# Analyse de réseaux

Une courte introduction

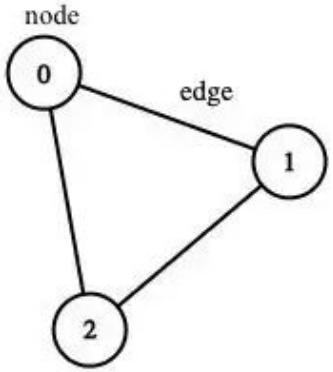
# Objectifs

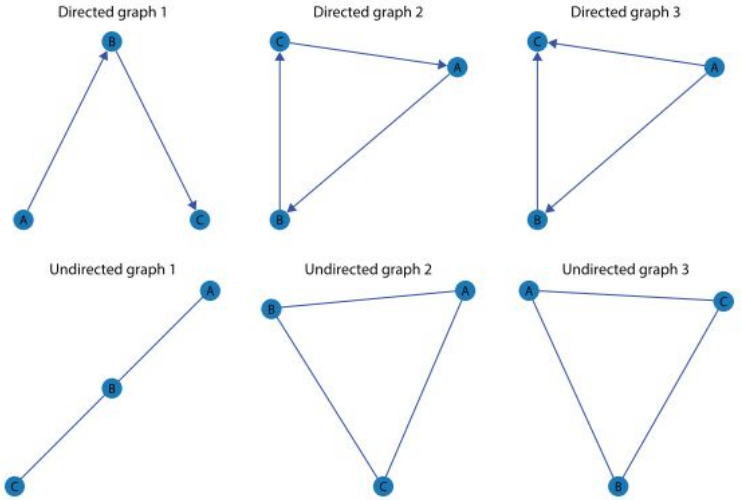
- Un peu de contexte sur le Social Network Analysis
- Les bases de vocabulaire pour en parler
- Quelques indicateurs importants à calculer
  - Centralité
  - Densité
  - Clustering
- Et puis on va s’y mettre…

# Au début : noeuds et liens

- Noeuds/Sommets
  - personnes
  - organisations
  - divers
- Liens
  - relations
  - orientés ?
- Des attributs
  - de noeuds
  - de liens

C’est un modèle



# De nombreux usages

| | |
|---|---|
| Physicists | phase transitions, universality |
| Computer Scientists | data / algorithm oriented, predictions |
| Applied Mathematicians | dynamical systems, diff. eq. |
| Statisticians | inference, consistency, covariates |
| Biologists | experiments, causality, molecules |
| Ecologists | observation, experiments, species |
| Sociologists | individuals, differences, causality |
| Political Scientists | rationality, influence, conflict |

| *domain* | *network* | *vertex* | *edge* |
|---|---|---|---|
| biological | metabolic network | metabolite | metabolic reaction |
| | protein-interaction network | protein | bonding |
| | gene regulatory network | gene | regulatory effect |
| | drug interactions | drug | *in vivo* health interaction |
| | connectome | neuron | synapse |
| | physiology | muscles and bones | physical attachment |
| | pollination network | plants and pollinators | pollination |
| | food web | species | predation or resource transfer |
| social | friendship network (offline) | person | friendship, trust, etc. |
| | friendship network (online) | account | "friendship," follow, etc. |
| | proximity network | person | physical proximity |
| | sexual network | person | intercourse |
| | coauthorships | authors | collaboration |
| | fictional | character | co-appearance |
| | animal behavior | animals | interaction |
| economic | hiring network | workers and jobs | hired into |
| | international trade | country | trade flow |
| | purchasing | users and items | purchased |
| | board of directors | directors and boards | sits on |
| | inventions | inventors and patents | authored |
| information | software | function | function call |
| | World Wide Web | web page | hyperlink |
| | documents | article, patent, legal case | citation |
| | artifacts | item, document, concept | relatedness or similarity |
| | language | word | adjacency in text |
| technological | Internet (1) | computer | IP network adjacency |
| | Internet (2) | autonomous system (AS) | GBP connection |
| | digital circuits | logic gates | wire |
| | power grid | generating or relay station | transmission line |
| transportation | rail system | rail station | railroad tracks |
| | road network (1) | intersection | pavement |
| | road network (2) | named road | intersection |
| | airport network | airport | non-stop flight |

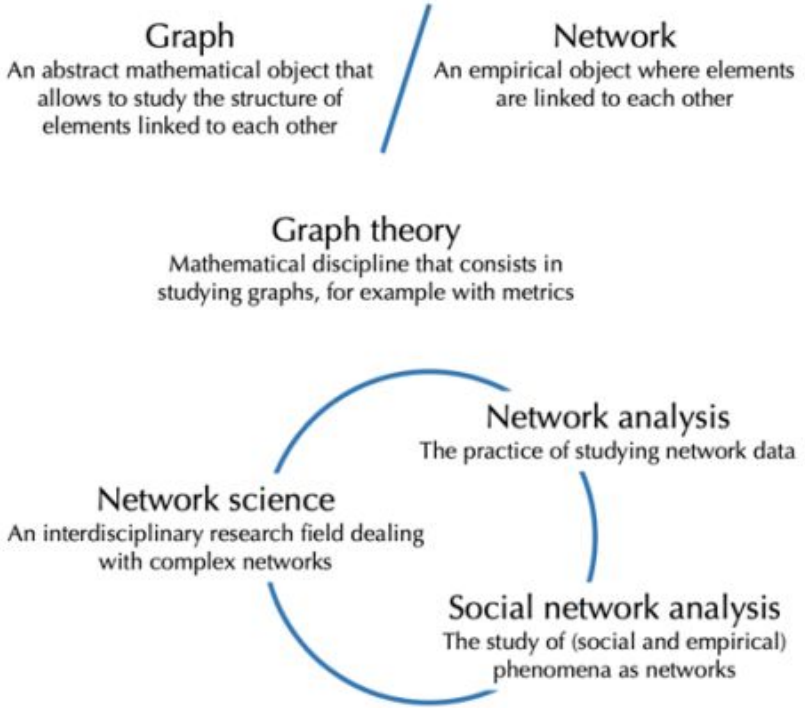
# Types of questions answered by SNA

Between holism (structures) and atomism (properties of entities)

- How does the information flew between actors ?
- Who is the most influential actor in a group ?
- Does different actors access to different entities ?
- Can we distinguish groups of entities ?
- How does the densities of exchanges of information evolve over time ?
- How close are two persons ?
- How many friends has this person ?

# Different objets/fields

- Social theories based on social network analysis
  - social capital
- But mainly methodological tools
- A whole area in mathematics : graph theory



# Producing the data

- definition of what count as an edge or a node
- the limit of the collection of data
  - what does sampling a network means ? For instance, if a bridge is out of the sample
- how to collect data:
  - transforming existing data (scientific publications)
  - using metadata (social media)
  - collecting (names generator and interview)

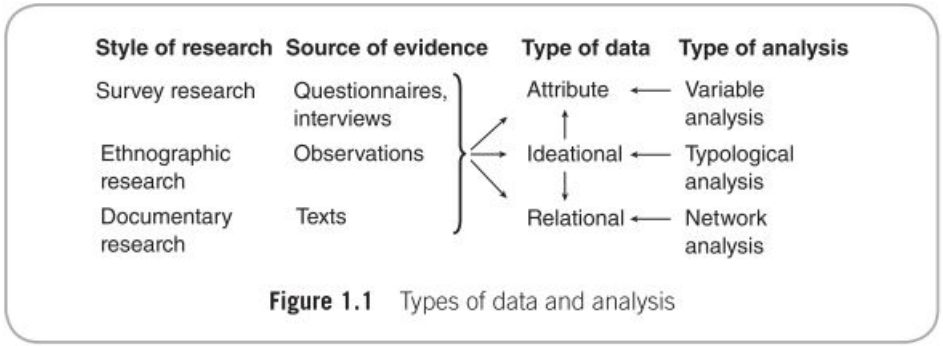



**Figure 1.1** Types of data and analysis

# Example : scientific collaborations

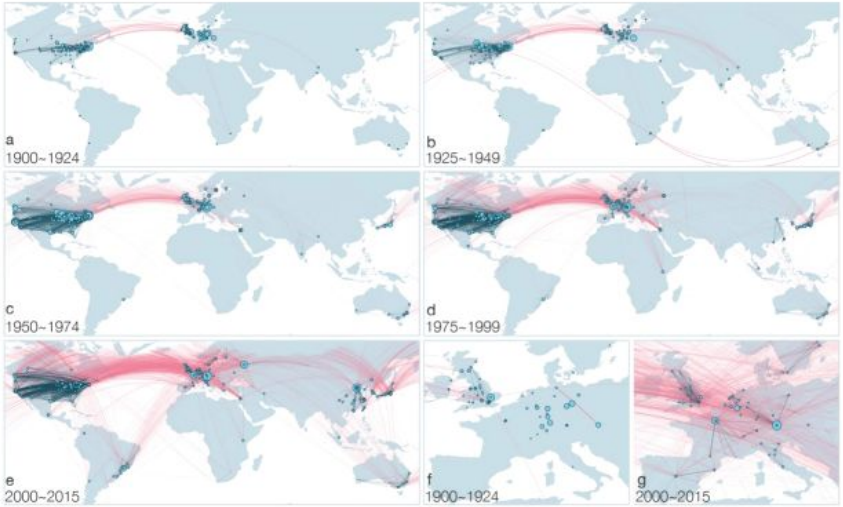
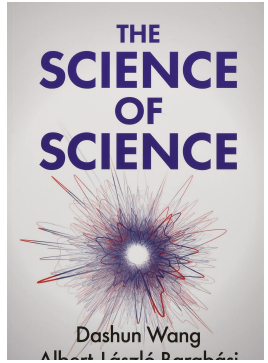



Figure 2.1.2 **A brief history of globalization of science collaborations**. The figure shows two types of collaborations: across institutions within the same country (green) and across institutions of different countries (blue). Between 1900–1924, collaborations across different institutions were prominent only i the US; international collaborations were mainly between the U.S. and the U.K. However, both types of collaborations were relatively weak. Between 1925–1949, international collaborations started to form between India and the U.K., as well as between Australia and the U.S. Due to World War II, and collaborations in Europe shrank during this period. Meanwhile, collaborative relationships in America were rapidly developing in the Midwest. Between 1950–1974, Israel and Japan started to engage in international teamwork. At the same time, the West Coast and the Southern United States became hubs scientific collaboration. Between 1975–1999, Africa began to develop relationships with Europe. Surprisingly, within-institution collaborations in the U.S. decreased relative to those in Europe, althoug the absolute number of collaborations grew substantially over time for all countries. In the 21st century, more and more countries have risen to the international stage to collaborate extensively with others. Aft Dong 2017 [13].



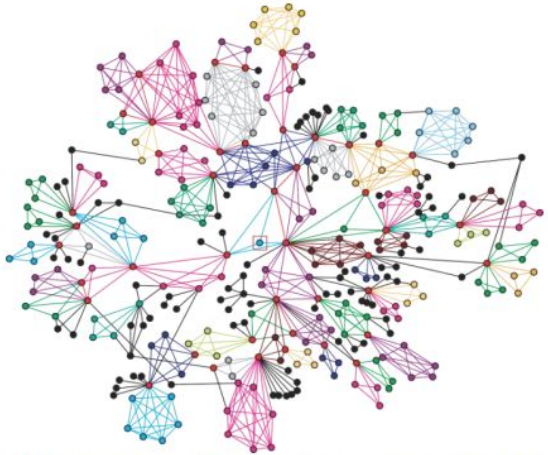
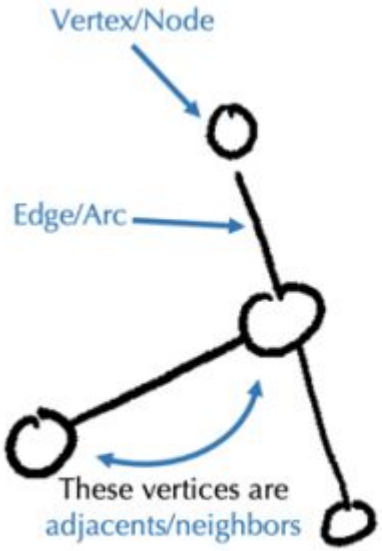



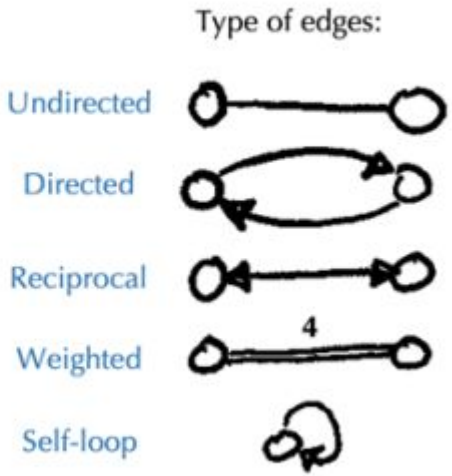
Figure 2.3.1 **Co-authorship network.** The figure shows the local structure of the co-authorship network between physicists in the vicinity of a randomly chosen individual (marked by a red frame). The network is constructed based on papers from Cornell University's archive server (cond-mat), the precursor of the widely used arXiv, containing at that time over 30,000 authors. Each node is a scientist, and links document collaborative relationships in the form of co-authored publications. The color-coded communities represent groups of collaborators that belong to locally densely interconnected parts within the network. Black nodes/edges mark those scientists that do not belong to any community. After Palla *et al.* [34].

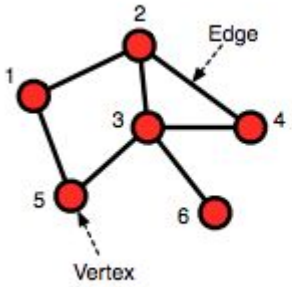
https://github.com/marcessz/Twenty-Years-of-Network-Science

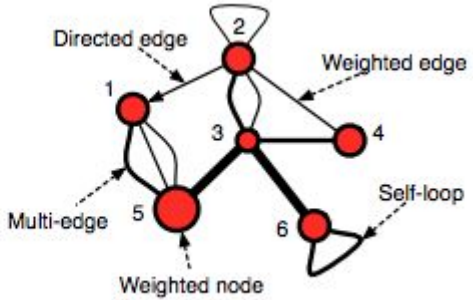
# Types of relations

At the end, it’s always a model, not the reality



# Types of networks

- **Sociocentric** : Whole networks
  - Creates one network
  - If your question is about different patterns of interaction within defined groups
  - The question is always : what will define the perimeter ?
- **Egocentric** : Personal networks
  - Creates many stand alone networks
  - If your question is about phenomena of or affecting individual entities across different settings
- one type of nodes : **1-mode**
- two types of nodes : **2-mode**, or bipartite
  - no connexion between nodes of the same set

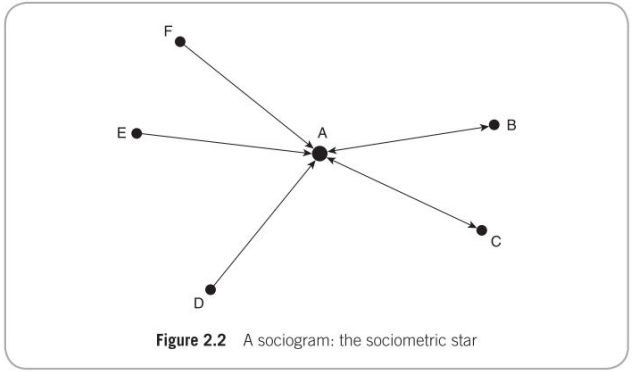


Figure 2.2 A sociogram: the sociometric star

# Building a graph from informations

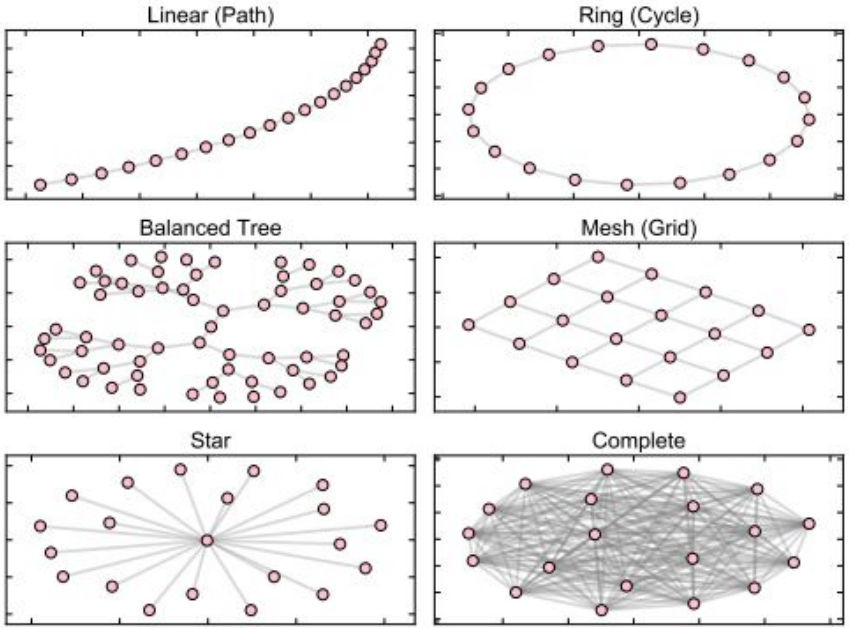
Basic situation : list of edges

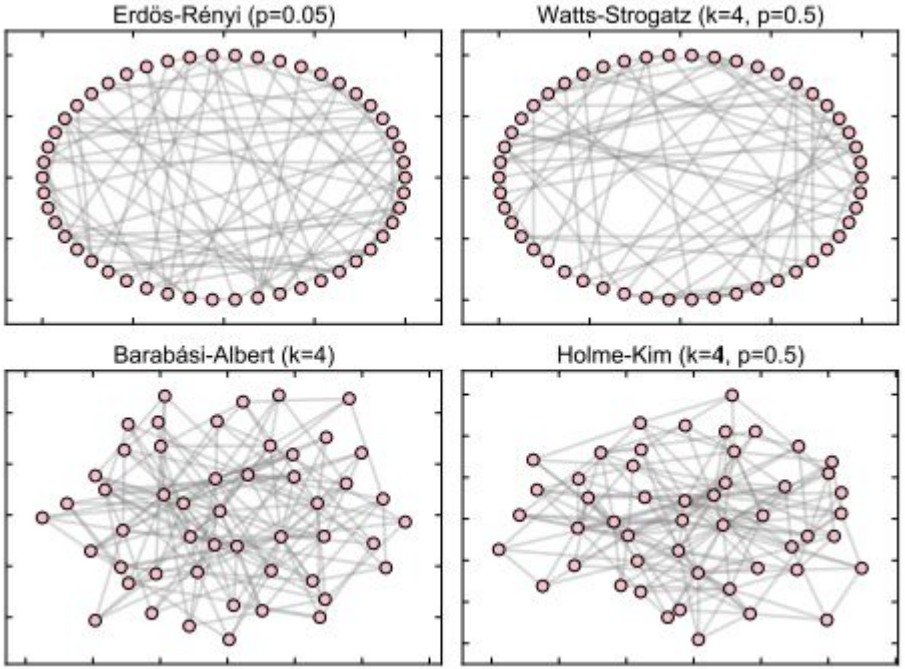
(for instance, interactions between users online)

(A;C) ; (C;D) ; (D;E) ; (E;A) ; (A;C) ; (F;G) ; (C ; F)

How do we build a network ?

# Synthetic networks (generated)



# Characteristics of the whole graph

- Type of the network
- Number of nodes
- Number of edges
- Number of components
- Diameter : max distance
- Density

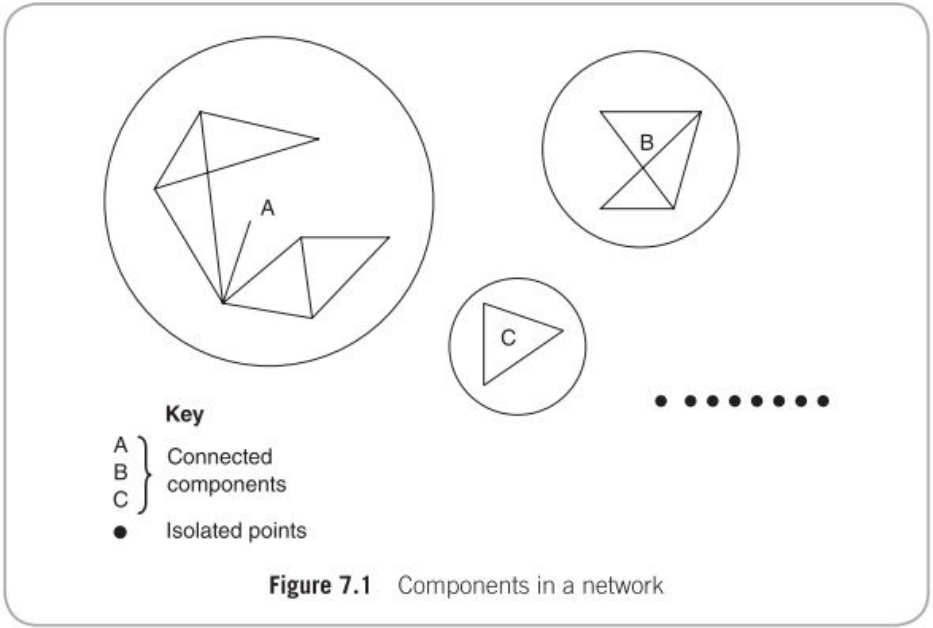



**Figure 7.1** Components in a network

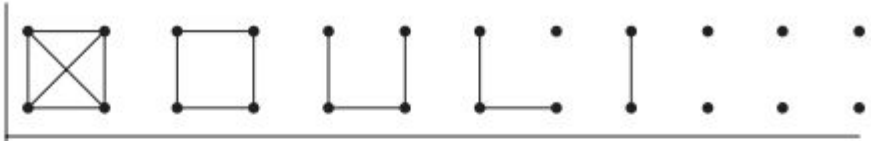
# Density

The density statistic represents the proportion of possible relationships in the network that are actually present.

The value ranges from 0 to 1, with the lower limit corresponding to networks with no relationships and the upper limit representing networks with all possible relationships.

**D (unorientated) = Total edges / Total possible edges = 2 * m / (n * (n-1))**

# Examples



| | | | | | | |
|---|---|---|---|---|---|---|
| No. of connected points | 4 | 4 | 4 | 3 | 2 | 0 |
| Inclusiveness | 1.0 | 1.0 | 1.0 | 0.7 | 0.5 | 0 |
| Sum of degrees | 12 | 8 | 6 | 4 | 2 | 0 |
| No. of lines | 6 | 4 | 3 | 2 | 1 | 0 |
| Density | 1.0 | 0.7 | 0.5 | 0.3 | 0.1 | 0 |

**Figure 5.4** Density comparisons

# The degree : properties at the level of the node/edges

Number of edges of a node (if weighted, the sum of the weights)

Degree is the most simple measure of the **centrality** of a node

The **degree distribution** is the probability distribution of these degrees over the whole network.

A well connected vertex is called a **hub**.

# Nodes/edges can have properties

- Weights/Categories for nodes
- Intensity/Categories for relations

Consequences on degree calculus / visualisation

- colors / size
- proximity

# Distance between 2 nodes

Number of edges between 2 nodes (geodesic)

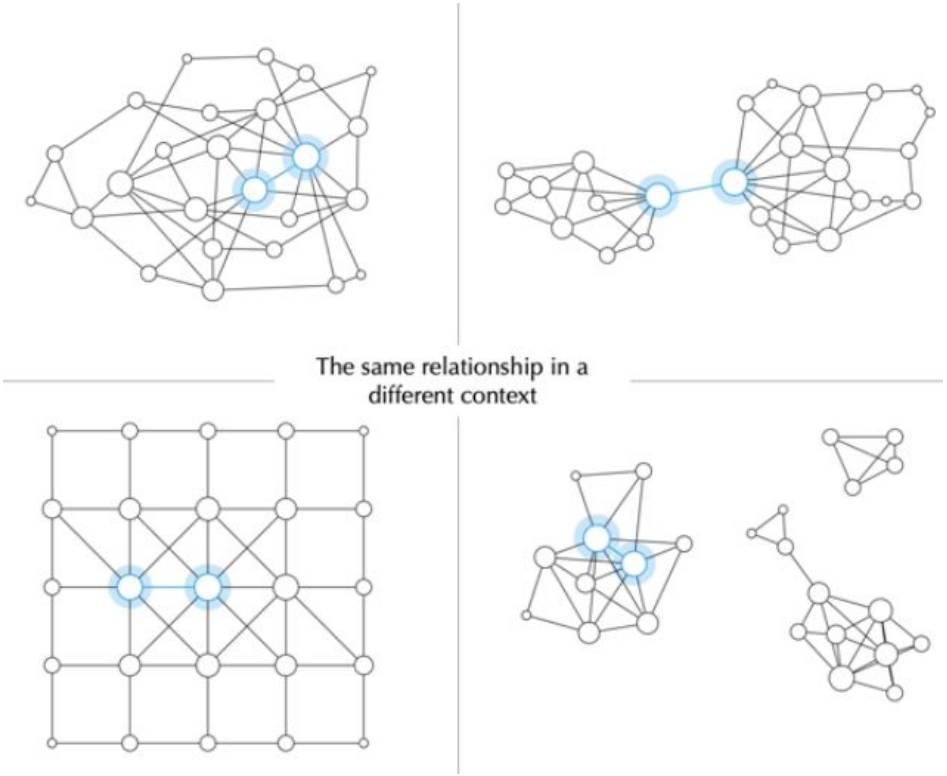
- 1 if connected
- 2 if there is a 2-edges path between the nodes
- …

# Networks have local properties

The importance of the context

(came number of nodes/edges)

Necessity of dedicated metrics



# The notion of clique

**Strict** : regions of the graph where all vertices are connected to each other and thus form groups of maximum density

**Flexible** (n-clique): a clique is a subset of a network in which the actors are more closely and intensely tied to one another than they are to other members of the network, with n the maximal distance between two nodes

Two specific clique:

- dyade
- tryade

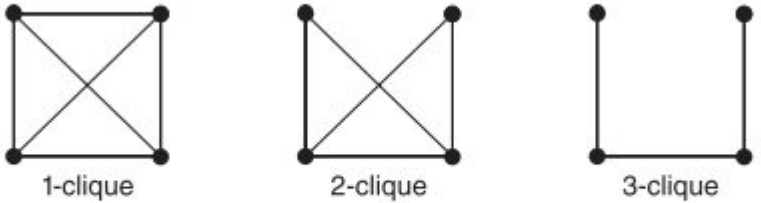



**Figure 7.11** *n*-cliques of size 4

# Clustering

a community/module is a group of nodes that connect to other groups in similar ways

There is different algorithm to gather nodes regarding their proximity. The most used : The Louvain method for community detection.

The question is always : how to select the parameters …

Journal of Statistical Mechanics: Theory and Experiment

## Fast unfolding of communities in large networks

Vincent D Blondel[1], Jean-Loup Guillaume[1,2], Renaud Lambiotte[1,3] and Etienne Lefebvre[1]

Published 9 October 2008 • IOP Publishing Ltd

Journal of Statistical Mechanics: Theory and Experiment, Volume 2008, October 2008

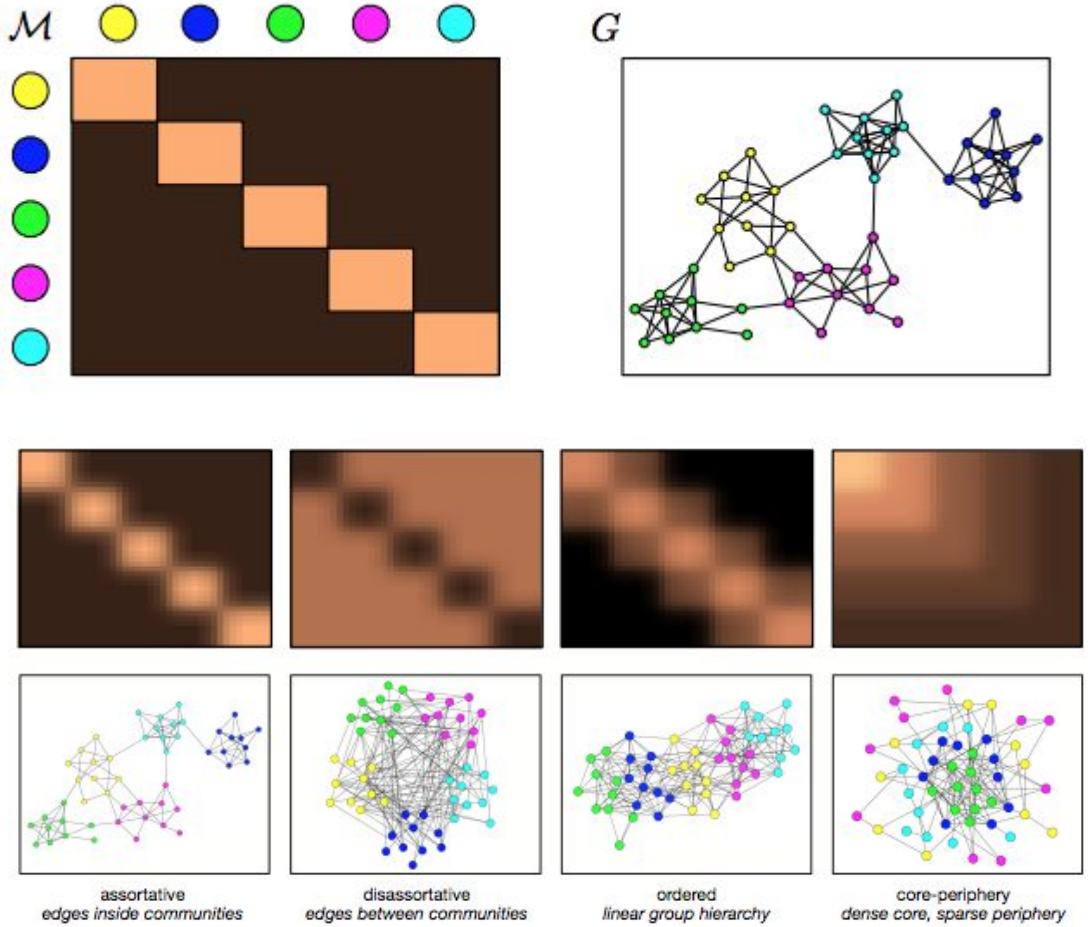
**Citation** Vincent D Blondel *et al J. Stat. Mech.* (2008) P10008

**DOI** 10.1088/1742-5468/2008/10/P10008

+ Article information

### Abstract

We propose a simple method to extract the community structure of large networks. Our method is a heuristic method that is based on modularity optimization. It is shown to outperform all other known community detection methods in terms of computation time. Moreover, the quality of the communities detected is very good, as measured by the so-called modularity. This is shown first by identifying language communities in a Belgian mobile phone network of 2 million customers and by analysing a web graph of 118 million nodes and more than one billion links. The accuracy of our algorithm is also verified on ad hoc modular networks.


$\mathcal{M}$
$G$
assortative
edges inside communities
disassortative
edges between communities
ordered
linear group hierarchy
core-periphery
dense core, sparse periphery

# Example of a practical use of SNA

Use of SNA to measure polarization on specific topics

Garimella K., Morales G.D.F., Gionis A., Mathioudakis M., 2018, « Quantifying Controversy on Social Media », ACM Transactions on Social Computing, 1, 1, p. 1-27.

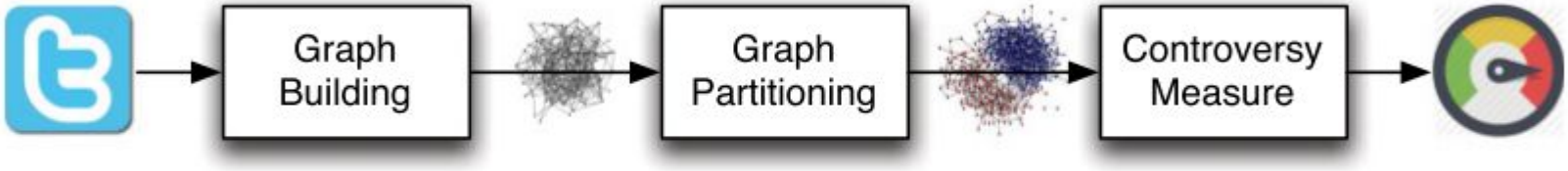



Fig. 1. Block diagram of the pipeline for computing controversy scores.

# Defining polarization

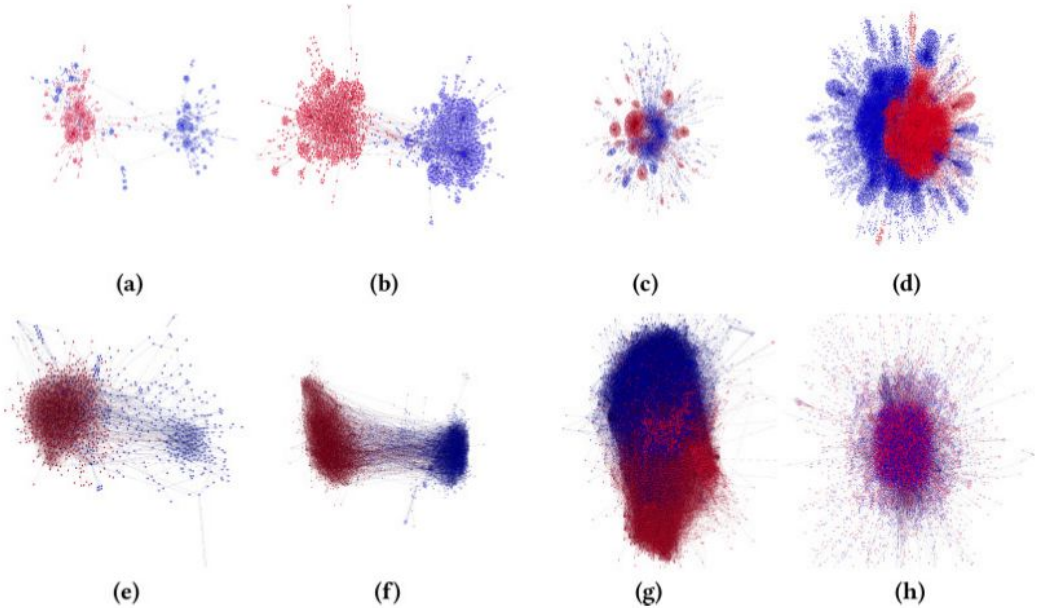


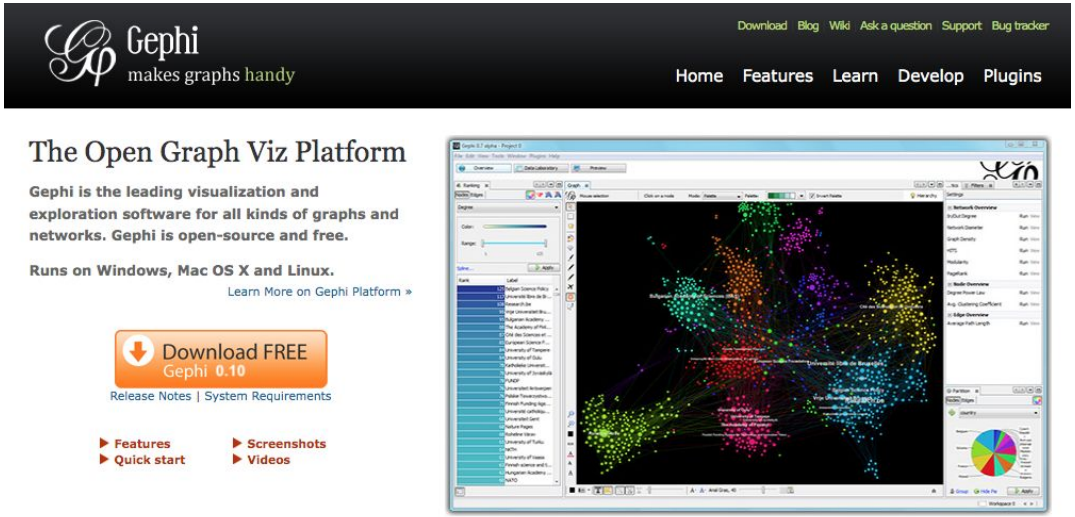

Fig. 3. Sample conversation graphs with retweet (top) and follow (bottom) aspects (visualized using the force-directed layout algorithm in Gephi). The left side is controversial, ((a) and (e)) #beefban and ((b) and (f)) #russia_march, while the right side is non-controversial, ((c) and (g)) #sxsw and ((d) and (h)) #germanwings. Only the largest connected component is shown.

# Bibliography


- L’analyse de réseau en sciences sociales. Petit guide pratique Laurent Beauguitte - https://hal.science/hal-04052709/document
- Introduction to Social Network Analysis: Basics and Historical Specificities. HNR+ResHist Conference 2021, Historical Network Research, 2021, Luxembourg, Luxembourg. ⟨10.5281/zenodo.5083036⟩. ⟨halshs-03351755⟩
  - Video: https://www.martingrandjean.ch/introduction-to-social-network-analysis/
- A complete course : https://aaronclauset.github.io/courses/5352/
- Scott J., 2017, Social network analysis.
- Zinoviev D., 2018, Complex Network Analysis in Python. Recognize → Construct → Visualize → Analyze → Interpret.
- Database of networks : https://icon.colorado.edu/#!/ ; http://snap.stanford.edu/data/index.html

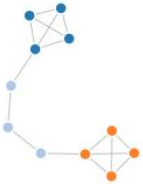
# Tools to analyze

- NodeXL (with Excel)
- Pajek (old)
- Gephi (state of the art)
- Python / R (a possibility)




NetworkX
Network Analysis in Python

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



## Software for complex networks

- Data structures for graphs, digraphs, and multigraphs
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g., text, images, XML records)
- Edges can hold arbitrary data (e.g., weights, time-series)
- Open source 3-clause BSD license
- Well tested with over 90% code coverage
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform