

M1 Info - Systemes Complexes Avances

# Cours 3 - Proprietes et modeles

Distribution des degres, densite locale, modele de configuration

Semestre Automne 2022-2023 - Université Côte D'azur

Christophe Crespelle

`christophe.crespelle@univ-cotedazur.fr`



# Rappel

## Quatre propriétés fondamentales des réseaux complexes

- Densité globale : faible
- Distances : courtes
- Distribution des degrés
- Densité locale

# Rappel

## Quatre propriétés fondamentales des réseaux complexes

- Densité globale : faible
- Distances : courtes
- Distribution des degrés
- Densité locale

## Jusqu'à maintenant

- Les deux premières propriétés
- Conçu un modèle  $(G_{n,m})$  qui les restitue

### III. Distribution des degres

#### Rappel (Degre)

Le **degre** d'un noeud  $u$  dans un reseau  $G$ , note  $d^\circ(u)$  est le nombre de voisins de  $u$  dans  $G$ .

### III. Distribution des degres

#### **Rappel (Degre)**

Le **degre** d'un noeud  $u$  dans un reseau  $G$ , note  $d^\circ(u)$  est le nombre de voisins de  $u$  dans  $G$ .

#### **Definition (Distribution des degres (PDF))**

La **distribution des degres** d'un reseau  $G$  est la fonction qui a chaque entier naturel  $k \geq 0$  associe le nombre (ou la proportion) de sommets de  $G$  qui ont degre exactement  $k$ .

PDF = Probability Distribution Function

### III. Distribution des degres

#### Rappel (Degre)

Le **degre** d'un noeud  $u$  dans un reseau  $G$ , note  $d^{\circ}(u)$  est le nombre de voisins de  $u$  dans  $G$ .

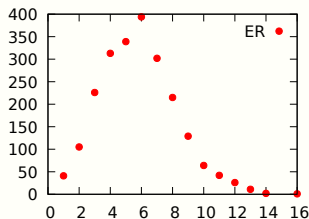
#### Definition (Distribution des degres (PDF))

La **distribution des degres** d'un reseau  $G$  est la fonction qui a chaque entier naturel  $k \geq 0$  associe le nombre (ou la proportion) de sommets de  $G$  qui ont degre exactement  $k$ .

PDF = Probability Distribution Function

#### Version en nombre de noeuds

figeys (2217)



### III. Distribution des degres

#### Rappel (Degre)

Le **degre** d'un noeud  $u$  dans un reseau  $G$ , note  $d^\circ(u)$  est le nombre de voisins de  $u$  dans  $G$ .

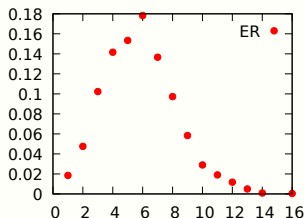
#### Definition (Distribution des degres (PDF))

La **distribution des degres** d'un reseau  $G$  est la fonction qui a chaque entier naturel  $k \geq 0$  associe le nombre (ou la proportion) de sommets de  $G$  qui ont degre exactement  $k$ .

PDF = Probability Distribution Function

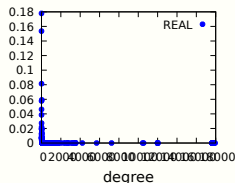
#### Version en proportion de noeuds (PDF)

figeys (2217)

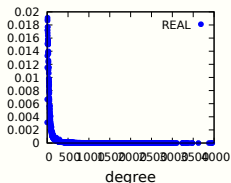


### III. Distribution des degres des reseaux reels

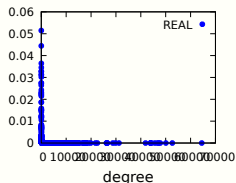
cnr-2000 (227058)



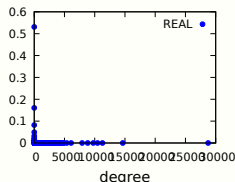
actor-col. (374511)



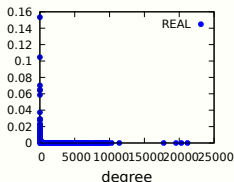
eu-2005 (835044)



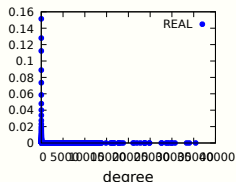
youtube (1134890)



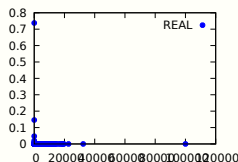
in-2004 (1148875)



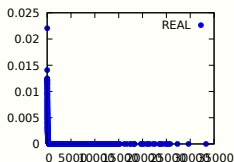
as-skitter (1694616)



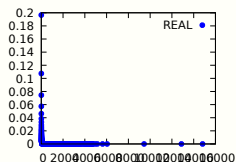
wiki-Talk (2388953)



orkut (3072441)



LiveJournal (3997962)

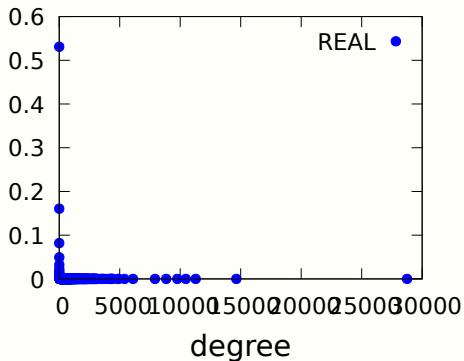




# Un probleme de visualisation

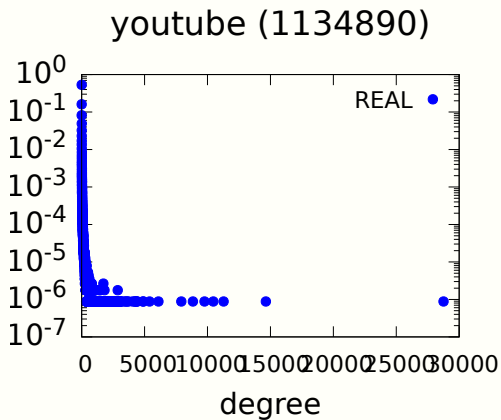
## Echelle classique : lin-lin

youtube (1134890)



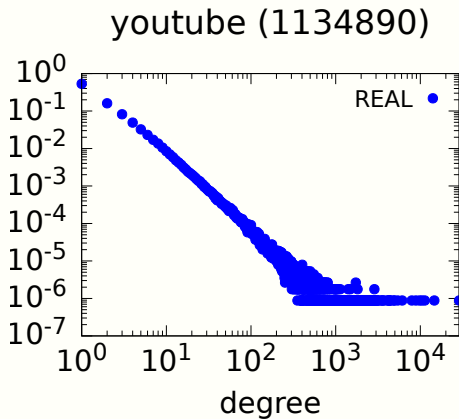
# Un probleme de visualisation

## Echelle lin-log



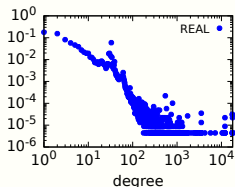
# Un probleme de visualisation

## Echelle log-log

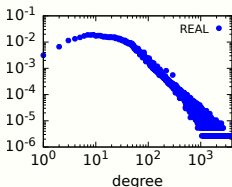


### III. Distribution des degres des reseaux reels

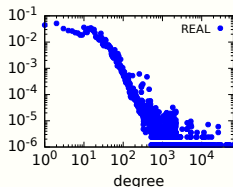
cnr-2000 (227058)



