary field of network science, with a particular focus on exploratory analysis and modeling with applications for the World Wide Web and online social media.
- Topics covered include graphic structures of networks, mathematical models of networks, common network topologies, the structure of large-scale graphs, community structures, epidemic spreading, PageRank and other centrality measures, dynamic processes in networks, and graphs visualization.
- Students will learn how to apply the basic principles of network science to perform CNA (Complex Network Analysis) tasks on real data, with Python and many different packages/libraries such as NetworkX, as well as advanced graph visualization tools as Gephi.

DataViz description and objectives

- **Visualizations are extremely powerful** ("a picture is worth a thousand words")
- However, not all the visualizations are designed or implemented correctly, and sometimes they misrepresent underlying data.
- Students will learn basic visualization design and evaluation principles and how to acquire, parse, analyze, and visualize large datasets.
- Students will learn techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data.
- Students will utilize Gephi, D3.js, Python, matplotlib, and many other tools to prototype these techniques on existing datasets.

UNIVERSITÀ DI TORINO

Introduction to Networks

Reading material

[ns1] **Chapter 0**

[ns2] **Chapter 1**

UNIVERSITÀ DI TORINO

Complex Systems

Complex != Complicated

Composed by many **interacting elements**

They give rise to emergent **collective phenomena**

Emergence: not directly related to individual phenomena

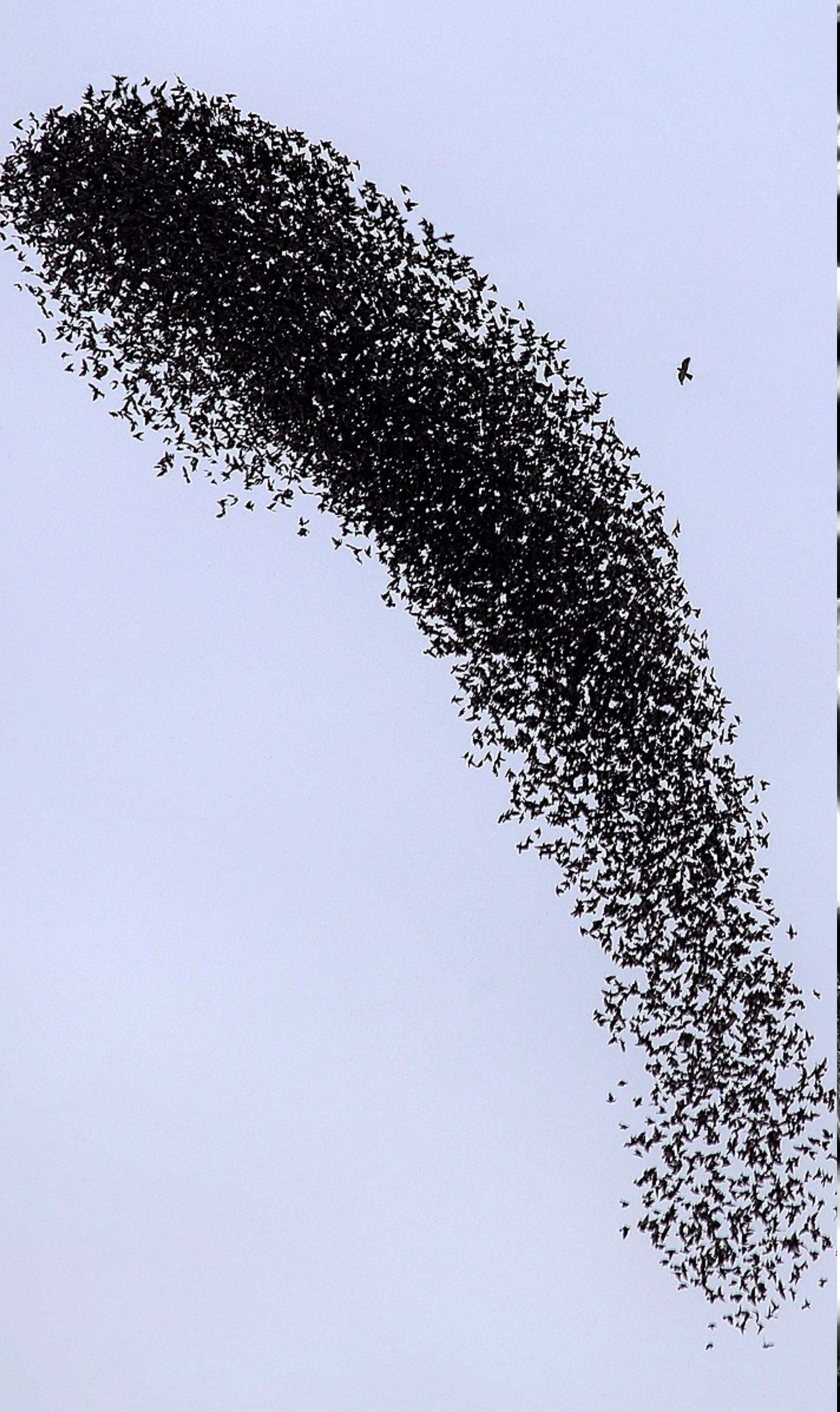
Linearity vs **non linearity**

Heterogeneous vs homogeneous

UNIVERSITÀ DI TORINO

An example of emergent synchronization

Watch this: https://www.youtube.com/watch?v=5v5eBf2KwF8



inversione
di marcia

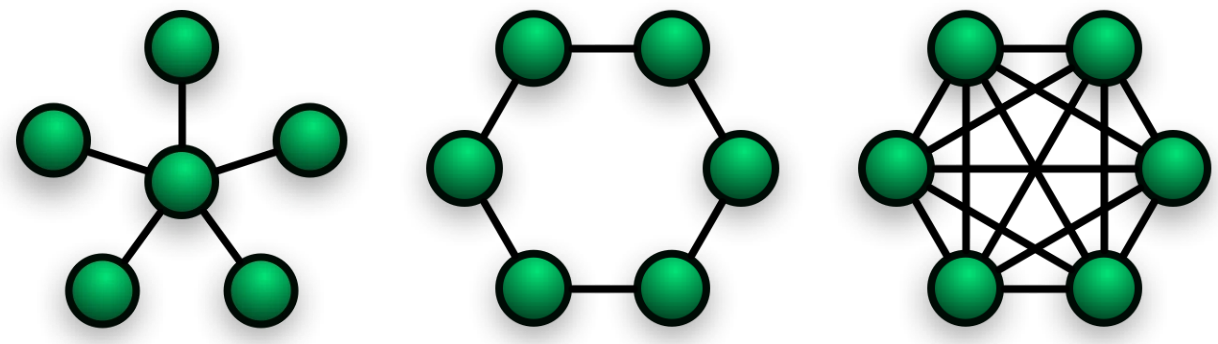
Networks are everywhere

- **Social Networks**
  - social ties, e.g., friendships, co-working, family members
  - political affiliations, sports
- **Information systems**
  - book, web page
  - citation, link
- **Economic systems**
  - loans, financial transactions
- **Technical systems**
  - power grid, water supply network, street network
- **Biological and ecological systems**
  - protein interaction, brain
  - food web
- ...

Aspects of Networks: Structure

Summary **representation** of a complex system

Simple networks: A few characteristics describe the network



We need a **language** and a **framework** to describe complex networks.

