Projet AGGP - Synthèse des articles

10 mars 2011

1 Mesures possibles sur réseaux

Degré d'un noeud : k
Représente le nombre de connexions branchées sur un noeud.

Degré moyen d'un graphe : $< k >$
Représente le nombre moyen de connexions branchées sur un noeud dans le graphe.

Distribution des degrés : $P(k)$
La distribution des degrés des noeuds du graphe donne des informations sur la nature du graphe : scale-free, random...

"Degree exponent" : γ
Ce paramètre n'a de sens que dans le cas de scale-free networks. Il traduit le rôle des hubs dans le réseau.

Plus petit chemin : l
Le plus petit chemin qui relie deux noeuds. Attention, dans le cas de graphes orientés, il est probable que $l_{AB} \neq l_{BA}$.

Plus petit chemin moyen : $< l >$
Moyenne des plus petits chemins qui relient deux à deux tous les noeuds du graphe. $< l >$ traduit la navigabilité du graphe : plus $< l >$ est petit, plus le chemin entre deux points du graphe est court.

Coeffcient de clustering : C_I
Nombre de "triangles" ABI passant par le noeud I, où A et B représentent deux noeuds du graphe. $C_i = \dfrac{2n_i}{k(k-1)}$, avec n_I le nombre de liens reliant les k voisins de I. C est la signature d'une potentielle modularité du graphe.

Coefficient de clustering moyen : $< C >$
Moyenne des $< C_I >$.

Distribution des coefficients de clustering : $P(C)$
Traduit le caractère hiérarchique du réseau : $C(k) \sim k^{-1}$

Important :
$< k >, < l >, < C >$ dépendent du nombre de noeuds et de liens (N, L) du graphe, alors que $P(k)$ et $C(k)$ n'en sont pas dépendant. Ces deux derniers paramètres permettent de caractériser le réseau : scale-free, random...

2 Réseau aléatoire

On construit un réseau aléatoire en connectant un nombre fixe N de noeuds de façon aléatoire. On sait que dans un réseau aléatoire, $P(k)$ suit une distribution de Poisson.
Les réseaux aléatoires ne peuvent pas expliquer les propriétés topologiques des réseaux biologiques. Dans de tels réseaux, $< l >= log(N)$: propriété de "petit monde".

3 Scale free networks

Dans ce type de réseau, on prend en compte la présence de noeuds très fortement branchés aux autres noeuds du graphes. De tels noeuds sont appelés des **hubs**.

Exemple :
- sur internet : Twitter, Google, Youtube, Facebook,...
- dans un réseau métabolique : ATP, ADN$_{pol}$, ...

La présence de ces hubs bouleverse la distribution de $< P(k) >$. On a alors $P(k) \sim k^{-\gamma}$ (**loi de puissance**), avec γ degree exponent, traduisant le rôles des hubs dans le réseau.

La valeur de γ est primordiale pour caractériser le réseau :
- $\gamma > 3$: pas de hubs dans le réseau. Ce n'est pas un scale-free,
- $2 < \gamma < 3$ **(le cas de la majorité des réseaux biologiques)** : hiérarchie de hubs et $< l > =$ log(log(N)) : propriété de (très) petit monde,
- $\gamma = 2$: présence d'un très gros hub.

On a trouvé que la majorité des données sont le mieux fittées si $\gamma \simeq 2.1$. Les réseaux cellulaires, métaboliques, sociaux ... sont scale-free.

4 Comparaison des types de réseaux

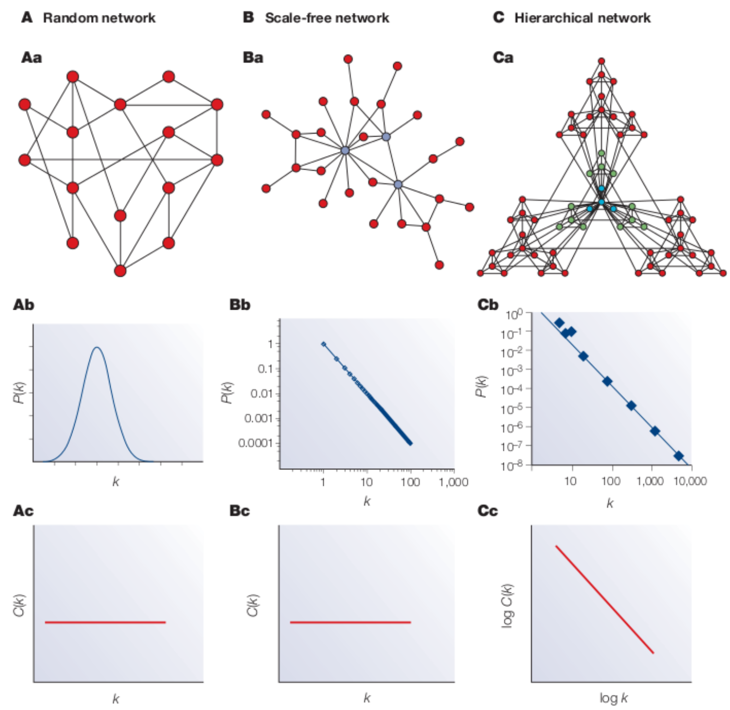

FIGURE 1 – Random (A), scale-free (B), hierarchic (C)

Réseaux aléatoires,
$< P(k) >_{max} \simeq < k >$.
Si p est la probabilité de connecter une paire de noeuds, on a environ $\frac{pN(N-1)}{2}$ liens.
$C(k)$ indépendant de k.

Scale free networks :
Les noeuds bleus sont statistiquement plus connectés que dans un réseau aléatoire : ce sont des hubs.
$< l > =$ log(log(N)) : effet des hubs.

Hierarchical networks :
Combinaisons de clusters générant des réseaux hiérarchiques.

5 Propriétés de réseaux biologiques

Un réseau biologique doit (si je ne me trompe pas) valider les conditions suivantes :

1. $< P(k) > \sim k^{-\gamma}$ (Power law) avec $2 < \gamma < 3$,
2. valider une propriété de petit monde (inclus dans la 1 ?),
3. avoir une forte resistance ($\simeq$80% des nodes sont "éjectables") (inclus dans la 1 ?)
4. formation de cliques/modules : contradictoire avec la propriété de petit monde => TRADE-OFF à déterminer.