Réseaux Sociaux

Licence 3 - Introduction aux Systèmes Complexes

Sébastien Verel
verel@i3s.unice.fr
www.i3s.unice.fr/∼verel/TEACHING/12-13/introSC/index.html

Équipe ScoBi - Université Nice Sophia Antipolis

28 février 2013

Plan

Commencons par un peu de littérature... in English bitte

"I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. I find it A) extremely comforting that we're so close, and B) like Chinese water torture that we're so close because you have to find the right six people to make the right connection... I am bound to everyone on this planet by a trail of six people."

John Guare, 1990

Et par une petite histoire

So small

- Je connais quelqu'un qui connait quelqu'un qui connait José Maria Aznar qui connait Georges Bush
- 4 personnes vous séparent d'un ancien président des Etats-Unis
- Sur cette Terre que 2 humains sont en moyenne reliés par une petite chaine de 6 personnes.

Tous trés proches ? A condition de connaitre cette chaine...

Et par une petite histoire

So small

- Je connais quelqu'un qui connait quelqu'un qui connait José Maria Aznar qui connait Georges Bush
- 4 personnes vous séparent d'un ancien président des Etats-Unis
- Sur cette Terre que 2 humains sont en moyenne reliés par une petite chaine de 6 personnes.

Tous trés proches ? A condition de connaitre cette chaine...

But

- Découvrir la structure des réseaux, en particulier sociaux

L'univers des réseaux

- Connaissez-vous des exemples de réseau ?
- Comment modéliser un réseau ?

L'univers des réseaux

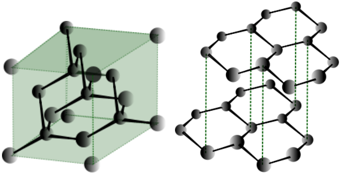
- Connaissez-vous des exemples de réseau ?
- Comment modéliser un réseau ?

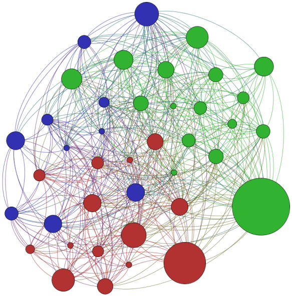
Réseaux

- Relation binaire :
  relation directe entre 2 entitées
- Graphe :
  - Ensemble de Noeuds ou sommet
  - Ensemble d'Arcs (orienté) ou arètes (non-orienté)
- Réseau : graphe + échange d'informations
  messages entre personnes, paquets entre machines, agent chimique entre gènes, voiture entre villes, etc.

$\Rightarrow$ un de ces universaux capables de modéliser de nombreuses situations existantes

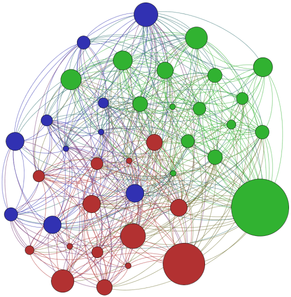
Réseau social





- Graphite, Diamant :
  pas d'échange d'information
  → "juste" graphe, statique, régulier.
- Relations amicales :
  Réseau, Raison d'exister est l'échange d'information
  - structure moins répétitive, sans régularité, très imbriquée
  - Arrivé / départ d'individus

Réseau social

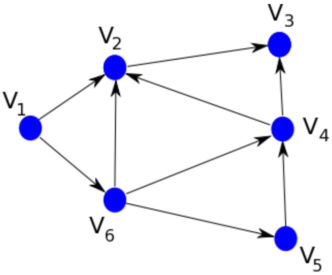


Réseau social

Réseau dont les relations ou les entités sont issue d'activités sociales

- Relations sociales :
  système complexe
- Relations inter-individuelles simples
  $\Rightarrow$ émergence de propriétés collectives
  (partage information, résistance virale, etc.)
- Science informatique et calculateurs :
  exploration de ces réseaux

Définitions basiques

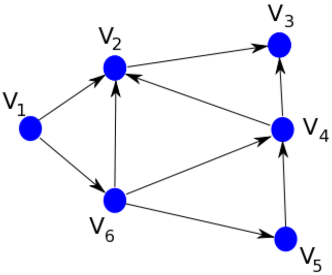


Graphe, réseau

Un graphe orienté G est un couple $G = (V, E)$ où :

- V est l'ensemble des noeuds
- E est l'arc des arcs $E \subset V^2$.

Définitions basiques



Degré sortant

$deg_+(v)$ est le nombre de noeuds w en relation depuis v : $(v, w) \in E$.

Degré entrant

$deg_-(v)$ est le nombre de noeuds w en relation vers v : $(w, v) \in E$.

La petite histoire du jour : Réunion de rentrée universitaire

Création de réseau de relations amicales

La situation

- Réunion de rentrée à l'université
- Une grande salle est prévue,
- Les étudiants arrivent les uns après les autres dans cette pièce.
- Au début un seul étudiant est présent dans la salle,
- Chaque nouvel étudiant décide ou non d'échanger quelques paroles et de se lier d'amitié avec une ou plusieurs personnes.

Selon la façon de lier amitié,
la structure du réseau est évidement fort différente.

Réunion de rentrée de première année

En cette première année universitaire,

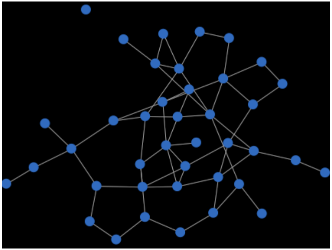
- les étudiants ont un degré de sociabilité similaire.
- Chacun engage une discussion avec un autre étudiant selon un *taux de sociabilité*.

taux de sociabilité 10% :
une chance sur dix de discuter avec un autre étudiant.

Les questions

- Combien de groupes d'étudiants vont se former ?
- Quelle sera la taille des groupes ?

Chemin et composante connexe



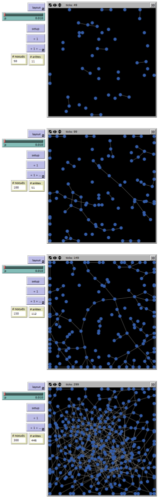
Chemin

Un chemin $P(v, w)$ est une suite de noeuds $(v_0, v_1, \dots, v_k)$ où :

- $v_0 = v$ et $v_k = w$
- $\forall i \in [|0, k-1|]$, $(v_i, v_{i+1}) \in E$

- S'il n'y qu'un seul groupe, tous les étudiants sont reliés par un chemin, le graphe G est dit *connexe*
- Un sous-graphe connexe est une *composante connexe*, i.e. un groupe d'étudiants

Formation du réseau



- Réseau non-régulier
- D'abord nombreux étudiants isolés
- Puis peu d'étudiants isolés, un groupe majoritaire
- A partir d'un certain seuil, un seul groupe

Graphe aléatoire

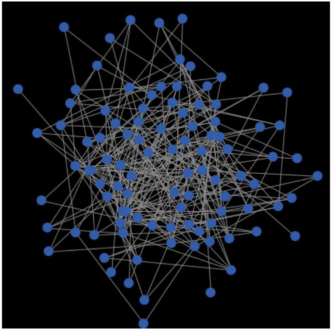
- Random graph : the easiest model for any network
  - Simple underlying asumption
  - but more realistic than regular graph
- First family of network studied ($\approx$ 1950)
  - exact results possible
- Basic idea :
  - Edges are added at random between a fixed number n of vertices
- Many models were developped, but the most important
  - the *Erdos-Rényi* random graph
  - the *Gilbert* random graph

Graphe aléatoire *Erdos-Rényi*

Définition

$\mathcal{G}_{n,m}$ ensemble des graphes à n noeuds et m arcs.
Probabilité uniforme de tirer un graphe dans cet ensemble.

Graphe aléatoire de *Gilbert*

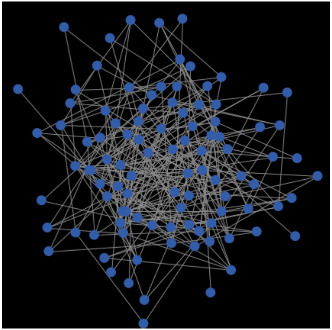


Définition

$G_{n,p}$ est un graphe à n noeuds dans lequel un arc (v, w) existe avec une probabilité p.

Facile à contruire itérativement

Graphe aléatoire de *Gilbert*



Définition

$G_{n,p}$ est un graphe à n noeuds dans lequel un arc (v, w) existe avec une probabilité p.

Facile à contruire itérativement

Relation entre les modèles

si $m \equiv p.n$ les deux modèles $\mathcal{G}_{n,m}$ et $G_{n,p}$ sont presque interchangeable.

Résultat sur les composantes connexes

Nombre et taille des groupes

- Lorsque $p < \frac{1}{n}$,
  grande probabilité que les composantes connexes du graphe aient une taille de l'ordre de $\log(n)$.
- Lorsque $\frac{1}{n} < p$,
  il existe presque surement une composante connexe dominante dont la taille est une fraction de n,
  et les autres composantes ont des tailles encore une fois très petite de l'ordre de $\log(n)$.

Résultat sur les composantes connexes

Seuil critique : changement de connexité rapide

Soit $\epsilon > 0$:

- Si $p > (1 + \epsilon)\frac{\log(n)}{n}$
  alors le graphe est connexe presque surement.
- Si $p < (1 - \epsilon)\frac{\log(n)}{n}$
  alors le graphe n'est pas connexe presque surement.

Distribution des degrés

nature de la distribution des degrés

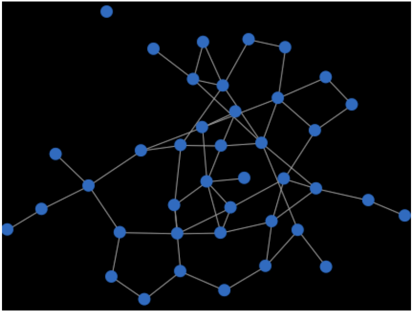
"En gros une bosse autour de la moyenne"

Plus précisement :

- Distribution binomial de paramètre $(n-1, p)$
- Probabilité d'avoir k relations : $\binom{n-1}{k} p^k (1-p)^{n-k}$,
- Moyenne : $p(n-1)$

Tout le monde a à peu près le même nombre de relation proche de la moyenne....

Distance de séparation



Longueur de chemin

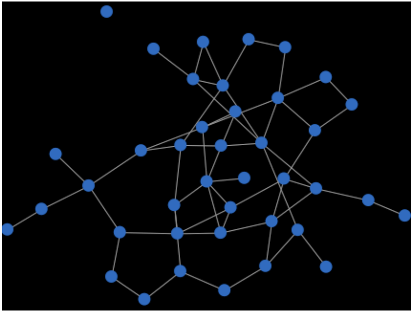
La longueur d'un chemin est le nombre d'arcs du chemin $P(v, w)$.

Distance

La distance entre 2 noeuds v et w est la longueur minimale chemin parmis les chemins reliant v et w.

Distance moyenne de séparation

Average Path length



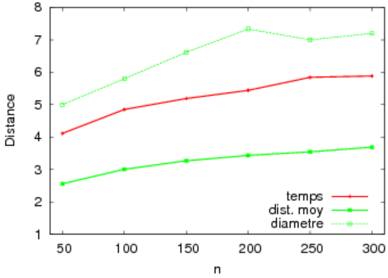
Longueur moyenne de chemin

La longueur moyenne de chemin est la moyenne des distance entre tout couple de noeuds du graphe.

Diamètre

Si le graphe est connexe, le diamètre est la plus grande distance entre les couples de noeuds.

Distance moyenne / Diamètre



Distance Moyenne graphe aléatoire

Lorsque celui-ci est connexe, la distance moyenne est $\frac{\log(n)}{\log(pn)}$.

