

Révision: les réseaux complexes



Sommaire

■ Révision

- Notions de base sur les graphes
- Rappel des propriétés des grands graphes de terrain (un peu de théorie)

■ Exemples d'application:

- Les réseaux économiques (CAC 40)
- La propagation des épidémies
- Les relations entre maladies
- Internet (pour mémoire)
 - Le réseau
 - Le WEB

■ Intérêt: comprendre, prévoir, agir

Introduction

De récentes avancées dans le domaine des systèmes complexes ont fait ressortir le rôle central que jouent les graphes dans de nombreux phénomènes. Ces **grands graphes** permettent de modéliser les interactions entre les différents acteurs de ces phénomènes complexes.

De manière surprenante ces graphes, appelés **grands graphes de terrain** ou **complex networks**, possèdent des **propriétés structurelles non triviales communes** qui ont fait l'objet de nombreuses et récentes études. Ces propriétés communes permettent d'envisager une nouvelle algorithmique des grands graphes de terrain. Les avancées qui en découlent ont des applications potentielles dans les divers domaines concernés.

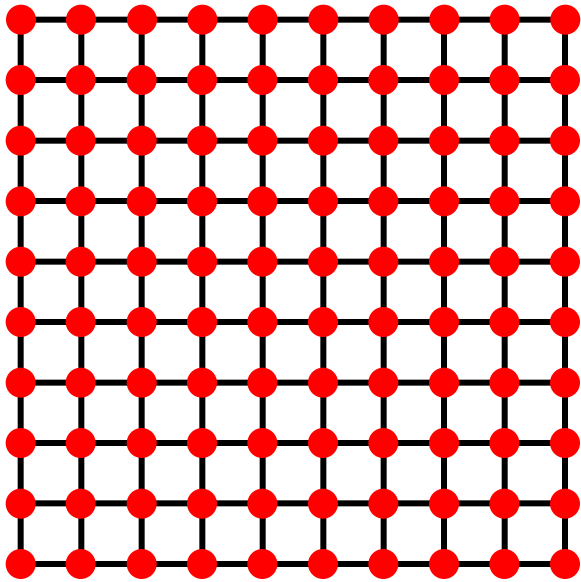
Graphes : deux exemples simples

Avant de rappeler les propriétés des grands graphes de terrain, les notions de base sont illustrées par deux exemples simples (« triviaux ») :

- 1. Graphe régulier**
- 2. Graphe « petit mondisé »**

NB: Les définitions de bas sur les graphes sont rappelées en annexe

Exemple 1 : Graphe régulier 100 sommets



**Le diamètre D est
proportionnel à $\sqrt{n}$**

Exemple de graphe régulier:

- $n = 10 \times 10 = 100$ sommets
- $m = 2 \times (9 \times 10) = 180$ arêtes

Nombre maximum d'arêtes :

- $= n(n-1)/2 = 4950$

Densité : $\delta = 180/4950 = 3.6 \%$

Diamètre: $D = 2(\sqrt{n}-1) = 18$

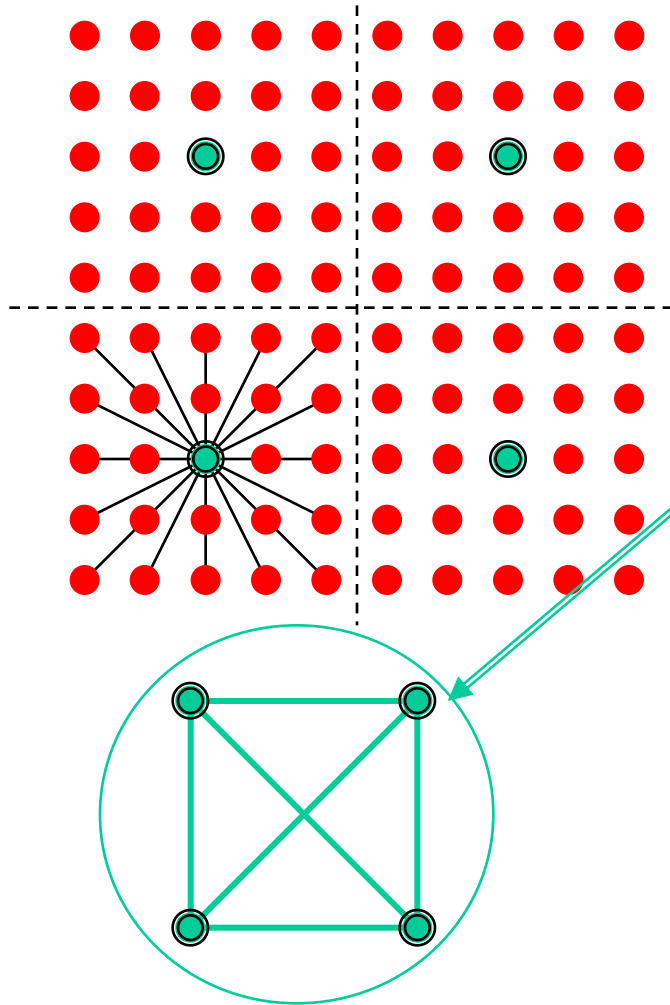
Distance moyenne: $L \approx 5$

Distribution des degrés:

- 4 sommets avec $k=2$
- 32 sommets avec $k=3$
- 64 sommets avec $k=4$

**Si n est grand, le degré tend
uniformément vers $k=4$**

Exemple 2 : Graphe « petit mondisé »



n = 10x10=100 sommets :

- 4 sommets de concentration (nœuds verts)
- Chaque nœud **rouge** a un lien direct avec le nœud **vert** le plus proche
- Un lien direct entre chaque nœud **vert**

m = 102 arêtes :

- 6 arêtes entre nœuds verts
- (25-1)x4 arêtes entre rouge et vert, soit 96 arêtes

Densité: $\delta = 102/4950 = 2.06 \%$

Diamètre: 3

Distance moyenne: $2 < L < 3$

Distribution des degrés

- 96 sommets avec k=1
- 4 sommets avec k=27

Caractéristiques des graphes de terrain

Les grands (*) graphes de terrain rencontrés dans différentes disciplines sont très différents des graphes réguliers. Ils ont des caractéristiques communes, en particulier :

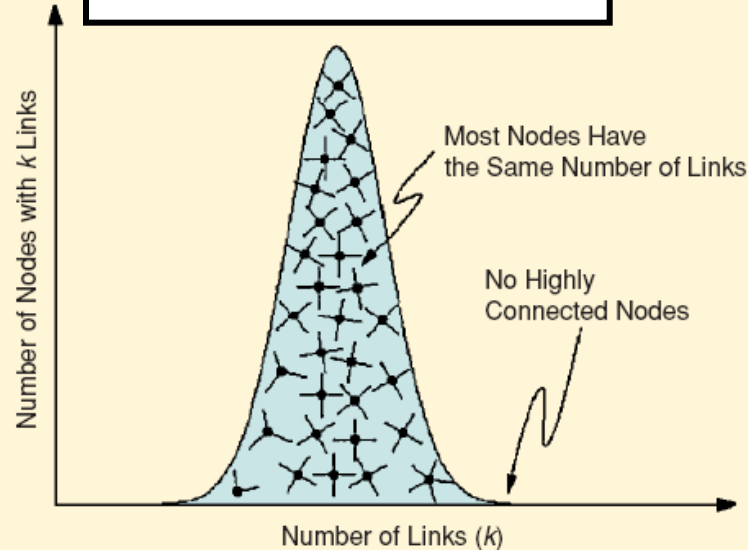
- un faible degré moyen et une forte hétérogénéité des degrés : distribution de puissance des degrés des sommets ;
- des chemins courts entre tous les sommets. C'est la propriété du petit monde ;
- une faible densité globale couplée à une forte densité locale : on parle de clustering ;

NB: Distribution de puissance et Petit Monde sont des propriétés liées

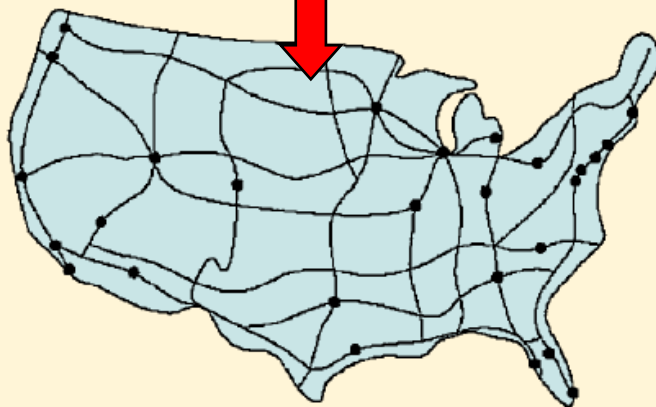
(*) Grands graphes: Internet: ~ 20 000 AS, $180 \cdot 10^6$ sites; cerveau: 10^{12} neurones

Propriétés des graphes (selon Barabasi)

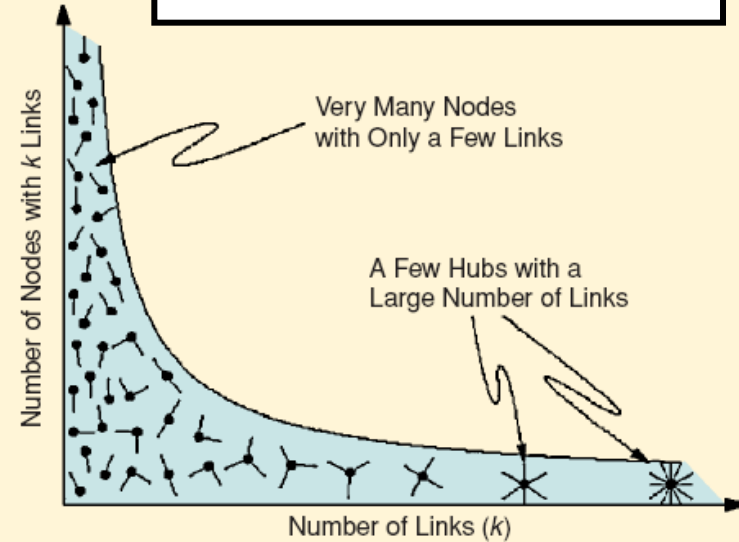
Distribution de Poisson



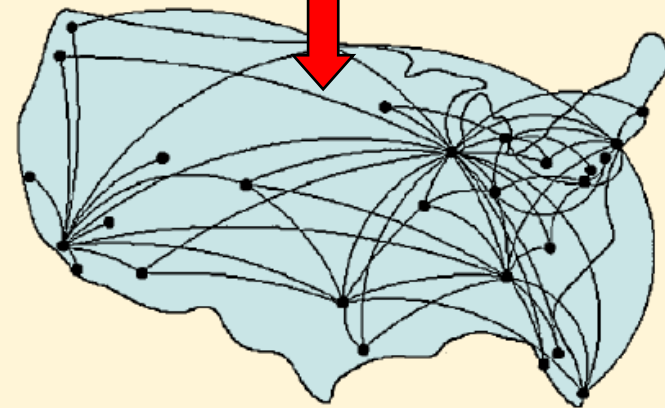
(a)



Distribution de Puissance



(b)

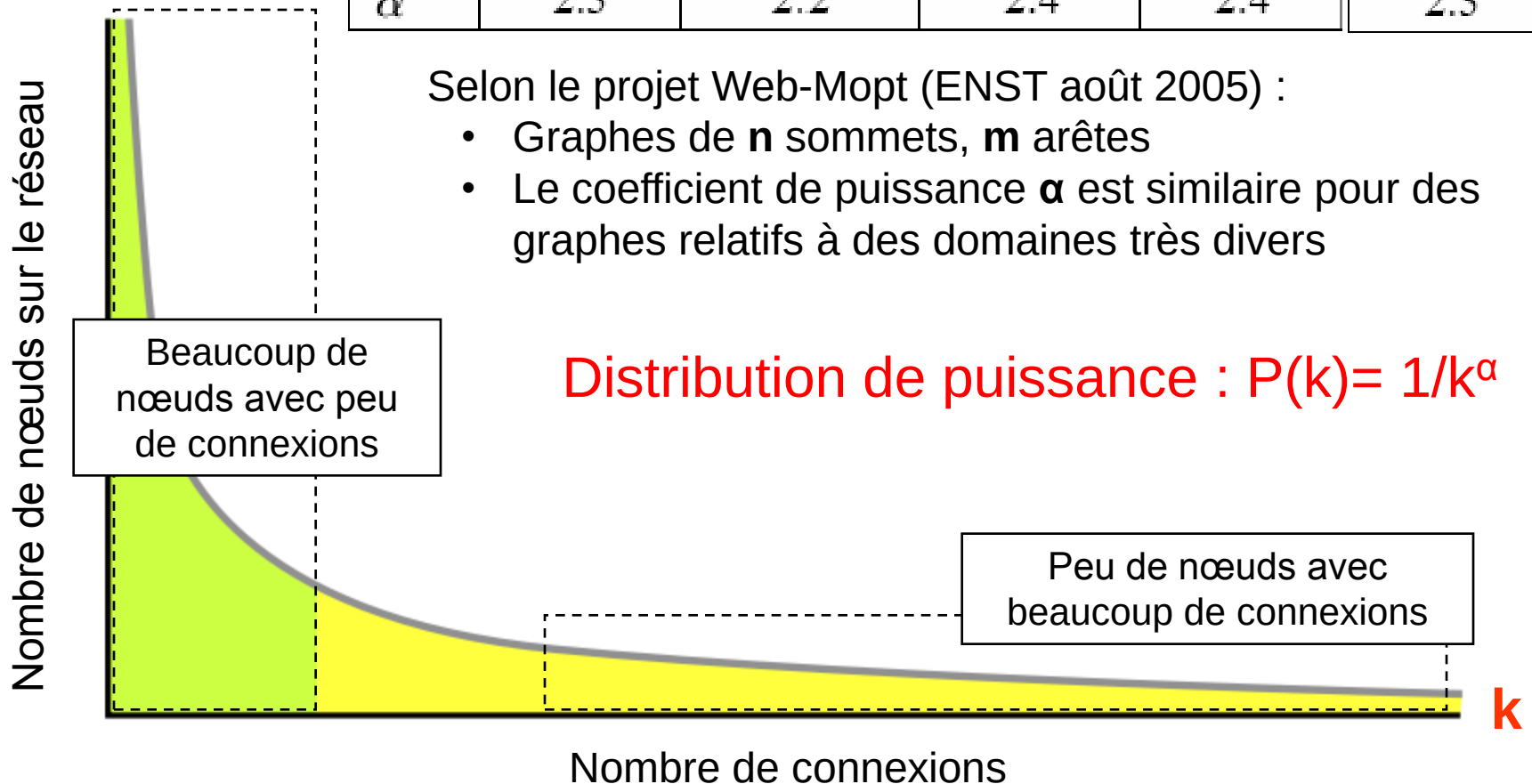


Exemples de loi de puissance

	Internet	Actors	Co-auth	Protein	Web
n	228263	392340	16401	2113	325729
m	320149	15038083	29552	2203	1090108
α	2.3	2.2	2.4	2.4	2.3

Selon le projet Web-Mopt (ENST août 2005) :

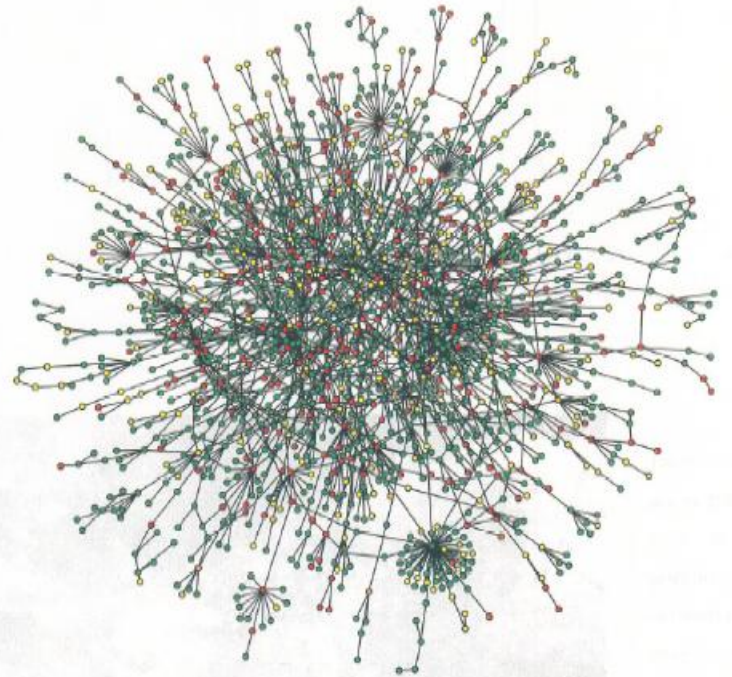
- Graphes de n sommets, m arêtes
- Le coefficient de puissance α est similaire pour des graphes relatifs à des domaines très divers



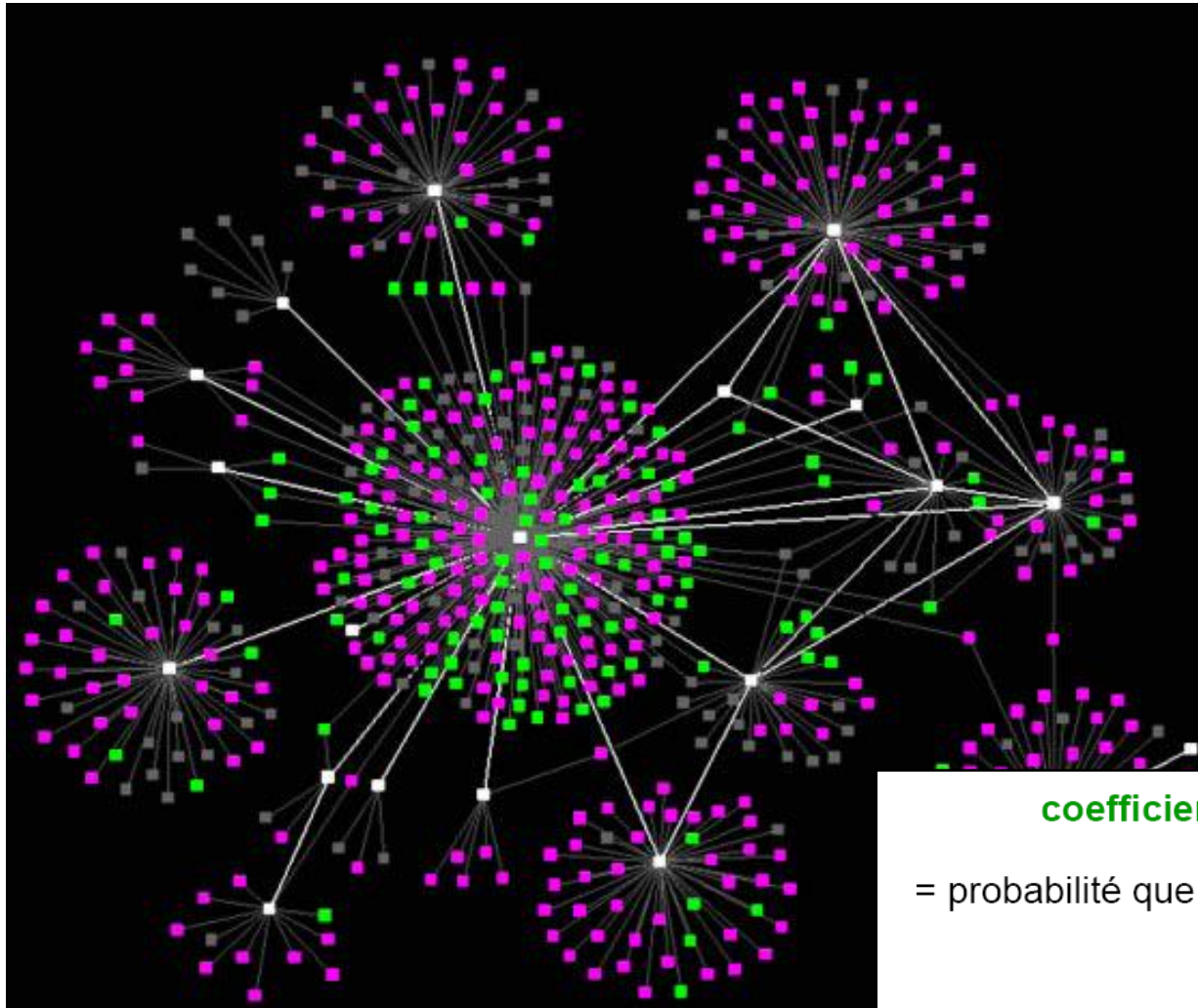
Exemples de Petit Monde

- **d (individus) = 6**
(Milgram)
- **d (IP) = 15**
(réseau physique d'Internet)
- **d (WWW) = 21**
(Liens hypertexte entre pages du WEB)

Routage Internet **BGP**:
15 sauts maximum pour
relier 2 AS, c.a.d. $d=15$



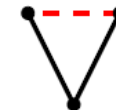
Le clustering



Propriétés
complémentaires du
petit monde: des
zones denses
faiblement
interconnectées: le
clustering

coefficient de clustering $cc(v)$

= probabilité que deux voisins de v soient reliés



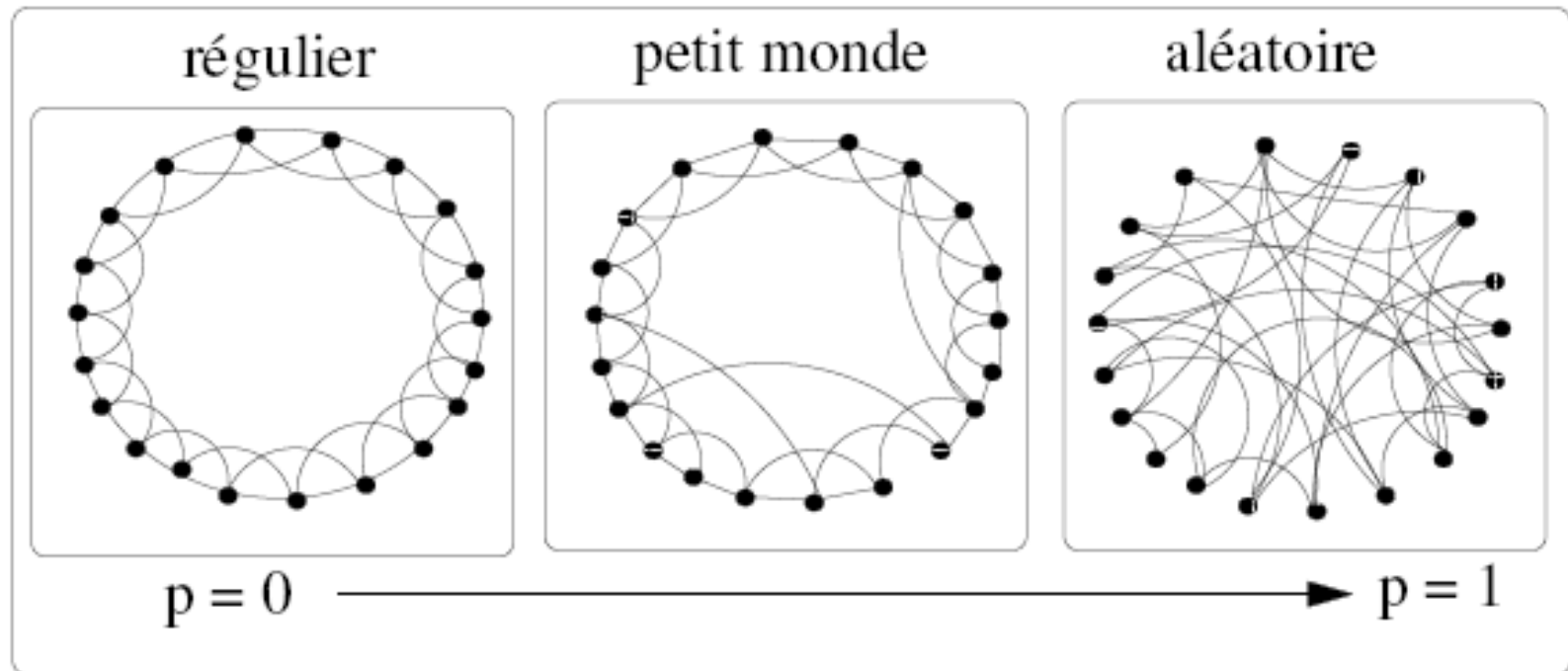
= densité locale

Génération des graphes de terrain

- Plusieurs modèles ont été imaginés pour simuler la constitution de graphes de terrain:
 - Génération aléatoire (Watts et Strogatz) ;
 - Rattachement préférentiel « **les riches deviennent plus riches** »: ce mécanisme explique la création de réseaux « **petit monde** », par exemple le réseau physique d'Internet ;
 - Phénomène de percolation: effet de seuil pour la connectivité ;
- Les propriétés des graphes de terrain peuvent être considérées comme une émergence résultant des interactions élémentaires entre les sommets.

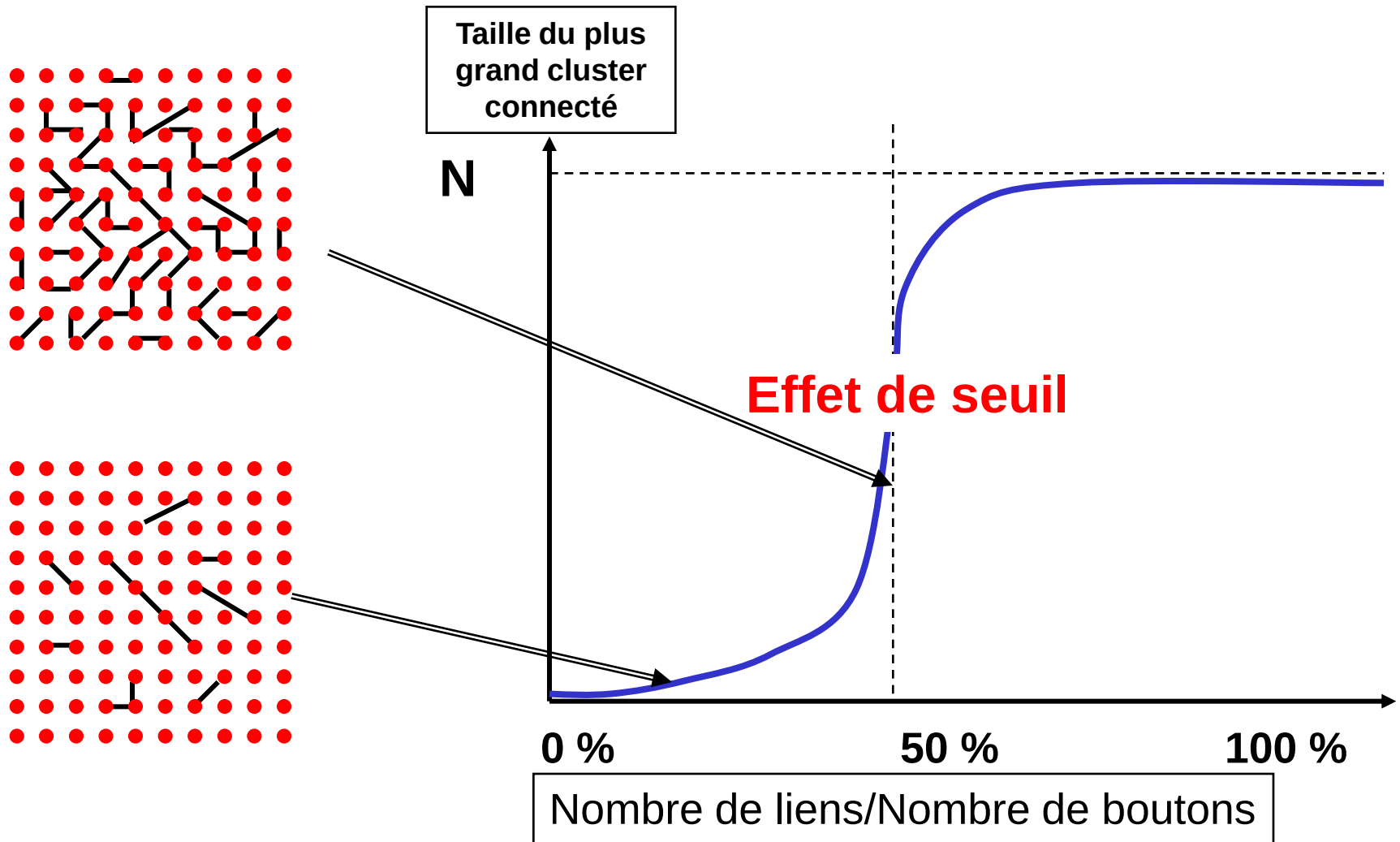
Modèle de Watts & Strogatz

Création d'un graphe régulier puis rebrassage de certains liens au hasard (probabilité **p**)

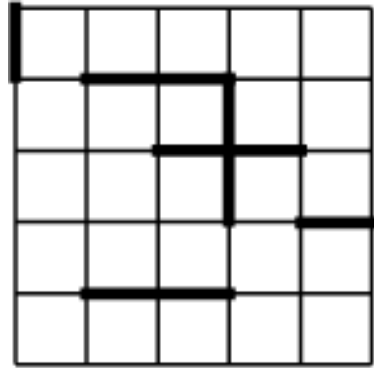


Watts [1999] a montré que pour un graphe aléatoire de presque toutes les topologies, on peut s'attendre à ce que la longueur moyenne de chemin tende vers la valeur $\log(n)/\log(k)$. Watts a également montré que le coefficient de **clustering** attendu est de k/n .

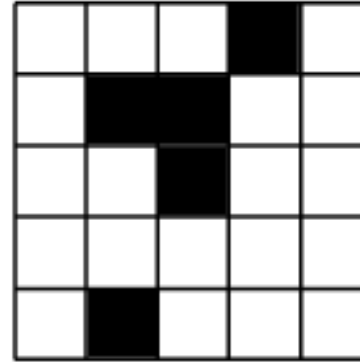
Phénomène de percolation



La percolation comme principe de base des contagions dans les réseaux



Percolation par liens



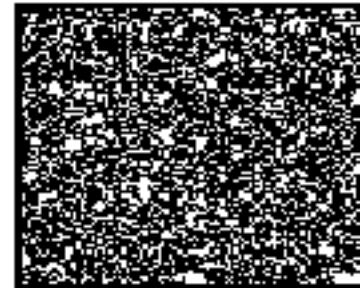
Percolation par Sites



Sous Critique



Critique

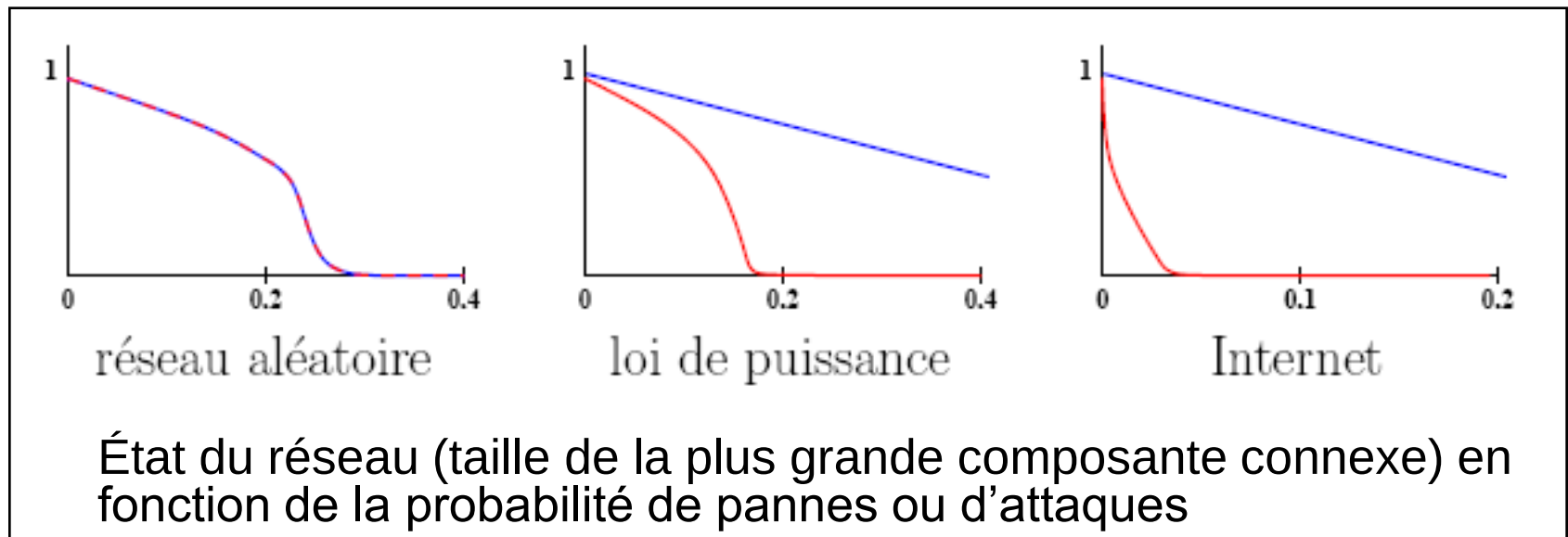


Sur Critique

La résilience : percolation à l'envers

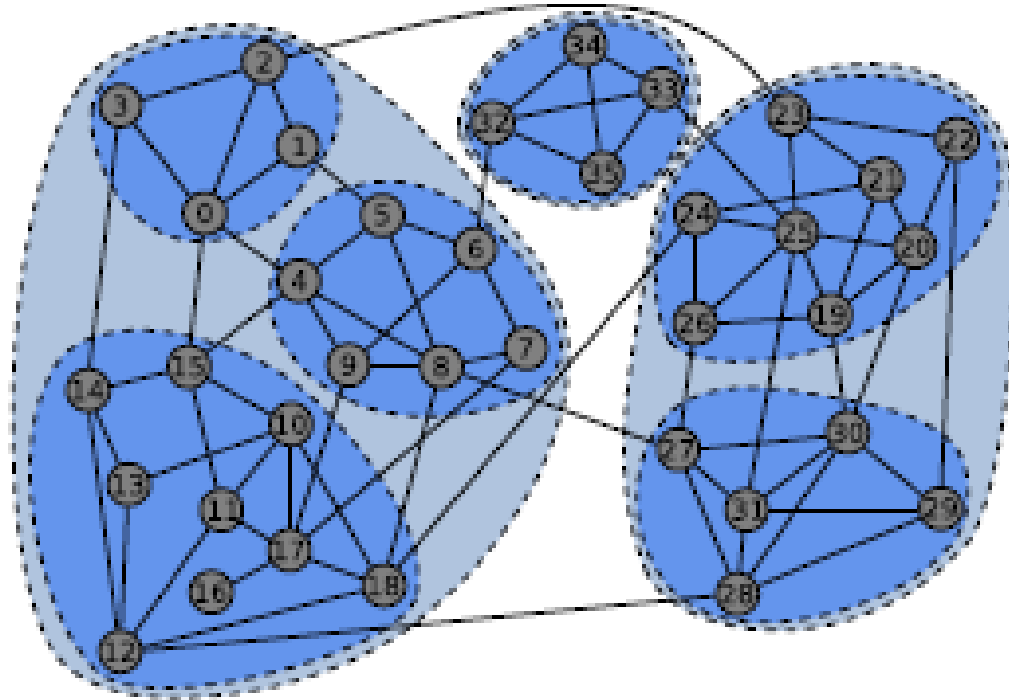
■ On distingue :

- **Panne** = suppression de sommets au hasard
- **Attaque** = suppression de sommets choisis (par exemple des sommets d'AS de niveau 1).



- Une panne peut être considérée comme une « percolation à l'envers »
- Par exemple, Internet est résistant aux pannes et sensible aux attaques

La détection de communautés



La détection de communautés permet de comprendre les structures et les fonctionnements macroscopiques des grands graphes de terrain.

Une telle méthode peut être utilisée comme brique de base pour l'élaboration d'algorithmes plus complexes comme la parallélisation, la visualisation ou la compression des données.

L'existence dans les grands graphes de terrain de zones plus densément connectées que d'autres constitue une des caractéristiques non triviales que l'on retrouve dans de nombreux cas. Ces zones sont appelées communautés (par analogie avec les réseaux sociaux) et correspondent intuitivement à des groupes de sommets plus fortement connectés entre eux qu'avec les autres sommets.

Source: Détection de communautés dans les grands graphes de terrain par Pascal Pons, thèse de doctorat 2007

Grandes catégories de graphes de terrain

1. **Les réseaux sociaux.** Ils constituent un champ d'application ancien et important dans lequel les acteurs sont des individus ou entités sociales (associations, entreprises, pays, etc). Les liens entre eux peuvent être de différentes natures :
 - les réseaux de connaissance (deux individus sont reliés s'ils se connaissent),
 - les réseaux de contact physique (deux individus sont reliés s'ils ont été physiquement en contact),
 - les réseaux de collaboration (deux individus sont reliés s'ils ont travaillé ensemble, en particulier de nombreux travaux ont étudié les collaborations scientifiques,
 - les réseaux d'appels téléphoniques (deux individus ou numéros de téléphones sont reliés s'il y a eu un appel entre eux),
 - les réseaux d'échanges (deux entités sont reliées si elles ont échangé un fichier ou un courrier électronique par exemple), etc.

Grandes catégories de graphes de terrain

2. **Les réseaux biologiques.** Il en existe plusieurs types parmi lesquels nous pouvons citer:

- les réseaux métaboliques (les sommets sont des gènes ou des protéines qui sont liés par leurs interactions chimiques),
- les réseaux de neurones (chaque neurone est connecté à plusieurs autres neurones),
- les réseaux trophiques (les espèces d'un écosystème sont reliées pour représenter les chaînes alimentaires).

3. **Les réseaux d'infrastructure.** Ils représentent des connections matérielles entre objets distribués dans un espace géographique. Nous pouvons citer

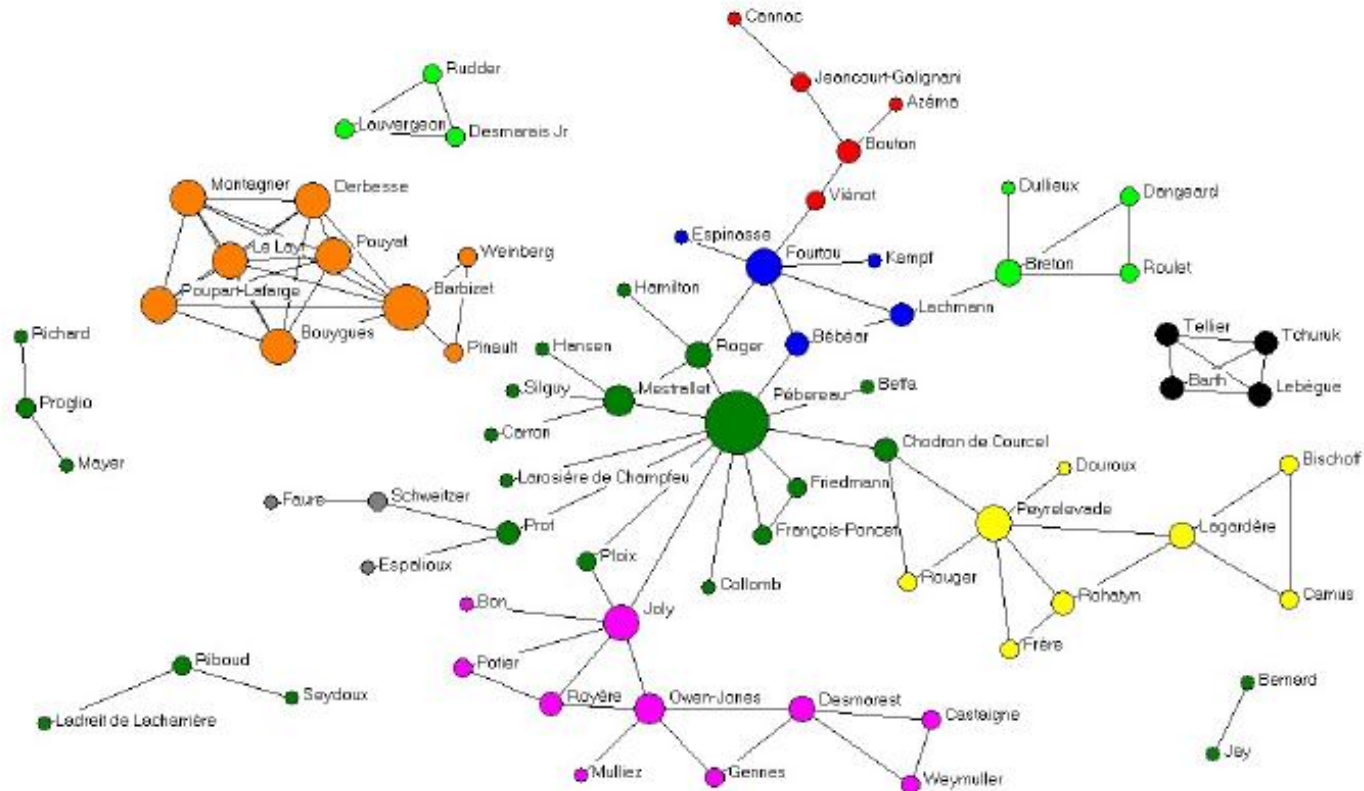
- les réseaux de transport (routes entre villes ou liaisons aériennes entre aéroports),
- les réseaux de distribution électrique (câbles entre les lieux de production et de consommation)
- le réseau physique de l'Internet (câbles entre ordinateurs).

