Parcours Intelligence Artificielle

Machine Learning for Network Modeling

Argyris Kalogeratos

name.surname@ens.paris-saclay.fr

MLMDA research group

Center Giovanni Borelli

ENS Paris-Saclay

Who’s talking?

Argyris Kalogeratos - kalogeratos@cmla.ens-cachan.fr

- **Researcher** at the **Center Giovanni Borelli***, ENS Paris-Saclay
- Background: Computer Science – Machine Learning
- Coordinating the “**Machine Learning for Graphs**” research theme at the MLMDA group

* ex. Center of Applied Mathematics – CMLA.

Why are we here?

Short course on **Machine Learning** for **Network** Modeling

Planning: 4 dense sessions, 2.5 hours each

1. Introduction to Graph Theory and Network Science
2. Network models - Static and dynamic graphs*
3. Structure and topology inference
4. Processes and signals over graphs

* Session 2 is going to be given by Fabian Tarissan, CNRS, ENS Paris-Saclay

fabien.tarissan@ens-paris-saclay.fr

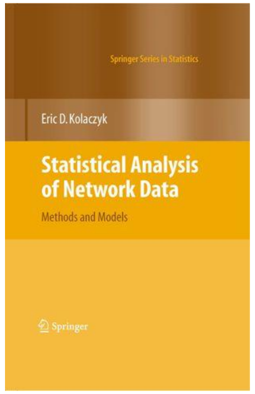
How we'll get through this?

Attend the courses

Do a short project

- It can be something around using the tools of the course for a problem of your main discipline or a thematic you'd like to pursue in the future
- The subject and perimeter of each project should be discussed
- Deliverables: report + codes (Matlab, R, Python, …)

Resources



Course's material

- *http://kalogeratos.com/psite/ai-ml-for-network-modeling/*

Books

- E.D. Kolaczyk (2009). *Statistical Analysis of Network Data: Methods and Models, Springer*, New York [download]
- M. Newman (2018). *Networks: An Introduction*, Oxford University Press
- A-L. Barabási (2016). *Network Science*. Cambridge University Press
- … material from our research in academia and industrial collaborations

Check also: https://github.com/briatte/awesome-network-analysis

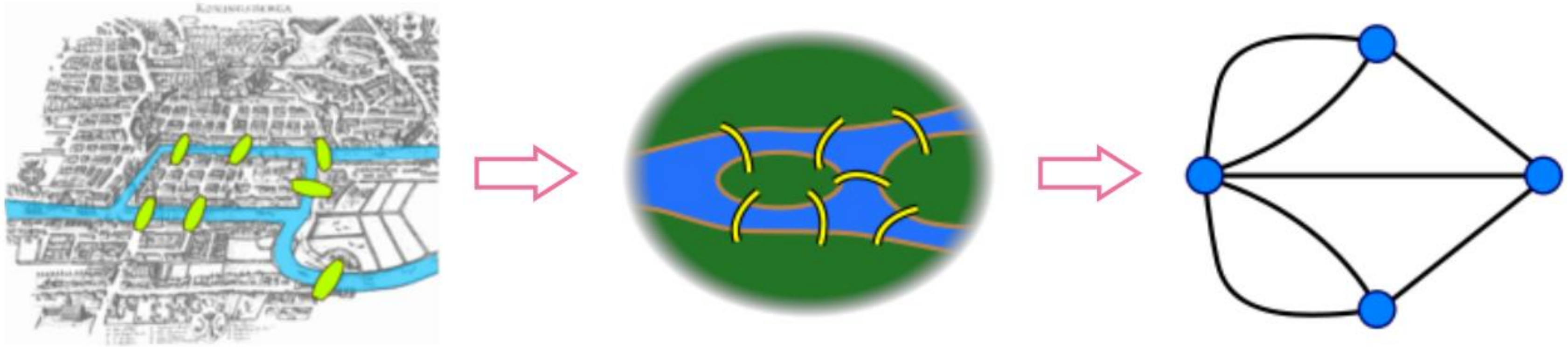
Introduction to Network Science and Graph Theory

.:: In this lecture

1. Motivation why to study networks
2. Why and how Statistics and Machine Learning can help
3. What’s behind Networks:
   Intro/review of graphs and related topics

Why Networks?

- Spoiler: behind networks there are **graphs**!!
- Graphs and graph theory come from the old days (recall Euler?)



The Königsberg Bridge Problem
[Leonhard Euler, 1736]

Why Networks?

- Until ~20 years ago, a field of study attracting mainly mathematicians
- Ever since, an increasing trend due to several reasons…
  - Simple models of `*reductionism*’ have been proven to be limiting to our view
  - Scientific tendency to find the right level of simplicity/complexity
  - System-level analysis has been gaining fans and space in science
  - Creation and storage of abundant and complex data in databases Exponential growth (recall Moore’s law)
  - Technological (and not only) globalization, Internet, Internet-of-Things, etc
- New terms: Network Science, Network (Data) Engineering, Graph-based ML

What is a ‘Network’?

Roughly… ***a collection of interconnected entities***

Entities of interest may be

- people
- species of the flora or fauna (e.g. plants, animals, …)
- organizations (e.g. states, airports, companies, …)
- computers (e.g. servers, mobile phones/PCs, sensors, …)
- geographic locations (e.g. places for weather forecasting)
- *… or generally interrelated variables of some multivariate environment/problem…*

What is a ‘Network’?

Roughly… ***a collection of interconnected entities***

We need to be careful as the term ‘network’ might be used to describe either one, two, or even all of the following

- the overall interconnected system (‘networked system’)
- the graph structure that represents that system
- and if a system evolves in time, ‘network’ might even imply that it is an object that encodes also the system’s time-varying nature…

What is then ‘network data’?

Roughly… ***a collection of interconnected entities***

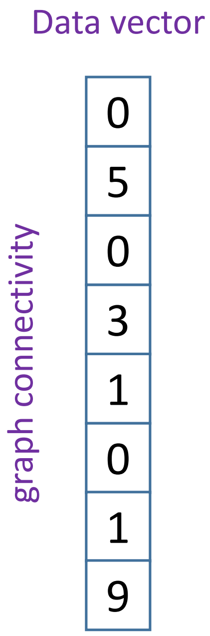
‘Networked system’ is a system conceptualized as a network

Then, ‘network data’ can be …

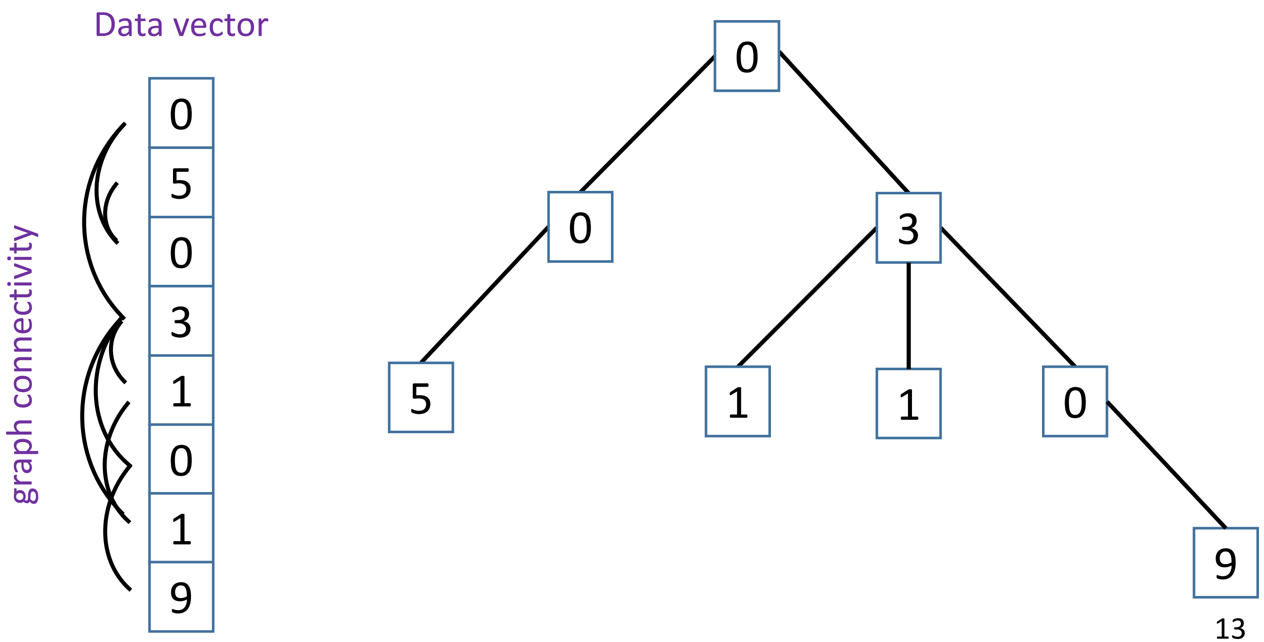
- either a set of measurements that describes the networked system (e.g. its organizational structure)
- or a set of measurements that come from the interconnected system itself,
- … (imagine we also have the time component)

Traditional vs Network-based methods

- Traditional methods see a set of individual variables (recall vectors)
- Network-based methods see interrelated variables depending on the 'structure' of the problem



vs



Where we aim at?

Conceptualize problems and systems as a networked environment

Modeling and statistical analysis of network data

ML and decision making (maybe interactive) in such environments

Some of the challenges

- the relations between entities give relational data
- sometimes (super) high-dimensional data and/or (super) big in size
- complex statistical dependencies
  - This is where special statistical methods and ML can give the lead

Examples of networks

The interest for a network-based perspective concerns broadly

- computational sciences
- humanities
- administration and management
- art!!

General application areas where we can see networks

- Technological
- Informational
- Social
- Biological

Example

Transportation

Air traffic network

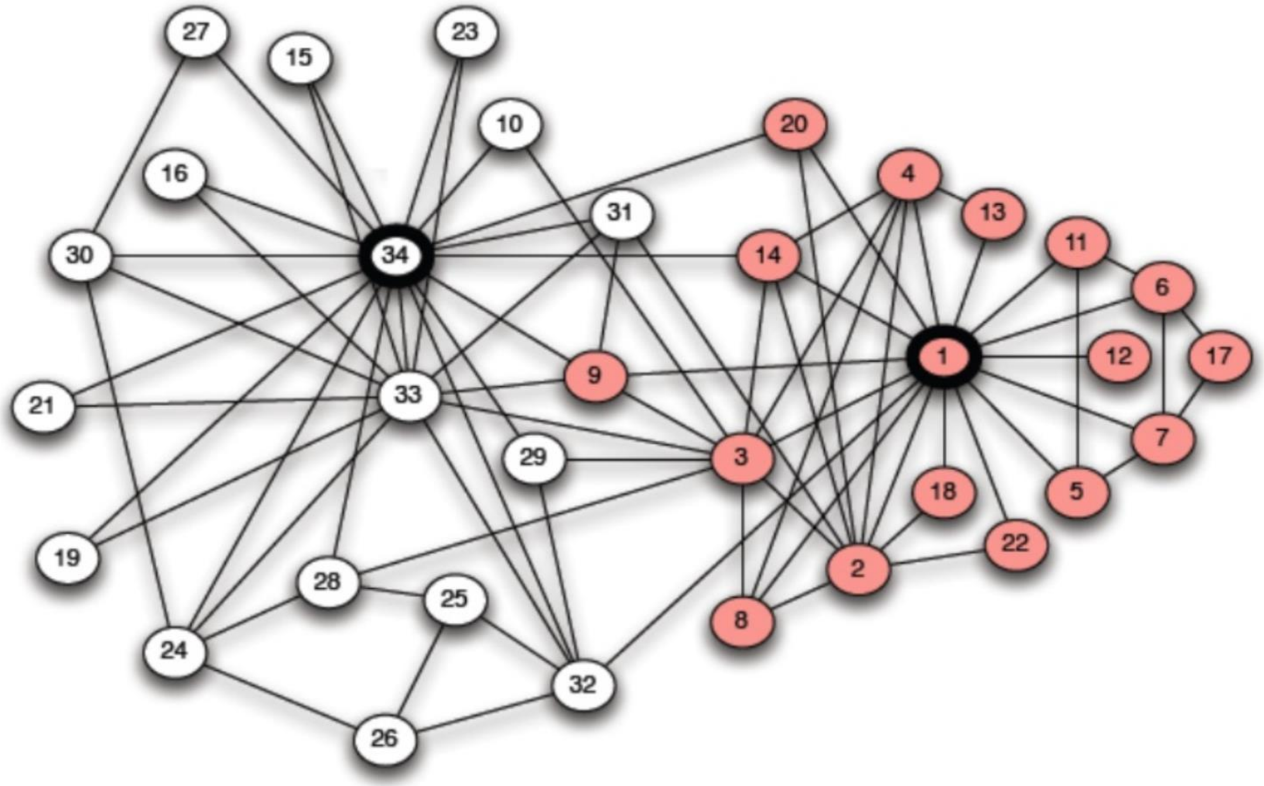
Brockmann et al. (2013). The Hidden Geometry of Complex, Network-Driven Contagion Phenomena, Science

Network inspection

Link Analysis vs Network Analysis

Qualitative vs Quantitative

Zachary's university karate club

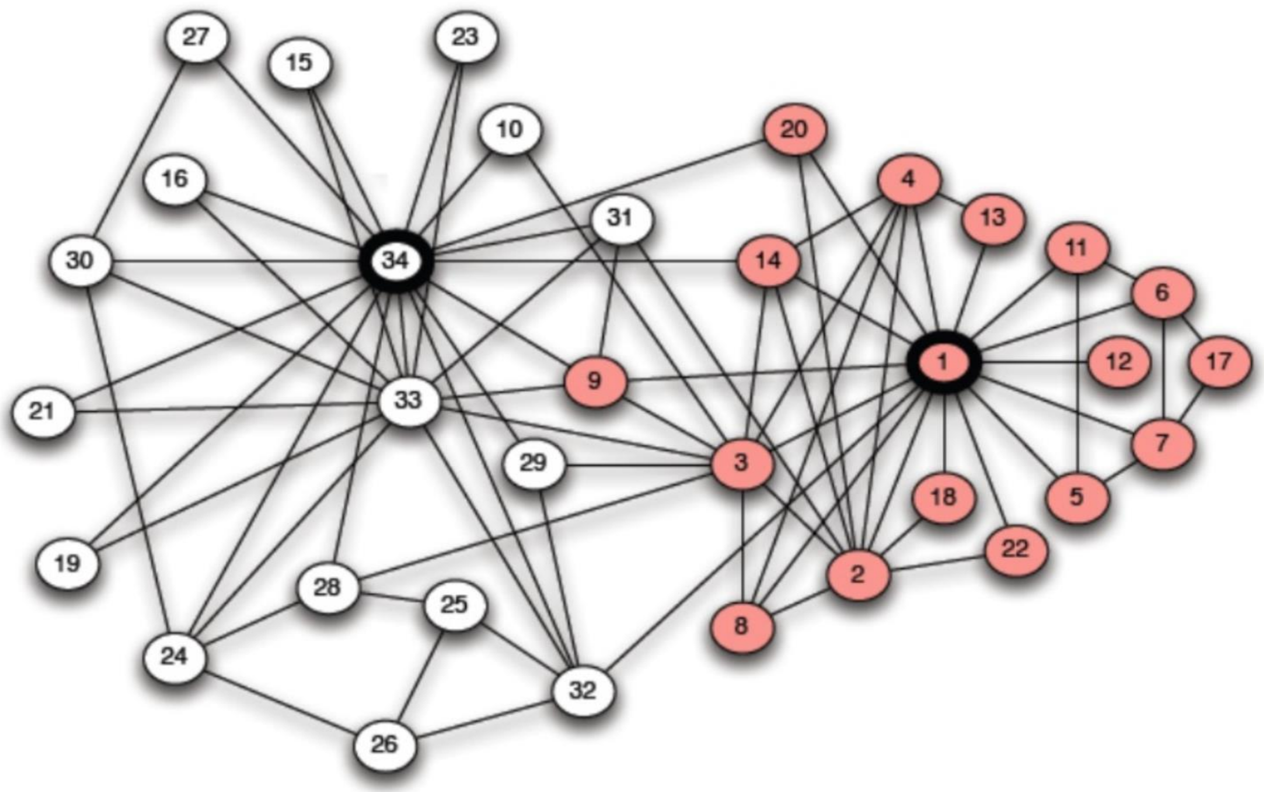


Network inspection

We can identify, visually and/or computationally, the roles of different graph nodes

- Edge density / connectivity
- Center vs Periphery
- Hub (or bridge) vs isolated nodes
- Communities
- Interface nodes between different communities
- k-core identification (subgraph of nodes that all have at least degree k)

Zachary's university karate club



Network inspection

For a large graph, which we may not be able to fully access, we are forced to work with 'samples'

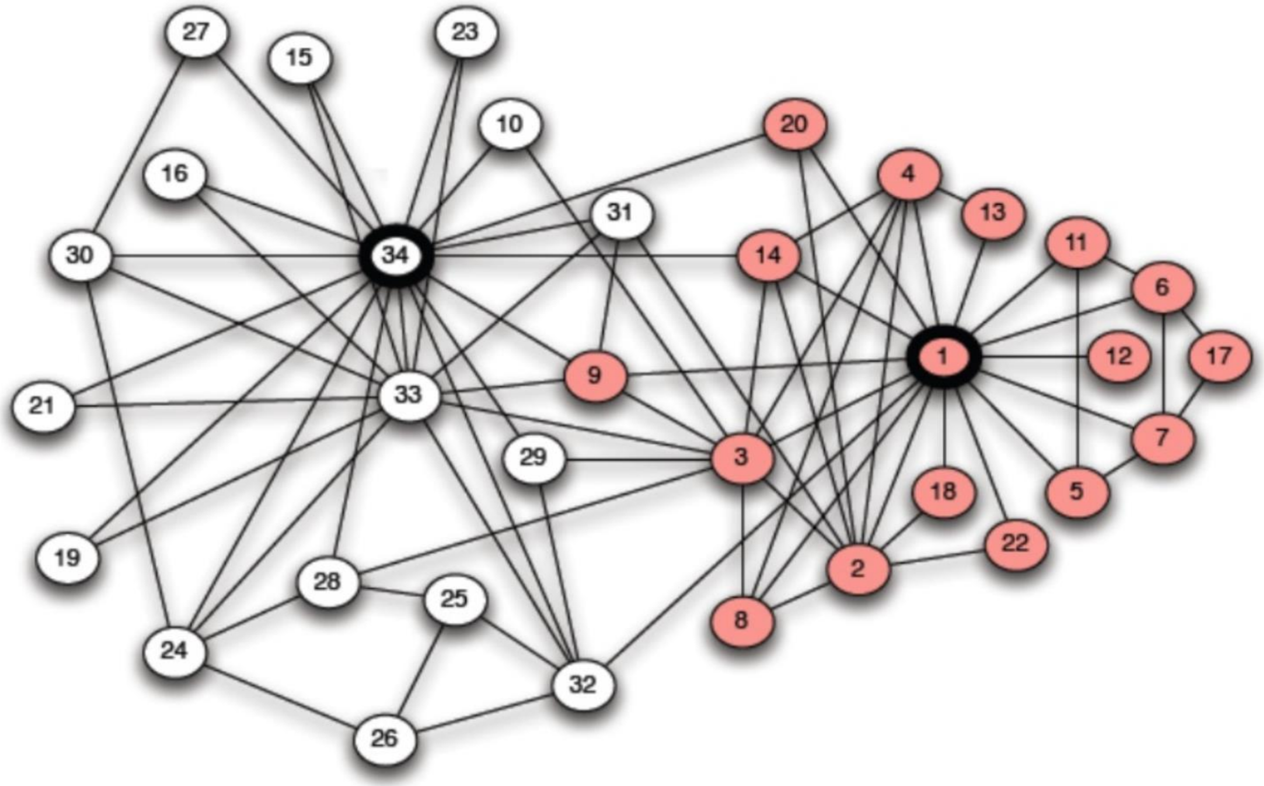
Sampling can be

- *passive* (we don't choose)
- *active* (we choose our sample)
- *active-corrective*

We need to know the statistical properties of the sampling scheme to deal with biases

Very challenging statistical problems!

Zachary's university karate club



Examples of questions on networked systems

How can I…

- visualize a network?
- extract features, and simplify its complex structure?
- compare two networks?
- realize the different roles of each node in the system?
- reveal functional attributes?
- reveal its vulnerabilities?
- evaluate its security against (static or dynamic) attacks?
- …

Examples of questions on networked systems

How can I…

- …
- take advantage of (and engineer on) its vulnerabilities or functional attributes for achieving a given goal?
- estimate the stability after dramatic changes in a network?
- monitor the system and detect events or outlying/erroneous behavior?
- automatically decode complex information using knowledge bases?
- predict next events of the evolution of a growing/changing graph in time?
- create random graph models that resemble real networks?
- …

Elements of Graphs and Graph Theory

Graphs

Network

A network provides a ‘structured’ space in which we can conceptualize and think of a problem/system!

Map

The analog is Geography and cartographic maps!

Graphs… are the basic mathematical models that allow analysis of networks

Next we will see

- Definition of a graph and concepts
- Graphs and matrix algebra
- Data structures for representing graphs and related algorithms

Definition of a graph

A graph $G = (V, E)$ is a mathematical structure of two sets:

- V containing *vertices* (or *nodes*)
- E containing connecting *edges* (or *links*), typically unordered pairs of vertices (u, v), with $u, v \in V$

Let $N_v = |V|$ and $N_e = |E|$ the size of these sets

Adjacency: two adjacent vertices have an edge connecting them
Analogously, adjacent edges have one mutual end vertex

Simple graph: no self-loops over a node, no parallel edges (multi-edges)

Variable terminology across domains

points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Direct vertex connectivity

non-connected

simply connected

one-way connected

two-way connected
(*reciprocity*)

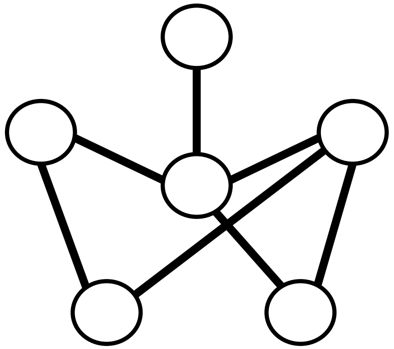
Subgraphs

A graph $g = (V_g, E_g)$ is a subgraph of another graph $G = (V, E)$ iff

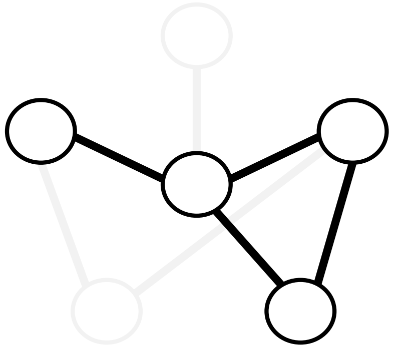
- $V_g \subseteq V$
- $E_g \subseteq E$

A graph $g = (V', E')$ is an induced subgraph of $G = (V, E)$ if

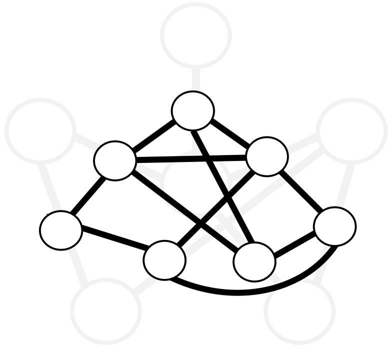
- first, a set of vertices $V' \subseteq V$ is given
- then, all edges connecting them are included in the $E' \subseteq E$



Graph

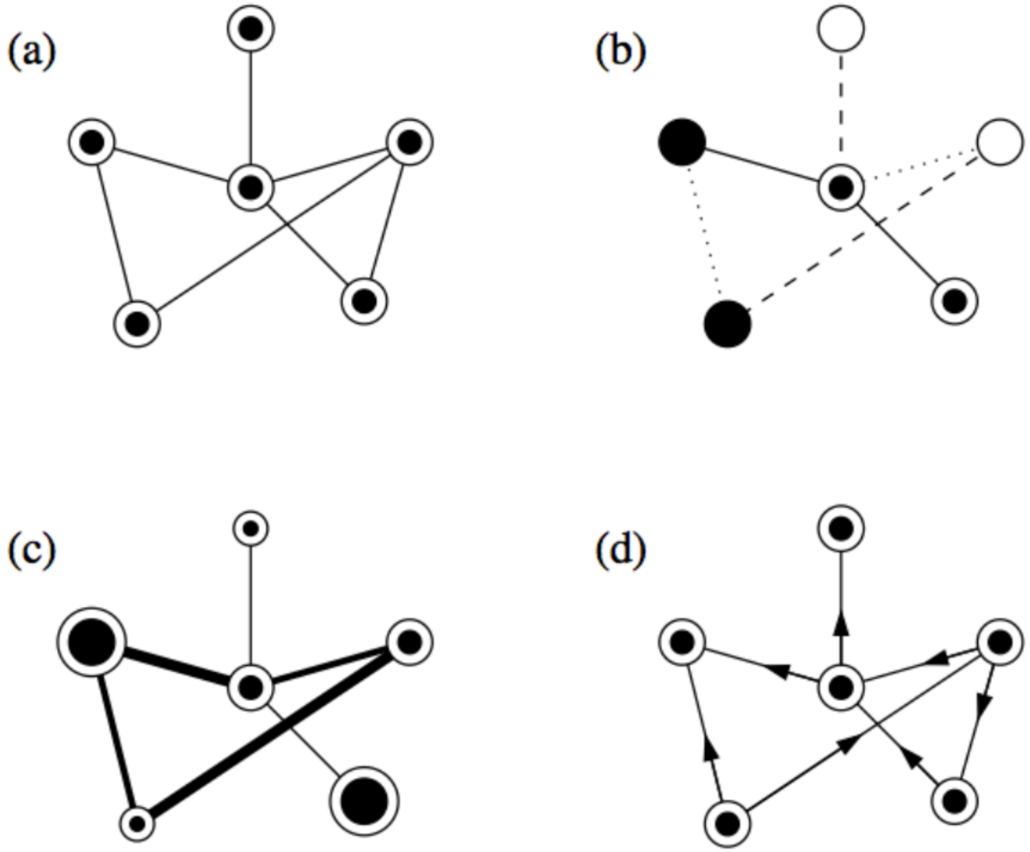


Subgraph



Dual graph

Types of graphs (some)



(a) unweighted, undirected

(b) discrete vertex and edge types, undirected

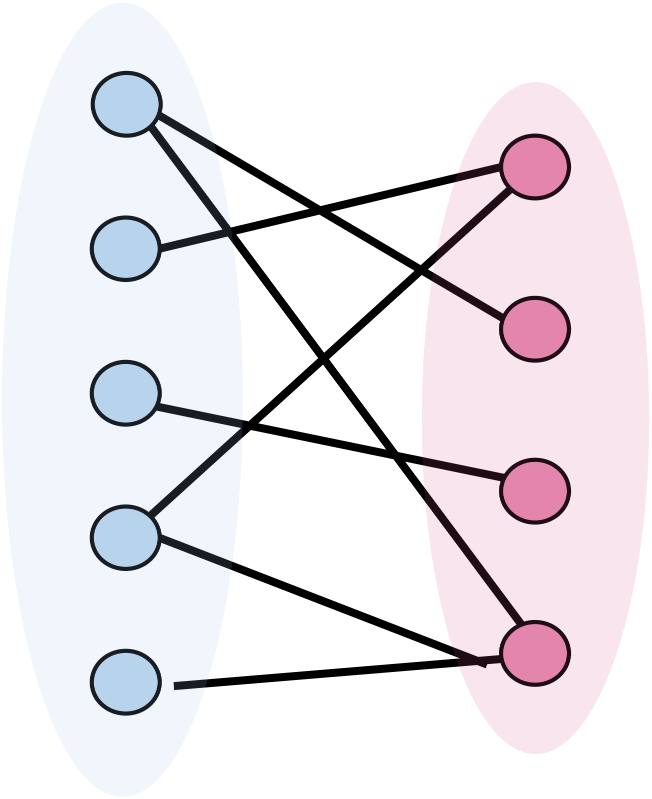
(c) varying vertex and edge weights, undirected

(d) directed

Source: Newman 2013

Types of graphs (some)

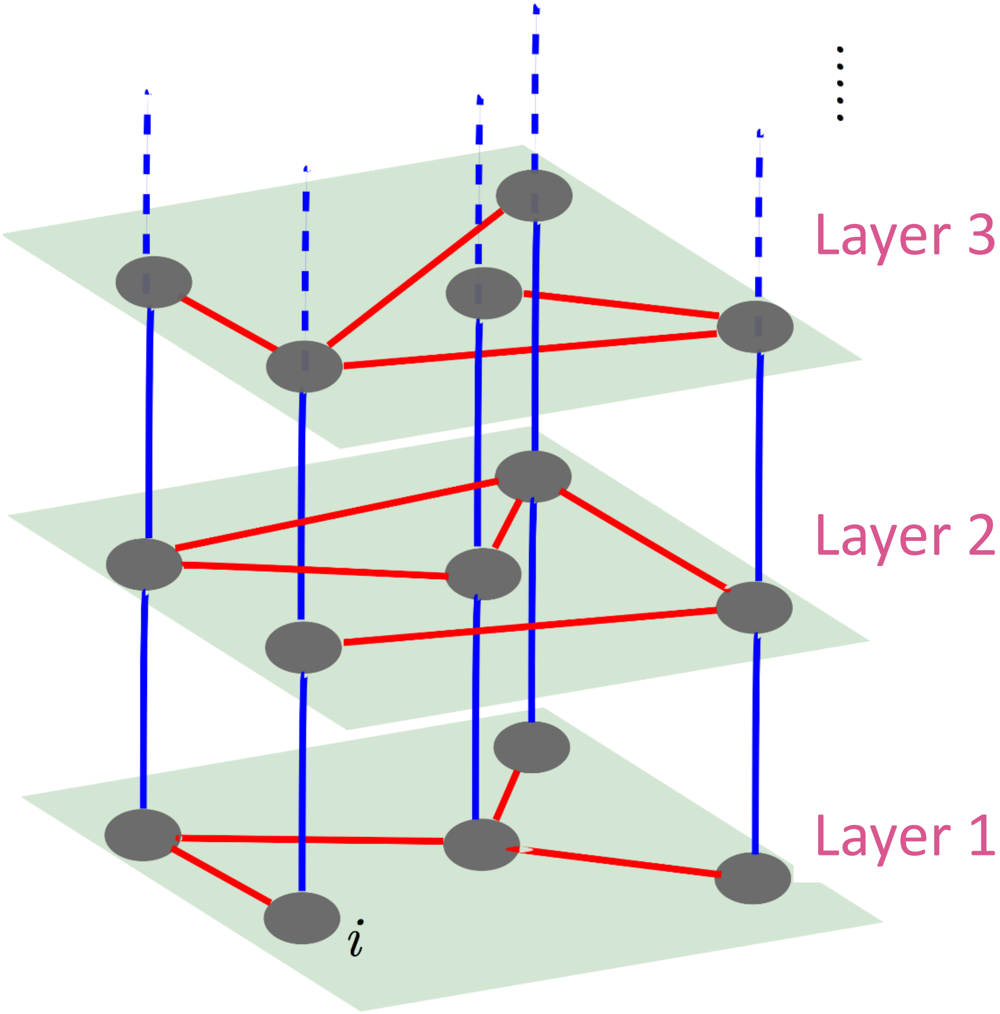
Bipartite graph



Part 1 Part 2

Types of graphs (some)

Multiplex graph



Degree

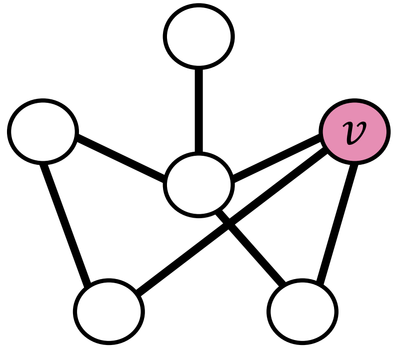
Degree d_v of a vertex v is the number of incident edges to it

The sum of the degrees: $\sum_{v=1}^{|N|} d_v = 2|E|$

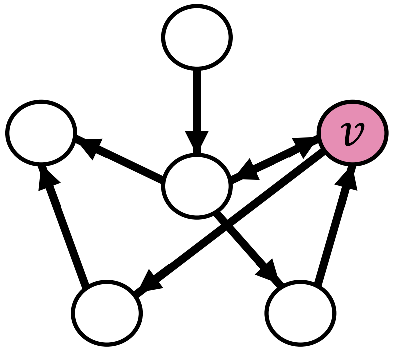
Degree sequence is the non-decreasing ordering of the all vertices' degrees in the graph: $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(|N|)}$

Degree distribution …

Directed graphs: we can define the in-degree and out-degree for a vertex



$d_v = 3$



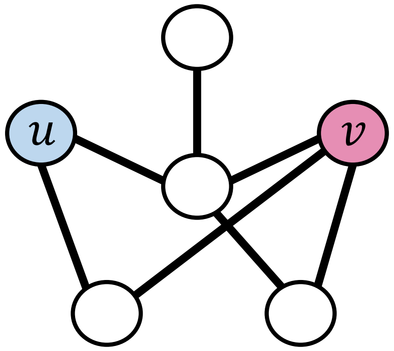
$d_v^{\text{in}} = 2$
$d_v^{\text{out}} = 3$

Movement / Reachability / Components

Walk on a graph from v_0 to v_l is any sequence $(v_0, e_0, v_1, e_1, \dots, v_{l-1}, e_{l-1}, v_l)$

There might exist several walks from v_0 to v_l

- Length of walk is l
- Trail is a walk without repeating edges
- Path is a walk without repeating vertices
- Circuit is trail that comes back to $v_0 = v_l$
- Cycle is path that comes back to $v_0 = vl$
- Distance from v_0 to v_l is the shortest path connecting them
- Diameter of a graph is the maximal distance between any pair of vertices



Graph

Movement / Reachability / Components

Vertex v is reachable from u if there is a path connecting them

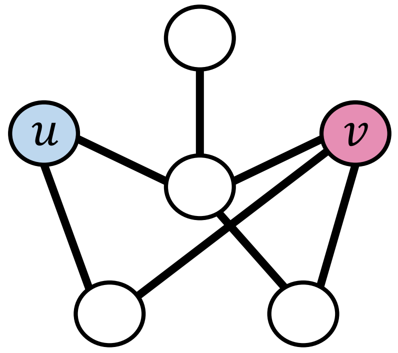
A graph is connected if all vertices are reachable to each other

A component is a maximally connected subgraph (also *strong* and *weak connectivity* for digraphs)

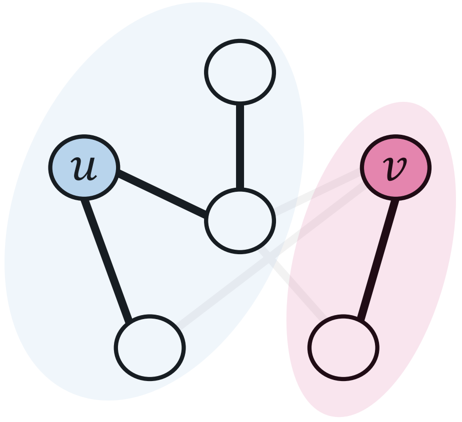
A regular graph's vertices have equal degree

A complete graph has $(N_e - 1)^2$ edges

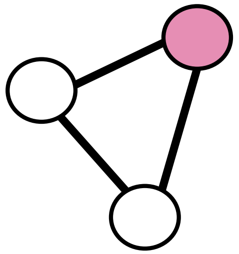
A clique is a complete graph of c vertices that is totally connected (i.e. complete)



Graph



2 components



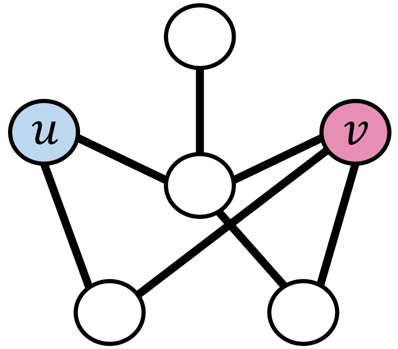
clique

Movement / Reachability / Components

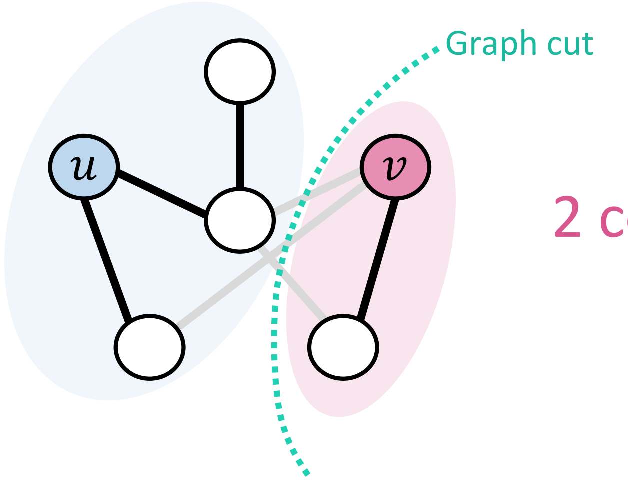
Graph cut is the set of edges that after removal they induce one or more disconnected components

A random walk starts from a vertex and follows randomly edges of the visited node (weighted random choices for weighted graphs)

- At the limit, during a very long random walk the frequency of visiting each vertex converges to a stationary distribution which is the degree distribution
- Random walks are very important tools



Graph



2 components

Adjacency matrix

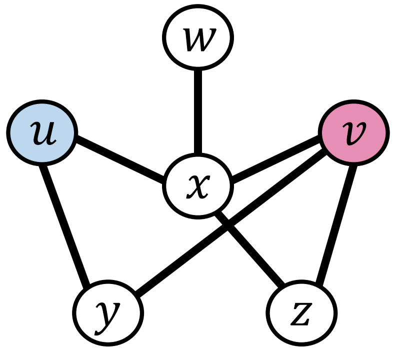
Representing the connectivity of a graph by a matrix is very convenient

Graph Theory + Matrix Algebra = …
… *Algebraic Graph Theory*

The adjacency matrix A of a graph G is a square binary matrix, (typically symmetric), where

$$A_{ij} = f(x) = \begin{cases} 1, & if\ \{i,j\} \in E \\ 0, & \text{otherwise} \end{cases}$$

- The non-zero weights can be bigger than 1 for weighted graphs
- Symmetricity does not generally hold for digraphs



Graph

	u	v	w	x	y	z
u	0	0	0	**1**	**1**	0
v	0	0	0	**1**	**1**	**1**
w	0	0	0	**1**	0	0
x	**1**	**1**	**1**	0	0	**1**
y	**1**	**1**	0	0	0	0
z	0	**1**	0	**1**	0	0

Adjacency matrix

Adjacency matrix

Various operations on matrix A

- Degree: $d_i = \sum_{j>i} A_{ij}$
- Number of walks: $(A^r)_{ij} = (A\,A \ldots A)_{ij}$
- Eigen-structure:

$$A\mathrm{v}_1 = \lambda_1 \mathrm{v}_1$$
$$A\mathrm{v}_2 = \lambda_2 \mathrm{v}_2$$
$$\ldots$$
$$A\mathrm{v}_{N_v} = \lambda_{N_v} \mathrm{v}_{N_v}$$

λ_{N_v} is always zero

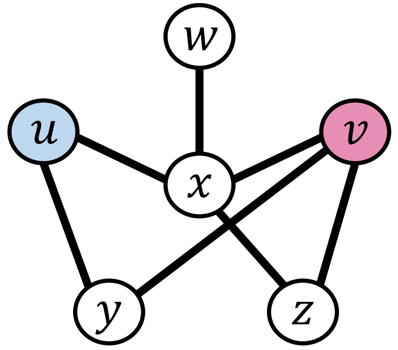
where v is an eigenvector and λ an associated eigenvalue in each case

The ordering of the eigenvalues exhibits also important properties

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_v}$$

Spectral radius: $\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_{N_v}|\}$

Example with paths



Graph

Adjacency matrix A

	u	v	w	x	y	z
u				1	1	
v				1	1	1
w				1		
x	1	1	1			1
y	1	1				
z		1		1		

Adjacency matrix A^2

	u	v	w	x	y	z
u	2	2	1			1
v	2	3	1	1		1
w	1	1	1			1
x		1		4	2	1
y				2	2	1
z	1	1	1	1	1	2

Adjacency matrix A^3

	u	v	w	x	y	z
u		1		6	4	2
v	1	2	1	7	5	4
w		1		4	2	1
x	6	7	4	2	1	5
y	4	5	2	1		2
z	2	4	1	5	2	2

Complexity of Matrix multiplication: N^3 = (M^2 * N), M are the edges

Graph Laplacian

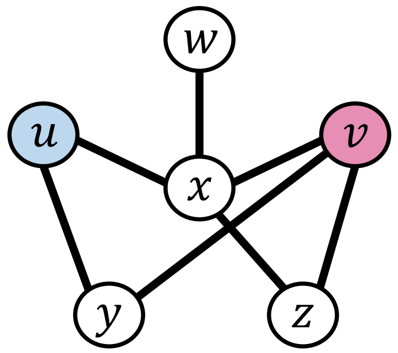
There are several definitions for the Laplacian matrix, the most common is the following N_v x N_v

$$L = D - A$$

where D is a matrix with the degrees of the graph vertices in its diagonal, and zero values everywhere else

Graph Laplacian

Graph



Laplacian matrix L

	u	v	w	x	y	z
u	2			-1	-1	
v		3		-1	-1	-1
w			1	-1		
x	-1	-1	-1	4		-1
y	-1	-1			2	
z		-1		-1		2

=

Degree matrix D

	u	v	w	x	y	z
u	2					
v		3				
w			1			
x				4		
y					2	
z						2

−

Adjacency matrix A

	u	v	w	x	y	z
u				1	1	
v				1	1	1
w				1		
x	1	1	1			1
y	1	1				
z		1		1		

Graph Laplacian

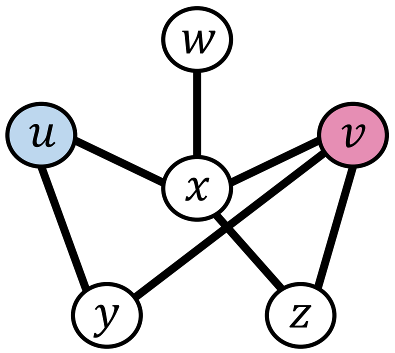
Basic properties

- Zero-sum rows and columns ==> zero-sum matrix L
- All negative values except in the diagonal
- Same off-diagonal zeros as A has information only for the directly connected pairs of vertices
- The input graph G cannot be a multigraph, edge weights are ignored (but could be considered)
- Like in multivariate calculus, for (a problem-specific) $x \in \mathrm{R}^{N_v}$ that comes from some function $f(G, \dots)$,

$$x^{\mathrm{T}} L x = \sum_{(i,j)\in E} (x_i - x_j)^2 = \sum_{i,j \in [1,\dots,N_v]} A_{ij} (x_i - x_j)^2$$

The closest $x^{\mathrm{T}} L x$ is to zero, the more smooth the x is with respect to the graph, and so for the f

Graph

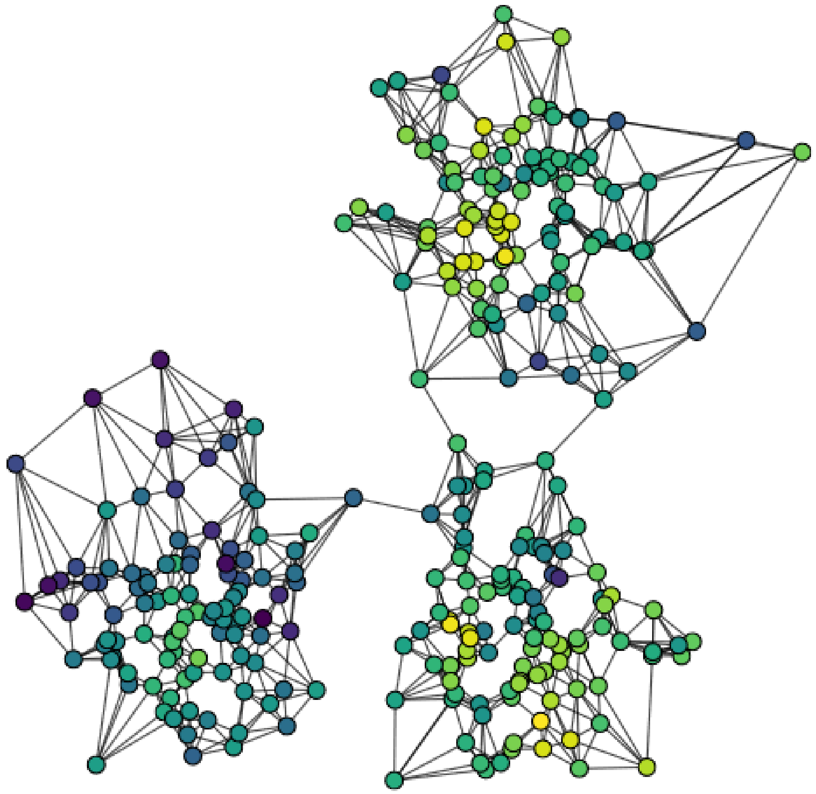


Laplacian matrix L

	u	v	w	x	y	z
u	2			-1	-1	
v		3		-1	-1	-1
w			1	-1		
x	-1	-1	-1	4		-1
y	-1	-1			2	
z		-1		-1		2

Graph Laplacian

Graph signals with increased smoothness (a) to (c)



(a) (b) (c)

Example of smoothness

Waldo Tobler's First Law of Geography:

"Everything is related to everything else, but near things are more related than distant things."

- It is the foundation of the fundamental concepts of spatial dependence and spatial autocorrelation
- It is the fundamental assumption used in all spatial analysis

Graph Laplacian

Alternative definition: the normalized Laplacian matrix

$$
\begin{aligned}
\tilde{L} &= D^{-\frac{1}{2}}\, L\, D^{-\frac{1}{2}} \\
&= D^{-\frac{1}{2}}\,(D - A)D^{-\frac{1}{2}} \\
&= (D^{-\frac{1}{2}}D - D^{-\frac{1}{2}}A)\, D^{-\frac{1}{2}} \\
&= D^{-\frac{1}{2}}DD^{-\frac{1}{2}} - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \\
&= I\; - D^{-\frac{1}{2}}\, A\, D^{-\frac{1}{2}}
\end{aligned}
$$

- More appropriate when there is degree inhomogeneity
- Constraints the eigenvalues in [0, 2]

Eigen-analysis of graph Laplacian

Generally, L's eigenvalues and eigenvectors yield a lot of interesting information for a graph regarding

- G's connectivity
  - The smallest eigenvalue is 0 with eigenvector **1**
  - If the second smallest eigenvalue is 0 then the graph is disconnected The multiplicity of that gives the # components
  - A connected graph of diameter δ has at least $\delta + 1$ distinct eigenvalues
  - The larger its non-trivial eigenvalues are, the more connected a graph is
- G's conductance (how fast does a random walk converge)
- …

Eigen-analysis of graph Laplacian

G's conductance or Cheeger constant … (how fast does a random walk converge)

- The conductance of a $\text{cut}(S, S')$, where $S \cap S' = \emptyset$ different node sets

$$c(S) = \frac{\sum_{i \in S, j \in S'} A_{ij}}{\min(A(S), A(S'))}$$

Where A_{ij} is the element of the adjacency matrix, and $A(S) = \sum_{i \in S, j \in V} A_{ij}$

- The conductance of all graph is the minimal for all possible sets S

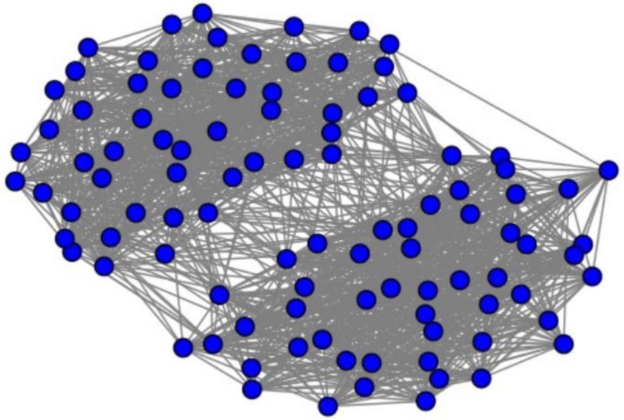
$$c(G) = \min_{S \in G} c(S)$$

Eigen-analysis of graph Laplacian

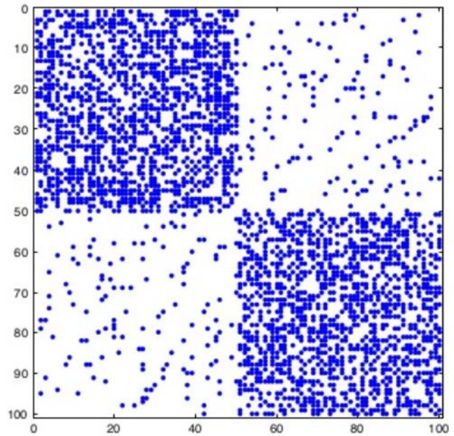
In the presence of graph structure of k communities

- the $\mathrm{rank}(A) \approx k$ and is related to the number of clusters/communities
- A can be rearranged (both row- and column-wise) in a way that a block-diagonal structure is revealed
- Each eigenvalue is associated with a different cluster, and the largest difference of eigenvalues (eigengap) can be measured between $\lambda_k - \lambda_{k+1}$

Community structure



B.D. Adjacency matrix



Data structures and algorithms

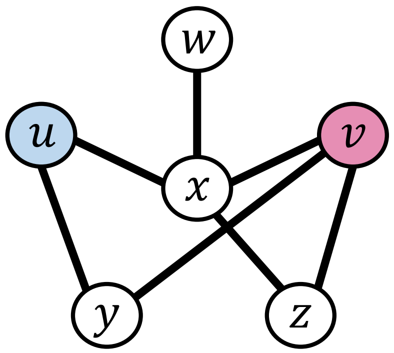
How we can process, manipulate, store, query graphs in practice?

Contributions mainly from Computer Science

Three main data structures

- Adjacency matrix (as we saw that earlier)
- Adjacency list
- Edge list
  - 2-column list with the vertices of all edges

Graph

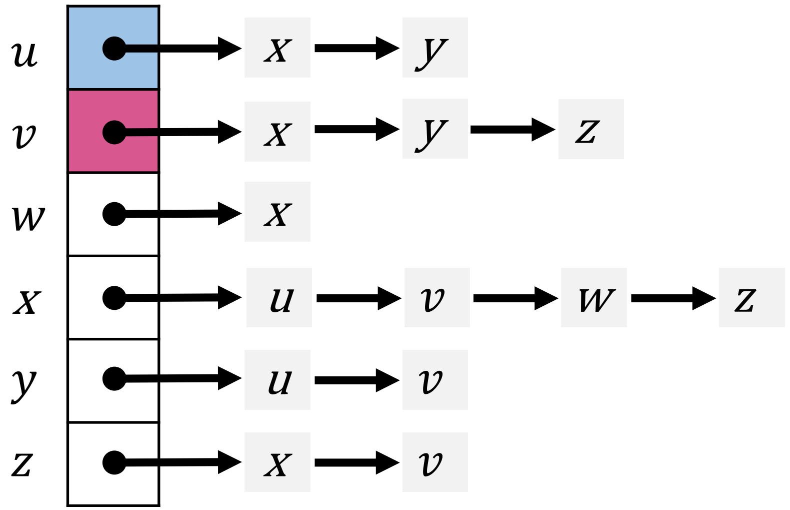


Edge list

u	x
u	y
v	x
v	y

…

Adjacency list



Data structures and algorithms

Space required to store a graph G with N_v vertices and N_e edges

- Adjacency matrix: $\mathrm{O}(N_v^2)$
- Adjacency list: $\mathrm{O}(N_v + N_e)$
- Column list: $\mathrm{O}(N_e)$

where $\mathrm{O}(\,.\,)$ measures order of magnitude

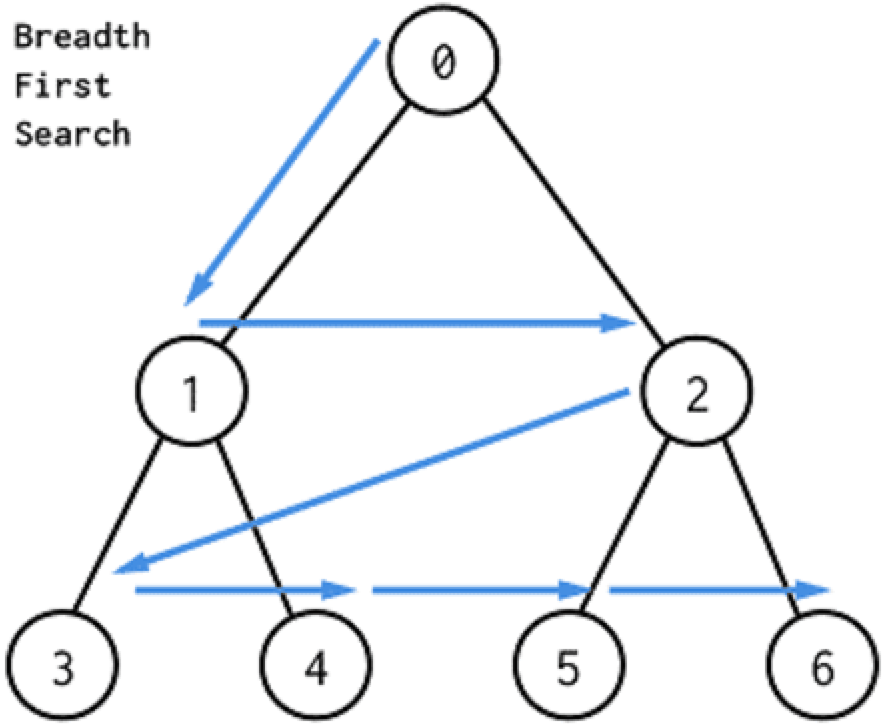
For sparse graphs there is a big reduction of storage requirements, but also algorithms get much faster!

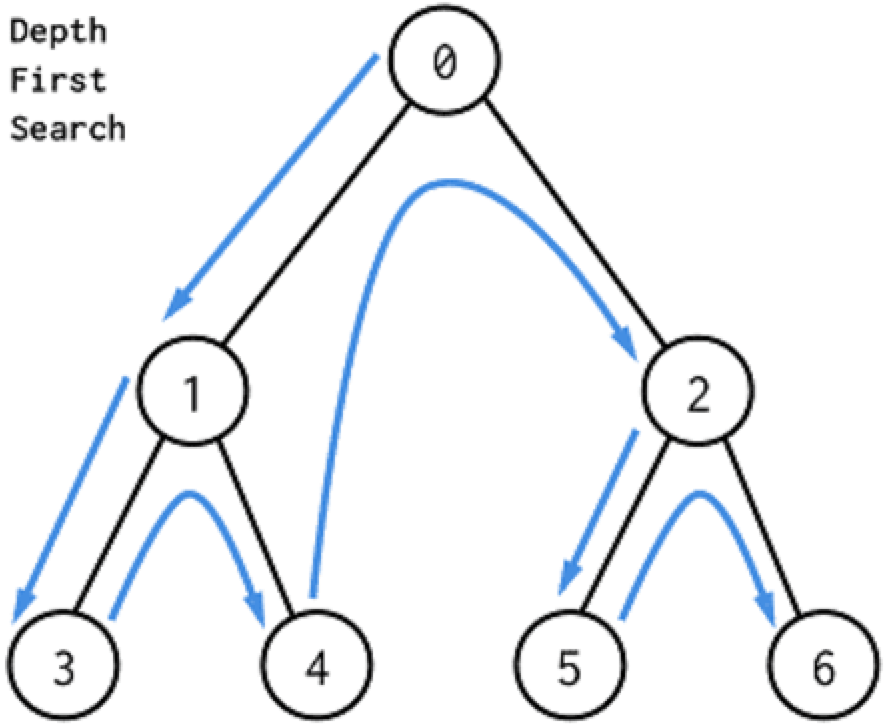
The data structure can be also chosen having in mind how it will affect the computational tasks we want to perform on the graph

Data structures and algorithms

Suppose a problem statement: *annotate the nodes of a given network*

Solutions: Breadth-First Search (BFS)   vs   Depth-First Search (DFS)

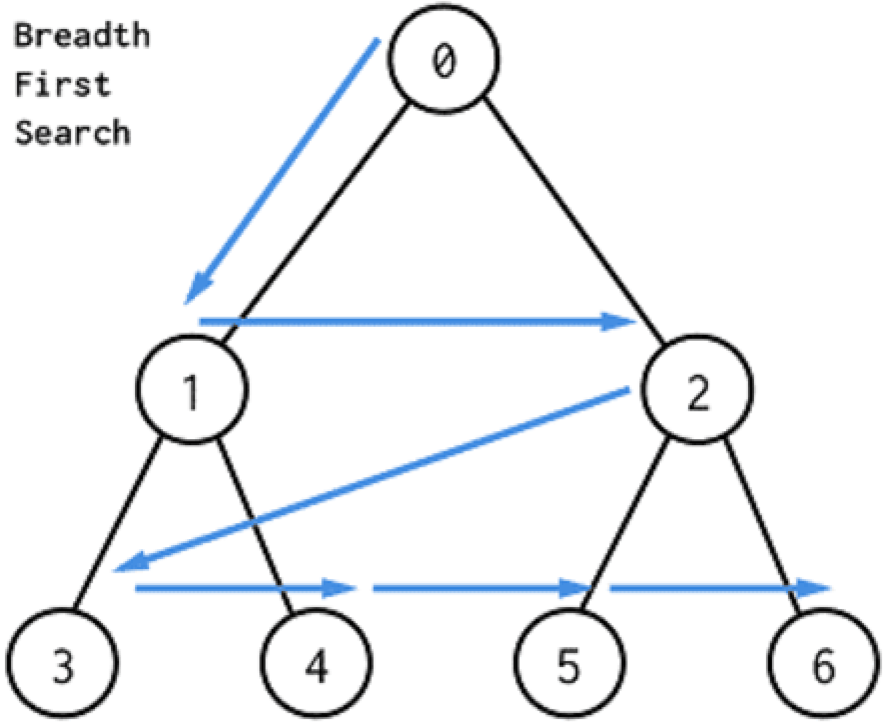




Data structures and algorithms

Suppose a problem statement: *annotate the nodes of a given network*

Breadth-First Search (BFS)



Implementation with a **queue** data structure (*First-in-First-Served)*

Queue: a list where *insert* (enqueue) goes to the front and *removal* (dequeue) takes an element from the front

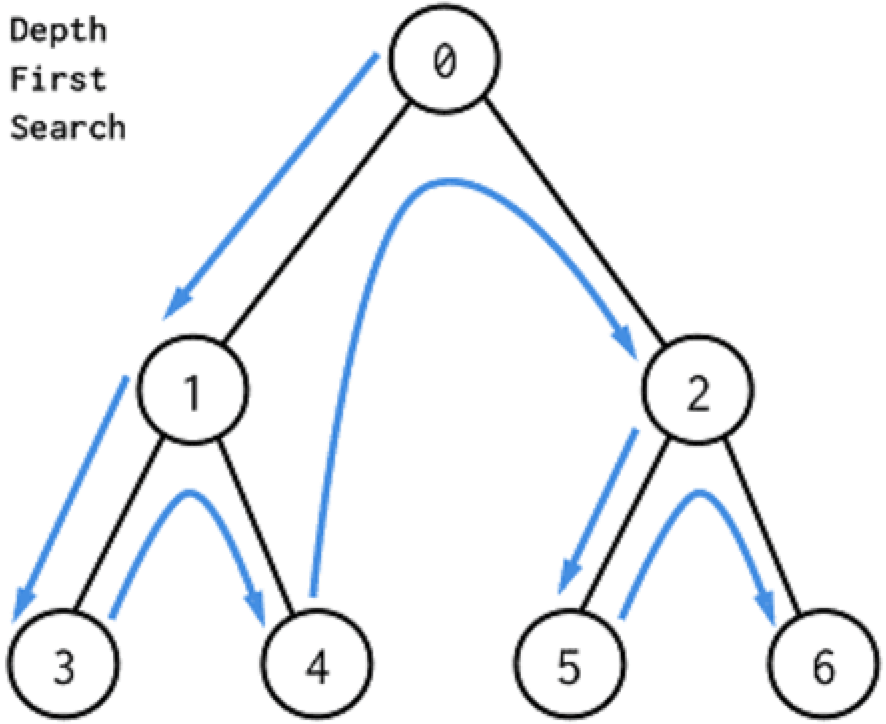
step 0: Insert 0 to empty Q	Q = { 0 }
step 1: Dequeue and insert children	Q = { ~~0~~, 1, 2 }
step 2: Dequeue and insert children	Q = { ~~1~~, 2, 3, 4 }
step 4: Dequeue and insert children	Q = { ~~2~~, 3, 4, 5, 6 }
step 5: Dequeue	Q = { ~~3~~, 4, 5, 6 }
step 6: Dequeue	Q = { ~~4~~, 5, 6 }
step 7: Dequeue	Q = { ~~5~~, 6 }
step 8: Dequeue	Q = { ~~6~~ }
	Q = empty

Output: 0, 1, 2, 3, 4, 5, 6

Data structures and algorithms

Suppose a problem statement: *annotate the nodes of a given network*

Depth-First Search (DFS)



Implementation with a **stack** data structure and a table noting which nodes have been **visited** (*First-in-Last-Served)*

Stack: a list where both *insert* (push) and *removal* (pop) operate on the front (top) of the list

step 0: Push 0 to empty S	S = { 0 }
step 1: Pop and re-push, push left child	S = { ~~0~~, 1, 0 }
step 2: Pop and re-push, push left child	S = { ~~1~~, 3, 1, 0 }
step 4: Pop	S = { ~~3~~, 1, 0 }
step 5: Pop and push next child	S = { ~~1~~, 4, 0 }
step 6: Pop	S = { 4, 0 }
step 7: Pop and re-push, push left child	S = { ~~0~~, 2 }
step 8: …	S = { ~~2~~, 5, 2 }
	S = { ~~5~~, 2 }
	S = { ~~2~~, 6 }
Output: 0, 1, 3, 4, 2, 5, 6	S = { ~~6~~ }
	S = empty

Data structures and algorithms

There are queries of variable difficulty. These are easy or doable

- Are two vertices i, j connected?

  Check $A(i,j)$; or traverse the $adjlist(i)$ to find j
- Compute the degree $d(i)$

  sum the i-th row or column of A; or measure the length of $adjlist(i)$
- Which is the shortest path between vertices i, j?

  Dikstra's algorithm finds all shortest paths from vertex i to all other vertices in $O(N_v^2 \log N_v + N_v N_e)$ time
- Identify connected components?

  DFS/BFS algorithms $\mathrm{O}(N_v + N_e)$ time
- Find the Minimum Spanning Tree

  Prim's algorithms which is $\mathrm{O}(N_e \log N_v)$

Data structures and algorithms

These queries are very hard or infeasible (i.e. NP-hard) for large graphs and in fact is where Machine Learning can help with approximations

- Find the maximal clique?
- Segment the graph in k parts in a way that minimizes the number of cross-edges between the parts
- Find a subgraph of G that is isomorfic to a given query graph
- Graph matching: best match two graphs
- Compute the similarity between two graphs by finding the optimal correspondance among their vertices
- Color a graph with minimal number of different colors, in a way that no adjacent vertex has the same color
- Layout problems for visualizing complex graphs

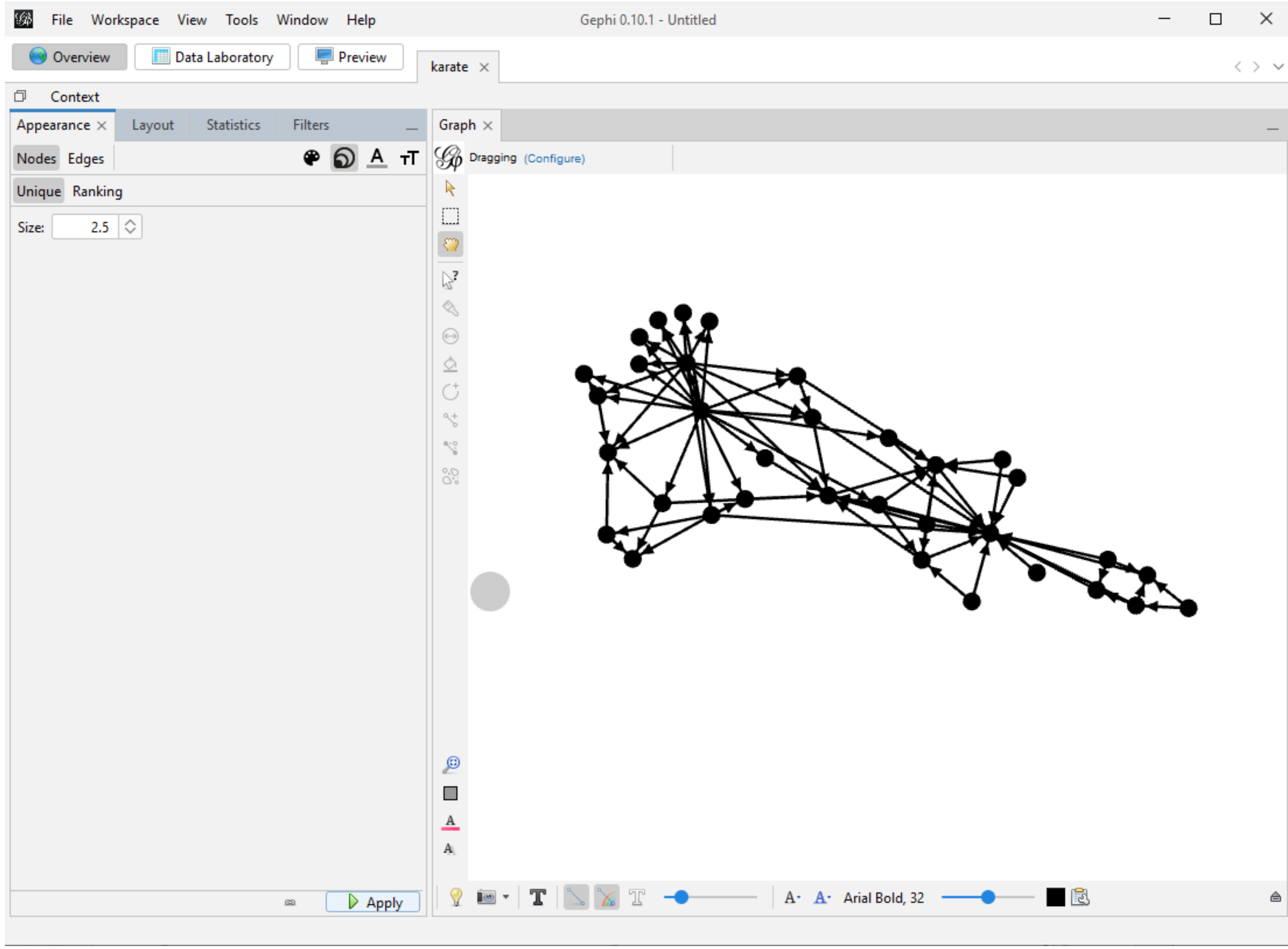
Vizualizing a graph

- Several algorithms apply heuristics to visualize graphs in the 2D space
- Force-directed graph drawing algorithms are a class of such algorithms that are based on attraction (and possibly repulsion) forces that tend to bring closer in the space a pair of nodes with high connection weight
- They initialize randomly the node positions and then they operate iteratively.
- They are useful and intuitive, but also non-deterministic and slow, …

- Suggested tool: **Gephi**

Let’s check an example

- Download Gephi from https://gephi.org/
- Download an example graph (e.g. Zachary’s karate club) from https://github.com/gephi/gephi/wiki/Datasets
- Produce a vizualization using a force-based algorithm
- Modify the appearance



Examples of Real Networks

Technological networks

Transportation, communication, sensor networks, energy …

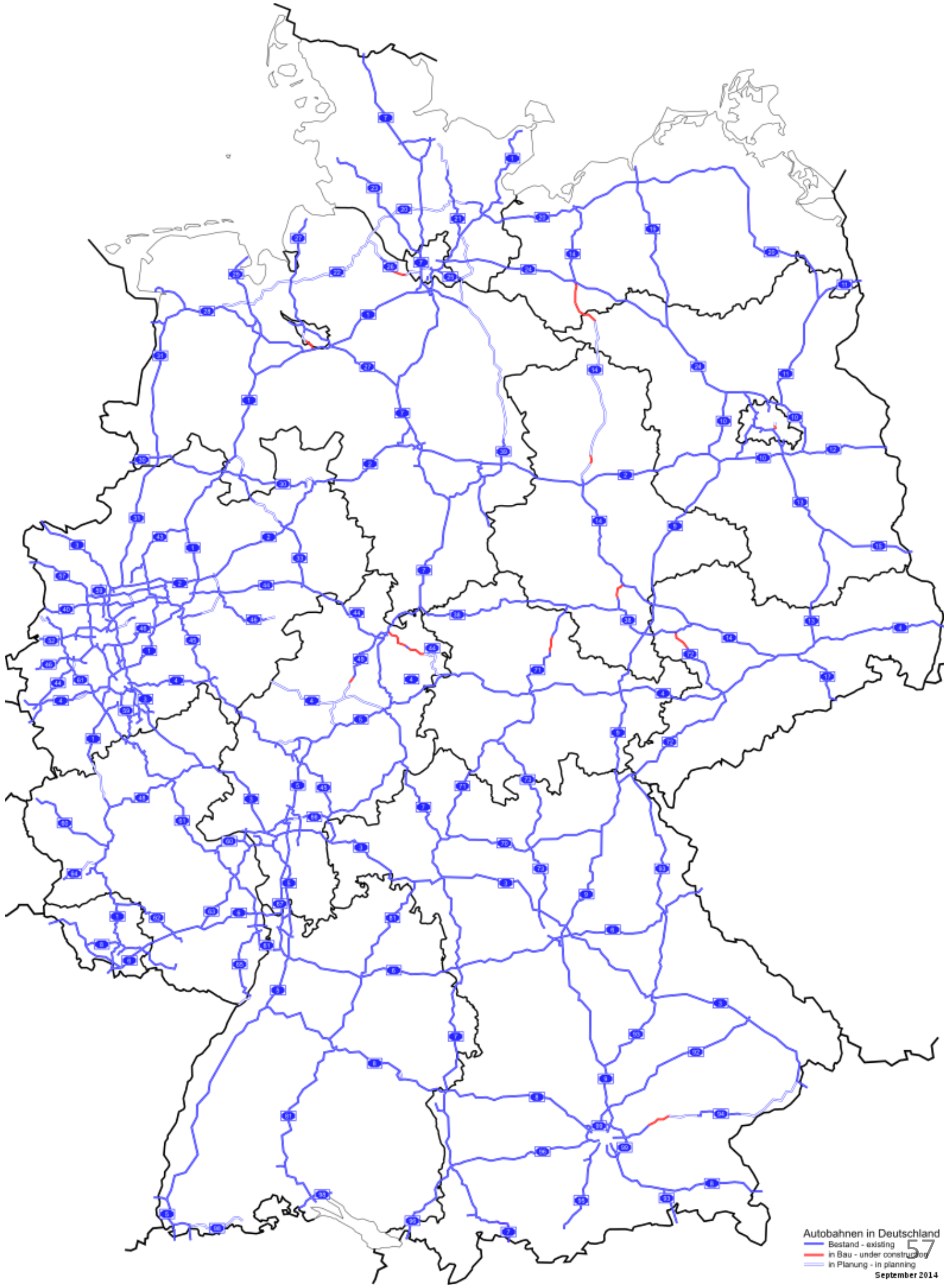
Air traffic network

Brockmann et al. (2013). The Hidden Geometry of Complex, Network-Driven Contagion Phenomena, Science

Technological networks

Transportation, communication, sensor networks, energy …

German Autobahn (high-way)

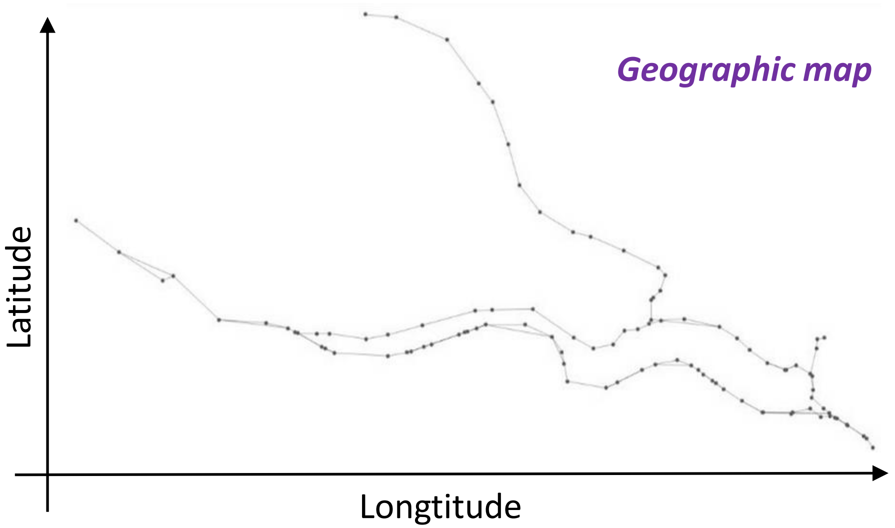


Source: Wikipedia,
[*https://en.wikipedia.org/wiki/Autobahn#/media/File:Autobahnen_in_Deutschland.svg*](https://en.wikipedia.org/wiki/Autobahn#/media/File:Autobahnen_in_Deutschland.svg)

