Analyse de réseaux complexes réels via des méthodes issues de la matrice de Google

Célestin Coquidé

THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE FRANCHE-COMTÉ

École doctorale n°553

Carnot-Pasteur

Doctorat de Physique

Par

M. Célestin Coquidé

Analyse de réseaux complexes réels via des méthodes issues de la matrice de Google

Thèse présentée et soutenue à l'Observatoire de Besançon, le 23 novembre 2020

Composition du Jury :

M. Borgnat Pierre	Directeur de Recherche CNRS École Normale Supérieure de Lyon	Président
M. Frahm Klaus	Professeur Université Paul Sabatier, Toulouse	Rapporteur
Mme Jaffrès-Runser Katia	Maître de Conférences Institut National Polytechnique de Toulouse	Rapporteuse
M. Joubert Pierre	Professeur Université Bourgogne Franche-Comté, Besançon	Codirecteur de thèse
M. Lages José	Maître de Conférences Université Bourgogne Franche-Comté, Besançon	Directeur de thèse
M. Queyroi François	Chargé de Recherche CNRS Université de Nantes	Examinateur

Remerciements

Trois années se sont écoulées depuis mon arrivée à Besançon et depuis le début de mon doctorat. Je remercie Pierre Borgnat d'avoir présidé mon jury de thèse, ainsi que tous les membres de ce jury pour leur attention et leurs remarques. Je remercie bien évidemment mes directeurs de thèse José Lages et Pierre Joubert pour leur soutien et leur aide. Je tiens à remercier Leonardo Ermann, Włodzimierz Lewoniewski, Klaus Frahm et Dima Shepelyansky, avec qui j'ai partagé d'intéressantes discussions et de riches collaborations scientifiques. Je suis friand d'une atmosphère chaleureuse, conviviale et pleine de vie au travail, je dois avouer que je n'ai pas été déçu de ce côté-là, le laboratoire UTINAM est vraiment exceptionnel. Je remercie tout le personnel et les chercheurs avec qui j'ai passé des moments aussi agréables qu'enrichissants. Je remercie aussi les deux Guillaume du laboratoire pour les discussions et l'aide qu'ils m'ont apportées. Ce n'est pas simple d'arriver dans une toute nouvelle ville et de ne connaître personne, pour cette raison, je voudrais remercier mes amis bisontins qui m'ont soutenu et m'ont fait découvrir la ville, et surtout, m'ont permis de me sentir chez moi. Un grand merci à Mikaël et à Cécile, ainsi qu'à mes compères de jeu (*Magic The Gathering*). Mon intérêt pour la science et plus particulièrement la physique vient d'une personne que je me dois de remercier : Loïc qui, malheureusement, succomba d'un cancer au tout début de mes années universitaires. Parfois, la famille et les amis sont indissociables, j'ai une pensée toute particulière pour Kurt-William, ainsi que Lynda, la reine du maquillage, qui m'ont aidé dans mes réflexions et dans la rédaction de ma thèse, à Darius mon compagnon de parkour et de chasse aux zombies, à Ysaline pour son soutien et à Remedios pour ses services de couturière et enfin à Marion qui m'a soutenu pendant la dernère ligne droite de mon doctorat. Je remercie mon père d'avoir été présent pour mon emménagement et d'avoir fait en sorte que je me sente au mieux chez moi. Je remercie ma mère pour sa confiance sans faille à mon égard. Je remercie les membres de ma famille qui m'ont demandé d'expliquer ce que je faisais pendant ma thèse, grâce à eux, j'ai pu prendre du recul et améliorer mes compétences en vulgarisation scientifique. Merci à mes parents, Christian et Christine, à mes sœurs, Amélie, Jennifer et Justine, à mes frères, Corentin et Raphaël, à ma belle-mère Catherine, merci aux cousines, aux cousins, aux oncles, aux tantes, à mes deux grands-mères, Jacqueline et Hélène, à mes nièces, Chloé, Rose, Youna, et Lucile la petite dernière. Enfin, j'ai une pensée pour Alain, mon défunt grand-père.

J'ai passé trois années vraiment formidables, j'ai adoré travailler au sein du laboratoire UTINAM et j'aimerais continuer dans la recherche. Je ne peux pas terminer ces remerciements sans vous parler de mon amour pour la nourriture, et pour ça, Besançon est vraiment une très bonne ville !

Table des matières

Chapitre 1

Introduction

Le terme réseau a une origine latine *retis* qui a donné naissance aux adjectifs réticulé et réticulaire, désignant un objet fait de filets. Cette idée se trouve également dans la définition anglaise *network*. La notion de réseau est utilisée dans des domaines scientifiques divers et variés et nous l'utilisons tous dans la vie de tous les jours sans forcément y penser. En sociologie, on parle de sociogramme, de réseau social, en biologie les interactions entre les protéines sont représentées sous forme de réseau, les réseaux de neurones artificiels sont à la base même du machine learning. Nous pouvons définir le réseau traduisant les interactions sociales d'un individu, où chaque personne de son entourage constitue un nœud du réseau et les interactions sociales des uns avec les autres forment les filaments de cette toile. Dans la société du numérique d'aujourd'hui, un grand nombre de personnes utilisent les réseaux sociaux tels que : Facebook, Twitter ou bien encore Instagram pour ne citer que les plus connus. Nous faisons intrinsèquement partie de plusieurs réseaux qui nous permettent de trouver un travail, rencontrer des nouvelles personnes, acheter un bien, ou même faire garder ses animaux lorsqu'on part en vacances.

Historiquement, le terme graphe, utilisé en mathématique, est apparu avant celui de réseau. L'un des fondateurs de la théorie des graphes, le mathématicien et physicien Leonhard Euler étudia, en 1736, la ville de Königsberg en intégrant le concept de graphe [1]. Dans le formalisme de la théorie des graphes, on note $G(V, E)$ un graphe G constitué d'un ensemble de sommets V interconnectés par un ensemble d'arêtes E. Dans son étude, il construit le graphe de la ville de Königsberg composée de deux îles centrales autour desquelles se trouve le reste de la ville. Chaque île, ainsi que les parties nord et sud de la ville, sont les sommets du graphe. Les sept ponts permettant de passer d'une zone à une autre, sont alors les arêtes de ce graphe. La question est alors de savoir s'il est possible de trouver un chemin nous permettant de visiter tous les lieux tout en empruntant une seule fois chaque pont [1]. Plus tard, au XX$^{\text{ème}}$ siècle, les communautés de physiciens et de biologistes préféreront utiliser le terme réseau. Tout au long de ce travail de thèse de doctorat, nous utiliserons les termes associés aux réseaux, à savoir : les sommets sont des nœuds et les arêtes sont des liens.

Il existe plusieurs types de réseaux dont les réseaux aléatoires, les réseaux invariants d'échelle, les arbres, les réseaux bipartis et aussi les réseaux multiplexes pour n'en citer que quelques uns. Dans chaque cas, les liens peuvent être dirigés, non dirigés, ou bien pondérés. Dans la suite de cette introduction, nous allons voir les définitions et représentations de ces différents réseaux ainsi que leurs caractéristiques. Ensuite, nous verrons un exemple de réseau réel, le World Wide Web (WWW), et nous présenterons le moteur de recherche Google permettant de classer les pages du WWW. Enfin, je préciserai le but de mon travail de thèse et je donnerai le plan de lecture de ce manuscrit.

1.1 Notions importantes associées aux réseaux

1.1.1 Le degré d'un nœud

En théorie des réseaux, on note N le nombre de nœuds et N_l le nombre de liens du système. On symbolisera le lien entre deux nœuds j et i, par une arête s'il est non dirigé et par une flèche, par exemple $j \to i$, s'il est dirigé. Dans le cas du réseau dirigé, la direction est importante, la flèche allant toujours du nœud source vers le nœud cible. Le degré $k(j)$ d'un nœud j est le nombre de nœuds avec lesquels j est lié, on parle aussi du nombre de voisins ou encore de la connectivité de j. Pour un réseau dirigé, on distingue alors deux types de degré, le degré entrant, $k_{\text{in}}(j)$, donnant le nombre de liens *entrants* et le degré sortant, $k_{\text{out}}(j)$, donnant le nombre de liens *sortants* du nœud j. On a alors, pour le nœud j, $k(j) = k_{\text{in}}(j) + k_{\text{out}}(j)$.

1.1.2 La distribution des degrés

Les degrés associés aux différents nœuds, suivent une distribution statistique spécifique au réseau étudié. On note $P(k)$ la probabilité de trouver dans un réseau de N nœuds, un nœud ayant k voisins. Le premier moment de cette distribution $\sum_k kP(k)$ est égale au degré moyen $\langle k \rangle$ et nous permet de retrouver le nombre de liens $N_l = \langle k \rangle \frac{N}{2}$. Bien évidemment, dans un réseau dirigé, nous avons deux distributions, $P_{\text{in}}(k_{\text{in}})$ et $P_{\text{out}}(k_{\text{out}})$ associées aux liens entrants et sortants. On peut aussi s'intéresser à la distribution couplée $P(k_{\text{in}}, k_{\text{out}})$, tel que $P(k_{\text{in}}) = \sum_{k_{\text{out}}} P(k_{\text{in}}, k_{\text{out}})$ et inversement $P(k_{\text{out}}) = \sum_{k_{\text{in}}} P(k_{\text{in}}, k_{\text{out}})$, nous permettant de retrouver la distribution de liens totale $P(k) = \sum_{k_{\text{in}}} P(k_{\text{in}}, k - k_{\text{in}}) = \sum_{k_{\text{out}}} P(k - k_{\text{out}}, k_{\text{out}})$.

1.1.3 La pondération des liens

On peut considérer que chaque lien d'un réseau possède un poids. On note $w(i, j)$, le poids du lien entre les nœuds i et j. Dans un réseau dirigé, l'égalité $w(i, j) = w(j, i)$ n'est pas forcement vrai. La pondération des liens est importante dans les réseaux où les interactions dépendent des nœuds sources et cibles. Ainsi, dans le cas d'un réseau d'échanges commerciaux, les liens représentant des flux de produits seront pondérés par les masses monétaires associées.

1.1.4 La notion de chemin

Le chemin menant de j vers i est une succession de liens, précisant les étapes de ce chemin. On mesure la longueur d'un chemin en comptant le nombre d'étapes, ou bien en sommant les poids associés aux liens qui composent le chemin. Il est intéressant de caractériser un réseau par son plus court chemin moyen, $\bar{l} = \sum_l lP(l)$, où $P(l)$ est la probabilité que le plus court chemin entre deux nœuds pris aléatoirement mesure l. Le plus court chemin moyen, $\bar{l}$, nous donne finalement une information sur la distance moyenne entre les nœuds d'un réseau. Dans le cadre d'un réseau de contact entre personnes, cette mesure nous permet de comprendre par exemple comment une épidémie peut se propager rapidement. Dans le cas d'une petite valeur de $\bar{l}$, on parle aussi de *petit monde*, phénomène qui sera défini plus tard.

1.1.5 Les communautés de nœuds

Dans un réseau, on définit une communauté de nœuds comme étant un sous-réseau ayant une forte connectivité interne. Dans un réseau, On définit un triangle fermé comme un sous-réseau complet, composé d'un triplet de nœuds (a, b, c), et un triangle ouvert comme un triangle fermé avec une paire de nœuds déconnectée. Le coefficient de cluster global [2], $C = \frac{\#\text{triangles fermés}}{\#\text{triangles fermés}+\#\text{triangles ouverts}}$, permet de caractériser la connectivité d'un réseau. La valeur maximale $C = 1$, est caractéristique d'un réseau où tous les nœuds sont liés entre eux. On parle aussi de graphe complet. Le coefficient de cluster local [3] associé au nœud

i, $C_i = \frac{2y_i}{k(i)(k(i)-1)}$, où y_i est le nombre de liens entre les voisins du nœud i, caractérise la connectivité du sous-réseau constitué du nœud i et de ses voisins.

1.1.6 La représentation matricielle

Outre la représentation graphique d'un réseau, sa représentation matricielle permet une modélisation plus efficace et permet par l'étude de son spectre d'avoir accès à des informations plus précises quant à la topologie du réseau. La représentation la plus simple est la matrice d'adjacence A, dont la composante A_{ij} dénote l'éventuelle existence d'un lien dirigé dont la source est le nœud j et la cible le nœud i. Si $A_{ij} = 1$, ce lien existe, si $A_{ij} = 0$, ce lien n'existe pas. Dans le cas d'un réseau non dirigé, nous avons pour tout couple (i, j) de nœuds, $A_{ij} = A_{ji}$. Dans le cas d'un réseau pondéré, $A_{ij} = w(i, j)$. Nous verrons d'autres matrices représentatives de réseaux dans le prochain chapitre.

1.1.7 La notion de centralité et ses mesures

Les nœuds d'un réseau ne jouant pas tous le même rôle, il y a par exemple des hubs, des intermédiaires entre clusters ou bien encore des nœuds périphériques faiblement connectés avec le reste du réseau, il est intéressant de mesurer l'importance relative de chaque nœud au sein du réseau, c'est ce qu'on appel la centralité [4]. Il existe beaucoup de mesures de centralité et chacune répond à des problématiques différentes. Les trois méthodes les plus connues sont : la centralité de degré (C_D), la centralité de proximité (C_C) et la centralité d'intermédiarité (C_B). La centralité C_D classe les nœuds par ordre décroissant de leur degré. Pour un nœud j, sa valeur de centralité de proximité, $C_C(j) = (\sum_i l(i,j))^{-1}$, représente l'inverse de la somme de tous les plus courts chemins entre j et tous les autres nœuds du réseau [5]. La mesure de centralité d'intermédiarité d'un nœud j est $C_B(j) = \sum_{s \neq t \neq j} \frac{\sigma_{st}(j)}{\sigma_{st}}$, où σ_{st} est le nombre de plus courts chemins entre les nœuds s et t, et où $\sigma_{st}(j)$ est le nombre de plus courts chemins reliant les nœuds s et t passant par le nœud j. Cette dernière mesure donne l'importance aux nœuds permettant de connecter les autres nœuds du réseau entre eux [6]. Ces trois mesures de centralité, illustrées à la Figure 1.1, permettent de classer, du plus central au moins central, les nœuds d'un réseau. Bien que corrélés entre eux, les classements obtenus à partir de différentes mesures de centralité sont *a priori* différents. Bien que ces mesures de centralité soient applicables aux réseaux dirigés, il existe d'autres mesures de centralité basées sur l'analyse spectrale de matrices décrivant les réseaux complexes dirigés. À la section 2.3 de ce manuscrit, nous décrirons en détail le vecteur PageRank $\mathbf{P}$ dont la composante P_i mesure la centralité du $i^{\text{ème}}$ nœud d'un réseau complexe.

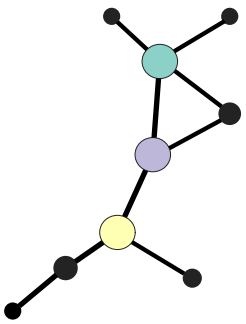

FIGURE 1.1 : Exemple d'un réseau non dirigé. Les nœuds colorés en vert, mauve, et jaune sont respectivement les nœuds les plus centraux du réseau d'après la centralité de degré (C_D), la centralité de proximité (C_C), et la centralité d'intermédiarité (C_B).

1.2 Les différents réseaux

Comme nous l'avons vu, un réseau est un ensemble de nœuds et de liens. Les nœuds peuvent se regrouper en communautés et ont des fonctions différentes au sein du réseau. Les liens peuvent être pondérés et aussi dirigés. La distribution des connexions dans un réseau est caractéristique du type de réseau (voir Figure 1.2). Il existe plusieurs méthodes pour construire un réseau. Dans la plupart d'entre elles, le nombre de nœuds est fixé. Une distribution des degrés est alors imposée, en suivant une loi de Poisson ou bien une loi exponentielle par exemple. Des liens sont ensuite créés en suivant cette distribution. Cette première approche, très simple, caractérise les réseaux aléatoires. Lorsque la distribution des degrés suit une loi de puissance, $P(k) \propto k^{-\gamma}$, le réseau est alors dit invariant d'échelle. Il existe encore d'autres types de réseaux, nous allons voir comment ils peuvent être modélisés et caractérisés.

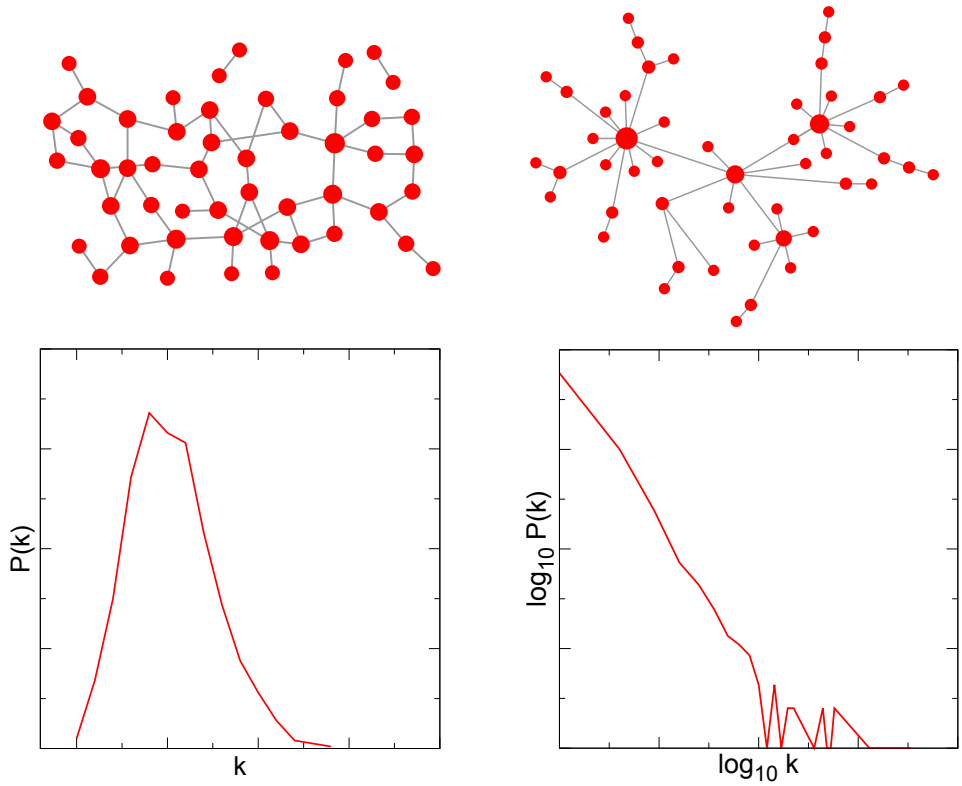

FIGURE 1.2 : Exemples d'un réseau aléatoire simple (à gauche) et d'un réseau invariant d'échelle (à droite). En dessous de chaque réseau se trouve la distribution des degrés associée : loi de Poisson (à gauche) et loi de puissance (à droite).

1.2.1 Les réseaux aléatoires

Un réseau aléatoire possède une distribution des degrés qui suit une loi de probabilité comme une distribution de Poisson ou bien une distribution exponentielle. Le modèle de Erdős–Rényi est un modèle simple permettant de générer ce type de réseau [7, 8]. L'ensemble, $G(N, p)$, représente tous les réseaux possibles avec N nœuds et une probabilité, p, que n'importe quel couple de nœuds soit connecté par un lien. Une première méthode de construction de réseau aléatoire est de tirer au hasard un réseau dans cet ensemble. Le nombre de liens attendu $\bar{N_L}$ pour un réseau aléatoire $G(N, p)$ est $pN(N-1)/2$ et le degré moyen est $\langle k \rangle = p(N-1)$. Une autre méthode consiste à générer tous les réseaux possédant N nœuds et M liens, ensemble noté $G(N, M)$, et d'en choisir un aléatoirement. Ces réseaux aléatoires se définissent comme réseau à l'équilibre. En effet le nombre de nœuds est fixé dès le début et les liens sont générés

par itérations, soit en respectant la probabilité de connexion p, soit en arrêtant le processus au bout de M liens. On peut cependant construire un réseau aléatoire hors équilibre de la manière suivante. On débute avec un petit nombre de nœuds, N_0, et à chaque étape de la construction, on rajoute un nœud et un lien. Dans cette dernière méthode, il y a un biais dans la distribution des degrés puisque plus le nœud est ancien plus il a de chance d'être connecté.

1.2.2 Les réseaux invariants d'échelle

Les réseaux invariants d'échelle possèdent une distribution des degrés suivant une loi de puissance en $k^{-\gamma}$. Il en résulte la présence de quelques nœuds ayant un très grand nombre de connexions : les hubs. Les réseaux invariants d'échelle montrent une forte résistance à la suppression aléatoire de nœuds et de liens [9, 10], mais ils sont fragile face à une épidémie [11]. Le modèle de Barabási-Albert [12, 13] basé sur le principe de l'attachement préférentiel, permet de générer de tel réseau. Pour ce faire, il faut commencer par construire un réseau contenant M_0 nœuds et dont les connexions entre eux peuvent être arbitraires, mais il faut tout de même que $k(i) > 0 \ \forall i \in \{1, \ldots, M_0\}$. Pour chaque itération du processus de construction, on ajoute un nœud que l'on connecte à $m \leq M_0$ nœuds déjà présents dans le réseau. La probabilité d'attachement, $P(i) = \frac{k(i)}{\sum_j k(j)}$, est la probabilité qu'un nouveau nœud soit connecté au nœud i. En procédant ainsi, les nœuds ajoutés seront préférentiellement connectés aux hubs du réseau. Ce type de réseau est hors équilibre puisque le nombre de nœuds et de liens augmentent avec le temps. De nombreux réseaux que l'on trouve dans la nature sont invariants d'échelle, le World Wide Web [14], les réseaux d'interactions de protéines [15] ou bien encore les réseaux de citations scientifiques [16].

1.2.3 Les réseaux petit monde

Lorsque le plus court chemin moyen, $\bar{l}$, est petit on parle de phénomène de petit monde. Un réseau d'interactions sociales est un exemple de réseau petit monde. L'expérience de Milgram montre que seulement 6 intermédiaires suffisent pour transmettre une lettre, de main à main, depuis la ville de Omaha dans le Nebraska, jusqu'à la ville de Sharon, proche de Boston dans le Massachusetts [17].

1.2.4 Les arbres

Un exemple très connu d'arbre est celui de l'arbre généalogique. Un arbre est un réseau où l'on distingue deux types de nœuds, les nœuds parents et les nœuds enfants. Un enfant peut aussi être parent. Un nœud terminal est un nœud qui ne possède pas d'enfants. Dans un arbre, il n'y a qu'un seul chemin possible entre deux nœuds. Tout arbre à N nœuds possède $N_l = N - 1$ liens. En effet, l'origine est le seul nœud de l'arbre à ne pas avoir de parents, tous les autres nœuds ont un unique parent et donc, ils comptent chacun pour un lien ce qui donne $N - 1$ liens. Un arbre binaire se compose de $N = n + p$ nœuds, où n est le nombre de nœuds parents ayant deux enfants et p, le nombre de nœuds terminaux. Dans un tel arbre, le nombre de nœuds terminaux est $p = \frac{N+1}{2}$ puisque $N = n + p$ et $N_l = N - 1 = 2n = 2(N - p)$, on aboutit à l'égalité $N + 1 = 2p$. On peut aussi, à partir d'un réseau quelconque, construire un arbre recouvrant. Il faut détruire les liens du réseau afin d'avoir un arbre tout en gardant les nœuds de départs. Dans le cas d'un réseau non dirigé à n nœuds et m liens, il faudra alors retirer $m - (n - 1)$ liens pour construire son arbre recouvrant.

1.2.5 Les réseaux bipartis

Dans le cas d'un réseau biparti, les nœuds sont classés en deux catégories, et le lien (j, i) n'est possible que si les nœuds j et i appartiennent à deux catégories différentes. Un très bon exemple

est le réseau électoral, où il y a deux types de nœuds, les électeurs et les candidats. Un lien représente alors le choix d'un électeur. On peut obtenir un réseau en projetant le réseau biparti sur l'un des deux sous-espace associé à une des catégories, dans le cas du réseau électoral, une projection sur l'espace des électeurs va lier les électeurs ayant voté pour un même candidat.

1.2.6 Les réseaux multiplexes

Un réseau multiplex est un réseau construit en couches successives de réseaux. Chaque couche représente un réseau et les nœuds appartenant à des couches différentes peuvent aussi être connectés. Les systèmes complexes tels que les interactions sociales ou bien encore les transports, ne peuvent être caractérisés par un simple réseau, mais par un réseau multiplex où chaque couche décrit un type d'interaction [18].

1.2.7 Les réseaux réels

Les réseaux réels sont des réseaux représentant des informations concernant un système complexe réels. Ils peuvent appartenir aux catégories définies plus haut. Pour la plupart, ils sont invariants d'échelle et montrent une forte décroissance de la distribution associée aux degrés forts [19]. Dans le cadre de mes travaux de thèse, je me suis intéressé à l'analyse de réseaux réels.

1.3 Le réseau de pages web et le moteur de recherche Google

Le réseau des pages web est un exemple contemporain où l'on peut appliquer la théorie des réseaux. Si nous voulons trouver l'information dont on a besoin, il faut d'abord classer les sources d'information. Dans une libraire, chaque livre est une source d'information et pour un accès rapide, les livres sont classés par thème et par nom d'auteur. Ainsi, le processus de recherche de l'information est facilité, il suffit de se rendre à l'étagère correspondante au thème et de retrouver l'auteur qui nous intéresse. L'ensemble des pages web est comme une bibliothèque numérique regroupant des milliards de livres, ces livres étant connectés les uns aux autres. Dans les premières études empiriques sur le World Wide Web (WWW), il a été montré que le réseau de pages web, dont le plus court chemin moyen est d'une vingtaine de clics [19], peut-être qualifié de petit monde. Pourtant la recherche d'information n'est pas si simple. Quand un utilisateur veut chercher une page web, il fait appel à un moteur de recherche. Il existe différents moteurs de recherche se basant sur différentes méthodologies de recherche. Les moteurs booléens utilisent une syntaxe logique comme par exemple "Voiture AND rouge", où les deux mots-clefs voiture et rouge sont attendus par l'utilisateur. Il suffit alors de donner à l'utilisateur la liste des pages web contenant ces deux informations. On imagine évidement, qu'en fonction de la langue utilisée et des différents synonymes et homonymes, il est difficile de répondre parfaitement à la requête de l'internaute. Le modèle vectoriel [20], développé dans les années 80, traduit un document, ici une page web et sa sémantique, en un vecteur. Le moteur de recherche va alors mesurer la pertinence sémantique entre la requête et les pages web existantes en assignant aux pages web un score compris entre 0 et 1. Les pages proposées sont classées par ordre décroissant de leur score. Le point commun entre ces moteurs de recherche est que le résultat est dépendant de la requête. En effet, selon ce que recherche l'internaute, les scores des pages web proposées seront différents. Dans le but d'optimiser la recherche d'information en temps de calcul, il est nécessaire d'avoir une méthode donnant un score aux pages ne dépendant pas de la requête. En 1998, la notion de popularité apparaît et permet d'avoir une méthode non dépendante de la requête. Le moteur de recherche le plus efficace et le plus utilisé par les internautes est celui de Google, dont l'algorithme fut pensé et développé par Sergey Brin et Larry Page. Il se distingue des moteurs recherche cités plus haut, par l'intégration d'un processus de marche aléatoire dans le WWW. En effet les fondateurs de Google, en 1999,

vont proposer de classer les pages web en utilisant une nouvelle mesure, le PageRank [21], permettant de mesurer la popularité d'une page. Plus le marcheur aléatoire a de chance de tomber sur une page web, plus cette page web est importante dans le réseau. Des chercheurs se sont intéressés à l'algorithme de Google et ont montré que, le PageRank pouvait être vu comme une mesure de centralité basée sur les vecteurs propres [22]. La matrice utilisée pour cela est appelée matrice de Google. Elle est la représentation matricielle de la marche aléatoire dans un réseau dirigé et donne accès à de nouvelles informations. Grâce à l'algorithme du PageRank qui classe les pages web par popularité, le temps pour obtenir une réponse à une requête est considérablement diminué puisque le classement est déjà fait. Nous allons voir plus précisément le fonctionnement de cet algorithme dans le prochain chapitre.

1.4 Le but de cette thèse

Nous allons nous intéresser à des réseaux dirigés réels et à l'utilisation d'outils mathématiques issus de la matrice de Google, avec notamment la méthode de la matrice de Google réduite. Cette dernière nous permet de quantifier les liens indirects entre des nœuds d'intérêts contenus dans un vaste réseau. J'ai choisi de vous présenter mes travaux de doctorat comme suit. Tout d'abord, une partie théorique sur la matrice de Google, l'algorithme du PageRank et la méthode de la matrice de Google réduite sera donnée au chapitre 2. Je vous présenterai alors mes recherches sur le réseau d'articles Wikipédia (chapitre 3), les réseaux économiques dont le réseau du commerce international (RCI) et le réseau mondial d'activités économiques (RMAE) (chapitre 4). Enfin, nous étudierons la propagation de crises économiques au sein du RCI et du réseau des transactions de Bitcoin (RTB) (chapitre 5).

Chapitre 2

Matrice de Google

Dans ce premier chapitre, nous allons voir la méthode de construction de la matrice de Google et l'algorithme du PageRank, élément au cœur du moteur de recherche Google. Nous verrons, aussi, les outils d'analyses de réseau qui en découlent et qui seront utilisés dans la suite de cette thèse.

2.1 Chaînes de Markov et matrices stochastiques

La méthode établie par Brin et Page, cofondateurs de Google, pour mesurer la popularité d'une page web est basée sur un calcul récursif [23]. Une page web peut être pointée, via des liens hypertextes, par d'autres pages web. La popularité de l'une d'elles dépend donc des autres

$$P_i = \sum_{j \in B_i} P_j/k_j \tag{2.1}$$

où P_i est la composante du vecteur PageRank associée à la page i, B_i est l'ensemble des pages web j pointant vers la page i, et k_j est le degré sortant de la page j. Le problème majeur est que nous ne savons pas à l'avance quelles sont les valeurs de PageRank des pages web. La solution la plus simple est donc de définir $P_i^{(0)} = 1/N \ \forall \ i$, un PageRank initial. On note $\mathbf{P}^{(0)}$ le vecteur de dimension N associé à ce PageRank. Il suffit alors de calculer itérativement le PageRank des pages web et d'arrêter quand le processus converge vers un unique vecteur $\mathbf{P}$. Symboliquement, le critère de convergence est

$$\left\| \mathbf{P}^{(k)} - \mathbf{P}^{(k-1)} \right\| = 0 \tag{2.2}$$

où $\mathbf{P}^{(k)}$ est le vecteur PageRank obtenu à la $k^{\text{ème}}$ itération. On peut réécrire (2.1) dans un formalisme d'algèbre linéaire, on obtient alors

$$\mathbf{P} = H\mathbf{P} \tag{2.3}$$

où la matrice d'hyperliens H est une matrice dont les éléments $H_{ij} = 1/k_j$ si $j \in B_i$, 0 sinon. H_{ij} représente la probabilité d'atteindre i depuis la page j. Finalement, le vecteur PageRank est simplement un vecteur propre de H correspondant à la valeur propre $\lambda = 1$. Cette matrice d'hyperliens est une matrice stochastique par colonne, c'est à dire que pour la colonne j, on a $\sum_i H_{ij} = 1$. Ainsi, la théorie des matrices stochastiques nous dit que le rayon spectral de H est unitaire, $\rho(H) = 1$. Le vecteur PageRank $\mathbf{P}$ est donc le vecteur propre dominant et décrit un vecteur de probabilité stationnaire.

De part l'existence de nœuds ballants (le nœud rouge sur la Figure 2.1), certaines colonnes de H peuvent ne pas être stochastique. La matrice H décrivant le réseau de la Figure 2.1, contiendra une colonne nulle en indice $j = 4$. En effet le nœud 4 ne pointe vers aucun nœud.

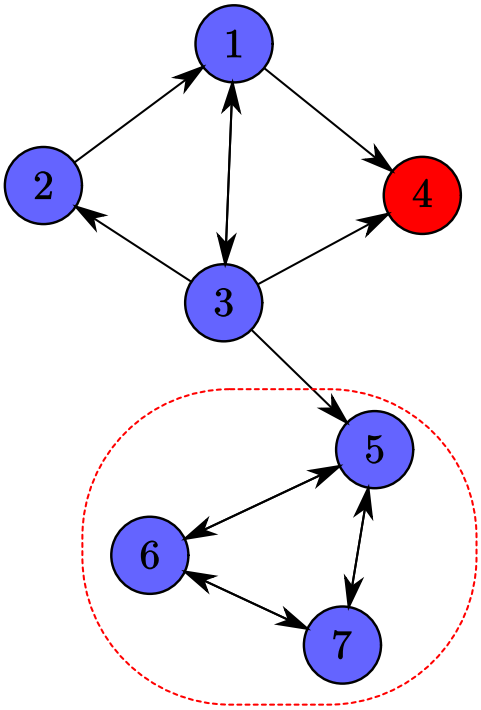

FIGURE 2.1 : Réseau représentant 7 pages web et 13 liens hypertextes, les nœuds sont les pages web et le lien $A \to B$ est l'existence de l'hyperlien présent dans la page A citant la page B. Le nœud rouge est associé à un nœud ballant, une impasse, et la zone en pointillé est ce qu'on peut appeler un évier, un sous-réseau de pages web ne citant pas de pages en dehors.

Dans ce cas nous avons $\rho(H) < 1$ et on ne peut plus trouver le vecteur propre recherché. Les auteurs ont alors considéré le problème d'une marche aléatoire d'un internaute dans le réseau de pages web. Soit S la matrice stochastique du système définie comme

$$S = H + \mathbf{a}\mathbf{e}^T \tag{2.4}$$

avec $\mathbf{a}$ un vecteur tel que $a_i = 1/N$ si i est un nœud ballant, $a_i = 0$ sinon, et $\mathbf{e}$ est un vecteur dont toutes les composantes sont égales à 1. Cela revient donc à changer une colonne de 0 en une colonne remplie de valeurs égales à $1/N$. De ce point de vue, quand un marcheur aléatoire, où plutôt un internaute aléatoire, se retrouve bloqué sur un nœud ballant, il peut toujours se diriger vers n'importe quelle page existante en utilisant son adresse url. Seulement, la nature stochastique de S ne suffit pas pour avoir une unique valeur propre localisée sur le cercle unitaire. Il faut pour cela avoir une matrice de Markov. La matrice S doit répondre à certaines propriétés importantes.

Un processus Markovien est un processus de marche aléatoire sans mémoire, autrement dit nous avons

$$Pr(X_{t+1} = x | X_1 = x_1, X_2 = x_2 ..., X_t = x_t) = Pr(X_{t+1} = x | X_t = x_t) \tag{2.5}$$

où $Pr(X_k = x_k)$ est la probabilité que la variable aléatoire X au temps k vaille x_k. L'unicité du vecteur dominant est respectée par deux contraintes, S doit être irréductible et apériodique.

L'irréductibilité d'une matrice

Il y a un lien fort entre l'irréductibilité d'une matrice et la forte connectivité du réseau qu'elle décrit. Soit M, une matrice carrée de taille $N \times N$, s'il existe au moins un opérateur de permutation P tel que $P^T M P = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, alors M est réductible. Ici A et C sont nécessairement des

matrices carrées. D'un point de vue de la théorie des graphes, la présence d'une sous-matrice nulle 0 montre que des nœuds ne sont pas atteignables. Or, la définition d'un réseau fortement connecté, est qu'il existe toujours un chemin entre n'importe quels nœuds i et j du réseau.

La périodicité d'une matrice
Une matrice carrée M est dite de période k, si $M^{k+1} = M$. Si tel n'est pas le cas, nous avons une matrice apériodique.

Ces deux points sont importants, car en ayant une matrice stochastique irréductible représentant une chaîne de Markov apériodique, nous avons une matrice dite primitive, c'est à dire qu'il existe une unique valeur propre λ se trouvant sur le rayon spectral. Ainsi dans le cadre de l'algorithme du PageRank, nous aurons un unique vecteur de probabilité stationnaire $\mathbf{P}$.

Afin d'aboutir à une telle matrice de Markov, il faut permettre au marcheur aléatoire de se téléporter. On aboutit alors à la matrice de Google, notée G

$$G = \alpha S + \frac{(1-\alpha)}{N}\mathbf{e}\mathbf{e}^T \tag{2.6}$$

où α, appelée *damping factor* ou encore coefficient d'amortissement en français, est un réel compris dans l'intervalle $[0.5, 1[$. La matrice de Google donne la possibilité au marcheur aléatoire de suivre la topologie du réseau avec une probabilité α ou bien de se téléporter n'importe où sur le réseau avec une probabilité $(1-\alpha)$. La matrice de Google G représente une chaîne de Markov et répond au théorème de Perron-Frobenius. En remplaçant H par G dans (2.3), nous obtenons la formule du PageRank

$$G\mathbf{P} = \mathbf{P} \tag{2.7}$$

On peut aussi récrire (2.6) en utilisant (2.4) et ainsi.

$$G = \alpha(H + \mathbf{a}\mathbf{e}^T) + (1-\alpha)\mathbf{e}\mathbf{e}^T/N \tag{2.8}$$

2.2 Opérateur de Perron-Frobenius

La méthode de construction de la matrice de Google, permet l'obtention d'une matrice stochastique apériodique et irréductible. Ces caractéristiques lui permettent de répondre positivement aux pré-requis nécessaires au théorème de Perron-Frobenius.

Théorème de Perron-Frobenius
Si une matrice A de taille $N \times N$ tel que $A \geq \mathbf{0}$ est irréductible alors

- $\mathrm{r} = \rho(A) > 0$.
- $\mathrm{r} \in \sigma(A)$ (r est une racine de A).
- La multiplicité algébrique de la racine r est égale à 1.
- Il existe un vecteur $\mathbf{x}$ non nul tel que $A\mathbf{x} = \mathrm{r}\mathbf{x}$, $\mathbf{x} > \mathbf{0}$[1] et $\|\mathbf{x}\|_1 = 1$.
- Ce vecteur, appelé vecteur de Perron est unique et $\|x\|_1 = 1$.
- r n'a pas besoin d'être la seule valeur propre localisée sur le cercle de rayon r.
- r maximise la formule de Collatz–Wielandt.

Afin d'obtenir une seule valeur propre sur le cercle unitaire, il nous faut en plus que A soit apériodique.

[1] Soit $x_j > 0 \; \forall \; j$.

2.3 Algorithme du PageRank

Nous allons voir maintenant comment mettre en place informatiquement l'algorithme du PageRank de Google. Le vecteur PageRank, noté $\mathbf{P}$, est le vecteur de Perron associé à G. Ainsi, il représente une distribution stationnaire de probabilité. Autrement dit, l'état stationnaire d'un marcheur aléatoire voyageant à travers un réseau dirigé est décrit par $\mathbf{P}$. La composante P_i est alors la probabilité de trouver le marcheur sur le nœud i après un parcours infini. Pour avoir accès aux vecteurs propres de G, il est possible de diagonaliser la matrice G. Cependant, la matrice G associée à un réseau réel est très souvent de très grande taille. Ces réseaux pouvant atteindre par exemple, des millions de nœuds dans le cas du réseau d'articles Wikipédia. Il est donc déconseillé d'utiliser des méthodes exactes tel que l'algorithme de Gauss. Finalement, en ne voulant que le vecteur dominant, le choix de la méthode des puissances est particulièrement judicieux, de par son efficacité et sa rapidité.

Méthode itérative des puissances

- Initialiser un vecteur de distribution de probabilité, $\mathbf{P}^{(0)} = \mathbf{e}/N$.
- Calculer $P^{(1)} = G\mathbf{P}^{(0)}$.
- Continuer de manière itérative $\mathbf{P}^{(k+1)} = G\mathbf{P}^{(k)}$.
- Arrêter lorsque $\mathbf{P}^{(k+1)} = \mathbf{P}^{(k)}$.

La taille de la matrice G est le facteur limitant puisqu'il peut être impossible de stocker un tel objet en mémoire. Heureusement en utilisant (2.8), on peut tout simplement utiliser la liste des liens (j, i) (nœud source, nœud cible) ainsi qu'une liste des nœuds ballants. Dans le cadre des travaux présentés dans ce manuscrit, l'algorithme 1 a été implémenté en C++. Il est bien entendu implémentable dans d'autres langages de programmation.

Data : T : tableau des liens (source, cible, poids) ; D : liste des nœuds ballants ; S : vecteur des poids associés aux liens sortants ; $\alpha \in [0.5, 1[$.
Result : $\mathbf{P}$: vecteur PageRank.
init $\mathbf{P}^{(0)} = \mathbf{e}/N$, $\mathbf{P} = \mathbf{0}$, $k = 0$;
while *test* = **FALSE do**
  for (j, i, w) *in* T **do**
    $P_i \mathrel{+}= P_j^{(0)} * \frac{w}{S_j}$;
  end
  for i *in* D **do**
    $k \mathrel{+}= P_i^{(0)}$;
  end
  for $i = 0$; $i < N$; i+=1 **do**
    $P_i = \alpha\left(P_i + k/N\right) + (1 - \alpha)/N$;
  end
  $test = conv(\mathbf{P}, \mathbf{P}^{(0)})$;
  $\mathbf{P}^{(0)} \leftarrow \mathbf{P}$;
  $\mathbf{P} = \mathbf{0}$;
end

Algorithme 1 : Algorithme du PageRank. Ici, la fonction $conv(\mathbf{P}, \mathbf{P}^{(0)})$ est un critère de convergence.

Le classement PageRank des nœuds d'un réseau s'obtient en classant dans l'ordre décroissant les composantes du vecteur $\mathbf{P}$. Nous définissons $K = \{K_1, K_2, K_3, ..., K_N\}$ les rangs

associés, avec $P_{K_1} \geq P_{K_2} \geq \ ... \ \geq P_{K_N}$. Le critère de convergence utilisé dans l'Algorithme 1 est nécessaire. En effet l'égalité $\mathbf{P}^{(k+1)} = \mathbf{P}^{(k)}$ n'est pas possible, cela étant dû à des erreurs de précisions et des artefacts numériques. Voici par exemple, deux critères de convergence

$$\mathcal{C}_1 : \left\| \mathbf{P}^{(k+1)} - \mathbf{P}^{(k)} \right\|_1 \leq \epsilon_1 \tag{2.9}$$

$$\mathcal{C}_2 : \max_j \left(\frac{|P_j^{(k+1)} - P_j^{(k)}|}{P_j^{(k+1)}} \right) \leq \epsilon_2 \tag{2.10}$$

où ϵ_1 et ϵ_2 sont arbitraires. Ceux deux critères de convergence peuvent être utilisés en même temps.

2.4 Algorithmes CheiRank et 2DRank

Comme décrit plus haut, le PageRank est une mesure de la centralité des nœuds d'un réseau. Cette mesure donne l'importance d'un nœud vis-à-vis de celle des nœuds pointant vers lui. Puisqu'un lien sortant d'un nœud i est aussi un lien en direction d'un nœud j, on pourrait penser que le PageRank suffit. Il se trouve que la distribution des liens sortants n'est pas toujours la même et donc il est intéressant de mesurer une centralité, réciproque, basée sur les liens sortants. Cette mesure, appelée CheiRank, est complémentaire au PageRank. Cette nouvelle mesure utilise les caractéristiques des liens sortants des nœuds [24]. On peut ensuite définir une mesure bidimensionnelle couplant PageRank et CheiRank [25, 26].

Le vecteur CheiRank $\mathbf{P}^*$ est tout simplement le résultat de l'algorithme du PageRank appliqué au réseau réciproque. Ce réseau réciproque est obtenu en inversant la direction de tous les liens du réseau original, de cette manière, nous mesurons l'importance d'un nœud vis-à-vis des liens sortants. Si A est la matrice d'adjacence du réseau de départ, alors A^T est la matrice d'adjacence du réseau réciproque. Les matrices S^* et G^* sont la matrice stochastique et la matrice de Google associées au réseau réciproque. On note $\{K_1^*, K_2^*, K_3^*, ..., K_N^*\}$ les rangs associés tel que $P^*_{K_1^*} \geq P^*_{K_2^*} \geq \ ... \ \geq P^*_{K_N^*}$. Le PageRank mesure l'influence, ou encore la popularité, d'un nœud dans le réseau tandis que le CheiRank mesure la communication d'un nœud. Dans le cadre d'une marche aléatoire, les nœuds à grand PageRank absorbent le marcheur tandis que ceux à grand CheiRank distribuent le marcheur dans le reste du réseau.

Imaginons maintenant que nous voulions trouver, au sein d'un réseau social, la personne propageant au mieux l'information. Il faut s'assurer que cette personne soit populaire mais aussi communicante. On peut quantifier cela avec le rang 2DRank ($\tilde{K}$). Cette méthode utilise les rangs PageRank et CheiRank, K et K^*. Sur la Figure 2.2, les nœuds d'un réseau sont repérés sur le plan (K, K^*). On forme les *carrés* successifs, $[1, K] \times [1, K]$ en faisant croître K de 1 à N. Sur le bord de chacun de ces *carrés*, il peut se trouver 0, 1 ou 2 nœuds. On classe les nœuds du réseau suivant la taille des *carrés*. Le nœud se trouvant sur le *carré* de plus petite (grande) dimension est le nœud le plus (moins) central selon la mesure 2DRank. Lorsque deux nœuds se trouvent sur le bord d'un même *carré*, c'est-à-dire que le PageRank de l'un des nœuds est égal au CheiRank de l'autre nœud, alors le nœud possédant le rang K^* le plus petit est classé premier. Pour obtenir ce classement, il suffit de disposer d'un tableau trois colonnes : le label du nœud, le rang K, et le rang K^*. On classe alors les lignes de ce tableau suivant les valeurs croissantes de $\max(K, K^*)$ puis on classe les lignes ex-æquo par valeurs de K^* croissantes.

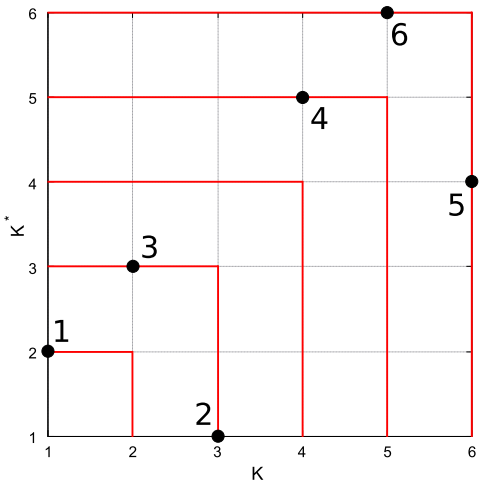

FIGURE 2.2 : Méthode de classement 2DRank. Les points sont les nœuds représentés dans le plan (K, K^*). Les labels $1, \ldots, 6$ indiquent le 2DRank de chaque nœud.

2.5 Vecteurs propres et valeurs propres de la matrice de Google G

Nous venons de voir que le PageRank et le CheiRank sont associés aux vecteurs propres dominants des matrices G et G^*. Le facteur α, utilisé dans la construction de ces matrices, nous permet d'avoir une seule valeur propre sur le cercle unitaire. Il est évident qu'en le faisant varier, il va de même pour les vecteurs propres associés et donc nous pouvons obtenir une variation du classement. En effet, pour $\alpha \approx 1$, une légère variation de α provoque d'importantes variations des composantes du vecteur PageRank $\mathbf{P}$ [22, p.59]. Il faut alors choisir une valeur de α permettant d'une part, un classement robuste, et d'autre part, un calcul rapide.

Les fondateurs de Google ont alors choisi une valeur $\alpha = 0.85$ comme un bon compromis entre la vitesse de calcul de $\mathbf{P}$ et sa stabilité lorsque α varie autour de 0.85 [21, 23].

La distribution $P(K)$ des probabilités de PageRank suit une loi de puissance avec un exposant de décroissance β

$$P(K) \approx 1/K^{\beta} \tag{2.11}$$

où $P(K)$ est la $K^{\text{ème}}$ plus grande composante du vecteur PageRank $\mathbf{P}$. Comme le PageRank est relié aux liens entrants (2.1), il est attendu que $P(K)$ suive une loi de puissance quand le réseau est invariant d'échelle. Afin de comparer les réseaux invariants d'échelle entre eux [22] et [25], il est intéressant de regarder la relation (2.12) entre β et μ, l'exposant caractéristique de la distribution des liens (entrants et/ou sortants) du réseau.

$$\beta = \frac{1}{\mu - 1}. \tag{2.12}$$

Outre le vecteur PageRank, les autres vecteurs propres de G peuvent donner d'importantes informations quant à la topologie du réseau. Les travaux [27, 28] présentent une méthode permettant d'accéder aux autres états propres du système.

La multiplicité de la valeur propre 1 pour la matrice stochastique S est due à des sous espace invariants de la matrice stochastique S. Il s'agit de l'ensemble des nœuds formant des sous-réseaux isolés, autrement dit en partant d'un de ces nœuds, on ne peut pas atteindre tous les nœuds du réseau. Ces sous-réseaux représentent $\approx 20 - 30\%$ des réseaux étudiés dans [27].

On peut récrire S en blocs

$$S = \begin{pmatrix} S_{\text{SS}} & S_{\text{SC}} \\ \mathbf{0} & S_{\text{CC}} \end{pmatrix} \tag{2.13}$$

où les indices S et C sont associées aux sous-espaces invariants (S) et à l'espace de cœur (C) du réseau. Comme on peut le voir dans (2.13), l'espace cœur C d'un réseau contient les nœuds pouvant atteindre tout le reste du réseau. Il est simple de voir que si un nœud peut atteindre un nœud ballant, alors, par construction de S, il peut atteindre tous les nœuds du réseau. En utilisant une méthode de diagonalisation basée sur la méthode Arnoldi, les auteurs ont montré que $\sigma(S_{\mathrm{CC}}) < 1$. On peut alors utiliser S_{CC} afin de récupérer les états propres de notre réseau sans passer par la construction de la matrice G.

L'article [28] présente une application de cette méthode au réseau d'articles Wikipédia. Il est montré que l'utilisation des modules des composantes des vecteurs propres, ψ_j de S_{SS}, permet de classer les articles Wikipédia. On observe, une similarité entre les articles les mieux classés. En effet, certains états propres ψ_j, se localisent sur des articles avec une thématique commune comme la Bible, les protéines ou un pays par exemple.

Autrement dit, les vecteurs propres, autres que le vecteur PageRank $\mathbf{P}$, nous permettent d'extraire des communautés d'articles. On peut aussi avoir accès au nombre d'éléments contenus dans ces communautés. En effet $\arg(\lambda_j)$ nous donne le nombre d'itérations i tel que, $S^i_{\mathrm{SS}}\psi_j = \psi_j$. Ainsi, le marcheur aléatoire initialement à l'état ψ_j, en suivant la topologie du réseau, induite par S_{SS}, reviendra à cet état d'origine en i itérations.

2.6 Sensibilité du PageRank

Nous venons de voir comment le vecteur PageRank $\mathbf{P}$ pouvait dépendre de α. Pour une valeur fixe de α, on peut faire varier le PageRank en modifiant les liens du réseau. Le phénomène du *link spam* est justement un problème pour les moteurs de recherches basés sur la méthode du PageRank. Le rang PageRank d'une page j, peut être augmenté par l'ajout de nouveaux liens entrants via la création d'hyperliens sur d'autres pages web. Ainsi, on augmente le nombre de liens entrants et pour peu qu'ils proviennent de pages web importantes, on peut grandement augmenter le score de la page j. Dans une démarche d'analyse physique des réseaux complexes réels, il est intéressant de voir comment la perturbation du lien $j \to i$ impacte le PageRank du nœud i. Cela revient à mesurer la sensibilité du PageRank, notée $\mathcal{D}$. Nous définissons deux types de sensibilités $\mathcal{D}$, la sensibilité diagonale $\mathcal{D}_{j\to i}(i)$, où on mesure la variation du score du nœud ciblé par le lien perturbé, et la sensibilité non diagonale $\mathcal{D}_{j\to i}(k)$, mesurant la variation des probabilités PageRank associées aux autres nœuds.

Le principe est simple, nous perturbons un ou plusieurs liens, ce qui modifie la matrice de Google associée (2.14). Ensuite, nous regardons la variation logarithmique des composantes du vecteur PageRank associées aux nœuds d'intérêts. Une fois le lien $j_0 \to i_0$ perturbé, les éléments de la matrice de Google sont

$$G(\delta)_{ij} = \begin{cases} \frac{G(0)_{ij}(1+\delta)}{1+G(0)_{i_0j_0}\delta} & \text{si } j = j_0 \text{ et } i = i_0 \\ \frac{G(0)_{ij}}{1+G(0)_{i_0j_0}\delta} & \text{si } j = j_0 \text{ et } i \neq i_0 \\ G(0)_{ij} & \text{sinon} \end{cases} \tag{2.14}$$

où $G(\delta)$ et $G(0)$ représentent respectivement la matrice de Google perturbée et non perturbée. On définit par δ le paramètre de perturbation. Les nœuds i_0 et j_0 sont respectivement les nœuds cible et source associés au lien perturbé. Afin que $G(\delta)$ reste stochastique par colonne, il faut renormaliser les colonnes contenant les perturbations, d'où la présence du dénominateur $(1 + G(\delta)_{i_0j_0}\delta)$ dans (2.14).

On peut calculer analytiquement la dérivée $dP_i/d\delta$ qui permet d'obtenir la sensibilité du nœud i à la variation infinitésimale du lien $j_0 \to i_0$ [29] On peut aussi, de manière numérique, faire converger (2.15) avec $\delta \to 0$

$$\mathcal{D}_{j\to i}(k) = (\mathbf{P}(\delta) - \mathbf{P}(0))_k/(P_k(0)\delta) \tag{2.15}$$

où $\mathcal{D}_{j\to i}(k)$ est la sensibilité du PageRank associé au nœud k par rapport à la perturbation du lien partant du nœud j et allant vers le nœud i.

2.7 La matrice de Google réduite G_{r}

La matrice de Google réduite, notée G_{r}, est une matrice de Google décrivant les interactions directes et indirectes, entre les nœuds appartenant à un sous-réseau d'intérêt enchâssé dans un vaste réseau. Les travaux [30, 31], décrivent cette méthode et la construction de cette matrice. Elle vient d'une analogie avec la théorie de la diffusion, utilisée notamment en physique nucléaire et mésoscopique [32, 33, 34, 35, 36].

Soit G la matrice de Google associée à un réseau de N nœuds et de N_l liens. Nous nous intéressons à un ensemble de nœuds d'intérêts, appelé ensemble des nœuds réduits N_{r}. Nous cherchons à construire la matrice de Google réduite G_{r} de taille $N_{\mathrm{r}} \times N_{\mathrm{r}}$. On peut récrire G sous forme de matrices blocs

$$G = \begin{pmatrix} G_{\mathrm{rr}} & G_{\mathrm{rs}} \\ G_{\mathrm{sr}} & G_{\mathrm{ss}} \end{pmatrix} \tag{2.16}$$

avec G_{rr} et G_{ss} représentant respectivement les interactions directes entre les nœuds réduits, et les interactions directes entre les autres nœuds du réseau. Les blocs hors diagonaux, G_{rs} et G_{sr}, décrivent les liens permettant de passer du sous-réseau d'intérêt au reste du réseau, et inversement.

Nous pouvons récrire le vecteur PageRank $\mathbf{P}$ comme

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{\mathrm{r}} \\ \mathbf{P}_{\mathrm{s}} \end{pmatrix}. \tag{2.17}$$

Il vient que

$$G_{\mathrm{rr}}\mathbf{P}_{\mathrm{r}} + G_{\mathrm{rs}}\mathbf{P}_{\mathrm{s}} = \mathbf{P}_{\mathrm{r}} \tag{2.18}$$
$$G_{\mathrm{sr}}\mathbf{P}_{\mathrm{r}} + G_{\mathrm{ss}}\mathbf{P}_{\mathrm{s}} = \mathbf{P}_{\mathrm{s}}. \tag{2.19}$$

On peut isoler $\mathbf{P}_{\mathrm{s}}$ dans (2.19)

$$\mathbf{P}_{\mathrm{s}} = (\mathbf{1} - G_{\mathrm{ss}})^{-1} G_{\mathrm{sr}}\mathbf{P}_{\mathrm{r}}. \tag{2.20}$$

En injectant (2.20) dans (2.18) on obtient

$$G_{\mathrm{rr}}\mathbf{P}_{\mathrm{r}} + G_{\mathrm{rs}}(\mathbf{1} - G_{\mathrm{ss}})^{-1} G_{\mathrm{sr}}\mathbf{P}_{\mathrm{r}} = \mathbf{P}_{\mathrm{r}}. \tag{2.21}$$

Les N_{r} nœuds du sous-réseau d'intérêt doivent conserver un classement PageRank identique, que ce soit par le vecteur PageRank $\mathbf{P}$, calculé à partir de G, ou bien par le vecteur PageRank $\mathbf{P}_{\mathrm{r}}$ associé à G_{r}. Compte tenu de cela, il vient de (2.21)

$$G_{\mathrm{r}} = G_{\mathrm{rr}} + G_{\mathrm{rs}}(\mathbf{1} - G_{\mathrm{ss}})^{-1} G_{\mathrm{sr}}. \tag{2.22}$$

L'inverse de la matrice, $\mathbf{1} - G_{\mathrm{ss}}$, est non négative. En effet, elle peut s'écrire comme une somme, $\sum_{l=0}^{\infty} G_{\mathrm{ss}}^{l}$. L'utilisation de méthode exacte pour le calcul de cette inverse peut être impossible lorsque $N_{\mathrm{s}} = N - N_{\mathrm{r}}$ est élevé. Il est donc plus intéressant d'utiliser la somme. Cette matrice inverse donne finalement l'ensemble des chemins de taille l permettant de circuler dans le sous-espace associé à N_{s}. Dans le cas où $N_{\mathrm{s}} \gg N_{\mathrm{r}}$, on peut estimer que la valeur propre λ_c associée au vecteur propre dominant de G_{ss}, est telle que $\lambda_c \approx 1$. On peut construire l'opérateur, $\mathcal{P}_c$, de projection sur l'espace propre associé à λ_c et $\mathcal{Q}_c$ l'opérateur complémentaire. On a

$$\mathcal{P}_c = \psi_{\mathrm{r}}\psi_L^T \tag{2.23}$$
$$\mathcal{Q}_c = \mathbf{1} - \mathcal{P}_c \tag{2.24}$$

où ψ_{r} et ψ_L^T sont respectivement les vecteurs propres dominants à droite et à gauche de G_{ss}. On a alors $G_{\mathrm{ss}}\psi_{\mathrm{r}} = \lambda_c\psi_{\mathrm{r}}$ et $\psi_L^T G_{\mathrm{ss}} = \lambda_c\psi_L^T$. Les vecteurs ψ_R et ψ_L sont normalisés à 1. La matrice de Google réduite G_{r} est la somme de la matrice G_{rr}, associée aux interactions directes, et de la matrice $G_{\mathrm{I}} = G_{\mathrm{rs}}(\mathbf{1} - G_{\mathrm{ss}})^{-1}G_{\mathrm{sr}}$, associée aux interactions indirectes. La composante G_{I} peut être décomposée en la somme $G_{\mathrm{pr}} + G_{\mathrm{qr}}$ [30, 31] avec

$$G_{\mathrm{pr}} = G_{\mathrm{rs}}\frac{\mathcal{P}_c}{1-\lambda_c}G_{\mathrm{sr}} \tag{2.25}$$

et

$$G_{\mathrm{qr}} = G_{\mathrm{rs}}\left(\mathcal{Q}_c\sum_{l=0}^{\infty}\tilde{G}_{ss}\right)G_{\mathrm{sr}} \tag{2.26}$$

où $\tilde{G}_{ss} = \mathcal{Q}_c G_{\mathrm{ss}}\mathcal{Q}_c$. On peut alors récrire (2.22) comme

$$G_{\mathrm{r}} = \underbrace{G_{rr}}_{\text{Liens directs}} + \underbrace{G_{\mathrm{pr}}}_{\text{Contribution du PageRank}} + \underbrace{G_{\mathrm{qr}}}_{\text{Liens cachés}} \tag{2.27}$$

où G_{pr} est une matrice dont les colonnes sont similaires à celles de la matrice $\mathbf{P}_{\mathrm{r}}\mathbf{e}^T$. La matrice G_{pr} décrit les liens indirects entre les N_{r} nœuds qui passent par les nœuds importants (au sens du PageRank) du réseau. Les éléments de la matrice G_{qr} représentent les liens indirects entre les nœuds d'intérêt qui passent par tous les chemins possibles, inclus dans le reste du réseau. À la différence de G_{pr}, la matrice G_{qr} décrit les liens indirects non-triviaux par diffusion dans le reste du réseau, dans le sens où ces liens ne passent pas par des nœuds forcément importants.