Grandes catégories de graphes de terrain

4. **Les réseaux d'information.** Ils représentent des liens abstraits de référencement entre des supports d'information. Parmi eux :
- les réseaux de citation d'articles
 - les graphes du Web (les sommets sont des pages Web liées par des liens hypertextes).
5. **Les réseaux linguistiques.**
- Ils relient les mots d'un langage donné et regroupent entre autres les réseaux des synonymie (deux mots sont reliés s'ils sont synonymes),
 - les réseaux de co-occurrences (deux mots sont reliés s'ils apparaissent dans une même phrase d'un ouvrage),
 - les réseaux de dictionnaires (deux mots sont liés si l'un est utilisé dans la définition de l'autre).

Le petit monde du CAC 40 (2003)

Réseaux sociaux des principaux dirigeants du CAC 40 au 31.12.2003

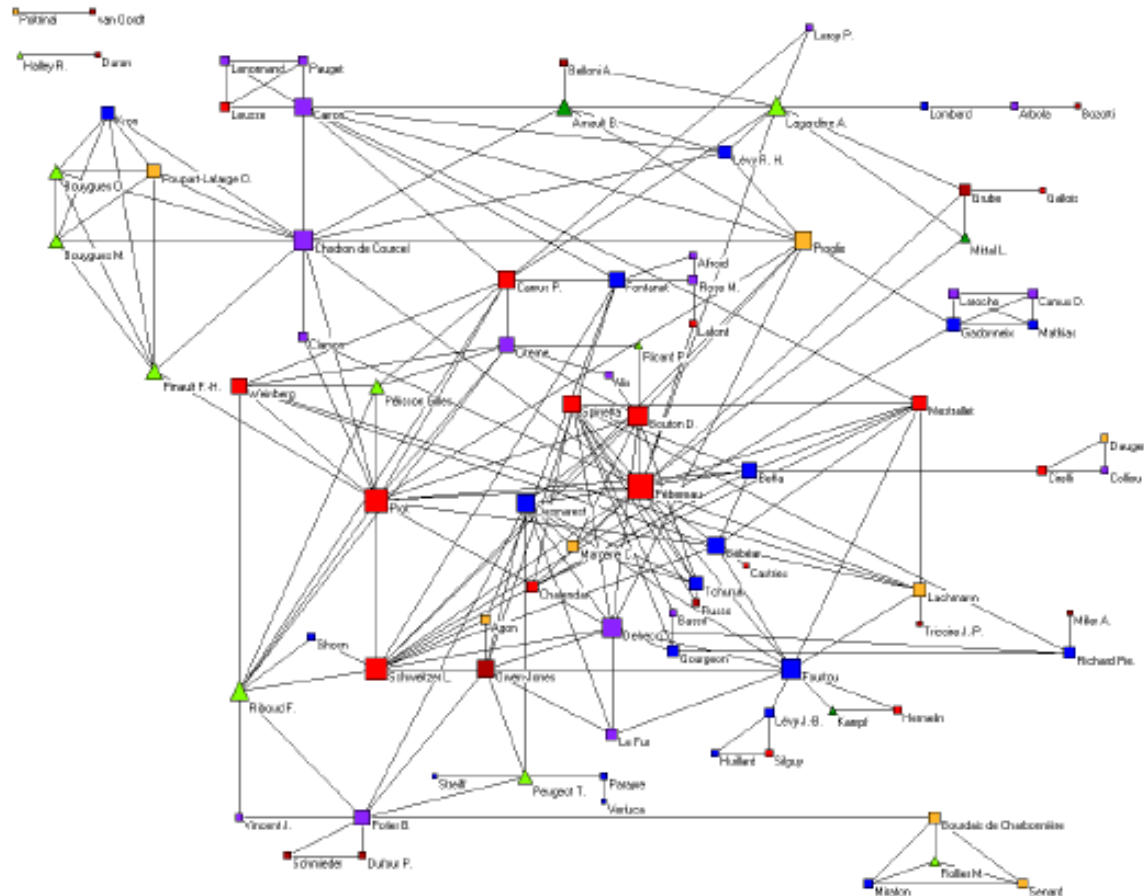


Légende

La taille des cercles indique le nombre de dirigeants avec lesquels qu'une personnalité côtoie au moins deux fois dans les sociétés du CAC 40. Analyse structurale dite par « faction ».

Les patrons du CAC 40 (2007)

Réseaux sociaux des patrons du CAC 40 au 31.12.2007



Légende

Origines scolaires : ■ ENA-ENS ■ X-MINES-PONTS-ENST ■ HEC-ESCP-ESSEC ■ Autres ■ Etranger ▲ Fondateur ▲ Héritier

Modèles d'épidémies

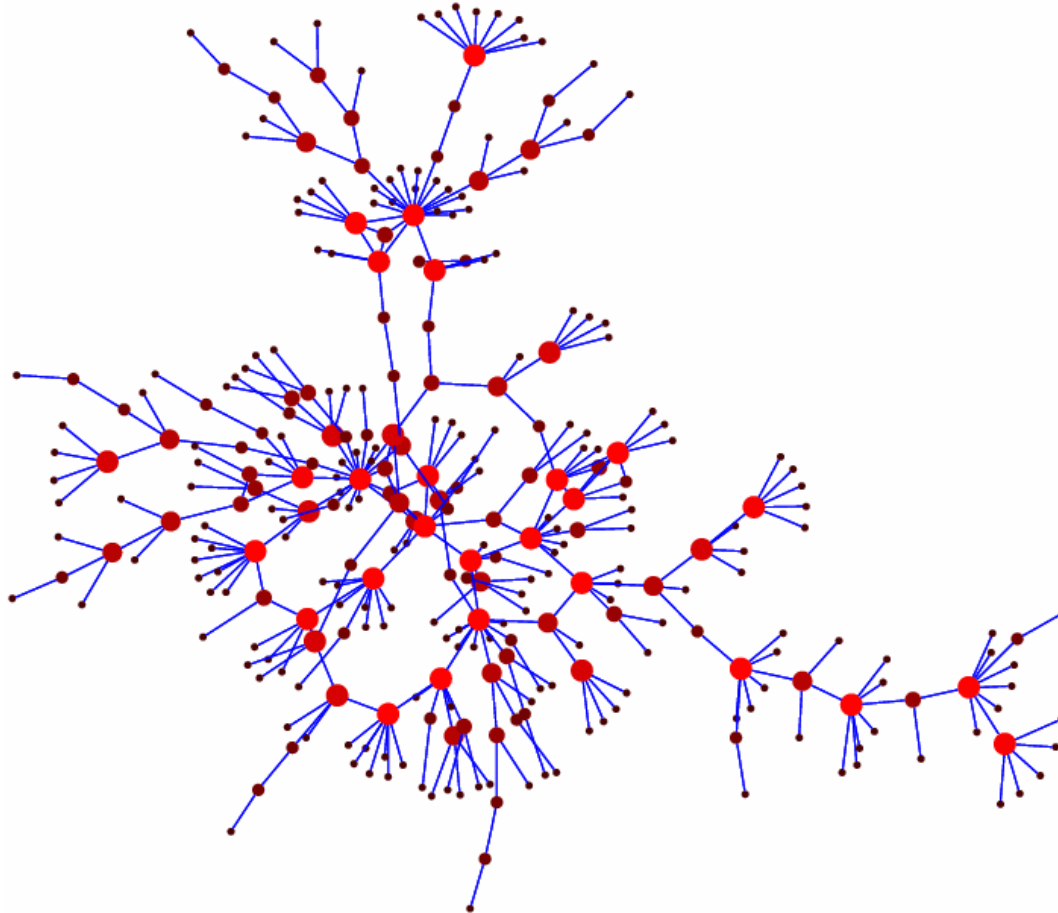
- Les petits mondes constituent un bon modèle pour étudier la propagation d'*influences* à travers un réseau. Par exemple, ces influences peuvent être des épidémies se propageant au sein d'une population ou des virus informatiques diffusés par messages électroniques. On distingue en général deux modes de propagation :
 - **Mode SIR** (Susceptible, Infected, Removed) : dans ce mode lorsqu'un individu a été infecté, il est ensuite retiré du réseau. **Ex: maladies ;**
 - **Mode SIS** (Susceptible, Infected, Susceptible) : dans ce mode les individus guérissent de la maladie et sont à nouveau susceptibles de l'attraper. **Ex: virus informatiques ;**

Épidémies : importance de la topologie du réseau de contacts

Virus sur ordinateurs	→	Topologie d'Internet
Computer worms	→	Topologie du réseau d'@mail
Épidémiologie	→	Topologie du réseau de transports

La structure petit monde du réseau de transport et de contacts joue un rôle majeur sur la propagation des épidémies.