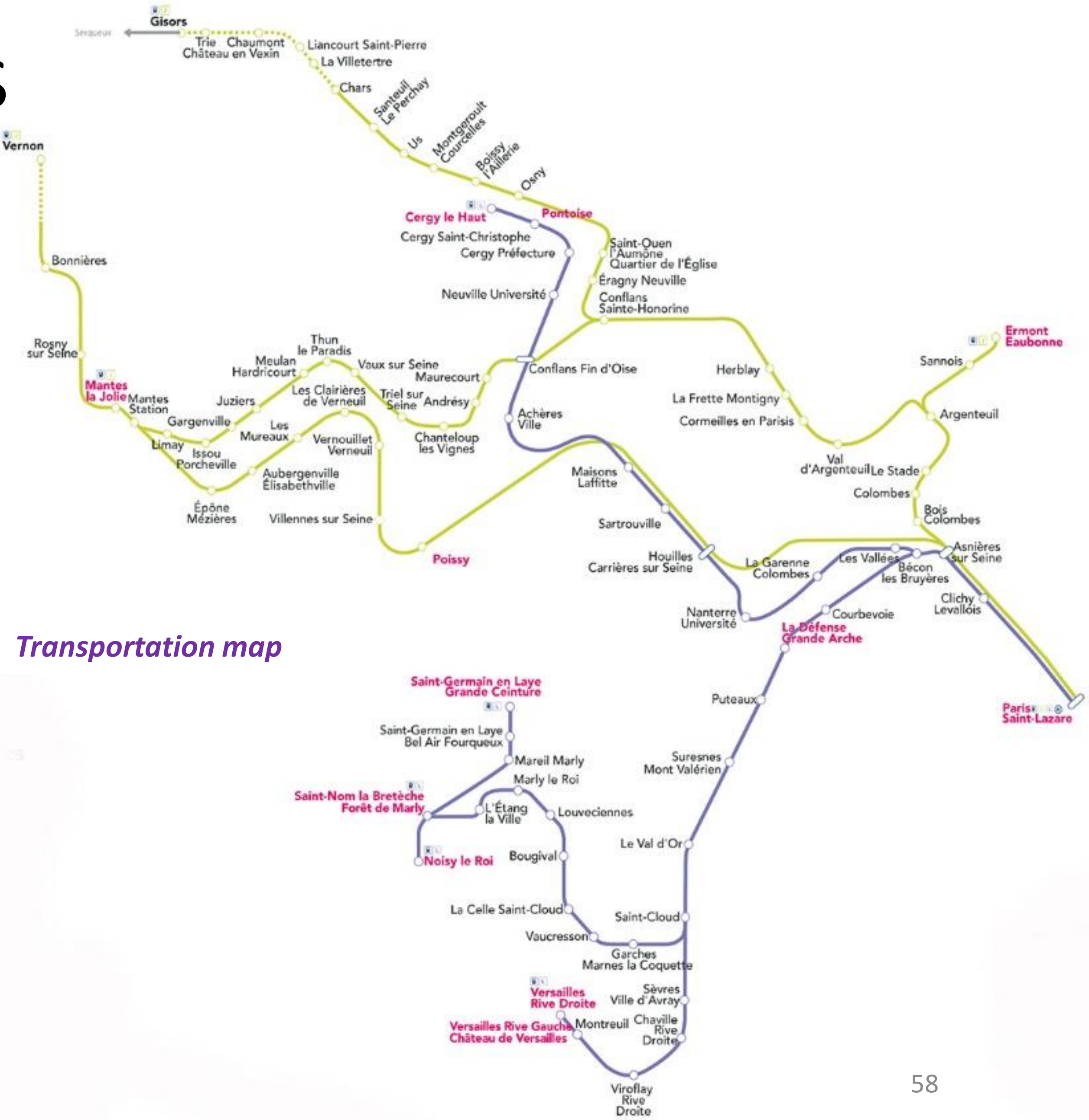
Technological networks

Transportation, communication, sensor networks, energy …

Line J and Line L of Transilien (Ile-de-France)



Geographic map



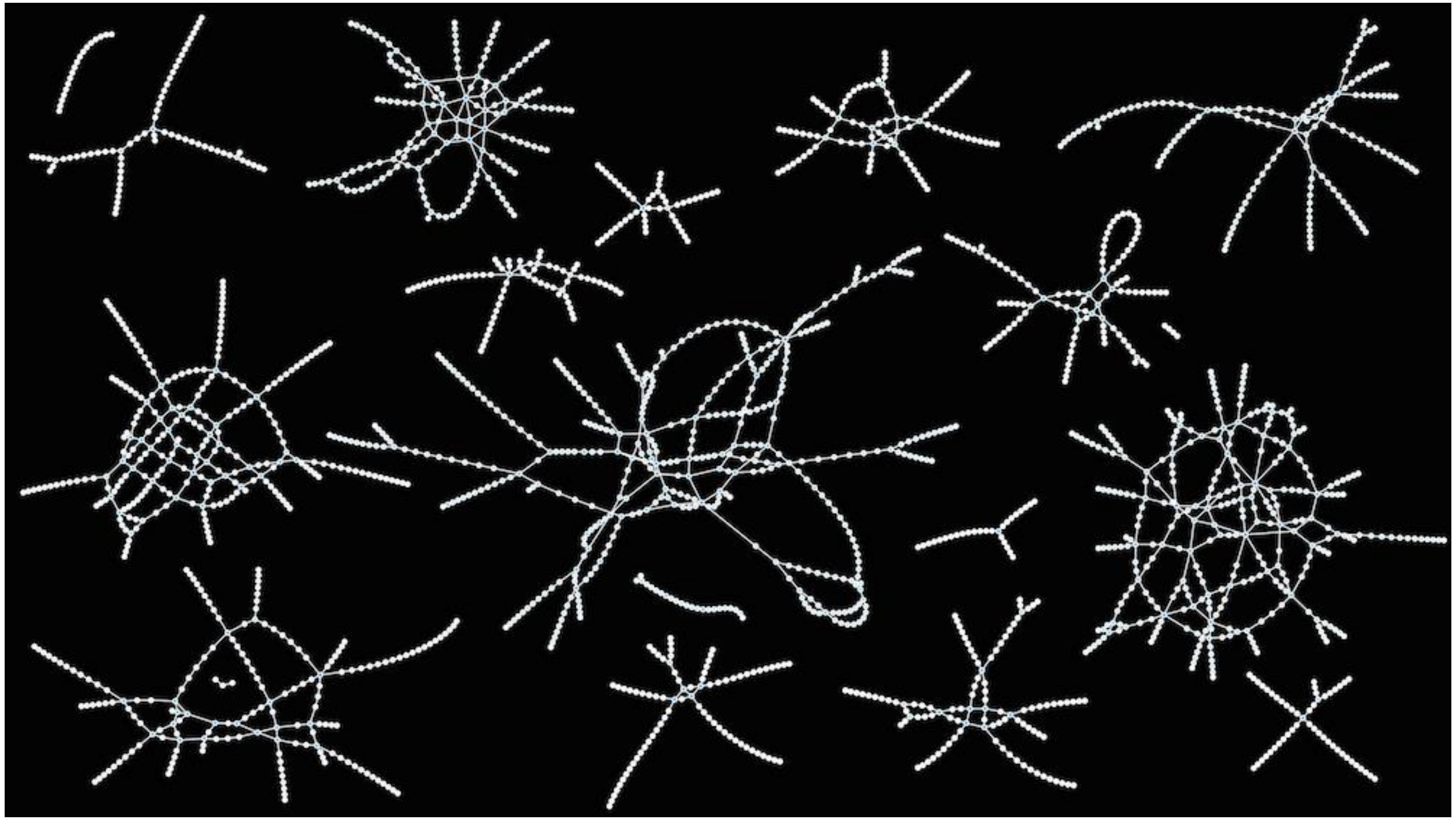
Transportation map

Source: MORANE project, SNCF-Center Borelli

Technological networks

Transportation, communication, sensor networks, energy …

Metro networks
Around the world



Source: http://www.martingrandjean.ch/metro-quiz-15-subway-networks/

Technological networks

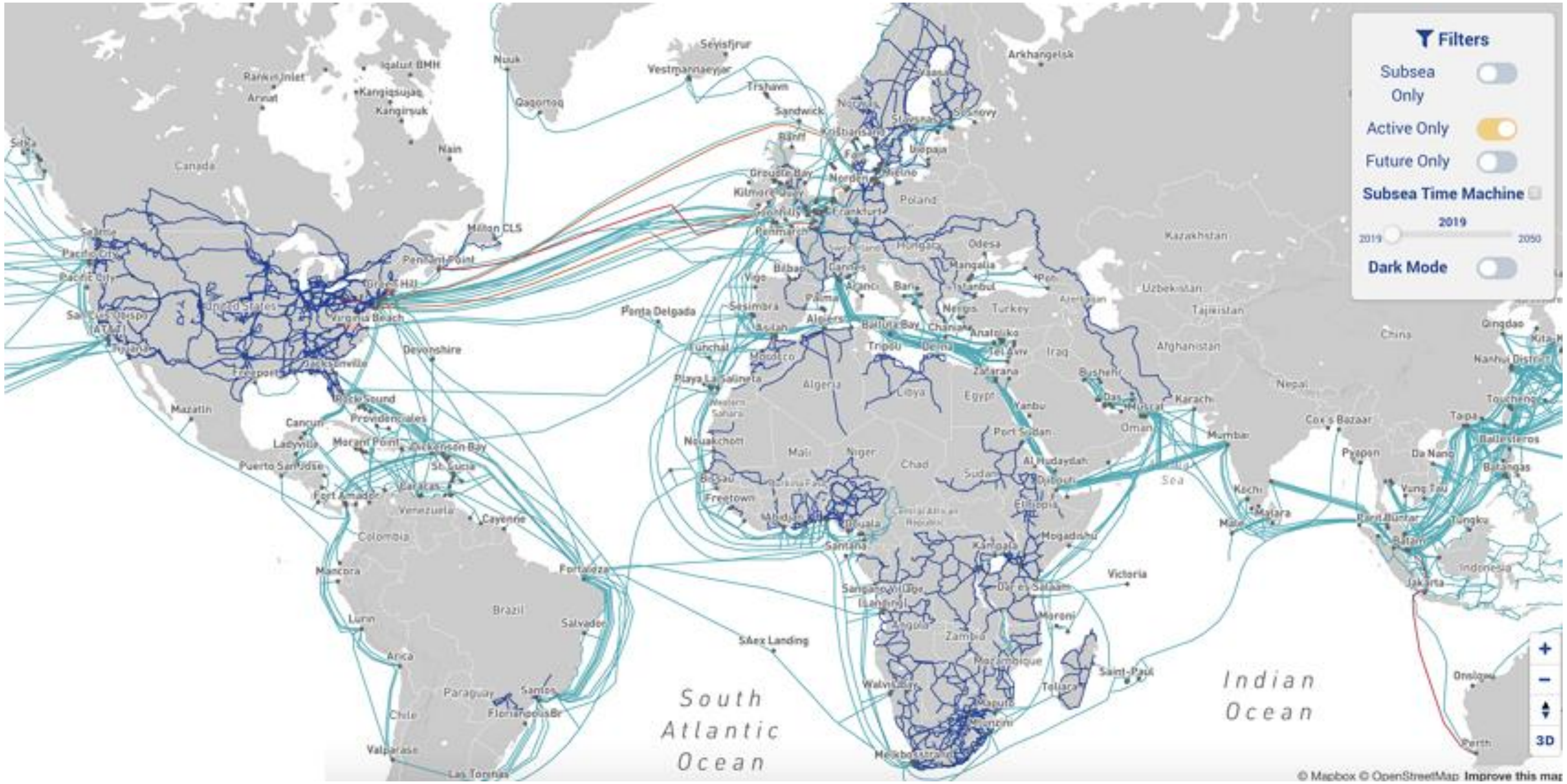
Transportation, communication, sensor networks, energy …

Communication antennas receiving messages from mobile devices

Technological networks

Transportation, communication, sensor networks, energy …

Network atlas of the global Internet infrastructure

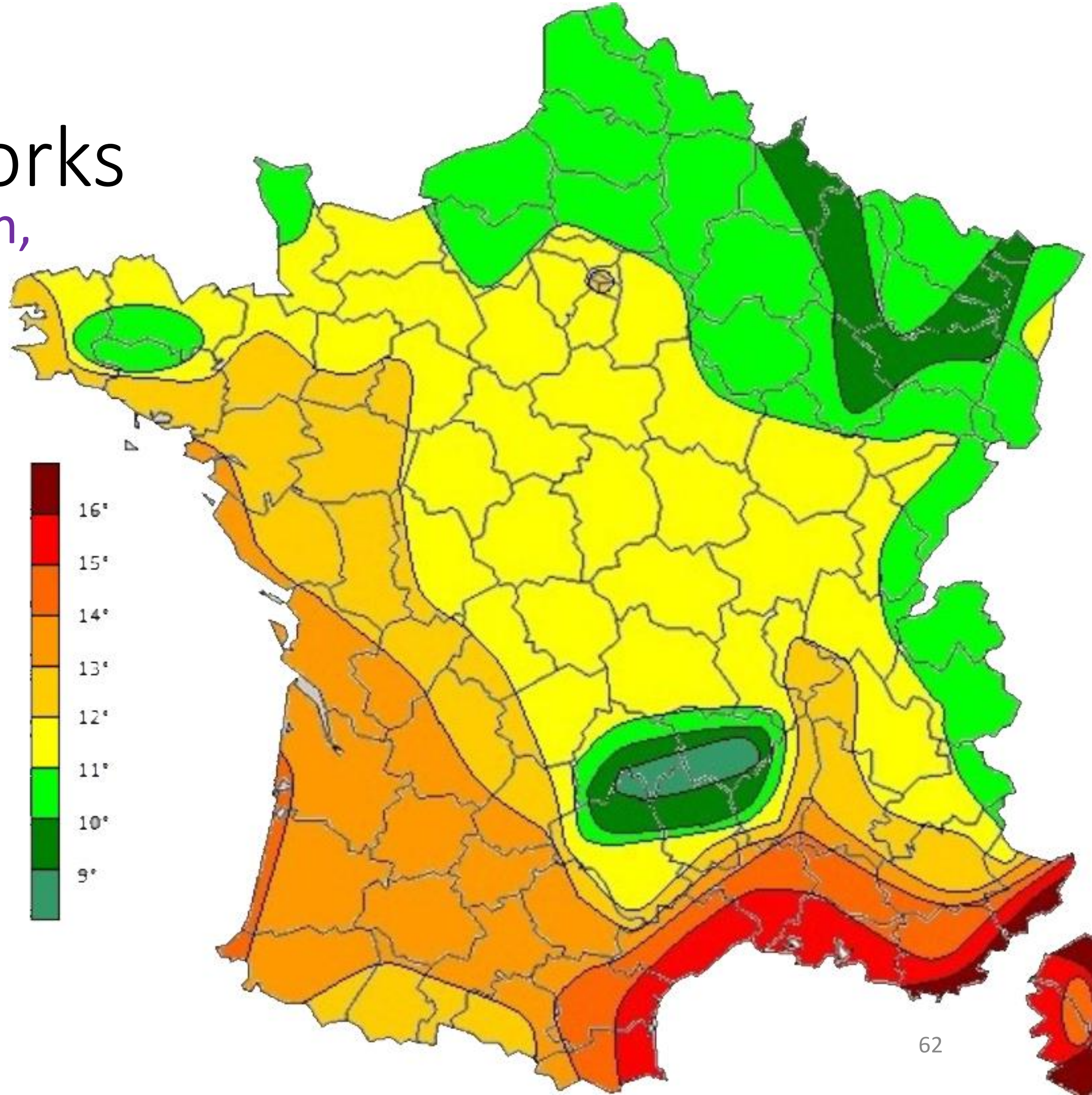


Source: Network Atlas, https://networkatlas.com/

Technological networks

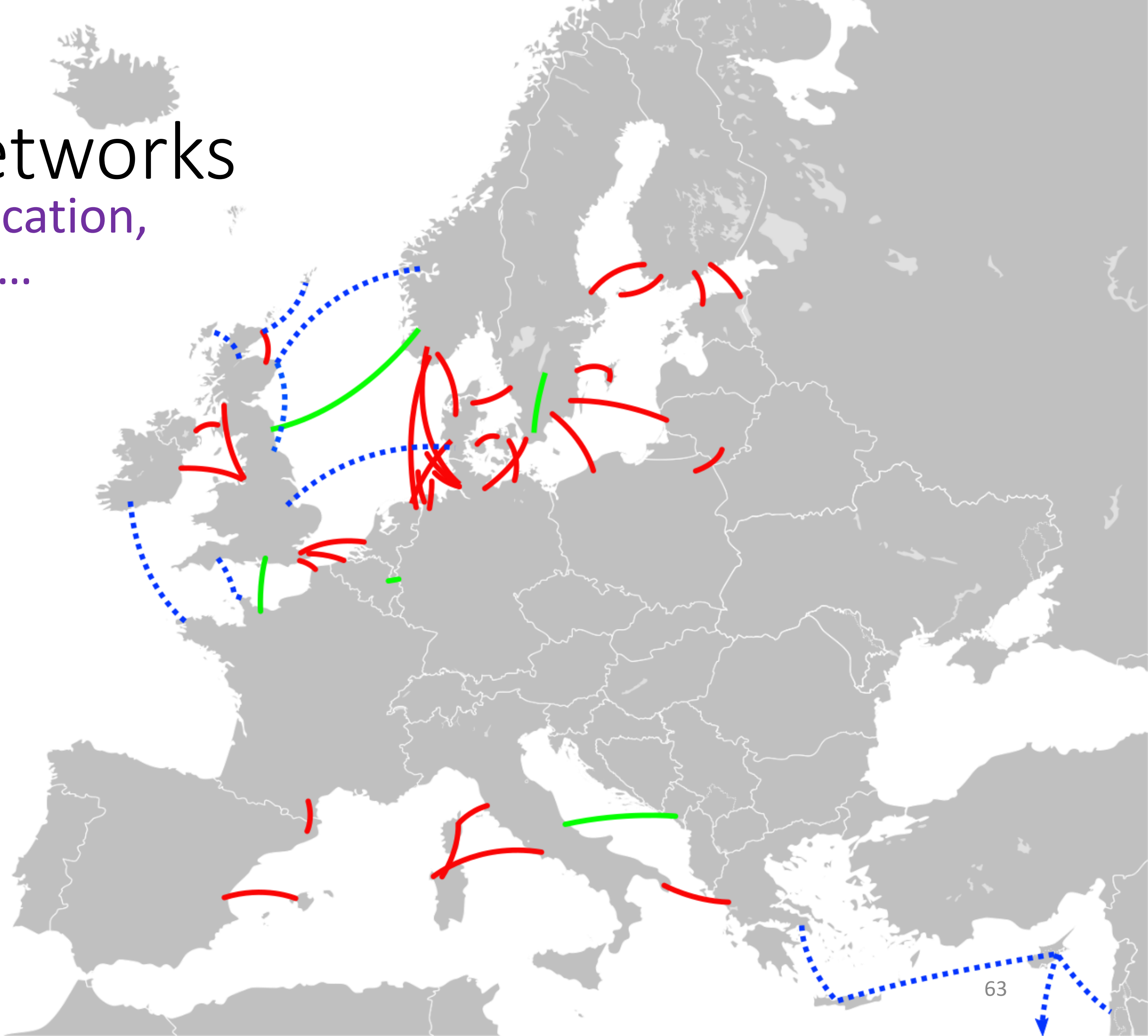
Transportation, communication, sensor networks, energy …

Sensors network measuring temperature or other meteo attributes



Technological networks

Transportation, communication, sensor networks, energy …



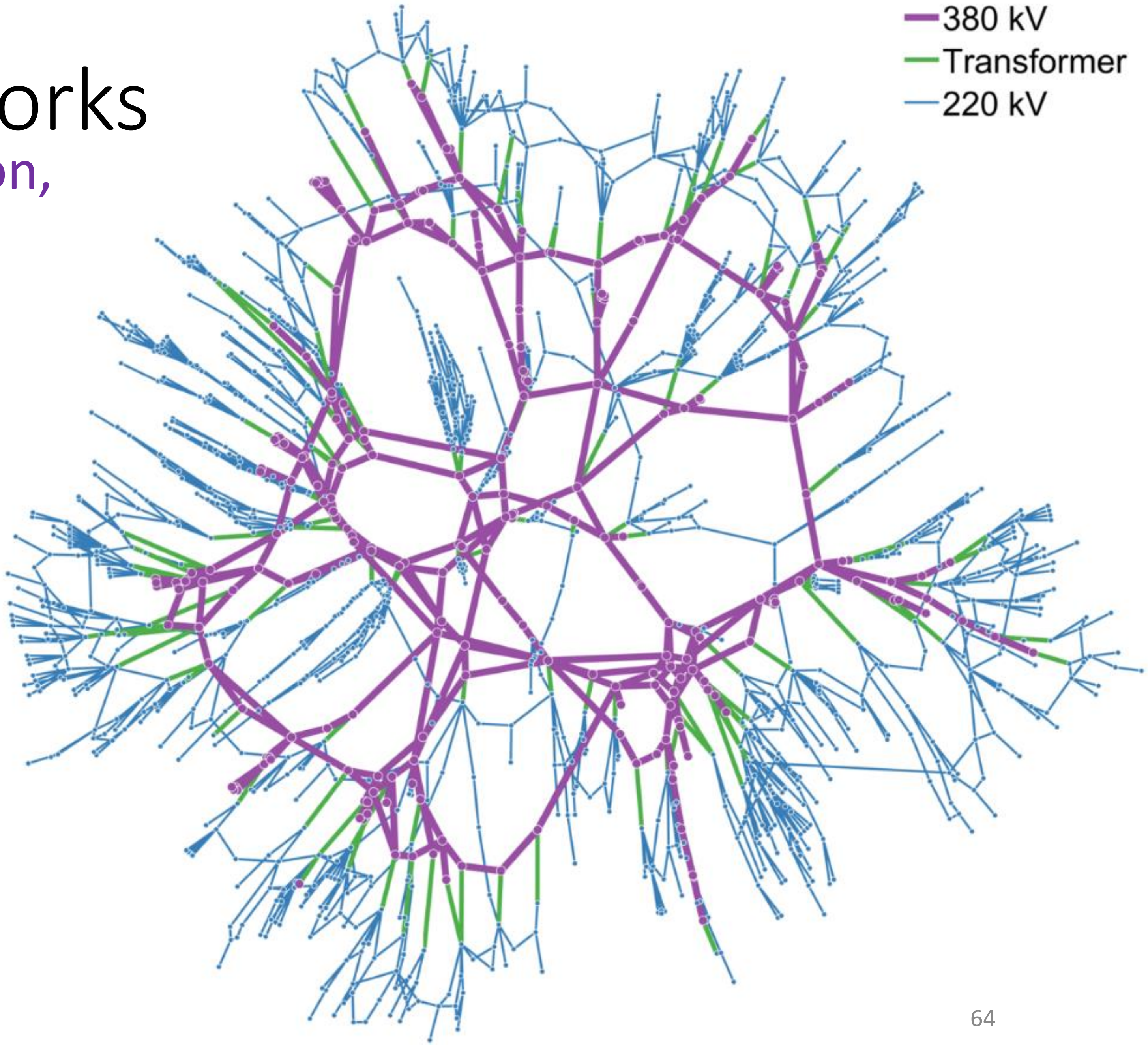
European high-voltage electrical grid

Source: Wikipedia, *https://en.wikipedia.org/wiki/Electrical_grid#/media/File:HVDC_Europe.svg*
See also: *https://ec.europa.eu/energy/infrastructure/transparency_platform/map-viewer/main.html*

Technological networks

Transportation, communication, sensor networks, energy …

Network diagram of a high-voltage transmission system (not physical geography)



Source: P. Cuffe et al. (2017). "Visualizing the Electrical Structure of Power Systems". IEEE Systems Journal.

