

UNIVERSITÉ DE MONTRÉAL

PHY 3030 - PROJET DE FIN D'ÉTUDES

La physique des réseaux

par :

Patrice Béchard

20019173

7 mai 2017

Table des matières

Résumé	2
1 Introduction	3
2 La théorie des réseaux en bref	3
2.1 Propriétés des réseaux	4
2.2 Réseau aléatoire : Modèle d'Erdős-Rényi	6
2.3 Invariance d'échelle et Modèle de Barabási-Albert	7
3 Réseau #1 : cercle d'amis Facebook	11
3.1 Cercle d'amis Facebook avec hôte	13
3.2 Cercle d'amis Facebook sans hôte	14
4 Réseau #2 : cartographier le <i>World Wide Web</i>	17
4.1 Domaine <i>phys.umontreal.ca</i>	18
4.2 Domaine <i>exoplanetes.umontreal.ca</i>	20
4.3 Domaine <i>lps.umontreal.ca</i>	22
4.4 Réseau incluant les trois domaines	22
5 Comparaison des réseaux étudiés avec d'autres réseaux réels	26
5.1 Réseau de neurones du ver <i>C. elegans</i>	27
5.2 Réseau d'acteurs d'Hollywood	28
6 Conclusion	29
Bibliographie	30

Résumé

L'étude de deux réseaux réels a été conduite. Le premier réseau étudié est un cercle d'amis Facebook et présente une distribution de degré suivant une loi de puissance avec $\gamma = 2.20 \pm 0.10$, caractéristique très courante dans les réseaux réels. Celui-ci présente aussi le phénomène des petits mondes, disant que deux noeuds choisis aléatoirement dans le réseau sont séparés par une petite distance, avec une distance moyenne de séparation de $\langle d \rangle \approx 2$. Le second réseau étudié est une carte de plusieurs sites internet. Celle-ci présente aussi une distribution de degré invariante d'échelle, avec $\gamma = 2.0 \pm 0.2$, ainsi que le phénomène des petits mondes, avec $\langle d \rangle = 2.052$. Ces réseaux ont été comparés avec d'autres réseaux réels étudiés précédemment et comportent plusieurs caractéristiques communes.

1 Introduction

Le début du 21e siècle a vu l'apparition de deux sciences grâce à l'avènement d'une grande puissance de calcul des ordinateurs. Tout d'abord, la science des données (*Data Science*) nous permet de traiter et de comprendre la quantité astronomique de données provenant d'une multitude de capteurs pour en retirer des tendances et pouvoir mieux prédire des situations. L'autre science majeure ayant fait son apparition est la science des réseaux (*Network Science*). Dans un monde de plus en plus interconnecté, il devient impératif de comprendre l'interaction entre la grande quantité de parties constituant un réseau. L'internet des choses, phénomène en effervescence de nos jours, combine des éléments de ces deux sciences, notamment en connectant des appareils ménagers et en les faisant communiquer dans le but de rendre une maison plus intelligente, par exemple.

Ce texte porte sur la science des réseaux et vise à mettre en lumière certaines caractéristiques récurrentes dans divers réseaux réels. La section 2 présentera tout d'abord une brève introduction à la théorie des graphes, pilier mathématique de la science des réseaux, et présentera deux modèles distincts de réseaux, soit le modèle d'Erdős-Rényi pour les réseaux aléatoires, et le modèle de Barabási-Albert pour les réseaux invariants d'échelle. Ensuite, un premier réseau réel, un réseau social Facebook, sera étudié à la section 3. Un second réseau réel, soit une partie du *World Wide Web*, sera ensuite abordé à la section 4. Finalement, une comparaison de ces deux réseaux avec d'autres réseaux réels sera faite à la section 5.

2 La théorie des réseaux en bref

La théorie des réseaux est une science très jeune et multidisciplinaire, regroupant des chercheurs en mathématiques, informatique, biologie, physique, et plusieurs autres domaines. Elle tire ses origines de la théorie des graphes, branche des mathématiques, ainsi que de certains travaux en sciences sociales sur le phénomène des *petits mondes* (d'où provient l'adage que toute personne sur Terre n'est éloignée de tout le reste des gens que de 6 personnes, aussi connu sous le nom de *six degrés de séparation*).

La théorie des graphes tire ses origines du problème classique des ponts de Königsberg (voir fig. 1), résolu pour la première fois par Euler, utilisant les graphes pour prouver qu'il était impossible de passer par les 7 ponts une seule fois et de revenir au point initial.

Cette section présentera tout d'abord quelques propriétés de base de la théorie des réseaux. Ensuite suivra une présentation de la théorie derrière les réseaux aléatoires, qui sont à la base de cette science. Finalement, une introduction aux réseaux invariants d'échelle ainsi qu'à un modèle permettant de les représenter sera abordée. La grande majorité de cette section est basée sur trois ouvrages, soit le livre écrit par Albert-László Barabási sur le sujet de la science des réseaux [4], un article écrit par Barabási et

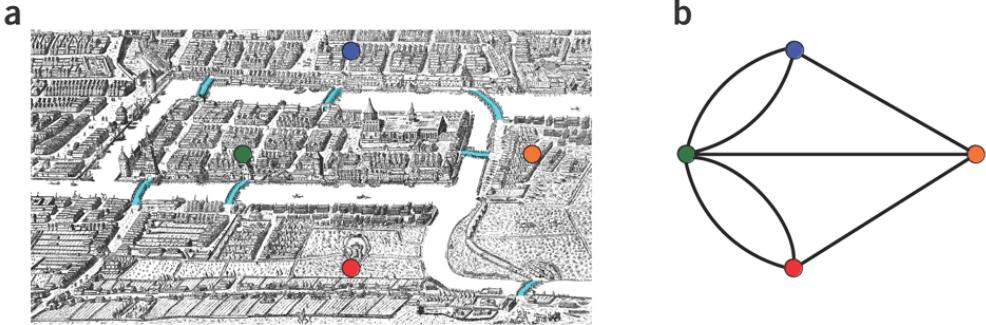


FIGURE 1 – a) Carte de la ville de Königsberg avec chaque partie de la ville identifiée par un point de couleur et chaque pont accentué en bleu. b) Représentation du problème sous forme de graphe, où chaque noeud représente une partie de la ville, et où chaque arête représente un pont reliant les parties de la ville. [6]

son étudiante de l'époque Réka Albert [1], ainsi qu'un article présentant le sujet en détails, par Mark EJ Newman [14].

2.1 Propriétés des réseaux

Un réseau est représenté par un ensemble de noeuds reliés entre eux par des liens (en théorie des graphes, on parle plutôt d'arêtes ou d'arcs, mais on omettra la distinction dans le présent document). Cette abstraction est essentielle à la compréhension d'un système complexe, pour comprendre comment les différentes composantes sont reliées entre elles. Le nombre total de noeuds d'un réseau (N) représente le nombre total de composantes d'un système. Les noeuds sont souvent identifiés selon leur indice ($1, 2, \dots, N$). Similairement, le nombre de liens d'un réseau (L) représente le nombre d'interactions entre chacun des noeuds. Un lien est identifié par les deux noeuds qu'il relie (ex. $(3, 4)$ relie les noeuds 3 et 4). Un réseau peut être *orienté* ou non (*directed* ou *undirected network* en anglais), signifiant dans quel sens il est possible de se déplacer sur un lien reliant deux noeuds ; dans un réseau orienté, pour le lien (n_1, n_2) , il est possible de se déplacer du noeud n_1 vers n_2 , mais l'inverse est impossible, alors que pour un réseau non-orienté, les deux trajets sont possibles.

Le degré k_i d'un noeud i du réseau représente le nombre de connections que ce noeud fait avec les autres. Il est possible de trouver des expressions pour le nombre total de liens ainsi que le degré moyen d'un réseau :

$$L = \frac{1}{2} \sum_{i=1}^N k_i \quad \langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N} \quad (1)$$

Un noeud dans un réseau orienté possède un *in-degré* ainsi qu'un *out-degré*, dénotant le nombre de liens pointant vers ce noeud, ainsi que le nombre de liens partant de celui-ci pour aller pointer vers un autre point. Le degré d'un noeud est l'addition de ces deux valeurs. Le nombre total de liens dans un

graphé orienté et le degré moyen sont alors donnés par les expressions suivantes :

$$L = \sum_{i=1}^N k_{in} = \sum_{i=1}^N k_{out} \quad \langle k_{in} \rangle = \langle k_{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in} = \frac{1}{N} \sum_{i=1}^N k_i^{out} = \frac{L}{N} \quad (2)$$

La distribution de degré p_k est la probabilité de choisir un noeud au hasard et que le degré de ce noeud soit k . Cette valeur doit bien entendu être normalisée. Le degré moyen d'un réseau peut aussi être trouvé avec cette distribution. La forme de la distribution de degré est une valeur très importante, déterminant la robustesse d'un réseau réel à des attaques extérieures (comme la propagation d'un virus, par exemple). Les expressions suivantes définissent mathématiquement ces concepts :

$$p_k = \frac{N_k}{N} \quad \sum_{k=1}^{\infty} p_k = 1 \quad \langle k \rangle = \sum_{k=0}^{\infty} k p_k \quad (3)$$

Le nombre de noeuds et de liens varie grandement d'un réseau à un autre. Pour un réseau de N noeuds, le nombre maximal de liens possibles est donné par $L_{max} = \frac{N(N-1)}{2}$. Un réseau possédant un tel nombre de noeuds est appelé un réseau *complet*. Les réseaux réels possèdent cependant un nombre de liens $L \ll L_{max}$.

La distance entre deux composantes est une caractéristique important des réseaux. Un *chemin* entre les noeuds n_i et n_j représente le trajet entre ces deux noeuds suivant les liens du réseau. La *longueur* de ce chemin est donné par le nombre de liens empruntés pour passer du noeud n_i au noeud n_j . Le chemin le plus court représente évidemment le chemin pour lequel la longueur du trajet est minimisée. Elle est appelée la *distance* d_{ij} entre les noeuds n_i et n_j . Dans un réseau non-orienté, $d_{ij} = d_{ji}$, mais cela n'est pas nécessairement le cas dans un réseau orienté. En pratique, ce chemin est facilement obtenu avec un algorithme de parcours en largeur (*Breadth First Search (BFS)*). Le *diamètre* d_{max} d'un réseau représente la plus grande distance entre deux noeuds du réseau. La *longueur de chemin moyen* d'un réseau représente la distance moyenne entre chaque paire de noeuds du réseau :

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1,N,i \neq j} d_{ij} \quad (4)$$

Deux noeuds n_i et n_j sont dits *connectés* si il est possible d'établir un chemin entre les deux noeuds (en d'autres mots, $d_{ij} \neq \infty$). Un réseau est connecté si tous ses noeuds sont connectés entre eux. Deux parties déconnectées d'un réseau sont appelés des *îlots* ou *regroupements* (*clusters* en anglais). Un seul lien bien placé peut nous permettre de connecter deux îlots ensemble.

La dernière propriété des réseaux abordée ici sera le coefficient de regroupement (*clustering coefficient*), représentant à quel point les voisins d'un noeud sont connectés entre eux, indicatif de la densité de liens locale. Ce coefficient est de 0 si aucun voisin n'est connecté et 1 si le voisinage forme un réseau complet. Le coefficient de regroupement moyen est tout simplement la moyenne des coefficients de regroupement de chaque noeud du réseau. Mathématiquement, le coefficient de regroupement d'un noeud n_i est donné

par :

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (5)$$

2.2 Réseau aléatoire : Modèle d'Erdős-Rényi

Les travaux de Paul Erdős et Alfréd Rényi sur les graphes aléatoires à la fin des années 1950 contribuèrent énormément à l'avènement de la théorie des réseaux [10]. Un réseau aléatoire est composé de N noeuds, où chaque paire de noeuds est connectée selon une probabilité p .