La matrice G_{r} est bien une matrice stochastique par colonne. En effet en sachant que $\mathbf{e}^T G = \mathbf{e}^T$ et en définissant $\mathbf{e}^T = (\mathbf{e}_{\mathrm{r}}^T, \mathbf{e}_{\mathrm{s}}^T)$, on trouve $\mathbf{e}_{\mathrm{r}}^T G_{\mathrm{r}} = \mathbf{e}_{\mathrm{r}}^T$. Ainsi,

$$\mathbf{e}^T\begin{pmatrix} G_{\mathrm{rr}} & G_{\mathrm{rs}} \\ G_{\mathrm{sr}} & G_{\mathrm{ss}}\end{pmatrix} = (\mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rr}} + \mathbf{e}_{\mathrm{s}}^T G_{\mathrm{sr}}, \mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rs}} + \mathbf{e}_{\mathrm{s}}^T G_{\mathrm{ss}}) \tag{2.28}$$

$$\mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rs}} = \mathbf{e}_{\mathrm{s}}^T(\mathbf{1} - G_{\mathrm{ss}}) \tag{2.29}$$

$$\mathbf{e}_{\mathrm{s}}^T G_{\mathrm{sr}} = \mathbf{e}_{\mathrm{r}}^T(\mathbf{1} - G_{\mathrm{rr}}) \tag{2.30}$$

en multipliant (2.22) à gauche par $\mathbf{e}_{\mathrm{r}}^T$ et en utilisant (2.29) et (2.30) on obtient

$$\begin{aligned}\mathbf{e}_{\mathrm{r}}^T G_{\mathrm{r}} &= \mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rr}} + \mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rs}}(\mathbf{1} - G_{\mathrm{ss}})^{-1}G_{\mathrm{sr}} \\ &= \mathbf{e}_{\mathrm{r}}^T G_{\mathrm{rr}} + \mathbf{e}_{\mathrm{s}}^T G_{\mathrm{sr}} \\ &= \mathbf{e}_{\mathrm{r}}^T.\end{aligned}$$

Pour chaque composante de G_{r}, on peut classer les valeurs d'une colonne j et ainsi, définir un classement par ordre d'importance, des liens directs et indirects sortants de j. Les matrices G_{rr} et G_{pr} sont, par construction, des matrices strictement positives. Afin que G_{r} soit stochastique, il est possible que des éléments de la matrice G_{qr} soient négatifs. Dans ce cas, nous pouvons prendre uniquement les éléments positifs associés à une colonne j de G_{qr}, afin de classer par ordre d'importance les liens sortants du nœud j. L'interprétation physique des termes négatifs présents dans la matrice G_{qr} n'est pas discuté dans les articles [30, 31], et pourrait être l'objet d'une étude ultérieure

La matrice de Google réduite, G_{r}, est un outil puissant, il nous permet de retrouver les liens directs et indirects entre les éléments d'un sous-réseau enchâssé dans un réseau plus vaste. En prenant en compte tous les chemins possibles entre les nœuds d'intérêts, par diffusion dans tout le réseau, on est en mesure de stocker l'information d'un vaste réseau dans un réseau beaucoup plus petit, et ainsi gagner en mémoire de stockage.

Chapitre 3

Réseau Wikipédia

Wikipédia est une encyclopédie en ligne et collaborative. Lancée le 15 janvier 2001 par Jimmy Wales et Larry Sanger, elle contient aujourd'hui 52 millions d'articles regroupés dans environ 300 éditions linguistiques différentes. L'édition la plus complète est l'édition anglaise, avec un total de 6.1 millions d'articles environ, soit 11% de l'ensemble des articles de Wikipédia. Les éléments de connaissance se trouvant dans cette bibliothèque numérique sont riches, variés, et dans certains cas exhaustifs, ce qui confère à cette base de données la qualité d'encyclopédie. L'accès à un article Wikipédia se fait simplement. En effet, pour consulter l'article "Univers" par exemple, il suffit de taper "Univers Wikipédia" sur Google, ou bien de taper "Univers" dans la barre de recherche de Wikipédia. Enfin, on peut passer d'un article à un autre en cliquant sur un lien. Par exemple, à partir de la page "Univers" de Wikipédia français, on peut lire l'article "cosmologie" en cliquant sur le lien présent dans "Univers" qui cite ce dernier. Ainsi, on peut construire, pour une édition linguistique donnée, le réseau des articles Wikipédia. Dans ce chapitre, nous allons voir deux études permettant d'extraire une information pertinente d'un tel réseau. Au vu du grand nombre d'articles et de liens présents dans Wikipédia, il peut être difficile d'avoir accès à certaines informations. Premièrement, nous allons voir comment l'application de la matrice de Google et de sa version réduite permet d'obtenir un classement mondial des universités. L'utilisation de la matrice de Google réduite et son analyse nous permet aussi de mesurer l'influence des différentes universités sur les pays et de pouvoir mettre en lumière les interactions directes et indirectes entre les meilleures universités mondiales. La seconde étude concerne l'utilisation des données relatives aux comportements des utilisateurs de Wikipédia telles que le nombre de clics sur les différents liens et le nombre de vues des articles. Grâce à ces informations, nous pouvons obtenir des classements d'articles Wikipédia reflétant les tendances sociales et culturelle actuelles.

3.1 Interactions et influences des universités dans le monde

3.1.1 Introduction

Le développement des sociétés humaines est fortement lié aux connaissances qu'elles détiennent et à la façon dont celles-ci sont transmises, notamment au travers du système académique. Il est alors important de pouvoir mesurer l'efficacité d'une université et aussi sa capacité à rayonner culturellement à travers le temps et l'espace. De nombreux outils ont été mis en place dans ce but. Parmi eux, les classements mondiaux d'universités sont de plus en plus nombreux [37]. Le classement le plus connu et le plus relayé dans les médias est le classement *Academic Ranking of World Universties* (ARWU), mis en place par l'université de Shanghai Jiao Tong depuis 2003.[1] Bon nombres de pays ont adapté leurs politiques d'enseignements supérieur et

[1] Academic Ranking of World Universities Site web : http://www.shanghairanking.com/. Accès en Juillet 2018.

de recherche afin que leurs universités soient mieux classées dans ces palmarès. En France, on note l'émergence de projets, plus ou moins dédiés, les *Laboratoires d'Excellences* (LabEx) et les *Initiative d'Excellences* (IDex).[2] En Russie, une source de financement est le *Russian Academic Excellence Project*.[3] Outre ARWU, il existe de nombreux autres classements d'universités.[4] Des auteurs se sont intéressés aux points forts et aux points faibles des méthodes de classement [38, 39, 40]. Ces classements se distinguent par leur méthodologie et les critères utilisés pour donner un score aux universités. L'utilisation de critères *a priori* induit des biais spécifiques à chaque classement. Récemment, un classement appelé Wikipedia Ranking of World Universities (WRWU) a été proposé [25, 41, 42]. L'idée est d'utiliser le réseau d'articles Wikipédia d'une édition linguistique donnée, afin d'obtenir un classement par popularité, au sens du PageRank, des articles Wikipédia associés aux universités du monde entier. Une première étude basée sur une édition anglaise de Wikipédia de 2009 [41] a montré l'efficacité d'une telle méthode. Une seconde étude a réitéré ce classement en utilisant 24 éditions linguistiques de Wikipédia de 2013. L'utilisation d'un grand nombre d'éditions linguistiques permet une vision multiculturelle [42]. D'autres classements ont été réalisés en utilisant cette méthode, notamment des classements de pays [41] ou bien encore des classements grandes figures historiques [43]. L'utilisation de la matrice de Google réduite pour l'analyse des interactions directes et indirectes entre les universités ainsi que pour l'étude de leur rayonnement culturel n'a pas été faite. Dans cette première partie de ce chapitre sur les réseaux Wikipédia, nous nous sommes intéressés au nouveau classement mondial des universités pour l'année 2017 (données Wikipédia de type XML datant de mai 2017[5]) mais aussi à l'application de la matrice de Google réduite [30] afin d'analyser les interactions directes et indirectes entre les meilleures universités mondiales. En effet, cette méthode nous permet de quantifier les liens indirects entre les éléments d'un sous-réseau en prenant en compte tous les chemins existants du réseau global. L'efficacité de l'application de la méthode de la matrice de Google réduite au réseau Wikipédia a été montrée dans différents domaines tels que l'analyse des liens entre les leaders politiques [31] et le terrorisme [44]. En dehors de Wikipédia, la méthode de la matrice de Google réduite a été appliquée dans d'autres domaines tels que la biologie avec l'analyse du réseau d'interactions entre protéines [45]. L'étude sur les interactions et influences des universités dans le monde est présentée de la manière suivante. Nous allons voir premièrement la méthode de construction de la matrice de Google associée à Wikipédia, la mesure du prestige académique et nous comparerons le classement WRWU avec le classement ARWU (section 3.1.2 et section 3.1.3). Ensuite, il sera donné les résultats sur la mesure du rayonnement culturel des universités, ainsi que sur les interactions entre grandes universités à travers l'analyse de matrices de Google réduites associées à 4 éditions : anglaise (EN), française (FR), allemande (DE) et russe (RU) (section 3.1.4). Enfin, nous étudierons le cas de la matrice de Google réduite effective apportant un point de vue multiculturel se basant sur 24 éditions linguistiques (section 3.1.5).

3.1.2 Construction de la matrice de Google

Nous nous intéressons aux réseaux d'articles Wikipédia correspondant à 24 éditions linguistiques dont les données ont été extraites à partir des *dumps* XML de mai 2017.[6] La liste des

[2] Enseignement supérieur et recherche, Investissements d'avenir. Site web : http://www.enseignementsup-recherche.gouv.fr/pid24578/investissements-d-avenir.html. Accès Juillet 2018.

[3] Russian Academic Excellence Project. Site web : http://5top100.ru/. Accès Juillet 2018.

[4] Times Higher Education World University Ranking https://www.timeshighereducation.com, U-Multirank de l'Union Européenne http://www.umultirank.org/ et l'IREG Observatory on Academic Ranking and Excellence.

[5] Données mises en ligne par K.M. Frahm et D.L. Shepelyansky Site web : http://www.quantware.ups-tlse.fr/QWLIB/24wiki2017. Accès Juillet 2018.

[6] Les dumps XML sont accessibles à l'adresse suivante https://dumps.wikimedia.org/enwiki/ où enwiki réfère à l'édition anglaise. Pour d'autres éditions, il suffit de changer [en]wiki par le code équivalent, par exemple frwiki, arwiki, svwiki, etc ...

éditions utilisées et leur nombre d'articles et de liens est affichée dans la Table 3.1. Le lien dirigé $A \rightarrow B$ traduit la présence d'un lien citant l'article B dans l'article A. Dans ces 24 réseaux, nous avons en moyenne 17 liens par nœud. Les boucles telles que $A \rightarrow A$ ne sont pas prises en compte. Le classement de ces éditions en fonction de leur nombre d'articles a changé entre 2013 et 2017. En effet, il est intéressant de voir que durant ces dernières années, la mise en place de plusieurs "bots" afin de créer/modifier des articles a permis une augmentation du nombre d'articles pour certaines éditions linguistiques. C'est le cas notamment pour l'édition suédoise, passant de 780 000 articles en 2013 à 3 800 000 articles en 2017.

Edition	Langue	N	Edition	Langue	N
EN	Anglais	5416537	ZH	Chinois	939625
SV	Suédois	3786455	FA	Perse	539926
DE	Allemand	2057898	AR	Arabe	519714
NL	Hollandais	1900222	HU	Hongrois	409297
FR	Français	1866546	KO	Koréen	380086
RU	Russe	1391225	TR	Turque	291873
IT	Italien	1353276	MS	Malaisien	289234
ES	Espagnol	1287834	DA	Danois	225523
PL	Polonais	1219733	HE	Hébreu	205411
VI	Vietnamien	1155932	EL	Grec	130429
JP	Japonais	1058950	HI	Indien	121503
PT	Portuguais	967162	TH	Thaïlandais	116495

TABLE 3.1 : Liste des 24 éditions utilisées pour la construction des réseaux d'articles Wikipédia (données obtenues en mai 2017). N est le nombre d'articles. D'après [46].

La matrice de Google se construit de la même manière que celle décrite dans le chapitre 2

$$G = \alpha S + (1 - \alpha)\frac{ee^T}{N} \tag{3.1}$$

avec, S une matrice stochastique telle que l'élément S_{ij} est la probabilité de cliquer sur le lien présent dans l'article j qui mène à l'article i. Dans le cas d'un nœud ballant, c'est-à-dire un article ne citant aucun autre, la colonne correspondante à cet article est équivalent au vecteur $(1/N, 1/N, ..., 1/N)^T$.

Nous avons choisi $\alpha = 0.85$, une valeur utilisée dans les études de réseaux issus du WWW [22].

3.1.3 Classement 2017 des universités (WRWU17)

Le classement des universités mondiales par le biais du réseau Wikipédia est simplement un sous-ensemble du classement de tous les articles d'une édition donnée. Pour une édition E et un algorithme de classement A, qui peut être le classement PageRank (PR), CheiRank (CR) ou bien 2DRank (2D), on note $\mathcal{R}_{E,A}$ l'ensemble des 100 meilleurs articles relatifs à des universités ou des écoles supérieures ayant un pôle de recherche. On définit alors le rang $\Theta_{U,A}$, d'une université U ainsi :

$$\Theta_{U,A} = \sum_{E}(101 - R_{U,E,A}) \tag{3.2}$$

où $R_{u,E,A}$ est le rang calculé avec l'algorithme A d'une université U dans l'édition E. Dans le cas où une université U' n'est pas présente dans l'ensemble $\mathcal{R}_{E,A}$, c'est-à-dire qu'elle n'est pas présente dans le top 100 obtenu pour l'édition E avec l'algorithme A, on fixe la valeur du score à $R_{U',E,A} = 101$.

On note WRWU17 le classement des universités obtenu avec l'algorithme PageRank. Le top 10 de ce classement et le top 10 ARWU17 sont listés dans la Table 3.2 et Table 3.3 respectivement. On observe déjà une différence dans le top 3. En effet, ARWU17 donne de l'importance aux universités américaines, avec un top 3 contenant deux d'entre elles, Harvard et Stanford. La troisième université selon ARWU17 est Cambridge, une université anglaise. Le classement WRWU17 donne les deux premières places à deux universités anglaises, Oxford suivie de Cambridge.

Rang	Θ_{PR}	N_a	Université	CC	LC	FC
1	2281	24	Oxford	UK	EN	11
2	2278	24	Cambridge	UK	EN	13
3	2277	24	Harvard	US	EN	17
4	2099	24	Columbia	US	EN	18
5	1959	23	Yale	US	EN	18
6	1917	24	Chicago	US	EN	19
7	1858	23	Princeton	US	EN	18
8	1825	21	Stanford	US	EN	19
9	1804	21	MIT	US	EN	19
10	1693	20	California, Berkeley	US	EN	19

TABLE 3.2 : Liste des 10 meilleures universités selon WRWU17. Le score Θ_{PR} est défini par (3.2) en utilisant l'algorithme PageRank, N_a est le nombre d'occurrences de l'université parmi les 100 des 24 éditions linguistiques, CC est le code pays, LC est le code linguistique et FC le siècle de fondation de l'établissement. D'après [46].

Rang	ARWU17	WRWU17
1	Harvard	-2
2	Stanford	-6
3	Cambridge	+1
4	MIT	-5
5	Californie, Berkeley	-5
6	Princeton	-1
7	Oxford	+6
8	Columbia	+4
9	CalTech	-13
10	Chicago	+4

TABLE 3.3 : Liste des 10 meilleures universités selon ARWU17 [47]. La dernière colonne montre la différence de rang entre ARWU17 et WRWU17. D'après [46].

Nous pouvons aussi mesurer la similarité entre deux classements en déterminant une valeur de recouvrement

$$\eta(j) = j_c/j \tag{3.3}$$

avec $0 \leq \eta(j) \leq 1$ et où j_c est le nombre d'éléments communs sur les j premiers comparés.

La Figure 3.1 montre ces valeurs de recouvrement pour différents couples de classements et allant jusqu'à $j = 100$. Le classement WRWU17 ne semble pas éloigné du classement ARWU17. Nous avons un recouvrement $\eta(100) = 60\%$ à l'échelle du top 100. Dans l'étude précédente [42], cette similarité était à peine plus élevée ($\eta(100) = 62\%$). Les classements WRWU13 et WRWU17 sont davantage similaires, avec une valeur $\eta(100) = 91\%$ par rapport à ARWU13

et ARWU17, qui ont entre eux une similarité de 84%. Le fait que le classement WRWU17 soit à la fois proche du classement ARWU17 et en même temps très proche du classement WRWU13 montre qu'il s'agit d'un classement cohérent et bien plus stable dans le temps que le classement ARWU. Une des spécificités du classement WRWU est l'incorporation de diverses cultures via la prise en compte de 24 éditions linguistiques différentes. Des différences culturelles peuvent être notées. Ainsi, dans la partie droite de la Figure 3.1, le top 10 obtenu avec l'édition allemande DEWIKI17 (DEWRWU17) a seulement 10% de similarité avec le top 10 ARWU17 tandis que les éditions FR et EN montrent une similarité d'environ 50% à l'échelle du top 10. Pour les valeurs de similarités obtenues en comparant les tops 100, on observe de nouveau de telles disparités. En effet, les tops 100 DE, FR et ENWIKI17 ont respectivement 34%, 43% et 60% de similarité avec le top 100 ARWU17. Comme discuté dans [42], le cas de l'édition allemande donne lieu à un classement qu'on pourrait qualifier de "patriotique", car il contient bien plus d'universités germanophones que d'universités étrangères.

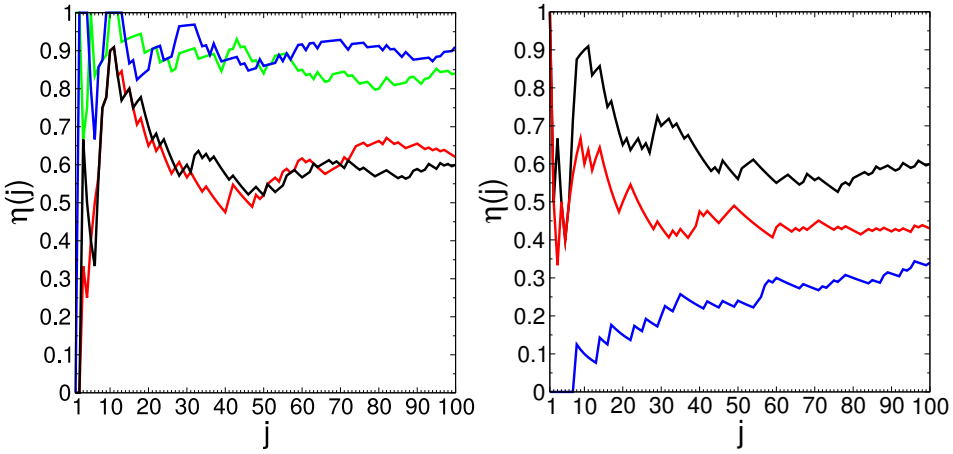

FIGURE 3.1 : Figure de gauche : Similarité $\eta(j) = j_c/j$ entre ARWU et WRWU en fonction du rang j. Le nombre j_c est le nombre d'universités en commun dans les deux tops j considérés. Les paires de classements présentées sont : ARWU2017 et WRWU2017 (noir), ARWU2013 et WRWU2013 (rouge), ARWU2013 et ARWU2017 (vert) et WRWU2013 et WRWU2017 (bleu), pour la partie gauche, et ARWU2017 et ENWRWU2017 (noir), ARWU2017 et FRWRWU2017 (rouge) et ARWU2017 et DEWRWU2017 (bleu), pour la partie droite. D'après [46].

En utilisant 24 éditions (Table 3.1), nous avons classé au total 1011 universités avec l'algorithme PageRank, et 1464 avec l'algorithme CheiRank. La distribution géographique de ces classements est présentée à la Figure 3.2. Dans le cas WRWU17, les pays ayant le plus d'universités classés sont les États-Unis, l'Inde, le Japon, l'Allemagne et la France. Pour l'algorithme CheiRank (WRWU17-CR), les pays ayant le plus d'universités communicatives[7] sont les États-Unis, l'Inde, le Japon, la France et la Chine.

Il est intéressant de voir les différences dans les distributions géographiques des 100 meilleures universités issues des classements ARWU17 et WRWU17. Comme le montre la Figure 3.3, le classement WRWU17 donne plus d'universités classées en Europe. D'un côté, la méthode de classement ARWU facilite la présence des universités américaines, anglaises et chinoises en tête du classement tandis que le top 100 WRWU17 donne plus de poids aux universités allemandes, surlignant ainsi leur importance historique.

La Figure 3.4 donne la distribution statistique des universités par pays pour les tops 100 ARWU17 et WRWU17 ainsi que pour les 1011 universités présentes dans le classement global

[7] Comme expliqué dans le Chapitre 2, le PageRank mesure l'efficacité des liens entrants et donne alors une mesure de popularité. À contrario, le CheiRank mesure l'efficacité des liens sortants et donc mesure la qualité de communication d'un nœud au sein du réseau.

WRWU17.[8]

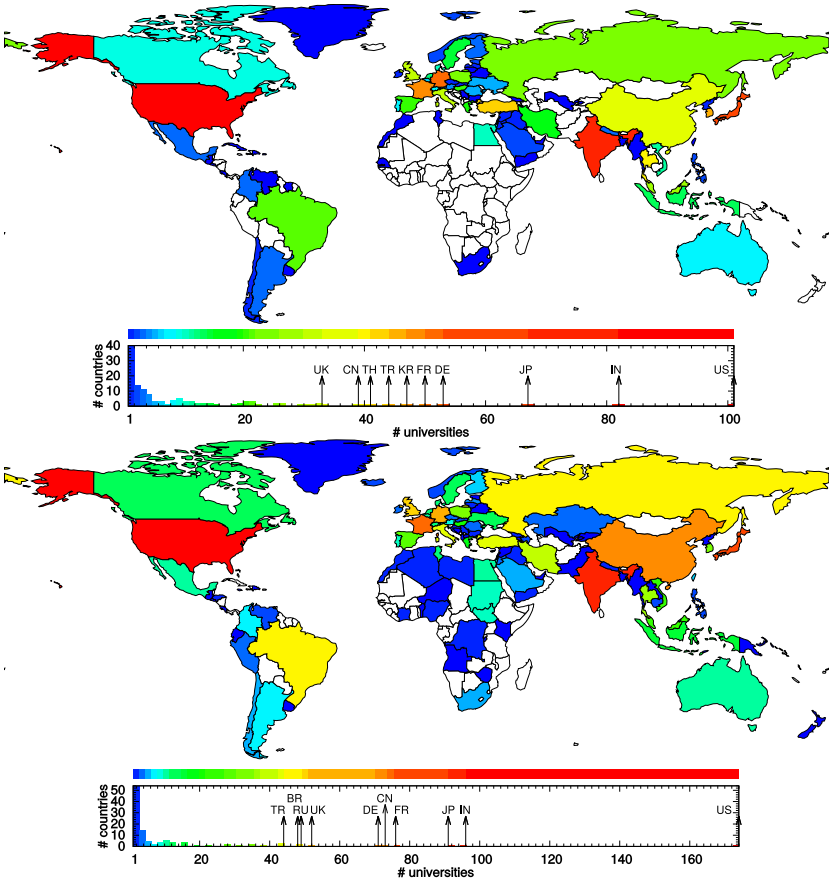

FIGURE 3.2 : Distribution géographique des universités issues du classement WRWU17 utilisant l'algorithme PageRank (en haut) et en utilisant l'algorithme CheiRank (en bas). Il y a 1011 (1464) universités présentes dans les classements obtenus avec la méthode du PageRank (CheiRank). Les universités américaines sont les plus présentes avec 101 (174) occurrences pour le classement PageRank (CheiRank) WRWU17. Les pays qui n'ont pas d'université dans les classements sont colorés en blanc. Une méthode de Jenks[48] a été utilisée pour produire les classes des histogrammes. D'après [46].

[8] Les classements des meilleures universités selon chaque édition mais aussi selon le PageRank, CheiRank et 2DRank sont accessibles depuis : http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/

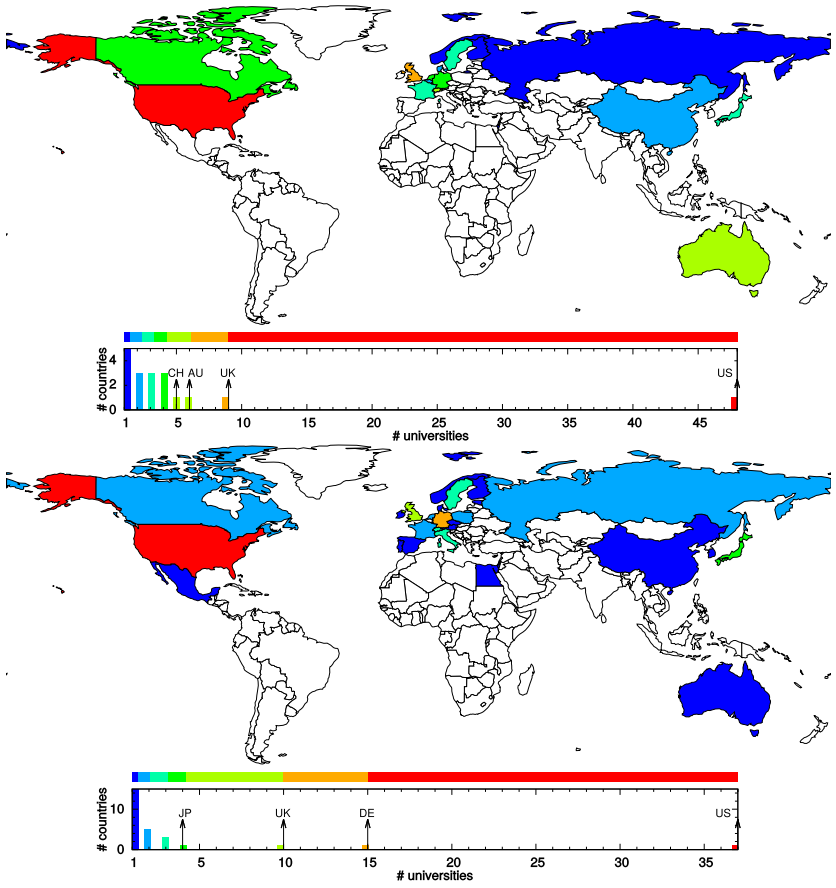

FIGURE 3.3 : Distribution géographique des 100 meilleures universités du classement ARWU17 (en haut) et du classement WRWU17 (en bas). Les universités US sont les plus représentées avec 48 pour ARWU17 et 37 pour WRWU17. Une méthode de Jenks[48] a été utilisée pour produire les classes des histogrammes. D'après [46].

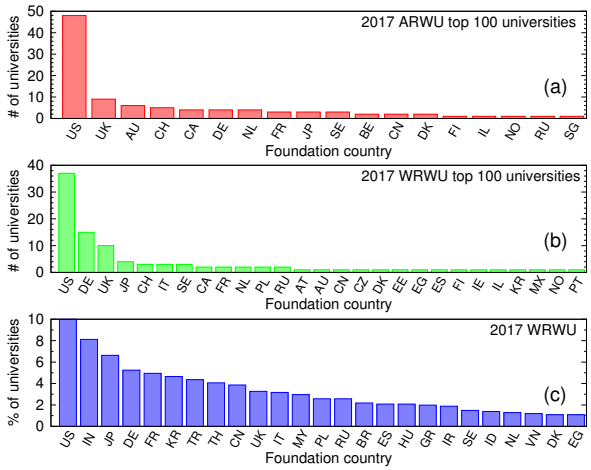

FIGURE 3.4 : Distribution statistique des universités pour (a) le top 100 ARWU17 et (b) le top 100 PageRank WRWU17. Le graphe (c) donne par pays le pourcentage d'universités présentes dans WRWU17. Les pays dont les universités ne représentent que moins de 10% du total des universités recensées dans WRWU17 ne sont pas représentés. Si deux pays sont ex-æquo alors ils sont interclassés par ordre alphabétique. Les codes pays ISO-3166-2 sont données dans la Table 3.5. D'après [46].

3.1.4 Influences mondiales et interactions des universités

Afin d'étudier les interactions entre les universités et leur influence culturelle sur le monde, nous avons utilisé la méthode de la matrice de Google réduite. L'application de cette méthode aux réseaux d'articles Wikipédia et à d'autres réseaux dirigés a montré son efficacité à mesurer les interactions indirectes entre des éléments d'intérêts d'un réseau. Pour plus de détails, le lecteur est invité à revoir la Section 2.7. Nous rappelons que $G_{\rm r} = G_{\rm rr} + G_{\rm pr} + G_{\rm qr}$ est la matrice de Google réduite de taille $N_{\rm r} \times N_{\rm r}$, où $N_{\rm r}$ est le nombre de nœuds réduits. Enfin, la matrice G_{qrnd} est définie telle que

$$G_{{\rm qrnd}_{ij}} = G_{{\rm qr}_{ij}}(1 - \delta_{ij}). \tag{3.4}$$

Influence culturelle des universités selon ENWIKI17

Afin de mesurer l'influence mondiale des universités les mieux classées du réseaux Wikipédia, nous nous sommes intéressés à l'édition anglaise de Wikipédia 2017 (ENWIKI17). Il s'agit de l'édition linguistique la plus complète, comptant plus de 5 millions d'articles. Nos nœuds d'intérêt sont les 20 meilleures universités provenant du classement $\mathcal{R}_{EN,PR}$ (voir Table 3.4). Nous avons choisi les 85 pays (voir Table 3.5) dont les universités sont représentées dans le classement WRWU17. La mesure de l'influence d'une université u dans le monde est obtenue en mesurant la sensibilité du PageRank, introduite à la section 2.6, pour chacun des 85 pays. Pour cela, nous perturbons le lien $u \to c$, avec c un pays donné, et mesurons la sensibilité $\mathcal{D}_{u\to c}(c) = d\ln(P_c)/d\delta$. L'efficacité de cette méthode pour déterminer les sensibilités des probabilités PageRank est montrée dans les articles [44] et [49] notamment. La matrice $G_{\rm r}$ et ses composantes sont représentées à la Figure 3.5. Nous définissons le poids $W_{\rm x}$ associé à la matrice $G_{\rm x}$ de taille $N_{\rm r} \times N_{\rm r}$ tel que $W_{\rm x} = \frac{1}{N_{\rm r}} \sum_{ij} G_{{\rm x}ij}$. Il est intéressant de voir que le poids $W_{\rm rr}$, associé à la matrice des liens directs, est plus petit que celui associé à $G_{\rm qr}$, $W_{\rm rr} < W_{\rm qr}$.

Rang	Université	Rang	Université
1	**Harvard**	11	**Michigan**
2	**Oxford**	12	**Cornell**
3	**Cambridge**	13	**Californie, Los Angeles**
4	**Columbia**	14	**Pennsylvanie**
5	**Yale**	15	**NYU**
6	**Stanford**	16	**Texas Austin**
7	**MIT**	17	**Florida**
8	**Californie, Berkeley**	18	**Edinburgh**
9	**Princeton**	19	**Wisconsin-Madison**
10	**Chicago**	20	**Californie du sud**

TABLE 3.4 : Liste des 20 meilleures universités selon le classement l'édition ENWIKI17 avec l'algorithme PageRank (ENWRWU17). Les universités britanniques sont en violet, les universités de l'Ouest américain sont en rouge, les universités situées dans la partie centrale des US sont en orange et les universités de l'Est américain sont en bleu. D'après [46].

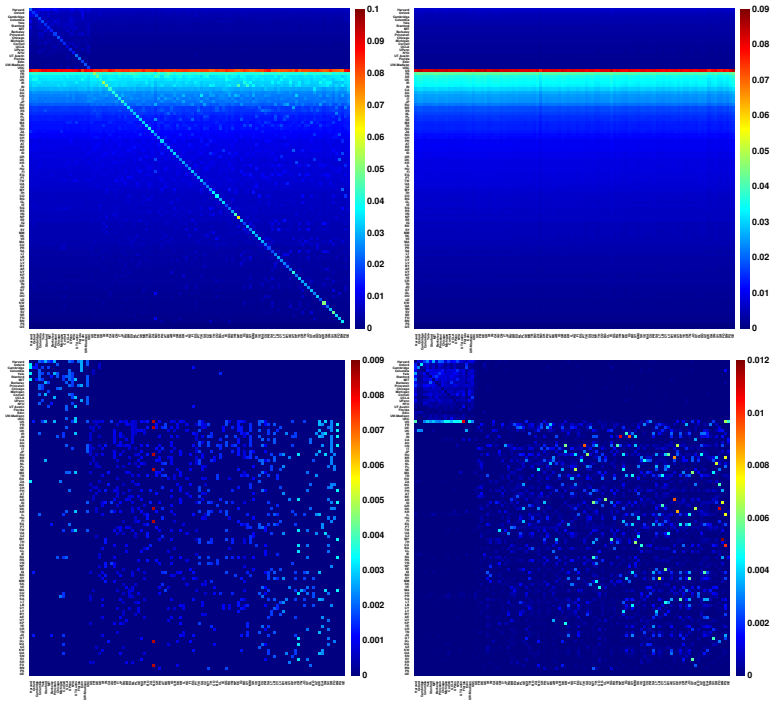

FIGURE 3.5 : Matrice de Google réduite G_{r} pour les 20 premières universités du classement ENWRWU17 (voir Table 3.4) et les 85 pays (voir Table 3.5). La matrice G_{r} est représentée en haut à gauche, G_{pr} en haut à droite, G_{rr} en bas à gauche et G_{qrnd} en bas à droite. Les poids associés à la matrice G_{r} et à ses composantes sont $W_{\mathrm{r}} = 1$, $W_{\mathrm{pr}} = 0.948273$, $W_{\mathrm{rr}} = 0.0144137$, et $W_{\mathrm{qr}} = 0.0373132$. Les colonnes et lignes sont classées selon l'ordre de ENWRWU17 (voir Table 3.4) puis selon le classement PageRank des pays avec ENWIKI17 (voir Table 3.5). D'après [46].

Rang	Pays	CC	Rang	Pays	CC
1	États-Unis	US	44	Chili	CL
2	France	FR	45	Irlande	IE
3	Allemagne	DE	46	Singapour	SG
4	Royaume-Unis	UK	47	Serbie	RS
5	Iran	IR	48	Vietnam	VN
6	Inde	IN	49	Népal	NP
7	Canada	CA	50	Estonie	EE
8	Australie	AU	51	Irak	IQ
9	Chine	CN	52	Bangladesh	BD
10	Italie	IT	53	Syrie	SY
11	Japon	JP	54	Myanmar	MM
12	Russie	RU	55	Slovaquie	SK
13	Brésil	BR	56	Venezuela	VE
14	Espagne	ES	57	Maroc	MA
15	Pays-Bas	NL	58	Cuba	CU
16	Pologne	PL	59	Puerto Rico	PR
17	Suède	SE	60	Arabie saoudite	SA
18	Mexique	MX	61	Lituanie	LT
19	Turquie	TR	62	Liban	LB
20	Roumanie	RO	63	Chypre	CY
21	Afrique du Sud	ZA	64	Lettonie	LV
22	Norvège	NO	65	Biélorussie	BY
23	Suisse	CH	66	United Arab Emirates	AE
24	Philippines	PH	67	Uruguay	UY
25	Autriche	AT	68	Corée du Nord	KP
26	Belgique	BE	69	Yémen	YE
27	Argentine	AR	70	Costa Rica	CR
28	Indonésie	ID	71	Tunisie	TN
29	Grèce	GR	72	Jordanie	JO
30	Danemark	DK	73	Guatemala	GT
31	Corée du Sud	KR	74	Groenland	GL
32	Israël	IL	75	République Dominicaine	DO
33	Hongrie	HU	76	Ouzbékistan	UZ
34	Finlande	FI	77	Koweït	KW
35	Egypte	EG	78	Qatar	QA
36	Portugal	PT	79	Sénégal	SN
37	Taïwan	TW	80	El Salvador	SV
38	Ukraine	UA	81	Suriname	SR
39	République Tchèque	CZ	82	Îles Féroé	FO
40	Malaisie	MY	83	Brunei	BN
41	Thaïlande	TH	84	Paléstine	PS
42	Colombie	CO	85	Géorgie	GE
43	Bulgarie	BG			

TABLE 3.5 : Liste des pays dont les universités sont présentes dans le classement WRWU17. Les pays sont classés via PageRank de ENWIKI17. D'après [46].

La Figure 3.6, montre la distribution géographique de l'influence culturelle de quatre universités sur les 85 pays sélectionnés. Les universités que nous avons testé sont celles de Harvard, Chicago, Stanford et Oxford. Pour Harvard, le pays c dont le PageRank est le plus impacté par une perturbation du lien Harvard $\rightarrow c$ est l'Afrique du Sud (ZA). On peut noter un gap conséquent entre la sensibilité de ZA (0.0085) et des autres pays. Dans la page Wikipédia anglaise de Harvard, il est fait mention d'un scandale faisant le lien entre Harvard et un investissement dans l'apartheid sud africain.[9] Le second pays le plus impacté, Puerto Rico (PR), est cité dans la page wiki mais dans une zone de l'article qu'on ne prend pas en compte dans la construction du réseau. En effet, ce pays est cité dans la légende d'une image et apparaît comme étant en relation avec les universités américaines les plus anciennes. Les trois autres pays hautement impactés sont la Géorgie (GE), Israël (IL) et l'Irlande (IR). Ces pays n'apparaissent pas dans la page Wikipédia anglaise de Harvard. Les pays les plus sensibles à Harvard ne sont pas cités directement dans l'article. Ces résultats sont dus aux liens indirects forts entre ces pays et Harvard. L'exemple de Chicago est intéressant car parmi les trois pays les plus impactés, seul le 3$^{\text{ème}}$ est cité dans l'article Wikipédia de l'université de Chicago. En effet, l'Inde est mentionnée en raison de la présence d'un campus de Chicago en Inde et du fait qu'un ancien membre de l'université fut président de la banque centrale indienne. Bien que la valeur de sensibilité maximale dans le cas de Chicago soit plus petite que celle provenant de l'analyse de l'influence de Harvard ($\approx 50\%$ plus faible), on observe encore un gap entre le pays le plus impacté et les autres. Le pays le plus sensible est Singapour (SG), suivi par PR. Stanford impacte l'Espagne (ES), pays mentionné dans sa page wiki. Les deux autres pays sont PR et ZA, dont l'influence est due aux liens indirects. La seule université britannique présente dans cette analyse est Oxford. Les deux premiers pays les plus influencés par Oxford sont la Jordanie (JO) et l'Irak (IQ). Tous deux sont mentionnés. Le personnage politique jordanien, Abdullah II fut un étudiant d'Oxford ainsi que T.E. Lawrence, écrivain britannique ayant écrit sur l'histoire jordanienne et irakienne.

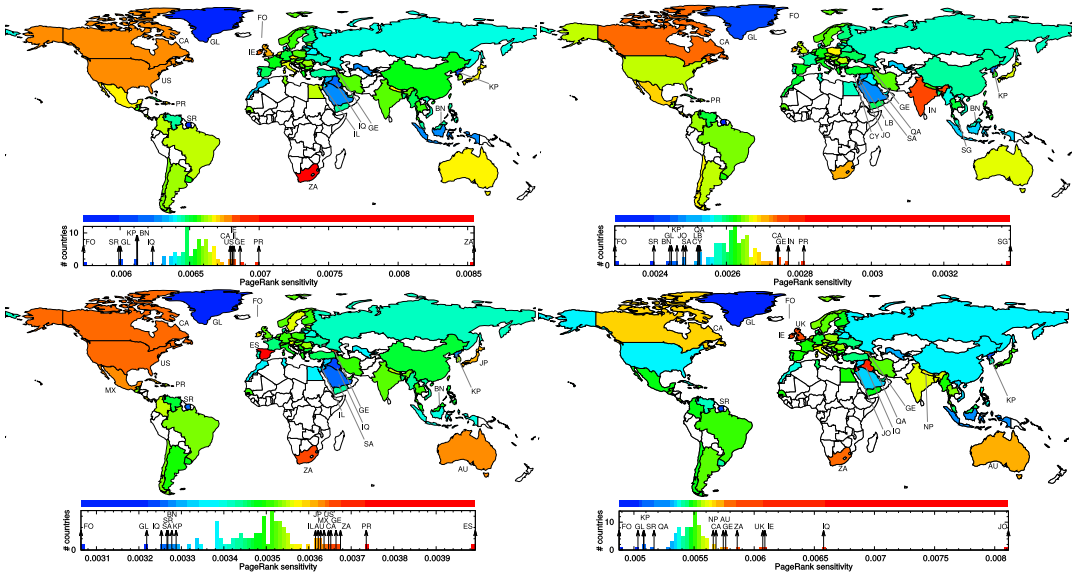

FIGURE 3.6 : Distribution géographique de sensibilité diagonale $\mathcal{D}_{u\rightarrow c}(c)$ vis à vis du lien $u \rightarrow c$ pour quatre universités : Harvard (en haut à gauche), Chicago (en haut à droite), Stanford (en bas à gauche) et Oxford (en bas à droite). Une méthode de Jenks[48] a été utilisée pour produire les classes des histogrammes. D'après [46].

[9] Dans les années 80, des étudiants activistes ont construit un bidonville dans la cour de l'université Harvard afin de protester contre les investissements de Harvard en Afrique du Sud visant à promouvoir indirectement l'apartheid.

La Figure 3.7 donne une distribution géographique pour les sensibilités non-diagonales d'autres universités. Dans le cas non-diagonal, la quantité $\mathcal{D}_{u\to c}(c')$ mesure la sensibilité d'un pays c vis à vis du lien $u \to c$. On étudie le cas de deux universités, $\mathcal{D}_{\text{Harvard}\to\text{US}}(c')$ et $\mathcal{D}_{\text{Oxford}\to\text{UK}}(c')$. Il est difficile de retrouver le chemin responsable d'une telle influence lorsqu'on regarde la sensibilité non-diagonale. Cependant, l'utilisation de la matrice de Google réduite permet d'effectuer très simplement une telle mesure dans un vaste réseau.

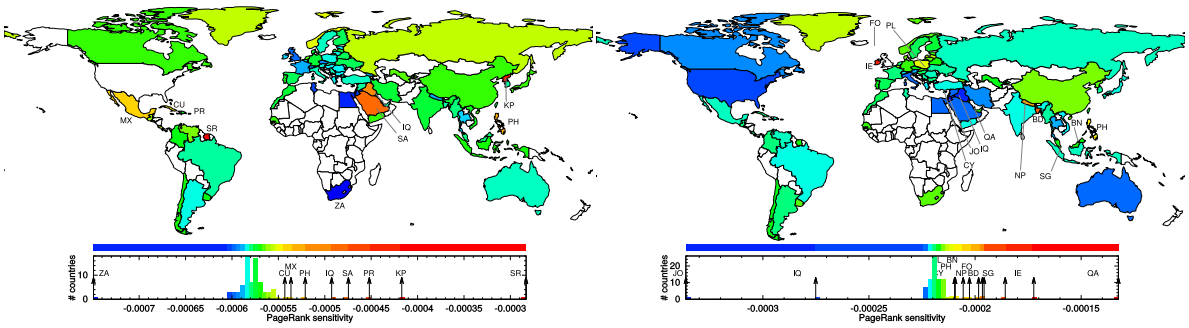

FIGURE 3.7 : Distribution géographique de la sensibilité non diagonale $\mathcal{D}_{u\to c}(c')$ vis à vis des liens Harvard→US (à gauche) et Oxford→UK (à droite). La couleur rouge (bleue) représente la valeur maximale (minimale) en valeur absolue. La sensibilité diagonale $\mathcal{D}_{u\to c}(c' = c)$ n'est pas affichée. Une méthode de Jenks[48] a été utilisée pour produire les classes des histogrammes. D'après [46].

L'utilisation des matrices G_{rr} et G_{qr} nous permet de définir la matrice G_{sum} représentant l'ensemble des liens directs et cachés. Elle se construit ainsi

$$G_{\text{sum}} = G_{\text{rr}} + G_{\text{qrnd}}. \tag{3.5}$$

À partir de G_{sum} (3.5), nous construisons le réseau réduit des relations entre universités et pays. La construction du réseau se fait de la manière suivante. Premièrement, nous définissons quatre catégories pour le top 20 des universités du classement ENWRWUI17 correspondant au code couleur de la Table 3.4. Le *leader* d'une catégorie est le nœud ayant le plus haut PageRank. Ensuite, pour chaque leader l, nous regardons les quatre éléments $G_{\text{sum}_{cl}}$ les plus grands,[10] avec c les indices lignes correspondant à des pays.

La Figure 3.8 représente le réseau d'influence université-pays. Chaque pays est coloré en fonction de sa langue dominante. Il est intéressant de voir l'apparition de liens entre des pays où l'on parle une langue distincte de celle parlée dans le pays où se situe l'université. Nous avons par exemple des pays de langue arabe (Jordanie et Iraq) autour d'Oxford, de langue hébraïque (Israël) pointé par Harvard et Stanford ou encore deux pays où l'on parle l'espagnol (Puerto Rico et Espagne) autour de Chicago et Stanford. Enfin, nous retrouvons Singapour (dont l'une des langues est le mandarin) et l'Inde (indien) qui sont pointés par Chicago. Parmi les 11 pays présents dans ce réseau réduit, seulement 4 sont anglophones (Irlande, Royaume-Unis, Afrique du Sud et US).

[10] Par construction, la matrice G_{qr} peut avoir des éléments négatifs. La somme des G_{rr}, G_{pr} et G_{qr} donne une matrice de Google. G_{sum} peut également avoir des entrées négatives, aussi on cherche à regarder les 4 entrées positives les plus élevées.

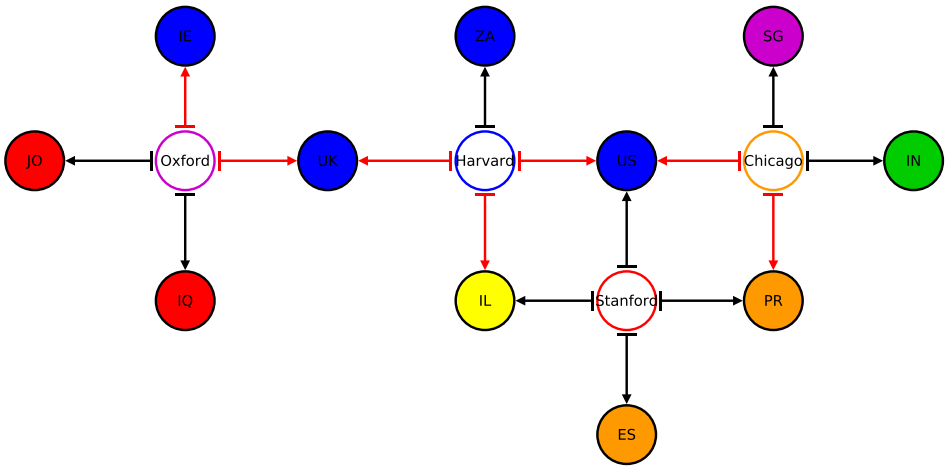

FIGURE 3.8 : Réseaux réduits en utilisant G_{sum} pour un ensemble de 20 universités (top 20 ENWRWU) et 85 pays. Pour chaque leader régional, Stanford, Chicago, Harvard et Oxford, les 4 liens sortants les plus forts pointant vers un pays sont affichés. Les universités (pays) sont représentées par des nœuds vides (pleins). Le code couleur pour les pays dépend de la langue parlée majoritairement : **bleu** pour l'anglais, **rouge** pour l'arabe, **orange** pour l'espagnol, **violet** pour le chinois, **vert** pour l'hindi, et **Jaune** pour l'hébreu. Les liens en rouge sont purement indirects et en noir, directs. D'après [46].

Réseaux réduits des universités

Nous nous intéressons maintenant, dans le contexte de 4 éditions linguistiques différentes (EN, FR, DE et RUWIKI), aux interactions entre les meilleures universités, classées selon le PageRank associé à chacune de ces éditions. Les listes présentant ces tops 20 sont disponibles dans la Table 3.4, la Table 3.6, la Table 3.7 et la Table 3.8. Ces listes sont différentes, bien que certaines universités peuvent apparaître dans plusieurs d'entre elles. Afin de voir si le choix de l'édition modifie les interactions directes et indirectes entre les universités, nous avons analysé une liste de 52 universités (voir Table 3.9. Cette liste est tout simplement l'union des 4 tops 20. Si une université n'apparaît pas dans une des 3 autres éditions, on la retire alors de la liste finale.

Top 20 ENWIKI

Les éléments des matrices G_{rr} et G_{qrnd} sont représentés à la Figure 3.9. Les interactions indirectes entre le top 20 ENWIKI sont plus importantes que leurs interactions directes. En effet, le poids W_{qrnd} est 50% plus élevé que le poids W_{rr}. De la même manière que pour le réseau réduit des relations entre universités et pays (voir la Figure 3.10), nous avons utilisé la matrice G_{sum}. Nous optons ici pour une méthode de construction du réseau réduit plus complexe, afin d'avoir le plus d'information possible.

1. Les leaders sont placés autour d'un cercle de rayon arbitraire.
2. On place les quatre liens sortants les plus forts (selon G_{sum}) pour chacun de ces leaders.
3. Pour chaque nœud cible qui n'est pas déjà présent dans le réseau en construction, on l'ajoute et on le place sur un cercle centré sur le nœud source (si possible de même catégorie).
4. Reprendre à l'étape 1 pour chaque nouveau nœud ajouté.

5. La construction est terminée lorsqu'il n'y a plus de nouveau nœud à ajouter.

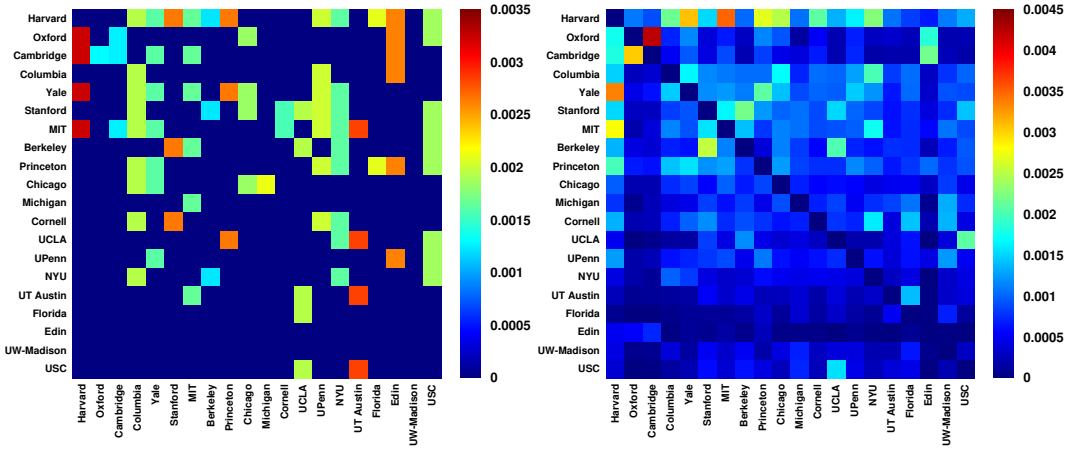

FIGURE 3.9 : Matrices G_{rr} (à gauche) et G_{qrnd} (à droite) pour le top 20 ENWRWU (voir Table 3.4). Les colonnes et lignes sont ordonnées selon cette même liste. Les poids des matrices sont $W_{\text{rr}} = 0.00877$ et $W_{\text{qrnd}} = 0.01381$. Le code couleur dépend des éléments des matrices. Les figures relatives au 3 autres éditions sont disponibles en annexe (voir la Figure A.1 pour FRWIKI, Figure A.2 pour DEWIKI et Figure A.3 pour RUWIKI). D'après [46].

Le réseau réduit d'amis pour ENWIK17 est présenté à la Figure 3.10. On peut y distinguer des liens purement cachés[11] ou bien indirects entre u et u', tels que $A_{u'u} = 0$, et des liens directs. Dans ce premier exemple, il est intéressant de voir que les différents leaders n'arrivent pas à capturer les universités de leur région. Celle de Chicago, leader de la région centrale, est isolée. La seule autre université de sa catégorie, UT Austin, est capturée par la communauté centrée sur Oxford. Harvard capture seulement deux universités de sa propre catégorie (MIT et Yale) et ne capture rien d'autre au delà du premier niveau de construction. Stanford capture une seule université de sa région (Berkeley) mais aussi certaines appartenant à la catégorie de l'est des USA (NYU, Cornell et Columbia). L'université la plus intéressante est Oxford. Premièrement, ce leader réussit à capturer toutes les universités de sa catégorie. Deuxièmement, on peut voir qu'Oxford capture également des universités de toutes catégories (Princeton, UCLA, USC et UT Austin). Les liens cachés (liens rouges) sont caractéristiques des liens entre universités issues de communautés différentes[12]. Ainsi, nous avons par exemple Stanford→MIT, Chicago→Harvard, Oxford→Harvard ou encore Princeton→Stanford. Par ailleurs, certains liens cachés viennent renforcer une relation entre universités. De cette façon, lorsque le lien depuis Edinburgh vers Cambridge est direct, le lien inverse est indirect. Il en va de même pour Oxford et Edinburgh.

[11] Un lien purement caché $j \to i$ est simplement la composante $G_{\text{sum}\,ij}$ telle que $A_{ij} = 0$. Dans ce cas, $G_{\text{rr}\,ij} = \frac{(1-\alpha)}{N}$ du fait de la construction de la matrice de Google associée au réseau global.

[12] Les nœuds réduits sont colorés selon des catégories tandis qu'une communauté est un cluster de nœuds gravitant autour d'un leader.

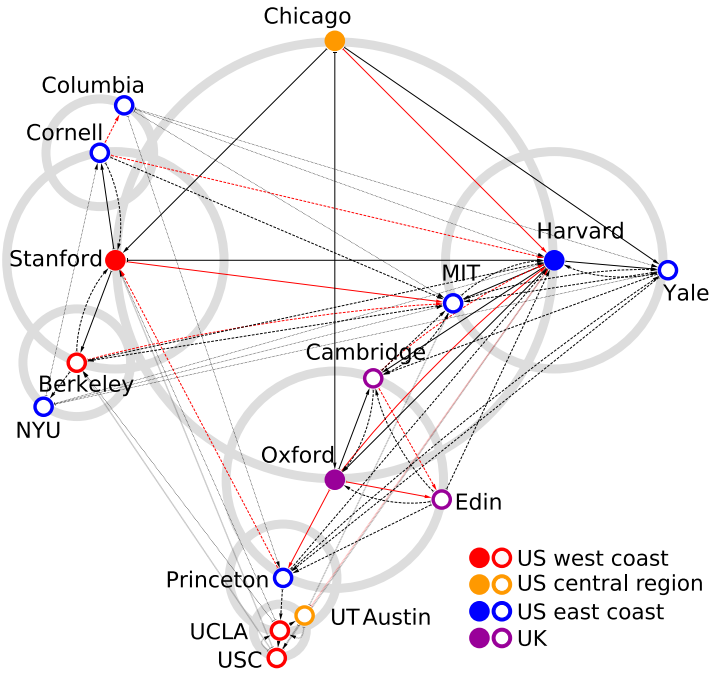

FIGURE 3.10 : Réseau réduit du top 20 ENWRWU à partir de G_{sum}. Les nœuds pleins représentent les leaders régionaux. Les liens rouges sont des liens purement indirects, i.e. sans existence dans la base de données. Il y a un total de 4 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, les lignes en traits sont pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour le $4^{\text{ème}}$. D'après [46].

Top 20 FRWIKI

Le réseau réduit des interactions entre les 20 meilleures universités selon FRWIKI17 est présenté à la Figure 3.11. Ces 20 universités listées dans la Table 3.6 sont situées dans 5 pays différents : US, Royaume-Uni (UK), France (FR), Canada (CA) et Belgique (BE). On observe des communautés très homogènes. En effet, les leaders des différents pays ont capturé des universités de même catégorie. À l'instar de ENWIKI17, les liens entre les communautés sont préférentiellement indirects. Les liens entre communautés se dirigent principalement vers des universités anglo-saxonnes. Par exemple, l'université de Montréal et de Laval (CA) pointent indirectement vers les communautés centrées sur Harvard (US) et Oxford (UK). De même pour les universités belges, avec UC Louvain et ULB qui respectivement ont des liens indirects avec la communauté de Harvard et celle d'Oxford. Il est également intéressant de noter la présence de liens en provenance de l'université de Montréal et de l'ULB se dirigeant vers la communauté des universités françaises, soulignant ainsi leur histoire linguistique commune. La communauté centrée autour de Polytechnique fait exception car ses liens sortants vers d'autres communautés sont uniquement des liens directs, tout en étant également dirigés vers les communautés anglo-saxonnes. La place de Princeton dans ce réseau est très intéressante. Princeton est pointée par des universités de toutes les communautés et la plupart de ces liens sont des liens indirects. La place de Princeton est bien plus centrale dans FRWIKI que dans ENWIKI. Une propriété intéressante de ce réseau réduit est le fait que l'édition française place les universités anglo-saxonnes dans un sous-espace invariant, dans lequel un surfeur aléatoire resterait bloqué indéfiniment.

Rang	Université	Rang	Université
1	**Harvard**	11	**Laval**
2	**Oxford**	12	**Panthéon-Sorbonne**
3	**École polytechnique**	13	**Princeton**
4	**Cambridge**	14	**Californie, Berkeley**
5	**École normale supérieure**	15	**Paris-Sorbonne**
6	**MIT**	16	**Université libre de Bruxelles**
7	**Yale**	17	**Montréal**
8	**Columbia**	18	**Université catholique de Louvain**
9	**Stanford**	19	**Paris-Nanterre**
10	**École pratique des hautes études**	20	**Chicago**

TABLE 3.6 : Liste des 20 meilleures universités selon le classement PageRank de l'édition FRWIKI17 (FRWRWU17). Le code couleur correspond aux différentes catégories : bleu pour US, violet pour UK, rouge pour FR, vert pour CA et jaune pour BE. D'après [46].