Information networks

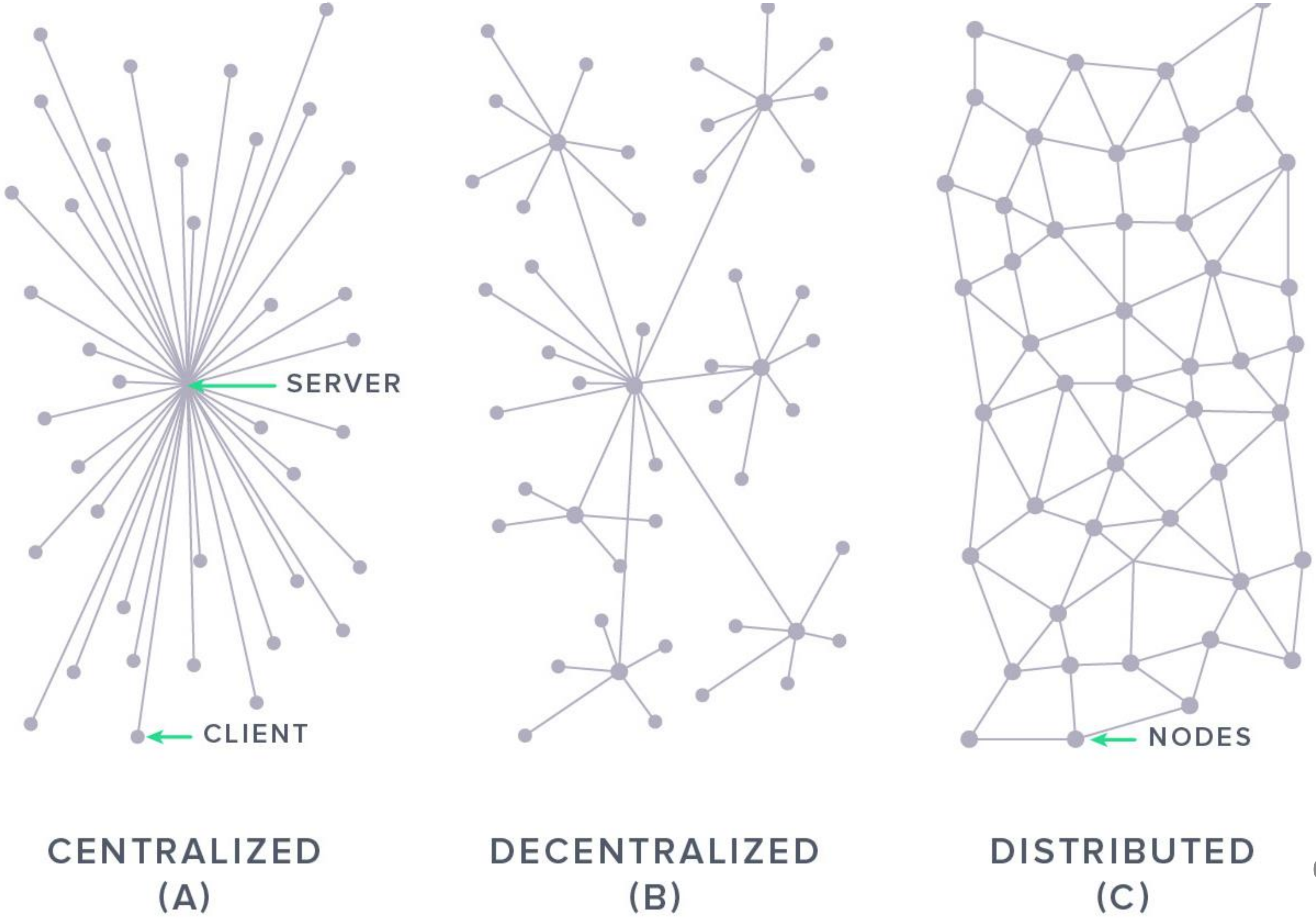
…can be also tech nets

Architecture and structure in computer networks

e.g. the computer system of a company

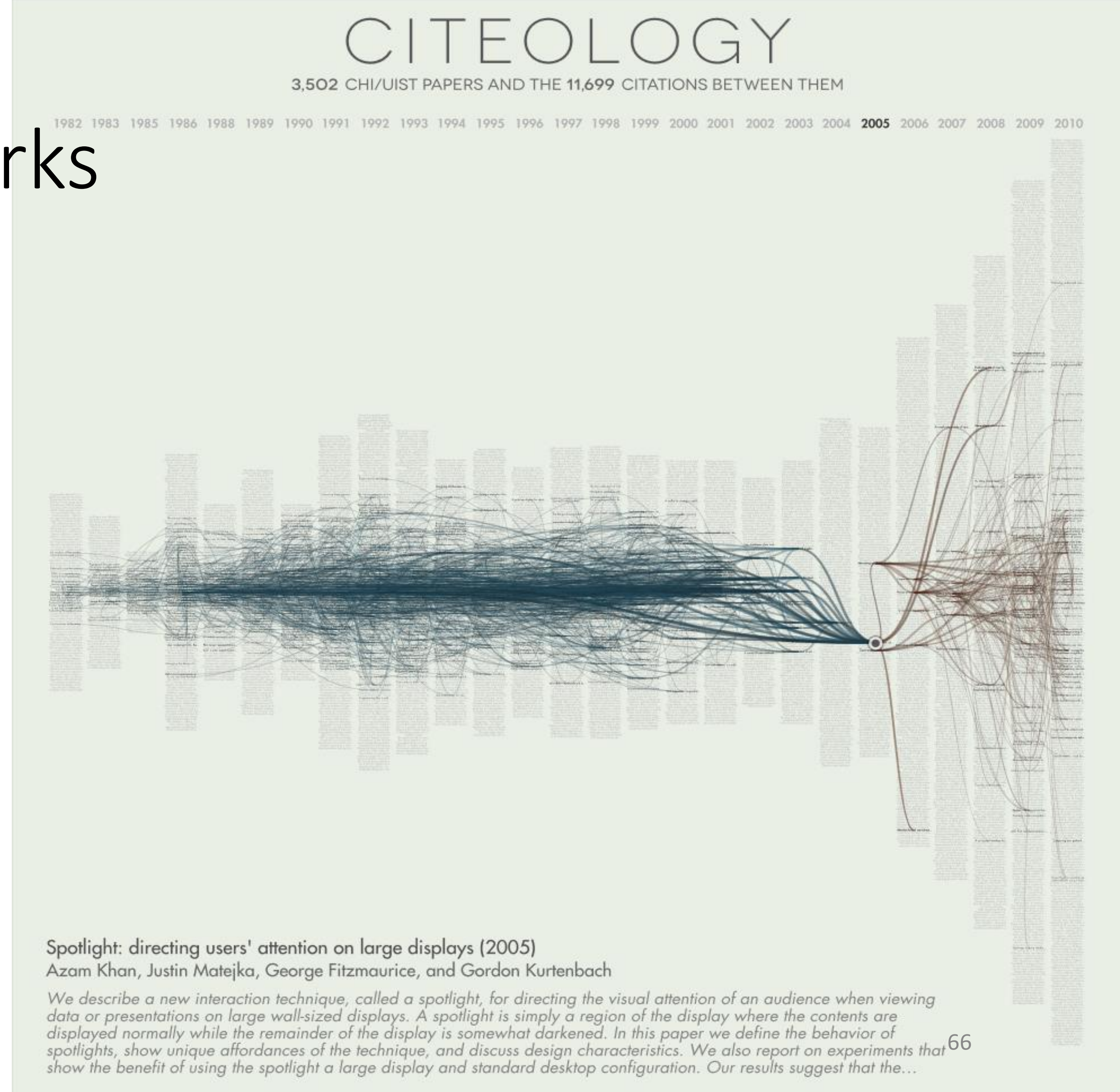
e.g. the computer system of an university campus

e.g. a peer-to-peer file sharing network



Information networks

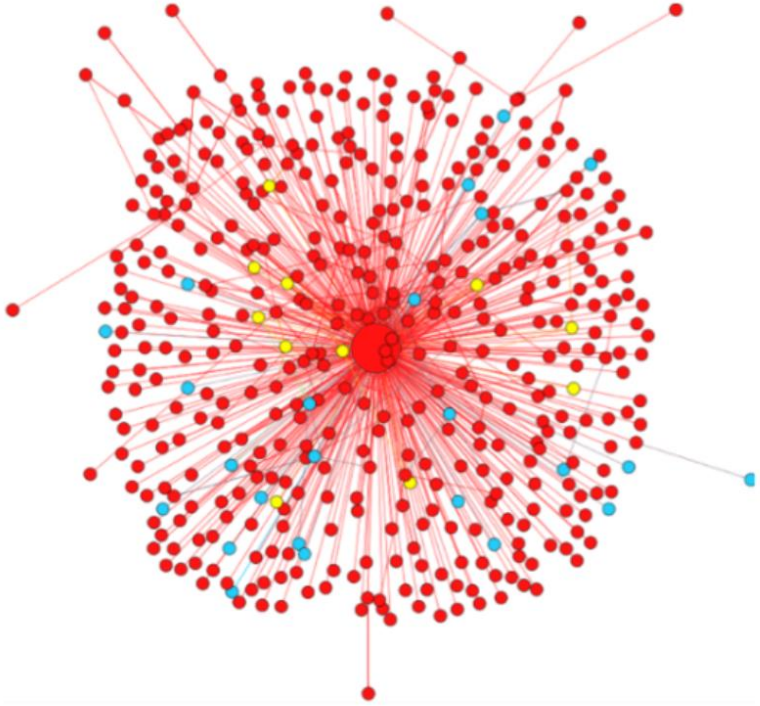
Paper-based citation network



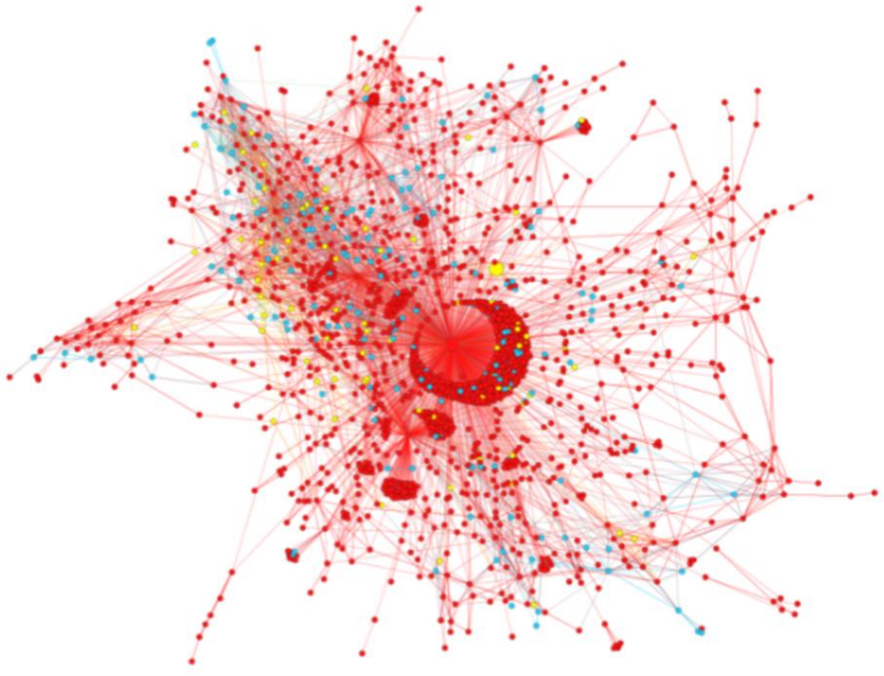
Source: Citeology: visualizing paper genealogy (A project by Autodesk)

Information networks
… over social networks

Hacked Associated Press twitter account spreading rumor about Obama injury (2013)



(a) 60 seconds after the hacked twitter account sent out the White House rumor there were already sufficient enquiry tweets (blue nodes).

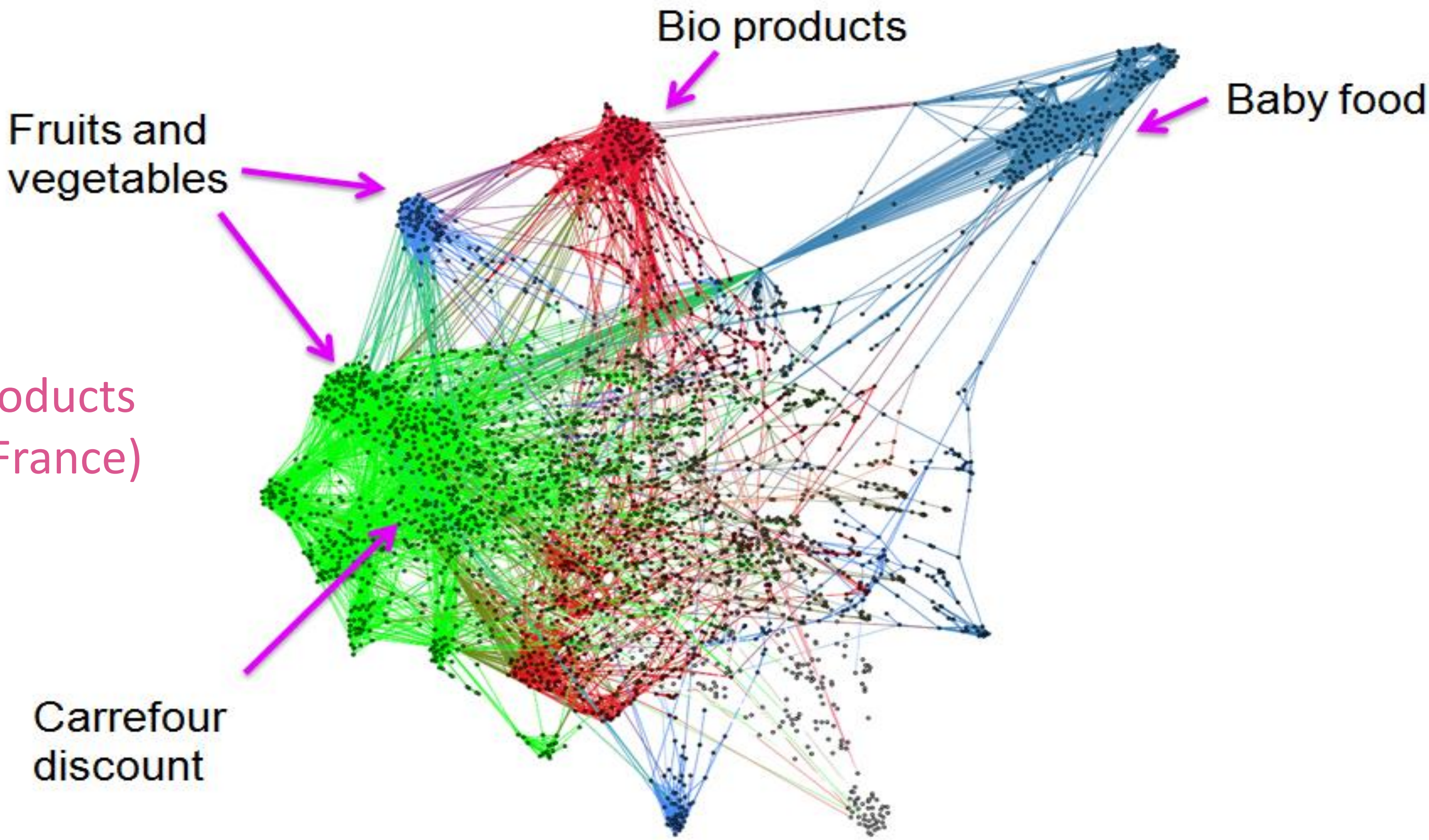


(b) Two seconds after the first denial from an AP employee and two minutes before the official denial from AP, the rumor had already gone viral.

Source: Z. Zhao, *et al. (2015).* Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts, *WWW.*

Information networks

… similarity nets



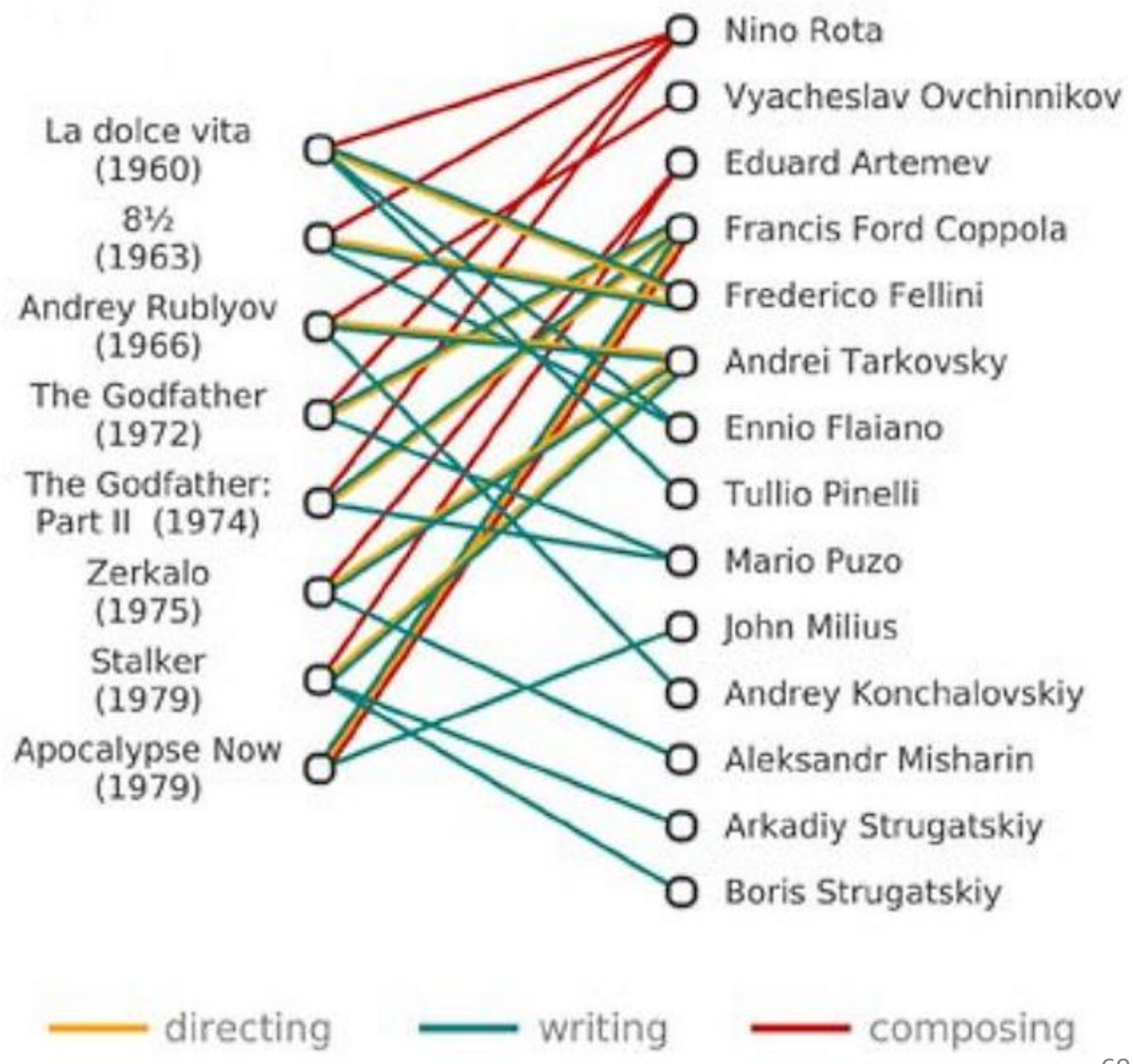
Often co-purchased products in Carrefour Markets (France)

Source: SODATECH project (Center Borelli)

Information networks

… preference nets

Movies and people in credits



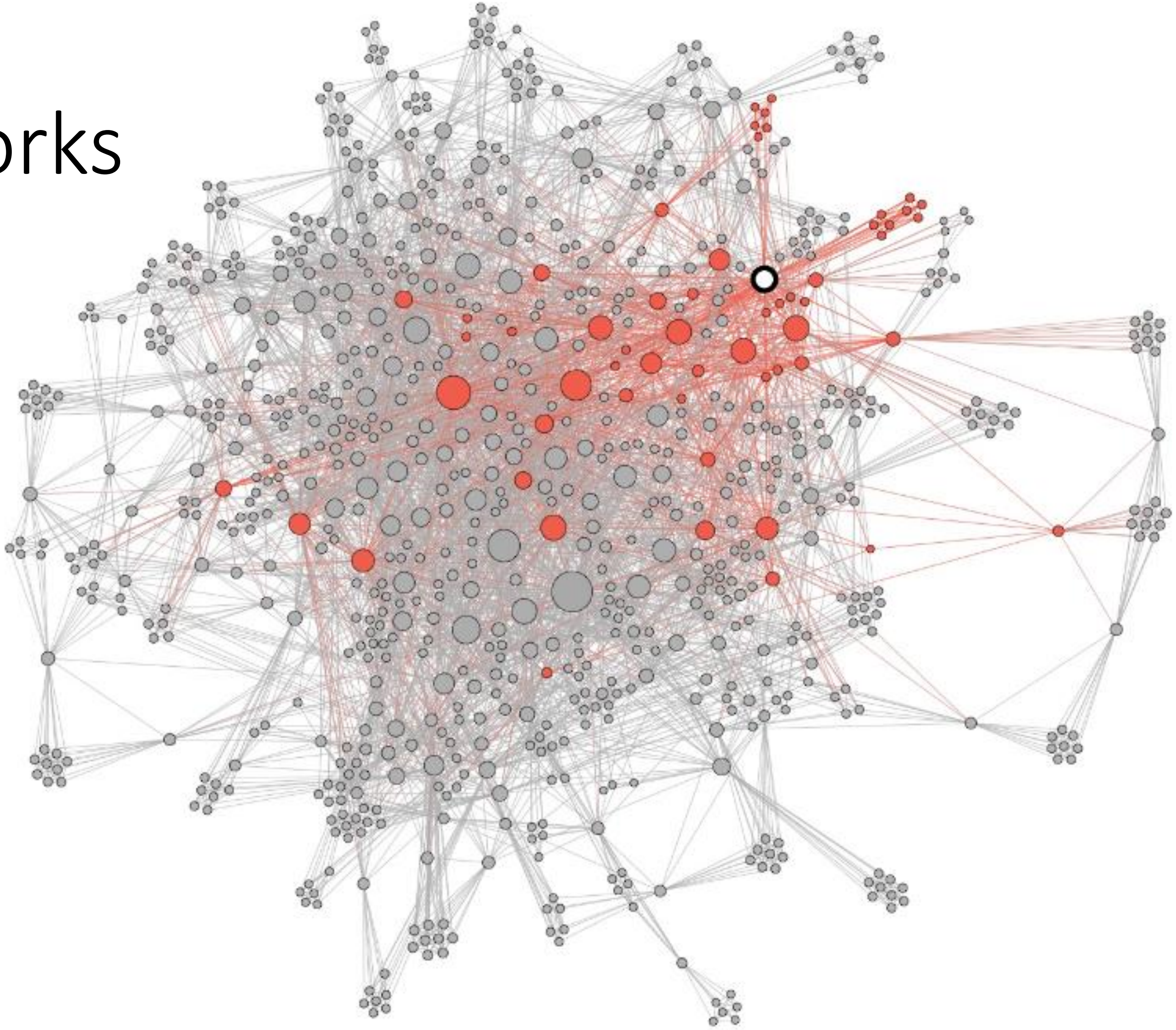
Source: A. Spitz et al. (2014). Measuring Long-Term Impact Based on Network Centrality: Unraveling Cinematic Citations

Information networks

… similarity nets

Network of similarity between authors of literature based on the number of readers' preference

Stephen King is in white and in red all other 'relevant' authors.