La modélisation de ce type de réseau est plutôt simple : il suffit de générer N noeuds isolés, puis de passer sur chacun des $N(N - 1)/2$ paires distinctes de noeuds en tirant un noeud aléatoire. Si ce nombre tiré est plus petit que la probabilité p , les deux noeuds sont alors connectés. Chaque réseau généré avec un N et un p donnés seront différents. Le nombre total de liens créés dans le réseau sera distribué selon une distribution binomiale, avec un nombre moyen de liens créés $\langle L \rangle$ et une déviation standard σ_L donnés par :

$$\langle L \rangle = p \frac{N(N - 1)}{2} \quad \sigma_L = \sqrt{\langle L^2 \rangle - \langle L \rangle^2} = \sqrt{p(1 - p)N} \quad (6)$$

Le nombre moyen de liens créé est donc tout simplement le nombre maximum de liens L_{max} multiplié par la probabilité que deux noeuds soient connectés. La distribution de degré d'un tel réseau suivra aussi une loi binomiale. Le degré moyen d'un tel réseau sera donné par :

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N - 1) \quad (7)$$

Dans l'approximation où $\langle k \rangle \ll N$ (ceci est tout à fait raisonnable dans la plupart des réseaux réels, puisqu'ils sont généralement peu connectés), la distribution binomiale peut être approximée comme une distribution de Poisson, selon l'équation :

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad \sigma_k = \langle k \rangle^{\frac{1}{2}} \quad (8)$$

Le phénomène des petits mondes, qui stipule que deux noeuds choisis aléatoirement dans un réseau sont séparés par une distance d courte, est présent dans les réseaux aléatoires [24]. Cette distance moyenne entre deux noeuds aléatoirement choisis peut être exprimée selon l'équation :

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle} \quad (9)$$

La distance moyenne ou le diamètre d'un réseau aléatoire dépend donc logarithmiquement de la taille du système. Le terme $\ln \langle k \rangle^{-1}$ implique que plus le réseau est dense, plus les distances séparant deux noeuds sont petites.

Un réseau suivant le modèle d’Erdős-Rényi de 250 noeuds avec un probabilité de connexion de 0.05 a été modélisé à la figure 2 et sa distribution de degré ainsi que plusieurs renseignements pertinents sont présentés à la figure 3.

Peu de noeuds dans le réseau possèdent une taille très petite ou un taille très grande. En fait, la majorité des noeuds possèdent une taille intermédiaire. Cette caractéristique s’aperçoit aussi très bien sur le graphique montrant la distribution de degré du réseau, où l’on voit très clairement la distribution de Poisson mentionnée plus tôt.

Bien que les réseaux aléatoires soient intéressants au niveau théorique, les réseaux réels ne sont pas distribués suivant une loi de Poisson. Cette distribution sous-estime grandement le nombre de pivots d’un réseau, soit le nombre de noeuds grandement connectés. Par exemple, certains sites comme Google et Facebook possèdent un grands nombres de liens entrant et sortants, et leur nombre est beaucoup plus grand que ne le prévoit la distribution de Poisson. Cette découverte à la fin des années 1990 est le pilier fondateur de la théorie des réseaux contemporaine. Celle-ci sera couverte dans la prochaine section.

2.3 Invariance d’échelle et Modèle de Barabási-Albert

Il pourrait être instinctif de penser que les réseaux réels sont aléatoires et suivent le modèle d’Erdős-Rényi. Cependant, plusieurs résultats sur le sujet à la fin des années 1990 ont montré que ceci n’était pas le cas, et que les réseaux réels n’étaient pas distribués en suivant une loi de Poisson, mais plutôt une loi de puissance. Cette découverte a été fait pour la toute première fois par Réka Albert, Hawoong Jeong et Albert-László Barabási lors de leur tentative de mapper le domaine complet de l’Université Notre Dame ([nd.edu](#)). En créant un robot se déplaçant de site internet en site internet (*webcrawler*), ils ont pu représenter un réseau comprenant 325 729 noeuds, reliés entre eux par 1 469 680 liens, présentant une distribution suivant une loi de puissance, signe que, malgré le fait que l’internet est à première vue un réseau aléatoire, où tout le monde peut créer une page et la lier à n’importe quelle autre, elle possède en fait une certaine structure [2]. Ils trouvèrent que le diamètre du web pouvait être approximé avec l’expression suivante :

$$\langle d \rangle = 0.35 + 2.06 \log_{10}(N) \quad (10)$$

Aujourd’hui, on estime la taille du *World Wide Web* à 10^{12} documents reliés entre eux par des URLs, en faisant ainsi le plus grand réseau créé par l’humain.

Un réseau suivant une loi de puissance $p_k \sim k^{-\gamma}$ est aussi appelé un réseau invariant d’échelle. La principale caractéristique des réseaux invariants d’échelle sont la présence de pivots (*hubs*), qui sont des noeuds possédant un grand degré. Selon le modèle de Erdős-Rényi, ce genre de noeud ne devrait pas exister, tout comme les noeuds possédant un très petit degré. Les réseaux invariants d’échelle comportent

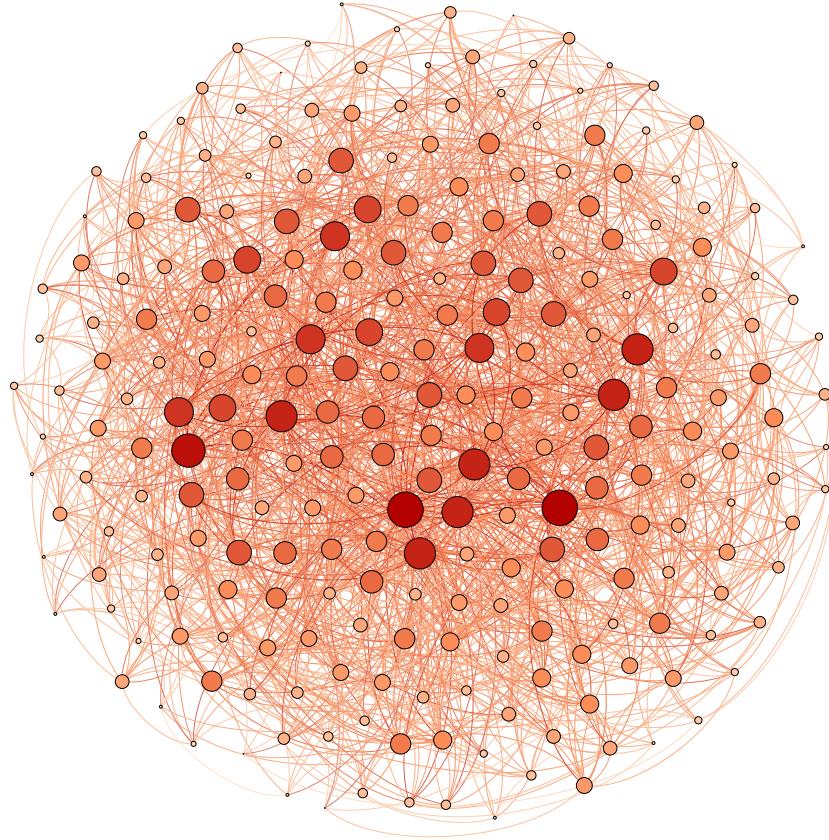


FIGURE 2 – Réseau aléatoire avec $N = 250$ et $p = 0.05$ présentés en fonction de leur degré. Le réseau possède un nombre de liens $L = 1586$ et le nombre moyen de liens attendu est $\langle L \rangle = 1556.25$

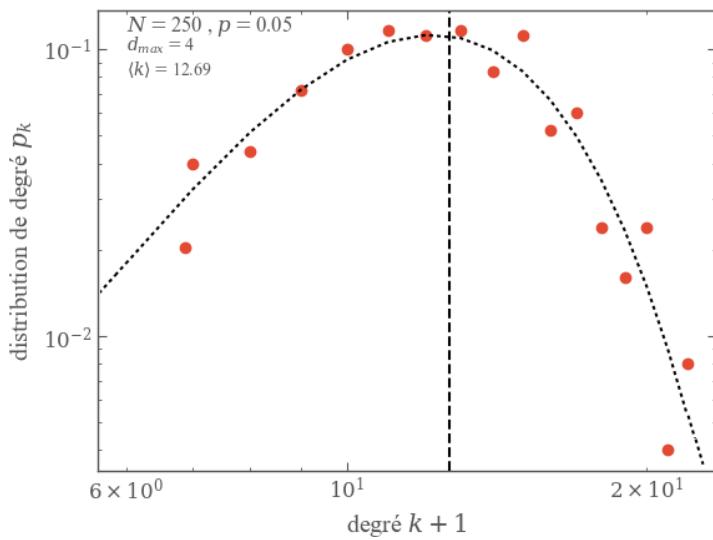


FIGURE 3 – Distribution de degré du réseau présenté ci-haut. La ligne verticale représente le degré moyen du réseau et la courbe pointillée représente la distribution de Poisson associée à la distribution de degré, selon l'équation. 8

cependant un grand nombre de noeuds possédant un petit degré, allant encore une fois à l'encontre du modèle des réseaux aléatoires.

La plupart des réseaux retrouvés dans la vraie vie sont invariants d'échelle. En effet, le *World Wide Web*, le réseau des collaborations scientifiques et le réseau d'interaction de protéines, pour n'en nommer quelques uns, sont tous des réseaux suivant une loi de puissance. Il est facile de voir si un réseau étudié présente une invariance d'échelle : il suffit de mettre en graphique log-log la distribution de degré du réseau et d'observer si cette distribution suit une droite. La pente de cette droite tracée correspond à l'exposant de degré γ . Il est aussi courant qu'un réseau ne montre pas d'invariance d'échelle, et ce, à cause de la taille du réseau étudié (incomplétude des données). Il est à noter que ce ne sont pas tous les réseaux réels qui suivent une loi de puissance, l'exemple le plus flagrant étant un réseau d'atomes dans un matériau [5].

Dans ce genre de réseau, à cause de la présence des pivots, le phénomène des petits mondes est affecté, de sorte que la distance entre deux noeuds choisis aléatoirement est encore plus petite que pour un réseau aléatoire. Cette distance moyenne dépend de la grandeur du réseau N ainsi que de l'exposant de degré γ du réseau. Lorsque $\gamma > 3$, la distance moyenne entre deux noeuds est toujours proportionnelle à $\ln N$. Dans ce cas, il n'y a pas assez de pivots pour influencer la distance entre les noeuds. Ce régime est appelé *régime aléatoire*. Lorsque $2 < \gamma < 3$, cette distance est plutôt proportionnelle à $\ln \ln N$ et le réseau se trouve en *régime invariant d'échelle*. Le dernier cas est lorsque $\gamma < 2$. Dans ce cas, le réseau est en *régime anormal* et le pivot principal est connecté à l'ensemble du réseau, de sorte que la distance moyenne entre deux noeuds est constante.

Un modèle simple permettant de construire un réseau invariant d'échelle est le modèle de Barabási-Albert. Ce modèle se base sur deux principes fondateurs : la **croissance** et l'**attachement préférentiel**. La croissance prend en compte l'évolution du réseau dans le temps. Les réseaux réels ne sont pas statiques ; ils sont au départ composés de quelques noeuds, et le réseau croît progressivement dans le temps. L'attachement préférentiel prend en compte le fait qu'un nouveau noeud créé aura tendance à se lier à un noeud déjà assez connecté. Un exemple révélateur permettant de bien comprendre ces concepts est celui du réseau des collaborations scientifiques, reliant chaque article à tous les articles que celui-ci cite. À chaque fois qu'un nouvel article s'écrit, il est ajouté au réseau. Les chances que cet article cite un article possédant déjà beaucoup de citations est plus envisageable que le fait qu'il cite un article possédant peu de citations et ayant peu été lu.

Le modèle de Barabási-Albert permet facilement d'implémenter un réseau possédant ces principes fondateurs. Tout d'abord, un nombre de noeuds initiaux m_0 est créé et des liens entre eux sont faits. À chaque pas de temps, un nombre $m < m_0$ de nouveaux noeuds est introduit (croissance). La probabilité qu'un de ces nouveaux noeuds se lie avec un noeud i préalablement présent dans le réseau dépend du

degré k_i du noeud, selon l'équation suivante :

$$\Pi_k = \frac{k_i}{\sum_j k_j} \quad (11)$$

Il est donc possible pour un nouveau noeud de se lier à tout noeud dans le réseau (si son degré n'est pas nul), et les noeuds possédant le plus grand degré (les pivots) se voient avantagés par cette méthode d'attribution, devenant de plus en plus connectés avec le temps. Sans modification, ce modèle retourne un exposant de degré $\gamma = 3$. Une analyse de l'évolution de la distribution de degré dans ce genre de réseau montre que plus le temps avance, moins les noeuds possédant peu de connections verront ce nombre grandir. Les noeuds apparaissant au début de l'évolution sont donc très avantagés pour devenir un pivot du réseau.

Il est pertinent de se demander si les deux principes fondateurs sur lesquels le modèle se base sont nécessaires pour l'apparition du phénomène d'invariance d'échelle dans le réseau. Pour ce faire, deux modèles différents seront étudiés, soit un omettant l'attachement préférentiel (modèle **A**), et l'autre omettant la croissance (modèle **B**).

Dans le modèle A, un nombre de noeuds $m < m_0$ est ajouté à chaque pas de temps. La probabilité que ces noeuds se lient à d'autre noeuds du réseau ne dépend plus du degré du noeud. Ce choix est donc aléatoire. La distribution de degré prédite par ce genre de réseau est donné par $p(k) \sim \exp(-k/m)$, qui décroît beaucoup plus rapidement qu'une loi de puissance. Dans ce genre de réseau, il n'y aura donc pas apparition de pivots. Dans le modèle B, le réseau commence avec N noeuds et ne change pas. Les liens entre les noeuds sont créés selon l'équation 11. Au début de l'évolution du réseau, celui-ci semble se comporter comme un réseau invariant d'échelle, mais puisque le réseau ne peut pas grandir, le degré moyen $\langle k \rangle$ ne cessera de grandir, et ce, jusqu'à atteindre la valeur maximale de $k_{max} = N - 1$, formant alors un réseau complètement connecté. Les deux conditions sont donc nécessaires pour que le modèle réplique un réseau invariant d'échelle.

L'équation 11 présente une corrélation linéaire entre la probabilité qu'un noeud se lie à un autre et le degré du second noeud. Il est possible de modifier ce modèle de sorte cette relation ne soit pas linéaire, seulement en introduisant un facteur α , de sorte que la probabilité qu'un noeud se lie à un autre devienne :

$$\Pi_k = \frac{k_i^\alpha}{\sum_j k_j} \quad (12)$$

Ce nouveau modèle est appelé le modèle de Barabási-Albert non-linéaire. Plusieurs régimes différents font leur apparition en fonction du paramètre α choisi. Pour $\alpha = 0$, l'attachement préférentiel est absent du modèle. Pour $0 < \alpha < 1$ (régime *sous-linéaire*), le biais influençant la connection entre des noeuds n'est pas assez grande pour retrouver une invariance d'échelle. Pour $\alpha = 1$, le modèle de Barabási-Albert standard est retrouvé. Finalement, lorsque $\alpha > 1$, les noeuds principaux vont être grandement avantagés

et le réseau sera dominé par quelques *super-pivots*. La distance maximale entre deux noeuds sera alors très petite.

La distance moyenne entre deux noeuds d'un réseau suivant le modèle de Barabási-Albert et le coefficient de regroupement moyen d'un tel réseau seront donnés par les expressions suivantes :

$$\langle d \rangle \approx \frac{\ln N}{\ln \ln N} \quad \langle C \rangle \approx \frac{(\ln N)^2}{N} \quad (13)$$

Comme mentionné plus tôt, la distance moyenne entre deux noeuds sera plus petite que pour un réseau aléatoire de la même taille. De plus, le coefficient de regroupement sera plus élevé que pour un réseau aléatoire, signe qu'un réseau invariant d'échelle sera plus dense qu'un réseau aléatoire de même taille.

Un réseau construit selon le modèle de Barabási-Albert de 250 noeuds est présenté à la figure 4 et sa distribution de degré ainsi que plusieurs renseignements pertinents sont présentés à la figure 5.

Ce réseau est très différent du réseau présenté à la figure 2. Tout d'abord, la grande majorité des noeuds possèdent une taille très petite, représentant leur degré associé, tandis que trois noeuds en particulier possèdent une taille beaucoup plus grande et sont plus foncés. Ce sont les pivots du réseau. La figure 5 présente également de façon très claire la distribution suivant une loi de puissance, tel que mentionné dans cette section. Il est à noter que le diamètre des deux réseaux est le même, mais cette caractéristique est due à la faible taille du réseau. Celle-ci a été gardée à 250 noeuds pour que les réseaux restent clairs lorsque mis en graphique.

Maintenant que les aspects techniques de la théorie des réseaux ont été abordées, il est temps de présenter l'analyse de réseaux réels étudiés dans le cadre du projet. Le premier réseau réel présenté est celui d'un cercle d'amis Facebook et est traité à la section suivante.

3 Réseau #1 : cercle d'amis Facebook

Le premier réseau réel étudié est un cercle d'amis Facebook, reliant l'hôte (l'auteur du présent texte) à tous ses amis, ainsi que ceux-ci entre eux si ils sont amis mutuels. L'analyse des réseaux sociaux date de longtemps, remontant aux travaux des psychologues américains Stanley Milgram et Jeffrey Travers au cours des années 1960 sur le phénomène des petits mondes [21]. Leur expérience consistait à sélectionner des gens au hasard dans l'état du Nebraska ainsi qu'à Boston et ils leur fallait transmettre une lettre, par l'intermédiaire de connaissances, à une personne cible au Massachusetts. Des 64 chaînes qui arrivèrent à la personne cible, 48% passèrent par 3 personnes avant de terminer leur course, et la longueur moyenne des chaînes était de 5.2, signe que le phénomène des petits mondes était bel et bien réel. Dans les années qui suivirent, plusieurs expériences semblables eurent lieu à plus grandes échelles, rapportant des conclusions similaires.

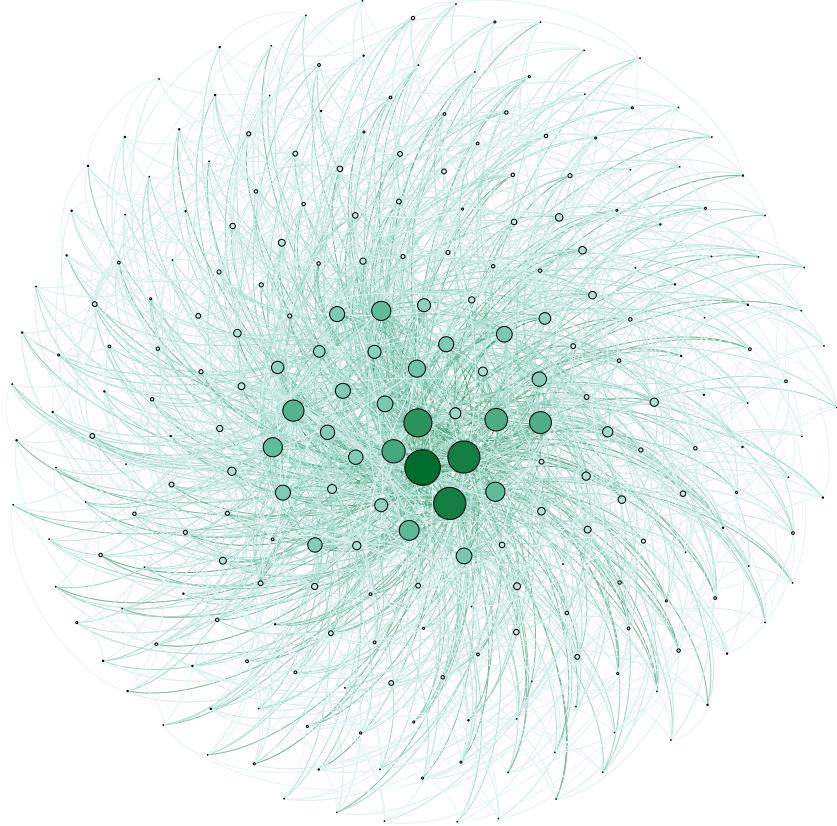


FIGURE 4 – Réseau invariant d'échelle avec $N = 250$ présentés en fonction de leur degré. Le réseau possède un nombre de liens $L = 1464$.

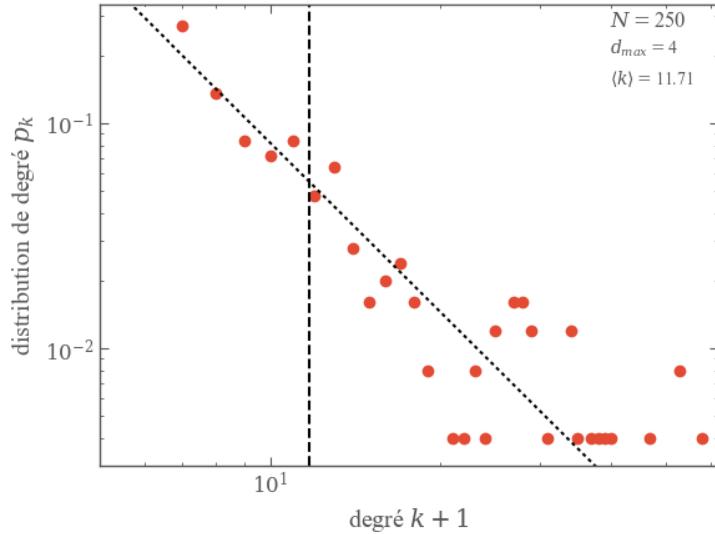


FIGURE 5 – Distribution de degré du réseau présenté ci-haut. La ligne verticale représente le degré moyen du réseau et la courbe pointillée représente la distribution de puissance associée à la distribution de degré, selon l'équation. Le coefficient γ est de 2.5.

Plus récemment, des chercheurs de Facebook ont conduit une expérience similaire via le réseau d'amis Facebook des utilisateurs de la plateforme et sont arrivés à la conclusion que parmi les 1.59 milliards d'utilisateurs de Facebook, une distance moyenne de 4.57 entre deux noeuds du réseau Facebook complet, représentant un *degré de séparation de 3.57* entre chaque personne, diminuant la précédente valeur obtenue de 3.74 cinq ans auparavant, alors qu'il n'y avait que la moitié des utilisateurs présents [8].

Depuis quelques années, Facebook a changé les permissions permettant d'accéder aux informations des comptes d'autrui, rendant la tâche de représenter son réseau d'amis de plus en plus difficile. Cependant, des développeurs de plus en plus astucieux réussissent à détourner ces nouvelles contraintes pour permettre l'exécution de la tâche. En date de la rédaction de ce projet, le programme utilisé pour modéliser le cercle d'amis Facebook d'un utilisateur donné est l'extension *Lost Circles*, disponible sur le fureteur *Chrome* [16]. Celle-ci retourne les données sous forme de fichier *.json*, facilement traitable avec un langage de programmation comme Python ou MATLAB. Le module NetworkX disponible sur Python permet facilement d'implémenter un réseau de toute sorte avec de nombreuses fonctions facilitant le traitement des données [20]. Finalement, le logiciel *Gephi* permet de représenter facilement un réseau construit à l'aide de NetworkX à l'aide de divers algorithmes de dessin basés sur des forces (*Force-directed algorithms*), permettant d'espacer les différents noeuds du réseau afin de les représenter de manière moins entassée [7].

3.1 Cercle d'amis Facebook avec hôte

Le premier réseau présenté comprends l'hôte, qui est connecté à chacun des noeuds du réseau. Celui-ci est présenté à la figure 6. Ce réseau social est composé de 707 noeuds, reliés entre eux par 26628 liens. Celui-ci est plutôt intéressant, étant séparé en deux entités très distinctes, représentant deux cercles d'amis avec peu de connections communes. En pratique, ces deux entités représentent les amis issus de la communauté universitaire de laquelle fait partie l'auteur (appelons celle-ci le regroupement A), ainsi que les amis issus de la ville natale de l'auteur (regroupement B). Les différentes couleurs de noeuds représentent des sous-groupes de noeuds plus connectés entre eux. Alors que le regroupement A est plutôt homogène, le regroupement B comporte lui-même quatre sous-groupes distincts, qui possèdent une plus grande connection entre eux qu'avec les autres groupes de couleur.

Puisque tous les noeuds du réseau sont directement connectés à l'hôte (grand noeud bleu), le diamètre du réseau est de deux, car si deux noeuds ne sont pas directement connectés entre eux, il le sont indirectement par l'entremise de l'hôte. Il est aussi facile de prédire que la distance moyenne entre deux noeuds de ce réseau sera située entre 1 et 2, à cause de cette propriété. De ce point de vue, ce réseau ne semble donc pas si intéressant. Cela dit, il est tout de même pertinent d'observer la distribution de degré de ce réseau pour voir si celui-ci est distribué selon une distribution mentionnée à la section 2. Celle-ci

est présentée à la figure 7.

Une distribution suivant une loi de puissance est observée pour les noeuds possédant un grand degré, comme prédit par le modèle de Barabási-Albert. Le coefficient γ est de 2.2 ± 0.10 pour ce réseau, ce qui est plutôt faible. La distance minimale moyenne obtenue pour le réseau est de 1.893, valeur se rapprochant grandement du seuil maximal de 2 pour ce réseau. Le degré moyen obtenu pour ce réseau est de 75.33, signe que la plupart des gens dans le réseau social possèdent un grand nombre d'amis en commun. Il était en effet envisageable que les noeuds possédant un petit degré ne soient pas très présents dans ce réseau social, car de nouvelles rencontres ont tendance à se faire par l'entremise d'amis communs. Il aurait donc été surprenant de voir un grand nombre de personnes posséder un petit degré. Finalement, cette dernière constatation est aussi visible dans le coefficient de regroupement moyen obtenu pour le réseau, qui est de 0.518, voulant dire que sur tous les noeuds adjacents à un noeud n , ceux-ci sont connectés ensemble avec une probabilité un peu plus grande que 0.5, signe que le réseau social est plutôt dense.

3.2 Cercle d'amis Facebook sans hôte

Il est intéressant de voir ce de quoi le réseau a l'air en enlevant le noeud principal, relié à chacun des noeuds. Ceci a pour conséquence que le diamètre du réseau et la distance moyenne entre les noeuds n'est plus triviale. Le réseau obtenu, contenant maintenant 704 noeuds et 26092 liens, est présenté à la figure 8 (les noeuds n'étant liés qu'à l'hôte ont aussi été omis pour que le graphe reste connexe). Il est possible de voir que les parties constituant le réseau sont à peu près les mêmes que celles d'avant, ce qui n'est pas du tout surprenant, puisque seulement le pivot principal a été enlevé du réseau et le réseau était déjà plutôt dense. Seule un petit groupe de noeuds (représentés en orange) semblent s'être détachés plus grandement du reste du réseau. Ceci est dû au fait que ce groupe n'est relié qu'à un noeud du regroupement B, alors qu'il l'était aussi avec l'hôte précédemment. Une comparaison des caractéristiques principales des deux réseaux est présentée au tableau 1 pour mieux pouvoir mettre en lumière ce que l'effet du retrait du noeud principal a sur la topologie du réseau.

Tout d'abord, il est à noter que le degré moyen du réseau a un peu changé suite au retrait du pivot principal, diminuant d'un peu plus de 1. En effet, le retrait d'un noeud étant relié à chaque autre noeud fait diminuer le degré de ceux-ci de 1 partout. De plus, le retrait de ce noeud, qui est le noeud possédant le plus grand degré, diminue encore la valeur moyenne.

Ensuite, la distance minimale moyenne $\langle d \rangle$ du graphique a augmenté d'environ 0.3, ce qui est un changement plutôt significatif. Le diamètre du réseau a lui aussi augmenté, passant de 2 à 6. Ces deux paramètres sont ceux étant les plus affectés par le retrait du noeud principal. Le chemin le plus court entre deux noeuds aléatoirement choisis ne peut plus passer par le noeud principal. Ce chemin ne peut

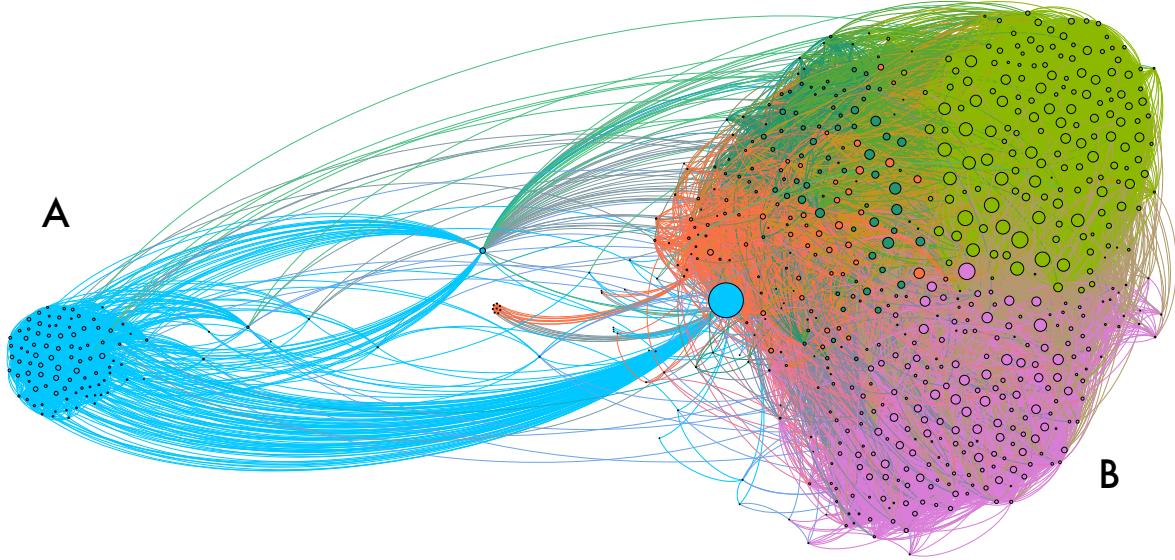


FIGURE 6 – Réseau d’amis Facebook comportant l’hôte du réseau (noeud le plus volumineux), relié à chacun des autres noeuds. Le réseau comporte deux entités distinctes (regroupement A et B), faiblement reliés entre eux.

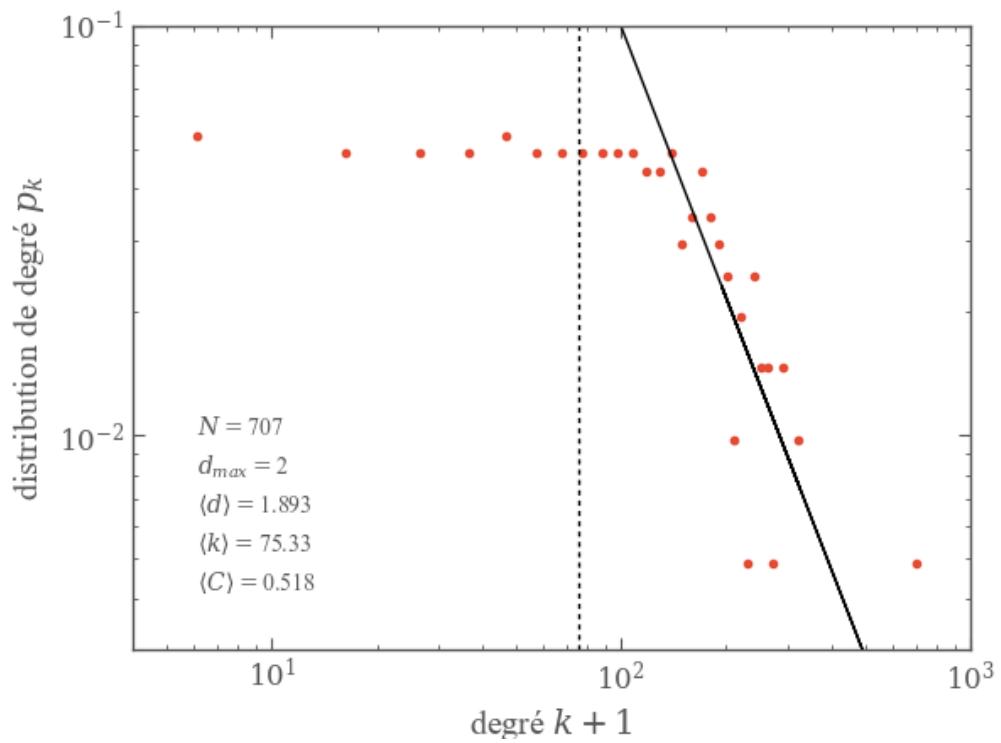


FIGURE 7 – Distribution de degré du réseau d’amis Facebook comportant l’hôte du réseau, relié à chacun des autres noeuds. Chaque point représente la probabilité de piger un noeud possédant un degré dans un intervalle de 10 de la position du point. Le réseau présente une distribution suivant une loi de puissance pour les noeuds possédant un grand degré, montrée par le trait plein. Le coefficient γ est de 2.2 ± 0.10 . Le trait pointillé montre le degré moyen du réseau.

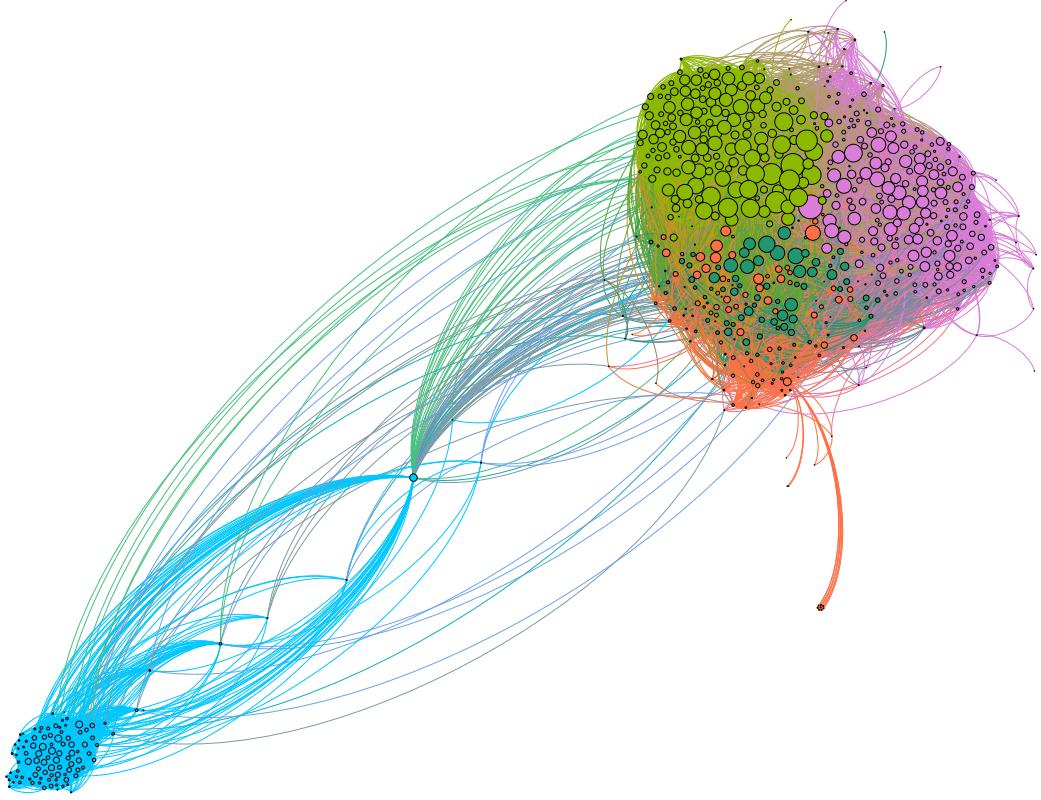


FIGURE 8 – Réseau d'amis Facebook auquel l'hôte a été soustrait. Le réseau comporte deux entités distinctes (regroupement A et B), faiblement reliés entre eux. L'allure globale du réseau est similaire à celle du réseau avec l'hôte encore présent (figure 6).

	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\langle C \rangle$	γ
Avec hôte	707	26628	75.33	1.893	2	0.512	2.20 ± 0.10
Sans hôte	704	26092	74.12	2.299	6	0.484	2.20 ± 0.10

TABLE 1 – Comparaison des caractéristiques principales des réseaux d'amis Facebook présentés aux figures 6 et 8.

donc qu'augmenter. Malgré ce changement, le phénomène des petits mondes est toujours observable dans le réseau. En effet, le nouveau diamètre du réseau est toujours très petit devant le nombre de noeuds total du réseau, malgré avoir triplé. De plus, le phénomène des très petits mondes (eq. 13) est encore plutôt présent, ayant comme distance minimale moyenne $\langle d \rangle = 2.299 \approx \ln \ln N$.

Le coefficient de regroupement moyen du réseau a diminué d'une petite quantité après le retrait de l'hôte. Cette variation était prévisible, puisque la présence du noeud principal augmentait inévitablement ce coefficient, étant relié à tous les noeuds du réseau. Cependant, même après son retrait, le coefficient de regroupement est toujours près de 0.5, signe que le réseau conserve sa densité de liens globalement. Finalement, le coefficient γ n'a pas changé après le retrait de ce noeud. Ce retrait représente la suppression du point dans le coin inférieur droit de la figure 7 (et seulement ce point), ce qui a peu d'influence sur la tendance de la distribution de degré globale.

Pour conclure l'analyse du réseau social Facebook conduite, il est intéressant de comparer les résultats obtenus à ceux obtenus Facebook Research sur l'ensemble du réseau social de ses utilisateurs [8]. Il a été mentionné plus tôt qu'une étude conduite en 2016 avait mené à la conclusion que la distance moyenne entre deux utilisateurs de Facebook choisis aléatoirement était de 4.57. Cette valeur est nettement supérieure à la valeur obtenue dans le réseau construit, et ce, dans les deux cas. Cependant, cette disparité semble être entièrement due à la taille du réseau étudié, puisque tous ces réseaux présentent les caractéristiques de très petits mondes. Conduire cette même étude sur un réseau Facebook plus volumineux, par exemple sur la page d'une personnalité publique ou d'un grand organisme, pourrait permettre de confirmer ces hypothèses.

4 Réseau #2 : cartographier le *World Wide Web*

Tel que mentionné à la section 2.3, la découverte que la distribution de degré d'un réseau réel se comportait comme une loi de puissance et non une loi de Poisson a été faite en cartographiant une partie du *World Wide Web* pour en étudier les propriétés. Une expérience similaire a ici été reproduite pour quelques domaines internet, soit le domaine `phys.umontreal.ca`, site du département de physique de l'Université de Montréal, ainsi que ceux de deux groupes de recherche associés au département, soit l'Institut de recherche sur les exoplanètes (`exoplanetes.umontreal.ca`) et le Groupe de physique des particules (`1ps.umontreal.ca`). Ces domaines ont été sondés à l'aide d'un *webcrawler*, robot se déplaçant sur la toile, effectuant une recherche en largueur du réseau en partant d'un site visité. Le réseau global, contenant tous les domaines, ainsi que les réseaux pour chacun de ceux-ci ont été étudiés séparément, de sorte à faire l'analyse de plusieurs tailles de réseaux internet et à identifier des propriétés émergentes des réseaux qui pourrait apparaître. Contrairement aux réseaux sociaux étudiés à la section précédente, les

réseaux du *World Wide Web* sont dirigés. De plus, certains fichiers, par exemple un fichier PDF vers lequel un lien pointe, auront un degré sortant étant nul. Ceci complique le calcul pour le diamètre du réseau (qui est en théorie infinie, puisqu'il est impossible de faire un chemin entre certains noeuds) ainsi que du coefficient de regroupement moyen, qui perd un peu de son sens dans un réseau dirigé. Ces caractéristiques ont donc été étudiées en prenant une version non-dirigée du réseau, pour tout de même recueillir certaines informations.

L'algorithme permettant de se déplacer sur la toile est un algorithme de parcours de graphe en largeur (*breadth-first search*). Il faut tout d'abord comprendre que la toile est un réseau ayant pour noeud des fichiers (le plus souvent, des fichiers HTML) étant reliés entre eux par des URLs (l'adresse de chaque fichier). L'algorithme consiste à partir d'un noeud et à sonder le code source de la page HTML, de sorte à trouver URLs pointant vers une page externe. Ces liens sont emmagasinés dans file, qui est une structure de données informatique linéaire fonctionnant selon le principe *First In, First Out*. Le premier élément inséré dans la file sera aussi le premier à en ressortir. Les éléments s'ajoutant ensuite vont «attendre leur tour» avant d'être visités. Il est possible de garder en mémoire une liste de tous les noeuds visités ainsi que leurs voisins pour construire le réseau voulu [11]. Le réseau est complètement parcouru lorsque tous les noeuds ont été visités et qu'il n'y a plus d'éléments en file.

Les représentations graphiques des réseaux présentent des noeuds de différente taille ainsi que diverses couleurs. La taille des noeuds représente le degré total (entrant et sortant) de ces noeuds. Les différentes couleurs représentent des communautés regroupés entre eux et séparés des autres. Les algorithmes permettant de séparer le réseau en communautés sont traités dans l'article de 2006 de Mark Newman [15].

4.1 Domaine *phys.umontreal.ca*

Le premier réseau étudié est celui du domaine *phys.umontreal.ca* sans l'inclusion des domaines des groupes de recherche. Ce réseau a été mis en graphique à la figure 9, et la distribution de degré associée à celui-ci est présentée à la figure 10. Ce réseau est constitué de 2943 noeuds, reliés entre eux par 194 276 liens, et possède donc un degré moyen $\langle k \rangle = 132.022$, selon l'équation 3. Les caractéristiques complètes du réseau sont présentées au tableau 2, comparant les quatre réseaux étudiés dans cette section.

La distribution de degré du réseau montre une tendance suivant une loi de puissance, comme attendu. Cependant, l'exposant de degré de la distribution est plutôt faible, soit de 1.5 ± 0.2 . Cette valeur indique que le réseau devrait se trouver en régime anormal et qu'un noeud devrait être connecté à l'ensemble du réseau. La figure 10 montre que le degré maximum possédé par un réseau est près de 3000, signe que c'est effectivement le cas. De plus, la figure montre que plus qu'un noeud possède cette caractéristique. Ceci est dû au fait que le site en question possède un menu, facilitant l'accès à certaines pages pour

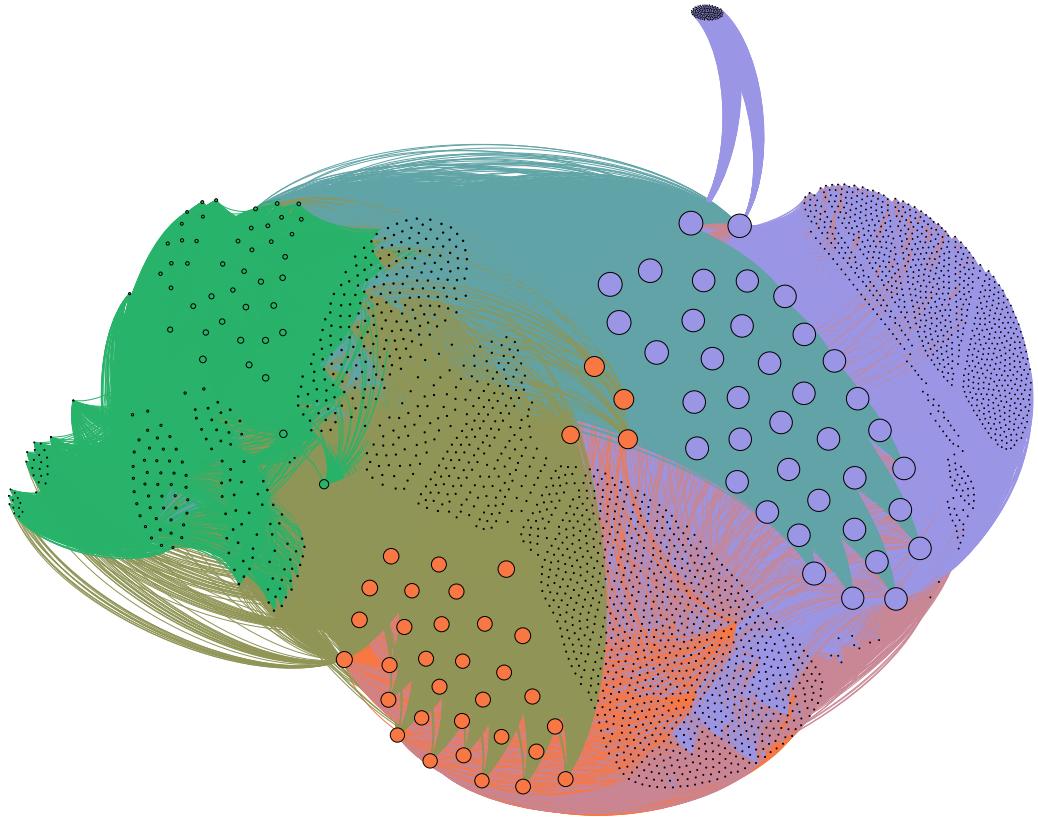


FIGURE 9 – Représentation du réseau du domaine *phys.umontreal.ca*. Les caractéristiques de ce réseau sont présentées au tableau 2.

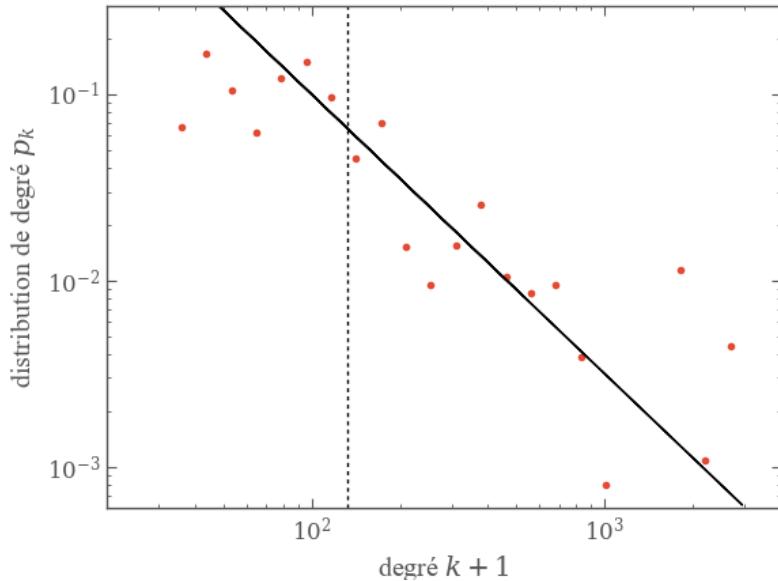


FIGURE 10 – Distribution de degré du réseau du domaine *phys.umontreal.ca* présenté à la figure 9. L'exposant de degré γ est de 1.5 ± 0.2 pour ce réseau. La ligne pointillée montre le degré moyen du réseau $\langle k \rangle = 132.022$.

l'utilisateur. Presque l'entièreté des pages alors visitées par le robot vont pointer vers ces pages. Celles-ci sont visibles directement sur la représentation graphique du réseau, où un grand nombre de noeuds violets de même taille sont tous amalgamés ensemble. Il en résulte que le diamètre du réseau n'est que de 4 et la distance moyenne séparant deux noeuds aléatoires est de 1.727, montrant clairement que seulement quelques noeuds permettent de connecter tous les autres ensemble.

La densité du réseau, soit la fraction du nombre total de liens dans le réseau sur le nombre maximal de liens ($L_{max} = N(N - 1)/2$) est de 0.022, signe que le réseau est plutôt épars, ce qui est habituellement le cas dans un réseau réel.

Le coefficient de regroupement moyen pour ce réseau est de 0.625. Cette valeur, plutôt élevée, est cohérente avec le phénomène des petits mondes observé dans ce réseau, et ce, dû au régime *anormal* dans lequel le réseau se trouve.

4.2 Domaine *exoplanetes.umontreal.ca*

Le second réseau présenté correspond à celui du domaine de l'institut de recherche sur les exoplanètes (IREX). Une représentation graphique de ce réseau est montrée à la figure 11 et la distribution de degré y étant associée est présentée à la figure 12. Ce réseau est beaucoup plus petit que le dernier réseau étudié, possédant environ la moitié du nombre de noeud ($N=1527$) et le quart du nombre de liens ($L=44367$) présents dans le réseau pour le domaine du département de physique. Le degré moyen du réseau est environ deux fois moins grand que ce dernier, avec $\langle k \rangle = 58.11$.

Cette fois-ci, à cause de la petite taille du réseau, la tendance suivant une loi de puissance est moins évidente que dans le cas précédent, semblant plus bruitée. Cela dit, il a tout de même été possible de tracer une droite montrant cette tendance. L'exposant de degré est identique à celui trouvé pour le réseau précédent, à 1.5 ± 0.2 , montrant encore une fois un réseau en régime anormal. Cette fois-ci, le diamètre du réseau est de 3 et la distance minimale moyenne entre deux noeuds est de 2.042. La densité de ce réseau est semblable à celle du réseau précédent, étant de 0.019. Le coefficient de regroupement moyen est $\langle C \rangle = 0.781$, ce qui est plus élevé que la valeur obtenue pour le réseau précédent. En somme, ces deux réseaux, malgré qu'ils possèdent différence importante en terme de taille, sont très semblables lorsque comparés caractéristique par caractéristique.

Le réseau associé au domaine de l'IREX possède cependant beaucoup moins de pivots beaucoup plus volumineux que les autres noeuds. La figure 11 montre que trois noeuds sont de vrais pivots dans ce réseau, alors qu'il y en avait plus d'une trentaine dans le réseau précédent.

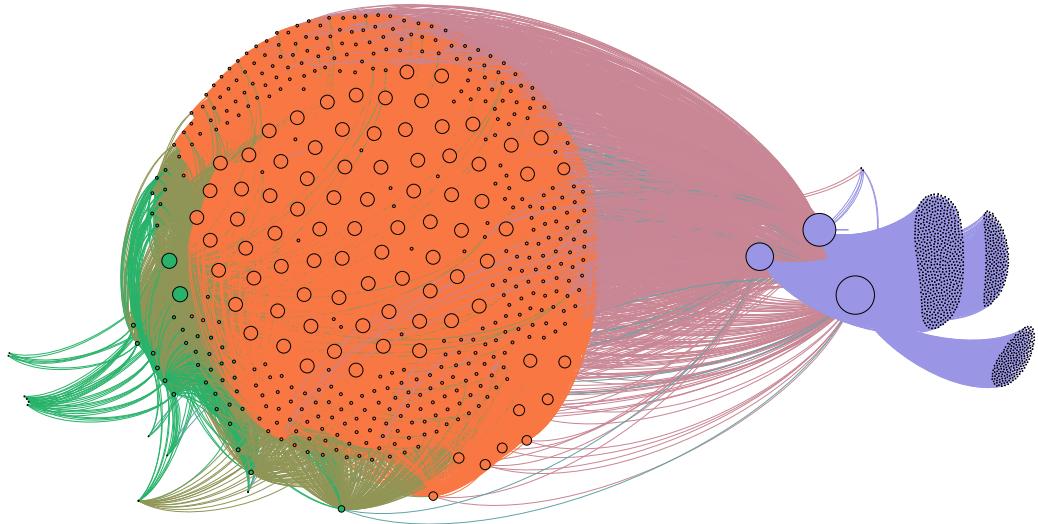


FIGURE 11 – Représentation du réseau du domaine exoplanetes.umontreal.ca. Les caractéristiques de ce réseau sont présentées au tableau 2.

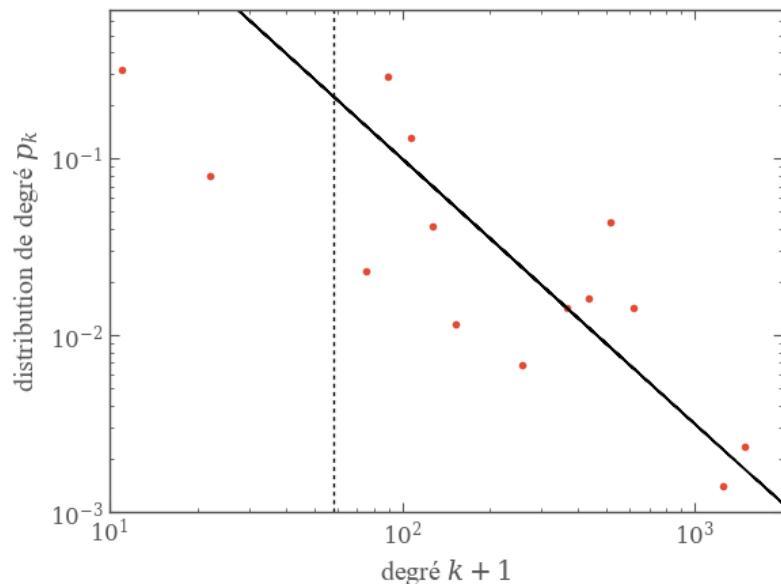


FIGURE 12 – Distribution de degré du réseau du domaine exoplanetes.umontreal.ca présenté à la figure 11. L'exposant de degré γ est de 1.5 ± 0.2 pour ce réseau. La ligne pointillée montre le degré moyen du réseau $\langle k \rangle = 58.11$.

4.3 Domaine *lps.umontreal.ca*

Le réseau représentant les connections entre les pages du site du groupe de recherche en physique des particules de l'Université de Montréal est le dernier réseau étudié avant que tous ceux-ci soient combinés pour l'étude du réseau complet. Ce réseau est constitué de 3595 noeuds, reliés entre eux par 97518 liens. Il s'agit donc d'un réseau de taille possédant un grand nombre de noeuds, mais un nombre relativement petit de liens. Une représentation graphique de ce réseau ainsi que sa distribution de degré associée sont présentés aux figure 13 et 14.

Le diamètre du réseau obtenu est de 4, ce qui est semblable aux valeurs trouvées précédemment, et la distance minimale moyenne entre deux noeuds aléatoirement choisis est de 1.809, aussi semblable aux valeurs précédemment trouvées. Ce réseau est cependant environ deux fois moins dense que les autres, possédant une densité de 0.008. Le coefficient de regroupement moyen du réseau est de 0.776, ce qui signifie que les noeuds voisins possèdent beaucoup de noeuds voisins en commun. La distribution de degré suivant une loi de puissance est ici très évidente. L'exposant de degré trouvé pour le réseau est de 1.65 ± 0.05 , signe que le réseau se trouve toujours en régime anormal, comme ce qui avait été obtenu précédemment.

L'entièreté des réseaux étudiés jusqu'à présent possèdent tous des caractéristiques très similaires, signe que des réseaux du *World Wide Web* de cette taille possèdent souvent des caractéristiques de ce genre. Le petit nombre de noeuds est peut-être la cause du régime anormal dans lequel tous ces réseaux se trouvent. Il est aussi à noter qu'aucun de ces réseaux ne possède une distance minimale moyenne arrivant près de la valeur estimée par Albert *et. al.* selon l'équation 10, et aucun des réseaux ne possède d'exposant de degré se rapprochant de 2.1 et 2.45, valeurs trouvées par les mêmes auteurs [2].

4.4 Réseau incluant les trois domaines

Il vient maintenant le temps de combiner les trois réseaux préalablement analysés en un grand réseau. Celui-ci compte 8068 noeuds étant reliés par 336 465 liens entre eux. Le degré moyen des noeuds de ce réseau est de 83.407, ce qui se trouve entre les valeurs précédemment obtenues pour les autres réseaux, résultat attendu. Les figures 15 et 16 montrent une représentation graphique de ce réseau, ainsi que la distribution de degré lui étant associée.

Le nouveau réseau créé possède un diamètre de 4, identique à celui de deux autres réseaux le composant, signe que ces domaines sont plutôt bien reliés entre eux. La distance minimale moyenne entre deux noeuds est de 2.052 pour ce réseau, étant donc légèrement supérieure à celle obtenue pour trois réseaux précédemment analysés. Celle-ci encore une fois nettement inférieure à la valeur attendue $\langle d \rangle_{expected} \approx 8.41$ prédite par l'équation 10. Cela dit, le réseau est encore beaucoup plus petit que ceux utilisés pour la rédaction de l'article mentionnant cette loi empirique, utilisant un réseau possédant jusqu'à 350 000 noeuds

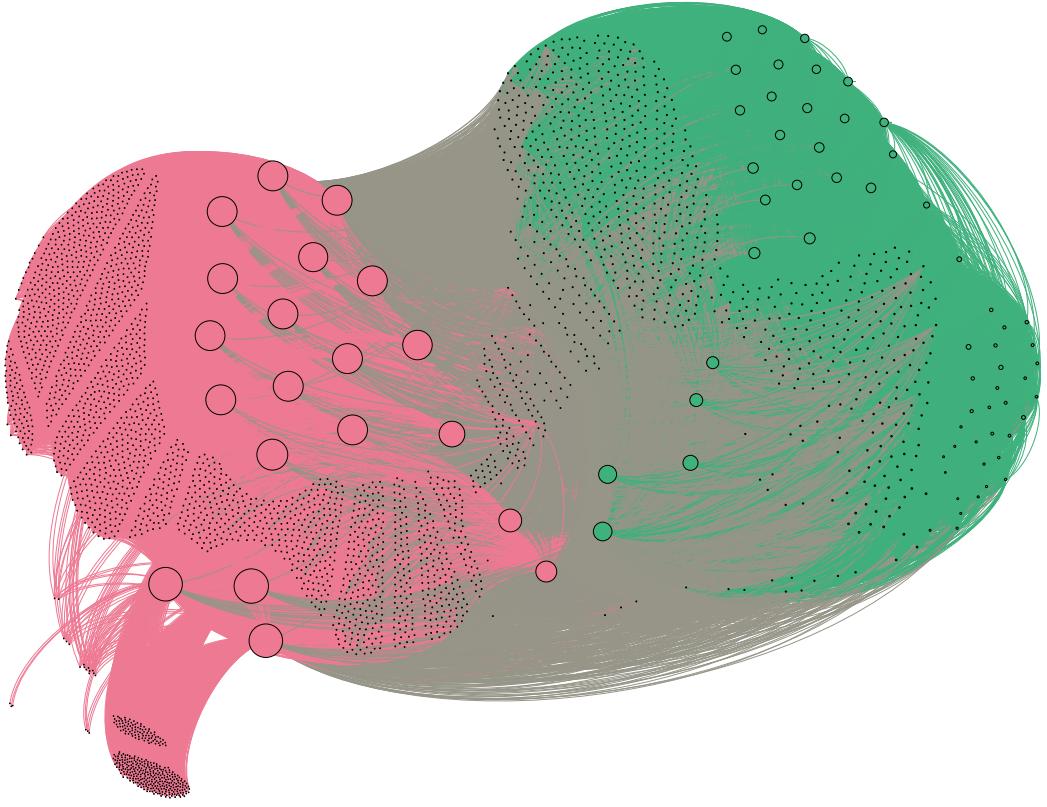


FIGURE 13 – Représentation du réseau du domaine *lps.umontreal.ca*. Les caractéristiques de ce réseau sont présentées au tableau 2.

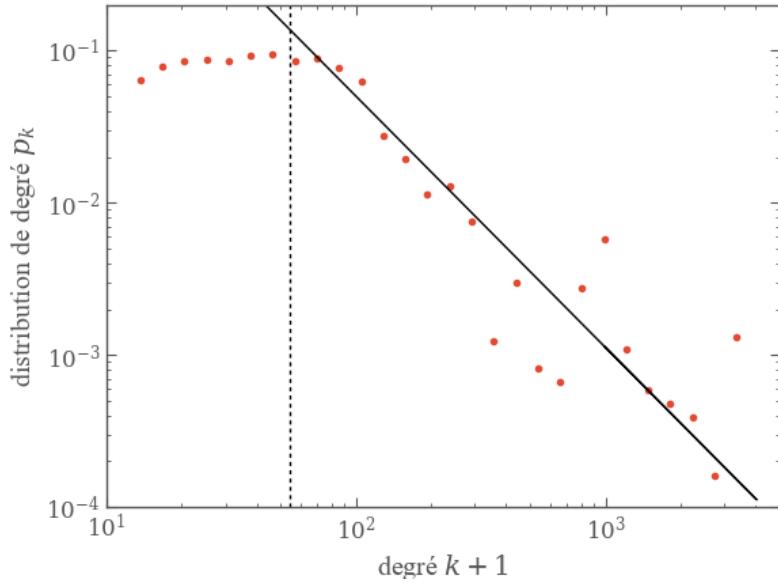


FIGURE 14 – Distribution de degré du réseau du domaine *lps.umontreal.ca* présenté à la figure 13. L'exposant de degré γ est de 1.65 ± 0.05 pour ce réseau. La ligne pointillée montre le degré moyen du réseau $\langle k \rangle = 54.252$.

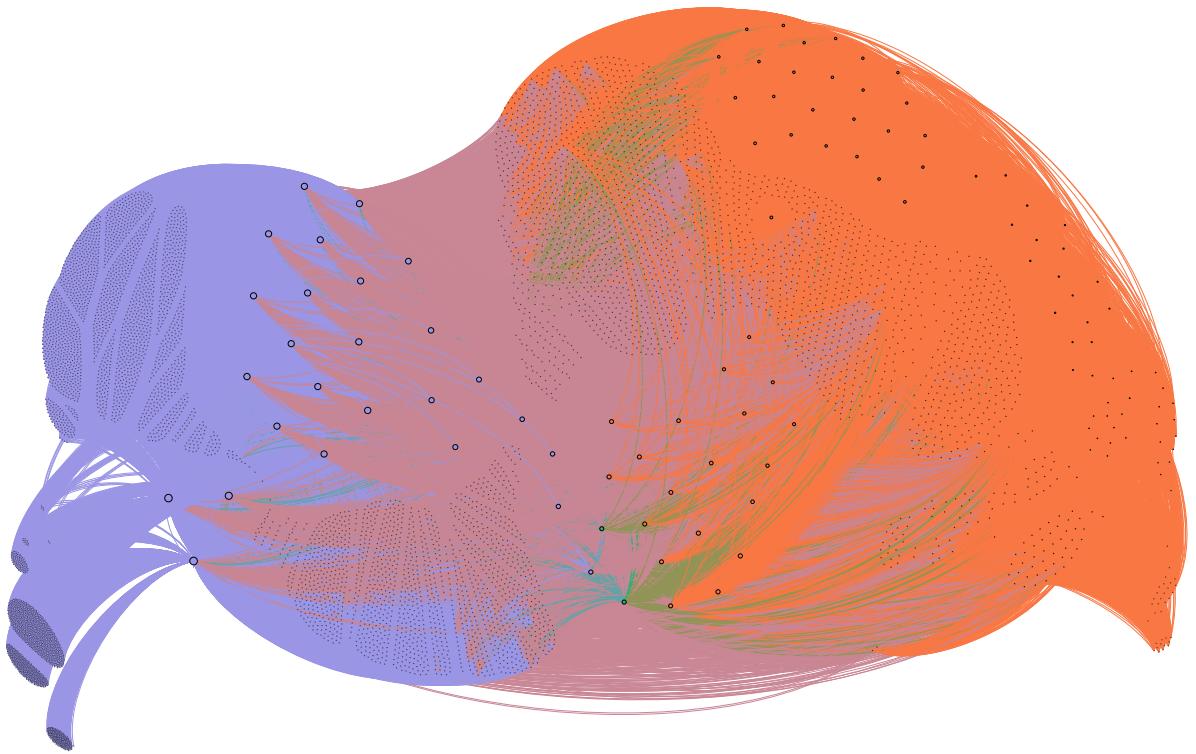


FIGURE 15 – Représentation du réseau comprenant les trois domaines mappés. Les caractéristiques de ce réseau sont présentées au tableau 2.

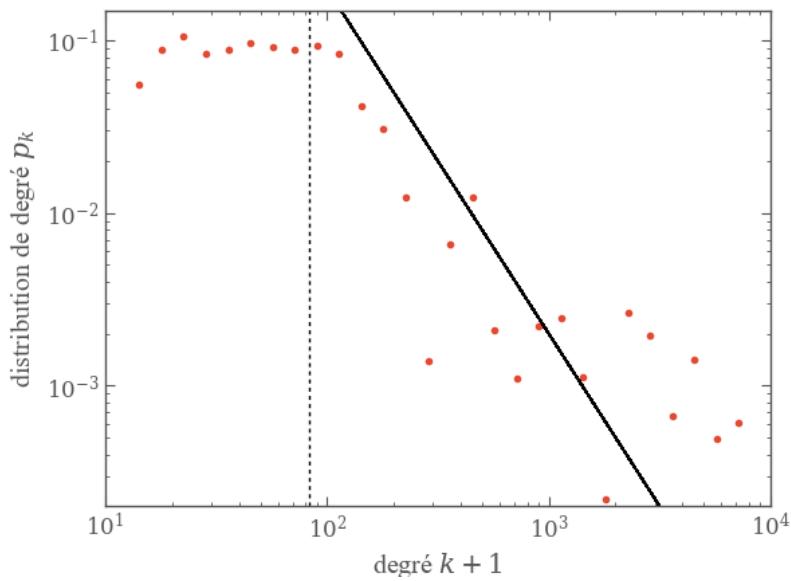


FIGURE 16 – Distribution de degré du réseau comprenant les trois domaines mappés présenté à la figure 15. L'exposant de degré γ est de 2.0 ± 0.2 pour ce réseau. La ligne pointillée montre le degré moyen du réseau $\langle k \rangle = 83.407$.

et 1 500 000 liens. La distance minimale moyenne trouvée est cependant très près de la valeur prédictée par un réseau en régime invariant d'échelle, prédisant une valeur se rapprochant de $\langle d \rangle_{pow} = \ln \ln N \approx 2.20$.

Le coefficient de regroupement moyen pour ce réseau est de 0.685, ce qui est encore une fois relativement élevé. Comme pour les réseaux précédents, le phénomène des petits mondes est observé dans ce réseau plus volumineux. Il se retrouve à la limite du régime anormal et du régime invariant d'échelle, avec un exposant de degré $\gamma=2.0 \pm 0.2$.

Étrangement, l'algorithme permettant de calculer la modularité du réseau ne détecte que deux communautés distinctes, malgré que ce réseau soit le fruit de trois réseaux combinés. Seulement deux couleurs distinctes sont présentes sur le réseau, soit les noeuds appartenant à la communauté "orange" et les autres appartenant à la communauté "mauve". La partie "saumon" du graphique ne représente que les liens liant les deux communautés. De plus, les composantes distinctes étant présentes dans les autres réseaux semblent avoir disparus s'étant entremêlées entre parties pour ne créer que deux groupes distincts. Une excroissance reliée à trois pivots est présente sur la figure 15. Une telle excroissance était aussi présente sur les réseaux pour le domaine de l'IREX, ainsi que le domaine du groupe de physique des particules (voir figures 11 et 13). Celle aperçue est cependant due au second. En effet, la partie de gauche du réseau présenté correspond au réseau du groupe de physique des particules. Les noeuds centraux orangés représentent la partie du réseau reliés au département de physique. Leur position dans le graphe est naturelle, puisqu'il s'agit d'une partie centrale, étant reliée aux divers groupes de recherches associés au département. Finalement, la partie supérieure droite du réseau est associée à l'institut de recherche sur les exoplanètes, possédant le regroupement de noeuds le moins volumineux des trois regroupements principaux. Celle-ci est plus éloignée du groupe de physique des particules, ayant peu de pages en commun mis-à-part des liens via les pages du département de physique.

Les caractéristiques principales des différents réseaux sont présentés de façon plus claire dans le tableau 2.

	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\langle C \rangle$	γ
phys.umontreal.ca	2943	194 278	132.022	1.727	4	0.625	1.5±0.2
exoplanetes.umontreal.ca	1527	44 367	58.110	2.042	3	0.781	1.5±0.2
lps.umontreal.ca	3595	97 518	54.252	1.809	4	0.766	1.65±0.05
Global	8068	336 465	83.407	2.052	4	0.685	2.0±0.2

TABLE 2 – Comparaison des caractéristiques principales des réseaux du *World Wide Web* présentés aux figures 9,11, 13 et 15.

En somme, le réseau final incluant les trois plus petits réseaux n'est toujours pas assez grand pour

montrer des caractéristiques bien différentes de ces derniers. Mis à part est au moins deux fois plus grand que tous les autres réseaux, celui-ci ne présente pas de caractéristique bien différentes que les autres. Tous les réseaux étudiés se retrouvent soit dans le régime anormal, soit à la frontière entre le régime anormal et le régime invariant d'échelle, faisant qu'ils diffèrent peu au niveau du diamètre d_{max} et de la distance minimale moyenne $\langle d \rangle$. Le coefficient de regroupement $\langle C \rangle$ est le seul paramètre ayant changé de façon remarquable, mais celui-ci ne semble pas être affecté par la taille du réseau de façon générale. L'exposant de degré γ semble cependant présenter une croissance en fonction de la taille du réseau, signe que l'étude d'un réseau beaucoup plus volumineux permettrait de mettre en lumière les caractéristiques globales d'un réseau aussi vaste que le *World Wide Web*.

5 Comparaison des réseaux étudiés avec d'autres réseaux réels

Les deux types de réseaux étudiés, soit un réseau social et un réseau de pages web, sont très différents l'un de l'autres. Tout d'abord, le premier est un réseau non-dirigé, puisque deux personnes sont amis entre eux, et le second est dirigé, puisqu'une page peut pointer vers une autre sans que l'autre face de même. L'écart de taille entre les deux types de réseaux étudiés est très grand, le réseau de pages web étant plus de dix fois plus volumineux que le réseau d'amis Facebook, dans un certain cas.

Cela dit ces types de réseaux sont tout de même très semblables au niveau de leurs caractéristiques principales. Dans tous les cas étudiés, le degré moyen est élevé, étant autour de 75 pour les réseaux sociaux et allant de 54 à 132 dans l'autre cas. Le degré moyen trouvé pour le réseau de pages web est plutôt impressionnant, ne s'approchant pas du tout des valeurs obtenues par Albert *et. al.*, qui sont plutôt autour de 4.6 [2]. Cependant, cette grande disparité semble être le fruit de la présence de menus sur les pages web, pointant systématiquement vers un grand nombre d'autres pages. La distance moyenne minimale entre deux noeuds choisis aléatoirement est aussi une caractéristique très similaire partagée par les deux types de réseaux, se trouvant autour de 2 pour tous les réseaux étudiés. Cette ressemblance semble être due à la taille limitée des réseaux étudiés, puisque tous ceux-ci se trouvaient à la limite entre le régime anormal et le régime suivant une loi de puissance. Le diamètre des réseaux en témoigne par sa faible valeur. Le seul réseau possédant un diamètre plus grand que 5 était le plus petit, ce qui est plutôt inhabituel. Celui-ci possède aussi la plus grande distance minimale moyenne, caractéristique normalement gouvernée par le nombre de noeuds présents dans le réseau. Pour chacun des réseaux étudiés, l'exposant de degré était très similaire, et ce, dû à la taille limitée des échantillons.

Il est intéressant de comparer ces deux types de réseaux avec d'autres réseaux retrouvés dans la nature. Deux réseaux historiquement importants seront ici abordés, soit le réseau de neurones du ver *Caenorhabditis elegans* (*C. elegans*), le premier réseau de neurones modélisé cartographié dans son entiereté, ainsi

que le réseau d'acteurs d'Hollywood, notamment connu d'après les six degrés de Kevin Bacon.

5.1 Réseau de neurones du ver *C. elegans*

L'intérêt pour la théorie des réseaux en neuroscience est capitale. Le cerveau humain est l'un des réseaux les moins bien compris des réseaux réels. Celui-ci est composé d'une centaine de milliards de neurones connectées entre elles, constituant donc un réseau de très grande taille [12]. C'est cette taille énorme qui rend la compréhension du cerveau si ardu. Le premier cerveau complètement cartographié est celui d'un ver nommé *C. elegans* (ce diagramme de connection est appelé *connectome*). Celui-ci est composé de seulement 302 neurones connectées entre elles par 6393 connections [22]. Une reproduction graphique du connectome de *C. elegans* est présenté à la figure 17.

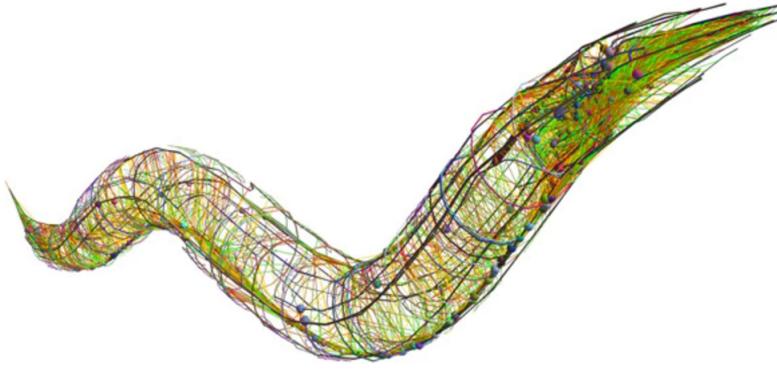


FIGURE 17 – Connectome du vers *C. elegans*, constitué de 302 neurones, connectées entre elles par 6393 connections [13].

En 2012, l'équipe de Dimitri Chklovskii a étudié en fond et en comble le réseau de neurones de *C. elegans* et a trouvé plusieurs paramètres d'intérêt pour ce réseau, présentés au tableau 3.

	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\langle C \rangle$	γ
<i>C. elegans</i>	302	6393	3.68	≈ 3	4	0.21	3.14 ± 0.13

TABLE 3 – Caractéristiques principales du réseau neuronal du ver *C. elegans* tel que découvert Varshney *et al.* [22].

Malgré sa petite taille, ce réseau présente tout de même une invariance d'échelle, avec un exposant de degré de 3.14 ± 0.13 . Cette valeur n'est pas non plus à la limite du régime anormal comme elle l'était pour les autres réseaux étudiés. Les noeuds de ce réseau possèdent des degrés variant de 1 à 40, avec un degré moyen de 3.68. Ce réseau présente aussi un phénomène des petits mondes, avec un diamètre de réseau égal à 4 et une distance minimale moyenne d'approximativement 3. Ces valeurs sont très semblables aux valeurs trouvées pour le réseau social ainsi que pour les domaines internet, malgré que la taille de ceux-ci

était largement supérieure au réseau de neurones de *C. elegans*. Finalement, le coefficient de regroupement moyen pour ce réseau est nettement inférieur à celui trouvé pour les réseaux étudiés, étant de 3 à 4 fois inférieur à ceux-ci, signe que le réseau est plutôt épars et que peu de neurones voisins sont connectés. Les valeurs trouvées plutôt semblent donc plutôt biaisées par un grand nombre de noeuds ne possédant qu'un lien connecté à un autre, possédant donc un coefficient de regroupement de 1, ce qui fait hausser la moyenne.

Même si la disparité entre la taille du réseau neuronal de *C. elegans* et les réseaux étudiés aux sections 3 et 4 sont grandes, ceux-ci présentent tout de même chacun une distribution suivant une loi de puissance ainsi qu'un phénomène des petits mondes.

5.2 Réseau d'acteurs d'Hollywood

Le réseau d'acteurs d'Hollywood est un réseau social de grande envergure, connectant 4.6 millions d'acteurs, reliés entre eux par 1.9 millions de films et de séries télé (on pose que deux acteurs sont voisins si ils ont déjà travaillé ensemble sur un projet). Il est possible, grâce à un *crawler* et à l'algorithme BFS, de trouver les différentes caractéristiques de ce réseau.

Une alternative amusante, populaire à la fin des années 1990, consiste à prendre comme centre du réseau l'acteur Kevin Bacon et trouver comment connecter un acteur au centre. La distance entre ces noeuds est appelée le nombre de Bacon. Avec Kevin Bacon au centre du réseau, la distance moyenne le séparant des 2 283 910 autres acteurs de Hollywood pour lesquels il est possible de se connecter (nombre moyen de Bacon) est de 3.024 et la distance maximale séparant un acteur du centre est 10 (seulement 7 acteurs se trouvent à cette distance). Il se trouve que le centre du réseau possédant une distance moyenne la plus courte entre tous les autres noeuds est l'acteur Eric Roberts, possédant une distance moyenne le séparant des autres acteurs de 2.82455. Quant à l'acteur le plus obscur d'Hollywood, sa distance moyenne est de 11.079 ! [17].

Un sous-ensemble de ce réseau massif a été analysé par Albert et Barabási en 1999 sur un échantillon de 212 250 noeuds. Cet échantillon, conjointement avec la mappe du domaine de l'université Notre Dame, est à l'origine de la découverte de l'invariance d'échelle dans les réseaux réels [5]. Plus récemment, la même analyse a été effectuée sur un échantillon de 702 388, étant donc plus de trois fois plus grand que l'échantillon original. Les caractéristiques de ce réseau sont présentées au tableau 4. Plusieurs caractéristiques ne sont malheureusement pas disponibles.

Même si ce sous-ensemble du réseau des acteurs d'Hollywood est nettement plus petit que le réseau dans son ensemble, il est tout de même beaucoup plus volumineux que les réseaux étudiés aux sections 3 et 4. Le réseau possède un degré moyen de 83.71, très semblable aux valeurs trouvées pour le réseau

	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\langle C \rangle$	γ
Acteurs	702 388	29 397 908	83.71	?	?	?	2.12

TABLE 4 – Caractéristiques principales du réseau d’acteurs de Hollywood tel que trouvé par Albert et Barabási [4].

d’amis Facebook ainsi que pour l’échantillon regroupant tous les domaines internet mappés. L’exposant de degré γ de 2.12 est aussi très rapproché des valeurs trouvées, se trouvant dans l’intervalle de confiance de celles-ci.

Malgré que la distance moyenne ne soit pas disponible pour ce réseau, le nombre moyen de Bacon de 3.024 peut porter à croire qu’un tel réseau présente aussi le phénomène des petits mondes, comme c’était le cas avec les réseaux étudiés.

6 Conclusion

En somme, les réseaux réels étudiés, soit un cercle d’amis Facebook ainsi qu’une mappe d’un domaine du *World Wide Web*, montrent plusieurs similitudes avec des réseaux réels précédemment analysés dans la littérature. Tout d’abord, les deux réseaux étudiés présentent une distribution de degré suivant une loi de puissance, caractéristique mis en lumière par Barabási et son équipe à la fin des années 1990. Ces deux réseaux montrent aussi le phénomène des petits mondes, stipulant que deux noeuds aléatoirement choisis dans un réseau sont séparés par une petite distance, phénomène aussi omniprésent dans les réseaux réels. Une comparaison de ces deux réseaux avec le réseau de neurones du ver *C. elegans* ainsi que le réseau d’acteurs de Hollywood a aussi présenté plusieurs similarités.

Plusieurs efforts sont présentement mis pour l’élaboration du connectome de plusieurs mammifères ainsi que des humains. La science des réseaux pourrait donc peut-être permettre la compréhension de ces gigantesque réseaux et même permettre la compréhension de l’intelligence. Certaines études du transport des signaux dans le cerveau ont déjà montré que ceux-ci suivaient une loi de puissance [9]. D’autres études ont suggéré que l’entraînement d’un réseau de neurones artificiels se faisait plus rapidement lorsque celui-ci présentait le phénomène des petits mondes [18]. Il serait intéressant d’appliquer le modèle de Barabási-Albert à ces réseaux de neurones artificiels pour voir comment cette structure influence l’apprentissage du réseau.

Bibliographie

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1) :47, 2002.
- [2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet : Diameter of the world-wide web. *nature*, 401(6749) :130–131, 1999.
- [3] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 371(1987) :20120375, 2013.
- [4] Albert-László Barabási. *Network science*. Cambridge University Press, 2016.
- [5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439) :509–512, 1999.
- [6] Phillip EC Compeau, Pavel A Pevzner, and Glenn Tesler. How to apply de bruijn graphs to genome assembly. *Nature biotechnology*, 29(11) :987–991, 2011.
- [7] The Gephi Consortium. Gephi. <https://gephi.org/>. Accessed : 2016-04-17.
- [8] Sergey Edunov, Carlos Diuk, Ismail Onur Filiz, Smriti Bhagat, and Moira Burke. Three and a half degrees of separation. *Research at Facebook*, 2016.
- [9] Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1) :018102, 2005.
- [10] Paul Erdős and Alfréd Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6 :290–297, 1959.
- [11] Michael T Goodrich, Roberto Tamassia, and Michael H Goldwasser. *Data structures and algorithms in Python*. John Wiley & Sons Ltd, 2013.
- [12] Suzana Herculano-Houzel. The human brain in numbers : a linearly scaled-up primate brain. *Frontiers in human neuroscience*, 3 :31, 2009.
- [13] Ferris Jabr. The connectome debate : Is mapping the mind of a worm worth it ?, 2012.
- [14] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2) :167–256, 2003.
- [15] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23) :8577–8582, 2006.
- [16] Arlind Nocaj, Jan Hildenbrand, Felix Schönenberger, and Alban Nocaj. Lost circles. <https://lostcircles.com/>. Accessed : 2016-04-17.

- [17] Patrick Reynolds. The oracle of bacon. <https://oracleofbacon.org/help.php>, 2016. Accessed : 2016-04-27.
- [18] D Simard, L Nadeau, and H Kröger. Fastest learning in small-world neural networks. *Physics Letters A*, 336(1) :8–15, 2005.
- [19] Annamaria Talas. Connected : The power of six degrees, 2008.
- [20] Benjamin Edwards *et. al.* Networkx. <https://networkx.github.io/>. Accessed : 2016-04-17.
- [21] Jeffrey Travers and Stanley Milgram. The small world problem. *Psychology Today*, 1 :61–67, 1967.
- [22] Lav R Varshney, Beth L Chen, Eric Paniagua, David H Hall, and Dmitri B Chklovskii. Structural properties of the *caenorhabditis elegans* neuronal network. *PLoS Comput Biol*, 7(2) :e1001066, 2011.
- [23] Duncan J Watts. *Six degrees : The science of a connected age*. WW Norton & Company, 2004.
- [24] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684) :440–442, 1998.