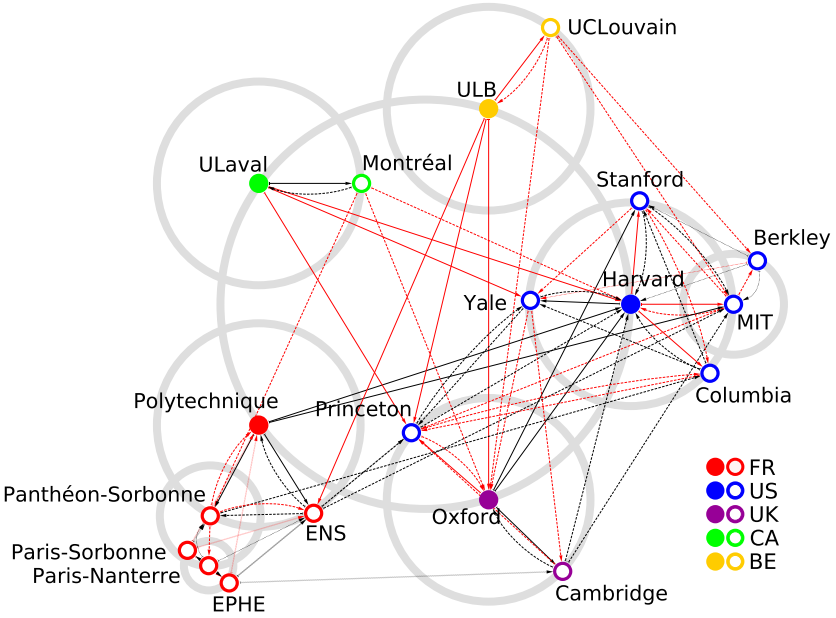

FIGURE 3.11 : Même figure que la Figure 3.10 mais dans le cas du top 20 FRWRWU (voir Table 3.6). Les nœuds pleins représentent les leaders régionaux. Les liens rouges sont des liens purement indirects, i.e. sans existence dans la base de données. Il y a un total de 4 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, lignes en traits pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour le $4^{\text{ème}}$. D'après [46].

Top 20 DEWIKI

Le réseau réduit des 20 meilleures universités vue par DEWIKI17, présentée à la Figure 3.12, montre le fort patriotisme allemand dont on a discuté plus haut, dans la section concernant le classement WRWU17. En effet, dans la liste de ces universités, présente dans la Table 3.7, seulement 15% de ces universités ne sont pas localisées en Allemagne. Parmi ces universités non-allemandes, on retrouve Harvard, Oxford et Cambridge. On distingue alors 4 catégories, correspondant à 4 pays : Autriche (AT), Allemagne (DE), UK et US. Les universités allemandes sont principalement placées dans la communauté centrée sur LMU Munich mais sont aussi capturées par chacun des autres leaders. Ainsi, Harvard capture Hambourg, Vienne capture Leipzig, Cologne, Bonn et Münster. Enfin, Oxford capture Tübingen, Marburg et GU Francfort. Le réseau réduit des interactions entre le top 20 des universités selon DEWIKI17

est majoritairement composé de liens indirects, que ce soit pour les interactions au sein d'une communauté qu'entre communautés.

Rang	Université	Rang	Université
1	**Munich**	11	**Fribourg**
2	**Humboldt, Berlin**	12	**Cologne**
3	**Göttingen**	13	**Münster**
4	**Heidelberg**	14	**Oxford**
5	**Université Libre de Berlin**	15	**Hambourg**
6	**Vienne**	16	**Frankfurt**
7	**Tübingen**	17	**Cambridge**
8	**Harvard**	18	**Marburg**
9	**Bonn**	19	**Kiel**
10	**Leipzig**	20	**Jena**

TABLE 3.7 : Liste des 20 meilleures universités selon le classement PageRank de l'édition DEWIKI17 (DEWRWU17). Le code couleur correspond aux différentes catégories : vert pour DE, bleu pour US, violet pour UK, noir pour AT. D'après [46].

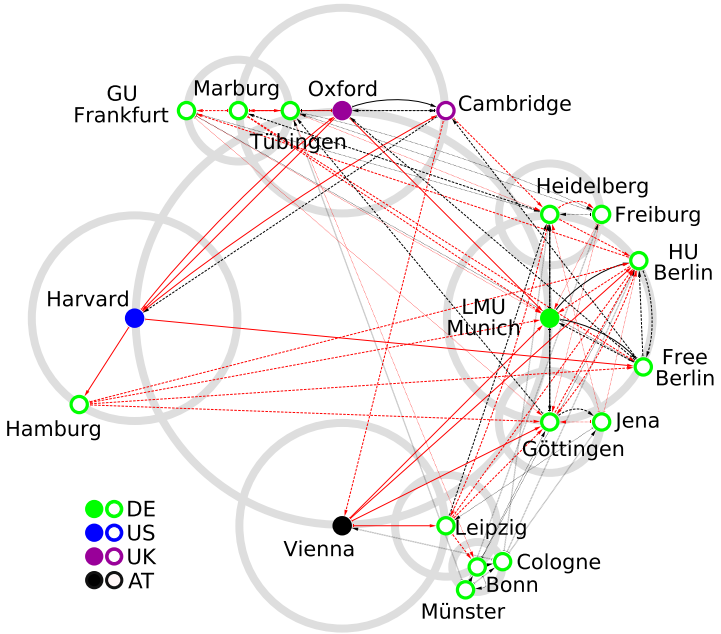

FIGURE 3.12 : Même figure que la Figure 3.10 mais dans le cas du top 20 DEWRWU (voir Table 3.4). Les nœuds pleins représentent les leaders régionaux. Les liens rouges sont des liens purement indirects, i.e. sans existence dans la base de données. Il y a un total de 4 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, les lignes sont en trait pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour le $4^{\text{ème}}$. D'après [46].

Top 20 RUWIKI

Le top 20 des universités selon RUWIKI17 (voir Table 3.8) contient plus d'universités situées dans des pays différents que les autres tops 20. Nous avons 6 catégories : Russie (RU), US, UK, DE, Ukraine (UA) et AT. Le réseau réduit des interactions entre ces universités est présenté à la Figure 3.13. Les communautés sont homogènes même si les universités américaines

sont distribuées dans trois communautés. On trouve Berkeley et Chicago autour de Vienne et enfin Stanford autour d'Oxford. Les liens entre communautés sont encore majoritairement issus de la composante G_{qr}. On observe tout de même des liens directs entre communautés, notamment pour Kiev→Moscou SU et Kiev→St. Pétersbourg. Ces universités appartenaient historiquement à l'URSS et il est donc intéressant de voir cette information émerger dans ce réseau ; le fait que ces liens soient directs ne semble pas être anodin dans le cas de l'édition russe. Les deux universités allemandes présentes, HU Berlin et Leipzig, étaient toutes deux localisées en République démocratique allemande dans le passé et ont ici des liens sortants dirigés vers la communauté centrée sur Moscou SU, soulignant ainsi l'importance de l'histoire du bloc soviétique dans RUWIKI.

Rang	Université	Rang	Université
1	**Moscou SU**	11	**Kazan FU**
2	**Saint-Pétersbourg SU**	12	**NU Kharkiv**
3	**Harvard**	th	**Stanford**
4	**Oxford**	14	**Princeton**
5	**Cambridge**	15	**Chicago**
6	**MIT**	16	**Higher School of Economics**
7	**Yale**	17	**Bauman**
8	**Columbia**	18	**Leipzig**
9	**Kiev**	19	**Vienne**
10	**Humboldt, Berlin**	20	**Californie, Berkeley**

TABLE 3.8 : Liste des 20 meilleures universités selon le classement PageRank de l'édition RUWIKI17 (RUWRWU17). Le code couleur correspond aux différentes catégories : rouge pour RU, bleu pour US, violet pour UK, vert pour DE, noir pour AT et jaune pour UA. D'après [46].

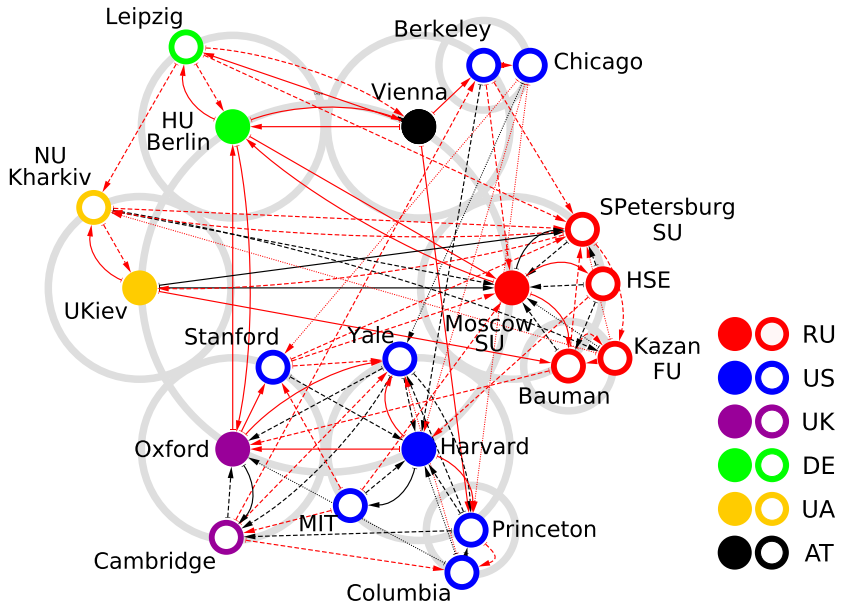

FIGURE 3.13 : Même figure que la Figure 3.10 mais dans le cas du top 20 RUWRWU17 (voir Table 3.8). Les nœuds pleins représentent les leaders régionaux. Les liens rouges sont des liens purement indirects, i.e. sans existence dans la base de données. Il y a un total de 3 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, les lignes sont en trait pour le second niveau et en pointillé pour le $3^{\text{ème}}$. D'après [46].

Comparaison multi-linguistique

La liste des 52 universités (voir Table 3.9) représentant l'union des 4 listes de 20 universités analysées plus haut, nous permet de mesurer qualitativement l'impact du choix de l'édition linguistique sur les interactions directes et indirectes obtenues avec REGOMAX. Nous allons comparer les réseaux réduits et matrices de Google réduites obtenus avec EN, FR, DE et RUWIKI17.

Rang	Université	Rang	Université
1	**Harvard**	27	**Göttingen**
2	**Columbia**	28	**Heidelberg**
3	**Yale**	29	**Université Libre de Berlin**
4	**Stanford**	30	**Tübingen**
5	**MIT**	31	**Bonn**
6	**Californie, Berkeley**	32	**Fribourg**
7	**Princeton**	33	**Cologne**
8	**Chicago**	34	**Münster**
9	**Michigan**	35	**Hambourg**
10	**Cornell**	36	**Francfort**
11	**Californie, Los Angeles**	37	**Marburg**
12	**Pennsylvanie**	38	**Kiel**
13	**NYU**	39	**Jena**
14	**Texas Austin**	40	**Oxford**
15	**Floride**	41	**Cambridge**
16	**Wisconsin–Madison**	42	**Edinburgh**
17	**Californie du sud**	43	**Laval**
18	**École polytechnique**	44	**Montréal**
19	**École normale supérieure**	45	**Moscou SU**
20	**École pratique des hautes études**	46	**Saint-Pétersbourg SU**
21	**Panthéon-Sorbonne**	47	**Kazan FU**
22	**Paris-Sorbonne**	48	**Bauman**
23	**Paris-Nanterre**	49	**Vienne**
24	**Humboldt, Berlin**	50	**Université Libre de Bruxelles**
25	**Leipzig**	51	**Kiev**
26	**Munich**	52	**NU Kharkiv**

TABLE 3.9 : Liste de 52 universités, union des top 20 EN, FR, DE et RUWRWU17. Code couleur : **US**, **FR**, **DE**, **UK**, **CA**, **RU**, **AT**, **BE** and **UA**. Les universités sont classées par pays et pour un même pays, les universités sont classées selon ENWRWU17. D'après [46].

Ces universités sont localisées dans 9 pays, US, FR, DE, UK, CA, RU, AT, BE et UA. Les matrices G_r pour chaque édition sont représentées dans la Figure 3.14. Les lignes remplies de poids forts ne sont pas les mêmes d'une édition à l'autre. Ce résultat est bien évidemment attendu puisque le classement des universités par PageRank n'est pas le même.

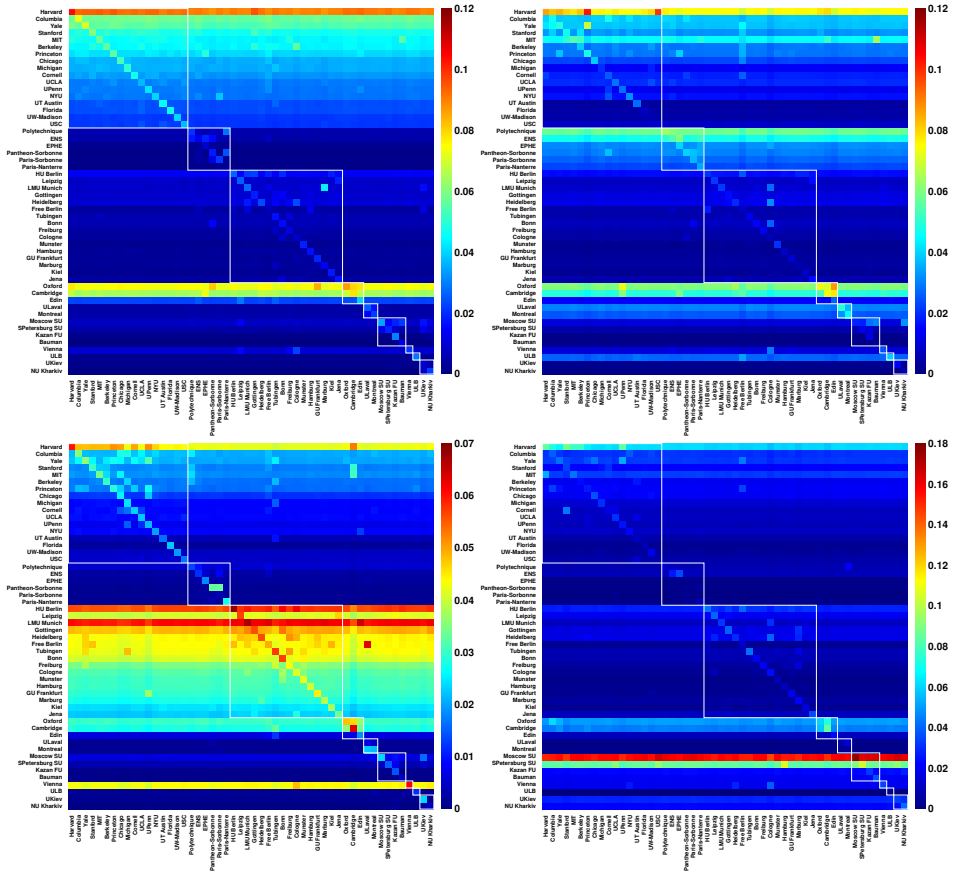

FIGURE 3.14 : Matrice de Google réduite G_{r} pour l'ensemble des 52 universités présentes dans la Table 3.9. À partir de ENWIKI (en haut à gauche), FRWIKI (en haut à droite), DEWIKI (en bas à gauche) et RUWIKI (en bas à droite) ; Les colonnes et lignes sont ordonnées selon la Table 3.9. D'après [46].

Il est intéressant de comparer les différentes matrices G_{qrnd}. En effet, selon la Figure 3.15, on observe une forte densité de poids forts dans les sous-régions G_{qr} représentant les interactions entre universités d'un même pays (carrés diagonaux blancs). En revanche, les interactions indirectes entre universités issues de pays différents sont bien plus faibles. L'édition russe donne une matrice G_{qrnd} avec un poids maximal 5 fois supérieur aux maxima des matrices G_{qrnd} associées aux autres éditions. Cette valeur maximale est reliée à l'interaction indirecte EPHE→ENS. On voit que les 4 éditions linguistiques considérées donnent des matrices G_{qrnd} relativement similaires. Certaines éditions permettent de mettre en lumière des relations cachées entre universités de pays différents dont les poids sont forts et la présence exclusive à l'édition. Ainsi, nous retrouvons GU Francfort→Harvard et Francfort→Columbia pour ENWIKI, Cologne→EPHE pour FRWIKI et aussi Kiev→Moscou SU et Kiev→St. Pétersbourg pour RUWIKI.

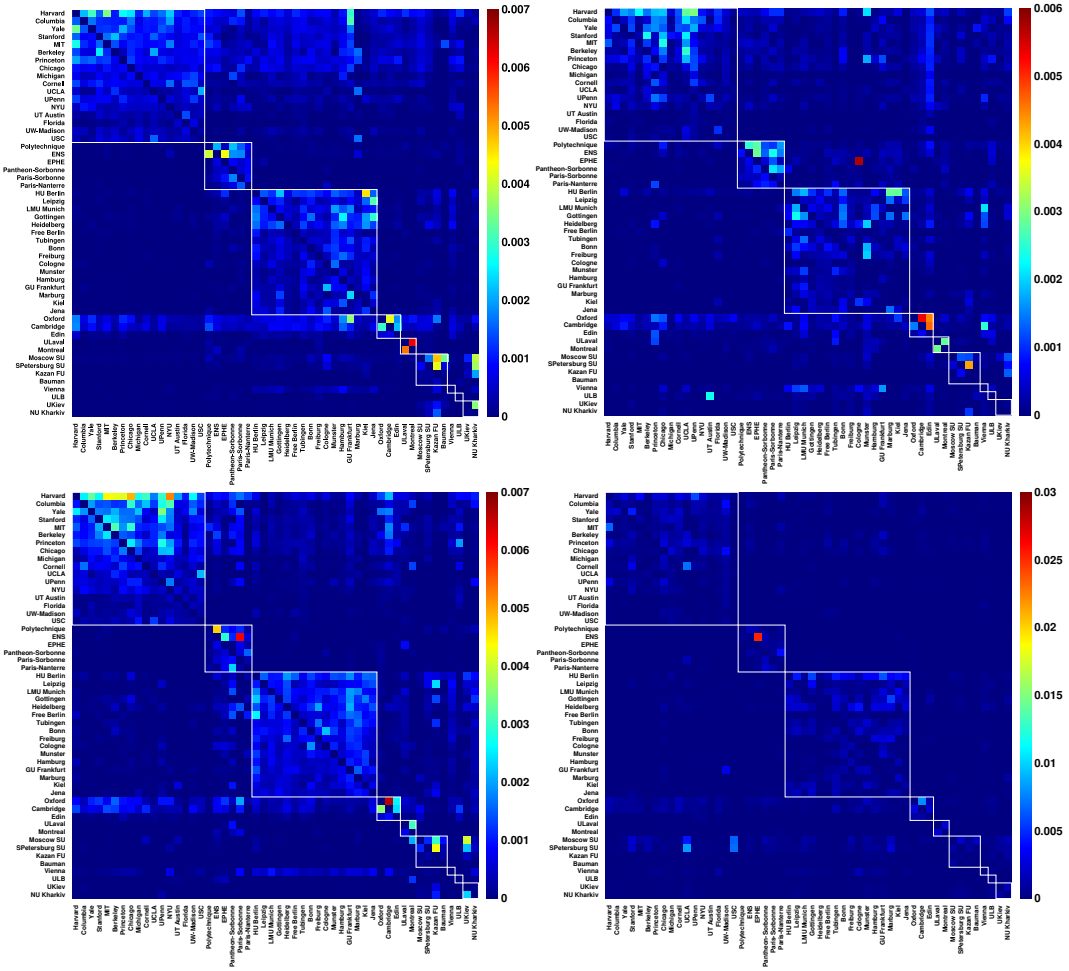

FIGURE 3.15 : Matrices G_{qrnd} pour les 52 universités de la Table 3.9. À partir de ENWIKI (en haut à gauche), FRWIKI (en haut à droite), DEWIKI (en bas à gauche) et RUWIKI (en bas à droite).

Le réseau réduit de ces 52 universités construit à partir ENWIKI est représenté à la Figure 3.16. Il compte un total de 33 universités sur les 52 possibles. Bien qu'issues de l'édition anglaise, on retrouve 6 universités allemandes contre 3 universités britanniques. La majorité des universités présentes dans ce réseau réduit sont américaines (13 universités). On observe que le réseau est constitué majoritairement de liens cachés. Nous avons la présence de deux communautés importantes (UK et US). Tout comme observé dans le cas du top 20 FRWIKI (voir la Figure 3.11), ces deux communautés forment un sous-espace invariant dont un surfeur aléatoire ne peut s'échapper. On retrouve aussi l'université de Chicago, isolée des autres universités américaines, comme observé pour ENWIKI17 et son top 20 d'universités (voir la Figure 3.10). En effet, elle est capturée par la communauté RU dans ce réseau réduit. Il y a une importante communauté centrée autour de HU Berlin et composée d'universités allemandes. Les deux seules universités françaises présentes dans ce réseau réduit sont dans la communauté centrée sur le leader de leur catégorie (Polytechnique). Par ailleurs, on peut noter que ces universités sont pointées uniquement par des universités françaises. En revanche, ENS et Polytechnique pointent préférentiellement vers des universités anglo-saxonnes. Ce réseau donne finalement une vision réaliste sur le faible rayonnement international des universités françaises face aux universités américaines, anglaises et allemandes.

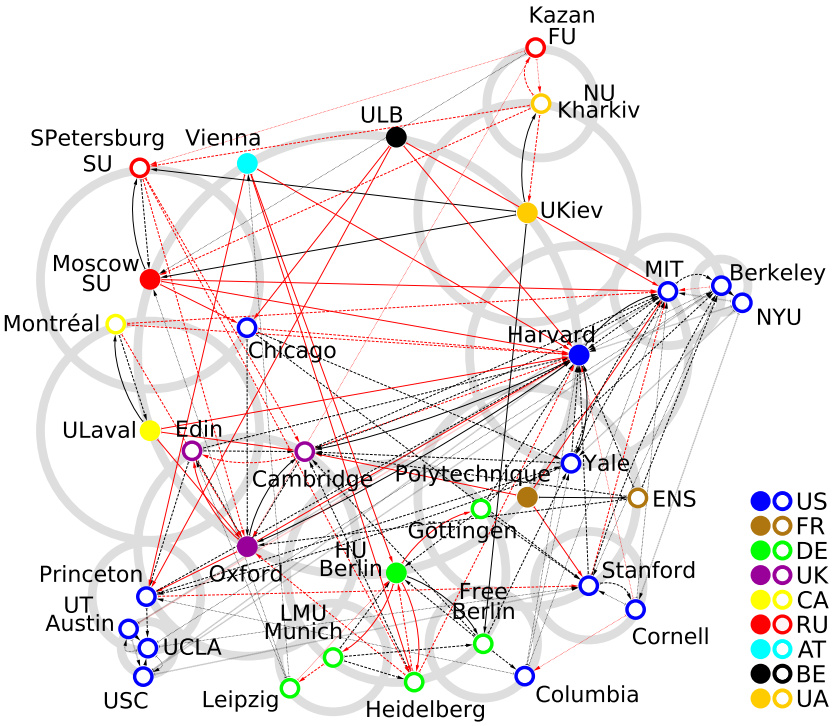

FIGURE 3.16 : Réseau réduit des 52 universités construit à partir de G_{sum} et de ENWIKI17. Les nœuds pleins représentent les leaders régionaux. Les liens rouges sont des liens purement indirects, i.e. sans existence dans la base de donnée. Il y a un total de 4 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, lignes en traits pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour le $4^{\text{ème}}$. D'après [46].

Le réseau réduit associé à FRWIKI est présenté à la Figure 3.17. On retrouve dans ce réseau 39 universités (sur les 52 possibles) avec une dominance marquée des universités allemandes, avec un total de 13 universités allemandes. Les liens indirects sont nombreux dans ce réseau réduit. Tout comme pour ENWIKI, on observe des communautés centrées autour des leaders US et UK, comme étant centrales et attractives.

La Figure 3.18 montre le réseau réduit dans le cas de DEWIKI. La dominance allemande est naturellement encore plus forte. On compte 14 universités DE. Les US sont deuxième en terme de nombre d'universités présentes dans le réseau avec 11 universités. Parmi les 52 universités possibles, 39 sont présentes dans ce réseau réduit. Les universités allemandes sont distribuées sur plusieurs communautés. On en retrouve autour des leaders AT, RU et DE. Les universités américaines sont distribuées dans trois communautés : US, RU et FR.

Enfin, le réseau réduit construit à partir de RUWIKI, présenté Figure 3.19, montre aussi une dominance allemande. On retrouve 16 universités DE et 10 US. Cette édition est celle qui donne le réseau réduit avec le plus grand nombre d'universités : 45. Dans ce réseau, on observe très clairement la présence de gros hubs. Un hub DE centré sur les universités HU Berlin, Heidelberg, et Göttingen, un hub US-UK centré sur les universités Harvard et Oxford, et un hub RU centré sur Moscou SU.

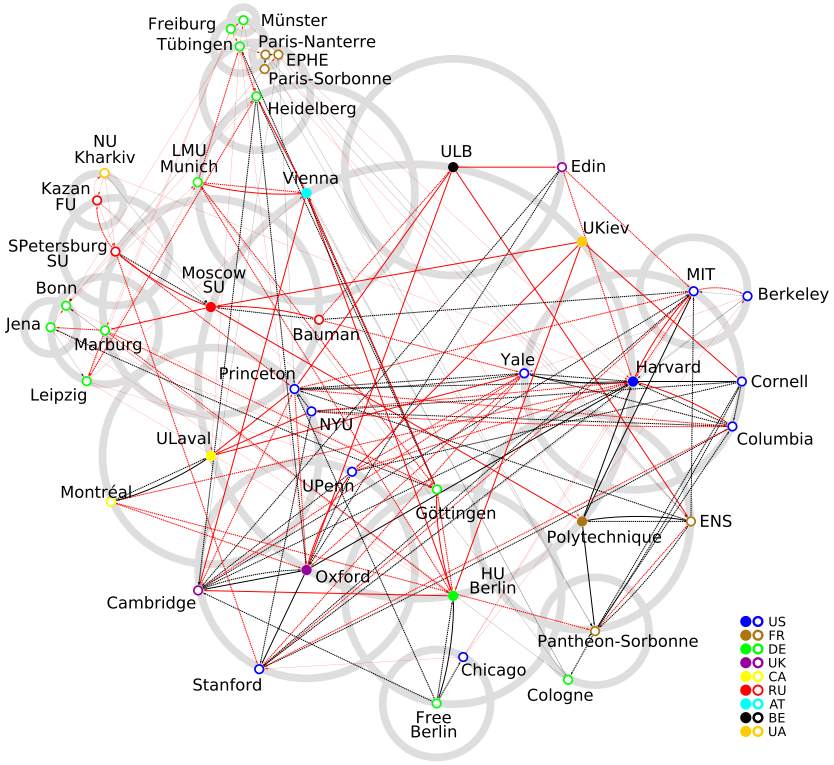

FIGURE 3.17 : Similaire à la Figure 3.16 mais appliqué à l'édition FRWIKI17. D'après [46].

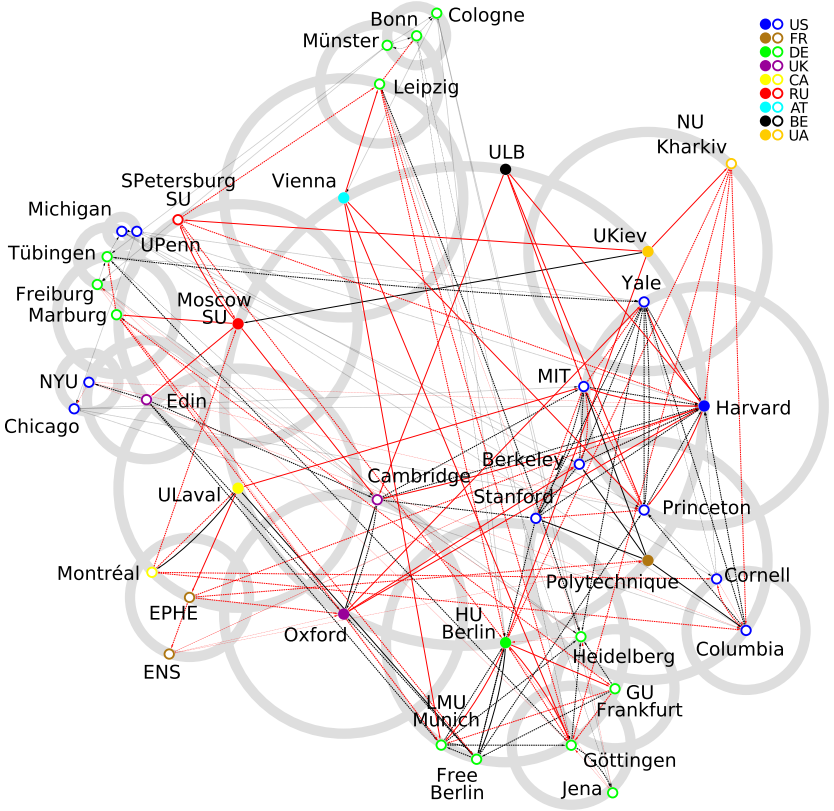

FIGURE 3.18 : Similaire à la Figure 3.16 mais appliqué à l'édition DEWIKI17. D'après [46].

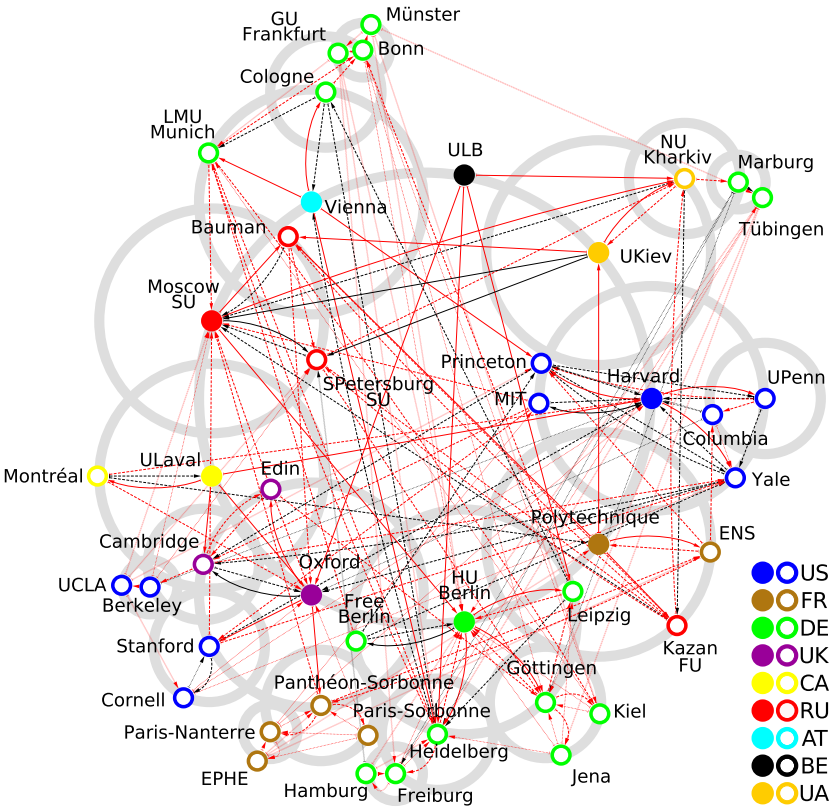

FIGURE 3.19 : Similaire à la Figure 3.16 mais appliqué à l'édition RUWIKI17. D'après [46].

3.1.5 Matrice de Google réduite multi-linguistique

Lorsqu'on analyse les réseaux Wikipédia d'éditions linguistiques distinctes séparément, des résultats caractéristiques sont observés. En analysant le réseau réduit de 52 universités pour chacune des 4 éditions étudiées plus haut, on note quelques similarités, comme si les interactions indirectes entre universités venait palier au manque d'information d'une édition donnée. Tout comme nous avons utilisé 24 éditions linguistiques afin de construire un classement mondial des universités, nous proposons de construire une matrice de Google réduite moyenne basée sur ces 24 mêmes éditions linguistiques. Cette matrice noté $\tilde{G}_{\mathrm{r}}$ est construite comme suit :

$$\tilde{G}_{\mathrm{r}} = \frac{1}{24} \sum_{E} G_{\mathrm{r}}^{(E)} \tag{3.6}$$

où E est une édition.

À partir de cette matrice de Google réduite moyenne, on peut construire $\tilde{G}_{\mathrm{sum}}$ et donc construire un réseau réduit des interactions entre les meilleures universités, en prenant en compte la richesse culturelle de 24 éditions linguistiques. Ainsi, les biais dus au manque d'informations d'une édition par rapport à d'autres sont limités. Nous prenons comme ensemble de nœuds réduits les 100 meilleures universités selon WRWU17 (voir la Table. 3.10). Il est important de choisir une même base pour les matrices de Google réduites relatives aux éditions utilisées. Pour cela, nous gardons l'ordre du classement PageRank WRWU17. Si une université u n'apparait pas dans une édition E, nous procédons ainsi

$$G_{\mathrm{pr}_{iu}}^{(\mathrm{E})} = \frac{1}{100} \forall i \tag{3.7}$$

$$G_{\mathrm{pr}_{uj}}^{(\mathrm{E})} = 0 \forall j \tag{3.8}$$

Le vecteur PageRank $\tilde{P}_{\mathrm{r}}$ associé à $\tilde{G}_{\mathrm{r}}$ est présenté à la Table 3.10. Le classement des universités obtenu avec $\tilde{G}_{\mathrm{r}}$ est différent de celui obtenu avec la méthode de classement WRWU17

(voir Table 3.2). Cependant, pour le top 10, nous retrouvons une similarité $\eta(10) = 100\%$ avec WRWU17 et 90% avec ARWU17. L'utilisation de $\tilde{G}_{\mathrm{r}}$ pour classer les universités ne donne plus les deux premières places à Oxford et Cambridge qui perdent chacune une place. Harvard devient l'université en tête de ce classement.

Rang	PageRank	Université	Rang	PageRank	Université
1	0.0633191	Harvard	51	0.00659655	Colorado Boulder
2	0.0528587	Oxford	52	0.00657266	Glasgow
3	0.0518905	Cambridge	53	0.00636839	Toronto
4	0.0339304	MIT	54	0.0063255	Stockholm University
5	0.0301911	Columbia	55	0.00624184	Tübingen
6	0.0283041	Yale	56	0.00609986	Texas Austin
7	0.0261455	Stanford	57	0.00593539	Virginia
8	0.024318	Californie, Berkeley	58	0.00584412	Imperial College London
9	0.0229394	Princeton	59	0.00582829	Carnegie Mellon
10	0.0215136	Chicago	60	0.00579437	Bonn
11	0.0197203	Copenhague	61	0.00570673	Minnesota
12	0.0168679	Humboldt, Berlin	62	0.00567465	Keio
13	0.0160439	Uppsala	63	0.00557384	Helsinki
14	0.0148231	Tokyo	64	0.00548871	King's College London
15	0.0135633	Moscou SU	65	0.0054485	Floride
16	0.0127305	Cornell	66	0.00538279	Zurich
17	0.0126064	HUJI[a]	67	0.00536546	Manchester
18	0.0125732	Pennsylvanie	68	0.00523928	McGill
19	0.0120329	Californie, Los Angeles	69	0.00507791	Université Libre de Berlin
20	0.011732	Leiden	70	0.00505635	Washington
21	0.011246	Caltech	71	0.00505447	Illinois U.-C.
22	0.0112404	NYU	72	0.00497258	Brown
23	0.0112273	Vienne	73	0.00491403	Wisconsin-Madison
24	0.0104997	Edinburgh	74	0.00485964	Northwestern
25	0.0103698	Jagiellonian	75	0.00480294	Coimbra
26	0.0101557	Bologne	76	0.00479832	Oslo
27	0.0100089	Göttingen	77	0.00477973	Padua
28	0.00987766	Heidelberg	78	0.00476805	Georgetown
29	0.00982921	Michigan	79	0.00475634	NAU Mexico
30	0.00974263	Lund	80	0.00468635	Boston
31	0.00929623	LSE[b]	81	0.0045985	Ohio SU
32	0.00918967	Johns Hopkins	82	0.00458516	Michigan SU
33	0.00909002	Varsovie	83	0.00452351	Genève
34	0.00902656	Séoul NU	84	0.00451385	Marburg
35	0.00877768	Leipzig	85	0.00433353	Salamanque
36	0.00832413	Munich	86	0.0042273	Fribourg
37	0.00791791	Waseda	87	0.00418341	Arizona
38	0.0076835	UC London	88	0.00417181	Jena
39	0.00751886	Duke	89	0.00415139	Martin Luther Halle-Wittenberg
40	0.00718132	Sapienza	90	0.00401368	St Andrews
41	0.00711981	ETH Zurich	91	0.00398415	TU Berlin
42	0.0071081	Californie du sud	92	0.00391916	Californie, Chapel Hill
43	0.00693105	École Polytechnique	93	0.00390789	Tartu
44	0.00692597	Pékin	94	0.00388656	TU Munich
45	0.00682986	Al-Azhar	95	0.00385376	Sydney
46	0.00682254	École Normale Supérieure	96	0.00384341	Californie, San Diego
47	0.00680075	Kyoto	97	0.00371085	Trinity College, Dublin
48	0.00666809	Charles	98	0.00368454	Indiana
49	0.00666454	Saint-Pétersbourg SU	99	0.00355122	Notre Dame
50	0.00662585	Utrecht	100	0.00353878	Kiel

[a] Université Hébraïque de Jérusalem, [b] London School of Economics,

TABLE 3.10 : Classements des universités du top 100 WRWU17 classées selon le PageRank associé à la matrice de Google réduite $\tilde{G}_r$. D'après [46].

La Figure 3.20 donne la matrice de Google réduite multi-linguistique et ses 3 composantes. Les structures sont similaires aux matrices obtenues avec une seule édition. Avec l'utilisation de 24 éditions linguistiques, nous avons $W_{\mathrm{rr}} > W_{\mathrm{qrnd}}$ et donc nous avons davantage de liens directs que de liens indirects. L'utilisation de 24 éditions donne une connaissance complète sur les universités, ainsi nous obtenons plus de liens directs. On observe aussi très nettement que la majorité des interactions indirectes obtenues avec $\tilde{G}_{\mathrm{qrnd}}$ ne sont pas dans $\tilde{G}_{\mathrm{rr}}$. Cela montre

l'importance de l'utilisation de plusieurs éditions afin de pouvoir isoler les liens indirects qui sont réellement non-triviaux, leur existence n'étant pas intuitive même avec les connaissances fournies par les 24 éditions.

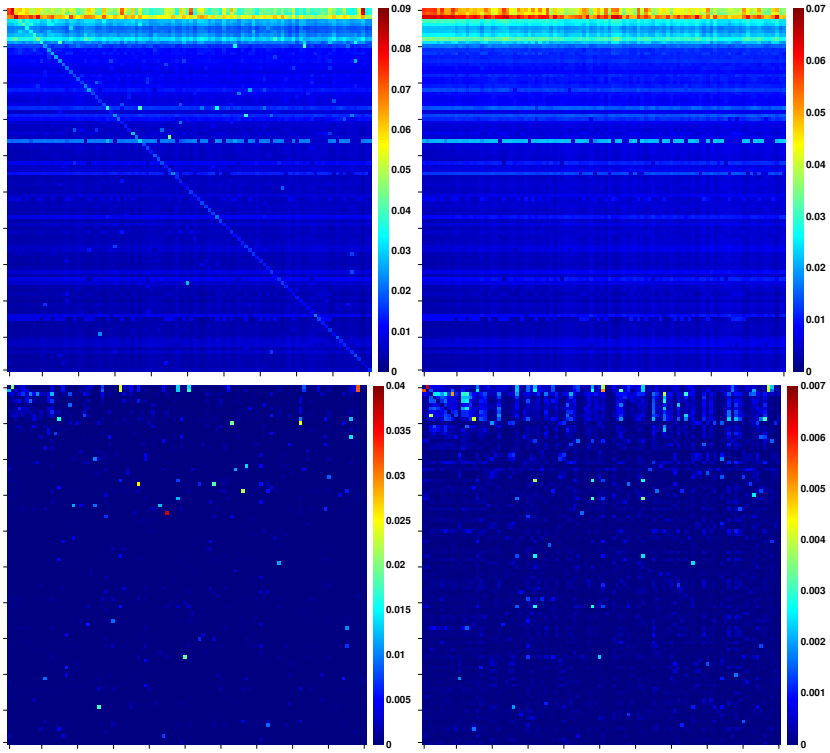

FIGURE 3.20 : Matrice de Google réduite $\tilde{G}_{\mathrm{r}}$ des universités du top 100 WRWU17 (voir Table 3.10) construit à partir de 24 éditions linguistiques. La matrice $\tilde{G}_{\mathrm{r}}$ en haut à gauche, $\tilde{G}_{\mathrm{pr}}$ en haut à droite, $\tilde{G}_{\mathrm{rr}}$ en bas à gauche et $\tilde{G}_{\mathrm{qrnd}}$ en bas à droite. Les poids associés sont $W_{\mathrm{r}} = 1$, $W_{\mathrm{pr}} = 0.957$, $W_{\mathrm{rr}} = 0.019$, $W_{\mathrm{qr}} = 0.024$ et $W_{\mathrm{qrnd}} = 0.015$. D'après [46].

Nous proposons deux réseaux réduits basés sur deux types de catégories : des périodes de fondation et des continents. La Figure 3.21 montre le réseau réduit d'interactions dans le cadre des catégories basées sur l'époque de fondation à travers les 10 derniers siècles. La communauté centrée sur Oxford, leader du groupe $[1000, 1300[$, est composée majoritairement d'universités britanniques et italiennes. Ces universités transmettent leurs influences via des liens dirigés vers la communauté centrée sur Copenhague, leader de la période $[1300, 1600[$. Cette communauté intègre des universités provenant de pays d'Europe du nord, incluant L'Écosse, le Danemark, l'Allemagne, la Suède et le Pays-Bas. Les quelques liens sortant de cette communauté sont dirigés vers la communauté d'Oxford mais aussi vers celle centrée sur Harvard, leader de la période $[1600, 1800[$. Harvard a capturé des universités US. Enfin la communauté centrée sur le MIT, leader de groupe d'université les plus récentes de notre top 100 $[1800, 2000]$, capture également des universités US. Les communautés les moins homogènes sont : la communauté centrée sur Oxford, ayant capturé des universités de toutes les époques et celle centrée sur Copenhague qui capture des universités qui lui sont contemporaines ou qui sont plus jeunes. Majoritairement, les liens partent d'universités anciennes et vont vers des universités récentes, ce qui semble cohérent. Les liens entre communautés sont également intéressants. Ils mettent en lumière d'importantes interactions entre universités récentes et anciennes et révèlent une absence de liens entre les communautés $[1800, 2000]$ et $[1300, 1600[$.

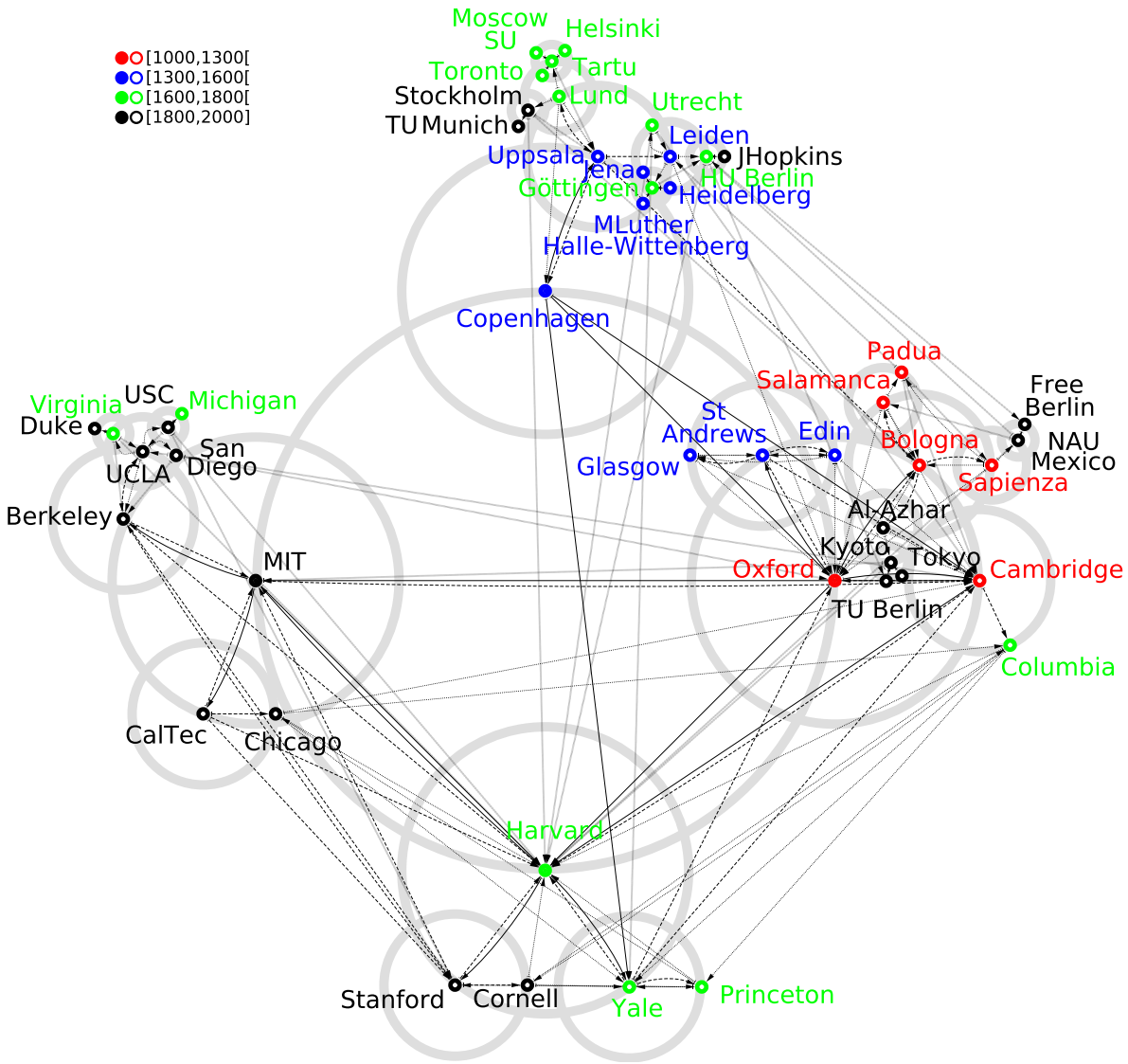

FIGURE 3.21 : Réseau réduits du top 100 WRWU avec catégories relatives à la période de fondation des universités, construit avec la matrice $\tilde{G}_{\text{sum}}$. La couleur des nœuds est caractéristique de leur date de fondation ; Les nœuds pleins sont les leaders de chaque période. Il y a 5 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, lignes en traits pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour les $4^{\text{ème}}$ et $5^{\text{ème}}$ niveaux. D'après [46].

L'analyse du réseau réduit des 100 meilleures universités selon WRWU17, d'un point de vue géographique est présentée à la Figure 3.22. On observe que la communauté des universités centrée sur le leader US (Harvard) est homogène. Les universités d'Europe ont aussi un groupement similaire. On remarque l'influence de l'Europe et de L'Amérique dans le monde. En effet, les universités d'Asie centrées sur Tokyo, capturent des universités provenant de ces continents. On note l'émergence de nouvelles universités asiatiques ayant de forts liens entre elles comme par exemple le triplet formé par Tokyo, Keio et Pékin.

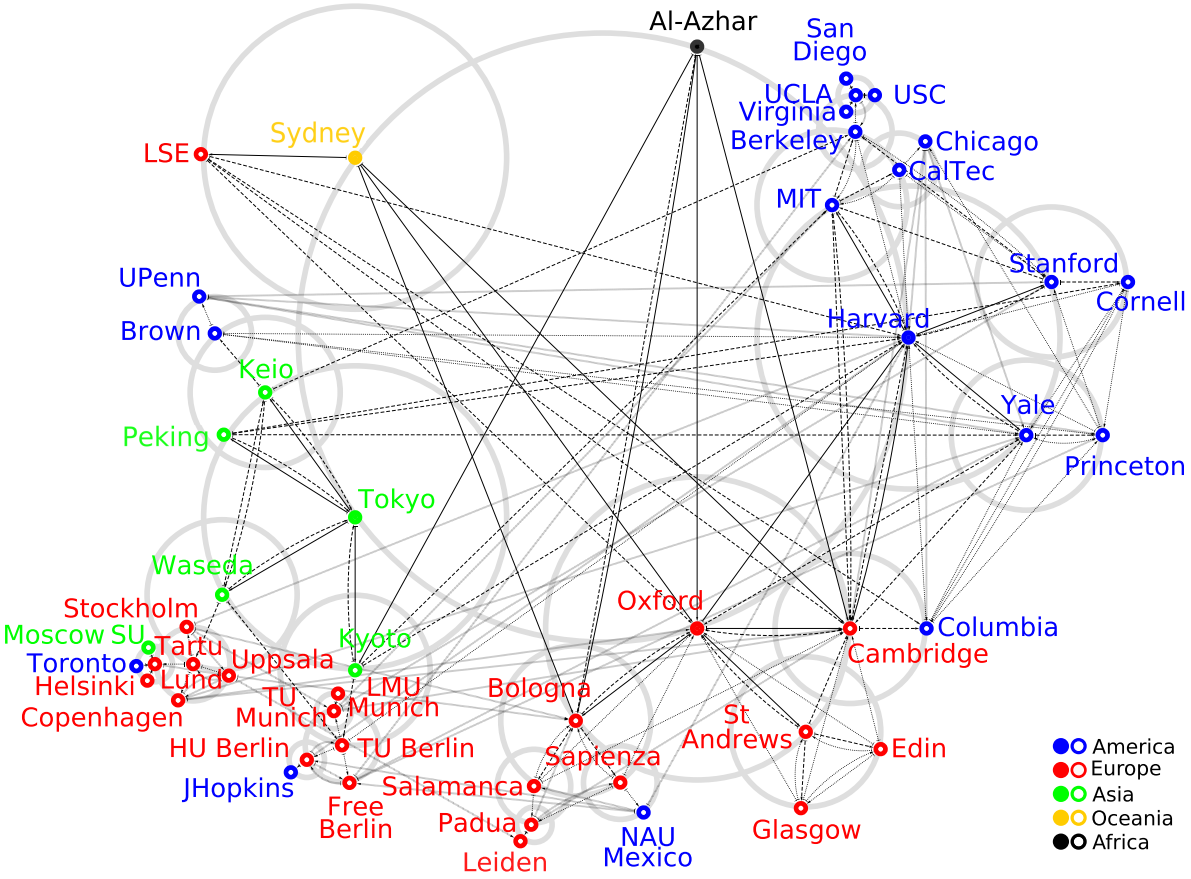

FIGURE 3.22 : Réseau réduit des 100 meilleurs universités mondiales selon WRWU avec des catégories basées sur les continents, construit avec $\tilde{G}_{\text{sum}}$. La couleur des nœuds définit le continent où se situe l'université ; les nœuds pleins sont les leaders. Il y a 5 niveaux d'amitiés représentés par des cercles gris. Les liens appartenant au premier niveau sont en ligne pleine, lignes en traits pour le second niveau, en pointillé pour le $3^{\text{ème}}$ et formées de "\" pour les $4^{\text{ème}}$ et $5^{\text{ème}}$ niveaux. D'après [46].

3.1.6 Conclusion

Cette étude basée sur les données de 24 éditions linguistiques de Wikipédia (mai 2017) et centrée sur l'étude des interactions entre universités et leurs rayonnement à travers le monde montre l'efficacité de l'utilisation de la matrice de Google et de l'algorithme PageRank dans le but de classer les universités, avec le classement Wikipédia Ranking of World Universities (WRWU) et l'efficacité de l'utilisation de la matrice de Google réduite pour récupérer des informations sur des liens directs et indirects entre universités. En effet, nous avons obtenu un classement WRWU17 ayant 60% de similarité avec l'Academic Ranking of World Universities (ARWU17) à l'échelle du top 100. L'utilisation de 24 éditions linguistiques et de la théorie des réseaux complexes nous permet de nous affranchir de règles subjectives comme le nombre d'articles publiés dans Nature ou encore le nombre de prix Nobel. L'analyse du réseau Wikipédia par l'utilisation de la méthode de la matrice de Google réduite permet de mettre en lumière les influences des universités sur le monde mais également de rendre compte des leurs interactions, que ce soit directement ou indirectement, par diffusion dans tout le réseau Wikipédia. Malgré des éditions prises séparément qui donnent des top 20 d'universités différentes et des interactions entre ces universités qui leur sont propres, l'utilisation d'une liste basée sur l'union de 4 tops 20 révèle que les éditions FR, DE, RU et EN donnent des structures similaires en ce qui concerne la distribution des liens indirects entre ces universités. En élaborant une

matrice de Google réduite à l'échelle de 24 éditions linguistiques, nous avons pu établir deux réseaux réduits d'interactions entre les 100 meilleures universités WRWU17. Le premier réseau réduit est basé sur des catégories relatives aux périodes de fondation des universités. Le second réseau réduit est basé sur des catégories relatives à un découpage géographique calqué sur les continents. L'utilisation de plusieurs éditions en même temps est importante afin ne pas avoir de biais culturels. En effet, l'étude [43] montre qu'un classement des personnages historiques selon ENWIKI place en haut du classement les présidents américains tandis que l'utilisation de 24 éditions procure davantage de profondeur et rend compte de personnalités historiques du monde entier. Les résultats obtenus avec la méthode de la matrice de Google réduite sur les interactions et les influences des universités sont en accord avec ce qu'on peut lire dans Wikipédia comme par exemple l'influence de Harvard sur L'Afrique du Sud vis à vis de l'apartheid. Toutefois, nous avons également découvert des influences qui ne sont pas directement visibles sur l'article relatif à l'université étudiée mais dues à des liens indirects.

3.2 Classement des articles de Wikipédia avec biais social

3.2.1 Introduction

Wikipédia referme une colossale quantité avec plus de 300 éditions linguistiques pour environ de 51 millions d'articles.[13] Cette diversité de sujets, présente dans Wikipédia, permet des études dans un grand nombre de domaines. L'algorithme PageRank est un outil puissant permettant de mesurer la centralité des éléments d'un réseau dirigé, son efficacité étant largement prouvée par la popularité du moteur de recherche Google [21]. La matrice de Google dont est issu le PageRank modélise une marche aléatoire dans un réseau [23]. Cependant, la construction de G ne prend pas en compte les comportements réels des utilisateurs web d'aujourd'hui. Avec les réseaux sociaux et des interfaces de plus en plus intuitives, les internautes deviennent conscients du réseau dans lequel ils naviguent. Wikipédia est l'encyclopédie la plus riche et étant libre d'accès, elle est très populaire. Nous proposons donc d'utiliser les comportements utilisateurs relatifs à Wikipédia, tels que les clics et le nombre de vues par article afin de modifier G et d'avoir une marche aléatoire plus réaliste. Les classements d'articles Wikipédia, obtenus avec la méthode du PageRank standard sont robustes à travers le temps. Par exemple, les pays ont un meilleur score PageRank que la plupart des articles et leur position dans le classement global reste stable dans le temps. L'utilisation des informations relatives aux comportements des lecteurs Wikipédia permet d'obtenir des classements capturant les tendances sociales. Nous avons utilisé les données Wikipédia d'octobre 2019 relatives aux 11 éditions linguistiques qui archivent, mensuellement, les clics (clickstream) et les vues (pageview) des différents articles. Nous proposons deux modèles : wikiclick (**wc**), basé uniquement sur le nombre de clics, et wikiclick-pageview (**wcpv**), basé sur le nombre clics et le nombre de vues des articles. Cette section se compose de la manière suivante, premièrement, nous allons faire état de la littérature récente sur l'utilisation du PageRank et des informations utilisateurs appliqués à Wikipédia et d'autres réseaux de connaissances (section 3.2.2). Ensuite, la méthode de construction des matrices de Google associées à ces modèles, ainsi que la méthode d'extractions des données Wikipédia, seront décrites (Section 3.2.3 et section 3.2.4). Enfin nous allons discuter les résultats obtenus avec ENWIKI19, en utilisant deux types de données, XML et SQL (section 3.2.5).

3.2.2 Études récentes

Les statistiques sur les articles Wikipédia, tels que le nombre de vues, le comportement des utilisateurs ou bien encore la composition sémantique des différents articles (longueur de la page, nombre de liens, nombre de figures) sont exploitées dans des domaines scientifiques va-

[13] Informations disponibles à l'adresse https://meta.wikimedia.org/wiki/List_of_Wikipédias.

riés. L'utilisation du nombre de vues permet de dévoiler une estimation statistique des sujets et articles les plus populaires dans Wikipédia, mais également d'afficher l'évolution des tendances. De cette manière, des auteurs se sont intéressés aussi bien à la prévision des mouvements de marchés boursiers [50], qu'aux succès cinématographiques [51, 52], aux services touristiques [53], à l'étude des cryptomonnaies et de la performance des marchés [54, 55]. Ces informations utilisateurs se révèlent également utiles en épidémiologie [56] et dans les classements de personnes célèbres [57]. Aussi, une mesure de la qualité des articles Wikipédia se base sur les compositions sémantiques de ces derniers autant que sur les informations statistiques telles que le nombre de vues [58, 59].

Comme vu précédemment, et également étayé dans d'autres études récentes, la matrice de Google associée au réseau d'articles Wikipédia est aussi étudiée. Ces études mènent à des classements de personnalités historiques [43] ou encore à des classements d'universités [42]. Une étude montre l'évolution de ces classements sur le temps, obtenus par PageRank, CheiRank et 2DRank, d'articles sur les pays, personnages célèbres, physiciens et joueurs d'échec. L'étude [41] montre que ces classements sont plutôt stables dans le temps. L'analyse spectrale de la matrice de Google associée à Wikipédia permet de faire ressortir des communautés d'articles [28]. La matrice de Google réduite, utilisé dans l'étude décrite plus haut et définit à la Section 2.7, permet de quantifier les interactions et influences entre universités et pays [46], l'étude des interactions entre les types de cancer [60] ou bien encore, l'étude des maladies infectieuses [61].

Récemment, une étude intègre à la fois le nombre de vues et les clics utilisateurs relatifs à un répertoire ontologique biomédical (BioPortal) [62]. Pour Wikipédia, des travux récents utilisant les données clickstream ont permis d'étudier l'espace des phases des navigations [63]. Ces études motivent notre choix d'utiliser les informations basées sur les vues et les clics dans le but de classer les articles Wikipédia. En procédant ainsi, nous espérons obtenir des classements permettant de refléter les tendances actuelles.

3.2.3 Les modèles wc et wcpv

La version standard de la matrice de Google associée au réseau Wikipédia, noté ici **nowc**, est basée sur la topologie du réseau d'article Wikipédia et sur un terme de téléportation homogène. Dans ce contexte, le marcheur aléatoire peut suivre les liens entre articles (avec liens sortants virtuels pour les nœuds ballants) avec une probabilité α. Il peut aussi se téléporter équiprobablement vers un article du réseau. Afin de prendre en compte les comportements réels des lecteurs Wikipédia, nous pouvons modifier les poids associés aux liens entre articles ou bien modifier le terme de téléportation. Le modèle **wc** intègre les nombres de clics dans les poids des liens entre articles Wikipédia. Le clickstream est le nombre de clics que reçoit un hyperlien présent dans Wikipédia. Ainsi l'article B, cité dans l'article A, est cliqué par l'utilisateur et chaque clic est archivé. Wikipédia enregistre seulement les articles qui sont cliqués plus de 10 fois. Ainsi si le lien $A \to B$ existe mais qu'il a été cliqué moins de 10 fois, nous définissons le poids de ce lien comme étant 1. Le second modèle, **wcpv**, est similaire à **wc** mais intègre le nombre de vues dans le terme de téléportation de la matrice de Google. Nous faisons l'hypothèse que le nombre de vues donne une information sur la tendance actuelle tandis que les clics donnent une informations sur l'intérêt culturel. En effet, si un utilisateur lit un article A et clique que le lien menant vers l'article B depuis l'article A, c'est que B complète la compréhension de A ou bien que B attire l'attention du lecteur.

