

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Une analyse du réseau de participations des entreprises belges

Vandemoortele, Gauthier

Award date:
2016

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

UNIVERSITÉ DE NAMUR
Faculté d'informatique
Année académique 2015–2016

**Une analyse du réseau de participations
des entreprises belges**

Gauthier Vandemoortele



Promoteur : _____ (Signature pour approbation du dépôt - REE art. 40)
Benoît Frenay

Co-promoteur : Renaud Lambiotte

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Informatiques.

Table des matières

Résumé	6
Remerciements	7
1 Introduction	8
2 Les données et leur contexte	11
2.1 Présentation générale	11
2.2 Fondements réglementaires, terminologie et conventions	12
2.3 Notions comptables	13
2.4 Origine et caractéristiques des données	15
2.5 Sémantique des liaisons	22
3 État de l’art sur les études de réseaux de participations	29
3.1 La séparation de la propriété et du contrôle	30
3.2 L’évaluation du contrôle	31
3.3 Les réseaux financiers comme cas particulier de réseau complexe	36
3.3.1 Les réseaux économiques et financiers en général	37
3.3.2 Les réseaux de participations	39
3.3.3 “ <i>The Network of Global Corporate Control</i> ”	42
4 Analyse	45
4.1 Distribution des composantes connexes	45
4.2 Représentation matricielle du réseau	47
4.3 Distribution des degrés	52
4.4 Assortativité	57
4.5 La structure en nœud papillon	57
5 Recherche de sous-ensembles significatifs	61
5.1 Terminologie	61
5.2 Rôles <i>vs</i> communautés	63
5.3 Analogie avec les réseaux trophiques	67
5.4 La similarité de rôle selon Cooper et Barahona	71
5.5 Application au réseau des participations	73
5.5.1 Sélection du nombre de clusters	78
5.5.2 Validation du nombre de clusters	88
5.5.3 Représentation condensée du réseau	91

6	Conclusions	100
7	Annexes	102
7.1	Exemple de calcul de participation	103
7.2	Le phénomène de «petit monde»	104
7.3	La structure en <i>bow-tie</i>	105
7.4	Groupes de sociétés, précisions terminologiques	107
7.5	Cooper et Barahona, visualisations	108

Table des figures

1.1	Le syndrome de la pelote de laine	9
2.1	Distinction entre les types de participations	12
2.2	Mention de participations dans les comptes annuels	14
2.3	Intégration des flux financiers résultant des participations	16
2.4	Nombre et composition des composantes du réseau	20
2.5	Distribution des participations selon leur poids	21
2.6	Exemple élémentaire de chaîne de participations.	23
2.7	Exemple de distinction entre propriété et contrôle	24
2.8	Exemple de structure d'actionnariat avec cycle	25
2.9	Exemple de structure de propriétés croisées de firmes islandaises	26
2.10	Structure de propriété et de contrôle cyclique	27
3.1	Exemple de calcul de participation intégrée (1)	34
3.2	Exemple de calcul de contrôle intégré	35
3.3	Graphes <i>one-mode</i> déduits d'un graphe bipartite	38
4.1	Caractéristiques des composantes pour divers réseaux	45
4.2	Distribution des participations selon la taille de la composante	48
4.3	Distribution des composantes	49
4.4	Représentation matricielle du réseau de Souma et al.	50
4.5	Représentation matricielle du réseau des sociétés belges	51
4.6	Sections de premier niveau des codes NACE	52
4.7	Transferts financiers entre secteurs NACE	53
4.8	Distribution des degrés dans le réseau observé par Souma et al.	54
4.9	Distribution de la <i>force</i> dans le réseau belge	55
4.10	Distribution des degrés dans le réseau belge	56
4.11	Corrélation de degré (assortativité)	59
4.12	Les 20 premières structures en nœud papillon présentes dans le réseau	60
5.1	Représentation des rôles et communautés dans un réseau de participations	64
5.2	Équivalence régulière et coloration	67
5.3	«Les gros poissons mangent les petits»	70
5.4	L'algorithme de Cooper et Barahona appliqué à un petit réseau	73
5.5	L'algorithme de Cooper et Barahona appliqué à un réseau trophique	74
5.6	Exemple de rôles différents selon le poids des chemins	75
5.7	Calcul du barycentre d'un profil	76

5.8	Distribution des sommets selon leur barycentre	77
5.9	Clustering hiérarchique du sous-ensemble SC	79
5.10	Profils moyens des clusters obtenus par k-means	80
5.11	PCA sur les données du réseau entier	83
5.12	Clustering hiérarchique du sous-ensemble SC (log.)	84
5.13	Profils pour le réseau entier, 6 clusters	85
5.14	Profils pour le réseau entier, 6 clusters	86
5.15	Évaluation du meilleur nombre de clusters	87
5.16	Positionnement des sources et des puits	91
5.17	Matrice d'adjacence du réseau entier, données brutes, 9 clusters.	93
5.18	Matrice d'adjacence du réseau entier, «centré-mis à l'échelle» 8 clusters.	94
5.19	Matrice d'adjacence du réseau entier, normalisé, 8 clusters.	95
5.20	Matrice d'adjacence du réseau entier, traitement logarithmique, 7 clusters.	96
5.21	Visualisations du graphe image avec Graphviz	97
5.22	Visualisations du graphe image avec Graphviz	98
5.23	Visualisations du graphe image avec Graphviz	99
7.1	Exemple de calcul de contrôle intégré	103
7.2	Représentation originale de la structure en bow-tie d'internet	105
7.3	Algorithme de Cooper et Barahona, $\beta = 1.0$	108
7.4	Algorithme de Cooper et Barahona, $\beta = 0.6$	109
7.5	Algorithme de Cooper et Barahona, $\beta = 0.4$	109
7.6	Algorithme de Cooper et Barahona, $\beta = 0.2$	110
7.7	Algorithme de Cooper et Barahona, $\beta = 0.05$	110

Résumé

Les liens de participations entre les sociétés belges figurant dans les données des comptes annuels déposés auprès de la Centrale des Bilans de la Banque Nationale permettent de constituer un réseau de grande taille (de l'ordre de 85000 sommets et 70000 liaisons). L'analyse de ses caractères topologiques (distributions des composantes, assortativité des degrés, etc.) a mis en évidence son caractère «sans échelle», en phase avec les résultats d'analyses similaires sur des données de même nature. Un procédé d'extraction de rôles expérimenté auparavant sur des réseaux trophiques a permis, moyennant adaptation, d'effectuer un clustering vérifiant en partie son caractère hiérarchique et d'en obtenir une représentation condensée.

Mots clés participations financières, groupes de sociétés, réseaux complexes, clustering, apprentissage automatique.

Abstract

The ownership relations between belgian companies, extracted from their annual accounts filed with the Central Balance Sheet Office, form a very large network of approximately 85,000 vertices and 70,000 edges. The analysis of its topological properties (distribution of connected components, assortativity of degrees and so on) highlights a scale-free behavior in line with the results of similar studies on data of the same nature. A method for grouping nodes based on the roles of the nodes, formerly tested on food-webs, is adapted to fit the distinctive features of the shareholding network; it allows to perform a clustering that points out its hierarchical nature and to display a condensed view of it.

Mots clés shareholding relations, business groups, complex networks, clustering, machine learning.

Remerciements

Je tiens d'abord à remercier (par ordre d'apparition. . .) mes promoteurs, Renaud Lambiotte pour son approche décomplexée du domaine intimidant des graphes et des réseaux, et Benoît Frénay qui m'a fourni les clés de compréhension de l'apprentissage automatique.

Je remercie sans pouvoir les citer tous les membres du personnel académique et administratif de l'UNamur pour offrir à ceux qui savent la saisir la chance de pouvoir reprendre le chemin de l'instruction malgré le fait d'être déjà absorbé par la vie professionnelle.

Je remercie mes proches qui ont supporté de me voir si peu disponible pendant ce long parcours d'études.

Enfin, je remercie les membres souvent anonymes de la communauté des développeurs de logiciels libres dont les efforts désintéressés ont apporté à l'humanité ce capital d'une portée inestimable.

1 Introduction

Le choix du sujet de ce mémoire s'est opéré de manière assez atypique. Il trouve son origine dans l'intérêt qu'avait éveillé chez moi le cours de théorie de graphes de M. Lambiotte ainsi que dans l'expérience enrichissante qu'avait représentée l'étude réalisée dans le cadre du cours de Projet personnel du cursus de bachelier. Sous la conduite du même professeur, elle avait porté sur le «petit monde», un phénomène observé dans les réseaux complexes et affectant de manière transversale des domaines aussi variés que la technologie, les sciences humaines, la biologie. Cette étude a constitué une première confrontation au passage de la théorie à l'application, de la théorie des graphes à celle des réseaux complexes ; les concepts parfois difficiles élaborés d'après des modèles abstraits se sont trouvés éclairés par le fait de les voir à l'œuvre dans des domaines familiers et inattendus : les relations entre les mots d'un dictionnaire, les liens entre les pages de sites webs¹...

Tout naturellement, c'est donc auprès du même professeur que j'ai sollicité un sujet de mémoire et c'est avec un plaisir mêlé d'appréhension que j'ai reçu sa proposition de trouver un sujet d'étude de mon choix permettant de poursuivre les investigations menées jusque là.

Comme j'étais alors affecté au service de la Centrale des Bilans de la Banque Nationale, j'avais accès dans le cadre de mon travail à un vaste ensemble de données comptables recueillies auprès des sociétés belges. Parmi celles-ci figurent les données relatives aux participations des entreprises et j'avais à l'occasion mis à profit les notions élémentaires acquises au cours de théorie des graphes pour en extraire des vues ponctuelles de groupes de sociétés².

La nature des données des participations est propre à susciter la curiosité. Comme il sera expliqué, ces relations de nature plurielle représentent à la fois des liens de pouvoir et des transferts de richesse ; elles sont la clé de compréhension des phénomènes de concentration économique qui transforment notre société et qui sont généralement ressentis comme une menace. Elles sont susceptibles d'apporter des éléments de réponse à une question élémentaire mais fondamentale : «Qui contrôle qui ?»

Bien qu'elle soit une science encore jeune, l'étude des réseaux complexes a fourni des outils permettant d'appréhender des problèmes de cet ordre ; elle a souvent permis de

1. Une part importante de cette étude a consisté à collecter des jeux de données réelles pour vérifier de quelle façon le phénomène y était observable.

2. Ces données sont en fait peu utilisées au quotidien dans ce service et, faute d'une application dédiée, leur consultation ne permet pas d'explorer récursivement les structures complexes de participations.

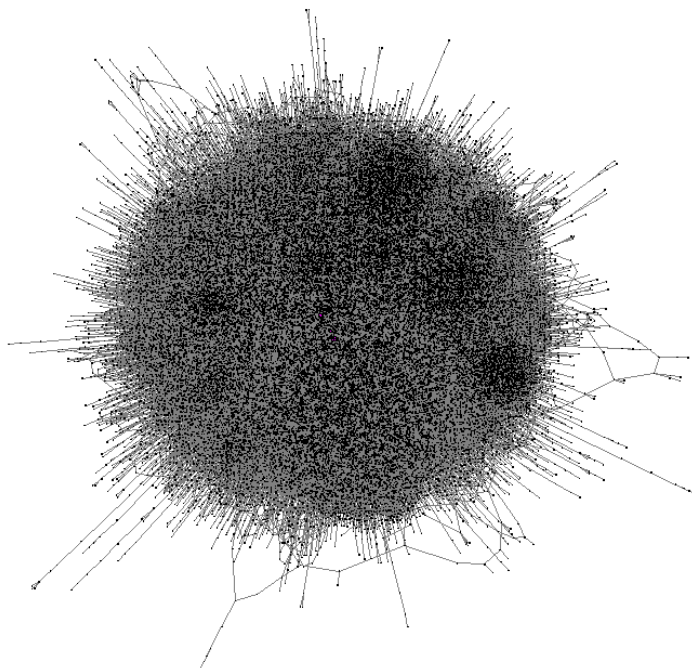


FIGURE 1.1 – Face à un réseau complexe de grande taille, les praticiens sont confrontés au syndrome de la pelote de laine (*hairball*). Comment en obtenir une réduction qui en révèle la structure ? (illustration *Linkurious SAS*©)

dépasser une vision statique et «plate» des données et de révéler leur dynamique, leur évolution, leur profondeur. Le jeu de données des participations apparaissait donc comme un candidat digne d'intérêt pour la mettre en œuvre.

Faute d'un cadre limitatif, la mise en train de ce travail s'est cependant révélée laborieuse. La recherche de la littérature existante a parfois suscité un sentiment d'égarement. Ce qu'on peut qualifier de «dimension réseau» de ce type de données, c.-à-d. la prise en compte de relations à plusieurs niveaux, a en effet été peu – ou tardivement – explorée ; les données de participations ne sont généralement pas accessibles librement³ ou lorsqu'elles le sont, demandent un travail décourageant avant d'être exploitables. En même temps, une littérature considérable a été consacrée à ce qui relève de la gouvernance d'entreprise, c.-à-d. aux implications de la structure de l'actionnariat envers les différentes parties prenantes. L'examen de cette littérature a demandé un temps assez important alors que sa contribution dans le cadre de l'orientation qu'a pris ce mémoire s'est avérée au final assez limitée⁴.

3. Je précise à ce sujet que l'accord pour l'utilisation de ces données s'est accompagné d'un engagement à ne pas les diffuser, celles-ci faisant l'objet d'une exploitation commerciale.

4. Cette remarque doit cependant être nuancée. Le mot d'ordre "*Know your data !*" est populaire dans le domaine de la fouille de données et effectivement, la familiarisation avec les concepts de la finance d'entreprise n'a pas été sans intérêt.

C'est au cours d'une discussion informelle avec M. Lambiotte qu'une impulsion dans une orientation plus précise a été donnée à ce travail : suite à ma réflexion que parmi les catégories les plus fréquemment citées, les réseaux qui se rapprochaient le plus de ceux de participations étaient les réseaux trophiques, M. Lambiotte m'a suggéré la lecture d'un article présentant un procédé de clustering appliqué sur ceux-ci avec succès.

Dès lors, des objectifs plus circonscrits se sont trouvés définis, consistant à vérifier si cette méthode de clustering permettrait de mettre en évidence le caractère hiérarchique du réseau ou, à tout le moins, d'en dégager un modèle simplifié qui faciliterait la compréhension de sa structure. Pour y parvenir, il convenait d'implémenter la méthode décrite, de l'adapter pour qu'elle exprime au mieux les spécificités des données, de trouver des critères permettant d'en déterminer les meilleurs paramètres.

À l'exception de la section 2, le plan de ce mémoire correspond approximativement au déroulement chronologique du cheminement du travail.

Dans la section 2, nous présenterons le contexte réglementaire et professionnel relatif à la collecte des données utilisées. Nous les compléterons par les notions élémentaires de comptabilité utiles à la compréhension de la nature des liaisons présentes dans le réseau que nous construisons, puis par l'exposé des caractéristiques générales des données. Nous avons jugé utile de nous étendre sur la sémantique des liens de participations ; le caractère pluriel de leurs significations est en effet assez particulier parmi les réseaux offerts à la curiosité du praticien.

La section 3 est dédiée à un état de l'art fort large montrant la variété des sujets associés à ce type de données (gouvernance d'entreprise, détermination du degré de contrôle et analyse des réseaux proprement dite).

La section 4 expose les résultats de l'examen de quelques aspects des données des sociétés belges, en les recoupant avec celles d'analyses antérieures (assez peu nombreuses) effectuées dans la perspective de leur dimension réseau.

La section 5 est consacrée à l'implémentation du clustering ; elle introduit la terminologie et les concepts théoriques de la recherche de rôles, puis ceux particuliers aux réseaux trophiques avant d'établir l'analogie avec les réseaux de participations. La description de la méthode de Cooper et Barahona, l'exposé des modifications qui y sont apportées ainsi que des résultats expérimentaux terminent cette section.

2 Les données et leur contexte

2.1 Présentation générale

Le capital social d'une entreprise représente les ressources financières mises à sa disposition par les actionnaires lors de sa création ou lors d'augmentations de capital ultérieures. Chaque associé ayant contribué au capital se voit délivrer un certain nombre d'actions (aujourd'hui sous une forme dématérialisée) qui représentent sa participation, c.-à-d. la fraction correspondante du capital¹. Cette participation confère à son détenteur la possibilité d'influer sur la gestion de l'entreprise au moyen de droits de votes qui y sont associés ainsi que de bénéficier de la partie des bénéfices que l'entreprise redistribue à ses actionnaires : les dividendes.

La relation entre l'actionnaire et la société détenue est donc de nature multiple : c'est à la fois une relation de propriété et de contrôle, mais elle correspond aussi à un flux de valeur en sens inverse. Du fait que les actions peuvent être détenues par des personnes physiques comme par diverses formes de sociétés qui elles-mêmes peuvent être détenues par d'autres, ces relations peuvent s'enchaîner ; leur ensemble détermine ainsi de vastes réseaux.

Dans les travaux consacrés à ces réseaux de participations, la convention la plus répandue est de les représenter par un graphe dirigé de l'actionnaire vers la société détenue, évoquant ainsi la position dominante du premier sur la seconde. La connaissance du taux de participation permet de pondérer les arcs de ce type de réseau par une valeur de pourcentage ou – de façon équivalente – par un rationnel entre 0 et 1.

Afin de clarifier la sémantique des relations au sein de ce réseau, nous commencerons par exposer les fondements réglementaires qui déterminent l'existence et la mention de ces participations ainsi que le minimum de notions comptables qui permettent d'établir le lien entre le taux de participation et la valeur des flux (propriétés, contrôle et cash-flow) qui y transitent. Nous terminerons ce chapitre par une description des données recueillies et de leurs limites.

1. Au sens réglementaire, le terme participation est plus restrictif : cette précision est apportée à la section 2.2.

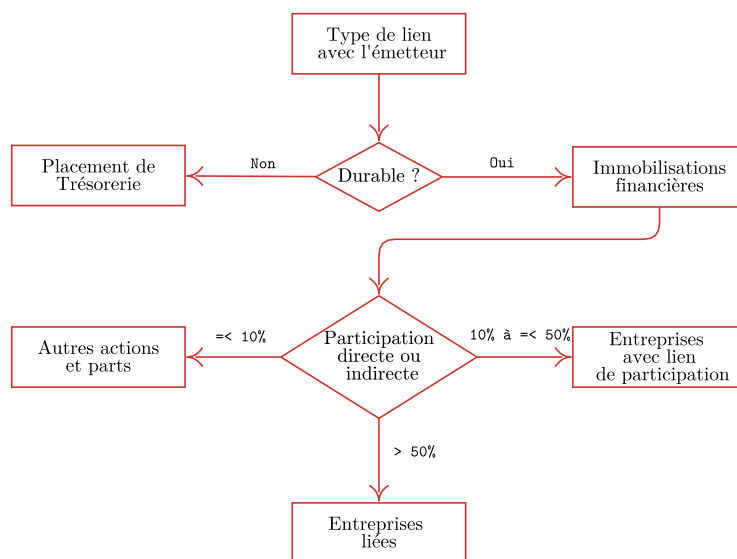


FIGURE 2.1 – Diagramme synthétique de distinction entre les types de participations d’après [Antoine et Cornil, 2002].

2.2 Fondements réglementaires, terminologie et conventions

En Belgique, c’est le *Code des sociétés*² qui réunit l’ensemble des dispositions relatives aux liens entre les sociétés.

La législation définit une **société mère** comme celle qui détient un pouvoir de contrôle sur une autre et une **filiale** comme une société à l’égard de laquelle un contrôle existe. La législation établit une distinction entre les participations détenues dans des **entreprises liées** ou des **entreprises avec lesquelles il existe un lien de participation**. Sommairement, les premières sont des sociétés sur laquelle un contrôle est exercé (le terme équivaut alors à celui de filiale), les secondes sont des sociétés dans lesquelles la société mère détient au moins 10% du capital. Le terme de participation ne s’applique au sens strict que lorsque les liens de détention atteignent ce seuil. Ces distinctions sont sans objet dans le cadre de ce mémoire aussi nous utiliserons les termes de **filiale** (directe ou indirecte) sans que cela induise un taux de participation minimal ou une notion de contrôle pour les entreprises situées en aval d’une autre dans le graphe, et de **société parente** pour celles situées en amont.

Le **réseau de participations** défini à la section précédente est dénommé, dans les

2. La version consolidée du code des sociétés est disponible sur le site du Moniteur Belge : http://www.ejustice.just.fgov.be/mopdf/1999/08/06_2.pdf ainsi qu’une version annotée [Benoit-Moury *et al.*, 2013] des éditions La Charte sur le site de la Banque Nationale : <https://www.nbb.be/doc/ba/jur/ent/code%20des%20societes.pdf>

articles de langue anglaise, *ownership network* ou *shareholding network*. Occasionnellement, on retrouve l'expression malheureuse et ambiguë *financial network* qui est également utilisée pour des données d'autres natures qui seront brièvement présentées à la section 3.3.1.

En anglais, les actionnaires sont dénommés *shareholders*, mais on trouve parfois l'expression *blockholders* pour désigner les actionnaires qui détiennent une part du capital suffisamment large pour s'assurer le contrôle de l'entreprise ou du moins une influence prépondérante par rapport aux autres [Barker, 2004].

L'ensemble formé par la société mère et ses filiales constitue ce qu'on nomme un **groupe de sociétés**. Deux éléments essentiels le caractérisent : il s'agit d'entités juridiques distinctes et l'une d'elles exerce un contrôle sur les autres et constitue l'organe de décision. Cette définition d'usage courant³ est suffisante dans le cadre de ce mémoire. Cependant la notion de *groupe d'entreprises* ainsi que l'usage d'autres expressions apparentées qu'on retrouve dans la littérature peuvent induire en confusion, aussi la section 7.4 dans les annexes apporte-t-elle quelques précisions terminologiques.

2.3 Notions comptables

Les comptes annuels d'où proviennent les données des participations comportent essentiellement le *bilan* et le *compte de résultats*.

Le bilan constitue un «instantané» du patrimoine de la société à la date où il est établi. Il comporte en vis-à-vis un actif et un passif dont les totaux concordent exactement ce qui étonne assez souvent le néophyte ; il est plus aisé de le comprendre lorsque l'actif et le passif sont qualifiés respectivement d'*emploi* et de *ressources*. Le passif décrit ainsi l'origine des moyens financiers dont dispose l'entreprise pour exercer son activité tandis que l'actif décrit à quel usage ils sont affectés.

On remarquera que l'ordre dans lequel les différents postes du bilan sont présentés n'est pas indifférent : ceux de l'actif sont classés par ordre de disponibilité, ceux du passif par ordre d'exigibilité.

Au sein du passif, les *capitaux propres* représentent la part des ressources qui permettent de financer l'activité sans recourir à l'emprunt. Ils se composent du *capital* proprement dit, c.-à-d. les fonds apportés par les actionnaires, lors de la constitution de l'entreprise (ou ultérieurement si nécessaire, on procède alors à une *augmentation de capital*) et les résultats cumulés de l'activité (bénéfices ou pertes). Contrairement aux emprunts, les fonds propres ne sont pas soumis à une obligation de remboursement, aussi leur part dans le bilan représente-t-elle un indice de la capacité de l'entreprise à faire face à des circonstances difficiles.

3. En fait, étonnamment, le droit belge ne connaît pas de définition globale relative aux groupes de sociétés. Cette notion n'apparaît que de façon indirecte dans le Code des Sociétés par les définitions relatives aux participations ainsi que celles concernant les comptes consolidés.

N°	BE 0423.898.710	C 5.5.1
----	-----------------	---------

INFORMATION RELATIVE AUX PARTICIPATIONS

PARTICIPATIONS ET DROITS SOCIAUX DÉTENUS DANS D'AUTRES ENTREPRISES

Sont mentionnées ci-après, les entreprises dans lesquelles l'entreprise détient une participation (comprise dans les rubriques 280 et 282 de l'actif) ainsi que les autres entreprises dans lesquelles l'entreprise détient des droits sociaux (compris dans les rubriques 284 et 51/53 de l'actif) représentant 10% au moins du capital souscrit.

DÉNOMINATION, adresse complète du SIÈGE et pour les entreprises de droit belge, mention du NUMÉRO D'ENTREPRISE	Droits sociaux détenus			Données extraites des derniers comptes annuels disponibles			
	directement		par les filiales	Comptes annuels arrêtés au	Code devise	Capitaux propres	Résultat net
	Nombre	%	%			(+ ou -) (en unités)	
SOLVAY BE 0403.091.220 Société anonyme Rue de Ransbeek 310 1120 Neder-Over-Heembeek BELGIQUE Ordinaire	25.578.267	30,2		31-12-2014	EUR	7.781.417.665	549.929.569

FIGURE 2.2 – Exemple d'information relative aux participations dans des comptes annuels : extrait des comptes de la société Solvac avec la mention des participations dans sa filiale Solvay.

Avec les résultats nets, c'est l'une des deux données financières qui sont fournies pour chaque participation mentionnée dans les comptes annuels (voir fig. 2.2). Si le législateur en a imposé la mention, c'est probablement parce que les fonds propres reflètent mieux que le total du bilan la «richesse» réelle de l'entreprise, et parce qu'associés aux résultats, ils fournissent l'un des ratios financiers⁴ les plus utilisés, le *ratio de rentabilité* (*“Return on Equity”*, ou ROE) :

$$\text{rentabilité nette des capitaux propres} = \frac{\text{bénéfice ou perte de l'exercice}}{\text{capitaux propres}} \times 100$$

Ce ratio exprime la capacité de l'entreprise à générer des profits à partir des capitaux ; il permet donc à l'actionnaire d'estimer le retour financier qu'il obtiendra en retour de son investissement. Ce rapport peut varier fortement selon le type d'activité (une société d'un secteur de l'industrie lourde demande des investissements plus conséquents qu'un cabinet d'audit) ou le moment retenu dans l'historique de l'entreprise (les années suivant la constitution demandent plus d'investissements et génèrent assez souvent des pertes).

Lorsqu'une entreprise détient des participations dans la perspective d'un lien durable, celles-ci sont reprises à l'actif dans les **immobilisations financières**⁵. Si au contraire des participations sont acquises pour en tirer un rendement financier sans envisager d'influer sur la gestion de cette entreprise, il s'agit de **placements de trésorerie**.

Contrairement au bilan, le **compte de résultats** traduit l'activité de l'entreprise au cours d'une période donnée. Il fournit la différence entre les produits et les charges de l'entreprise, se traduisant par une perte ou un bénéfice. Le compte se termine par l'**affectation** des résultats. C'est à ce stade qu'est effectuée la liaison entre compte de résultats et bilan (figure 2.3 p.16) ; le bénéfice ou la perte est cumulée avec le résultat de l'exercice précédent puis en partie affecté au bilan (rubrique «Bénéfice (Perte) reporté(e)», en partie distribué aux actionnaires. La partie affectée au bilan s'ajoutera aux capitaux propres en cas de bénéfices. Inversement, un résultat en perte diminuera le total des capitaux propres et peut mener à des prélèvements sur les réserves, voire nécessiter une augmentation du capital proprement dit.

2.4 Origine et caractéristiques des données

Le Code des Sociétés définit les obligations des sociétés en ce qui concerne l'établissement des comptes annuels. Cela recouvre le contenu des comptes annuels, les catégories de

4. Les ratios financiers sont des rapports arithmétiques entre des masses financières extraites des comptes annuels et qui ont été définis pour fournir des clés d'interprétation simples à calculer et pour résumer l'information financière [Cerrada, 2006].

5. Une distinction est en plus effectuée entre les «entreprises liées» et «les entreprises avec lesquelles il existe un lien de participation» en ce qui concerne la rubrique à laquelle la valeur est mentionnée. Cette différence de traitement reflète l'intensité du lien qui existe entre la société détentrice et celle dans laquelle elle a investi. Mais cela est sans incidence dans le cadre de ce mémoire.

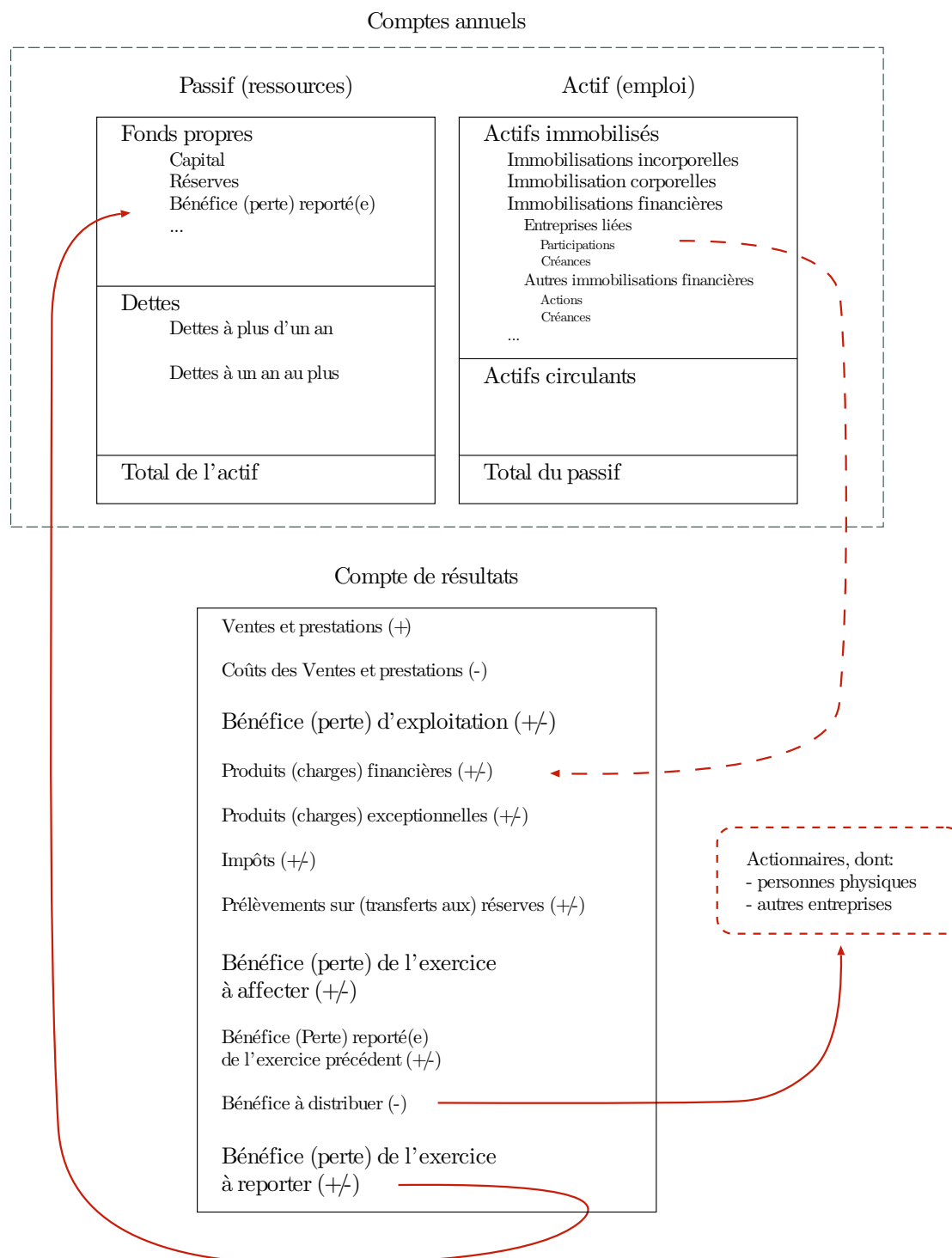


FIGURE 2.3 – Intégration des flux financiers résultant des participations au sein des comptes annuels. Lorsqu'une entreprise détient des participations, les bénéfices distribués par ses filiales sont intégrés au niveau des produits financiers. Lorsqu'elle-même distribue une part de ses bénéfices, ils sont soustraits des bénéfices avant de reporter ceux-ci au bilan.

sociétés qui sont soumises à l'obligation de dépôt, les «formalités de publicités» c.-à-d. les procédures relatives au dépôt auprès de la Banque Nationale, etc. Les éléments qui présentent un intérêt dans le cadre de ce mémoire sont les suivants :

- La plupart des entreprises belges dont la responsabilité des actionnaires ou associés est limitée à leur apport doivent publier leurs comptes annuels en les déposant auprès de la Centrale des bilans de la Banque nationale.
- La majorité des entreprises soumises à l'obligation de dépôt utilisent l'un des deux *modèles normalisés* mis à disposition par la Banque⁶ :
 - le modèle complet pour les grandes entreprises et les entreprises cotées en bourse ;
 - le modèle abrégé pour les petites entreprises non cotées en bourse.

Cela, adjoint au fait que le dépôt s'effectue majoritairement par voie électronique, permet à la Centrale des Bilans d'en automatiser l'extraction des données et d'établir des statistiques qui globalisent les différentes rubriques des comptes annuels rassemblées par secteurs d'activités ainsi que des distributions et moyennes de ratios financiers⁷.

La Centrale des Bilans fournit de manière gratuite les données individuelles des comptes annuels via un service web, sous forme de fichiers pdf ou *xbrl*, un dialecte xml spécialement développé pour l'échange de rapports financiers. Ce service web n'est cependant pas destiné à permettre des téléchargements automatisés et massifs ; la Banque se réserve le droit de limiter de telles procédures en bloquant l'accès à l'application. Ainsi, s'il est possible de prendre connaissance des participations d'une ou quelques entreprises par la consultation de leurs comptes annuels (*cf.* figure 2.2), la disponibilité de l'ensemble des participations n'est possible que via un service annuel payant incluant l'ensemble des données des comptes annuels (ce service est destiné essentiellement aux grandes institutions financières ou institutionnelles).

Les données relatives aux participations collectées par la Banque Nationale présentent malheureusement plusieurs limites quant à l'exploitation qu'on peut en faire :

- La première est que par définition, étant composées de données concernant les participations de sociétés belges, elles ne comprennent pas les participations que les sociétés étrangères détiennent dans les sociétés belges, ce qui dans l'analyse des caractéristiques topologiques du graphe impactera les rapports entre liens entrants et sortants.
- La seconde tient à la possibilité pour une société d'émettre des actions de différents types, associés à des droits de votes différents ou dépourvus de droits de votes⁸. Le type d'action est mentionné par les sociétés mères dans leurs comptes

6. <https://www.nbb.be/fr/centrale-des-bilans/etablir/modeles-des-comptes-annuels/modeles-pour-entreprises>

7. <https://www.nbb.be/fr/centrale-des-bilans/produits-de-la-centrale-des-bilans/statistiques>

8. L'existence de ces types d'actions relève des «*Control Enhancing Mechanisms*», mécanismes ayant pour objectif de contrer le principe général de proportionnalité entre propriété et contrôle qui sera abordé à la section 3.1.

annuels de façon non standardisée et en différentes langues. Pour corriger ce problème, il n’y aurait d’autre solution que de consulter les statuts de chaque société concernée pour connaître les caractéristiques liées aux types de participations, ce qui n’est pas envisageable dans le cadre de ce mémoire⁹. Les conséquences apparaissent de façon évidente lorsqu’une entreprise détient des parts de plusieurs types dans une même filiale : cette situation a pour effet que le graphe contient des arcs parallèles et que certaines entreprises présentent une somme des poids des arcs entrants supérieure à 100%. Cependant, un examen rapide des différentes dénomination semble conforter l’hypothèse que lorsqu’un seul type de participation est mentionné pour une filiale, il s’agit d’actions ordinaires procurant de manière égale droits de vote et propriété.

- Alors que la plupart des rubriques des comptes annuels normalisés sont l’objet de contrôles arithmétiques et logiques automatisés¹⁰ qui en vérifient la cohérence, aucun recoupement n’est effectué pour les mentions des participations. Comme elles sont rédigées dans un format libre, elles nécessitent d’être encodées manuellement. S’agissant de données transmises par un intermédiaire, on conçoit aisément que leur fiabilité ne soit pas entière (nous avons pu constater que lorsqu’une entreprise était filiale de plusieurs, il arrive que les données financières mentionnées ne concordent pas). Même en l’absence d’erreurs, les sociétés mères ne sont tenues que de fournir «les données extraites des derniers comptes disponibles» de leurs filiales.
- Nous ne connaissons pas non plus les secteurs économiques auxquels appartiennent les sociétés étrangères, alors que les sociétés belges sont assorties d’un code dit NACE (Nomenclature statistique des Activités économiques dans la Communauté Européenne) qui définit les secteurs sous une forme hiérarchique et dont il sera fait usage au chapitre 4, p.47. Les deux données financières fournies par les sociétés étrangères sont mentionnées dans la devise du pays d’origine ; pour des raisons pratiques, nous avons dû nous limiter à effectuer la conversion en euro à des taux définis à une date commune à toutes, sans tenir compte de la date précise d’établissement de ces comptes.

Pour toutes ces raisons, il serait hasardeux d’avancer des affirmations telles que celles relatives à l’identification des sociétés exerçant le contrôle le plus important sans les recouper avec d’autres sources ni effectuer un travail considérable de vérification qui n’est pas envisageable dans le cadre d’un mémoire. Néanmoins, l’exploitation de ces données en l’état nous a semblé intéressante et suffisamment significative pour en étudier des caractéristiques générales d’ordre topologique.

9. Le nombre de formes distinctes enregistrées en base de donnée est supérieur à 5000 pour un seul exercice. L’impossibilité de les traiter même «au moyen de filtres sophistiqués» avait été soulignée par [Becht et Chapelle, 1997] ce qui les avait amené à ne pas les exploiter et à limiter leur analyse aux sociétés cotées et à leur voisinage.

10. La procédure de dépôt des comptes sous forme numérique comprend une série de contrôles élémentaires qui prévient à ce stade l’introduction de données fautives. Des contrôles plus affinés effectués lors de leur traitement peuvent donner lieu à l’envoi de «lettres d’anomalies» auprès des entreprises afin d’obtenir des données corrigées.

Il faut relever que la majorité des travaux réalisés sur les réseaux de participations (sections 3.3.2 à 3.1) l'ont été à partir d'informations extraites des déclarations des sociétés cotées en bourse sur la composition de leur actionnariat auprès de leurs autorités de contrôle¹¹. Les données étaient donc établies *de bas en haut*, en partant d'un échantillon défini de sociétés détenues. Ces données utilisées pouvaient comprendre comme actionnaires à la fois des personnes physiques et morales, mais seulement celles détenant des participations importantes. Ce mode de collecte amène généralement à une surreprésentation des institutions financières au sein des actionnaires.

À l'inverse, les données utilisées dans ce mémoire sont établies *du haut vers le bas*, au départ d'un ensemble beaucoup plus large délimité seulement par des critères réglementaires (les sociétés tenues de publier leurs comptes annuels). Elles comprennent donc une multitude de petites entreprises détenant de façon majoritaire une seule ou quelques filiales, dont on peut s'attendre à ce que l'actionnariat ne varie que très occasionnellement. Comme on le verra, le réseau qui en résulte est extrêmement fragmenté en dehors d'une composante principale.

D'autre part, comme les sociétés publiant leurs comptes annuels sont des sociétés belges,

- nous avons connaissance des sociétés étrangères détenues par des sociétés belges ;
- mais pas des sociétés étrangères détenant des participations dans les sociétés belges (ni d'ailleurs dans les sociétés de quelque nationalité qui soit).

Par conséquent, les chemins partant d'une société belge vers une société étrangères n'auront jamais une longueur supérieure à 1.

En résumé, alors que les études antérieures concernaient généralement des sous-ensembles nationaux du réseau global des participations limitées à celles détenues par et/ou détenant des sociétés cotées en bourse, mais d'un poids supérieur à un seuil déterminé, nous disposons d'un réseau plus vaste, beaucoup plus hétérogène (tant pour le type des sociétés que pour le poids des liaisons), déséquilibré en amont (par l'absence d'actionnaires étrangers) et souffrant de quelques approximations citées auparavant.

En pratique, les données avec lesquelles nous travaillons constituent un réseau composé de 85857 sommets et 71654 arcs sur la base des comptes déposés en 2012¹². La majorité des sommets appartiennent à des composantes de petite taille puisque la réunion des composantes de taille 2,3 et 4 rassemble déjà plus de 50% des sommets et celle des composantes d'une taille jusqu'à 16 plus de 75%. La composante principale ne comporte

11. Dans un but de protection des consommateurs de produits financiers (les acheteurs d'actions potentiels), la plupart des pays ont institué une autorité de contrôle dont la mission est – parmi d'autres – de permettre à ceux-ci d'avoir connaissance de l'actionnariat des sociétés cotées. Les actionnaires dont les participations franchissent certains seuils (5% par exemple) sont tenus de le déclarer auprès de cette autorité. En Belgique, il s'agit depuis 2011 de la FSMA (pour *Financial Services and Markets Authority*) : <http://www.fsma.be>.

12. On a éliminé du réseau 641 arcs parallèles provenant des actions de types différents en retenant – arbitrairement, il est vrai – la participation relative la plus élevée, leur nombre ne permettant pas une sélection fine, comme expliqué plus haut.

TC	#C	#S	%S	%FP	%B
2	13649	27298	31,7947	3,950	4,205
3	4083	12249	14,2667	3,973	2,301
4	1684	6736	7,8456	2,012	1,730
5	885	4425	5,1539	3,576	2,262
6	543	3258	3,7947	0,858	0,546
7	331	2317	2,6987	0,719	1,275
8	207	1656	1,9288	0,595	0,661
9	167	1503	1,7506	0,555	0,327
10	124	1240	1,4443	0,382	0,450
⋮					
178	1	178	0,2073	0,005	0,009
14610	1	14610	17,0167	69,183	70,729

FIGURE 2.4 – Nombre et composition des composantes du réseau. TC, #C : taille et nombre de composantes. #S : nombre de sommets. %S : pourcentage du total des sommets. %FP, %B : pourcentage en valeur des fonds propres positifs et des bénéfices.

que 17% des sommets, mais en valeur monétaire, ceux-ci totalisent 69% des valeurs propres positives et 71% des bénéfices.

La distribution des arcs selon leur poids montre des densités importantes autour des valeurs de 0, 50 et 100%, de façon moins aigüe autour de 10, 25, 33, 75 et 90% (figure 2.5).

On peut s'étonner de la présence importante d'arcs de poids nul (2% du total des arcs ont un poids égal à 0 – par ailleurs 5% ont un poids inférieur à 0,1%, près de 9% un poids inférieur ou égal à 1%). La présence de ceux-ci s'explique pour plusieurs raisons :

- l'imprécision des sources : les déposants ne fournissent pas une information complète, ne mentionnent par exemple que le nombre d'actions détenues et pas leur pourcentage. Une valeur de 0 est alors encodée ;
- la participation est exclusivement indirecte. Une valeur de 0 est alors encodée pour la participation directe ;
- l'actionnaire détient une seule action et celle-ci ne représente qu'une fraction minimale qui n'est pas représentable avec 2 décimales.

Cette dernière situation peut provenir de l'obligation pour une société de compter un nombre minimal d'associés selon sa forme juridique¹³. Ce minimum n'est plus que de 2 actuellement pour une société anonyme, mais il a été de 7 par le passé. Cela peut également expliquer l'existence de liens croisés entre maison mère et filiale.

13. Cette obligation montre également que la présomption d'un lien durable entre actionnaire et société détenue ne peut être déduite du pourcentage d'actions détenues. Des sociétés détenant des actions à titre de placement de trésorerie peuvent effectivement présenter de faibles participations volatiles, mais a contrario la détention d'une seule action peut signifier un lien fort et durable dans ce contexte.

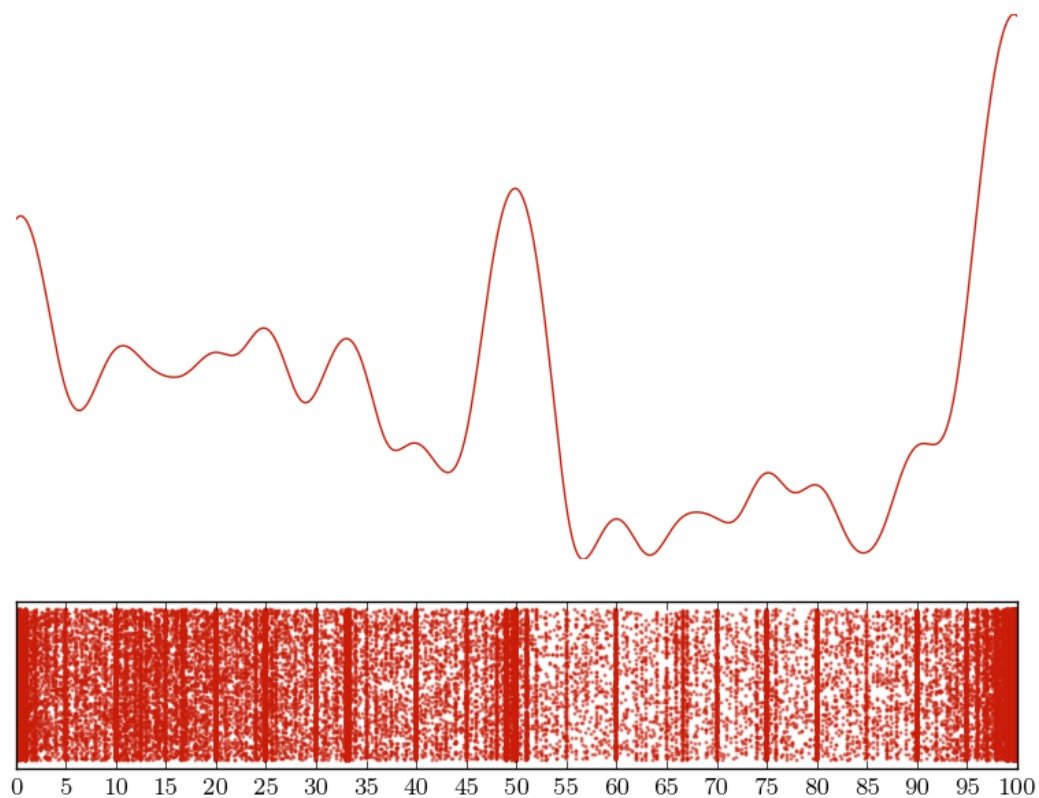


FIGURE 2.5 – Distribution des participations selon leur poids. Le nuage de points représente les participations positionnées de 0 à 100 sur l'axe horizontal et de façon aléatoire sur l'axe vertical, de manière à en rendre perceptibles les zones de forte densité. La courbe supérieure établie au moyen d'un noyau gaussien (méthode `sklearn.neighbors.KernelDensity` de Scikit Learn) fournit une autre visualisation de cette densité.

Au cours de l'analyse de ces données, nous utiliserons généralement 3 ensembles :

- le réseau entier comprenant toutes les entreprises (**RE** en abrégé) ;
- celui, faiblement connexe, constitué par la composante principale (**CP**) comprenant 14610 sociétés et 18501 relations de participation ;
- celui établi à partir des sociétés cotées en bourse et de l'ensemble de leurs descendantes et ancêtres, obtenu pour chacune par une recherche en largeur ordinaire et inversée (**SC**), composé de 2865 sociétés et 3326 relations.

Outre le fait que ce dernier ensemble permet de disposer d'un échantillon de taille réduite pour tester divers algorithmes, les sociétés cotées en bourse sont susceptibles de présenter plusieurs caractéristiques particulières. Leurs actions étant disponibles sur le marché, elles présentent en général un actionnariat plus dispersé (bien que les personnes physiques qui représentent les petits porteurs ne figurent pas dans nos données). Inversement, on trouve parmi elles des sociétés d'envergure disposant de filiales étrangères. On s'attend donc à y trouver une plus large proportion de sommets présentant des degrés élevés¹⁴.

Une description plus détaillée des caractéristiques de ces trois jeux de données sera effectuée au début de la section 4.

Ces graphes ont été enregistrés au format *graphml* [Brandes *et al.*, 2002], un dialecte *xml* permettant de stocker pour les sommets comme pour les arcs des propriétés de toute nature et qui s'est imposé comme un standard exploitable par exemple par la librairie NetworkX en Python ou le logiciel de visualisation Gephi.

2.5 Sémantique des liaisons

Nous avons brièvement mentionné que la relation de l'actionnaire vers la société détenue est à la fois une relation de propriété et de contrôle, et qu'elle correspond de plus à un flux de valeur (*cash-flow*) en sens inverse. Sa modélisation s'effectue donc par un graphe dirigé et le choix généralement adopté est d'orienter les arcs de l'actionnaire vers les sociétés détenues.

Pour être complet, l'on peut ajouter à ces relations clairement orientées des relations de nature moins quantifiable et non dirigée. On peut en effet estimer que les liens de participation induisent entre les parties une transmission mutuelle d'information. De même, la propagation du risque financier s'effectue de façon non dirigée : la faillite d'une filiale peut affecter les autres filiales de la société mère, quand bien même il n'existe pas de chemin de l'une aux autres en termes de théorie des graphes. Nous négligerons ces aspects de la relation mais nous reviendrons sur l'opportunité d'utiliser un graphe non orienté dans le cadre de la recherche de communautés (section 5.2).

14. La liste des sociétés cotées en bourse provient du site de la FSMA. Malheureusement, ce site ne fournit pas l'historique de cette liste qui varie avec le temps, des introductions en bourse ou des retraits de ce marché ayant lieu occasionnellement. Par conséquent, ce jeu de données présente une certaine imprécision, la liste utilisée étant postérieure aux données bilantaires.

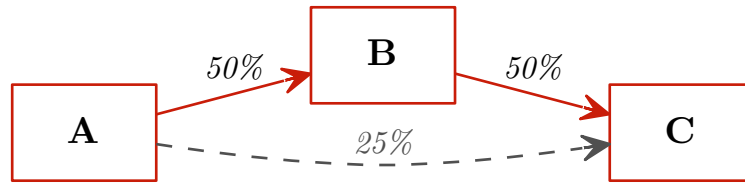


FIGURE 2.6 – Exemple élémentaire de chaîne de participations.

La nature de cette relation différencie fortement les réseaux de participations d'autres réseaux «de terrain» analysés dans la littérature, en particulier de ceux où le poids des arcs exprime une distance. Cette différence provient de la façon dont on doit interpréter la succession de plusieurs arcs constituant un chemin. Alors que les distances constitutives d'un chemin s'additionnent, les participations successives se multiplient. Dans l'exemple élémentaire de la figure 2.6, l'actionnaire *A* qui détient la moitié de la société *B* détient au final le quart de la société *C*, soit le produit de 0,5 par 0,5.

Il en découle également que *plus un parcours est long, plus faible est son poids* ; comme le réseau comporte de nombreuses participations très réduites, lorsqu'elles sont incorporées dans des chaînes de participation, celles-ci s'en trouvent réduites à des taux souvent négligeables.

La situation se complexifie lorsque plusieurs chemins joignent deux sociétés ; il faut alors additionner les participations correspondant à chaque chemin pour déterminer la propriété. L'exemple de la figure 2.7, outre qu'il présente le calcul de propriété dans un tel cas, introduit le problème de la distinction entre le droit de propriété représenté par les parts du capital et les droits de vote qui y sont associés. On y montre comment, au moyen de participations indirectes, une société *Y* détient plus de la moitié du capital d'une société *C* sans disposer de droits de vote permettant de la contrôler alors que sa concurrente *X* détient la majorité des droits de vote sans en détenir la moitié du capital.

Lorsqu'une structure en étages de l'actionnariat permet ainsi de détenir un contrôle supérieur à la propriété, on la qualifie de *pyramide*. La divergence qui peut subvenir entre propriété et contrôle provient de l'effet de seuil inhérent au contrôle. On ne doit cependant pas déduire de l'exemple de la figure 2.7 que le contrôle sur une entreprise soit défini de façon binaire («*X* détient/ne détient pas le contrôle sur *Y*») ni qu'il soit nécessaire pour l'obtenir d'atteindre un seuil de 50% de propriété. Lorsque les parts de l'actionnaire le plus important n'atteignent pas ce seuil mais qu'aucune coalition ne se forme entre les autres actionnaires, il peut exercer un contrôle de fait sur l'entreprise à partir d'un seuil bien moins élevé. Le degré de contrôle qu'un actionnaire exerce peut ainsi prendre différentes valeurs et sa mesure a fait l'objet de multiples recherches aboutissant à des conclusions qui ne sont pas unanimes. Ces problématiques ont été abondamment étudiées et un relevé de cette littérature sera effectué dans la section 3.1.

Indépendamment de cette divergence, l'estimation du taux de propriété n'est pas évidente en présence de cycles. Ainsi, dans l'exemple de la figure 2.8 les sociétés intermé-

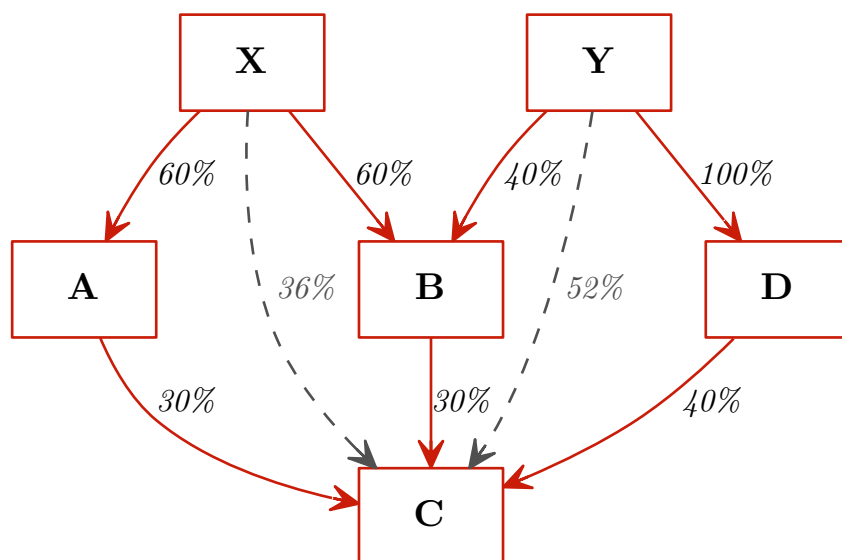


FIGURE 2.7 – «*X* a le contrôle indirect de *C*. Même s’il ne détient indirectement que 36% de ses parts de capital : $(60\% \times 30\%) + (60\% \times 30\%)$, *X* obtient ce contrôle par le biais de ses deux filiales *A* et *B* en combinant leurs droits lors de l’assemblée générale de *C* : $30\% + 30\% = 60\%$. D’un autre côté, *Y* détient $52\% = (40\% \times 30\%) + (100\% \times 40\%)$ de *C*, mais ne le contrôle pas, du fait que les droits de vote dont il dispose lors de l’assemblée générale ne s’élèvent qu’à 40% des parts de celui-ci. Autrement dit, les droits de vote résultant des chemins $X \rightarrow A \rightarrow C$ et $X \rightarrow B \rightarrow C$ doivent être cumulés pour déterminer le contrôle effectif» [Eurostat, 2010].

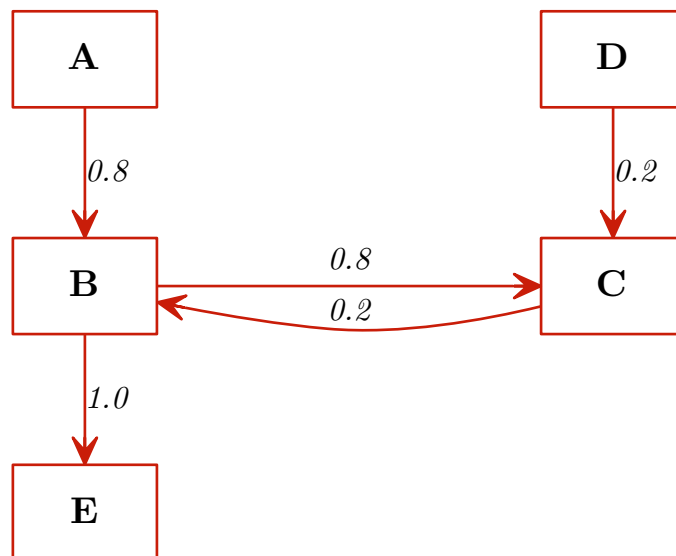


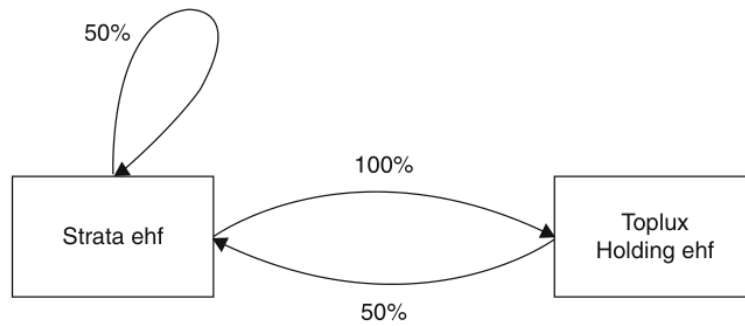
FIGURE 2.8 – Exemple de structure d’actionnariat avec cycle. Les chaînes menant de A et D vers E valent respectivement $0,8 \times 1,0 = 0,8$ et $0,2 \times 0,2 \times 1,0 = 0,04$. Comme les sociétés intermédiaires B et C ne détiennent rien en propre, le solde de $0,16$ reste à répartir entre A et B .

diaires B, C étant intégralement détenues par des sociétés internes au réseau, on peut considérer qu’elles ne détiennent rien de E par elles-mêmes et que la propriété de E est intégralement répartie entre A et B . On pourrait penser à première vue que cette répartition correspond aux chaînes sortant de A et B , soit respectivement 80% et 4%, mais ce n’est pas le cas. Comme B a une participation en C , A détient aussi une part de C , mais comme celui-ci détient également une part de B , une fraction en supplémentaire de B reviendra à A et ainsi de suite.

La résolution d’un tel problème fait appel au calcul matriciel et sera explicitée à la section 3.1, mais on peut la comprendre par une approche intuitive ; il suffit d’imaginer qu’une quantité quelconque d’une «charge» émanant d’un nœud soit répartie proportionnellement en empruntant les arcs à rebours¹⁵ et qu’on procède de façon récursive pour chaque fraction déposée dans un nouveau nœud jusqu’à ce que toute la charge ait abouti dans les nœuds sans arcs entrants. En présence de cycles, il est évident que le processus devrait être poursuivi indéfiniment, mais plus on le répétera, plus le résultat tendra vers la répartition espérée.

Dans notre exemple, les participations de A et D dans E sont respectivement de 95,2% et 4,8% (le détail du calcul figure en annexe 7.1). La présence de cycles induit ainsi des écarts contre-intuitifs par rapport aux participations apparentes. De telles situations sont

15. On pourrait aussi envisager ce réseau comme un ensemble de canalisations de capacités variées et dotées de clapets anti-retour par lesquelles le flot se répartit en fonction de la résistance qu’il rencontre.



Source: Icelandic Enterprise Registry, Bjarnadottir and Hansen, 2010, *Investigation into the Cross Ownership and Bank Credit to Related Parties*. (Report of the Special Investigation Commission (SIC), Vol. 9. Appendix 2, p. 74. [Title in Icelandic: Rannsókn á kross eignartengslum og útlánnum bankanna til tengdra aðila] Reykavík: Special Investigation Commission).

FIGURE 2.9 – Exemple de structure de propriétés croisées de firmes islandaises en 2008, d’après [Johnsen, 2014]. De manière anormale, aucune personne physique n’est détentrice d’une part de ces firmes, qui sont intégralement propriétaires d’elles-mêmes.

d’ailleurs l’objet de restrictions ; le Code des Sociétés prescrit que «*les sociétés filiales ne peuvent posséder ensemble des actions ou parts bénéficiaires de leur société mère ayant pris la forme de société anonyme [...] représentant plus de 10% des voix attachées à l’ensemble des titres émis par ladite société mère.*»(Art.631). Si une telle situation survient, les droits de vote associés sont suspendus et les titres acquis par méconnaissance doivent être aliénés dans un délai d’un an. La nécessité d’une telle régulation peut être illustrée par l’exemple de la figure 2.9 où l’auteur décrit l’emballement qui a mené à l’effondrement du système bancaire islandais. Dans le chapitre intitulé «*The Web of Ownership*», il décrit comment le recours excessif aux structures en pyramide ainsi que le rachat d’actions propres a opacifié la propriété réelle des entreprises et mené à ce que 91 d’entre elles deviennent propriétaires d’elles-mêmes à 100%.

Pour souligner le caractère pathologique des structures cycliques, on considérera également la figure 2.10 où la seule participation dans les sociétés du noyau central qui ne soit pas détenue par l’une d’elles est celle (minime) émanant de 1. La *totalité* de la propriété et du contrôle des sociétés 2 à 6 lui revient donc.

Le flux de valeur représenté par la relation constitue ainsi le reflet inversé de la relation de propriété mais deux précisions importantes concernent celui-ci :

1. On ne doit pas perdre de vue qu’il existe une dimension *temporelle* au cheminement du cash-flow dans le réseau. Toute société ne peut transmettre à sa maison mère que les dividendes qu’elle a reçus de ses propres filiales *au terme de l’exercice précédent*. Ainsi, si la société *E* distribue une quantité 100 de dividendes lors de l’exercice *t*, *B* les intégrera dans ses résultats lors de l’exercice *t + 1* et *A* n’inté-

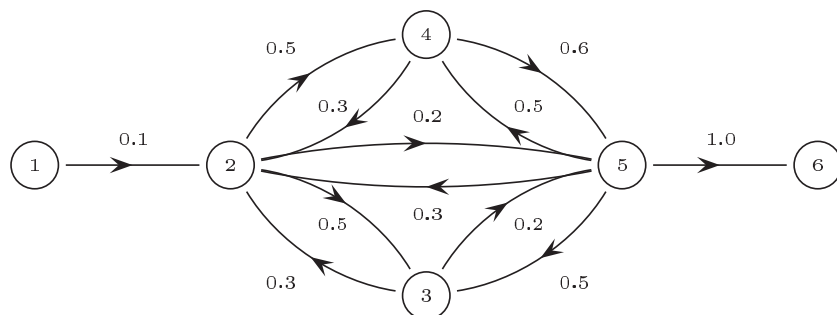


FIGURE 2.10 – Structure de propriété et de contrôle cyclique, d’après [Glattfelder, 2013]. Dans cette situation extrême, les sommets du noyau central détiennent mutuellement la totalité de leurs parts, à l’exception du 2 dont 10% sont détenus par 1, ce qui permet en théorie à ce dernier de contrôler l’ensemble. Cette interprétation a été nuancée par [Vitali *et al.*, 2011], voir section 3.3.3.

- grera 80 qu’au cours de l’exercice $t + 2$. Au cours de l’exercice $t + 4$ seulement, elle pourra intégrer $20 \times 0,8 \times 0,8 = 12,8$ supplémentaires, soit les dividendes qui auront transité par C .
2. La relation de cash-flow ne correspond qu’à un dividende *potentiel* ! D’une part, les sociétés ne distribuent des dividendes que si elles réalisent des bénéfices. D’autre part, la décision de distribuer ceux-ci ainsi que la détermination de leur montant reviennent à l’assemblée générale ; en fonction des circonstances, celle-ci peut juger que ces fonds seront plus utilement affectés à l’investissement ou au financement de la société. Dans une structure d’actionnariat à plusieurs niveaux, il serait erroné de croire que les sociétés situées aux niveaux intermédiaires transmettent automatiquement les dividendes du bas vers le haut ; les dividendes qu’elles reçoivent sont intégrés dans leurs résultats au même titre que d’autres revenus et c’est la globalisation de tous ceux-ci (voir figure 2.3) qui inspire la décision de l’assemblée générale.

Ces précisions amènent à relativiser l’évaluation théorique du niveau de propriété. Si l’on reprend l’exemple de la figure 2.8, on voit que la société A peut encaisser la part de 80% des dividendes de E dans un délai de deux ans, mais que les fractions suivantes qui permettront de parvenir au total théorique de 95,2% ne lui parviendront qu’après 4 ans, puis 6, 8, etc. À l’échelle économique, il s’agit de délais durant lesquels la perspective que les résultats des filiales intermédiaires – ainsi que les taux de participations en celles-ci – restent constants n’est évidemment guère assurée.

Enfin, il faut souligner que le réseau ainsi construit à partir des participations des personnes morales occulte dramatiquement le rôle des acteurs humains dans les processus qui l’affectent. Le réseau que nous examinons comprend une large part de sociétés dont

la totalité de l'actionnariat n'est pas connu¹⁶, soit parce que l'information est incomplète, soit parce que la part non mentionnée est composée d'actionnaires qui ne sont pas des sociétés. Lorsqu'on ne dispose pour seule information au sujet de l'actionnariat d'une société *A* que d'une participation de 40% par une autre société *B*, cela ne signifie évidemment pas que les 60% restants soient la propriété de *A* et qu'elle conserve pour elle les dividendes qui y correspondent. De même, lorsqu'on identifie une société comme actionnaire ultime d'une pyramide, ce n'est pas cette société en tant que telle qui en détient le contrôle, mais bien, «un étage au-dessus», les personnes physiques qui en sont actionnaires.

16. Pour donner l'ordre de grandeur de ce qui échappe ainsi à notre connaissance dans l'examen du réseau des sociétés belges, le cumul de toutes les participations non répertoriées des sociétés ayant au moins un arc entrant représente 63,6% de la valeur de leurs fonds propres.

3 État de l’art sur les études de réseaux de participations

Bien avant que l’étude des réseaux complexes ne devienne un champ d’étude à part entière et ne fournisse un cadre d’analyse commun à des réseaux de toute nature, une littérature abondante a été consacrée aux participations des entreprises selon une approche qui relève de ce qu’on nomme à présent la gouvernance d’entreprise, c-à-d. la construction d’un cadre, d’un système qui permettrait de respecter de manière équitable les intérêts des diverses parties prenantes (*stakeholders*).

Cette littérature consacrée à *la séparation de la propriété et du contrôle*, qui n’intègre pas encore la dimension réseau des participations, sera abordée dans la section 3.1.

Nous consacrerons ensuite une section à la littérature relative à l’évaluation du contrôle exercé par les actionnaires qui, bien que n’envisageant pas l’étude de larges réseaux de participation, a nécessité de prendre en considération le réseau à l’échelle locale et de recourir à la théorie des graphes et au calcul matriciel. Comme nous en ferons usage dans la partie applicative de ce mémoire, nous en détaillerons le mode de calcul.

Les participations ont aussi été étudiées dans le cadre plus général de celui des réseaux complexes, dont l’étude a mis en évidence des caractéristiques qui sont communes à la majorité des réseaux quel que soit leur domaine d’application. Aussi, les participations sous la forme de réseau ont-elles été étudiées au même titre que d’autres – de façon assez occasionnelle – pour y souligner la récurrence de certains phénomènes. Un relevé en est effectué dans la section 3.3.

Nous entamerons celle-ci par une mise en contexte de la recherche sur les réseaux complexes suivie de quelques éclaircissements relatifs à la terminologie, puis nous exposerons brièvement les différents types de réseaux à caractère économique rencontrés dans la littérature avant d’en venir aux réseaux de participations proprement dits.

Nous terminerons ce chapitre par une section consacrée aux travaux de Vitali, Battiston et Glattfelder, à la fois parce qu’ils constituent une synthèse des différentes problématiques relatives aux participations et pour l’impact inhabituel qu’une telle étude a pu avoir en dehors du domaine académique.

3.1 La séparation de la propriété et du contrôle

Cette problématique se pose du fait que lorsque des actionnaires disposent d'un degré de contrôle supérieur à celui de leur propriété, ils auront tendance à orienter la stratégie de l'entreprise vers la satisfaction de leurs intérêts particuliers au détriment de ceux des autres actionnaires, ce qui constitue une véritable forme d'*expropriation*.

Elle apparaît typiquement dans deux types de configurations qui ont chacune suscité des études spécifiques :

- celle d'entreprises isolées mais à actionnariat dispersé où le problème d'agence se pose entre actionnaires externes (*outsiders*) et internes, les managers (*insiders*) ;
- celle d'entreprises insérées dans des groupes où le problème se pose entre actionnaires disposant du contrôle ou non selon la configuration du réseau, en particulier les pyramides.

Pour les premières, l'ouvrage fondateur est celui de [Berle et Means, 1932] ; ceux-ci ont exposé en 1932 les mécanismes qui – selon eux et dans le contexte contemporain aux États-Unis – mènent inévitablement au fait que ceux qui détiennent la propriété de l'entreprise voient le contrôle effectif leur échapper :

1. le pouvoir économique se concentre dans des entreprises de plus en plus grandes ;
2. la propriété de ces entreprises se répartit entre un nombre grandissant d'actionnaires ;
3. le contrôle effectif des opérations est délégué à des managers.

[Becht, 1997] a cependant relevé que cette interprétation historique de la séparation entre propriété et contrôle concerne essentiellement une configuration particulière aux États-Unis (où l'actionnariat est généralement dispersé et la structure «plate») et qu'elle n'est pas appropriée au contexte européen. La séparation telle que définie par Berle et Means se manifeste lorsque l'actionnaire le plus important ne dispose que d'une fraction minimale du capital (1%) et qu'aucun n'est donc en mesure d'influer sur l'orientation du management.

De nombreux auteurs ont étudié les conséquences de la caractérisation des entreprises selon la distribution des parts détenues par les investisseurs, la nature de ceux-ci (privés ou publics, groupes familiaux, actionnaires stables ou occasionnels, etc.) sur les performances des entreprises et le retour financier que pouvaient en attendre les investisseurs. L'interrogation centrale est alors celle de la concentration de l'actionnariat et de la façon de l'évaluer. [Demsetz, 1983, Holderness, 2003, Edwards et Weichenrieder, 2009, Sjögren *et al.*, 2012].

[La Porta *et al.*, 1999], par exemple, posent les questions suivantes :

1. Quelle est la proportion d'entreprises à capital dispersé en comparaison de celles où un actionnaire dispose d'un degré de contrôle significatif ?

2. Lorsqu'un tel actionnaire existe, quel est sa nature (familiale, institution financière, pouvoirs publics, autres entreprises) ?
3. Quel mécanisme leur permet de maintenir ce contrôle ?
4. Qu'est-ce qui explique les différences entre les patterns de structure selon les pays ?

En relation avec ce dernier point, il ressort de la majorité des études qu'on peut distinguer de fortes orientations régionales, en particulier entre les pays dits de *common-law* et de *code-law*, l'influence du pouvoir judiciaire prédominant pour les premiers, du pouvoir législatif pour les seconds. [Leuz *et al.*, 2003] identifient ainsi trois groupes de pays :

1. les économies de type *outsider* dotées de marchés boursiers importants, à actionnariat dispersé, où les droits des investisseurs sont forts (ex. le Royaume-uni et les États-Unis (bien que cela ait été contesté par [Holderness, 2009]) ;
2. les économies de type *insider* dotées de marchés boursiers plus faibles, à actionnariat concentré, faible protection de l'investisseur mais fort cadre réglementaire (ex. Allemagne et Suède) ;
3. les économies de type *insider* à faible cadre réglementaire (ex. Italie et Inde).

Dans le cas des sociétés intégrées dans des groupes, la structure en pyramide a été considérée comme une forme des mécanismes (qualifiés de *Control Enhancing Mechanisms*) permettant à des actionnaires de s'affranchir de la proportionnalité entre propriété et contrôle et qu'on présente sous l'expression consacrée de "*One share, one vote*" [Adams et Ferreira, 2008, ISSE *et al.*, 2007]. Les autres principaux mécanismes sont le recours aux participations croisées et les actions de différentes classes (*dual class equity*) [Bebchuk *et al.*, 2000].

3.2 L'évaluation du contrôle

Une proposition d'évaluation dénommée principe du lien le plus faible (*Weakest link principle*, *WLP*) a été présentée par [La Porta *et al.*, 1999] puis reprise entre autres par [Claessens *et al.*, 2000, Faccio et Lang, 2002]. Elle pose que, si entre un actionnaire et une entreprise existe une chaîne de participations dont le lien le plus faible dépasse un seuil déterminé, ce seuil est considéré comme le niveau de contrôle. Dans leur étude comparant la structure de propriétés de grandes corporations dans 27 pays, LaPorta et al. utilisent par exemple les seuils de 10% ou 20% de contrôle. Lorsqu'un actionnaire détient plus de 20% de droits de vote, il est qualifié de *ultimate owner*. Lorsqu'une entreprise ne dispose pas d'un tel actionnaire, elle est qualifiée de *widely held* (à actionnariat dispersé).

Cette méthode empirique a été critiquée dans une synthèse effectuée par [Edwards et Weichenrieder, 2009] ; ils y énumèrent différentes configurations où son application mène à des résultats manifestement non significatifs.

Une méthodologie plus solide permettant de prendre en considération les chemins multiples de propriété qui peuvent survenir dans les structure en pyramide a été développée autour d'Arianne Chapelle et d'autres dans une série d'articles [Becht *et al.*, 1999, Biebuyck *et al.*, 2005, Chapelle et Szafarz, 2005, Chapelle et Szafarz, 2007]. Elle est fondée sur un théorème de la théorie des graphes¹ qui veut que l'entrée a_{ij}^n d'une matrice A_G^n obtenue par la puissance n de la matrice d'adjacence A_G soit égale au nombre de chemins $i \rightarrow j$ de longueur n . Lorsque le graphe ne comporte pas de cycle, il vient un moment où l'exponentiation de cette matrice ne comporte plus que des 0 ; dans le cas contraire, l'exponentiation peut se poursuivre et la matrice comportera des valeurs croissant vers l'infini. Lorsqu'un réseau de participation est représenté par une telle matrice et que les 1 initiaux sont remplacés par les poids des participations, la combinaison de leurs produits le long des différents chemins rendra fidèlement le niveau de propriété dans les filiales directes et indirectes.

La première étape de la méthode consiste à calculer à partir de A la matrice Y de propriété intégrée qui est la somme des matrices de puissances successives correspondant aux chemins de longueurs croissantes. Elle est résolue par l'équation :

$$Y = \sum_{\alpha=1}^{\infty} A^{\alpha} = A(I - A)^{-1}$$

Dans l'exemple de la figure 3.1, on constate qu'une telle matrice introduit malheureusement un double comptage des participations détenues directement par un intermédiaire (le sommet 4 dans cet exemple) et indirectement par les détenteurs ultimes, ce qui a pour effet que la somme de la colonne 5 dépasse 100%.

Pour éliminer ces doubles comptages, Chapelle établit une matrice D qui est la diagonale de $(I - \bar{A})$ avec $\bar{A} = [a_j = \sum_i a_{ij}]$. Autrement dit, D contient sur la diagonale, pour chacun des sommets, les participations qui ne sont pas détenues à l'intérieur du système. Sur la figure 3.1e, on voit par exemple que 5, détenu à 90% par 1 et 4, est associé à une valeur de 10% dans cette matrice.

La matrice de participation intégrée est alors $V = DY = D(I - A)^{-1}A$. On peut constater qu'à la figure 3.1f, la somme de chaque colonne ne dépasse plus 100%. *«En pré-multipliant Y par la matrice diagonale des participations hors-système, on considère seulement les parts détenues aux niveaux successifs par les détenteurs externes, et on efface (c.-à-d. qu'on multiplie par zéro) les participations déjà prises en compte aux niveaux des détenteurs internes»* [Chapelle *et al.*, 2002].

La finalité de cette étape est de ré-allouer le contrôle détenu par les intermédiaires au niveau des détenteurs supérieurs et, au final, de déterminer celui des détenteurs ultimes. À ce stade du processus, on n'a cependant pas établi de distinction entre la propriété et le contrôle, c.-à-d. qu'on n'a pas tenu compte des effets de seuil qui permettent de

1. Mentionné par exemple par [Skiena, 1990], p.193.

contrôler intégralement une société détenue à plus de 50%. Pour y parvenir, la matrice A utilisée jusqu'à présent est remplacée par une matrice $AC = [ac_{ij}]$ dite de contrôle

direct, obtenue de la façon suivante à partir de A : $ac_{ij} = \begin{cases} 100 & \text{si } ac_{ij} \geq 50, \\ 0 & \text{si } \exists k \neq i \text{ et } a_{kj} \geq 50, \\ a_{ij} & \text{autrement} \end{cases}$

Une matrice IC , dite de contrôle intégrée, est alors calculée à partir de AC comme V l'avait été à partir de A . La figure 7.1 montre le résultat sur le même réseau que celui utilisé pour la propriété intégrée.

Nous clôturons cette section en remarquant que malgré le fait que la problématique du contrôle ait été l'objet de nombreuses recherches, les méthodes d'évaluation du taux de contrôle ne sont pas, à notre connaissance, d'une utilité réelle dans la pratique de la gestion des sociétés. Il serait intéressant de disposer de témoignages sur le déroulement des assemblées générales qui sont le lieu où se dénoue en théorie l'équilibre des parties en présence, mais il est probable que dans la majorité des cas, «les jeux sont faits» et l'issue bien connue des participants et représentants des sociétés détentrices des parts majoritaires, qui se concertent au préalable au sein des conseils d'administration². De temps à autre, des assemblées générales houleuses sont médiatisées³ du fait de la mobilisation de petits actionnaires floués, mais ces barouds d'honneur ne changent généralement rien au cours des événements.

En fait, les méthodes d'évaluation du contrôle ne sont pas destinées à l'examen particulier d'une société ; leur intérêt est essentiellement de pouvoir évaluer de façon statistique l'impact du cadre réglementaire régional et d'autres variables régionales ou conjoncturelles sur la gouvernance des sociétés. Les travaux qui y ont recours examinent généralement la distribution d'un rapport entre propriété et contrôle ainsi que selon le type d'actionnaires.

Dans le cadre de la partie applicative de ce mémoire (5.5), nous n'en ferons d'ailleurs pas usage, le seul recours à une partie de cette méthode sera l'utilisation des poids des participations dans l'exponentiation de la matrice d'adjacence pour obtenir les participations indirectes.

2. Les recherches portant sur le *petit monde* des conseils d'administration (voir section 3.3.1) constituent sans doute une piste plus alléchante pour qui serait tenté de chercher la localisation du pouvoir économique.

3. Certaines sont mémorables et font le bonheur des polémistes.

«Du jamais vu ! L'assemblée générale de LVMH était convoquée à 10h30 hier matin au Carrousel du Louvre. Contrairement à ce qui se passait les années précédentes, LVMH a organisé une ségrégation entre ses actionnaires : les petits actionnaires (auxquels on avait remis un badge blanc) n'avaient pas accès à la salle de l'assemblée générale (ouverte seulement aux porteurs de badges rouges ou verts) et étaient relégués dans un local annexe doté de deux écrans audiovisuels. Plus grave : il ne leur a pas été possible de poser des questions, aucun micro n'étant disponible dans ce local, aucune liaison n'étant établie entre ce local et la salle, dont l'accès leur demeurerait interdit même pour poser des questions.»

“LVMH muselle ses petits actionnaires”, 19 avril 2013, G. Beisson, sur le site de Mediapart.

<https://blogs.mediapart.fr/georges-beisson/blog/190413/lvmh-muselle-ses-petits-actionnaires>

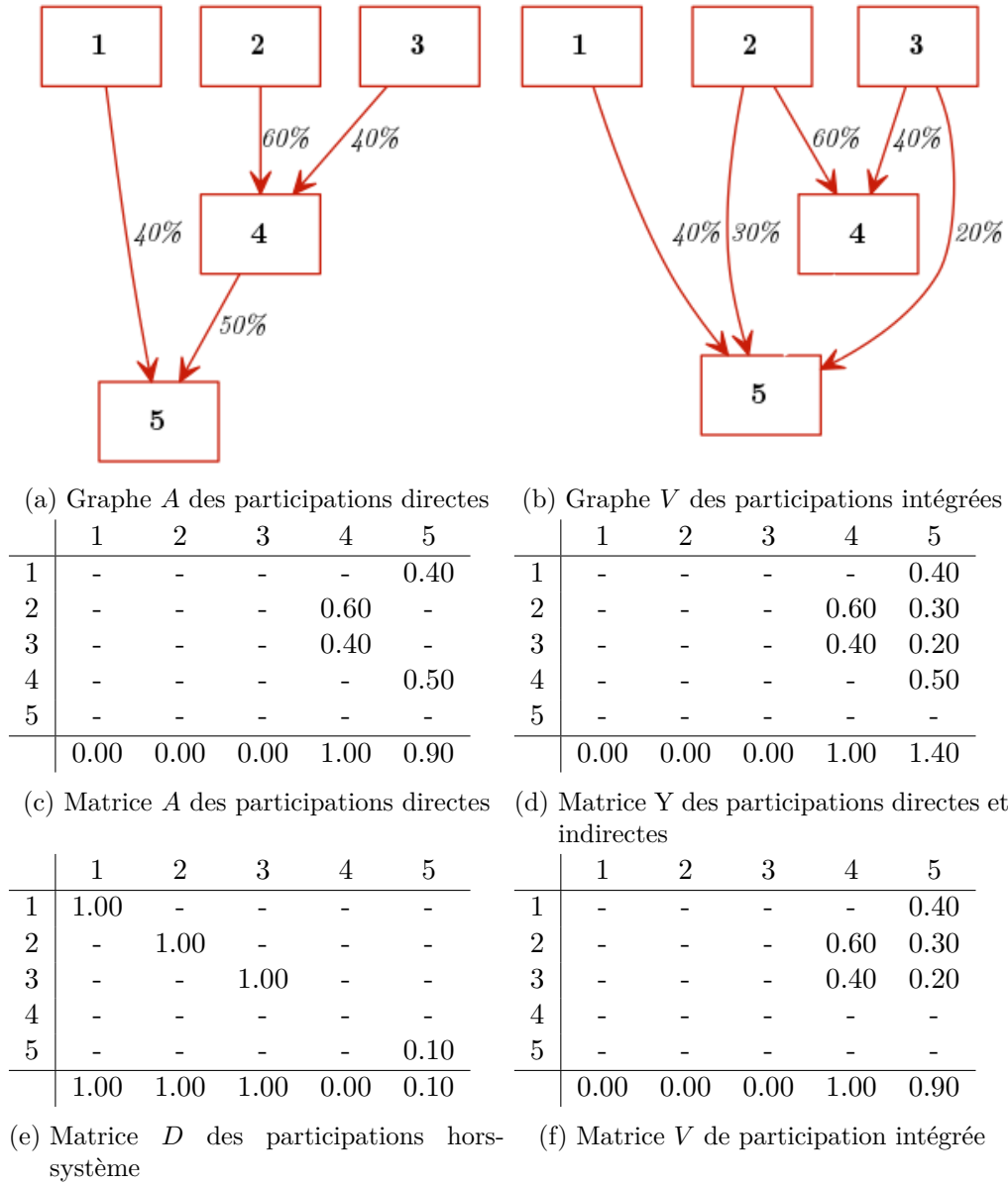


FIGURE 3.1 – Exemple de calcul de participation intégrée. En (d) le double comptage des parts détenues par les intermédiaires et les détenteurs ultimes entraîne un total supérieur à 100% pour la colonne 5. En (f) on a retiré les participations de 4, qui est intégralement détenu par d'autres acteurs du système, comme on le voit sur la diagonale de D .

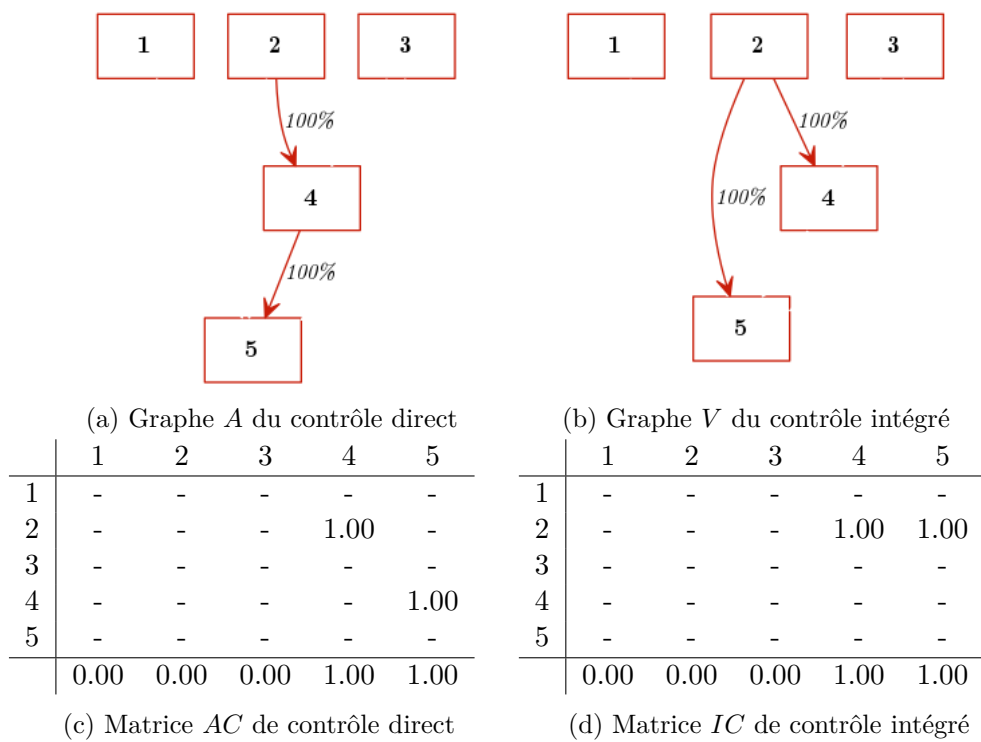


FIGURE 3.2 – Exemple de calcul de contrôle intégré.