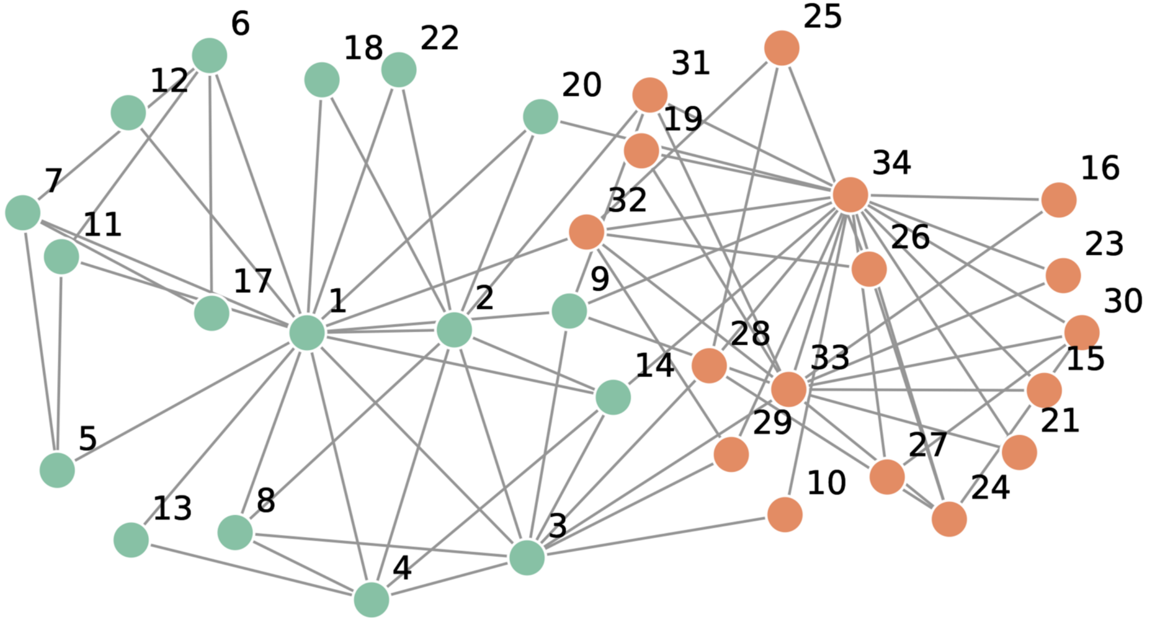
Example: Zachary's Karate Club

No simple description is possible.

However, there are some **regularities**.

Some individuals are more **central** than others.

Does it resemble some kind of **randomness**?