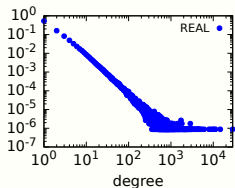
actor-col. (374511)



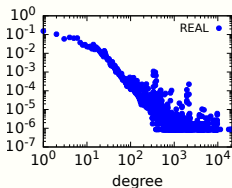
eu-2005 (835044)



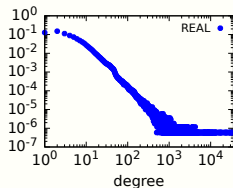
youtube (1134890)



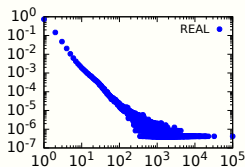
in-2004 (1148875)



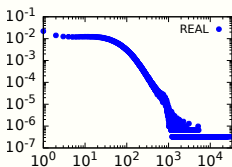
as-skitter (1694616)



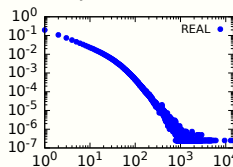
wiki-Talk (2388953)



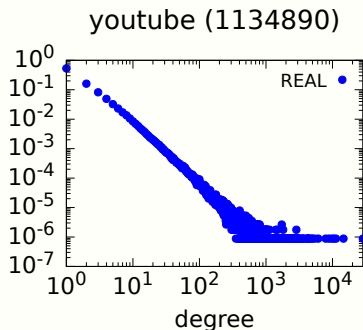
orkut (3072441)



LiveJournal (3997962)

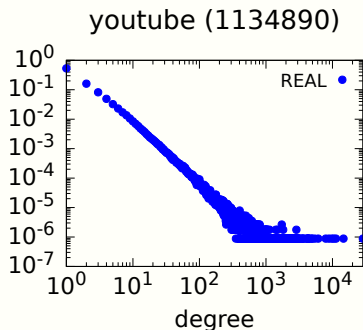


### III. Distribution des degres des reseaux reels



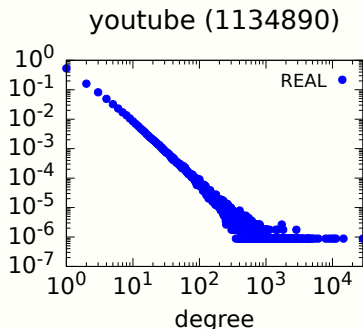
- Distribution heterogene

### III. Distribution des degres des reseaux reels



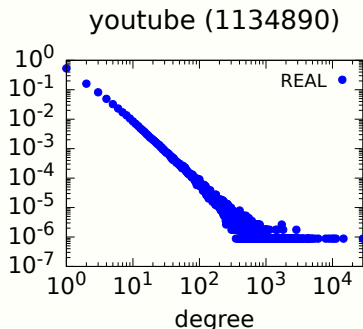
- Distribution heterogene
  - ▶ Large intervalle de valeurs possibles

### III. Distribution des degres des reseaux reels



- Distribution heterogene
  - ▶ Large intervalle de valeurs possibles
  - ▶ Decroissante lente (polynomiale) de la distribution

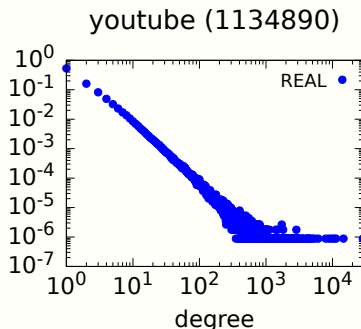
### III. Distribution des degres des reseaux reels



- Distribution heterogene
  - ▶ Large intervalle de valeurs possibles
  - ▶ Decroissante lente (polynomiale) de la distribution
- Consequence : la valeur moyenne n'est pas representative



### III. Distribution des degres des reseaux reels

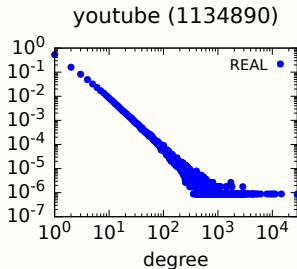


- Distribution heterogene
  - ▶ Large intervalle de valeurs possibles
  - ▶ Decroissante lente (polynomiale) de la distribution
- Consequence : la valeur moyenne n'est pas representative

Question ???

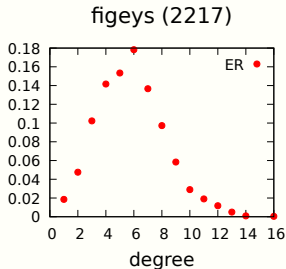
*Dans une distribution heterogene, laquelle est la plus grande : la moyenne ou la mediane ?*

# Distribution heterogene vs. homogene



Distribution heterogene

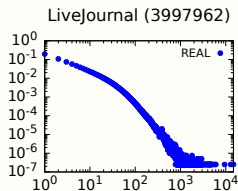
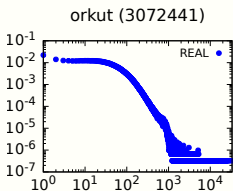
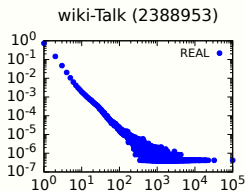
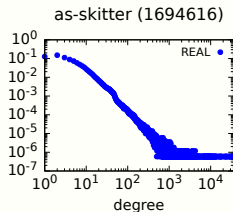
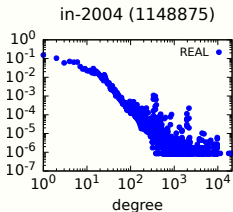
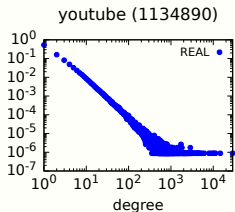
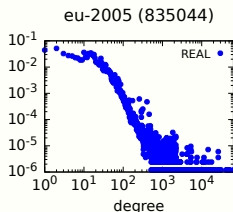
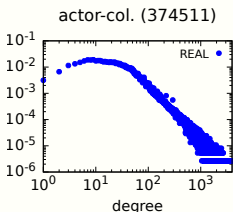
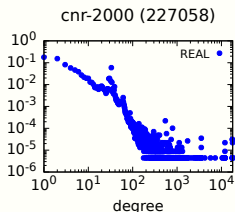
- Large intervalle de valeurs possibles
- Decroissante lente (polynomiale) de la distribution
- La valeur moyenne n'est pas representative
- Archetype : powerlaw  $P(k) \sim k^{-\alpha}$



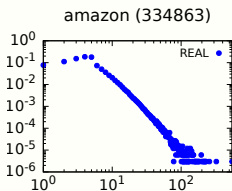
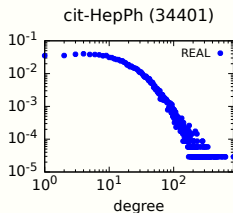
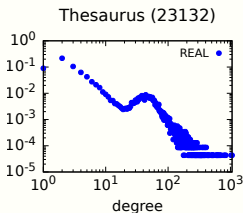
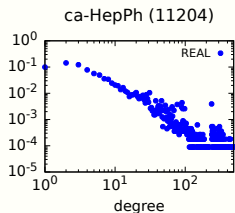
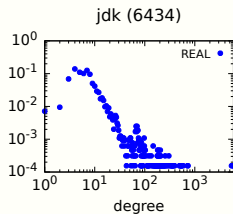
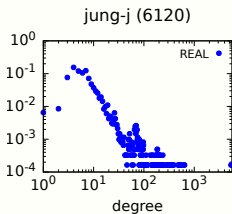
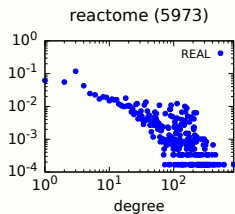
Distribution homogene

- Valeurs concentrees autour de la moyenne
- Decroissante rapide (exponentielle) de la distribution
- La valeur moyenne est representative
- Archetypes :
  - loi normale  $P(k) \sim e^{-k^2}$  ou
  - loi de Poisson  $P(k) \sim \frac{\lambda^k}{k!}$

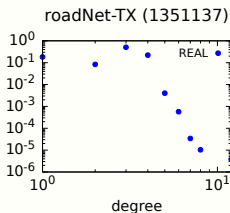
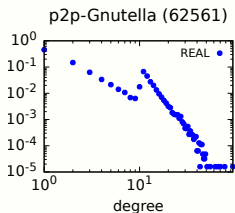
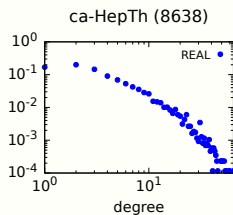
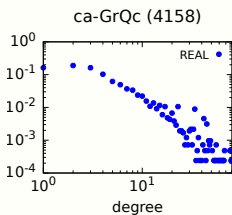
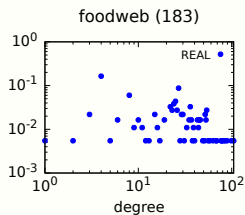
### III. Distribution des degres des reseaux reels



### III. Distribution des degres des reseaux reels - bis



### III. Distribution des degres des reseaux reels - exceptions



**Conclusion**

# Distribution de degres des graphes d'Erdős-Rényi

## Question ???

*La distribution de degres des modeles  $G_{n,m}$  et  $G_{n,p}$  est elle heterogene comme celle des reseaux reels ou plutot homogene ?*

## Reponse par le calcul dans $G_{n,p}$

- On regarde pour les graphes a degre myoen fixe  $\lambda$
- on a alors  $p = \frac{\lambda}{n-1}$
- d'ou la proba  $P(k)$  d'avoir degre exactement  $k$  pour un sommet

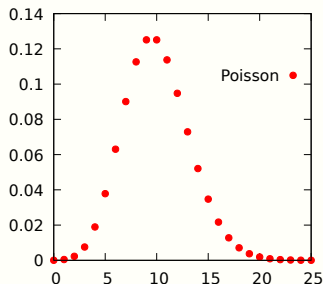
$$\begin{aligned}P(k) &= \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= \frac{A_n^k}{k!} \left(\frac{\lambda}{n-1}\right)^k \left(1 - \frac{\lambda}{n-1}\right)^{n-1-k} \\ &= \frac{A_n^k}{(n-1)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n-1}\right)^{-k} \left(1 - \frac{\lambda}{n-1}\right)^{n-1}\end{aligned}$$

$k$  fixe,  $n \rightarrow +\infty$

$$P(k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

# Distribution de Poisson

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$



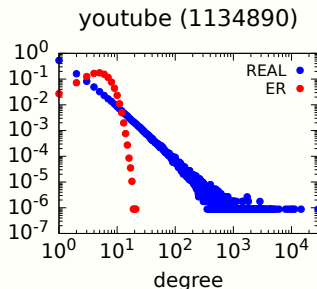
## Conclusion

- Distribution de degres de  $G_{n,p}$  et  $G_{n,m}$  : homogene
- Le modele d'Erdős-Rényi **ne reproduit pas correctement** la distribution des degres observee pour les reseaux reels (heterogene)

# Verification expérimentale

## Protocole expérimental

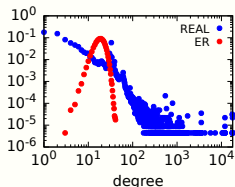
- Pour chaque resau reel  $G$  de notre collection
- On genere un graphe  $G'$  avec le  $G_{n,m}$ , de meme parametre  $m$  que  $G$
- On compare la distribution de  $G$  et du graphe genere



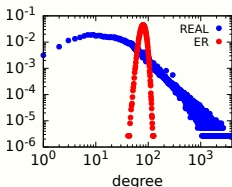


### III. Distribution des degres des reseaux reels

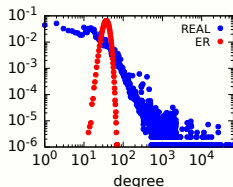
cnr-2000 (227058)



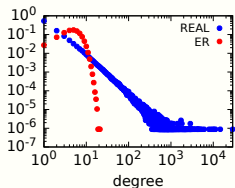
actor-col. (374511)



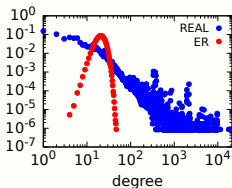
eu-2005 (835044)



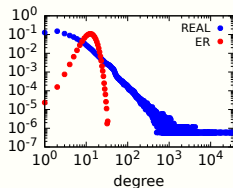
youtube (1134890)



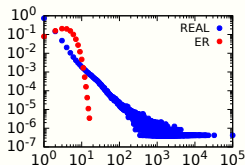
in-2004 (1148875)



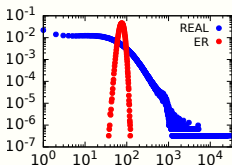
as-skitter (1694616)



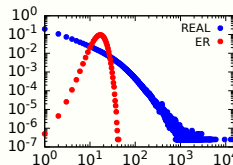
wiki-Talk (2388953)



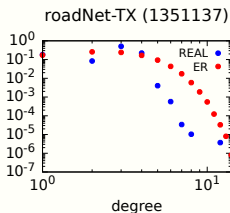
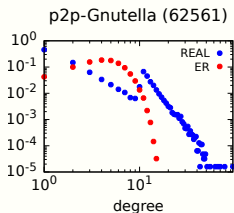
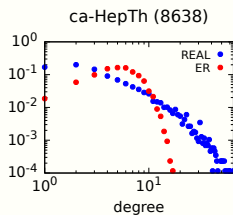
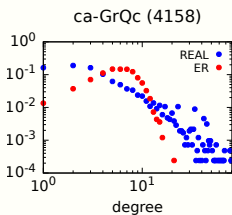
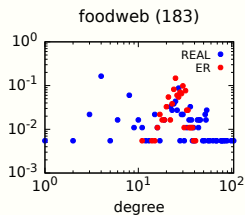
orkut (3072441)



LiveJournal (3997962)



### III. Distribution des degres des reseaux reels - exceptions



**Conclusion**

## Un deuxième modèle : le modèle de configuration

- Le modèle d'ER reproduit correctement :
  - ▶ La densité globale (paramètre)
  - ▶ Les distances courtes
- Mais pas la distribution des degrés hétérogène des réseaux réels
- $\implies$  Il faut un nouveau modèle !