Nous définissons A_{wc} comme étant une matrice d'adjacence pondérée, dont l'élément $A_{\text{wc}ij} = W_{ij}$, avec W_{ij} le nombre de clics que reçoit la citation de l'article i depuis l'article j. Comme expliqué plus haut, si $W_{ij} = 0$ et que $A_{ij} = 1$, où A et la matrice d'adjacence associée au modèle **nowc**, alors on a $A_{\text{wc}ij} = 1$.

Soit la matrice stochastique S_{wc} telle que

$$S_{\mathrm{wc}ij} = \begin{cases} \frac{A_{\mathrm{wc}ij}}{\sum_{i'} A_{\mathrm{wc}i'j}} & \text{si } \sum_{i'} A_{\mathrm{wc}i'j} \neq 0 \\ 1/N & \text{sinon.} \end{cases} \tag{3.9}$$

La matrice de Google G associée

$$G_{ij} = \begin{cases} \alpha S_{\mathrm{wc}ij} + (1-\alpha)/N & \text{pour } \mathbf{wc} \\ \alpha S_{\mathrm{wc}ij} + (1-\alpha)\tilde{v}_i & \text{pour } \mathbf{wcpv} \end{cases} \tag{3.10}$$

où le vecteur préférentiel $\tilde{\mathbf{v}}$, encode le nombre de vues tel que, $\tilde{v}_i = v_i/v_{\mathrm{tot}}$, avec v_i le nombre de vues associé à l'article i et v_{tot}, le nombre total de vues. La valeur de α utilisée pour **wc**, **wcpv** et **nowc**, est $\alpha = 0.85$ [22].

3.2.4 Les données XML et SQL relatives à Wikipédia 2019

Afin de pouvoir utiliser les données XML et SQL relatives à Wikipédia, nous avons dû utiliser une méthode d'extraction du réseau d'articles différente de celle utilisée dans l'étude de la section 3.1.

- **XML** - Langage wikicode des articles Wikipédia.
- **SQL** - Banque de données représentant les liens source-destination de Wikipédia.

Dans le langage wikicode, un article dont le titre est "Titre" sera citée dans un autre article via l'utilisation de double crochets : "[[Titre]]". Au sein de Wikipédia, il existe plusieurs catégories d'articles ou *namespace* (ns). Les articles qui nous intéressent sont les articles *ns0*. Aussi un article peut avoir des *redirects*, ce sont des articles doublons dont les titres peuvent être vus comme des synonymes ou des acronymes et ils permettent de se rediriger vers l'article principal. Par exemple, l'article "USA" est un *redirect* qui redirige vers l'article "États-Unis". Dans la construction du réseau d'articles Wikipédia, il est important de pouvoir retrouver l'article principal, correspondant à un *redirect*, lorsque un *redirect* est cité.

Le SQL est simplement une banque de données donnant les connexions entre les pages Wikipédia. Elle donne plus d'information que l'utilisation du XML. En effet, nous pouvons avoir, par exemple, l'information relative à la citation d'un article au sein d'une légende d'image ou bien encore dans des *infoboxes*.

Seulement 11 éditions linguistiques archivent à la fois des données relatives au clickstream, et les données relatives au nombre de vues. La Table 3.11 répertorie ces éditions ainsi que leur nombre brut de liens, avant traitement des données, et leur nombre final de liens, après traitement (fusions des *redirects* et suppressions des articles non *ns0*). Au total nous avons

Langage	XML		SQL	
	Brut	Final	Brut	Final
de	86 242 247	63 618 326	111 288 696	108 762 081
en	228 373 266	165 832 345	500 144 739	479 163 241
es	54 878 393	39 369 961	53 948 827	50 625 623
fa	15 298 097	7 427 045	74 249 078	71 867 560
fr	80 719 270	61 576 083	156 691 399	153 108 004
it	52 731 642	40 857 564	117 826 190	115 641 441
ja	63 112 674	50 122 887	92 626 350	90 901 975
pl	36 838 878	27 240 200	76 958 914	76 318 086
pt	31 311 443	22 167 152	61 269 416	58 843 986
ru	52 646 408	37 922 206	99 995 034	95 706 281
zh	30 253 747	18 718 463	86 343 098	83 272 015

TABLE 3.11 : Liste des éditions archivant les données utilisateurs. Les dumps sont datés d'octobre 2019. D'après [64].

extrait plus de 500 millions de liens pour XML et 1 milliard pour SQL.

3.2.5 Application à ENWIKI19

Nous allons comparer les résultats obtenus avec le modèle **wcpv** et **wc** avec différents classements de référence (benchmark). Ces classements benchmarks sont ceux issus du modèle **nowc** mais aussi des classements statistiques basés sur le clickstream **cR** et le pageview **vR**.

Données XML

Le top 10 PageRank des articles Wikipédia avec le modèle **wcpv** est présenté dans la Table 3.12. Nous avons regardé les rangs respectifs de ces articles K_{wc}, K_{nowc}, vR et cR. Nous pouvons noter que comme pour un PageRank standard de Wikipédia, il y a présence d'articles dédiés aux pays dans les premières places tels que "United States", "United Kingdom", "Germany", "India" et "France". Ils sont classés respectivement 1^{er}, $4^{\mathrm{ème}}$, $7^{\mathrm{ème}}$, $8^{\mathrm{ème}}$ et $9^{\mathrm{ème}}$. On peut voir que les différences majeures entre **wc** et **nowc** sont des changements d'ordres. On voit par exemple que 7 éléments sur les 10 présentés sont inclus dans le top 10 PageRank **nowc**. Seulement 5 sont dans le top 10 PageRank **wc**. On note moins de correspondances avec les classements **cR** et **vR**. En effet, aucun élément du top 10 PageRank **wcpv** n'est dans le top 10 PageRank **cR** et seulement 2 articles du top 10 PageRank **wcpv** sont présents dans celui associé à **vR**. Il est intéressant de noter la présence d'éléments non-triviaux avec l'utilisation du modèle **wcpv**. En effet, 3 éléments ne sont pas présents dans les tops 10 PageRank **wc** et **nowc** et sont de plus très mal classés. Au mieux, le $2^{\mathrm{ème}}$ élément du classement PageRank **wcpv** et $3542^{\mathrm{ème}}$ dans le classement PageRank **nowc**, dans le pire des cas le $3^{\mathrm{ème}}$ élément descend de 5 millions de places avec les modèles **wc** et **nowc**. Parmi ces trois éléments, deux viennent de l'utilisation des données pageview, en effet les $2^{\mathrm{ème}}$ et $3^{\mathrm{ème}}$ articles du classement PageRank **wcpv** sont dans le top 2 **vR**. En revanche, l'article "Queen of The South (TV series)" est $10^{\mathrm{ème}}$ du classement PageRank **wcpv** et est bien moins classé dans le classement PageRank **nowc** ($744237^{\mathrm{ème}}$) et le classement **vR** ($5871^{\mathrm{ème}}$).

Name	K_{wcpv}	K_{wc}	K_{nowc}	K_{cR}	K_{vR}
United States	1	1	1	15	24
Wikipédia	2	11665	3542	25013	1
List of Queen of the South episodes	3	5170889	5128933	4455336	2
United Kingdom	4	9	5	81	63
New York City	5	23	10	150	139
World War II	6	12	3	181	78
Germany	7	7	7	727	118
India	8	10	9	138	68
France	9	5	2	1432	197
Queen of the South (TV series)	10	166342	744237	28297	5871

TABLE 3.12 : Le top 10 PageRank associé au modèle **wcpv** pour l'édition anglaise de Wikipédia 2019. D'après [64].

On note **nowc***, **wc*** et **wcpv***, les modèles **nowc**, **wc** et **wcpv**, lorsque l'algorithme utilisé est le CheiRank. La Table 3.13, donne le top 10 CheiRank associé au modèle **wcpv***. Dans le cas du modèle **nowc***, les articles relatif à des listes d'articles sont préférentiellement en tête du classement CheiRank. Dans le cas **wcpv***, on observe deux éléments comparables dont un seul apparaît dans le top 10 CheiRank **nowc***. Le $10^{\mathrm{ème}}$ élément du CheiRank **nowc*** est $2^{\mathrm{ème}}$ dans le CheiRank **wcpv*** et est une liste dont le titre est "List of deaths by year". Le second élément du classement CheiRank **nowc*** qui est présent dans le top 10 CheiRank **wcpv*** est classé $382^{\mathrm{ème}}$ alors qu'il est $4^{\mathrm{ème}}$ avec le modèle **wcpv***. Il s'agit aussi d'une liste, "2019 in film". À la différence du modèle **nowc***, l'utilisation des comportements d'utilisateurs donne un top 10 contenant des articles relatifs à l'année 2019. De façon similaire au classement

PageRank, le classement CheiRank obtenu avec le modèle **wcpv*** donne de nouveaux éléments. Ce top 10 est encore plus éloigné de celui associé à **nowc*** que dans le cas de l'algorithme PageRank. Nous obtenons des articles qui ne sont pas des listes, comme par exemple "It (2017 film)", "Joker (2019 film)", "Mindhunter (TV series)" et "2019 FIBA Basketball World Cup". Néanmoins, ces articles peuvent être vus comme une liste d'intérêt ou bien des points d'entrées de Wikipédia. Autrement dit un utilisateur, sujet à la tendance culturelle actuelle, va préférentiellement débuter son voyage bibliographique, sur Wikipédia, via des articles traitant de faits d'actualité, de films récents, etc. Par exemple, le film Joker, récent au moment de la récolte des données, attire les lecteurs et peut être vu comme une liste d'articles intéressants. Les utilisateurs peuvent se documenter sur la vie des acteurs, les lieux du tournage et autres informations contenues dans les articles Wikipédia accessibles depuis l'article "Joker (2019 film)".

Name	K^*_{wcpv}	K^*_{wc}	K^*_{nowc}	K_{cR}	K_{vR}
Deaths in 2019	1	2	909	406	4
Lists of deaths by year	2	1	10	55939	7874
It Chapter Two	3	10	334575	2	12
2019 in film	4	9	382	365	34
List of Bollywood films of 2019	5	12	19249	640	54
Wikipédia	6	421	11031	25013	1
It (2017 film)	7	19	106231	18	37
Joker (2019 film)	8	20	95145	33	10
Mindhunter (TV series)	9	17	310462	147	35
2019 FIBA Basketball World Cup	10	11	20498	44	13

TABLE 3.13 : Le top 10 CheiRank associé au modèle **wcpv*** pour l'édition anglaise de Wikipédia 2019. D'après [64].

Nous proposons deux mesures permettant de quantifier la ressemblance entre deux classements d'articles Wikipédia. La similarité d'ordre η_{O} (3.11) et la similarité de présence η_{N} (notée η dans (3.3)). La similarité d'ordre $\eta_{\mathrm{O}}(j)$ entre deux tops j est définit comme suit

$$\eta_{\mathrm{O}}(j) = \frac{\sum_{i=1}^{j} f(i)}{j} \tag{3.11}$$

où, $f(i) = 1$ si les $i^{\text{èmes}}$ éléments des listes A et B sont identiques. La Figure 3.23, partie de gauche, montre l'évolution de ces similarités en fonction du top j (avec $j \leq 100$) pour différents couples de classements PageRank et CheiRank. Les similarités entre classements PageRank et CheiRank, avec les modèles **wcpv** et **wcpv***, **wc** et **wc***, et les classements **vR** et **cR**, sont montrés sur la figure de droite. Comme on peut le voir $\eta_{\mathrm{O}} \approx 0$ pour presque tous les couples de classements PageRank (ou CheiRank)/classements statistiques (**vR** ou **cR**). Dans la partie gauche de la figure, on observe que les courbes relatives aux classements CheiRank sont presque nulles. Aussi, sur la figure de gauche, les courbes relatives aux classements PageRank commencent à décroître à partir de $j = 20$ et les valeurs à $j = 100$ montrent que **wcpv** donne des résultats éloignés du benchmark **nowc**. En effet, nous avons $\eta_{\mathrm{O}}(100) = 0.05$ (0.02) pour **nowc** vs. **wc** (**nowc** vs. **wcpv**). $\eta_{\mathrm{N}} >> \eta_{\mathrm{O}}$ pour tous les couples étudiés. Les modèles **wc** et **wcpv** permettent d'obtenir des tops 100 qui sont majoritairement des réordonnements. Pour la partie haute des classements ($j \leq 20$), nous obtenons les valeurs maximales $\eta_{\mathrm{N}} = 0.75$ (**nowc** vs. **wc**), 0.7 (**nowc** vs. **wcpv**). Les classements CheiRank sont plus éloignés que les classements PageRank, en effet on obtient $\eta_{\mathrm{N}} = 0.35$ (0.15) pour **nowc*** vs. **wc*** (**nowc*** vs. **wcpv***). Les seules mesures de similarité de même ordre de grandeur, sont celles relatives aux couples (**wc**,**wcpv**) et (**wc***,**wcpv***). L'intégration du comportement utilisateur dans notre

modèle de matrice de Google tend à produire des classements CheiRank bien différents et le modèle **wcpv** semble être celui qui donne les classements les plus éloignés de **nowc**. Pour $j = 100$ nous avons des similarités de présence, η_N, associées aux classements PageRank et aux classements CheiRank **wcpv** (**wcpv***) vs. **nowc** (**nowc***), valant respectivement 53% et 6%. Dans la partie droite de la figure, les similarités, η_N, sont plus élevées quand on compare **cR** et **vR** avec les classements CheiRank plutôt qu'avec avec les classements PageRank. Le modèle le plus proche de **vR** est **wcpv** (resp. **wcpv***), avec $\eta_O(100) = 0.24$ (resp. 0.5). Ces valeurs de similarité montrent que **wcpv** est un bon candidat, en effet, il donne des classements éloignés du benchmark **nowc** mais aussi en accord avec les classements issus des statistiques, **vR** et **cR**.

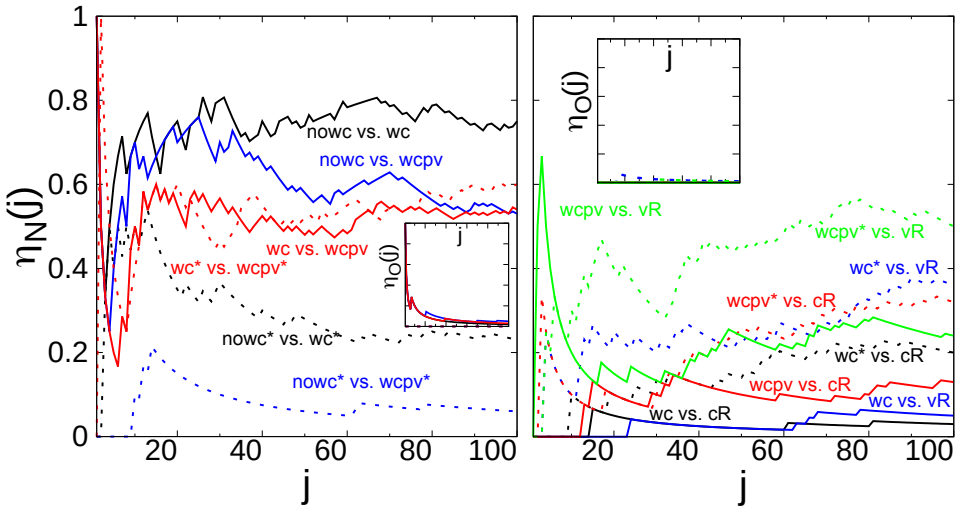

FIGURE 3.23 : Similarité η_N en fonction du rang j entre classements PageRank/CheiRank associés aux modèles **wc**, **wcpv** and **nowc** (à gauche) et associées à **wc**, **wcpv**, **vR** et **cR** (à droite). Les encarts représentent la similarité η_O. Les courbes pleines sont relatives à des couples de classements PageRank, les courbes en pointillé à des couples de classements CheiRanks. D'après [64].

Afin de voir, à l'échelle globale, les conséquences de l'intégration des données utilisateurs dans la matrice de Google, nous avons regardé la distribution des articles Wikipédia dans le plan (K, K^*), où K et K^* sont les rangs des classements PageRank et CheiRank. Les densités d'articles $W(K, K^*) = \frac{dn}{dKdK^*}$ contenus dans ce plan sont présentées à la Figure 3.24 pour les différents modèles. Pour le modèle **nowc**, les articles ont une tendance à avoir un meilleur rang CheiRank que PageRank, en effet, le nuage de cellules avec densité non nulle s'étire vers la zone $K^* < K$ et donne une matrice de densités non symétrique. On peut voir que, parmi les articles en dehors de ce cluster d'articles, seulement 5 sont présents dans le top 100 **vR** et/ou **cR**. Parmi ces articles, 4 appartiennent au top **vR** ("World War II", "United Kingdom", "China" et "India") et 2 au top **cR** ("United Kingdom" et "James VI and I"). Les modèles **wc** et **wcpv** placent les articles issus des tops 100 **vR** et **cR** en dehors du cluster. L'asymétrie de la matrice de densité est beaucoup moins présente dans le cas **wcpv** et le cluster se concentre sur la diagonale. Cette propriété pourrait venir de l'utilisation d'un vecteur préférentiel, qui est non homogène et permet au marcheur aléatoire de se téléporter, vers un article de préférence, de la même manière au sein du réseau dont les liens sont inversés et dans le cas du réseau de base.

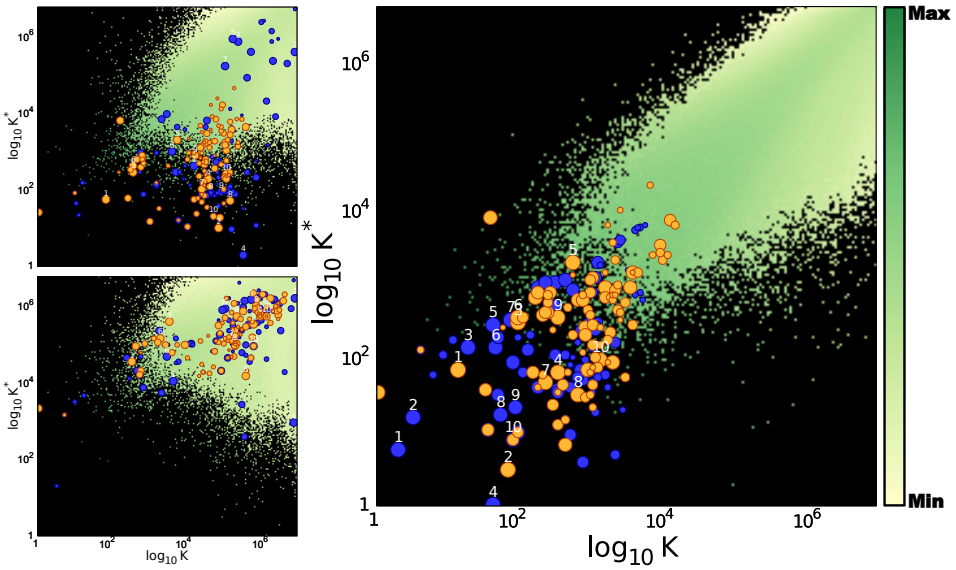

FIGURE 3.24 : Distribution de densité des articles Wikipédia $W(K, K^*) = dn/dKdK^*$ pour **nowc** (en bas à gauche), **wc** (en haut à gauche) et **wcpv** (à droite). Le plan (K, K^*) est subdivisé en 200×200 cellule de taille logarithmique. Pour chaque cellule d'aire $dKdK^*$, nous calculons la densité d'articles. L'échelle de couleur est une échelle logarithmique. Les cercles bleus et oranges représentent respectivement les 100 meilleurs articles par **vR** et **cR**. Les rayons des cercles décroit avec le rang. Les labels de 1 à 10 sont placés sur les articles présent dans les tops 10 **vR** et **cR**. D'après [64].

Données SQL

L'utilisation des données SQL requiert une attention particulière. En effet, il peut arriver que des articles *ns0* appartiennent à une autre catégorie dans certaines éditions. Par exemple, la page d'accueil de Wikipédia pour l'édition anglaise "Main page", est un article *ns0* mais son équivalent dans l'édition française est "Wikipédia :Accueil_ principal" n'est pas *ns0*. Nous avons décider de comparer les résultats relatifs aux réseaux construits avec et sans cette page d'accueil. Les pages d'accueil Wikipédia proposent, notamment, une liste d'articles pouvant intéresser les lecteurs.

Comme le montre la Table 3.14 et la Table 3.15, en enlevant l'article Main Page, des éléments très intéressants peuvent disparaître du top 10 **wcpv**. En effet, les articles "Brexit" et "Impeachment inquiry against Donald Trump" ne sont plus présents dans le top 10 **wcpv** lorsqu'on ne prend pas en compte la page d'accueil. Le top 10 **wcpv** ne contient plus l'article "QR Code" quand on ne garde pas la page d'accueil.

	Name	K_{wcpv}	K_{wc}	K_{nowc}	K_{cR}	K_{vR}
Avec "Main Page"	Main Page	1	2451	127099	7096	1
	Deaths in 2019	2	5732	2930554	2	5
	Brexit	3	6601	7066	245	85
	Impeachment inquiry against Donald Trump	4	19841	340242	180	569
	International Standard Book Number	5	2	1	8313	215
	Wikipédia	6	372	533	3392	2
	2019 Southeast Asian haze	7	110186	1178409	1651	2393
	List of Queen of the South episodes	8	116413	5051712	1	3
	2I/Borisov	9	82476	263658	1977	1103
	United States	10	4	7	5	25
Sans "Main Page"	International Standard Book Number	1	2	1	8304	214
	List of Queen of the South episodes	2	116418	5051707	1	2
	United States	3	4	7	4	24
	Wikipédia	4	371	533	25382	1
	Geographic coordinate system	5	1	2	40532	1212
	WorldCat	6	3	4	105828	6682
	Virtual International Authority File	7	5	3	165154	6990
	Library of Congress Control Number	8	8	5	132532	5746
	Digital object identifier	9	14	6	78831	683
	International Standard Name Identifier	10	9	8	178206	8145

TABLE 3.14 : Le top 10 des articles selon l'algorithme PageRank relatif au modèle **wcpv** et à l'utilisation du dump SQL de Wikipédia anglais 2019. En haut, le cas avec page d'accueil et en bas, sans cet article.

	Name	K^*_{wcpv}	K^*_{wc}	K^*_{nowc}	K_{cR}	K_{vR}
Avec "Main Page"	Main Page	1	1	329787	7096	1
	Wikipédia	2	2	25621	3392	2
	English Wikipédia	3	12	233157	208196	577
	Main	4	552	167855	4615950	65903
	QR code	5	891	161385	105728	1649
	It Chapter Two	6	6	605410	4	13
	Wiki	7	861	182305	47081	53
	2019 in film	8	4	12751	122	35
	List of Wikipédias	9	1453	88042	29205	38175
	Deaths in 2019	10	3	860	2	5
Sans "Main Page"	Deaths in 2019	1	1	860	263	4
	It Chapter Two	2	6	605408	3	12
	Lists of deaths by year	3	2	142	39666	7874
	2019 in film	4	4	12750	121	34
	Joker (2019 film)	5	10	152567	33	10
	It (2017 film)	6	12	167372	18	37
	List of Bollywood films of 2019	7	7	5337	652	54
	Wikipédia	8	327	25737	25382	1
	2019	9	3	2891	14578	553
	September 11 attacks	10	9	24288	32	15

TABLE 3.15 : Le top 10 des articles selon l'algorithme CheiRank relatif au modèle **wcpv*** et à l'utilisation du dump SQL de Wikipédia anglais 2019. En haut, le cas avec page d'accueil et en bas, sans cet article.

3.3 Conclusion

Cette étude présente deux nouveaux modèles basés sur l'utilisation des données utilisateurs. Les modèles **wcpv** (wikiclick-pageview) et **wc** (wikiclick), donnent des classements PageRank et CheiRank qui reflètent les tendances actuelles, tandis que la matrice de Google standard **nowc** donne des classements d'articles Wikipédia stables avec le temps. L'utilisation de différentes banques de données, XML et SQL, requiert une méthodologie de travail différente. D'un côté les données SQL, beaucoup plus complètes que les données XML, sont difficilement filtrables, tandis que les données XML permettent de garder les citations d'articles présents uniquement dans le corps principal d'un article. Le couplage entre les liens d'articles, source-destination, caractéristiques de vérité ou bien de lien logique entre des articles, et les données utilisateurs permettant la construction d'un modèle de marche aléatoire plus réaliste, permet d'obtenir des classements d'articles Wikipédia caractéristiques du contexte culturelle contemporain. Bien que nous montrons ici une utilisation banale de ce modèle avec des comparaisons entre les classements issus de nos modèles et des classements benchmark tels que **nowc**, **vR** (classement par vues) et **cR** (classement par clics), nous pensons que ces modèles peuvent donner lieu à des applications telles que, l'étude de l'évolution dans le temps des tendances. Ces modèles sont facilement implémentables pour l'analyse d'autres réseaux de connaissances, basés sur le concept de wiki-page. Une application dans le domaine commercial est aussi envisageable, une entreprise pourrait, par exemple, mesurer l'évolution des demandes clients et mettre en valeur les produits tendances sur le marché.

Chapitre 4

Réseaux économiques

4.1 Impact du pétrole et du gaz étrangers sur l'Europe des 27

4.1.1 Introduction

Les données économiques provenant de la base de données *United Nations Commodity Trade Statistics Database*[1] (Comtrade), ainsi que, le rapport statistique de l'OMC pour l'année 2018,[2] montrent l'importance capitale du commerce international pour la santé économique et le développement des pays [65]. La banque de données Comtrade est une collection des transactions commerciales mettant en jeu plus de 294 pays pour plus de 10 000 produits. La théorie des réseaux complexes et les récents outils mathématiques utilisés pour l'analyse de leurs structures [66] sont applicables à l'étude du commerce international. Par exemple dans [67, 68], l'utilisation de la matrice de Google, ainsi que des algorithmes PageRank et CheiRank [22, 23], permettent une analyse du réseau du commerce international sur plusieurs périodes. Ces méthodes ont été utilisées dans l'étude d'autres réseaux dirigés [69]. La construction de la matrice de Google associée au commerce international permet une étude équitable en mettant tous les pays sur un même pied d'égalité. En effet, la matrice de Google étant un opérateur de Perron, chaque colonne est normalisée. Ainsi, les pays riches et les pays pauvres sont pris en compte de la même manière, leurs capacités totales d'importation et d'exportation respectives étant normalisées. L'utilisation des vecteurs PageRank et CheiRank, relatifs au réseau du commerce international, donne une mesure de l'importance économique pour l'importation et pour l'exportation. La complexité d'un tel réseau requiert l'analyse des liens directs et indirects entre pays. La méthode de la matrice de Google réduite, introduite dans [30], appliquée au réseau d'articles Wikipédia [31, 46] et aux interactions entres protéines [45], est efficace. Cette méthode (voir section 2.7), permet d'extraire le sous-réseau constitué des nœuds d'intérêts avec leurs liens directs ainsi que d'extraire les liens indirects entre ces nœuds. Dans l'étude présentée ici, nous avons utilisé cette dernière méthode pour mesurer l'impact économique du pétrole et du gaz issus d'exportateurs n'appartenant pas à l'Union européenne sur les pays membres de l'Union européenne des 27 (UE27), membres de l'Union européenne avant l'intégration de la Croatie. Nous nous sommes concentrés sur les sous-réseaux de transactions commerciales entre les pays de l'UE27 et des grands exportateurs hors UE tels que, la Russie (RU), les États-Unis (US), l'Arabie Saoudite (SA) et la Norvège (NO). Nous avons utilisé ces sous-réseaux afin de mesurer l'impact économique de l'UE27 face à un changement du prix du pétrole et du gaz en provenance de ces exportateurs. Dans cette étude, nous analysons le commerce international à l'échelle de 277 pays, pour 61 produits et pour les années 2004, 2008, 2012 et 2016. L'étude est présentée comme suit. En premier, nous allons voir la méthode de construction de la matrice

[1] Ces données sont accessibles depuis le site http://comtrade.un.org/db/.

[2] L'article est en ligne et est téléchargeable gratuitement https://www.wto.org/english/res_e/statis_e/wts2018_e/wts2018_e.pdf.

de Google G, associée au commerce international (section 4.1.2), ensuite, nous présenterons les réseaux réduits des transactions commerciales pour le pétrole et le gaz (section 4.1.3). Finalement, nous présenterons les résultats relatifs à la sensibilité économique de l'UE27 face au pétrole russe, saoudien et américain, ainsi que la sensibilité face au gaz russe et norvégien (section 4.1.4).

4.1.2 Matrice de Google et commerce international multi-produits

Le réseau du commerce international multi-produits est composé de $N_{\rm c} = 227$ pays (voir Table A.2 située en annexes) échangeant jusqu'à $N_{\rm p} = 61$ produits (voir Table A.1 située en annexes). Les produits pris en compte sont issus de la classification du commerce international standard (SITC), première révision. En suivant la méthode utilisée dans [68], pour chaque archive relative à une année de transactions commerciales, nous avons $N_{\rm p}$ matrices monétaires M^p. La valeur $M^p_{c,c'}$ est la masse monétaire, mesurée en dollar américain, de l'exportation du produit p à partir du pays c' vers le pays c. On définit les volumes d'importation (V_c^p) et d'exportation (V_c^{p*}) du produit p par le pays c comme

$$V_c^p = \sum_{c'} M^p_{c,c'} \quad \text{et} \quad V_c^{p*} = \sum_{c'} M^p_{c',c}. \tag{4.1}$$

On définit l'ImportRank ($\hat{\mathbf{P}}$) et l'ExportRank ($\hat{\mathbf{P}}^*$) tels que

$$\hat{P}_{cp} = \frac{V_c^p}{V} \quad \text{et} \quad \hat{P}^*_{cp} = \frac{V_c^{p*}}{V} \tag{4.2}$$

ici, la quantité $V = \sum_{cp} V_c^p = \sum_{cp} V_c^{p*}$ est le volume total de marchandise échangée sur une année pour le commerce international et $\hat{P}_{cp}$ ($\hat{P}^*_{cp}$) est la quantité relative de produit importé (exporté) par le pays c.

La matrice de Google G associée au réseau du commerce international, contenant $N = N_{\rm c}N_{\rm p} = 13847$ nœuds, correspondants à des couples pays-produit (cp), est construite selon la méthodologie décrite dans [68]. Soit les deux matrices G et G^*, associées respectivement au réseau du commerce international direct et indirect (avec inversion de la direction des liens, i.e. le lien $j \to i$ devient $i \to j$), et dont les éléments sont

$$\begin{array}{rcl} G_{cp,c'p'} & = & \alpha S_{cp,c'p'} + (1-\alpha)\, v_{cp}, \\ G^*_{cp,c'p'} & = & \alpha S^*_{cp,c'p'} + (1-\alpha)\, v^*_{cp} \end{array} \tag{4.3}$$

Les vecteurs $\mathbf{v}$ et $\mathbf{v}^*$ sont appelés vecteurs préférentiels et sont normalisés tels que $\sum_i v_i = \sum_i v_i^* = 1$. Tout comme pour l'étude [68], nous avons choisi $\alpha = 0.5$. Dans les réseaux du commerce international, le lien $(cp) \to (c'p)$ traduit le flux de produit p exporté depuis le pays c vers le pays c'. Ainsi, l'élément P_{cp} du vecteur PageRank $\mathbf{P}$ mesure la capacité du pays c à importer tandis que l'élément P^*_{cp} du vecteur CheiRank $\mathbf{P}^*$ mesure la capacité à exporter du pays c. On construit les matrices stochastiques S et S^* à partir des $N_{\rm p}$ matrices monétaires $M^{\rm p}$. On a

$$\begin{array}{l} S_{cp,c'p'} = \left\{ \begin{array}{cl} M^p_{c,c'}\delta_{pp'}/V^{p*}_{c'} & \text{si } V^{p'*}_{c'} \neq 0 \\ 1/N & \text{si } V^{p'*}_{c'} = 0 \end{array} \right. \\ S^*_{cp,c'p'} = \left\{ \begin{array}{cl} M^p_{c',c}\delta_{pp'}/V^{p}_{c'} & \text{si } V^{p'}_{c'} \neq 0 \\ 1/N & \text{si } V^{p'}_{c'} = 0. \end{array} \right. \end{array} \tag{4.4}$$

Les matrices stochastiques sont presque des matrices blocs diagonales. Pour une colonne relative à un nœud ballant, c'est-à-dire un pays ayant un volume d'export en produit p nul (V_c^{p*}), les éléments de la colonne sont égaux à $1/N$.

Les vecteurs préférentiels, utilisés dans (4.3) pour la première itération des matrices de Google G et G^* sont

$$v_{cp} = \frac{V_c^p}{N_c \sum_{p'} V_c^{p'}} \quad \text{et} \quad v_{cp}^* = \frac{V_c^{p*}}{N_c \sum_{p'} V_c^{p'*}}. \tag{4.5}$$

En utilisant ces vecteurs préférentiels, la première itération de la matrice de Google permet de placer sur un pied d'égalité les pays.[3] On construit la matrice de Google finale en utilisant de nouveaux vecteurs préférentiels dans (4.3). Ces nouveaux vecteurs préférentiels sont eux-même construits à partir des vecteurs PageRank et CheiRank obtenus à la première itération, soit $\tilde{\mathbf{v}}$ et $\tilde{\mathbf{v}}^*$ tels que

$$\tilde{v}_{cp} = \frac{P_p}{N_c} \quad \text{et} \quad \tilde{v}_{cp}^* = \frac{P_p^*}{N_c} \tag{4.6}$$

où $P_p = \sum_c P_{cp}$ et $P_p^* = \sum_c P_{cp}^*$ sont les projections des vecteurs PageRank et CheiRank, obtenus à la première itération, sur l'espace des pays. De cette manière, le marcheur aléatoire peut se téléporter préférentiellement sur un couple (cp), avec p un produit central du réseau de départ.

4.1.3 Réseaux réduits

Nous allons analyser à l'échelle des 27 pays de l'Union européenne et des 10 plus gros exportateurs hors UE, déterminés selon le volume d'export en pétrole (code $p = 33$) et en gaz (code $p = 34$), le réseau du commerce international. Les classements de ces 37 pays selon les classements PageRank, CheiRank, ImportRank et ExportRank, pour le pétrole et le gaz pour l'année 2016, se trouvent dans la Table 4.1 et Table 4.2.

Dans le cas du pétrole (Table 4.1), la Russie (RU) occupe la première position en CheiRank et en ExportRank. Pour l'importation, les États-Unis (US) sont classés premiers par PageRank et ImportRank. On observe des différences entre les classements standards (ImportRank et ExportRank) et les classements issus du PageRank et du CheiRank. Ces différences viennent du fait que le volume d'export et le volume d'import ne sont pas les seuls paramètres pris en compte pour le CheiRank et le PageRank. Par exemple, l'Arabie Saoudite (SA), classée $2^{\text{ème}}$, pour son volume d'exportation en pétrole, est classée $6^{\text{ème}}$ par la méthode du CheiRank. L'Arabie Saoudite oriente principalement son commerce de pétrole vers les États-Unis, cette faible diversité le rend donc moins important dans le réseau. En revanche, la capacité de Singapour (SG) à importer du pétrole au sein du réseau est très importante, SG passe alors de $4^{\text{ème}}$ (ImportRank) à $2^{\text{ème}}$ (PageRank). Singapour est donc un pays central pour l'importation de pétrole au sein du commerce international. Les Pays-Bas (NL) sont le premier pays de l'UE27 dans tous les classements relatifs au marché du pétrole, on y voit alors son rôle important dans l'import et l'export de pétrole, autant sur le marché européen que sur le marché international. Son rôle capital est notamment dû à ses connections maritimes avec le reste du monde.

Concernant le commerce du gaz, la Table 4.2 montre que les Pays-Bas sont encore le premier pays de l'UE27, mais uniquement pour les classements PageRank et CheiRank. Les classements statistiques standards placent la France (FR) comme pays importateur de gaz numéro 1 et la Belgique (BE) comme $8^{\text{ème}}$ meilleur exportateur de gaz, et aussi, le premier pays de l'UE27 dans ce classement. La France, l'Italie (IT) et le Royaume-Uni (GB) occupent les trois premières places en ImportRank et ils perdent tous deux places dans le classement PageRank, les deux premières du classement PageRank sont prises par les Pays-Bas et la Belgique. Le classement PageRank place 7 pays de l'UE27 devant tous les autres pays, la $8^{\text{ème}}$ place étant occupée par les États-Unis (US). On voit alors que l'UE27 et plus particulièrement les Pays-Bas et la Belgique ont une place centrale dans le réseau du commerce international, en

[3] Dans le cas où un pays a un volume d'export ou d'import total nul, on remplace les composantes des vecteurs **v** et **v'** correspondantes par $\frac{1}{N_p N_c}$.

terme d'importation de gaz. Le Qatar (QA), premier exportateur de gaz, selon le classement ExportRank, est 4$^{\text{ème}}$ dans le classement CheiRank. Ainsi, comme pour l'Arabie Saoudite et le pétrole, le Qatar possède une faible diversité de ses partenaires commerciaux pour le gaz.

Rang	PageRank	CheiRank	ImportRank	ExportRank	Rang	PageRank	CheiRank	ImportRank	ExportRank
1	US	RU	US	RU	20	PT	SE	LV	FI
2	SG	US	NL	SA	21	RO	PT	MT	LT
3	NL	AE	IN	US	22	BG	RO	CZ	DK
4	IN	IN	SG	AE	23	SK	DK	DK	PL
5	FR	SG	DE	NL	24	GR	BG	LT	PT
6	DE	SA	IT	CA	25	MT	LT	RO	RO
7	ES	NL	FR	IQ	26	SA	PL	IE	BG
8	GB	BE	GB	SG	27	RU	HU	HU	SK
9	IT	GR	BE	KW	28	LT	AT	SK	AT
10	BE	NG	ES	NG	29	IE	SK	SA	LV
11	CA	IT	CA	IN	30	CY	LV	BG	MT
12	AE	DE	SE	GB	31	DK	MT	SI	HU
13	NG	CA	PL	BE	32	FI	CZ	RU	CZ
14	PL	IQ	NG	DE	33	LV	SI	EE	EE
15	SI	KW	AE	IT	34	LU	CY	LU	SI
16	CZ	GB	GR	ES	35	IQ	EE	CY	IE
17	AT	ES	FI	FR	36	EE	IE	IQ	CY
18	SE	FR	AT	GR	37	KW	LU	KW	LU
19	HU	FI	PT	SE					

TABLE 4.1 : Classements des pays de l'UE27 (bleu) et des 10 plus grands exportateurs de pétrole hors UE (rouge). Classements effectués avec les données UN Comtrade pour l'année 2016. D'après [70].

Rang	PageRank	CheiRank	ImportRank	ExportRank	Rang	PageRank	CheiRank	ImportRank	ExportRank
1	NL	US	FR	QA	20	MY	PL	SE	IT
2	BE	CA	IT	NO	21	CZ	ES	BG	PL
3	FR	RU	GB	RU	22	SE	AT	LT	SE
4	IT	QA	US	US	23	AU	PT	RO	HU
5	GB	NO	DE	AU	24	IE	HU	LV	DK
6	ES	AU	BE	DZ	25	AE	IE	SI	RO
7	HU	NL	ES	MY	26	LT	SK	AU	SI
8	US	GB	NL	BE	27	NO	LT	RU	PT
9	DE	DZ	AE	CA	28	AT	RO	DK	GR
10	PT	AE	CA	AE	29	DK	CZ	EE	LT
11	BG	BE	ID	ID	30	CY	SI	NO	LU
12	SK	DE	CZ	NL	31	EE	LV	LU	LV
13	PL	IT	SK	GB	32	MT	BG	FI	FI
14	SI	FR	PT	DE	33	LV	FI	AT	MT
15	CA	SE	HU	FR	34	LU	LU	CY	IE
16	RO	ID	PL	ES	35	FI	MT	MT	EE
17	ID	DK	MY	AT	36	QA	EE	QA	BG
18	GR	MY	IE	SK	37	DZ	CY	DZ	CY
19	RU	GR	GR	CZ					

TABLE 4.2 : Classements des pays de l'UE27 (bleu) et des 10 plus grands exportateurs de gaz hors UE (rouge). Classements effectués avec les données UN Comtrade pour l'année 2016. D'après [70].

Matrices de Google réduites de l'UE27 + RU pour le pétrole

La Figure 4.1 présente les matrices de Google réduites G_{r} et G_{r}^{*} et leurs composantes associées aux transactions directes et indirectes entre l'UE27 et la Russie (RU) pour le pétrole. Les indices (cp) des colonnes et des lignes de ces matrices sont classées selon l'ordre du classement

PageRank de la Table 4.1. Le poids W_{pr} de la matrice G_{pr} est bien plus grand que les poids associés aux matrices G_{qr} et G_{rr}. Contrairement au réseau Wikipédia (voir section 3.1), où $W_{\mathrm{pr}} \approx 0.95$ et les 0.05 restants sont presque également répartis entre les composantes G_{rr} et G_{qr}, le réseau du commerce international a une structure bien différente. En effet, nous avons $W_{\mathrm{pr}} = 0.651568$, $W_{\mathrm{rr}} = 0.30849$, $W_{\mathrm{qr}} = 0.039942$ et $W_{\mathrm{qrnd}} = 0.036512$. Il en découle que la matrice G_{qr} a une contribution moins importante dans le réseau réduit, ce dernier est donc dominé par les liens directs, $W_{\mathrm{rr}} \gg W_{\mathrm{qr}}$. Ce résultat vient du fait que pour le commerce international, la matrice stochastique S représentant les transitions de probabilités entre les nœuds pays-produit, est plus dense que celle associée à Wikipédia. De même pour la matrice de Google réduite G^* associée au réseau inversé, nous avons une dominance caractérisée par les liens directs. Nous avons $W^*_{\mathrm{pr}} = 0.6051$, $W^*_{\mathrm{rr}} = 0.34379$, $W^*_{\mathrm{qr}} = 0.05111$ et $W^*_{\mathrm{qrnd}} = 0.047$.

La méthode de la matrice de Google réduite appliquée au commerce de pétrole entre les pays de l'UE27 et RU nous permet d'extraire les transactions directes et indirectes importantes. L'élément associé à la $j^{\text{ème}}$ colonne et $i^{\text{ème}}$ ligne des matrices de la partie gauche de la Figure 4.1 décrit le lien $j \to i$, le flux de produit passant de j à i. Le lien Irlande (IE) $\to$ GB, correspondant à une importation de pétrole depuis IE vers GB, est le lien le plus fort selon les matrices G_{r} et G_{rr}. L'importance de cette transaction est due au fait que IE et GB ont, tous deux, des territoires en île d'Irlande. Le second lien le plus important, selon les matrices G_{r} et G_{rr}, est Danemark (DK) $\to$ Suède (SE). Les liens indirects forts, observés à l'aide de la matrice G_{qrnd}, mettent en relation des pays ayant une frontière commune. La transaction indirecte la plus importante est Portugal (PT) $\to$ Espagne (ES) suivie par les transactions Roumanie (RO) $\to$ Bulgarie (BG) et Chypre (CY) $\to$ IT. Les valeurs maximales de la matrice G_{pr} sont localisées sur les lignes associées à NL, FR, Allemagne (DE), ES, GB et IT, ce qui correspond bien aux pays les mieux classés du classement PageRank des importateurs de pétrole. On note que pour les matrices relatives au réseau non inversé, la contribution de RU est nulle. Puisque la matrice de Google réduite G_{r} est construite à partir du réseau non inversé et de G_{pr}, dont les colonnes sont proches du PageRank, il vient que le rôle de fournisseur de pétrole de RU est masqué par le commerce intra-européen. Nous avons aussi observé, pour le cas du pétrole saoudien et américain, une contribution de SA et US masquée par les transactions entre membres de l'UE27 (voir la Figure A.4, pour le pétrole saoudien, et la Figure A.5, pour le pétrole américain, en annexes).

Les 4 matrices présentes en bas de la Figure 4.1 décrivent le réseau inversé. Les liens ayant été inversé, le lien $j \to i$ est maintenant décrit à la colonne i, ligne j. Les contributions du pétrole russe sont importantes ici, les pays ciblés par l'exportation de pétrole russe, selon la matrice de Google réduite G^*_{r}, sont dans l'ordre décroissant en poids, la Lettonie (LV), la Lituanie (LT), la Finlande (FI), BG, la Pologne (PL) et l'Estonie (EE). Du fait que ces pays soient géographiquement proches de RU, ces transactions sont plus importantes. On note aussi que NL et BE participent aussi à d'importantes exportation de pétrole, depuis NL vers BE et depuis BE vers le Luxembourg (LU).

Matrices de Google réduites de l'UE27 + RU pour le gaz

Les matrices de Google réduites G_{r} et G^*_{r}, ainsi que leurs composantes, associées à UE27 et RU pour le gaz, sont présentées à la Figure 4.2. Tout comme pour le cas du pétrole, les poids associés aux liens indirects sont bien plus faibles que ceux associés aux liens directs. Nous avons $W_{\mathrm{pr}} = 0.634069$, $W_{\mathrm{rr}} = 0.3008960$, $W_{\mathrm{qr}} = 0.056971$ ($W_{\mathrm{qrnd}} = 0.051085$), $W^*_{\mathrm{pr}} = 0.611761$, $W^*_{\mathrm{rr}} = 0.322066$ et $W^*_{\mathrm{qr}} = 0.066173$ ($W^*_{\mathrm{qrnd}} = 0.058111$). On note aussi que ces poids sont proches de ceux obtenus pour le cas précédent de l'UE27 et RU pour le pétrole.

La partie haute de la Figure 4.2 représente les matrices associées à G. Il y a plus d'éléments à poids forts dans la matrice G_{r} pour la gaz que pour le pétrole (voir Figure 4.1). Ces éléments décrivent les liens CY $\to$ IT, IE $\to$ GB, DK $\to$ SE, BE $\to$ FR, LT $\to$ PL et ES $\to$ PT. La matrice G_{pr} montre que NL est un efficace importateur européen de gaz. Les éléments de

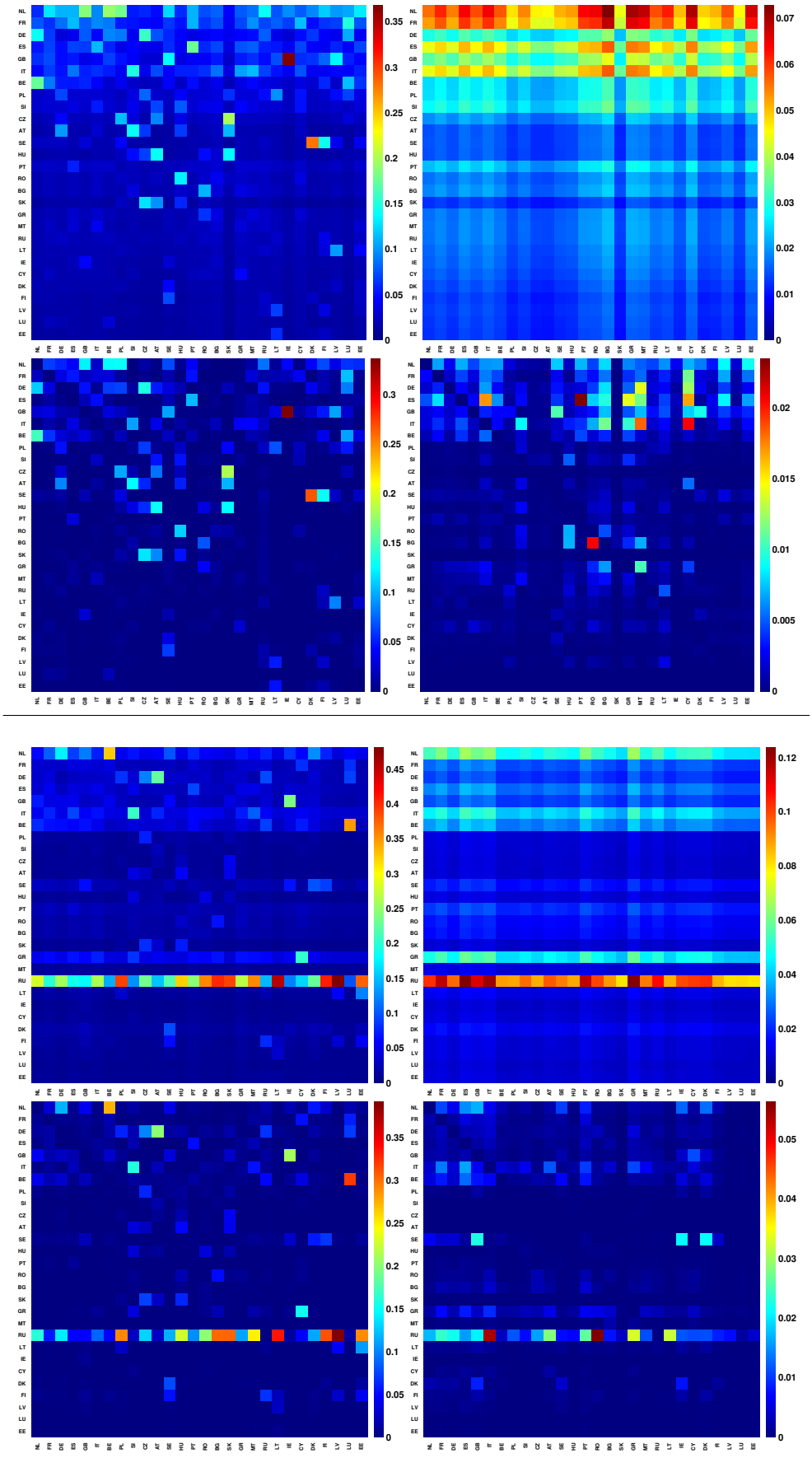

FIGURE 4.1 : Matrices de Google réduite associées à G (4 matrices du haut) et G^* (4 matrices du bas), pour l'année 2016 et le pétrole russe. Pour chaque ensemble de 4 matrices, la matrice de Google réduite G_{r} est en haut à gauche et ses composantes sont G_{pr} (en haut à droite), G_{rr} (en bas à gauche) et G_{qrnd} (en bas à droite). Les lignes et colonnes sont classées suivant le classement PageRank de la Table 4.1. D'après [70].

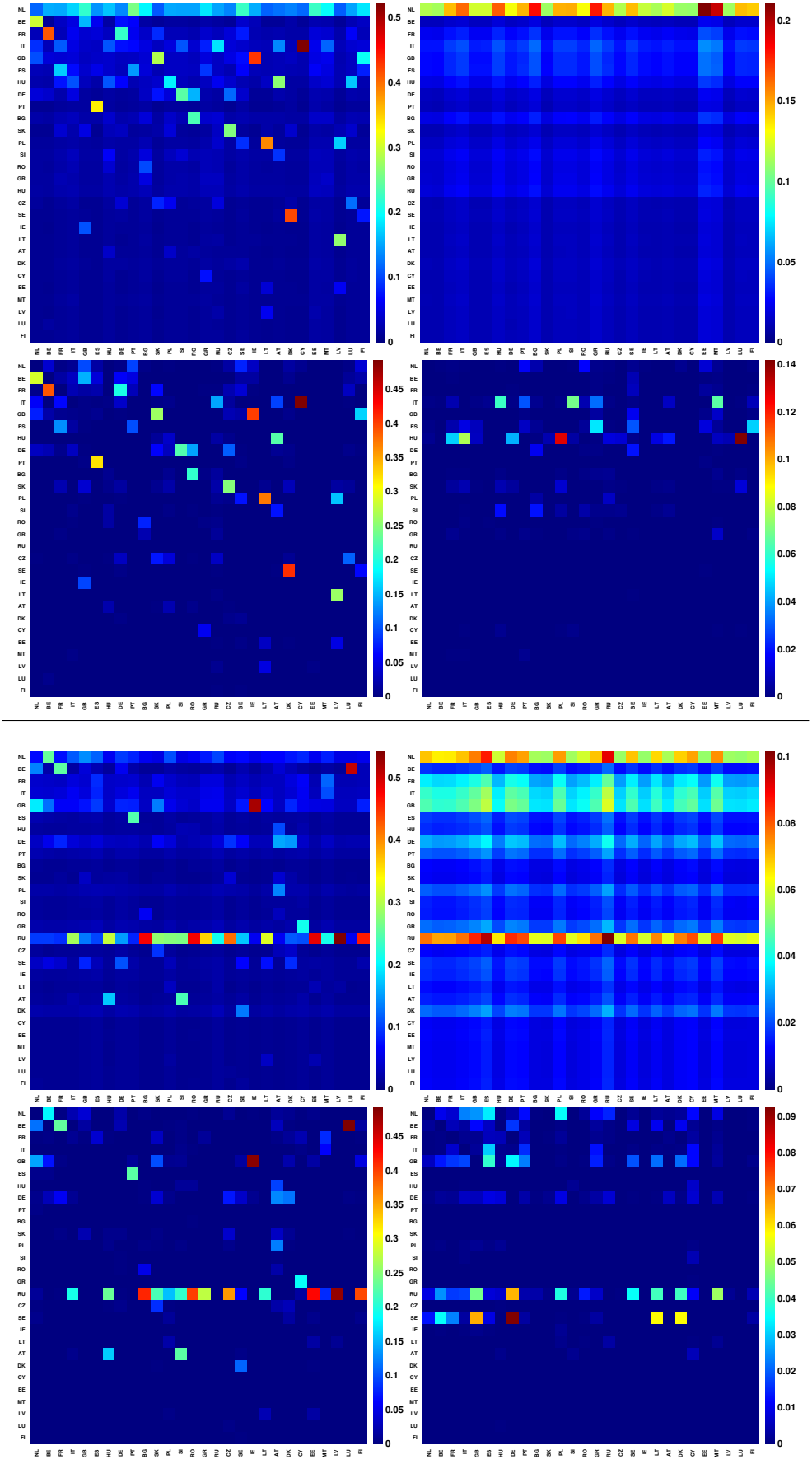

FIGURE 4.2 : Matrices de Google réduites et leurs composantes pour le gaz russe et l'année 2016. La disposition des matrices est identique à celle de la Figure 4.1. Les lignes et colonnes sont classées suivant le classement PageRank de la Table 4.2. D'après [70].

G_{pr}, associés à l'importation de gaz de NL sont supérieurs, d'au moins un ordre de grandeur, aux éléments relatifs au reste de l'UE27. Les éléments de la matrice G_{pr} associés à FR, bien qu'il soit le premier pays importateur de gaz selon classement ImportRank de la Table 4.2, ne sont pas importants. Parmi les liens indirects les plus forts, donnés par G_{qrnd}, nous avons LU $\rightarrow$ HU, PL $\rightarrow$ HU et IT $\rightarrow$ HU. La partie du bas de la Figure 4.2 associée au réseau inverse montre la forte contribution de RU dans l'alimentation en gaz de l'UE27. Nous avons les liens partant de RU et se dirigeant vers LV, EE, FI, BG et RO qui sont parmi les plus forts de la matrice G^*_{r}. Les transactions BE $\rightarrow$ LU et GB $\rightarrow$ IE sont aussi importantes et font de BE et GB d'importants exportateurs de gaz. La matrice G^*_{pr} montre une compétition entre NL et RU, en effet, les éléments présents dans les lignes correspondantes sont de même ordre de grandeur. La matrice G^*_{qrnd} permet de voir l'importance des transactions indirectes à destination de DE et en provenance de SE et RU. Le Royaume-Uni (GB) est aussi impliqué dans des transactions indirectes importantes de gaz en provenance de SE et RU.

Réseaux réduit pour le commerce du gaz et du pétrole en Europe

Au vue de la faible densité en poids fort de G_{qr} par rapport à G_{rr}, nous allons construire les réseaux réduits en utilisant la matrice de Google réduite G_{r}, pour l'import et G^*_{r} pour l'export. Pour chaque nœud (pays/pétrole), nous traçons les 4 liens sortant les plus forts selon la matrice de Google réduite considérée. La Figure 4.3, donne les réseaux réduits des 27 membres de l'UE et de la Russie pour l'import du pétrole (à gauche) et l'export de pétrole (à droite), pour l'année 2016. Parmi les 6 pays européens les mieux classés en terme de PIB en 2016 (DE, GB, FR, IT, ES et NL), FR et NL sont les pays les plus centraux du réseau réduit de la partie gauche de la Figure 4.3. On observe aussi des pays ayant beaucoup de liens entrants et sortants dans ce réseau. Ils peuvent être alors considérés comme des portes d'entrée du pétrole russe en Europe. Nous avons DE, FR, NL et IT qui jouent ce rôle. Il est intéressant de noter la présence de boucles fermées, mettant en jeu des pays voisins. Nous avons par exemple, DE $\longleftrightarrow$ AT, République Tchèque (CZ) $\longleftrightarrow$ Slovaquie (SK), DE $\longleftrightarrow$ PL, AT $\longleftrightarrow$ HU, AT $\longleftrightarrow$ SK, PT $\longleftrightarrow$ ES, ES $\longleftrightarrow$ IT et SE $\longleftrightarrow$ FI. La partie droite de la Figure 4.3 représente le réseau réduit des échanges pétroliers entre l'UE27 et la Russie, construit à partir de la matrice de Google réduite G^*_{r}. Tout comme pour le réseau construit à partir de G_{r}, nous plaçons pour les 4 liens sortants les plus importants pour chaque pays considéré. Enfin, nous inversons la direction des liens car, comme expliqué plus haut, la matrice G^*_{r} est construite à partir de G^*. Nous inversons les liens après construction du réseau réduit. On retrouve bien le fait que RU est la première source de pétrole en Europe, il a le plus de liens sortants $k_{\text{out}} = 27$. Les Pays-Bas (NL) apparaissent comme le second meilleur exportateur de pétrole en Europe. De façon moins notable, on retrouve GR, IT, BE, GB, SE et DE comme étant de grands exportateurs de pétrole en UE.

La Figure 4.4, montre les réseaux réduits des échanges commerciaux de gaz entre UE27 et RU pour l'année 2016. À gauche se trouve le réseau réduit construit via G_{r}. On peut observer que NL est le pays le plus central de ce réseau. À la différence du pétrole, on peut remarquer que HU et CZ participent plus significativement à l'alimentation de l'UE en gaz. En plus de ces deux pays, on retrouve encore DE, FR, IT et NL comme pays jouant le rôle de porte d'entrée du marché européen. On trouve aussi des boucles fermées entre pays, cependant les pays mis en jeu sont moins proche géographiquement que avec l'exemple du pétrole (voir Figure 4.3). Le Royaume-Uni est au cœur des échanges GB $\longleftrightarrow$ NL, GB $\longleftrightarrow$ BE, GB $\longleftrightarrow$ IE et GB $\longleftrightarrow$ SK, montrant ainsi sa grande diversité de partenaires commerciaux pour la gaz. On retrouve aussi des boucles entre pays voisins telles que FR $\longleftrightarrow$ ES, ES $\longleftrightarrow$ PT, BG $\longleftrightarrow$ RO, NL $\longleftrightarrow$ BE et LV $\longleftrightarrow$ LT. La partie droite de la Figure 4.4, montre le réseau réduit d'export (construit avec G^*_{r}). Comme pour la partie droite de la Figure 4.3, on voit que RU exporte du gaz à la majorité de l'UE27. Les seconds meilleurs exportateurs de ce réseau réduit sont NL et GB. FR, IT et DE sont aussi, en moindre mesure, de grands exportateurs de gaz dans ce réseau réduit.

