Structure et dynamique des réseaux

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

prenom.nom@lip6.fr

16 septembre 2014

Fonctionnement du cours

Enseignants

Clémence Magnien `clemence.magnien@lip6.fr`
Lionel Tabourier `lionel.tabourier@lip6.fr`
Fabien Tarissan `fabien.tarissan@lip6.fr`

Page du cours

`http://www-rp.lip6.fr/~magnien/sdr.html`

Évaluation

- Contrôle continu : 40%
  - deux TP notés
- Examen : 60%

Outline

Outline

Réseaux / graphes

informatique : internet, web, pair-à-pair, usages, etc

mais aussi :

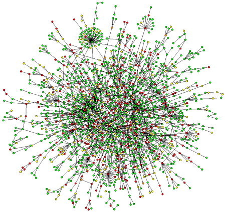
sciences sociales : collaboration, amitié, contacts sexuels, échanges, économie, etc

biologie : cerveau, gènes, protéines, écosystèmes, etc

linguistique : synonymie, co-occurrence, etc

transport : routier, aérien, électrique, etc

etc, etc



réseaux de relations

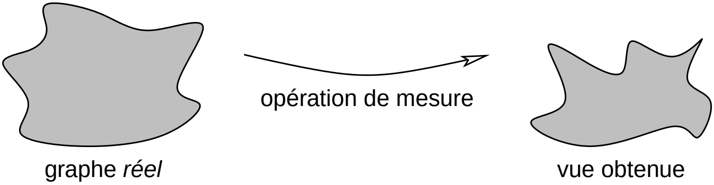
contextes très différents

Réseaux / graphes

Contextes différents, mais

- Propriétés communes
  $\rightarrow$ Ressemblance
- Problématiques communes

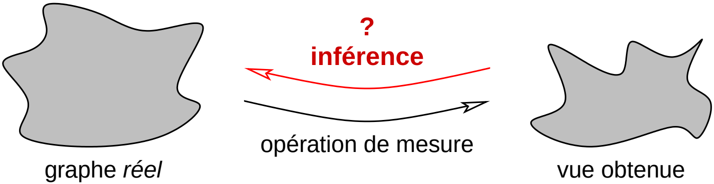
Problématique – Mesure



Impossible de mesurer l'objet complet :

- Taille
- Contraintes techniques
- ...

Problématique – Métrologie



que peut on dire sur l'objet réel à partir de la mesure ?

Comparer aux instituts de sondage

- Vue représentative : complexe
- Ici, choix limité sur la procédure de mesure

impact sur les propriétés observées ?
impact des propriétés sur la vue?
mesures ciblant certaines propriétés?

Problématique – Analyse

décrire

extraire de l'information pertinente

statistiques

densité
degrés
densité locale
corrélations
...

structure

Problématique – Analyse

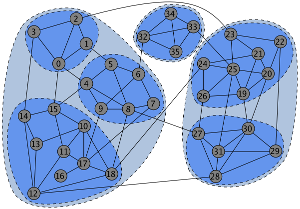
décrire

extraire de l'information pertinente

statistiques

densité
degrés
densité locale
corrélations
...

structure



Problématique – Modélisation

Modèle : générateur de graphes

génération de graphes *réalistes*
(*i.e.* ayant les propriétés observées)

motivations : approches formelles, simulation, explication

Problématique – Algorithmique

Taille

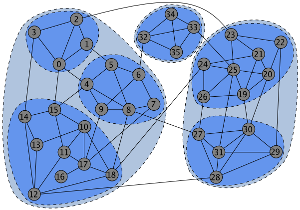
$\rightarrow$ problèmes classiques à revisiter
$\rightarrow$ restrictions en espace

Problématique – Algorithmique

Taille

→ problèmes classiques à revisiter
→ restrictions en espace

+ problèmes spécifiques



Dynamique

Apparition/disparition au fil du temps

- de nœuds
- de liens

Dans ce cours

- Panorama des questions du domaine
  - Modèles
  - Communautés
  - Métrologie
  - ...
- Trois cas d'intérêt particuliers
  - Internet
  - Réseaux de contacts
  - P2P

Outline

Graphe

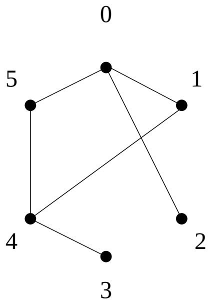
On modélise les réseaux par des *graphes*

Un graphe $G = (V, E)$ est un couple d'ensembles.

- V est l'ensemble des *sommets* (ou *nœuds*)
- $E \subseteq (V \times V)$ est l'ensemble des *arêtes* (ou *liens*).

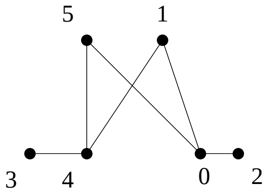
Exemple

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



Exemple

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



Attention à ne pas confondre le graphe et le dessin !

Notations

On note :

- $n = |V|$ le nombre de sommets
- $m = |E|$ le nombre d'arêtes

u et v sont voisins s'il y a une arête entre eux.

Degré : $d°(v)$: nombre de voisins de v

Degré moyen, densité

degré moyen du graphe, $d^\circ(G)$

moyenne des degrés de tous les sommets

Degré moyen, densité

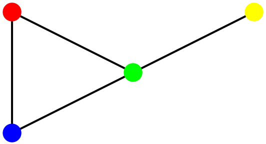
densité du graphe, δ

= probabilité d'existence de tout lien

= à quel point tout le monde est lié

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

Degré moyen, densité



degrés : **2, 2, 3, 1** ; degré moyen 2

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

Graphes orientés/non orientés

Graphe non-orienté : $(u, v) = (v, u)$
Graphe orienté : $(u, v) \neq (v, u)$

Degré entrant : $d^+(v)$
Degré sortant : $d^-(v)$

Sauf mention contraire : graphes non-orientés

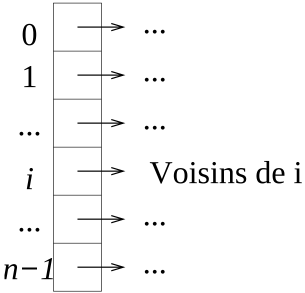
Matrice d'adjacence

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

Case (i, j) :

- 1 s'il y a une arête entre i et j
- 0 sinon

Listes d'adjacence



Avantages, inconvénients

Existence d'une arête
Lister les voisins d'un sommet
Taille

En général, les graphes de terrain sont peu denses
$\rightarrow \mathcal{O}(m) << \mathcal{O}(n^2)$

Avantages, inconvénients

	Matrice	Listes
Existence d'une arête	$\mathcal{O}(1)$	$\mathcal{O}(d^\circ(v))$
Lister les voisins d'un sommet	$\mathcal{O}(n)$	$\mathcal{O}(d^\circ(v))$
Taille	$\mathcal{O}(n^2)$	$\mathcal{O}(m)$

En général, les graphes de terrain sont peu denses

$$\rightarrow \mathcal{O}(m) << \mathcal{O}(n^2)$$

Stockage compact

On stocke :

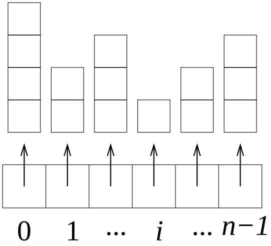
- le nombre de sommets

$\boxed{n}$

- leurs degrés dans un tableau

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

- les voisins dans un tableau



Stockage compact

On stocke :

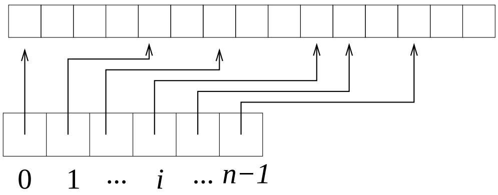
- le nombre de sommets

$\boxed{n}$

- leurs degrés dans un tableau

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

- les voisins dans un tableau



Allocation contigue : un seul gros tableau pour tous les voisins

Outline

Densité

En pratique, densité faible pour les graphes de terrain.

Faible ?

Degré moyen très faible par rapport à n

Connexité

Chemin de u à v : suite d'arêtes $(u, v_1), (v_1, v_2), \ldots, (v_{k-1}, v)$
Longueur = k (nombre d'arêtes)

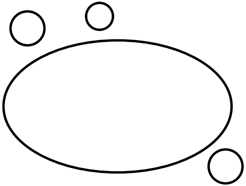
Composante connexe : ensemble maximal de sommets t. q. $\exists$ un chemin entre toutes les paires de sommets.

Graphe connexe : une seule composante connexe

Connexité

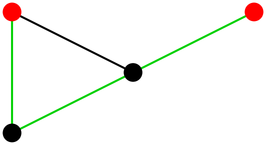
Pour les graphes de terrain

En général, composante géante
$\rightarrow$ Contient la plupart des sommets



Distance

chemin de u à v = suite d'arêtes u...v

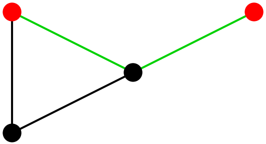


un chemin de longueur 3

Distance

chemin de u à v = suite d'arêtes u...v

distance $d(u, v)$ = longueur d'*un* plus court chemin



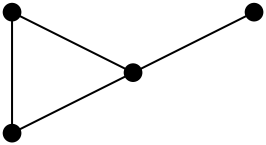
un plus court chemin ; longueur 2 $\Rightarrow$ distance = 2

Distance

chemin de u à v = suite d'arêtes u...v

distance $d(u,v)$ = longueur d'*un* plus court chemin

diamètre Δ = plus grande distance



diamètre = 2

Distances et connexité

Distance moyenne : moyenne de la distance pour toutes les paires de sommets

Connexité ?

Distance définie pour deux sommets de la même composante connexe

En pratique

Distance moyenne, diamètre :
→ dans la plus grande composante connexe

Distances et connexité

Distance moyenne : moyenne de la distance pour toutes les paires de sommets

Connexité ?

Distance définie pour deux sommets de la même composante connexe

En pratique

Distance moyenne, diamètre :
$\rightarrow$ dans la plus grande composante connexe

Distance

Pour les graphes de terrain

En général, distances courtes ($\sim \log(n)$)

Expérience de Milgram
“Six degrés de séparation”
Kevin Bacon game

Distribution des degrés (1/2)

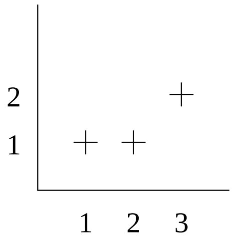
Distribution des degrés :
4 nœuds, degrés : 2 3 3 1

Distribution des degrés (1/2)

Distribution des degrés :
4 nœuds, degrés : 2 3 3 1

Distribution : combien de nœuds ont degré k, en fonction de k.

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



Loi de puissance (power-law)

Loi de puissance

- $N_k \sim k^{-\alpha}$
- droite en échelle log-log

Distribution hétérogène : proche d'une loi de puissance

Loi de puissance

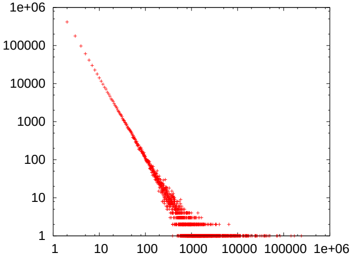
- droite en échelle log-log sur plusieurs ordres de grandeur

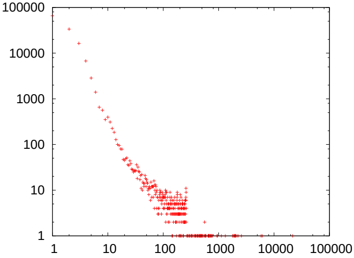
$\neq$

Hétérogène

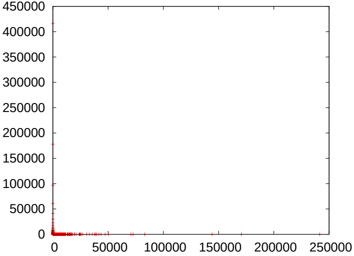
- plusieurs ordres de grandeur
- proche d'une droite en échelle log-log

Exemple

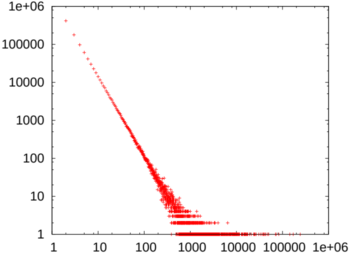

1e+06
100000
10000
1000
100
10
1
1
10
100
1000
10000
100000
1e+06


100000
10000
1000
100
10
1
1
10
100
1000
10000
100000

Distributions hétérogènes : échelle log-log

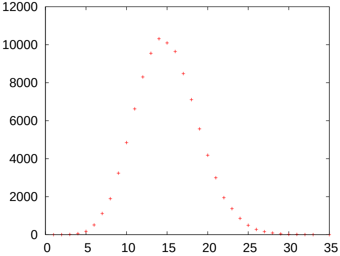


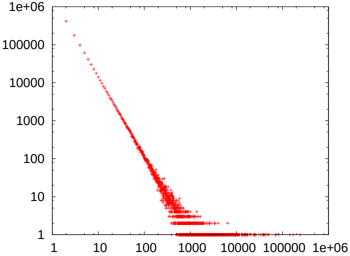
échelle linéaire



échelle logarithmique

Distributions hétérogènes vs homogènes





Homogène

Notion de normalité (et d'exceptions)

Hétérogène

Tous les comportements existent

→ pas de notion de normalité

Distribution des degrés (2/2)

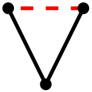
Pour les graphes de terrain

En général, distributions des degrés hétérogènes

Coefficient de clustering

coefficient de clustering $\mathrm{cc}(v)$

= probabilité que deux voisins de v soient reliés

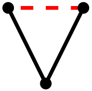


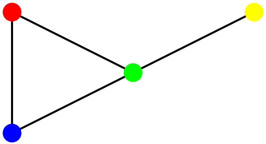
= # paires de voisins reliés / # paires de voisins
= densité locale

Coefficient de clustering

coefficient de clustering cc(v)

= probabilité que deux voisins de v soient reliés





coefficients de clustering : 1, 1, $\frac{1}{3}$, indéfini

Coefficient de clustering

Coefficient de clustering du graphe :
moyenne sur tous les sommets de degré ≥ 2

Pour les graphes de terrain

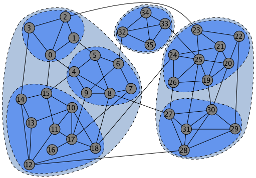
En général, clustering fort

Plusieurs ordres de grandeur au dessus de la densité

Définir les communautés

But

Rechercher une structure interne au graphe.



Définition

- intuitive: personnes partageant un intérêt commun, pages web au contenu similaire...
- structurelle: zone du graphe dense en liens

Propriétés communes – conclusion

La plupart des graphes de terrain ont des propriétés communes :

densité	faible
connexité	comp. géante
distances	faibles
degrés	hétérogènes
clustering	fort
communautés	oui