### Le modèle de configuration [Molloy & Reed 1995]

- Idée principale : le paramètre n'est plus  $m$  mais la distribution des degrés
- Remarque : la distribution des degrés fixe  $m!!!$
- Dans le modèle de configuration, on fixe davantage la structure du réseau.

# Le modele de configuration [Molloy & Reed 1995]

- Re-rappel sur ce qu'on appelle un modele (ici)
  - ▶ Un procede aleatoire qui genere un graphe
  - ▶ Avec des proprietes prescrites
  - ▶ Tire uniform. aleat. parmi les graphes ayant ces proprietes
- Ce qu'on voudrait faire (a ce stade)
  - ▶ Tirer uniformement aleatoirement un graphe parmi ceux ayant
    - ▶  $n$  sommets
    - ▶  $m$  aretes
    - ▶ une distance moyenne  $l$
    - ▶ une distribution de degres  $P(k)$
  - ▶ ex :  $n = 10$ ,  $m = 17$ ,  $l = 4.68$ ,  
 $P(1) = 0.3$ ,  $P(2) = 0.1$ ,  $P(3) = 0.1$ ,  $P(4) = 0.2$ ,  $P(6) = 0.3$
- Ce qu'on va faire (parce que c'est plus facile !)
  - ▶ Tirer uniformement aleatoirement un graphe parmi ceux ayant
    - ▶  $n$  sommets
    - ▶ une distribution des degres  $P(k)$
  - ▶ On a  $m$  fixe par  $P(k)$  car  $m = \frac{n}{2} \cdot \sum_{k \geq 0} k \cdot P(k)$
  - ▶ On va verifier a posteriori que la distance moyenne est faible.

## Le modele de configuration [Molloy & Reed 1995]

- Le modele  $G_{n,m}$ 
  - ▶ On part d'un graphe a  $n$  sommets
  - ▶ Pour chaque  $k$ , on choisit  $n \cdot P(k)$  sommets qui auront degres exactement  $k$
  - ▶ Chaque sommet recoit  $k$  demi-liens, avec  $k$  son degre choisi au point precedent
  - ▶ On tire unif. alea. (et sans remise)  $m$  couples de demi-liens parmi les  $2m$  demi-liens disponibles
- Ex :  $n = 10$ ,  
 $P(1) = 0.3, P(2) = 0.1, P(3) = 0.1, P(4) = 0.2, P(6) = 0.3$

# Le modele de configuration [Molloy & Reed 1995]

## Un petit probleme technique

- Lors du choix aleatoire des couples de demi-liens, on peut etre amene a former :
  - ▶ Des boucles
  - ▶ Des liens multiples
  - ▶ Pas acceptable! car on veut generer des graphes simples

## Question ???

*Quel est la complexite, lors du tirage d'un couple de demi-liens, de verifier si c'est une boucle? de verifier si le lien existe deja?*

## Question ???

*Si on rejette tous les tirages non-valides, est-ce que ca regle le probleme?*

# Le modele de configuration [Molloy & Reed 1995]

## Solution

- On ne verifie rien
- A la fin, on retire toutes les boucles et aretes multiples

## Consequences

- La distribution des degres est legerement modifiee
- Le nombre d'aretes est un peu moindre

## Question

- Est-ce que ca modifie beaucoup la distribution et le nombre d'aretes ?
  - ▶ Cela depend peut-etre de la distribution...

## Question ???

*Quid dans le cas des distributions de reseaux reels ?*

# Le modele de configuration [Molloy & Reed 1995]

Reseaux reels						Mod. config.	
Context	Network	n	m	$\rho$	$d^0$	m'	%perte
SPECIES	foodweb	183	2434	$1 \cdot 10^{-1}$	26.6	2031	16.6
CO-OCCUR	bible-names	1707	9059	$6 \cdot 10^{-3}$	10.6	8713	3.8
PROTEIN	figeys	2217	6418	$3 \cdot 10^{-3}$	5.8	5897	8.1
CO-AUTHOR	ca-GrQc	4158	13422	$2 \cdot 10^{-3}$	6.5	13343	0.6
PROTEIN	reactome	5973	145778	$8 \cdot 10^{-3}$	48.8	140886	3.4
SOFTWARE	jung-j	6120	50290	$3 \cdot 10^{-3}$	16.4	40692	19.1
SOFTWARE	jdk	6434	53658	$3 \cdot 10^{-3}$	16.7	43931	18.1
INTERNET	as2000	6474	12572	$6 \cdot 10^{-4}$	3.9	11292	10.2
CO-AUTHOR	ca-HepTh	8638	24806	$7 \cdot 10^{-4}$	5.7	24733	0.3
CO-AUTHOR	ca-HepPh	11204	117619	$2 \cdot 10^{-3}$	21.0	113705	3.3
CO-AUTHOR	ca-AstroPh	17903	196972	$1 \cdot 10^{-3}$	22.0	195992	0.5
CO-AUTHOR	ca-CondMat	21363	91286	$4 \cdot 10^{-4}$	8.5	91169	0.1
WORD-REL.	Thesaurus	23132	297094	$1 \cdot 10^{-3}$	25.7	294462	0.9
CITATION-SCI.	cora	23166	89157	$3 \cdot 10^{-4}$	7.7	89001	0.2
INTERNET	as-caida2007	26475	53381	$2 \cdot 10^{-4}$	4.0	48707	8.8
CITATION-SCI.	cit-HepTh	27400	352021	$9 \cdot 10^{-4}$	25.7	349539	0.7
SOFTWARE	linux	30817	213208	$4 \cdot 10^{-4}$	13.8	194248	8.9
CITATION-SCI.	cit-HepPh	34401	420784	$7 \cdot 10^{-4}$	24.5	419772	0.2
INTERNET	topology	34761	107720	$2 \cdot 10^{-4}$	6.2	98340	8.7
P2P-CONNECT.	p2p-Gnutella	62561	147878	$8 \cdot 10^{-5}$	4.7	146211	1.1
WORD-REL.	wordnet	145145	656230	$6 \cdot 10^{-5}$	9.0	655440	0.1
WWW	cnr-2000	227058	2187201	$8 \cdot 10^{-5}$	19.3	2039226	6.8
CO-AUTHOR	dblp	317080	1049866	$2 \cdot 10^{-5}$	6.6	1049248	0.1
CO-SOLD	amazon	334863	925872	$2 \cdot 10^{-5}$	5.5	925674	0.0
CITATION-SCI.	citeseer	365154	1721981	$3 \cdot 10^{-5}$	9.4	1721023	0.1
CO-ACTOR	actor-col.	374511	15014839	$2 \cdot 10^{-4}$	80.2	14971396	0.3
WWW	eu-2005	835044	15718784	$5 \cdot 10^{-5}$	37.6	15326454	2.5
SOCIAL	youtube	1134890	2987624	$5 \cdot 10^{-6}$	5.3	2927671	2.0
WWW	in-2004	1148875	12281937	$2 \cdot 10^{-5}$	21.4	12061006	1.8
ROAD	roadNet-TX	1351137	1879201	$2 \cdot 10^{-6}$	2.8	1869140	0.5
INTERNET	as-skitter	1694616	11094209	$8 \cdot 10^{-6}$	13.1	10932180	1.5
COMMUNIC.	wiki-Talk	2388953	4656682	$2 \cdot 10^{-6}$	3.9	4148927	10.9
SOCIAL	orkut	3072441	117185083	$2 \cdot 10^{-5}$	76.3	117148832	0.0
CITATION-PAT.	cit-Patents	3764117	16511740	$2 \cdot 10^{-6}$	8.8	16504810	0.0
SOCIAL	LiveJournal	3997962	34681189	$4 \cdot 10^{-6}$	17.3	34672804	0.0



# Le modele de configuration [Molloy & Reed 1995]

## Question

- Est-ce que ca modifie beaucoup la distribution et le nombre d'aretes ?
  - ▶ Cela depend peut-etre de la distribution...

## Question ???

*Quid dans le cas des distributions de reseaux reels ?*

## Reponse

- C'est variable
  - ▶ Souvent negligeable
  - ▶ Parfois avec un vrai impact

## Question ???

*Quels sont les reseaux pour lesquels l'impact est sensible ?*

# Distance moyenne dans le modele de configuration ?

Reseaux				Densite		Dist. moy.		
Context	Network	n	m	$\rho$	$d^{\circ}$	Reel	$G_{n,m}$	modc
SPECIES	foodweb	183	2434	$1 \cdot 10^{-1}$	26.6	2.1	1.9	2.1
CO-OCCUR	bible-names	1707	9059	$6 \cdot 10^{-3}$	10.6	3.4	3.4	3.1
PROTEIN	figeys	2217	6418	$3 \cdot 10^{-3}$	5.8	3.8	4.1	3.6
CO-AUTHOR	ca-GrQc	4158	13422	$2 \cdot 10^{-3}$	6.5	6.0	5.7	4.1
PROTEIN	reactome	5973	145778	$8 \cdot 10^{-3}$	48.8	4.2	3.0	2.8
SOFTWARE	jung-j	6120	50290	$3 \cdot 10^{-3}$	16.4	2.1	2.1	2.7
SOFTWARE	jdk	6434	53658	$3 \cdot 10^{-3}$	16.7	2.1	2.1	2.7
INTERNET	as2000	6474	12572	$6 \cdot 10^{-4}$	3.9	3.7	4.7	3.9
CO-AUTHOR	ca-HepTh	8638	24806	$7 \cdot 10^{-4}$	5.7	6.0	5.7	4.7
CO-AUTHOR	ca-HepPh	11204	117619	$2 \cdot 10^{-3}$	21	4.7	3.9	3.2
CO-AUTHOR	ca-AstroPh	17903	196972	$1 \cdot 10^{-3}$	22	4.2	3.6	3.4
CO-AUTHOR	ca-CondMat	21363	91286	$4 \cdot 10^{-4}$	8.6	5.3	5.1	4.3
WORD-REL.	Thesaurus	23132	297094	$1 \cdot 10^{-3}$	25.7	3.5	3.4	3.4
CITATION-SCL.	cora	23166	89157	$3 \cdot 10^{-4}$	7.7	5.9	5.1	4.4
INTERNET	as-caida2007	26475	53381	$2 \cdot 10^{-4}$	4.0	3.9	5.3	4.0
CITATION-SCL.	cit-HepTh	27400	352021	$9 \cdot 10^{-4}$	25.7	4.3	3.5	3.2
SOFTWARE	linux	30817	213208	$4 \cdot 10^{-4}$	13.8	3.2	3.4	3.1
CITATION-SCL.	cit-HepPh	34401	420784	$7 \cdot 10^{-4}$	24.5	4.3	3.6	3.4
INTERNET	topology	34761	107720	$2 \cdot 10^{-4}$	6.2	3.8	4.8	3.7
P2P-CONNECT.	p2p-Gnutella	62561	147878	$8 \cdot 10^{-5}$	4.7	5.9	7.2	5.8
WORD-REL.	wordnet	145145	656230	$6 \cdot 10^{-5}$	9.0	5.5	5.6	4.3
WWW	cnr-2000	227058	2187201	$8 \cdot 10^{-5}$	19.3	9.3	4.7	3.3
CO-AUTHOR	dblp	317080	1049866	$2 \cdot 10^{-5}$	6.6	6.8	7.4	5.4
CO-SOLD	amazon	334863	925872	$2 \cdot 10^{-5}$	5.5	11.9	8.1	6.4
CITATION-SCL.	citeseer	365154	1721981	$3 \cdot 10^{-5}$	9.4	6.5	5.7	4.7
CO-ACTOR	actor-col.	374511	15014839	$2 \cdot 10^{-4}$	80.2	3.7	3.3	3.1
WWW	eu-2005	835044	15718784	$5 \cdot 10^{-5}$	37.6	4.6	3.7	3.1
SOCIAL	youtube	1134890	2987624	$5 \cdot 10^{-6}$	5.3	5.3	6.6	4.5
WWW	in-2004	1148875	12281937	$2 \cdot 10^{-5}$	21.4	8.8	5.3	3.5
ROAD	roadNet-TX	1351137	1879201	$2 \cdot 10^{-6}$	2.8	415.7	16.3	17.4
INTERNET	as-skitter	1694616	11094209	$8 \cdot 10^{-6}$	13.1	5.1	5	3.7
COMMUNIC.	wiki-Talk	2388953	4656682	$2 \cdot 10^{-6}$	3.9	3.9	5.6	4
SOCIAL	orkut	3072441	117185083	$2 \cdot 10^{-5}$	76.3	4.2	3.9	3.5
CITATION-PAT.	cit-Patents	3764117	16511740	$2 \cdot 10^{-6}$	8.8	8.2	7.2	6.1
SOCIAL	LiveJournal	3997962	34681189	$4 \cdot 10^{-6}$	17.4	5.6	5.7	4.6

# Le modele de configuration est-il un bon modele ?

- $n$  : parametre
- Distribution des degres : parametre
- $m$  : fixe par la distribution
  - ▶ Bemol de la perte d'arete parfois significative
  - ▶ Mais souvent negligeable
  - ▶ Existe d'autres methodes de tirage
- Distance moyenne courte et similaire aux reseaux reels : OK

## Conclusion

- Le modele de configuration est un tres bon modele
- Jusqu'a maintenant...
- Mais il reste la quatrieme propriete a examiner

## IV. Densité locale

### Idee generale

On a plus de chance de trouver une arête entre deux sommets  $u$  et  $v$  si ils sont "proches dans le graphe".

⇒ notion de densité locale

### Definition (Coefficient de clustering d'un sommet)

Le **coefficient de clustering** d'un sommet  $u$  d'un graphe  $G$  est défini comme  $cc(u) = |\{v_1 v_2 \in E \mid v_1, v_2 \in N(u)\}|$ .

### Definition (Coefficient de clustering (n°1) d'un graphe)

Le **coefficient de clustering**, dit local, d'un graphe  $G$  est défini comme la moyenne des coef. de clustering de ses sommets :

$$cc_{loc}(G) = \frac{\sum_{u \in V} cc(u)}{n}.$$

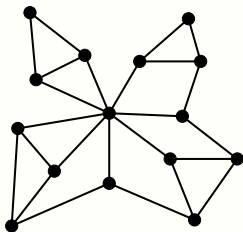
## IV. Densité locale

### Definition (Coefficient de clustering ( $n^{\circ}1$ ) d'un graphe)

Le **coefficient de clustering**, dit local, d'un graphe  $G$  est défini comme la moyenne des coef. de clustering de ses sommets :

$$cc_{loc}(G) = \frac{\sum_{u \in V} cc(u)}{n}.$$

### Exemple :



## IV. Densité locale

### Definition (Patte-d'oie et triangles)

Une **patte-d'oie** est un triplet de sommets  $(a, b, c)$  tel que  $ab \in E$  et  $bc \in E$ . Leur nombre est noté  $\#pdo(G)$ .

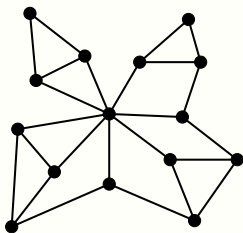
Un **triangle** est un triplet de sommets  $(a, b, c)$  tel que  $ab \in E$  et  $bc \in E$  et  $ac \in E$ . Leur nombre est noté  $\#tri(G)$ .

### Definition (Coefficient de clustering ( $n^{\circ}2$ ) d'un graphe)

Le **coefficient de clustering**, dit global, d'un graphe  $G$  est défini comme la probabilité qu'une patte d'oie soit aussi un triangle :

$$cc_{glob}(G) = \frac{\#tri(G)}{\#pdo(G)}.$$

### Exemple :



# Deux coefficients

- Le coefficient local
  - ▶ Idee : densite dans le voisinage d'un sommet
  - ▶ Pb des sommets ayant un unique voisin
  - ▶ Pb de l'importance relative des noeuds
  - ▶ Avantage d'un coef. par sommet
  
- Le coefficient global
  - ▶ Idee : transitivite (les amis de mes amis sont mes amis)
  - ▶ Vision plus equilibree de la totalite du reseau
  
- Une alternative : le nombre de triangle
  - ▶ Avantage : mesure plus neutre
  - ▶ Inconvenients :
    - ▶ Ce n'est pas une densite
    - ▶ Ne peut pas etre utilise pour comparer des reseaux differents

## Deux coefficients

### Question ???

*Quel ordre de grandeur pour les deux coefficients de clustering pour les reseaux reels ?*

### Question ???

*Comparaison a la densite globale ?*



# Densite locale des reseaux reels

Reseaux				Densite		Distances		Clustering		
Context	Network	n	m	$\rho$	$d^{\circ}$	dist. moy.	$cc_{glob}$	$cc_{loc}$	#triangles	
SPECIES	foodweb	183	2434	$1 \cdot 10^{-1}$	26.6	2.1	0.332	0.324	11292	
CO-OCCUR	bible-names	1707	9059	$6 \cdot 10^{-3}$	10.6	3.4	0.162	0.710	19936	
PROTEIN	figeys	2217	6418	$3 \cdot 10^{-3}$	5.8	3.8	0.008	0.071	897	
CO-AUTHOR	ca-GrQc	4158	13422	$2 \cdot 10^{-3}$	6.5	6.0	0.629	0.665	47779	
PROTEIN	reactome	5973	145778	$8 \cdot 10^{-3}$	48.8	4.2	0.606	0.649	4187395	
SOFTWARE	jung-j	6120	50290	$3 \cdot 10^{-3}$	16.4	2.1	0.011	0.680	182139	
SOFTWARE	jdk	6434	53658	$3 \cdot 10^{-3}$	16.7	2.1	0.011	0.675	194842	
INTERNET	as2000	6474	12572	$6 \cdot 10^{-4}$	3.9	3.7	0.010	0.399	6584	
CO-AUTHOR	ca-HepTh	8638	24806	$7 \cdot 10^{-4}$	5.7	6.0	0.281	0.580	27869	
CO-AUTHOR	ca-HepPh	11204	117619	$2 \cdot 10^{-3}$	21	4.7	0.659	0.690	3357890	
CO-AUTHOR	ca-AstroPh	17903	196972	$1 \cdot 10^{-3}$	22	4.2	0.318	0.669	1350014	
CO-AUTHOR	ca-CondMat	21363	91286	$4 \cdot 10^{-4}$	8.6	5.3	0.262	0.696	171051	
WORD-REL.	Thesaurus	23132	297094	$1 \cdot 10^{-3}$	25.7	3.5	0.040	0.098	409174	
CITATION-SCI.	cora	23166	89157	$3 \cdot 10^{-4}$	7.7	5.9	0.117	0.307	78791	
INTERNET	as-caida2007	26475	53381	$2 \cdot 10^{-4}$	4.0	3.9	0.007	0.333	36365	
CITATION-SCI.	cit-HepTh	27400	352021	$9 \cdot 10^{-4}$	25.7	4.3	0.120	0.329	1478698	
SOFTWARE	linux	30817	213208	$4 \cdot 10^{-4}$	13.8	3.2	0.003	0.141	170862	
CITATION-SCI.	cit-HepPh	34401	420784	$7 \cdot 10^{-4}$	24.5	4.3	0.146	0.296	1276859	
INTERNET	topology	34761	107720	$2 \cdot 10^{-4}$	6.2	3.8	0.049	0.421	554749	
P2P-CONNECT.	p2p-Gnutella	62561	147878	$8 \cdot 10^{-5}$	4.7	5.9	0.004	0.010	2024	
WORD-REL.	wordnet	145145	656230	$6 \cdot 10^{-5}$	9.0	5.5	0.096	0.655	1144648	
WWW	cnr-2000	227058	2187201	$8 \cdot 10^{-5}$	19.3	9.3	0.009	0.635	17701650	
CO-AUTHOR	dblp	317080	1049866	$2 \cdot 10^{-5}$	6.6	6.8	0.306	0.732	2224385	
CO-SOLD	amazon	334863	925872	$2 \cdot 10^{-5}$	5.5	11.9	0.205	0.430	667129	
CITATION-SCI.	citeseer	365154	1721981	$3 \cdot 10^{-5}$	9.4	6.5	0.050	0.212	1350310	
CO-ACTOR	actor-col.	374511	15014839	$2 \cdot 10^{-4}$	80.2	3.7	0.166	0.781	346728049	
WWW	eu-2005	835044	15718784	$5 \cdot 10^{-5}$	37.6	4.6	0.015	0.638	193930463	
SOCIAL	youtube	1134890	2987624	$5 \cdot 10^{-6}$	5.3	5.3	0.006	0.172	3056386	
WWW	in-2004	1148875	12281937	$2 \cdot 10^{-5}$	21.4	8.8	0.096	0.731	452887093	
ROAD	roadNet-TX	1351137	1879201	$2 \cdot 10^{-6}$	2.8	415.7	0.060	0.058	81195	
INTERNET	as-skitter	1694616	11094209	$8 \cdot 10^{-6}$	13.1	5.1	0.005	0.296	28769842	
COMMUNIC.	wiki-Talk	2388953	4656682	$2 \cdot 10^{-6}$	3.9	3.9	0.002	0.201	9203514	
SOCIAL	orkut	3072441	117185083	$2 \cdot 10^{-5}$	76.3	4.2	0.041	0.170	627584181	
CITATION-PAT.	cit-Patents	3764117	16511740	$2 \cdot 10^{-6}$	8.8	8.2	0.067	0.092	7514922	
SOCIAL	LiveJournal	3997962	34681189	$4 \cdot 10^{-6}$	17.4	5.6	0.125	0.354	177820130	

# Quid de nos deux modeles ?

## Question ???

*Quelle est l'esperance du coefficient de clustering global dans le modele  $G_{n,p}$  ?*

# Quid de nos deux modeles ?

## Question ???

*Quelle est l'esperance du coefficient de clustering global dans le modele de configuration ?*

- Proba que  $u$  ait  $k$  autres voisins que  $w$  :  
 $q(k) = P(d^\circ(u) = k + 1)$
- Sachant que  $u$  a  $k$  autres voisins que  $w$  et que  $v$  en a  $k'$ ,  
proba d'un lien entre  $u$  et  $v$  :  $\frac{kk'}{\langle k \rangle N}$
- Au total, proba d'un lien entre  $u$  et  $v$  sachant qu'ils sont tous deux voisins de  $w$  :

$$\sum_{k_i} \sum_{k_j} \frac{k_i k_j}{\langle k \rangle N} q(k_i) q(k_j)$$

# Densite locale des modeles d'ER et de configuration

Reseaux				Densite		Coef. clustering global		
Context	Network	n	m	$d^o$	$\rho$	REAL	MODC	ER
SPECIES	foodweb	183	2434	26.6	$1 \cdot 10^{-1}$	0.332	0.246	0.145
CO-OCCUR	bible-names	1707	9059	10.6	$6 \cdot 10^{-3}$	0.162	0.048	0.006
PROTEIN	figeys	2217	6418	5.8	$3 \cdot 10^{-3}$	0.008	0.060	0.003
CO-AUTHOR	ca-GrQc	4158	13422	6.5	$2 \cdot 10^{-3}$	0.629	0.010	0.002
PROTEIN	reactome	5973	145778	48.8	$8 \cdot 10^{-3}$	0.606	0.061	0.008
SOFTWARE	jung-j	6120	50290	16.4	$3 \cdot 10^{-3}$	0.011	0.023	0.003
SOFTWARE	jdk	6434	53658	16.7	$3 \cdot 10^{-3}$	0.011	0.022	0.003
INTERNET	as2000	6474	12572	3.9	$6 \cdot 10^{-4}$	0.010	0.018	0.001
CO-AUTHOR	ca-HepTh	8638	24806	5.7	$7 \cdot 10^{-4}$	0.281	0.003	0.001
CO-AUTHOR	ca-HepPh	11204	117619	21	$2 \cdot 10^{-3}$	0.659	0.056	0.002
CO-AUTHOR	ca-AstroPh	17903	196972	22	$1 \cdot 10^{-3}$	0.318	0.010	0.001
CO-AUTHOR	CO-CondMat	21363	91286	8.6	$4 \cdot 10^{-4}$	0.262	0.002	0.000
WORD-REL.	Thesaurus	23132	297094	25.7	$1 \cdot 10^{-3}$	0.040	0.016	0.001
CITATION-SCL.	cora	23166	89157	7.7	$3 \cdot 10^{-4}$	0.117	0.003	0.000
INTERNET	as-caida2007	26475	53381	4.0	$2 \cdot 10^{-4}$	0.007	0.016	0.000
CITATION-SCL.	cit-HepTh	27400	352021	25.7	$9 \cdot 10^{-4}$	0.120	0.011	0.001
SOFTWARE	linux	30817	213208	13.8	$4 \cdot 10^{-4}$	0.003	0.013	0.000
CITATION-SCL.	cit-HepPh	34401	420784	24.5	$7 \cdot 10^{-4}$	0.146	0.004	0.001
INTERNET	topology	34761	107720	6.2	$2 \cdot 10^{-4}$	0.049	0.041	0.000
P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	$8 \cdot 10^{-5}$	0.004	0.000	0.000
WORD-REL.	wordnet	145145	656230	9.0	$6 \cdot 10^{-5}$	0.096	0.002	0.000
WWW	cnr-2000	227058	2187201	19.3	$8 \cdot 10^{-5}$	0.009	0.006	0.000
CO-AUTHOR	dblp	317080	1049866	6.6	$2 \cdot 10^{-5}$	0.306	0.000	0.000
CO-SOLD	amazon	334863	925872	5.5	$2 \cdot 10^{-5}$	0.205	0.000	0.000
CITATION-SCL.	citeseer	365154	1721981	9.4	$3 \cdot 10^{-5}$	0.050	0.001	0.000
CO-ACTOR	actor-col.	374511	15014839	80.2	$2 \cdot 10^{-4}$	0.166	0.006	0.000
WWW	eu-2005	835044	15718784	37.6	$5 \cdot 10^{-5}$	0.015	0.007	0.000
SOCIAL	youtube	1134890	2987624	5.3	$5 \cdot 10^{-6}$	0.006	0.005	0.000
WWW	in-2004	1148875	12281937	21.4	$2 \cdot 10^{-5}$	0.096	0.016	0.000
ROAD	roadNet-TX	1351137	1879201	2.8	$2 \cdot 10^{-6}$	0.060	0.000	0.000
INTERNET	as-skitter	1694616	11094209	13.1	$8 \cdot 10^{-6}$	0.005	0.005	0.000
COMMUNIC.	wiki-Talk	2388953	4656682	3.9	$2 \cdot 10^{-6}$	0.002	0.024	0.000
SOCIAL	orkut	3072441	117185083	76.3	$2 \cdot 10^{-5}$	0.041	0.001	0.000
CITATION-PAT.	cit-Patents	3764117	16511740	8.8	$2 \cdot 10^{-6}$	0.067	0.000	0.000
SOCIAL	LiveJournal	3997962	34681189	17.4	$4 \cdot 10^{-6}$	0.125	0.000	0.000

# Densite locale des modeles d'ER et de configuration

Context	Reseaux			Densite		Coef. clustering local		
	Network	n	m	d°	$\rho$	REAL	MODC	ER
SPECIES	foodweb	183	2434	26.6	$1 \cdot 10^{-1}$	0.324	0.262	0.145
CO-OCCUR	bible-names	1707	9059	10.6	$6 \cdot 10^{-3}$	0.710	0.063	0.006
PROTEIN	figeys	2217	6418	5.8	$3 \cdot 10^{-3}$	0.071	0.114	0.003
CO-AUTHOR	ca-GrQc	4158	13422	6.5	$2 \cdot 10^{-3}$	0.665	0.010	0.002
PROTEIN	reactome	5973	145778	48.8	$8 \cdot 10^{-3}$	0.649	0.064	0.008
SOFTWARE	jung-j	6120	50290	16.4	$3 \cdot 10^{-3}$	0.680	0.225	0.003
SOFTWARE	jdk	6434	53658	16.7	$3 \cdot 10^{-3}$	0.675	0.217	0.003
INTERNET	as2000	6474	12572	3.9	$6 \cdot 10^{-4}$	0.399	0.115	0.001
CO-AUTHOR	ca-HepTh	8638	24806	5.7	$7 \cdot 10^{-4}$	0.580	0.003	0.001
CO-AUTHOR	ca-HepPh	11204	117619	21	$2 \cdot 10^{-3}$	0.690	0.062	0.002
CO-AUTHOR	ca-AstroPh	17903	196972	22	$1 \cdot 10^{-3}$	0.669	0.010	0.001
CO-AUTHOR	ca-CondMat	21363	91286	8.6	$4 \cdot 10^{-4}$	0.696	0.003	0.000
WORD-REL.	Thesaurus	23132	297094	25.7	$1 \cdot 10^{-3}$	0.098	0.017	0.001
CITATION-SCI.	cora	23166	89157	7.7	$3 \cdot 10^{-4}$	0.307	0.003	0.000
INTERNET	as-caida2007	26475	53381	4.0	$2 \cdot 10^{-4}$	0.333	0.102	0.000
CITATION-SCI.	cit-HepTh	27400	352021	25.7	$9 \cdot 10^{-4}$	0.329	0.013	0.001
SOFTWARE	linux	30817	213208	13.8	$4 \cdot 10^{-4}$	0.141	0.119	0.000
CITATION-SCI.	cit-HepPh	34401	420784	24.5	$7 \cdot 10^{-4}$	0.296	0.005	0.001
INTERNET	topology	34761	107720	6.2	$2 \cdot 10^{-4}$	0.421	0.127	0.000
P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	$8 \cdot 10^{-5}$	0.010	0.000	0.000
WORD-REL.	wordnet	145145	656230	9.0	$6 \cdot 10^{-5}$	0.655	0.002	0.000
WWW	cnr-2000	227058	2187201	19.3	$8 \cdot 10^{-5}$	0.635	0.090	0.000
CO-AUTHOR	dblp	317080	1049866	6.6	$2 \cdot 10^{-5}$	0.732	0.000	0.000
CO-SOLD	amazon	334863	925872	5.5	$2 \cdot 10^{-5}$	0.430	0.000	0.000
CITATION-SCI.	citeseer	365154	1721981	9.4	$3 \cdot 10^{-5}$	0.212	0.001	0.000
CO-ACTOR	actor-col.	374511	15014839	80.2	$2 \cdot 10^{-4}$	0.781	0.006	0.000
WWW	eu-2005	835044	15718784	37.6	$5 \cdot 10^{-5}$	0.638	0.038	0.000
SOCIAL	youtube	1134890	2987624	5.3	$5 \cdot 10^{-6}$	0.172	0.014	0.000
WWW	in-2004	1148875	12281937	21.4	$2 \cdot 10^{-5}$	0.731	0.031	0.000
ROAD	roadNet-TX	1351137	1879201	2.8	$2 \cdot 10^{-6}$	0.058	0.000	0.000
INTERNET	as-skitter	1694616	11094209	13.1	$8 \cdot 10^{-6}$	0.296	0.022	0.000
COMMUNIC.	wiki-Talk	2388953	4656682	3.9	$2 \cdot 10^{-6}$	0.201	0.101	0.000
SOCIAL	orkut	3072441	117185083	76.3	$2 \cdot 10^{-5}$	0.170	0.001	0.000
CITATION-PAT.	cit-Patents	3764117	16511740	8.8	$2 \cdot 10^{-6}$	0.092	0.000	0.000
SOCIAL	LiveJournal	3997962	34681189	17.4	$4 \cdot 10^{-6}$	0.354	0.000	0.000

# Densite locale des modeles d'ER et de configuration

Reseaux				Densite		Nombre de triangles		
Context	Network	n	m	$d^\circ$	$\rho$	REAL	MODC	ER
SPECIES	foodweb	183	2434	26.6	$1 \cdot 10^{-1}$	11292	5344	3113
CO-OCCUR	bible-names	1707	9059	10.6	$6 \cdot 10^{-3}$	19936	4613	200
PROTEIN	figeys	2217	6418	5.8	$3 \cdot 10^{-3}$	897	4584	35
CO-AUTHOR	ca-GrQc	4158	13422	6.5	$2 \cdot 10^{-3}$	47779	737	46
PROTEIN	reactome	5973	145778	48.8	$8 \cdot 10^{-3}$	4187395	383441	19399
SOFTWARE	jung-j	6120	50290	16.4	$3 \cdot 10^{-3}$	182139	78372	729
SOFTWARE	jdk	6434	53658	16.7	$3 \cdot 10^{-3}$	194842	83661	784
INTERNET	as2000	6474	12572	3.9	$6 \cdot 10^{-4}$	6584	5808	11
CO-AUTHOR	ca-HepTh	8638	24806	5.7	$7 \cdot 10^{-4}$	27869	274	30
CO-AUTHOR	ca-HepPh	11204	117619	21	$2 \cdot 10^{-3}$	3357890	251807	1558
CO-AUTHOR	ca-AstroPh	17903	196972	22	$1 \cdot 10^{-3}$	1350014	42342	1793
CO-AUTHOR	ca-CondMat	21363	91286	8.6	$4 \cdot 10^{-4}$	171051	1606	105
WORD-REL.	Thesaurus	23132	297094	25.7	$1 \cdot 10^{-3}$	409174	152803	2814
CITATION-SCI.	cora	23166	89157	7.7	$3 \cdot 10^{-4}$	78791	1829	69
INTERNET	as-caida2007	26475	53381	4.0	$2 \cdot 10^{-4}$	36365	42111	11
CITATION-SCI.	cit-HepTh	27400	352021	25.7	$9 \cdot 10^{-4}$	1478698	123116	2830
SOFTWARE	linux	30817	213208	13.8	$4 \cdot 10^{-4}$	170862	340674	437
CITATION-SCI.	cit-HepPh	34401	420784	24.5	$7 \cdot 10^{-4}$	1276859	38946	2425
INTERNET	topology	34761	107720	6.2	$2 \cdot 10^{-4}$	554749	253557	40
P2P-CONNECT.	p2p-Gnutella	62561	147878	4.7	$8 \cdot 10^{-5}$	2024	198	18
WORD-REL.	wordnet	145145	656230	9.0	$6 \cdot 10^{-5}$	1144648	23980	122
WWW	cnr-2000	227058	2187201	19.3	$8 \cdot 10^{-5}$	17701650	5874109	1206
CO-AUTHOR	dblp	317080	1049866	6.6	$2 \cdot 10^{-5}$	2224385	1470	54
CO-SOLD	amazon	334863	925872	5.5	$2 \cdot 10^{-5}$	667129	195	31
CITATION-SCI.	citeseer	365154	1721981	9.4	$3 \cdot 10^{-5}$	1350310	17047	142
CO-ACTOR	actor-col.	374511	15014839	80.2	$2 \cdot 10^{-4}$	346728049	11423272	85861
WWW	eu-2005	835044	15718784	37.6	$5 \cdot 10^{-5}$	193930463	65833626	8912
SOCIAL	youtube	1134890	2987624	5.3	$5 \cdot 10^{-6}$	3056386	1943389	23
WWW	in-2004	1148875	12281937	21.4	$2 \cdot 10^{-5}$	452887093	60610327	1607
ROAD	roadNet-TX	1351137	1879201	2.8	$2 \cdot 10^{-6}$	81195	1	3
INTERNET	as-skitter	1694616	11094209	13.1	$8 \cdot 10^{-6}$	28769842	19522290	356
COMMUNIC.	wiki-Talk	2388953	4656682	3.9	$2 \cdot 10^{-6}$	9203514	47172837	10
SOCIAL	orkut	3072441	117185083	76.3	$2 \cdot 10^{-5}$	627584181	8226644	74184
CITATION-PAT.	cit-Patents	3764117	16511740	8.8	$2 \cdot 10^{-6}$	7514922	1403	119
SOCIAL	LiveJournal	3997962	34681189	17.4	$4 \cdot 10^{-6}$	177820130	297064	890