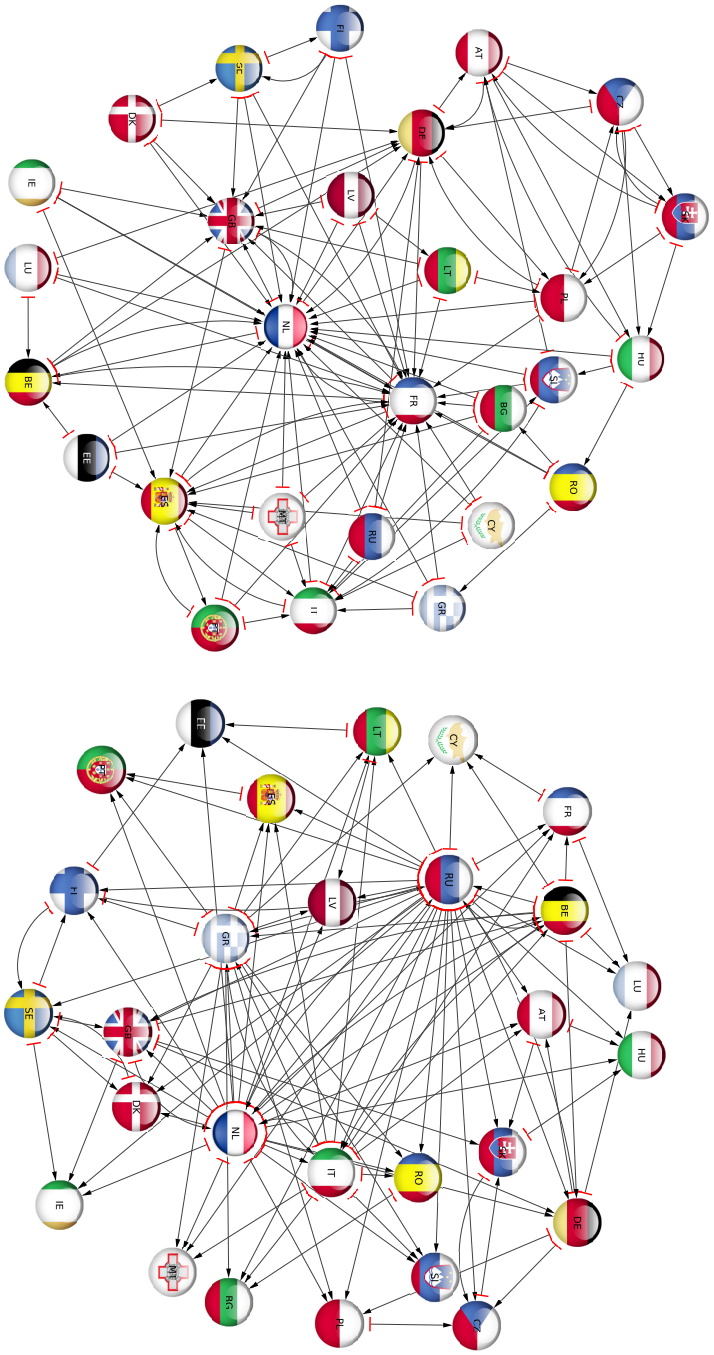

FIGURE 4.3 : Réseaux réduits des transactions de pétrole entre l'UE27 et la Russie pour l'année 2016. La partie gauche présente le réseau réduit d'importation construit à partir de G_r, la partie droite, le réseau d'exportation construit avec G_r^*. Pour chaque pays, les 4 liens sortants les plus forts, au regard des colonnes correspondantes dans G_r et G_r^*, sont affichés. La direction des liens indiquent la destination du pétrole. D'après [70].

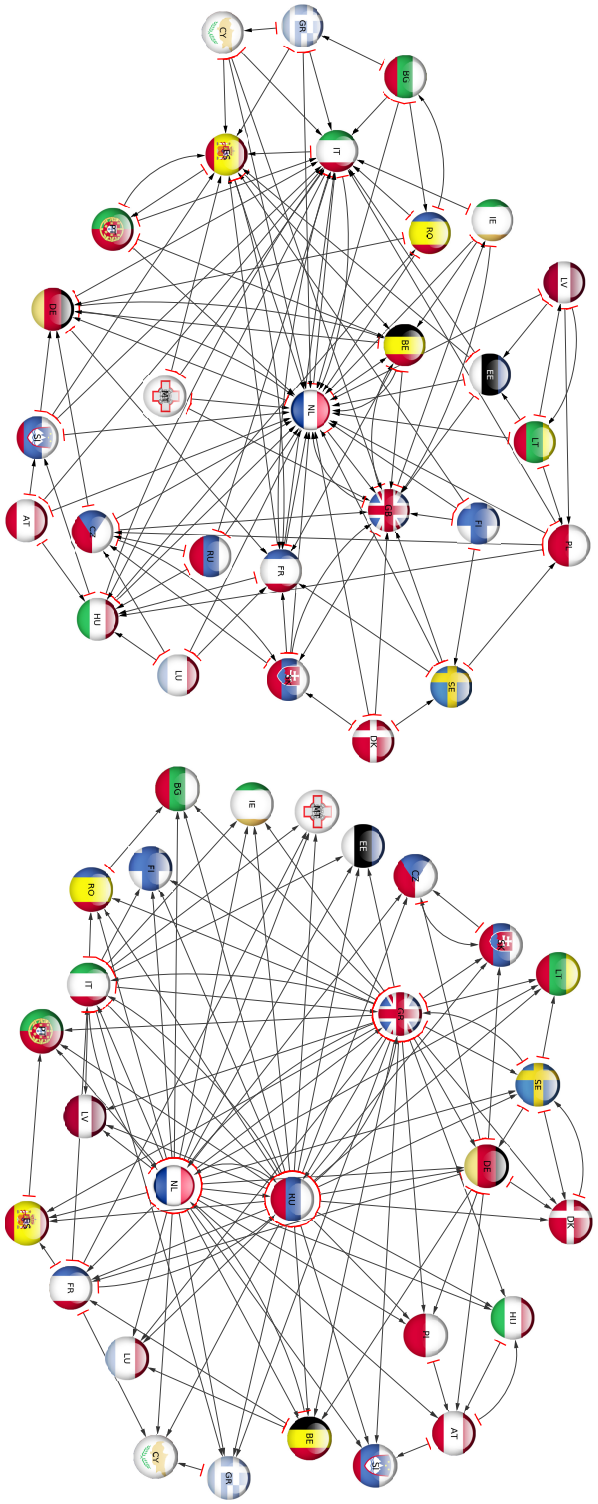

FIGURE 4.4 : Réseaux réduits des transactions de gaz entre l'UE27 et la Russie pour l'année 2016. La légende est identique à celle de la Figure 4.3. D'après [70].

4.1.4 Impacts économiques

Afin de mesurer l'impact économique sur l'UE27 du pétrole et du gaz importés depuis l'extérieur de l'Europe, nous avons utilisé une balance économique **B** basée sur le vecteur PageRank et le vecteur CheiRank, appelée balance PageRank-CheiRank. La démarche est la suivante, nous perturbons la matrice monétaire,[4] plus précisément nous multiplions les transactions de pétrole/gaz depuis un exportateur hors UE vers l'UE27 par un facteur $(1+\delta)$, où δ est un paramètre de perturbation. La balance PageRank-CheiRank d'un pays c est alors

$$B_c = \frac{P_c^* - P_c}{P_c^* + P_c} \tag{4.7}$$

avec les quantités $P_c = \sum_p P_{cp}$ et $P_c^* = \sum_p P_{cp}^*$ mesurant respectivement la capacité d'importation et d'exportation du pays c. La balance PageRank-CheiRank est comprise dans l'intervalle $[-1, 1]$, une valeur négative, relative à une trop grande capacité importatrice, est caractéristique d'un pays en déficit économique tandis qu'une valeur positive indique une bonne santé économique. Enfin, nous mesurons l'impact économique comme la sensibilité de la balance PageRank-CheiRank. La sensibilité de la balance PageRank-CheiRank du pays c vis-à-vis d'une variation infinitésimale du flux de pétrole/gaz du pays c' vers le pays c est

$$\mathcal{D}_{c'\to c}(c) = \frac{dB_c}{d\delta}. \tag{4.8}$$

Une valeur négative de la sensibilité économique peut être interprétée comme une décroissance économique. Au contraire, une valeur positive de $\mathcal{D}$ montre la résistance économique d'un pays face à un changement du prix du pétrole et ou de gaz. Dans la construction des réseaux réduits des transactions pétrole et gaz de l'UE27 et des grands exportateurs, nous avons utilisé comme nœuds réduits, l'ensemble des 27 pays de l'UE pour le pétrole et ou le gaz plus l'exportateur de pétrole et ou de gaz. Afin d'avoir le plus d'information possible, nous avons choisi d'utiliser l'ensemble des 61 produits pour l'UE27 et uniquement le pétrole et ou le gaz pour l'exportateur. Ainsi nous avons une matrice de Google réduite basée sur $N_{\mathrm{r}} = 27 \times 61 + 1 = 1648$ nœuds. Pour chaque pays européen, nous aurons donc les 61 nœuds pays-produit présents dans notre ensemble de nœuds réduits auquel nous rajoutons 1 nœud représentant l'exportateur non-UE pour le produit dont on veut mesurer l'impact. On va donc pouvoir perturber directement G_{r} au lieu de modifier la matrice monétaire.

Sensibilité de l'UE face au pétrole étranger

La Figure 4.5 montre la distribution géographique pour l'année 2016 de l'impact économique du pétrole russe (en haut à gauche), du pétrole en provenance d'Arabie Saoudite (SA) (en bas à gauche) et du pétrole américain (en bas à droite) sur l'UE27. Le code couleur va du rouge, pour les sensibilités les plus négatives, au le bleu, pour les plus positives. Comme on peut le voir, les sensibilités positives sont rares. Nous obtenons de telles sensibilités pour l'impact économique du pétrole russe sur la Finlande (FI) ($\mathcal{D} = 2.10^{-4}$). La Grèce (GR) est aussi, en 2016, positivement impactée par le pétrole saoudien avec une valeur de $\approx 6.10^{-5}$. Quelque soit le fournisseur de pétrole, nous observons que les Pays-Bas (NL) sont le pays le plus impacté. Ce résultat montre son rôle important dans la distribution de pétrole au sein de l'UE. Pour le pétrole russe, d'autres pays sont impactés comme l'Italie (IT), GR, la Bulgarie (BG), la Pologne (PL), la Lituanie (LT) et la Lettonie (LV). Le pétrole saoudien impacte, en plus de NL, l'Espagne (ES), tandis qu'une partie de l'UE27 semble plutôt stable devant le pétrole américain. Les effets du pétrole saoudien sur l'UE sont d'un ordre de grandeur de 3 à 4 fois plus petit que ceux associées au pétrole russe. En effet, nous avons par exemple, l'Allemagne

[4] Nous faisons l'hypothèse que dans un tel réseau, si nous perturbons d'un facteur $(1+\delta)$ la masse monétaire de pétrole importé par un pays c, cela revient à changer le prix du pétrole/gaz associé à cette transaction.

(DE) qui est 5 fois plus impactée par le pétrole russe que par le pétrole saoudien. En revanche l'impact du pétrole en provenance des US est deux fois plus important que celui du pétrole russe. L'économie allemande est impactée de la même manière par le pétrole russe et américain (avec respectivement $\mathcal{D} = -5.10^{-4}$ et -6.10^{-4}). La partie de l'Europe des 27 située à l'est reste très stable devant une variation du prix du pétrole américain. La partie en haut à droite de la Figure 4.5 montre la distribution géographique des sensibilités économiques associées à l'UE27 face au pétrole russe lorsqu'on utilise la balance économique standard $\hat{B}$ qui est définie par le volume total d'importation et d'exportation. Dans ce cas, LV est le pays le plus impacté, suivi par LT. Ces deux pays, de part leur histoire avec RU et l'URSS, ont des échanges commerciaux étroits avec RU. On observe aussi que l'ouest de l'Europe est plus robuste que lorsqu'on utilise la balance PageRank-CheiRank B. La différence, entre la méthode usuelle en économie et celle que nous proposons dans cette étude, vient principalement du fait que la méthode de la matrice de Google réduite permet la prise en compte des transactions indirectes en plus des transactions directes. Aussi, comme on peut le voir en comparant les panneaux, en haut à gauche et en haut à droite de la Figure 4.5, relatifs à ces deux mesures de balance économique, on observe que la

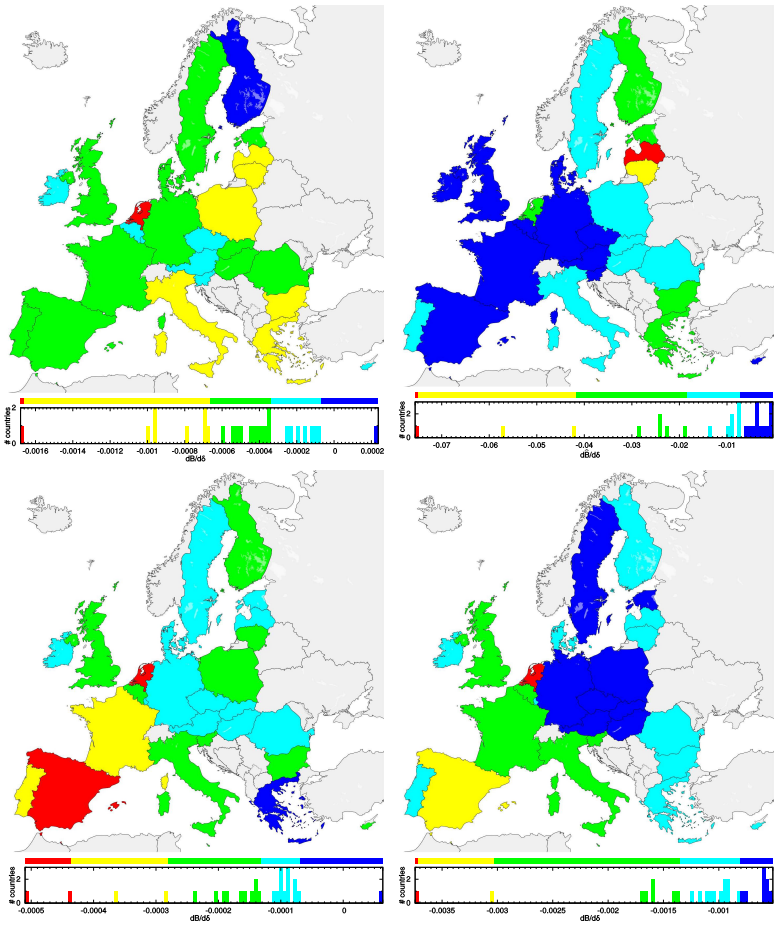

FIGURE 4.5 : Distribution géographique de la dérivée de la balance économique $dB/d\delta$ associée à l'UE27 et induite par une hausse du prix du pétrole pour l'année 2016. En haut à gauche, le cas du pétrole russe, en bas à gauche, le pétrole saoudien et américain en bas à droite. En haut à droite est présentée la variation de la balance économique standard $\hat{B} = \frac{\hat{P}^* - \hat{P}}{\hat{P}^* + \hat{P}}$ pour l'année 2016 et le pétrole russe. Le code couleur va du rouge, pour les variations les plus négatives, au bleu, pour les variations les plus positives. Les catégories ont été calculer avec la méthode de Jenks [48]. D'après [70].

balance PageRank-CheiRank donne plus de nuances, ainsi que des résultats plus proches de la réalité. Pour cette raison, nous utiliserons uniquement la balance PageRank-CheiRank B dans la suite de cette étude.

Les sensibilités mesurées pour les années 2004, 2008 et 2012 sont montrées à la Figure 4.6. La colonne de gauche montre les distributions géographiques de l'impact économique résultant du pétrole russe, la colonne centrale montre le cas du pétrole saoudien et la colonne de droite, le cas du pétrole américain. De bas en haut, se trouvent les résultats pour 2004, 2008 et 2012. En 2004, le pétrole russe touche fortement NL, IT, et la République de Chypre (CY). En 2008, NL et IT sont les pays les plus sensibles. En 2012 et en 2016 (voir la Figure 4.5 pour 2016), les NL sont très impactés. De 2004 à 2012, la valeur de sensibilité la plus négative passe de -0.0016 à -0.0029 et enfin vaut -0.0037 en 2012. Pour l'année 2016, la sensibilité minimum retrouve une valeur semblable à l'année 2004. On voit que le pétrole russe impacte beaucoup plus l'UE27 avant 2016. Nous pouvons observer les effets de la chute du prix du pétrole, dus à la crise financière de 2007-2008. Nous observons aussi ce scénario pour le pétrole saoudien et américain. En 2004, le pétrole saoudien impacte de façon importante NL et GR. NL et IT sont les plus impactés en 2008. NL et ES le sont en 2012 et en 2016 (voir la Figure 4.5). Les

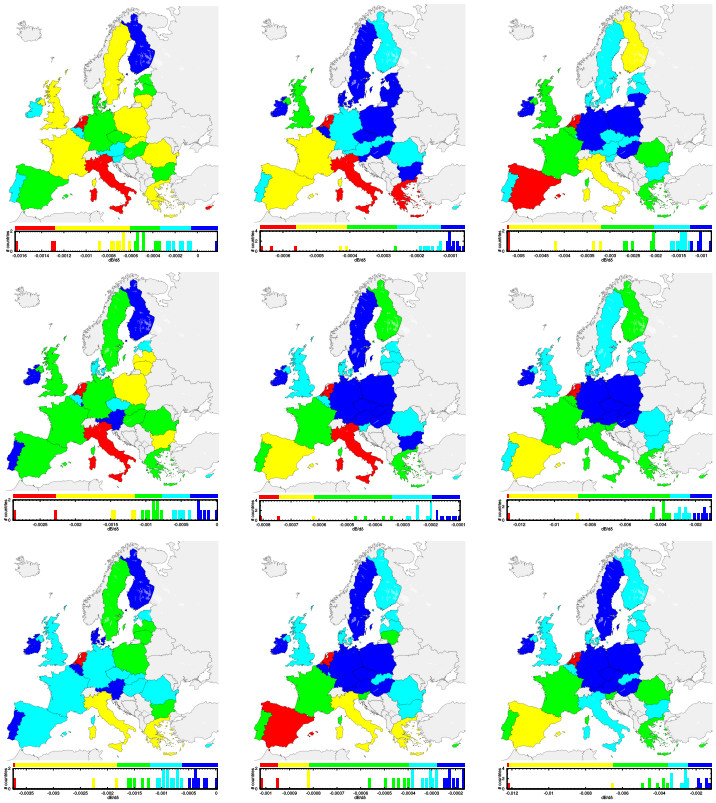

FIGURE 4.6 : Distribution géographique de la dérivée de la balance économique pour l'UE27 dans le cas du pétrole en provenance de la Russie (colonne de gauche), de l'Arabie Saoudite (colonne centrale) et des États-Unis (colonne de droite). Plusieurs années ont été analysées, 2004 (en haut), 2008 (milieu) et 2012 (en bas). D'après [70].

sensibilités les plus négatives sont égales à -0.0006 (2004), -0.0008 (2008), -0.001 (2012) et -0.0005 (2016). Le pétrole américain impacte grandement NL et ES en 2004 et NL en 2008, 2012 et 2016 (voir la Figure 4.5). De même que pour SA et RU, la valeur de sensibilité la plus négative, associée au pétrole américain, augmente entre 2004 et 2012 et revient, en 2016, à une valeur similaire à celle de l'année 2004. Nous avons -0.0052 en 2004, -0.0127 en 2008, -0.0122 en 2012 et -0.0037 en 2016. Les anciens pays de l'URSS membres de l'UE27 ainsi

qu'une partie du centre de l'UE montrent une faible sensibilité face au pétrole américain et saoudien. Le rôle de fournisseur européen de NL, mais aussi son importance dans le commerce extérieur, rendent sa santé économique fragile devant un changement du prix du pétrole russe, saoudien et américain. L'impact économique du pétrole américain, de 2 à 3 fois plus important que celui associé au pétrole russe, vient du haut PageRank de l'US. On note aussi que FI est positivement impactée par le pétrole russe en 2004.

Sensibilité de l'UE face au gaz étranger

De même que pour l'étude de l'impact économique de l'UE27 face au pétrole étranger, nous nous intéressons à $N_r = 61 \times 27 + 1$ nœuds réduits. Nous avons étudié le cas de deux grands exportateurs de gaz, RU et la Norvège (NO). La Figure 4.7 montre l'impact de la hausse du prix du gaz russe (à gauche) et du gaz norvégien (à droite) sur l'économie de l'UE27 pour l'année 2016. On peut remarquer que l'Europe est moins impactée par le gaz russe que par le pétrole russe (d'environ un ordre de grandeur). Le gaz russe impacte grandement IT et de façon moins conséquente la Hongrie (HU). La partie ouest de l'Europe est stable face à un changement du prix du gaz. Tout comme pour le pétrole russe, le gaz en provenance de RU impacte préférentiellement les pays voisins de RU. La partie droite de la Figure 4.7 montre comment le gaz norvégien impacte l'UE27. Contrairement au gaz russe, on observe une sensibilité économique positive pour un pays. En effet, SE a une sensibilité $\mathcal{D} = 4.4.10^{-4}$. DE est le pays le plus impacté par le gaz norvégien, suivie par GB et BE. On voit aussi que 14 pays de l'UE27 ont une sensibilité presque nulle (de l'ordre de -10^{-7}).

La Figure 4.8 et la Figure 4.9 montrent l'évolution dans le temps de l'impact économique du gaz russe et norvégien. L'Europe de l'ouest, de PT à DE, reste très stable face à une hausse du prix du gaz russe. IT est cependant fortement impactée depuis 2012 et FR est impactée uniquement en 2004. Les pays les plus impactés par le gaz russe sont HU, en 2004 et en 2008, IT et SK en 2012 et seulement IT en 2016.

La Figure 4.9 montre le cas du gaz norvégien. On voit que, contrairement au gaz russe, le gaz norvégien n'impacte pas les pays de l'est de l'UE. Les pays les plus impactés sont FR, BE et DE en 2004, BE en 2008, BE, NL et DE en 2012 et finalement DE en 2016. L'économie suédoise bénéficie de la hausse du prix du gaz norvégien puisqu'elle a une sensibilité positive en 2004, 2012 et en 2016. GB est aussi un pays positivement impacté par le gaz en provenance de NO (en 2008 seulement).

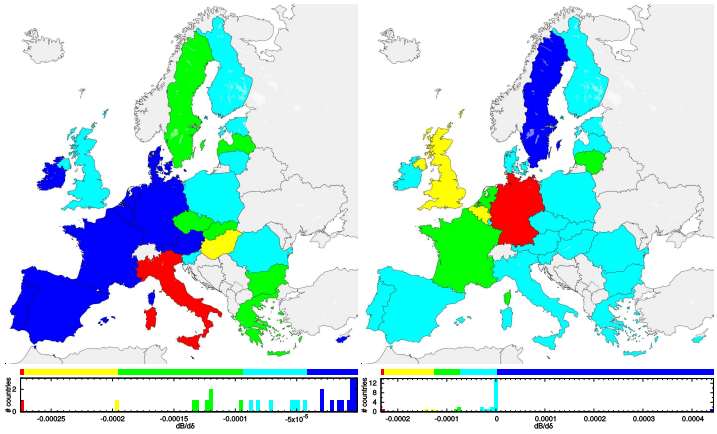

FIGURE 4.7 : Distribution géographique de la dérivée de la balance économique pour l'UE27 dans le cas du gaz exporté par la Russie (à gauche) et par la Norvège (à droite) en 2016. D'après [70].

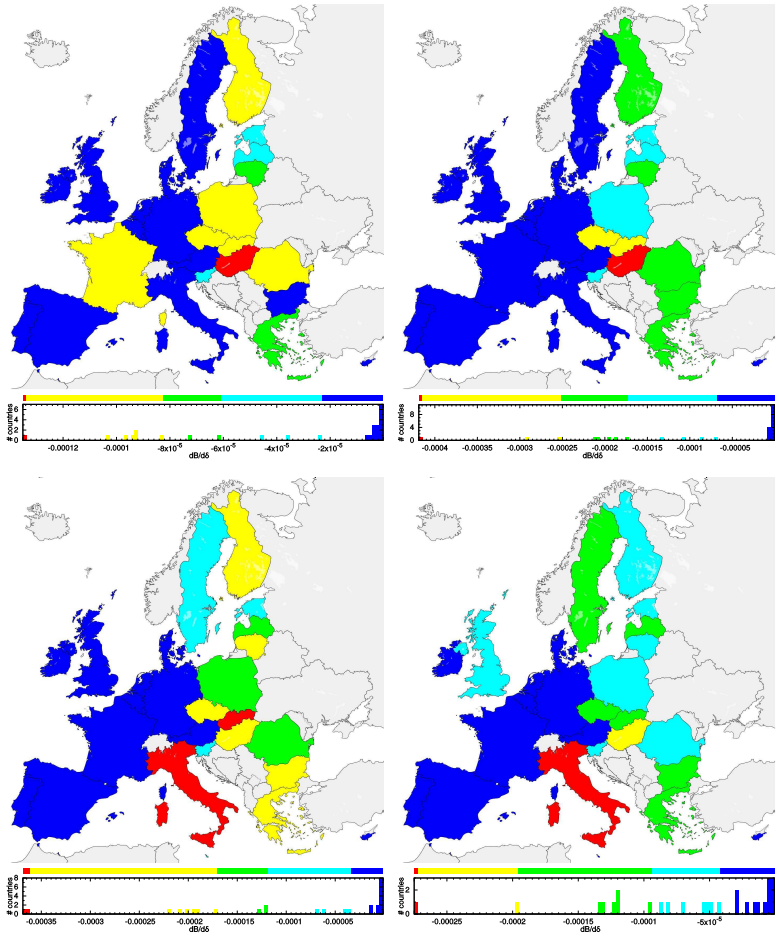

FIGURE 4.8 : Distribution géographique de la dérivée de la balance économique pour l'UE27 dans le cas du gaz russe, en 2004 (en haut à gauche), 2008 (en haut à droite), 2012 (en bas à gauche) et 2016 (en bas à droite). D'après [70].

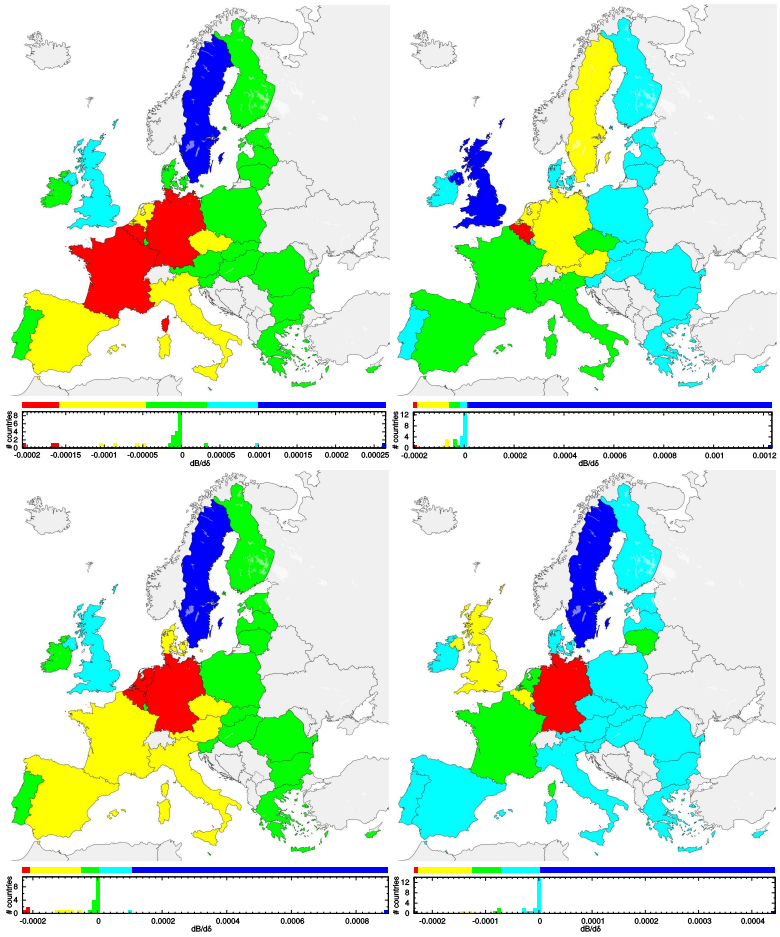

FIGURE 4.9 : Distribution géographique de la dérivée de la balance économique pour l'UE27 dans le cas du gaz norvégien, en 2004 (en haut à gauche), 2008 (en haut à droite), 2012 (en bas à gauche) et 2016 (en bas à droite). D'après [70].

4.1.5 Conclusion

Nous avons utilisé la méthode de la matrice de Google réduite afin de pouvoir étudier les transactions directes et indirectes entre les membres de l'Europe des 27 (UE27) sans la Croatie. Les données du commerce international utilisées, relatives aux années 2004, 2008, 2012 et 2016, proviennent de la base de données Comtrade des Nations Unis. Les matrices de Google réduites G_{r} et G_{r}^{*} nous ont permis d'étudier l'impact de la hausse du prix du pétrole et du gaz étrangers sur l'économie de l'UE27, sur la période 2004-2016. Nous avons constaté que les économies des pays de l'UE27 se trouvaient impactées différemment selon la source des ces transactions. En effet, la Russie (RU) et son pétrole n'impactent pas les mêmes pays que le pétrole saoudien ou encore américain. Le pétrole russe touche préférentiellement les pays voisins de la Russie et le centre de l'UE27 tandis que les États-Unis (US) et l'Arabie Saoudite (SA) influencent l'économie de l'Europe de l'ouest. Cette étude montre le rôle important des Pays-Bas (NL) dans l'import et l'export de pétrole. On voit aussi que l'économie de NL est la plus impactée, quelque soit l'exportateur de pétrole, pour toute la période 2004-2016. De la même manière que pour son pétrole, RU impacte l'économie de l'Europe de l'est, et plus particulièrement des anciens membres de l'URSS, par le gaz qu'elle exporte. Les pays de l'UE27 qui sont les plus robustes à une variation du prix du gaz russe sont situés à l'ouest de l'Europe, à l'exception de l'Italie (IT). La Norvège (NO), second fournisseur de gaz en Europe, impacte très peu les économies des pays de l'UE27. Quelques pays sont néanmoins impactés. L'Allemagne (DE) est impactée sur la période 2004-2016, la France (FR) en 2004, la Belgique (BE) en 2004 et 2012 et NL en 2012. Cette étude montre aussi que certains pays bénéficient de l'augmentation du prix du pétrole et du gaz étrangers. En effet, la Finlande (FI) est positivement impactée par le pétrole russe en 2004 et en 2016, La Grèce (GR) bénéficie de la variation du pétrole saoudien en 2016. Le gaz en provenance de NO impacte positivement 3 pays : la Suède (SE) en 2004 et de 2012 à 2016, le Royaume-Uni (GB) de 2004 à 2012, et le Danemark (DK) en 2004. La méthode de la matrice de Google réduite permet une analyse détaillée et réaliste du commerce international, en effet, avec cette méthode nous avons aussi accès aux transactions indirectes, par diffusion dans tous le réseau. Nous allons, dans un prochain chapitre, montrer comment construire à l'aide de la balance économique basée sur le vecteur PageRank et le vecteur CheiRank, un modèle de propagation de crise économique, et seront présentés les résultats relatifs à l'application de ce modèle au réseau du commerce international.

4.2 Le réseau mondial d'activités économiques (RMAE) et interdépendance des secteurs économiques

4.2.1 Introduction

Les données provenant de UN Comtrade décrites dans l'étude précédente fournissent des informations sur les échanges commerciaux entre un grand nombre de pays, cependant elle présente une économie mondiale sans prendre en compte la complexité de fabrication des produits. En effet, le cycle de production peut dépendre de plusieurs secteurs industriels. La banque de données OECD-WTO TiVA, provenant de l'Organisation de coopération et de développement économiques (OCDE) et de l'Organisation Mondiale du Commerce (OMC), fournit des informations sur les échanges entre les secteurs de plusieurs pays en valeur ajouté. Nous nous sommes intéressés à l'analyse et à la comparaison des interactions entre secteurs d'activités économiques de différentes puissances économiques. L'utilisation de la méthode de la matrice de Google réduite, introduite dans [30], a permis d'étudier les interactions directes et indirectes entre les secteurs d'un pays. Les interactions indirectes sont importantes, elles capturent les informations relatives aux chaînes de transformation des produits. L'efficacité de la matrice de Google réduite est montrée dans de nombreuses études impliquant le réseau d'articles Wikipédia [31, 46, 60, 61, 71, 72], le réseau d'interaction entre protéines [45] et aussi

le réseau du commerce international [70]. Une multitude de travaux sur le réseau du commerce international existent, notamment les études [73, 74, 75, 76, 77, 78], mais n'utilisent pas de matrice de Google. Une étude récente [79] donne l'analyse de la matrice de Google associée au réseau mondial des activités économiques. Nous avons décidé d'étendre cette dernière étude avec l'utilisation de la méthode de la matrice de Google réduite, et de s'intéresser à l'interdépendance des secteurs d'activités économiques. Dans la suite de ce chapitre, nous présenterons notre étude de la manière suivante. Nous commencerons par la présentation des données OECD-WTO TiVA ainsi que la méthode de construction de la matrice de Google pour le réseau mondial des activités économiques (section 4.2.2). Ensuite, nous présenterons les résultats relatifs à l'interdépendance des secteurs économiques des États-Unis (USA) et à celle des secteurs économiques de la Chine (CHN) (section 4.2.3). La sensibilité économique des secteurs de productions face à des variations de prix des produits provenant d'autres secteurs, notamment des secteurs de produits issus du pétrole, sera présenté (section 4.2.4). Enfin, nous caractériserons les économies de la Russie (RUS), des USA et de la CHN par la construction de réseaux réduits de transaction entre secteurs économiques.

4.2.2 La matrice de Google pour le réseau mondial des activités économiques

Nous avons utilisé les données provenant de la base OECD-WTO TiVA datant de l'année 2013. Ces données représentent les échanges économiques en valeur ajouté entre 37 secteurs d'activités, présentés dans la Table 4.3, pour un total de 58 pays[5] et pour les années 1995, 2000, 2005, 2008 et 2009. La liste de ces pays est présentée dans la Table A.3 située en annexes. Les secteurs sont classés selon la *International Standard Industrial Classification* (ISIC) $3^{\text{ème}}$ révision. Les secteurs $s = 1, \ldots, 21$ représentent les secteurs de production tandis que les secteurs $s = 22, \ldots, 37$ représentent les secteurs de service. Pour chaque année, nous construisons à partir des données OECD-WTO TiVA une matrice d'échanges économiques M. L'entrée $M_{cc',ss'}$ de cette matrice donne la masse monétaire associée au volume d'export depuis le secteur s' du pays c' vers le secteur s du pays c en dollar US. Dans le réseau mondial des activités économiques, les nœuds représentent un couple pays-secteur et peuvent être vus comme les nœuds pays-produit du réseau du commerce international.

Nous construisons la matrice de Google du réseau mondial des activités économiques en suivant la méthode présentée dans les études [68, 79]. Nous avons un total de $N = 58 \times 37$ nœuds pays-secteur (cs). La matrice de Google G est construite en deux itérations et avec l'utilisation de deux vecteurs préférentiels. Nous excluons les échanges $M_{cc,ss}$ décrivant, pour un pays donné, les échanges au sein d'un même secteur. Comme il a été expliqué plus haut, les secteurs sont en quelque sorte similaire au produit du réseau du commerce international, ainsi, la construction de G et G^* se fait en remplaçant les indices p de (4.4), (4.5) et (4.6), par les indices s associés aux secteurs.

[5] Le $58^{\text{ème}}$ pays est labellisé *ROW* et désigne le reste du monde.

	OECD ICIO Category	ISIC Rev. 3 correspondence
1	C01T05 AGR	01 - Agriculture, hunting and related service activities 02 - Forestry, logging and related service activities 05 - Fishing, operation of fish hatcheries and fish farms ; service activities incidental to fishing
2	C10T14 MIN	10 - Mining of coal and lignite ; extraction of peat 11 - Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction excluding surveying 12 - Mining of uranium and thorium ores 13 - Mining of metal ores 14 - Other mining and quarrying
3	C15T16 FOD	15 - Manufacture of food products and beverages 16 - Manufacture of tobacco products
4	C17T19 TEX	17 - Manufacture of textiles 18 - Manufacture of wearing apparel; dressing and dyeing of fur 19 - Tanning and dressing of leather ; manufacture of luggage, handbags, saddlery, harness and footwear
5	C20 WOD	20 - Manufacture of wood and of products of wood and cork, except furniture ; manufacture of articles of straw and plaiting materials
6	C21T22 PAP	21 - Manufacture of paper and paper products 22 - Publishing, printing and reproduction of recorded media
7	C23 PET	23 - Manufacture of coke, refined petroleum products and nuclear fuel
8	C24 CHM	24 - Manufacture of chemicals and chemical products
9	C25 RBP	25 - Manufacture of rubber and plastic products
10	C26 NMM	26 - Manufacture of other non-metallic mineral products
11	C27 MET	27 - Manufacture of basic metals
12	C28 FBM	28 - Manufacture of fabricated metal products, except machinery and equipment
13	C29 MEQ	29 - Manufacture of machinery and equipment n.e.c.
14	C30 ITQ	30 - Manufacture of office, accounting and computing machinery
15	C31 ELQ	31 - Manufacture of electrical machinery and apparatus n.e.c.
16	C32 CMQ	32 - Manufacture of radio, television and communication equipment and apparatus
17	C33 SCQ	33 - Manufacture of medical, precision and optical instruments, watches and clocks
18	C34 MTR	34 - Manufacture of motor vehicles, trailers and semi-trailers
19	C35 TRQ	35 - Manufacture of other transport equipment
20	C36T37 OTM	36 - Manufacture of furniture ; manufacturing n.e.c. 37 - Recycling
21	C40T41 EGW	40 - Electricity, gas, steam and hot water supply 41 - Collection, purification and distribution of water
22	C45 CON	45 - Construction
23	C50T52 WRT	50 - Sale, maintenance and repair of motor vehicles and motorcycles ; retail sale of automotive fuel 51 - Wholesale trade and commission trade, except of motor vehicles and motorcycles 52 - Retail trade, except of motor vehicles and motorcycles ; repair of personal and household goods
24	C55 HTR	55 - Hotels and restaurants
25	C60T63 TRN	60 - Land transport ; transport via pipelines 61 - Water transport 62 - Air transport 63 - Supporting and auxiliary transport activities ; activities of travel agencies
26	C64 PTL	64 - Post and telecommunications
27	C65T67 FIN	65 - Financial intermediation, except insurance and pension funding 66 - Insurance and pension funding, except compulsory social security 67 - Activities auxiliary to financial intermediation
28	C70 REA	70 - Real estate activities
29	C71 RMQ	71 - Renting of machinery and equipment without operator and of personal and household goods
30	C72 ITS	72 - Computer and related activities
31	C73 RDS	73 - Research and development
32	C74 BZS	74 - Other business activities
33	C75 GOV	75 - Public administration and defense ; compulsory social security
34	C80 EDU	80 - Education
35	C85 HTH	85 - Health and social work
36	C90T93 OTS	90 - Sewage and refuse disposal, sanitation and similar activities 91 - Activities of membership organizations n.e.c. 92 - Recreational, cultural and sporting activities 93 - Other service activities
37	C95 PVH	95 - Private households with employed persons

TABLE 4.3 : Liste des 37 secteurs économiques considérés dans la base de données OECD-WTO TiVA et leur équivalent avec la classification ISIC 3$^{\text{ème}}$ révision de l'ONU. Les descriptions des secteurs sont en anglais. D'après [80].

4.2.3 Interdépendance des secteurs économiques

Nous avons construit pour les États-Unis (USA) et la Chine (CHN), la matrice de Google réduite G_{r} associée au réseau mondial des activités économiques et la matrice de Google réduite G_{r}^* associée au réseau dont la direction des liens est inversée ($cs \to c's'$ devient $c's' \to cs$). Nous nous intéressons uniquement aux dépendances entre $N_{\mathrm{r}} = 21$ secteurs de production, $s \in \{1, \ldots, 21\}$.

Interdépendance des secteurs américains

Nous nous intéressons maintenant aux dépendances entre secteurs de productions américains pour l'année 2008. La Figure 4.10 et Figure 4.11 donnent respectivement les matrices de Google réduites G_{r} et G_{r}^* ainsi que leurs composantes G_{pr}, G_{rr}, G_{qr}, G_{pr}^*, G_{rr}^* et G_{qr}^*. Tout comme pour le commerce international, les poids associés aux matrices G_{qr} et G_{qr}^*, matrices représentant les transactions indirectes entre secteurs, sont plus faibles que les poids associés aux matrices G_{rr} et G_{rr}^* relatives aux liens directs, $W_{\mathrm{rr}} > W_{\mathrm{qr}}$ et $W_{\mathrm{rr}}^* > W_{\mathrm{qr}}^*$. On note que les transactions indirectes sont 3 à 10 fois plus faibles que les transactions décrites par les matrices G_{rr} et G_{rr}^*. Les éléments les plus importants de la matrice G_{r} représentent les liens [C01T05 AGR (secteur agricole)]→[C15T16 FOD (secteur alimentaire)] et [C10T14 MIN (secteur minier)]→[C23 PET (secteur pétrolier)]. La transaction depuis le secteur agricole vers le secteur alimentaire est plutôt triviale au vue de l'importance de l'alimentation dans un pays. En revanche, le second lien nous montre la place importante du secteur pétrolier dans l'économie américaine. Un autre élément de G_{r} traduit un volume d'export important depuis le secteur de fabrication des métaux basiques [C27 MET] vers le secteur lié à la fabrication de produits métalliques [C28 FBM], ce lien reste cohérent dans la mesure où les USA sont un pays développé. Ces trois importantes transactions sont aussi présentes dans la matrice G_{rr}. Il y a la présence de liens directs dont l'importance est moins forte dans G_{rr} que dans G_{r}. C'est le cas des transactions depuis les secteurs [C21T22 PAP] (produits à base de papier) et [C25 RBP] (produits en caoutchouc et en plastique), vers le secteur alimentaire [C15T16 FOD]. La contribution des liens indirects dans ce réseau mondial d'activités économiques renforce ces liens. Il est cohérent que le secteur alimentaire utilise de tels produits, notamment pour les emballages. La matrice G_{pr} donne des informations sur les secteurs économiques américains centraux, du point de vue du PageRank. Les éléments dominants de cette matrice se trouvent sur les lignes relatives aux secteurs [C15T16 FOD], [C34 MTR] (véhicules motorisés) et [C24 CHM] (produits chimiques). La matrice encodant les liens indirects G_{qr} donne des dépendances *cachées* entre certains secteurs économiques des USA. Les liens indirects forts sont [C15T16 FOD]→[C21T22 PAP], [C23 PET]→[C15T16 FOD] et [C23 PET]→[C21T22 PAP]. Naturellement, l'industrie du papier et l'industrie alimentaire utilisent des produits à base de pétrole, que ce soit pour les procédés de fabrication ou bien l'emballage par exemple. Le lien indirect partant du secteur alimentaire et allant vers l'industrie du papier peut s'expliquer par l'utilisation de sylvicultures dans la fabrication de papiers ou encore par le recyclage de matières organiques.

La matrice de Google réduite G_{r}^*, associée au réseau dont les liens sont inversés, et ses composantes sont présentées à la Figure 4.11. Du fait de l'inversion des liens, l'élément situé à la ligne j et à la colonne i de ces matrices traduit le lien $j \to i$. D'après la matrice G_{r}^*, les poids les plus importants sont attribués aux liens [C25 RBP]→[C24 CHM], [C17T19 TEX (produits textiles)]→[C24 CHM] et [C23 PET]→[C10T14 MIN]. La matrice G_{rr}^* montre aussi l'importance de ces liens. Les matrices G_{pr}^* et G_{qr}^* soulignent, encore une fois, l'importance du secteur pétrolier et le fait que ce secteur soit un fournisseur important pour la majorité des secteurs de production américains.

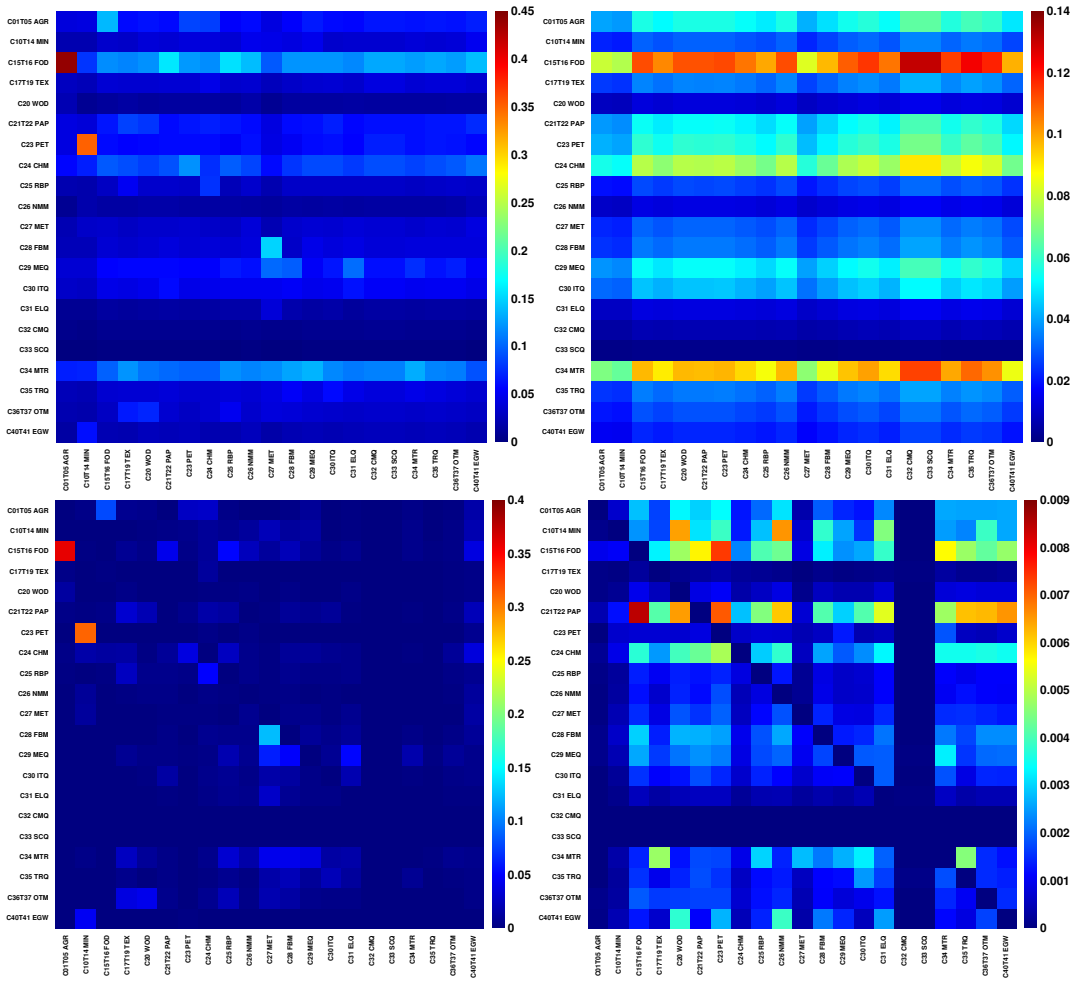

FIGURE 4.10 : Matrice de Google réduite G_{r} et ses 3 composantes pour les $N_{\mathrm{r}} = 21$ secteurs économiques de production américains et l'année 2008. Les poids des matrices G_{r} (en haut à gauche), G_{pr} (en haut à droite), G_{rr} (en bas à gauche) et G_{qrnd} (en bas à droite) sont $W_{\mathrm{pr}} = 0.813817$, $W_{\mathrm{rr}} = 0.155258$, $W_{\mathrm{qr}} = 0.030925$ et $W_{\mathrm{qrnd}} = 0.027383$. D'après [80].

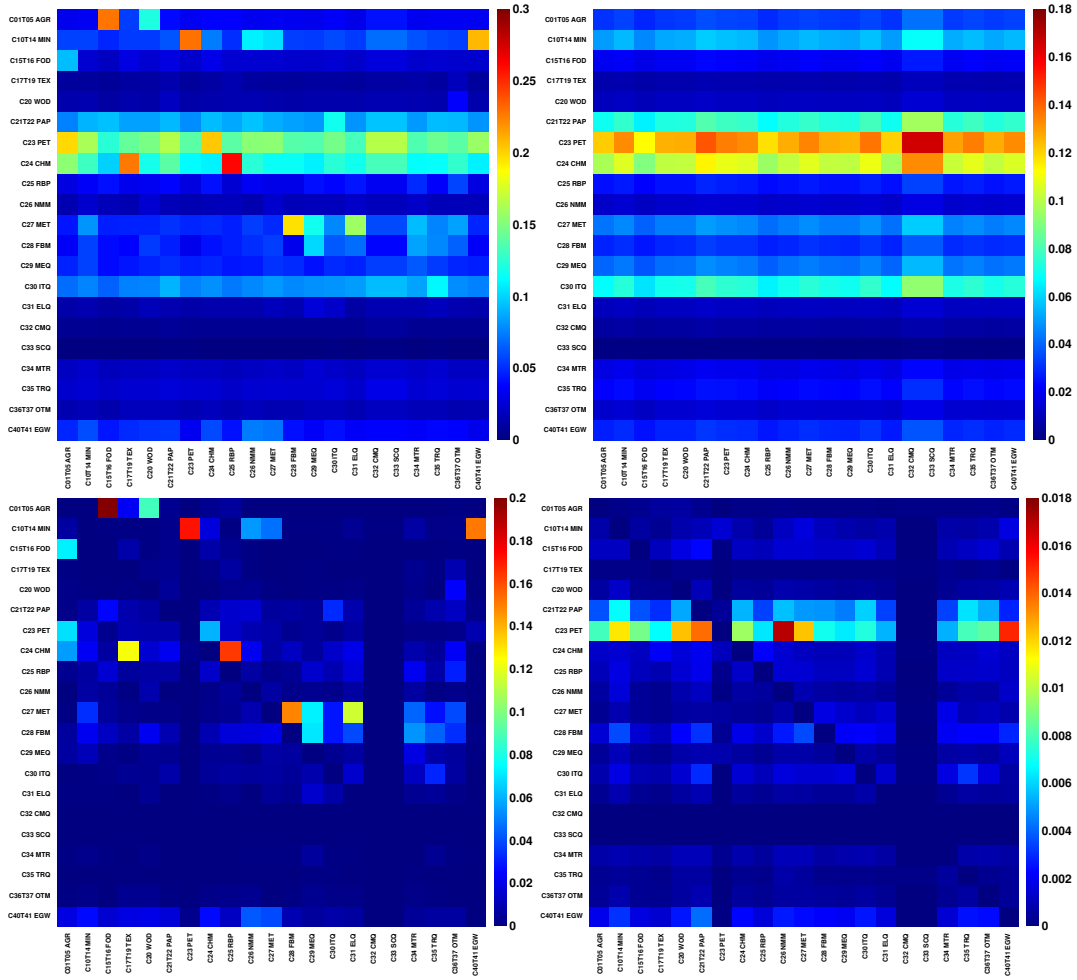

FIGURE 4.11 : Matrice de Google réduite G_{r}^{*} et ses 3 composantes pour les $N_{\mathrm{r}} = 21$ secteurs économiques de production américains et l'année 2008. Les poids associés aux matrices G_{r}^{*} (en haut à gauche), G_{pr}^{*} (en haut à droite), G_{rr}^{*} (en bas à gauche) et G_{qrnd}^{*} sont $W_{\mathrm{pr}}^{*} = 0.813817$, $W_{\mathrm{rr}}^{*} = 0.155258$, $W_{\mathrm{qr}}^{*} = 0.030925$ et $W_{\mathrm{qrnd}}^{*} = 0.027383$. D'après [80].

Interdépendance des secteurs chinois

L'interdépendance des 21 secteurs de production pour la Chine (CHN), en 2008, est présentée à la Figure 4.12 et à la Figure 4.13. La matrice de Google réduite G_{r} et ses composantes sont présentées à la Figure 4.12. De même que pour l'interdépendance des secteurs américains, les liens directs jouent un rôle plus important que les liens indirects, $W_{\mathrm{rr}} > W_{\mathrm{qr}}$ et $W_{\mathrm{rr}}^{*} > W_{\mathrm{qr}}^{*}$. L'élément le plus fort de la matrice G_{r} souligne les transactions partant du secteur [C01T05 AGR] et allant vers le secteur [C15T16 FOD]. Les autres éléments importants de la matrice G_{r} montrent la force de production de la CHN dans le secteur de l'audiovisuel, de l'informatique et de la communication via le lien [C23 CMQ (équipements de communication)]→[C30 ITQ (machines informatiques)]. Les éléments forts de la matrice G_{pr} associée à l'économie chinoise sont distribués sur beaucoup plus de lignes que dans le cas de l'économie américaine (voir Figure 4.10). L'interdépendance des secteurs de productions chinois est donc plus importante, du moins pour l'importation, que celle des secteurs américains. En plus du secteur [C15T16 FOD], les secteurs [C24 CHM], [C29 MEQ] (machinerie et équipements) et [C32 CMQ] sont des secteurs centraux (au sens du PageRank). Les éléments de la matrice G_{qr} sont faibles devant ceux des autres matrices. L'élément le plus fort de cette matrice décrit le lien [C32 CMQ]→[C30 ITQ].

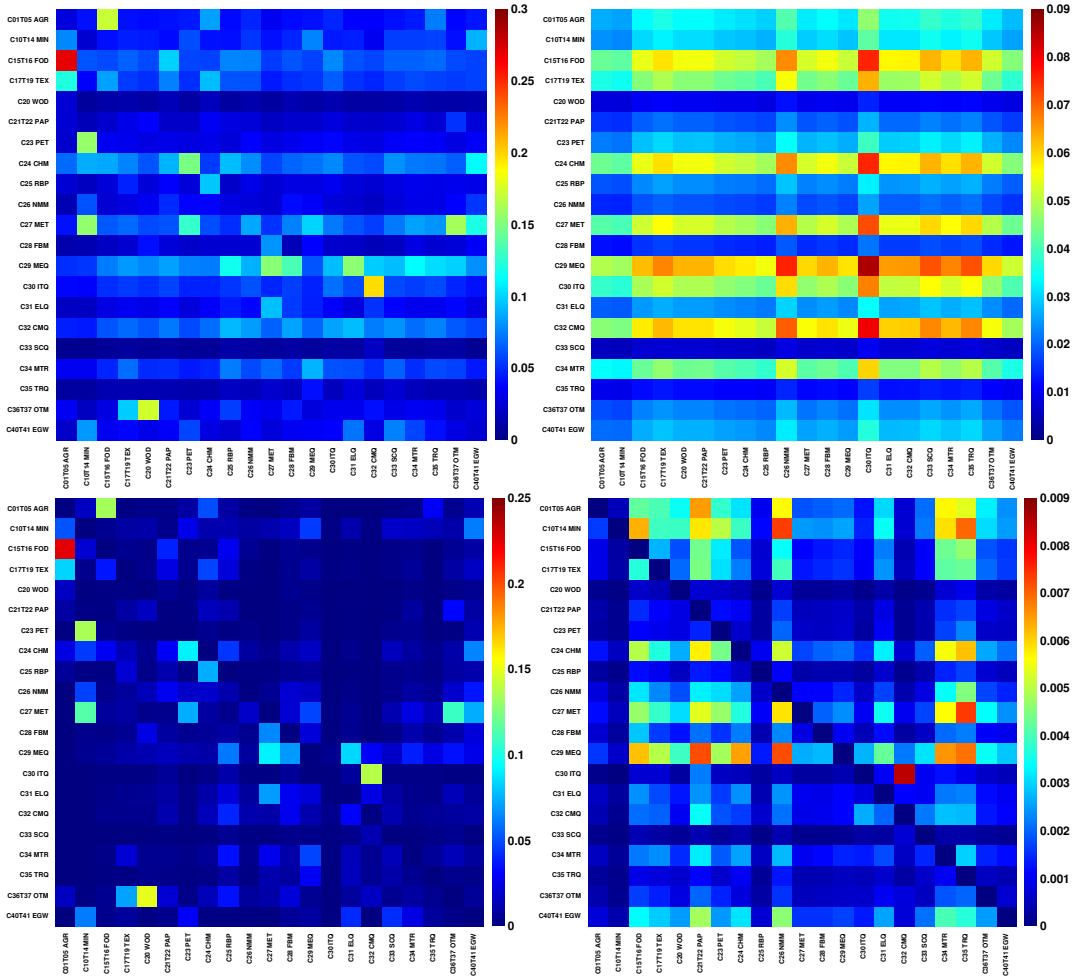

FIGURE 4.12 : Matrice de Google réduite G_{r} et ses 3 composantes pour les $N_{\mathrm{r}} = 21$ secteurs économiques de production chinois et l'année 2008. Les poids des matrices G_{r} (en haut à gauche), G_{pr} (en haut à droite), G_{rr} (en bas à gauche) et G_{qrnd} sont $W_{pr} = 0.698164$, $W_{rr} = 0.263683$, $W_{qr} = 0.038153$ and $W_{qrnd} = 0.035547$. D'après [80].

L'interdépendance des secteurs économiques de la CHN, du point de vue de l'exportation, est donnée par la matrice G_{r}^{*} et ses composantes présentées à la Figure 4.13. Tout comme pour les USA, la matrice de Google réduite G_{r}^{*} associée à CHN nous montre de fortes transactions depuis le secteur [C23 PET] vers le secteur [C10T14 MIN] et depuis [C25 RBP] vers [C24 CHM]. On note aussi que les liens [C15T16 FOD]→[C01T05 AGR], [C25 FBM]→[C27 MET] et [C24 CHM]→[C23 ELQ (machinerie et appareils électriques)] sont significatifs. La transaction indirecte depuis le secteur lié aux machines informatiques vers le secteur des équipements de communication correspond à l'élément de G_{qr}^{*} le plus fort.