3.3 Les réseaux financiers comme cas particulier de réseau complexe

Alors que la théorie des graphes est construite au départ d'un concept simple et abstrait – un graphe est un ensemble de sommets reliés par des liens – la théorie des réseaux en est une descendance tardive fondée en grande partie sur l'examen de réseaux du monde réel dont l'avènement a été permis par la disponibilité de vastes ensembles de données sous forme numérique ([Barrat, 2013] qualifie les deux dernières décennies d'«ère de l'avalanche de données»). En plus de présenter des propriétés absentes des réseaux aléatoires théoriques, ces réseaux empiriques ont révélé le phénomène d'*émergence*. Ce terme décrit une propriété observable au niveau macroscopique d'un système et qui découle des interactions entre les éléments ou des sous-ensembles réduits de ce système alors que ceux-ci ne la présentent pas. La présence de telles propriétés est un trait caractérisant la complexité d'un réseau[Boschetti *et al.*, 2005]. Une définition proposée par Barrat est qu'un réseau complexe est «un système auto-organisé dont l'évolution fait émerger des propriétés non prévues au départ»⁴, mais en pratique, on qualifie ainsi la plupart des réseaux empiriques comprenant un nombre élevé de nœuds et de liens. La littérature francophone utilise parfois l'expression *graphe de terrain*.

Les réseaux empiriques avaient cependant été étudiés de longue date dans le domaine de la sociologie; on en trouve dès 1934 dans Moreno sous forme de *sociomatrices* ou de *sociogrammes* qui correspondent, dans la terminologie propre à cette discipline, à la matrice d'adjacence et à la représentation sagittale d'un graphe. Le développement de l'analyse des réseaux sociaux (*SNA*, *Social Network analysis*) a été mené sur des données de taille réduite⁵ mais a permis d'établir l'usage d'une méthodologie décrite dans les ouvrages de référence tels que [Wasserman et Faust, 1994, Hanneman et Riddle, 2005b, Knoke et Yang, 2008], résumée par exemple dans [Butts, 2008]. Les principaux acquis en sont une panoplie de mesures qui ne reposent plus sur des caractéristiques intrinsèques aux agents observés mais relèvent de la *topologie* de leurs relations. De par cette généralité, elles ont pu être transposées aux réseaux complexes quelle qu'en soit la nature. Un ensemble de mesures est ainsi généralement appliqué lors de l'analyse d'un réseau [Costa *et al.*, 2007, Takemoto et Oosawa, 2012, Zafarani *et al.*, 2014]; on en trouvera également un exemple associé à la collection de réseaux de la Koblenz Network Collection [Kunegis, 2013].

On peut considérer que la littérature consacrée aux réseaux complexes relève de deux approches opposées sans être antagonistes; dans un cas les réseaux ont constitué un objet de recherche par eux-mêmes et l'on a cherché des cas d'application de leurs modèles dans diverses disciplines; dans l'autre des praticiens ont intégré cette science nouvelle des réseaux dans leurs travaux comme un outil supplémentaire enrichissant la recherche

4. rapportée par L. Beauguitte sur <http://groupefmr.hypotheses.org/>

5. Les auteurs recourent fréquemment à quelques réseaux bien connus dans la littérature : celui des familles florentines de la Renaissance utilisé par Padgett, le Karaté Club de Zachary, les *High-tech managers* de Krackardt, les organisations bureaucratiques de Knoke, etc.

spécifique à leur discipline.

Pour le premier cas, dans l'un des articles les plus abondamment cités et qui passe en revue l'ensemble des développements de cette discipline [Newman, 2003b], on trouve un relevé de quatre catégories assez lâches de réseaux du monde réel : les réseaux sociaux, les réseaux d'information, les réseaux technologiques et les réseaux biologiques. On n'y mentionne aucun sujet économique ou financier dans une bibliographie pourtant vaste. Il en est de même pour ceux de [Albert et Barabási, 2002] et [Boccaletti *et al.*, 2006].

Les disciplines économiques et financières semblent ainsi avoir été nettement moins abordées que d'autres, ou plus tardivement, vraisemblablement à cause d'une moindre disponibilité des données. Nous présentons rapidement quelques réseaux issus de celles-ci dans la section suivante avant de passer à ceux de participations proprement dits.

3.3.1 Les réseaux économiques et financiers en général

Parmi les types de réseaux qu'on rencontre dans les domaines économique et financier, [Battiston *et al.*, 2010] établissent une distinction entre les réseaux d'*interaction* et ceux de *similarité*. Dans les premiers, un lien entre deux sommets dénote une interaction effective telle qu'une relation entre créancier et débiteur, un transfert de marchandises ou de cash-flow, etc. Dans les seconds, c'est le fait de présenter un caractère commun qui détermine la présence d'un lien, sans que les sommets ne soient nécessairement liés par une activité. Ce type de graphe implique généralement d'effectuer des choix méthodologiques quant à la mesure de similarité et au niveau (*threshold*) à retenir pour établir un lien. Notons qu'à partir d'un graphe d'interaction, il est souvent possible d'en induire un autre de similarité. C'est en particulier une façon de passer d'un graphe bipartite à un autre où tous les sommets sont de même nature (voir plus loin le cas des réseaux de conseils d'administration).

1. **Réseaux de corrélation de valeurs boursières** (*stock correlation networks, financial networks*). Les fluctuations des prix auxquels les valeurs sont échangées sur les marchés boursiers sont enregistrées sur une période donnée et leur corrélation détermine une distance entre les valeurs. Un clustering (éventuellement sous forme d'un arbre couvrant minimal) est effectué sur cette base [Bonanno *et al.*, 2004, Mandere, 2009].
2. **Réseaux des conseils d'administration** (*corporate interlock networks*). Ces recherches concernent la présence d'administrateurs communs aux conseils de plusieurs sociétés. Techniquement, au départ d'un graphe bipartite rassemblant des personnes physiques et les entreprises où elles siègent, on construit des graphes non dirigés en reliant par des arêtes :
 - d'une part les administrateurs qui siègent ensemble dans un ou plusieurs conseils (les administrateurs d'un même conseil forment alors une clique) ;
 - de l'autre les sociétés qui ont en commun un ou plusieurs administrateurs. Cf. figure 3.3.

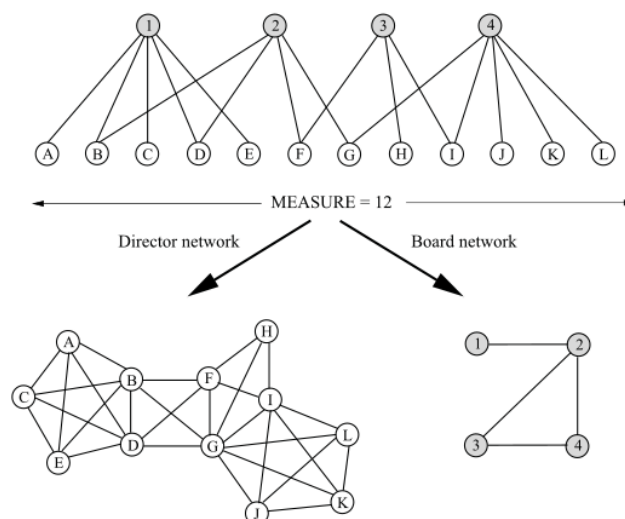


FIGURE 3.3 – Les administrateurs ($A \dots L$) siègent aux conseils des sociétés $1 \dots 4$. On déduit du graphe bipartite les graphes *one-mode* des administrateurs et des sociétés [Caldarelli, 2007].

Généralement associées au thème du «petit monde», ces recherches assez abondantes se situent à la frontière de la sociologie, de la gouvernance d'entreprise et de la politique du fait qu'elles touchent à l'hypothèse controversée (bien qu'il s'agisse d'une interrogation légitime) qu'une élite sociale restreinte concentrerait le pouvoir économique [Scott *et al.*, 1992, Mizruchi, 1996, Kogut et Walker, 2001].

3. **Réseaux de dettes inter-bancaires** Les prêts mutuels entre les banques ont été étudiés dans un esprit de prévention des risques systémiques et de la prévention des contagions [Minoiu et Reyes, 2013]. On trouve une compilation sur le sujet dans [Wims *et al.*, 2011] ou [Hüser, 2015].
4. **Réseaux d'échanges commerciaux.** Le réseau mondial des échanges commerciaux (*WTN*, *World Trade Network*) peut être construit en assignant à chaque sommet un pays et en ajoutant des arcs pour les échanges, soit de façon binaire, soit éventuellement pondérés par les volumes ou les valeurs échangées [Garlaschelli et Loffredo, 2004, De Benedictis et Tajoli, 2011]. Des réseaux plus particuliers (*input-output networks*) concernent les échanges entre secteurs d'une économie au sein d'un pays. Dans [Blöchl *et al.*, 2011], l'examen des indices de centralité révèle les secteurs principaux pour chaque pays et permet par un procédé de clustering de rassembler les pays présentant des profils de production similaires.

3.3.2 Les réseaux de participations

Un des premiers articles consacrés à l’analyse des réseaux de participation semble être celui de [Berkowitz *et al.*, 1979] qui en 1979, au départ d’un ensemble de 5306 sociétés, tente d’en dégager des regroupements en combinant les liaisons de participations et celles entre les membres des conseils d’administration. Datant d’une époque où les réseaux complexes n’étaient pas encore abondamment explorés, il s’agit plutôt d’un regard de sociologue essayant de déterminer ce qui constitue les *patterns* de contrôle entre les sociétés et de les quantifier. On y trouve une ébauche des méthodes de calcul matriciel permettant de déterminer le contrôle au travers d’une ou plusieurs *chaînes de contrôles* telles qu’elles seront développées par la suite dans les recherches sur la séparation entre propriété et contrôle mais la méthode exposée ici ne semble pas avoir eu de postérité. L’examen simultané des liens de participations et de ceux des conseils d’administrations sera cependant repris par d’autres [Piccardi *et al.*, 2010, Ferraro *et al.*, 2012].

Dans le contexte de l’éconophysique, [Souma *et al.*, 2004b, Souma *et al.*, 2006b, Aoyama *et al.*, 2010] ont effectué une large série d’observations sur un ensemble de sociétés cotées sur le marché japonais et leurs participations, soit un ensemble de 3152 sommets et 23064 arcs [Souma *et al.*, 2004a]. Ils ont étudié les motifs apparaissant dans la matrice d’adjacence, la distribution des degrés, la corrélation entre degré et coefficient de clustering. Dans des articles ultérieurs, ils ont procédé de même pour un réseau plus réduit établi à partir des entreprises du secteur automobile japonais [Souma *et al.*, 2005, Souma *et al.*, 2006a] en y ajoutant l’examen de la corrélation entre le degré et l’âge des sociétés ou la valeur de leurs actifs, abordant ainsi la dynamique du réseau, dont l’évolution est liée à la croissance des sociétés. La même démarche a été adoptée concernant comme précédemment l’ensemble construit à partir des sociétés cotées japonaises [Fujiwara *et al.*, 2006]. Tous ces articles mettent en évidence l’existence d’une *loi sans échelle* ; l’introduction du dernier fournit d’ailleurs une analogie particulièrement éclairante du phénomène :

“If a firm’s size were to be regarded as human height, anyone could easily see that a population of firms would look quite peculiar and striking. From a distance, only a few giants can be identified, while a large number of dwarves are visible only faintly. Removing the giants from the population and looking closely, one would have a similar landscape with a rescaled dimension.”

Cette loi sans échelle (*power law* ou *scaling law*) est caractérisée par une fonction de la forme $y = kx^\gamma$ qui, dans une représentation à l’échelle logarithmique (qualifiée communément de *log-log*), prend une forme rectiligne dont la pente est caractérisée par l’exposant γ (le facteur k n’est pas significatif et ne fait que décaler la droite). Elle affecte des domaines de nature extrêmement variée, y compris économiques. [Glabaix, 2008] mentionne notamment la distribution de la taille des entreprises ou des villes, les salaires des dirigeants (proportionnels à S^γ où S est la taille de l’entreprise), certaines caractéristiques de l’activité des marchés boursiers, etc.

[Souma, 2007] a également étudié les relations entre entreprises en construisant ce qu’il dénomme un *multigraphe*⁶ : partant de l’idée que ces relations sont de nature diverse, il en a rassemblé quatre représentant les participations, l’existence d’administrateurs communs (*interlockings*), les transactions et le partage de brevets. Il en résulte un graphe non-dirigé où le poids des arêtes varie entre 1 et 4 selon le nombre de relations entre les entreprises concernées. Quoiqu’assez peu formel, ce graphe comprenant 3576 entreprises donne un aperçu de la force des relations qui existent entre elles au sein de différents secteurs économiques et permet d’établir quelques indices dont les propriétés sont à rapprocher de celles d’un graphe constitué uniquement de participations.

Par leur approche générique, ces travaux constituent un fil conducteur que nous suivrons lors de l’analyse des données dont nous disposons ; nous reviendrons donc en détail sur l’ensemble de leurs résultats dans cette section (pages 45 et suivantes).

[Caldarelli *et al.*, 2004, Garlaschelli *et al.*, 2005] ont semblablement étudié le caractère sans échelle de trois réseaux composés de sociétés cotées sur les marchés boursiers italien (MIB) et américain (NYSE et NASDAQ). À la différence de l’article précité, ils se sont basés sur un caractère strictement topologique : la distribution des degrés entrants et sortants. Ils ont avancé l’hypothèse que ce caractère découle d’un attribut non topologique des sommets qui les rend attractifs – dénommé *fitness* – et qu’on peut associer intuitivement à la perspective plus ou moins bonne de récolter des dividendes en retour de l’investissement.

Pour expliquer une telle distribution, deux modèles principaux ont été avancés. Le premier est celui de *l’attachement préférentiel* [Barabási *et al.*, 1999] où les sommets ajoutés à un réseau se raccordent de préférence à ceux disposant déjà d’un degré élevé. Ce modèle parfois qualifié de “*rich gets richer*” s’applique au réseau de liens du WWW ainsi qu’à celui des citations entre articles scientifiques. Le second est le *fitness model*, *hidden variable* ou “*good get richer*” [Garlaschelli *et al.*, 2005] où, en l’absence de croissance, une probabilité de connexion entre les sommets s’effectue en fonction d’un caractère non topologique⁷.

Les auteurs ont jugé que l’attachement préférentiel n’était pas pertinent en l’absence de croissance du réseau – du moins pas au rythme requis par ce modèle – et parce qu’il aurait impliqué que les actions introduites sur le marché aient été acquises par les actionnaires détenant les portefeuilles les plus étendus.

Dans leur interprétation⁸, le degré $(k_{out})_i$ correspond au nombre de participations dans

6. L’acception usuelle de ce terme est un graphe présentant des arêtes parallèles, c-à-d. que deux sommets peuvent éventuellement être reliés par plus d’une arête. Ici, les arêtes parallèles sont fusionnées pour en faire un graphe simple pondéré.

7. [Garlaschelli *et Loffredo*, 2004, Garlaschelli *et Loffredo*, 2005] avaient appliqué la même analyse au réseau international des échanges commerciaux et avaient attribué le rôle de variable cachée au produit intérieur brut (*gross domestic product*). On trouvera par ailleurs un résumé de la problématique de la cause d’un caractère sans échelle dans la section “*Models of network growth*” de [Newman, 2003b] ou dans “*Graph Generators*” [Chakrabarti *et Faloutsos*, 2006].

8. Où les arcs sont dirigés de la société détenue vers les actionnaires. La *portfolio diversification* est en fait le degré entrant dans l’article original mais ce choix étant l’inverse de la majorité des études, nous

des sociétés différentes et est dénommé *portfolio diversification*. Le poids des arcs W_{ij} étant le pourcentage détenu par i dans j et C_j étant la capitalisation boursière de j (le produit de l'ensemble des actions en circulation par leur valeur sur le marché), ils définissent une quantité $v_i = \sum_j W_{ij} C_j$ comme le *portfolio volume* ou la richesse (*wealth*) de i . Les valeurs γ et α étant les exposants associés respectivement aux distributions de k_{out} et v et obtenus de façon empirique, ils remarquent qu'une valeur d'échelle $\beta = (1 - \alpha)/(1 - \gamma)$ détermine la relation entre k_{out} et v . La valeur de *fitness* déterminant le caractère sans échelle est donc la valeur financière détenue au moyen des participations directes.

Les mesures de centralité sont des indices fréquemment utilisés pour déterminer le «pouvoir» d'un sommet, particulièrement en analyse des réseaux sociaux [Hanneman et Riddle, 2005a]. Cependant, comme les variantes en sont nombreuses et qu'il n'y pas d'unanimité au sujet de la notion de pouvoir, le choix de l'indice et son interprétation doivent être menés avec prudence. [D'Errico et al., 2009] ont étudié la centralité dans le réseau de participations de 223 sociétés cotées italiennes et de leurs filiales. Dans un premier temps, ils définissent ce réseau comme une matrice $W = [w_{ij}]$ où w_{ij} n'est pas la fraction du capital de j détenue par i (comme il est généralement établi) mais son produit par la valeur de marché de j . Mais ils remarquent que les liens de participation peuvent véhiculer de l'*information*. Lorsqu'on les considère sous ce rôle, ils ne doivent être associés ni à une direction ni à un poids; autrement dit, le réseau est alors un graphe non dirigé qui peut être représenté par une matrice d'adjacence symétrique

$$A = [a_{ij}] \begin{cases} a_{ij} = 1 \Leftarrow w_{ij} > 0 \\ a_{ij} = 0 \Leftarrow w_{ij} = 0 \end{cases}$$

La *betweenness centrality* qui est basée sur le nombre de chemins les plus courts qui passent par un sommet révèle les sociétés où transite le maximum d'information, par où se propageront les chocs affectant le réseau⁹.

Dans un second temps, ils relèvent que le fait d'intégrer la valeur des filiales dans w_{ij} permet de quantifier au niveau des arcs une *capacité* du *flot* remontant vers les sociétés parentes. Il en découle logiquement un recours à la *flow betweenness centrality* des sommets. Cette mesure de centralité utilise un modèle assimilant le réseau à un réseau électrique où le courant transitant par un sommet dépend de la résistance des chemins entrants et sortants; prenant en considération l'ensemble des chemins transitant par un sommet, elle reflète le fait que le cash-flow n'emprunte pas nécessairement le chemin le plus court. Les auteurs assimilent cette centralité au contrôle, négligeant les mécanismes de seuil qui induisent les différences entre propriété et contrôle qui ont été exposés à la section 3.1.

avons inversé leur notation pour rester cohérent avec les autres,

9. La nature de ces chocs est la volatilité des actions sur le marché. Les sociétés centrales amortissent cette volatilité et absorbent les chocs, sans doute du fait de leur capitalisation plus importante; les 4 sociétés présentant la plus haute centralité associée à une faible volatilité sont des banques.

3.3.3 “The Network of Global Corporate Control”

[Vitali *et al.*, 2011] ont innové en présentant un article qui présentait pour la première fois une investigation sur l’architecture globale du réseau des participations. Alors que les travaux qui avaient précédé n’abordaient que des ensembles nationaux et souffraient de l’absence d’une méthodologie appropriée pour évaluer le contrôle, Vitali et al. ont synthétisé et comparé différents modèles de contrôle et, en les appliquant au réseau global, en ont tiré des observations qui ont engendré des débats d’une ampleur inattendue.

Pour construire le réseau en question, ils ont extrait de la base de données Orbis du Bureau Van Dijk¹⁰ les entreprises considérées comme transnationales selon la définition de l’OECD¹¹ en se limitant à celles qui détenaient des parts supérieures ou égales à 10% dans deux états ou plus. Il en résulte une liste de 43060 TNCs, parmi lesquelles 5675 TNCs cotées sur des marchés boursiers. Au départ de cette liste, ils ont parcouru de façon récursive les liens vers les filiales au moyen d’une recherche en largeur et procédé de même – en parcourant les liens entrants à rebours – pour les sociétés parentes. Le réseau obtenu est composé de trois niveaux dont les membres sont qualifiés de

1. *Shareholders* (SH) ;
2. *Transnational Corporations* (TNC) ;
3. *Participated Companies* (PC).

Les pyramides inférieures et supérieures peuvent évidemment comprendre des éléments communs, donnant naissance à des composantes (faiblement) connexes de tailles diverses.

Une composante géante se dégage et présente une structure de type *bow-tie* telle que celle qui a été décelée par [Broder *et al.*, 2000] dans le graphe du web. Ce type de structure se caractérise par un découpage en différentes parties définies en terme de «joignabilité» (*reachability*) autour d’un noyau central fortement connexe (il existe un chemin joignant toute paire de ses sommets)¹². La structure du réseau global des participations présente cependant des caractéristiques qui la distinguent de celle du web :

- Elle est fortement asymétrique, le rapport entre le nombre de sommets de la section des liens entrants et celui de la section des liens sortant étant de $\frac{5616}{324561} = 0.017$;
- Le noyau fortement connexe que les auteurs dénomment *cœur* est de taille réduite par rapport à l’ensemble de la partie connexe $\frac{1318}{463006} = 0.0037$;
- Ce cœur est également très dense en liaisons, ses sommets présentant en moyenne 20 liaisons avec les autres.

10. <http://www.bvdinfo.com/en-gb/our-products/company-information/international-products/orbis>

11. “[...] comprise companies and other entities established in more than one country and so linked that they may coordinate their operations in various ways, while one or more of these entities may be able to exercise a significant influence over the activities of others, their degree of autonomy within the enterprise may vary widely from one multinational enterprise to another. Ownership may be private, state or mixed.”

12. Voir en annexe (section 7.3) une description plus formelle de ce type de structure.

Vitali et al. ont récapitulé les méthodes existantes pour estimer le contrôle détenu par un sommet du réseau sur ses successeurs :

1. Dans le modèle LM (*Linear Model*), ils considèrent qu'il n'y a pas d'écart entre propriété et contrôle. Par conséquent, les valeurs de la matrice L sont les mêmes que celles de la matrice de propriété : $L_{ij} = W_{ij}$;
2. Dans le modèle TM (*Threshold Model*), ils considèrent que le contrôle est totalement détenu par une société-mère lorsque ses parts dépassent un seuil t déterminé (50% selon [Chapelle et Szafarz, 2005], 10% ou 20% selon [La Porta et al., 1999]) :

$$T_{ij} = \begin{cases} 1 & \text{si } W_{ij} \geq t, \\ 0 & \text{si } \exists k \neq i \text{ et } W_{kj} \geq t, \\ W_{ij} & \text{autrement} \end{cases}$$

3. Dans le modèle RM, le contrôle est basé sur la part relative que chaque société-mère détient par rapport aux autres¹³, définie par : $R_{ij} = \frac{W_{ij}^2}{\sum_{l=1}^{k_{in}^j} W_{lj}^2}$

Ils y ont apporté des corrections pour corriger la surestimation du contrôle conféré aux sommets de la partie *IN*. En effet, on a vu (figure 2.10) que dans les structures dont le noyau comprend des participations croisées, tout le contrôle peut revenir aux sommets entrants. Cette interprétation repose cependant sur une estimation des flux dans une perspective de temps infinie, ce qui n'est guère réaliste.

Les auteurs ont alors établi une comparaison

- entre la « valeur de portefeuille » détenue par les sociétés, les « valeur sous contrôle », selon les trois modèles précités, en triant les sociétés par ordre décroissant (selon leur *rang*) ;
- et la fraction cumulée de la valeur totale détenue par les sociétés jusqu'au rang n .

Quel que soit le modèle de contrôle, il s'avère que le contrôle détenu par les sociétés les plus importantes dépassait leur valeur de portefeuille, ce qui signifie que le contrôle est plus concentré que la richesse ; 80% de la valeur des TNCs se trouve détenue par seulement 737 sociétés essentiellement financières localisées dans le cœur de la structure. Plus précisément, 40% de la valeur est située sous le contrôle d'un groupe de 147 TNCs qui « *par un réseau complexe de relations de participations, dispose d'un contrôle presque total sur lui-même* ».

13. L'indice ainsi obtenu a été précédemment expliqué par [Glattfelder et Battiston, 2009] et s'inspire de la théorie des jeux. Un détenteur i d'une participation W_{ij} minoritaire peut disposer d'un certain contrôle dans la mesure où les autres participations sont dispersées et permettent difficilement de former une alliance pour le dépasser. Inversement, sa position est beaucoup moins forte si le second détenteur le suit de peu et peut s'allier avec un seul autre détenteur quelconque pour y parvenir.

Les auteurs concluent sur une sorte de mise en garde relative aux conséquences d'une telle situation. Ils s'interrogent sur ses effets en terme de stabilité du système financier global et sur la concurrence sur les marchés, la plupart des TNCs couvrant des domaines d'activités multiples.

Cet article a connu une diffusion exceptionnelle ; l'essayiste Paul Jorion a exprimé sur son blog son inquiétude devant les conclusions de l'article en intitulant son billet ¹⁴ de manière lapidaire «Les maîtres du monde», ce qui a, selon l'expression consacrée, «enflammé la blogosphère». En dehors de multiples billets qui trouvaient dans cet article une caution scientifique à des opinions voisinant avec le complotisme, l'article a divisé les éditorialistes de la presse conventionnelle et a suscité des attaques envers la méthodologie suivie qui ont amené l'un des coauteurs, James Glattfelder, à présenter des arguments complémentaires sur son blog et à y souligner la neutralité de leur démarche «*dénuée de tout agenda politique ou de volonté de promouvoir des théories conspiratrices*» [Glattfelder, 2013].

On retiendra des interventions et commentaires publiés dans la presse ceux contestant un plan délibéré de concentration : «La forte concentration du réseau de détention ne reflète pas le contrôle des entreprises par la finance, mais le fonctionnement des marchés financiers : les investisseurs, pour minimiser le risque, préfèrent suivre le flux des investissements déjà réalisés» ¹⁵. Celui-là parmi d'autres évoquait sans les nommer les principes d'attachement préférentiel décrits précédemment.

14. <http://www.pauljorion.com/blog/2011/09/07/les-maitres-du-monde-a-propos-de-the-network-of-global-corporate-control-par-s-vitali-j-b-glattfelder-s-battiston/> Il a publié sur son site une traduction en français de l'article et est revenu sur le sujet dans un essai ultérieur [Jorion, 2016] : «*Cette étude, à la méthodologie scientifique irréprochable [...] offrait les éléments permettant de confirmer une hypothèse qui, bien que susceptible d'éclairer les anomalies constatées dans le fonctionnement des démocraties, paraissait jusque-là extrêmement aventureuse, à savoir que les structures de pouvoir effectives sont exécutées aujourd'hui par ces compagnies transnationales, le pouvoir des nations souveraines étant désormais limité à la part que celles-ci sont encore disposées à leur déléguer.*»

15. Antoine Reverchon, «Un noeud de 147 sociétés au coeur de l'économie mondiale», Une du «Monde Economie» du mardi 29 novembre 2011.

4 Analyse

Dans cette section, nous effectuerons essentiellement le relevé de divers caractères topologiques du réseau utilisé communément pour caractériser un graphe. Nous suivrons en particulier les travaux de Souma et al. qui sont à notre connaissance les premiers à avoir appliqué à un graphe de participations les enseignements relatifs aux réseaux complexes.

4.1 Distribution des composantes connexes

Le réseau complet se caractérise par une abondance de composantes de très petite taille (de 2 à 5 sommets) et une composante géante comportant 14610 sommets. Dans une composante *fortement connexe*, il existe un chemin joignant chaque paire de sommets ; les composantes dont il est question ici sont *faiblement connexes*, c-à-d. que chaque sommet n'est assuré d'en joindre un autre que si l'on fait abstraction de la direction des arcs. Cependant, considérer de telles composantes est «*intéressant d'un point de vue économique parce que le flot de connaissance et d'information n'est pas contraint par la direction des liens*» [Vitali et al., 2011].

L'existence d'une composante géante dans les graphes aléatoires a constitué un sujet d'étude important ; on a en particulier mis en évidence un phénomène de *transition de phase*, une composante géante unique surgissant de façon subite lorsque le degré moyen des sommets dépasse 1.

[Easley et Kleinberg, 2010] ont expliqué de façon informelle pourquoi, lorsqu'un réseau comporte une composante géante, il n'en contient quasi certainement qu'une seule. L'émergence d'une telle composante survient lorsque toute connexion ajoutée au réseau aura vraisemblablement une de ses extrémités dans la composante ; dans ces conditions,

Réseau	$ V $	$ E $	$ PGCC $	r	$ CC $
Réseau entier	85857	71654	14610	0,17	22218
Sociétés cotées et connexes	2865	3326	1623	0,56	68
Vitali et al	600508	1006987	463006	0,77	23825
Graphe aléatoire	85857	71569	58039	0,81	20318

FIGURE 4.1 – Caractéristiques des composantes pour divers réseaux. $|V|$: nombre de sommets $|E|$: nombre d'arcs $|PGCC|$: taille de la plus grande composante r : ratio entre cette composante et l'ensemble $|CC|$ nombre de composantes.

si tant est que deux composantes géantes aient jusque-là coexisté, il est presque inévitable qu'une liaison apparaîtra entre les deux et les fusionnera.

Le réseau des sociétés belges présente une telle composante, mais elle est comparative-ment plus réduite que celle apparaissant dans des graphes aléatoires comparables¹, dans des variantes reconnectées du réseau² ou dans le réseau étudié par Vitali et al. Comme on l'a souligné précédemment, le nombre de composantes de très petite taille est important ; on le constate d'autant plus vis-à-vis du réseau de Vitali et al. qui ne présente guère plus de composantes alors qu'il est près de 10 fois plus important. D'autre part, lorsqu'on observe l'évolution du réseau des sociétés belges depuis 1993 (figure ??), on constate que le rapport entre l'ensemble et la composante géante est remarquablement stable parallèlement au fait que le nombre de sociétés a plus que doublé.

Une hypothèse peut être avancée, qui tient à la façon dont ce réseau s'agrandit : alors que dans un graphe aléatoire, la densité augmente par l'ajout progressif d'arcs à un nombre de sommets invariable, le réseau des sociétés s'accroît en grande partie par l'ajout *simultané* de deux sociétés et de la participation qui les réunit³. Lorsqu'il s'agit de petites sociétés non financières qui ne cherchent ni à diversifier un portefeuille ni à augmenter leur capital par la prise de participations d'autres sociétés, il est fort probable que cette petite composante restera stable ou n'augmentera que par l'ajout d'une ou deux participations dans des filiales créées à leur initiative.

Malgré cette particularité, le caractère sans échelle de la distribution des tailles des composantes est parfaitement en phase avec les résultats de Vitali et al. Ils trouvent un γ de 2,13, celui des sociétés belges est de 2,25. Si l'on compare le profil du réseau entier avec celui d'un réseau aléatoire, celui-ci présente un γ plus élevé (3,4 en moyenne), une composante géante beaucoup plus importante et, corollairement, un nombre de tailles distinctes des composantes plus réduit (de l'ordre de 14).

Cependant, le profil du sous-ensemble des sociétés cotées est celui qui se rapproche le plus de celui tracé par Vitali et al., avec une amorce de courbe proche de l'horizontale – le γ demande ici une plus grande part de subjectivité – signifiant que les populations de composantes de tailles les plus faibles sont en nombres assez proches. Or ces deux réseaux sont en fait une sélection de sommets⁴ qui a scindé les composantes du réseau d'origine ; la distribution qui en résulte est donc biaisée de la même manière.

Des questions peuvent être posées quant aux caractères non topologiques des sociétés en fonction des composantes auxquelles elles appartiennent. Si la présence d'une compo-

1. On a utilisé le modèle d'Erdős-Rényi implémenté dans NetworkX par la méthode `fast_gnp_random_graph(n,p)` où n est le nombre de sommets et p la probabilité que deux sommets soient réunis ; les nombres figurant dans le tableau sont la moyenne de 100 itérations avec $p = 9,71 \times 10^{-6}$.

2. Cette variante consiste à redistribuer aléatoirement les liaisons entre les sommets tout en conservant pour ceux-ci les degrés entrants et sortants. On a veillé à conserver le poids des arcs entrants de façon à ce que la somme de ceux-ci ne dépasse pas 100% pour respecter leur sémantique.

3. Dans la réalité de la vie économique, le phénomène inverse se produit également : une participation et une société peuvent disparaître lorsque la société-mère réalise une « fusion par absorption ».

4. Cette sélection est effectuée de la même manière en partant d'un ensemble de sociétés et en extrayant leurs ancêtres et leurs descendantes.

sante géante s'explique par des phénomènes d'attachement préférentiel et de croissance qu'on peut associer aux sociétés présentant une dynamique de connexion plus active, *a contrario* les sociétés figurant dans les très petites composantes pourraient représenter une population où les relations sont stables et les liens plus importants. À priori, on peut s'attendre à ce qu'une société créant une filiale maintienne une participation majoritaire alors qu'une société diversifiant ses placements verra ses participations fluctuer selon les opportunités.

La figure 4.2 vérifie une telle tendance. De la même façon qu'à la figure 2.5, on y représente une courbe de densité des participations selon leur poids, mais en distinguant celles présentes dans les toutes petites composantes (de taille 2 ou 3), celles de la composante géante, et toutes les tailles intermédiaires. Les courbes sont remarquablement similaires, mais présentent un décalage qui s'inverse selon que les participations sont inférieures ou supérieures à 50%. Les sociétés des petites composantes ont ainsi tendance à présenter moins de petites participations, celles de la composante géante un peu moins de participations étendues.

4.2 Représentation matricielle du réseau

[Souma *et al.*, 2006b] représentent un réseau de 3152 sociétés sous forme de matrice carrée asymétrique où les rangées sont les sociétés-mères et les colonnes les filiales. Un point M_{ij} est tracé lorsque la société i détient une participation dans j . Les sociétés sont classées sur les deux axes selon un code qui définit une catégorie économique ; la densité des points décelable visuellement leur permet d'observer que :

- les participations sont distribuées dans toutes les catégories lorsque la société-mère appartient au secteur des banques ou des assurances ;
- la densité est relativement élevée dans chaque carré correspondant aux participations intra-catégorie, à l'exception des secteurs des banques et assurances.

Cette représentation du réseau sous forme de matrice d'adjacence constitue ainsi un moyen sommaire mais intuitif de mettre en évidence une forme d'assortativité par secteur d'activité.

Nous avons appliqué le même traitement au réseau des entreprises belges (sans leurs filiales étrangères, pour lesquelles on ne dispose pas de mention du secteur). Pour les sociétés belges, on dispose du code NACE qui définit les secteurs économiques selon une forme hiérarchique par une lettre de A à U suivie de 5 chiffres de telle sorte que :

- les *sections* sont indicées de A à U ;
- les *divisions* sont représentées par les 2 premiers chiffres du code ;
- les *groupes* selon les 3 premiers chiffres ;
- les *classes* selon les 4 premiers chiffres ;
- les *sous-classes* selon les 5 chiffres.

Les résultats corroborent ceux de Souma et al. D'une manière générale, il apparaît que :

- les zones denses horizontales correspondent à l'existence d'un secteur qui effectue des placements dans un éventail large d'activités («il ratisse large»);
- les zones denses verticales à l'existence d'un secteur où tous les autres investissent («il est très courtisé»).

L'observation la plus évidente est que le secteur qui investit de la façon la plus diversifiée est le secteur financier (K, à l'exception de la division «Activités auxiliaires de services financiers et d'assurance»). La sous-division des assurances (K66), à l'inverse des résultats de Souma et al., est nettement assortative.

On distingue de plus une diversification forte pour le secteur immobilier (L) et les «Activités des sièges sociaux, conseil de gestion» (sous-division de M. Les autres sous-divisions de M sont assortatives).

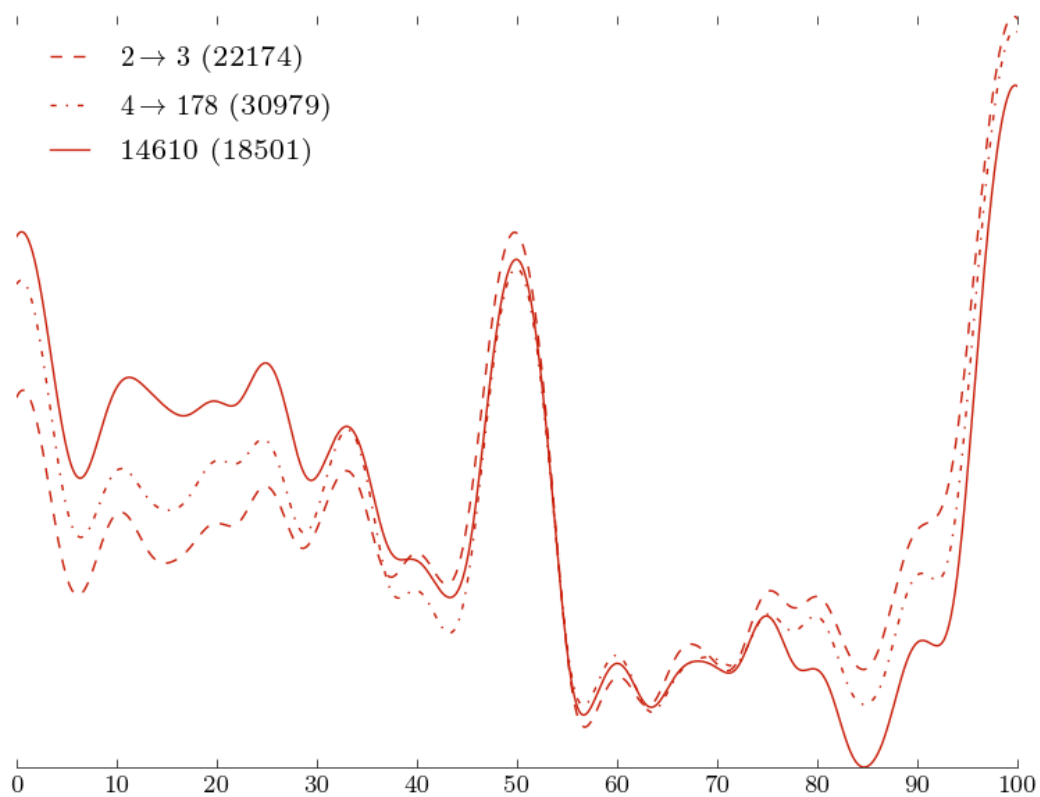
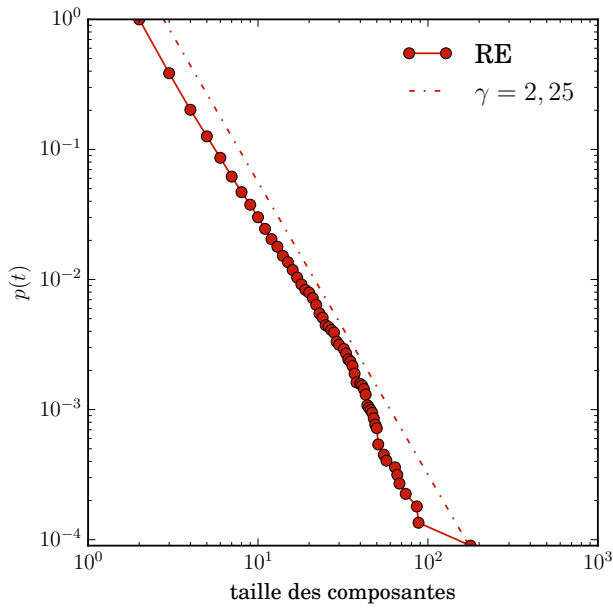
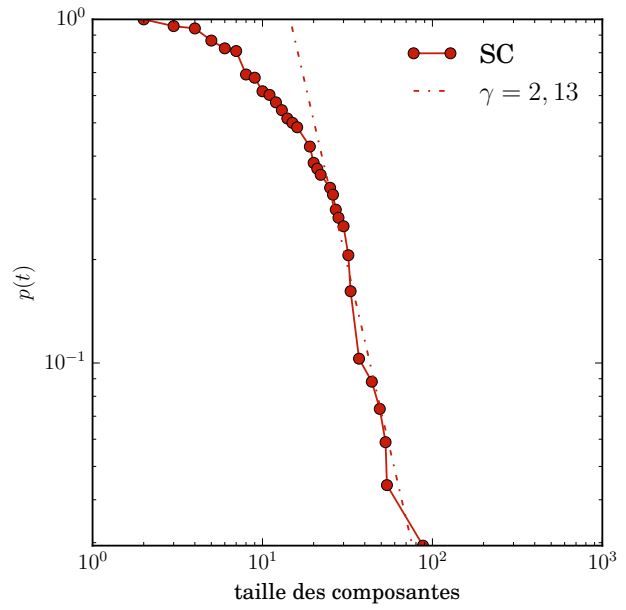


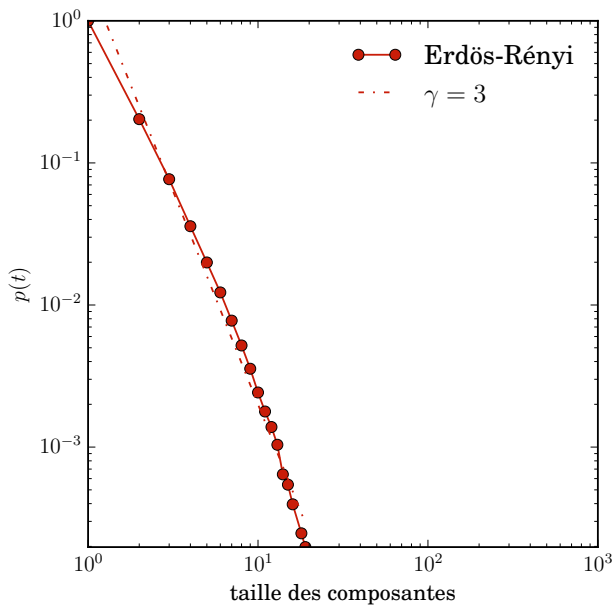
FIGURE 4.2 – Variations de la densité des différents degrés de participations au sein des composantes. La légende précise les tailles des composantes et le nombre de sociétés y correspondant. On observe une tendance à compter moins de faibles participations parmi les sociétés des petites composantes.



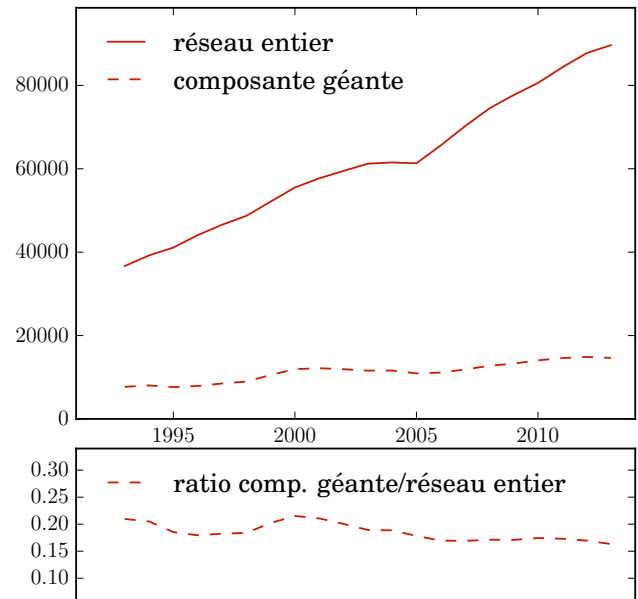
(a) Réseau entier



(b) Sociétés cotées et connexes



(c) Graphe aléatoire



(d) Évolution du réseau

FIGURE 4.3 – a) à c) Distribution cumulative des composantes connexes en fonction de leur taille. Dans les 3 cas, on a négligé le point représentant la composante géante du fait qu'il est complètement hors-continuité. d) Évolution de la taille du réseau des sociétés belges et de sa composante géante de 1993 à 2013.

Le niveau de détail et l'étendue des données permettent cependant des observations supplémentaires⁵ :

- comme le secteur immobilier (L) est aussi sensiblement plus dense sur l'axe vertical, on peut supposer que dans tous les secteurs, des sociétés détiennent des filiales gérant leur patrimoine immobilier propre (bureaux, entrepôts, etc) ;
- le secteur de la construction de bâtiments (F) manifeste une nette assortativité en investissant en lui-même, ainsi que dans celui des activités immobilières (L). Inversement, ce dernier investit aussi dans la construction ;
- le secteur du commerce automobile investit dans celui des activités de location ;
- une zone verticale légère mais sensible correspond à «M 71 Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques» dépendant de secteurs tels que l'industrie, la construction, les services financiers.

Comme nous disposons des données extraites des comptes annuels, nous pouvons compléter l'information brute fournie par le graphe au moyen d'une visualisation des transferts financiers réels d'un secteur à un autre⁶ ; la figure 4.7 représente ceux-ci sous forme de *heatmap* où les lignes représentent les secteurs d'où proviennent les transferts ; les colonnes, ceux où ils aboutissent et la couleur de l'intersection la valeur transférée. C'est

5. Nous avons implémenté une version interactive de cette matrice en javascript qui permettait de déceler des détails invisibles sur la version imprimée.

6. Il s'agit de la rubrique 694 (Rémunération du capital), partie de 694/6 (Bénéfice à distribuer) du Compte de résultats, multipliée par la part de la société détentrice.

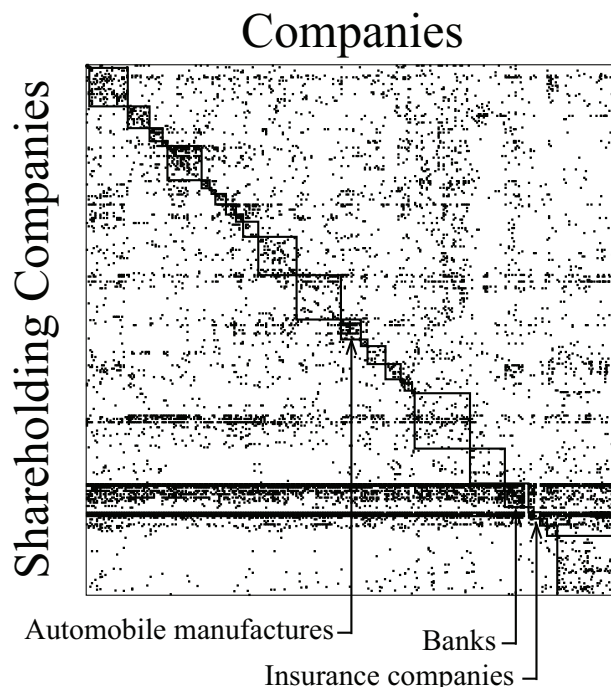


FIGURE 4.4 – Représentation matricielle du réseau de [Souma *et al.*, 2006b]

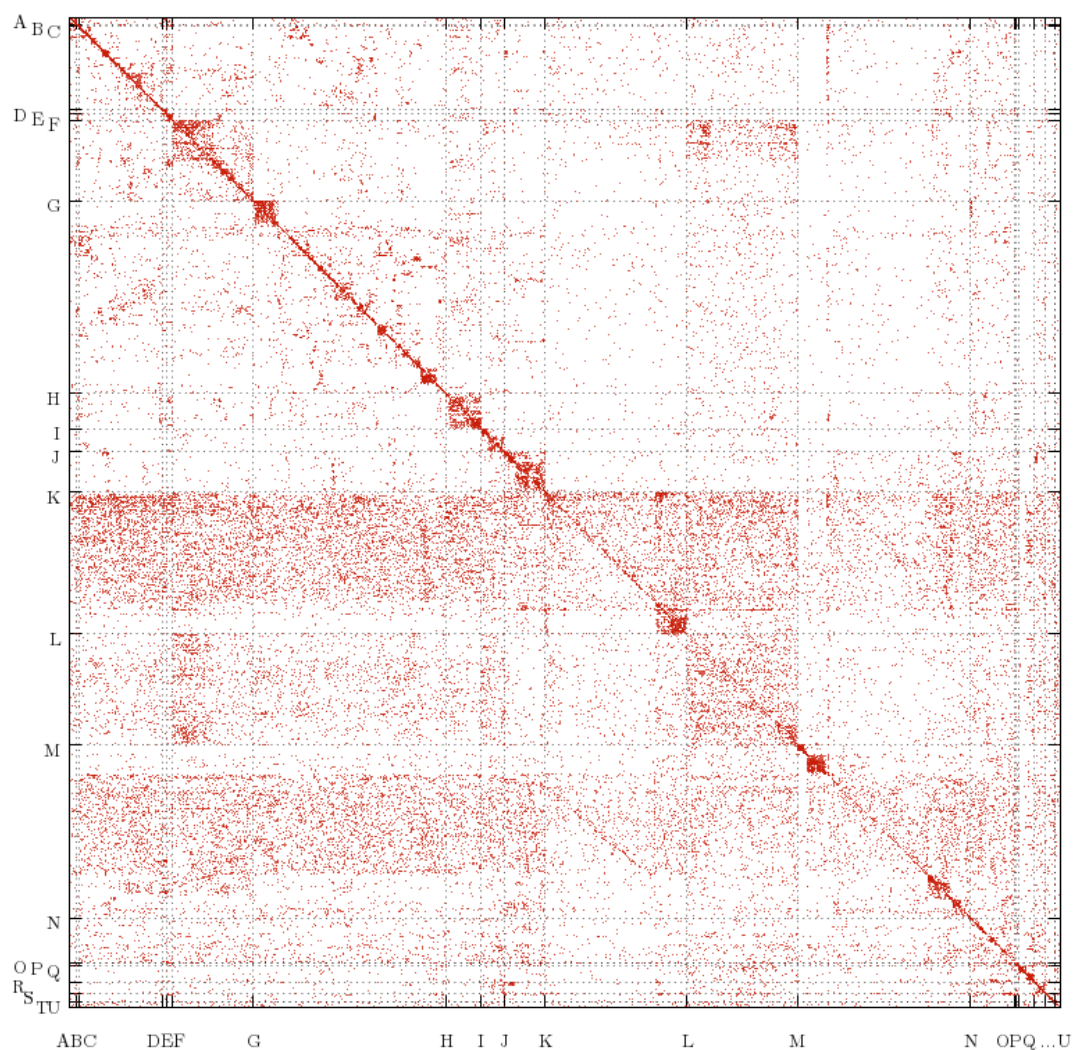


FIGURE 4.5 – Représentation matricielle du réseau des sociétés belges. Les sociétés mères (les sources des arcs) correspondent aux lignes de la matrice, les filiales directes (les cibles) aux colonnes. Les dénominations des secteurs NACE sont détaillées à la figure 4.6

évidemment le secteur financier et des assurances (K) qui en est le principal bénéficiaire et celui de l'industrie manufacturière (C) qui en est le principal fournisseur. Celui-ci est alimenté par lui-même, par le secteur financier ainsi que par les activités spécialisées (M). Ces dernières sont majoritairement alimentées par les 2 premières. Ces trois secteurs concentrent ainsi en valeur une large part des transferts alors que le secteur manufacturier est peu perceptible sur la figure précédente. Mais la perception visuelle d'un

A	Agriculture, sylviculture et pêche
B	Industries extractives
C	Industrie manufacturière
D	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E	Production et distribution d'eau ; assainissement, gestion des déchets et dépollution
F	Construction
G	Commerce de gros et de détail ; réparation de véhicules automobiles et de motocycles
H	Transports et entreposage
I	Hébergement et restauration
J	Information et communication
K	Activités financières et d'assurance
L	Activités immobilières
M	Activités spécialisées, scientifiques et techniques
N	Activités de services administratifs et de soutien
O	Administration publique et défense ; sécurité sociale obligatoire
P	Enseignement
Q	Santé humaine et action sociale
R	Arts, spectacles et activités récréatives
S	Autres activités de services
T	Activités des ménages en tant qu'employeurs ; activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre
U	Activités des organismes extra-territoriaux

FIGURE 4.6 – Sections de premier niveau des codes NACE

secteur sur celle-ci tient aussi au nombre de ses sociétés ; il ne faut donc pas s'étonner que le secteur industriel, plus concentré, y soit si peu visible.

4.3 Distribution des degrés

Le degré sortant d'un sommet (noté d^+) correspond au nombre de sociétés dans lesquelles un actionnaire détient des participations et constitue une mesure rudimentaire de la diversification de son portefeuille (*portfolio*). Le degré entrant (d^-) correspond au nombre de détenteurs de parts et donne une estimation de la dispersion de l'actionnariat [Vitali *et al.*, 2011].

[Souma *et al.*, 2005, Souma *et al.*, 2006a] interprètent le réseau des participations comme un graphe dirigé et en examinent la distribution des degrés sortants au cours d'un ou plusieurs exercices. Par contre, dans [Souma *et al.*, 2006b], ce réseau est considéré comme non-dirigé. Ils en tracent un graphique où on associe à chaque degré k la probabilité qu'un sommet présente un degré égal ou supérieur à celui-ci. On observe une probabilité égal à 1 que tout sommet soit du degré minimal (1 dans le cas non-dirigé puisqu'il n'y a pas de sommet isolé dans le graphe). La probabilité qu'un sommet pris au hasard soit de

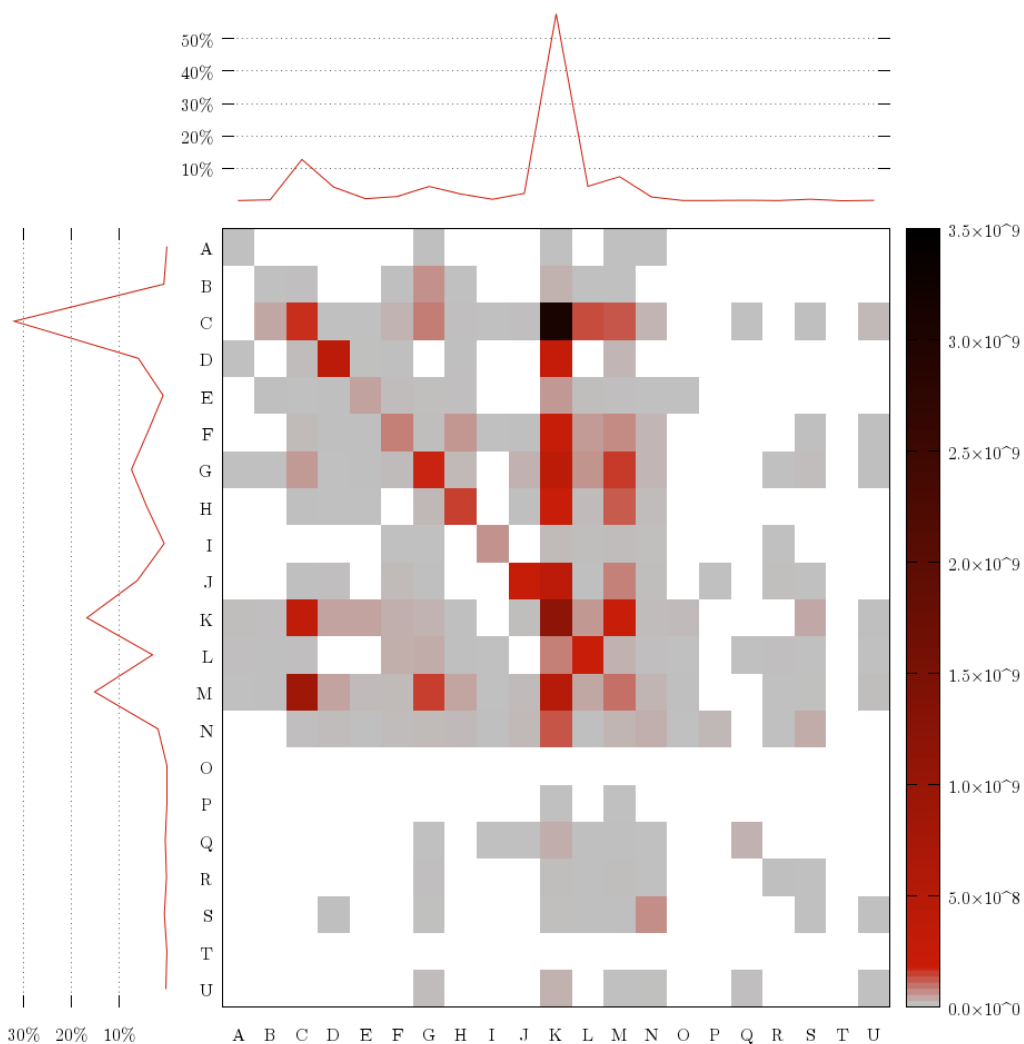


FIGURE 4.7 – Transferts financiers entre secteurs NACE. Au croisement de chaque ligne avec une colonne, la couleur indique la valeur absolue des dividendes (les bénéfices répartis distribués aux actionnaires) provenant des sociétés du secteur de la ligne vers celles du secteur de la colonne. À gauche et au sommet, un graphique montre la part relative des montants (respectivement) émis et reçus par chaque secteur par rapport à la globalité des montants transférés.

degré k ou plus diminue à mesure qu'on prend un degré plus élevé pour devenir égale à $1/N$ pour le degré le plus élevé lorsqu'il correspond bien à un unique sommet.

Sur un tel graphique, une pente plus raide signifie donc que les degrés élevés sont concen-

trés en un nombre plus réduit d'individus.

Dans tous les cas, ils observent une distribution sans échelle (*power law distribution*) caractérisée par un γ ⁷ qui définit la pente de la portion rectiligne de la courbe sur un graphique log-log. Cette courbe montre cependant une portion finale où la pente présente une inflexion (*“an exponential cutoff”*). Lorsque la courbe est représentée sur un graphique d'échelle semi-log, cette partie présente alors un tracé rectiligne. Cette partie finale est composée majoritairement d'institutions financières et, inversement, la partie sans échelle d'institutions non-financières.

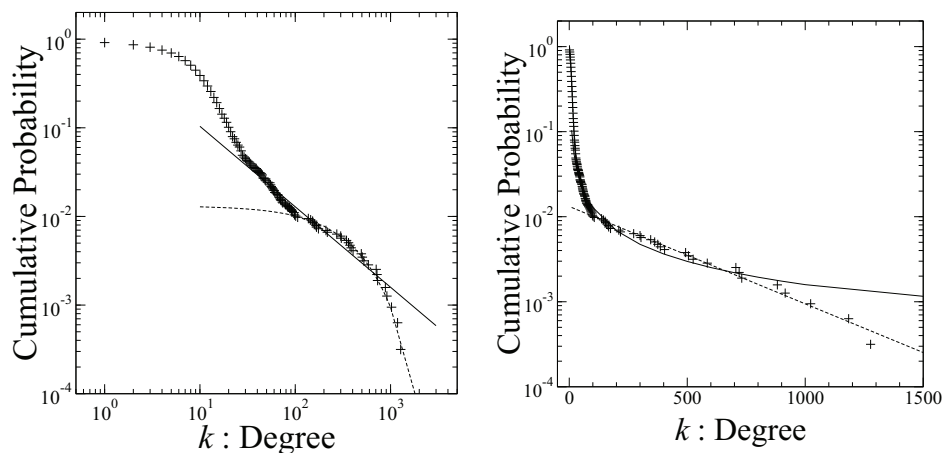


FIGURE 4.8 – Distribution des degrés dans le réseau observé par [Souma *et al.*, 2006b]

Nous l'avons pour notre part considéré comme un graphe dirigé pour pouvoir distinguer les distributions des degrés entrants, sortants et leur combinaison⁸. Pour les trois ensembles considérés, la pente de la courbe des degrés entrants est nettement plus forte et «tire» vers le bas celle des degrés combinés. Souma *et al.* n'avaient pas examiné les degrés entrants car ils avaient limité la collecte des actionnaires aux 20 plus importants ce qui biaisait leur représentativité. Dans notre cas, il ne s'agit pas d'un parti-pris méthodologique, mais un biais similaire provient de l'absence des société-mères étrangères ainsi que des personnes physiques ; les degrés entrants sont donc sous-estimés.

On obtient pour la pente du degré combiné des γ de 2,1, 1,9 et 1,3 alors que Souma obtenait 1,8. [Vitali *et al.*, 2011] obtenaient 1,15 (ceux-ci ne mentionnaient pas l'inflexion exponentielle). Les trois sont visibles sur la figure 4.10 ainsi que le résultat d'un graphe aléatoire de type Erdős–Rényi de même ordre et même taille que le réseau complet.

Conformément à la démarche de Souma, la figure 4.10e d'échelle semi-log montre la

7. Ce γ est en fait négatif mais les auteurs le mentionnent indifféremment avec ou sans le signe.

8. Plus précisément, nous avons considéré pour chaque sommet le cardinal de l'ensemble de ses prédécesseurs, de ses successeurs et de leur union (autrement dit leur nombre total de voisins distincts, indépendamment du sens des arcs). De cette façon, les voisins pouvant éventuellement être à la fois prédécesseurs et successeurs, le cardinal de leur union élimine le double comptage qu'on obtiendrait en additionnant les degrés entrants et sortants.

courbe de probabilité cumulée (relative aux degrés sortants uniquement) et la façon dont elle s'associe à une fonction de puissance en deçà d'un certain degré (environ 50) et à une partie exponentielle pour les degrés supérieurs, rendue ici par un tracé rectiligne. Néanmoins, on doit relever que cette dernière partie ne concerne qu'un nombre minime de sociétés ; les sommets de degré jusque 8 représentent déjà 99% de l'ensemble, ceux de degré supérieur à 50 ne sont plus que 19, soit moins de 1‰ aussi cette hypothèse doit-elle être considérée avec précaution. Il y a bien une tendance à compter une proportion plus élevée de sociétés du secteur financier au-dessus d'un degré sortant donné (figure 4.10f), mais on ne décèle pas de seuil manifeste au passage à la courbe exponentielle.

Vitali et al examinaient également la distribution de ce qu'ils dénomment la *force* de chaque société : $\sum_j W_{ij}$ c-à-d. la somme des degrés de participations, autrement dit les poids des arcs sortants, indépendamment de la valeur absolue des filiales concernées, et qu'ils présentent comme une mesure de la «force de connectivité» des sociétés. Le résultat sur le réseau belge donne un profil très comparable au leur mais avec un γ légèrement supérieur : 2 au lieu de 1,62. On distingue sur cette courbe (figure 4.9) un très long palier correspondant aux valeurs inférieures à 10%, puis les décrochements correspondant aux seuils de 50%, 100% (cette valeur étant affectée à 3913 sociétés) et 200%, enfin la partie sans échelle. Vu le nombre de sociétés présentant des participations de 0 ou négligeables, on a inclus également celles sans degré sortant, estimant qu'elles présentaient des forces comparables ; cela explique que la courbe démarre à 0,4, la probabilité de rencontrer une société de force supérieure à 0 étant déjà amputée de ce nombre considérable (50428).

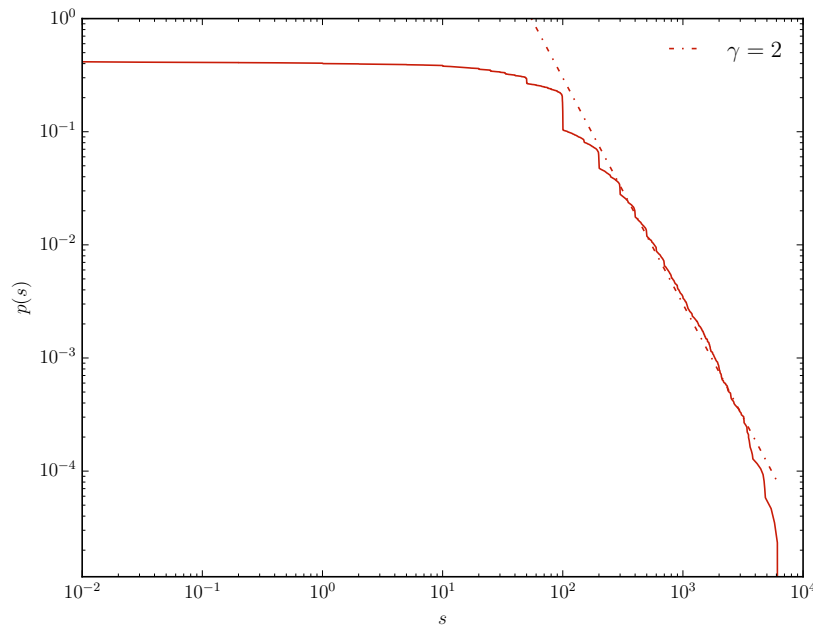
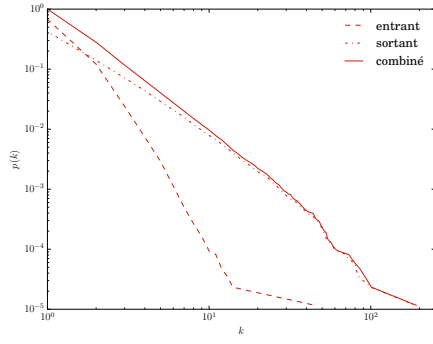
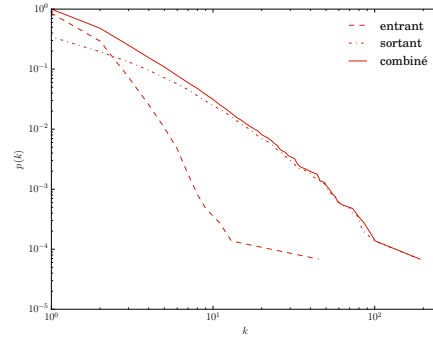


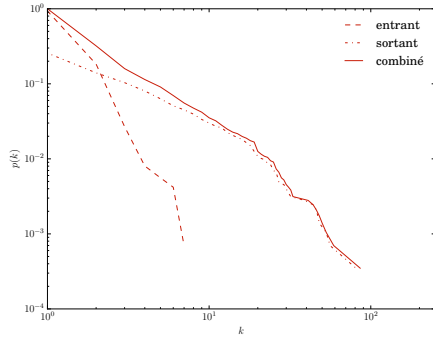
FIGURE 4.9 – Distribution de la *force* dans le réseau belge



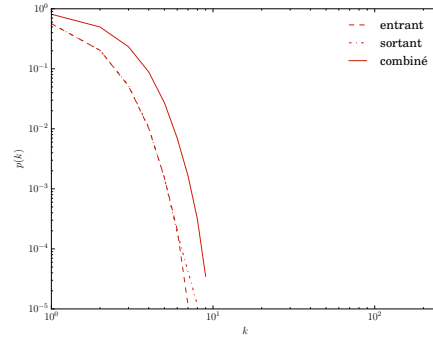
(a) Réseau entier



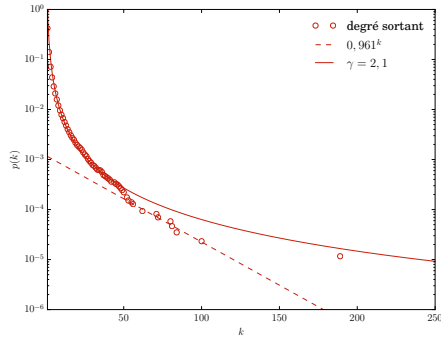
(b) Composante principale



(c) Sociétés cotées et connexes



(d) Graphe aléatoire



(e) Probabilité cumulée de degré sortant, (f) Ratio des sociétés financières parmi celles ayant $d^+ \geq k$

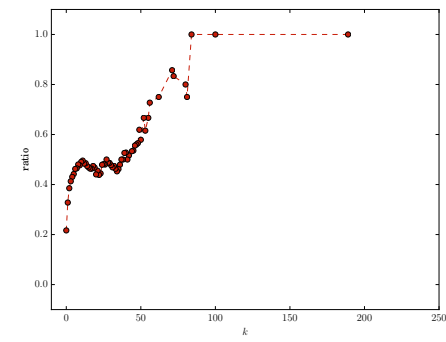


FIGURE 4.10 – a) à d) Distribution des degrés entrants (trait interrompu), sortants (trait mixte) et combinés (trait plein) en échelle log-log. L'axe horizontal représente les degrés k sur une échelle de 1 à 250, l'axe vertical la probabilité qu'un sommet soit de degré k ou plus. e) Inflexion exponentielle de la courbe au delà de $k = 50$ environ. f) La proportion de sociétés financières parmi celles de degré sortant supérieur ou égal à k augmente avec celui-ci, mais sans effet de seuil très marqué.

4.4 Assortativité

Un réseau est qualifié d'*assortatif* lorsque les sommets tendent à choisir des voisins présentant des caractéristiques semblables.

Cela apparaît de façon évidente dans des réseaux sociaux où l'on observe par exemple des regroupements par tranche d'âge (dans ce contexte, on parle aussi d'*homophilie*). L'assortativité peut aussi être établie de façon topologique sur base du degré des sommets ; on l'observe dans les réseaux sociaux lorsque «les individus grégaires sont amis avec d'autres individus grégaires et les ermites avec d'autres ermites» [Newman, 2010]. Le concept a été précisé par [Newman, 2003a] qui a établi pour l'assortativité par degré une mesure qui utilise les coefficients de Corrélation de Pearson entre les degrés des sommets concernés, et qui varie de 1 pour un réseau assortatif à -1 pour un réseau disassortatif.

Pour visualiser cette caractéristique, Souma et al se sont inspirés de [Pastor-Satorras *et al.*, 2001] où la structure d'Internet était examinée au moyen d'une mesure de connectivité moyenne :

$$\langle k_{nn} \rangle = \sum_{k'} k' p_c(k'|k)$$

$p_c(k'|k)$ est la probabilité qu'un lien existe d'un sommet de degré k vers un autre de degré k' . $\langle k_{nn} \rangle$ est en fait la connectivité moyenne des voisins d'un sommet. On associe sur un graphique de type *log-log* le degré des sommets avec ce $\langle k_{nn} \rangle$. Souma et al. obtiennent un γ de 0,8 (voir figure 4.11a) et observent que «*les sommets de degré élevé présentent un $\langle k_{nn} \rangle$ peu élevé, ce qui signifie que les hubs – les sommets de degré élevé – ne sont pas directement liés les uns aux autres*», une caractéristique des réseaux disassortatifs tels que le web ou les réseaux biologiques.

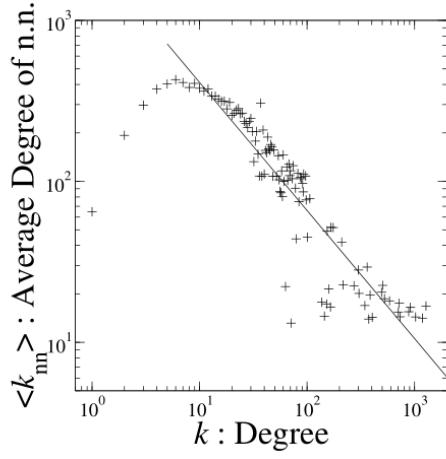
Pour les données du réseau belge, il y a une progression de la disassortativité en passant du réseau entier au sous-ensemble des sociétés cotées (figures 4.11b à 4.11d). L'assortativité plus importante du réseau entier s'explique aisément par le nombre important de composantes comprenant une société mère et une unique filiale ou à tout le moins une structure très réduite. La disassortativité du dernier s'explique par le fait que les sociétés cotées sont à la fois susceptibles d'être détenues par de nombreux actionnaires et d'être des groupes importants disposant de nombreuses filiales.

4.5 La structure en nœud papillon

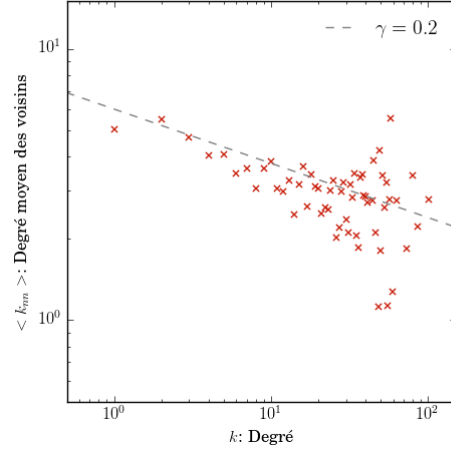
La structure en nœud papillon du type de celle qui caractérise internet est au centre de l'interprétation de Vitali et al. Une définition plus formelle que celle qui avait cours depuis son introduction par [Broder *et al.*, 2000] a été fournie par [Yang *et al.*, 2011],

ainsi qu'un processus algorithmique de décomposition permettant d'en extraire les sous-parties (voir annexe 7.3). Nous l'avons appliqué aux plus importantes des composantes fortement connexes du graphe complet mais la plus grande de celles-ci ne présente que 9 sommets en son centre. Si l'on considère pour chacune de ces structures le nombre de ses sommets (colonne « $|V - OTHERS|$ » du tableau), seule l'une d'elle en compte assez pour représenter plus de 1% du total du réseau, et encore n'atteint-elle ce niveau que grâce au nombre important de sommets appartenant aux *in-tendrils* (les sommets joignables depuis la partie *IN* sans passer par *SCC* ni permettant de joindre *OUT*). Or, l'intérêt de cette structure dans le sens que lui donnaient Vitali et al. tenait essentiellement à l'observation du flux qui en traversait le noyau et seul le regroupement «*IN + SCC + OUT*» importe alors ; les parties secondaires (*IT*, *OUTT*, *TUBES*) n'y jouaient aucun rôle. Si l'on s'en tient à ce groupement «*IN + SCC + OUT*», les plus importantes avoisinent à peine la centaine de sommets.

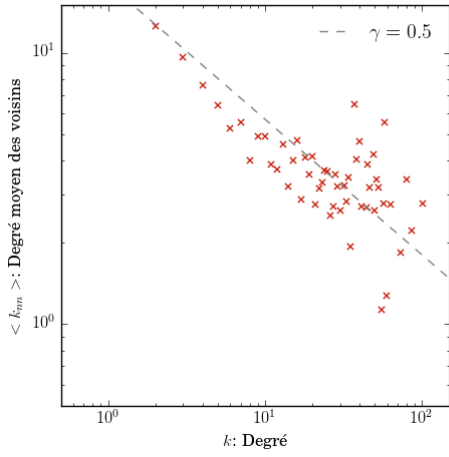
Au vu des résultats, il apparaît que ce type de structure permet, pour le réseau belge, de capturer le voisinage particulier de petits cœurs connexes (qui montrent la même asymétrie vers la partie sortante que la structure globale analysée par Vitali et al) mais ne reflète pas une organisation affectant l'ensemble du réseau. La législation nationale qui interdit les participations croisées supérieures à 10% joue sans doute un rôle à ce sujet. La question de savoir si l'ajout des actionnaires étrangers (directs et indirects, jusqu'à un niveau à déterminer) des sociétés belges mettrait en évidence un noyau plus important reste ouverte.



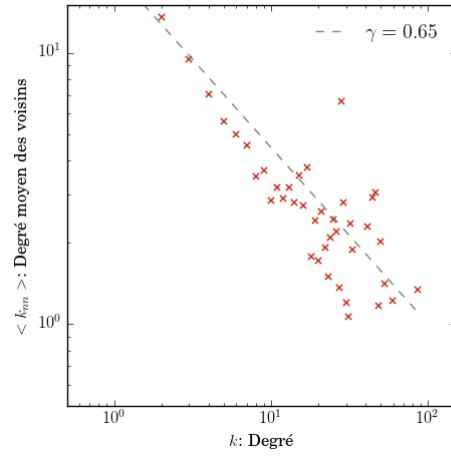
(a) [Souma *et al.*, 2006b] $\gamma = 0,8$



(b) Réseau entier



(c) Composante géante



(d) Société cotées et associées

FIGURE 4.11 – Corrélation de degré (assortativité).

$ SCC $	$ IN $	$ OUT $	$\frac{ IN+SCC+OUT }{ V }$	$ IT $	$ OUTT $	$ TUBES $	$ V - OTHERS $	$\frac{ V-OTHERS }{ V }$
9	0	20	0.0003	3	0	0	32	0.0004
8	0	25	0.0004	0	0	0	33	0.0004
7	1	6	0.0002	0	0	0	14	0.0002
7	4	44	0.0006	13	4	0	72	0.0008
6	5	83	0.0011	96	15	0	205	0.0024
6	3	84	0.0011	72	5	0	170	0.0020
6	1	14	0.0002	0	0	0	21	0.0002
5	12	3	0.0002	0	13	0	33	0.0004
5	0	18	0.0003	14	0	0	37	0.0004
5	7	43	0.0006	8	221	0	284	0.0033
5	0	23	0.0003	3	0	0	31	0.0004
5	4	57	0.0008	14	678	8	766	0.0089
5	7	2	0.0002	0	1	0	15	0.0002
5	0	2	0.0001	1	0	0	8	0.0001
5	0	2	0.0001	0	0	0	7	0.0001
4	4	11	0.0002	0	9	0	28	0.0003
4	23	77	0.0012	56	1323	0	1483	0.0173
4	0	10	0.0002	2	0	0	16	0.0002
4	0	23	0.0003	1	0	0	28	0.0003
4	0	0	0.0000	0	0	0	4	0.0000

FIGURE 4.12 – Les 20 premières structures en nœud papillon présentes dans le réseau. V : l'ensemble des sommets du réseau. SCC : le cœur fortement connexe. IN et OUT : les parties d'où l'on peut atteindre SCC et joignables depuis celui-ci, respectivement. $IT, OUTT$: les rameaux entrants et sortants. $TUBES$: les branches joignant IN à OUT . $OTHERS$: la partie du réseau qui n'est atteignable par aucun sommet de la structure.

5 Recherche de sous-ensembles significatifs

Regrouper les éléments qui se ressemblent au sein d'un ensemble plus vaste pour le rendre intelligible est une démarche spontanée et répandue. Dans une perspective de *machine learning*, le terme d'usage pour la désigner est celui de *clustering*, les regroupements qualifiés de *clusters* sont effectués en recherchant des régularités (*hidden patterns*) parmi les caractéristiques des objets comparés, lesquelles sont généralement représentées sous la forme de vecteurs de mesures. Le clustering relève du domaine de l'*apprentissage non-supervisé* du fait qu'on ne dispose pas d'une connaissance préalable des regroupements recherchés ou d'éléments individuels qui les caractérisent. Il s'agit typiquement d'une démarche *exploratoire* «destinée à voir si des regroupements naturels existent dans les données. Si de tels regroupements émergent, ils peuvent être nommés et leurs propriétés résumées» [Webb, 2003].

Son application au domaine des réseaux n'en est qu'un cas particulier mais elle s'avère particulièrement fertile¹. Il règne cependant une certaine confusion sur le plan terminologique²; comme le relève [Fortunato, 2010], “*The 1st problem in graph clustering is to look for a quantitative definition of community. No definition is universally accepted. As a matter of fact, the definition often depends on the specific system at hand and/or application one has in mind.*”

Nous entamerons donc ce chapitre par une section destinée à éclaircir la terminologie, nous poursuivrons par une autre plus étendue expliquant les concepts-clés de rôles et communautés. Pour bien les illustrer, nous effectuerons un détour par un type de réseau particulier, celui des réseaux trophiques ou *food-webs* et montrerons comment l'analogie entre ceux-ci et les réseaux de participations nous ont amené à la lecture de l'article de Cooper et Barahona explicité dans la section suivante. Finalement, nous exposerons comment nous avons adapté leur algorithme pour dégager des regroupements du réseau des participations.

5.1 Terminologie

La finalité du *clustering* est de rassembler des individus en sous-ensembles de façon à ce que les individus présents dans un sous-ensemble soient relativement similaires

1. Pour une compilation des concepts et algorithmes relatifs à ce sujet, voir [Gaetler, 2005, Schaeffer, 2007, Porter *et al.*, 2009, Aggarwal et Wang, 2010].

2. C'est pourquoi l'intitulé de ce chapitre s'en tient prudemment au terme neutre de «sous-ensemble».

entre eux et présentent moins de similarité avec ceux figurant dans les autres sous-ensembles (voir par exemple [Webb, 2003]). Cette similarité est associée à la notion de distance : un cluster est «un groupe de points de données dont les distances inter-point sont petites comparativement aux distances relatives à celles aux points situés à l’extérieur du cluster» [Bishop, 2006]. Cette distance est généralement la distance euclidienne entre les vecteurs de mesures des individus comparés.

Dans le cas des réseaux, une définition est celle qui a été fréquemment utilisée par [Girvan et Newman, 2002] qui observent que les réseaux présentent une *structure en communautés* c.-à-d. une division en groupes au sein desquels les connexions sont denses, tandis qu’elles sont plus rares entre ces groupes. Ils reprenaient cette définition du domaine des sciences sociales ([Wasserman et Faust, 1994, Scott, 2000]) où il apparaît de façon intuitive que les populations humaines se subdivisent en groupes où les relations sont plus denses. Notons que Girvan et Newman utilisent le terme de communauté plutôt que celui de *cluster* pour éviter toute confusion avec la propriété de *clustering* relative à la probabilité que deux sommets adjacents à un autre soient eux-mêmes adjacents entre eux³.

Cette définition est cohérente avec celle plus générale qui précède si l’on considère qu’une distance entre les sommets peut être évaluée sur base du nombre de voisins partagés ou, autrement dit, sur le recouvrement (*overlapping*) de leurs voisinages.

En dehors de l’utilisation du terme *community* par les physiciens plutôt que celui de *cluster*, ou de *cohesive subgroup* par les praticiens de l’analyse de réseaux sociaux⁴, une certaine confusion tient à la distinction entre *graph partitioning* et *graph clustering*. D’après [Browet, 2014], le premier terme consiste à dégager des regroupements en retirant progressivement des liaisons en respectant des contraintes telles que la minimisation du nombre de retraits. Le second vise plus particulièrement le respect d’une forme d’homogénéité au sein des regroupements. Ce dernier recouvre donc la recherche de communautés, mais pas exclusivement. Le même auteur donne en exemple le réseau physique d’internet dont l’architecture en couches comprend des routeurs de différents niveaux. Au sein d’un niveau, les routeurs ne sont pas connectés entre eux ; ils n’en déterminent pas moins la structure du réseau bien que celui-ci ne présente pas de communautés denses. Le processus de clustering consiste en ce cas en une identification de rôles, qui sera développée dans la section suivante.

D’après [Parthasarathy et Faisal, 2013], la distinction entre partitionnement et clustering tient à ce que

- le partitionnement couvre la totalité des éléments et que les éléments ne peuvent appartenir qu’à un seul sous-ensemble ;
- le clustering permet éventuellement qu’un élément appartienne à plusieurs sous-ensembles. On considère parfois qu’il est exclusif par convention et on précise alors lorsqu’il ne l’est pas (*overlapping clustering*). On peut également effectuer

3. Propriété utilisée dans la définition du phénomène de «petit monde». Voir l’annexe 7.2 à ce sujet.

4. En français, [Beauguitte, 2011] a abordé ces précisions terminologiques.

une distinction entre un clustering *total* ou *partiel* selon qu'il partage la totalité des éléments ou que certains restent isolés, ainsi qu'entre un clustering équilibré ou non (*balanced*, *skewed*) si les sous-ensembles sont de tailles approximativement égales ou pas.

Il distingue la recherche de communautés par le fait que l'intérêt est porté plus particulièrement sur les sous-ensembles denses et que les autres sommets peuvent être négligés, ce qui procure un gain important en termes de calcul.

Avant de terminer cette section, il nous faut revenir sur un aspect qui distingue la plupart des techniques de clustering au sens large de celles portant sur les réseaux :

- Comme on l'a vu, les premières sont fondées sur des distances établies sur base de propriétés *intrinsèques* aux individus, indépendamment des relations entre eux-ci. Dans le cas qui nous concerne, nous pourrions ainsi utiliser des postes du bilan comptable et regrouper des entreprises présentant des profils comparables selon plusieurs dimensions telles que l'endettement, les immobilisations, divers ratios financiers, etc.
- À l'opposé, le clustering de réseaux est fondé sur des caractéristiques *topologiques* des éléments, c.-à-d. leurs positions relatives les uns par rapport aux autres, indépendamment de leurs propriétés intrinsèques. Dans ce dernier cas, la notion de distance peut être absente des algorithmes ; ceux-ci se répartissent le plus souvent en processus divisifs, agglomératifs ou fondés sur l'optimisation d'un critère.

On peut cependant envisager une voie hybride consistant à établir pour les sommets d'un réseau des vecteurs de valeurs fondées sur des données topologiques (telles que l'identité des sommets situés en deçà d'une distance) et évaluer en fonction de ceux-ci les distances ou la similarité entre les sommets pour effectuer le clustering ; tel sera le cas de la procédure de Cooper et Barahona exposée à la section 5.4.

5.2 Rôles *vs* communautés

Un partitionnement en zones denses n'est pas l'unique manière de dégager la structure d'un réseau. Pour s'en convaincre, on peut observer la représentation imaginaire d'un petit réseau de participations dans la figure 5.1.

On y a reproduit une structure étagée en pyramide s'élevant sur trois niveaux. Dans la vue «par-dessus», on distingue deux composantes correspondant aux filiales directes et indirectes des deux «têtes de groupes» *A* et *B*. C'est le type de structure qui peut être mise en évidence par des algorithmes de recherche de communautés, bien qu'une difficulté puisse provenir du fait que les ensembles des filiales de *A* et *B* se recouvrent et que la plupart de ces algorithmes effectuent des partitionnements exclusifs.

Un tel partitionnement rassemble des sociétés qui jouent des rôles économiques différents. Les têtes de groupes exercent le contrôle mais n'exercent pas nécessairement une activité

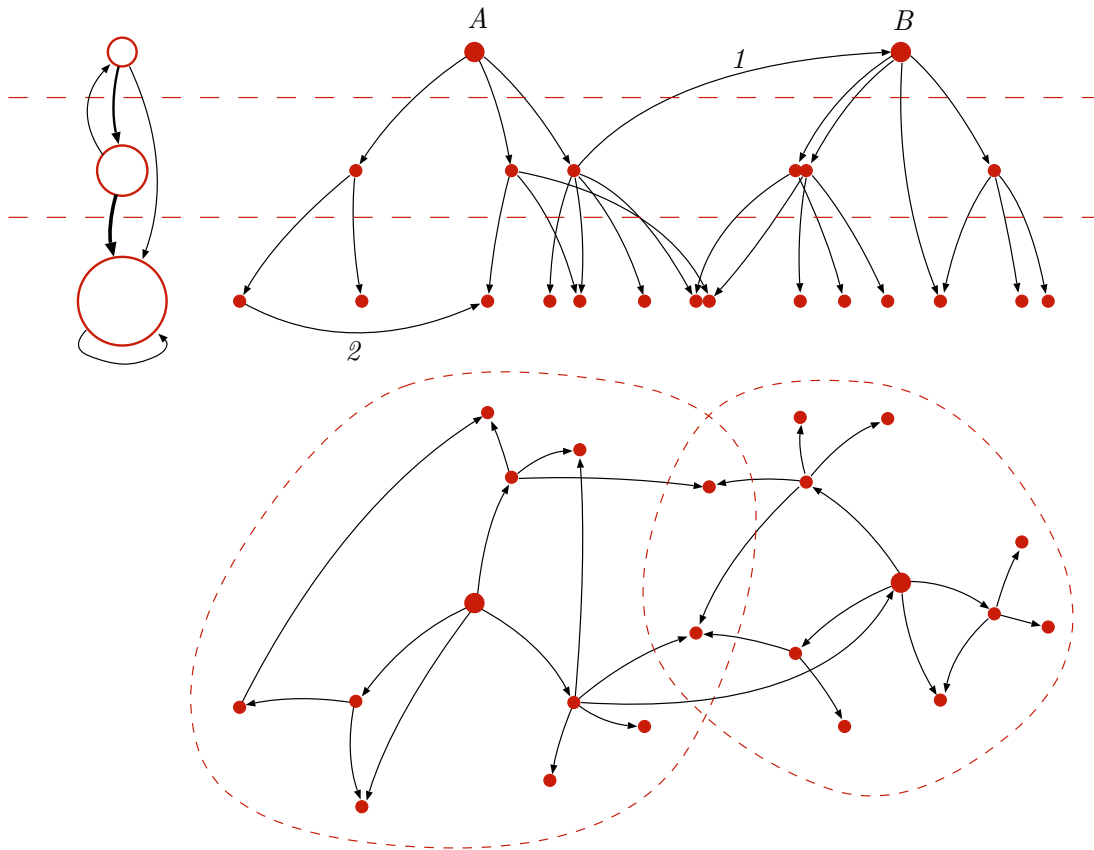


FIGURE 5.1 – Représentation des rôles et communautés dans un réseau imaginaire de participations. Une vue «latérale» (en haut à droite) y souligne les similarités de rôles tandis qu’une vue «de dessus» en montre les communautés. En haut à gauche, un «graphe image» en donne une vue condensée.

de production ; les filiales en fin de chemin sont vraisemblablement dans la situation inverse.

Une autre manière de structurer un tel réseau est de rassembler les sociétés qui jouent le même *rôle*, tel qu’on peut l’observer dans la vue «de côté» en fonction du «niveau» dans la pyramide. Les sommets *A* et *B* seront alors regroupés, bien qu’il n’existe pas de chemin de l’un vers l’autre (des sommets appartenant à des parties totalement déconnectées d’un réseau peuvent d’ailleurs partager le même rôle).

L’assignation de chaque sommet à un rôle n’est cependant pas aussi triviale qu’elle est présentée ici. La présence des arcs 1 et 2 qui relient des sommets de même niveau ou remontent vers un niveau supérieur vient en perturber l’interprétation ; le sommet *B* peut-il par exemple être considéré comme un actionnaire ultime alors qu’il comporte un arc entrant ?

Cette notion de rôle a été particulièrement abordée dans le domaine de l'analyse des réseaux sociaux.

«In an intuitive way, we would say that two actors have the same “position” or “role” to the extent that their pattern of relationships with other actors is the same⁵» [Hanneman et Riddle, 2005b].

Elle est étroitement liée à celle d'*équivalence* : les acteurs qui sont connectés de la même manière au reste du réseau sont qualifiés d'équivalents et occupent la même position. L'objectif est de répartir les acteurs dans des classes mutuellement exclusives d'acteurs présentant un *pattern* similaire de relations. Une première manière d'envisager l'équivalence est celle où elle est qualifiée de *structurelle* ; elle a été définie par [Lorrain et White, 1971] : «*a est structurellement équivalent à b si a est en relation avec tout objet x exactement comme c'est le cas pour b. Du point de vue de la logique de la structure, a et b sont alors absolument équivalents, ils sont substituables*». [Wasserman et Faust, 1994] ont reformulé cette définition en précisant que «*deux acteurs sont structurellement équivalents s'ils ont des liaisons identiques depuis et vers tous les autres acteurs dans le réseau*». On peut en souligner certaines conséquences :

- Comme des sommets structurellement équivalents sont connectés exactement aux mêmes voisins, ils ont les mêmes propriétés telles que la centralité, l'excentricité, le degré, le prestige, etc. L'inverse n'est pas vrai : les sommets qui partagent ces attributs ne sont pas nécessairement équivalents.
- Comme on ne prend en considération que les sommets adjacents, l'équivalence entre des sommets n'est pas affectée si des changements apportés au réseau le sont à une distance supérieure à 1.
- Des sommets structurellement équivalents appartiennent toujours à la même composante (connexe) d'un graphe.

La propriété d'équivalence structurelle peut être observée dans la matrice d'adjacence du réseau : les sommets équivalents doivent présenter dans celle-ci les mêmes éléments à la fois dans leurs lignes et leurs colonnes. Cette propriété est cependant extrêmement stricte et ne permet en pratique que de dégager des classes de très petite taille. Cependant, des sommets peuvent partager *en partie* un certain nombre de voisins. On peut donc envisager l'équivalence comme une mesure continue en fonction du nombre de voisins partagés. Cette façon de *relâcher* l'exigence de la définition fait appel à une matrice de similarité ou de distance pour toutes les paires d'acteurs en présence. Il existe de nombreuses façons d'établir cette similarité, telles que le nombre de concordances dans les matrices binaires (distances de Jaccard ou de Hamming), les distances euclidiennes ou *cosine*, le coefficient de Pearson...

5. La distinction précise entre ces termes dans le domaine de la *SNA* est la suivante : la *position* est la localisation d'un acteur ou d'une classe d'acteurs dans un système social de relations, une *attente* correspond à un standard lié à l'occupant d'une position (comportement, normes, droits...) et un *rôle* est un ensemble d'attentes associées à une position [Alhajj et Rokne, 2014].

Une seconde manière d’envisager l’équivalence est celle où elle est qualifiée de *régulière* : deux acteurs sont régulièrement équivalents s’ils présentent le même *pattern* de liaisons avec des acteurs qui sont eux-mêmes régulièrement équivalents [Borgatti et Everett, 1992]. Cette définition peut paraître déroutante, mais elle s’applique pourtant à des situations qui nous sont familières. [Hanneman et Riddle, 2005b] prennent l’exemple de mères qui sont rattachées à leurs maris et à un certain nombre d’enfants, bien que ces maris et enfants soient évidemment distincts pour chacune. Pour cette raison, elles ne peuvent être structurellement équivalentes (interchangées), mais elles occupent pourtant des positions comparables. [Johnson *et al.*, 2001] ont illustré cette méthode en recourant à une coloration des sommets : trouver des classes d’équivalence revient à attribuer aux sommets des couleurs de telle sorte que les sommets d’une couleur donnée soient tous reliés de la même façon aux sommets d’autres couleurs (voir figure 5.2).

La recherche de classes d’équivalence régulière ne connaît pas de solution unique. Les solutions triviales consistant à englober tous les sommets dans une seule classe ou au contraire à les isoler dans des classes dont ils sont les seuls représentants en sont deux exemples extrêmes. D’autre part le problème paraît circulaire : attribuer un sommet à une classe nécessite de connaître celles de ses voisins, lesquelles ne peuvent être connues que par celles de leurs propres voisins. Il existe pourtant une solution itérative décrite par [Borgatti et Everett, 1993]. L’algorithme consiste à attribuer un score de similarité initialement égal à 1 à tous les sommets 2 à 2. Puis, à examiner les voisinages de chaque paire de sommets et réévaluer ce score pour chacun d’eux en fonction de ceux de leurs voisins⁶. Lors des itérations suivantes, le calcul est répété mais en tenant compte des scores obtenus auparavant. En pratique, le processus réduit de façon différenciée la similarité en fonction de la distance qui sépare chaque sommet des *sources* et des *puits* du réseau⁷.

La relation d’équivalence régulière permet de proposer un «graphe image» du réseau d’origine. En agrégeant les sommets qui appartiennent à une même classe, on obtient une forme condensée du réseau qui met en évidence les relations les plus fréquentes du réseau d’origine et en facilitent la compréhension. Une autre façon d’obtenir ce résultat est de pratiquer le *blockmodeling*. Ce procédé, notablement exploré par [Doreian *et al.*, 2005], consiste à effectuer des permutations sur les rangées et les colonnes d’une matrice d’adjacence binaire de manière à rassembler des régions homogènes de 1 ou de 0. Les blocs non homogènes peuvent être distingués selon une typologie assez complexe d’après la manière dont les valeurs s’y organisent. Nous le citons parce qu’il existe une relation entre ce procédé et les notions d’équivalence régulière et structurelle : les blocs pour une équivalence régulière sont soit nuls (composés de 0), soit comprennent au moins une valeur 1 pour chacune des rangées ou des colonnes, les blocs pour une équivalence

6. Sans entrer dans les détails, on compte pour chaque sommet le nombre de relations sortantes, de relations entrantes et de relations «réciproques» et on attribue des points en fonction du nombre de relations de chaque type qu’ils partagent.

7. Dans un graphe dirigé, les sources sont les sommets sans arcs entrants, les puits ceux sans arcs sortants.

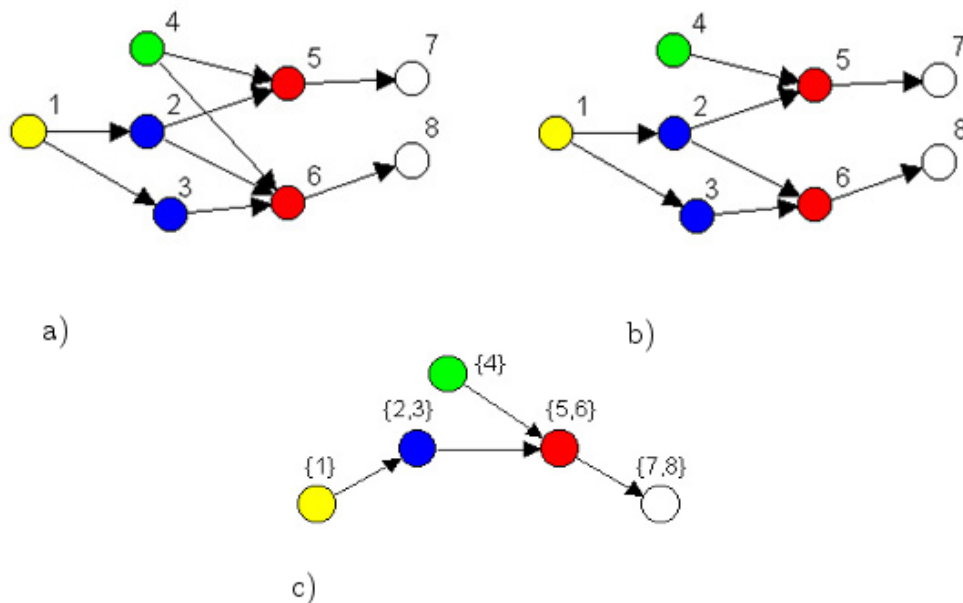


FIGURE 5.2 – L'équivalence régulière conçue comme une coloration des sommets, extrait de [Johnson *et al.*, 2001]. a) Coloration régulière. b) Coloration non régulière (les sommets rouges diffèrent par la liaison avec un sommet vert). c) «Graphe image» du réseau a).

structurelle sont nuls ou complets.

Cette section peut être clôturée par l'observation que la notion de rôle, malgré l'intérêt qu'elle présente pour permettre la compréhension de réseaux complexes, paraît nettement moins suivie dans la littérature récente que ne le sont les travaux sur la recherche de communautés. Il ne semble pas que des librairies largement répandues permettent de la mettre en pratique; l'algorithme REGE n'a été implémenté à notre connaissance que dans des logiciels qui ne permettent pas le traitement de graphes très importants, et malheureusement à code source non ouvert. Ce dernier point n'est sans doute pas étranger à sa désaffection.

5.3 Analogie avec les réseaux trophiques

La biologie a également recours aux graphes pour modéliser et simuler les systèmes écologiques. Dans un écosystème, les espèces interagissent de différentes façons constituant autant de réseaux entremêlés; on distingue ainsi des relations de compétition, de parasitisme, de mutualité ou de type proie-prédateur. Cette dernière détermine ce qu'on nomme un réseau trophique (*food web*) où des liens sont établis de i vers j si j consomme

i. La direction des arcs représente donc un flux de *ressources* de la proie vers le prédateur.

[Cohen *et al.*, 1990] définissent une espèce trophique (*trophic species*) comme une collection d'organismes partageant les mêmes régimes alimentaires et prédateurs. Une telle espèce peut occasionnellement se confondre avec une espèce biologique au sens usuel du terme, mais le plus souvent, elle en comporte plusieurs, ainsi que des représentants d'espèces à des stades particuliers de développement (l'asticot et la mouche, le têtard et la grenouille...). C'est pourquoi le terme de *trophic guild* est parfois préféré.

Cette définition est en relation directe avec les concept d'équivalence :

“An additional point of concern when using the concept of trophic guild is that organisms may play similar roles in the food web to the extent that they have similar links to analogous (but not identical!) species. In former concepts, two species became members of the same trophic guild if and only if they had trophic links to (almost) the same set of other trophic groups. Stimulated by work in social sciences, this limitation was overcome by distinguishing between structural equivalence (same links) and regular equivalence (analogous but not identical links).” [Gaedke, 2008]

Les espèces présentes dans le réseau peuvent se trouver à son sommet (aucune autre ne les consomme), à sa base (elles n'en consomment aucune autre) ou à un niveau intermédiaire (elles sont à la fois consommées et consommatrices). Une caractérisation sommaire d'un réseau trophique peut être établie selon le nombre d'espèces de chaque catégorie. La distinction entre ces espèces permet d'en déduire une autre entre les liens en fonction des sommets réunis : basal-intermédiaire, basal-sommital, intermédiaire-intermédiaire, intermédiaire-sommital.

Une chaîne est un chemin menant d'une espèce basale à une espèce du sommet, et sa longueur – comme dans la terminologie usuelle des graphes – est le nombre de liens qui la constituent. Les organismes dont la nourriture est obtenue depuis le niveau de base (celui des *producteurs* ou organismes autotrophes, c.-à-d. capables de consommer de la matière non organique) par un même nombre d'étapes appartiennent au même *niveau trophique*. Au-dessus des producteurs, on trouve celui des *premiers consommateurs* ou herbivores, puis les *seconds consommateurs* ou carnivores₁, etc.

Cohen a dégagé 5 lois ou «généralisations empiriques» des réseaux trophiques :

1. les cycles sont rares ;
2. les chaînes sont courtes. Sur les 113 réseaux qu'il a collationnés, la médiane des longueurs maximales est 4 et la longueur la plus grande – 10 – n'apparaît qu'une seule fois ;
3. la proportion des espèces trophiques basales, sommitales et intermédiaires est un *invariant d'échelle*. Elle est à peu près constante et indépendante du nombre d'espèces ;
4. la proportion entre les différentes catégories de liens est également un invariant d'échelle ;

5. la proportion entre le nombre de liens et d'espèces est également un invariant d'échelle.

Des rapprochements ont été effectués entre la position d'une espèce dans un tel réseau et celle d'un acteur dans un réseau social : [Johnson *et al.*, 2001] ont ainsi souligné que le concept d'*espèce trophique*, qui rassemble des espèces dotées de proies et de prédateurs similaires, est identique à la notion d'équivalence régulière abondamment développée en *SNA* et affirment que

“the striking convergence to date of the fields of ecology and sociology via independent invention of network concepts suggests that there may be considerable value in cross-fertilization of the two fields.”

De la même manière, l'analogie des réseaux trophiques avec ceux de participations vient assez naturellement. Les deux sont des graphes dirigés et présentent une structure hiérarchique ; celle-ci n'est pas immédiatement perceptible mais on perçoit bien que certains individus occupent une position supérieure aux autres. Dans les deux cas, des ressources sont puisées dans le « milieu » extérieur et transmises depuis les niveaux inférieurs vers les supérieurs⁸. Le parcours de l'énergie dans la chaîne alimentaire rappelle celui du cash-flow ; une partie en est affectée à l'activité physique (*vs* les dépenses de fonctionnement), une autre au développement des organismes (*vs* à l'investissement), une partie sera consommée par le niveau supérieur (*vs* distribuée). Dans un réseau trophique, il existe un déclin dans la quantité d'énergie passant d'un niveau à l'autre et au final, elle est dissipée. On a estimé que pour cette raison, le nombre de niveaux excède rarement cinq. Dans les réseaux de participations, la multiplication des niveaux se traduit de la même façon par une perte d'efficacité dans la transmission du cash-flow.

D'une certaine manière, les prédateurs exercent également un contrôle sur leurs proies. [Granovetter, 2005], en décrivant les *business groups* au sein de leur environnement, cite [Hannan et Freeman, 1993] en observant que « *Les organisations sont des compétiteurs pour des ressources dans des niches qui ne peuvent supporter une occupation illimitée* ».

Mais il faut se garder des rapprochements abusifs. En premier lieu, les nœuds des réseaux trophiques ne sont pas des individus mais des populations d'espèces, alors que chaque sommet du réseau des participations est un individu particulier. En ce sens, un réseau trophique est un modèle constituant déjà une forme de contraction de la réalité qui est absente de celui des participations. D'autre part, le contrôle exercé par les prédateurs sur les proies est d'une autre nature et doit se comprendre en termes de population ; ce contrôle est par ailleurs mutuel puisqu'un équilibre est nécessaire entre leurs populations respectives.

8. Les arcs des réseaux trophiques peuvent être de type binaire (présence ou absence de flux) ou pondérés par une estimation du transfert de carbone ou d'azote mais les analyses mettant en œuvre ces derniers sont rares. “Flow-based methods are also time consuming, expensive, and methodologically involved. Thus, highly and evenly resolved flow-based or isotope based food webs are still exceedingly rare due to the massive effort involved in assembling these types of quantitative food webs. [...] Still, the degree to which binary links reflect actual energy flow patterns in food webs is poorly understood” [Carscadden *et al.*, 2012].



FIGURE 5.3 – Dans l'inconscient collectif, les rapport économiques sont souvent assimilés à d'impitoyables rapports de prédation. «Les gros poissons mangent les petits», gravure de Pieter van der Heyden, d'après Pieter Brueghel l'ancien (1525–1569).

Aussi, bien qu’une comparaison plus poussée entre les propriétés de ces deux types de réseaux puisse s’avérer enrichissante⁹, nous nous en tiendrons à leur caractère hiérarchique.

[Girvan et Newman, 2002], en testant leur méthode de recherche de communauté¹⁰, l’ont appliquée à deux réseaux dont la structure n’était pas clairement connue : un réseau de collaboration et un réseau trophique en milieu marin¹¹. Dans ce dernier cas, ils en ont extrait deux communautés d’égale importance ainsi qu’un reliquat de sommets non rattachés. La division en communautés correspondait assez précisément à une division entre organismes proches ou moyennement éloignés de la surface et ceux – qualifiés de *benthiques* – résidant sur les fonds. Ce qui nous intéresse particulièrement est qu’ils ont relevé que « *ces résultats contrastent avec d’autres techniques utilisées pour analyser les réseaux trophiques qui tendent à rassembler les espèces selon leur niveau trophique plutôt que leur habitat* ».

Nous sommes ainsi tentés d’établir une analogie entre les regroupements de sommets opérés dans un réseau trophique et ceux effectués dans le graphe des participations. De même que les méthodes fondées sur des regroupements hiérarchiques rassemblent des niveaux de prédation qu’on peut apparenter à des rôles dans le sens sociologique, et que celles fondées sur la densité locale du graphe rassemblent des espèces par proximité d’habitat, on peut supposer que les premières dégageront du graphe des participations les entreprises selon leur degré de contrôle ou de concentration des revenus, tandis que les secondes les rassembleront par *marché*. Nous choisissons délibérément ce terme assez souple parce qu’il peut selon le cas se rapporter à un secteur géographique, une clientèle, un produit ou une activité ; selon la stratégie des sociétés contrôlantes, il est vraisemblable qu’on observera des regroupements selon l’un ou l’autre de ces axes.

5.4 La similarité de rôle selon Cooper et Barahona

Les réseaux non dirigés peuvent se voir appliquer des méthodes assez standardisées de regroupements basées sur la densité des liaisons. En comparaison, les méthodes appropriées aux réseaux dirigés sont moins répandues ; [Cooper et Barahona, 2010] ont présenté une méthode fondée sur des motifs de flux entrants et sortants permettant de regrouper les sommets de réseaux dirigés présentant des rôles comparables même lorsqu’ils ne sont

9. À titre d’exemples, les réseaux trophiques apportent des enseignements sur la stabilité d’un système. Dans quelle mesure la disparition d’une espèce se répercute-t-elle à l’ensemble ? L’impact des distances entre les nœuds dans le réseau et de la densité des liaisons peut être comparé avec des recherches semblables sur la contagion des risques dans les systèmes financiers.

Les réseaux trophiques sont également caractérisés par un degré d’*omnivory* ; une espèce omnivore se nourrit dans plus d’un niveau et consomme « les proies de ses proies » [Pimm, 1982]. La comparaison avec la structure de portefeuille de certaines entreprises est ici également attrayante.

10. Leur méthode innovait en ce sens qu’au lieu de rechercher les cœurs denses de communautés, elle déconnecte progressivement ces dernières en retirant les arêtes dotées de la *betweenness centrality* la plus réduite, c.-à-d. par où passent le moins de chemins.

11. Celui de la baie de Cheseapeake, compilé par Baird et Ulanowicz et composé de 33 sommets.

pas eux-mêmes connectés entre eux. Elle offre l'avantage d'être aisément compréhensible ainsi que de mettre en évidence l'importance des flux comme élément caractérisant les réseaux dirigés, en particulier *«ceux dont le caractère dominant est le transfert d'une quantité sous-jacente telle que l'information, l'énergie, la matière. Cette représentation basée sur les rôles est appropriée dans des domaines tels que l'écologie, l'économie, les sciences sociales et le métabolisme cellulaire, où elle peut contribuer [...] à établir des relations fonctionnelles entre des éléments du réseau à première vue distants»*.

Essentiellement, cette méthode consiste à établir pour chaque sommet un vecteur dont une moitié des valeurs est composée du nombre de chemins entrants de longueur $1, 2, \dots, n$ et l'autre moitié de façon similaire pour les chemins sortants. Les vecteurs obtenus sont rassemblés dans une «matrice de flot» de dimensions $n \times 2k$ où n est le nombre de sommets et k la longueur maximale des chemins considérés ; cette matrice est alors l'objet d'un processus de clustering au choix pour rapprocher les rôles similaires. De manière évidente, un tel clustering rapprochera les nœuds-sources n'ayant aucun chemin entrant, et par ailleurs les «puits» (*sinks*) sans chemins sortants.

Dans le détail, les longueurs de chemins ne sont pas inscrites de façon brute dans le vecteur, mais sont modulées par un facteur qui permet de leur accorder un poids plus ou moins important selon la distance.

Pour obtenir les vecteurs des longueurs de chemins, on procède par des puissances successives de la matrice d'adjacence A du graphe ; les longueurs sont extraites au moyen du produit de la matrice ou de sa transposée (respectivement pour les chemins sortants et entrants) par un vecteur-colonne unitaire (noté $\mathbf{1}$) ce qui donne un vecteur du nombre de chemins pour chaque sommet :

- les chemins sortants du sommet i de longueur k sont $[A^k \times \mathbf{1}]_i$;
- les chemins entrants du sommet i de longueur k sont $[(A^T)^k \times \mathbf{1}]_i$.

Les vecteurs ainsi extraits sont concaténés à cet effet sous forme de colonnes dans la matrice de flot, indicés par k , et multipliés par un facteur β^k ¹². Ce facteur β est égal à α/λ_1 , λ_1 étant la plus grande valeur propre de la matrice A et $0 \leq \alpha \leq 1$. L'utilisation de la valeur propre assure la convergence de la séquence des colonnes tandis qu' α permet de moduler le poids relatif des chemins en fonction de leur longueur.

Cooper et Barahona effectuent le clustering de la matrice de flot en utilisant une distance de type *cosine* et un algorithme spectral tout en précisant que le résultat ne dépend pas fondamentalement du choix de l'algorithme.

Sur les données de démonstration qu'ils fournissent, on peut voir l'impact du facteur β (figure 5.4) :

- Lorsqu'il est très petit, l'information relative aux chemins longs s'estompe et les sommets intermédiaires ont alors un profil dominé par la présence d'un chemin

12. Autrement dit, β peut être utilisé à la place de la valeur 1 dans la matrice d'adjacence avant exponentiation.

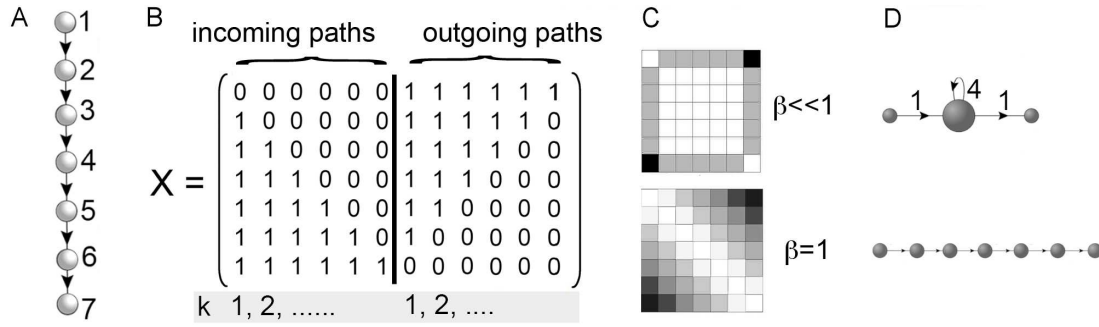


FIGURE 5.4 – Illustration extraite de [Cooper et Barahona, 2010]. A) le réseau de démonstration. B) la matrice de flot sans mise à l’échelle, le nombre de chemins y apparaît de façon brute. C) la matrice de similarité où les sommets sont comparés 2 à 2 et colorés en fonction de leur similarité. D) le clustering qui en résulte selon le choix de β .

- entrant et sortant. Par conséquent, ils sont regroupés en un seul cluster, l’impact de l’environnement local prédomine.
- Lorsque β vaut 1, l’information relative aux chemins de toutes longueurs est préservée et chaque sommet s’individualise.

On trouvera à l’annexe 7.5 p.108 l’application de la méthode sur le même réseau mais avec des valeurs intermédiaires de β , l’utilisation d’un dendrogramme et d’une *heatmap* pour la matrice de flot qui souligne visuellement la modulation apportée aux profils par ce paramètre.

Cette méthode a été testée par ses auteurs sur trois types de réseaux de données réelles «où les flots constituent la nature intrinsèque du système» : un réseau économique des échanges commerciaux dans l’industrie métallurgique, un réseau trophique et un réseau métabolique. Le réseau trophique est reproduit en figure 5.5. Pour celui-ci, les auteurs ne précisent pas comment la spatialisation du réseau d’origine a été établie ni l’étagement hiérarchique du réseau condensé. On ignore en particulier si c’est le clustering obtenu qui a déterminé la disposition du réseau entier ou s’il a confirmé une spatialisation qui avait été obtenue par d’autres moyens¹³. Nous reviendrons sur cet aspect dans notre adaptation de la méthode.

5.5 Application au réseau des participations

Nous avons vu que le procédé de Cooper et Barahona a été appliqué avec succès sur un réseau trophique et qu’il en souligne la structure étagée. Nous sommes donc tentés de le reprendre pour le réseau des participations, mais son application directe escamote le poids relatif des actionnaires et retourne une image imparfaite de sa structure. Pour s’en

13. L’image du réseau ne figurait pas dans les documents cités en référence.

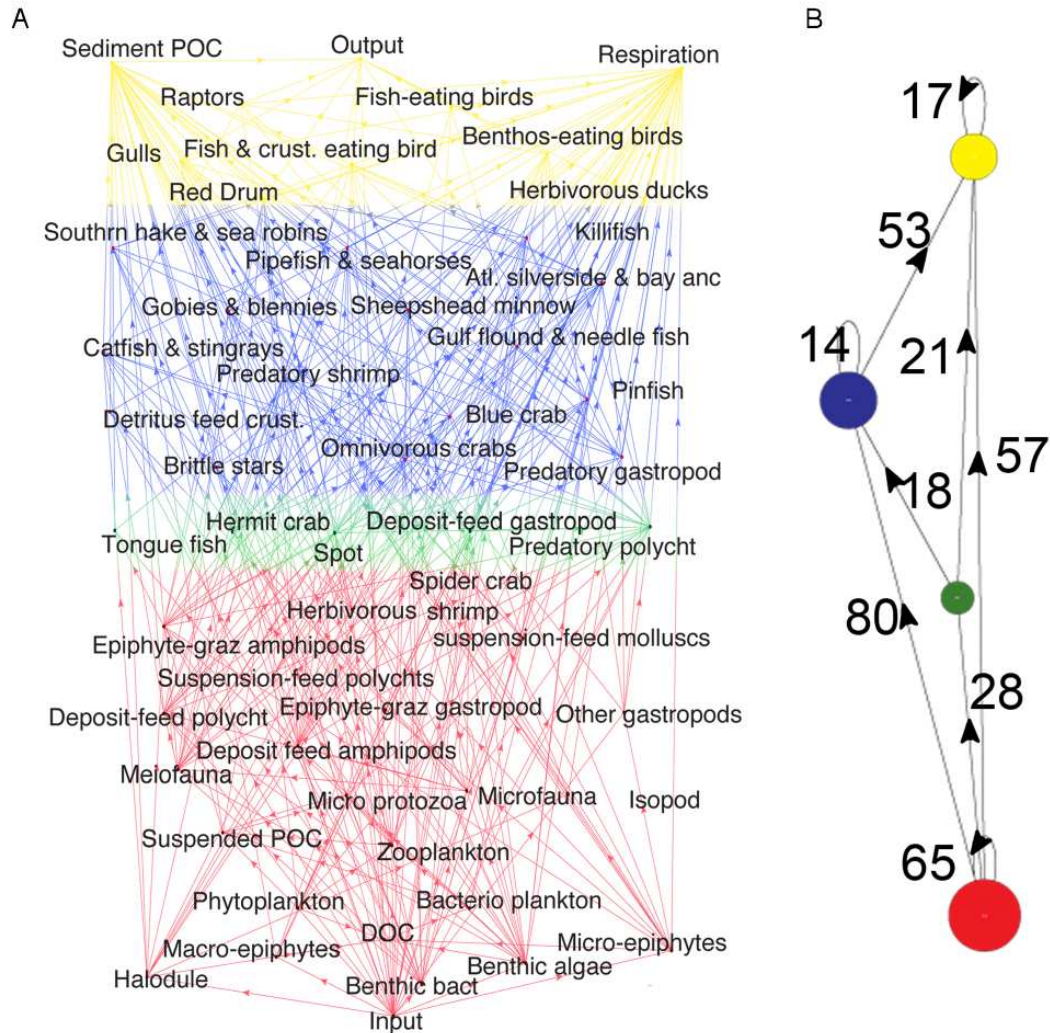


FIGURE 5.5 – Illustration extraite de [Cooper et Barahona, 2010]. À droite, la condensation du réseau où chaque cluster obtenu constitue un sommet, «indiquant le flot de carbone au travers du réseau». Les auteurs ne le précisent pas explicitement mais le poids des arcs est selon toute vraisemblance la somme des poids des arcs joignant les sommets de chaque cluster.

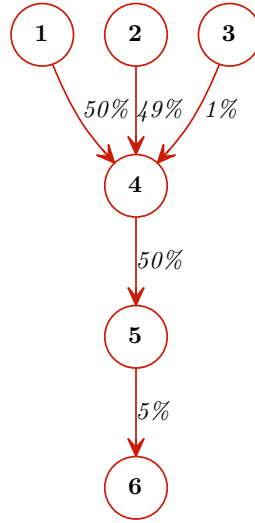


FIGURE 5.6 – Exemple de rôles différents selon le poids des chemins. 3 se distingue de 1 et 2 malgré une position semblable.

convaincre, on observera la figure 5.6 où l'un des actionnaires situés au sommet de la pyramide joue manifestement un rôle différent des deux autres.

Bien que les sommets 1 à 3 présentent la même chaîne de participations sortantes, le sommet 3, avec sa participation directe minime, ne dispose que de participations négligeables dans les niveaux inférieurs. Son rôle est représentatif des petits actionnaires vis-à-vis des *blockholders* qui disposent d'un contrôle effectif sur leurs filiales.

Barahona et Cooper utilisent une matrice d'adjacence binaire dont les valeurs sont multipliées par un facteur β global pour moduler la part des chemins distants dans le profil. On remarque que les puissances successives effectuées sur cette matrice rappellent le mode de calcul de la propriété intégrée par Chapelle. Pour moduler le profil des chemins non pas de façon globale mais de façon différenciée selon les participations détenues, on remplace donc la valeur de 1 correspondant à un arc dans la matrice d'adjacence par son poids effectif entre 0 et 1. À titre d'exemple, on peut observer le résultat du partitionnement sur le réseau de la figure précédente. Avec un partitionnement en 4 clusters, l'application directe de Cooper et Barahona répartit exactement les sommets selon leur étagement, les sommets 1 à 3 étant réunis sans distinction de leurs participations. En utilisant celles-ci dans l'évaluation des profils, le sommet minoritaire 3 se trouve réuni avec le 6. Bien qu'il puisse sembler étonnant de rapprocher deux sommets situés aux niveaux extrêmes de la pyramide, il faut réaliser que dans les deux cas, les profils sont composés de valeurs nulles à l'exception d'une seule valeur très petite ; dans la réalité économique, les participations de cet ordre, qu'elles soient entrantes ou sortantes, ne différencient guère par elles seules les parties concernées.

En plus de cette modification, nous avons aussi modifié l'ordre des colonnes de la matrice

$$\bar{x} = \frac{\sum_{i=0}^{2k-1} V[i].X[i]}{\sum_{i=0}^{2k-1} V[i]} = \frac{-k.V[0] + -(k-1).V[1] \cdots + (k-1).V[2k-2] + k.V[2k-1]}{V[0] + V[1] \cdots + V[2k-1]}$$

FIGURE 5.7 – Calcul du barycentre d’un profil représenté par un vecteur V et un vecteur associé X des longueurs de chemins. $X[i]$ vaut de $-k$ à -1 pour les chemins entrants, de 1 à k pour les chemins sortants, $V[i]$ est la valeur calculée pour cette longueur de chemin.

de façon à la symétriser : les valeurs correspondant aux chemins sortants et entrants de longueur 1 sont ainsi contigües et de part et d’autre de l’axe central de la matrice, ceux de longueur 2 sont à 2 colonnes d’écart de l’axe central, et ainsi de suite. Outre que cette disposition nous semble visuellement plus suggestive, elle permet aussi d’associer à chaque profil une valeur correspondant à la localisation d’un *centre de gravité*.

En considérant chaque valeur du profil comme une masse et les longueurs de chemins comme des coordonnées allant de $-k$ pour les plus longs chemins entrants à $+k$ pour les plus longs chemins sortants, les profils composés uniquement de chemins entrants auront un centre de gravité de coordonnée négative et ceux composés uniquement de chemins sortants auront un centre de gravité de coordonnée positive. Ceux disposant de chemins entrants et sortants s’étageront le long de cet axe selon leurs poids respectifs. Ces valeurs fournissent une indication¹⁴ de la position des sommets (ou de l’ensemble d’un cluster en calculant la moyenne de ses éléments) dans la structure hiérarchique du réseau ; leur distribution (figure 5.8 p.77 selon le dataset est en relation avec celle des clusters (ainsi qu’avec leur taille). Elles seront également utilisées pour le positionnement des sommets du réseau dans sa forme condensée.

Dans une première phase exploratoire, nous avons associé dans un même graphique la matrice des profils, une matrice carrée représentant les distances mutuelles entre ces profils et un dendrogramme¹⁵. Ces 3 composantes du graphique sont synchronisées, c.-à-d. que le tri effectué sur les feuilles du dendrogramme est répercuté sur les matrices de façon à y regrouper les profils les plus similaires.

L’intérêt de cette représentation à ce stade est double :

- le nombre de clusters ne doit pas être prédéterminé. On peut le retenir *a posteriori* par observation du dendrogramme en le sectionnant à une hauteur au choix¹⁶ ;

14. Elles présentent cependant un appauvrissement par rapport à l’information offerte par le vecteur du profil ; des profils différents peuvent présenter une même localisation de leur barycentre.

15. Le dendrogramme est la représentation d’un clustering hiérarchique agglomératif où l’on regroupe de façon itérative les clusters les plus proches, ces clusters étant au départ les éléments isolés, puis des regroupements de plus en plus importants.

16. La librairie Scipy offre dans son module `scipy.cluster.hierarchy.dendrogram` la possibilité de visualiser ce sectionnement par une option `color_threshold` qui a pour effet de colorer différemment les branches de l’arbre sous le niveau indiqué.

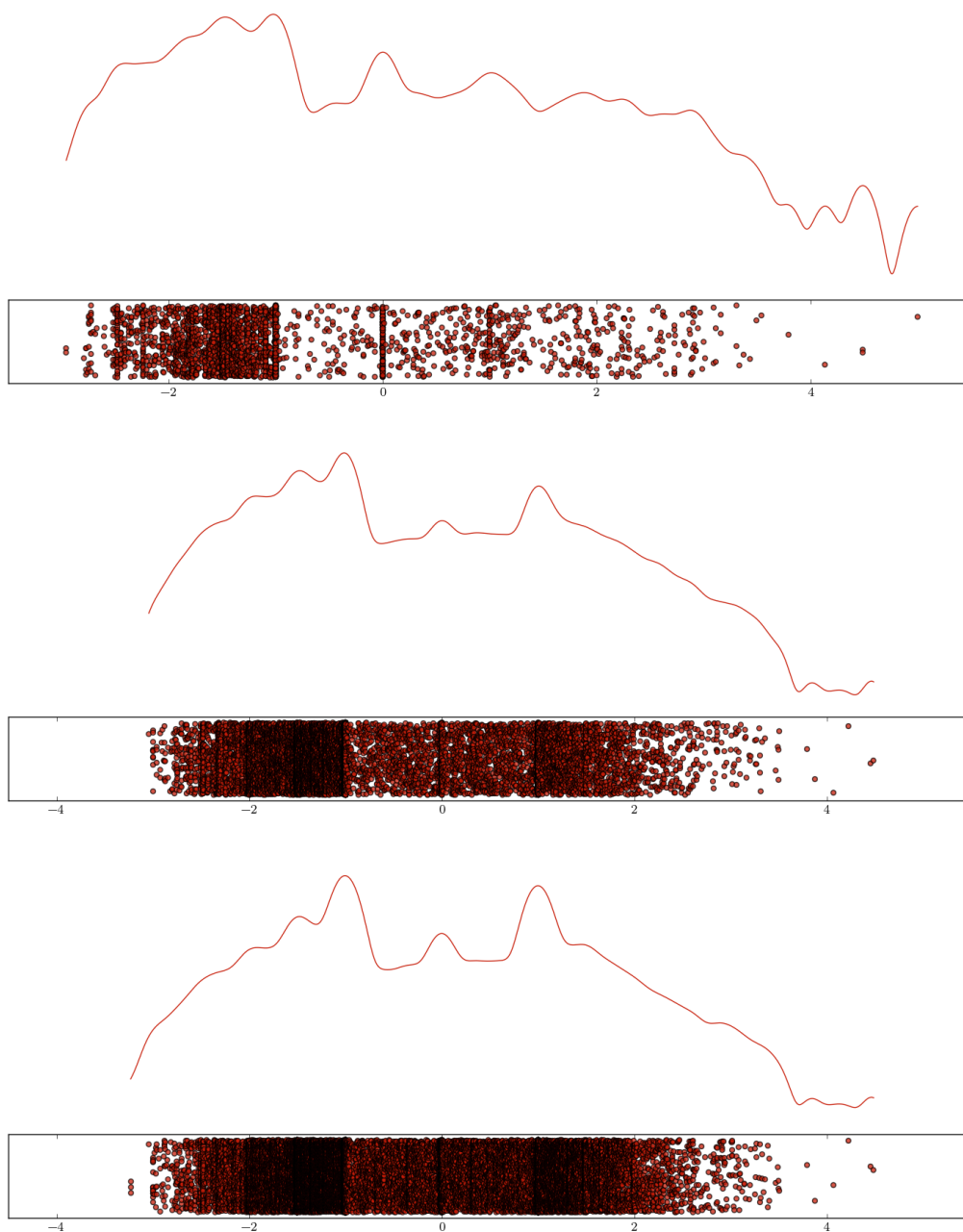


FIGURE 5.8 – Distribution des sommets sur l'axe horizontal selon leur barycentre, associé à une dispersion aléatoire sur l'axe vertical pour en faciliter la lecture, ainsi qu'une courbe de densité. De haut en bas : sociétés cotées et associées, composante principale, réseau entier. Pour ce dernier, l'importance des chemins de longueur 1, entrants et sortants, dus à l'abondance des composantes de petite taille est manifeste.

- la présence de clusters fortement homogènes est mise en évidence par la présence de *blocs* de couleur sombre uniforme sur la diagonale de la matrice de similarité.

Appliquée au sous-ensemble des sociétés cotées, elle souligne – ce qui était assez attendu – la présence majoritaire de sommets sans chemins sortants et dotés de chemins entrants peu profonds, ainsi qu’une proportion assez réduite de sociétés disposant de longs chemins sortants (figure 5.9). Elle montre également la structure fort déséquilibrée du dendrogramme, les distances (c.-à-d. la dissimilarité des profils) entre les sociétés dotées de longs chemins sortants étant nettement plus élevées qu’entre les filiales. Cela a pour effet que si on cherche à scinder les clusters des niveaux inférieurs pour obtenir plus de détails à ce niveau (en situant plus bas le niveau de sectionnement), le nombre de clusters se multiplie au niveau supérieur au point que les sommets y sont quasi isolés. Autrement dit, il est difficile de rassembler en un même cluster les sociétés des niveaux supérieurs sans perdre en détails dans les niveaux inférieurs.

Ce type de clustering ne permettant pas de traiter des ensembles de données importants, nous avons ensuite utilisé l’algorithme *k-means* pour effectuer un clustering sur l’ensemble des données. Cet algorithme consiste en un processus itératif où l’on forme des clusters qui minimisent les distances entre leurs éléments et un prototype représentatif qui est lui-même adapté en fonction des éléments ajoutés ou retirés du cluster, jusqu’à stabilisation. L’intérêt du *k-means* réside à la fois dans sa rapidité et l’existence de ces prototypes qui en permettent une interprétation aisée.

Nous avons choisi arbitrairement un nombre de clusters de 5¹⁷, observé quel était le profil moyen de chacun en le traçant sur un graphique et appliqué une standardisation des valeurs des profils de façon à ce qu’ils présentent une moyenne nulle et une variance de 1¹⁸. La figure 5.10 reprend les résultats pour les trois ensembles de données. On y observe que les profils moyens présentent une forme en cloche dont les sommets se déplacent de la gauche vers la droite à mesure que le cluster se situe plus haut dans la hiérarchie. La forme de ces courbes est relativement fluide sauf pour les clusters comportant une majorité écrasante de sommets présentant quasi exclusivement des chemins entrants ou sortants de longueur 1 ; la part de tels sommets étant considérable dans le graphe complet (voir la distribution en figure 5.8), on y repère les clusters 2 et 3 présentant ces pics et qui sont également les plus importants en nombre de sommets.

5.5.1 Sélection du nombre de clusters

Le choix d’un nombre approprié de clusters est un sujet de recherche à part entière. Les méthodes proposées dans la littérature (par exemple [Chiang et Mirkin, 2010] pour le *k-means*) ne semblent cependant pas avoir été communément implémentées et aucune

17. À première vue, il semblait que cet ordre de grandeur suffisait pour décrire l’étagement hiérarchique des sociétés. Barahona et Cooper utilisent 4 niveaux dans leur analyse de food-web. On verra plus loin pourquoi nous avons envisagé par la suite un nombre de clusters plus élevé.

18. Au moyen de la fonction `sklearn.preprocessing.scale`.

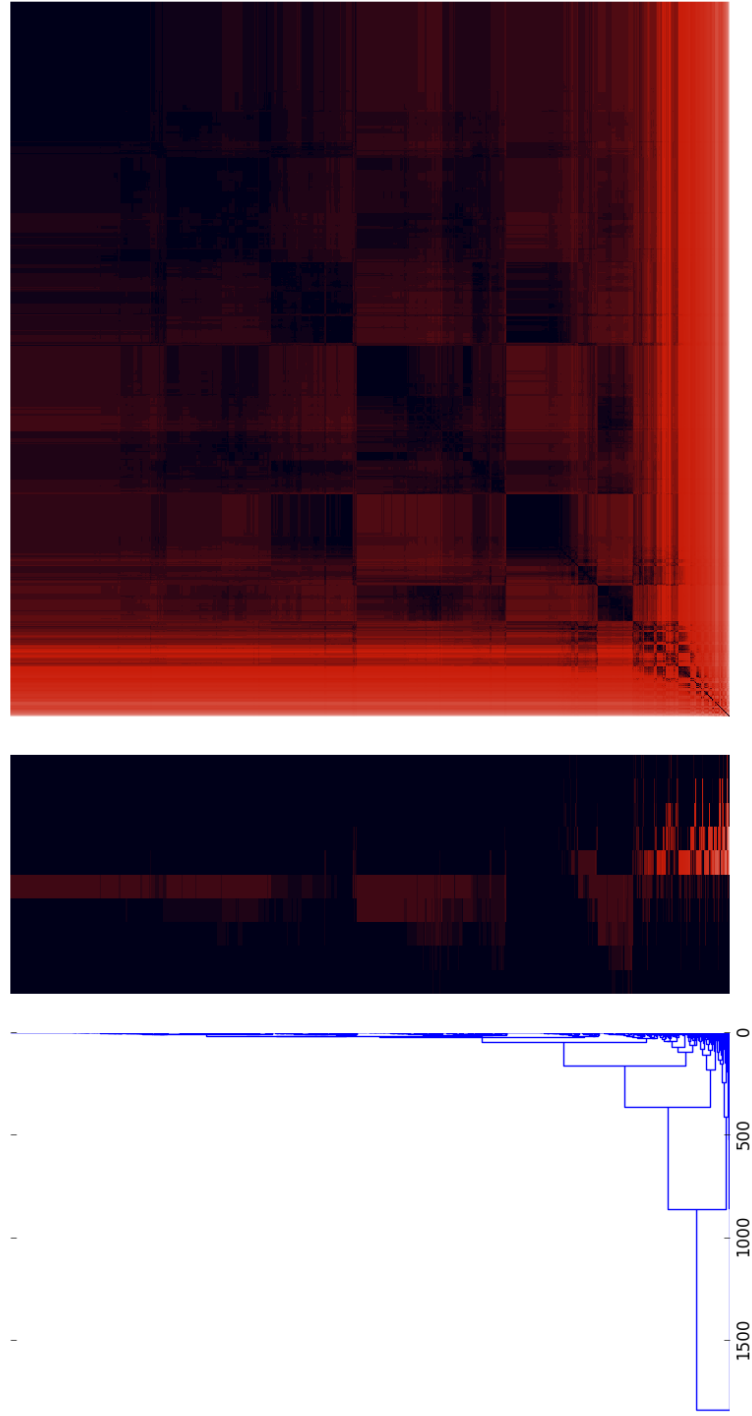


FIGURE 5.9 – Clustering hiérarchique du sous-ensemble des sociétés cotées.

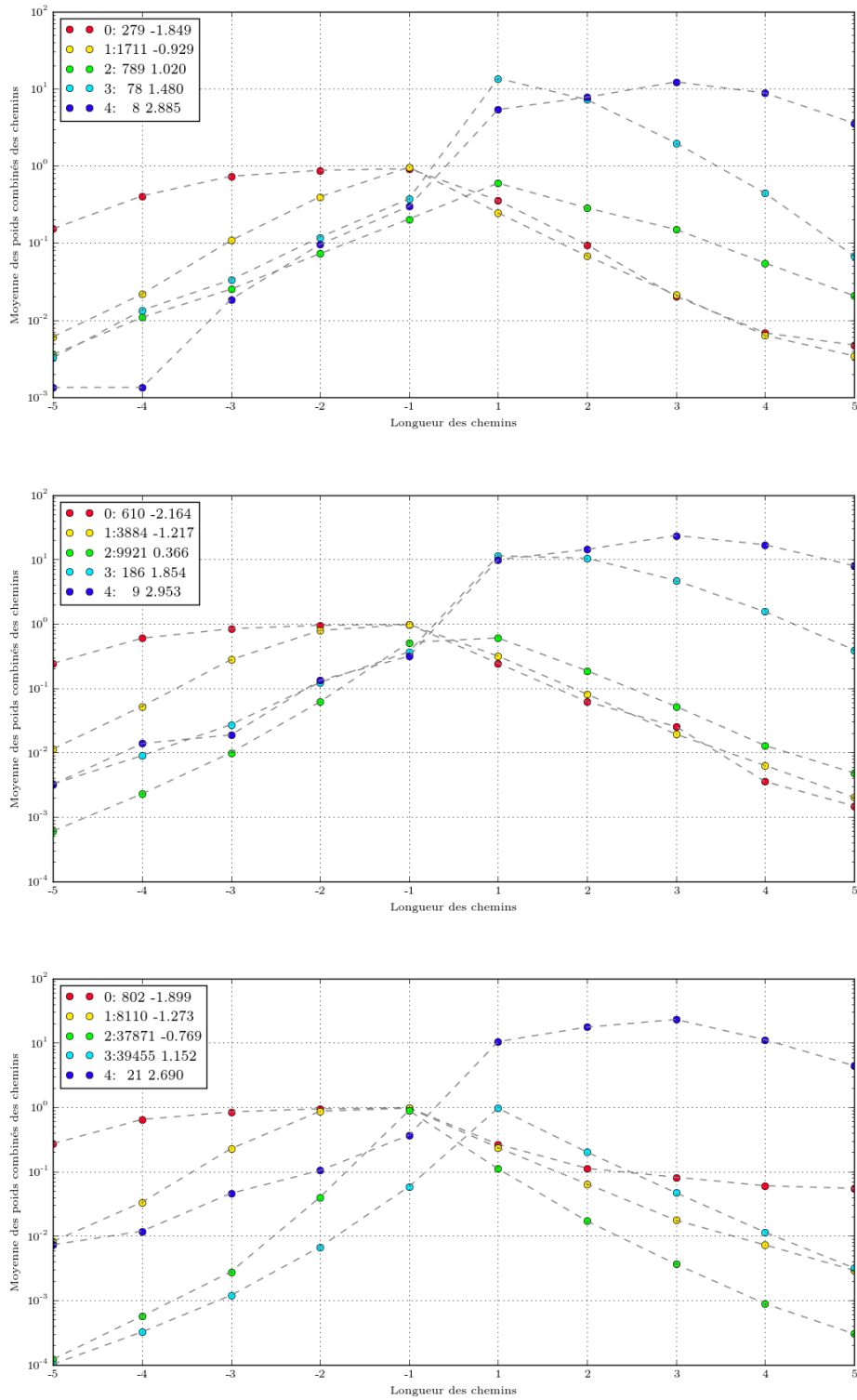


FIGURE 5.10 – Profils moyens des clusters obtenus par *k-means*. L'échelle sur l'axe vertical est logarithmique. Pour chaque cluster, la légende précise l'index (0 étant le plus bas), le nombre de sommets, le barycentre.

ne peut être proposée comme une solution universelle. Vu les délais impartis, nous nous sommes contenté de recourir à des solutions offertes par la librairie *Scikit-Learn* :

- l'examen de l'*inertie* du clustering : il s'agit de la somme des distances des instances aux centres des clusters auxquels ils sont assignés ;
- l'usage des *silhouettes* proposé par [Rousseeuw, 1987].

Pour la première, il s'agit de la fonction à minimiser lors de la recherche itérative des centres des clusters par l'algorithme ; le problème est qu'elle ne peut être utilisée en valeur absolue pour déterminer le meilleur nombre de clusters puisqu'elle est d'autant plus réduite que le nombre de clusters augmente (à la limite, si chaque instance constitue un cluster par elle-même, l'inertie devient nulle). On trace donc un graphique de l'inertie en fonction du nombre de clusters et on cherche à repérer une inflexion (*elbow*) dans cette courbe dénotant qu'en deçà d'un certain seuil, elle ne diminue plus de façon significative.

Pour la seconde, la méthode consiste à comparer la distance moyenne entre chaque instance et celles qui font partie du même cluster (a) et avec celles qui font partie du cluster le plus proche (b). Le coefficient de silhouette pour une instance est alors $(b - a) / \max(a, b)$ et permet d'évaluer si la situation aurait été meilleure en affectant l'instance à cet autre cluster.

Le problème s'additionnait dans notre cas à celui de déterminer la meilleure façon de prétraiter la matrice des valeurs fournie par l'algorithme de Barahona et Cooper ; en dehors de la normalisation évoquée dans la section précédente, nous avons aussi tenté, pour chaque sommet, de diviser :

- les valeurs des chemins sortants par le nombre total de ses descendants ;
- les valeurs des chemins entrants par le nombre total de ses ancêtres.

Cette manière de normaliser les valeurs avait pour objectif de rapprocher les sommets présentant des profils de chemins comparables indépendamment de leur nombre, et en particulier de pouvoir rassembler les sociétés quel que soit le nombre de leurs filiales, directes et indirectes, pour autant qu'elles soient reliées par une même structure de participations.

Malheureusement, l'application de ces deux méthodes n'a pas permis de faire ressortir un choix particulier, celui d'une méthode de normalisation comme celui d'un nombre de clusters donnant des résultats peu réguliers selon l'ensemble de données retenu.

D'autre part, nous avons signalé précédemment (p.78) que l'augmentation du nombre de clusters ne permettait pas de scinder de la façon attendue les clusters de grande taille ; le résultat en était plutôt de disperser les plus petits. L'explication tient à ce que la majorité des profils ne comprennent que des valeurs très petites alors que quelques-uns présentent des valeurs à la fois nettement plus élevées et plus variées (d'une colonne à l'autre). Pour y remédier, nous avons alors pré-traité les profils en remplaçant les valeurs par leur logarithme. Le résultat en est remarquable à la fois par la redistribution des sommets entre les clusters et par la visualisation des données au moyen des composantes principales.

La méthode des composantes principales (*PCA*) permet de réduire le nombre de dimensions d'un dataset en représentant les données dans un nouveau système de coordonnées où la variance est maximisée. Cette méthode montre fréquemment ses limites lorsqu'une des dimensions d'origine offre un poids nettement supérieur aux autres et tend à rester dominante dans le nouveau système. C'est en effet ce qui est advenu lors de la tentative d'effectuer une *PCA* sur les données brutes, centrées ou normalisées : la visualisation obtenue reflétait grossièrement le nuage de points obtenu en utilisant comme coordonnées les degrés sortants et entrants des sommets du graphe. Une *PCA* réalisée avec les données logarithmiques offre par contre une visualisation tout à fait différente en faisant apparaître des regroupements très distincts (voir figure 5.11). Cette visualisation nous a amené à reconsidérer le nombre de clusters à rechercher dans la mesure où il s'est avéré qu'augmenter ce nombre de clusters tendait à les faire correspondre plus ou moins aux regroupements de la *PCA*.

Le clustering hiérarchique sur les données logarithmiques (figure 5.12, p.84) peut être comparé avec celui effectué précédemment (figure 5.9, p.79) ; on observe que la largeur des branches du dendrogramme est mieux équilibrée, que les distances y sont réduites et qu'une coupe sur celui-ci est susceptible de fournir des clusters de tailles moins inégales. La visualisation de la matrice des flots sous forme de *heatmap* fournit une explication à ce comportement ; les écarts entre les valeurs ont été « tassés » et l'existence d'une valeur différente de 0 dans une cellule de cette matrice, plus que sa valeur, devient prépondérante. Autrement dit, cette matrice devient quasi binaire, le clustering tend à regrouper les sommets en fonction de la longueur maximale de leurs chemins entrants et sortants, sans tenir compte des nuances apportées par le nombre de chemins intermédiaires.

Pour des raisons de scalabilité, il n'a pas été possible d'obtenir un dendrogramme du réseau entier. Mais les figures 5.13 et 5.14 permettent de comparer les profils obtenus à partir des données brutes ou de leurs logarithmes. Dans les deux cas, on a représenté pour chaque cluster l'ensemble des profils ainsi que leur moyenne (en couleur). On peut voir que pour les données brutes, la masse importante de sommets de degré entrant 1 et de degré sortant 1 ont été intégrés dans les clusters 0 et 1 ; ils influencent notablement ces 2 profils moyens en créant un pic aigu aux valeurs de -1 et 1. Les clusters supérieurs sont beaucoup moins importants. Néanmoins, tous les profils moyens ont des profils assez souples du fait qu'ils comportent des valeurs non nulles pour la plupart des colonnes.

Avec les données logarithmiques, un effet de seuil a réduit à zéro les colonnes dont les valeurs étaient très faibles. Le cluster 0 rassemble ainsi les sommets ayant des chemins entrants de longueur 2, le 2 les sommets de chemins entrants 1, le 3 des chemins entrants et sortants de longueur 2, le 4 des chemins sortants de longueur 1. Les 2 autres comportent des profils plus variés. Le résultat est que les sommets associés à des chemins courts (les sociétés des petites composantes) sont affectés à des clusters propres.

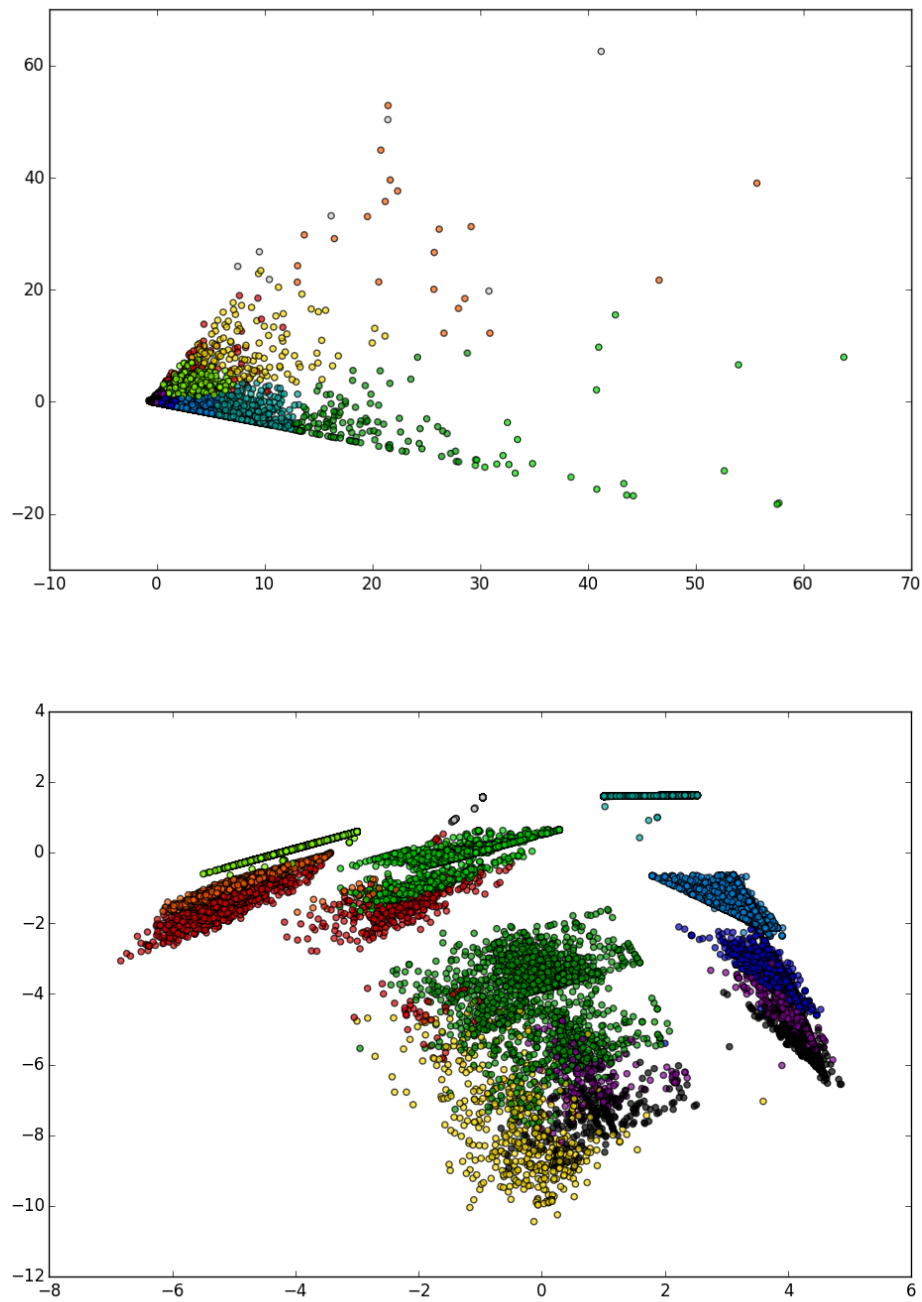


FIGURE 5.11 – Analyse en composantes principales sur les données du réseau entier, au-dessus avec les données brutes, au-dessous avec pré-traitement logarithmique. Les couleurs correspondent à une division en 12 clusters ; dans le second cas, ils se rapprochent des sous-ensembles isolés par la PCA.

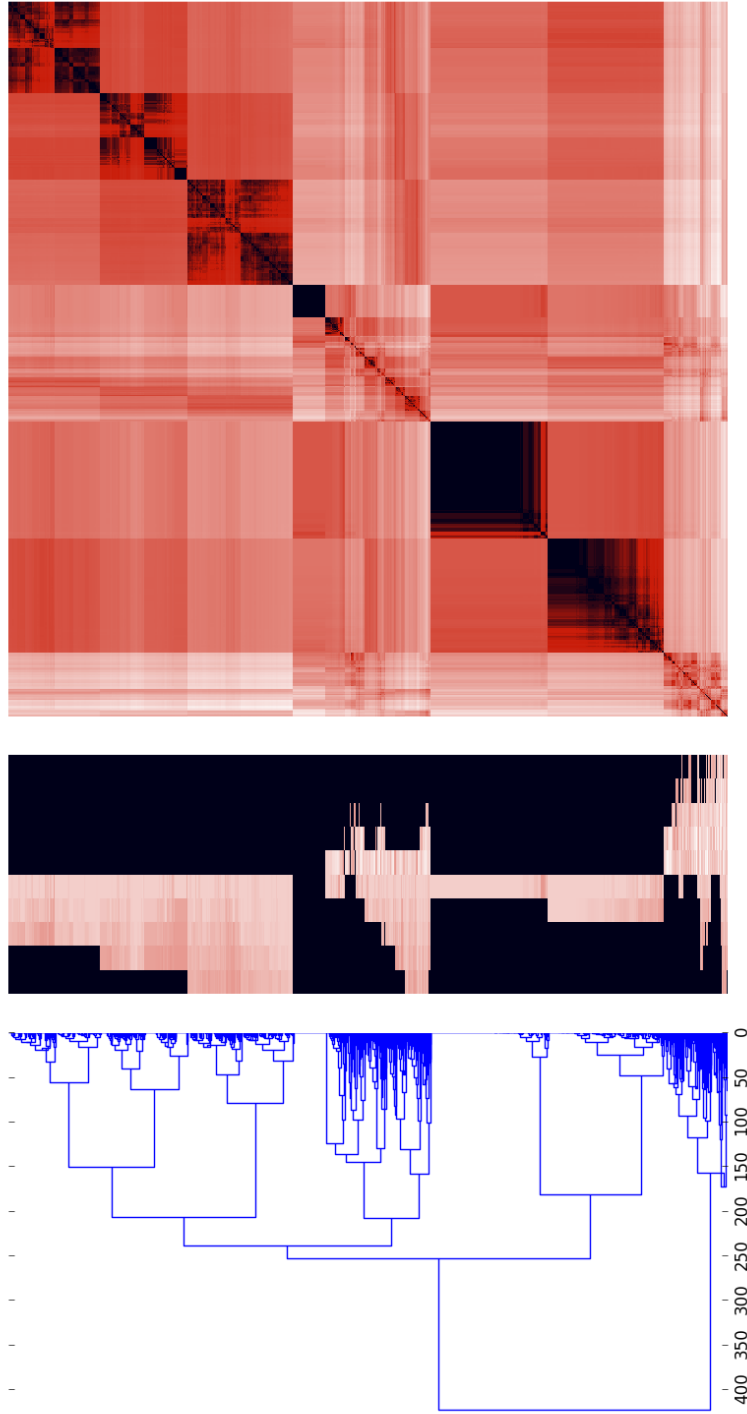
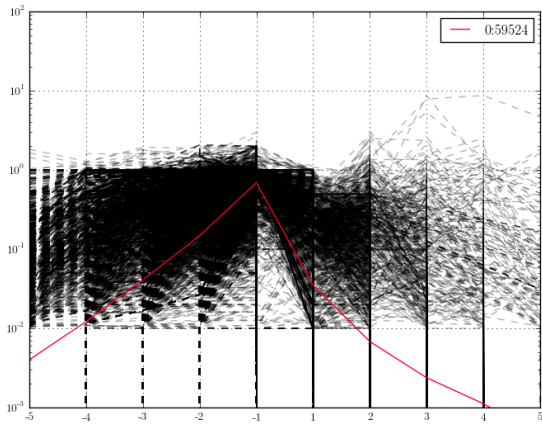
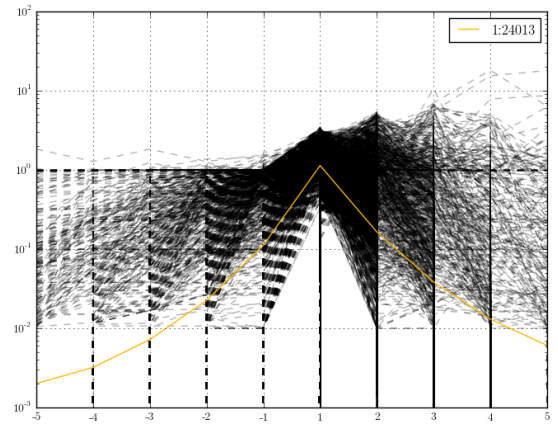


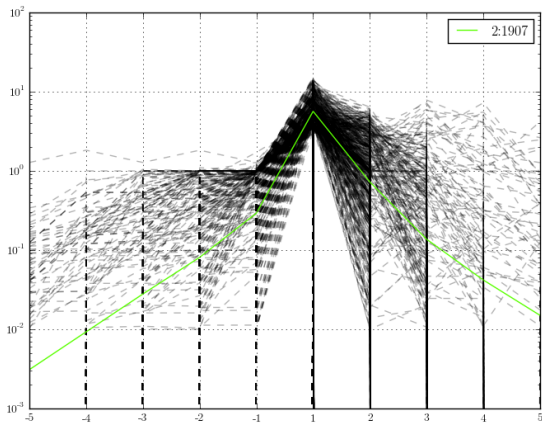
FIGURE 5.12 – Clustering hiérarchique du sous-ensemble des sociétés cotées avec pré-traitement logarithmique.



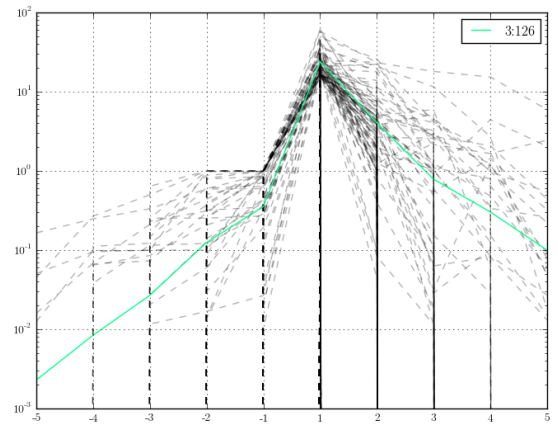
(a) Cluster 0



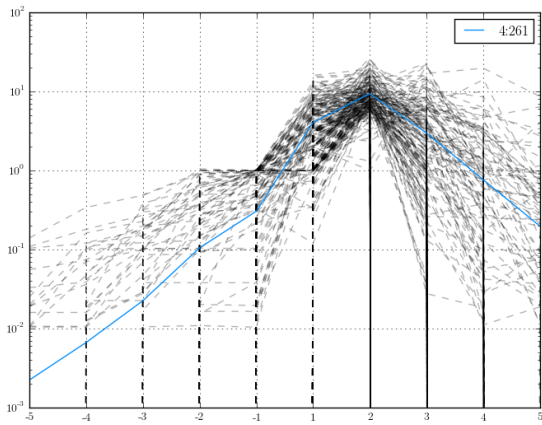
(b) Cluster 1



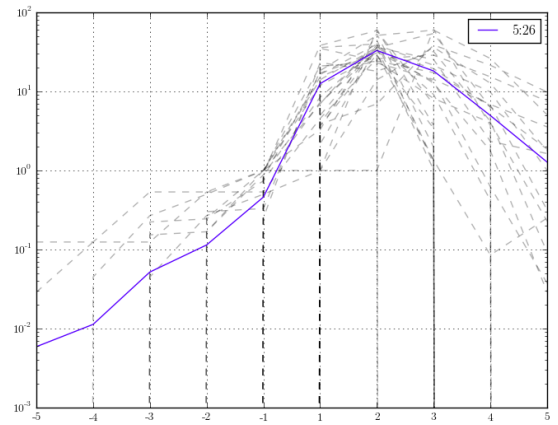
(c) Cluster 2



(d) Cluster 3

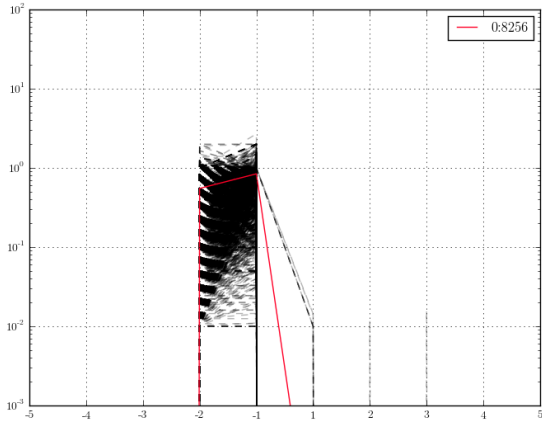


(e) Cluster 4

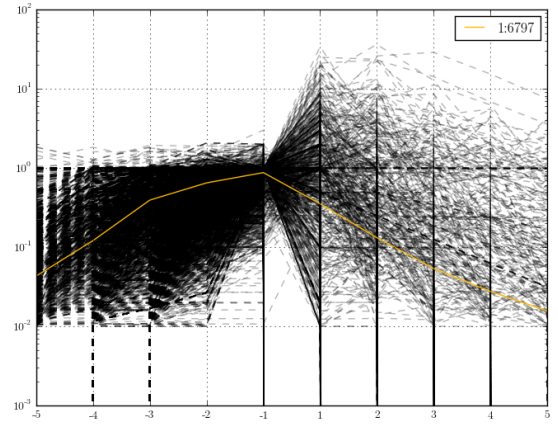


(f) Cluster 5

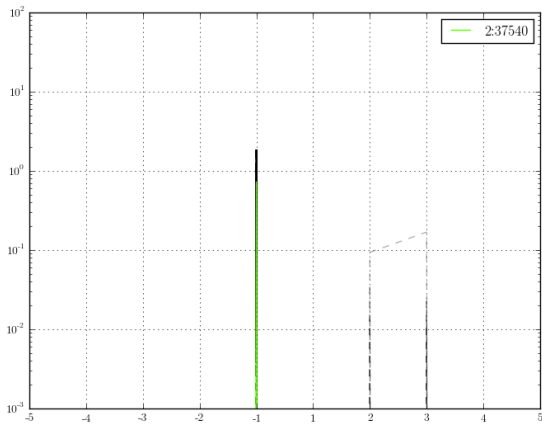
FIGURE 5.13 – Profils pour le réseau entier, 6 clusters, données brutes.



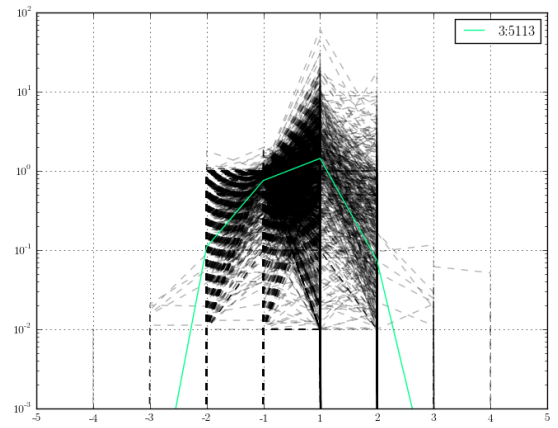
(a) Cluster 0



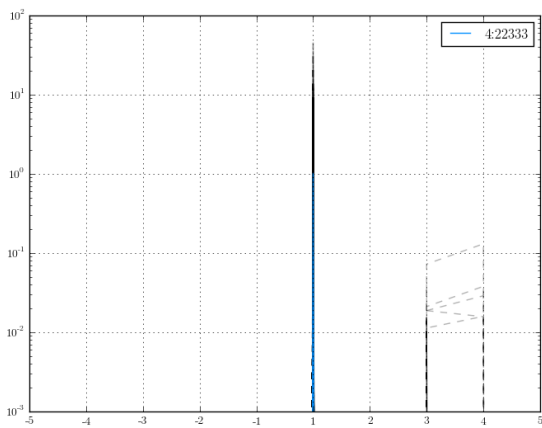
(b) Cluster 1



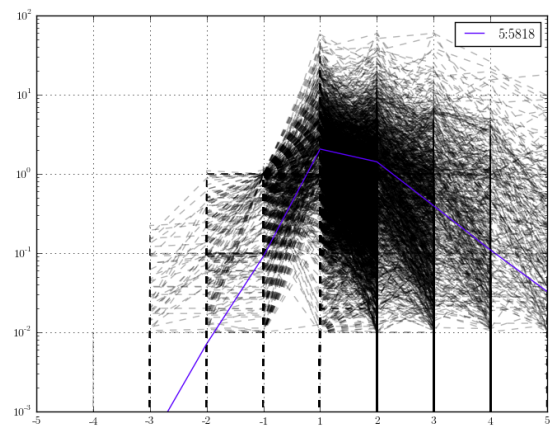
(c) Cluster 2



(d) Cluster 3

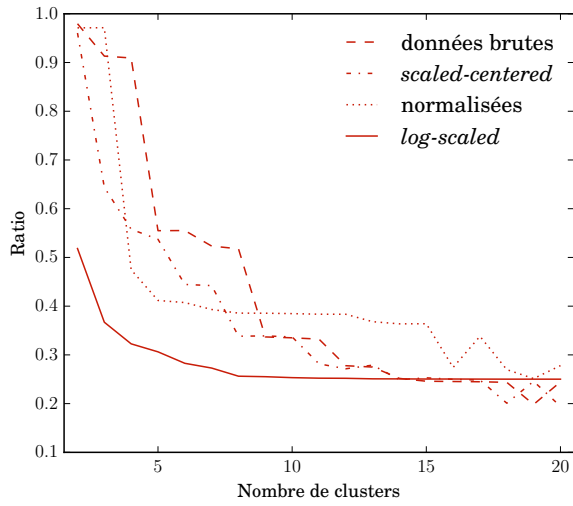


(e) Cluster 4

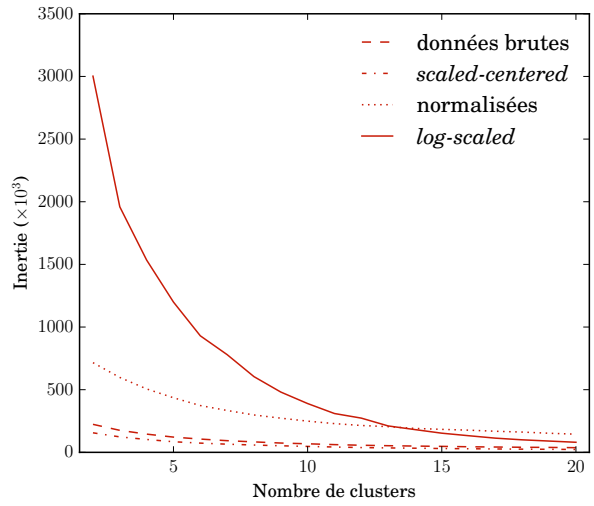


(f) Cluster 5

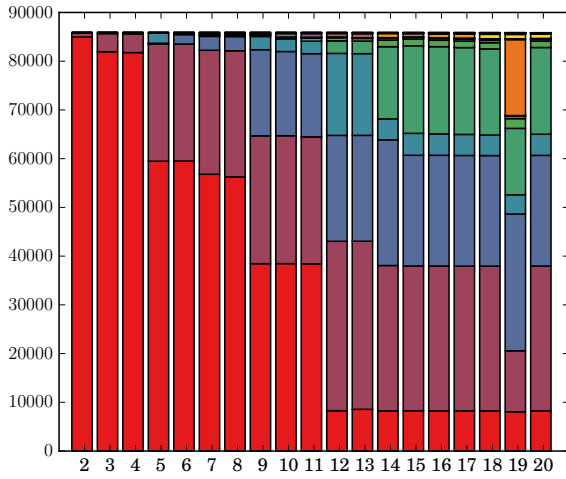
FIGURE 5.14 – Profils pour le réseau entier, 6 clusters, valeurs logarithmiques.



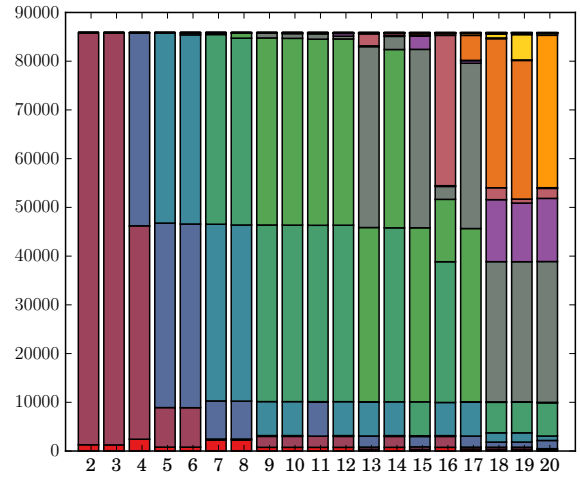
(a) Proportion d'arcs intra-clusters, RE



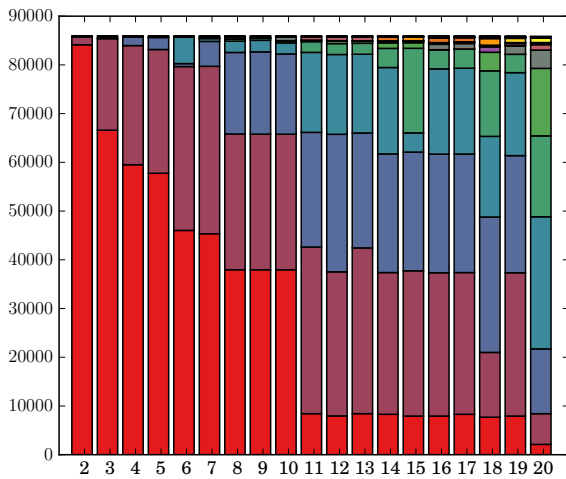
(b) Inertie, RE



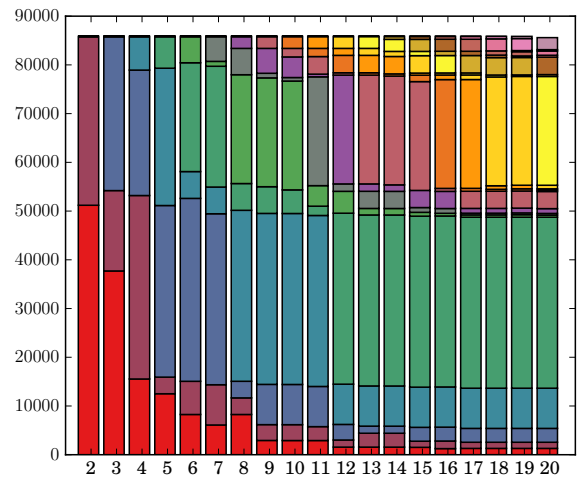
(c) Données brutes



(d) Données "scaled-centered"



(e) Données normalisées



(f) Données logarithmiques

FIGURE 5.15 – a) Proportion d'arcs intra-clusters. b) L'inertie ne suggère aucun niveau de clustering. c) à f) Clusterings de 2 à 20 pour le réseau entier et les 4 manières de traiter les données de la matrice de flot.

5.5.2 Validation du nombre de clusters

Les méthodes de recherche du nombre de clusters le plus approprié par l’inertie ou les silhouettes n’ayant pas fourni de résultats concluants, c’est encore une analogie avec les réseaux trophiques qui nous a suggéré d’examiner les résultats du clustering de la façon suivante. On a vu que dans ces réseaux, les niveaux trophiques correspondaient au nombre d’étapes qui les séparaient du niveau le plus bas. La plupart des espèces consomment des proies situées dans le niveau inférieur à celui qu’elles occupent elles-mêmes. Du fait de l’existence d’espèces omnivores qui puisent dans plus d’un niveau, il n’est pas possible de combiner les chaînes alimentaires pour former un réseau dont les niveaux soient indiscutablement définis¹⁹, mais il devrait en principe présenter un minimum de liaisons au sein d’un même niveau.

On obtient aisément pour un clustering déterminé le nombre d’arcs ayant leur source et leur cible au sein du même cluster ; on calcule alors le rapport entre ce nombre et le total des arcs du réseau. Le résultat pour des clusterings s’étageant de 2 à 20 et pour 4 choix de traitement de la matrice de flot est visible sur la figure 5.15a. Il apparaît immédiatement que la courbe basée sur l’utilisation des grandeurs logarithmiques des profils se distingue des autres. Non seulement elle démarre avec un rapport nettement moins défavorable que les autres, mais l’augmentation du nombre de clusters impacte la courbe de façon progressive et celle-ci se stabilise rapidement. À l’opposé, les trois autres présentent des décrochements brutaux à des seuils qui ne concordent pas.

Pour bien comprendre l’origine de ces décrochements, nous avons effectué une visualisation des clusterings successifs sous la forme de diagrammes en bâtons empilés (figures 5.15c à 5.15f). Ceux-ci sont apparus très révélateurs de la manière dont les clusters se scindent au cours de l’augmentation de leur nombre.

Ainsi, pour les données brutes, les décrochements à 5, 9 et 12 clusters correspondent à des seuils où les plus gros clusters se divisent ou répartissent différemment leurs éléments. La vue simultanée des clusterings permet aussi de remarquer des séquences où l’augmentation du nombre de clusters affecte une portion limitée du diagramme (qui se subdivise) sans en influencer le reste. On remarque bien que pour les clusterings de 9 à 11 ou de 15 à 18, les limites entre les plus gros clusters restent constantes et que si l’on cherche à minimiser le nombre de clusters, le premier clustering d’une telle séquence est un meilleur choix.

Pour la version “*scaled-centered*”, une division a lieu au clustering 4 à un niveau qui ne bouge quasiment pas jusqu’au clustering 16. À partir de 5 et de la création d’un petit cluster inférieur, un second niveau de scission aussi stable s’ajoute. De là jusqu’à 13-14, il n’y a pas de corrélation évidente entre ce diagramme et la baisse progressive mais un peu hachée du nombre de liens intra-clusters. À partir de 16 apparaît une forme d’instabilité

19. “In general, there appears to be no discrete mapping of real populations to integral trophic levels” [Ulanowicz, 1995]. L’affectation exacte d’une espèce à un niveau trophique semble avoir divisé les biologistes : “The trophic concept is not just wrong at its edges : it is erroneous in fundamental ways” [Cousins, 1985].

associée à la création des clusters intermédiaires entre les deux plus importants.

La version normalisée montre une séquence stable à partir de 11 sans qu'on puisse dire clairement jusqu'où elle s'étend, les clusterings 18 et 20 semblant être assez atypiques alors que le 19 est similaire aux précédents. Curieusement, le nombre de liens intra-clusters montre un palier très stable à partir de 5 alors que cela ne se distingue en rien sur le diagramme en bâtons.

Pour le traitement logarithmique des données, l'essentiel de la répartition a été opéré entre deux gros clusters encadrant une zone intermédiaire à partir de 8 ; les augmentations ultérieures n'en modifient guère les limites. Comme ce clustering correspond au plancher de la courbe des liens intra-clusters, c'est celui qui semble le plus intéressant. On a également observé que les barycentres des 2 plus gros clusters (donnée non reprise sur les figures) se stabilisaient à des valeurs quasi exactes de 1 et -1 à partir de 7 clusters, puis précisément sans aucune décimale à partir de 10 – ce sont d'ailleurs les seules valeurs entières.

Une autre forme de visualisation est susceptible de fournir un indice de qualité des clusterings pour un graphe ; il s'agit d'une matrice d'adjacence réordonnée de façon symétrique selon les clusters. Dans le cas d'une recherche de communautés, la densité plus forte des arcs au sein de celles-ci entraîne la formation de blocs rassemblés sur la diagonale (de façon comparable à la visualisation par secteur NACE qui montrait les participations intra-catégories). À l'opposé, une division par rôles entraîne pour ceux n'ayant pas de relations avec eux-mêmes la présence de blocs vides sur la diagonale. L'interprétation visuelle de ces matrices constitue en quelque sorte une application informelle du *block-modeling* mentionné précédemment.

Pour bien interpréter une telle matrice, on doit avoir à l'esprit que les lignes correspondent aux destinations des arcs partant du cluster et les colonnes à leurs origines. Sachant que les clusters ont été ordonnés en fonction du barycentre moyen de leurs sommets, si le clustering obtenu reflète bien une forme de hiérarchie, on peut espérer obtenir le comportement idéal suivant :

- Sur une ligne de la matrice, les blocs de la diagonale devraient être de faible densité lorsque le taux d'arcs intra-cluster est faible ;
- Comme les arcs devraient aller en majorité d'un cluster à un autre plus bas dans la hiérarchie, les blocs à droite de la diagonale sur une ligne devraient être plus denses ;
- Inversement, les blocs à gauche de la diagonale sur une ligne devraient l'être moins (dans le cas contraire, cela signifie que des arcs remontent les niveaux hiérarchiques).

Les résultats, visibles dans les figures 5.17 à 5.20, s'en écartent assez bien. On peut faire les observations suivantes :

- Dans tous les cas, la présence des gros clusters liés aux sommets des petites composantes est prépondérante ;

- La version logarithmique présente des clusters aux profils plus distincts les uns des autres ; inversement, celle des données brutes présente des lignes dont les densités des sections sont moins tranchées ;
- Le cluster le plus élevé présente toujours un fort taux de liaisons internes ainsi que des liens remontant des clusters inférieurs jusqu'à lui (il semble donc erroné de croire en l'existence d'un noyau de «détenteurs ultimes» dominant sans partage le reste du réseau) ;
- Il subsiste toujours une certaine proportion de clusters présentant une forte densité de liens intra-clusters (sur la diagonale des matrices).

Dans tous les cas, le résultat du clustering ne correspond donc pas à un étagement hiérarchique strict comme on peut l'observer sur le réseau trophique étudié par Cooper et Barahona, où les liens à contre-sens étaient absents. Les liens intra-clusters y étaient cependant présent dans une proportion comparable (0,27, ce qui est égal au résultat des données logarithmiques).

Au regard des résultats, la recherche d'une division stricte en niveaux hiérarchiques n'est-elle pas illusoire ? On a vu que même pour les réseaux trophiques, l'existence de niveaux n'est pas acceptée de façon unanime. [Johnson *et al.*, 2001], lorsqu'ils ont proposé l'équivalence régulière pour l'analyse des réseaux trophiques, mentionnent ces critiques à l'égard des niveaux :

“Dissatisfied with early approaches to the trophic level concept, which are now seen as too coarse, broad, or even arbitrary to be useful, some ecologists have sought to develop new trophic concepts that better capture the inherent complexity of systems of trophic interactions (Cousins 1987; Polis 1991). These newer approaches often refer to the functional role (Cummins, 1974) of a species (or sets of related species) in a community, and also tend to see these roles as “basic building blocks” (Hawkins and MacMahon, 1989) of communities and therefore aids in reducing the complexity of an ecological system.”

À la suite de quoi ils recommandent le concept de rôle associé à l'équivalence régulière pour dégager les «rôles fonctionnels» des systèmes écologiques et aider à la réduction de leur complexité.

Le barycentre moyen des clusters obtenu par notre application ne doit donc pas être considéré comme un critère strict d'étagement mais comme une information complémentaire (la section suivante en propose une ébauche d'application). L'importance des sommets provenant des très petites composantes souligne par ailleurs un problème posé par l'interprétation de ces niveaux : une paire de sommets isolée est constituée d'une source et d'un puits, lesquels présentent des profils avec des barycentres de 1 et -1. Ils sont donc situés à mi-hauteur et séparés des sources et des puits des autres clusters. Bien que simple dans son énoncé, la question de savoir où positionner les extrémités des chemins dans un tel modèle ne connaît pas de réponse évidente (figure 5.16)

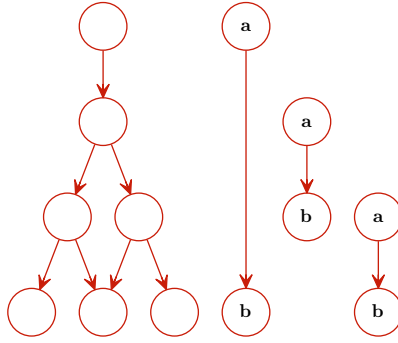


FIGURE 5.16 – Faut-il positionner a et b au niveau des sources et des puits des pyramides à niveaux multiples ?

5.5.3 Représentation condensée du réseau

Dans cette section, nous présentons les résultats de quelques essais pour visualiser le graphe image correspondant au résultat d'un clustering (celui que nous avons retenu est celui en 7 clusters sur les données logarithmiques). Sa représentation sous forme de matrice d'adjacence est en effet peu intuitive.

Le logiciel Graphviz [Gansner et North, 2000] offre des algorithmes éprouvés de visualisation de graphes et présente l'avantage d'accepter comme format d'entrée un langage aisé à produire. Il a ainsi été possible de spécifier certains éléments du *layout* de la façon suivante :

- associer le diamètre des sommets au nombre de sociétés dans chaque cluster et la largeur des arcs au nombre d'arcs existant entre les sommets du graphe d'origine ;
- associer le diamètre des sommets à la somme des fonds propres des sociétés de chaque cluster et la largeur des arcs à une évaluation du cash-flow transféré.

Pour ce dernier cas, le sens des arcs a été inversé pour souligner le sens du flot. Le poids des arcs est établi à partir de la somme des produits des résultats nets des sources (positifs uniquement) par la valeur des participations (le poids des arcs du réseau d'origine).

Dans les deux cas, un paramètre a été utilisé pour éliminer les arcs d'un poids au-dessous d'un certain seuil.

Les figures 5.21 et 5.22 montrent le positionnement obtenu avec *dot*, l'algorithme hiérarchique de Graphviz. Ce dernier contraint les sommet à se positionner sur des niveaux de valeur entière selon des heuristiques assez complexes : ici, on a imposé le positionnement des clusters 1 et 7 à la base et au sommet ce qui a forcé les positions intermédiaires. Le problème avec *dot* est que cette façon d'imposer les *sources* et les *puits* ne garantit pas

que l'étagement des niveaux intermédiaires sera respecté. D'autre part, *dot* dispose d'un algorithme apte à tracer élégamment les arcs en évitant de les superposer aux sommets mais il donne des résultats parfois bizarres lorsqu'on force la position des sommets.

Les figures suivantes montrent les résultats en utilisant un autre algorithme de graphviz – *fdp* – qui permet d'imposer des coordonnées pour positionner tous les sommets de façon absolue ; la valeur du barycentre des clusters est utilisée pour y et (alternativement) 1 et -1 pour x . On laisse à *fdp* la possibilité d'étendre le graphe pour que les sommets ne se recouvrent pas. Malheureusement, le recouvrement de ceux-ci par les arcs n'est pas évité.

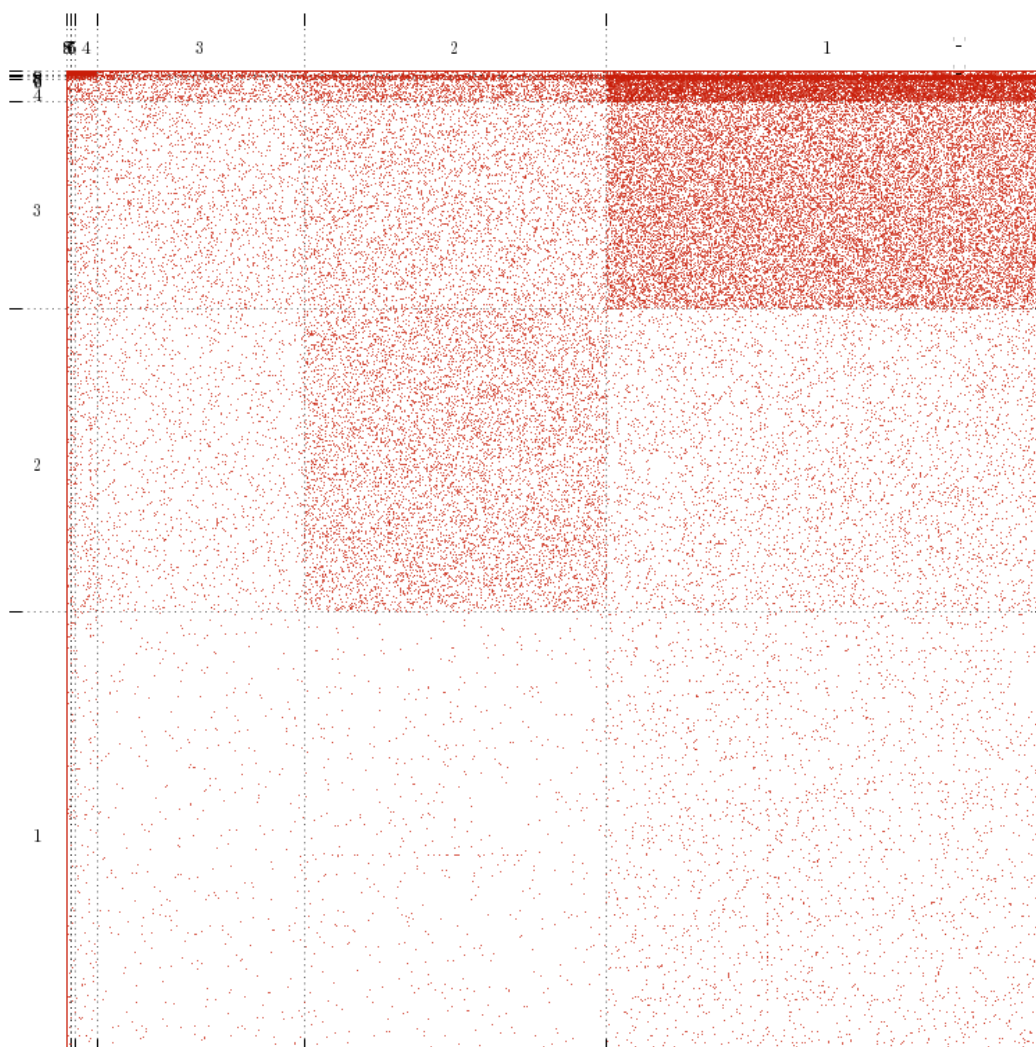


FIGURE 5.17 – Matrice d’adjacence du réseau entier, données brutes, 9 clusters.

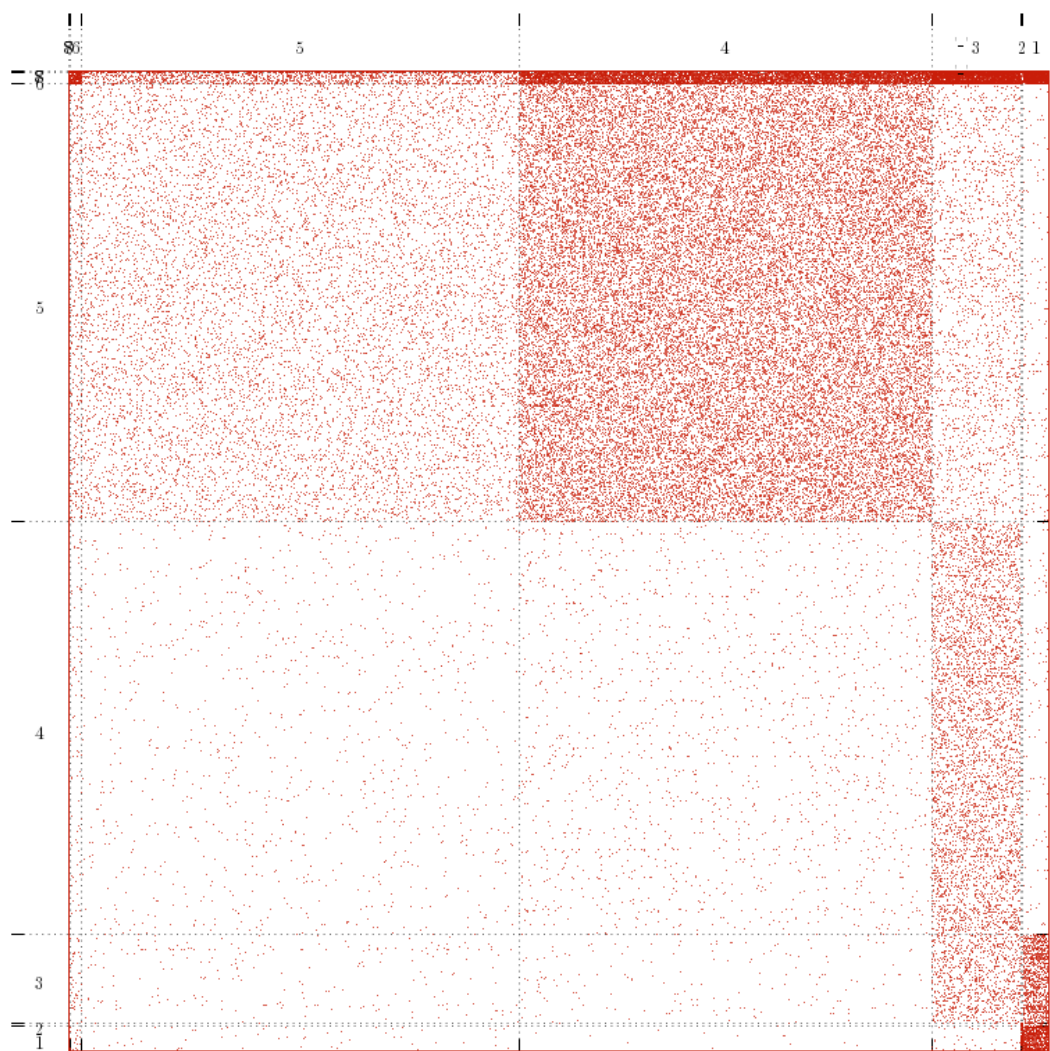


FIGURE 5.18 – Matrice d’adjacence du réseau entier, «centré-mis à l’échelle» 8 clusters.

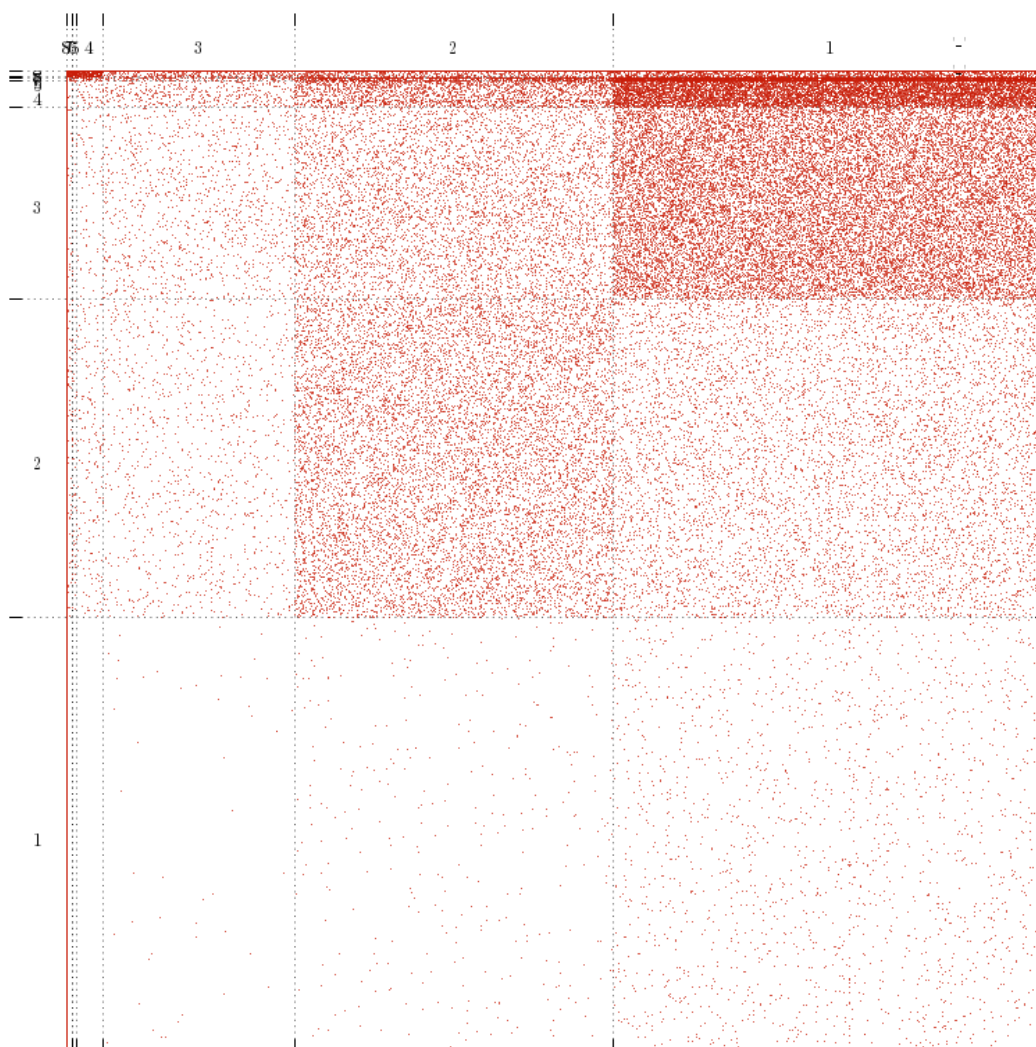


FIGURE 5.19 – Matrice d'adjacence du réseau entier, normalisé, 8 clusters.

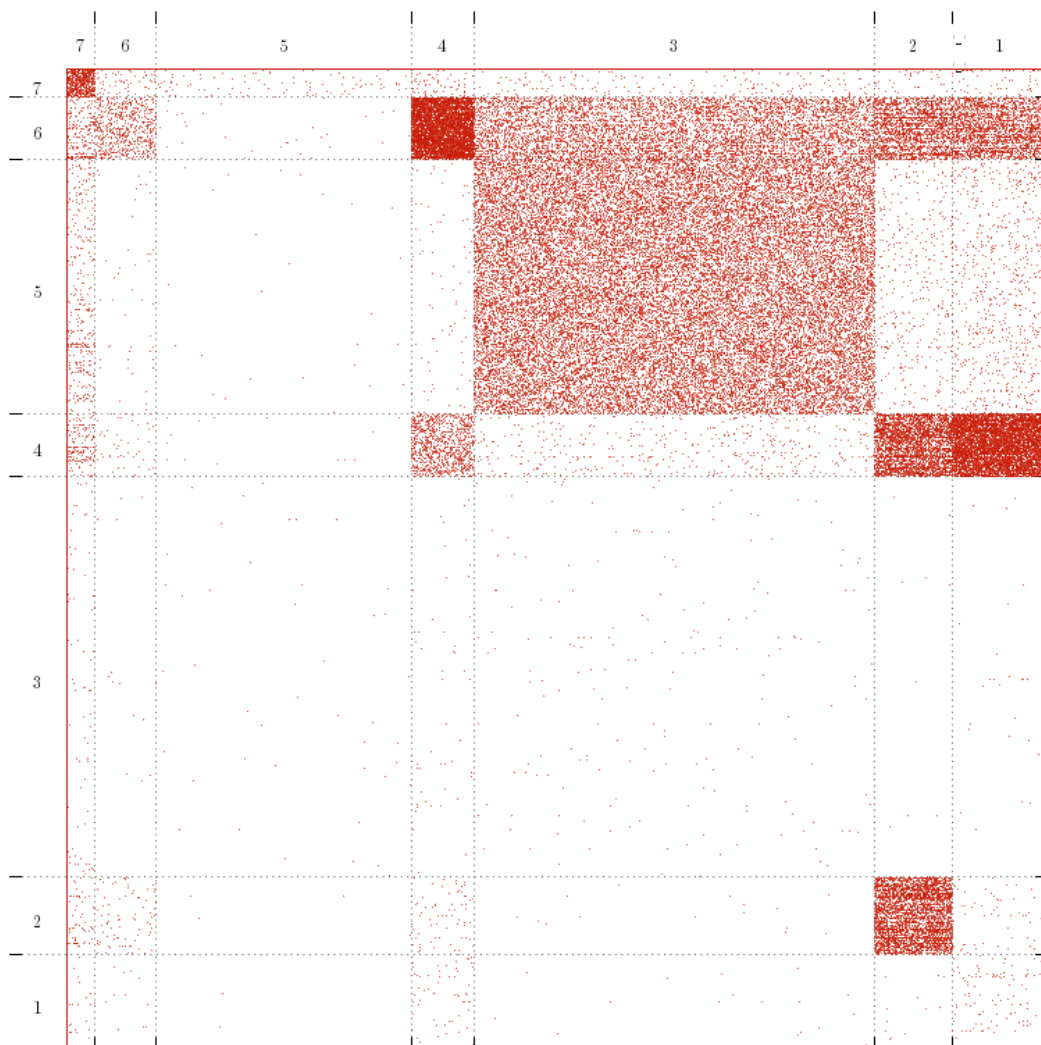


FIGURE 5.20 – Matrice d’adjacence du réseau entier, traitement logarithmique, 7 clusters.

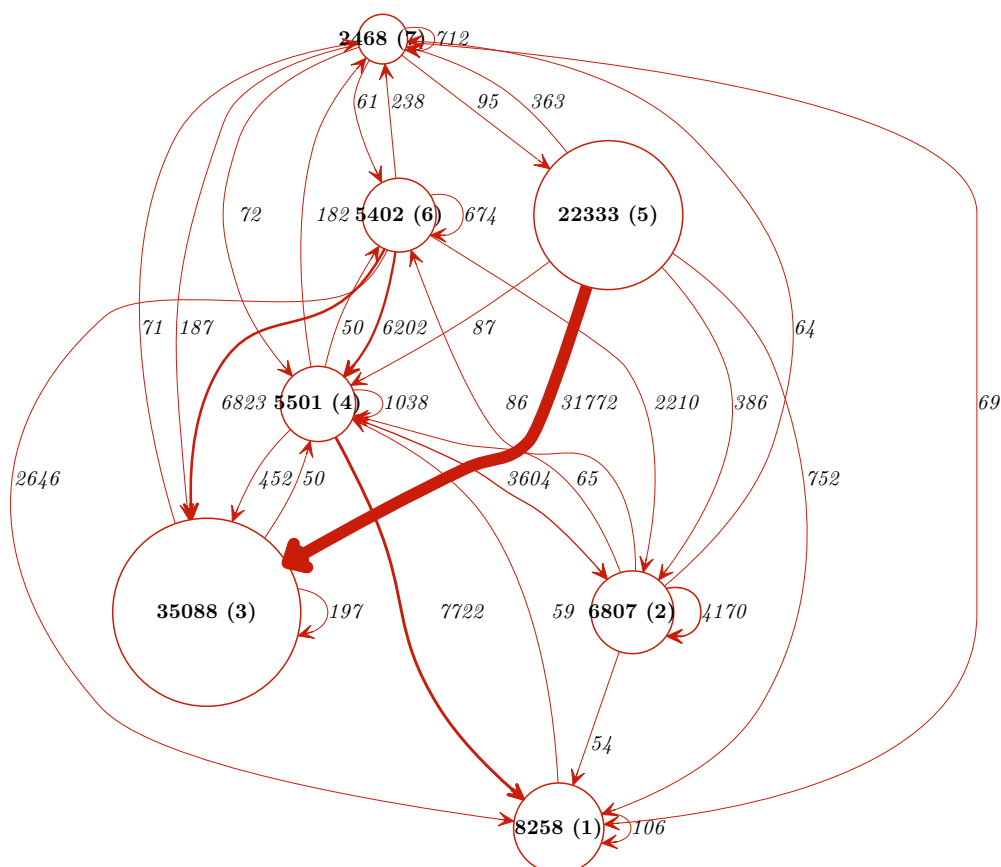


FIGURE 5.21 – Visualisations du graphe image avec Graphviz, positionnement par *dot*. Taille des sommets et des arcs d'après le nombre de sociétés et de participations. Les arcs représentant moins de 50 participations ne sont pas représentés.

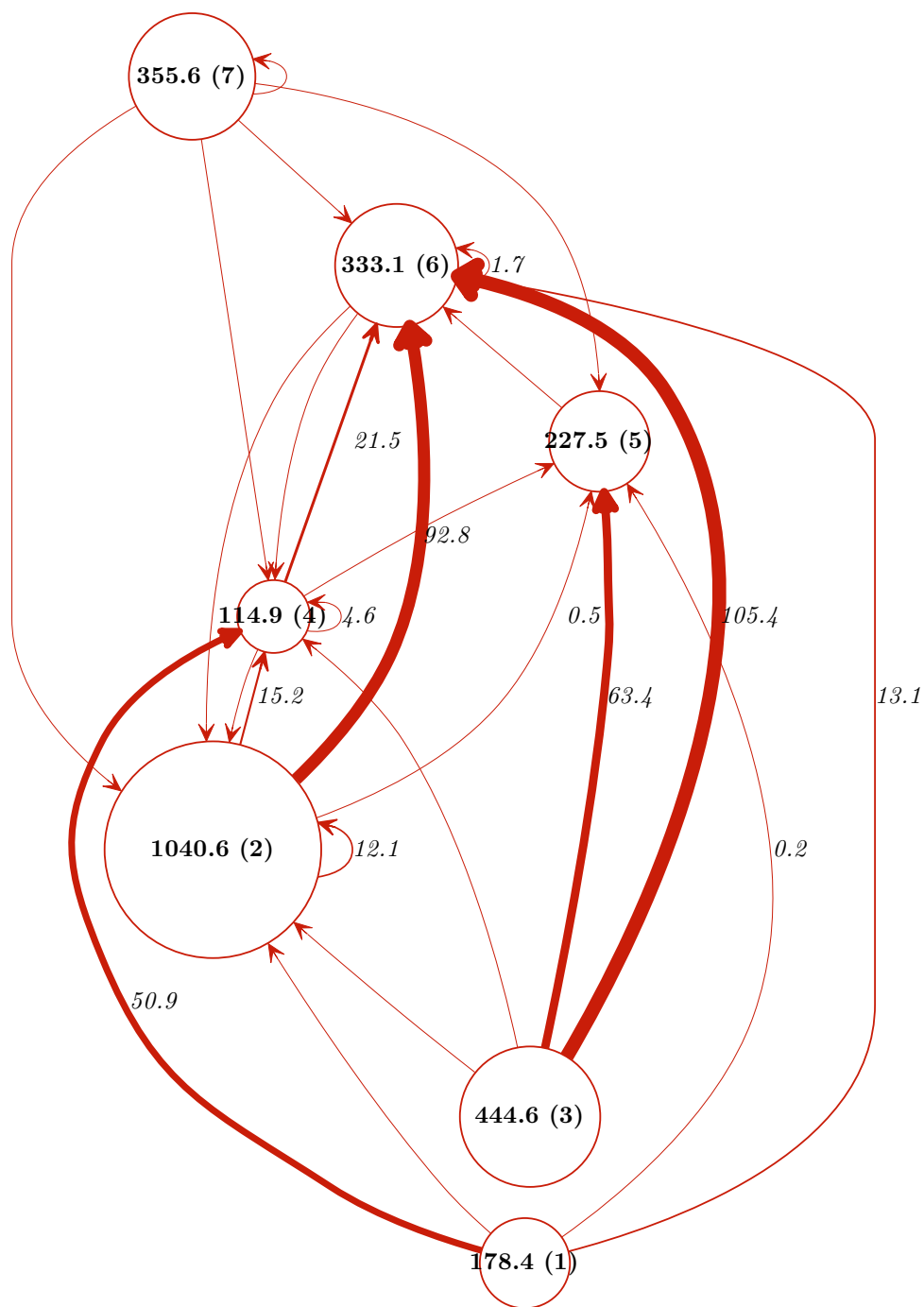
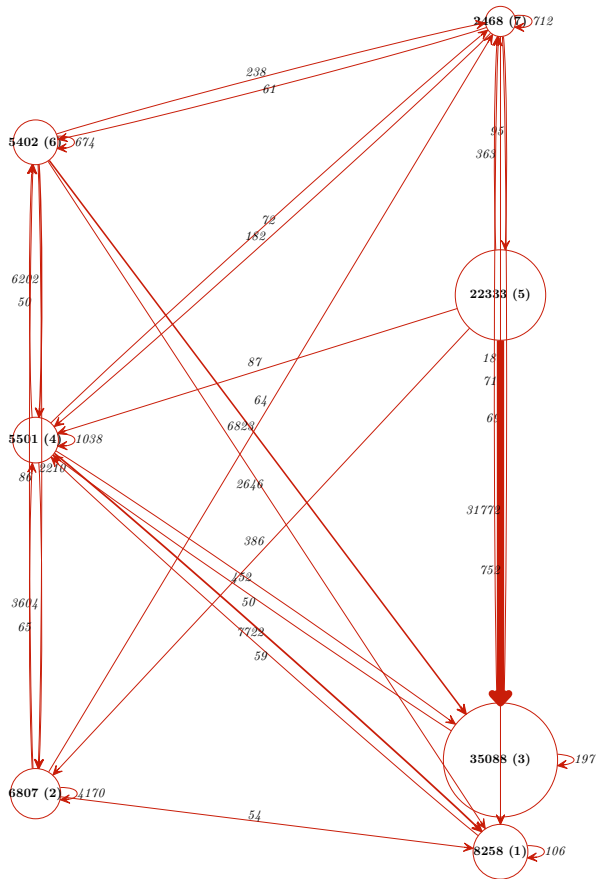
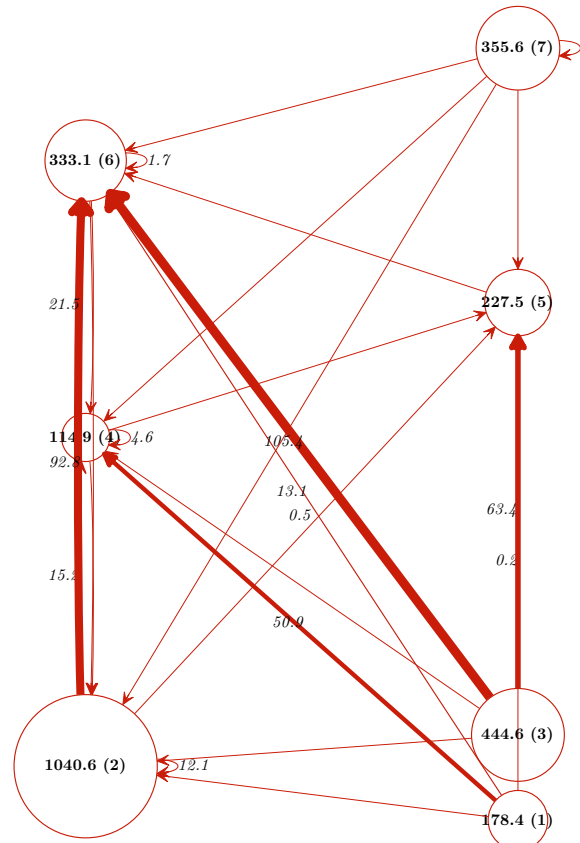


FIGURE 5.22 – Visualisations du graphe image avec Graphviz, positionnement par *dot*. Taille des sommets et des arcs d'après le total des fonds propres et des transferts de cash-flow, les valeurs sont exprimées en milliards d'euros. Les transferts de cash-flow inférieurs à 100000000 ne sont pas représentés.



(a) Nombre de sociétés et de participations



(b) Fonds propres et cash-flow

FIGURE 5.23 – Visualisations du graphe image avec Graphviz, positionnement contraint par coordonnées avec y =barycentre et $x = [-1|1]$. (a) Les arcs d'un poids inférieur à 50 ne sont pas représentés. (b) Les valeurs sont exprimées en milliards d'euros. Les transferts de cash-flow inférieurs à 100000000 ne sont pas représentés.

6 Conclusions

La littérature scientifique concernant les réseaux de participations a traité essentiellement des ensembles de données limitées à un seul pays et aux sociétés cotées, celles-ci étant les seules pour lesquelles des données sont aisément accessibles. Les travaux de Vitali et al ont pour la première fois étudié un ensemble international de sociétés, néanmoins construit à partir d’une sélection de sociétés transnationales.

Pour ce mémoire, nous disposons d’un jeu de données assez différent puisque, malgré le fait qu’il ne couvre qu’un ensemble limité à la Belgique – enrichi cependant d’une proportion non négligeable de filiales étrangères – il comprend un nombre de sociétés beaucoup plus large que ceux considérés par les études antérieures à Vitali et al. ; parmi celles-ci figure une quantité très importante de sociétés de moindre poids économique, aux profils hétérogènes et différents de celles constituant le corps des travaux antérieurs (on peut en effet supposer que leur actionnariat est moins ouvert et volatile). Il était donc intéressant de vérifier si les observations établies lors de ceux-ci se trouvaient confirmées.

Conformément à ces travaux, il s’avère que le caractère sans échelle affecte les attributs topologiques les plus caractéristiques du réseau (la distribution des composantes connexes par taille, la distribution des degrés) ainsi que ceux des sociétés elles-mêmes (fonds propres, force, âge des sociétés). L’ordre de grandeur des γ observés dans tous les cas ne diffère guère des travaux antérieurs, les résultats obtenus avec notre sous-ensemble des sociétés cotées et connexes s’en montrant plus proche. Le caractère sans échelle s’expliquant par des mécanismes d’attachement sélectif liés à un processus de croissance, il est somme toute assez logique qu’un sous-ensemble de sociétés appelées par nature à une dynamique de connexion plus importante montre une tendance plus marquée en ce sens.

Par la suite, nous avons relevé nombre de parallèles entre les caractéristiques de ce réseau de participations et celles des réseaux trophiques ; tous deux sont des réseaux où les flots transitant par les arcs constituent la nature intrinsèque du système et présentent une structure hiérarchique dont l’existence est évidente mais la lecture difficile. Ces comparaisons enrichissantes nous ont amené à examiner les méthodes utilisées pour dégager la structure des réseaux trophiques et les représenter sous une forme simplifiée. Nous avons retenu la méthode de Cooper et Barahona après l’avoir adaptée de deux façons :

- en intégrant le poids total des chemins considérés pour ainsi mieux individualiser les profils de chaque société ;
- en symétrisant la matrice de flot de façon à obtenir une estimation du niveau

hiérarchique des sommets sur base de l'importance – en poids et en longueur – de leurs chemins entrants et sortants.

La détermination d'un nombre approprié de clusters (et donc de niveaux) ainsi que des paramètres influant sur le résultat du clustering s'est révélée difficile ; un étagement hiérarchique se dégageait effectivement mais l'on ne disposait guère de critères pour évaluer sa pertinence. Le recours à diverses techniques de visualisations s'est montré d'un grand secours. Une PCA a souligné l'influence particulière d'une transposition logarithmique des valeurs de la matrice de flot et la visualisation de celle-ci a permis de comprendre qu'un phénomène de «quasi-binarisation» s'y effectuait.

À ce stade, le doute s'est cependant instillé quant à la pertinence de la méthode. La simplification de la matrice de flot ne constituait-elle pas un appauvrissement des résultats, et le clustering qui en résulte, où la longueur maximale des chemins devient l'élément prépondérant, reflète-t-il bien la hiérarchie du réseau ? C'est encore une fois la littérature relative aux réseaux trophiques qui apporté les éclaircissements nécessaires. Celle-ci témoigne du fait que la séparation de tels réseaux en niveaux trophiques ne fait pas l'unanimité, car ils imposent un degré de simplification auquel il n'est pas possible de faire correspondre la réalité. Les travaux de Luczkovich et Borgatti utilisant l'équivalence régulière pour définir et mesurer les rôles trophiques font le constat que les images du réseau qu'ils obtiennent semblent mieux définir sa structure.

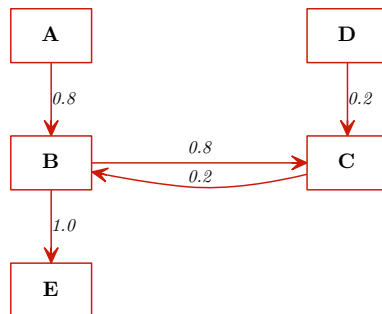
Par conséquent, le niveau hiérarchique obtenu par la méthode adaptée de Cooper et Barahona ne doit pas être considéré de façon stricte, mais comme une information pouvant enrichir une analyse des rôles dans un tel réseau. Il peut être utilisé dans la représentation du graphe image pour y rendre plus lisibles les positions.

Les limites de ce travail tiennent à l'absence de validation objective des résultats du clustering qui est due essentiellement à des questions de calendrier. La littérature relative au blockmodeling présente par exemple des critères de qualité dont nous n'avons pas eu le temps de prendre connaissance. De même, il aurait été intéressant d'examiner les caractéristiques économiques des sociétés regroupées pour identifier des corrélations entre celles-ci et leur position.

Nous envisagions également ce travail avec le présupposé que les particularités d'un réseau de participations rendraient improbable la possibilité d'appliquer à d'autres types de réseaux les enseignements que nous pourrions en tirer. Mais la lecture de l'article de Cooper et Barahona a suggéré qu'ils pouvaient être appliqués à tout réseau caractérisé par un flux et dont le poids des arcs pouvait être exprimé par un ratio de transmission (celui-ci impliquant la multiplication des poids des arcs successifs). L'intégration du poids des arcs à la méthode initiale en constitue alors un affinement et un développement naturel.

7 Annexes

7.1 Exemple de calcul de participation



(a) Graphe A des participations directes

	A	B	C	D	E
A	-	0.80	-	-	-
B	-	-	0.80	-	1.00
C	-	0.20	-	-	-
D	-	-	0.20	-	-
E	-	-	-	-	-
	0.00	1.00	1.00	0.00	1.00

(b) Matrice A de participations directes

	A	B	C	D	E
A	-	0.95	0.76	-	0.95
B	-	0.19	0.95	-	1.19
C	-	0.24	0.19	-	0.24
D	-	0.05	0.24	-	0.05
E	-	-	-	-	-
	0.00	1.43	2.14	0.00	2.43

(c) Matrice Y de participations directes et indirectes

	A	B	C	D	E
A	1.00	-	-	-	-
B	-	-	-	-	-
C	-	-	-	-	-
D	-	-	-	1.00	-
E	-	-	-	-	-
	1.00	0.00	0.00	1.00	0.00

(d) Matrice D des participations hors-système

	A	B	C	D	E
A	-	0.95	0.76	-	0.95
B	-	-	-	-	-
C	-	-	-	-	-
D	-	0.05	0.24	-	0.05
E	-	-	-	-	-
	0.00	1.00	1.00	0.00	1.00

(e) Matrice V de participation intégrée

FIGURE 7.1 – Exemple de calcul de contrôle intégré avec cycle. De telles situations peuvent mener à des anomalies dans le calcul des participations indirectes, comme celle supérieure à 100% de B dans E , qui tient à l'addition de celle qu'il détient lui-même et de celle qu'il détient via son parent C . Elle n'apparaît plus dans le calcul intégré.

7.2 Le phénomène de «petit monde»

Le phénomène de «petit monde» a été caractérisé à partir de l'observation que les réseaux présentent de manière généralement l'une de ces deux propriétés :

- soit ils montrent une forte densité locale de liens, ce qui a pour conséquence que deux voisins d'un sommet présentent une forte probabilité d'être eux-mêmes reliés par un lien ;
- soit ils offrent des chemins courts permettant de joindre deux sommets quelconques en un minimum de *sauts*.

Ces deux propriétés s'avèrent le plus souvent contradictoires. On s'est cependant aperçu qu'il existe des réseaux qui les présentent de manière simultanée. Watts dans [Watts, 1999] a formalisé le phénomène en définissant, pour la première de ces propriétés, le *coefficient de clustering* et, pour la seconde, la *longueur de chemin caractéristique*. Si un sommet v d'un graphe a k voisins immédiats, il peut exister entre eux au maximum $k_v(k_v - 1)/2$ arêtes. Le coefficient local C_v est donc la proportions des arêtes effectivement présentes par rapport à ce maximum. Le coefficient global pour le réseau est alors la moyenne des coefficients locaux. La longueur de chemin caractéristique s'obtient en effectuant la moyenne de tous les chemins reliant les sommets deux à deux :

$$L = \frac{1}{1/2n(n+1)} \sum_{i>j} d_{ij}$$

Sur cette base, Watts a défini un «petit monde» comme un réseau décentralisé de grande taille, de faible densité (le degré maximal étant *beaucoup* plus petit que le nombre de sommet : $1 \ll k_{max} \ll n$), présentant une longueur de chemin caractéristique proche de celle d'un réseau aléatoire équivalent ($L \approx L_{\text{random}}$) en même temps qu'un coefficient de clustering beaucoup plus grand ($C \gg C_{\text{random}}$).

Watts et Strogatz [Watts et Strogatz, 1998] ont créé un modèle théorique de réseau répondant à cette définition de cette manière : au départ d'un réseau régulier où chaque sommet est connecté à un nombre constant de voisins immédiats, il suffit de déconnecter de manière probabiliste un certain nombre de ces liens directs et de les reconnecter au hasard à un autre sommet. En faisant varier la probabilité de reconnexion tel un curseur, on passe ainsi d'un réseau régulier où le coefficient de clustering est maximal mais les chemins très longs à un réseau totalement aléatoire où le coefficient de clustering est quasiment nul mais les chemins caractéristiques très courts.

L'intérêt de ce modèle est de montrer qu'en augmentant le degré de probabilité, on arrive rapidement à diminuer fortement la longueur des chemins caractéristiques sans affecter notablement le degré de cohésion locale. Autrement dit, il suffit d'ajouter une faible part de désordre à un réseau régulier pour en faire un petit monde.

Une autre façon de caractériser un petit monde est celle avancée par Newman dans [Newman, 2003b] : un réseau présente cette propriété si la longueur des chemins carac-

téristiques augmente de façon logarithmique – voire moins – que la taille du réseau pour un degré moyen donné.

7.3 La structure en *bow-tie*

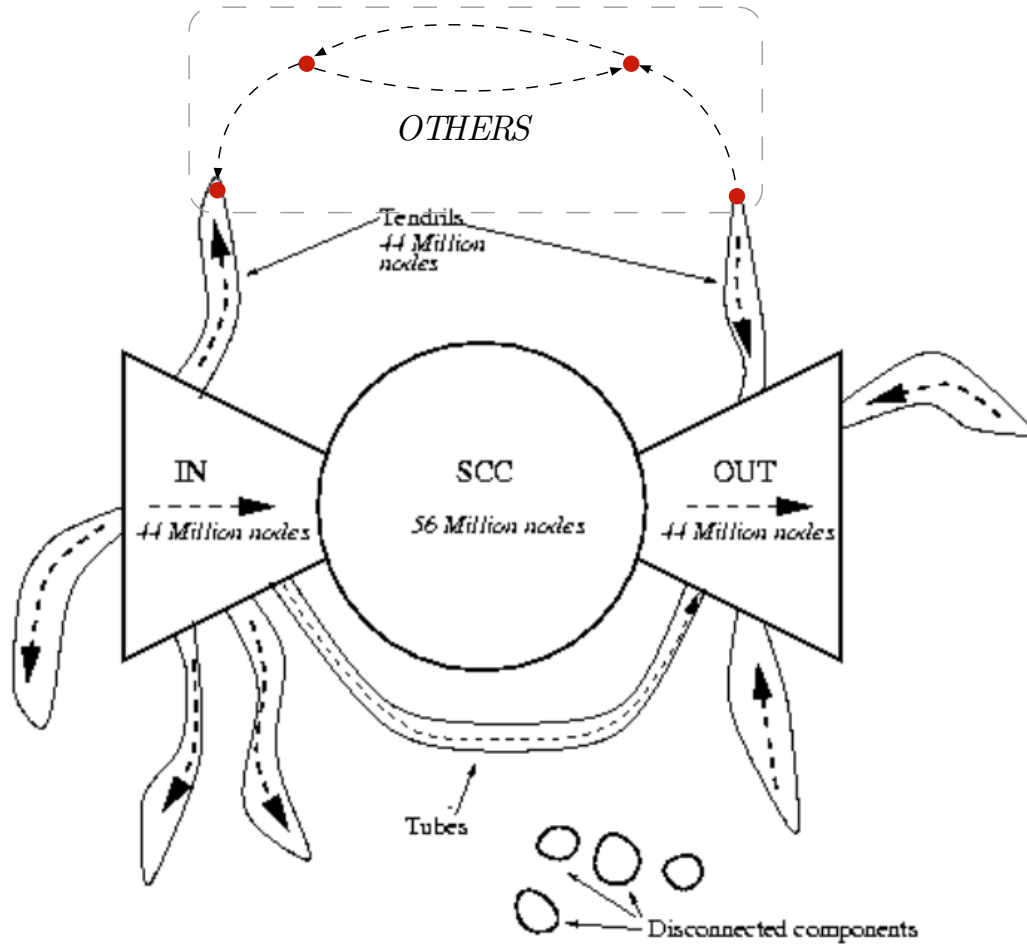


FIGURE 7.2 – Représentation originale de la structure en bow-tie d'internet, par [Broder *et al.*, 2000]. On y a ajouté une composante OTHERS montrant, conformément à Yang et al., qu'il pouvait exister une composante dont les sommets ne sont pas accessibles depuis la structure, tout en n'étant pas déconnectée.

Cette section expose certains des éclaircissement apportés par [Yang *et al.*, 2011] au sujet des structures en *bow-tie*.

Ils effectuent une distinction entre les *tendrils*¹ entrants et sortants, se différenciant en cela des travaux antérieurs. Ils soulignent aussi que la partie *OTHERS* n'est pas nécessairement constituée de composantes totalement déconnectées de la structure ; par exemple, un sommet d'un *out-tendril* peut comporter des chemins sortants menant à des sommets qui n'appartiennent pas à *OUT* (figure 7.2).

Les apports de Yang et al. au sujet de la structure en *bow-tie* sont fondés sur la notion d'accessibilité étendue (*reachability*) entre des sommets $v, w, \dots$ et différents sous-ensembles de sommets $S, T, \dots$, définie ainsi :

- S est accessible depuis v s'il existe $w \in S$ tel que w est accessible depuis v ;
- v est accessible depuis S s'il existe $w \in S$ tel que v est accessible depuis w ;
- S est accessible depuis T s'il existe $v \in T$ et $w \in S$ tel que w est accessible depuis v .

Soit $G = (V, E)$ un graphe dirigé et S un sous-ensemble fortement connexe de G . En adoptant la notation $S \rightsquigarrow v$ pour signifier que v est accessible depuis S , la *décomposition* de G relativement à S consiste dans les ensembles de sommets suivants :

- $SCC = S$;
- $IN = \{v \in (V - S) | v \rightsquigarrow S\}$;
- $OUT = \{v \in V - S | S \rightsquigarrow v\}$
- $TUBES = \{v \in V - S - IN - OUT | IN \rightsquigarrow v \wedge OUT \rightsquigarrow v\}$
- $INTENDRILS = \{v \in V - S | IN \rightsquigarrow v \wedge \neg(v \rightsquigarrow OUT)\}$
- $OUTTENDRILS = \{v \in V - S | \neg(IN \rightsquigarrow v) \wedge v \rightsquigarrow OUT\}$
- $OTHERS = \{S - IN - OUT - TUBES - INTENDRILS - OUTTENDRILS\}$.

Ces définitions sont directement transposables en un processus algorithmique. Le fait d'être atteignable depuis un sommet revient à faire partie des sommets accessibles par un parcours en largeur ; la condition « S est accessible depuis v » équivaut à vérifier que l'intersection de S et des descendants de v n'est pas vide. Ainsi, abstraction faite de toute optimisation, avec la librairie *NetworkX*, l'assertion $(v \rightsquigarrow OUT)$ équivaut à

```
(len(OUT.inter(set(networkx.descendants(G,v)))) > 0)
```

La notion d'accessibilité selon Yang et al., parce qu'elle requiert seulement l'*existence* d'un sommet pour être assurée, ne met cependant pas en évidence la relation particulière entre les sommets des parties *IN* et *OUT*. Du fait que *SCC* est fortement connexe, *tout* sommet de *SCC* est accessible depuis *tout* sommet de *SCC*. Par conséquent,

- *tout* sommet de *OUT* est accessible depuis *tout* sommet de *SCC* ;
- *tout* sommet de *SCC* est accessible depuis *tout* sommet de *IN* ;
- *tout* sommet de *OUT* est accessible depuis *tout* sommet de *IN*.

En terme de participations, cela signifie donc que chaque sommet de *IN* possède une part et dispose d'un degré de contrôle de chacun des sommets de *OUT*.

1. Ce terme désigne en botanique les vrilles au moyen desquelles les plantes grimpantes s'accrochent, ainsi que des objets sans utilité qu'on conserve sans raison.

7.4 Groupes de sociétés, précisions terminologiques

[Granovetter, 2005] définit le **groupe d'entreprises** (*business group*) comme un ensemble de firmes ayant des personnalités juridiques distinctes mais liées d'une manière persistante, formelle ou informelle. Il le distingue du **conglomérat** dans le sens où dans ce dernier, des filiales sont rassemblées sous une tête de groupe sans qu'il n'existe de réels liens opérationnels ou personnels entre elles, ce qui est généralement le cas lorsque l'intérêt financier importe d'avantage à la tête de groupe qu'une stratégie commune à long terme (cela allant de pair avec une faible identité commune). Granovetter reconnaît toutefois que cette distinction est souvent malaisée. Il définit également six *dimensions* variables caractérisant la forme d'un *business group*, parmi lesquelles figure la structure de propriété².

Le groupe d'entreprises ainsi défini constitue un réseau à lui seul ; les réseaux de participations étudiés dans la suite de ce mémoire comportent des sous-graphes qui peuvent éventuellement correspondre à un groupe tel que défini par Granovetter mais inversement, sa définition incorpore des groupes qui ne s'inscrivent pas nécessairement au sein d'un réseau de participations.

Cette définition qui a cours en sociologie souligne la multiplicité des types de liens qui peuvent constituer le réseau dans lequel une entreprise est «encastrée»³.

[Cuervo-Cazurra, 2006] clarifie la notion de groupe d'entreprises et en analyse les différents types ; il en restreint la définition à ceux qui sont placés sous un contrôle et une propriété commune mais insiste sur la diversification des activités exercées (*unrelated diversification*).

L'expression *business network* se rapporte à «une structure de relations entre des acteurs hétérogènes interagissant pour la réalisation de projets communs»[Todeva, 2006]. Dans un tel réseau, les éléments interconnectés sont des acteurs, des ressources et des activités.

2. Les autres étant les sources de solidarité, l'économie morale, la structure d'autorité, le rôle des institutions financières et la relation à l'État.

3. Ce terme – *embeddedness* – est utilisé par convention dans cette discipline pour expliquer comment ces réseaux de nature non-économique affectent notablement les performances d'une entreprise en fournissant mutuellement de l'information ce qui réduit l'incertitude, renforce les contrats et permet d'identifier les opportunités [Granovetter, 2005].

«L'encastrement est un système d'échanges qui offre des opportunités uniques vis-à-vis du marché et les firmes organisées en réseaux ont des chances de survie supérieures» [Uzzi, 1996].

7.5 Cooper et Barahona, visualisations

Cette section présente les variations obtenues dans le clustering du réseau de démonstration de Cooper et Barahona. Contrairement à leur illustration reprise à la figure 5.4 :

- la matrice carrée est une matrice de *distance* où ce sont les valeurs les plus faibles (les plus sombres) qui représentent les sommets les plus similaires ;
- la matrice rectangulaire de flot est symétrisée conformément aux explications données à la section 5.5 ;
- les lignes des deux matrices sont synchronisées et se rapportent à un même sommet.

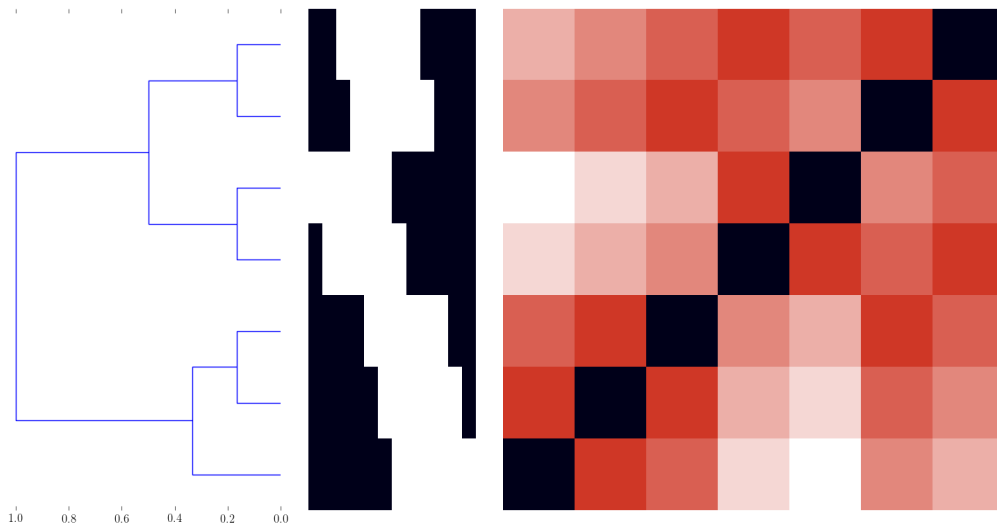


FIGURE 7.3 – $\beta = 1.0$. Dans la matrice de flot, chaque longueur pour laquelle un chemin est présent est rendue de façon égale par une valeur de 1 (en blanc). Sur la matrice des distances, on peut observer qu’ hormis la distance minimale (en noir) de chaque sommet avec lui-même, les distances les plus faibles concernent les sommets voisins dans la chaîne, puisque leurs vecteurs respectifs dans la matrice de flot ont 10 valeurs communes ($5 \times 0, 5 \times 1$).

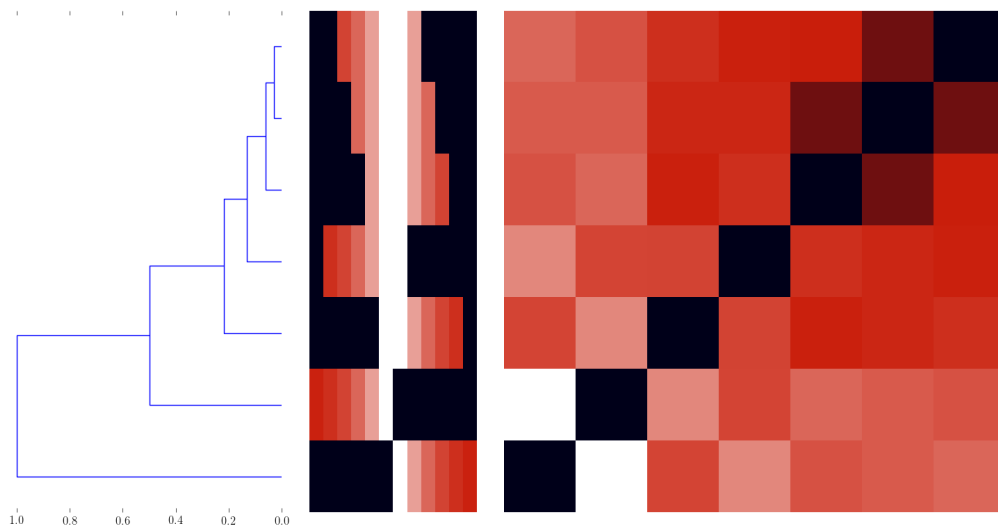


FIGURE 7.4 – $\beta = 0.6$. Dans la matrice de flot, les valeurs décroissent (s’assombrissent) en relation avec la longueur des chemins. Dans la matrice des distances, l’éloignement des sommets extrêmes (rassemblés dans le coin inférieur gauche) est souligné.

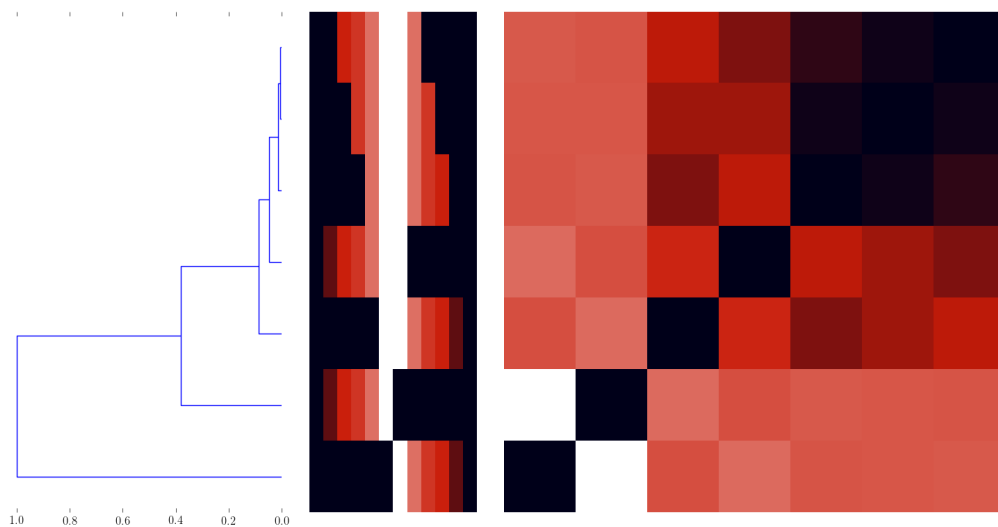


FIGURE 7.5 – $\beta = 0.4$. Dans la matrice de flot, l’assombrissement des valeurs pour les chemins longs s’accroît. Dans la matrice des distances, le cluster rassemblant les sommets internes de la chaîne se dégage par une valeur plus sombre.

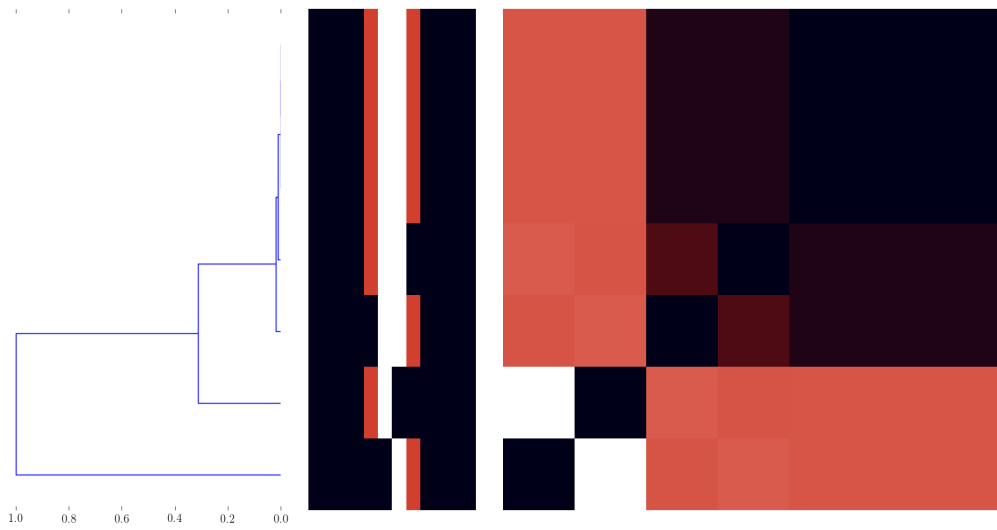


FIGURE 7.6 – $\beta = 0.2$. L'information se restreint de plus en plus au voisinage local.

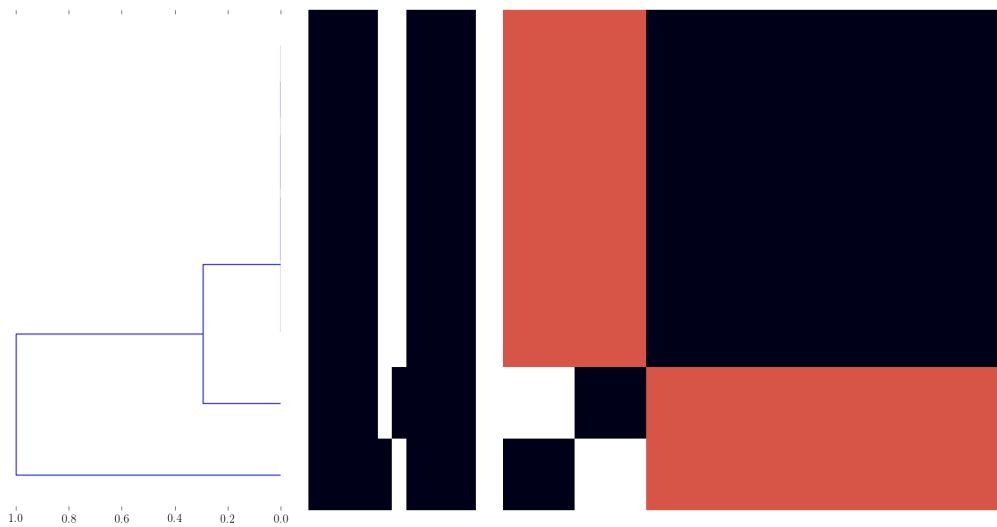


FIGURE 7.7 – $\beta = 0.05$. Seule l'information locale subsiste, la matrice de flot reflète seulement les degrés entrants et sortants. Tous les nœuds intermédiaires présentent donc un profil semblable.

Bibliographie

- [Adams et Ferreira, 2008] ADAMS, R. et FERREIRA, D. (2008). One share-one vote : The empirical evidence. *Review of Finance*, 12(1):51–91.
- [Aggarwal et Wang, 2010] AGGARWAL, C. et WANG, H. (2010). A survey of clustering algorithms for graph data. In AGGARWAL, C. C. et WANG, H., éditeurs : *Managing and Mining Graph Data*, volume 40 de *Advances in Database Systems*, chapitre 9, pages 275–301. Springer US.
- [Albert et Barabási, 2002] ALBERT, R. et BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47.
- [Alhajj et Rokne, 2014] ALHAJJ, R. et ROKNE, J. (2014). *Encyclopedia of Social Network Analysis and Mining*. Encyclopedia of Social Network Analysis and Mining. Springer New York.
- [Antoine et Cornil, 2002] ANTOINE, J. et CORNIL, J. (2002). *Lexique thématique de la comptabilité : dictionnaire spécialisé explicatif*. Comptabilité, contrôle et finance. De Boeck.
- [Aoyama et al., 2010] AOYAMA, H., FUJIWARA, Y., IKEDA, Y., IYETOMO, H. et SOUMA, W. (2010). Complex business networks. In *Econophysics and Companies : Statistical Life and Death in Complex Business Networks*, chapitre 4, pages 99–151. Cambridge University Press.
- [Barabási et Albert, 1999] BARABÁSI, A.-L. et ALBERT, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- [Barker, 2004] BARKER, R. (2004). Shareholders, blockholders and stackholders. In *The Handbook of International Corporate Governance : A Definitive Guide*, An Institute of Directors publication, chapitre 2, pages 42–47. IOD.
- [Barrat, 2013] BARRAT, A. (2013). La notion de réseau complexe : du réseau comme abstraction et outil à la masse de données des réseaux sociaux en ligne. *Communication & Organisation*, 43(1):15–24.
- [Battiston et al., 2010] BATTISTON, S., GLATTFELDER, J. B., GARLASCHELLI, D., LILLO, F. et CALDARELLI, G. (2010). The structure of financial networks. In ESTRADA, E., FOX, M., HIGHAM, D. J. et OPPO, G.-L., éditeurs : *Network Science, Complexity in Nature and Technology*, chapitre 7, pages 131–163. Springer London.
- [Beauguitte, 2011] BEAUGUITTE, L. (2011). Cliques, communautés et dérivées. working paper or preprint.

- [Bebchuk *et al.*, 2000] BEBCHUK, L. A., KRAAKMAN, R. et TRIANTIS, G. (2000). Stock pyramids, cross-ownership, and dual class equity : the mechanisms and agency costs of separating control from cash-flow rights. *In Concentrated corporate ownership*, pages 295–318. University of Chicago Press.
- [Becht, 1997] BECHT, M. (1997). Strong blockholders, weak owners and the need for european mandatory disclosure. *In The separation of ownership and control : A survey of 7 European countries*, volume 1. European Corporate Governance Network.
- [Becht et Chapelle, 1997] BECHT, M. et CHAPELLE, A. (1997). Ownership and control in belgium. *In The separation of ownership and control : A survey of 7 European countries*, volume 2. European Corporate Governance Network.
- [Becht *et al.*, 1999] BECHT, M., CHAPELLE, A. et RENNEBOOG, L. (1999). Shareholding cascades : the separation of ownership and control in belgium.
- [Benoit-Moury *et al.*, 2013] BENOIT-MOURY, A., CAPRASSE, O. et TILLEMANN, B. (2013). *Droit des sociétés : code des sociétés annoté avec législation spéciale y compris les associations et les fondations*. La Charte.
- [Berkowitz *et al.*, 1979] BERKOWITZ, S. D., CARRINGTON, P. J., KOTOWITZ, Y. et WEVERMAN, L. (1979). The determination of enterprise groupings through combined ownership and directorship ties. *Social Networks*, 1.
- [Berle et Means, 1932] BERLE, A. et MEANS, G. (1932). *The Modern Corporation and Private Property*. Transaction Publishers.
- [Biebuyck *et al.*, 2005] BIEBUYCK, T., CHAPELLE, A. et SZAFARZ, A. (2005). Les leviers de contrôle des actionnaires majoritaires. *Revue gouvernance*, 1:52–70.
- [Bishop, 2006] BISHOP, C. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.
- [Blöchl *et al.*, 2011] BLÖCHL, F., THEIS, F. J., VEGA-REDONDO, F. et FISHER, E. O. (2011). Vertex centralities in input-output networks reveal the structure of modern economies. *Physical Review E*, 83(4):046127.
- [Boccaletti *et al.*, 2006] BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M. et HWANG, D.-U. (2006). Complex networks : Structure and dynamics. *Physics reports*, 424(4):175–308.
- [Bonanno *et al.*, 2004] BONANNO, G., CALDARELLI, G., LILLO, F., MICCICHÈ, S., VANDEWALLE, N. et MANTEGNA, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):363–371.
- [Borgatti et Everett, 1992] BORGATTI, S. P. et EVERETT, M. G. (1992). Notions of position in social network analysis. *Sociological methodology*, 22(1):1–35.
- [Borgatti et Everett, 1993] BORGATTI, S. P. et EVERETT, M. G. (1993). Two algorithms for computing regular equivalence. *Social Networks*, 15(4):361–376.
- [Boschetti *et al.*, 2005] BOSCHETTI, F., PROKOPENKO, M., MACREADIE, I. et GRISOGONO, A.-M. (2005). Defining and detecting emergence in complex networks. *In*

- KHOSLA, R., HOWLETT, R. et JAIN, L., éditeurs : *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3684 de *Lecture Notes in Computer Science*, pages 573–580. Springer Berlin Heidelberg.
- [Brandes *et al.*, 2002] BRANDES, U., EIGLSPERGER, M., HERMAN, I., HIMSOLT, M. et MARSHALL, M. S. (2002). Graphml progress report structural layer proposal. In MUTZEL, P., JÜNGER, M. et LEIPERT, S., éditeurs : *Graph Drawing : 9th International Symposium, GD 2001 Vienna, Austria, September 23–26, 2001 Revised Papers*, pages 501–512. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Broder *et al.*, 2000] BRODER, A., KUMAR, R., MAGHOUL, F., RAGHAVAN, P., RAJAGOPALAN, S., STATA, R., TOMKINS, A. et WIENER, J. (2000). Graph structure in the web. *Computer networks*, 33(1):309–320.
- [Browet, 2014] BROWET, A. (2014). *Algorithms for community and role detection in networks*. Thèse de doctorat, UCL.
- [Butts, 2008] BUTTS, C. T. (2008). Social network analysis : A methodological introduction. *Asian Journal of Social Psychology*, 11(1):13–41.
- [Caldarelli, 2007] CALDARELLI, G. (2007). Financial networks. In *Scale-Free Networks : Complex Webs in Nature and Technology*, OUP Catalogue, chapitre 11, pages 231–250. Oxford University Press.
- [Caldarelli *et al.*, 2004] CALDARELLI, G., BATTISTON, S., GARLASCHELLI, D. et CATANZARO, M. (2004). Emergence of complexity in financial networks. In BEN-NAIM, E., FRAUENFELDER, H. et TOROCZKAI, Z., éditeurs : *Complex Networks*, volume 650 de *Lecture Notes in Physics*, pages 399–423. Springer Berlin Heidelberg.
- [Carscallen *et al.*, 2012] CARSCALLEN, W. M. A., VANDENBERG, K., LAWSON, J. M., MARTINEZ, N. D. et ROMANUK, T. N. (2012). Estimating trophic position in marine and estuarine food webs. *Ecosphere*, 3(3):1–20.
- [Cerrada, 2006] CERRADA, K. (2006). *Comptabilité et analyse des états financiers : principes et applications*. Comptabilité, contrôle & finances. De Boeck.
- [Chakrabarti et Faloutsos, 2006] CHAKRABARTI, D. et FALOUTSOS, C. (2006). Graph mining : Laws, generators, and algorithms. *ACM Computing Surveys (CSUR)*, 38(1):2.
- [Chapelle et Szafarz, 2005] CHAPELLE, A. et SZAFARZ, A. (2005). Controlling firms through the majority voting rule. *Physica A : Statistical Mechanics and its Applications*, 355(2):509–529.
- [Chapelle et Szafarz, 2007] CHAPELLE, A. et SZAFARZ, A. (2007). Control consolidation with a threshold : an algorithm. *IMA Journal of Management Mathematics*, 18(3):235–243.
- [Chapelle *et al.*, 2002] CHAPELLE, A., SZAFARZ, A. *et al.* (2002). Ownership and control : dissecting the pyramid. Rapport technique, ULB–Université Libre de Bruxelles.
- [Chiang et Mirkin, 2010] CHIANG, M. M.-T. et MIRKIN, B. (2010). Intelligent choice of the number of clusters in k-means clustering : an experimental study with different cluster spreads. *Journal of classification*, 27(1):3–40.

- [Claessens *et al.*, 2000] CLAESSENS, S., DJANKOV, S. et LANG, L. H. (2000). The separation of ownership and control in east asian corporations. *Journal of Financial Economics*, 58(1–2):81 – 112. Special Issue on International Corporate Governance.
- [Cohen *et al.*, 1990] COHEN, J., PALKA, Z., BRIAND, F. et NEWMAN, C. (1990). *Community Food Webs : Data and Theory*. Biomathematics. Springer Berlin Heidelberg.
- [Cooper et Barahona, 2010] COOPER, K. et BARAHONA, M. (2010). Role-based similarity in directed networks. *arXiv preprint arXiv :1012.2726*.
- [Costa *et al.*, 2007] COSTA, L. d. F., RODRIGUES, F. A., TRAVIESO, G. et VILLAS BOAS, P. R. (2007). Characterization of complex networks : A survey of measurements. *Advances in Physics*, 56(1):167–242.
- [Cousins, 1985] COUSINS, S. (1985). Ecologists build pyramids again. *New Scientist*, 106(1463):50–54.
- [Cuervo-Cazurra, 2006] CUERVO-CAZURRA, A. (2006). Business groups and their types. *Asia Pacific Journal of Management*, 23(4):419–437.
- [De Benedictis et Tajoli, 2011] DE BENEDICTIS, L. et TAJOLI, L. (2011). The world trade network. *The World Economy*, 34(8):1417–1454.
- [Demsetz, 1983] DEMSETZ, H. (1983). The structure of ownership and the theory of the firm. *JL & Econ.*, 26:375.
- [D’Errico *et al.*, 2009] D’ERRICO, M., GRASSI, R., STEFANI, S. et TORRIERO, A. (2009). Shareholding networks and centrality : An application to the italian financial market. In NAIMZADA, A., STEFANI, S. et TORRIERO, A., éditeurs : *Networks, Topology and Dynamics*, volume 613 de *Lecture Notes in Economics and Mathematical Systems*, pages 215–228. Springer Berlin Heidelberg.
- [Doreian *et al.*, 2005] DOREIAN, P., BATAGELJ, V. et FERLIGOJ, A. (2005). *Generalized blockmodeling*, volume 25. Cambridge university press.
- [Easley et Kleinberg, 2010] EASLEY, D. et KLEINBERG, J. (2010). *Networks, Crowds, and Markets : Reasoning About a Highly Connected World*. Cambridge University Press.
- [Edwards et Weichenrieder, 2009] EDWARDS, J. S. et WEICHENRIEDER, A. J. (2009). Control rights, pyramids, and the measurement of ownership concentration. *Journal of Economic Behavior & Organization*, 72(1):489 – 508.
- [Eurostat, 2010] EUROSTAT (2010). Business registers. recommendations manual. Rapport technique, European Commision.
- [Faccio et Lang, 2002] FACCIO, M. et LANG, L. H. (2002). The ultimate ownership of western european corporations. *Journal of financial economics*, 65(3):365–395.
- [Ferraro *et al.*, 2012] FERRARO, F., CORRADO, R., SCHNYDER, G., HEEMSKERK, E. et DEL VECCHIO, N. (2012). Structural breaks and governance networks in western europe. In KOGUT, B., éditeur : *The Small Worlds of Corporate Governance*, chapitre 5, pages 151–182. MIT Press.
- [Fortunato, 2010] FORTUNATO, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5):75–174.

- [Fujiwara *et al.*, 2006] FUJIWARA, Y., AOYAMA, H. et SOUMA, W. (2006). Growth of firms and networks. In CHAKRABARTI, B., CHAKRABORTI, A. et CHATTERJEE, A., éditeurs : *Econophysics and Sociophysics : Trends and Perspectives*, chapitre 4, pages 99–129. John Wiley & Sons.
- [Gaedke, 2008] GAEDKE, U. (2008). Ecological networks. In BJÖRN, J. H. et FALK, S., éditeurs : *Analysis of Biological Networks*, Wiley Series in Bioinformatics, pages 283–304. John Wiley & Sons, Inc.
- [Gaetler, 2005] GAETLER, M. (2005). Clustering. In *Network Analysis : Methodological Foundations (Lecture Notes in Computer Science)*, chapitre 8, pages 178–215. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Gansner et North, 2000] GANSNER, E. R. et NORTH, S. C. (2000). An open graph visualization system and its applications to software engineering. *SOFTWARE - PRACTICE AND EXPERIENCE*, 30(11):1203–1233.
- [Garlaschelli *et al.*, 2005] GARLASCHELLI, D., BATTISTON, S., CASTRI, M., SERVEDIO, V. D. et CALDARELLI, G. (2005). The scale-free topology of market investments. *Physica A : Statistical Mechanics and its Applications*, 350(2–4):491 – 499.
- [Garlaschelli et Loffredo, 2004] GARLASCHELLI, D. et LOFFREDO, M. I. (2004). Fitness-dependent topological properties of the world trade web. *Physical review letters*, 93(18):188701.
- [Garlaschelli et Loffredo, 2005] GARLASCHELLI, D. et LOFFREDO, M. I. (2005). Structure and evolution of the world trade network. *Physica A : Statistical Mechanics and its Applications*, 355(1):138–144.
- [Girvan et Newman, 2002] GIRVAN, M. et NEWMAN, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.
- [GlabaiX, 2008] GLABAIX, X. (2008). Power laws. In *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK.
- [Glattfelder, 2013] GLATTFELDER, J. B. (2013). *Decoding Complexity*. Springer Theses. Springer Berlin Heidelberg.
- [Glattfelder et Battiston, 2009] GLATTFELDER, J. B. et BATTISTON, S. (2009). Backbone of complex networks of corporations : The flow of control. *Physical Review E*, 80(3):036104.
- [Granovetter, 2005] GRANOVETTER, M. (2005). Business groups and social organization. In *The Handbook of Economic Sociology*, chapitre 19, pages 429–450. Princeton University Press.
- [Hannan et Freeman, 1993] HANNAN, M. et FREEMAN, J. (1993). *Organizational Ecology*. Harvard University Press.
- [Hanneman et Riddle, 2005a] HANNEMAN, R. et RIDDLE, M. (2005a). Centrality and power. In *Introduction to Social Network Methods*. University of California.
- [Hanneman et Riddle, 2005b] HANNEMAN, R. et RIDDLE, M. (2005b). *Introduction to Social Network Methods*. University of California.

- [Holderness, 2003] HOLDERNESS, C. G. (2003). A survey of blockholders and corporate control. *Economic policy review*, 9(1).
- [Holderness, 2009] HOLDERNESS, C. G. (2009). The myth of diffuse ownership in the united states. *Review of Financial studies*, 22(4):1377–1408.
- [Hüser, 2015] HÜSER, A.-C. (2015). Too interconnected to fail : A survey of the inter-bank networks literature. Safe working paper, Research Center SAFE - Sustainable Architecture for Finance in Europe, Goethe University Frankfurt.
- [ISSE *et al.*, 2007] ISSE, I. S. S. E., STERLING, S. . et ECGI (2007). Proportionality between ownership and control in eu listed companies. Rapport technique MSU-CSE-05-29, European Commission.
- [Johnsen, 2014] JOHNSEN, G. (2014). *Bringing Down the Banking System : Lessons from Iceland*. EBL-Schweitzer. Palgrave Macmillan US.
- [Johnson *et al.*, 2001] JOHNSON, J. C., BORGATTI, S. P., LUCZKOVICH, J. J. et EVERETT, M. G. (2001). Network role analysis in the study of food webs : an application of regular role coloration. *Journal of Social Structure*, 2(3).
- [Jorion, 2016] JORION, P. (2016). *Le dernier qui s'en va éteint la lumière : Essai sur l'extinction de l'humanité*. Documents. Fayard.
- [Knoke et Yang, 2008] KNOKE, D. et YANG, S. (2008). Basic methods for analyzing networks. In *Social Network Analysis*, numéro 154 de Quantitative Applications in the Social Sciences, chapitre 4. SAGE Publications.
- [Kogut et Walker, 2001] KOGUT, B. et WALKER, G. (2001). The small world of germany and the durability of national networks. *American sociological review*, pages 317–335.
- [Kunegis, 2013] KUNEGIS, J. (2013). Konect : The koblenz network collection. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13 Companion*, pages 1343–1350, New York, NY, USA. ACM.
- [La Porta *et al.*, 1999] LA PORTA, R., LOPEZ-DE-SILANES, F. et SHLEIFER, A. (1999). Corporate ownership around the world. *The Journal of Finance*, 54(2):471–518.
- [Leuz *et al.*, 2003] LEUZ, C., NANDA, D. et WYSOCKI, P. D. (2003). Earnings management and investor protection : an international comparison. *Journal of financial economics*, 69(3):505–527.
- [Lorrain et White, 1971] LORRAIN, F. et WHITE, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80.
- [Mandere, 2009] MANDERE, E. O. (2009). *Financial Networks and Their Applications to the Stock Market*. Thèse de doctorat, Bowling Green State University.
- [Minoiu et Reyes, 2013] MINOIU, C. et REYES, J. A. (2013). A network analysis of global banking : 1978–2010. *Journal of Financial Stability*, 9(2):168–184.
- [Mizruchi, 1996] MIZRUCHI, M. S. (1996). What do interlocks do ? an analysis, critique, and assessment of research on interlocking directorates. *Annual review of sociology*, pages 271–298.

- [Newman, 2003a] NEWMAN, M. E. (2003a). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.
- [Newman, 2003b] NEWMAN, M. E. (2003b). The structure and function of complex networks. *SIAM review*, 45(2):167–256.
- [Newman, 2010] NEWMAN, M. E. (2010). *Networks : An Introduction*. Oxford University Press, Inc., New York, NY, USA.
- [Parthasarathy et Faisal, 2013] PARTHASARATHY, S. et FAISAL, S. (2013). Network clustering. In *Data Clustering : Algorithms and Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, chapitre 17, pages 415–445. CRC Press.
- [Pastor-Satorras et al., 2001] PASTOR-SATORRAS, R., VÁZQUEZ, A. et VESPIGNANI, A. (2001). Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87: 258701.
- [Piccardi et al., 2010] PICCARDI, C., CALATRONI, L. et BERTONI, F. (2010). Communities in italian corporate networks. *Physica A : Statistical Mechanics and its Applications*, 389(22):5247–5258.
- [Pimm, 1982] PIMM, S. (1982). *Food Webs*. Population and Community Biology. Springer Netherlands.
- [Porter et al., 2009] PORTER, M. A., ONNELA, J.-P. et MUCHA, P. J. (2009). Communities in networks. *Notices of the AMS*, 56(9):1082–1097.
- [Rousseeuw, 1987] ROUSSEEUW, P. J. (1987). Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.
- [Schaeffer, 2007] SCHAEFFER, S. E. (2007). Graph clustering. *Computer Science Review*, 1(1):27–64.
- [Scott, 2000] SCOTT, J. (2000). *Social Network Analysis : A Handbook*. SAGE Publications.
- [Scott et al., 1992] SCOTT, J. et al. (1992). Corporate interlocks and social network analysis. Appears in Collections : Social Sciences Research Centre : Occasional Papers.
- [Sjögren et al., 2012] SJÖGREN, S., OVERLAND, C. et MAVRUK, T. (2012). Keeping it real or keeping it simple? ownership concentration measures compared. *The 4th workshop of the Nordic Corporate Governance Network, Reykjavik University June 15, 2012*.
- [Skiena, 1990] SKIENA, S. (1990). *Implementing discrete mathematics : combinatorics and graph theory with Mathematica*. Advanced book program. Addison-Wesley.
- [Souma, 2007] SOUMA, W. (2007). Networks of firms and the ridge in the production space. In CHATTERJEE, A. et CHAKRABARTI, B. K., éditeurs : *Econophysics of Market and Business Networks*, pages 149–158. Springer Milan.
- [Souma et al., 2004a] SOUMA, W., FUJIWARA, Y. et AOYAMA, H. (2004a). Random matrix approach to shareholding networks. *Physica A : Statistical Mechanics and its Applications*, 344(1–2):73 – 76. Applications of Physics in Financial Analysis 4 (APFA4).

- [Souma *et al.*, 2004b] SOUMA, W., FUJIWARA, Y. et AOYAMA, H. (2004b). Visualization of business networks. In *The Application of Econophysics*, pages 293–298. Springer.
- [Souma *et al.*, 2005] SOUMA, W., FUJIWARA, Y. et AOYAMA, H. (2005). Shareholding networks in japan. *arXiv preprint physics/0503177*.
- [Souma *et al.*, 2006a] SOUMA, W., FUJIWARA, Y. et AOYAMA, H. (2006a). Change of ownership networks in japan. In *Practical Fruits of Econophysics*, pages 307–311. Springer.
- [Souma *et al.*, 2006b] SOUMA, W., FUJIWARA, Y. et AOYAMA, H. (2006b). Heterogeneous economic networks. In NAMATAME, A., KAIZOUJI, T. et ARUKA, Y., éditeurs : *The Complex Networks of Economic Interactions*, Lecture Notes in Economics and Mathematical Systems, pages 79–92. Springer.
- [Takemoto et Oosawa, 2012] TAKEMOTO, K. et OOSAWA, C. (2012). Introduction to complex networks : Measures, statistical properties and models. In *Statistical and Machine Learning Approaches for Network Analysis*, pages 45–76. John Wiley & Sons, Inc.
- [Todeva, 2006] TODEVA, E. (2006). Introduction. In *Business networks : strategy and structure*, chapitre 1, pages 1–17. Routledge.
- [Ulanowicz, 1995] ULANOWICZ, R. (1995). Ecosystem trophic foundations : Lindeman exonerata. *Complex Ecology : the part-whole relation in ecosystems*. Prentice Hall, Englewood Cliffs, NJ, pages 549–550.
- [Uzzi, 1996] UZZI, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations : The network effect. *American sociological review*, pages 674–698.
- [Vitali *et al.*, 2011] VITALI, S., GLATTFELDER, J. B. et BATTISTON, S. (2011). The network of global corporate control. *PloS one*, 6(10):e25995.
- [Wasserman et Faust, 1994] WASSERMAN, S. et FAUST, K. (1994). *Social Network Analysis : Methods and Applications*. Cambridge University Press.
- [Watts, 1999] WATTS, D. J. (1999). Networks, dynamics, and the small-world phenomenon 1. *American Journal of sociology*, 105(2):493–527.
- [Watts et Strogatz, 1998] WATTS, D. J. et STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442.
- [Webb, 2003] WEBB, A. (2003). *Statistical Pattern Recognition*. Wiley InterScience electronic collection. Wiley.
- [Wims *et al.*, 2011] WIMS, G., MARTENS, D. et de BACKER, M. (2011). Network models of financial contagion : A definition and literature review. Working papers of faculty of economics and business administration, ghent university, belgium, Ghent University, Faculty of Economics and Business Administration.
- [Yang *et al.*, 2011] YANG, R., ZHUHADAR, L. et NASRAOUI, O. (2011). Bow-tie decomposition in directed graphs. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–5.

[Zafarani *et al.*, 2014] ZAFARANI, R., ABBASI, M. A. et LIU, H. (2014). Network measures. *In Social Media Mining : An Introduction*, chapitre 3, pages 73–104. Cambridge University Press, New York, NY, USA.