Réseau de contacts sexuels humains

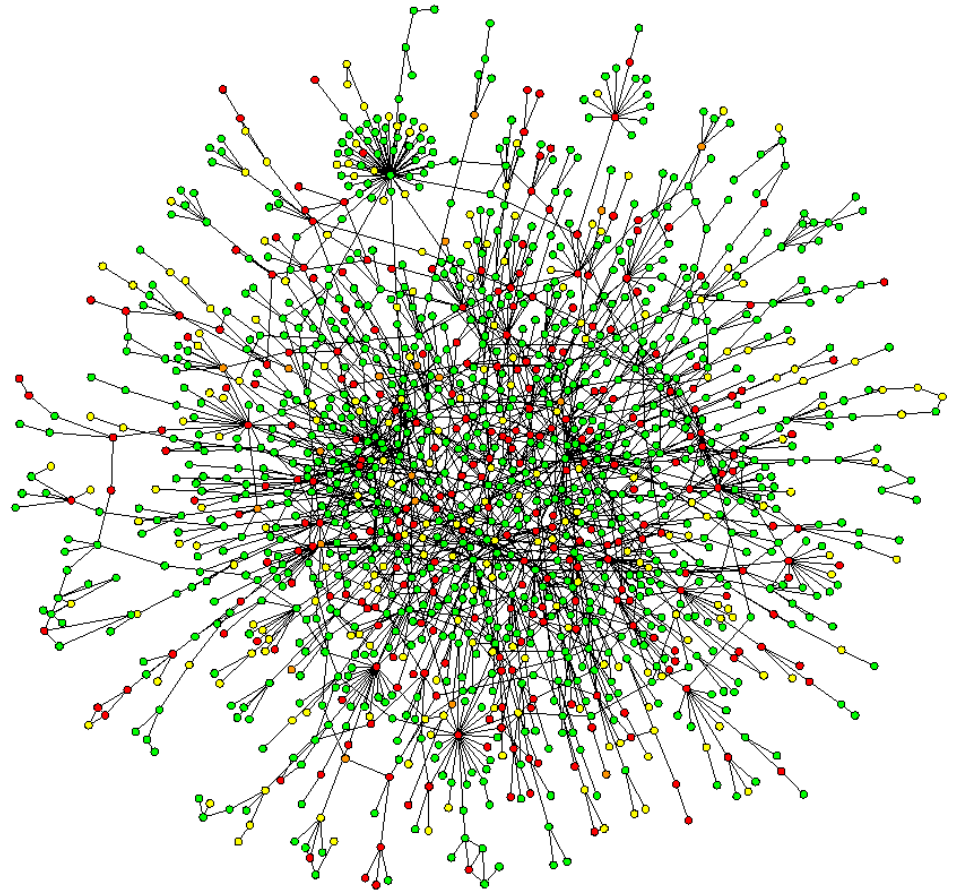


The structure and function of complex networks,
M. E. J. Newman, SIAM Review 45, 167-256 (2003)

Réseau d'interaction des protéines

Noeuds : protéines

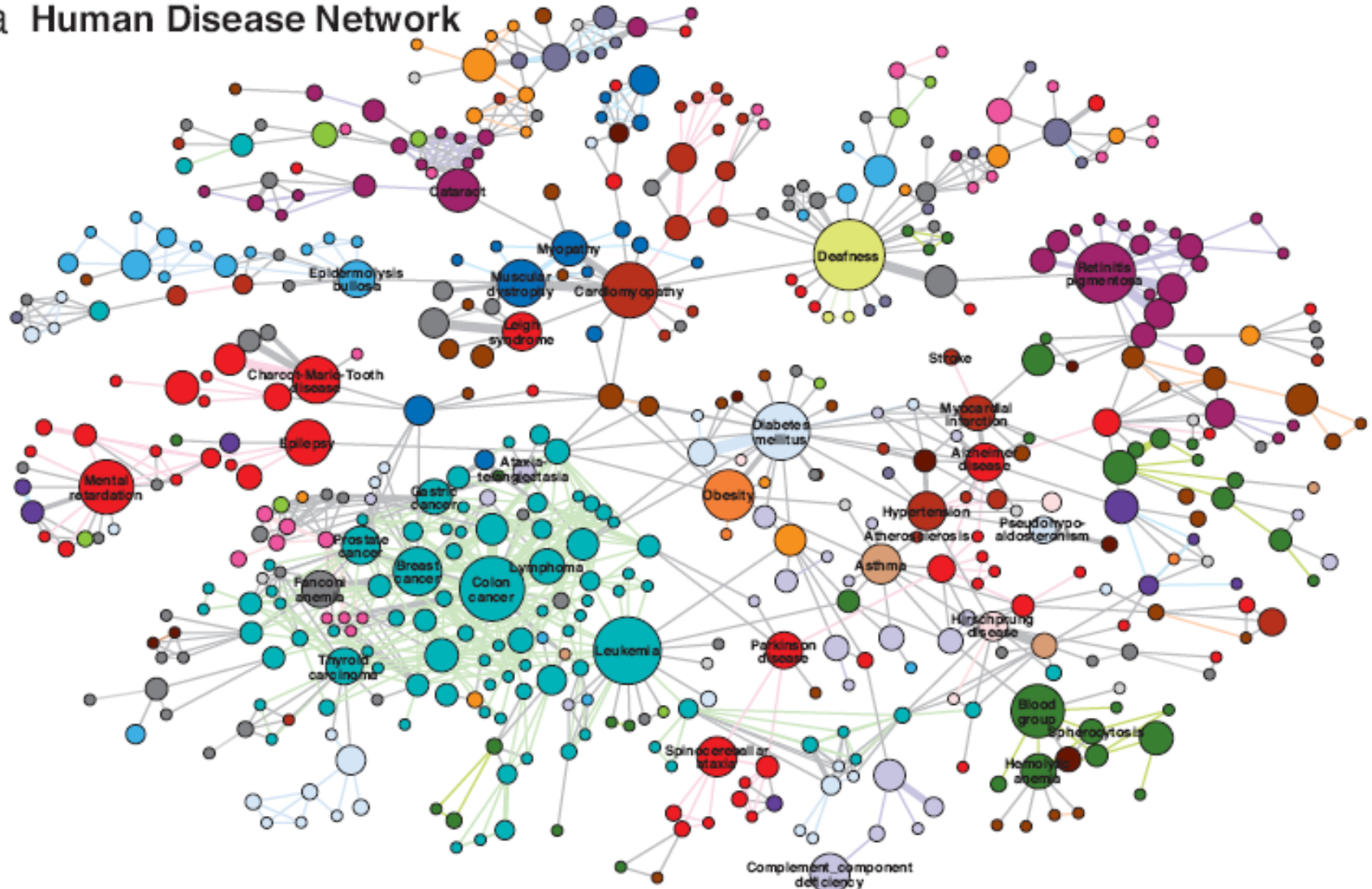
Arêtes: interactions



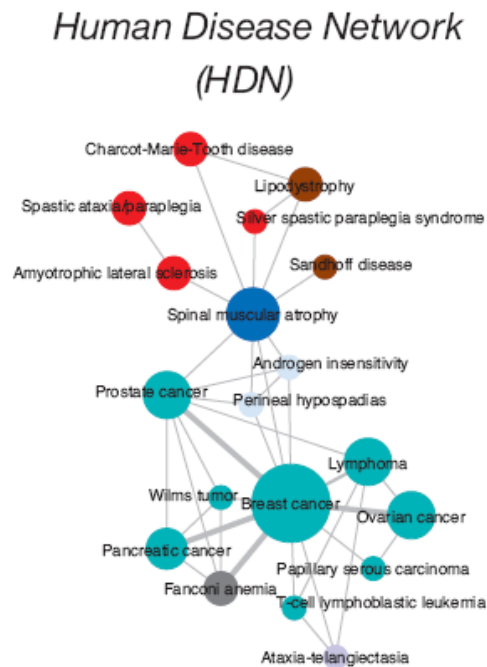
Jeong, H., Mason, S., Barabasi, A.-L. & Oltvai, Z. N., "Lethality and centrality in protein networks", *Nature*, 411, pp 41, 2001.

Réseau des maladies (selon Barabasi)

a Human Disease Network

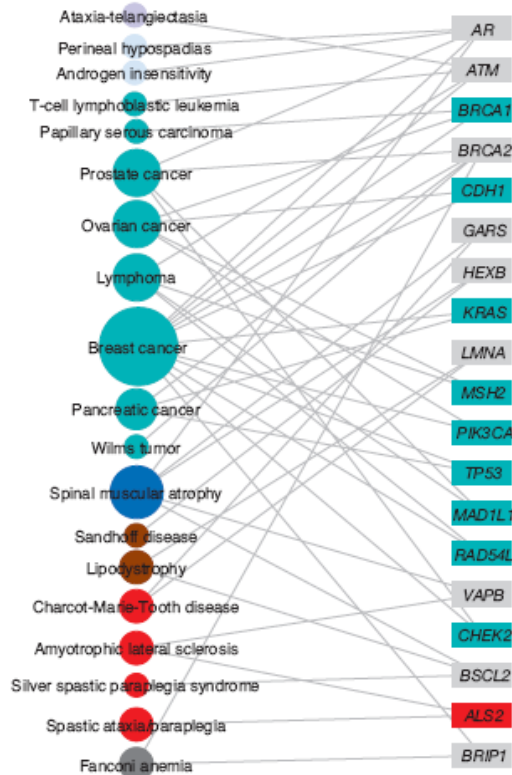


Les réseaux des maladies et des gènes humains

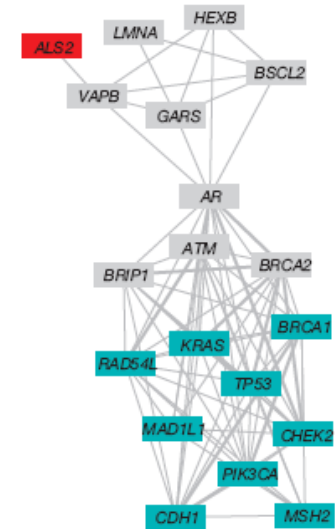


DISEASOME

disease phenotype disease genome



Disease Gene Network (DGN)



Conclusion : à quoi tout cela sert-il ?

■ Pour la science :

- Comprendre et théoriser
- Modéliser, prévoir

■ Agir :

- Algorithmes
 - routage
 - data-mining
- Fiabilisation
- Épidémiologie
- Interactions biologiques
- Etc.

Annexes

- **Définitions des graphes**
- **Rappel : cas d'Internet**

Définitions de base des graphes (1)

- Un graphe est composé de **n sommets** et de **m arêtes** ;
- L'**ordre** du graphe **n** est le nombre de sommets présents dans le graphe ;
- Le **degré** **k** d'un sommet est le nombre d'arêtes dont ce sommet est une extrémité ;
- La **longueur** d'une chaîne est le nombre d'arêtes qui la composent ;
- La **distance** **d** entre deux sommets est la plus courte longueur des chaînes qui la relient. La moyenne de ces distances est la longueur caractéristique **L** ;
- Le **diamètre** **D** d'un graphe est la plus grande distance entre deux sommets ;

Définitions de base des graphes (2)

- **degré moyen** d'un graphe:

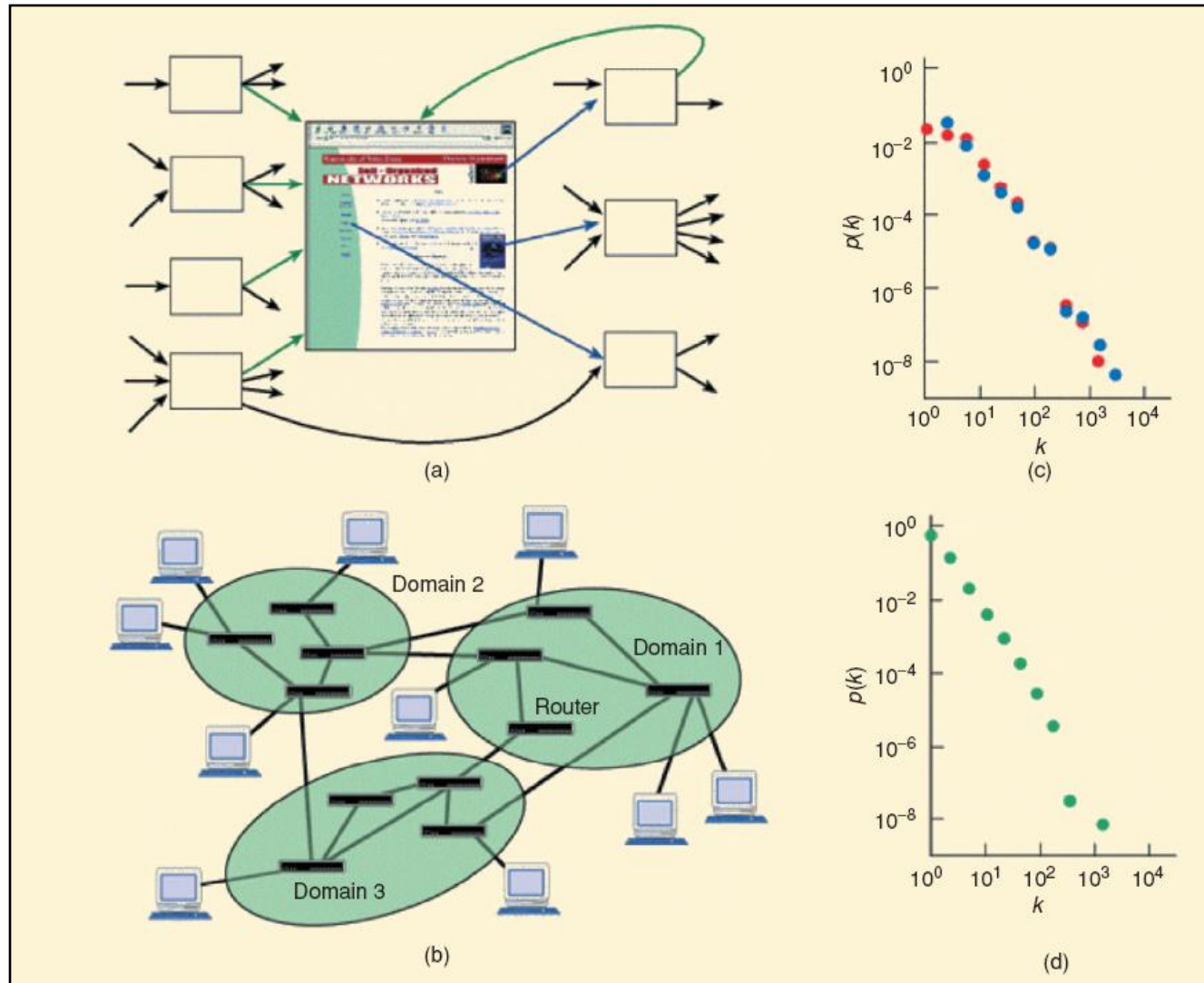
$$k=2m/n$$

- **densité** d'un grand graphe : proportion de **m** par rapport au nombre total possible **$m_{\max}=n(n-1)/2$** :

$$\delta=m/m_{\max} = 2m/(n(n-1))$$

- **connexité** d'un graphe: un graphe est complet si deux nœuds quelconques peuvent être reliés par une chaîne ;

Distribution de puissance dans Internet



WEB_{in}

WEB_{out}

Routers