

Introduction à l'analyse de réseaux complexes

1 - La notion de réseau complexe (partie B)

Lionel Tabourier

LIP6 – Université Pierre et Marie Curie / CNRS

Cours à l'Université Babeş-Bolyai
Mai 2017

Outline

- 1 La distance entre deux nœuds
- 2 Les petits-mondes
- 3 Le clustering
- 4 Notion de communautés

Plan

- 1 La distance entre deux nœuds
- 2 Les petits-mondes
- 3 Le clustering
- 4 Notion de communautés

Qu'est-ce que la distance entre deux nœuds?

Définition: un **chemin** de i à j est une séquence de liens permettant d'aller d'un nœud i à un nœud j du graphe

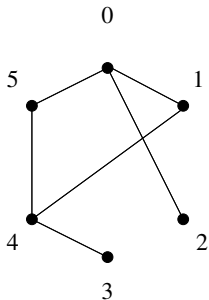
Définition: la **longueur** d'un chemin est le nombre de liens de ce chemin

Définition: la **distance** entre deux nœuds est la longueur du plus court chemin

remarque : n'a de sens que dans une composante connexe

Qu'est-ce que le diamètre un graphe?

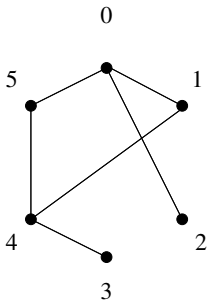
Définition: le **diamètre** d'un graphe est la plus grande distance entre deux nœuds d'un graphe



Sur cet exemple, diamètre =

Qu'est-ce que le diamètre un graphe?

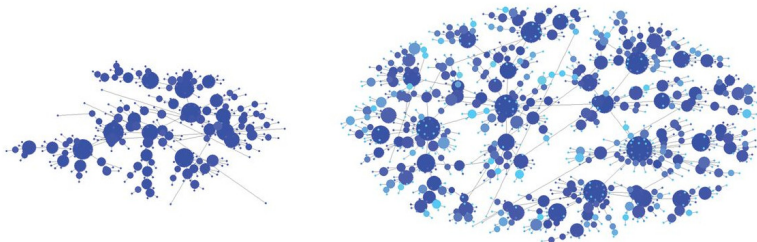
Définition: le **diamètre** d'un graphe est la plus grande distance entre deux nœuds d'un graphe



Sur cet exemple, diamètre = 4

Croissance d'un réseau

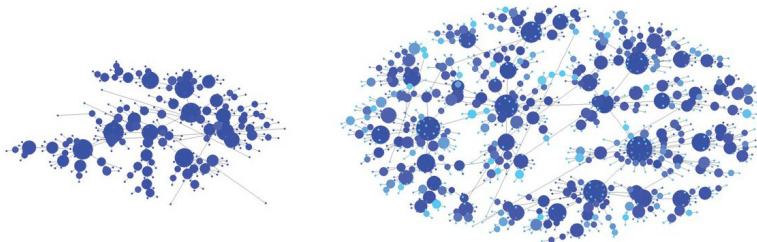
Allure typique d'un réseau (de collaborations) croissant:



⇒ **évolution du diamètre avec N ?**
croissance logarithmique

Croissance d'un réseau

Allure typique d'un réseau (de collaborations) croissant:



⇒ **évolution du diamètre avec N ?**
croissance logarithmique

Plan

- 1 La distance entre deux nœuds
- 2 Les petits-mondes**
- 3 Le clustering
- 4 Notion de communautés

L'expérience des petits-mondes de Milgram

il existe des chemins courts entre deux individus quelconques

En 1967: depuis au moins le début du XX^{ème} siècle
l'idée de petit-monde existe (Frigyes Karinthy)
→ approche théorique de Sola Pool et Kochen

Stanley Milgram: spécialiste de la psychologie sociale
(célèbre pour une expérience sur l'obéissance à l'autorité)
réalise une série d'expériences de validation

L'expérience des petits-mondes de Milgram

il existe des chemins courts entre deux individus quelconques

En 1967: depuis au moins le début du XX^{ème} siècle
l'idée de petit-monde existe (Frigyes Karinthy)
→ approche théorique de Sola Pool et Kochen

Stanley Milgram: spécialiste de la psychologie sociale
(célèbre pour une expérience sur l'obéissance à l'autorité)
réalise une série d'expériences de validation

L'expérience des petits-mondes de Milgram

Les premières expériences des petits-mondes

The small-world problem, *Psychology today*, 1967

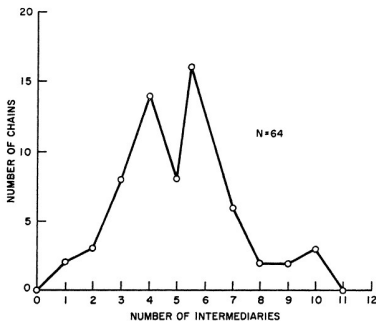
Deux études, même protocole:

- plusieurs sources (Wichita, Kansas / Omaha, Nebraska)
- une destination (Cambridge / Sharon, Massachusetts)
- informations: nom, métier (femme au foyer / courtier)
- **règle:** passer à une unique personne de son réseau de connaissances (*acquaintance*) uniquement

on suit qui a relayé le paquet dans ce protocole

Une première image des petits-mondes

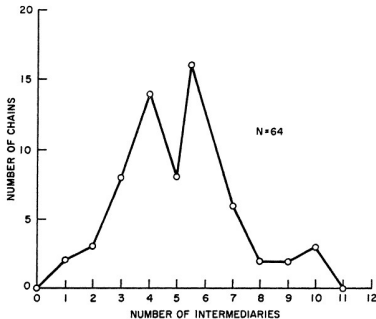
Résultat le plus célèbre: médiane de 5 intermédiaires
(contre-intuitif?)



Progression quasi-géométrique du nombre de relations
⇒ “Six degrés de séparation” (pas de Milgram)

Une première image des petits-mondes

Résultat le plus célèbre: médiane de 5 intermédiaires
(contre-intuitif?)



Progression quasi-géométrique du nombre de relations
⇒ “Six degrés de séparation” (pas de Milgram)

Analyse plus poussée des expériences

Fort taux de perte de messages

- nombre chaînes complétées / nombre total de chaînes:
64/296 (exp Nebraska)
- pour quelles raisons? coût pour les participants? pas de chemin évident?
- longueur maximum : 11
→ **biais de survie** tend à raccourcir les chaînes

Analyse plus poussée des expériences

Géographie des trajectoires

- **grande distance** entre villes départ/arrivée
Nebraska, Kansas: “in the plains out there” (M.)
- **sauts irréguliers** entre sources et destination
cas typiques: fort rapprochement géographique, puis circulation autour de la destination
- beaucoup de chaînes “meurent” proches de la destination

Analyse plus poussée des expériences

Pas seulement une distance ...

- **capacité de “routage humain”**
“depends on the wisdom of the choice” (M.)
- **des nœuds plus importants que d’autres:**
(exp. Nebraska) 50 % des chaînes terminées par M.Jacobs, M. Brown ou M. Jones
 - Jacobs dans le cercle dans le lieu de résidence
 - Jones dans le cercle du travail

→ même principe que ce qui se passe sur Internet

Analyse plus poussée des expériences

Pas seulement une distance ...

- **capacité de “routage humain”**
“depends on the wisdom of the choice” (M.)
- **des nœuds plus importants que d’autres:**
(exp. Nebraska) 50 % des chaînes terminées par M.Jacobs, M. Brown ou M. Jones
 - Jacobs dans le cercle dans le lieu de résidence
 - Jones dans le cercle du travail

→ même principe que ce qui se passe sur Internet

Analyse plus poussée des expériences

Choix du correspondant

- **biais sur le genre:**
femmes envoient 3x plus vers les femmes
hommes envoient 4x plus vers les hommes
- **communautés et isolation:**
une des questions de départ de l'expérience
des cercles isolés?

Expériences complémentaires selon ethnies source/destination

Analyse plus poussée des expériences

Choix du correspondant

- **biais sur le genre:**
femmes envoient 3x plus vers les femmes
hommes envoient 4x plus vers les hommes
- **communautés et isolation:**
une des questions de départ de l'expérience
des cercles isolés?

Expériences complémentaires selon ethnies source/destination

Faire l'expérience des petits-mondes aujourd'hui

Réseaux de collaborations scientifiques: *nombre d'Erdős*

Réseaux de collaborations d'acteurs: *oracle of Bacon*
données IMDB, calcul de plus court chemin

Sur Facebook: distance moyenne entre 2 comptes:
en 2011: 4.50 (0.72 milliards de comptes actifs)
en 2016: 4.57 (1.59 milliards de comptes actifs)

Mais **pas du routage** dans ces mesures, pas de prise en compte de la force des liens

Recherche d'information sur le web: *Wikispeedia*

Faire l'expérience des petits-mondes aujourd'hui

Réseaux de collaborations scientifiques: *nombre d'Erdős*

Réseaux de collaborations d'acteurs: *oracle of Bacon*
données IMDB, calcul de plus court chemin

Sur Facebook: distance moyenne entre 2 comptes:
en 2011: 4.50 (0.72 milliards de comptes actifs)
en 2016: 4.57 (1.59 milliards de comptes actifs)

Mais **pas du routage** dans ces mesures, pas de prise en compte de la force des liens

Recherche d'information sur le web: *Wikispeedia*

Faire l'expérience des petits-mondes aujourd'hui

Réseaux de collaborations scientifiques: *nombre d'Erdős*

Réseaux de collaborations d'acteurs: *oracle of Bacon*
données IMDB, calcul de plus court chemin

Sur Facebook: distance moyenne entre 2 comptes:
en 2011: 4.50 (0.72 milliards de comptes actifs)
en 2016: 4.57 (1.59 milliards de comptes actifs)

Mais **pas du routage** dans ces mesures, pas de prise en compte de la force des liens

Recherche d'information sur le web: *Wikispeedia*

Faire l'expérience des petits-mondes aujourd'hui

Réseaux de collaborations scientifiques: *nombre d'Erdős*

Réseaux de collaborations d'acteurs: *oracle of Bacon*
données IMDB, calcul de plus court chemin

Sur Facebook: distance moyenne entre 2 comptes:
en 2011: 4.50 (0.72 milliards de comptes actifs)
en 2016: 4.57 (1.59 milliards de comptes actifs)

Mais **pas du routage** dans ces mesures, pas de prise en compte de la force des liens

Recherche d'information sur le web: *Wikispeedia*

Faire l'expérience des petits-mondes aujourd'hui

Réseaux de collaborations scientifiques: *nombre d'Erdős*

Réseaux de collaborations d'acteurs: *oracle of Bacon*
données IMDB, calcul de plus court chemin

Sur Facebook: distance moyenne entre 2 comptes:
en 2011: 4.50 (0.72 milliards de comptes actifs)
en 2016: 4.57 (1.59 milliards de comptes actifs)

Mais **pas du routage** dans ces mesures, pas de prise en compte de la force des liens

Recherche d'information sur le web: *Wikispeedia*

Plan

- 1 La distance entre deux nœuds
- 2 Les petits-mondes
- 3 Le clustering**
- 4 Notion de communautés

Qu'est-ce que le clustering?

“Les amis de mes amis sont mes amis”

Point de vue graphe: densité de **triangles** dans le graphe.

Plusieurs mesures possibles:

- *clustering*:

$$cc(x) = \frac{3\Delta(x)}{\wedge(x)}$$

- *transitivité*:

$$tr(G) = \frac{3\Delta(G)}{\wedge(G)}$$

remarque: moyenne du clustering \neq transitivité

Qu'est-ce que le clustering?

“Les amis de mes amis sont mes amis”

Point de vue graphe: densité de **triangles** dans le graphe.

Plusieurs mesures possibles:

- *clustering*:

$$cc(x) = \frac{3\Delta(x)}{\wedge(x)}$$

- *transitivité*:

$$tr(G) = \frac{3\Delta(G)}{\wedge(G)}$$

remarque: moyenne du clustering \neq transitivité

Qu'est-ce que le clustering?

“Les amis de mes amis sont mes amis”

Point de vue graphe: densité de **triangles** dans le graphe.

Plusieurs mesures possibles:

- *clustering*:

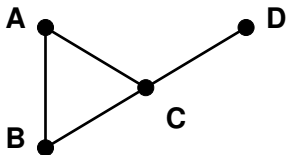
$$cc(x) = \frac{3\Delta(x)}{\wedge(x)}$$

- *transitivité*:

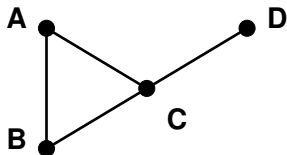
$$tr(G) = \frac{3\Delta(G)}{\wedge(G)}$$

remarque: moyenne du clustering \neq transitivité

Qu'est-ce que le clustering?

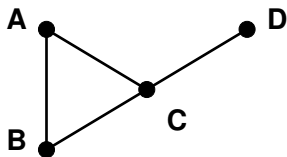


Qu'est-ce que le clustering?



$$cc(A) = cc(B) = 1, cc(C) = \frac{1}{3}, cc(D) \text{ non-défini}$$
$$\Rightarrow \overline{cc} = \frac{7}{9}$$

Qu'est-ce que le clustering?



$$cc(A) = cc(B) = 1, cc(C) = \frac{1}{3}, cc(D) \text{ non-défini}$$
$$\Rightarrow \overline{cc} = \frac{7}{9}$$

$$tr(G) = \frac{1}{5}$$

Interpréter statistiquement le clustering

Le clustering est faible/fort par rapport à quoi?

- **probabilité qu'un lien existe: densité**
- **probabilité qu'un lien existe entre deux nœuds voisins: clustering**

⇒ on parle parfois de **densité locale**

Interpréter statistiquement le clustering

Le clustering est faible/fort par rapport à quoi?

- **probabilité qu'un lien existe:** densité
- **probabilité qu'un lien existe entre deux nœuds voisins:** clustering

⇒ on parle parfois de **densité locale**

Clustering dans les réseaux réels

Quelques exemples:

- Internet (CAIDA): 0.0073 (densité: 0.00015)
- Social (gowalla): 0.024 (densité: 0.00005)
- métabolique (levure): 0.055 (densité: 0.0013)

clustering \gg densité

en général, nœuds de fort degré ont faible clustering

Petit-monde: faible distance moyenne + forte densité locale

Collective dynamics of “small-world” networks
Watts & Strogatz, *Nature* 1998

Clustering dans les réseaux réels

Quelques exemples:

- Internet (CAIDA): 0.0073 (densité: 0.00015)
- Social (gowalla): 0.024 (densité: 0.00005)
- métabolique (levure): 0.055 (densité: 0.0013)

clustering \gg densité

en général, nœuds de fort degré ont faible clustering

Petit-monde: faible distance moyenne + forte densité locale

Collective dynamics of “small-world” networks
Watts & Strogatz, *Nature* 1998

Clustering dans les réseaux réels

Quelques exemples:

- Internet (CAIDA): 0.0073 (densité: 0.00015)
- Social (gowalla): 0.024 (densité: 0.00005)
- métabolique (levure): 0.055 (densité: 0.0013)

clustering \gg densité

en général, nœuds de fort degré ont faible clustering

Petit-monde: faible distance moyenne + forte densité locale

Collective dynamics of “small-world” networks
Watts & Strogatz, *Nature* 1998

Propriétés largement partagées dans les réseaux complexes.

Mais l'image du petit-monde ne suffit pas:

- allure de la distribution de degré?
- pas d'explication simple au fait que certains nœuds jouent un rôle particulier
- forte densité locale ne se résume pas au clustering
- et d'autres selon le contexte ...

Propriétés largement partagées dans les réseaux complexes.

Mais l'image du petit-monde ne suffit pas:

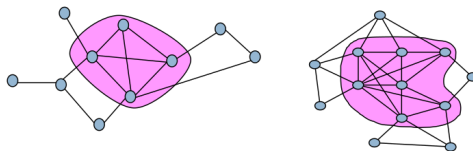
- allure de la distribution de degré?
- pas d'explication simple au fait que certains nœuds jouent un rôle particulier
- forte densité locale ne se résume pas au clustering
- et d'autres selon le contexte ...

Plan

- 1 La distance entre deux nœuds
- 2 Les petits-mondes
- 3 Le clustering
- 4 Notion de communautés**

Rechercher des groupes cohésifs

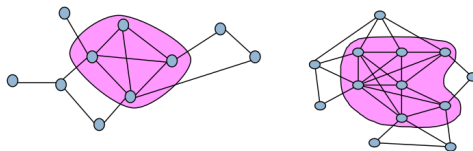
Exemple: cliques (rappel: sous-graphe complet)



définition souvent trop rigide

Rechercher des groupes cohésifs

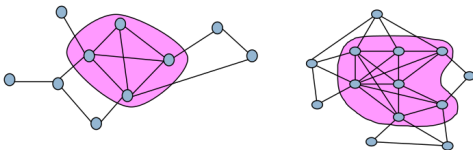
Exemple: cliques (rappel: sous-graphe complet)



définition souvent trop rigide

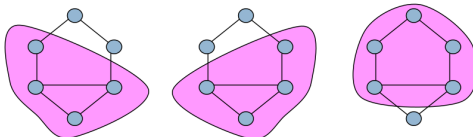
Rechercher des groupes cohésifs

Exemple: cliques (rappel: sous-graphe complet)



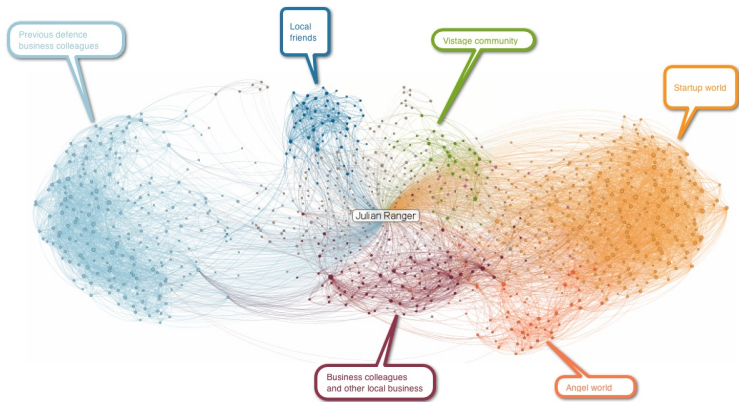
définition souvent trop rigide

Des variantes: n -club ...



Quelques pistes de réflexion

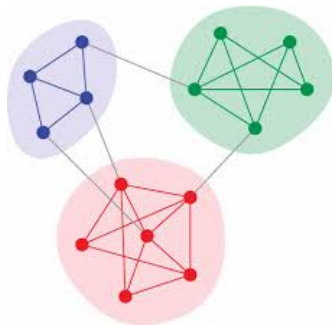
mais on préfère une définition plus flexible encore . . .



Densité locale

des structures denses à plus grande échelle

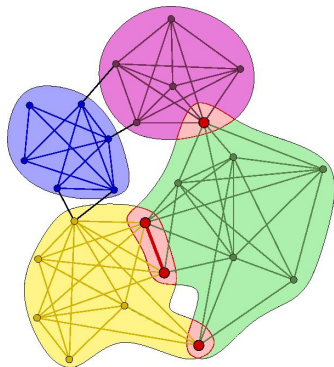
Notions de communautés



Partition communautaire

Densité locale

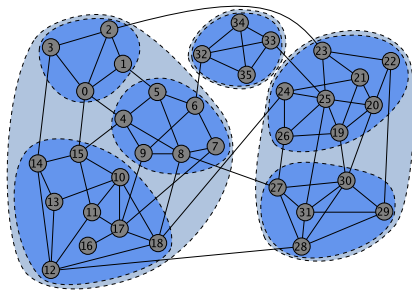
des structures denses à plus grande échelle
Notions de communautés



Communautés chevauchantes

Densité locale

des structures denses à plus grande échelle
Notions de communautés



Communautés hiérarchiques

Quelques pistes de réflexion

Intuition: une communauté est un groupe plus connecté vers l'intérieur que vers l'extérieur

Définir une quantité **maximum** lorsque c'est le cas

⇒ **problème d'optimisation**

Exemple:

nombre de liens à l'intérieur / nombre de liens vers l'extérieur

⇒ **pas approprié** (car une seule grande communauté)

Quelques pistes de réflexion

Intuition: une communauté est un groupe plus connecté vers l'intérieur que vers l'extérieur

Définir une quantité **maximum** lorsque c'est le cas
⇒ **problème d'optimisation**

Exemple:

nombre de liens à l'intérieur / nombre de liens vers l'extérieur
⇒ **pas approprié** (car une seule grande communauté)

Quelques pistes de réflexion

Intuition: une communauté est un groupe plus connecté vers l'intérieur que vers l'extérieur

Définir une quantité **maximum** lorsque c'est le cas

⇒ **problème d'optimisation**

Exemple:

nombre de liens à l'intérieur / nombre de liens vers l'extérieur

⇒ **pas approprié** (car une seule grande communauté)

Résumé

Propriétés typiques

densité	faible ($\ll 1$)
connexité	une composante géante
distances	faibles
degré	distribution hétérogène
clustering	élevé (\gg densité)
communautés	oui