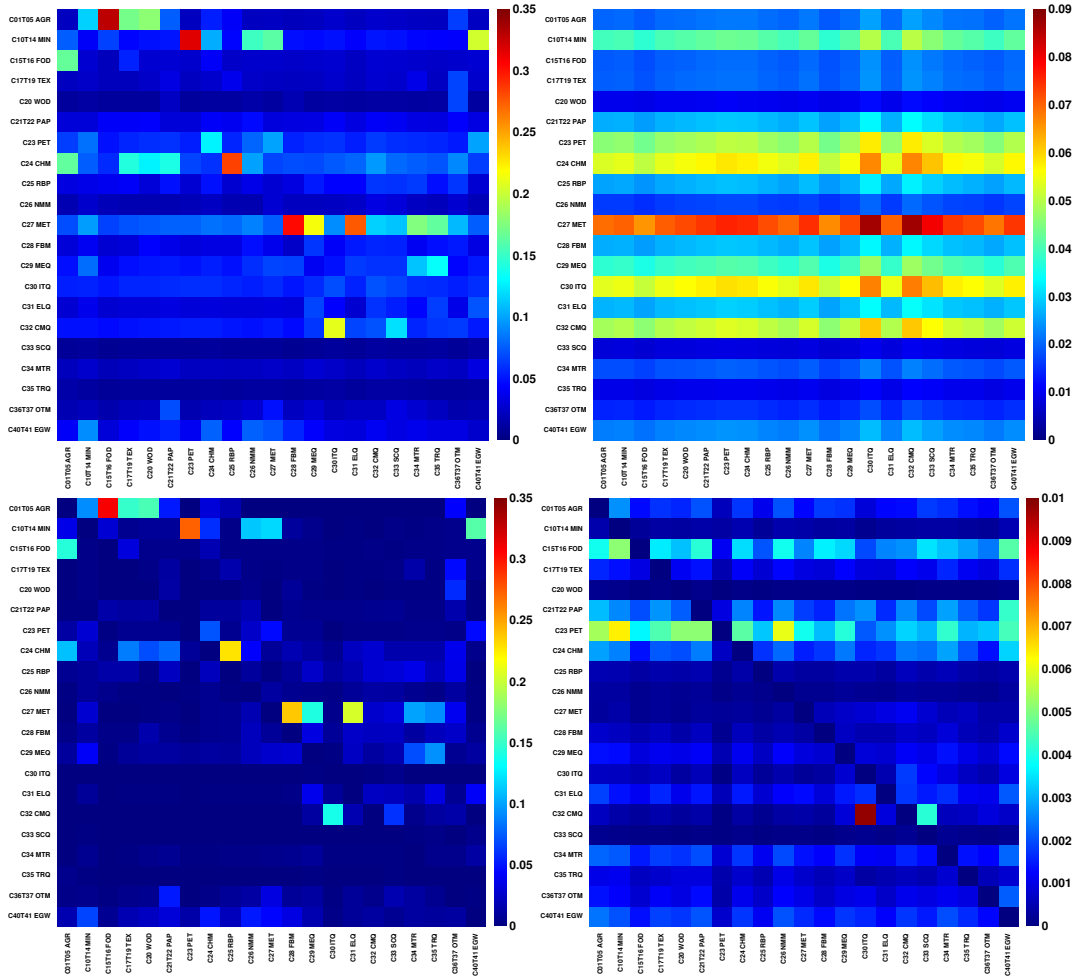

FIGURE 4.13 : Matrice de Google réduite $G^*_{\rm r}$ et ses 3 composantes pour les $N_{\rm r} = 21$ secteurs économiques de production chinois et l'année 2008. Les poids des matrices $G^*_{\rm r}$ (en haut à gauche), $G^*_{\rm pr}$ (en haut à droite), $G^*_{\rm rr}$ (en bas à gauche) et $G^*_{\rm qrnd}$ sont $W^*_{pr} = 0.647087$, $W^*_{rr} = 0.326402$, $W^*_{qr} = 0.026511$ and $W^*_{qrnd} = 0.024648$. D'après [80].

4.2.4 Sensibilités économiques entre secteurs

Nous avons mesuré la sensibilité économique des secteurs de production face à un changement du prix des produits exportés par d'autres secteurs. Nous nous sommes intéressés à 10 pays (USA, Russie (RUS), CHN, Allemagne (DEU), France (FRA), Italie (ITA), Royaume-Uni (GBR), Japon (JAP), Corée du Sud (KOR) et Inde (IND)) et aux années 1995 et 2008. Pour chaque pays, les matrices de Google réduites $G_{\rm r}$ et $G^*_{\rm r}$ correspondantes sont construites. La sensibilité économique $\mathcal{D}_{cs'\to cs}(cs) = dB_{cs}/d\delta$ d'un secteur s face à une perturbation δ des exports du secteur s' est similaire à la sensibilité de la balance économique définit dans (4.7). La balance PageRank-CheiRank pour le secteur s associée au pays c est

$$B_{cs} = \frac{P^*_{cs} - P_{cs}}{P^*_{cs} + P_{cs}} \tag{4.9}$$

où, P^*_{cs} et P_{cs} désignent respectivement la probabilité CheiRank et la probabilité PageRank du couple pays-secteur cs. Pour un pays c, la sensibilité économique de son secteur s par rapport au secteur s' se mesure en modifiant les éléments $G_{{\rm r}\,sc,s'c}$ et $G^*_{{\rm r}\,s'c,sc}$ des matrices de Google réduites associées au pays c, puis en mesurant la variation de la balance économique $(dB_{cs}/d\delta)$ pour une perturbation infinitésimale δ.

La Figure 4.14 montre pour 1995 et 2008, les sensibilités économiques de 20 secteurs de production par rapport au secteur [C23 PET]. Entre 1995 et 2008, la sensibilité la plus négative augmente d'un facteur 3. Il y a alors augmentation de la dépendance en pétrole des secteurs de production. On interprète ce changement par l'augmentation du prix du pétrole, de 3 à 5 fois supérieur entre 1995 et 2008. Certains secteurs ne sont pas impactés par le secteur pétrolier, c'est le cas des secteurs [C31 ELQ], [C33 TRQ] (équipements pour le transport) et [C20 WOD]. Les secteurs [C24 CHM], [C27 MET], [C40T41 EGW] (électricité, gaz et chauffage), [C15T16 FOD] et [C01T05 AGR] sont les plus impactés par un changement du prix des produits en provenance du secteur pétrolier. Le secteur agricole diminue sa sensibilité économique face au secteur pétrolier entre 1995 et 2008. Ces résultats sont cohérents puisque ces secteurs utilisent des produits issus du pétrole. Pour la majorité des pays considérés, le secteur minier est stable. Cependant, le secteur minier russe est impacté et sa sensibilité passe de $\mathcal{D} = -0.0045$ à -0.0012 entre 1995 et 2008, soulignant ainsi la forte dépendance du secteur minier russe au pétrole russe.

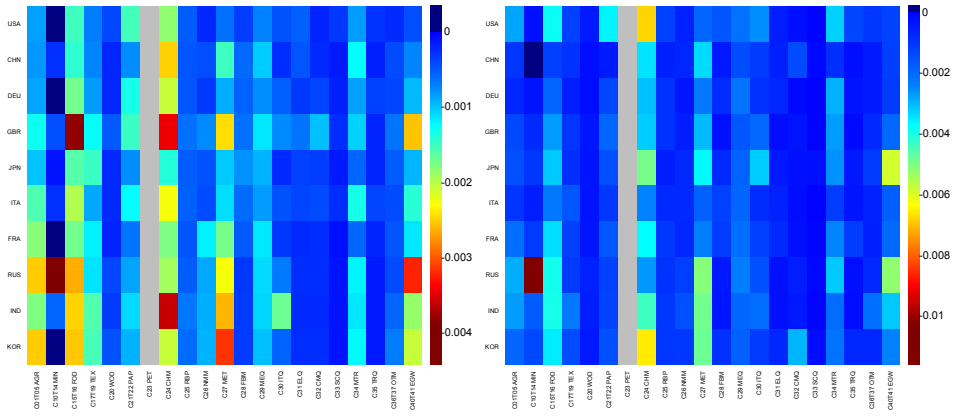

FIGURE 4.14 : Sensibilité des secteurs économiques de production face à la hausse des prix des produits provenant du secteur [C23 PET] pour l'année 1995 (à gauche) et 2008 (à droite). L'axe horizontal donne les secteurs $s \in \{1, \ldots, 21\}$ de la Table 4.3 et l'axe vertical les pays (dans l'ordre PageRank). Le code couleur va du rouge, pour les sensibilités les plus négatives, au bleu, pour les sensibilités les plus positives. La sensibilité du secteur pétrolier sur lui même n'est pas montrée (partie grisée). D'après [80].

La Figure 4.15 est identique à la Figure 4.14 mais permet de mieux observer l'évolution entre 1995 et 2008 des sensibilités économiques. Pour cela nous avons utilisé la même échelle de couleur pour les années 1995 et 2008. On voit que la sensiblité du secteur [C24 CHM] au secteur [C23 PET] augmente d'un facteur 3 pour les USA et la KOR. On note aussi une augmentation, moins importante, de la sensibilité des secteurs [C27 MET] et [C15T16 FOD] au secteur [C23 PET], entre 1995 et 2008. Pour le JAP et la RUS, le secteur [C40T41 EGW] est plus sensible au secteur [C23 PET] en 2008 qu'en 1995. Enfin la majorité des secteurs économiques de production allemands, italiens, et de façon moins notable, des secteurs de production français, britanniques et chinois sont résistants face à une augmentation des prix des produits issus du secteur pétrolier.

La Figure 4.16 montre les effets des secteurs [C24 CHM], [C27 MET], [C34 MTR] et [C10T14 MIN] sur les secteurs de production pour l'année 2008. Le secteur [24 CHM] impacte le plus les autres secteurs de production. Les secteurs [24 CHM] et [C27 MET] impacte sérieusement le secteur [C10T14 MIN], pour la RUS, et le secteur [C34 MTR], pour DEU. L'économie allemande, et plus particulièrement les secteurs [C29 MEQ], [C15T16 FOD] et [C24 CHM], sont les plus touchés par le secteur [C34 MTR]. Le secteur [C10T14 MIN] impacte l'économie provenant des secteurs [C23 PET] pour les 10 pays considérés, mais plus particulièrement le secteur pétrolier russe et le secteur pétrolier américain ainsi que le secteur minier

russe.

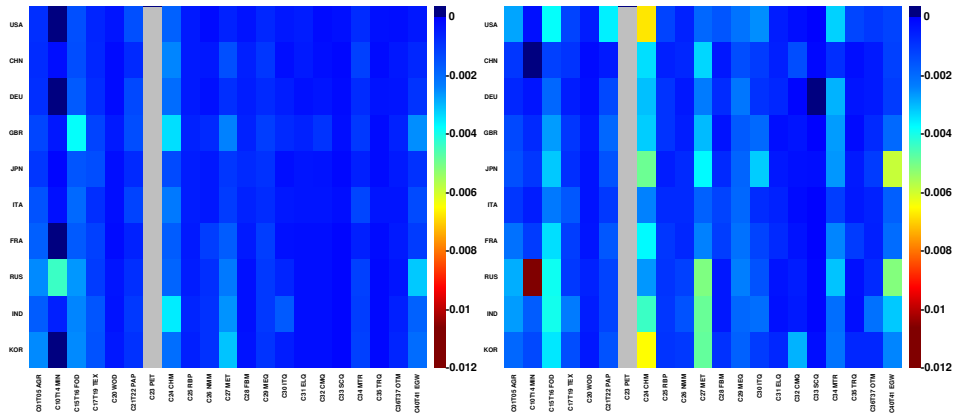

FIGURE 4.15 : Sensibilité des secteurs économiques de production face à la hausse des prix des produits provenant du secteur [C23 PET] pour l'année 1995 (à gauche) et 2008 (à droite). La légende est la même que celle de la Figure 4.14, en revanche le code couleur pour les année 1995 et 2008 sont les mêmes ici. D'après [80].

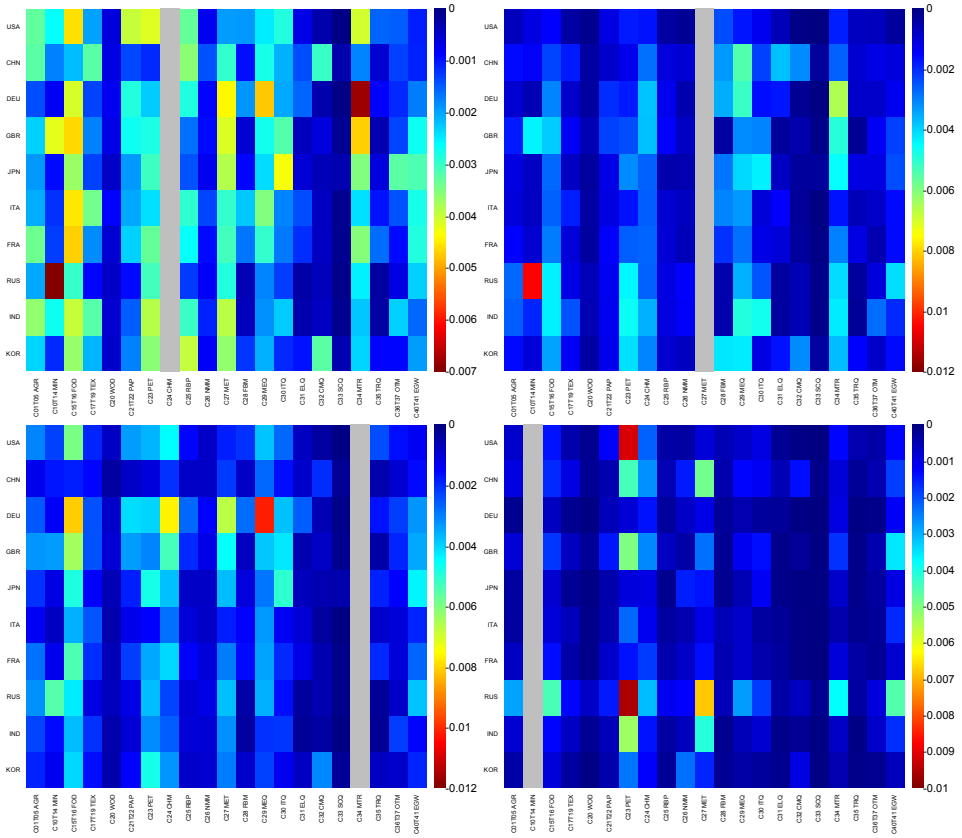

FIGURE 4.16 : Sensibilité des secteurs économiques de production face à la hausse des prix des produits provenant des secteurs [C24 CHM] (en haut à gauche), [C27 MET] (en haut à droite), [C34 MTR] (en bas à gauche) et [C10T14 MIN] (en bas à droite) pour l'année 2008. La légende est la même que celle de la Figure 4.14. Les parties grisées sont pour les sensibilités d'un secteur sur lui-même $\mathcal{D}_{cs\to cs}(s)$, non montrées ici. D'après [80].

4.2.5 Réseaux réduits des secteurs économiques

Nous avons vu plus haut, l'utilisation de la méthode de la matrice de Google réduite pour l'étude du réseau mondial des activités économiques. Nous avons déterminé la signature économique de différents pays par l'extraction des dépendances directes et indirectes entre 21 secteurs de production. À l'aide des matrices G_{sum} et G^*_{sum} construites à partir des sommes $G_{\text{rr}} + G_{\text{qrnd}}$ et $G^*_{\text{rr}} + G^*_{\text{qrnd}}$, nous allons construire le réseau réduit d'interactions entre secteurs de production pour les États-Unis (USA), la Russie (RUS) et la Chine (CHN). La Figure 4.17 représente ces réseaux réduits en tenant compte des données OCDE-WTO TiVA de 2009. Pour chaque nœud cs, nous avons tracé les 4 liens sortants les plus forts selon les colonnes des matrices G_{sum} et G^*_{sum} relatives à ces nœuds.

Le réseau réduit des interactions économiques entre les secteurs de production américains construit à partir de G_{sum} est présenté en haut à gauche de la Figure 4.17 et le réseau réduit construit à partir de G^*_{sum} se trouve en haut à droite de la Figure 4.17. Pour le réseau construit à partir de G_{sum}, on note que le secteur [C24 CHM] importe une grande diversité de produits provenant de plusieurs secteurs de production. En effet, ce secteur a le degré entrant k_{in} le plus grand, il possède des liens provenant de 13 secteurs différents sur les 20 possibles. Les 4 autres secteurs avec les plus hauts degrés entrants sont [C15T16 FOD] ($k_{\text{in}} = 10$), [C21T22 PAP] ($k_{\text{in}} = 10$), [C29 MEQ] ($k_{\text{in}} = 9$) et [C01T05 AGR] ($k_{\text{in}} = 7$). À la vue du réseau réduit concernant l'export, les secteurs fournisseurs les plus centraux, en terme de degré entrant,[6] sont les secteurs [C26 CHM] ($k_{\text{in}} = 13$), [C28 FBM] ($k_{\text{in}} = 13$), [C40T41 EGW] ($k_{\text{in}} = 10$) et [C27 MET] ($k_{\text{in}} = 9$). Le secteur de produits chimiques a une place importante dans l'économie américaine, autant il est un hub pour l'importation, autant il est aussi un fournisseur important. On remarque que certains secteurs sont uniquement liés aux autres via des liens indirects[7] (liens colorés en bleu), c'est le cas des secteurs relatifs aux équipements avec [C32 CMQ] et [C33 SCQ] (équipements médicaux, instruments de précision et d'optique et horlogerie).

Au centre de la Figure 4.17 sont présentés les réseaux réduits des interactions entre secteurs économiques pour la RUS, le réseau construit à partir de G_{sum} est à gauche et celui construit à partir de G^*_{sum} est à droite. Les secteurs importateurs majeurs sont [C10T14 MIN] avec 18 liens entrants sur 20 possibles, [C40T41 EGW] avec 12 liens entrants et [C15T16 FOD] avec 11 liens entrants. Contrairement à l'exemple des USA, le secteur importateur russe le plus central possède beaucoup plus de liens entrants. Le réseau de droite montre que les secteurs [C40T41 EGW] ($k_{\text{in}} = 20$), [C27 MET] ($k_{\text{in}} = 13$), [C23 PET] ($k_{\text{in}} = 13$), [C23 CHM] ($k_{\text{in}} = 12$) et [C01T05 AGR] ($k_{\text{in}} = 9$) sont des fournisseurs importants. Le secteur de l'énergie russe est le seul qui a 100% de ses exports dirigés vers tous les autres secteurs de production, il est un maillon important de la chaîne économique russe. À l'instar des USA, les réseaux réduits d'import et d'export entre secteurs de production russe montrent la présence de transactions indirectes impliquant les secteurs [C32 CMQ] et [C33 SCQ]. De plus, l'exemple de la Russie montre la présence de nouveaux liens indirects dédiés aux importations et aux exportations des secteurs [C31 ELQ], [C28 FBM] et [C35 TRQ].

Les réseaux réduits d'interactions entre secteurs de production chinois sont présentés en bas de la Figure 4.17. Le secteur le plus central du réseau construit à partir de G_{sum} (en bas à gauche de la Figure 4.17) est [C29 MEQ] avec un degré entrant $k_{\text{in}} = 11$, ensuite nous avons les secteurs [C24 CHM] avec $k_{\text{in}} = 10$, [C27 MET] avec $k_{\text{in}} = 10$ et [C10T14 MIN] avec $k_{\text{in}} = 8$. Dans le réseau de droite, les secteurs fournisseurs les plus importants (en terme de degré entrant) sont les secteurs [C24 CHM] et [C40T41 EGW] avec respectivement $k_{\text{in}} = 12$ et $k_{\text{in}} = 10$. Tout comme pour les USA, la CHN focalise son économie de production sur le secteur lié aux produits chimiques, ce secteur est un hub important autant pour l'importation que l'exportation. On note aussi l'inexistence de liens indirects dans l'exemple de l'économie

[6] On rappelle que pour le réseau réduit construit à partir des composantes de G^*_{r}, un lien $j \to i$ représente un export du secteur i vers le secteur j.

[7] Il n'y a aucune transactions directes traduisant ces liens dans la banque de donnée.

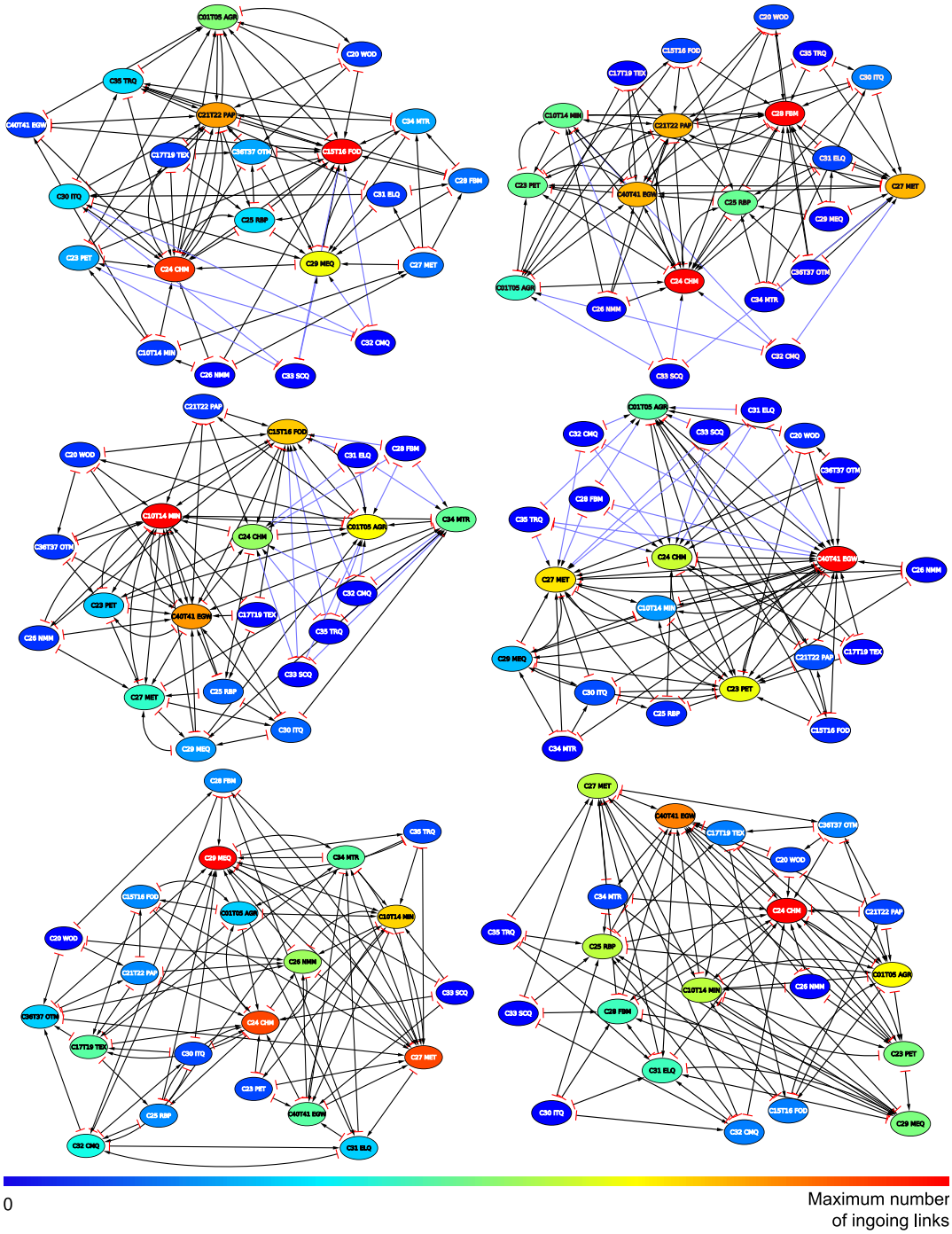

FIGURE 4.17 : Réseaux réduits des interactions entre secteurs économiques de production pour l'année 2009. Les réseaux relatifs aux secteurs de production américains sont présentés en haut, aux secteurs russes sont présentés au milieu et aux secteur chinois sont présentés en bas. Les réseaux construits à partir des matrices G_{sum} correspondantes (relatives aux importations) sont sur la colonne de gauche et ceux construits à partir des matrices G^*_{sum} (relatives aux exportations) sont sur la colonne de droite. Pour chaque pays, on place les nœuds relatifs aux secteurs de production (listés dans la Table 4.3) et on trace leurs 4 liens sortants les plus importants, selon les matrices G_{sum} et G^*_{sum}. Les liens purement indirects qui ne sont pas présents dans la banque de données sont tracés en bleu. Le code couleur des nœuds dépend de leur degré entrant, il va du bleu, pour les degrés les plus bas, au rouge, pour les degrés les plus hauts. D'après [80].

de la CHN montrant alors la forte dépendance entre les secteurs de production chinois.

4.2.6 Conclusion

Nous avons montré l'efficacité de la méthode de la matrice de Google réduite, à extraire des informations à partir de données sur les échanges économiques entre 37 secteurs d'activités. Nous nous sommes intéressés à l'économie de trois puissances économiques, les États-Unis (USA), la Russie (RUS) et la Chine (CHN), ainsi qu'à la mesure des sensibilités économiques de 21 secteurs de production pour 10 pays : Les USA, la RUS, la CHN, l'Allemagne (DEU), la France (FRA), l'Italie (ITA), le Royaume-Uni (GBR), le Japon (JAP), la Corée du Sud (KOR) et l'Inde (IND)). Les matrices de Google réduites G_{r} et G_{r}^{*} et leurs composantes associées aux liens directs et indirects permettent de pouvoir observer l'interdépendance des secteurs de production, que ce soit du point de vue de l'importation ou de l'exportation. Dans l'économie chinoise et américaine, les secteurs agricole [C01T05 AGR] et alimentaire [C15T16 FOD] sont fortement dépendants l'un de l'autre. Ce résultat est assez naturel puisque le secteur de l'agriculture est de première importance pour l'alimentation de la population. Ces deux puissances économiques partagent aussi une dépendance forte entre le secteur minier [C10T14 MIN] et le secteur relatif aux produits issus du pétrole [C23 PET]. On note aussi qu'en fonction du pays, les dépendances entre secteurs d'activités économiques varient. Ainsi, l'économie américaine montre une dépendance entre le secteur de métaux basiques [C27 MET] et le secteur de produits métalliques [C23 FBM], entre le secteur de produits plastiques et en caoutchouc [C25 RBP], le secteur du textile [C17T19 TEX] et le secteur relatif aux produits chimiques [C24 CHM]. L'économie chinoise se caractérise plutôt par une dépendance forte entre les secteurs d'équipements de communication [C32 CMQ] et le secteur informatique [C30 ITQ] et entre le secteur des machines et appareils électriques [C23 ELQ] et [C27 MET]. L'étude de la sensibilité économique de 21 secteurs de production, pour les années 1995 et 2008, face à une perturbation des transactions en provenance d'autres secteurs nous montre que pour la majorité des pays, le secteur [C23 PET] impacte fortement l'économie des secteurs [C24 CHM], [C27 MET], [C40T41 EGW] et [C15T16 FOD]. En revanche, on observe que les secteurs [C23 ELQ], [C32 CMQ], les secteurs d'équipements de transport [C35 TRQ] et le secteur de production de produits à base de bois [C20 WOD] restent insensibles au secteur pétrolier. Nous avons aussi étudié les sensibilités économiques dues aux secteurs [C24 CHM], [C27 MET], [C34 MTR] et [C10T14 MIN] pour l'année 2008. Le secteur de production de produits chimiques est celui qui touche le plus de secteurs différents pour la majorité des pays étudiés. On note des différences pour les sensibilités économiques des secteurs de production selon le pays pris en compte. En effet, pour la Russie, les secteurs [24 CHM] et [C27 MET] agissent principalement sur le secteur minier, pour l'Allemagne le secteur [C34 MTR] sera le plus touché par ces deux secteurs. Une hausse du prix des produits en provenance du secteur [C34 MTR] impacte préférentiellement l'économie allemande et plus particulièrement ses secteurs [C29 MEQ], [C15T16 FOD] et [C24 CHM]. Enfin, le secteur minier touche particulièrement l'économie russe, principalement les secteurs [C23 PET] et [C27 MET], et le secteur pétrolier américain. Enfin nous avons utilisé les matrices G_{sum} et G_{sum}^{*} associées aux USA, à la RUS et à la CHN dans le but de résumer les interactions économiques importantes entre leurs secteurs de production. Nous avons montré que l'économie russe se caractérise par un hub économique (import et export) du secteur [C40T41 EGW] tandis que les économies américaine et chinoise placent le secteur [C24 CHM] au centre des échanges économiques. Enfin, les méthodes d'analyse matricielle utilisées dans l'étude des risques financiers, dont l'efficacité est montrée dans les travaux [81, 82, 83], motivent nos perspectives de recherche. Ainsi, nous aimerions appliquer la méthode de la matrice de Google réduite à l'étude des transactions entre banques et plus généralement au domaine de l'éconophysique.

Chapitre 5

Propagation de crises économiques

Nous venons de voir l'utilisation de la méthode de la matrice de Google réduite appliquée au réseau du commerce international afin de mesurer l'impact économique des pays de l'Union européenne face à la variation du prix de ressources énergétiques en provenance de l'extérieur. Nous avons aussi utilisé les données de la base OECD-WTO TiVA et la méthode de la matrice de Google réduite dans le but d'extraire les interdépendances de secteurs de production pour différents pays et aussi afin de caractériser les échanges économiques importants entre ces secteurs. Dans ce chapitre, nous allons nous intéresser à la modélisation de crise économique et à l'étude de sa propagation dans deux réseaux : le réseau du commerce international et le réseau des transactions de Bitcoin. La propagation de crises économiques dans un réseau peut-être vue comme un problème de percolation qui à chaque étape altère le réseau. Des modifications aléatoires, ou non aléatoires, de la topologie d'un réseau modifient ses propriétés. La question de la fragilité d'un réseau face à des altérations de sa topologie est importante, surtout dans notre société où nous sommes des acteurs de différents réseaux. Le World Wide Web (WWW) et internet, utilisés pratiquement partout et par tous, sont-ils robustes face à des attaques informatiques ou bien face à des routeurs dysfonctionnants ? Un autre exemple, bien plus récent, est celui de la pandémie actuelle et de la manière dont nous propageons celle-ci. Aussi, ce chapitre est divisé en deux partie. La première partie fera l'introduction de la théorie de percolation et de ses applications en théorie des réseaux complexes (section 5.1), elle comprendra aussi une introduction à l'épidémiologie sur réseau, autrement dit l'étude de propagation d'épidémie dans un réseau (section 5.2) et enfin nous présenterons un modèle de contagion de crise économique basé sur la balance PageRank-CheiRank (section 5.3). La seconde partie de ce chapitre présentera deux applications de notre modèle de contagion de crise économique, au réseau du commerce international (section 5.4), et au réseau de transaction de Bitcoin (section 5.5).

5.1 Théorie de la percolation

La théorie de la percolation est l'étude de la capacité d'un système complexe à transmettre une information d'un bout à l'autre de celui-ci. Historiquement, cette théorie fut pensée afin de comprendre comment l'eau transite dans un matériau poreux, plongé dans de l'eau, et si elle peut atteindre le centre de ce matériau.

Les prémices de cette théorie ont été introduits par les physiciens, Flory (1941) et Stockmayer (1943), pour l'étude d'un phénomène biochimique, la polymérisation [84, 85]. La théorie de la percolation a été formalisée par les mathématiciens Broadbent et Hammersley (1957). Ces derniers ont appliqué cette théorie à l'étude du phénomène de diffusion à travers un milieu aléatoire [86] . Ils se sont intéressés à des particules diffusant via des chemins imposés par un milieu plutôt que par des chemins choisis par les particules.

Dans le cadre de la théorie de la percolation, le milieu de diffusion est considéré comme un

réseau aléatoire. Le but est de voir si le système est macroscopiquement connecté ou non. Deux types de percolations existent : la percolation de liens, où l'on retire aléatoirement des liens, et la percolation de site, où les nœuds sont retirés aléatoirement. On définit p la probabilité qu'un nœud, ou un lien, soit gardé, et la probabilité $(1-p)$ qu'un nœud, ou un lien, soit retiré. En faisant varier p, le réseau subit une transition de phase passant d'un réseau constitué de clusters isolés ($p \approx 0$) à un réseau macroscopiquement connecté ($p \to 1$). On définit le point critique p_c traduisant la transition entre ces deux phases.

Super-résilience des réseaux invariants d'échelle

Les attaques, aléatoires ou ciblées, des nœuds d'un réseau sont deux exemples simples de l'application de la théorie de la percolation. Dans le cas d'une attaque aléatoire sur un réseau, nous supprimons aléatoirement, avec une probabilité p, les nœuds d'un réseau tandis que lors d'une attaque ciblée, tous les nœuds ayant une forte connectivité sont supprimés. On dit d'un réseau qui montre une résistance à des attaques qu'il est résiliant.

L'étude [9], menée par Albert, Jeong et Barabasi en 2000, montre différentes résiliences entre les réseaux invariants d'échelle et les réseaux dont la distribution des degrés suit une loi exponentielle. Trois caractéristiques topologiques sont mises en avant dans cette étude : le plus court chemin moyen $\bar{l}$, associé à la composante connectée du réseau, la taille relative W de la composante connectée, et enfin, la taille moyenne $\bar{w}$ des autres clusters du réseau. Les auteurs de [9] montrent pour les deux types de réseau que, pour une probabilité critique p_c et une fraction de nœuds supprimés f_c, une transition de phase apparaît. Alors qu'un réseau, dont la distribution des degrés est exponentielle, voit ses caractéristiques topologiques impactées de façon similaire lorsque les attaques sont aléatoires ou ciblées, les réseaux invariants d'échelle réagissent différemment suivant le type d'attaque. En effet, les quantités $W(f)$, $\bar{w}(f)$ et $\bar{l}(f)$ n'ont pas le même comportement (voir Figure 5.1). On observe un point critique f_c uniquement dans le cas d'attaques ciblées. Cette *super-résilience* aux attaques aléatoires est une caractéristique des réseaux invariants d'échelle. Leur topologie et plus particulièrement la présence de hubs, semblent jouer un rôle important dans cette *super-résilience*. Il est intéressant de constater que les réseaux invariants d'échelle sont très présents dans la nature et que leur résistance aux attaques aléatoires expliquerait pourquoi on les observe si fréquemment. Allant plus loin dans la compréhension de ce phénomène, des auteurs [10, 87], ont montré que cette caractéristique est due à la divergence du second moment de la distribution de degrés d'un réseau $\langle k^2 \rangle$. Le critère de Molloy-Reed (1995) [88], définissant la condition pour qu'une composante connectée géante émerge dans un réseau, est

$$\langle k^2 \rangle - 2\langle k \rangle = \sum_k k(k-2)P(k) > 0 \tag{5.1}$$

où k, $\langle k \rangle$ et $P(k)$ décrivent respectivement le degré, le degré moyen, et la probabilité de trouver un nœud de degré k dans le réseau.

Dans le cas où $\langle k^2 \rangle$ diverge, la condition (5.1) est toujours respectée. Le point critique p_c, pour lequel il y a segmentation du réseau en composantes isolées, est

$$p_c = z_1/z_2 \tag{5.2}$$

où z_1 et z_2 sont le nombre moyen de premier voisins et le nombre moyen de second voisins des nœuds du réseau.

Si $\langle k^2 \rangle$ diverge, alors il y a infiniment plus de second voisins que de premier voisin. On a alors $z_2 \gg z_1$ et donc $p_c \to 0$. L'attaque aléatoire des nœuds d'un réseau invariant d'échelle ne cause pas de segmentation du réseau. Il faudrait supprimer la presque totalité des nœuds du réseau pour aboutir à une segmentation du réseau, en clusters isolés.

Le second moment de la distribution des degrés d'un réseau invariant d'échelle ($P(k) \propto k^\gamma$) diverge pour $\gamma \leq 3$. Il est aussi montré que, pour des réseaux construits de façon déterministe

et dont la distribution des degrés est discrète, si le second moment des degrés diverge alors ils sont super-résilient aux attaques aléatoires [89].

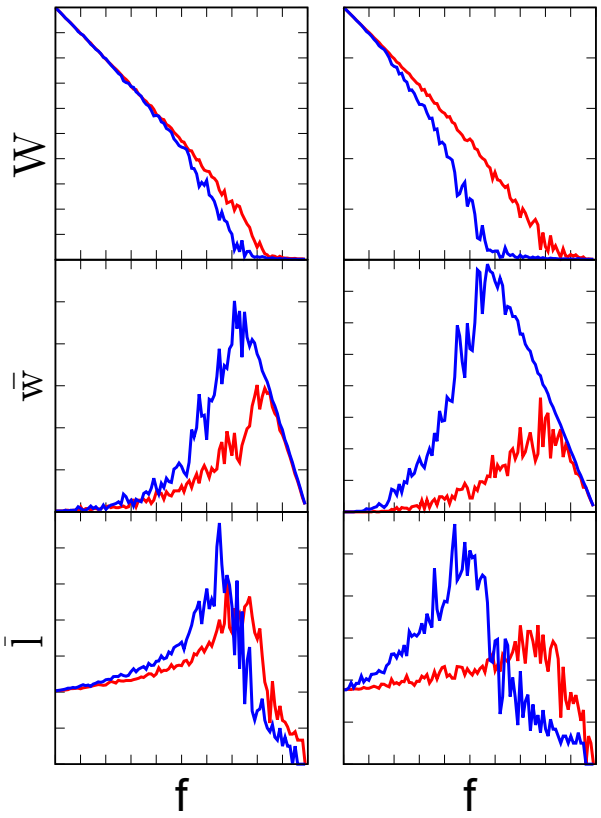

FIGURE 5.1 : Attaques aléatoires et attaques ciblées sur réseau aléatoire et invariant d'échelle. Évolution de la taille relative W de la plus grande composante connectée, de la taille moyenne $\bar{w}$ des autres clusters et du plus court chemin moyen $\bar{l}$ en fonction de la fraction f de nœuds retirés. Les attaques aléatoires sont représentées par des lignes rouges et les attaques ciblées par des lignes bleues. La colonne de gauche (droite) correspond à un réseau aléatoire produit à partir du modèle Erdős–Rényi (à un réseau invariant d'échelle produit avec le modèle Barabási–Albert).

5.2 Épidémiologie sur réseaux

L'épidémiologie est l'étude de la propagation des maladies. Elle cherche à déterminer les causes et les circuits de propagation ainsi que les états critiques et les états stables des épidémies. Deux modèles mathématiques très utilisés dans la littérature sont le modèle *Susceptible Infective Susceptible* (SIS) et le modèle *Susceptible Infective Removed* (SIR). L'utilisation de la théorie des réseaux complexes et de l'épidémiologie permet de prendre en compte les interactions entre individus ainsi que leurs déplacements, en y intégrant le réseau des transports en commun par exemple. L'épidémiologie sur réseau peut être vue comme un problème de percolation, où le matériau poreux serait les interactions sociales, et où le liquide serait la maladie qui se propage.

5.2.1 Le modèle *Susceptible Infective Susceptible* (SIS)

Les nœuds du réseau peuvent être soit dans l'état susceptible soit dans l'état infecté. Un nœud infecté peut alors transmettre son état à ses voisins avec un taux β ou bien devenir susceptible à nouveau avec un taux μ.

On peut considérer le problème en $k_{\max}$ équations différentielles. Ces équations modélisent la variation de populations relatives aux nœuds de degré k. Soit $i_k(t) = I_k(t)/N_k$ la proportion

de nœuds de degré k infectés au temps t. Ainsi $(1 - i_k(t))$ est la proportion de nœuds sains de degré k. La population i_k augmente avec un taux β en fonction des voisins infectés, autour de chaque nœud sain, et diminue avec un taux μ lorsque des nœuds infectés repassent à l'état susceptible. On a

$$di_k/dt = \beta k(1 - i_k)\theta_k - \mu i_k \tag{5.3}$$

où θ_k est la probabilité de trouver un nœud infecté dans le voisinage d'un nœud sain de degré k. Soit i, le nombre total de nœuds infectés tel que

$$i = \int_0^{k_{\max}} i_k p_k dk \tag{5.4}$$

où p_k est la probabilité de trouver un nœud de degré k dans le réseau.

On peut voir une différence entre réseaux aléatoires et réseaux invariants d'échelle. Soit deux mesurables $\lambda = \beta/\mu$, le taux de propagation de l'épidémie qui est une observable biologique, relative à un virus, et $\mathcal{T} = \frac{\langle k \rangle}{\beta\langle k^2 \rangle - \mu\langle k \rangle}$, le temps de propagation caractéristique d'un virus dans un réseau. Il y a propagation du virus si $\mathcal{T} > 0$.

Réseaux aléatoires

Dans le cas d'un réseau aléatoire nous avons $\langle k^2 \rangle = \langle k \rangle \langle k + 1 \rangle$, ainsi

$$\mathcal{T}^{\text{SIS}}_{\text{Aléat.}} = (\beta\langle k + 1 \rangle - \mu)^{-1} > 0 \tag{5.5}$$

soit,

$$\lambda > \frac{1}{\langle k + 1 \rangle}. \tag{5.6}$$

Pour un virus ayant un taux de propagation $\lambda_c = 1/\langle k + 1 \rangle$, il y a transition de phase vers un état d'épidémie endémique, c'est à dire qu'une fraction fini de la population sera impactée pour $\lambda > \lambda_c$.

Réseaux invariants d'échelle

Nous avons

$$\mathcal{T}^{\text{SIS}}_{\text{Inv. Éch.}} = \frac{\langle k \rangle}{\beta\langle k^2 \rangle - \mu\langle k \rangle} > 0 \tag{5.7}$$

soit,

$$\lambda > \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{5.8}$$

Comme $\langle k^2 \rangle \to \infty$, $\lambda_c = 0$, une épidémie peut survenir instantanément quelque soit le taux de propagation du virus.

Il est intéressant de constater que les réseaux invariants d'échelle, très présents dans la nature, soient super-résilients aux attaques aléatoires mais très fragiles aux épidémies. Le modèle SIS peut être utilisé dans d'autres domaines tels que la propagation d'informations, la propagation d'opinions [90, 91] ou bien encore l'étude du trafic de drogue [92].

5.3 Modèle de propagation de crises économiques

Une crise économique est un évènement pouvant toucher un ou plusieurs acteurs économiques. Elle peut survenir à plusieurs échelles, comme une ville, un pays ou bien au niveau mondial. Les crises bancaires, les crises monétaires ou encore les crises boursières sont appelées crises financières ou économiques.

Un exemple récent d'une crise financière mondiale est la crise des *subprimes* qui s'est déroulée pendant la période 2007-2008. Les causes de cette crise sont multiples. La conjonction de la hausse des taux directeurs de la réserve fédérale américaine et de la baisse du prix de l'immobilier a poussée les agences de prêts américaines dans une situation de crise économique puisqu'elles avaient des difficultés à se faire rembourser des clients non solvables. La *titrisation*, se définissant par le rachat de créances en tant que titres financiers, fut la cause majeure de la propagation de cette crise. Au vue de la perte économique qu'une telle crise financière peut engendrer, il est vital de pouvoir en comprendre les mécanismes afin de l'anticiper.

Tout comme pour l'étude des épidémies, l'étude de la propagation des crises économiques peut être couplée à la théorie des réseaux complexes [93, 94, 95]. Les modèles utilisés sont majoritairement issus des modèles épidémiologiques. Nous allons voir comment la matrice de Google et les algorithmes du PageRank, et du CheiRank, permettent l'élaboration d'un modèle de propagation de crises économiques.

5.3.1 Balance PageRank-CheiRank

Considérons un réseau de transactions économiques constitué de N nœuds représentant N acteurs économiques et de N_{t} liens dirigés et pondérés représentant les transactions entre les acteurs économiques. Nous définissons la matrice monétaire M telle que l'entrée M_{ij} représente la valeur monétaire totale des biens échangés de l'acteur j vers l'acteur i.

Comme nous l'avons vu dans le Chapitre 2, le vecteur PageRank $\mathbf{P}$ mesure l'importance des nœuds du réseau et le vecteur CheiRank $\mathbf{P}^*$ mesure la communicativité des nœuds. Dans le cas d'un réseau de transactions économiques, un lien entrant représente une acquisition de bien tandis qu'un lien sortant est relatif à une vente. On interprète alors le PageRank comme une mesure de la capacité à importer et le CheiRank la capacité à exporter.

En économie, la balance économique est une mesure simple de santé économique qui se définit symboliquement comme

$$B = [\textit{Volume des exportations}] - [\textit{Volume des importations}]. \tag{5.9}$$

On peut définir une version basée sur les vecteurs PageRank $\mathbf{P}$ et CheiRank $\mathbf{P}^*$, nous avons alors la balance PageRank-CheiRank

$$B_a = \frac{P_a^* - P_a}{P_a^* + P_a} \tag{5.10}$$

où a est l'indice relatif à l'acteur économique a. Par définition cette balance est bornée $B_a \in [-1, 1]$. Une valeur de B_a négative (positive) caractérisera un acteur principalement importateur (exportateur).

5.3.2 Contagion de crise économique

Nous définissons un seuil de crise économique κ de sorte que, pour un acteur économique a, si $B_a < -\kappa$, il est considéré en crise économique. Afin de stabiliser son déficit économique, cet acteur doit réguler son volume d'importation. De ce fait, il vient que les acteurs économiques qui exportent habituellement des biens à l'acteur a vont être économiquement impactés. La régulation du volume d'importation des acteurs économiques en crise implique de modifier les entrées de la matrice monétaire M relatives à leurs exportations. On note $\tau = 0$ l'étape initiale de la contagion de crise économique. À la fin de l'étape τ du processus, on construit l'ensemble $\mathcal{C}_\tau$ des acteurs qui sont en crise économique. Au début de la prochaine itération du processus, les éléments de la matrice monétaire M sont modifiés telles que

$$M_{ij}(\tau) = 0 \text{ si } i \in \mathcal{C}_{\tau-1} \tag{5.11}$$

avec $\tau > 0$.

Ce modèle de crise économique basé sur la matrice de Google et les vecteurs PageRank et CheiRank permet de mesurer le coup budgétaire d'une crise économique, mais aussi d'étudier la dynamique de la contagion pour différents seuil κ. Nous verrons également, à la suite de cette première partie, que l'on peut construire un réseau de contagion constitué des acteurs économiques responsables de la propagation de la crise économique.

5.4 Contagion de crise économique dans le réseau du commerce international (RCI)

5.4.1 Introduction

L'impact mondial de la crise financière de 2007-2008 a montré que le phénomène de contagion est, aussi, caractéristique d'une crise économique. Dans cet exemple, la crise s'est propagée à travers le réseau mondial des banques [93, 94, 95]. Ce phénomène de propagation de crise est aussi observable dans le commerce international, vulnérable aux crises énergétiques causées, notamment, par le commerce du pétrole et du gaz. Dans l'étude présentée ici, nous modélisons la contagion de crise économique au sein du réseau du commerce international construit à partir des données Comtrade de l'ONU. Le modèle de contagion utilisé est celui décrit, plus haut, à la section 5.3. Cette étude est basée sur l'analyse de la matrice de Google [22, 23, 69] du réseau du commerce international introduite dans les travaux [67, 68]. Contrairement à la méthode standard qu'est l'analyse des volumes d'imports et d'exports, intégrant uniquement les transactions directes entre les pays, la méthode de la matrice de Google prend aussi en compte les interactions indirectes entre ces pays. Ces transactions indirectes sont importantes. Même si deux pays n'ont pas de transactions directes entre eux, leurs économies peuvent être corrélées par l'intermédiaire de partenaires commerciaux communs. On retrouve dans la littérature des études sur les propriétés statistiques du réseau du commerce international, avec notamment, les travaux [73, 74, 75, 76, 77, 78, 96]. Cependant, il y a peu d'études concernant la contagion de crise économique pour ce réseau. Notre modèle de contagion est basé sur deux scénarios de crise. Le premier scénario, le modèle A, interdit à un pays en crise économique d'importer des produits à l'exception du pétrole et du gaz. Le second scénario, modèle B, interdit toutes importations aux pays en crise économique. La suite de cette section est présentée de la manière suivante. Premièrement, les données et la méthode utilisés seront rapidement présentés (section 5.4.2), puis nous discuterons les transitions de phases observées pour la contagion de crise économique dans le cas des modèles A et B (section 5.4.3). Nous nous intéresserons, ensuite, aux rôles joués par les pays dans cette crise, avec une distribution géographique des pays touchés (section 5.4.4). Enfin, nous présenterons les réseaux de contagion de crise économique (section 5.4.5).

5.4.2 Données et méthodes

Dans cette étude, nous avons utilisé les données Comtrade du commerce international entre $N_c = 227$ pays échangeant $N_p = 61$ produits pour les années 2004, 2008, 2012 et 2016. Nous nous sommes intéressés à la modélisation de la crise économique et à sa propagation. La méthode de construction des matrices de Google, G et G^*, associées au réseau du commerce international est présentée à la section 4.1.2. Nous mesurons la santé économique d'un pays c à l'aide de la balance PageRank-CheiRank B_c (BPC) définie par (5.10). Un pays ayant une BPC fortement négative doit diminuer son volume d'importation et le restreindre à ce qui est vital pour son économie. Des restrictions commerciales peuvent être imposées par une organisation supranationale dans le but d'éviter que la crise économique devienne globale. Nous considérons un pays c en banqueroute si $B_c \leq -\kappa$ où $\kappa \geq 0$ est un seuil de banqueroute. L'Algorithme 2

décrit le fonctionnement du modèle de contagion, défini à la section 5.3, appliqué au réseau du commerce international. Soit $\mathcal{C}$, l'ensemble des N_c pays et $\mathcal{P}$, l'ensemble des $N_p = 61$ produits. À l'étape initiale $\tau = 0$, nous utilisons les matrices de Google $G_0 = G$ et $G_0^* = G^*$, associées au réseau du commerce international pour calculer les BPC des pays. Nous construisons l'ensemble des pays $\mathcal{B}_0 = \{c \in \mathcal{C} | B_c \leq -\kappa\}$ en banqueroute. Ces pays restent en banqueroute à toutes les étapes $\tau \geq 1$ de la contagion. Nous considérons deux scénarios : les imports d'un pays en crise sont interdits à l'exception du pétrole et du gaz (modèle A) ou tous les imports sont bannis (modèle B). Lorsque des pays sont en crise économique, l'interdiction des imports entraîne la suppression de liens du réseau du commerce international. Ainsi, la matrice monétaire M est modifiée

Data : matrice monétaire M et seuil de banqueroute κ.
Result : ensemble des pays $\mathcal{B}_\tau$ en banqueroute à l'étape τ de la contagion.
$\tau = 0, \mathcal{B}_{-1} = \emptyset$
while $\tau \neq \tau_\infty$ **do**
  $\mathcal{B}_\tau = \emptyset$
  Construire G, G^*, $\mathbf{P}$ et $\mathbf{P}^*$ avec M
  for $c \in \mathcal{C} - \cup_{i=0}^{\tau-1} \mathcal{B}_i$ **do**
    if $B_c \leq -\kappa$ **then**
      $\mathcal{B}_\tau = \mathcal{B}_\tau + c$
      if *modèle A* **then**
        for $c'p \in \mathcal{C} \times \tilde{\mathcal{P}} | M_{cc'}^p \neq 0$ **do**
          $M_{cc'}^p = 0$
        end
      end
      if *modèle B* **then**
        for $c'p \in \mathcal{C} \times \mathcal{P} | M_{cc'}^p \neq 0$ **do**
          $M_{cc'}^p = 0$
        end
      end
    end
  end
  $\tau = \tau + 1$
  if $\mathcal{B}_\tau = \emptyset$ **then**
    $\tau_\infty = \tau$
  end
end

Algorithme 2 : Modèle de contagion de crise économique dans le réseau du commerce international.

$$M_{cc'}^p = 0, \forall c' \in \mathcal{C}, \forall c \in \mathcal{B}_0, \begin{cases} \forall p \in \tilde{\mathcal{P}} & \text{(modèle A)} \\ \forall p \in \mathcal{P} & \text{(modèle B)} \end{cases} \tag{5.12}$$

où $\tilde{\mathcal{P}} = \mathcal{P} - \{\text{pétrole, gaz}\}$. À la fin de l'étape τ, la matrice de Google $G_{\tau+1}$ est construite à partir de la matrice monétaire M modifiée (5.12). La contagion s'arrête à l'étape τ_∞ lorsqu'il n'y a plus de nouveaux pays en banqueroute. Nous définissons la proportion $\eta(\tau, \kappa)$ de pays en banqueroute à l'étape τ pour un seuil de banqueroute κ, $\eta(\tau, \kappa) \in [0, 1]$. Il n'y a aucun pays en banqueroute pour $\eta = 0$ et tous les pays sont en banqueroute pour $\eta = 1$. On définit le coût $C(\kappa, \tau)$ de la crise à l'étape τ pour un seuil κ tel que

$$C(\kappa, \tau) = \sum_{c \in \cup_{\tau'=0}^{\tau} \mathcal{B}_{\tau'}} \sum_{p \in \tilde{\mathcal{P}} \text{ ou } p \in \tilde{\mathcal{P}}} V_{cp} \tag{5.13}$$

où V_{cp} est le volume total des transactions d'imports du pays c en produit p. Lorsque le processus de contagion se termine (τ_∞), le coût total de la crise économique est $C_\infty(\kappa) = C(\kappa, \tau_\infty)$, il mesure le volume total des transactions avortées en raison des pays en crise.

La Figure 5.2 donne une représentation du réseau du commerce international de l'année 2016 tenant compte des transactions entre pays, tous produits confondus.

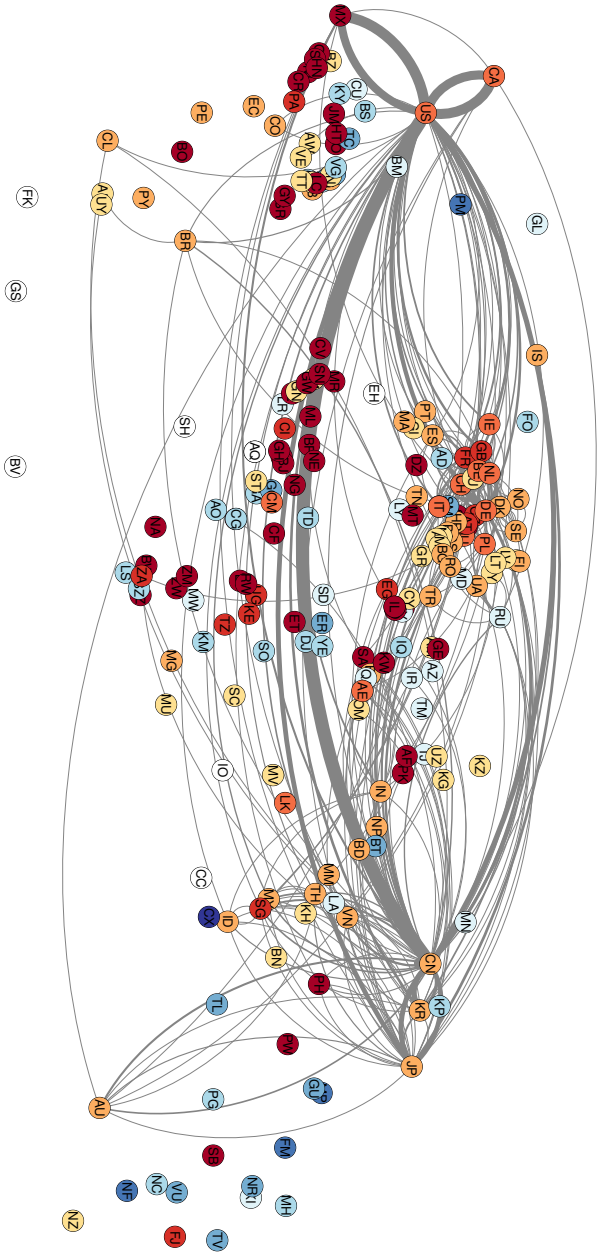

FIGURE 5.2 : Réseau du commerce international de l'année 2016. La direction du lien entre le pays A et le pays B est donnée par sa courbure, pour cela il faut suivre dans le sens horaire la courbure du lien $(A \frown B)$. Ce lien représente le volume total des produits exportés, par le pays A, au pays B. L'épaisseur du lien est proportionnelle à la masse monétaire de la transaction. Seules les transactions supérieures à 10^{10} dollars US sont représentées. Le code couleur des nœuds va du rouge pour les pays en banqueroute à $\tau = 0$ au bleu pour les pays en banqueroute à τ_∞. Les nœuds blancs représentent les pays survivant à la crise. Le modèle de contagion utilisé, ici, est le modèle A avec un seuil de banqueroute $\kappa = 0.1$. D'après [97].

5.4.3 Transition de phase

La Figure 5.3 donne l'évolution de la proportion $\eta(\kappa)$ de pays en banqueroute en fonction du seuil de banqueroute κ et pour différentes étapes τ de la contagion. Les données relatives à l'année 2016 ont été utilisées. Nous avons utilisé deux valeurs du paramètre α pour la construction des matrices de Google associées au réseau du commerce international. Les panneaux du haut donnent les graphiques $\eta(\kappa)$ pour $\alpha = 0.5$ et les panneaux du bas pour $\alpha = 0.85$, valeur standard pour le calcul du PageRank utilisée dans le papier original de Brin and Page [23]. Le modèle A est présenté dans les panneaux de gauche, le modèle B est présenté dans les panneaux de droite. Nous observons pour chacun de ces graphiques une transition de phase à partir d'un seuil critique κ_c. Nous passons d'un régime de contagion locale ($\kappa > \kappa_c$) à un régime de contagion globale ($\kappa < \kappa_c$). La valeur standard $\alpha = 0.85$ (figures du bas) donne une transition beaucoup moins abrupte. La valeur $\alpha = 0.5$ (figures du haut) donne une transition de phase que l'on pourrait qualifier de "tout ou rien". Le seuil de banqueroute critique κ_c est différent selon le modèle de restrictions commerciales, mais aussi selon le paramètre α utilisé. En effet, nous avons pour $\alpha = 0.5$ les points critiques $\kappa_c \approx 0.15$ (modèle A) et $\kappa_c \approx 0.175$ (modèle B). Pour une valeur $\alpha = 0.85$, nous obtenons les points critiques $\kappa_c \approx 0.18$ avec le modèle A et $\kappa_c \approx 0.25$ avec le modèle B. Dans le cas $\alpha = 0.5$, l'écart entre le point critique associé au modèle A et celui du modèle B est moins grand qu'avec $\alpha = 0.85$. Plus α est faible, plus la crise est susceptible de se propager. Cela explique pourquoi la transition est plus abrupte dans le cas $\alpha = 0.5$. Nous pensons que le modèle A, scénario où les pays en crise ne peuvent importer que du pétrole et du gaz, est plus réaliste que le modèle B. En effet, lorsqu'un pays est en banqueroute, il est important pour son économie de continuer à pouvoir produire dans le but d'exporter et de se stabiliser. Le pétrole et le gaz sont importants pour la production. On observe aussi que, le modèle B conduit à une contagion globale plus importante $\eta \approx 1$ que le modèle A. Ce résultat est compréhensible puisque, le modèle A protège les pays exportateurs de pétrole et de gaz. Pour le reste de cette section, nous présenterons les résultats associés au modèle A et avec le paramètre $\alpha = 0.5$.

La Figure 5.4 montre que pour les année 2004, 2008, 2012 et 2016, la transition de phase est toujours présente. Le régime de contagion localisée ($\kappa > \kappa_c$) touche moins de 10% des pays, tandis que le régime de contagion globale affecte $\approx 90\%$ des pays. Les points critiques associés aux années 2004-2016 sont assez proches, $\kappa_c \approx 0.14{-}0.175$. Nous observons que la courbe $\eta(\kappa)$ relative à la contagion de crise économique pour l'année 2016, contrairement aux autres courbes, a un profil non monotone dans l'intervalle $[0, \kappa_c]$. Cette irrégularité provient de l'altération du réseau du commerce international à chaque étape τ du processus de contagion, ainsi que de la protection des principaux exportateurs de pétrole et de gaz, produits toujours importés par les pays en crise. Si un important fournisseur de pétrole et de gaz est en banqueroute pour certains seuils κ, le nombre de pays en crise économique augmente fortement. La Russie est justement un grand fournisseur de pétrole et de gaz, elle est en banqueroute en 2016 pour $\kappa \in [0.04, 0.11]$ puis $\kappa = 0.14$. Les pics et les effondrements de la courbe $\eta(\kappa)$ sont en accords avec les seuils κ pour lesquels, la Russie passe de pays sain à pays en banqueroute. Nous observons aussi que la crise économique de la Russie entraîne la majorité des pays en banqueroute, $\eta \approx 1$, pour $\kappa = 0.04$. Les listes des pays survivants, pays qui ne sont pas banqueroute a τ_∞, pour $\kappa = 0.1$ et le modèle A, sont données aux Tables A.4, A.5, A.6 et A.7 situées en annexes. À l'exception de l'année 2016, les pays survivant à la contagion ont des volumes d'exports en pétrole et en gaz importants. C'est le cas pour le Nigeria (en 2004), l'Arabie Saoudite (en 2004), la Russie (en 2004, 2008 et 2012) ou encore le Timor Oriental (en 2008). De plus, la plupart des pays survivants sont des îles, ces pays peuvent être impliqués dans de petits réseaux insulaires de commerce, ce qui peut expliquer qu'ils soient résistants à la contagion.