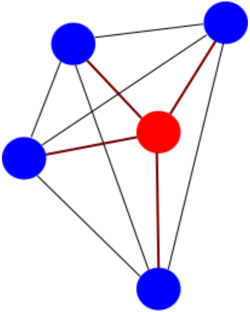
Les distances sont courtes !...

"Vous vous connaissez ?"

- Est-ce que mes amis sont amis ?
- Mes voisins sont-ils voisins entre eux ?
- Quelle est la densité du réseau ?

Clustering Coefficient, coefficient d'agglomération

- Introduit par Watts and Strogatz en 1998 (des physiciens !)
- Rapport entre le nombre de triangles de connaissance et le nombre maximum de triangles possibles ($k(k-1)/2$)
- Donne une bonne indication de la cohésion locale entre les individus.

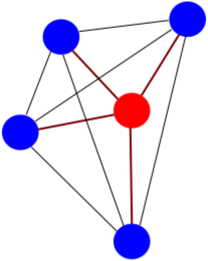


Coefficient d'agglomération

Définition (réseau non orienté)

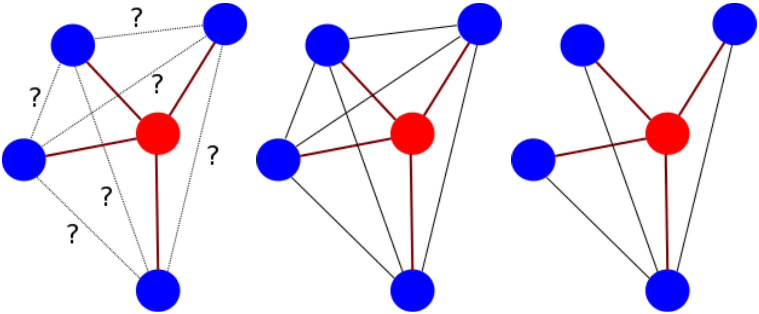
$$cc(v) = \frac{e(N(v))}{deg(v)(deg(v)-1)/2}$$

$N(v)$ ensemble des noeuds voisins
$e(N(v))$ nombre d'arcs dans le voisinage

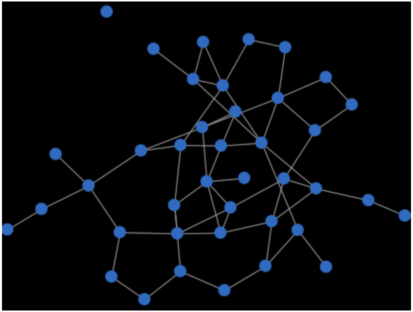


"Nombre de triangles"

Clustering Coefficient : exemples



Coefficient d'agglomération d'un graphe aléatoire



Graphe aléatoire

Coefficient d'agglomération est p

Très très peu dense, bizarre parce qu'on n'arrète pas de me répéter les mêmes trucs...

Pause NetLogo

Lien non orienté entre agent mobile en NetLogo

- `links` : ens. des agents liens (orienté ou non) entre turtles
- Création d'un lien non orienté entre *i* et *j* :

```
ask turtle i [
  create-link-with turtle j
]
```

- Création de liens non orienté depuis *i* :

```
ask turtle i [
  create-links-with turtles with [ ... ]
]
```

- Ensemble des noeuds voisins (agent set) de l'agent courant :

```
  link-neighbors
```

Pour que cela soit visible plus facilement

```
to layout
  let len (max-pxcor - min-pxcor) / (2 * sqrt count turtles)
  repeat 3 [
    layout-spring (turtles with [any? link-neighbors]) links 0.5 len 2
    display
  ]
end
```

Toujours des histoires

Incroyable ! Eurêka !

- L'histoire de la violoncielliste du pub irlandais
- L'histoire de la tondeuse mécanique, non électrique.

Toujours des histoires

Incroyable ! Eurêka !

- L'histoire de la violoncielliste du pub irlandais
- L'histoire de la tondeuse mécanique, non électrique.

- Comment penser rencontrer une collègue à qui on a jamais parlé dans un lieu de détente éloigné de l'ambiance du travail ?
- Comment imaginer en demandant à un nombre limité de personnes trouver un objet devenu rare ?

Ces exemples arrivent souvent dans notre monde d'humain sociaux

The small world problem, Milgram

Question : What is the distance between people in the social network ?

Experiment of 1967 :

- send a letter to 60 people randomly selected from Omaha (Nebraska)
- explain the study to people
- Ask them to send the letter to a person in Boston (Massachusetts) with only the name, the profession and the city.

(1933 - 1984)

http://maps.google.fr/maps?f=d&source=s_d&saddr=Omaha,+Nebraska&daddr=Sharon,+Massachusetts&hl=fr&geocode=&mra=ls&sll=47.15984,2.988281&sspn=15.664847,39.375&ie=UTF8&ll=42.182965,-83.56269&spn=34.

The small world problem, Milgram

- Some letters goes to the right person !
- When it arrives, only 6 persons between source and destination
- The diameter of the social network is small

The small world problem, Milgram

- Some letters goes to the right person !
- When it arrives, only 6 persons between source and destination
- The diameter of the social network is small

$\longrightarrow$ six degrees of separation

The small world problem, Milgram

- Some letters goes to the right person !
- When it arrives, only 6 persons between source and destination
- The diameter of the social network is small

$\longrightarrow$ six degrees of separation

As usual in science, there is some critisms :

- The effect of motivation increase the rate of succes :
  Is the social network changed with money ?
- The sociology of people has some effects on the succes rate :
  The social network have some small network,
  but it is not sure all the social network is small.

Six degrees of separation, Guare, 1990

American playwright, John Guare, 1990.

read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. I find it A) extremely comforting that we're so close, and B) like Chinese water torture that we're so close because you have to find the right six people to make the right connection... I am bound to everyone on this planet by a trail of six people.

Small Word phenomenon back to physicians

- from the experiment of Milgram, no convincing network model generating a network with high Clustering Coefficient or average short path length
- Watts and Strogatts started three king of real networks
- small world network : neither grid-like network neither full random network

Examples

- actor collaboration graph (Tjaden, 1997) :

Examples

- actor collaboration graph (Tjaden, 1997) :
  2 actors are connected by an undirected edge if they have acted in at least one film

Examples

- actor collaboration graph (Tjaden, 1997) :
  2 actors are connected by an undirected edge if they have acted in at least one film
- The power grid of the United States (Phadke and Thorp, 1988) :
- Neural network of the nematode C. elegans (White *et al.* 1992) :

Examples

Diameter and Clustering Coefficient of the 3 real networks compared to random graph

	D	D random	CC	CC random
actor collaboration	3.65	2.99	0.79	0.00027
power grid	18.7	12.4	0.08	0.005
Neural network	2.65	2.25	0.28	0.05

Small World Network

A small world network is a network with a dense local structure and a diameter comparable to a random graph with the same numbers of nodes and edges.

From random to order networks
Watts-Strogatz models

Small world Network

Neither :

- Grid-like networks : regularity and locality, but a high average path length and diameter ;
- Random graph : clustering coefficient of p

Watts and Strogatz proposed a mixture of both

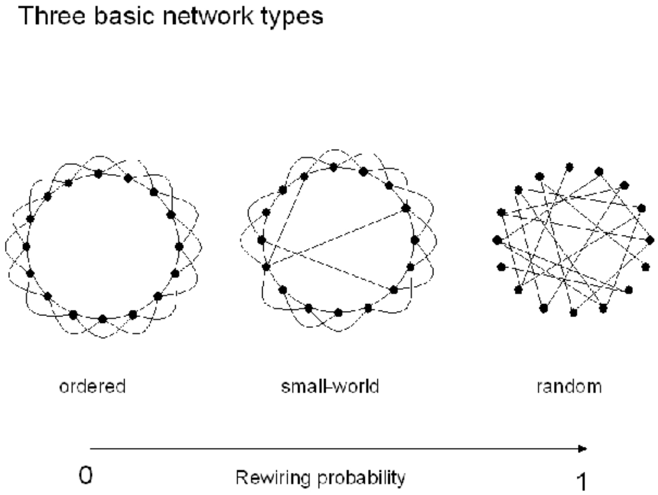
From random to order networks

Watts-Strogatz models

Generative Watts-Strogatz model :

- Build a ring of n vertices and connect each vertex with its k clockwie neighbors on the ring
- Draw a random number between 0 and 1 for each edge
- Rewire each edge with probability p :
  - if the edge's random nimber is smaller than p : keep the source vertex of the edge fixed, and choose a new target vertex uniformly at random from all other vertices.

From random to order networks
Watts-Strogatz models

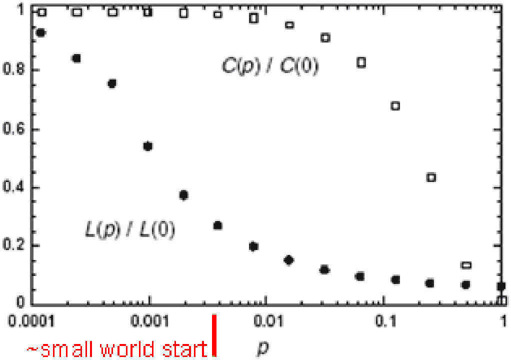


From random to order networks

Watts-Strogatz models

- For $p = 0$, network is totaly regular.
  - CC is approaching $3/4$ for large k
  - diameter in $O(n)$
- For $p = 1$, network is regular random network
  - CC is approaching p for large k
  - diameter in $O(\log n)$

Watts-Strogatz models



Characteristic path length $L(p)$ and clustering coefficient $C(p)$ for the family of randomly rewired graphs described in Fig. 1. Here L is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices. The clustering coefficient $C(p)$ is defined as follows. Suppose that a vertex v has k_v neighbours; then at most $k_v(k_v - 1)/2$ edges can exist between them (this occurs when every neighbour of v is connected to every other neighbour of v). Let C_v denote the fraction of these allowable edges that actually

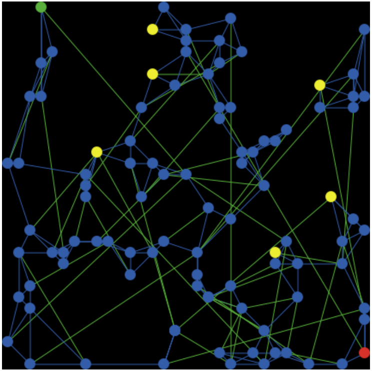
Réunion de rentrée universitaire de deuxiéme année

- Professeur en retard mais salle est fermée.
- Tous les étudiants entrent en même temps à l'ouverture, et se répartissent dans la salle.
- Chacun décide de parler avec les personnes les plus proches de lui (les k plus proches).
- Beaucoup se connaissent déjà dans la salle et veulent certainement échanger quelques nouvelles.
- Plus les étudiants sont éloignés, moins ils peuvent se parler.
- Probabilité d'établir une relation 'longue distance' entre a_i et a_j : $1/d(a_i, a_j)^r$.
  paramètre r : tendance à se parler loin

- Quelles sont les propriétés de ce réseau ?

Exemple de réseau de deuxième année

100 nœuds, $k = 3$, $q = 1$ et $r = 2.0$:



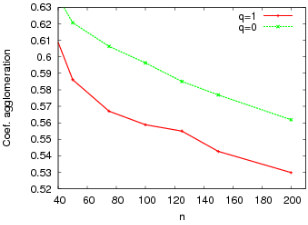
Coefficient d'agglomération 0.35, distance moyenne 4.18.

Principe de formation du réseau :

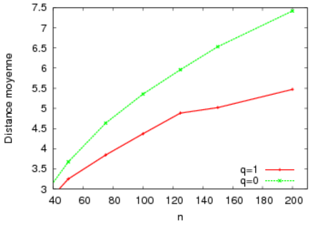
- En bleu : les liens courte distance avec les k plus proches étudiants
- En vert : les liens longue distance

Réseaux petit monde

Coefficient d'agglomération :



Distance moyenne :



Observations :

- Forte agglomération, Courte distance

$\Rightarrow$ Réseau petit monde

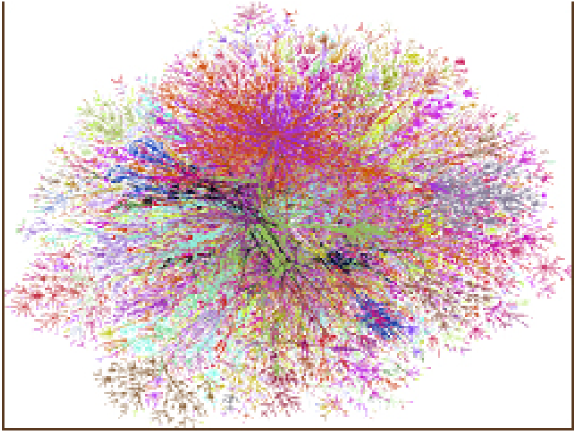
Quelques écart à la réalité dans certains cas

Les deux types de réseaux, les noeuds ont des degrés similaires :

- Pour le réseau aléatoire,
  distribution binomiale autour de la moyenne
- Pour les réseaux petit monde,
  distribution normale autour du degré k

Pourtant, dans beaucoup de réseaux sociaux ce n'est pas le cas...

Carte du trafic sur l'internet

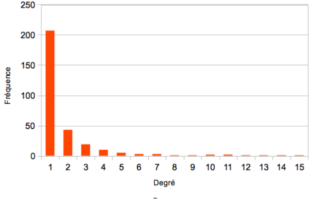


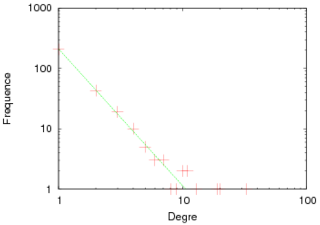
Il existe des noeuds avec un très grand degré : les hubs

Définition

Réseau sans échelle caractéristique

La distribution des degrés suit une loi de puissance :
Le nombre de noeuds de degré k est proportionel à k^α





- A toutes les échelles d'observation, on "voit" le même réseau
- Beaucoup de noeuds peu connecté
- Un très petit nombre de noued fortement connecté

Exemples

- Traffic (nombre de passager, lignes) des aéaroport
- Collaborations scientifique
- Proximité sémantique des mots
- Les partenaires sexuels pour les humains
- Les intéraction de réseau de protéines

Question

Quels modèles de formation pour ces réseaux ?

Encore une réunion de rentrée universitaire, la troisième

- Relation : en fonction des étudiants déjà présents.
- Toujours tendance à vouloir bavarder entre eux,
- Comment faire pour un étudiant pour participer facilement à de nombreuses discussions ?

Encore une réunion de rentrée universitaire, la troisième

- Relation : en fonction des étudiants déjà présents.
- Toujours tendance à vouloir bavarder entre eux,
- Comment faire pour un étudiant pour participer facilement à de nombreuses discussions ?
- En s'approchant préférentiellement de ceux qui mènent déjà une discussion avec de nombreux étudiants,
- Chaque étudiant va donc engager une conversation avec quelques étudiants de préférence avec ceux qui parlent déjà le plus...

- Est-il pour autant facile de faire circuler rapidement une nouvelle dans la salle pour les étudiants comme dans les réseaux petit monde ?

