

CHAPITRE IV

Petits mondes

EN 1967, le psychologue de l'université de Harvard Stanley Milgram publia dans la revue populaire *Psychology Today* (Milgram, 1967) les résultats d'une expérience pour le moins originale. Après avoir tiré au sort les sujets de son expérience dans les annuaires téléphoniques du Kansas et du Nebraska, Milgram leur fit parvenir ses instructions. Il leur était demandé de faire parvenir une lettre à un destinataire commun habitant à Boston (soit à plus de 2000 km de chez eux) avec la contrainte suivante : les sujets ne pouvaient transmettre la lettre qu'à une personne qu'ils connaissaient personnellement, recommandant à celle-ci de faire de même afin que, de proche en proche, la lettre atteigne son destinataire. La stratégie à suivre, tant pour les sujets choisis par Milgram que pour tous les autres passeurs, était donc d'identifier dans leur entourage quelqu'un plus proche qu'eux-mêmes (géographiquement ou socialement) du destinataire final de la lettre.

Un nombre assez restreint de lettres atteignirent le destinataire, et Milgram étudia les chaînes de personnes ainsi constituées. De manière fort surprenante, ces chaînes étaient très courtes, de l'ordre de 6 personnes.

Ce que fit Milgram lors de son expérience fut de sonder le *réseau d'accointances* des habitants des États-Unis, et en particulier d'en évaluer la *longueur caractéristique*. La notion de réseau est présente dans un grand nombre de disciplines scientifiques : en sociologie (nous venons d'en voir un exemple), mais aussi en biologie (nous en verrons de nombreux exemples dans ce chapitre), en informatique (internet est sans doute le premier exemple qui vient à l'esprit de beaucoup à l'évocation du mot réseau), et bien sûr en linguistique (le chapitre suivant y sera consacré).

Les réseaux ont été largement étudiés par les mathématiciens sous le nom de graphes¹. La première partie de ce chapitre aura pour objectif de définir formellement les graphes et d'introduire des notions clés. L'utilisation des graphes pour décrire des phénomènes réels est ancienne, et la seconde partie présentera le modèle de graphes aléatoires d'Erdős et Rényi longtemps considéré comme référence pour la modélisation de réseaux. Dans la troisième partie, nous verrons que ce modèle n'est pas adapté à la description des réseaux que nous sommes à même de rencontrer dans la nature. Ces réseaux ont en effet des propriétés très particulières qui font d'eux des *petits mondes*, souvent *sans échelle*, une classe de graphes très différents de ceux du modèle Erdős et Rényi. Après avoir explicité ces propriétés, nous présenterons quelques modèles plus appropriés pour construire des graphes proches des

¹Deux terminologies coexistent : une approche mathématique de ces objets aura tendance à les appeler *graphes*, alors qu'une approche plus appliquée aura tendance à les nommer *réseaux*. Certaines notions définies à partir de ces objets peuvent avoir des noms différents dans l'une ou l'autre de ces terminologies. Quand nous définirons une notion appelée différemment selon l'approche, nous expliciterons les deux termes. Nous utiliserons la terminologie autour de *réseau* lorsque nous considérerons des objets réels, et celle autour de *graphe* pour des considérations sur les objets mathématiques.

réseaux que l'on peut observer. Enfin, nous clorons ce chapitre par la présentation des résultats de notre étude du réseau fonctionnel cérébral. La raison d'être essentielle de ce chapitre est d'introduire les outils et les concepts qui nous permettront d'aborder au prochain chapitre l'étude de réseaux lexicaux.

1 Éléments de théorie des graphes

1.1 Les ponts de Königsberg

La ville de Königsberg (aujourd'hui Kaliningrad, dans l'enclave russe de même nom, coincée entre la Pologne et la Lituanie) est construite sur deux îles. La première est reliée au continent par quatre ponts, la seconde par deux. De plus, un septième pont relie les deux îles entre elles (Figure 4.1). Au 18^e siècle les habitants de la ville se posèrent la question suivante : est-il possible de partir d'un endroit quelconque de la ville, de traverser une et une seule fois chaque pont, et de revenir au point de départ. Ce problème était loin d'être trivial, puisqu'il a fallu l'intervention d'Euler lui-même pour y apporter une réponse (c'était impossible). Ce faisant, il avait jeté les fondations de la théorie des graphes (Euler, 1736). Sa solution reposait sur l'idée de représenter la ville par un objet mathématique constitué de sommets et d'arêtes, un graphe (Figure 4.2).

Dans cette section nous définissons tous les concepts que nous utiliserons dans la suite de ce chapitre et dans le suivant. Dans un premier temps nous nous concentrerons sur des définitions très générales sur les graphes, puis nous détaillerons différentes mesures utilisées pour les caractériser.

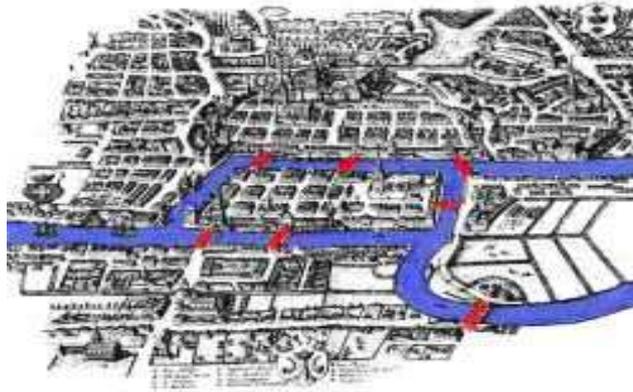


Figure 4.1 – Plan de la ville de Königsberg avec ses 7 ponts au 18^e siècle

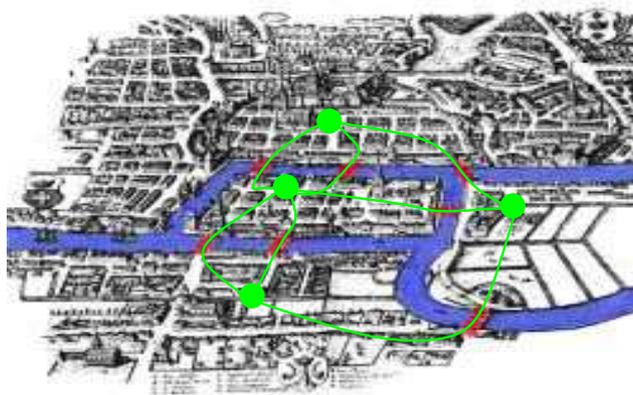


Figure 4.2 – Graphe construit à partir des ponts de Königsberg

1.2 Généralités sur les graphes

Définition 4.1 (Graphe, sommets et arêtes) Un graphe orienté G est constitué d'un ensemble S de sommets et d'un ensemble $A \subset S \times S$ de couples de sommets, les arcs.

Un arc (s_i, s_j) est constitué d'un sommet initial, s_i et d'un sommet final, s_j .

Si l'orientation des arcs n'est pas pertinente, on parlera alors de graphe non-orienté et de l'arête $\{s_i, s_j\}$

Un graphe est aussi appelé un réseau, et l'on parlera alors de nœud pour désigner les sommets et de liens pour les arcs ou les arêtes.

Pour plus de clarté, les graphes sont souvent représentés sous une forme graphique où les sommets sont des points et les arêtes des segments entre ces points (voir Figure 4.3).

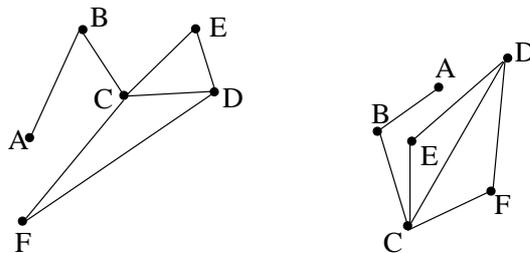


Figure 4.3 – Deux représentations graphiques du même graphe défini par $S = \{A, B, C, D, E, F\}$ et $A = \{\{A, B\}, \{B, C\}, \{C, E\}, \{C, D\}, \{C, F\}, \{D, E\}, \{D, F\}\}$

Définition 4.2 (Sous-graphe) Un sous-graphe de $G = (S, A)$ est un graphe $G' = (S', A')$ tel que $S' \subset S$ et $A' = A \cap S' \times S'$

Définition 4.3 (Voisins) L'ensemble des voisins d'un sommet s_i est l'ensemble des sommets s_j tels que (s_i, s_j) est une arête du graphe.

Définition 4.4 (Chemin) *On appelle chemin une suite de sommets $P = (s_0, \dots, s_n)$ telle que s_i et s_{i+1} sont voisins pour tout $i < n$.*

On appelle longueur du chemin P et on note $|P|$ le nombre de sommets de P .

On appelle plus court chemin entre s_i et s_j , le chemin entre les sommets s_i et s_j composé du plus petit nombre de sommets. Ce chemin n'est pas nécessairement unique.

Définition 4.5 (Graphe connexe, composante connexe) *Un graphe est dit connexe si pour tout couple de sommets s_i et s_j il existe un chemin entre s_i et s_j .*

Si G est un graphe non connexe, on nomme composantes connexes les sous-graphes G'_i de G tels que :

1. G'_i est connexe
2. Il n'existe pas de sous-graphe G' tel que G' est connexe et G'_i est un sous-graphe de G'

On appelle plus grande composante connexe d'un graphe G non connexe la composante connexe constituée du plus grand nombre de sommets. Elle n'est pas nécessairement unique.

1.3 Caractérisation des graphes

La définition d'un graphe est très générale et les objets y répondant peuvent être très différents les uns des autres. Les définitions que nous avons données nous permettent déjà de distinguer plusieurs classes de graphes : orientés ou non, connexes ou non. Ces distinctions sont intéressantes mais pas assez fines pour caractériser correctement les différents graphes auxquels nous allons être confrontés. Nous allons

présenter ici des mesures portant sur les graphes qui vont nous apporter la finesse nécessaire.

Définition 4.6 (Degré d'un sommet) *On appelle degré du sommet s_i , et on note k_i , le nombre de voisins du sommet s_i . On notera $\langle k \rangle_G$ la moyenne des degrés des sommets d'un graphe G .*

Nous nous intéresserons plutôt à la distribution des degrés d'un graphe G , c'est-à-dire la probabilité $P_G(k)$ qu'un de ses nœuds soit de degré k .

Définition 4.7 (Longueur caractéristique et diamètre) *Soit un graphe G . On nomme longueur caractéristique et on note ℓ_G la moyenne des plus courts chemins entre toutes les paires de sommets de G . On appelle diamètre et l'on note D_G le plus grand des plus courts chemins que l'on peut trouver dans G .*

Un graphe ayant un petite longueur caractéristique est un graphe dans lequel on peut rapidement passer de proche en proche d'un nœud à un autre.

Définition 4.8 (Densité) *On nomme densité du graphe $G = (S, A)$ la grandeur*

$$d_G = \frac{2 \cdot |A|}{|S| \cdot (|S| - 1)}$$

La densité d'un graphe est le rapport entre son nombre d'arêtes et le nombre d'arêtes qu'aurait un graphe complètement interconnecté ayant le même nombre de sommets.

Définition 4.9 (Coefficient de regroupement) *Soient s un sommet d'un graphe G , k_s son nombre de voisins et n le nombre d'arêtes entre les k_s voisins de s . La grandeur $C_s = \frac{2 \cdot n}{k_s \cdot (k_s - 1)}$ est nommée le coefficient de regroupement du sommet s .*

On note C_G , coefficient de regroupement du graphe G , la moyenne des coefficients de regroupement de tous les sommets de G

Alors que d est une mesure globale de la densité d'un graphe, le coefficient de regroupement est une mesure locale. Elle indique la tendance qu'ont les sommets d'un graphe à former des amas fortement interconnectés.

Les graphes sont des outils puissants de modélisation des réseaux. Encore faut-il pouvoir construire un graphe aux propriétés concordantes avec le phénomène observé. Euler a, sans aucun doute, un mérite immense à avoir su forger la notion de graphe pour représenter la ville de Königsberg et ses ponts. Cependant, il est trivial de construire ce graphe car il est de taille très modeste. Mais pour des réseaux plus complexes, tels que le réseau d'acointances des États-Unis étudié par Milgram, la construction d'un graphe exact relève de l'impossible : Milgram aurait par exemple eu à interroger les 200 millions d'individus qui peuplaient alors le pays sur l'identité des quelques dizaines de concitoyens qu'ils comptaient dans leurs relations. Pour contourner ce problème, la démarche adoptée est de construire un graphe non pas représentant exactement le réseau que l'on cherche à étudier, mais partageant avec lui suffisamment de propriétés. Le modèle longtemps cité en référence pour la construction de ces graphes est celui d'Erdős et Rényi que nous allons à présent examiner.

1.4 Le modèle de graphes aléatoire d'Erdős et Rényi

Lorsque l'on a à faire à un graphe particulièrement complexe dont la structure nous est inconnue, celui-ci semble aléatoire. Aussi, c'est par des graphes aléatoires que l'on a longtemps modélisé les réseaux complexes observables dans la nature. L'étude des graphes aléatoires est née avec les travaux de Erdős & Rényi (1959, 1960).

Dans le modèle d'Erdős et Rényi, partant d'un nombre N de sommets et d'une probabilité p , on construit un graphe en créant chacune des $\frac{N(N-1)}{2}$ arêtes possibles avec une probabilité p (Figure 4.4).

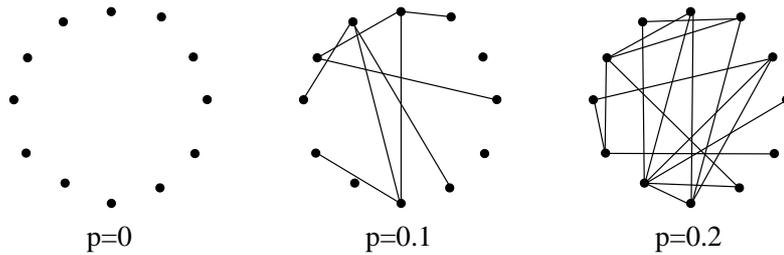


Figure 4.4 – Graphes aléatoires construits selon le modèle d'Erdős et Rényi avec $N = 12$ et $p \in \{0, 0.1; 0.2\}$

Un tel graphe aléatoire G a donc en moyenne $n = p \frac{N(N-1)}{2}$ arêtes, d'où :
 $d_G \simeq p$.

1.4.1 Distribution de degrés

Connaissant le nombre N de sommets et le nombre n d'arêtes d'un graphe, on peut déduire le degré moyen des sommets du graphe :

$$\langle k \rangle = \frac{2n}{N} = p(N-1) \simeq pN \quad (4.1)$$

D'après cette relation, un graphe aléatoire peut aussi bien être donné par les paramètres N et p que par les paramètres N et $\langle k \rangle$.