Une transition de phase est aussi observable pour l'évolution du temps de contagion τ_∞ en fonction du seuil de banqueroute κ, présentée à la Figure 5.5 (partie de gauche), ainsi que pour le coût total C_∞ de la crise économique en fonction de κ, présenté à la Figure 5.5 (partie de

droite). Le temps de contagion τ_∞ augmente pour $\kappa \in [0, \kappa_c]$ passant de $\tau_\infty = 4$ à $\tau_\infty \approx 12-16$. La contagion est très courte ($\tau_\infty \lesssim 5$) et peut s'arrêter après 1 à 2 étapes lorsqu'on est dans la région $\kappa > \kappa_c$. Il est intéressant de remarquer que quelque soit l'année étudiée et le modèle utilisé, les courbes $\tau_\infty(\kappa/\kappa_c)$ sont très similaires. Le coût total C_∞ de la crise économique est stable pour la région $\kappa < \kappa_c$ puis, diminue drastiquement pour $\kappa < \kappa_c$. Le coût total C_∞ d'une crise économique nous permet de quantifier l'ampleur de cette crise. Le régime de contagion globale provoque une perte s'élevant à environ 80-90% du total des transactions du commerce international, tandis que le régime de contagion locale coûte moins de 5% du total des transactions. Nous pensons qu'un tel graphique peut aider une organisation supranationale à limiter le coût d'une crise économique. Elle pourrait imposer un seuil de tolérance pour les pays en déficit économique, tel que le coût de la crise soit moindre. Pour $\kappa \gtrsim 1.5\kappa_c$, le coût de la crise est inférieur à 1% du total des transactions. On observe des différences selon l'année considérée uniquement après $1.5\kappa_c$. On note aussi que pour l'année 2008, le coût total C_∞ de la crise pour $\kappa \gtrsim 2.5\kappa_c$ est plus grand que pour les autres années.

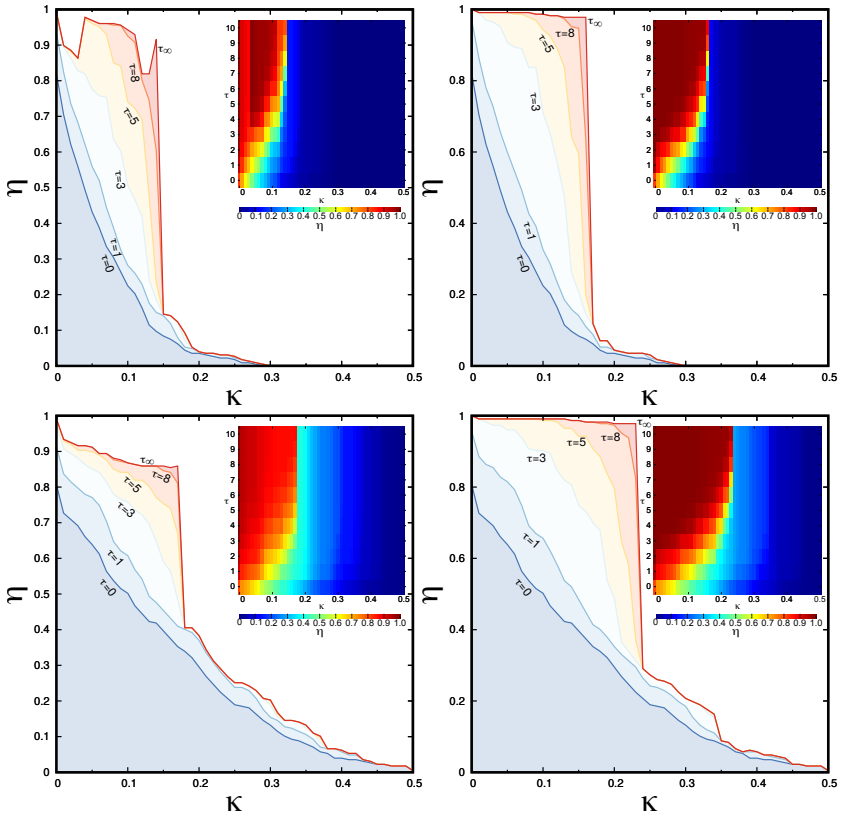

FIGURE 5.3 : Proportion de pays en banqueroute en 2016. Proportion η de pays en banqueroute en fonction du seuil κ pour différentes étapes τ de la contagion de crise économique dans le réseau du commerce international. Le modèle A (modèle B) est présenté sur la colonne de gauche (de droite). Deux valeurs du paramètre α sont utilisés, $\alpha = 0.5$ en haut et $\alpha = 0.85$ en bas. Les encarts représentent les proportions η de pays en banqueroute dans le plan (τ, κ). La couleur rouge correspond à $\eta = 1$ et bleu à $\eta = 0$. D'après [97].

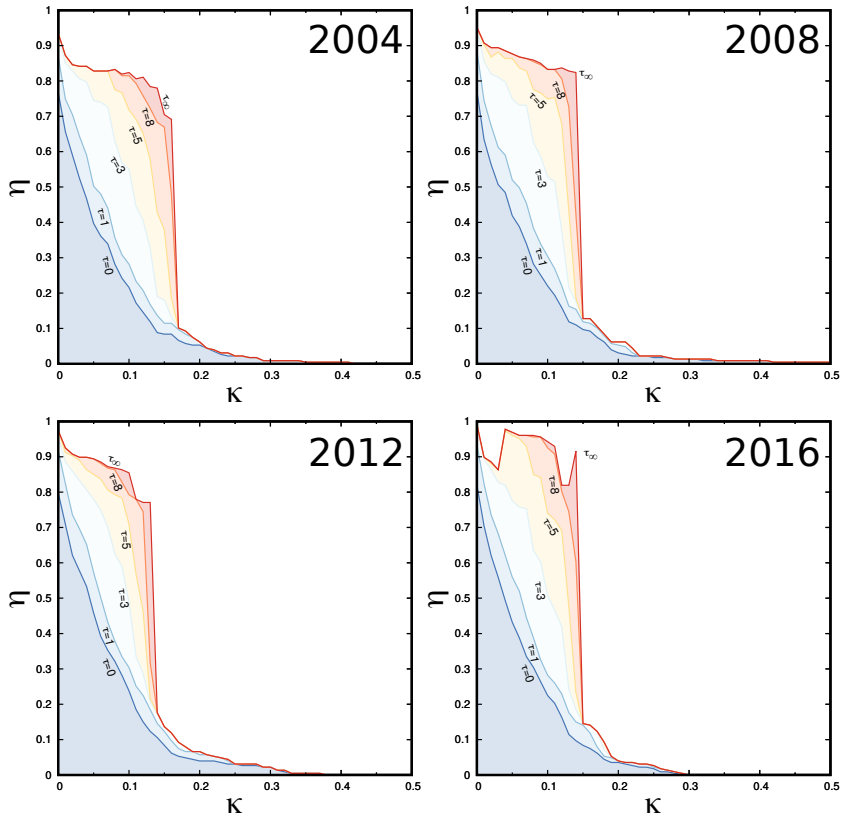

FIGURE 5.4 : Proportion de pays en banqueroute pour les année 2004, 2008, 2012 et 2016. Proportion η de pays en banqueroute en fonction du seuil κ pour différentes étapes τ de la contagion (modèle A et $\alpha = 0.5$). D'après [97].

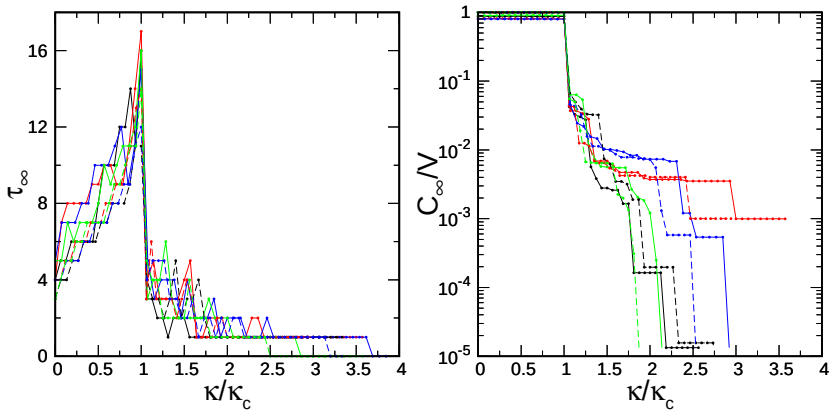

FIGURE 5.5 : Temps de contagion de la crise économique τ_∞ (à gauche) et coût total C_∞ de la crise (à droite) en fonction du seuil de banqueroute κ. Le coût total de la crise économique est calculé en utilisant 5.13. Les données du commerce international relatives aux années 2004 (courbes noires), 2008 (courbes rouges), 2012 (courbes bleues) et 2016 (courbes vertes) ont été utilisées. Les lignes pleines correspondent à la contagion de crise économique selon le modèle A et les lignes en pointillé selon le modèle B. Le volume total V des transactions en dollars US est $V = 9.43 \times 10^{12}$ en 2004, 1.68×10^{13} en 2008, 1.85×10^{13} en 2012 et 1.62×10^{13} en 2016. Le paramètre $\alpha = 0.5$ a été utilisé. D'après [97].

5.4.4 Distribution géographique des balances PageRank-CheiRank

Nous allons, à partir de maintenant, nous intéresser aux rôles joués par les pays dans ces crises économiques. La Figure 5.6 donne la distribution géographique des balances économiques B à l'étape initiale ($\tau = 0$) des contagions de crise économique avec un seuil $\kappa = 0.1$ pour les années 2004, 2008, 2012 et 2016. On appelle *graine de contagion*, les pays $c \in \mathcal{B}_0$ tels que $B_c < -\kappa$. Parmi ces graines de contagion (pays colorés en magenta), nous avons des pays d'Afrique et plus précisément de la région subsaharienne comme le Mali (2004-2016), le Niger (2004-2016), le Burkina-Faso (2004-2016), la République Démocratique du Congo (2004-2016) et la Zambie (2004-2016). Il y a aussi des pays d'Amérique centrale tels que le Mexique (en 2004, 2008 et 2016), la République Dominicaine (2004-2016). On retrouve, également, des pays du Moyen-Orient : Israël (2004, 2012 et 2016), Égypte (2012), Syrie (2004 et 2012), Irak (2004-2012) et l'Arabie Saoudite. On trouve des graines de contagion en Asie avec l'Afghanistan (2004-2016), le Pakistan (2004, 2012 et 2016), Papouasie-Nouvelle-Guinée (2012), le Bangladesh (2012) et les Philippines (2016), ainsi que des pays de l'Europe de l'est, la Pologne, la Slovaquie, les anciens membres de la Yougoslavie (2004 et 2012), la Grèce (2008) et la Géorgie (2004-2016). Les pays dont la balance économique $B_c \lesssim 0$ risquent d'être en banqueroute au tout début du processus de contagion. Pour chacune des années considérées, ces pays sont systématiquement les pays

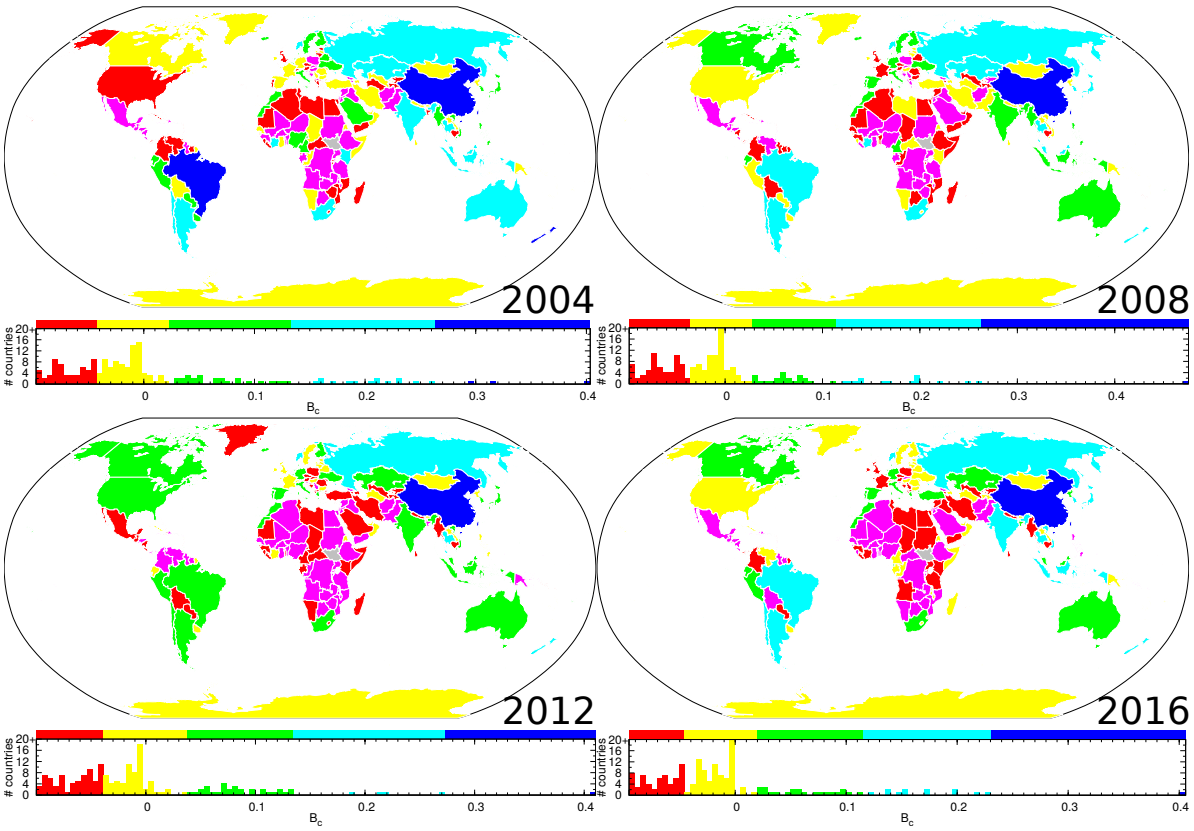

FIGURE 5.6 : Distribution géographique des balances PageRank-CheiRank à l'étape initiale ($\tau = 0$) de la contagion de crise économique pour un seuil de banqueroute $\kappa = 0.1$. Pour chacune des années, les pays colorés en rouge ont une balance PageRank-CheiRank fortement négative et les pays en bleu ont une balance positive. Les catégories de couleurs ont été calculées avec la méthode de Jenks [48]. Les pays considérés en banqueroute sont colorés en magenta. Le paramètre $\alpha = 0.5$ a été utilisé. D'après [97].

d'Afrique, avec l'exception du Maroc et de l'Afrique du Sud, les pays du Moyen-Orient, le Laos, le Cambodge, la Papouasie-Nouvelle-Guinée, les pays d'Amérique centrale, des Caraïbes, les

pays du nord de l'Amérique du sud, la Bolivie et le Paraguay. On retrouve également des pays riches comme les États-Unis (sauf en 2012), la France, Le Royaume-Uni, l'Irlande, ou encore la Suisse. Les pays ayant une très bonne santé économique sont les membres du BRICS (Brésil, Russie, Inde, Chine et Afrique du Sud), avec $0.1 \lesssim B_c \lesssim 0.4$.

Afin de mesurer la résistance d'un pays face à la contagion de crise économique au sein du réseau du commerce international, nous définissons $\kappa_{\max}$, la plus grande valeur κ pour laquelle le pays tombe en crise, au plus tard à τ_∞. Autrement dit, pour $\kappa > \kappa_{\max}$ le pays n'est jamais en banqueroute, $(B_c > -\kappa, \forall\tau)$. La Figure 5.7 donne la distribution géographique de $\kappa_{\max}$ pour les années 2004, 2008, 2012 et 2016. On observe que la Russie est le pays le plus résistant avec $\kappa_{\max} \approx -0.2$ en 2004, 2008 et 2012. Comme $\kappa > 0$, la Russie ne tombe jamais en crise économique pour cette période. En 2004, l'Arabie Saoudite, le Nigeria et le Kenya sont les seconds pays les plus insensibles, avec $\kappa_{\max} \lesssim 0.1$. La super résistance de ces

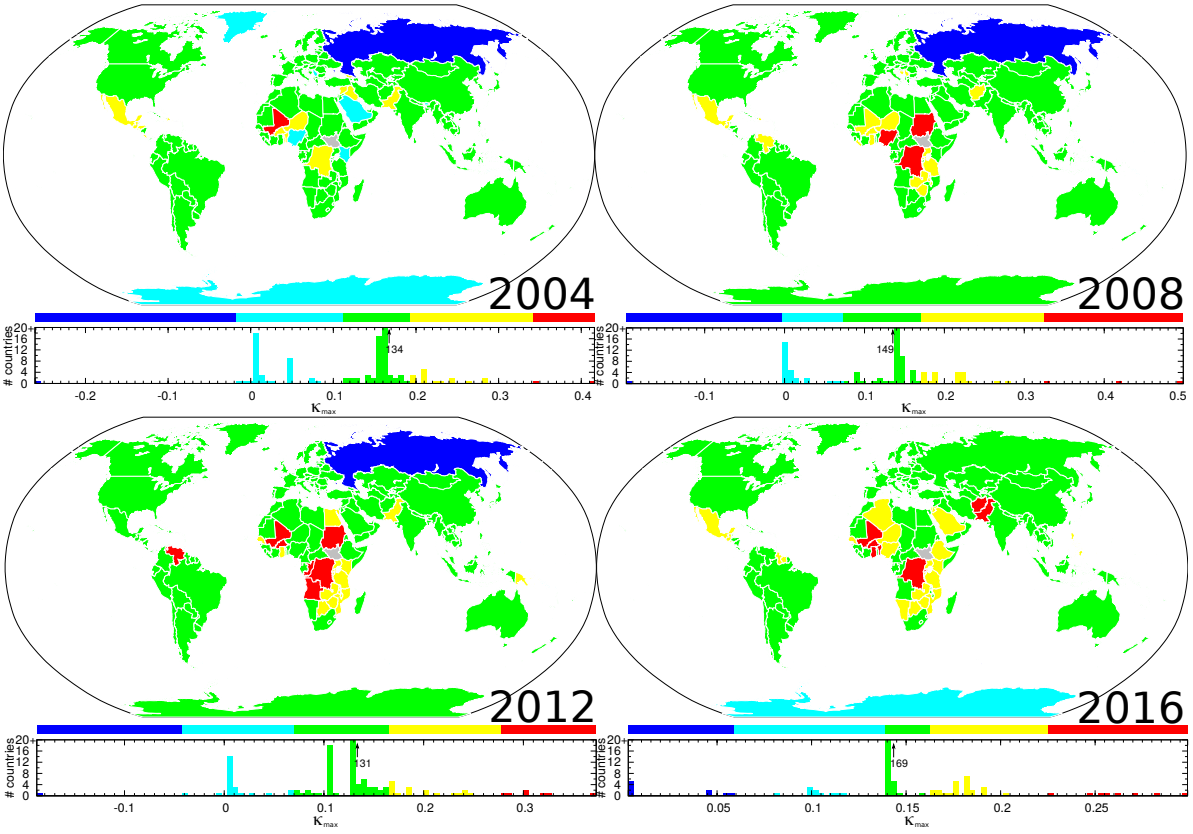

FIGURE 5.7 : Distribution géographique du seuil maximum $\kappa_{\max}$ pour lequel un pays est en banqueroute. Les pays ayant une valeur $\kappa_{\max}$ élevée sont en bleu, les pays avec un $\kappa_{\max}$ faible sont en rouge. Les catégories de couleurs ont été calculées avec la méthode de Jenks [48]. Le modèle A de contagion de crise économique et le paramètre $\alpha = 0.5$ ont été utilisés. D'après [97].

pays face à la propagation d'une crise économique est due à d'importants exports de pétrole et de gaz. On observe un pic dans la distribution des pays aux valeurs de $\kappa_{\max} \approx 0.175$ en 2004, $\kappa_{\max} \approx 0.15$ en 2008, $\kappa_{\max} \approx 0.14$ en 2012 et $\kappa_{\max} \approx 0.15$ en 2016. On remarque que ces valeurs sont très proches des points critiques κ_c obtenus avec les graphes $\eta(\kappa)$ (voir Figure 5.4). Les pays vulnérables, avec $\kappa_{\max} \gtrsim 0.2$, sont des pays d'Amérique centrale et d'Amérique du sud (2004,2008 et 2016) dont le Mexique (2004, 2008 et 2016), le Guatemala (2004, 2008 et 2016), le Salvador (2004 et 2016), le Honduras (2004), le Costa Rica (2004), la République Dominicaine (2004 et 2008), le Venezuela (2008,2012), la Guyane (2016), le Suriname (2016).

On retrouve aussi de tels pays en Afrique subsaharienne avec le Mali (2004-2016), le Burkina-Faso (2004-2016), le Togo (2004), le Bénin (2004 et 2016), le Niger (2004 et 2016), la République Démocratique du Congo (2004-2016), le Liberia (2008), le Ghana (2008-2016), le Nigeria (2008), le Soudan (2008 et 2012), l'Ouganda (2008-2016), le Rwanda (2008-2016), la Tanzanie (2008-2016), la Zambie (2008-2016), le Zimbabwe (2008-2016), le Malawi (2008 et 2012), le Sénégal (2012 et 2016), l'Égypte (2012), la République du Congo (2012), l'Angola (2012), le Burundi (2012 et 2016), le Kenya (2012 et 2016), le Mozambique (2012 et 2016), le Botswana (2012 et 2016), l'Éthiopie (2016), l'Algérie (2016) et la Namibie (2016). On retrouve aussi des pays situés au Moyen-Orient avec la Syrie (2004), l'Irak (2004), la Géorgie (2004), Israël (2016), la Jordanie (2016) et l'Arabie Saoudite (2016), des pays de l'Europe, la Slovénie (2008), la Bosnie-Herzégovine (2008) et la Serbie (2008), ainsi que quelques pays d'Asie comme le Pakistan (2004, 2008, 2016), l'Afghanistan (2008, 2016), les Philippines (2016) et la Papouasie-Nouvelle-Guinée (2012).

Nous venons de voir que les pays les plus fragiles, et aussi, les graines de contagion se trouvent principalement en Afrique subsaharienne, en Amérique du Sud et Amérique centrale, au Moyen-Orient et en Europe de l'est. Quand ces pays tombent en crise économique, la restriction de leur import fait que les pays en bonne santé économique ont de moins en moins de volume d'export. Nous montrons à la Figure 5.8 la distribution géographique de la fraction de produits ne pouvant être exportés à la fin de la contagion ($\tau = \tau_\infty$) avec un seuil $\kappa = 0.1$ et pour les années 2008 et 2016. Pour l'année 2008 (partie gauche de la Figure 5.8), les effets de la crise économique sont moins sévères que pour l'année 2016 (panneau de droite). La plupart des pays de l'ouest ont moins de 17% de leur produits bloqués en fin de contagion. Une situation similaire est obsevée pour d'autres pays, c'est le cas pour l'Ukraine, la Biélorussie, la Moldavie, la Bulgarie, le Kazakhstan, la Turquie, la Chine, l'Inde, la Thaïlande, Taïwan, la Corée du Sud, Singapour et l'Indonésie. Bien que la Russie ne soit pas en banqueroute en 2008 ($\kappa_{\max} \approx -0.2$), plus de 87% de ses produits ne peuvent plus être exportés à τ_∞. La Russie survit à la crise, puisque le modèle A permet aux pays banqueroute d'importer du pétrole et du gaz. De plus, en 2008, 60% de l'exportation de la Russie est liée au pétrole et au gaz. Pour l'année 2016, les effets de la crise économique sont plus importants. En effet, presque tous les pays ont plus de 90% de leurs produits bloqués. Les quelques pays qui sont les moins impactés par la crise sont, le Royaume-Uni, l'Afrique du Sud et la Nouvelle-Zélande, avec moins de 30% de produits non exportables.

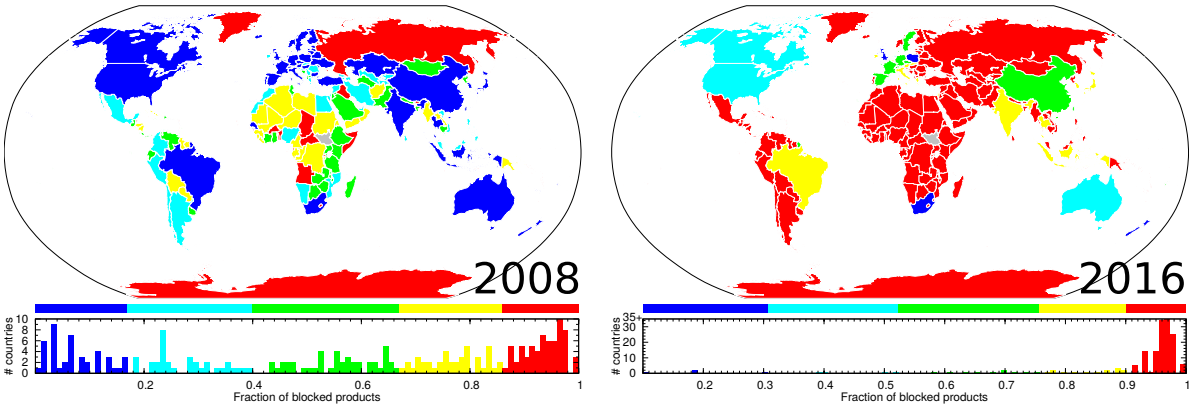

FIGURE 5.8 : Fraction de produits non exportables en raison de pays en banqueroute à τ_∞. Les pays colorés en bleu sont les pays ayant le plus d'exports et les pays en rouge sont ceux qui en ont le moins. Les catégories de couleurs ont été calculées avec la méthode de Jenks [48]. Le modèle A de contagion de crise économique avec un seuil $\kappa = 0.1$ et avec un paramètre $\alpha = 0.5$ a été utilisé. D'après [97].

5.4.5 Réseau de contagion

Nous nous intéressons, ici, au déroulement de la contagion et plus précisément à la construction de réseaux de causalités. Soit c et c' deux pays, le lien dirigé partant de c et allant vers c' existe dans le réseau de contagion, si c' est banqueroute juste après c, et si le pays c importe habituellement des produits venant du pays c'. Deux versions de réseaux de contagion sont présentées dans cette section : un réseau global de contagion, où le liens dirigé depuis c vers c' est courbé et dont l'épaisseur est proportionnelle à la masse monétaire $M_{cc'} = \sum_{\mathrm{p} \in \tilde{\mathcal{P}}} M^{p}_{cc'}$ des transactions que c ne peut plus importer, et un réseau hiérarchique de contagion, où on ne garde que les transactions telles que $M_{cc'} > 10^{10}$ dollars US. La couleur d'un nœud dépend de l'étape τ à laquelle le pays est en banqueroute, en partant du rouge ($\tau = 0$) au bleu (τ_∞), enfin, la couleur d'un lien est celle du nœud source.

La Figure 5.9 présente le réseau global de contagion pour les années 2004, 2008, 2012 et 2016 avec seuil de banqueroute $\kappa = 0.1$. La direction du lien allant de c vers c' est donnée en suivant la courbure dans le sens antihoraire ($c \smile c'$). On voit que les graines de contagion, en rouge, sont des pays situés en Afrique subsaharienne, au Moyen-Orient, en Amérique centrale et en Europe de l'est. Très peu de pays d'Asie sont parmi les graines de contagion. La direction des premiers liens (premiers imports interdits) est significative de la façon dont la contagion va évoluer. Pour l'année 2004, certains de ces liens partent de l'Europe de l'est et de l'Amérique centrale pour se diriger vers l'Amérique du nord. On voit que la crise économique des États-Unis, arrivant à l'étape $\tau = 1$, est principalement due à la chute du Mexique, de pays de l'Amérique centrale et de l'Europe de l'est. Une fois tombé, les États-Unis propage la crise économique dans les pays de l'Europe de l'ouest et initie le début de la contagion en Asie. À $\tau = 3$, le Japon et la Corée du Sud sont en banqueroute et provoquent la crise économique en Chine et en Australie. Ce même schéma de contagion arrive en 2008. Pour l'année 2012, les États-Unis sont en banqueroute plus tard ($\tau = 3$) après que le Mexique, la Corée du Sud, Singapour et la France soient en crise économique. La crise économique se propage en Asie via Singapour et la Corée du Sud. En 2016, les États-Unis sont en crise à l'étape $\tau = 3$ de la contagion, sa crise économique étant provoquée par Singapour, le Royaume-Uni et la France. La chute des États-Unis entraîne celle de la Chine, de la Corée du Sud et du Japon.

La Figure 5.10 présente les réseaux hiérarchiques de contagion pour la crise économique des années 2004, 2008, 2012 et 2016, avec un seuil de banqueroute $\kappa = 0.1$. Pour l'année 2004, les graines de contagion impliquées dans d'importantes transactions d'importations sont le Mexique et Israël. Tous deux contribuent à la crise économique des États-Unis (US) qui tombe à $\tau = 1$. Deux chemins de propagation de crise économique partent des US. Le chemin le plus important, en terme de volume de transactions bloqués, est le chemin US $\rightarrow$ Japon $\rightarrow$ (autres pays d'Asie et Australie). Le second chemin contribue aux crises économiques des pays de l'Europe. La crise de l'Europe est grandement due aux principaux pays européens tels que la France, l'Espagne, l'Autriche, le Royaume-Uni, qui tombent à $\tau = 1$. La crise économique de la Chine est due à la Corée du Sud, au Japon mais aussi à des pays d'Europe. On note que la France et le Royaume-Uni (UK) contribuent aussi à la crise économique du Japon. Pour l'année 2008, on observe un scénario de contagion similaire à celui de l'année 2004. Cependant, il y a plus de graines de contagion importantes, notamment, la Pologne et la Slovaquie, qui conduisent la crise économique à travers l'Europe, ainsi que le Venezuela, l'Arabie Saoudite et le Mexique, qui contribuent à la crise économique des US. De plus, à la différence du réseau hiérarchique de contagion de l'année 2004, le Canada tombe juste après les US et le lien entre ces deux pays est, par ailleurs, le plus fort de tout le réseau. Les pays du BRIC sont les moins impactés par la contagion (aussi en 2012 et 2016). Le réseau hiérarchique de contagion pour l'année 2012 diffère énormément des deux précédents. Il y a plus de sources de contagion aux étapes $\tau = 0$ et $\tau = 1$ de la contagion comme le Mexique, le Panama, le Costa Rica (Amérique centrale), l'Autriche, la France, la Slovaquie (Europe), l'Irak, la Turquie (Moyen-Orient), Singapour et la

Corée du Sud (Asie). C'est le premier réseau de contagion où l'on observe la présence de pays d'Asie en banqueroute dès le début ($\tau = 1$). La crise économique se propage très rapidement et sur de nombreux continents, à l'exception de l'Océanie. La chute des US arrive plus tard qu'en 2004 et 2008 ($\tau = 2$), elle est majoritairement due aux pays d'Amérique centrale, d'Asie et à la France. La crise des US contribue à celle de la majorité des pays présents dans ce réseau. La crise économique en Asie est principalement due à celle du Japon qui est induite par celle de la Corée du Sud, de Singapour et de la France. La crise en Europe est due principalement à la France, l'Autriche, le Mexique, la Turquie, la Corée du Sud et Singapour. En 2016, nous avons une remonté vers les $\tau \geq 3$ des pays du continent américain. Il y a une seule graine de contagion importante, la Slovaquie, elle propage la crise économique en Europe à travers la République Tchèque. La crise des pays d'Europe est majoritairement due à celle de la France et de UK. Singapour est le seul pays d'Asie de ce réseau à être tombé tôt ($\tau = 1$) et contribue à la crise économique des US. La crise économique des US est majoritairement due aux principaux importateurs européens, la France et UK. La crise en Asie est majoritairement due aux US. Enfin la Russie est le dernier pays à être tombé, à $\tau = 6$, en raison de la crise économique de la Biélorussie ($\tau = 5$). Il est intéressant de voir comment évolue la position des zones colorées, relatives aux continents. On remarque que de 2004 à 2016, la zone des pays d'Asie descend, tandis que la zone des pays d'Amérique remonte. Alors qu'au départ, la crise économique s'installe en Asie par le biais de pays asiatiques, en 2016, elle vient des pays d'Amérique.

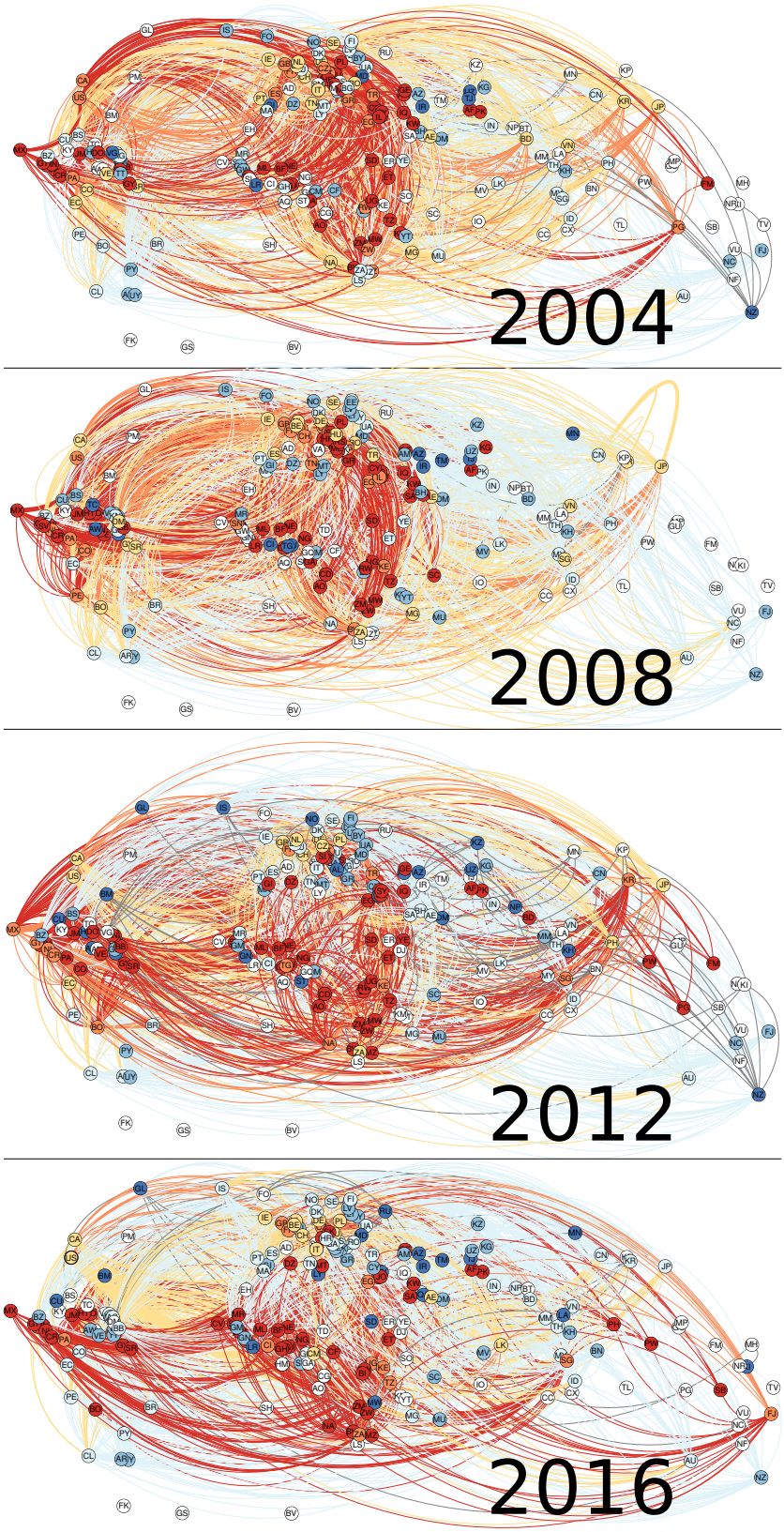

FIGURE 5.9 : Réseau global de contagion de crise économique pour les années 2004, 2008, 2012 et 2016. Les couleurs des nœuds vont du rouge pour les pays en banqueroute à $\tau = 0$, au bleu pour les pays en banqueroute à τ_∞. Les nœuds colorés en blanc sont les pays qui n'ont jamais été en banqueroute pendant la contagion. Le lien dirigé partant de A et allant vers B suit la courbure ($A \smile B$) dans le sens antihoraire, il signifie que le pays A est tombé en crise juste avant le pays B. La couleur des liens est la même que les nœuds sources. L'épaisseur des liens est proportionnelle au volume d'exports en provenance de B. Le modèle A de contagion de crise économique avec un seuil $\kappa = 0.1$ et un paramètre $\alpha = 0.5$ a été utilisé. D'après [97].

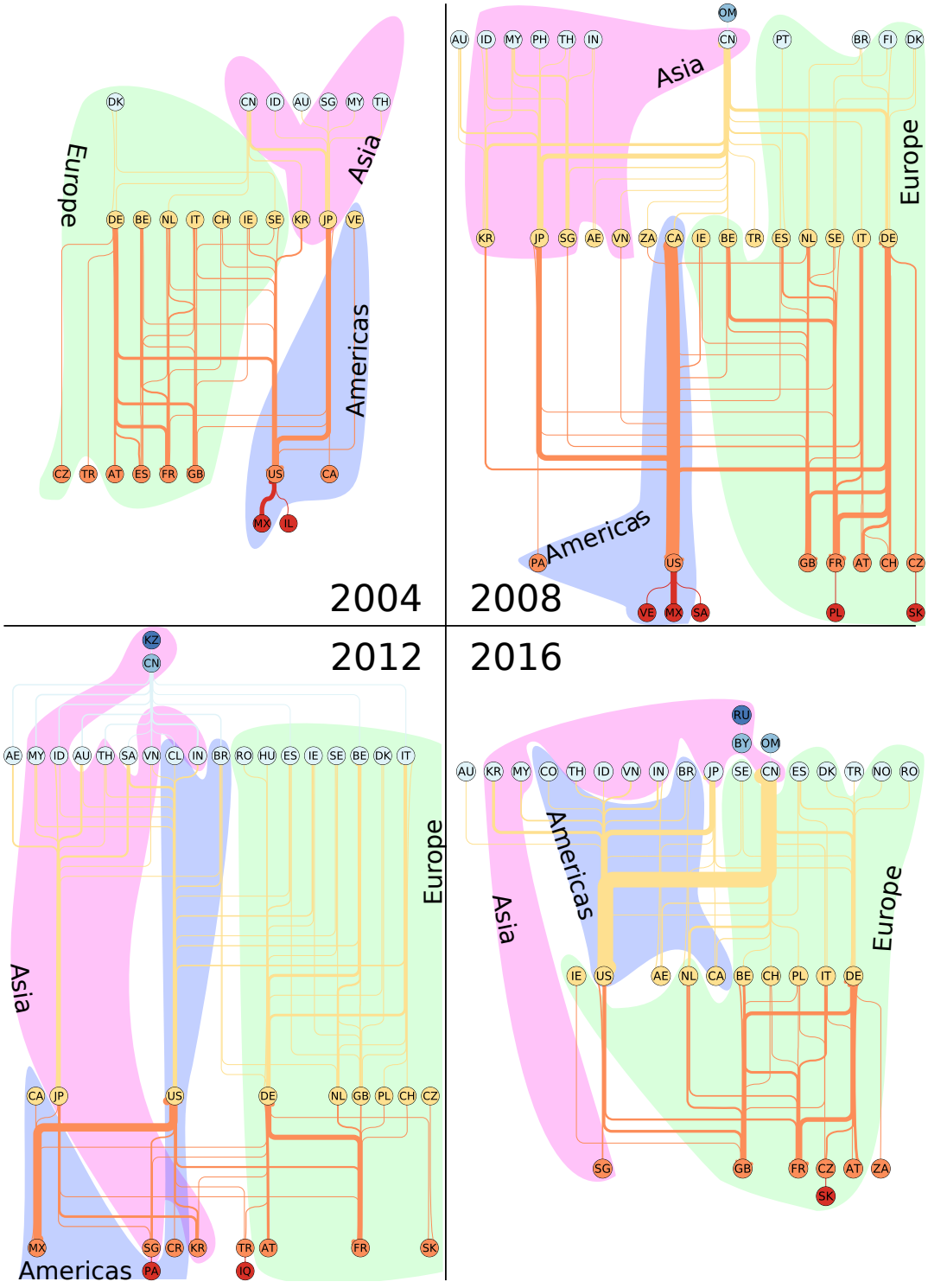

FIGURE 5.10 : Réseau hiérarchique de contagion de crise économique pour les années 2004, 2008, 2012 et 2016. Seules les transactions supérieures à 10^{10} dollars US sont représentées. Les pays qui sont en banqueroute à la même étape τ de contagion sont alignés sur une même ligne. Le code couleur pour les nœuds est le même que celui utilisé à la Figure 5.9. De bas en haut, les lignes correspondent aux étapes de contagion allant de $\tau = 0$ à $\tau = 3$ pour 2004, à $\tau = 4$ pour 2008, à $\tau = 5$ pour 2012 et 2016. L'épaisseur du lien partant du pays c, en banqueroute à τ, vers le pays c', en banqueroute à $\tau' = \tau + 1$, est proportionnelle au volume des transactions que le pays c ne peut plus importer. Les zones colorées regroupent les pays présents dans un même continent, zone verte pour les pays européens, bleue pour les pays américains et rose pour les pays asiatiques. Le modèle A de contagion de crise économique avec un seuil $\kappa = 0.1$ et un paramètre $\alpha = 0.5$ a été utilisé. D'après [97].

5.4.6 Conclusion

Nous venons de voir comment obtenir un modèle de contagion de crise économique à partir de la méthode de la matrice de Google. Ce modèle basé sur la balance économique PageRank-CheiRank B permet de tenir compte des transactions directes et indirectes entre les pays. Deux scénarios de contagion ont été proposés, le premier scénario laisse la possibilité aux pays en banqueroute de pouvoir continuer d'importer du pétrole et du gaz (modèle A), tandis que le second scénario interdit toutes importations à un pays, quand celui-ci est en banqueroute (modèle B). Lorsque la balance B_c PageRank-CheiRank d'un pays c est inférieure à un seuil de banqueroute $-\kappa$, le pays est considéré en banqueroute. Pour les deux modèles, nous observons une transition de phase pour $\kappa = \kappa_c$, entre un régime de contagion localisée ($\kappa > \kappa_c$) avec 10% des pays impactés et un coût total C_∞ de la crise représentant moins de 5% du volume commercial total, et un régime de contagion globale ($\kappa < \kappa_c$), avec 80 à 90% des pays touchés. Nous nous sommes intéressés à la propagation de crise économique au sein du réseau du commerce international des années 2004 à 2016 en utilisant le modèle A avec un seuil de banqueroute $\kappa = 0.1$. Nous avons observé qu'en début de contagion ($\tau = 0$) des *graines de contagion* apparaissent. Ces graines sont des pays dont les volumes d'imports et d'exports sont faibles. Ces pays sont situés principalement en Afrique subsaharienne, en Amérique centrale et Amérique du sud, au Moyen-Orient et en Europe de l'est. Pour avoir une économie stable, l'importation de ces pays est restreinte au pétrole et au gaz (modèle A). De ce fait, leur chute entraîne celle d'autres pays à la prochaine étape de contagion. Ainsi, de grands exportateurs comme les US ou encore les pays d'Europe de l'ouest vont être impactés et tomber en crise économique. Par exemple, la France, en 2004, 2012 et 2016, est en banqueroute à $\tau = 1$ en raison des crises économiques de petits importateurs dont les transactions avec la France s'élève à moins de 10^{10} dollars US. Le Royaume-Uni est aussi impacté, de façon similaire, par la contagion en 2004, 2008 et 2016. La construction de réseaux hiérarchiques de contagion, représentant les étapes de la contagion, permet de déterminer les pays transmetteurs de la crise économique, ainsi que les transactions majeures ($> 10^{10}$ dollars US) misent en cause. Nous avons observé que les pays d'Europe et d'Amérique sont des transmetteurs importants de crise économique. L'Asie reste résistante face à la contagion, elle est majoritairement impactée par des pays d'Asie comme, le Japon, la Corée du Sud et Singapour. Pour la période de 2004 à 2016, les pays d'Asie tels que la Chine, l'Inde, l'Indonésie, la Malaisie et la Thaïlande, sont parmi les derniers à être impactés. Les BRIC (Brésil, Russie, Inde et Chine) tombent pendant les dernière étapes de la contagion. Le modèle choisit pour notre étude, le modèle A, est plus réaliste que le modèle B, mais il est discutable. En effet, on pourrait laisser la possibilité aux pays en banqueroute d'importer d'autres produits comme le métal et les composants chimiques. Nous proposons, par exemple, d'utiliser une table d'entrée/sortie telle que OECD-WTO TiVA[1] étudiée dans les travaux [79, 80]. Elle archive les transactions entre secteurs d'activités économiques, en valeur ajouté. En utilisant ces données, nous pourrions restreindre les imports des pays en banqueroute aux produits les plus importants pour leurs secteurs de production. Ce nouveau scénario de propagation de crise économique peut-être l'objet de futures recherches. Aussi, nous proposons comme futures plans de recherche, l'étude de la propagation de crise économique, avec la possibilité que les pays en banqueroute puissent reprendre leurs imports après un délai $\Delta\tau$, ainsi que l'étude de crise énergétique induite par une augmentation du prix du pétrole et/ou du gaz à $\tau = 0$.

[1] Cette banque de données provient de l'Organisation de coopération et de développements économiques (OCDE) et de l'Organisation Mondiale du Commerce (OMC).

5.5 Contagion de crise économique dans le réseau de transactions de Bitcoin (RTB)

5.5.1 Introduction

Nous présentons, ici, une autre application du modèle de contagion de crise économique que nous proposons. Nous nous intéressons au réseau de transactions de Bitcoin qui, contrairement aux données sur les transactions bancaires, sont accessibles publiquement. Le Bitcoin est la première cryptomonnaie, elle est fondée par Nakamoto Satichi et lancée en 2008. Le fonctionnement de cette cryptomonnaie est décrite dans l'article [98]. Les premières analyses du réseau de transactions de Bitcoin sont présentées dans les travaux [99, 100]. Une vue générale du système du Bitcoin, ses atouts et ses désavantages, sont donnés dans l'article [101]. La différence majeure entre une cryptomonnaie et une monnaie nationale est l'utilisation de la *blockchain*. La *blockchain* peut être vue comme l'historique de toutes les transactions, depuis la création de la monnaie. Cet historique assure l'anonymat des participants, mais aussi il empêche la fraude. L'analyse de la matrice de Google associée au réseau de transactions de Bitcoin est présentée dans l'article [102], elle montre qu'une petite proportion d'utilisateurs, pouvant avoir plusieurs portefeuilles Bitcoin, sont responsables de la santé économique du réseau entier. Nous allons appliquer le modèle de propagation de crise économique basé sur la balance économique PageRank-CheiRank d'un utilisateur u (portefeuille) au réseau de transactions de Bitcoin, construit à partir des données disponibles sur le site `http://blockchain.info/` et extraites par la méthode d'Ivan Brugere [103]. Dans un premier temps, nous allons rapidement présenter les données de transactions de Bitcoin et la méthode de construction de la matrice de Google G associée au réseau de transaction de Bitcoin (section 5.5.2). Enfin, nous discuterons des résultats concernant la contagion de crise économique dans le réseau de transactions de Bitcoin (section 5.5.3).

5.5.2 Les données Bitcoin

Nous avons utilisé les données de transactions de Bitcoin, décrites dans [102], depuis son lancement jusqu'en 2013. Les données relatives aux différentes années sont découpées en trimestres. La Table 5.1 donne le nombre de nœuds N (utilisateurs) et le nombre de liens N_l (transactions) des réseaux construits à partir de ces 12 banques de données. Le lien dirigé $u \to u'$ représente le volume de Bitcoin échangé, depuis l'utilisateur u vers l'utilisateur u'. La matrice de Google G associée au réseau de transaction de Bitcoin est construite selon la méthode standard (méthode de construction décrite dans [102]). Nous avons $G_{ij} = \alpha S_{ij} + (1-\alpha)/N$, avec $\alpha = 0.85$. Contrairement à l'exemple précédent (voir la section 5.4), nous n'avons qu'un seul scénario de crise : si un utilisateur u a une balance PageRank-CheiRank $B_u < -\kappa$, avec κ un seuil de banqueroute, toutes les transactions entrantes de l'utilisateur sont bloquées. Les résultats discutés dans cette section sont ceux obtenus pour le réseau de transactions de Bitcoin, associé au premier quartier de l'année 2013 (BC2013Q1).

Réseau	N	N_l	Réseau	N	N_l	Réseau	N	N_l
BC2010Q3	37818	57437	BC2011Q3	1546877	2857232	BC2012Q3	3742174	8381654
BC2010Q4	70987	111015	BC2011Q4	1884918	3635927	BC2012Q4	4671604	11258315
BC2011Q1	204398	333268	BC2012Q1	2186107	4395611	BC2013Q1	5997717	15205087
BC2011Q2	696948	1328505	BC2012Q2	2645039	5655802	BC2013Q2	6297009	16056427

TABLE 5.1 : Liste des nombres de nœuds N (utilisateurs) et des nombres de liens N_l (transactions) associés aux réseaux de transactions de Bitcoin. Le label BCaQq est le nom associé au réseau construit à partir du jeu de données du $q^{\text{ème}}$ quartier de l'année a. D'après [104].

5.5.3 Transition de phase

L'évolution de la proportion d'utilisateurs en banqueroute $N_u(\tau)/N$ en fonction du seuil de banqueroute κ pour différentes étapes τ est présentée à la Figure 5.11. On observe une transition de phase au point critique $\kappa_c \approx 0.1$ entre un régime, où la majorité des utilisateurs sont en banqueroute pour $\kappa < \kappa_c$, et un régime de contagion moins sévère pour $\kappa_c \approx 0.1 < \kappa < 0.55$, avec 50 à 70% des utilisateurs touchés. Enfin, pour $\kappa > 0.55$ la contagion affecte presque personne. On voit que, contrairement à la contagion de crise économique dans le réseau du commerce international, la contagion de crise dans le réseau de transactions de Bitcoin est caractérisée par un seuil de banqueroute κ_c critique plus faible, ainsi que par une transition de phase beaucoup moins abrupte.

Nous avons placé les utilisateurs dans le plan (K, K^*) associé aux rangs PageRank K et CheiRank K^*. La Figure 5.12 donne l'évolution des états banqueroute (en rouge) ou sain (en bleu) des utilisateurs avec un seuil $\kappa = 0.15$ (en haut), $\kappa = 0.3$ (au milieu) et $\kappa = 0.6$ (en bas) aux étapes $\tau = 1$, 2 et 3 de la contagion. Pour le cas $\kappa = 0.15$, on voit que les utilisateurs, avec des rangs $K, K^* \sim 1$ sont rapidement en banqueroute. Les utilisateurs situés sous la diagonal $K = K^*$ partent avec une balance positive, mais arrivent rapidement dans l'état banqueroute. On retrouve une évolution similaire pour $\kappa = 0.3$, avec bien évidement un nombre diminué d'utilisateurs en banqueroute. La distribution dans le plan (K, K^*) des utilisateurs en banqueroute et des utlisateurs sains reste stable pour le cas $\kappa = 0.6$. La région colorée en blanc, représentant une proportion d'utilisateurs en banqueroute proche de la proportion d'utilisateurs sains, évolue avec τ.

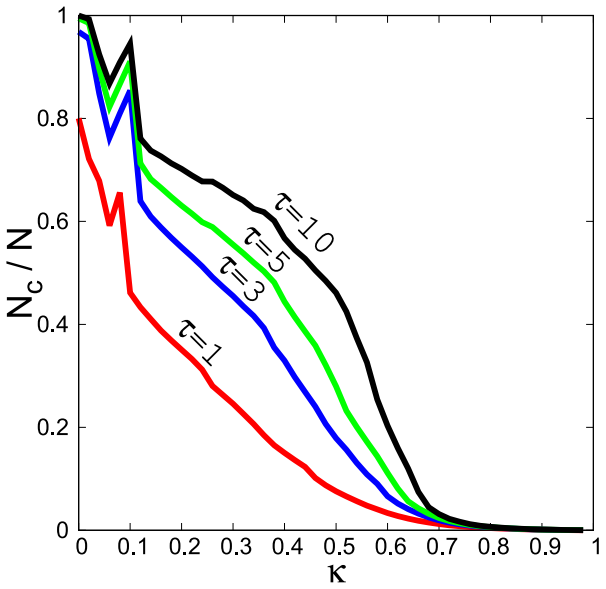

FIGURE 5.11 : Proportion d'utilisateurs (N_u/N) en banqueroute en fonction du seuil κ pour différentes étapes τ de la contagion. Les données BC2013Q1 ont été utilisées. D'après [104].

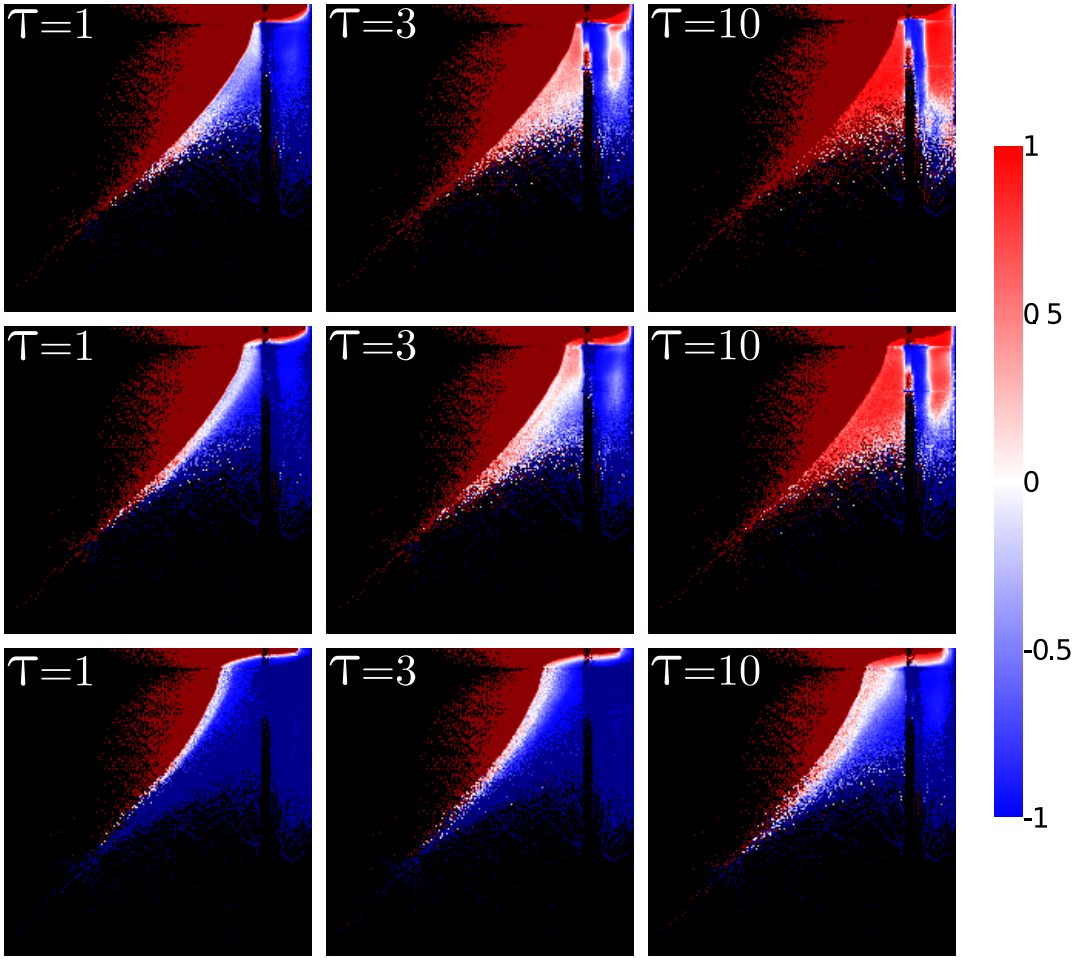

FIGURE 5.12 : Distribution des utilisateurs dans le plan (K, K^*) pour BC2013Q1. Le plan (K, K^*) est découpé en 200×200 cellules carrés de taille logarithmique. Le nombre d'utilisateurs dans une cellule donnée est N_{cell}. La proportion d'utilisateurs en banqueroute, à l'étape τ de la contagion, dans une cellule est $N_{\text{u,cell}}$. Les cellules sont colorés en fonction de la valeur de $(2N_{\text{u,cell}} - N_{\text{cell}})/N_{\text{cell}}$. La couleur rouge signifiant que la cellule est majoritairement remplie d'utilisateurs en banqueroute, le bleu qu'elle est majoritairement composée d'utilisateurs sains et blanc s'il y a autant d'utilisateurs sains que d'utilisateurs en banqueroute. Différentes valeurs κ ont été utilisées, $\kappa = 0.15$ (figures du haut), $\kappa = 0.3$ (figures du milieu) et $\kappa = 0.6$ (figures du bas). D'après [104].

5.5.4 Conclusion

L'application du modèle de propagation de crise économique au réseau de transactions de Bitcoin, nous montre que les utilisateurs avec $K, K^* \sim 1$ entrent en état banqueroute très rapidement. Un nombre important de transactions entre ces utilisateurs peut être responsable de ce phénomène. L'analyse du réseau des interactions directes et indirectes entre ce groupe d'utilisateurs est présentée dans l'article [104]. Enfin, on pourrait étendre cette étude avec l'intégration des adresses IP associées aux utilisateurs et ainsi étudier la contagion de crise économique dans le réseau de transaction de Bitcoin géographiquement.

Chapitre 6

Conclusion et perspectives

Dans ce manuscrit de thèse, nous vous avons présenté diverses applications de la méthode de la matrice de Google réduite, ainsi que d'autres outils issus de la matrice de Google. Cela montre à quel point le domaine de la science des réseaux et plus généralement des systèmes complexes sont des domaines qui étudient une grande diversité d'objets. Dans le monde d'aujourd'hui, où nous avons de plus en plus de données générées et par conséquent de plus en plus de données stockables et analysables, il est nécessaire d'optimiser l'extraction d'informations relatives à celles-ci, que ce soit au niveau du temps de calcul, ou bien de la mémoire allouée à cet effet. Pour cela, la méthode de la matrice de Google réduite, introduite dans [30], est un excellent outil qui permet, à partir d'un sous-réseau d'intérêt, l'extraction des interactions directes, et indirectes par diffusion dans tous le réseau, entre les nœuds d'intérêts. L'efficacité de cette méthode a été démontrée par son application à de nombreux réseaux tels que les réseaux d'encyclopédies avec l'exemple de l'encyclopédie en ligne Wikipédia ou encore les réseaux économiques comme le réseau du commerce international et le réseau mondial des activités économiques. Nous avons présentés la balance PageRank-CheiRank qui est une balance économique construite à partir des vecteurs PageRank et CheiRank, ces derniers vecteurs donnant respectivement les capacités d'importation et d'exportation relatives aux acteurs économiques qui constituent les réseaux économiques. Dans un contexte beaucoup plus récent, celui des épidémies et leur contagion à travers le monde, nous avons construit un modèle de propagation de crises économiques basé sur la balance PageRank-CheiRank. Deux types de scénarios ont été proposés, soit le pays en crise économique ne peut plus rien importer (modèle B), soit il peut uniquement importer du pétrole et du gaz (modèle A). Nous pensons que le modèle A est le plus représentatif de la réalité puisque ces produits sont importants dans l'industrie et la production des pays et permettent de rendre une stabilité économique aux pays en crise. Seulement, à la vue des données sur le commerce international et sur les échanges d'activités économiques entre secteurs, nous sommes capables d'améliorer ce modèle. En effet, une perspective de recherche serait d'aboutir à un scénario *sur mesure*, plus réaliste, avec l'utilisation couplée de ces deux bases de données. En déterminant les besoins importants des secteurs de productions d'un pays, nous pourrions déterminer un seuil de banqueroute κ ainsi qu'une liste de produits dont l'importation serait tolérée en cas de crise économique d'un pays donné, et ainsi freiner la contagion et diminuer le coût financier d'une crise économique mondiale.