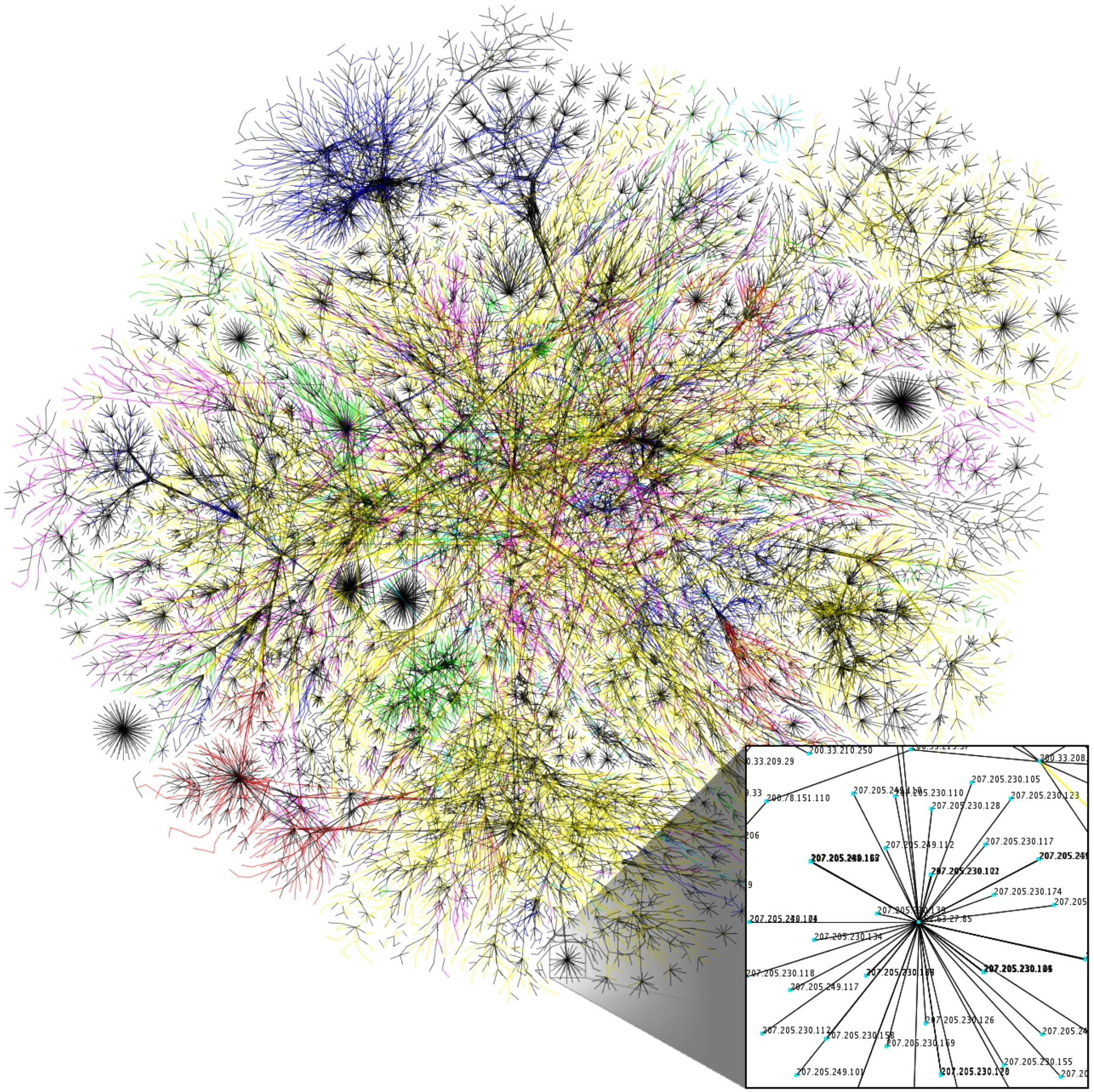
Example: the Internet

Some **order** emerges from chaos.

Core vs. periphery structure.

Other networks can show **hierarchical** structures.

We need a **language** to describe such regularities.



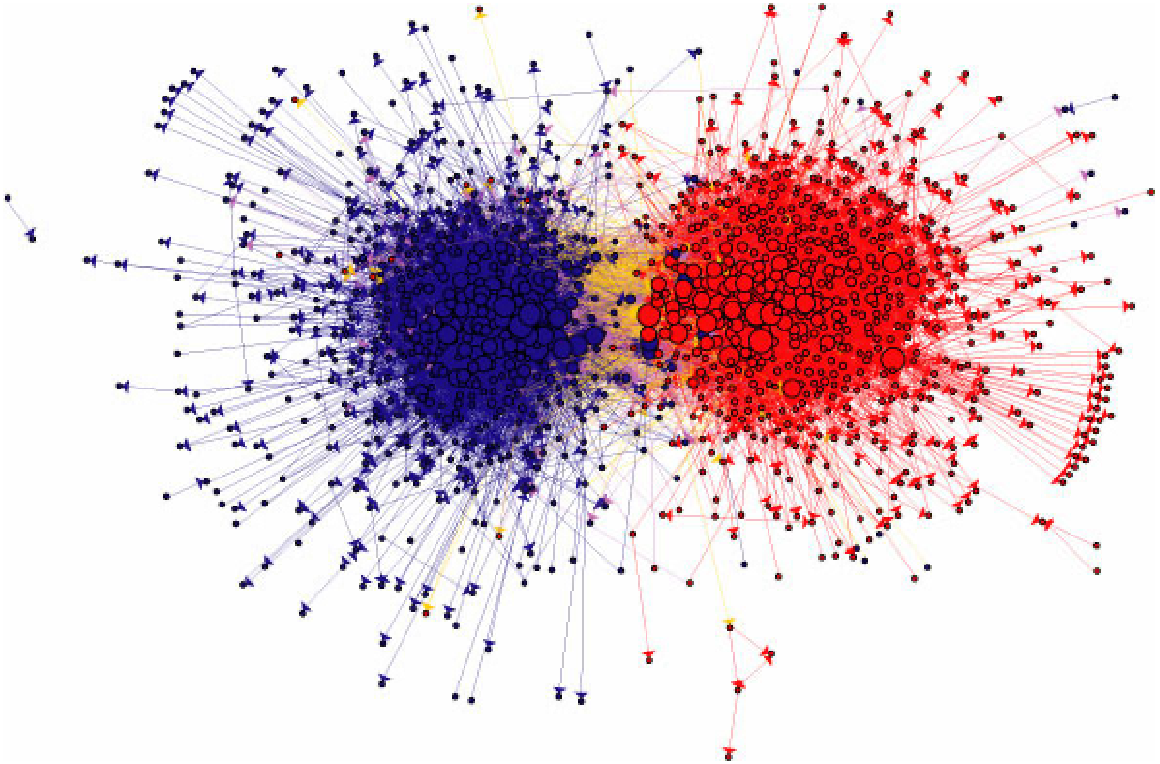
Example: retweet networks

communities or clusters (**partition**)

echo chambers

homophily : tendency of individuals to link with similar ones

warning: no trivial linear relationships but **interplay**



Visualization is not enough

- Visualizing a network suggests some **inherent** complexity
- this says a lot about the **structure** , but almost nothing about the **reasons**
- It is difficult to summarize succinctly the whole network
    - core vs periphery structure
    - hierarchical structure
    - tightly-linked regions
    - ...
- we need a **language** !

UNIVERSITÀ DI TORINO

Aspects of Networks: Behavior and Dynamics

- Connectedness
  - structure
  - behavior
- each **individual's actions have consequences** for the outcomes of everyone in the system
- we need a **framework for reasoning** about behavior and interactions in networks
- we need to take into account a **rational behavior** that sometimes lead to **strategies**
- actions are not evaluated at individual level (in isolation), but at a **network level**

From data to networks

We have **massive network datasets**

data-driven models allow you to make predictions BUT sometimes these predictions cannot be generalized: networks can be **different**

Individual vs Population level

- when a large group of people is tightly interconnected, they often respond in complex ways that can be observable only at the **population level**:
    - rich get richer
    - winner takes it all
    - viral ideas
    - ...
- At **individual level** : sometimes an information goes viral or one person becomes very popular, some other times this does not happen!
    - **Why**? no prediction at this level yet

Main topics in this course

Graph Theory and Complex Network Analysis

Game Theory (only basic concepts)

Information Networks

Network Dynamics

Graph Theory

paths/distances

clustering coefficient

degree and degree distribution

centralities

groups/clusters/partitions/communities

structural balance

homophily and spatial segregation

Individuals vs groups

We can find clues to the latent schism that eventually split the group into two separate clubs.

We can have different roles.

Game Theory

- a framework where our decisions depend on others' decisions
- **strategy** -> **pay-off**
- ex. in transportation networks, choice of a route can bring to **congestion** (Braess' Paradox)
- **Equilibrium** : a state that is self-reinforcing.
  - no individual has an incentive to unilaterally change their strategy, even if those individuals know how others will behave

Information Networks

E.g.: the Web

Interplay: if you define a measure (e.g., PageRank), the application of such measure may change the system

React to maximize rank.

Web pages authors

UNIVERSITÀ DI TORINO

Network dynamics: population effects

In large populations we have the emergence of new ideas, beliefs, opions, innovations, technologies, products, and social conventions

When and how are they established as **social practices**?

Individuals can **influence** or be influenced by others.

At the surface, humans tend to **conform**

Why ?

UNIVERSITÀ DI TORINO

Networks Dynamics: Structural Effects

- Populations vs local behaviors
- Individuals may have an incentive to adopt the behavior of their neighbors: **cascading behavior**
  - social contagion
  - epidemics
- The structure of the network has a role

Social contagion

- Influence: the adoption of a new product or innovation can cascade through the network structure
- ex.: e-mail recommendations for a Japanese graphic novel
- We have:
    - spreaders
    - closed communities that can stop the contagion

Epidemics

Epidemic disease is another form of cascading behavior in a network.

ex.: tuberculosis outbreak

Network-level dynamics are similar, and insights from the study of biological epidemics are also helpful in thinking about the process by which things spread in other networks (e.g., rumors and fake news)

Tools and Software

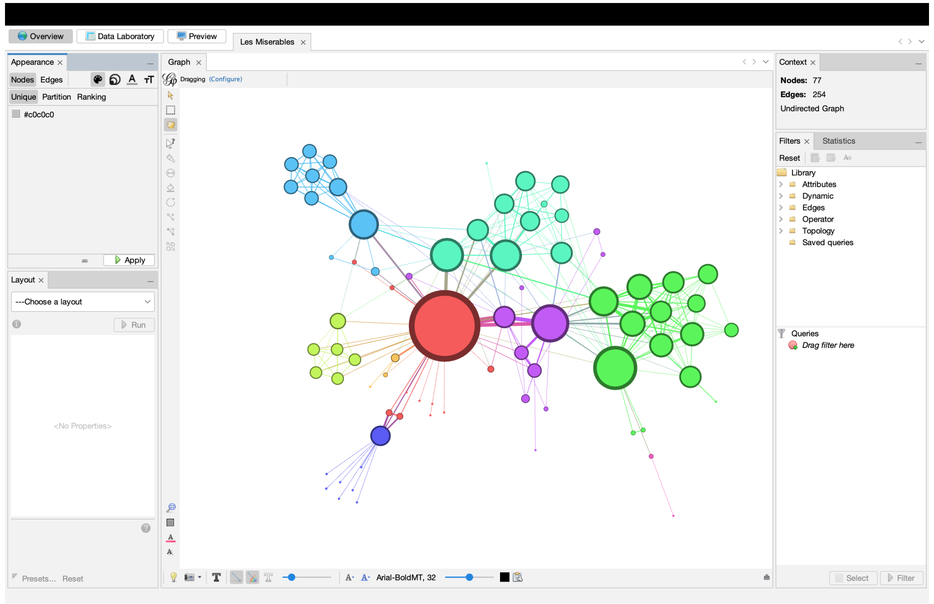
Network Analysis software

- **Python 3 + Project Jupyter/JupyterLab**
  - Alternatively: R
  - The use of the R ecosystem is out of the scope of this course; however, you can replicate all the practical sessions with alternative computational frameworks
- Students can pick the platform of their preference:
  - local distribution: Anaconda or base Python + pip
  - cloud distributions (many free): e.g., Google Colab, Binder Kaggle Kernels, Azure Notebooks,Datalore
- **networkX**
  - Even if there are other very valid alternative packages, e.g., igraph, graph-tool, this course will provide solutions to the practical sessions using networkX.

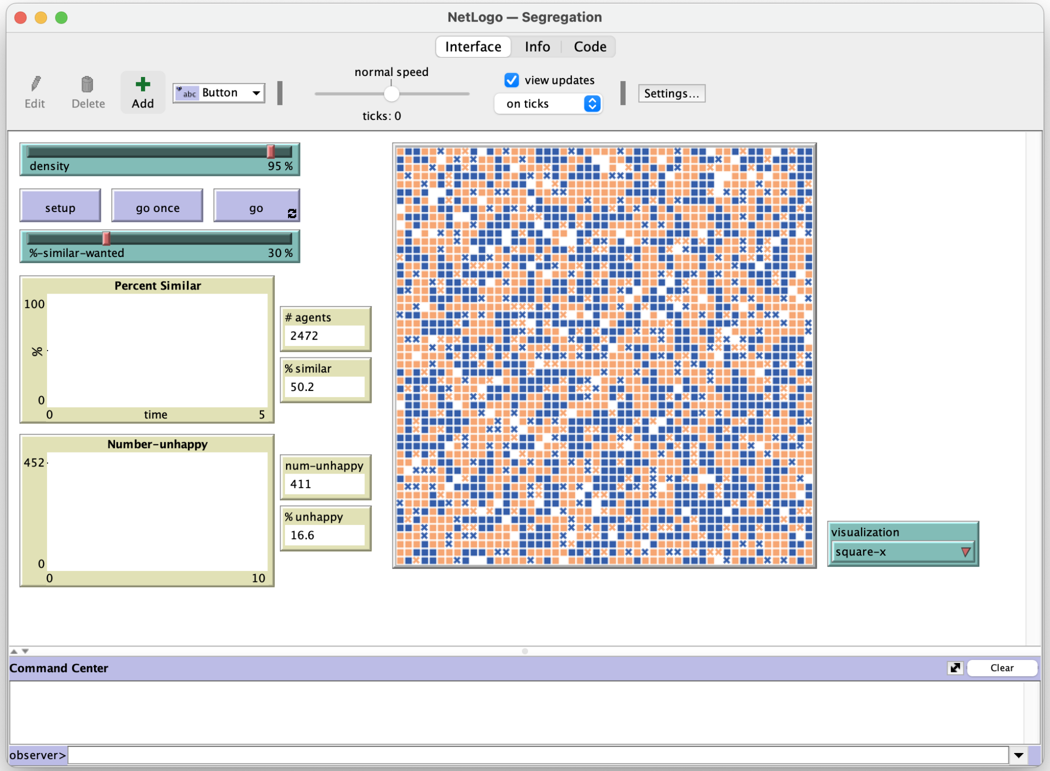
UNIVERSITÀ DI TORINO

Other software and tools

- Gephi
  - Tutorials and learning resources



- NetLogo
  - Beginner's guide



UNIVERSITÀ DI TORINO

Q&A