Calculons la probabilité qu'un sommet donné s_i ait comme degré k . Pour cela, considérons le mécanisme de construction du graphe du point de vue de s_i . Pour chacune des $N-1$ arêtes potentielles de s_i , cette arête sera créée avec la probabilité p . La création d'une arête étant indépendante de celle des autres, le nombre k_i de

voisins du sommet s_i est donné par une loi binomiale :

$$P(k_i = k) = \binom{N-1}{k} p^k (p-1)^{N-1-k} \quad (4.2)$$

La limite d'une distribution binomiale lorsque N tend vers l'infini et p vers 0 est une distribution de Poisson. Pour N suffisamment grand, on a donc (Figure 4.5) :

$$P(k_i = k) = \langle k \rangle^k \cdot \frac{e^{-\langle k \rangle}}{k!} \quad (4.3)$$

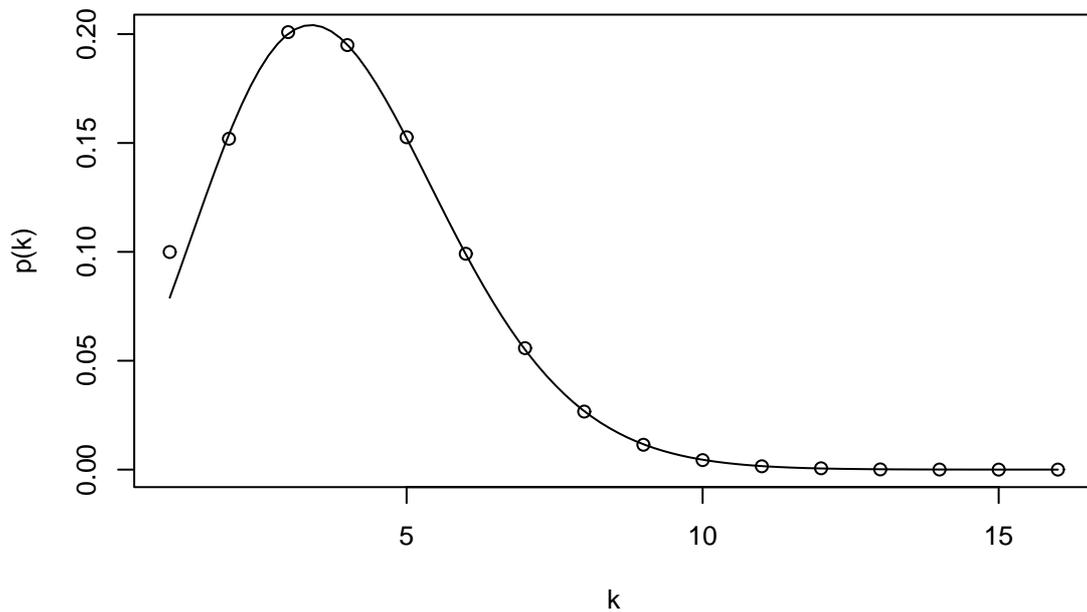


Figure 4.5 – Distribution de degrés d'un graphe aléatoire avec $N=70\,975$ et $\langle k \rangle = 3.9$ (ronds) et distribution théorique avec une loi de Poisson de paramètre 3.9 (ligne continue).

1.4.2 Coefficient de regroupement

Le coefficient de regroupement indique la tendance qu'ont les voisins d'un sommet à être voisins entre eux. Étant donné un sommet quelconque d'un graphe aléatoire, deux de ses voisins seront eux-mêmes voisins avec la même probabilité que n'importe quelle autre paire de sommets du graphe, c'est-à-dire p . On a donc :

$$C_{aléa} = p = \frac{\langle k \rangle}{N} \quad (4.4)$$

1.4.3 Longueur caractéristique

De nombreux auteurs ont étudié le diamètre des graphes aléatoires pour différentes valeurs de N et p . Chung & Lu (2001) reprennent les principaux résultats et les complètent. Ils montrent que si $\langle k \rangle \rightarrow +\infty$ quand $N \rightarrow +\infty$ alors le diamètre D est très probablement égal à $\ln(N)/\ln(\langle k \rangle)$. L'idée est que si le graphe est suffisamment peu dense, en partant d'un sommet donné, on a $\langle k \rangle^l$ sommets à une distance l . Pour atteindre tous les sommets du graphe en partant de ce sommet, il faut donc parcourir une distance l telle que $\langle k \rangle^l = N$, soit $l = \ln(N)/\ln(\langle k \rangle)$.

La longueur caractéristique est du même ordre de grandeur que le diamètre. On a donc :

$$\ell_{aléa} \sim \frac{\ln(N)}{\ln(\langle k \rangle)} \quad (4.5)$$

1.4.4 Les graphes aléatoires comme modèle des réseaux naturels

Le modèle d'Erdős et Rényi de graphes aléatoires a longtemps été utilisé comme modèle des réseaux naturels complexes. Néanmoins, à la fin des années 90, deux

articles (Watts & Strogatz, 1998 ; Barabási & Albert, 1999) ont montré, en étudiant les grandeurs définies ci-dessus, que ces réseaux naturels divergeaient grandement des graphes aléatoires. La section suivante expose quelques exemples de réseaux naturels complexes afin de mettre en lumière ces différences. La connaissance de ces exemples nous confèrera le recul nécessaire pour considérer correctement les réseaux lexicaux que nous étudierons au chapitre suivant.

2 Réseaux complexes naturels

2.1 Nombre d'Erdős

Erdős fut un des mathématiciens les plus surprenants mais aussi les plus prolifiques du 20^e siècle, avec quelques 1 500 articles écrits avec 507 collaborateurs différents. L'influence du personnage est telle que les mathématiciens s'attribuent ludiquement un *nombre d'Erdős*, défini de la façon suivante : Erdős lui-même a comme nombre d'Erdős 0. Les 507 mathématiciens ayant cosigné avec Erdős ont comme nombre d'Erdős 1. Tout mathématicien ayant cosigné un article avec l'un des 507 a comme nombre d'Erdős 2. D'une manière générale, le nombre d'Erdős d'un mathématicien est défini par récurrence comme étant le plus petit nombre d'Erdős parmi l'ensemble de ses collaborateurs passés, augmenté de 1.

On peut voir ce calcul d'une autre façon : les mathématiciens forment un réseau dont ils sont les nœuds. Dans ce réseau, il existe une arête entre deux mathématiciens s'ils ont cosigné ensemble. Ainsi le numéro d'Erdős d'un mathématicien est la longueur du plus court chemin qui le relie à Erdős dans ce graphe de collaboration. Existe-t-il des mathématiciens avec un grand numéro d'Erdős ? Quelles sont plus généralement les caractéristiques de ce réseau ? Est-il modélisable par un graphe aléatoire ? Dans les sections suivantes nous allons répondre à ces questions, pour ce réseau et un certain nombre d'autres : nous considérerons des réseaux sociaux, des réseaux technologiques ou encore biologiques. En particulier, nous détaillerons

nos propres résultats sur le réseau fonctionnel cérébral. Nous verrons que tous ces réseaux ont des caractéristiques communes, différentes de celles des graphes aléatoires.

2.2 Réseaux sociaux

Les réseaux sociaux sont des réseaux qui représentent des relations entre individus. Au début de ce chapitre, nous avons évoqué le graphe d'accointances des États-Unis dans lequel deux individus sont reliés entre eux s'ils se connaissent. Ce type de réseau social est ce que l'on peut imaginer de plus général. D'autres réseaux, représentant d'autres types de relations, ont été étudiés. Nous décrirons ici des réseaux de collaborations scientifiques, tels que celui qui conduit au calcul du nombre d'Erdős, et un réseau de collaborations cinématographiques.

2.2.1 Collaborations scientifiques

Il existe de nombreuses bases de données recensant les publications scientifiques, rendant facilement disponibles les données nécessaires à l'analyse des réseaux de collaborations. Aussi un certain nombre de champs scientifiques ont été étudiés de cette façon. Les collaborations mathématiques ont été étudiées par Barabási et al. (2002). Ces auteurs se sont concentrés sur des articles publiés entre 1991 et 1998. Le réseau qu'ils ont construit était constitué de 70 975 nœuds (mathématiciens). Le degré moyen des nœuds était $\langle k \rangle_{math} = 3.5$. La longueur caractéristique et le coefficient de regroupement mesurés étaient respectivement $\ell_{math} = 9.5$ et $C_{math} = 0.59$. Pour comparaison, la longueur caractéristique et le coefficient de regroupement d'un graphe aléatoire ayant le même nombre de nœuds et le même

degré moyen sont : $\ell_{aléa} = 8.2$ et $C_{aléa} = 5.5 \cdot 10^{-5}$. On a donc $\ell_{math} \simeq \ell_{aléa}$ et $C_{math} \gg C_{aléa}$. Par ailleurs, la distribution de degrés, c'est-à-dire du nombre de collaborateurs par auteur, diffère de celle d'un graphe aléatoire. Elle suit une loi de puissance (voir Figure 4.7) :

$$P(k) \sim k^{-\gamma} \quad (4.6)$$

L'observation d'une telle distribution nous indique qu'il y a beaucoup d'auteurs avec peu de collaborateurs, peu d'auteurs avec beaucoup de collaborateurs (Erdős est donc bien une exception) et surtout qu'il n'y a pas de nombre typique de collaborateurs. Ici, $\gamma_{math} = 2.5$.

Dans le même article, Barabási et al. (2002) étudient une autre base de données, concernant des publications en neurosciences. Le réseau construit était constitué de 209 293 nœuds et avait pour caractéristiques : $\langle k \rangle_{neuro} = 11.5$, $\ell_{neuro} = 6$ et $C_{neuro} = 0.76$. Si l'on compare ces valeurs avec celle d'un graphe aléatoire de même taille et de même densité, on obtient : $\ell_{neuro} \simeq \ell_{aléa} = 5.01$ et $C_{neuro} \gg C_{aléa} = 5.5 \cdot 10^{-5}$. De plus, la distribution de degrés de ce réseau suit elle aussi une loi de puissance avec $\gamma_{neuro} = 2.1$

Newman (2001b) a aussi étudié des réseaux de collaborations scientifiques, en utilisant 4 bases de données : MEDLINE qui a trait à la recherche biomédicale, Los Alamos e-Print Archive, qui contient des articles de physique théorique, SPIRES qui se concentre sur la physique des hautes énergies et NCSTRL, spécialisée en informatique. Les caractéristiques de ces réseaux sont données dans le Tableau 4.1. Ces 4 réseaux présentent eux aussi la particularité $\ell \simeq \ell_{aléa}$ et $C \gg C_{aléa}$. En revanche les distributions de degrés qu'observe Newman (2001b) diffèrent de celles obser-

vées par Barabási et al. (2002). Si elles suivent une loi de puissance pour les petites valeurs de k , pour les grandes la distribution décroît de manière exponentielle, soit plus rapidement qu’une loi de puissance (voir Figure 4.7). La forme générique de ces distributions en loi de puissance avec coupure exponentielle est :

$$P(k) \sim k^{-\gamma} e^{-\frac{k}{k_c}} \quad (4.7)$$

2.2.2 Collaborations cinématographiques

Si les mathématiciens s’attribuent des nombres d’Erdős, les acteurs hollywoodiens s’attribuent sur le même principe un nombre de Bacon. Le principe est identique, l’acteur Kevin Bacon jouant le rôle central (et ayant donc comme nombre de Bacon 0), et l’apparition de 2 acteurs dans un même film jouant le rôle de la

Réseaux	N	$\langle k \rangle$	ℓ	$\ell_{aléa}$	C	$C_{aléa}$	γ	k_c
MEDLINE	1 388 989	14.8	4.4	5.24	0.072	$1.06 \cdot 10^{-5}$	2.5	7300
Los Alamos	52 909	9.7	4.8	5.06	0.43	$1.8 \cdot 10^{-4}$	1.3	52.9
SPIRES	56 627	173	4.0	5.15	0.726	0.003	1.03	1200
NCSTRL	10 998	3.59	9.7	7.28	0.496	$3.26 \cdot 10^{-4}$	1.3	10.7

Tableau 4.1 – Nombres de nœuds (N), degrés moyens ($\langle k \rangle$), distances caractéristiques (ℓ), coefficients de regroupement (C) et exposants des lois de puissance (γ) suivies par les distributions de degrés des réseaux extraits par Newman (2001b) à partir des bases de données MEDLINE, Los Alamos e-Print Archive, SPIRES et NCSTRL. A titre de comparaison nous avons indiqué les distances caractéristiques ($\ell_{aléa}$) et coefficients de regroupement ($C_{aléa}$) de graphes aléatoires de taille et de degré moyen identiques.

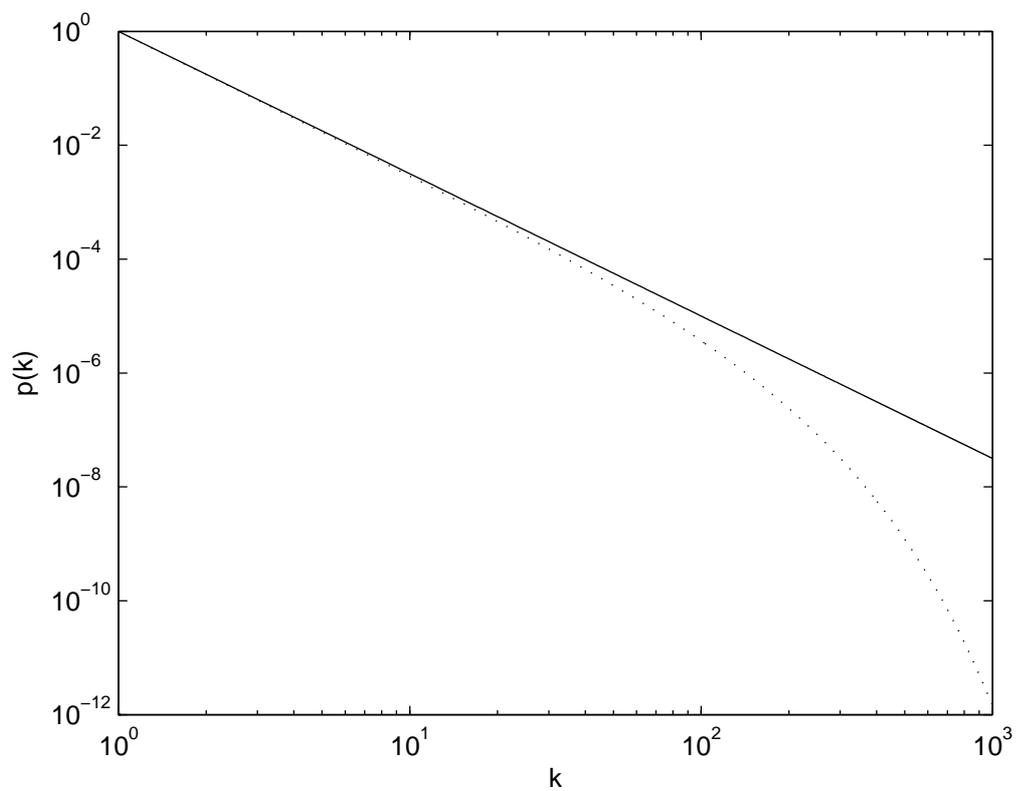


Figure 4.6 – Loi de puissance (ligne continue) avec $\gamma = 2.5$, et loi de puissance avec coupure exponentielle (ligne pointillée), avec $\gamma = 2.5$ et $k_c = 100$. Les axes du graphique ont des échelles logarithmiques. La loi de puissance apparaît comme une droite. La loi de puissance avec coupure exponentielle décroît plus rapidement.

collaboration scientifique ¹. Ce réseau a été étudié par Watts & Strogatz (1998) qui ont déterminé que ce réseau vérifiait $\ell \simeq \ell_{aléa}$ et $C \gg C_{aléa}$, et par Barabási & Albert (1999) qui ont montré que la distribution de degrés suivait une loi de Puissance. Amaral, Scala, Barthélémy, & Stanley (2000) ont néanmoins tempéré les affirmations de Barabási & Albert (1999) en montrant que cette distribution avait une coupure exponentielle.

2.3 Réseaux technologiques

Les réseaux sociaux ne sont pas les seuls réseaux que l'on peut observer. Cette section s'attache à décrire des réseaux de communication et de transport d'énergie.

2.3.1 Web

Le Web est constitué de pages hypertextes ayant la fabuleuse propriété de pouvoir être liées entre elles. Le Web est donc un réseau dans lequel les nœuds sont des pages et les arêtes les liens hypertextes entre ces pages. C'est même le plus grand réseau actuellement observable : le nombre de nœuds a dépassé le milliard à la fin de l'année 1999 et on évalue à 11.5 milliards le nombre de pages indexées par les moteurs de recherche début 2005 (Gulli & Signorini, 2005). De par sa taille, mais aussi et surtout de par son aspect très dynamique (de nouvelles pages apparaissent à chaque instant, tandis que d'autres disparaissent), il est quasi impossible de saisir ce réseau dans son ensemble. Les résultats portant sur le Web sont donc souvent des extrapolations à partir d'études portant sur une fraction de celui-ci.

¹Paul Erdős, apparaissant dans un documentaire sur sa vie, *N is a number*, dans lequel apparaissent également des acteurs de métier, a un nombre de Bacon : 4. En revanche, Kevin Bacon n'ayant jamais publié le moindre article mathématique, n'a pas nombre d'Erdős.

A la différence des réseaux que nous avons présentés jusqu'à présent, le graphe du Web est un graphe orienté : ce n'est pas parce qu'une page pointe vers une autre que la réciproque est vraie. Cette spécificité implique une redéfinition de la notion de degré. Dans un graphe orienté, le degré d'un sommet est constitué de 2 composantes qui doivent être considérées séparément : le degré entrant, k^{out} , qui correspond au nombre d'arcs dont le sommet est le sommet initial, et le degré sortant, k^{in} , qui correspond au nombre d'arcs dont le sommet est le sommet final. Les notions de degré moyen et de distribution de degrés sont étendues de la même façon.

Plusieurs études se sont penchées sur la distribution de degrés du Web, et toutes concluent que ses deux composantes suivent des lois de puissance : $P(k^{in}) \sim k^{-\gamma_{web}^{in}}$ et $P(k^{out}) \sim k^{-\gamma_{web}^{out}}$.

Albert, Barbasi, & Jeong (1999), en se basant sur un sous-graphe du Web de 325 729 pages ont mesuré $\gamma_{web}^{in} = 2.1$ et $\gamma_{web}^{out} = 2.45$. Kumar, Raghavan, Rajalopagan, & Tomkins (1999), à partir d'un échantillon de 40 millions de pages recueilli en 1997 ont obtenu $\gamma_{web}^{in} = 2.1$ et $\gamma_{web}^{out} = 2.38$. Broder et al. (2000) ont utilisé un sous-graphe de 200 millions de pages et ont déterminé $\gamma_{web}^{in} = 2.1$ et $\gamma_{web}^{out} = 2.72$.

Une vision alternative du Web a été adoptée par Adamic (1999) : plutôt que de considérer un graphe où chaque sommet représente une page, elle s'est placée au niveau des sites. Chaque sommet de ce graphe représente donc un site, et il existe un arc entre 2 sommets si l'une des pages du site correspondant au sommet initial pointe vers l'une des pages du site correspondant au sommet final de l'arc. Avec un graphe de 260 000 sommets, elle a obtenu $\gamma^{in} = 1.94$.

Le caractère orienté du graphe du Web ne permet pas de calculer le coefficient de regroupement tel que nous l'avons défini ci-dessus. Pour dépasser ce problème, Adamic (1999) a symétrisé un graphe de 153 127 sites et a calculé $C_{Web} = 0.1078$,

qui est très supérieur au coefficient de regroupement d'un graphe aléatoire de dimension et de densité similaires : $C_{aléa} = 2.3 \cdot 10^{-5}$.

La longueur caractéristique du Web, c'est-à-dire le nombre moyen de clics de souris nécessaires pour passer d'une page quelconque du Web à une autre, a été mesurée pour la première fois par Albert et al. (1999). Ils ont mesuré $\ell_{nd.edu} = 11.2$ pour leur sous-graphe de $N = 325\,729$ nœuds et ont extrapolé $\ell_{Web} \simeq 19$. Broder et al. (2000) ont mesuré $\ell_{Web} = 16$ sur leur échantillon de 200 millions de pages. Au niveau des sites, Adamic (1999) a mesuré $\ell_{sites} = 3.1$. Toutes ces mesures proposent des valeurs du même ordre de grandeur qu'un graphe aléatoire de dimension et de densité équivalentes.

2.3.2 Internet

Si le Web est un réseau de pages ou de sites, les ordinateurs sur lesquels reposent ce réseau forment eux aussi un réseau, le réseau Internet proprement dit. Il n'y a aucune raison de penser que ces deux réseaux, le Web et Internet, sont corrélés : ce n'est pas parce qu'une page pointe vers une autre que l'ordinateur sur lequel se trouve la première est relié directement à celui où se trouve la seconde. D'ailleurs, si le Web est un graphe orienté, il n'en est rien d'Internet : si un ordinateur peut communiquer avec un autre, alors la réciproque est vraie.

Comme son nom l'indique, Internet est en fait la mise en réunion de réseaux informatiques, nommés domaines. Donc, de la même manière que l'on peut observer le Web au niveau des pages ou au niveau des sites, il est possible de considérer Internet soit au niveau des ordinateurs qui le constituent, soit au niveau des domaines. Faloutsos, Faloutsos, & Faloutsos (1999) ont étudié Internet à ces deux niveaux. Dans

les deux cas, la distribution des degrés suit une loi de puissance avec $\gamma_{ordi} = 2.2$ et $\gamma_{domaines} = 2.48$.

Yook, Jeong, & Barabási (2002) ont étudié le coefficient de regroupement et la longueur caractéristique au niveau des domaines, et ont déterminé qu'en novembre 1999 on avait : $\ell_{domaines} = 3.77$ ($\ell_{aléa} = 6.18$) et $C_{domaines} = 0.3$ ($C_{aléa} = 0.001$).

2.3.3 Réseaux électriques

L'information n'est pas seule à voyager le long de fils électriques. C'est aussi le cas de l'énergie. Les réseaux que constituent les centrales électriques, les transformateurs et autres postes électriques reliés entre eux par les lignes électriques ont été étudiés. Watts & Strogatz (1998) se sont penchés sur le réseau électrique de l'ouest des États-Unis, constitué de $N = 4\,941$ nœuds, ayant en moyenne $\langle k \rangle_{elec} = 2.67$ voisins. Ils ont calculé que $\ell_{elec} = 18.7$ et $C_{elec} = 0.08$. Pour un graphe aléatoire avec le même nombre de nœuds et le même nombre d'arêtes, on a : $\ell_{aléa} = 12.4$ et $C_{aléa} = 0.005$. La distribution de degrés suit une loi de puissance selon Barabási & Albert (1999) ou une loi de puissance avec coupure exponentielle selon Amaral et al. (2000).

2.4 Réseaux biologiques

2.4.1 Réseaux métaboliques

Le métabolisme d'un organisme est l'ensemble des réactions chimiques qui s'effectuent en son sein afin de produire la matière et l'énergie qui lui sont nécessaires pour vivre. Chacune de ces réactions chimiques est coordonnée par une enzyme

qui transforme une ou plusieurs molécules dites *substrats* en nouvelles molécules dites *produits*. Le produit d'une réaction peut être le substrat d'une autre. Certaines réactions sont réversibles, d'autres non. Sur la base de ces considérations, il est naturel de concevoir le métabolisme d'un organisme comme un réseau dont les nœuds sont les composants qui participent aux réactions et dont les liens sont les réactions elles-mêmes (Barabási & Oltvai, 2004).

Jeong, Tombor, Albert, Oltvai, & Barabási (2000) ont étudié le réseau métabolique de 43 organismes de trois domaines de la vie différents (6 archaea, 32 bactéries et 5 eukaryotes). Ces réseaux étaient orientés et constitués $N = 430 (\pm 130)$ nœuds. Les distributions de degrés entrants et sortants de tous les réseaux suivent des lois de puissance semblable : $\gamma^{in} = 2.14 (\pm 0.07)$ et $\gamma^{out} = 2.22 (\pm 0.05)$.

A. Wagner & Fell (2001) ont considéré uniquement la bactérie *Escherichia coli*. Ils n'ont pas tenu compte de la possible réversibilité des réactions et ont donc obtenu un réseau non orienté. Celui-ci est constitué de $N = 275$ nœuds. La distribution de leur degré suit de nouveau une loi de puissance ($\gamma_{E. coli} = 1.77$). Ces auteurs ont aussi calculé le coefficient de regroupement ($C_{E. coli} = 0.48 \gg C_{aléa} = 0.017$) et la longueur caractéristique ($\ell_{E. coli} = 3.88 \simeq \ell_{aléa} = 3.78$).

2.4.2 Réseaux écologiques

Les espèces qui cohabitent au sein d'un écosystème se nourrissent les unes des autres, constituant ainsi un réseau alimentaire, propre à chaque écosystème, dans lequel les nœuds sont les espèces et où un lien entre 2 nœuds indique une relation proie / prédateur entre les espèces correspondantes.

Solé & Montoya (2002) ont étudié les réseaux alimentaires de trois des écosystèmes les mieux documentés : l'estuaire d'Ythan (en construisant 2 réseaux, à partir de 2 sources de données différentes ; $N_{Ythan_1} = 93$ et $N_{Ythan_2} = 134$), le parc de Silwood ($N_{Silwood} = 154$) et le lac de Little Rock ($N_{Little\ Rock} = 182$). Pour ce dernier écosystème, du fait des données, seulement 31% des nœuds du graphe construit par Solé & Montoya (2002) représentent des espèces. Les autres nœuds représentent des genres ou des niveaux encore plus élevés dans la hiérarchie taxonomique.

Tous ces réseaux présentent un coefficient de regroupement supérieur à celui d'un graphe aléatoire de même taille et de même densité :

- $C_{Ythan_1} = 0.21 > C_{aléa} = 0.09$;
- $C_{Ythan_2} = 0.22 > C_{aléa} = 0.06$;
- $C_{Silwood} = 0.15 > C_{aléa} = 0.03$;
- $C_{Little\ Rock} = 0.35 > C_{aléa} = 0.14$.

En revanche, leur longueur caractéristique est comparable à celle qu'aurait un graphe aléatoire de même taille et de même densité :

- $\ell_{Ythan_1} = 2.28 \simeq \ell_{aléa} = 2.09$;
- $\ell_{Ythan_2} = 2.43 \simeq \ell_{aléa} = 2.26$;
- $\ell_{Silwood} = 3.4 \simeq \ell_{aléa} = 3.23$;
- $\ell_{Little\ Rock} = 2.22 \simeq \ell_{aléa} = 1.6$.

Les réseaux alimentaires de l'estuaire d'Ythan et du parc de Silwood ont une distribution de degrés qui suit une loi de puissance :

- $\gamma_{Ythan_1} = 0.94$
- $\gamma_{Ythan_2} = 1.04$
- $\gamma_{silwood} = 1.13$

La distribution de degré du réseau alimentaire de lac de Little Rock ne ressemble

en revanche à aucune distribution particulière. Les auteurs interprètent ce fait par le grand nombre de nœuds de ce réseau ne correspondant pas à une espèce.

2.4.3 Réseaux neuronaux

Caenorhabditis elegans est un petit ver qui, entre autre du fait de sa totale transparence, a été activement étudié. *C. elegans* dispose d'un petit réseau de 302 neurones dont on connaît complètement la connectivité (Yamamoto & Achacoso, 1991). L'information neuronale étant, au niveau synaptique, transférée d'un neurone à l'autre de manière unidirectionnelle, ce réseau est orienté.

Watts & Strogatz (1998) en ont étudié la longueur caractéristique et le coefficient de regroupement. Ils ont déterminé $C_{C. elegans} = 0.8 \gg C_{aléa} = 0.05$ et $\ell_{C. elegans} = 2.65 \simeq \ell_{aléa} = 2.25$. Les distributions des degrés entrants et sortants ont été calculées par Amaral et al. (2000). Ces deux distributions sont presque identiques et suivent non pas une loi de puissance, mais une décroissance exponentielle :

$$P(k) \sim e^{-\beta k} \quad (4.8)$$

Lorsque l'on considère des espèces au système nerveux plus élaboré que *C. elegans*, il devient quasi impossible de recueillir des données sur la connectivité des neurones eux-mêmes, non seulement du fait de leur nombre ($6 \cdot 10^8$ chez le chat, $3 \cdot 10^9$ chez le macaque et $2 \cdot 10^9$ chez l'homme), mais surtout du fait du nombre de connexions entre ces neurones ($4.5 \cdot 12^8$ chez le chat, $2.2 \cdot 10^{13}$ chez le macaque et $1.5 \cdot 10^{14}$ chez l'homme ; Johansson & Lansner, 2004). Néanmoins, l'accroissement de la complexité des systèmes nerveux d'une espèce à l'autre ne se traduit pas uniquement en terme de nombre de neurones et de nombre de synapses. L'or-

ganisation structurelle reflète aussi cette complexité. La neuroanatomie décompose en effet les cerveaux en aires cérébrales et identifie des connexions entre ces aires cérébrales. Des réseaux anatomiques d'aires cérébrales ont ainsi été décrits, en particulier chez le chat (Scannell, Blakemore, & Young, 1995 ; Scannell, Burns, Hilgetag, O'Neil, & Young, 1999) et chez le macaque (Felleman & Van Essen, 1991 ; Young, 1993). Les propriétés de ces réseaux (orientés) ont été étudiées par Sporns & Zwi (2004 ; Sporns, Chialvo, Kaiser, & Hilgetag, 2004). Ces auteurs se sont plus particulièrement penchés sur le réseau du cortex visuel du macaque (CVM, 30 aires et 311 connexions), le réseau cortical du macaque (CM, 71 aires et 746 connexions) et le réseau cortical du chat (CC, 52 aires et 820 connexions). Ils ont déterminé $C_{CVM} = 0.5313$, $C_{CM} = 0.4614$, $C_{CC} = 0.5514$ (à chaque fois supérieur au coefficient de regroupement d'un graphe aléatoire comparable) et $\ell_{CVM} = 1.7256$, $\ell_{CM} = 2.3769$, $\ell_{CC} = 1.8114$ (à chaque fois comparable à la longueur caractéristique d'un graphe aléatoire comparable). Sporns & Zwi (2004) n'observent en revanche pas de loi de puissance dans les distributions de degrés de ces réseaux, sans toutefois fournir plus de détails sur ces distributions.

2.5 Caractéristiques des grands réseaux naturels

Dans la section précédente, nous avons passé en revue les propriétés de réseaux complexes provenant de domaines variés, de la biologie moléculaire à Internet, en passant par les collaborations scientifiques et artistiques et le transport d'énergie. De prime abord, on aurait pu s'attendre à ce que ces réseaux, d'origines aussi différentes, s'ils avaient eu quelque chose à partager, auraient partagé leur caractère quelconque, aléatoire. Au contraire, il se dégage, au regard de leurs propriétés, que

ce qui rend tous ces réseaux similaires est ce qui les démarque du modèle de graphe aléatoire d'Erdős et Rényi. Si leur longueur caractéristique est toujours du même ordre que celle d'un graphe aléatoire ayant le même nombre de nœuds et le même nombre d'arêtes, leur coefficient de regroupement est toujours supérieur (voire très largement supérieur), à celui d'un graphe aléatoire. De plus, dans la plupart des cas, les réseaux naturels ont une distribution de degrés suivant une loi de puissance, avec cependant, parfois, une coupure exponentielle. Il arrive aussi que ces distributions suivent des lois exponentielles (comme le réseau neuronal de *C. elegans*) ou encore d'autres types de distributions (Amaral et al., 2000). Mais jamais la distribution de leurs degrés ne répond à la loi de Poisson à laquelle obéissent les distributions de degrés des graphes aléatoires.

Ce sont Watts & Strogatz (1998) qui ont les premiers souligné que les réseaux complexes naturels ont comme particularité de posséder simultanément une petite longueur caractéristique et un grand coefficient de regroupement. Ils ont introduit le terme de *petit monde* (*small world*)¹ pour désigner les réseaux possédant ces deux propriétés. Quant à l'observation de la quasi-universalité de la loi de puissance dans la distribution de degrés des réseaux naturels, elle est due à Barabási & Albert (1999), qui ont introduit le terme de *réseaux sans échelle* pour les désigner.

Ces deux articles concluent à l'inadéquation du modèle de graphe aléatoire d'Erdős et Rényi pour modéliser les réseaux complexes naturels, et proposent des

¹Le terme de *petit monde* fait référence à l'article de Milgram (1967), et au fait que "le monde est petit", c'est-à-dire qu'il est relativement fréquent de rencontrer une personne avec laquelle, de manière inattendue, on partage une connaissance. Si "le monde est petit", c'est parce que, comme l'a mis en lumière Milgram, les chaînes d'accointances sont courtes. Ainsi, peu de connaissances nous séparent d'un individu quelconque, et lorsque que l'on rencontre quelqu'un, la probabilité d'avoir une connaissance en commun, c'est-à-dire la probabilité de n'être séparé que par une chaîne d'accointances de longueur 2, n'est pas négligeable.

Il est peut-être néanmoins malheureux d'avoir nommé ces réseaux du nom de la seule propriété qu'ils partagent avec les graphes aléatoires.

2.5 Caractéristiques des grands réseaux naturels

modèles alternatifs rendant compte pour le premier des hauts coefficients de regroupement, et pour le second de la distribution de degrés en loi de puissance. La section suivante présente ces deux modèles.

3 Graphes "petits mondes" et graphes "sans échelle"

3.1 Graphes petits mondes

Watts & Strogatz (1998) sont les premiers à avoir démontré que le modèle de graphe aléatoire d'Erdős et Rényi ne rend pas compte des propriétés observées dans les réseaux naturels. Leur argumentaire repose sur la notion de densité locale, qu'ils ont pu quantifier en introduisant la notion de coefficient de regroupement.

Les réseaux naturels sont peu denses : parmi tous les liens possibles (si chaque nœud était relié à tous les autres), peu existent réellement. Par contre, lorsque l'on considère un nœud d'un réseau naturel et l'ensemble de ses voisins, on obtient un sous-réseau dense, c'est-à-dire un sous-réseau dont les nœuds ont fortement tendance à être reliés, ce qui se traduit par un haut coefficient de regroupement. À l'inverse, comme nous l'avons vu à la section 1.4.2, la densité locale des graphes aléatoires est identique à leur densité globale. Les graphes aléatoires sont donc incapables de rendre compte des structures locales denses des réseaux naturels.

Il existe un autre type de graphes qui peuvent quant à eux présenter une forte densité locale : les graphes réguliers. On appelle *graphe régulier* un graphe construit par simple réplication d'un motif de base (figure 4.7). En revanche, ces graphes ont une grande longueur caractéristique.

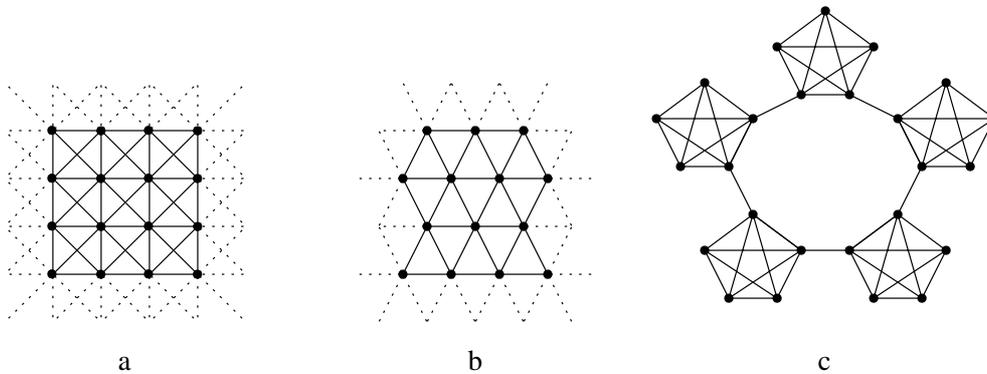


Figure 4.7 – Trois exemples de graphes réguliers. $C_a = 0.43$; $C_b = 0.6$; $C_c = 0.84$

Les réseaux naturels sont donc dotés de la petite longueur caractéristique des graphes aléatoires, et de la haute densité locale des graphes réguliers. L'idée du modèle de Watts & Strogatz (1998) est d'extrapoler entre l'ordre des réseaux réguliers et le désordre des réseaux aléatoires. Entre les deux, résident les petits mondes.

Le graphe régulier sur lequel s'appuie le modèle de Watts & Strogatz (1998) est un anneau de N sommets, où chaque sommet est connecté à ses k plus proches voisins. Ensuite, avec une probabilité p , l'une des extrémités de chaque arête est redirigée vers un sommet quelconque du graphe (figure 4.8).

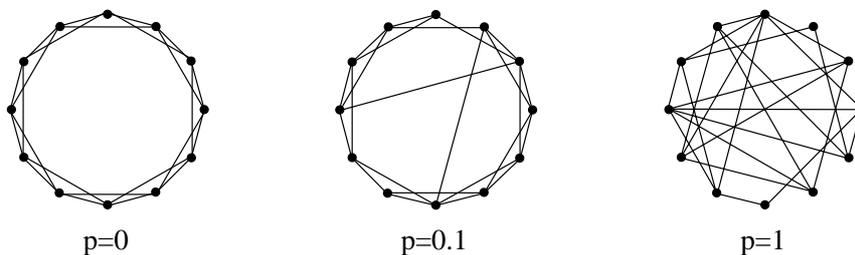


Figure 4.8 – Graphes construits selon le modèle de Watts et Strogatz avec $N = 12$, $k = 4$ et $p \in \{0, 0.1, 1\}$

Un graphe du modèle de Watts & Strogatz (1998) est donc défini par trois para-

mètres : son nombre de sommets N , son degré moyen k , et une probabilité de rebranchement p . Si $p = 0$, le graphe est un anneau régulier où chaque nœud a comme degré k . Son coefficient de regroupement est alors élevé, et sa longueur caractéristique est grande. Si $p = 1$, le graphe est un graphe aléatoire (au sens d'Erdős et Rényi) de degré moyen k . Son coefficient de regroupement est alors faible, et sa longueur caractéristique courte. La figure 4.9 présente l'évolution de ces deux grandeurs en fonction de p . On y observe que rapidement, la longueur caractéristique décroît fortement tandis que le coefficient de regroupement reste à peu près constant. La raison est qu'il suffit de quelques rebranchements pour créer des raccourcis d'une partie à l'autre du graphe, qui réduisent considérablement les distances entre les sommets. Néanmoins, si ces raccourcis entraînent des modifications radicales sur la structure globale du graphe, leur faible nombre modifie à peine sa structure locale. Il reste localement dense, conservant un haut coefficient de corrélation.

Ce modèle permet donc de construire des graphes exhibant des propriétés observées dans les réseaux complexes naturels, courte longueur caractéristique et haut coefficient de regroupement, dépassant ainsi le modèle d'Erdős et Rényi. Néanmoins, la distribution de degrés ainsi construite suit, comme les graphes aléatoires, une distribution de Poisson. Ces graphes ne parviennent donc pas à capturer toute la richesse des réseaux naturels.

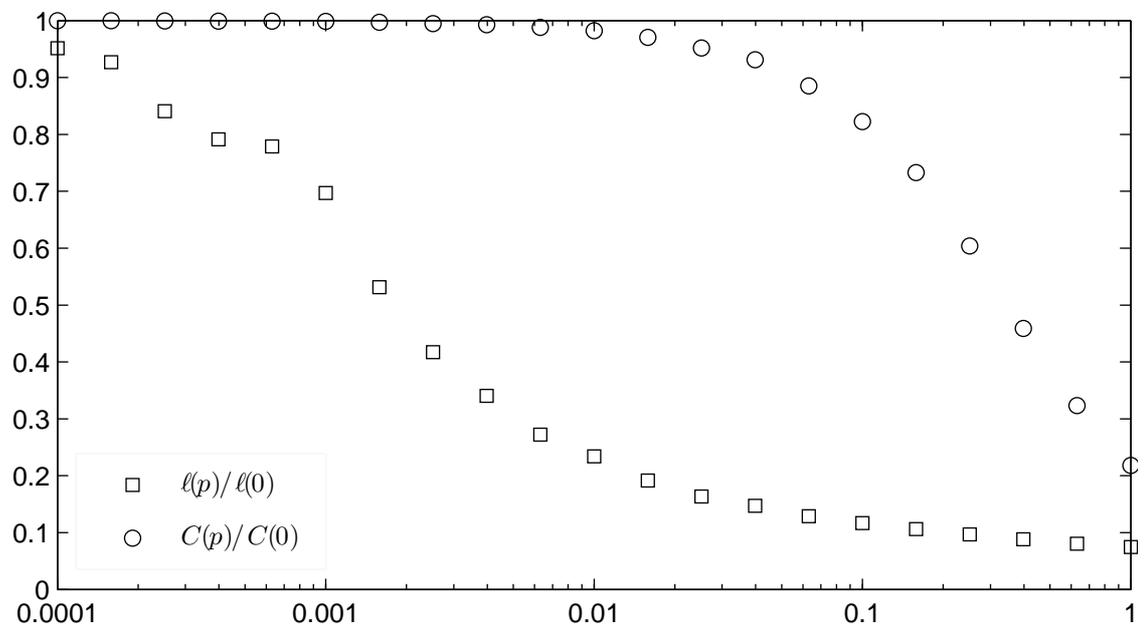


Figure 4.9 – Longueur caractéristique $\ell(p)$ et coefficient de regroupement $C(p)$ en fonction de la probabilité de rebranchement p , normalisés par leur valeur pour $p = 0$. Valeurs moyennes pour 15 graphes avec $N = 500$ et $k = 10$.

3.2 Graphes sans échelle

3.2.1 Le modèle de Barabási et Albert

La distribution de degrés des réseaux naturels en loi de puissance a été remarquée par Barabási & Albert (1999). Les graphes aléatoires comme les graphes petits mondes ont un degré moyen $\langle k \rangle$, caractéristique du graphe, autour duquel se répartissent les degrés des nœuds (voir la figure 4.5 ; la distribution atteint un maximum pour $k = \langle k \rangle$). Au contraire, dans les réseaux naturels, la distribution de degrés en loi de puissance implique qu'il n'y a aucun degré caractéristique du réseau (figure 4.6 ; il n'y a aucun pic dans la distribution). L'absence d'un tel degré caractéristique a conduit Barabási & Albert (1999) à introduire, à l'image de nombreux phénomènes eux aussi dépourvus de grandeur caractéristique, le terme de *réseaux sans échelle* (*scale free networks*).

Pour rendre compte de ce phénomène, Barabási & Albert (1999) ont adopté une perspective pour le moins pertinente : les réseaux naturels, quels qu'ils soient, ne sont pas construits de but en blanc, et les mécanismes qui sous-tendent leur construction ont sans doute des conséquences sur leurs propriétés. Ils ont donc proposé un modèle *constructiviste* de graphes. A la différence du modèle d'Erdős et Rényi et de celui de Watts et Strogatz, dans lesquels on se donne d'emblée un certain nombre de sommets entre lesquels on place des arêtes, dans le modèle de Barabási & Albert (1999) un graphe est construit incrémentalement, en y ajoutant un sommet à chaque pas de temps. Une autre différence de fond entre le modèle des graphes sans échelle et ceux des graphes aléatoires et graphes petits mondes, porte sur la probabilité de relier deux sommets par une arête. Dans les modèles d'Erdős et Rényi et de Watts

et Strogatz, cette probabilité est uniforme, identique pour tous les couples de sommets. Barabási & Albert (1999) proposent au contraire le principe d'*attachement préférentiel*, selon lequel un nouveau nœud entrant dans un réseau a plus de chance de s'associer à des nœuds ayant déjà un grand nombre de voisins : un nouvel acteur a toutes les chances de jouer son premier film avec des acteurs ayant participé à plusieurs autres films, et ayant donc déjà joué avec plusieurs autres acteurs.

Le mécanisme complet de construction est le suivant : initialement, le graphe est constitué de m_0 sommets. Chaque nouveau sommet est inséré dans le graphe en le connectant à $m < m_0$ sommets déjà présents. La probabilité que le sommet s_i , de degré k_i , soit choisi pour être un de ces m sommets est :

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (4.9)$$

Barabási & Albert (1999) ont calculé analytiquement la distribution de degrés des graphes ainsi construits :

$$P(k) = \frac{2m^2}{k^3} \quad (4.10)$$

Cette distribution est bien une loi de puissance, avec $\gamma = 3$.

La longueur caractéristique a aussi été étudiée. Il a été établi que $\ell_{B.A.} \sim \ln(N)/\ln(\ln(N)) < \ell_{aléa}$ (Bollobás & Riordan, 2004). On ne dispose en revanche d'aucune expression analytique du coefficient de regroupement des graphes sans échelle. Albert & Barabási (2002) ont effectué une estimation numérique pour différentes tailles de réseaux vérifiant $\langle k \rangle = 4$. Ils ont évalué que le coefficient de regroupement était alors 5 fois supérieur à celui de graphes aléatoires de même taille et de même degré moyen. Néanmoins ce coefficient décroît avec la taille des graphes ($C \sim N^{0.74}$), tandis qu'il est constant pour les graphes du modèle de Watts et Strogatz.

3.2.2 Élaborations autour du modèle de Barabási et Albert

Le modèle de Barabási et Albert permet de construire des graphes ayant des distributions de degrés en loi de puissance. Les mécanismes de construction incrémentale et d'attachement préférentiel sur lesquels il repose sont des plus séduisants, en tant qu'heuristiques intéressantes pour expliquer la formation de réseaux naturels. Il ne parvient néanmoins pas à rendre compte de la variété des distributions de degrés observées dans les réseaux naturels : comme les nombreux exemples que nous avons donnés l'attestent, ces lois de puissance ont des exposants variés et sont parfois accompagnées de coupure exponentielle. Plusieurs types d'amélioration ont été proposés afin de tenter de reproduire ces phénomènes observés.

Attachement préférentiel

Le modèle de Barabási & Albert (1999) implémente l'idée d'un attachement préférentiel en reliant le sommet nouvellement inséré au sommet i déjà présent dans le graphe, une probabilité $\Pi(k_i)$ proportionnelle au degré k_i du sommet i . L'hypothèse selon laquelle la probabilité $\Pi(k_i)$ dépend linéairement de k_i a pu être testée empiriquement à partir de réseaux en évolution pour lesquels on peut connaître la date à laquelle les nouveaux nœuds sont intégrés.

Newman (2001a), à partir de deux bases de données de collaborations scientifiques, a observé la présence d'attachement préférentiel linéaire pour des valeurs de k inférieures à une valeur critique qui dépend du réseau. Au delà, $\Pi(k)$ semble être constant. Newman note par ailleurs que la valeur critique de k semble être celle où apparaît la coupure exponentielle dans la distribution de degrés.

Pastor-Satorras, Vazquez, & Vespignani (2001) ont étudié l'attachement préférentiel d'Internet, au niveau des ordinateurs, et ont déterminé que $\Pi(k)$ dépend linéairement de k , ce qui conforte ce que nous avons vu à la section 2.3 : la distribution de degrés de ce réseau suit une loi de puissance parfaite.

Jeong, Neda, & Barabasi (2003) ont étudié l'attachement préférentiel dans 4 réseaux : un réseau de collaboration scientifique, un réseau de citations (les nœuds sont des articles, et il y a un lien entre deux articles si l'un cite l'autre), un réseau d'acteurs de cinéma et Internet au niveau des domaines. Ces quatre réseaux ont des distributions de degrés en loi de puissance. Pour Internet et le réseau de citations, les auteurs ont déterminé que $\Pi(k) \propto k$. En revanche, pour les deux autres réseaux, ils concluent que $\Pi(k) \sim k^{-0.8}$. Barabási et al. (2002) ont observé une probabilité identique avec des réseaux de co-citations en mathématiques et en neurosciences.

Krapivsky, Redner, & Leyvraz (2000) ont montré que si l'on a $\Pi(k) \sim k^\alpha$, la distribution de degré suit en loi une puissance uniquement si $\alpha = 1$, c'est-à-dire si $\Pi(k) \sim ak$; l'exposant de la loi de puissance dépend alors de a . Si $\alpha < 1$, $P(k)$ décroît de manière exponentielle. Si $\alpha > 1$, on obtient un réseau en étoile où les sommets n'ont qu'un seul même voisin, connecté à tous les sommets.

Il semble que la linéarité de la probabilité de l'attachement ne soit pas universelle. Dans certains cas, elle paraît suivre elle aussi une loi de puissance, avec un exposant α , $0 < \alpha < 1$. Mais comme l'ont montré Krapivsky et al. (2000), une telle probabilité induit une distribution de degré qui ne suit pas une loi de puissance mais une décroissance exponentielle. On peut noter que l'observation de distribution de degrés en loi de puissance avec une coupure exponentielle peut alors être due à une probabilité d'attachement préférentiel $\Pi(k) \sim k$ pour $k < k_{critique}$ et $\Pi(k) \sim k^\alpha$, $\alpha < 1$ pour $k > k_{critique}$. Néanmoins, Barabási et al. (2002) et Jeong

et al. (2003) ont observé des distributions de degré en loi de puissance sans coupure exponentielle, avec des probabilités $\Pi(k) \sim k^\alpha, \alpha < 1$, semblant contredire les résultats de Krapivsky et al. (2000). La conclusion qui s'impose donc est que la probabilité d'attachement n'est pas le seul facteur influant sur la distribution de degrés.

Modifications internes

Le modèle de Barabási & Albert (1999) ne considère l'évolution des réseaux que par l'ajout de nouveaux nœuds. Or, ce mécanisme est loin d'être le seul observé dans les réseaux naturels. Des extensions ont donc été proposées pour prendre en compte l'ajout, la suppression ou le rebranchement d'arêtes entre les sommets déjà présents dans le graphe.

Albert & Barabasi (2000) ont exploré une version de leur modèle dans laquelle à chaque pas de temps, soit m nouvelles arêtes sont ajoutées avec une probabilité p , soit le rebranchement de m arêtes est effectué avec une probabilité q , soit un nouveau sommet est ajouté et relié à m sommets déjà présents, avec une probabilité $1 - p - q$. Dans les trois cas, le principe de l'attachement préférentiel est appliqué avec $\Pi(k) \propto k$. Ils montrent que, dans ce cas, la distribution de degrés suit une loi de puissance si $q < q_{max}$, avec q_{max} dépendant de p et m . L'exposant de la loi de puissance peut alors varier entre 2 et $+\infty$. Si $q > q_{max}$, la distribution de degré se rapproche d'une distribution exponentielle. Ils ont montré que ce modèle rend compte avec une grande précision de la distribution de degrés du réseau des acteurs, pour lequel l'évolution se fait fréquemment par ajout de liens internes, sans introduction de nouveau nœud (à chaque fois qu'un film réunit des acteurs n'ayant jamais joué ensemble, mais, n'en étant pas à leur première apparition au grand écran).

Contraintes d'évolution

Le modèle de Barabási & Albert (1999) présuppose que le degré d'un sommet peut croître indéfiniment. Manifestement, cette hypothèse ne se vérifie pas dans les réseaux naturels. Dans le réseau des acteurs par exemple, il y a une contrainte d'âge : passé un certain âge, un acteur cesse de jouer, et le nœud lui correspondant dans le réseau ne peut donc plus développer de nouvelles connexions. Ou encore, dans le réseau de transport aérien, on trouve une contrainte de capacité : le nombre de connexions que peut accueillir un aéroport est borné par sa taille. Il en est de même pour les réseaux neuronaux, un neurone ne pouvant établir un nombre illimité de synapses. Amaral et al. (2000) ont étudié l'influence de ces contraintes sur la distribution de degrés, interdisant à tout nœud ayant dépassé un certain âge ou atteint un degré critique de développer de nouvelles connexions. Ils ont observé l'apparition de coupures exponentielles dans les distributions de degrés.

3.3 Modélisation des réseaux naturels

Nous avons vu que les réseaux naturels présentent des caractéristiques récurrentes qui laissent présupposer l'existence d'un paradigme général permettant d'appréhender de manière globale tous ces phénomènes, pourtant d'origines diverses, qui produisent des réseaux. L'incapacité du modèle d'Erdős et Rényi à rendre compte de ces propriétés démontre en effet que tous ces réseaux ne sont pas aléatoires.

Comme le modèle de Barabási & Albert (1999) et ses multiples variantes le suggèrent, la similarité observée entre les réseaux naturels semble provenir de principes sous-jacents communs à tous les processus qui conduisent à la construction de ces

réseaux. Néanmoins, on ne dispose à l'heure actuelle d'aucun modèle permettant de proposer une vision unifiée des réseaux naturels. En particulier, les modèles d'évolution de réseaux ne semblent pas apporter de justification claire à la densité locale et à la distance caractéristique des réseaux naturels. Toutefois, chacune de ces variantes, lorsqu'elle parvient à reproduire la distribution de degrés du réseau qu'elle est censée modéliser, apporte des arguments forts en faveur des principes évolutifs qu'elle implémente.

Pour clore ce chapitre, nous présentons une étude que nous avons réalisée afin de déterminer les propriétés du réseau fonctionnel cérébral.

4 Réseau fonctionnel cérébral

4.1 Réseaux anatomiques et réseaux fonctionnels

Nous avons examiné, à la section 2.4.3, des réseaux neuronaux : le réseau des neurones de *C. elegans*, et des réseaux d'aires cérébrales du chat et du macaque. Ces réseaux sont construits sur des considérations anatomiques : les liens entre les nœuds sont soit des connexions synaptiques pour *C. elegans*, soit des connexions inter-aréales pour le chat et le macaque. Ils représentent, à différentes échelles, la structure des systèmes nerveux de ces organismes. De ce fait, ces réseaux en sont des représentations *statiques*. Or, cette perspective statique, basée sur la connectivité anatomique, ne saurait rendre compte à elle-seule de la richesse du fonctionnement cérébral. Tant au niveau des neurones, dont l'activation dépend de l'information qui leur est transmise par leur semblables, qu'au niveau des aires cérébrales, qui fonctionnent de concert, cette richesse se dévoile dès lors que l'on considère les aspects *dynamiques* de l'activité neuronale ou cérébrale. A toute échelle, les corrélations temporelles d'activation définissent une connectivité fonctionnelle. Bien que la connectivité anatomique contraigne cette connectivité fonctionnelle, et qu'en retour cette dernière puisse façonner la première (par renforcement des poids synaptiques, ou, sur une plus grande échelle temporelle, par l'établissement de nouvelles connexions entre divers sous-systèmes sous l'effet d'une pression évolutive), ces deux connectivités ne sont pas nécessairement identiques : les activités de deux

neurones ou de deux aires peuvent être corrélées en l'absence de lien anatomique direct, par la réception d'informations d'une source commune (Sporns, 2003).

Du fait de la dimension dynamique qu'elle apporte par rapport à la connectivité anatomique, la connectivité fonctionnelle permet de concevoir un réseau cérébral bien plus représentatif du fonctionnement cérébral que les réseaux anatomiques. C'est l'étude des propriétés de ce réseau fonctionnel cérébral que nous nous proposons d'exposer ici.

4.2 Réseaux de voxels

Les réseaux fonctionnels cérébraux sont encore peu étudiés. A notre connaissance, seuls Eguíluz, Chialvo, Cecchi, Baliki, & Apkarian (2005) ont jusqu'à présent proposé une étude quantitative des propriétés de ces réseaux. La raison en est sans doute, qu'à la différence des réseaux anatomiques, il n'existe pas de données directement accessibles sur ces réseaux. Eguíluz et al. (2005) (tout comme nous, mais d'une manière différente, voir section 4.3), ont utilisé des données provenant d'imagerie cérébrale pour inférer ces réseaux.

Les techniques d'imagerie cérébrale telles que la tomographie par émission de positons (TEP) ou l'imagerie par résonance magnétique fonctionnelle (IRMf) procèdent par l'observation des variations locales de débit sanguin pour mesurer l'activité cérébrale. En effet, les neurones voient leur consommation d'énergie augmenter après avoir été excités. Cette énergie leur étant apportée par le sang, il y a, dans les aires cérébrales activées, une augmentation concomitante du débit sanguin. Ces variations de débit sanguin sont mesurées dans des petits cubes, les voxels, dont la taille, variant entre le mm^3 et le cm^3 , est plus dictée par des contraintes techniques

que des considérations anatomiques ou physiologiques.

Pour construire leurs réseaux fonctionnels, Eguíluz et al. (2005) ont utilisé des données d'IRMf acquises en faisant réaliser des tâches par des sujets : 1 sujet qui oppose le pouce et l'index 10 s, puis relâche pendant 10 s ; 7 sujets qui tapotent du doigt, avec 3 différents protocoles pour marquer le début et la fin des périodes de tapotement (instructions verbales, point vert ou rouge sur un écran ou bien écran vert ou rouge) ; un nombre indéterminé de sujets qui écoutent de la musique. Les sommets des réseaux étaient les voxels, et un lien était établi entre deux voxels, si la corrélation temporelle des activités qui y étaient enregistrées, dépassait un seuil fixé r_c . Ils obtenaient ainsi un réseau de voxels par sujet et par tâche.

Bien que l'article de Eguíluz et al. (2005) soit globalement confus et peu rigoureux, il ressort que les réseaux ainsi construits sont des petits mondes : leur coefficient de regroupement est supérieur de plusieurs ordres de grandeur à celui de graphes aléatoires de dimensions comparables, tandis que leur distance caractéristique est équivalente (du moins pour les grandes valeurs de r_c). Eguíluz et al. (2005) prétendent que les réseaux qu'ils ont construits ont des distributions de degrés suivant des lois de puissance, mais la faiblesse de leur argumentation sur ce point est telle qu'il paraît délicat de tenir ce résultat pour acquis.

Indépendamment du manque de rigueur de leur étude, les réseaux construits par Eguíluz et al. (2005) souffrent d'au moins deux limitations qui réduisent la portée de leur intérêt. En premier lieu, l'étude de réseaux de voxels soulève quelques interrogations. On l'a vu, les voxels sont des volumes dont la raison d'être à trait à des contraintes d'ordre technologique, qui font que ces voxels n'ont aucune réalité anatomique, ni fonctionnelle. L'art de l'analyse de données d'imagerie cérébrale consiste au contraire à dépasser l'échelle du voxel pour identifier les aires anatomo-

fonctionnelles où a effectivement lieu l'activité observée. En second lieu, la technique employée par Eguíluz et al. (2005) pour construire leurs réseaux conduit à des réseaux spécifiques à une seule tâche (et qui plus est, dépendants du sujet), ne dévoilant de fait qu'une partie du réseau fonctionnel global.

Le réseau que nous allons étudier à présent dépasse ces deux limitations : c'est un réseau d'aires anatomo-fonctionnelles, non pas spécifique à une tâche, mais représentant le fonctionnement global du cerveau.

4.3 Construction du réseau

L'immense majorité des expériences d'imagerie cérébrale sont calquées sur le même schéma, que l'on peut grossièrement décrire ainsi : on cherche à déterminer quelles sont les zones cérébrales impliquées dans telle ou telle tâche. Pour ce faire, on sélectionne un groupe de sujets auxquels on fait effectuer la tâche en question, ainsi qu'une tâche dite de contrôle, supposée impliquer les mêmes processus cognitifs que la première, sauf sur le point précis qui est l'objet de l'étude (par exemple, la tâche contrôle implique les mêmes activités sensorimotrices que la tâche étudiée, mais un traitement cognitif des stimuli différent). On enregistre l'activité des cerveaux des sujets durant la réalisation de ces deux tâches, et l'on compare ensuite ces enregistrements : les zones qui sont plus activées dans la tâche étudiée que dans la tâche contrôle sont celles spécifiquement engagées dans cette tâche. Ce que les dizaines de milliers d'études de ce type nous ont appris, c'est que lors de la réalisation d'une tâche un tant soit peu complexe, ce n'est pas une, mais une ensemble d'aires anatomo-fonctionnelles que l'on voit fonctionner de concert (Friston, 1994).

Le réseau fonctionnel cérébral que nous avons cherché à construire et à étudier

est le réseau dont les nœuds sont ces aires anatomo-fonctionnelles, reliées entre elles si elles participent conjointement à la réalisation d'une tâche. Il n'existe malheureusement aucune base de données regroupant ce type d'informations. Pour y accéder, nous avons donc réuni tous les résumés des articles publiés entre août 1992 et juin 2004 dans la revue *Neuroimage*, spécialisée en imagerie cérébrale. Nous avons alors fait la supposition suivante : les auteurs rapportent dans le résumé de leur article les aires cérébrales qu'ils ont vues activées dans leur étude. Nous avons donc construit un réseau dont les nœuds sont les aires cérébrales, reliées ensemble si elle sont mentionnées dans le même résumé.

Le problème majeur d'une telle approche est l'absence de normalisation dans la communauté neuroscientifique et la synonymie qui en découle : une aire anatomo-fonctionnelle peut avoir plusieurs noms. Afin de pas construire un réseau dans lequel les aires pouvaient apparaître plusieurs fois sous des noms différents, nous avons utilisé la base de données *NeuroNames* (D. M. Bowden & Martin, 1995 ; D. Bowden & Dubach, 2003). *NeuroNames* est une entreprise pour standardiser la terminologie neuro-anatomique. Elle propose un découpage hiérarchique du cerveau de l'Homme et du primate non humain en 855 structures ¹, dont 129 sont des super-structures, c'est-à-dire divisées en sous-structures. 18 de ces structures sont indiquées comme étant propres au macaque, et ont été retirées de la base qui comportait donc au final 837 structures. Pour chacune des ces structures, *NeuroNames* propose, et c'est là tout son intérêt, une liste des synonymes rencontrés dans la littérature. Pour les 837 structures que nous avons utilisées, la base fournissait 3501 synonymes, soit 4.18 termes par structure. Nous avons pu ainsi identifier dans les résumés de manière

¹Nous donnons ici les chiffres de la version avec laquelle nous avons travaillé, légèrement différente de celle décrite dans D. Bowden & Dubach (2003).

unique chaque aire, quel que soit le nom sous lequel elle apparaissait.

Sur les 1891 résumés originaux, nous n'avons conservé que les 576 mentionnant au moins deux aires. Chacun d'entre eux mentionnait en moyenne 3.21 aires. Au total, 161 aires étaient citées dans ces résumés. Chaque aire était mentionnée en moyenne dans 13.94 résumés.

4.4 Propriétés du réseau fonctionnel cérébral

Le réseau que nous avons construit était donc constitué de 150 nœuds et 1214 liens. Pour la suite des analyses de ce réseau, nous n'avons considéré que la plus grande partie connexe du réseau, constitué de 148 nœuds et 1213 liens.

4.4.1 Coefficient de regroupement et longueur caractéristique

Nous avons mesuré le coefficient de regroupement et la longueur caractéristique de ce réseau. Nous avons trouvé $C_{NeuroImage} = 0.62$ et $\ell_{NeuroImage} = 2.23$. Un graphe aléatoire de même taille et de même densité aurait $C_{aléa} = 0.122$ et $\ell_{aléa} = 1.4210$. Le réseau que nous avons obtenu est donc un petit monde.

4.4.2 Distribution de degrés

Nous avons aussi étudié la distribution de degrés de ce réseau. Le degré maximum est $k_{max} = 79$ et le degré moyen $\langle k \rangle = 16.39$. La figure 4.10 représente la distribution de degrés sur une échelle logarithmique. Si elle suit une loi de puissance d'exposant γ , elle doit apparaître comme une droite de pente γ sur un tel graphique. C'est sur la base d'une telle observation, que Eguíluz et al. (2005) ont conclu que

les distributions de degrés de leurs réseaux suivaient des lois de puissance. Une telle représentation est pourtant peu fiable pour tirer la moindre conclusion (Newman, à paraître). Ceci est en partie causé par le bruit que l'on observe pour les valeurs élevée de k , bruit qui est dû à un faible échantillonnage pour ces valeurs. Dans notre cas, il n'est pas évident d'évaluer dans quelle mesure il est possible d'ajuster nos données avec une droite.

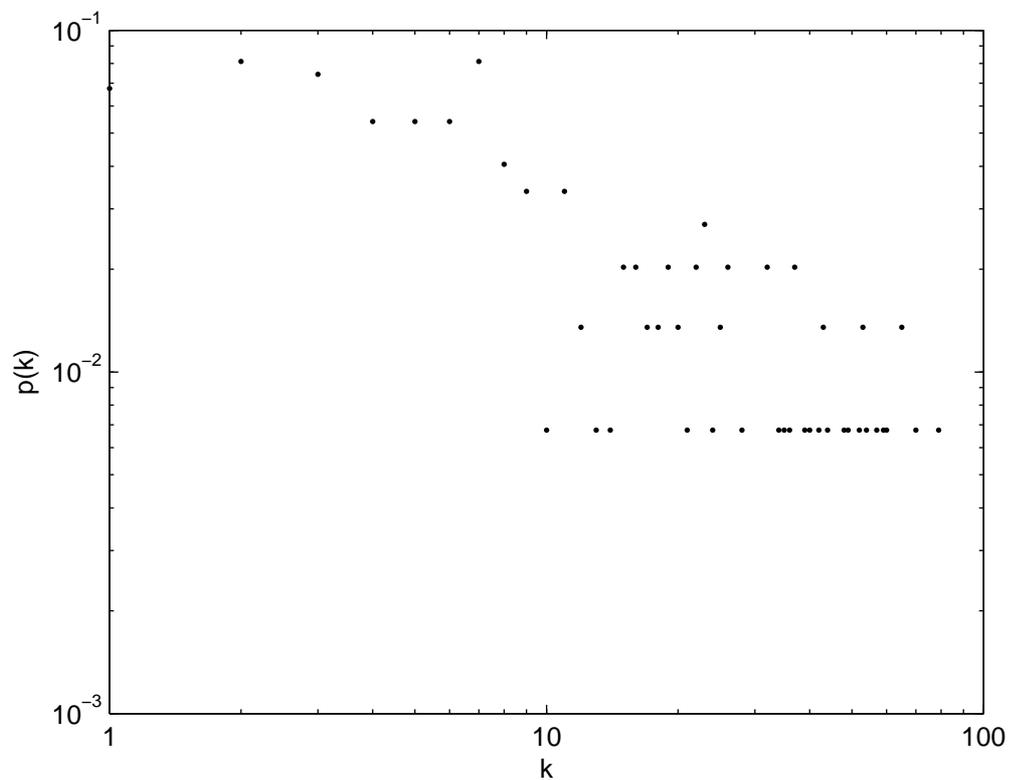


Figure 4.10 – Distribution de degrés du réseau fonctionnel cérébral avec une échelle logarithmique. Si cette distribution suit une loi de puissance, elle doit apparaître comme une droite avec une telle échelle.

Une méthode nettement plus fiable pour décider, ou non, de la présence d'une loi de puissance est d'observer non pas directement la distribution de degrés, mais

la distribution cumulative, c'est-à-dire $p(x \geq k)$ (Newman, à paraître). Cette distribution est représentée sur la figure 4.11. Si la distribution de degrés était une loi de puissance de degré γ , alors la distribution cumulative serait aussi une loi de puissance, de degré $\alpha = \gamma - 1$. Elle devrait donc apparaître aussi comme une droite sur une échelle logarithmique. Ce n'est manifestement pas le cas, la distribution de degrés de notre réseau ne suit pas une loi de puissance.

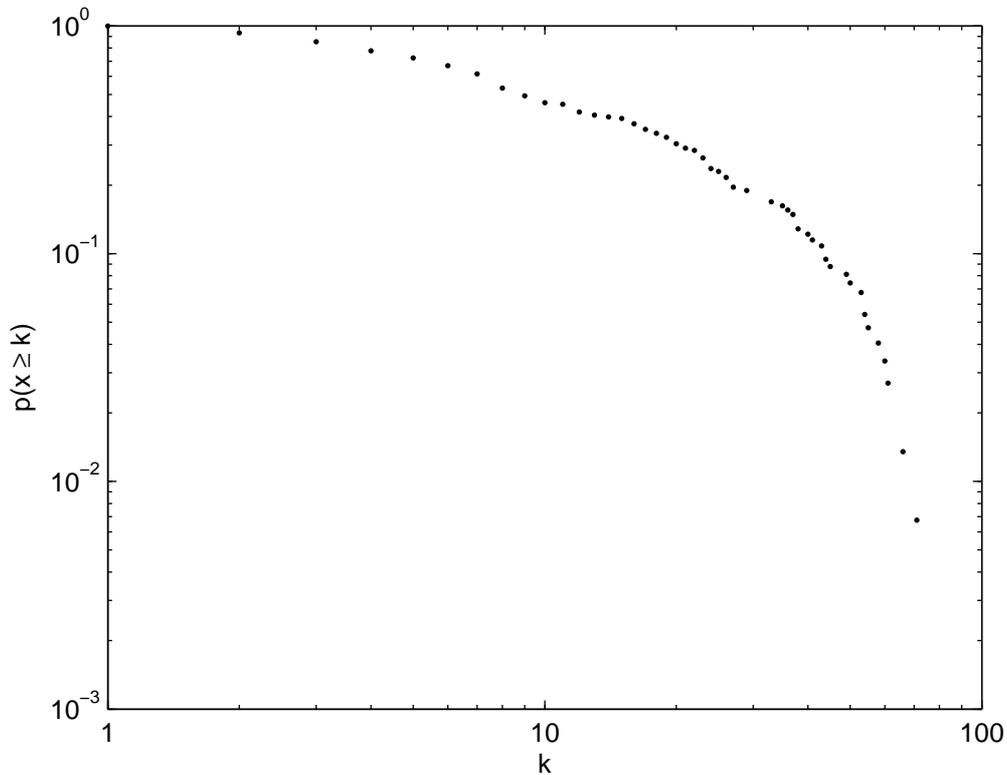


Figure 4.11 – Distribution cumulative des degrés du réseau fonctionnel cérébral avec une échelle logarithmique. Si la distribution des degrés est une loi de puissance d'exposant γ , la distribution cumulative est aussi une loi de puissance, d'exposant $\alpha = \gamma - 1$.

La nature de la distribution de degrés apparaît plus clairement lorsqu'on la trace sur un graphique semi-logarithmique (figure 4.12). Elle présente deux régimes, ca-

ractérisés par des droites. Une distribution apparaissant comme une droite sur un graphique semi-logarithmique est une distribution exponentielle (eq. 4.8)¹. Les deux régimes sont $p(k) \sim e^{-0.05k}$ pour $k < 54$ et $p(k) \sim e^{-0.12k}$ pour $k \geq 54$.

Il est remarquable de noter que Amaral et al. (2000) ont justement déterminé que le réseau neuronal de *C. elegans* a des distributions de degrés entrants et sortants qui suivent des lois exponentielles.

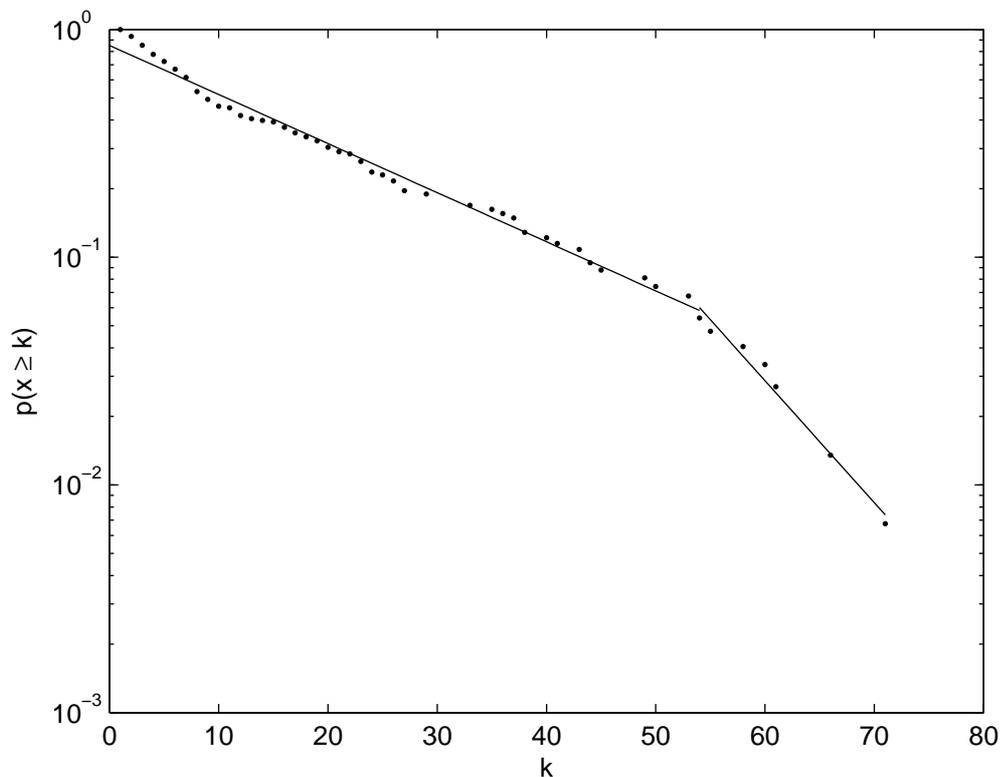


Figure 4.12 – Ajustement de la distribution des degrés du réseau fonctionnel cérébral avec avec une distribution exponentielle. Les points représentent la distribution cumulative des degrés. Les deux lignes continues représentent les deux régimes, $\beta = 0.05k$ et $\beta = 0.12$.

¹Si $p(k)$ est une distribution exponentielle d'exposant γ , alors la distribution cumulative $p(x \geq k)$ est aussi une distribution exponentielle, de même exposant. Elle apparaît donc aussi comme une droite sur un graphique semi-logarithmique.

4.5 Origines des propriétés du réseau extrait de *NeuroImage*

Nous prétendons que le réseau que nous avons construit à partir des résumés de *NeuroImage* est le réseau fonctionnel cérébral, et que l'on peut donc attribuer à ce dernier les propriétés que nous avons mesurées. De telles affirmations sont-elles légitimes ? Se pourrait-il, au contraire, que les cooccurrences de noms d'aires cérébrales dans un résumé soient de mauvais indices de relation fonctionnelle, et que les propriétés de notre réseau soit un artefact de construction ?

Ces questions revêtent une importance d'autant plus grande que le réseau que nous étudions ici est la version monopartite d'un graphe bipartite, et que l'on sait que de telles projections peuvent, sous certaines conditions, être des graphes petits mondes et / ou sans échelle (Guillaume & Latapy, 2004 ; Newman, Watts, & Strogatz, 2002). Cette section a donc pour objectif de justifier notre affirmation que les propriétés du réseau que nous avons construit sont bien celles du réseau fonctionnel cérébral que nous prétendons étudier. Nous allons dans un premier temps définir les graphes bipartites et expliquer comment, sous certaines conditions, ils permettent de fabriquer des graphes petits mondes et / ou sans échelle. Nous pourrons alors jeter un nouveau regard sur les propriétés de notre réseau, qui nous permettra de trancher entre une origine artefactuelle de ces propriétés et une origine dans l'architecture fonctionnelle cérébrale.

4.5.1 Graphes bipartites

Nous avons construit notre graphe en assignant un sommet à chaque aire anatomofonctionnelle rencontrée dans les résumés des articles que nous avons traités, et en créant une arête entre deux sommets si l'on trouve une cooccurrence des noms de ces aires dans au moins un des résumés.

Dans ce processus, il y a implicitement un second graphe que nous n'avons pas évoqué jusqu'à présent. Ce second graphe est constitué de deux types de sommets : les uns sont les même que ceux du premier graphe, les autres représentent les résumés. Dans ce graphe, une arête n'est possible qu'entre un sommet représentant une aire et un sommet représentant un résumé, et indique alors que l'aire est nommée dans le résumé (figure 4.13).

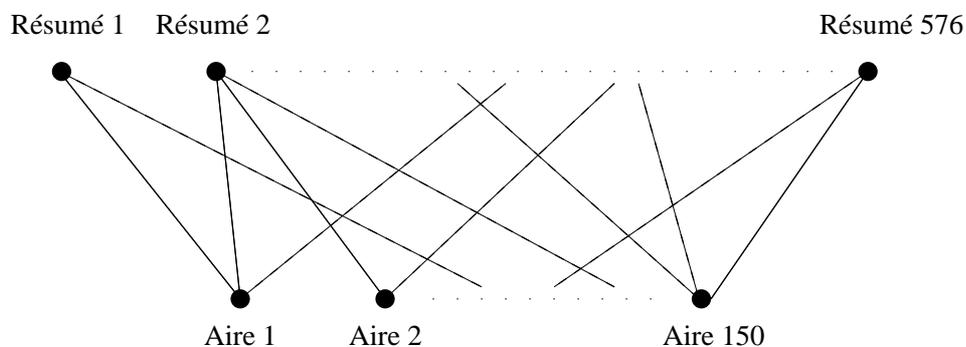


Figure 4.13 – Structure bipartite sous-jacente du réseaux fonctionnel cérébral construit.

Voici la définition formelle d'un graphe bipartite :

Définition 4.10 (Graphe bipartite) Un graphe bipartite G est un triplet (\top, \perp, E) , où \top et \perp sont deux ensembles de sommets disjoints, et $E \in \top \times \perp$ l'ensemble des arêtes.

Deux distributions de degrés sont associées à un graphe bipartite :

- la distribution des degrés des sommets de \top , que l'on notera $P_{\top}(k)$
- la distribution des degrés des sommets de \perp , que l'on notera $P_{\perp}(k)$

A partir d'un graphe bipartite, il est possible de former deux graphes non orientés, appelés projections monopartites.

Définition 4.11 (Projection monopartite) *Un graphe bipartite G a deux projections monopartites :*

- *La projection haute : $G^{\top} = (\top, E^{\top})$, où $s_i, s_j \in E^{\top}$ si et seulement si il existe un sommet de \perp auquel sont reliés s_i et s_j dans G .*
- *La projection basse : $G^{\perp} = (\perp, E^{\perp})$, où $s_i, s_j \in E^{\perp}$ si et seulement si il existe un sommet de \top auquel sont reliés s_i et s_j dans G .*

La figure 4.14 présente un graphe bipartite et ses deux projections monopartites.

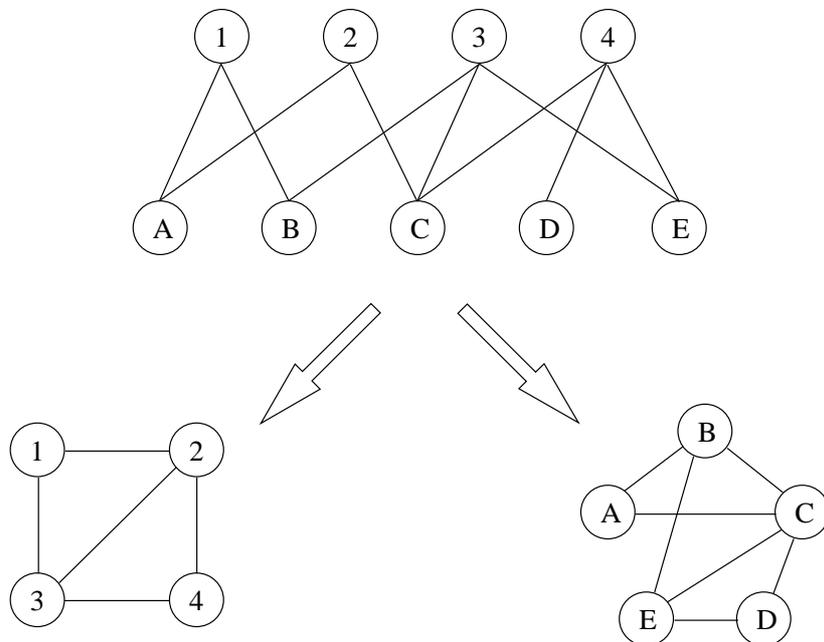


Figure 4.14 – Un graphe bipartite G (en haut) et ses deux projections monopartites : G^{\top} (en bas à gauche) et G^{\perp} (en bas à droite).

Guillaume & Latapy (2004) ont démontré trois théorèmes importants pour nous sur les projections monopartites¹ de graphes bipartites (leurs résultats convergent avec ceux de Newman et al. (2002) proposés indépendamment). Ces trois théorèmes concernent respectivement la distribution de degrés, la longueur caractéristique et le coefficient de regroupement des projections monopartites.

Le premier théorème affirme que si la distribution de degrés $P_{\perp}(k)$ suit une loi de puissance, alors la distribution de degrés de la projection monopartite suit aussi une loi de puissance, de même exposant.

Le second théorème affirme que si la distribution de degrés $P_{\perp}(k)$ suit une loi de puissance, alors la longueur caractéristique de la projection monopartite est de l'ordre de $\ln(|\perp|)$, c'est-à-dire du même ordre de grandeur que celle d'un graphe aléatoire de même densité (cf. équation 4.5).

Enfin, le dernier théorème affirme que si $P_{\perp}(2)$ est indépendant de $|\perp|$ et que $P_{\top}(2)$ ne tend pas vers 1 quand la taille du réseau augmente, alors le coefficient de regroupement de la projection monopartite est constant. Pour des graphes de taille suffisamment grande, il est donc très supérieur à celui d'un graphe aléatoire de même taille (cf. équation 4.4).

Le réseau que nous avons construit est la projection monopartite (basse) d'un graphe bipartite où \top est l'ensemble des résumés et \perp est l'ensemble des aires² (figure 4.13). Étant donné que la distribution de degrés de notre réseau ne suit pas une loi de puissance, en vertu du premier théorème de Guillaume & Latapy

¹Par symétrie, tout ce que l'on peut dire sur les projections monopartites basses est valable pour les projections monopartites hautes et *vice-versa*. Dans la suite nous ne considérerons plus que les projections basses des graphes bipartites.

²C'est le cas de nombreux réseaux, comme le réseau des collaborations cinématographiques, où \top est l'ensemble des films et \perp est l'ensemble des acteurs, le réseau des collaborations scientifiques, où \top est l'ensemble des articles et \perp est l'ensemble des scientifiques, ... Guillaume & Latapy (2004) argumentent en fait que tout réseau complexe a une structure bipartite sous-jacente.

(2004), nous pouvons affirmer que la distribution des degrés bas $P_{\perp}(k)$ du graphe monopartite non plus ne suit pas une loi de puissance. Ceci est confirmé par la figure 4.15 qui représente cette distribution. Cette distribution suit en revanche une distribution exponentielle : $P_{\perp}(k) \sim e^{-0.03k}$ (figure 4.16). La distribution des degrés hauts est elle aussi exponentielle $P_{\top}(k) \sim e^{-0.65k}$ (figure 4.17)

Nous ne sommes donc pas dans un cas d'application des deux premiers théorèmes. En revanche, étant donné que $P_{\perp}(2) \sim e^{-0.03 \cdot 2}$ est indépendant de $|\perp|$ et

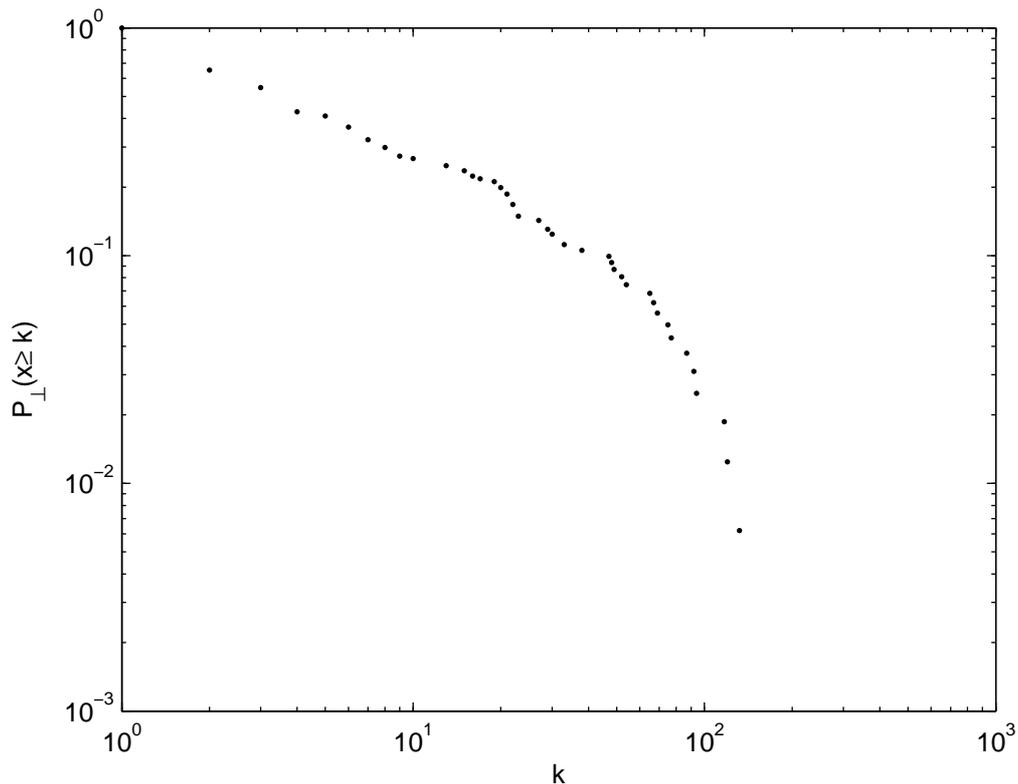


Figure 4.15 – Distribution cumulative des degrés bas du graphe monopartite sous-jacent avec des échelles logarithmiques. Si cette distribution suivait une loi de puissance, elle apparaîtrait comme une droite.

que $P_T(2) \sim e^{-0.65 \cdot 2}$ est constant, le troisième théorème s'applique.

Mais indépendamment du fait que ces théorèmes s'appliquent formellement ou non, ce qu'ils nous apprennent, c'est que les propriétés d'un réseau ayant une structure bipartite sous-jacente peuvent découler des propriétés de cette structure. Le réseau que nous étudions ici est, du fait de l'approche employée pour le construire, non pas le réel réseau fonctionnel cérébral, mais une approximation (un nouvel article,

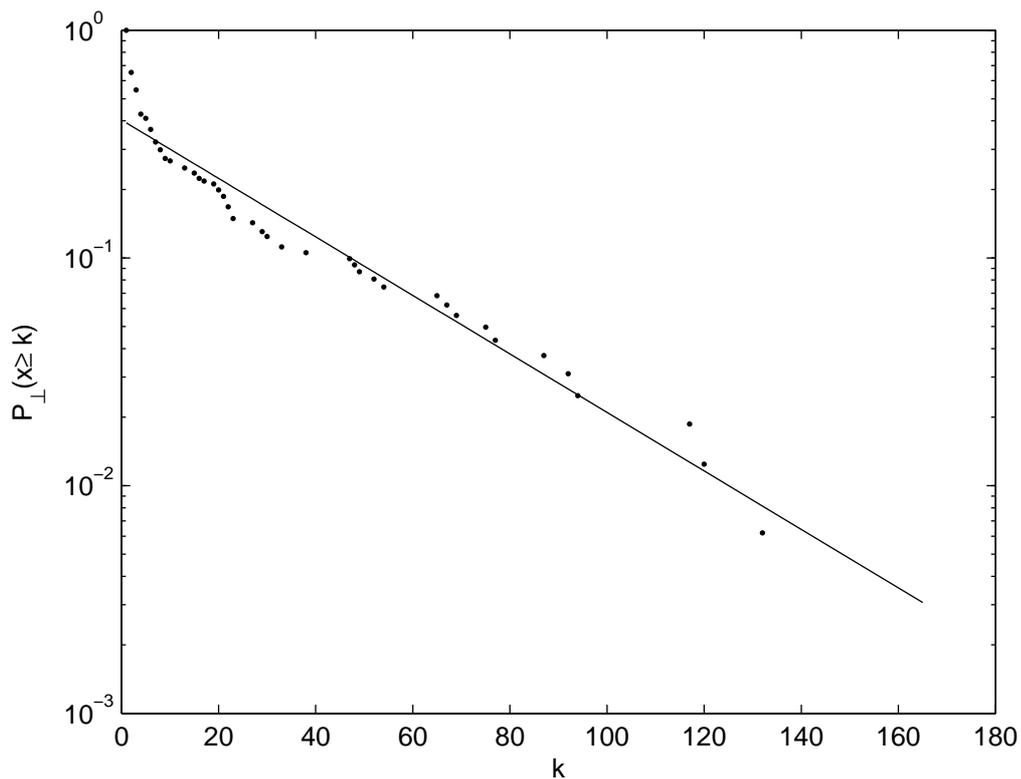


Figure 4.16 – Ajustement de la distribution cumulative des degrés bas du graphe monopartite sous-jacent. Cette distribution suit une droite de pente -0.03 sur une échelle semi-logarithmique, c'est donc une distribution exponentielle avec $\beta = 0.03$.

mettant en lumière des co-activations non encore observées, conduirait à construire un réseau légèrement différent, qui serait une meilleure approximation). Notre approximation est-elle une bonne approximation ? Les propriétés que nous avons détaillées aux sections précédentes approximent-elles le réseau fonctionnel cérébral réel, ou bien sont-elles les conséquences du graphe bipartite par lequel nous avons dû passer pour construire notre réseau ?

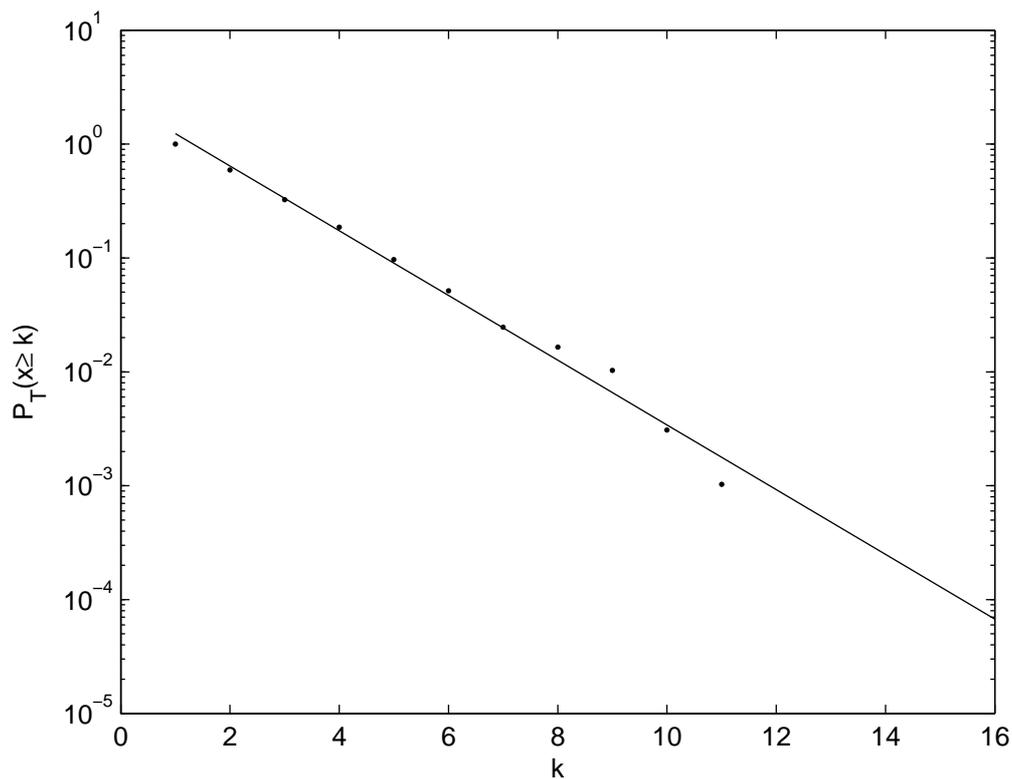


Figure 4.17 – Ajustement de la distribution cumulative des degrés hauts du graphe monopartite sous-jacent. Cette distribution suit une droite de pente -0.65 sur une échelle semi-logarithmique, c'est donc une distribution exponentielle avec $\beta = 0.65$.

4.5.2 Simulations numériques des propriétés des projections monopartites de graphes bipartites

Afin d'évaluer l'influence du graphe bipartite qui sous-tend notre réseau, nous souhaiterions connaître précisément les propriétés des projections monopartites de graphes bipartites tels que celui-ci. Pour cela, les théorèmes de (Guillaume & Laptay, 2004) ne vont pas nous être d'un grand secours : nous avons vu que, pour ce qui était de la distribution de degrés et de la longueur caractéristique, ils ne pouvaient s'appliquer dans notre cas. En ce qui concerne le coefficient de regroupement, leur théorème prédit seulement que, notre réseau doit avoir un coefficient de regroupement très supérieur à celui d'un graphe aléatoire comparable.

Pour obtenir des informations plus précises, nous avons créé 100 graphes bipartites aléatoires, ayant le même nombre de sommets que le graphe bipartite de notre réseau, et les mêmes distributions de degrés $P_{\top}(k)$ et $P_{\perp}(k)$ que celui-ci. Nous avons ensuite construit les projections monopartites basses de ces graphes et en avons mesuré la distribution de degrés, la longueur caractéristique et le coefficient de regroupement. S'il s'avère que les propriétés de notre réseau sont comparables à celles des graphes aléatoires ainsi construits, alors force sera de conclure que ses propriétés ont comme origine sa structure bipartite sous-jacente. Si elles sont différentes, alors elles ne seront pas dues, du moins en grande partie, à cette structure.

Longueur caractéristique et coefficient de regroupement

Les longueurs caractéristiques des 100 graphes aléatoires se distribuent selon une loi normale de moyenne $\mu_{\ell} = 2.14$ et de déviation standard $\sigma_{\ell} = 0.03$. La probabilité que la longueur caractéristique $\ell_{NeuroImage} = 2.23$ provienne de cette

distribution est 0.06 (figure 4.18).

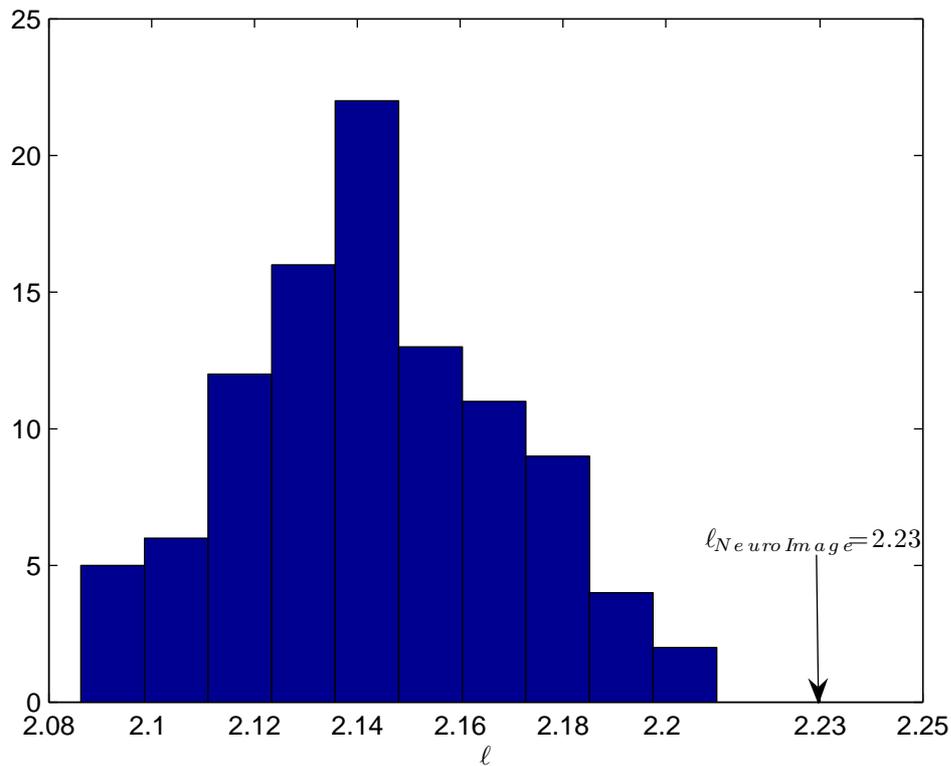


Figure 4.18 – Histogramme des longueurs caractéristiques des 100 graphes aléatoires. La probabilité d’un graphe aléatoire d’avoir comme longueur caractéristique $\ell_{NeuroImage} = 2.23$ est 0.06.

Les coefficients de regroupement des 100 graphes aléatoires se distribuent selon une loi normale de moyenne $\mu - C = 0.5$ et de déviation standard $\sigma_C = 0.01$. la probabilité que la longueur caractéristique $C_{NeuroImage} = 0.62$ provienne de cette distribution est de l’ordre de 10^{-17} (figure 4.19).

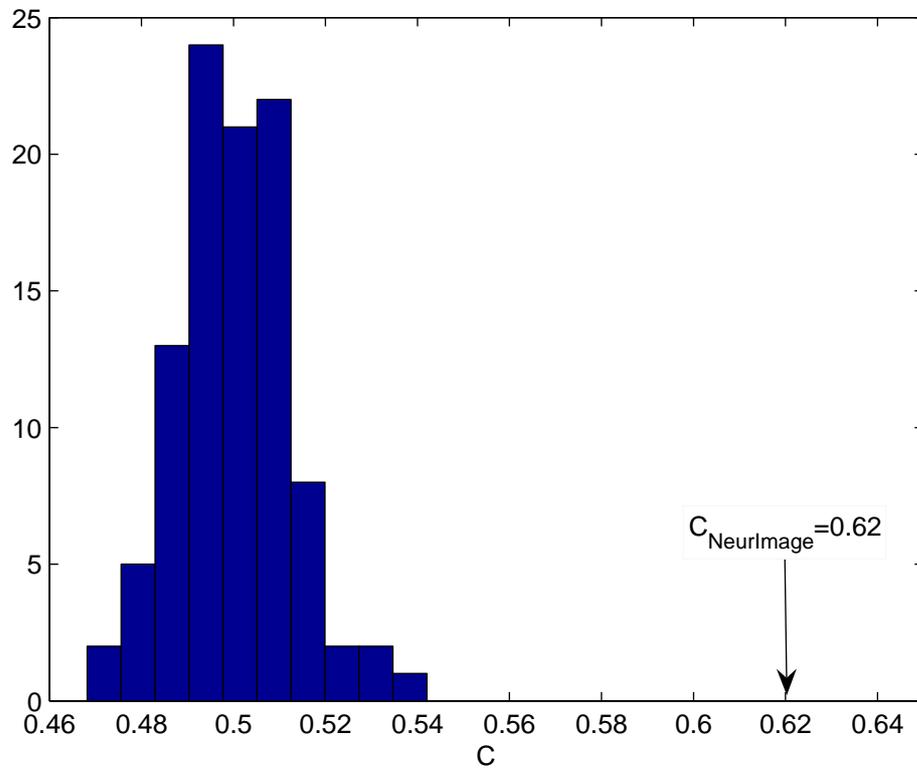


Figure 4.19 – Histogramme des coefficients de regroupement des 100 graphes aléatoires. La probabilité d'un graphe aléatoire d'avoir comme coefficient de regroupement $C_{NeuroImage} = 0.62$ est de l'ordre 10^{-17} .

Distributions de degrés

La figure 4.20 montre la distribution cumulative moyenne des degrés (ainsi les variations autour de cette moyenne) des 100 graphes aléatoires et la distribution cumulative des degrés de notre réseau. À l'image de cette dernière, les distributions de degrés des graphes aléatoires semblent suivre une distribution exponentielle avec deux régimes distincts. Néanmoins, alors que pour les petits degrés la distribution de notre réseau semble coïncider avec celle des graphes aléatoires, pour les degrés les plus grands elle s'en éloigne en décroissant nettement plus rapidement.

Conclusion sur l'origine des propriétés de notre réseau fonctionnel cérébral

Les simulations numériques que nous avons effectuées nous permettent d'affirmer que les projections monopartites de graphes bipartites aléatoires ayant les mêmes distributions de degrés que le graphe bipartite qui sous-tend notre réseau fonctionnel cérébral, sont des petits mondes. En effet, ils ont une petite longueur caractéristique, en moyenne de 2.14, et un haut coefficient de regroupement, en moyenne de 0.5. Par ailleurs, leur distribution de degrés semble suivre une loi exponentielle.

Dit autrement, si l'on reprenait tous les résumés de *NeuroImage* que nous avons utilisés, si dans chacun d'entre eux on remplaçait chaque occurrence d'aire cérébrale par le nom d'une autre, choisie aléatoirement tout en respectant le nombre de résumés dans lesquels chaque aire est citée, et si l'on construisait un réseau d'aires cérébrales sur la base de leurs cooccurrences dans les résumés, celui-ci serait, comme le nôtre dont nous affirmons qu'il représente le réseau fonctionnel cérébral, un petit monde avec une distribution exponentielle de ses degrés.

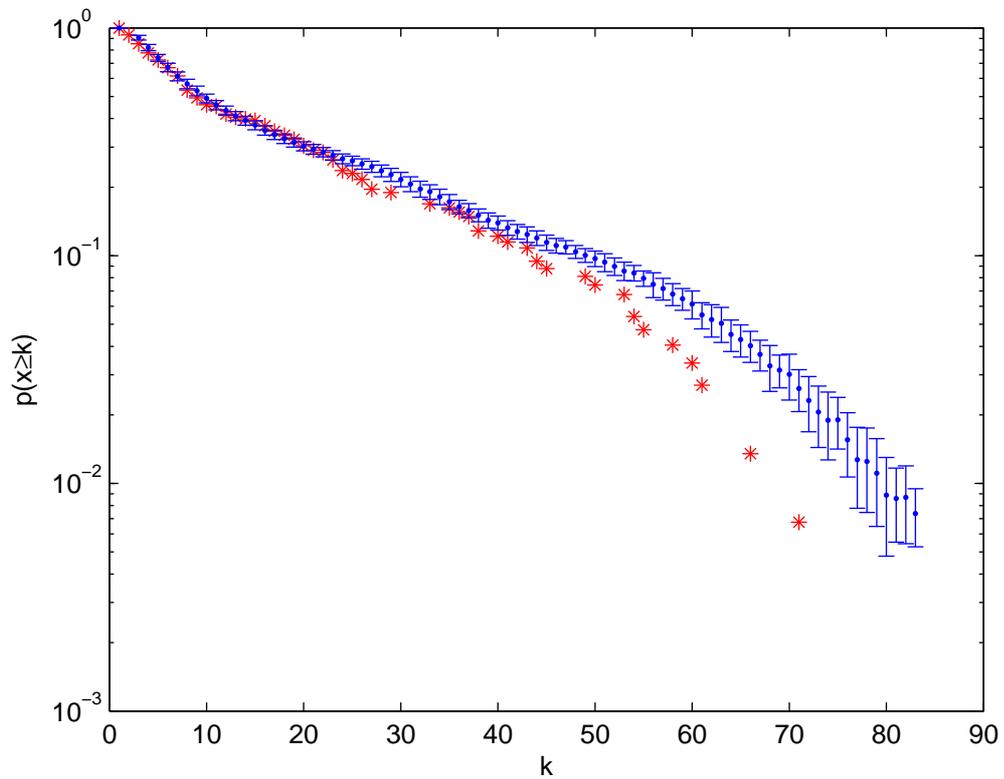


Figure 4.20 – En bleu : distribution cumulative moyenne des degrés de 100 projections monopartites de graphes bipartites aléatoires ayant les mêmes distributions de degrés $P_{\top}(k)$ et $P_{\perp}(k)$ que notre réseau. En rouge : distribution cumulative des degrés de notre réseau. Alors que les deux distributions coïncident pour les petites valeurs de k , la distribution de notre réseau s'éloigne de celles des graphes aléatoires pour les grandes valeurs.

Ceci étant dit, notre réseau a des propriétés très différentes de celles de ces graphes aléatoires. Sa longueur caractéristique et surtout son coefficient de regroupement sont très supérieurs à ceux de ces derniers. De même, si les distributions de degrés de notre réseau et des graphes aléatoires semblent qualitativement comparables, d'un point de vue quantitatif, notre réseau s'éloigne considérablement des ces derniers.

La conclusion de ces comparaisons est que notre réseau est très différent des graphes aléatoires. En tous points, ses propriétés sont différentes des leurs. Dès lors, il paraît difficile de les attribuer à sa structure bipartite sous-jacente. Plus précisément, cette structure bipartite n'est pas quelconque, et c'est sa singularité qui confère à notre réseau ses propriétés. Ce graphe bipartite représente la manière dont sont mentionnées les aires cérébrales dans les résumés d'articles relatant des expériences d'imagerie cérébrale fonctionnelle. Ces résumés étant censés énoncer les résultats importants de l'étude qu'ils décrivent, la singularité du graphe bipartite qui soutend notre réseau provient des patterns d'activations cérébrales observés lors de ces expériences. Il s'ensuit que nous pouvons à présent affirmer que le réseau que nous avons construit est une bonne représentation du réseau fonctionnel cérébral, et que les propriétés que nous en avons données sont bien les propriétés du réseau fonctionnel cérébral. Celui-ci est donc un petit monde, dont la distribution de degrés est composée de deux régimes exponentiels différents.

5 Conclusions

Un réseau représente un ensemble d'entités entretenant des relations les unes avec les autres. La généralité de cette définition est telle qu'elle peut s'appliquer à un nombre incalculable de cas de figure, dans sans doute tous les domaines de la Science. Néanmoins, la généralité de cette notion ne la rend pas caduque pour autant : mathématiquement formalisée par la théorie des graphes, elle permet d'analyser puissamment des problèmes issus de la sociologie, de l'informatique, de la biologie, . . . De plus, l'action combinée de la puissance et de la généralité de la notion de réseau a permis de mettre en évidence des similarités entre des phénomènes semblant pourtant très différents les uns autres. La prise de conscience de ce dernier point, à la fin du siècle dernier, constitue sans doute un tournant dans de nombreuses disciplines.

On pensait auparavant que les grands ensembles d'entités en interaction que l'on observait, répondaient à des schémas d'interactions propres à chacun, dont les principes organisateurs étaient si complexes, qu'au final ces schémas d'interactions étaient aléatoires. Le modèle de graphes aléatoires introduit par Erdős et Rényi, accompagné des développements théoriques de ces derniers et de leurs successeurs, s'imposaient alors comme des outils de prédilection pour appréhender ces grands réseaux complexes.

Les travaux de Watts & Strogatz (1998) et Barabási & Albert (1999), sans doute rendus possibles par des puissances de calcul qui nous permettent désormais de

manipuler ces réseaux, en ont radicalement changé notre vision. Nous savons maintenant que ces grands réseaux complexes, que l'on peut observer partout autour de nous, sont non seulement très différents des graphes aléatoires, mais surtout très similaires les uns des autres. Trois propriétés, apparemment indépendantes, se retrouvent quasi systématiquement dans tous ces réseaux : mêmes si elles ne sont pas en contact, deux entités sont toujours proches l'une de l'autre ; ces réseaux présentent un contraste entre une structure globale peu dense, c'est-à-dire présentant peu d'interactions au vu du nombre d'entités présentes, et une structure locale très dense, les entités ayant tendance à se regrouper en petites communautés au sein desquelles les interactions sont très fréquentes ; enfin, le nombre d'interactions de chaque entité est une grandeur irrégulièrement distribuée : tandis que la majorité interagissent avec très peu d'autres, quelques unes interagissent avec un grand nombre. On appelle "petits mondes" les réseaux vérifiant les deux premières de ces propriétés, et "réseaux sans échelle" ceux qui vérifient une version un peu plus spécifique de la troisième. Ces propriétés semblent quasi-universelles, tant elles sont récurrentes dans les réseaux que l'on observe.

Nous avons voulu vérifier la présence de ces propriétés dans un réseau très peu étudié quoique fascinant, le réseau fonctionnel cérébral. Dans ce réseau, les entités sont les aires cérébrales, et leurs interactions sont marquées par leur collaboration à la réalisation des activités cognitives. Nous avons utilisé des données d'imagerie cérébrale fonctionnelle pour construire ce réseau et avons montré qu'il était un petit monde, confirmant, une fois de plus, l'omniprésence de ce type de structure dans la nature.

Tout au long de ce chapitre, nous n'avons pas évoqué le langage. Pourtant il semble séduisant de vouloir l'observer en adoptant une perspective d'un ensemble

4.5 Origines des propriétés du réseau extrait de NeuroImage

d'entités en interaction. Peut-on construire un ou des réseaux à partir d'une langue ? Quelles en seraient les propriétés ? En particulier, peut-on retrouver dans le langage les propriétés universelles des réseaux complexes ? Qu'est-ce que ces réseaux pourraient nous apprendre sur les changements sémantiques ? Le chapitre suivant est consacré à ces questions, et aux réponses que l'on peut y apporter.