Annexe A

Figures et Tables annexes

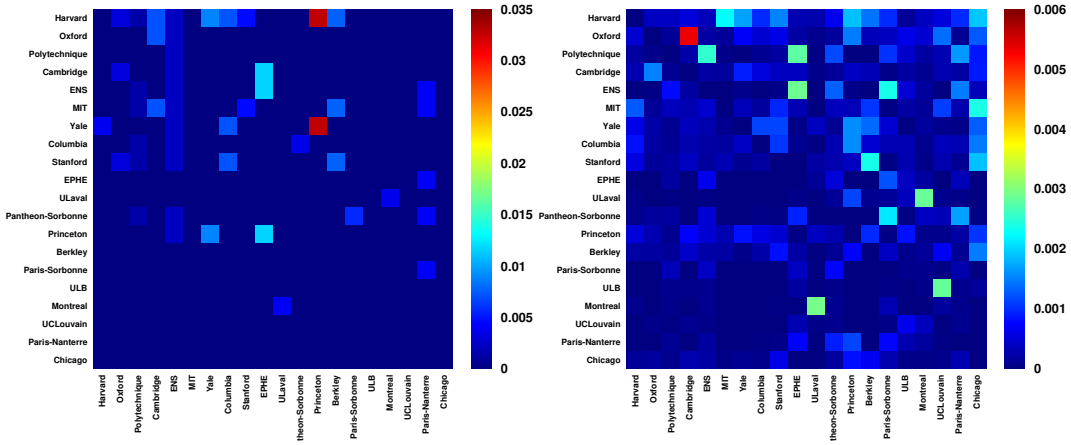

FIGURE A.1 : Matrices G_{rr} (à gauche) et G_{qrnd} (à droite) pour le top 20 FRWRWU (voir Table 3.6). Les colonnes et lignes sont ordonnées selon cette même liste. Les poids des matrices sont $W_{\mathrm{rr}} = 0.01404$ et $W_{\mathrm{qrnd}} = 0.00746$. Le code couleur dépend des éléments des matrices.

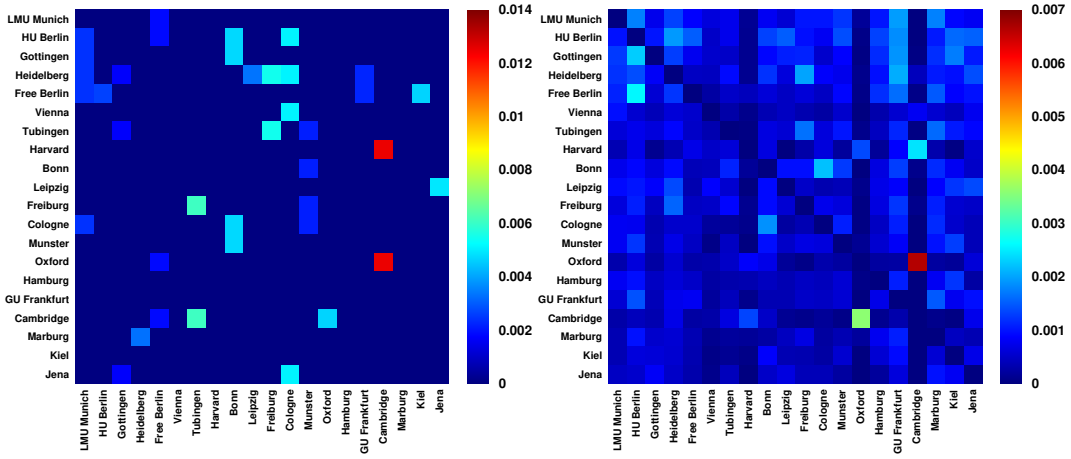

FIGURE A.2 : Matrices G_{rr} (à gauche) et G_{qrnd} (à droite) pour le top 20 DEWRWU (voir Table 3.7). Les colonnes et lignes sont ordonnées selon cette même liste. Les poids des matrices sont $W_{\mathrm{rr}} = 0.00746$ et $W_{\mathrm{qrnd}} = 0.0128$. Le code couleur dépend des éléments des matrices.

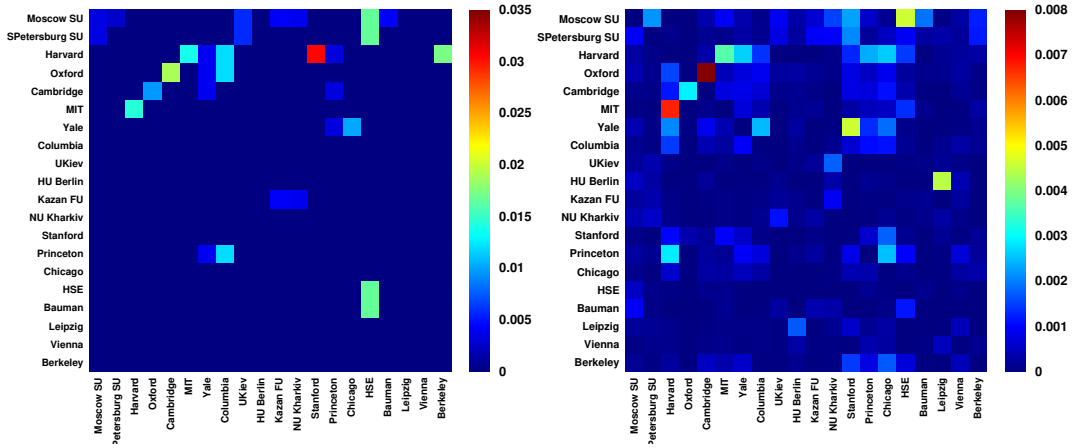

FIGURE A.3 : Matrices G_{rr} (à gauche) et G_{qrnd} (à droite) pour le top 20 RUWRWU (voir Table 3.8). Les colonnes et lignes sont ordonnées selon cette même liste. Les poids des matrices sont $W_{\mathrm{rr}} = 0.014$ et $W_{\mathrm{qrnd}} = 0.008$. Le code couleur dépend des éléments des matrices.

SITC	Produit	SITC	Produit
00	Live animals	54	Medicinal and pharmaceutical products
01	Meat and meat preparations	55	Perfume materials, toilet and cleansing preptions
02	Dairy products and eggs	56	Fertilizers, manufactured
03	Fish and fish preparations	57	Explosives and pyrotechnic products
04	Cereals and cereal preparations	58	Plastic materials, etc.
05	Fruit and vegetables	59	Chemical materials and products, nes
06	Sugar, sugar preparations and honey	61	Leather, lthr. Manufs., nes and dressed fur skins
07	Coffee, tea, cocoa, spices & manufacs. Thereof	62	Rubber manufactures, nes
08	Feed. Stuff for animals excl. Unmilled cereals	63	Wood and cork manufactures excluding furniture
09	Miscellaneous food preparations	64	Paper, paperboard and manufactures thereof
11	Beverages	65	Textile yarn, fabrics, made up articles, etc.
12	Tobacco and tobacco manufactures	66	Non metallic mineral manufactures, nes
21	Hides, skins and fur skins, undressed	67	Iron and steel
22	Oil seeds, oil nuts and oil kernels	68	Non ferrous metals
23	Crude rubber including synthetic and reclaimed	69	Manufactures of metal, nes
24	Wood, lumber and cork	71	Machinery, other than electric
25	Pulp and paper	72	Electrical machinery, apparatus and appliances
26	Textile fibres, not manufactured, and waste	73	Transport equipment
27	Crude fertilizers and crude minerals, nes	81	Sanitary, plumbing, heating and lighting fixt.
28	Metalliferous ores and metal scrap	82	Furniture
29	Crude animal and vegetable materials, nes	83	Travel goods, handbags and similar articles
32	Coal, coke and briquettes	84	Clothing
33	Petroleum and petroleum products	85	Footwear
34	Gas, natural and manufactured	86	Scientif and control instrum, photogr gds, clocks
35	Electric energy	89	Miscellaneous manufactured articles, nes
41	Animal oils and fats	91	Postal packages not class. According to kind
42	Fixed vegetable oils and fats	93	Special transact. Not class. According to kind
43	Animal and vegetable oils and fats, processed	94	Animals, nes, incl. Zoo animals, dogs and cats
51	Chemical elements and compounds	95	Firearms of war and ammunition therefor
52	Crude chemicals from coal, petroleum and gas	96	Coin, other than gold coin, not legal tender
53	Dyeing, tanning and colouring materials		

TABLE A.1 : Liste des $N_{\mathrm{p}} = 61$ produits présents dans la banque de données UN Comtrade. La colonne CC indique les codes des produits selon la classification du commerce standard (SITC) $1^{\text{ère}}$ révision.

Pays	CC	Pays	CC
Afghanistan	AF	Albania	AL

TABLE A.2: Liste des $N_{\mathrm{c}} = 227$ pays présents dans la banque de données UN Comtrade. La colonne CC indique le code des pays selon le format ISO-3166-2.

Pays	CC	Pays	CC
Algeria	DZ	American Samoa	AS
Andorra	AD	Angola	AO
Anguilla	AI	Antarctica	AQ
Antigua and Barbuda	AG	Argentina	AR
Armenia	AM	Aruba	AW
Australia	AU	Austria	AT
Azerbaijan	AZ	The Bahamas	BS
Bahrain	BH	Bangladesh	BD
Barbados	BB	Belarus	BY
Belgium	BE	Belize	BZ
Benin	BJ	Bermuda	BM
Bhutan	BT	Bolivia	BO
Bosnia and Herzegovina	BA	Botswana	BW
Bouvet Island	BV	British Indian Ocean Territory	IO
British Virgin Islands	VG	Brazil	BR
Brunei	BN	Bulgaria	BG
Burkina Faso	BF	Burundi	BI
Cambodia	KH	Cameroon	CM
Canada	CA	Cape Verde	CV
Cayman Islands	KY	Central African Republic	CF
Chad	TD	Chile	CL
China	CN	Christmas Island	CX
Cocos (Keeling) Islands	CC	Colombia	CO
Comoros	KM	Republic of the Congo	CG
Cook Islands	CK	Costa Rica	CR
Ivory Coast	CI	Croatia	HR
Cuba	CU	Cyprus	CY
Czech Republic	CZ	North Korea	KP
Democratic Republic of the Congo	CD	Denmark	DK
Djibouti	DJ	Dominica	DM
Dominican Republic	DO	Ecuador	EC
Egypt	EG	El Salvador	SV
Equatorial Guinea	GQ	Eritrea	ER
Estonia	EE	Ethiopia	ET
Faroe Islands	FO	Falkland Islands	FK
Fiji	FJ	Finland	FI
France	FR	French Polynesia	PF
Micronesia	FM	Gabon	GA
The Gambia	GM	Georgia	GE
Germany	DE	Ghana	GH
Gibraltar	GI	Greece	GR
Greenland	GL	Grenada	GD
Guam	GU	Guatemala	GT
Guinea	GN	Guinea-Bissau	GW
Guyana	GY	Haiti	HT
Heard Island and McDonald Islands	HM	Vatican	VA
Honduras	HN	Hungary	HU

TABLE A.2: Liste des $N_c = 227$ pays présents dans la banque de données UN Comtrade. La colonne CC indique le code des pays selon le format ISO-3166-2.

Pays	CC	Pays	CC
Iceland	IS	India	IN
Indonesia	ID	Iran	IR
Iraq	IQ	Ireland	IE
Israel	IL	Italy	IT
Jamaica	JM	Japan Ryukyu Island	JP
Jordan	JO	Kazakhstan	KZ
Kenya	KE	Kiribati	KI
Kuwait	KW	Kyrgyzstan	KG
Laos	LA	Latvia	LV
Lebanon	LB	Lesotho	LS
Liberia	LR	Libya	LY
Lithuania	LT	Luxembourg	LU
Madagascar	MG	Malawi	MW
Malaysia	MY	Maldives	MV
Mali	ML	Malta	MT
Marshall Islands	MH	Mauritania	MR
Mauritius	MU	Mayotte	YT
Mexico	MX	Mongolia	MN
Montenegro	ME	Montserrat	MS
Morocco	MA	Mozambique	MZ
Myanmar	MM	Northern Mariana Islands	MP
Namibia	NA	Nauru	NR
Nepal	NP	Netherlands Antilles	AN
Netherlands	NL	New Caledonia	NC
New Zealand	NZ	Nicaragua	NI
Niger	NE	Nigeria	NG
Niue	NU	Norfolk Islands	NF
Norway	NO	State of Palestine	PS
Oman	OM	Pakistan	PK
Palau	PW	Panama	PA
Papua New Guinea	PG	Paraguay	PY
Peru	PE	Philippines	PH
Pitcairn	PN	Poland	PL
Portugal	PT	Qatar	QA
South Korea	KR	Moldova	MD
Romania	RO	Russia	RU
Rwanda	RW	Saint Helena	SH
Saint Kitts and Nevis	KN	Saint Lucia	LC
Saint Pierre and Miquelon	PM	Saint Vincent and the Grenadines	VC
Samoa	WS	San Marino	SM
Sao Tome and Principe	ST	Saudi Arabia	SA
Senegal	SN	Serbia	RS
Seychelles	SC	Sierra Leone	SL
Singapore	SG	Slovakia	SK
Slovenia	SI	Solomon Islands	SB
Somalia	SO	South Africa	ZA
South Georgia and the South Sandwich Islands	GS	Spain	ES
Sri Lanka	LK	Sudan	SD

TABLE A.2: Liste des $N_c = 227$ pays présents dans la banque de données UN Comtrade. La colonne CC indique le code des pays selon le format ISO-3166-2.

Pays	CC	Pays	CC
Suriname	SR	Swaziland	SZ
Sweden	SE	Switzerland	CH
Syria	SY	Tajikistan	TJ
Macedonia	MK	Thailand	TH
Timor-Leste	TL	Togo	TG
Tokelau	TK	Tonga	TO
Trinidad and Tobago	TT	Tunisia	TN
Turkey	TR	Turkmenistan	TM
Turks and Caicos Islands	TC	Tuvalu	TV
Uganda	UG	Ukraine	UA
United Arab Emirates	AE	United Kingdom	GB
Tanzania	TZ	United States Minor Outlying Islands	UM
Uruguay	UY	United States	US
Uzbekistan	UZ	Vanuatu	VU
Venezuela	VE	Vietnam	VN
Wallis and Futuna	WF	Western Sahara	EH
Yemen	YE	Zambia	ZM
Zimbabwe	ZW		

TABLE A.2: Liste des $N_c = 227$ pays présents dans la banque de données UN Comtrade. La colonne CC indique le code des pays selon le format ISO-3166-2.

Pays	CC	Pays	CC
Australia	AUS	Austria	AUT
Belgium	BEL	Canada	CAN
Chile	CHL	Czech Republic	CZE
Denmark	DNK	Estonia	EST
Finland	FIN	France	FRA
Germany	DEU	Greece	GRC
Hungary	HUN	Iceland	ISL
Ireland	IRL	Israel	ISR
Italy	ITA	Japan	JPN
Korea, Republic of	KOR	Luxembourg	LUX
Mexico	MEX	Netherlands	NLD
New Zealand	NZL	Norway	NOR
Poland	POL	Portugal	PRT
Slovakia	SVK	Slovenia	SVN
Spain	ESP	Sweden	SWE
Switzerland	CHE	Turkey	TUR
United Kingdom	GBR	United States	USA
Argentina	ARG	Brazil	BRA
China	CHN	Taiwan, Province of China	TWN
India	IND	Indonesia	IDN
Russian Federation	RUS	Singapore	SGP
South Africa	ZAF	Hong Kong	HKG
Malaysia	MYS	Philippines	PHL
Thailand	THA	Romania	ROU
Viet Nam	VNM	Saudi Arabia	SAU
Brunei Darussalam	BRN	Bulgaria	BGR
Cyprus	CYP	Latvia	LVA
Lithuania	LTU	Malta	MLT
Cambodia	KHM	Rest Of the World	ROW

TABLE A.3 : Liste des $N_c = 48$ pays (dont *Rest Of the World*) présents dans la banque de données OECD-WTO TiVa. Les colonnes CC indiquent les codes pays selon le format ISO-3166-3.

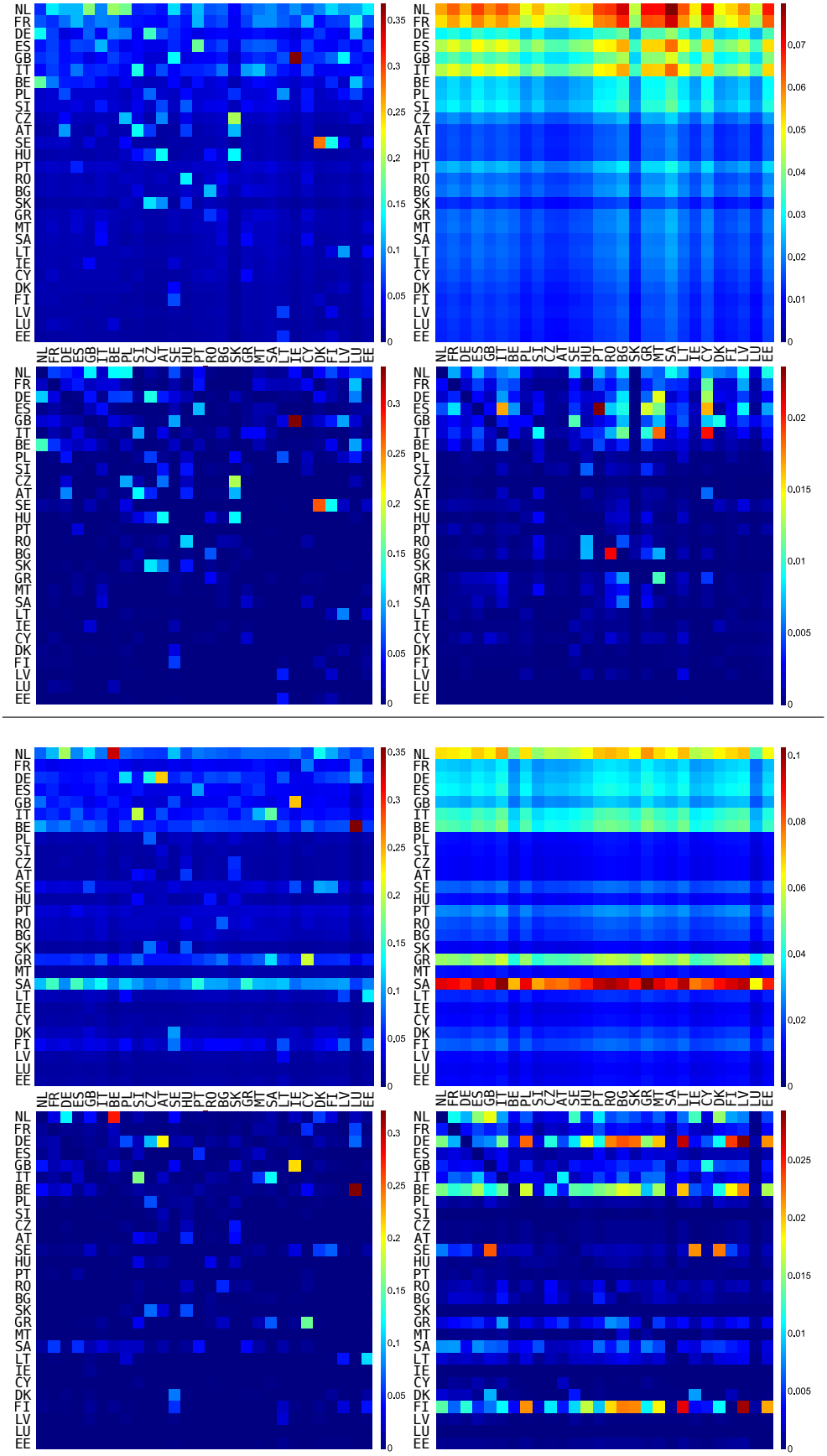

FIGURE A.4 : Matrices de Google réduite associées à G (4 matrices du haut) et G^* (4 matrices du bas), pour l'année 2016 et le pétrole saoudien. Pour chaque ensemble de 4 matrices, la matrice de Google réduite G_{r} est en haut à gauche et ses composantes sont G_{pr} (en haut à droite), G_{rr} (en bas à gauche) et G_{qrnd} (en bas à droite). Les lignes et colonnes sont classées suivant le classement PageRank de la Table 4.1. D'après [70].

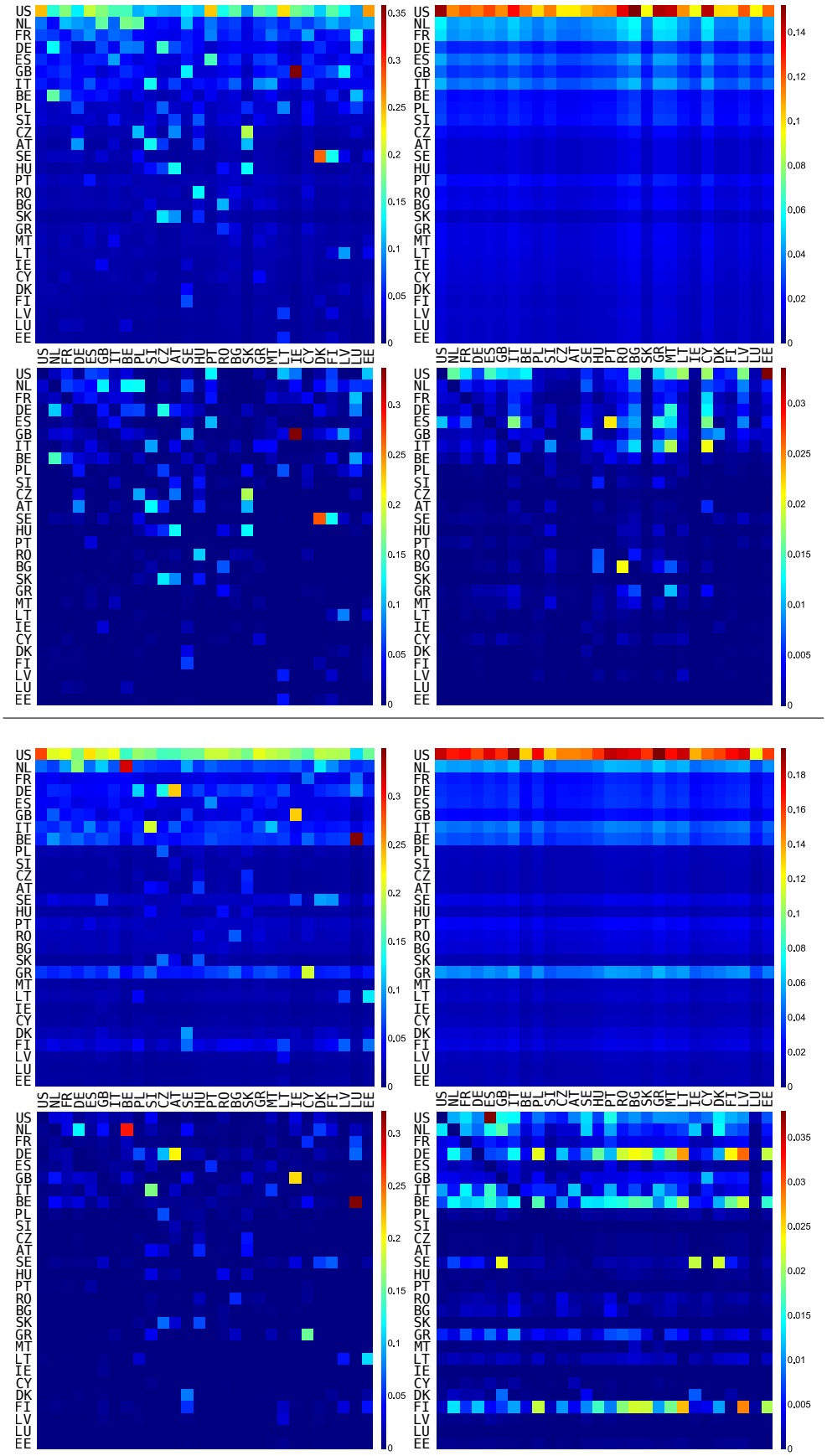

FIGURE A.5 : Matrices de Google réduite associées à G (4 matrices du haut) et G^* (4 matrices du bas), pour l'année 2016 et le pétrole américain. Pour chaque ensemble de 4 matrices, la matrice de Google réduite G_{r} est en haut à gauche et ses composantes sont G_{pr} (en haut à droite), G_{rr} (en bas à gauche) et G_{qrnd} (en bas à droite). Les lignes et colonnes sont classées suivant le classement PageRank de la Table 4.1. D'après [70].

Pays				Exportation			
CC	SC	LT	Name	Total (10^6USD)	Gas (‰)	Petroleum (‰)	Gas & petroleum (‰)
NG	NG	M	Nigeria	34582.35	49.45	910.87	960.32
SA	SA	M	Saudi Arabia	110667.14	34.88	787.04	821.92
RU	RU	M	Russia	225850.70	51.36	447.20	498.56
NR	NR	I	Nauru	17.58	28.38	425.38	453.76
PN	UK	I	Pitcairn Islands	20.02	0.00	448.97	448.97
TL	TL	I	East Timor	155.31	21.89	274.94	296.83
SC	SC	I	Seychelles	520.47	0.02	194.06	194.08
GU	US	I	Guam	68.83	1.60	167.09	168.69
KE	KE	M	Kenya	3507.94	0.65	139.40	140.05
BM	UK	I	Bermuda	189.97	0.00	74.22	74.22
MH	MH	I	Marshall Islands	651.83	0.00	57.53	57.53
KY	UK	I	Cayman Islands	674.49	0.00	46.41	46.41
VA	VA	M	Vatican	3.19	0.00	18.58	18.58
MP	US	I	Northern Mariana Islands	18.97	0.00	10.81	10.81
SH	UK	I	Saint Helena, Ascension and Tristan da Cunha	15.52	0.00	9.99	9.99
TC	UK	I	Turks and Caicos Islands	30.78	9.52	0.01	9.53
AS	US	I	American Samoa	22.63	0.00	8.62	8.62
FK	UK	I	Falkland Islands	136.20	0.00	4.53	4.53
IO	UK	I	British Indian Ocean Territory	3.23	0.00	3.07	3.07
TV	TV	I	Tuvalu	2.12	1.95	0.01	1.95
TK	NZ	I	Tokelau	20.41	0.00	1.48	1.48
SM	SM	M	San Marino	53.24	0.00	0.92	0.92
SB	SB	I	Solomon Islands	193.42	0.00	0.84	0.84
GL	DK	M	Greenland	538.59	0.00	0.26	0.26
PW	PW	I	Palau	22.80	0.00	0.10	0.10
CK	CK	I	Cook Islands	17.01	0.00	0.04	0.04
BT	BT	M	Bhutan	57.88	0.00	0.01	0.01
UM	US	I	United States Minor Outlying Islands	33.55	0.00	0.00	0.00
CX	AU	I	Christmas Island	15.54	0.00	0.00	0.00
PM	FR	I	Saint Pierre and Miquelon	7.80	0.00	0.00	0.00
CC	AU	I	Cocos (Keeling) Islands	6.05	0.00	0.00	0.00
NF	AU	I	Norfolk Island	4.01	0.00	0.00	0.00
GS	UK	I	South Georgia and the South Sandwich Islands	3.50	0.00	0.00	0.00
EH		M	Western Sahara	2.11	0.00	0.00	0.00
AQ		M	Antarctica	2.05	0.00	0.00	0.00
NU	NU	I	Niue	1.88	0.00	0.00	0.00
HM	AU	I	Heard and McDonald Islands	0.95	0.00	0.00	0.00
BV	NO	I	Bouvet Island	0.25	0.00	0.00	0.00

TABLE A.4 : Liste des 38 pays sains en fin de contagion de crise économique τ_∞ avec un seuil de banqueroute $\kappa = 0.1$ pour l'année 2004 (modèle A). La colonne CC représente les codes pays selon le format ISO-3166-2, la colonne SC donne les pays souverain dont les pays font partis, la colonne LT donne le type de terre, avec I pour les îles et M pour les pays continentaux. Enfin, les quatre dernières colonnes donnent les volumes d'export total, en gaz, en pétrole, et en pétrole + gaz. Les pays sont classés suivant leur volume d'export total puis selon le volume d'export gaz + pétrole. D'après [97].

Pays				Exportation			
CC	SC	LT	Name	Total (10^6USD)	Gas (‰)	Petroleum (‰)	Gas & petroleum (‰)
TL	TL	I	East Timor	169.40	816.28	0.28	816.56
RU	RU	M	Russia	570605.26	44.12	550.37	594.49
BV	NO	I	Bouvet Island	41.07	0.00	357.29	357.29
CK	CK	I	Cook Islands	33.11	0.00	239.31	239.31
BM	UK	I	Bermuda	1849.83	3.54	104.77	108.30
UM	US	I	United States Minor Outlying Islands	24.16	0.00	68.68	68.68
SZ	SZL	M	Eswatini	1058.15	0.00	61.57	61.57
GS	UK	I	South Georgia and the South Sandwich Islands	1.62	0.00	21.96	21.96
AD	AD	M	Andorra	175.92	7.70	9.86	17.56
NF	AU	I	Norfolk Island	4.04	0.00	15.57	15.57
EH		M	Western Sahara	11.31	0.00	14.72	14.72
TV	TV	I	Tuvalu	4.02	0.00	11.10	11.10
SH	UK	I	Saint Helena, Ascension and Tristan da Cunha	43.99	0.00	3.38	3.38
NU	NU	I	Niue	9.67	0.00	2.77	2.77
GU	US	I	Guam	77.22	0.00	2.33	2.33
AQ		M	Antarctica	2.55	0.00	1.75	1.75
GL	DK	M	Greenland	753.16	0.00	1.38	1.38
WS	WS	I	Samoa	89.95	0.00	1.28	1.28
WF	FR	I	Wallis and Futuna	18.64	0.00	0.98	0.98
VU	VU	I	Vanuatu	568.61	0.00	0.67	0.67
LS	LS	M	Lesotho	873.69	0.27	0.39	0.66
KI	KI	I	Kiribati	14.02	0.00	0.47	0.47
NR	NR	I	Nauru	126.51	0.00	0.33	0.33
FM	FM	I	Federated States of Micronesia	28.57	0.00	0.13	0.13
AS	US	I	American Samoa	70.09	0.00	0.08	0.08
BT	BT	M	Bhutan	688.82	0.00	0.03	0.03
IO	UK	I	British Indian Ocean Territory	8.25	0.00	0.02	0.02
SB	SB	I	Solomon Islands	383.50	0.00	0.01	0.01
PW	PW	I	Palau	29.04	0.00	0.01	0.01
FK	UK	I	Falkland Islands	196.72	0.00	0.00	0.00
GW	GW	M	Guinea-Bissau	135.22	0.00	0.00	0.00
CX	AU	I	Christmas Island	52.61	0.00	0.00	0.00
CC	AU	I	Cocos (Keeling) Islands	29.68	0.00	0.00	0.00
PM	FR	I	Saint Pierre and Miquelon	17.42	0.00	0.00	0.00
MP	US	I	Northern Mariana Islands	12.64	0.00	0.00	0.00
PN	UK	I	Pitcairn Islands	9.38	0.00	0.00	0.00
VA	VA	M	Vatican	2.52	0.00	0.00	0.00
HM	AU	I	Heard and McDonald Islands	0.53	0.00	0.00	0.00

TABLE A.5 : Liste des 38 pays sains en fin de contagion de crise économique τ_∞ avec un seuil de banqueroute $\kappa = 0.1$ pour l'année 2008 (modèle A). La colonne CC représente les codes pays selon le format ISO-3166-2, la colonne SC donne les pays souverain dont les pays font partis, la colonne LT donne le type de terre, avec I pour les îles et M pour les pays continentaux. Enfin, les quatre dernières colonnes donnent les volumes d'export total, en gaz, en pétrole, et en pétrole + gaz. Les pays sont classés suivant leur volume d'export total puis selon le volume d'export gaz + pétrole. D'après [97].

Pays				Exportation			
CC	SC	LT	Name	Total (10^6USD)	Gas (‰)	Petroleum (‰)	Gas & petroleum (‰)
RU	RU	M	Russia	640181.69	82.16	565.46	647.63
TC	UK	I	Turks and Caicos Islands	94.84	0.00	585.21	585.21
GU	US	I	Guam	146.08	7.94	519.57	527.51
AS	US	I	American Samoa	91.34	6.58	379.91	386.49
BV	NO	I	Bouvet Island	55.78	0.00	288.63	288.63
AQ		M	Antarctica	150.45	0.00	243.27	243.27
KY	UK	I	Cayman Islands	622.35	0.00	19.92	19.92
HM	AU	I	Heard and McDonald Islands	245.23	0.00	11.25	11.25
VA	VA	M	Vatican	7.71	0.00	1.45	1.45
NU	NU	I	Niue	3.59	0.00	1.17	1.17
SH	UK	I	Saint Helena, Ascension and Tristan da Cunha	19.12	0.00	0.99	0.99
MP	US	I	Northern Mariana Islands	3.65	0.01	0.86	0.86
NR	NR	I	Nauru	100.46	0.00	0.58	0.58
GS	UK	I	South Georgia and the South Sandwich Islands	4.03	0.00	0.46	0.46
SB	SB	I	Solomon Islands	572.38	0.00	0.40	0.40
YT	FR	I	Mayotte	28.29	0.00	0.16	0.16
AI	UK	I	Anguilla	10.29	0.00	0.15	0.15
CK	CK	I	Cook Islands	51.41	0.00	0.14	0.14
FO	DK	I	Faroe Islands	1001.96	0.00	0.10	0.10
CX	AU	I	Christmas Island	38.71	0.00	0.05	0.05
IO	UK	I	British Indian Ocean Territory	32.05	0.00	0.05	0.05
UM	US	I	United States Minor Outlying Islands	20.42	0.00	0.03	0.03
TK	NZ	I	Tokelau	43.62	0.00	0.03	0.03
VU	VU	I	Vanuatu	454.29	0.00	0.01	0.01
FK	UK	I	Falkland Islands	210.46	0.01	0.00	0.01
SM	SM	M	San Marino	128.83	0.00	0.00	0.00
ER	ER	M	Eritrea	47.89	0.00	0.00	0.00
PM	FR	I	Saint Pierre and Miquelon	7.12	0.00	0.00	0.00
CC	AU	I	Cocos (Keeling) Islands	7.05	0.00	0.00	0.00
PN	UK	I	Pitcairn Islands	6.71	0.00	0.00	0.00
NF	AU	I	Norfolk Island	3.85	0.00	0.00	0.00
WF	FR	I	Wallis and Futuna	1.30	0.00	0.00	0.00

TABLE A.6 : Liste des 32 pays sains en fin de contagion de crise économique τ_∞ avec un seuil de banqueroute $\kappa = 0.1$ pour l'année 2012 (modèle A). La colonne CC représente les codes pays selon le format ISO-3166-2, la colonne SC donne les pays souverain dont les pays font partis, la colonne LT donne le type de terre, avec I pour les îles et M pour les pays continentaux. Enfin, les quatre dernières colonnes donnent les volumes d'export total, en gaz, en pétrole, et en pétrole + gaz. Les pays sont classés suivant leur volume d'export total puis selon le volume d'export gaz + pétrole. D'après [97].

Pays				Exportation			
CC	SC	LT	Name	Total (10^6USD)	Gas (‰)	Petroleum (‰)	Gas & petroleum (‰)
GS	UK	I	South Georgia and the South Sandwich Islands	0.34	0.00	83.01	83.01
AQ		M	Antarctica	10.28	0.00	11.67	11.67
NU	NU	I	Niue	2.30	0.00	10.50	10.50
FK	UK	I	Falkland Islands	257.30	0.00	5.87	5.87
CC	AU	I	Cocos (Keeling) Islands	4.59	0.00	0.29	0.29
EH		M	Western Sahara	8.92	0.00	0.01	0.01
BV	NO	I	Bouvet Island	0.86	0.00	0.00	0.00
IO	UK	I	British Indian Ocean Territory	20.16	0.00	0.00	0.00
SH	UK	I	Saint Helena, Ascension and Tristan da Cunha	26.64	0.00	0.00	0.00
PN	UK	I	Pitcairn Islands	1.29	0.00	0.00	0.00
HM	AU	I	Heard and McDonald Islands	0.12	0.00	0.00	0.00

TABLE A.7 : Liste des 11 pays sains en fin de contagion de crise économique τ_∞ avec un seuil de banqueroute $\kappa = 0.1$ pour l'année 2016 (modèle A). La colonne CC représente les codes pays selon le format ISO-3166-2, la colonne SC donne les pays souverain dont les pays font partis, la colonne LT donne le type de terre, avec I pour les îles et M pour les pays continentaux. Enfin, les quatre dernières colonnes donnent les volumes d'export total, en gaz, en pétrole, et en pétrole + gaz. Les pays sont classés suivant leur volume d'export total puis selon le volume d'export gaz + pétrole. D'après [97].

Annexe B

Liste des publications

- Célestin Coquidé, José Lages, and Dima L. Shepelyansky. Interdependence of Sectors of Economic Activities for World Countries from the Reduced Google Matrix Analysis of WTO Data. *Entropy* 22(12) :1407, 2020.

- Célestin Coquidé, José Lages, and Dima L. Shepelyansky. Crisis contagion in the world trade network. *Appl. Netw. Sci.* 5, 67, 2020.

- Célestin Coquidé and Włodzimierz Lewoniewski. Novel version of PageRank, CheiRank and 2drank for wikipedia in multilingual network using social impact. In Witold Abramowicz and Gary Klein, editors, *Business Information Systems*, Lecture Notes in Business Information Processing, pages 319–334. Springer International Publishing, 2020.

- Célestin Coquidé, José Lages, and Dima L. Shepelyansky. Contagion in bitcoin networks. In Witold Abramowicz and Rafael Corchuelo, editors, *Business Information Systems Workshops*, Lecture Notes in Business Information Processing, pages 208–219. Springer International Publishing, 2019.

- Coquidé, Célestin, Ermann, Leonardo, Lages, José, and Shepelyansky, Dima L. Influence of petroleum and gas trade on eu economies from the reduced google matrix analysis of un comtrade data. *Eur. Phys. J. B*, 92(8) :171, 2019.

- Célestin Coquidé, José Lages, and Dima L. Shepelyansky. World influence and interactions of universities from Wikipedia networks. *Eur. Phys. J. B*, 92(1) :3, 2019.

- Célestin Coquidé, Bertrand Georgeot and Olivier Giraud. Distinguishing humans from computers in the game of go : A complex network approach. *EPL (Europhysics Letters)*, 119 48001, 2017.

Bibliographie

[1] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. U. Petrop.*, 8 :128–140, 1736.

[2] A. Barrat and M. Weigt. On the properties of small-world network models. *Eur. Phys. J. B*, 13(3) :547–560, February 2000.

[3] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684) :440–442, June 1998.

[4] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3) :215–239, January 1978.

[5] Alex Bavelas. Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, 22(6), November 1950.

[6] Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 1977.

[7] Paul Erdős and Alfréd Rényi. On random graphs, i. *Publ. Math. Debrecen*, 6 :290–297, 1959.

[8] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publication of the mathematical institute of the hungarian academy of sciences*, 1960.

[9] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794) :378–382, July 2000.

[10] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.*, 85(21), November 2000.

[11] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14) :3200–3203, April 2001.

[12] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439) :509–512, October 1999.

[13] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A : Statistical Mechanics and its Applications*, 272(1) :173–187, October 1999.

[14] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the Internet topology, August 1999.

[15] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833) :41–42, May 2001.

[16] S. Redner. How popular is your paper ? An empirical study of the citation distribution. *Eur. Phys. J. B*, 4(2) :131–134, July 1998.

[17] Stanley Milgram. The small world problem. *Psychology Today*, 1(1) :61–67, 1967.

[18] Emanuele Cozzo, Guilherme Ferraz de Arruda, Francisco A. Rodrigues, and Yamir Moreno. Multilayer Networks : Metrics and Spectral Properties. In Antonios Garas, editor, *Interconnected Networks*, Understanding Complex Systems, pages 17–35. Springer International Publishing, Cham, 2016.

[19] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the World-Wide Web. *Nature*, 401(6749) :130–131, September 1999.

[20] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., USA, 1986.

[21] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking : Bringing Order to the Web., November 1999.

[22] A.M. Langville and C.D. Meyer. *Google's PageRank and beyond : the science of search engine ranking*. Princeton University Press, 2006.

[23] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7) :107–117, April 1998.

[24] A. D. Chepelianskii. Towards physical laws for software architecture. http://arxiv.org/abs/1003.5455, 2010.

[25] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky. Two-dimensional ranking of Wikipedia articles. *Eur. Phys. J. B*, 77(4) :523–531, October 2010.

[26] L. Ermann, A. D. Chepelianskii, and D. L. Shepelyansky. Toward two-dimensional search engines. *J. Phys. A : Math. Theor.*, 45(27) :275101, June 2012.

[27] K. M. Frahm, B. Georgeot, and D. L. Shepelyansky. Universal emergence of PageRank. *J. Phys. A : Math. Theor.*, 44(46), November 2011.

[28] Leonardo Ermann, Klaus M. Frahm, and Dima L. Shepelyansky. Spectral properties of Google matrix of Wikipedia and other networks. *Eur. Phys. J. B*, 86(5) :193, April 2013.

[29] José Lages. *Classical and quantum complex systems*. Habilitation à diriger des recherches, COMUE Université Bourgogne Franche-Comté, November 2019.

[30] K. M. Frahm and D. L. Shepelyansky. Reduced Google matrix. https://arxiv.org/abs/1602.02394v1, February 2016. arXiv :1602.02394 [physics.soc-ph].

[31] Klaus Frahm, Katia Jaffres-Runser, and Dima L. Shepelyansky. Wikipedia mining of hidden links between political leaders. *European Physical Journal B : Condensed Matter and Complex Systems*, 89 :269, December 2016.

[32] V. V Sokolov and V. G Zelevinsky. Collective dynamics of unstable quantum states. *Annals of Physics*, 216(2) :323–350, June 1992.

[33] C. W. J. Beenakker. Random-matrix theory of quantum transport. *Rev. Mod. Phys.*, 69(3), July 1997.

[34] Thomas Guhr, Axel Müller–Groeling, and Hans A. Weidenmüller. Random-matrix theories in quantum physics : common concepts. *Physics Reports*, 299(4) :189–425, June 1998.

[35] Pierre Gaspard. Quantum chaotic scattering. *Scholarpedia*, 9(6) :9806, June 2014.

[36] Rodolfo A. Jalabert. Mesoscopic transport and quantum chaos. *Scholarpedia*, 11(1) :30946, January 2016.

[37] Ellen Hazelkorn. *Rankings and the Reshaping of Higher Education - The Battle for World-Class Excellence*. Palgrave macmillan edition, 2015.

[38] Heike Jöns and Michael Hoyler. Global geographies of higher education : The perspective of world university rankings. 46 :45–59, May 2013.

[39] Andrejs Rauhvargers. *Global university rankings and their impact : report II*. European University Association, 2013.

[40] Domingo Docampo and Lawrence Cram. On the internal dynamics of the shanghai ranking. 98(2) :1347–1366, February 2014.

[41] Young-Ho Eom, Klaus M. Frahm, András Benczúr, and Dima L. Shepelyansky. Time evolution of Wikipedia network ranking. *Eur. Phys. J. B*, 86(12) :492, December 2013.

[42] José Lages, Antoine Patt, and Dima L. Shepelyansky. Wikipedia ranking of world universities. *Eur. Phys. J. B*, 89(3) :69, March 2016.

[43] Young-Ho Eom, Pablo Aragón, David Laniado, Andreas Kaltenbrunner, Sebastiano Vigna, and Dima L. Shepelyansky. Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions. *PLOS ONE*, 10(3) :e0114825, March 2015.

[44] Samer El Zant, Klaus M. Frahm, Katia Jaffrès-Runser, and Dima L. Shepelyansky. Analysis of world terror networks from the reduced google matrix of wikipedia. *Eur. Phys. J. B*, 91(1) :7, January 2018.

[45] José Lages, Dima L. Shepelyansky, and Andrei Zinovyev. Inferring hidden causal relations between pathway members using reduced google matrix of directed biological networks. *PLoS ONE*, 13(1) :e0190812, January 2018.

[46] Célestin Coquidé, José Lages, and Dima L. Shepelyansky. World influence and interactions of universities from Wikipedia networks. *Eur. Phys. J. B*, 92(1) :3, January 2019.

[47] Academic ranking of world universities edition 2017. http://www.shanghairanking.com/. Accès : Juillet 2018.

[48] Jenks George F. The data model concept in statistical mapping. *International Yearbook of Cartography*, 7 :186–190, 1967.

[49] Samer El Zant, Katia Jaffrès-Runser, and Dima L. Shepelyansky. Capturing the influence of geopolitical ties from wikipedia with reduced google matrix. *PLOS ONE*, 13(8) :e0201397, August 2018.

[50] Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y Kenett, H Eugene Stanley, and Tobias Preis. Quantifying wikipedia usage patterns before stock market moves. *Scientific reports*, 3 :1801, May 2013.

[51] Márton Mestyán, Taha Yasseri, and János Kertész. Early prediction of movie box office success based on wikipedia activity big data. *PloS one*, 8(8) :e71226, August 2013.

[52] Muhammad Hassan Latif and Hammad Afzal. Prediction of movies popularity using machine learning techniques. *International Journal of Computer Science and Network Security (IJCSNS)*, 16(8) :127, December 2016.

[53] Pejman Khadivi and Naren Ramakrishnan. Wikipedia in the tourism industry : forecasting demand and modeling usage behavior. In *Twenty-Eighth IAAI Conference*, 2016.

[54] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia : Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3 :3415, December 2013.

[55] Abeer ElBahrawy, Laura Alessandretti, and Andrea Baronchelli. Wikipedia and cryptocurrencies : Interplay between collective attention and market performance. *Front. Blockchain*, 2, October 2019.

[56] Kyle S Hickmann, Geoffrey Fairchild, Reid Priedhorsky, Nicholas Generous, James M Hyman, Alina Deshpande, and Sara Y Del Valle. Forecasting the 2013–2014 influenza season using wikipedia. *PLoS computational biology*, 11(5) :e1004239, May 2015.

[57] Amy Zhao Yu, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A. Hidalgo. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Sci Data*, 3(1) :1–16, January 2016.

[58] Włodzimierz Lewoniewski, Krzysztof Węcel, and Witold Abramowicz. Multilingual ranking of wikipedia articles with quality and popularity assessment in different topics. *Computers*, 8(3) :60, August 2019.

[59] Włodzimierz Lewoniewski. Measures for quality assessment of articles and infoboxes in multilingual wikipedia. In *International Conference on Business Information Systems*, pages 619–633. Springer, 2019.

[60] Guillaume Rollin, José Lages, and Dima L. Shepelyansky. Wikipedia network analysis of cancer interactions and world influence. *PLOS ONE*, 14(9) :e0222508, September 2019.

[61] Guillaume Rollin, José Lages, and Dima L. Shepelyansky. World Influence of Infectious Diseases From Wikipedia Network Analysis. *IEEE Access*, 7 :26073–26087, February 2019.

[62] Lisette Espín-Noboa, Florian Lemmerich, Simon Walk, Markus Strohmaier, and Mark Musen. HopRank : How Semantic Structure Influences Teleportation in PageRank (A Case Study on BioPortal). In *The World Wide Web Conference*, WWW '19, pages 2708–2714, New York, NY, USA, 2019. ACM. event-place : San Francisco, CA, USA.

[63] Patrick Gildersleve and Taha Yasseri. Inspiration, captivation, and misdirection : Emergent properties in networks of online navigation. In Sean Cornelius, Kate Coronges, Bruno Gonçalves, Roberta Sinatra, and Alessandro Vespignani, editors, *Complex Networks IX*, Springer Proceedings in Complexity, pages 271–282. Springer International Publishing, 2018.

[64] Célestin Coquidé and Włodzimierz Lewoniewski. Novel version of PageRank, CheiRank and 2drank for wikipedia in multilingual network using social impact. In Witold Abramowicz and Gary Klein, editors, *Business Information Systems*, Lecture Notes in Business Information Processing, pages 319–334. Springer International Publishing, 2020.

[65] Paul Krugman. *International Economics : Theory and Policy, Global Edition*. Pearson, 11 edition, 2018.

[66] Sergey Dorogovtsev. *Lectures on Complex Networks*. Oxford University Press, 2010.

[67] L. Ermann and D.L. Shepelyansky. Google matrix of the world trade network. *Acta Physica Polonica A*, 120(6) :A–158–A–171, December 2011.

[68] Leonardo Ermann and Dima L. Shepelyansky. Google matrix analysis of the multiproduct world trade network. *Eur. Phys. J. B*, 88(4) :84, April 2015.

[69] Leonardo Ermann, Klaus M. Frahm, and Dima L. Shepelyansky. Google matrix analysis of directed networks. *Rev. Mod. Phys.*, 87(4) :1261–1310, November 2015.

[70] Coquidé, Célestin, Ermann, Leonardo, Lages, José, and Shepelyansky, Dima L. Influence of petroleum and gas trade on eu economies from the reduced google matrix analysis of un comtrade data. *Eur. Phys. J. B*, 92(8) :171, August 2019.

[71] Denis Demidov, Klaus M. Frahm, and Dima L. Shepelyansky. What is the central bank of wikipedia ? *Physica A : Statistical Mechanics and its Applications*, 542 :123199, March 2020.

[72] Guillaume Rollin, José Lages, Tatiana S. Serebriyskaya, and Dima L. Shepelyansky. Interactions of pharmaceutical companies with world countries, cancers and rare diseases from wikipedia network analysis. *PLOS ONE*, 14(12) :e0225500, December 2019.

[73] M. Ángeles Serrano, Marián Boguñá, and Alessandro Vespignani. Patterns of dominant flows in the world trade web. *J Econ Interac Coord*, 2(2) :111–124, December 2007.

[74] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. World-trade web : Topological properties, dynamics, and evolution. *Phys. Rev. E*, 79(3) :036115, March 2009.

[75] Jiankui He and Michael W. Deem. Structure and response in the world trade network. *Phys. Rev. Lett.*, 105(19) :198701, November 2010.

[76] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. The evolution of the world trade web : a weighted-network analysis. *J Evol Econ*, 20(4) :479–514, August 2010.

[77] Matteo Barigozzi, Giorgio Fagiolo, and Diego Garlaschelli. Multinetwork of international trade : A commodity-specific analysis. *Phys. Rev. E*, 81(4) :046104, April 2010.

[78] Luca De Benedictis and Lucia Tajoli. The world trade network. *The World Economy*, 34(8) :1417–1454, August 2011.

[79] Vivek Kandiah, Hubert Escaith, and Dima L. Shepelyansky. Google matrix of the world network of economic activities. *Eur. Phys. J. B*, 88(7) :186, July 2015.

[80] Coquidé C., Lages J., and Shepelyansky D.L. Interdependence of sectors of economic activities for world countries from the reduced google matrix analysis of wto data. *Entropy*, 22(12) :1407, December 2020.

[81] Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing : From Statistical Physics to Risk Management*. Cambridge University Press, 2^{nd} edition, 2003.

[82] Michael C. Münnix, Rudi Schäfer, and Thomas Guhr. A random matrix approach to credit risk. *PLOS ONE*, 9(5) :e98030, May 2014.

[83] Marco Bardoscia, Giacomo Livan, and Matteo Marsili. Statistical mechanics of complex economies. *J. Stat. Mech.*, 2017(4) :043401, April 2017.

[84] Paul J. Flory. Molecular Size Distribution in Three Dimensional Polymers. I. Gelation1. *J. Am. Chem. Soc.*, 63(11), November 1941.

[85] Walter H. Stockmayer. Theory of Molecular Size Distribution and Gel Formation in Branched Polymers II. General Cross Linking. *J. Chem. Phys.*, 12(4), April 1944.

[86] S. R. Broadbent and J. M. Hammersley. Percolation processes : I. Crystals and mazes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 53(3) :629–641, July 1957.

[87] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility : Percolation on random graphs. *Phys. Rev. Lett.*, 85 :5468–5471, Dec 2000.

[88] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6, 1995.

[89] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Pseudofractal scale-free web. *Phys. Rev. E*, 65(6), June 2002.

[90] Jiyoung Woo and Hsinchun Chen. Epidemic model for information diffusion in web forums : experiments in marketing exchange and political dialog. *SpringerPlus*, 5(1) :66, January 2016.

[91] Luís M. A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chávez. The power of a good idea : Quantitative modeling of the spread of ideas from epidemiological models. *Physica A : Statistical Mechanics and its Applications*, 364 :513–536, May 2006.

[92] Michael Montagne. The Social Epidemiology of International Drug Trafficking : Comparison of Source of Supply and Distribution Networks*. *International Journal of the Addictions*, 25(5), January 1990.

[93] Prasanna Gai and Sujit Kapadia. Contagion in financial networks. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 466(2120) :2401–2423, August 2010.

[94] Matthew Elliott, Benjamin Golub, and Matthew O. Jackson. Financial Networks and Contagion. *American Economic Review*, 104(10) :3115–3153, October 2014.

[95] Kilian Fink, Ulrich Krüger, Barbara Meller, and Lui-Hsian Wong. The credit quality channel : Modeling contagion in the interbank market. *Journal of Financial Stability*, 25 :83–97, August 2016.

[96] Tsuyoshi Deguchi, Katsuhide Takahashi, Hideki Takayasu, and Misako Takayasu. Hubs and authorities in the world trade network using a weighted HITS algorithm. *PLOS ONE*, 9(7) :e100338, July 2014.

[97] Célestin Coquidé, José Lages, and Dima L. Shepelyansky. Crisis contagion in the world trade network. *Appl Netw Sci*, 5(1) :1–20, September 2020.

[98] Satoshi Nakamoto. Bitcoin : A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf.

[99] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In Ahmad-Reza Sadeghi, editor, *Financial Cryptography and Data Security*, Lecture Notes in Computer Science, pages 6–24. Springer, 2013.

[100] Alex Biryukov, Dmitry Khovratovich, and Ivan Pustogarov. Deanonymisation of clients in bitcoin p2p network. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 15–29. Association for Computing Machinery, 2014.

[101] John Bohannon. The bitcoin busts. *Science*, 351(6278) :1144–1146, March 2016.

[102] Leonardo Ermann, Klaus M. Frahm, and Dima L. Shepelyansky. Google matrix of Bitcoin network. *Eur. Phys. J. B*, 91(6) :127, June 2018.

[103] Ivan Brugere. Bitcoin transaction networks extraction. https://github.com/ivanbrugere/Bitcoin-Transaction-Network-Extraction, 2013. Accès : Octobre 2017.

[104] Célestin Coquidé, José Lages, and Dima L. Shepelyansky. Contagion in bitcoin networks. In Witold Abramowicz and Rafael Corchuelo, editors, *Business Information Systems Workshops*, Lecture Notes in Business Information Processing, pages 208–219. Springer International Publishing, 2019.

ECOLE DOCTORALE
CARNOT - PASTEUR

Titre : Analyse de réseaux complexes réels via des méthodes issues de la matrice de Google
Mots clés : Réseaux complexes, Big data, Matrice de Google réduite, Contagion, Chaînes de Markov.

Résumé : Dans une époque où Internet est de plus en plus utilisé et où les populations sont de plus en plus connectées à travers le monde, notre vie quotidienne est grandement facilitée. Un domaine scientifique très récent, la science des réseaux, dont les prémices viennent des mathématiques et plus précisément de la théorie des graphes a justement pour objet d'étude de tels systèmes complexes. Un réseau est un objet mathématique fait de nœuds et de connexions entre ces nœuds. Dans la nature, on retrouve une multitude de phénomènes pouvant être vus ainsi, par exemple, le mycélium qui est un réseau souterrain capable d'avoir accès à courtes et moyennes distances aux ressources organiques propices à sa survie, ou bien encore le réseau vasculaire sanguin. À notre échelle, il existe aussi des réseaux dont nous sommes les nœuds. Dans cette thèse, nous allons nous intéresser aux réseaux réels, réseaux construits à partir de banques de données, afin de les analyser, puis d'extraire des informations difficilement accessibles dans des réseaux pouvant contenir, parfois, des millions de nœuds et cent fois plus de connexions. Les réseaux étudiés sont aussi dirigés, autrement dit, les liens ont une direction. On représente une marche aléatoire dans un tel réseau à l'aide d'une matrice stochastique appelée matrice de Google. Elle permet notamment de mesurer l'importance des nœuds d'un réseau à l'aide de son vecteur propre dominant, le vecteur PageRank. À partir de la matrice de Google, nous pouvons aussi construire une matrice de Google de taille réduite représentant toutes les connexions entre les éléments d'un sous-réseau d'intérêt, le réseau réduit, mais aussi et surtout de pouvoir quantifier les connexions indirectes entre ces nœuds, obtenues par diffusion à travers tout le reste du réseau. Cette matrice de Google réduite permet, en plus de réduire considérablement la taille du réseau et de la matrice de Google associée, d'extraire des liens indirects non-triviaux entre les nœuds d'intérêts, appelés *liens cachés*. À l'aide d'outils construits à partir de la matrice de Google, notamment la matrice de Google réduite, nous allons, à travers le réseau Wikipédia, identifier les interactions entre les universités et leurs influences sur le monde, et utiliser des données de comportements utilisateurs Wikipédia afin de mesurer les tendances culturelles actuelles. À partir de réseaux économiques, nous allons mesurer la résistance économique de l'Union européenne face à une hausse des prix liés au pétrole et au gaz extérieurs, mais aussi établir les interdépendances entre secteurs de production propres à quelques puissances économiques comme les États-Unis ou encore la Chine. Enfin, nous allons établir un modèle de propagation de crise économique et l'appliquer au réseau du commerce international et au réseau de transactions de Bitcoin.

Title: Methods from The Google Matrix for Real Complex Networks Analysis
Keywords: Complex Networks, Big Data, Reduced Google Matrix, Contagion, Markov Chains

Abstract: In a current period where people use more and more the Internet and are connected worldwide, our lives become easier. The Network science, a recent scientific domain coming from graph theory, handle such connected complex systems. A network is a mathematical object consisting in a set of interconnected nodes and a set of links connecting them. We find networks in nature such as networks of mycelium which grow underground and are able to feed their cells with organic nutrients located at low and long range from them, as well as the circulation system transporting blood throughout the human body. Networks also exist at a human scale where humans are nodes of such networks. In this thesis we are interested in what we call real complex networks which are networks constructed from databases. We can extract information which is normally hard to get since such a network might contain one million of nodes and one hundred times more links. Moreover, networks we are going to study are directed meaning that links have a direction. One can represent a random walk through a directed network with the use of the so-called Google matrix. The PageRank is the leading eigenvector associated to this stochastic matrix and allows us to measure nodes importance. We can also build a smaller Google matrix based on the Google matrix and a subregion of the network. This reduced Google matrix allows us to extract every existing links between the nodes composing the subregion of interest as well as all possible indirect connections between them by spreading through the entire network. With the use of tools developed from the Google matrix, especially the reduced Google matrix, considering the network of Wikipedia's articles we have identified interactions between universities of the world as well as their influence. We have extracted social trends by using data related to actual Wikipedia's users behaviour. Regarding the World Trade Network, we were able to measure economic response of the European Union to external petroleum and gas price variation. Regarding the World Network of economical activities we have figured out interdependence of sectors of production related to powerhouse such as The United States of America and China. We also built a crisis contagion model we applied on the World Trade Network and on the Bitcoin transactions Network.

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon