Networks Structure and Dynamics
1. Introduction

Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

first_name.last_name@lip6.fr

September 20th 2016

How it works?

Teachers

Lionel Tabourier lionel.tabourier@lip6.fr
Fabien Tarissan fabien.tarissan@lip6.fr

Webpage

http://lioneltabourier.fr/teaching_en.html/

Language

Course in English, questions in both English and French

Assessment

- Assignments : 40%
  1 project mid-November, 1 practical work mid-December
- Exam : 60%

Outline

Outline

Networks / graphs

computer science : internet, web, peer-to-peer, usages, etc

but also :

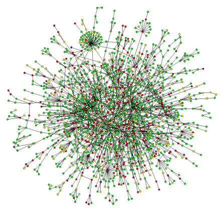
social sciences : collaboration, friendships, sexual contacts, trades, etc

biology : brain, genes, proteins, ecosystems, etc

linguistic : synonyms, co-occurrences, etc

transport : road-network, flights, power, etc

etc, etc



relationship networks

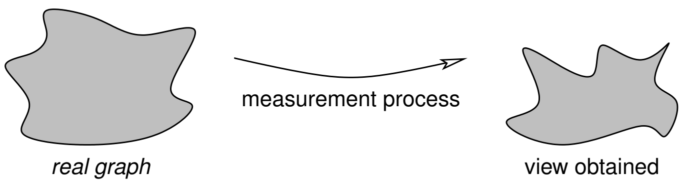
very different contexts

Networks / graphs

Various contexts but...

- Common properties
  → Similarities
- Common problems

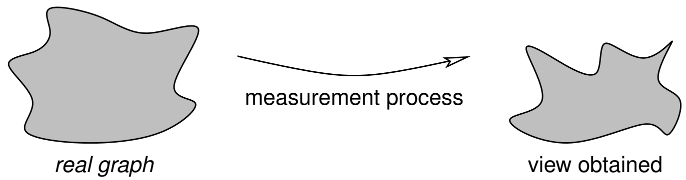
Problematic – Measurement



Comprehensive measurement of the object impossible because of :

- Size
- Technical constraints
- . . .

Problematic – Metrology



What can we say about the real object from the measurement?

Comparing to polling organization

- Significant view: difficult issue...
- Limited choice of the measurement processes

impact on the properties of the view?
measures targeting specific properties? etc

Problematic – Analysis

describe

extract meaningful information

statistics **structural properties**

density
degrees
local density
correlations
...

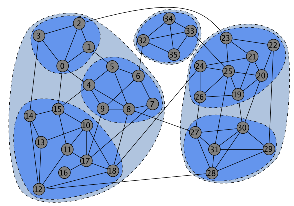
Problematic – Analysis

describe

extract meaningful information

statistics **structural properties**

density
degrees
local density
correlations
...



Problematic – Modelling

Model : generating graph

***realistic* graphs generation**
(*i.e.* having observed properties)

motivations : formal approaches, simulations, explanations

Problematic – Algorithmics

Size

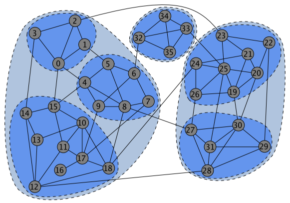
In general, large graphs (typically from 10^3 to 10^9 nodes)
$\rightarrow$ classical properties to be reconsidered
$\rightarrow$ space constraints

Problematic – Algorithmics

Size

In general, large graphs (typically from 10^3 to 10^9 nodes)
$\rightarrow$ classical properties to be reconsidered
$\rightarrow$ space constraints

+ specific problems



Dynamics

Appearance/disappearance through time

- of nodes
- of links

In this course...

- Panorama of the questions of the field
  - Models
  - Communities
  - Metrology
  - ...
- More focused on some cases:
  - Internet
  - Contact networks

Outline

Graph

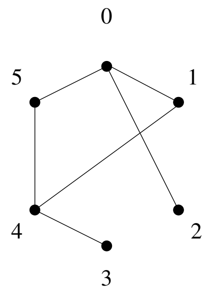
Networks will be represented by *graphs*

A graph $G = (V, E)$ is a couple of sets.

- V is a set of *vertices* (or *nodes*) ***fr:*** *sommet ou nœud*
- $E \subseteq (V \times V)$ is a set of *edges* (or *links*) ***fr:*** *lien, arête*

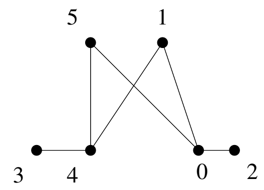
Example

- $V = \{0, 1, 2, 3, 4, 5\}$
- $E = \{(0, 1), (0, 2), (3, 4), (4, 5), (5, 0), (1, 4)\}$



Example

- $V = \{0, 1, 2, 3, 4, 5\}$
- $E = \{(0, 1), (0, 2), (3, 4), (4, 5), (5, 0), (1, 4)\}$



Warning: The graph should not be confused with its drawing !

Notations

We note :

- $n = |V|$ the number of nodes
- $m = |E|$ the number of edges

u and v are neighbours if there is an edge between them.

Degree : $d^\circ(v)$: number of neighbours of v *(**fr:** degré)*

Average degree, density

graph average degree, $d^\circ(G)$

average degree of all the nodes

Average degree, density

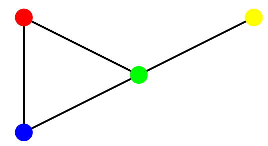
graph density, δ *(fr: densité du graphe)*

= probability that a link exists

= how connected the nodes are

$$\delta = \frac{2m}{n(n-1)}$$

Average degree, density



degree : **2, 2, 3, 1** ; average degree 2

$n = 4$, $m = 4$, $\delta = \frac{8}{12} = 0.66..$

Directed and undirected graphs

Undirected graph *(graphe non-orienté)* : $(u, v) = (v, u)$
Directed graph *(graphe orienté)* : $(u, v) \neq (v, u)$

In-degree *(degré entrant)* : $d^+(v)$
Out-degree *(degré sortant)* : $d^-(v)$

Except if otherwise specified: undirected graphs

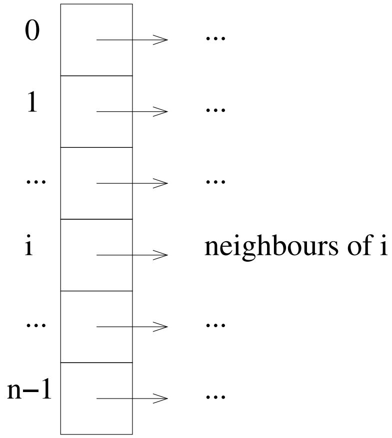
Adjacency matrix

$$\begin{array}{c} \\ 0 \\ 1 \\ \dots \\ n-1 \end{array} \begin{array}{c} 0 \quad 1 \quad \dots \quad n-1 \\ \left(\begin{array}{ccc} & & \\ & \dots & \\ & & \end{array} \right) \end{array}$$

Box (i, j) :

- 1 if there is an edge between i and j
- 0 otherwise

Adjacency list



Advantages, drawbacks

Temporal and spatial complexities

	Matrix	List
Existence of an edge		
List the neighbours of a node		
Size		

In general, real world networks have low density (sparse)
$\rightarrow \mathcal{O}(m) << \mathcal{O}(n^2)$

Advantages, drawbacks

Temporal and spatial complexities

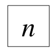
	Matrix	List
Existence of an edge	$\mathcal{O}(1)$	$\mathcal{O}(d^\circ(v))$
List the neighbours of a node	$\mathcal{O}(n)$	$\mathcal{O}(d^\circ(v))$
Size	$\mathcal{O}(n^2)$	$\mathcal{O}(m)$

In general, real world networks have low density (sparse)
$\rightarrow \mathcal{O}(m) << \mathcal{O}(n^2)$

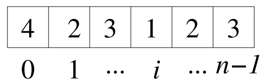
Compact storage

We store :

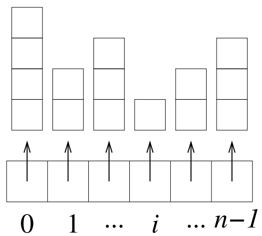
- the number of nodes



- degree stored in a table



- neighbours in a table



Compact storage

We store :

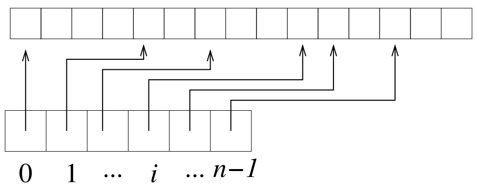
- the number of nodes

n

- degree stored in a table

4	2	3	1	2	3
0	1	...	i	...	$n-1$

- neighbours in a table



Contiguous storage : one big table for all neighbours

Outline

Density

In practice, low density for complex networks.

What means low here?

Average degree very low compared to n

Connectedness

Path *(chemin)* from u to v : sequence of edges $(u, v_1), (v_1, v_2), \ldots, (v_{k-1}, v)$

Length *(longueur)* = k (number of edges)

Connected component *(composante connexe)* : maximal set of nodes such that $\exists$ a path between all pairs of nodes.

Connected graph : only one connected component

Connectedness of directed graphs

Directed path *(chemin orienté)* from u to v : sequence of arcs $(u, v_1), (v_1, v_2), \ldots, (v_{k-1}, v)$

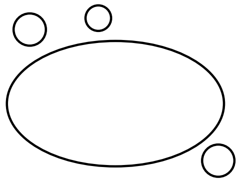
Strongly connected component *(composante fortement connexe)* : maximal set of nodes such that $\exists$ a directed path from any node to any other of the set.

Weakly connected component *(composante faiblement connexe)* : maximal set of nodes such that $\exists$ a path between all pairs of nodes in the graph where arcs are replaced by undirected links.

Connectedness

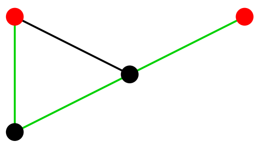
For complex networks

In general, a giant component
→ contains most nodes



Distance

Path from u to v = sequence of edges u...v

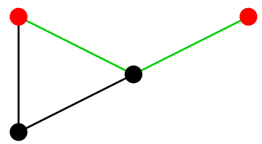


a path of length 3

Distance

Path from u to v = sequence of edges u...v

distance $d(u, v)$ = length of *one* shortest path



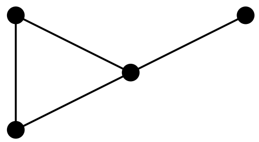
shortest path of length 2 $\Rightarrow$ distance = 2

Distance

Path from u to v = sequence of edges u...v

distance $d(u, v)$ = length of *one* shortest path

diameter Δ = longest distance between all pairs of nodes



diameter = 2

Distances and connectedness

Average distance : average distance for all pairs of nodes

Connectedness?
Distance defined for two nodes of the same connected component

In practice
Average distance, diameter :
$\rightarrow$ in the largest connected component

Distances and connectedness

Average distance : average distance for all pairs of nodes

Connectedness?
Distance defined for two nodes of the same connected component

In practice
Average distance, diameter :
$\rightarrow$ in the largest connected component

Distance

For complex networks
In general, short distances ($\sim \log(n)$)

Milgram experience
“Six degree of separation”
Kevin Bacon game
Erdös number

Degree distribution (1/2)

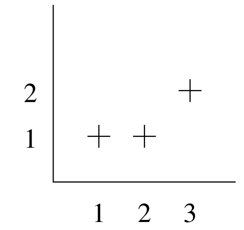
Degree distribution :
4 nodes, degrees : 2 3 3 1

Degree distribution (1/2)

Degree distribution :
4 nodes, degrees : 2 3 3 1

Distribution: how many nodes have degree k, as function of k.

$1 \rightarrow 1$, $2 \rightarrow 1$, $3 \rightarrow 2$



Power-law

Power-law *(loi de puissance)*

- $N_k \sim k^{-\alpha}$
- straight line in log-log scale

Heterogeneous distribution : close to a power-law (in our cases)

Power law

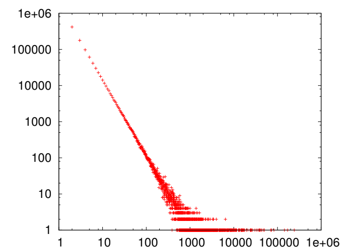
- line in log-log scale on several order of magnitude

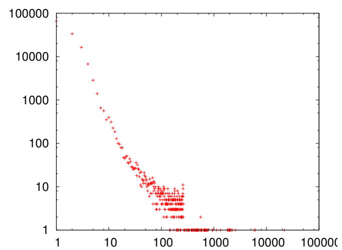
$\neq$

Heterogenous

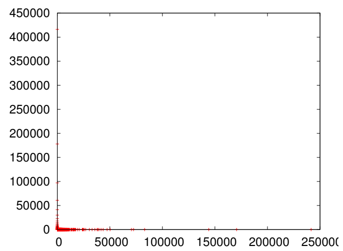
- on several order of magnitude
- close to linear in log-log scale

Example

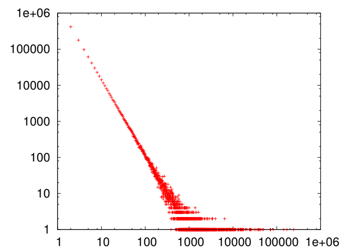




Heterogeneous distributions : log-log scale

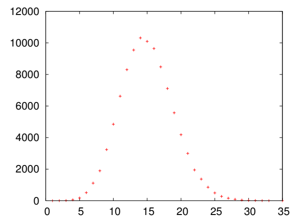


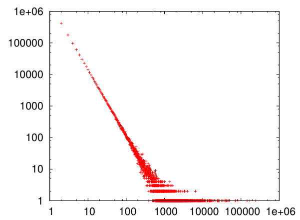
linear scale



logarithmic scale

Heterogeneous vs homogeneous distributions





Homogeneous

Idea of normality (and exceptions)

Heterogeneous

Any kind of behaviour exists $\rightarrow$ no simple idea of normality

Degree distributions (2/2)

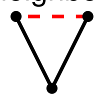
For complex networks

In general, degree distributions are heterogeneous

Clustering coefficient

clustering coefficient $cc(v)$

= probability that two neighbours of v are connected

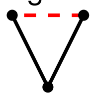


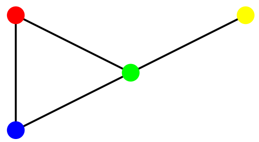
= # pairs of connected neighbours / # pairs of neighbours
= local density

Clustering coefficient

clustering coefficient cc(v)

= probability that two neighbours of v are connected





clustering coefficient : 1, 1, $\frac{1}{3}$, undefined

Clustering coefficient

Clustering coefficient of a graph :
average on all the nodes of degree ≥ 2

For complex networks

In general, high clustering

meaning several orders of magnitudes greater than density

Transitive ratio

Another coefficient to measure the local density of a graph:
transitive ratio

$$\mathrm{tr}(G) = \frac{3N_\Delta}{N_\vee}$$

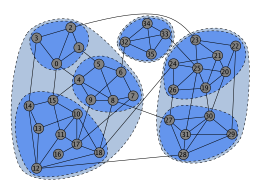
N_Δ : number of triangles of the graph
$N_\vee$: number of connected triplets

Careful: definitions vary in the literature
(sometimes called *global clustering*)

Defining communities

Goal

Looking for an internal structure of the graph.



Definition

- **intuitive:** people sharing a common interest, web pages having similar contents...
- **structural:** zone of the graph with a high density of links

Common properties – conclusion

Most complex networks share common properties :

density	low
connectivity	giant component
distances	low
degree distribution	heterogeneous
clustering	high
communities	yes