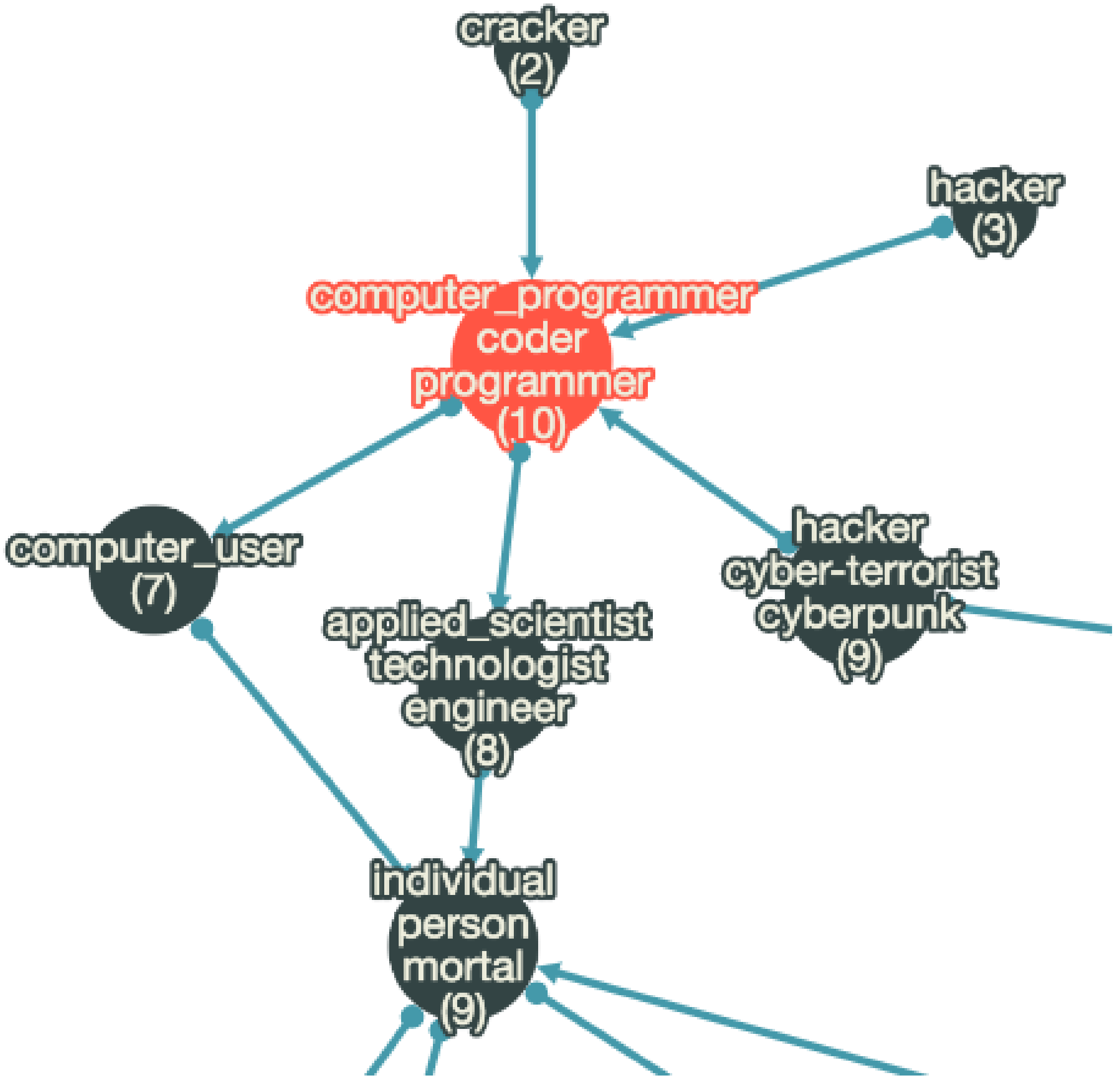


Source: http://www.martingrandjean.ch/mytoptenbooks-reseau-litteraire/

Information networks

… ontologies and semantic nets

A language term represented as part of a semantic network of connected and interrelated senses

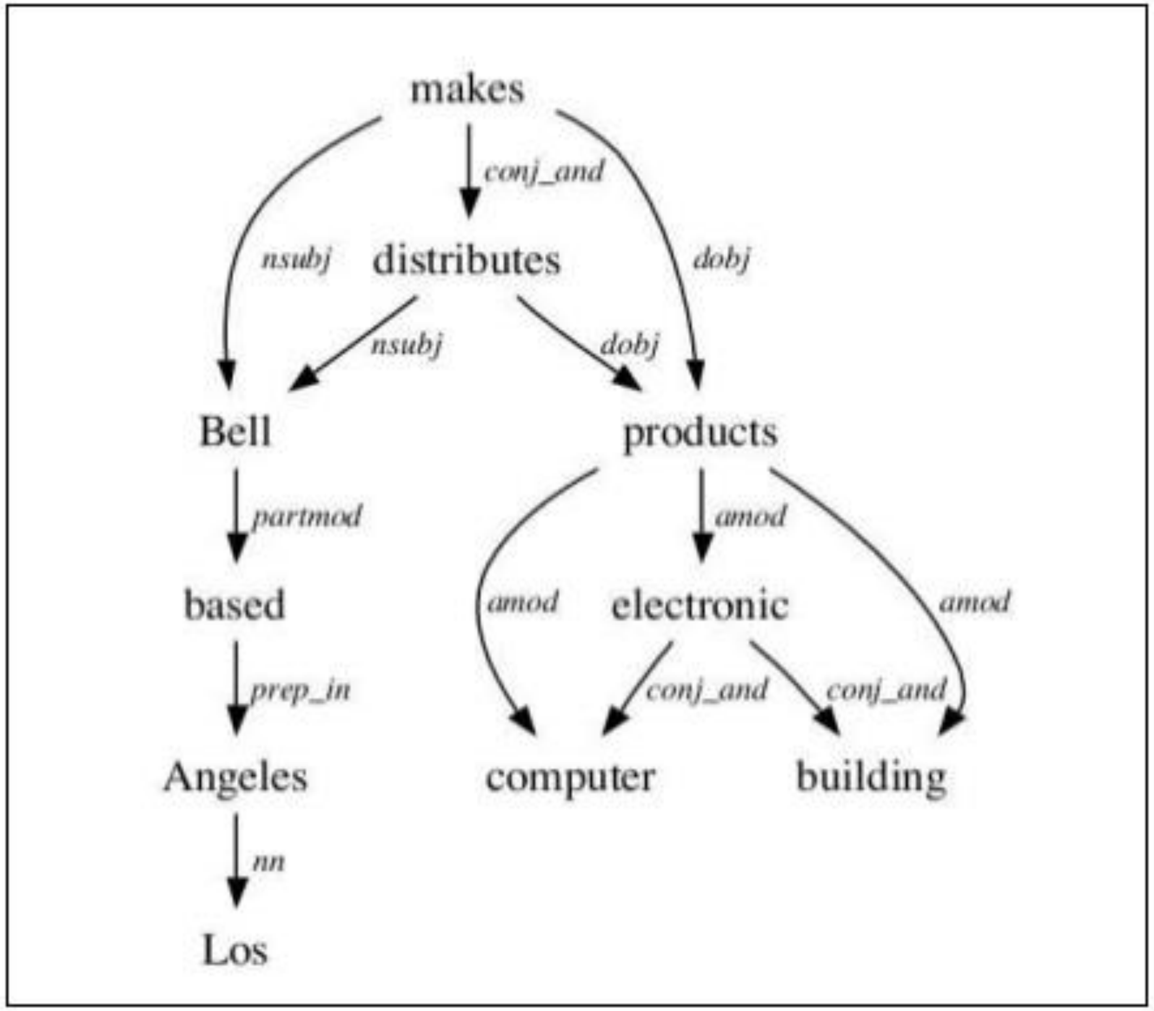


Source: WordNet Viewer, https://wordnet-viewer.herokuapp.com/

Information networks

… ontologies and semantic nets

Natural Language Processing (NLP):
A sentence represented as a dependency parsing graph

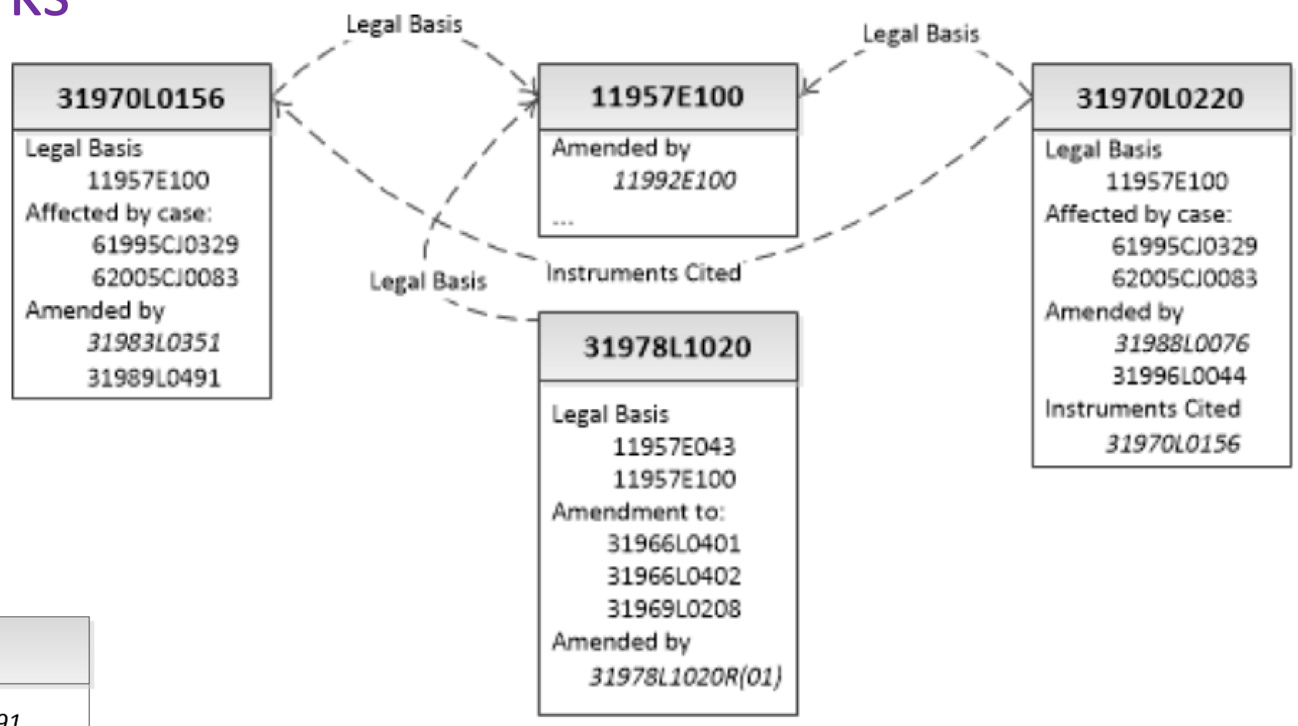


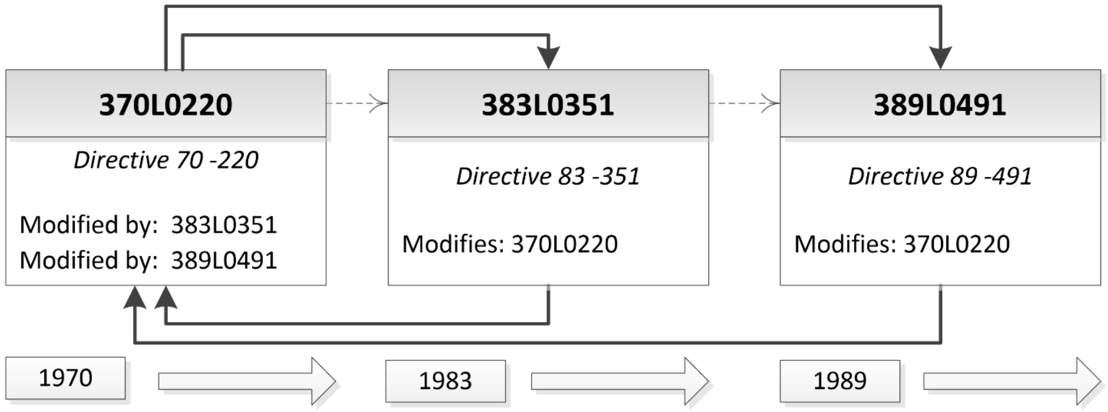
Source: WordNet Viewer, https://wordnet-viewer.herokuapp.com/

Information networks

… dependency and event networks

Legal network:
Cross-reference links
between legal documents
in the EUR-lex dataset





Source: M. Koniaris (2017). Network Analysis in the Legal Domain: A complex model for European Union legal sources

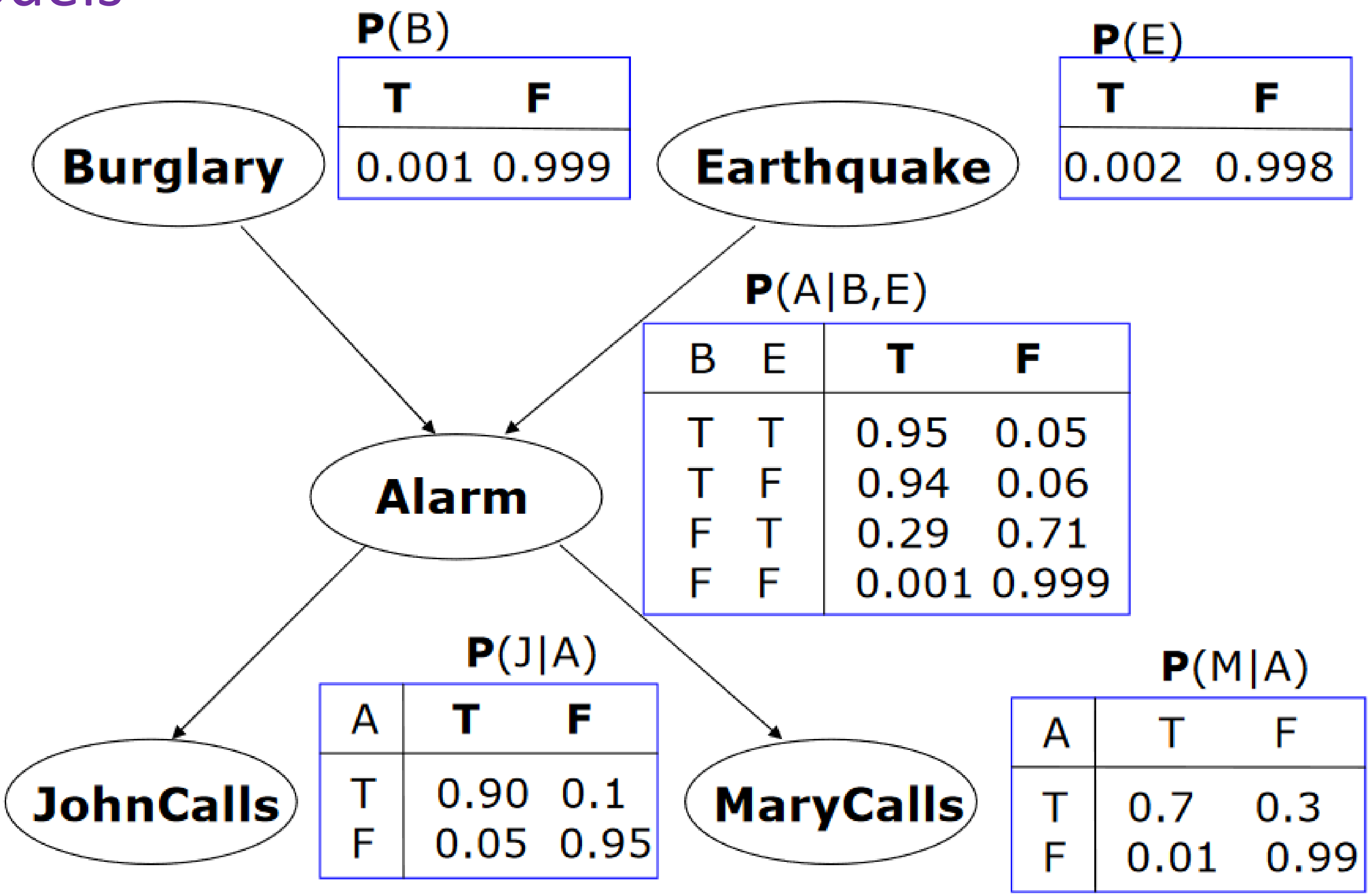
Information networks

… probabilistic graphical models

Modeling uncertainty with probabilities

Here a Bayesian belief network: directed acyclic graph.

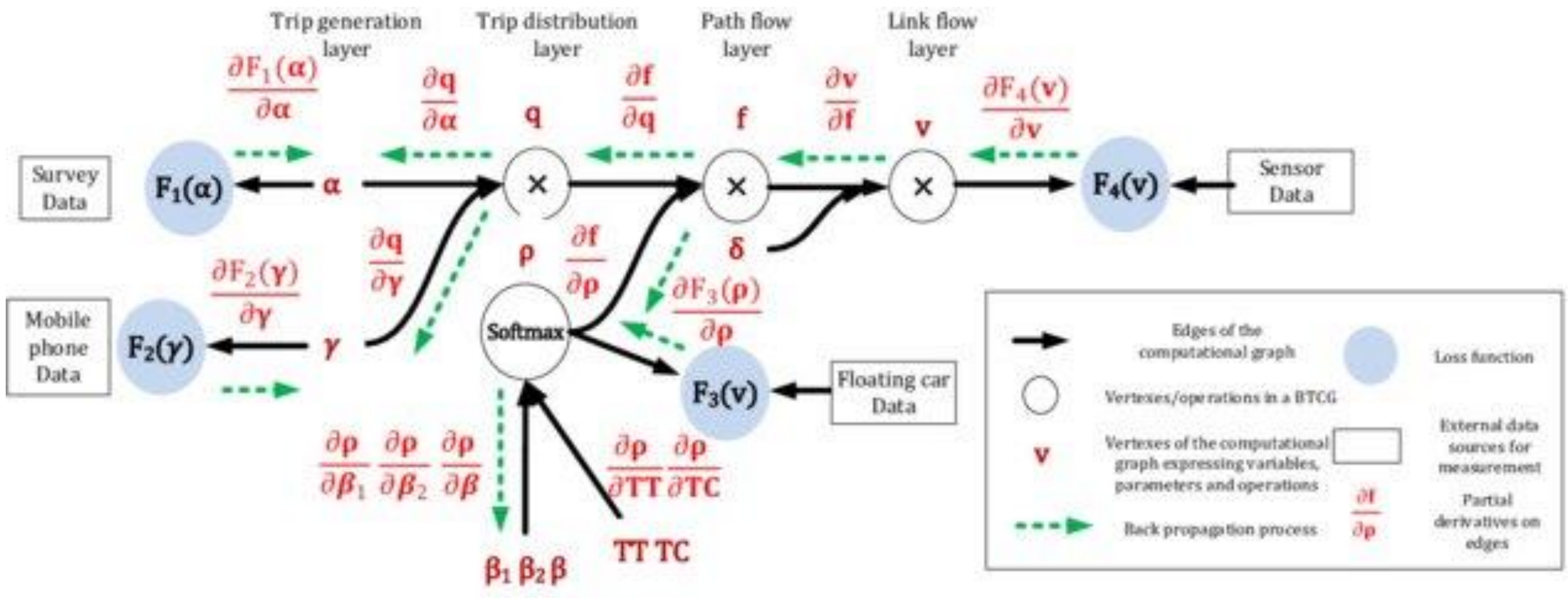
There are also Dependency, Gibs, Markovian PGMs…



Source: https://people.cs.pitt.edu/~milos/courses/cs3710/Lectures/Class3.pdf

Information networks
… computational (flow) networks

Computation graph of a TensorFlow application

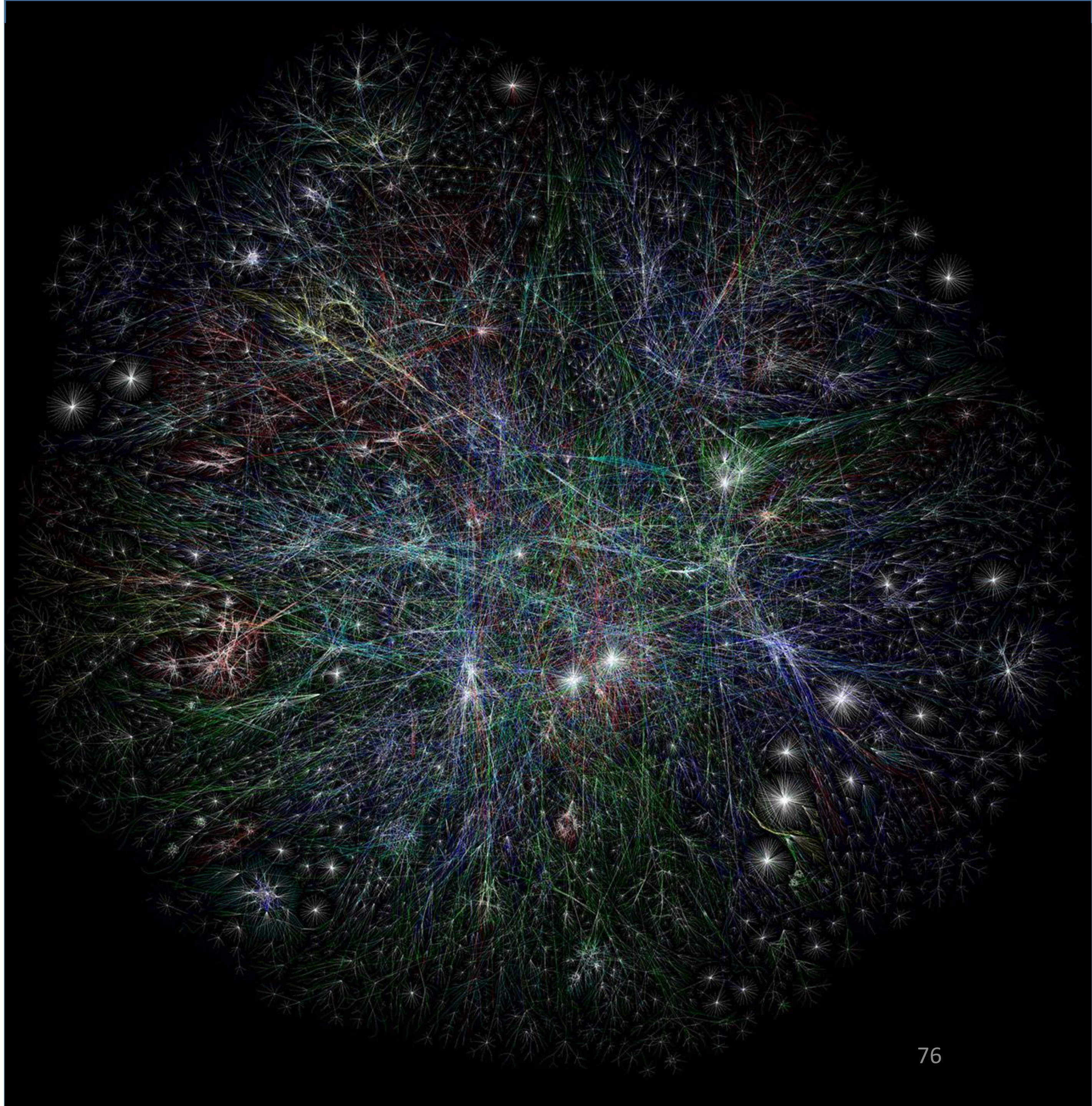


Source: Xin Wu (2018), *Data flow programming model to express traffic demand estimation problem based on Tensorflow*

Social networks

…through which information gets diffused

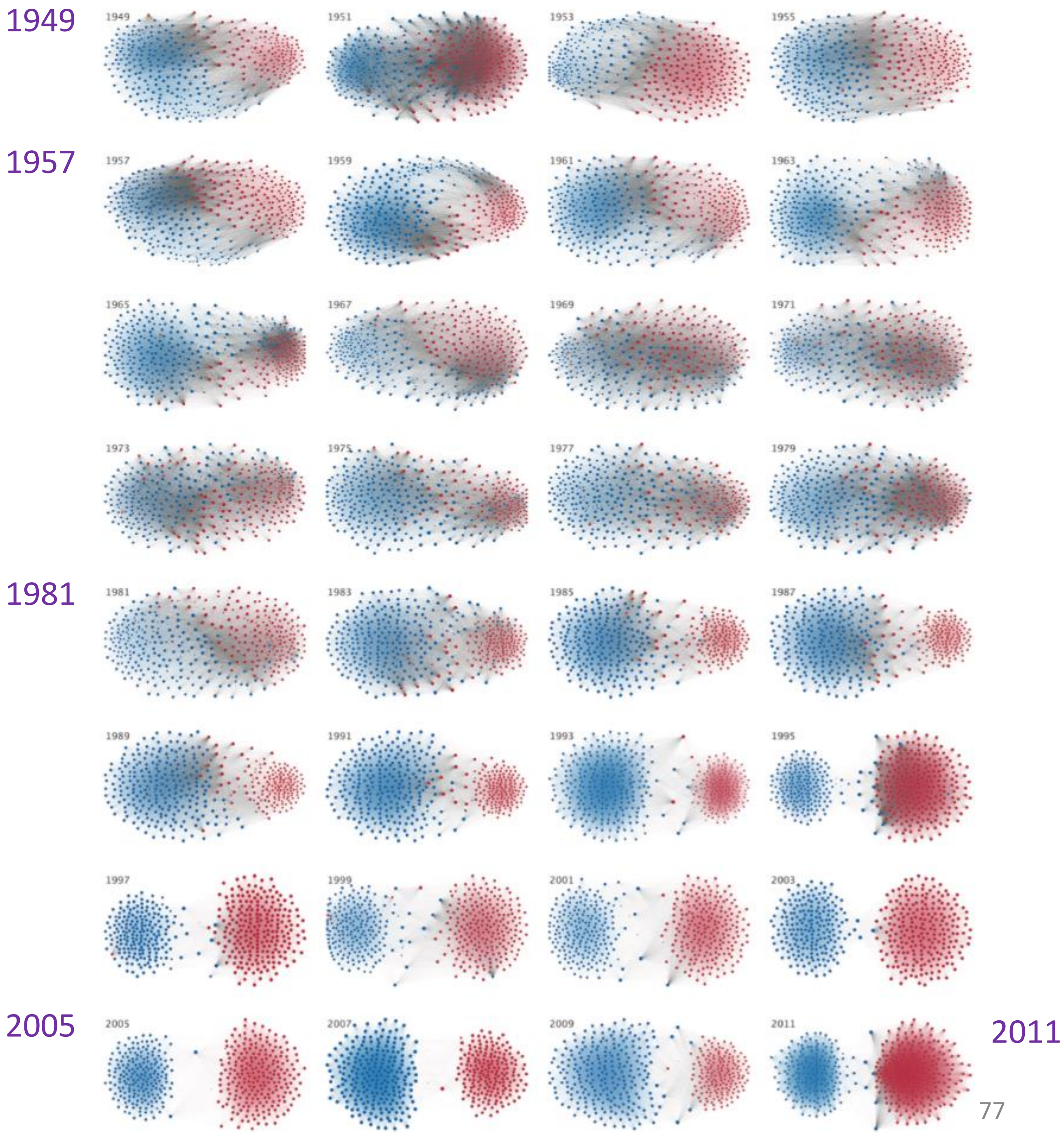
A map of the Facebook network



Social networks

… collaboration nets

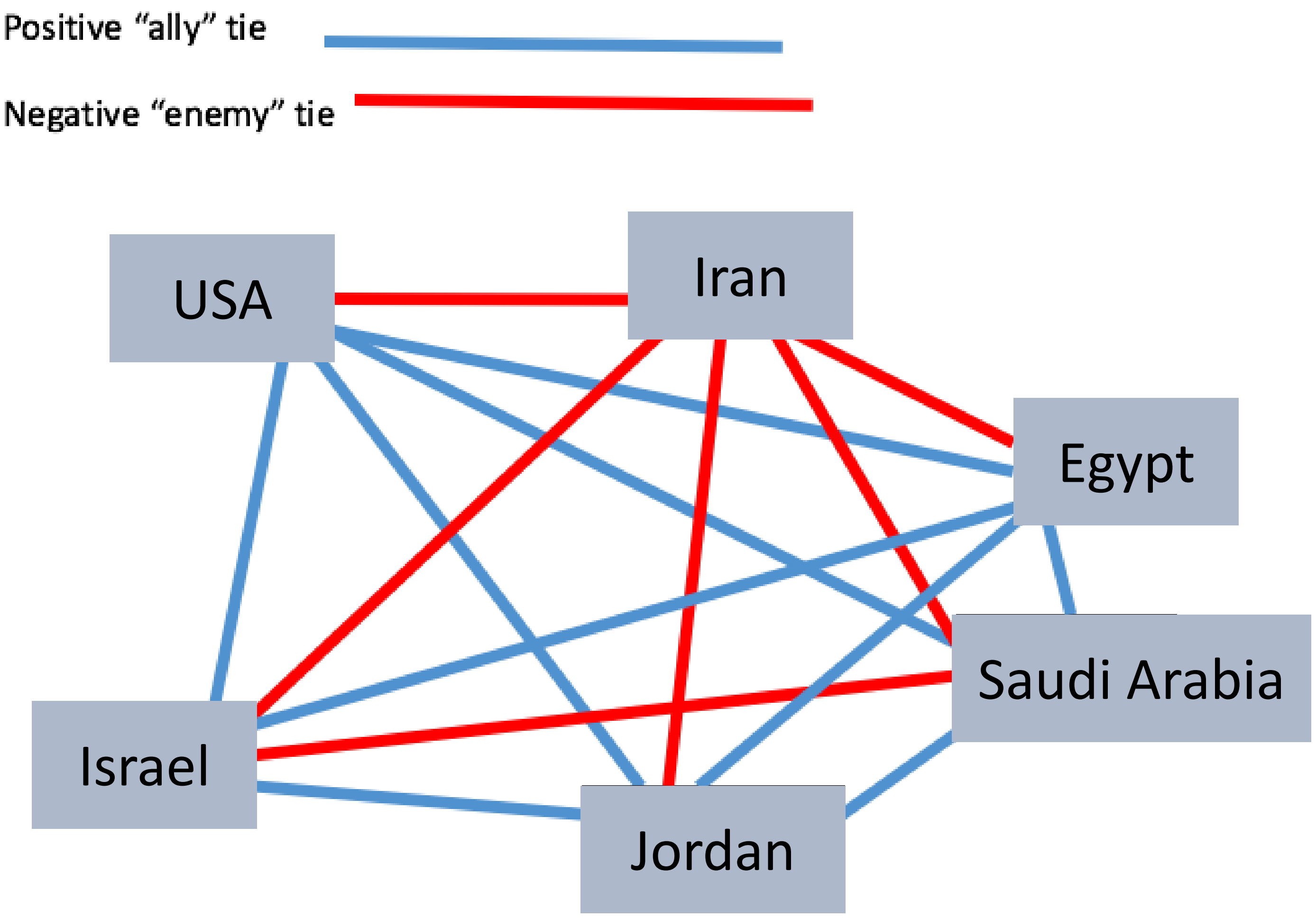
Collaboration vs partisanship
in US politics through time



Source: C. Andris, et al. (2015). The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives, PLOS ONE

Social networks

… or conflict nets



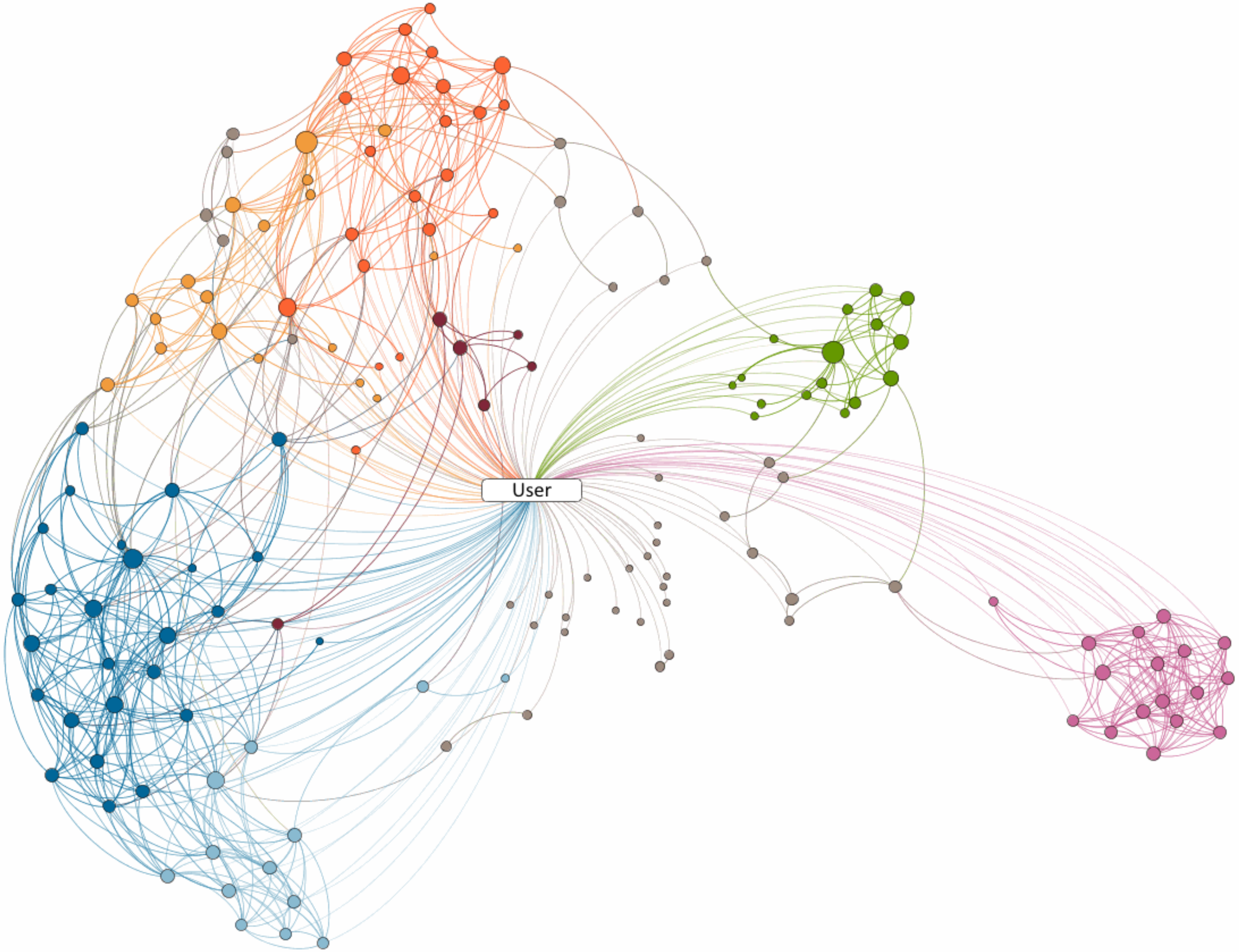
Political alliances and conflicts in Middle East

Source: https://www.rferl.org/a/25103887.html

Social networks

… through which information and ideas get diffused

Local social network around a LinkedIn user

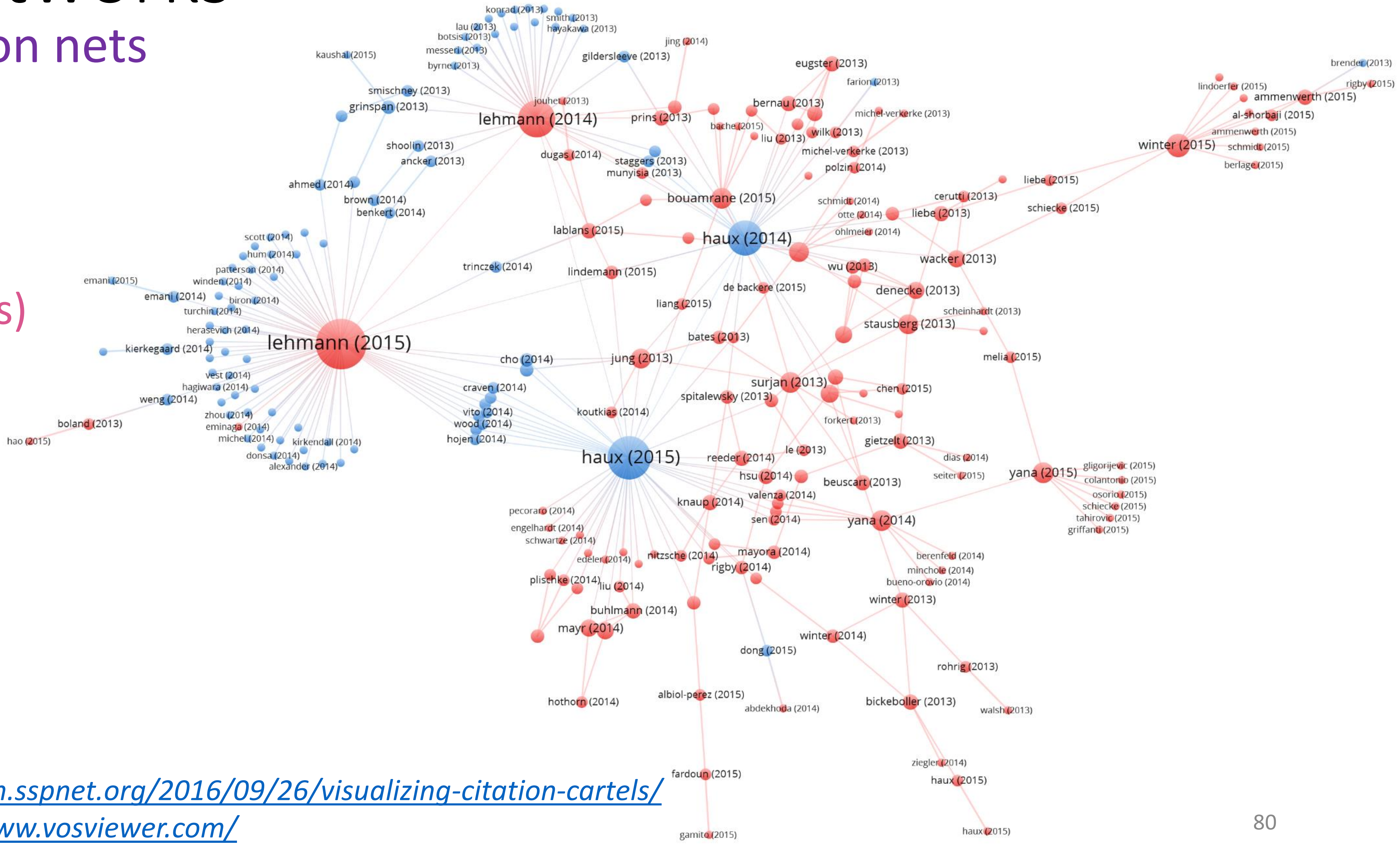


Source: https://cvo22.wordpress.com/2014/03/18/social-network-diagram-analysis-for-work-and-play/

Social networks

… collaboration nets

Author-based
citation network
(and citation cartels)

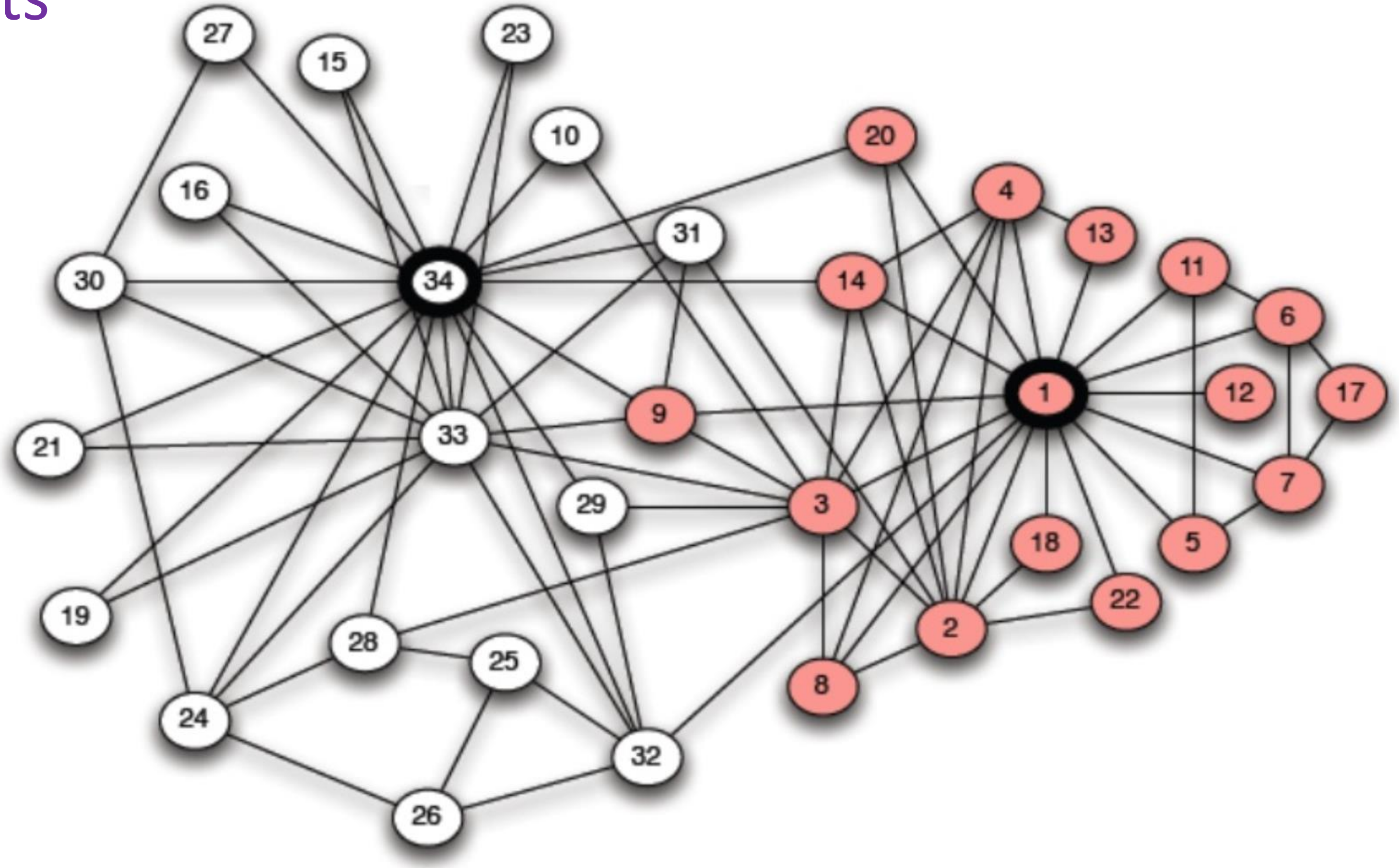


Source: https://scholarlykitchen.sspnet.org/2016/09/26/visualizing-citation-cartels/
See also: VOSviewer, https://www.vosviewer.com/

Social networks

… contact and friendship nets

Zachary's university karate club (frienships)

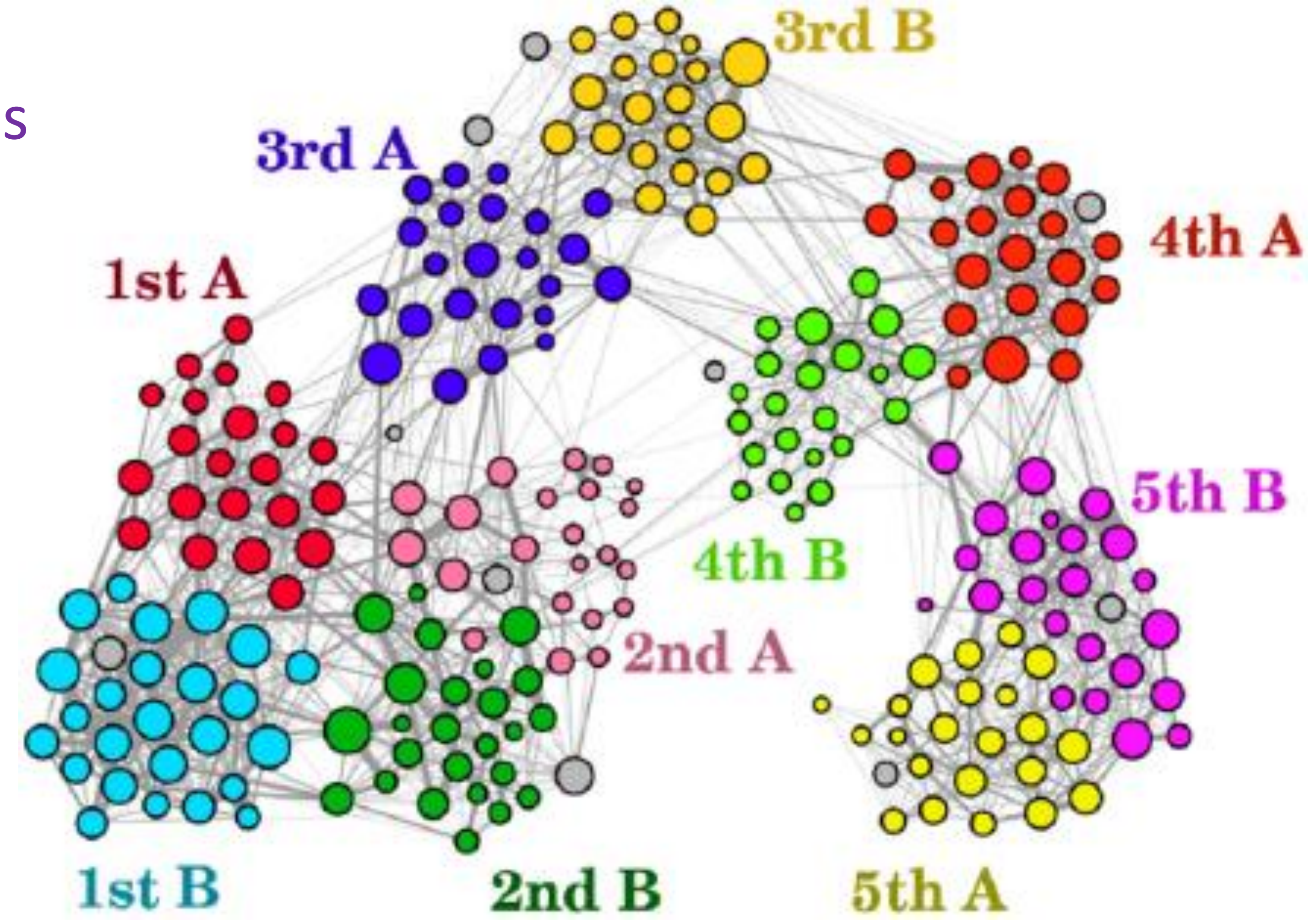


Source: W.W. Zachary, (1977). An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research.

Social networks

… contact and friendship nets

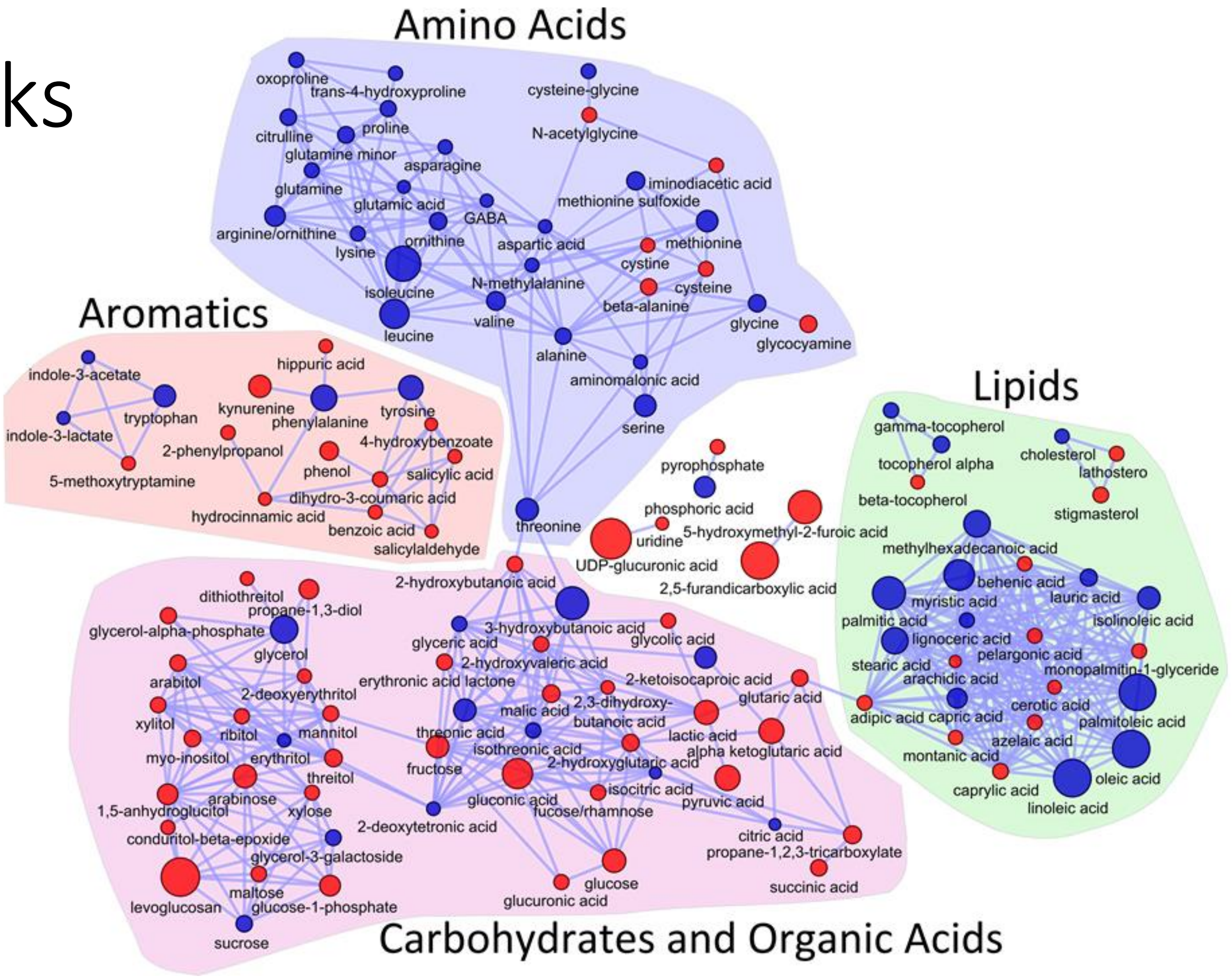
Face-to-face contacts among pupils of a French primary school



Source: J. Stehlé, (2011). High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLOS One

Biological networks

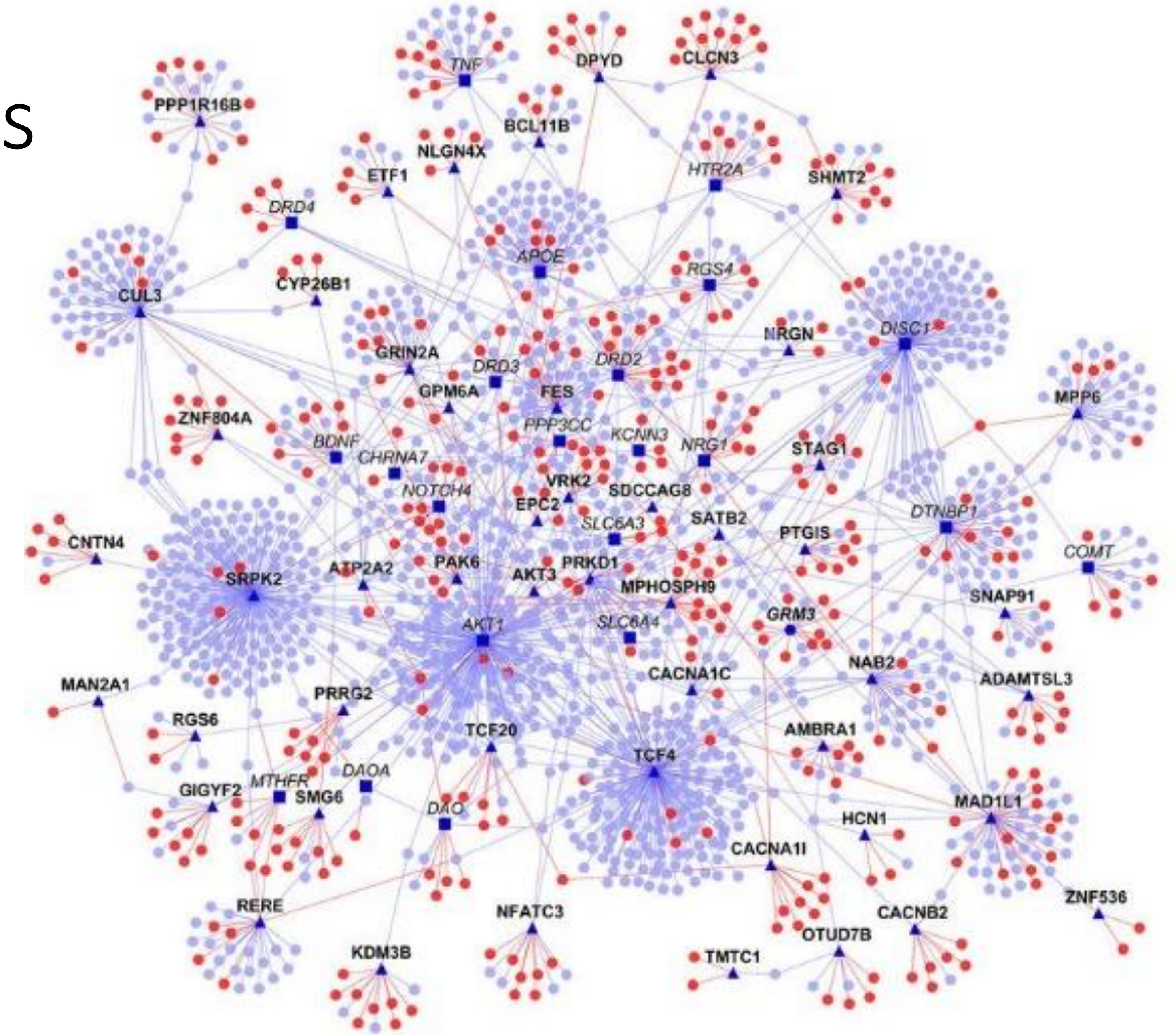
A biochemical network



Source: http://tagteam.harvard.edu/hub_feeds/1981/feed_items/155325
See also: Cytoscape, https://cytoscape.org/

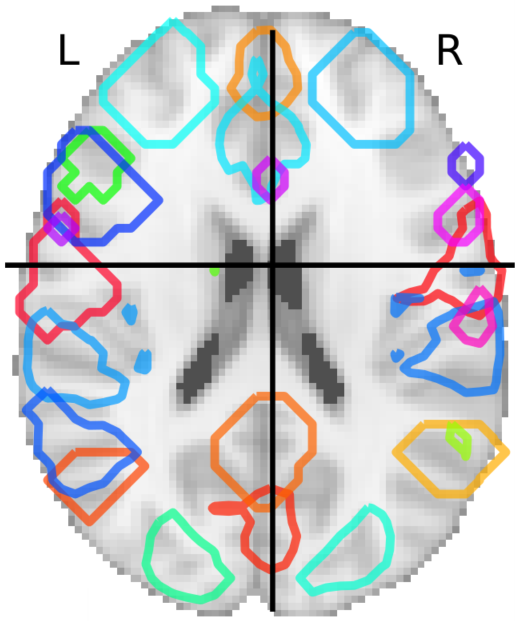
Biological networks

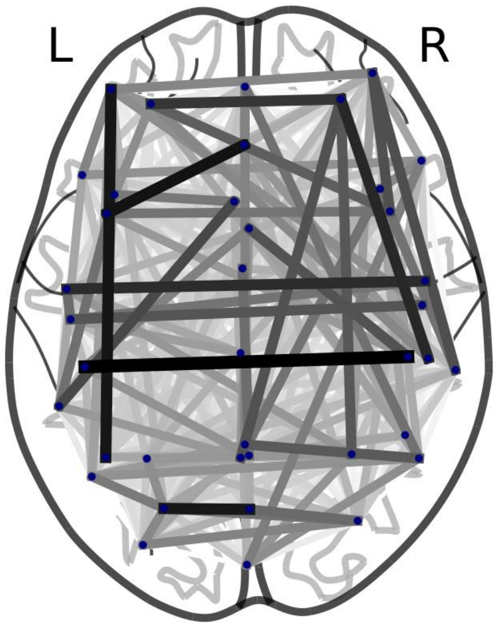
Protein-to-Protein interaction (PPI) network for Schziophrenia



Source: https://en.wikipedia.org/wiki/Interactome#/media/File:Schziophrenia_PPI.jpg

Biological networks



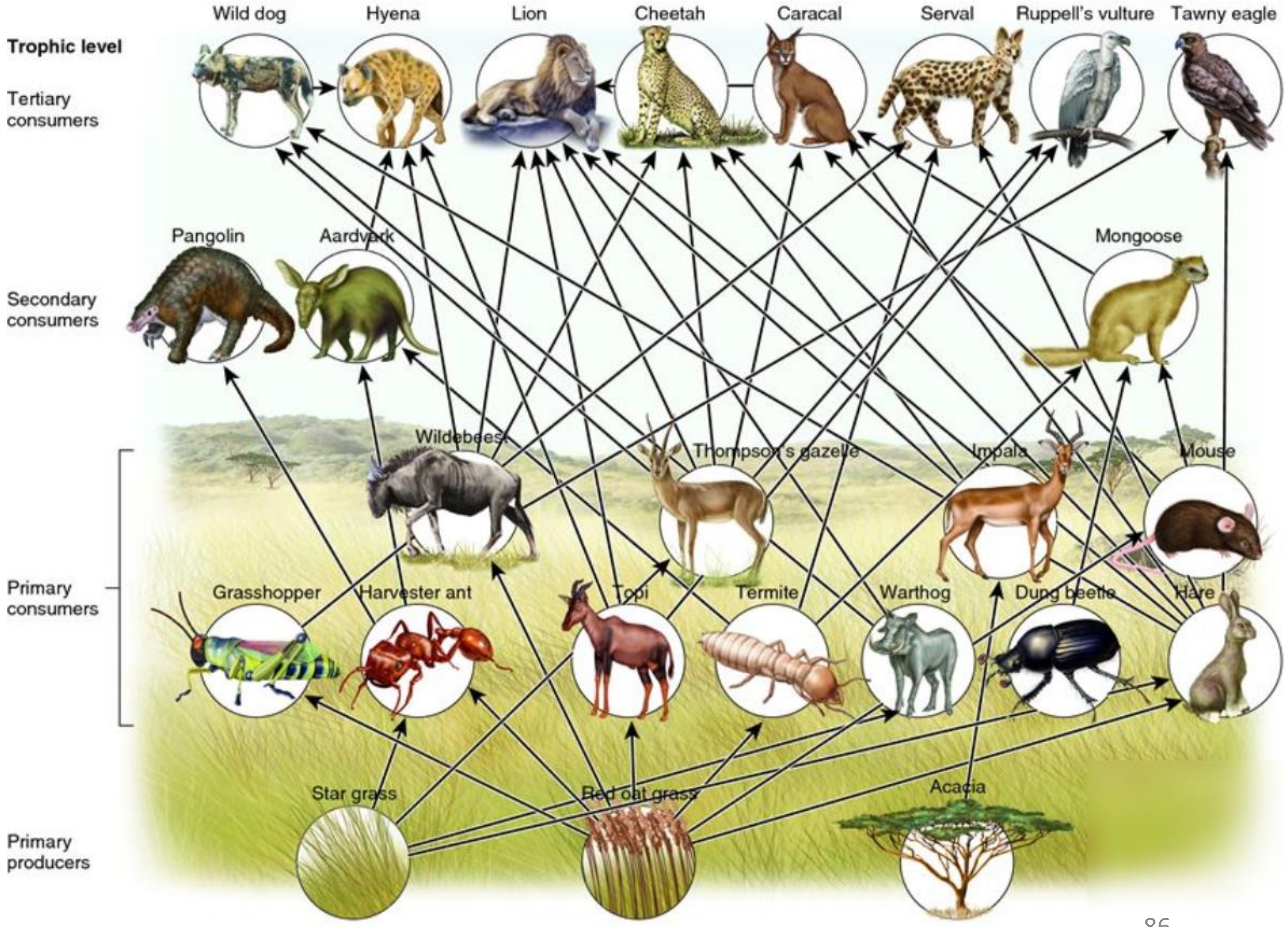


Interaction network of among different areas of the human brain (top) based on fMRI data

Source: on going work at ENS Paris-Saclay

Biological networks

Ecological network:
Land's food web

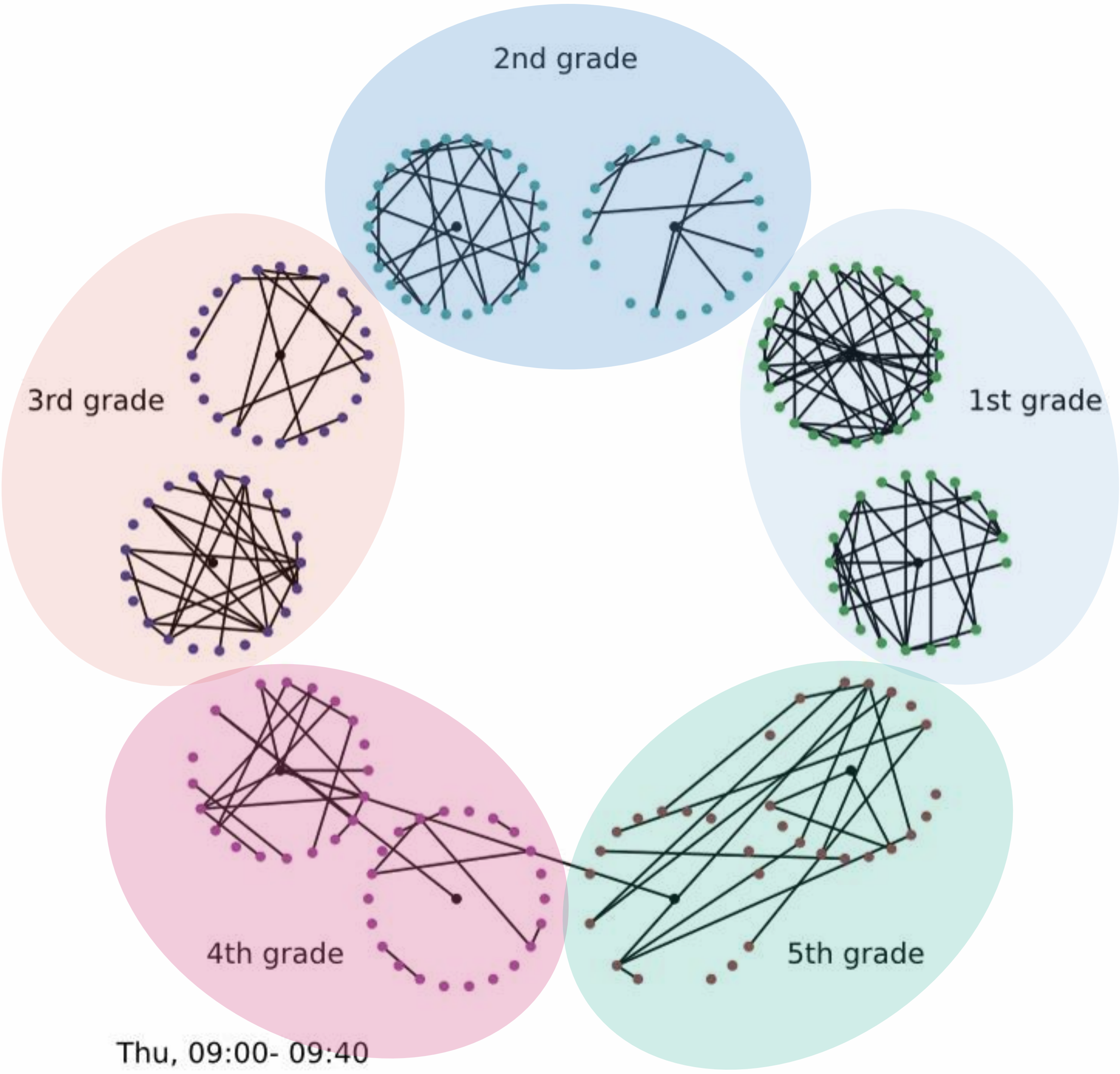


Source: https://www.thinglink.com/scene/731525498533838848

Dynamic networks

… contact and friendship nets

Face-to-face contacts among
pupils of a French primary school
… **during a school day**

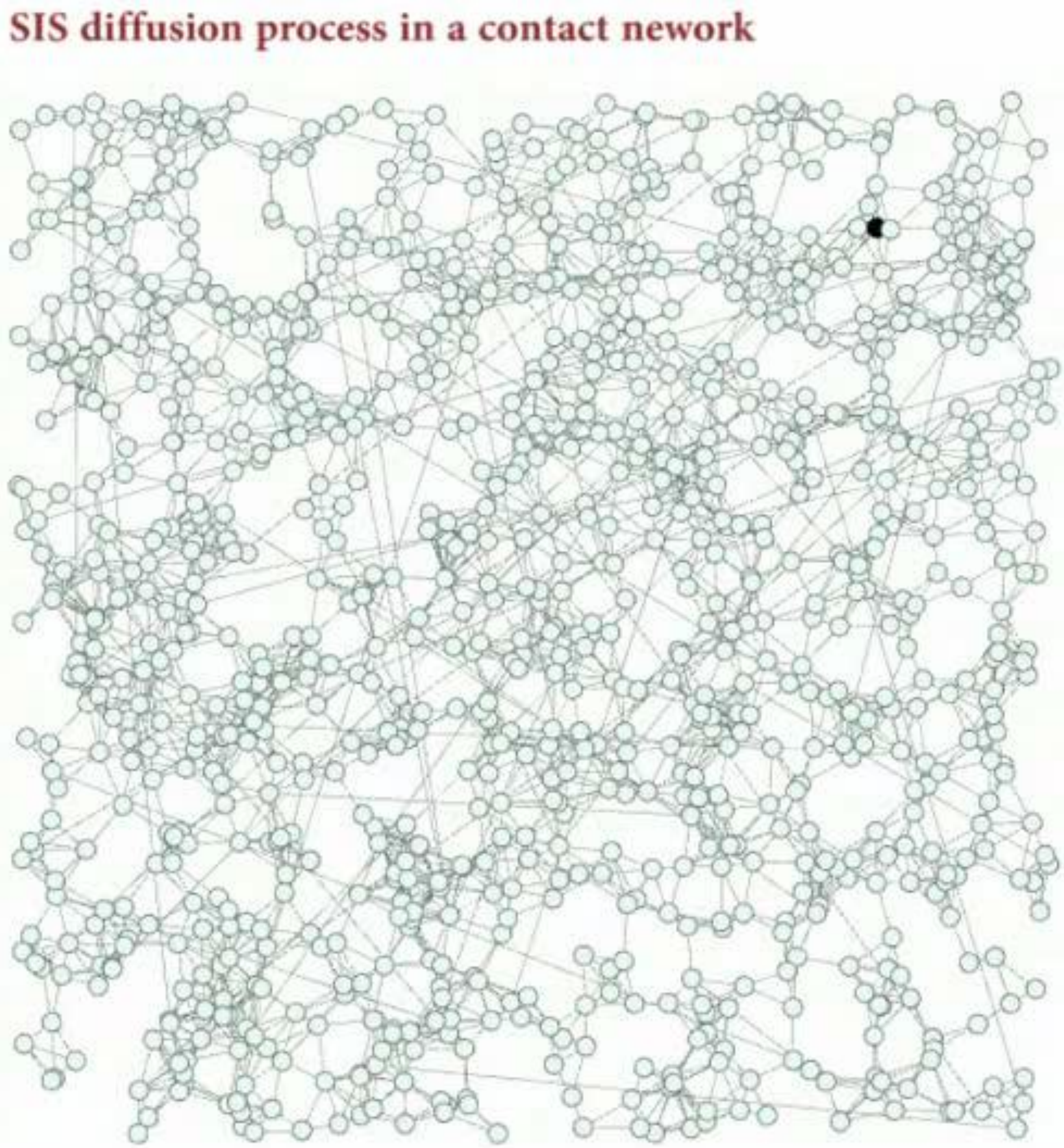


Source: J. Stehlé, (2011). High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLOS One [*https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023176*](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023176)

Dynamic networks
… diffusion networks

A dynamic process on a static network

An epidemic simulation of an recurrent epidemic (SIS)
… during a period of time

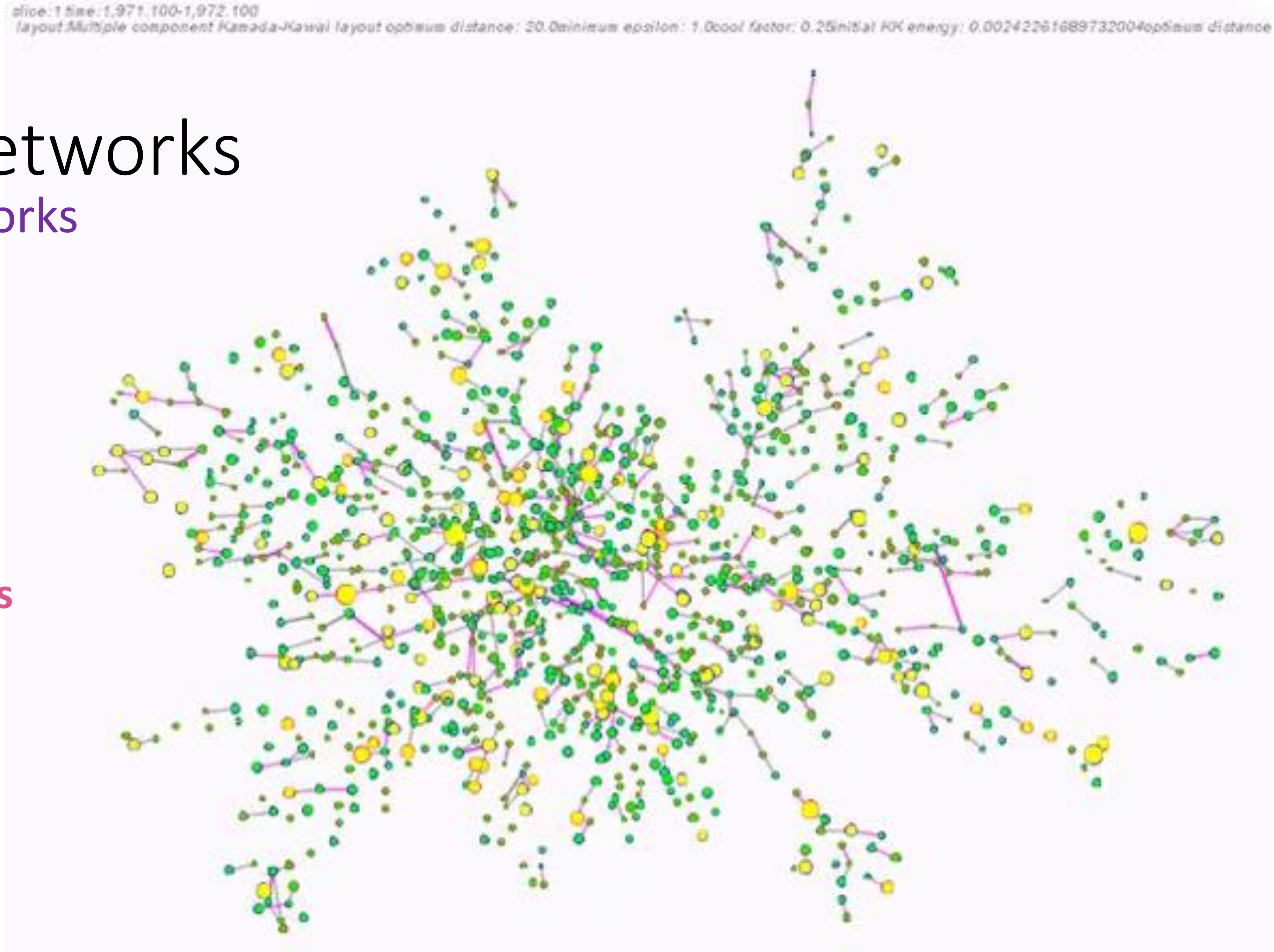


Source: research work at CMLA

Dynamic networks

… diffusion networks

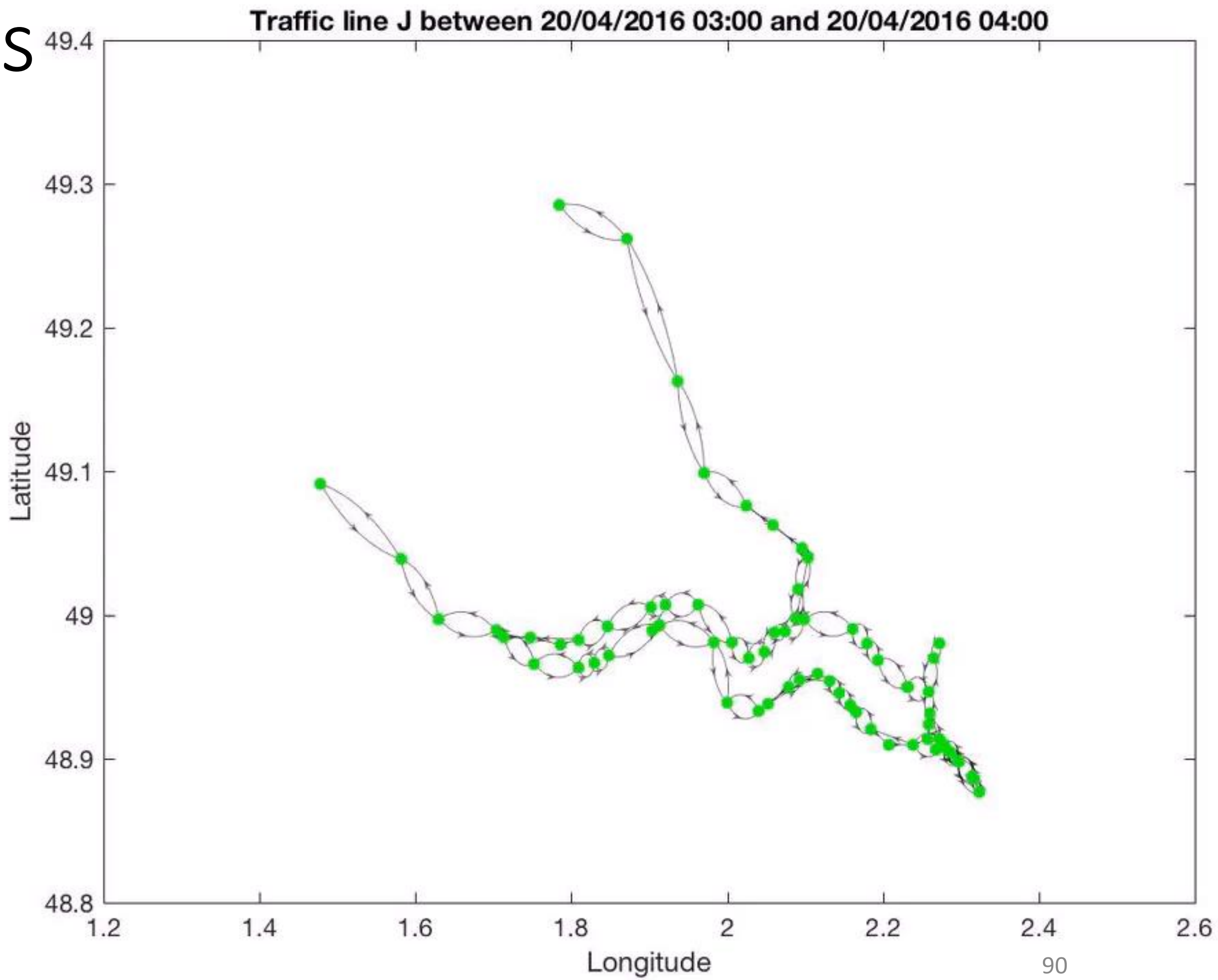
Obesity as a contagion
In a social network
… in period of 32 years



Source: N.A. Christakis et al. (2007). The Spread of Obesity in a Large Social Network over 32 Years, N.E.J. of Medicine

Dynamic networks
… transportation networks

Line J of Transilien
… during one day



Source: MORANE project, SNCF-Center Borelli

Discussion

Q & A