Statistical Analysis of Networks

Lecture 1 - Course Introduction

DOMENICO DE STEFANO ddestefano@units.it

SUSANNA ZACCARIN susanna.zaccarin@deams.units.it

MARCH 6 2024

Course schedule

- Wednesday 9 - 11 am
- Thursday 13 – 15 pm
- Friday 15 – 17 pm

Office hours:

D. De Stefano, **Tuesday 14.30 – 16.30 pm,** Building A, room 432

S. Zaccarin**, Thursday 11 am - 13 pm**, Building D, room 2.15

(write an-email for other days/hours)

Next lecture: March 8

Course organization combines basic theoretical concepts with software and practical applications with R

COURSE SYLLABUS

- Basic analytical concepts in network analysis
- Network data collection
- Network visualization
- Descriptive analysis and network indices
- Network decomposition
- Modeling network structure

The course will include practical examples and hands-on computer laboratories based on the analysis of real-life relational data.

In the laboratories, the emphasis will be on the analysis of social networks in structured social and economic settings such as, for example, business companies, and organizations.

Software tools in R

TEXTBOOKS

1. Kolaczyk E.D. (2009) Statistical Analysis of Network Data. Methods and Models, Springer, New York.

2. Lusher D., Koskinen J. and Robins G. (eds.) (2013) Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press (selected chapters).

Recommended readings:

3. Hanneman R.A. and Riddle M. (2005) Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/).

Additional materials, lecture notes and information will be available at the course web page and via moodle2 e-learning platform.

COURSE PREREQUISITES

Students are required to have basic knowledge in inferential statistics and should be familiar with linear and logistic regression models.

Basic knowledge of software R will be also required.

EXAM

The final exam will consist in the presentation and discussion, in groups of 2/3 students, on the analysis (description and model fitting) of a real network dataset, explaining the working steps and the obtained results.

The writing of a short report is also requested, with the commented R code.

During the presentations, few questions will be asked to assess the individual contributions and preparation on the topics of the course.

During the course, homeworks and discussion on specific topics will be proposed.

SOFTWARE TO BE INSTALLED FOR NEXT WEEK CLASSES

R packages:

- statnet (suite including several useful NA packages)
- igraph

Other packages will be included in the R script provided for the upcoming R labs

WHAT IS A NETWORK?

a collection of interconnected things

***things**:* individuals, blogs, digital devices, genes, organizations, academic papers, words, …

***interconnected**:* one (or more) relations among them

Relations can be of any kind (in general, to be defined):

contacts - acquaintance, friendship, contagion - blogger interactions, wired or wireless communications, regulatory behaviour of genes, collaboration in management, co-citations, semantic links, …

WHAT IS A NETWORK?

A (conceptual and formal) model to represent a set of connections, tangible or ideal among objects (entities/ units)

that can be applied in many fields: physics, chemistry, computer science, sociology, economics, etc.

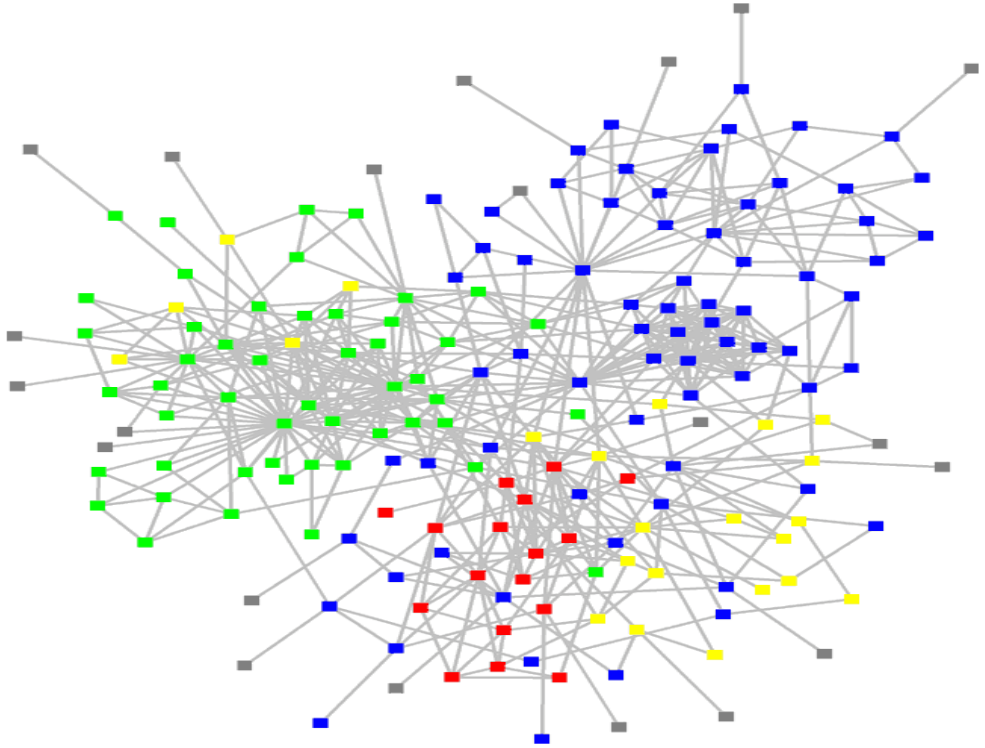
- phenomena exhibiting a relational nature can be properly formalized by network(s)

Examples in the next slides

The “Social Graph” behind Facebook (2017)

Structure of an organization

(note graph visualization!)



: departments

: consultants

: external experts

www.orgnet.com

The students' network (peer support) of Hogwarts - first book

Harry Potter and the Philosopher's Stone

Circles indicate that the person is attending Hogwarts in the school year 1991-1992, rectangles indicate future students.

The colors represent the houses:

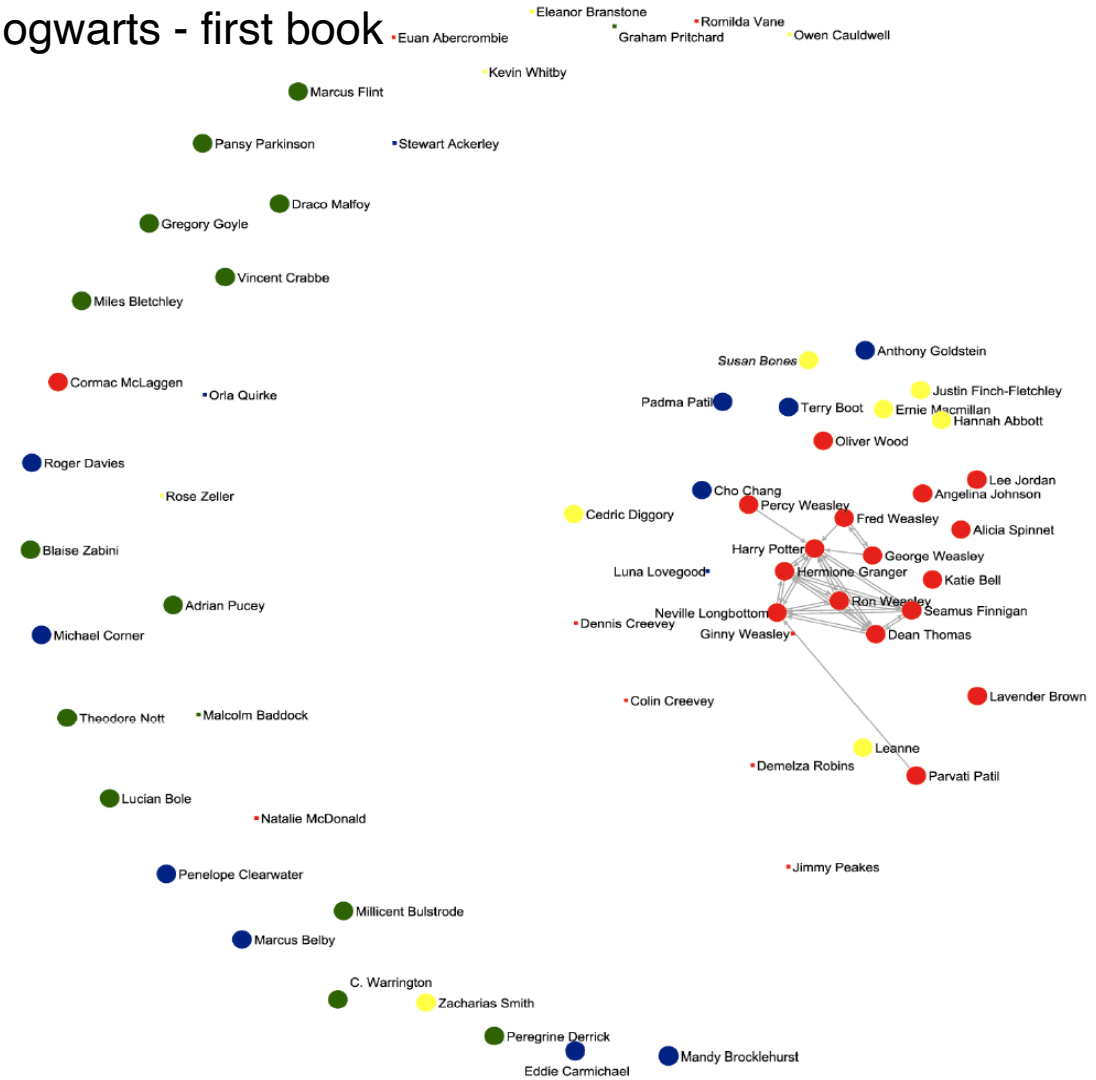
red = Gryffindor

yellow = Hufflepuff

blue = Ravenclaw

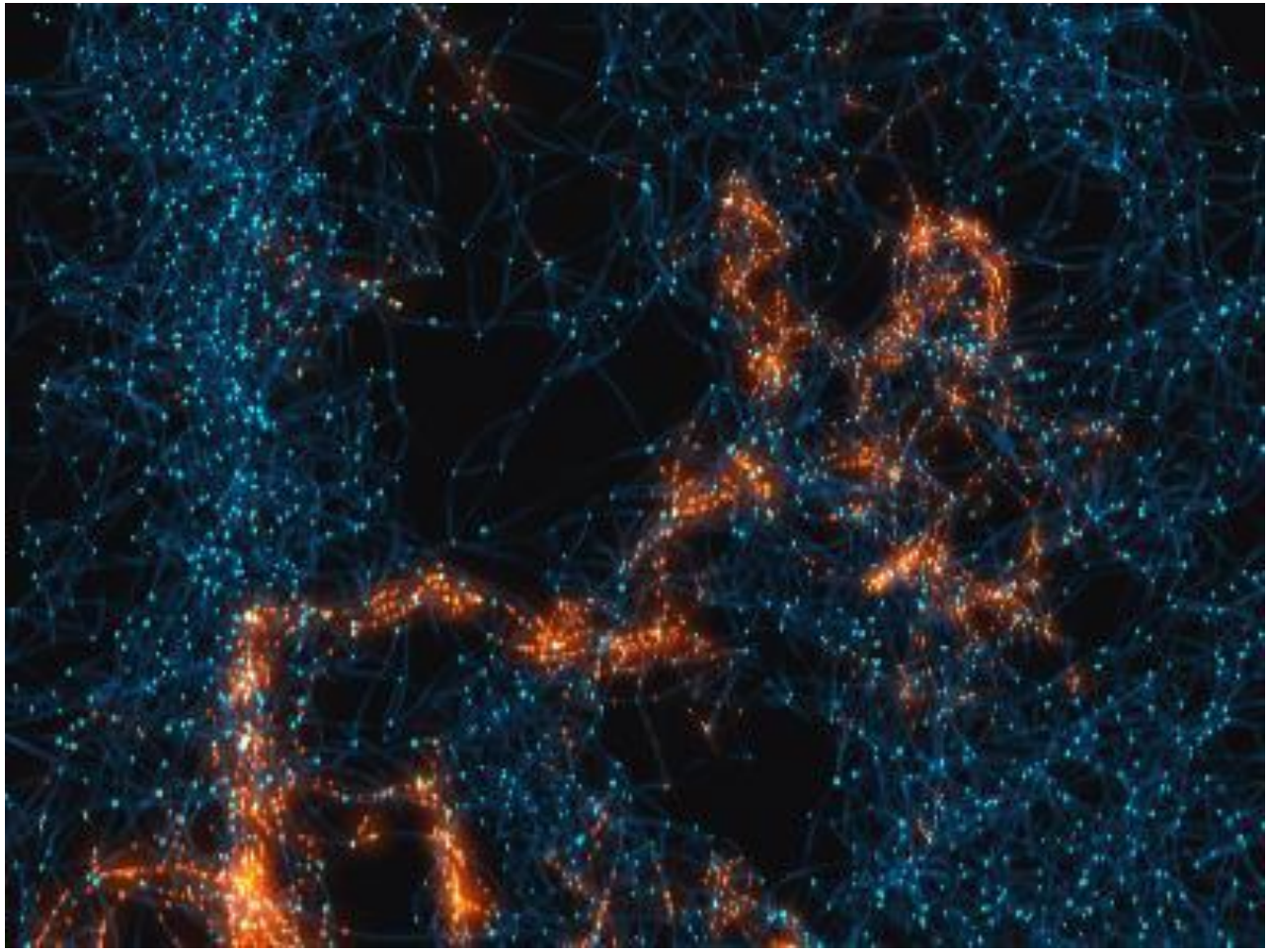
green = Slytherin

Peer support relation: emotional, instrumental, providing information and other kind of support



Open Journal of Applied Sciences, 2013, 3, 174-185 doi:10.4236/ojapps.2013.32024 Published Online June 2013
(http://www.scirp.org/journal/ojapps)

Neuron network of brain



http://nxxcxx.github.io/Neural-Network/

Red areas (nodes and links) indicate brain activity (simulation)

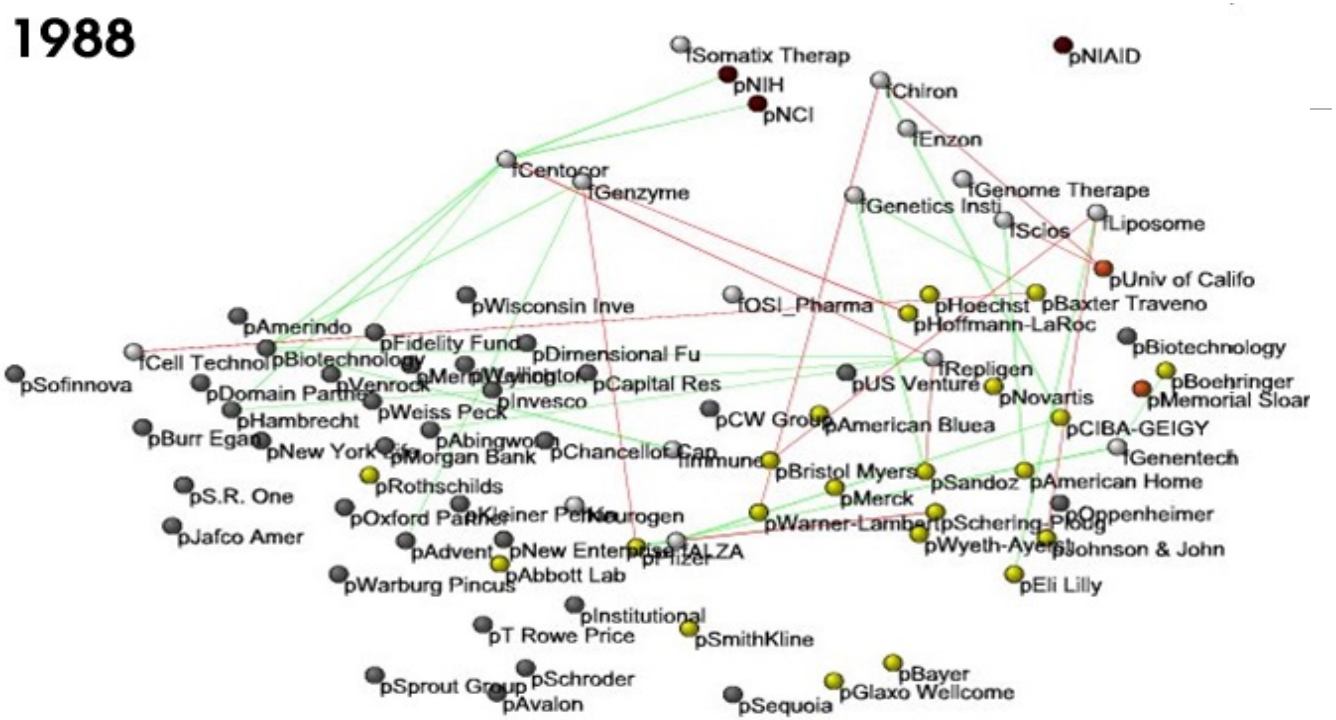
US biotech-industry – business relations in 1988 - 1991

Nodes:

- Companies
- Investment
- Pharma
- Research Labs
- Public
- Biotechnology

Links:

- Collaborations
- Financial
- R&D



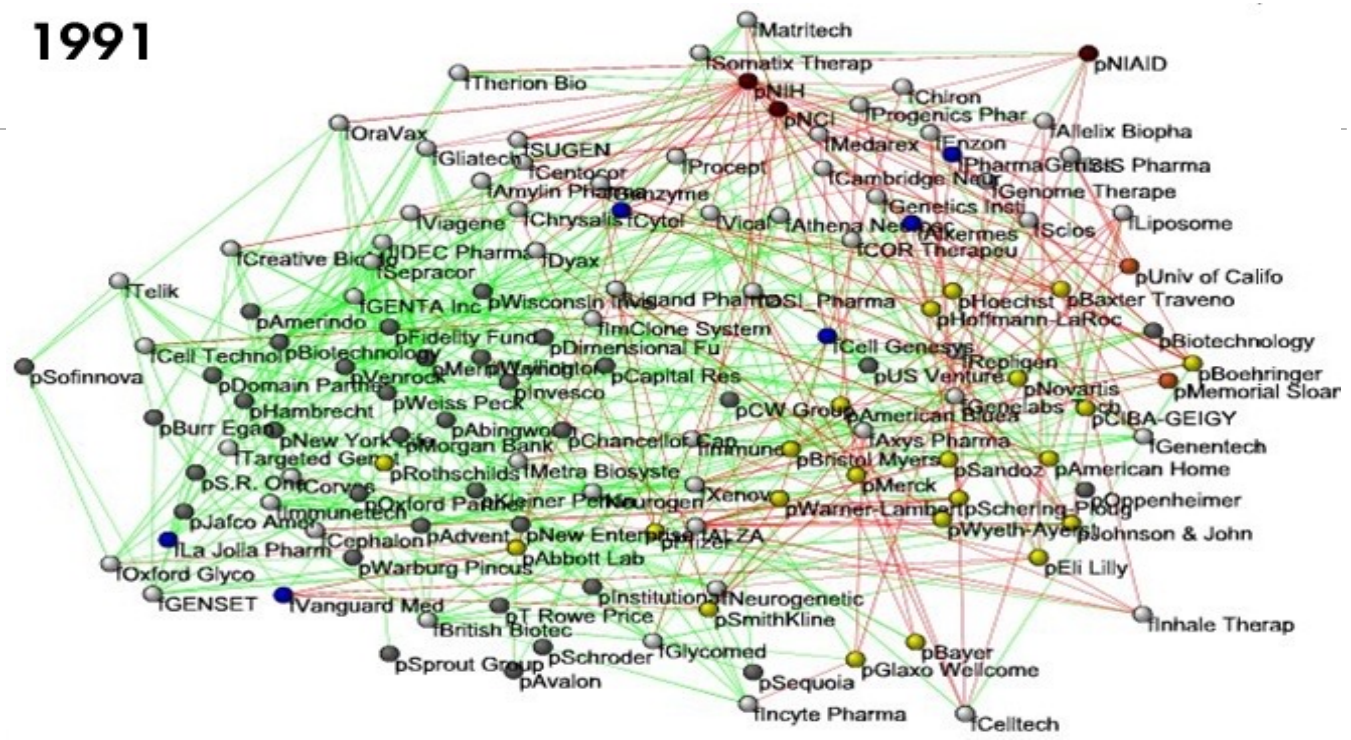
US biotech-industry – business relations in 1988 - 1991

Nodes:

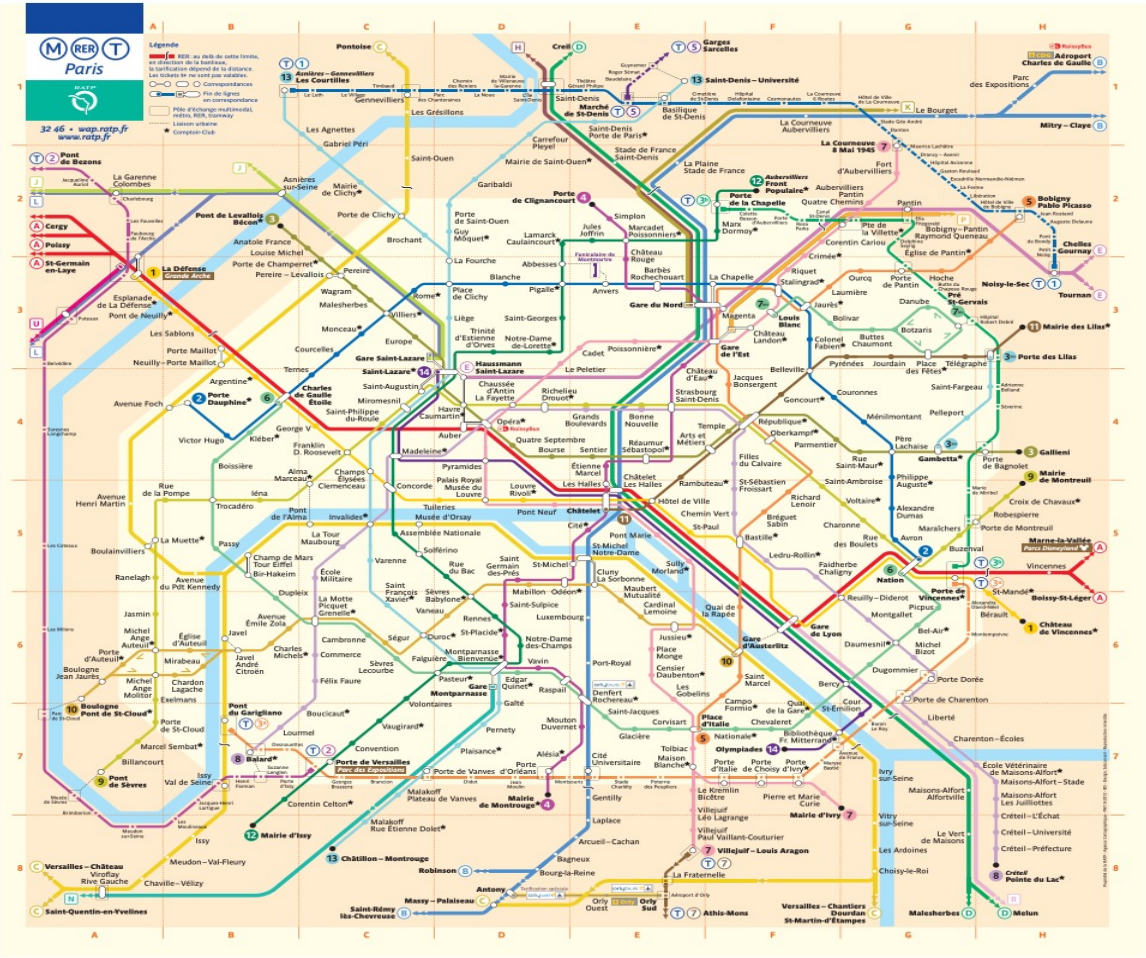
- Companies
- Investment
- Pharma
- Research Labs
- Public
- Biotechnology

Links:

- Collaborations
- Financial
- R&D



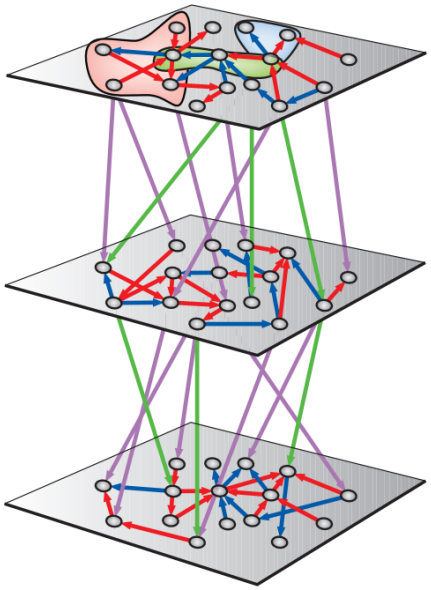
Paris metro



14 (+ 2) lines
303 stations

The immune system as a social network (MULTILAYER Network)

Andreas Bergthaler & Jörg Menche *Nature Immunology,* 18, 5, 2017



Network component	Social network	Immune system
Actors	Members of society	Immune-cell types
Roles	Same profession; similar age; shared interests	Shared function—e.g., early defense mechanisms; humoral immunity
Context	Family; professional; spare time	Organ; tissue; activation status
Intra-layer communication	Personal communication; e-mails; phone calls	Direct cell–cell interactions; receptor–ligand interactions
Inter-layer communication	Similar to intra-layer communication	Signaling across organs, tissues

Figure 1 The immune system as a multi-layered social network. The complex interplay among immune cells resembles in many ways a social network. Immune cells need to act in a highly cooperative and coordinated fashion when they respond to diverse external and internal threats. Communication

The immune system employs a multitude of molecules, cells and organs that act together throughout the entire body to guard human health.

Much like in a social network, immune cells can exert full functionality only through effective collaboration and communication.

Humans genes

Homo
Sapiens

Drosophila
Melanogaster

Points: elements of each organism's genetic network
Dotted lines: interactions between elements

Complex systems

Made of many non-identical **elements** connected by diverse **interactions**.

On-line Sources

Stanford Large Network Dataset Collection
https://snap.stanford.edu/data/

Cambridge Networks Network
https://www.cnn.group.cam.ac.uk/Resources/resources

Covert networks
https://sites.google.com/site/ucinetsoftware/datasets/covert-networks

WHAT IS A NETWORK?

Behind each complex system there is a network,

that defines the interactions between the components

WHY TO DO NETWORK ANALYSIS ?

Some good reasons:

- to study *whether the enviroment/system (network) affects unit outcomes*

- to study *whether occupying specific position* (central, intermediary, peripheral) in the system (network) have different outcomes

- to study *how single units affect* the system

- to study processes/mechanisms sustaining the system (network) structure

- to study intertwinning of unit outcomes and the system (network) structure

- to study the global system (network) effectiveness

(visualization is important but by itself it can not be sufficient to provide a definitive answer)

WHY NETWORK SCIENCE?

(TERM USED TO ADDRESS NETWORK RESEARCH)

Interdisciplinary nature:

Network research crossed many disciplinary boundaries.

Network science offers a language through which different disciplines can seamlessly interact with each other.

Indeed, cell biologists and computer scientists alike are faced with the task of characterizing the wiring diagram behind their system, extracting information from incomplete and noisy datasets, and the need to understand their systems' robustness to failures or deliberate attacks.

WHY NETWORK SCIENCE?

(TERM USED TO ADDRESS NETWORK RESEARCH)

Empirical, data driven nature:

The tools of network science have their roots in **graph theory**, a fertile field of mathematics.

What distinguishes network science from graph theory is its empirical nature, i.e. its focus on data (real networks) and utility (insights mathematical tools can offer about a system's structure or evolution).

WHY NETWORK SCIENCE?

(TERM USED TO ADDRESS NETWORK RESEARCH)

Quantitative and mathematical nature:

To contribute to the development of network science, it is essential to master the mathematical tools behind it.
The tools of network science borrowed the formalism to deal with graphs from graph theory and the conceptual framework to deal with randomness.
The field also is benefiting from concepts borrowed from engineering, control and information theory, statistics and data mining, helping us extract information from incomplete and noisy datasets.

WHY NETWORK SCIENCE?

(TERM USED TO ADDRESS NETWORK RESEARCH)

Computational nature:

Given the size of many networks, and the exceptional amount of data behind them, network science offers a series of computational challenges.
Hence, the field has a strong computational character, actively borrowing from algorithms, database management and data mining.
A series of software tools have been developed in the last years to analyze networks (also for practitioners).

INDIVIDUAL-BASED *VS* NETWORK-BASED RESEARCH

Individual-based (usual approach): interest in individual outcomes

- definition of the (statistical) population
- selection of a (random) sample and data collection
- inference (from the sample to population) based on (conditional) independency condition among units

Network-based: interest in individual and system outcomes

- network *boundary:* units to be considered in the network
- unit as well as relation *dependency* (outcomes affected by a structure of relations)
- network synthesis: *adaptation* of usual statistical measures

HOMEWORKS – 1^ WEEK

Readings:

Barabasi A. (2012) The network takeover, *Nature Physics*, 8, pp. 14-16.

Brandes U., Robins G., McCrannie A., Wassermann S. (2013) What is network science? *Network Science*, 1, pp.1-15.

Carrington P.J., Scott J. (2011) *Introduction,* in Carrington P.J., Scott J. (Eds.) The Sage Handbook of Social Network Analysis, Sage Publications.

(Second Edition: McLevey J., Scott J., Carrington P.J. (2023) *The Sage Handbook of Social Network Analysis SECOND EDITION, Sage Publications*)