Principe d'attachement préférentiel

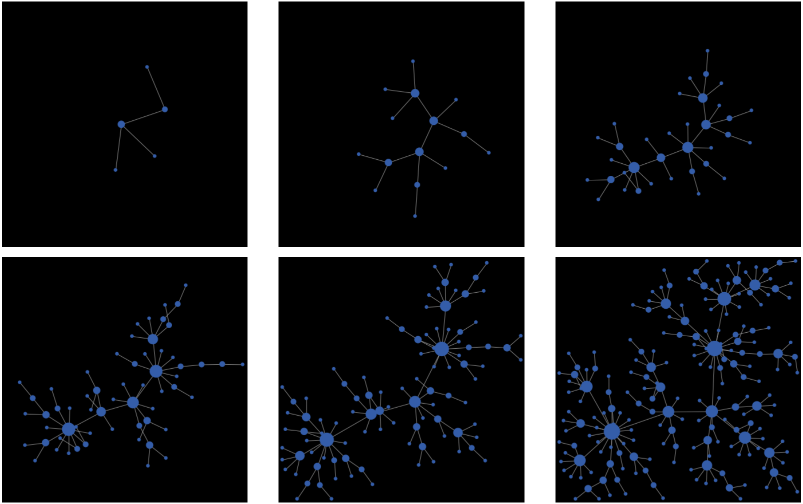
Attachement préférentiel

Pour tout étudiant u,

$$\forall v \in V, \qquad Pr(\{u, v\} \in E) = \frac{deg(v)}{\sum_{w \in V} deg(w)}$$

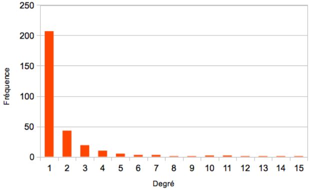
- Les riches deviennent plus riches

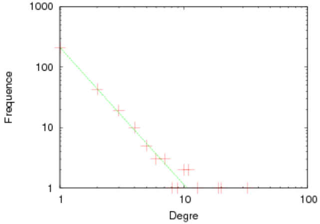
Exemple de formation



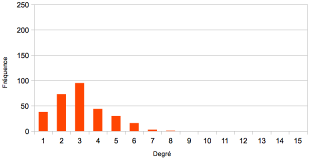
Comparaison des distributions des degrés

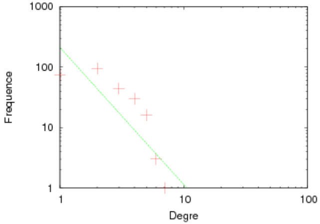
Réseau de troisième année (scale free)





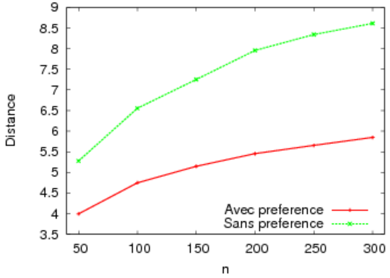
Réseau de première année (aléatoire Erdös-Rényi)





Distance moyenne

Attachement
avec et sans préférence



- Les distances sont plus courtes avec attachement préférentiels
- Les hubs jouent le rôle de raccourcis

Conclusions

Différents modes de formation (règles locales)

- Attachement uniforme : Réseaux aléatoires
- Création de lien courte et longue distance : Réseaux petit monde
- Attachement préférentiel : réseaux sans échelle caractéristiques

Trois propriétés principales

- Distance moyenne (et diamètre)
- Coefficient d'agglomération
- Distribution des degrés

Selon la structure, différentes propriétés globales... à exploiter

Conclusions

Sujet en plein développement

- De nombreuses applications sont encore à inventer ou à exploiter !
- Problématique : Réseaux dynamiques, grand réseaux, enchevetrement de réseaux, sécurité, etc.
- Propriétés globales : pharmacologie, information, offre de services, politique de santé, etc.

Bibliography

- D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), 440–442.
- D. J. Watts, *Networks, Dynamics, and the Small-World Phenomenon*, American Journal of Sociology **105** 2, (1998), 493–527.
- A-L. Barabasi and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), 509–512.
- D. J. Watts, *The "new" science of networks*, Annual Review of Sociology **30** (2004), 243–270.

Bibliography

Katharina Anna Lehmann, Michael Kaufmann.
Random Graphs, Small-Worlds and Scale-Free Networks.
Peer-to-Peer Systems and Applications (2005) 57-76

F. Vazquez, V.M. Eguiluz, and M. San Miguel.
Generic Absorbing Transition in Coevolution Dynamics
In Phys. Rev. Lett. 100, 108702 (2008)

J. Kleinberg,
"Navigation in a small world."
Nature 406 (2000)