

**Revue de littérature des méthodes de repérage des filières industrielles basées  
sur les tableaux d'entrées-sorties**

---



## Identification des grappes : deux types d'approches

---

Il existe deux grands types d'approches pour repérer les grappes industrielles : l'approche qualitative et l'approche quantitative. L'approche qualitative repose principalement sur l'opinion d'experts et s'intéresse généralement à l'ensemble des interactions liant des organisations situées à proximité. Pour sa part, l'approche quantitative a recours à d'importants ensembles de données statistiques (tels que les tableaux d'entrées-sorties) pour identifier les secteurs industriels fortement liés par des échanges commerciaux, généralement à travers un territoire relativement vaste (une province, un État ou un pays). Nombre d'études combinent jusqu'à un certain point les deux approches, mettant l'emphase sur un angle ou l'autre en fonction des buts visés.

### L'approche qualitative

Comme nous l'avons mentionné en introduction, les travaux de Porter ont suscité un regain d'intérêt pour le concept des grappes industrielles au début des années 1990. Selon Porter (1990), l'avantage compétitif d'une nation repose sur quatre ensembles de conditions (le fameux « diamant ») :

**Les facteurs de production** tels que la main-d'œuvre et l'infrastructure de transport, vus dans une optique *dynamique* : c'est la capacité d'une nation de *produire* une main-d'œuvre spécialisée, une infrastructure de transport de pointe, etc., qui importe.

**La demande domestique**, ou plus précisément son niveau de sophistication, qui force les compagnies à innover et qui leur permet d'anticiper les besoins futurs à l'échelle mondiale.

**La structure de l'industrie et la compétition domestique**, qui font entre autres référence aux modes de gestion des entreprises. Les nations excellent généralement dans des industries pour lesquelles les styles de gestion « domestiques » sont particulièrement adéquats (ex. : le système de gestion très discipliné de l'Allemagne donne un avantage à celle-ci dans les industries de l'optique et de la chimie, pour lesquelles la précision, la qualité du produit fini et le service après-vente comptent beaucoup).

**Les industries de soutien**, c'est-à-dire la présence locale de fournisseurs ou d'industries complémentaires compétitives à l'échelle mondiale. La proximité géographique peut ici jouer un rôle important selon Porter :

« Suppliers and end-users located near each other can take advantage of short lines of communication, quick and constant flow of information, and an ongoing exchange of ideas and innovations. Companies have the opportunity to influence their suppliers' technical efforts and can serve as test sites for R&D work, accelerating the pace of innovation. »<sup>4</sup>

---

<sup>4</sup> Porter (2000), p. 80.

Plusieurs pays ont pris le modèle de Porter comme cadre de référence pour analyser la compétitivité de leurs systèmes de production locaux et repérer des groupes d'entreprises fortement réseautées entre elles. Ces études sont habituellement de type « micro » : elles s'intéressent aux relations liant un groupe restreint d'entreprises (ou de secteurs industriels) au sein d'un petit territoire. La majorité sont de nature qualitative, c'est-à-dire qu'elles reposent principalement sur les connaissances d'experts régionaux; l'information est colligée par le biais d'entrevues et de groupes de discussion. Ces experts - des agents de développement économique du secteur public, des membres des chambres de commerce et des chefs d'entreprises, par exemple - possèdent une bonne connaissance des forces et des faiblesses de l'économie de leur région, de même que de ses principaux acteurs.

Les **avantages** de cette méthode sont les suivants :

- Les coûts sont relativement peu élevés;
- Les grappes identifiées peuvent contenir des institutions de soutien telles que des chambres de commerce, des centres de transfert technologique et des universités en plus des entreprises privées, et il ne faut donc pas se limiter aux seuls échanges de biens et services entre clients et fournisseurs;
- Le processus d'identification peut amener plusieurs types d'acteurs (entreprises privées, centres de recherche publics, associations diverses, etc.) à se rencontrer.

Les **inconvénients** sont toutefois importants et viennent du fait que la méthode inclut une bonne part d'arbitraire :

- Elle ignore les nombreux outils analytiques disponibles pour repérer les regroupements industriels à l'aide de *critères objectifs*;
- Quelques « voix fortes » et influentes de la région peuvent dominer le processus et ainsi biaiser l'identification des grappes en fonction de leurs intérêts;
- Les résultats ne sont pas généralisables et il est donc difficile de comparer les grappes entre elles.

Soulignons que ces études font souvent usage des *coefficients de localisation* afin d'identifier dans un premier temps les industries de la région qui sont les plus susceptibles d'exporter, se basant sur le postulat que ces dernières constituent les moteurs de la croissance économique de la région. Le calcul du coefficient de localisation d'une industrie  $i$  ( $CL_i$ ) se fait à l'aide de données sur l'emploi :

$$CL_i = \frac{e_i / \sum_i e_i}{E_i / \sum_i E_i}$$

où  $e_i$  est l'emploi dans le secteur  $i$  de la région et  $E_i$  est l'emploi dans le secteur  $i$  pour l'ensemble du pays ou de l'unité géographique qui sert de référence.

Une industrie, affichant un coefficient de localisation supérieur à l'unité, est davantage concentrée pour ce qui est des emplois dans la région que dans l'ensemble du pays et serait donc susceptible de desservir des marchés extérieurs. Souvent, seules ces industries sont retenues aux fins de l'identification des grappes; elles sont regroupées en fonction de relations économiques probables, telles que perçues par les experts régionaux (Anderson, 1994). Les coefficients de localisation reposent toutefois implicitement sur l'hypothèse que la productivité d'une industrie est la même au niveau régional qu'au niveau national, et que les profils de consommation sont également homogènes d'une région à l'autre. De plus, le calcul des coefficients ne tient pas compte du fait que la concentration d'emplois dans une industrie en région peut être liée à l'existence d'une seule grande entreprise plutôt qu'à un regroupement d'entreprises.

## **L'approche quantitative**

L'approche quantitative pour repérer les grappes industrielles fait référence à l'exploitation d'importantes bases de données contenant des statistiques sur les échanges entre secteurs industriels. L'analyse se fait généralement à un niveau davantage « macro » que l'analyse qualitative, puisqu'elle porte sur les *secteurs industriels* plutôt que sur les *entreprises* et couvre habituellement un territoire plus vaste. Les tableaux d'entrées-sorties constituent la base de données la plus fréquemment utilisée. Seules quelques études ont à ce jour exploité une autre base de données, soit des matrices d'innovation reflétant entre autres l'intensité en recherche et développement des échanges interindustriels (voir Baldwin et Peters, 2001 et OCDE, 2001).

Comme on le verra au chapitre suivant, les tableaux d'entrées-sorties contiennent de l'information détaillée sur la valeur des échanges en biens et services des secteurs d'une économie. L'exploitation de cette information permet de repérer les groupes de secteurs qui sont directement liés dans une chaîne de production ou encore qui ont des liens complémentaires du fait qu'ils prennent leurs intrants des mêmes secteurs industriels ou vendent leur production à des secteurs similaires. Divers types d'analyse statistique peuvent être utilisés à ces fins, tels que l'analyse factorielle et l'analyse de classification.

Les regroupements de secteurs identifiés par une analyse statistique des tableaux d'entrées-sorties constituent des *filiales*, soit les noyaux des grappes industrielles. En effet, les tableaux d'entrées-sorties ne permettent pas d'étudier les liens que les secteurs industriels entretiennent avec les institutions de soutien telles que des chambres de commerce et des centres de recherche. Ces liens doivent être établis a posteriori pour compléter les grappes, à l'aide d'une information davantage qualitative.

Soulignons que les filières industrielles sont établies à l'échelle du territoire couvert par les tableaux d'entrées-sorties utilisés. On peut étudier leur distribution spatiale et notamment vérifier si elles sont davantage concentrées sur une partie du territoire que sur une autre en examinant la répartition géographique des établissements ou des emplois qui y sont rattachés.

Résumons les **avantages** de la méthode quantitative :

- Se basant sur des critères objectifs et mesurables, elle est neutre et permet la comparaison des filières entre elles;
- Elle a recours à l'information la plus détaillée possible sur l'ensemble des échanges commerciaux entre les secteurs industriels d'une économie (soit les tableaux d'entrées-sorties).

Ses principaux **inconvénients** sont les suivants :

- La méthode ne permet pas d'étudier les liens avec les institutions de soutien; ceux-ci doivent être établis a posteriori;
- Les tableaux d'entrées-sorties étant le fruit de la compilation de données provenant de multiples sources (enquêtes statistiques et données administratives), ils sont souvent datés : la version la plus récente au moment d'effectuer une analyse remonte souvent à quelques années. À titre d'exemple, pour le Québec, ce sont les données de 2000 qui sont les plus récentes au printemps 2004.

## Le contenu des tableaux d'entrées-sorties

### Vue d'ensemble

Les tableaux d'entrées-sorties constituent une des composantes du Système de comptabilité nationale (SCN). Dans les faits, on retrouve généralement trois tableaux principaux avec les entrées-sorties. Le tableau des entrées, également appelé matrice de l'utilisation, présente la valeur des biens et des services qui sont utilisés au cours d'une année par les diverses branches d'activité de l'économie en tant que facteurs de production. Le tableau des sorties, ou matrice de production, contient quant à lui une estimation de la valeur des biens et des services produits par chaque branche d'activité. Enfin, le tableau de la demande finale présente les dépenses en biens et services qui sont effectuées par les utilisateurs *finaux*. Les dépenses des ménages, par exemple, apparaissent dans le tableau de la demande finale. Les importations et les exportations y figurent également, de même que la variation des stocks des entreprises et les dépenses  *finales*  de ces dernières, c'est-à-dire celles qui sont considérées comme des investissements devant être amortis sur plus d'un an (soit les dépenses concernant les machines et matériel ou la construction).

Selon le modèle des entrées-sorties, la production totale d'un secteur au cours d'une année doit évaluer la somme de l'ensemble de ses ventes intermédiaires auprès des secteurs industriels et de l'ensemble de ses ventes finales auprès des agents économiques :

$$X_i = \sum_j z_{ij} + \sum_l y_{il}$$

où  $z_{ij}$  est la vente intermédiaire du secteur  $i$  au secteur  $j$  et  $y_{il}$  est la vente finale du secteur  $i$  à l'agent  $l$ .

Composée des éléments  $z_{ij}$  pour les  $n$  branches d'activité économique, la matrice  $Z$  (de dimensions  $n \times n$ ) représente les flux interindustriels de l'économie : chaque colonne donne la répartition des achats intermédiaires d'un secteur auprès de chacun des secteurs, alors que chaque ligne contient la répartition des ventes intermédiaires d'un secteur auprès de chacune des secteurs. Riche en information sur les échanges en biens et services des secteurs, c'est généralement cette matrice qui est utilisée pour le repérage des filières industrielles.

### Les liens économiques étudiés pour le repérage des filières

Certaines méthodes de repérage des filières mettent l'emphase sur les liens directs entre les secteurs industriels alors que d'autres examinent surtout à quel point leurs profils d'achats ou de ventes se ressemblent : prennent-ils leurs intrants auprès des mêmes sources? Desservent-ils plusieurs secteurs communs? Que ce soit les liens verticaux (les chaînes d'approvisionnement) ou les liens horizontaux (la similitude des profils d'achats ou de ventes) qui sont privilégiés, les données utilisées représentent

souvent les liens économiques qui sont analysés pour former les filières. Dans la revue de la littérature, on retrouve quatre types de liens économiques.

**Les coefficients d'achat** : il s'agit, pour chaque secteur, de la proportion de ses intrants intermédiaires qui proviennent de chacun des autres secteurs. Chaque colonne de la matrice des flux interindustriels, où la valeur est exprimée en pourcentage plutôt qu'en valeur absolue, constitue le vecteur des coefficients d'achat d'un secteur.

**Les coefficients de vente** : il s'agit, pour chaque secteur, de la proportion des ventes intermédiaires qui sont effectuées auprès des autres secteurs. Chaque ligne de la matrice des flux interindustriels, où la valeur est exprimée en pourcentage plutôt qu'en valeur absolue, constitue le vecteur des coefficients de vente d'un secteur.

**Les coefficients de « flux » totaux** : pour chaque paire d'industries, le coefficient de flux total correspond à la somme des achats intermédiaires que l'une fait auprès de l'autre, divisée par la somme du total des achats intermédiaires des deux industries.

**Les coefficients techniques de production** : ces coefficients sont calculés à partir de la matrice des entrées plutôt qu'à partir de la matrice des flux interindustriels. Pour chaque secteur, ils correspondent à la consommation relative de chaque produit et service pour une unité de production. Chaque colonne de la matrice des entrées, où la valeur est exprimée en pourcentage plutôt qu'en valeur absolue, constitue le vecteur des coefficients techniques de production d'un secteur

Il est important de rappeler que la matrice des flux interindustriels ne comprend pas les transactions interindustrielles en capital (machines, matériel et construction), qui sont comptabilisées dans la matrice de la demande finale. Par exemple, la vente d'un avion par le secteur productif « Fabrication de produits aérospatiaux et de leurs pièces » au secteur « Transport aérien » n'apparaît pas dans la matrice des flux interindustriels.

## Méthodes de repérage quantitatives des filières

---

Plusieurs types d'analyse statistique ont été utilisés pour extraire et traiter de l'information des tableaux d'entrées-sorties afin de grouper les secteurs industriels en filières. Dans le présent chapitre, cinq types d'analyse statistique sont présentés : l'analyse des composantes principales (ACP)/l'analyse factorielle (AF), l'analyse de classification, la théorie des graphes, la méthode du Maxima et la méthode du *Consensus Clustering*<sup>5</sup>. Pour chacune, nous donnons d'abord un aperçu général des concepts théoriques sous-jacents. Suivent une description des usages faits par certains auteurs et, enfin, une liste des avantages et des inconvénients propres à la méthode. Soulignons que l'on trouve un tableau synthèse des travaux ayant fait l'objet de ce chapitre à la fin du présent document et un autre en l'annexe.

### L'analyse des composantes principales et l'analyse factorielle

#### Aperçu des concepts théoriques

L'analyse des composantes principales et l'analyse factorielle visent à *réduire* un grand nombre de variables à un plus petit nombre de dimensions cachées ou latentes, qui reflètent la structure des liens entre les variables originelles. Plus concrètement, ces techniques tentent d'expliquer la portion la plus importante de la variance des variables par la création de nouvelles variables appelées « composantes » dans le cas de l'analyse en composantes principales et « facteurs » dans le cas de l'analyse factorielle. L'analyse factorielle (AF) ne s'intéresse qu'à la variance commune des variables alors que l'analyse en composantes principales (ACP) tient compte de la totalité de la variance des variables (soit leur variance commune et leur variance unique).

Dans le cas de la problématique qui nous intéresse, l'ACP et l'AF permettent de classer les secteurs industriels en un nombre restreint de groupes, les composantes ou facteurs générés constituant nos filières industrielles. Ce regroupement se fait sur la base des similitudes observées à partir des tableaux d'entrées-sorties. La première étape consiste à manipuler ces tableaux (soit généralement la matrice des flux interindustriels décrite au chapitre précédent) afin de générer la matrice qui sera soumise à l'analyse. Certains auteurs calculent les coefficients d'achats des secteurs, d'autres calculent leurs coefficients de ventes, d'autres, encore, procèdent à un autre type de manipulation qui reflètera les liens interindustriels sous un autre angle. Habituellement, on transforme ensuite cette matrice en une matrice de corrélation, dont chaque cellule exprime la corrélation d'un secteur par rapport à une autre en fonction du type de liens considérés (achats, ventes, échanges intermédiaires totaux, etc).

On procède en deuxième lieu à l'extraction des composantes ou des facteurs (nos filières) à l'aide d'un logiciel d'analyse statistique. Vient en troisième lieu l'interprétation des résultats. Typiquement, ces derniers présentent la liste des composantes extraites

---

<sup>5</sup> Nous n'avons pas retenu la méthode utilisée par Boyer, Dagenais et Martin (1980) avec les tableaux d'entrées-sorties du Québec de 1973 car les auteurs avaient défini *a priori* le « noyau » de leurs trois filières, soit le pétrole, la chimie ainsi que le fer et l'acier. Les auteurs expliquent leur choix par le fait que ces industries « [...] correspondent traditionnellement à des groupes industriels ayant déjà, dans des études antérieures, été retenus comme industries centrales ». (p.4)

présentées en ordre décroissant de variance expliquée (ce qui est donné par la *valeur propre*<sup>6</sup>) : la première composante explique la plus forte proportion de la variance des variables originelles, et ainsi de suite. Le chercheur doit décider du nombre de composantes à garder et peut se référer à divers critères pour ce faire : le pourcentage cumulatif de la variance commune expliquée (on pourrait garder les « n » premières composantes totalisant 75 % de la variance commune, par exemple), le critère de Kaiser-Guttamn, qui stipule qu'une composante devrait avoir une valeur propre au moins supérieure à l'unité, ou encore le scree test de Cattell, qui permet de discriminer le nombre de composantes à garder à l'aide d'une représentation graphique de leurs valeurs propres<sup>7</sup>.

Le tableau qui suit montre un exemple de résultats provenant de l'extraction de facteurs, tirés des travaux de Feser et Bergman (2000) qui portent sur les filières américaines (nous présentons ces travaux un peu plus loin). On voit que l'analyse a extrait 28 facteurs (ou filières), dont le nom donné à chacun par les auteurs apparaît dans la colonne *Interpretation*. Au total, ces 28 facteurs expliquent 100 % de la variance commune des données analysées, comme l'illustre la dernière colonne du tableau. On remarque toutefois que les trois premiers facteurs expliquent à eux seuls la moitié du total de cette variance.

Tableau 3.1 Extrait des résultats de Feser et Bergman (2000) présentant la liste des filières américaines, *Table 1. Summary results : principal components factor analysis*

Factor	Interpretation	Eigenvalue	% total variance	% common variance	Cumulatif, % de la variance commune
1	Metalworking	90,50	25,0	28,0	28,0
2	Vehicule manufacturing	40,27	11,1	12,4	40,4
3	Chemicals and rubber	30,86	8,5	9,5	49,9
4	Electronics and computers	22,91	6,3	7,1	57,0
5	Packaged food products	18,30	5,1	5,7	62,7
6	Printing and publishing	15,96	4,4	4,9	67,6
7	Wood products	14,35	4,0	4,4	72,0
8	Knitted goods	12,49	3,4	3,9	75,9
9	Fabricated textile products	7,54	2,1	2,3	78,2
10	(unreported)	6,05	1,7	1,9	80,1
11	Nonferrous metals	5,64	1,6	1,7	81,8
12	Canned and bottled goods	5,37	1,5	1,7	83,5
13	Leather goods	5,36	1,5	1,7	85,2
14	Aerospace	4,91	1,4	1,5	86,7
15	Food products	4,37	1,2	1,4	88,1
16	Platemaking and typesetting	4,33	1,2	1,3	89,4
17	Aluminium	3,83	1,1	1,2	90,6
18	(unreported)	3,77	1,0	1,2	91,8

<sup>6</sup> La valeur propre, ou *Eigen value* en anglais, correspond à la somme des poids factoriels au carré de toutes les variables pour une composante ou un facteur donnés.

<sup>7</sup> Pour un exemple du test du scree plot de Cattell, voir ACITS, The University of Texas at Austin (1995).

19	Brake and wheel products	3,59	1,0	1,1	92,9
20	Concrete, cement and brick	3,53	1,0	1,1	94,0
21	Earthenware products	3,09	0,9	1,0	95,0
22	Tobacco products	2,82	0,8	0,9	95,9
23	(unreported)	2,67	0,7	0,8	96,7
24	(unreported)	2,47	0,7	0,8	97,5
25	Dairy products	2,37	0,7	0,7	98,2
26	Petroleum	2,32	0,6	0,7	98,9
27	Meat products	2,06	0,6	0,6	99,5
28	(unreported)	1,90	0,5	0,6	100,1

Note : la dernière colonne du tableau est un ajout fait par l'ISQ.

Pour chaque variable à l'étude ou secteur industriel dans notre cas les résultats de l'analyse indiquent le degré de corrélation avec chaque composante extraite. Cette mesure se nomme « poids factoriel » ou *factor loadings*. Généralement, le chercheur associera une variable (soit un secteur industriel) à la composante ou au facteur (c'est-à-dire à la filière industrielle) avec lequel elle est le plus fortement corrélée. Un seuil minimal est généralement utilisé : un secteur n'enregistrant aucun poids factoriel supérieur à ce seuil ne sera associé avec aucune filière. Soulignons que le tableau 3.3 du présent chapitre montre le poids factoriel de certains des secteurs associés à la deuxième filière de Feser et Bergman (2000) présentée ci-dessus, soit la filière *Vehicle manufacturing*.

### Utilisations de la méthode

L'analyse factorielle a été utilisée pour repérer des filières industrielles avec les tableaux d'entrées-sorties dès les années 1970. **Roepke et autres (1974)** se sont basés sur les tableaux de l'Ontario de l'année 1965 pour retracer ce qu'ils ont appelé des « complexes industriels » :

« An industrial complex contains a base group of industries that have similar patterns of transactions, and it also includes other industries which are major suppliers or markets for those within the group». <sup>8</sup>

Les auteurs ont travaillé avec les tableaux d'entrées-sorties agrégés à 44 secteurs, ne contenant pas les secteurs des services et des transports. Ils ont appliqué l'analyse en composantes principales sur trois matrices de corrélation. La première matrice se base sur la somme des coefficients d'achats de chaque paire d'industries (pour une analyse « agrégée »). La deuxième matrice est élaborée à partir des coefficients d'achats des industries (pour une analyse selon « l'origine ») alors que la troisième se fonde sur leurs coefficients de ventes (pour une analyse selon « la destination »).

<sup>8</sup> Roepke et autres (1974), page 15.

Les composantes extraites ayant une valeur propre de 1 ou plus ont été retenues comme filière et le seuil minimal des poids factoriels pour associer un secteur à une filière est fixé à 0,5. D'après les auteurs, les résultats des trois analyses se sont avérés similaires à bien des égards, soit par rapport au nombre de filières extraites (13), à la proportion de la variance totale expliquée (environ 85 %) et aux industries identifiées comme étant reliées entre elles. La composition de chaque composante ou filière industrielle, toutefois, varie quelque peu d'une analyse à l'autre. À la lumière des résultats, les auteurs jugent que l'analyse agrégée est davantage appropriée lorsque l'on s'intéresse à l'identification de filières contenant des industries directement liées entre elles, alors que les analyses selon l'origine ou la destination sont plus pertinentes lorsque l'on souhaite mettre en lumière les industries constituant ou desservant un même marché.

**Czamanski (1974)** a également eu recours à l'analyse des composantes principales pour regrouper des industries sur la base de leurs profils d'achats et de ventes. La composition de sa matrice de corrélation diffère toutefois de celle de Roepke et autres (1974). Pour chaque paire d'industries (i,j), Czamanski a calculé les quatre coefficients de ventes et d'achats :

$$\begin{array}{ll} a_{ij} = z_{ij} / s_i & a_{ji} = z_{ji} / s_j \\ b_{ij} = z_{ij} / p_i & b_{ji} = z_{ji} / p_j \end{array}$$

où  $z_{ij}$  ( $z_{ji}$ ) est la valeur des flux en biens et services intermédiaires de l'industrie i (j) vers l'industrie j (i) au cours de l'année;

$s_i$  ( $s_j$ ) est la valeur des ventes intermédiaires totales de l'industrie i (j);

$p_i$  ( $p_j$ ) est la valeur des achats intermédiaires totaux de l'industrie i (j).

Le coefficient  $a_{ij}$  ( $a_{ji}$ ) mesure la proportion des ventes intermédiaires de l'industrie i (j) auprès de l'industrie j (i) par rapport aux ventes intermédiaires totales de l'industrie i (j). Un coefficient  $a_{ij}$  ( $a_{ji}$ ) élevé révèle que l'industrie i (j) dépend de l'industrie j (i) pour une bonne partie de ses ventes.

Le coefficient  $b_{ij}$  ( $b_{ji}$ ) mesure quant à lui la proportion des achats intermédiaires de l'industrie i (j) effectués auprès de l'industrie j (i) par rapport aux achats intermédiaires totaux de l'industrie i (j). Un coefficient  $b_{ij}$  ( $b_{ji}$ ) élevé montre que l'industrie i (j) dépend de l'industrie j (i) pour une bonne partie de ses intrants.

Czamanski a par la suite calculé, pour chaque paire d'industries, la valeur du coefficient de corrélation entre ces quatre coefficients. Ainsi, quatre coefficients de corrélation peuvent décrire la similitude entre les entrées-sorties de deux industries, k et l :

$r(a_k, a_l)$  : mesure le degré de similitude entre les profils de ventes des industries k et l;

$r(b_k, b_l)$  : mesure le degré de similitude entre les profils d'achats des industries k et l;

$r(a_k, b_l)$  : mesure le degré de similitude entre les profils de ventes de l'industrie k et les profils d'achats de l'industrie l. Un coefficient élevé indique que l'industrie l achète une bonne partie de ses intrants auprès d'industries qui sont desservies par l'industrie k.

$r(b_k, a_l)$  : mesure le degré de similitude entre les profils d'achats de l'industrie  $k$  et les profils de ventes de l'industrie  $l$ . Un coefficient élevé indique que l'industrie  $k$  achète une bonne partie de ses intrants auprès d'industries qui sont desservies par l'industrie  $l$ .

Pour former sa matrice de corrélation finale, Czamanski n'a gardé que le coefficient de corrélation le plus élevé, pour chaque paire d'industries. Soumise à l'analyse factorielle, cette matrice (dérivée des tableaux d'entrées-sorties des États-Unis pour l'année 1963 avec 172 secteurs représentant tous les secteurs d'activité) a donné 16 filières industrielles.

Latham (1976) et Harrigan (1982) ont critiqué les travaux de Roepke et autres (1974) et de Czamanski (1974) en affirmant que l'analyse en composantes principales et l'analyse factorielle ne groupaient pas les industries liées *verticalement*, c'est-à-dire selon la chaîne de production. Dans un article paru en 1984, **O hUallachain (1984)** a voulu rétablir la pertinence de ces méthodes statistiques en insistant sur l'importance des relations *complémentaires* dans la formation de filières industrielles et sur le fait que l'ACP et l'AF sont extrêmement utiles pour décoder ce type de relations. Reconnaisant que les filières générées par l'ACP n'incluent pas nécessairement l'industrie (ou les industries) à la source de la complémentarité du groupe (c'est-à-dire à la base des achats ou des ventes de ce dernier), O hUallachain a affirmé qu'il est relativement facile d'identifier ces industries par un simple examen *a posteriori* des profils d'achats ou de ventes des industries groupées, à partir de la matrice des flux interindustriels.

O hUallachain a appuyé ses dires par une analyse en composantes principales de deux matrices de corrélation dérivées des tableaux d'entrées-sorties de l'État de Washington pour l'année 1963. Ces tableaux étaient constitués de 49 secteurs regroupant toutes les branches d'activité de l'économie. La première matrice se base sur les coefficients d'achats des secteurs et la deuxième, sur leurs coefficients de ventes. Les deux analyses ont chacune généré 15 composantes ayant une valeur propre égale ou supérieure à l'unité, expliquant respectivement 83 % et 84 % de la variance totale des données analysées. La valeur minimale pour associer un secteur à une composante a été fixée à 0,4. Afin de vérifier si la source de la complémentarité des industries groupées faisait partie des filières, O hUallachain a calculé un indice en amont (*backward linkage*) et un indice en aval (*forward linkage*) pour chaque filière, mesurant respectivement la proportion des achats et des ventes des industries groupées qui étaient effectuées au sein même de la filière. Ainsi, l'indice en amont (évalué pour les composantes identifiées à l'aide de la matrice de corrélation des coefficients d'achats) s'élève à 72 % pour la filière « bois et forêt » mais à 7 %, seulement, pour la filière contenant les secteurs « métaux lourds, moteurs non électriques et véhicules motorisés » (voir le tableau 3.2). Un examen du profil d'achats de ces trois secteurs, à l'aide de la matrice des flux interindustriels, a toutefois permis de constater que ces derniers étaient liés par l'industrie « fer et acier », qui représente globalement 46 % de leurs achats.

Tableau 3.2 Extrait traduit des résultats de O hUchallain (1984), Components derived from R-mode analysis of 49 x 49 matrix of regional interindustrial - purchase coefficients (common purchasing patterns)

Filière	Valeur propre	Industries	Poids factoriel	Indice en amont (%)	Source de complémentarité	Prop. des achats de la filière venant de la source de complémentarité (%)
II	9,6	Métaux lourds	0,49	7,03	Fer et acier	45,67
		Moteurs non élect.	0,44			
		Véhicules motorisés	0,47			
VIII	4,0	Forêt	0,65	72,4	Forêt	56,52
		Bois	0,72			

O hUallachain souligne l'importance, selon lui, d'analyser les liens interindustriels sous l'angle des achats et sous l'angle des ventes *séparément*, c'est-à-dire sans combiner les deux types d'information comme l'ont fait Roepke et autres (1974) et Czamanski (1974). L'interprétation des résultats s'en trouve facilitée, puisque l'on sait quel type de liens réunit les industries associées à une même filière. La comparaison des deux ensembles de résultats de O hUallachain révèle une meilleure « définition » des filières identifiées sur la base des ventes : en général, les ventes « intrafilières » sont proportionnellement supérieures aux achats « intrafilières » et les liens des secteurs avec les filières, tels que mesurés par les poids factoriels, sont plus tranchés.

À la fin des années 1990, **Feser et Bergman (2000)** ont repris les coefficients de corrélation de Czamanski pour identifier les filières industrielles dans le secteur manufacturier des États-Unis<sup>9</sup>. Ils ont utilisé les tableaux d'entrées-sorties de l'année 1987 à un niveau très désagrégé (362 secteurs ne représentant que les secteurs manufacturiers<sup>10</sup>). L'objectif était d'identifier les filières à l'échelle nationale pour ensuite étudier leur présence relative au sein de l'État de la Caroline du Nord, au moyen de données très fines sur l'emploi et la valeur ajoutée. L'analyse a extrait 28 filières, dont 5 ont été exclues car elles ne contenaient qu'un seul secteur industriel fortement lié. Reconnaissant que dans les faits il n'est pas rare qu'un secteur industriel soit apparenté à plus d'une filière, les auteurs ont défini les critères suivants pour associer un secteur à une filière :

- un secteur est considéré **fondamental** pour une filière (*primary*) s'il obtient un poids factoriel d'au moins 0,60 sur cette filière et qu'il s'agit là de son poids factoriel le plus élevé;

<sup>9</sup> Voir également Bergman et Feser (1999).

<sup>10</sup> Feser et autres (2000) ont également fait l'exercice en incluant les secteurs des services, mais ont alors obtenu un plus petit nombre de filières, plus larges, comprenant des industries très hétérogènes.

- un secteur est considéré **secondaire** pour une filière (*secondary*) si son poids factoriel se situe entre 0,35 et 0,60 (à moins, s'il surpasse 0,60, qu'il n'obtienne un score plus élevé sur un autre facteur).

Le tableau qui suit présente un extrait des résultats concernant la filière no 2, regroupant les industries liées à l'industrie automobile (*vehicule manufacturing cluster*). On trouve dans les colonnes L1, L2 et L3 le numéro des filières auxquelles chaque secteur est associé, en ordre décroissant d'importance; on voit que les secteurs secondaires sont d'abord et avant tout associés à une autre filière que la filière no 2 (colonne L1).

Tableau 3.3 Extrait des résultats de Feser et autres (2000), Table 4. Sectoral composition: US vehicle manufacturing cluster

	SIC*	Description	Cluster ID			Factor loading
			L1	L2	L3	
<i>Primary sectors</i>	3716	Motor homes	2			0.96
	3694	Electrical equip. for internal combustion eng.	2			0.93
	301	Tires & inner tubes	2			0.93
	3651	Household audio & video equip.	2	4		0.91
	306	Fabricated rubber products, n.e.c.	2	3		0.89
	2952	Asphalt felts & coatings	2			0.89
	302	Rubber & plastics footwear	2	9		0.88
	3645-8	Lighting fixtures & equip.	2	1		0.88
	(...)					
<i>Secondary sectors</i>	2522	Office furniture, except wood	1	2		0.66
	3823-4, 3829	Mechanical measuring devices	4	2		0.62
	3291	Abrasive products	3		2	0.59
	3465	Automotive stampings	1		2	0.59
	(...)					

\* Le SIC fait référence au Système de classification industrielle.

La comparaison des filières en matière d'emploi, de rémunération, de valeur ajoutée ou d'autres statistiques est possible lorsque l'on ne tient compte que des secteurs fondamentaux, puisque les filières sont alors mutuellement exclusives. Soulignons que le professeur Feser et ses collègues ont subséquentement étudié la concentration spatiale de l'emploi au sein d'une filière à l'aide de la statistique G de Getis et Org, qui prend en considération les juridictions voisines d'une région afin de transcender les frontières administratives découpant les statistiques régionales sur l'emploi (Feser et autres, 2001a, Feser et autres, 2001b). Des travaux ont également été menés pour vérifier si les établissements membres d'une filière industrielle étaient plus susceptibles d'être à proximité géographique que les établissements n'en faisant pas partie (Feser et Sweeney, 2000).

Encore plus récemment, Rey et Mattheis (2000) ont suggéré d'utiliser l'analyse factorielle pour identifier les filières industrielles, en conjonction avec d'autres techniques statistiques. Leur méthode, baptisée *Consensus Clustering Method*, est présentée à la fin de ce chapitre.

## Avantages et inconvénients

L'analyse des composantes principales/analyse factorielle comporte plusieurs **avantages** :

- La méthode identifie les secteurs qui ont des liens complémentaires, c'est-à-dire qui partagent des bassins d'intrants ou des marchés similaires. Ces liens sont souvent sous-jacents aux agglomérations industrielles;
- Bien que la source de complémentarité des secteurs groupés par la méthode ne fasse pas toujours partie d'une filière, il est possible de l'identifier en inspectant la matrice des flux interindustriels;
- Les valeurs propres et les poids factoriels donnent une mesure du degré d'importance des liens des secteurs groupés;
- Les poids factoriels indiquent l'importance de la corrélation d'un secteur avec *chacune* des filières identifiées. Dans les faits, il est pratique d'avoir des filières mutuellement exclusives, mais il est néanmoins utile de connaître le niveau d'association d'un secteur avec chacune des filières identifiées;
- Il existe des logiciels d'analyse statistique sophistiqués (tels que SAS et SPSS) pour effectuer ce genre d'analyse.

L'on peut voir deux **inconvénients** principaux à l'ACP / AF :

- La méthode est davantage appropriée pour identifier les industries complémentaires que pour regrouper les industries liées verticalement. Il faut inspecter la matrice des flux interindustriels *a posteriori* pour identifier les industries étant à la source des regroupements (sur le plan des achats ou des ventes);
- L'interprétation des résultats peut être difficile si l'on a peu ou trop de composantes ou de facteurs extraits.

## L'analyse de classification

### Aperçu des concepts théoriques

Tout comme l'analyse en composantes principales ou l'analyse factorielle, l'analyse de classification (*clustering analysis*) fait partie des outils d'analyse de covariance multivariée. À l'inverse de l'ACP ou de l'AF, toutefois, qui distribuent la variance d'une variable entre plusieurs composantes ou facteurs, l'analyse de classification attribue la totalité de la variance d'une variable au regroupement avec lequel cette dernière est le plus corrélée. Concrètement, dans le cadre de la

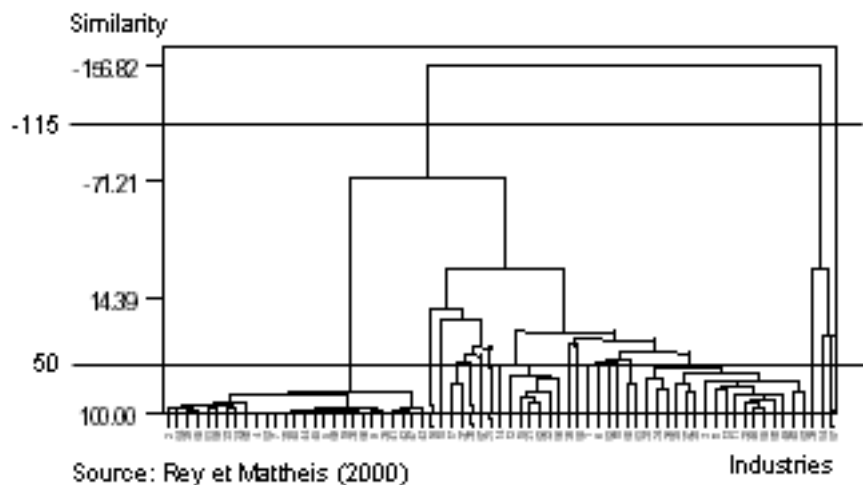
problématique qui nous intéresse, cela signifie qu'un secteur industriel se voit associé à une seule et même *classe*, ou filière industrielle. Ces dernières sont donc d'emblée mutuellement exclusives.

L'analyse de classification regroupe les secteurs industriels en fonction de leurs similitudes (ou, moins fréquemment, de leurs différences). Une « matrice de similitude » doit d'abord être préparée et ensuite soumise à l'analyse. Cette matrice peut contenir, par exemple, les coefficients d'achats des secteurs ou encore leurs coefficients techniques de production.

La méthode *hiérarchique ascendante* de l'analyse de classification commence par calculer la distance euclidienne séparant chaque secteur (ou plus concrètement chaque vecteur de la matrice de similitude)<sup>11</sup> et regroupe en filières les deux secteurs les plus « proches ». D'une façon itérative, la méthode calcule ensuite le centre de gravité des nouvelles filières et agrège les deux filières ou combinaisons filière-secteur étant les plus « proches », jusqu'à ce que tous les secteurs originels soient ultimement regroupés en une seule et même filière. Différents algorithmes peuvent être utilisés à ces fins (Centroid, Median et Ward, par exemple), chacun mesure, à sa façon, la distance entre deux filières.

Les résultats de l'analyse de classification sont présentés sous la forme d'un dendrogramme, dont on trouve un exemple à la figure suivante.

Figure 3.1 Exemple de dendrogramme



Le dendrogramme représente la « hiérarchie » des filières en fonction d'une échelle de similarité. Plus le niveau d'agrégation est élevé (c'est-à-dire plus le nombre de filières est petit), plus la similarité entre les secteurs constituant une filière est faible. Par exemple, le dendrogramme de la figure 3.1 montre qu'à un niveau de similarité d'environ 50, il existe un grand nombre de filières; au seuil d'environ -115, il n'en reste que deux (dont l'une contient la presque totalité des secteurs). Il revient au chercheur d'analyser

<sup>11</sup> La distance euclidienne entre des vecteurs de production A et B correspond à la racine carrée de la somme des différences observées entre A et B pour chaque coefficient technique de production, élevées au carré.

les résultats et de choisir le niveau de similarité donnant d'après lui la meilleure combinaison nombre et taille de filières.

### Utilisations de la méthode

Rey et Mattheis (2000) s'étonnent du fait que l'analyse de classification n'ait encore été que peu utilisée pour repérer les filières industrielles. Les travaux effectués dans le domaine semblent effectivement avoir davantage porté sur divers aspects méthodologiques de l'agrégation des tableaux d'entrées-sorties que sur l'identification *per se* de filières industrielles. Ainsi, **Caber et autres (1991)** se sont penchés sur les conditions permettant de discriminer le type d'algorithme à utiliser (entre ceux de Ward, Centroid et Median) pour que les filières formées respectent certaines propriétés. Ils ont eu recours aux tableaux d'entrées-sorties de 1985 de l'Espagne (à un niveau d'agrégation de 56 secteurs) et ont effectué des tests avec la matrice inverse de Léontief et la matrice des coefficients techniques de production. **Blin et Cohen (1977)** ont quant à eux examiné les biais d'agrégation des tableaux d'entrées-sorties avec les algorithmes de Ward et Centroid, appliqués aux coefficients techniques de production des États-Unis (avec 83 secteurs, pour l'année 1967).

**Abbott et Andrews (1990)** se sont servis de l'analyse de classification à des fins très pratiques. Ils ont effectué une analyse des secteurs manufacturiers des États-Unis tels qu'identifiés avec le Système de classification industrielle (SIC) à 4 chiffres, soit 449 secteurs. L'objectif était de regrouper les secteurs en fonction des technologies utilisées par chacun (telles qu'exprimées par les coefficients techniques de production) en faisant rouler l'algorithme de Ward jusqu'à ce qu'un nombre « n » de filières soient formées, « n » représentant le nombre de classes du SIC à 3 chiffres. Les auteurs ont ensuite comparé la composition des filières générées avec celles des classes du SIC; ils voulaient vérifier si le schème d'agrégation du Système de classification industrielle prenait bien en compte les technologies utilisées par les secteurs. L'analyse a révélé plusieurs différences entre les deux types d'agrégation.

D'après Rey et Mattheis (2000), l'utilisation de l'analyse de classification en vue de repérer les filières industrielles devrait faire l'objet de recherches supplémentaires. Eux-mêmes incorporent ce type d'analyse dans leur méthodologie de repérage des filières, la *Consensus clustering method*.

### Avantages et inconvénients

Voici les principaux **avantages** de l'analyse de classification :

- La méthode identifie les secteurs qui ont des liens complémentaires, c'est-à-dire qui partagent des bassins d'intrants ou des marchés similaires;
- La méthode génère des filières qui sont mutuellement exclusives;
- La présentation des résultats sous la forme d'un dendrogramme facilite leur interprétation, puisqu'on peut visualiser le niveau de similitude des secteurs formant une filière;

- Il existe des logiciels d'analyse statistique sophistiqués pour effectuer ce genre d'analyse.

En voici les **inconvénients** :

- La méthode est davantage appropriée pour repérer les secteurs complémentaires que pour regrouper ceux liés verticalement. Il faut inspecter la matrice des flux interindustriels *a posteriori* pour repérer les secteurs étant à la source des regroupements (sur le plan des achats ou des ventes);
- Dans les faits, un secteur industriel est souvent associé à plus d'une filière (bien qu'aux fins de la description de ces dernières, on souhaite que leur composition soit mutuellement exclusive). Or, la méthode ne relie un secteur qu'à une seule filière; contrairement à l'ACP/AF, on n'obtient pas une mesure de l'importance du lien qu'a une industrie avec chacune des filières identifiées;
- À la lumière des travaux de Rey et Mattheis (2000) présentés plus loin, il semble que l'analyse de classification forme peu de filières potentiellement très grosses, lorsqu'elle est utilisée avec les matrices de coefficients d'achats et de ventes;
- Jusqu'à ce jour, la méthode a semble-t-il été peu utilisée pour l'identification *per se* de filières industrielles. Nous ne pouvons donc nous référer qu'à un nombre restreint de résultats.

## La théorie des graphes

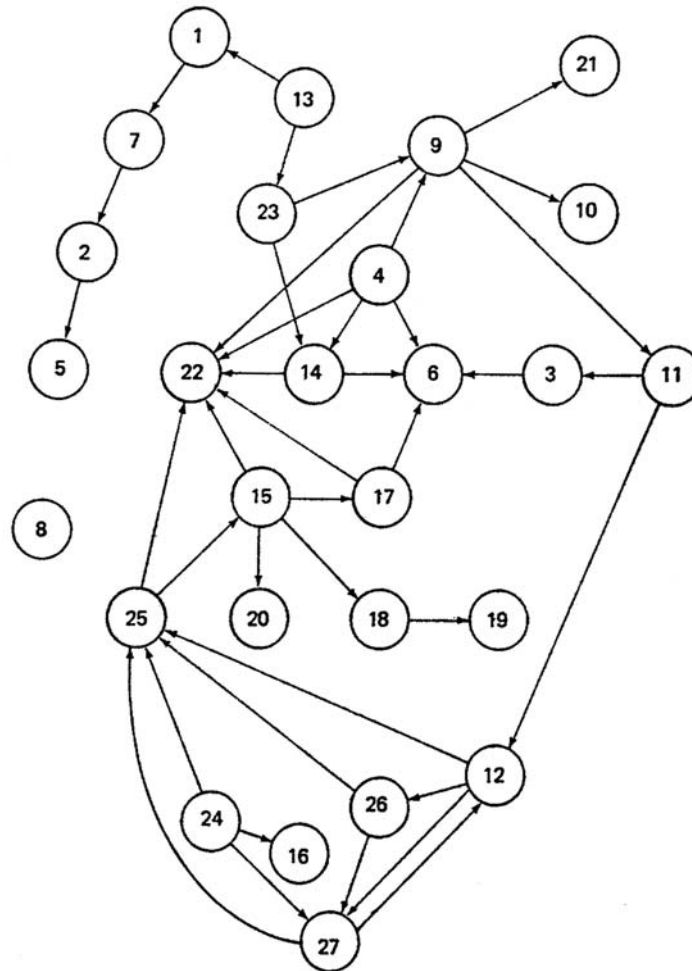
### Aperçu des concepts théoriques

Contrairement à l'ACP/l'AF et à l'analyse de classification qui s'intéressent aux liens horizontaux des secteurs en groupant ces derniers sur la base de la similitude de leurs profils d'achats et de ventes, la méthode de la théorie des graphes repère les filières en regroupant les secteurs qui sont liés entre eux par des liens verticaux. En effet, la méthode se base sur l'importance des achats et des ventes qui sont directement effectués entre les secteurs.

En tout premier lieu, on procède à l'examen de la matrice des flux interindustriels afin d'établir les liens de vente et d'achat. Un lien de vente existe entre deux secteurs si le premier vend une proportion suffisamment élevée de sa production au deuxième (lecture horizontale de la matrice). De façon similaire, un lien d'achat est établi si un secteur achète une proportion significative de ses intrants auprès d'un autre (lecture verticale de la matrice). Cela conduit à l'élaboration d'une *matrice d'adjacences*, composée de « 0 » et de « 1 » : les « 1 » indiquent l'existence d'un lien de vente ou d'achat significatif pour chaque paire de secteurs et les « 0 », l'absence de tels liens.

On transpose ensuite cette matrice en un graphe orienté ou *digraphe*, dont les arcs représentent les flux de biens et services entre les secteurs. La flèche est dirigée vers le secteur « client », à partir du secteur « fournisseur ». La figure qui suit est un exemple de digraphe tiré des travaux de Campbell (1972). On y voit les relations significatives d'achat et de vente entre les 27 secteurs agrégés de l'État de Washington en 1963.

Figure 3.2 Exemple de digraphe provenant des travaux de Campbell (1972), Fig.2. *Digraph of linked industries in the Washington economy, 1963*



La dernière étape consiste à analyser le digraphe afin de déceler la structure industrielle de l'économie. L'analyse peut se faire sous divers angles. On peut viser à identifier les *composantes connexes* (*weak components*) et les *composantes fortement connexes* (*strong components*). Deux secteurs font partie d'une composante connexe s'il existe au moins un chemin qui les relie entre eux quelle que soit l'orientation de ce chemin, alors qu'ils font partie d'une composante fortement connexe s'il existe au moins deux chemins les menant de l'un à l'autre. Il n'y a qu'une composante fortement connexe dans le digraphe ci-dessus, formée des secteurs 12, 26 et 27. Ces secteurs pourraient être « condensés » en un seul, comme l'explique Campbell :

« An important implication of the existence of this strong component is that the three industries within it might be treated as a unit for some planning purposes. This results from the fact that the impact of the expansion of any industry within the strong component will be passed along, via the backward linkage chain, to all of the other industries within the strong component.»<sup>12</sup>

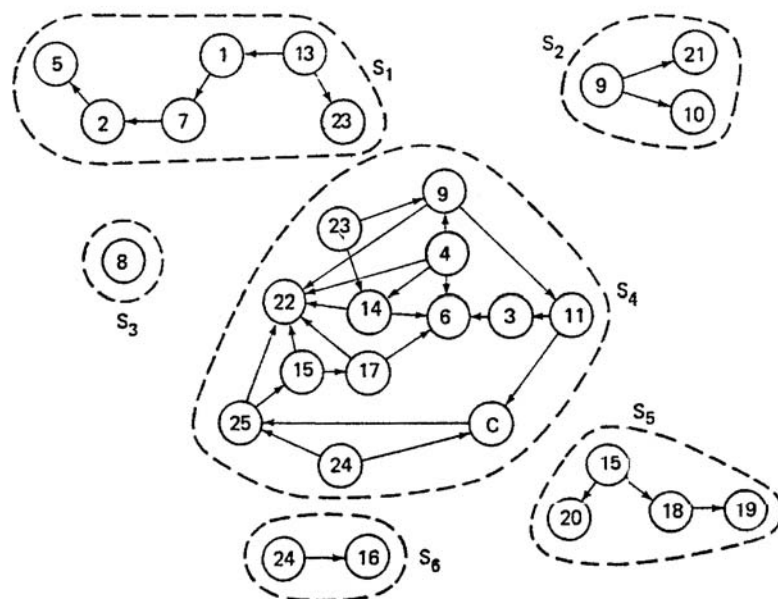
<sup>12</sup> Campbell (1972), page 103.

On peut également décortiquer le digraphe pour identifier les points d'articulation, c'est-à-dire les industries qui une fois retirées (avec leurs arcs incidents) scindent le digraphe en deux ensembles distincts. Il est ainsi possible de décomposer tout le digraphe en blocs, un bloc étant un sous-ensemble du graphe ne contenant aucun point d'articulation.

### Utilisations de la méthode

C'est **Campbell (1972)** qui, le premier, a utilisé la théorie des graphes pour repérer les filières industrielles d'une économie. Il a eu recours aux tableaux d'entrées-sorties de 1963 de l'État de Washington, agrégés à un niveau de 27 industries représentant l'ensemble des secteurs industriels. Campbell a considéré qu'un secteur était lié à un autre par un lien d'achat si il s'y procurait au moins le vingt-septième de ses intrants (en valeur); de façon similaire, il a établi un lien de vente entre deux secteurs si le premier vendait au moins le vingt-septième de sa production au deuxième. Il a construit une matrice d'adjacences sur ces bases, qui est représentée par le digraphe reproduit à la page précédente. Il a ensuite repéré les points d'articulation de ce digraphe, aux fins de la détermination des filières. En tout, neuf secteurs constituent de tels points d'articulation (soit les secteurs 1, 2, 7, 9, 13, 15, 18, 23 et 24). Le retrait de quatre de ces points et de leurs arcs incidents réduit le plus possible le « cœur » du digraphe et mène à l'identification de cinq filières industrielles, comme le montre la figure qui suit. Soulignons que ce sont les secteurs 9, 15, 23 et 24 qui constituent ces points charnières et qui sont d'ailleurs présentes à la fois dans une filière « propre » et dans la filière « centrale ».

Figure 3.3 Extrait des résultats de Campbell (1972), Fig. 5. *Articulation components of the digraph of linked industries in the Washington economy, 1963 (based on fig. 2, formed by removal of articulation points 9, 15, 23 and 24). C = Condensation of strong component-industries 12, 26 and 27.*

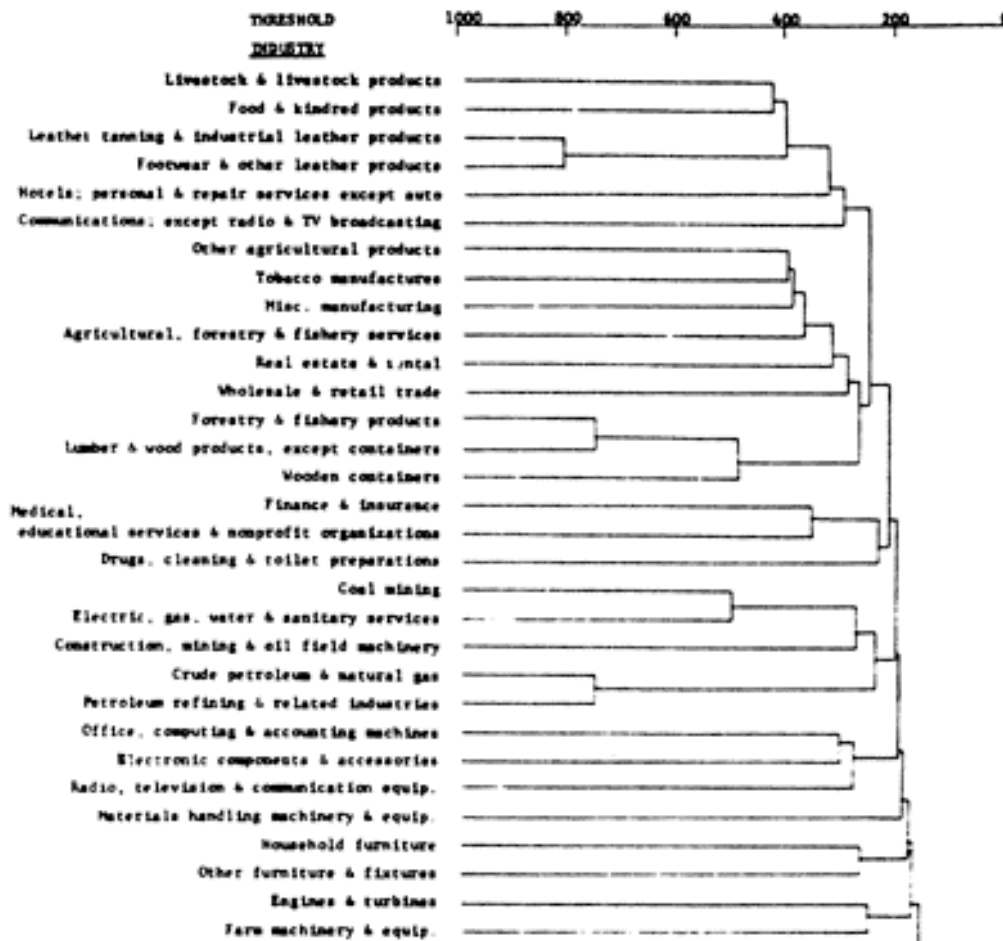


Peu après Campbell, **Slater (1977)** a appliqué la théorie des graphes aux tableaux d'entrées-sorties des États-Unis, retenant 75 des 86 secteurs des tableaux de 1967 (il a entre autres exclu le secteur public, les importations et les « industries spéciales »). Il a

commencé par standardiser la matrice des flux interindustriels, au moyen d'une procédure d'ajustement itératif proportionnel modifiant la valeur des transactions, de façon à ce que les achats et les ventes de chaque secteur totalisent le même montant. Ainsi, la matrice ajustée contient l'estimation des transactions qui auraient lieu si tous les secteurs avaient le même niveau de consommation et de vente intermédiaires. (Cela revient à utiliser les coefficients d'achats ou de ventes plutôt que la valeur nominale des transactions, mais permet de combiner les angles « achat » et « vente » dans la même matrice).

Ensuite, au lieu d'utiliser un unique seuil comme Campbell pour établir les liens d'achat et de vente significatifs, il en a utilisé plusieurs de façon séquentielle, commençant par la plus forte valeur de la matrice ajustée. Les deux secteurs impliqués dans cette transaction ont été regroupés. Slater a ensuite identifié la deuxième transaction la plus importante et a joint les deux secteurs impliqués (ou le secteur et la paire de secteurs déjà regroupés). Abaisant ainsi graduellement le seuil, Slater a pu observer à quel niveau (c'est-à-dire à quelle valeur ajustée de la matrice des flux interindustriels) les regroupements de secteurs se formaient, jusqu'à ce que tous les secteurs fassent éventuellement partie d'une seule grande filière. Les résultats se présentent sous la forme d'un dendrogramme, comme pour l'analyse de classification. Campbell a analysé le graphique sur le plan des composantes connexes : pour un seuil donné (en « coupant » le dendrogramme) on voit quels sont les regroupements de secteurs qui sont tous reliés entre eux par un chemin, quelle que soit l'orientation de ce dernier (c'est-à-dire sans distinction par rapport au point d'origine ou de destination des liens formant le chemin, la matrice ajustée des flux interindustriels combinant à la fois l'angle « achat » et l'angle « vente »).

Figure 3.4 Extrait des résultats de Slater (1977), Fig.1 [note : le dendrogramme est tronqué par manque d'espace]



Plus récemment, **Hauknes (1999)** a repéré les filières industrielles de la Norvège en étudiant les transactions entre les secteurs à partir de tableaux d'entrées-sorties désagrégés (161 secteurs) pour l'année 1993. Nous classons sa méthode dans la catégorie de la théorie des graphes, bien que l'auteur n'y fasse pas directement référence. Hauknes a utilisé deux critères pour établir les liens significatifs : d'une part, il n'a retenu que les achats et les ventes qui représentaient au moins 10 % de la valeur totale des achats et des ventes des secteurs impliqués. D'autre part, il a éliminé d'emblée toutes les transactions des « petits secteurs », c'est-à-dire des secteurs dont la valeur des transactions intermédiaires totalisait moins de 0,1 % du total des transactions intermédiaires de l'économie.

L'analyse des liens significatifs a conduit au repérage de six filières industrielles. Hauknes mentionne que les « frontières » des filières ont été établies de façon à minimiser les échanges inter filières. Il souligne également avoir dû poser un jugement qualitatif par rapport au contenu sectoriel des filières afin de s'assurer que ces dernières étaient plausibles d'un point de vue économique.

### **Avantages et inconvénients**

Le principal **avantage** de la méthode basée sur la théorie des graphes est de lier les secteurs ayant directement des liens entre eux. Cela est utile pour établir des filières sur la base des chaînes de production.

Les **inconvénients** semblent toutefois importants :

- On perd beaucoup d'information puisque l'on ne retient que les transactions effectuées directement entre deux secteurs et, qui plus est, que les transactions jugées suffisamment importantes;
- À la lumière des travaux des auteurs couverts ci - dessus, il semble que le chercheur doive souvent poser un jugement qualitatif sur la composition des filières repérées et plus particulièrement sur les « frontières » de celles-ci. Cela introduit une part d'arbitraire dans le processus.

### **La méthode du « Maxima »**

#### **Aperçu des concepts théoriques**

Certains auteurs ont récemment utilisé une méthode appelée « méthode du Maxima » pour identifier les filières industrielles, qui semble être une variante de l'approche de la théorie des graphes employée par Campbell dans les années 1970. En effet, la méthode du Maxima se base également sur l'importance des achats et des ventes qui sont directement effectués entre les secteurs. Elle tire son nom du fait qu'elle ne retient aux fins des regroupements que LE lien d'achat ou de vente le plus élevé, pour chaque secteur.

La méthode est elle aussi généralement appliquée avec des restrictions : pour être considérée, une transaction « maximale » doit représenter un certain pourcentage des transactions totales des secteurs impliqués (Hoen, 2002). Les paragraphes qui suivent

décrivent la procédure à suivre lorsque l'on examine les liens en aval, c'est-à-dire les ventes des secteurs.

Faisant une lecture des lignes de la matrice des flux intermédiaires, le chercheur identifie d'abord, pour chaque secteur  $i$ , quelle est la plus importante vente (en valeur absolue) qui est effectuée auprès des autres secteurs (excluant le sien, les éléments de la diagonale étant mis à zéro). Il vérifie ensuite si cette vente est substantielle, c'est-à-dire si elle excède, par rapport aux ventes intermédiaires totales du secteur  $i$ , un seuil  $k$  fixé à l'avance. Si c'est le cas, le secteur « client » associé est reconnu comme étant le « meilleur acheteur » du secteur  $i$ . L'exercice mène à la production d'une matrice binaire  $[0,1]$  : la matrice contient un « 1 » pour chaque « meilleur acheteur » et des « 0 » pour toutes les autres cellules.

Le chercheur vérifie ensuite si les transactions retenues sont également substantielles du point de vue des secteurs clients. Pour ce, il lit la matrice des flux interindustriels à la verticale : pour chaque « meilleur acheteur » identifié précédemment, il vérifie si la valeur de la transaction est supérieure à un seuil  $l$  déterminé à l'avance, lorsque divisée par la valeur totale des achats dudit secteur. Cela conduit à la production d'une deuxième matrice binaire, les « 1 » indiquant que la vente au « meilleur acheteur » est également substantielle du point de vue de ce dernier.

Enfin, on procède à l'addition des deux matrices binaires. La résultante est une matrice contenant des « 2 », « 1 » et « 0 » qui permet, au moyen d'un algorithme, de regrouper les secteurs liés par les ventes. Plus les seuils  $k$  et  $l$  sont élevés, plus les filières seront nombreuses mais petites; plus les seuils sont faibles, et plus le niveau d'agrégation des filières sera élevé. La détermination de la valeur de  $k$  et de  $l$  se fait par des essais et erreurs. Le jugement du chercheur est généralement requis pour valider la composition finale des filières.

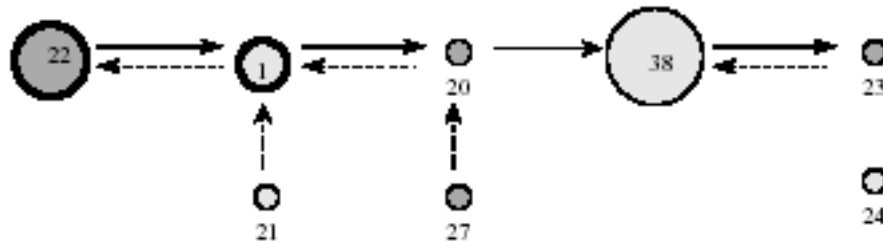
## Utilisations de la méthode

**Peeters et autres (2001)** ont eu recours à la méthode du Maxima pour repérer ce qu'ils ont appelé les *mega-clusters* de la Flandre (Belgique) et de la Suisse. Le niveau d'agrégation des tableaux d'entrées-sorties utilisés est élevé (51 secteurs dans le cas de la Flandre et 46 dans le cas de la Suisse). Concrètement, ils ont analysé séparément les liens en aval et en amont des secteurs à partir de la matrice des flux interindustriels, avec des restrictions par rapport à l'importance relative des transactions (ils ne précisent toutefois pas quels sont les seuils  $k$  et  $l$  utilisés pour chaque analyse). Deux matrices composées de « 0 », « 1 » et « 2 » ont ainsi été produites, l'une résumant les liens de ventes maximums et l'autre, les liens d'achats maximums. Les auteurs les ont additionnées et ont repéré les filières au moyen de cette matrice agrégée, en posant un jugement qualitatif sur les regroupements dans le cas de la Flandre. En tout, cinq *mega-clusters* ont été repérés pour la Flandre et six pour la Suisse.

La figure 3.5 présente la composition de la filière *Agro-food* pour la Flandre, regroupant sept secteurs liés par des liens en amont et en aval. Soulignons que le secteur 24 (produits du tabac) a été ajouté *a posteriori* par les auteurs, l'importance de ses liens avec les autres membres de la filière étant en deçà des seuils utilisés par l'algorithme.

Les flèches pleines représentent les liens maximum en aval (ex. : [38] est le « meilleur acheteur » de [20]) alors que les flèches en pointillé représentent les liens maximum en amont (ex. : [1] est le « meilleur fournisseur » de [21] et de [20]).

Figure 3.5 Extrait des résultats de Peeters et autres (2001), Figure 2. Structure of the Agro-food cluster



Key: [1] Agricultural, forestry and fishery products; [20] Meats and meat products; [21] Milk and dairy products; [22] Other food products; [23] Beverages; [24] Tobacco products; [27] Leathers, leather and skin products, footwear; [38] Lodging and catering services.

S'intéressant à la structure industrielle des Pays-Bas, **Roelandt et autres (1999)** se sont aussi servis de la méthode du Maxima pour tenter de déterminer les filières de l'économie. En plus d'examiner les liens en amont et en aval des secteurs à partir de la matrice des flux interindustriels, ils ont étudié la matrice des sorties et celle des entrées afin de « coupler » le plus important fournisseur d'un produit ou service avec le plus gros consommateur dudit produit ou service. Ils se sont servis de matrices très désagrégées : leur matrice des flux interindustriels comprend 214 secteurs et celles des entrées et sorties, 650 catégories de produits et services. La valeur de  $k$ , déterminant la proportion minimale que devait représenter une transaction par rapport aux ventes ou aux achats intermédiaires totaux d'un secteur pour être considérée substantielle, a été fixée à 0,20 dans le cas des ventes et à 0,15 dans le cas des achats. En tout, 12 filières industrielles ont été repérées par les auteurs, dont 10 gravitent autour de 2 filières reliées aux services (une filière de services commerciaux et une filière de services non commerciaux).

### Avantages et inconvénients

Comme pour la théorie des graphes, le principal avantage de la méthode du Maxima est de lier les secteurs ayant directement des liens entre eux, et, qui plus est, des liens très forts. Toutefois, la méthode ignore encore plus de renseignements que la théorie des graphes, sans produire pour autant des filières dont les frontières sont davantage « tranchées ». Le jugement du chercheur est encore souvent requis pour déterminer la composition finale de chaque filière.

### La méthode du *Consensus Clustering*

#### Aperçu des concepts théoriques

Comme on peut le constater à la lumière des sections précédentes, plusieurs outils statistiques ont été utilisés pour repérer quantitativement les filières industrielles sur la base des tableaux d'entrées-sorties. Les types de liens économiques considérés varient également selon les auteurs : certains ayant privilégié les achats, d'autres les ventes, d'autres une combinaison des deux et d'autres encore, les coefficients techniques de

production. Voulant comparer les résultats obtenus avec diverses méthodologies, **Rey et Mattheis (2000)** ont testé quelque 171 applications concernant l'analyse de classification ou l'analyse des composantes principales/analyse factorielle avec les tableaux d'entrées-sorties de la Californie (pour l'année 1995, à un niveau d'agrégation de 61 secteurs comprenant toutes les branches de l'économie). En ce qui concerne l'analyse de classification, ils ont utilisé quatre algorithmes, chacun appliqué sur cinq matrices de similitude différentes. Pour chaque implantation de l'analyse, ils ont eu recours à huit différents seuils de similarité pour « couper » le dendrogramme afin d'étudier la sensibilité du nombre final de filières par rapport à cette décision que doit prendre le chercheur.

Ils se sont servis de six matrices de corrélation pour ce qui a trait à l'analyse en composantes principales et ont utilisé six valeurs minimales pour les poids factoriels, dans chaque cas, afin d'examiner l'impact du choix de ces valeurs sur la détermination du nombre de secteurs associés aux filières identifiées. Le tableau qui suit résume les diverses méthodes testées.

Tableau 3.4 Description des méthodologies testées par Rey et Mattheis

Analyse de classification (135 applications)	ACP-AC (36 applications)
<p>Matrices utilisées :</p> <ul style="list-style-type: none"> <li>Proportion des flux intermédiaires<sup>1</sup></li> <li>Matrice inverse de Léontief</li> <li>Coefficients d'achats</li> <li>Coefficients de ventes</li> <li>Transactions intermédiaires</li> </ul> <p>Types d'analyse et algorithmes :</p> <ul style="list-style-type: none"> <li>Hiérarchique ascendante, Ward</li> <li>Hiérarchique ascendante, Average</li> <li>Hiérarchique « divisive », Dianna<sup>2</sup></li> <li>Partitioning around the Medoids (PAM)<sup>3</sup></li> </ul> <p>Niveau de similarité pour « couper » le dendrogramme : 8 valeurs différentes (sauf dans le cas de PAM<sup>3</sup>)</p>	<p>Matrices utilisées :</p> <ul style="list-style-type: none"> <li>Coefficients d'achats</li> <li>Coefficients de ventes</li> <li>Proportion des flux intermédiaires<sup>1</sup></li> <li>Maximum des 4 coeff. de Czamanski calculés par rapport à la production totale plutôt que par rapport aux achats et ventes intermédiaires</li> <li>Maximum des 4 coeff. de Czamanski</li> <li>Matrice inverse de Léontief</li> </ul> <p>Seuil minimal des poids factoriels : 6 valeurs différentes</p>
<p><sup>1</sup> Il s'agit, pour une paire d'industries, de la somme des achats que les secteurs font auprès de l'autre, divisée par la somme des ventes totales et des achats totaux de deux secteurs.</p> <p><sup>2</sup> À la connaissance des auteurs, c'est la première fois qu'une analyse de classification d'un autre type que l'analyse hiérarchique ascendante était utilisée pour identifier des filières industrielles.</p> <p><sup>3</sup> Les résultats de la méthode du PAM nécessitent de la part du chercheur de spécifier <i>a priori</i> un nombre de filières. Les résultats de ce type d'analyse ne sont pas présentés sous la forme d'un dendrogramme.</p>	

Rey et Mattheis ont comparé les résultats obtenus selon trois types de mesure :

- La taille des filières définies. Les auteurs ont évalué le nombre d'emplois et l'importance de la production de chaque filière. De même, la taille des filières a été évaluée en fonction du nombre de secteurs industriels associés à chacune. Enfin, les auteurs ont noté le nombre total de secteurs ayant été attribués à une filière (parmi les 61 secteurs originels) par chacune des méthodes;

- L'importance des liens internes. Les proportions des achats et des ventes effectués par les membres d'une filière auprès des autres membres ont été calculées;
- L'importance des liens externes. Les auteurs ont mesuré l'importance des liens en amont et en aval des membres d'une filière avec tous les autres secteurs industriels de l'économie en élaborant deux indices. Lorsque supérieur à l'unité, l'indice des liens externes en *amont* révèle que l'augmentation de la demande finale pour les secteurs d'une filière génère une demande pour l'économie régionale supérieure à la moyenne<sup>13</sup>. De même, un indice des liens externes en *aval* supérieur à l'unité montre que les membres d'une filière sont davantage sensibles que la moyenne à une augmentation de la demande finale pour l'ensemble de l'économie régionale.

Le tableau qui suit présente un extrait des résultats obtenus avec l'analyse en composantes principales, avec le seuil minimal des poids factoriels fixé à 0,50. On peut en tirer un certain nombre de conclusions (soulignons que ces dernières s'appliquent aux tableaux d'entrées-sorties de la Californie pour 1995 et ne sont pas nécessairement généralisables). D'abord, les deux types de coefficients de Czamanski (calculés soit par rapport à la production totale, soit par rapport aux échanges intermédiaires seulement) donnent des résultats similaires. Ensuite, la matrice constituée des proportions de flux intermédiaires de même que la matrice inverse de Léontief groupent peu de secteurs (respectivement 17 et 18 sur un total de 61), bien que ce soit cette dernière matrice qui forme le plus grand nombre de filières (9). Ce sont les matrices basées sur les coefficients de Czamanski qui associent le plus grand nombre de secteurs à une filière (42). Enfin, les liens internes des filières repérées à l'aide des coefficients d'achats sont les plus élevés, de même que le lien externe de sensibilité en aval.

Soulignons que Rey et Mattheis n'ont pas comparé la composition sectorielle des filières identifiées par chaque méthode.

Tableau 3.5 Extrait des résultats de Rey et Mattheis concernant l'analyse en composantes principales (seuil minimal des poids factoriels fixé à 0,50)

Matrice analysée	Taille			Liens internes		Liens externes	
	Nombre de filières	Nombre de secteurs « groupés »	Nombre moyen de secteurs par filière	Prop. achats intra-filière	Prop. ventes intra-filière	Dispersion en amont	Sensibilité en aval
Coefficients d'achats	6	34	5,67	0,58	0,53	0,96	1,42
Coefficients de ventes	6	33	5,50	0,46	0,50	0,98	1,29
Prop. des flux interm.	8	17	2,13	0,32	0,34	0,98	1,39
Coeff. de Czamanski*	7	42	6,00	0,48	0,44	0,96	1,35
Coeff. de Czamanski	7	42	6,00	0,48	0,44	0,96	1,35
Inverse de Léontief	9	18	2,00	0,21	0,20	1,01	1,18

\* Coefficients calculés par rapport à la production totale plutôt que par rapport aux achats et ventes intermédiaires.

La comparaison de l'ensemble des résultats révèle que l'analyse de classification a tendance à générer un petit nombre de « grosses » filières, lorsque appliquée aux matrices de coefficients d'achats, de ventes ou de transactions intermédiaires (entre 1 et

<sup>13</sup> Le calcul des liens externes de Rey et Mattheis est présenté à l'annexe 2.

4 seulement, selon le seuil utilisé pour « couper » le dendrogramme). Rey et Mattheis soulignent que les filières identifiées à l'aide de l'analyse de classification enregistrent de plus forts liens internes, alors que celles qui proviennent de l'analyse en composantes principales ont des liens externes plus importants.

Ne trouvant pas qu'une méthode soit supérieure à une autre et croyant que chacune, avec ses forces et ses faiblesses, apporte un éclairage particulier sur les liens interindustriels d'une économie, Rey et Mattheis ont élaboré la « méthodologie du consensus » (*consensus clustering method*) pour incorporer les résultats de plusieurs types d'analyse :

« [...] rather than relying on a single method that may focus on only one type of interindustry linkage in deciding whether two industries should be placed in the same cluster, the consensus clustering approach considers how each pair of industries was viewed by multiple identification methods. This allows for a multidimensional view of the nature of interindustry linkages within clusters. »<sup>14</sup>

La méthode repose sur l'élaboration d'une matrice de similitude qui reflète la propension de chaque paire de secteurs à être groupée dans une même filière par diverses méthodes. Une analyse de classification hiérarchique ascendante est appliquée à cette matrice, avec l'algorithme de Ward.

### **Utilisation de la méthode**

Se servant de la « méthode du consensus », Rey et Mattheis ont repéré 35 filières pour l'État de la Californie avec des tableaux d'entrées-sorties désagrégés pour l'année 1997. Ils ont limité leur analyse aux seuls secteurs affichant un coefficient de localisation supérieur à 1 (soit 149 industries), voulant mettre l'emphase sur les secteurs étant plus susceptibles d'exporter à l'extérieur de l'État. En tout, ils ont appliqué 81 variations de l'analyse de classification et de l'analyse en composantes principales en vue de bâtir leur matrice de similitude finale. Les combinaisons analytiques qu'ils ont utilisées sont décrites au tableau qui suit.

---

<sup>14</sup> Rey et Mattheis (2000), vol.I, p. iii.

Tableau 3.6 Méthodes employées par Rey et Mattheis pour bâtir la matrice de similitude de la méthode du consensus

Analyse de classification (69 variations)	ACP-AC (12 variations)
<p>Matrices utilisées :</p> <ul style="list-style-type: none"> <li>Maximum des 4 coeff. de Czamanski</li> <li>Maximum des coeff. d'achats et de ventes</li> <li>Proportion des flux intermédiaires</li> </ul> <p>Types d'analyse et algorithmes :</p> <ul style="list-style-type: none"> <li>Hiérarchique ascendante, Ward</li> <li>Hiérarchique ascendante, Average</li> <li>Hiérarchique ascendante, Complete</li> <li>Hiérarchique descendante</li> <li><i>Partitioning</i></li> </ul> <p>Niveau de similarité pour « couper » le dendogramme : 5 valeurs différentes</p>	<p>Matrices utilisées :</p> <ul style="list-style-type: none"> <li>Maximum des 4 coeff. de Czamanski</li> <li>Maximum des coeff. d'achats et de ventes</li> <li>Proportion des flux intermédiaires</li> </ul> <p>Seuil minimal des poids factoriels :</p> <p>0,35</p> <p>Aucun (association du secteur à la filière pour laquelle le poids factoriel est le plus élevé)</p>

Rey et Mattheis ont « coupé » le dendogramme des résultats finaux à un niveau de similarité qui leur semblait générer un bon équilibre entre le nombre de filières et la grosseur de ces dernières. Les filières ne contenant qu'un seul secteur n'ont pas été gardées. En tout, 114 secteurs ont été associés à l'une des 35 filières retenues.

### Avantages et inconvénients

La méthode du consensus permet d'étudier le regroupement des secteurs sous plusieurs angles, ce qui constitue un avantage très séduisant par rapport aux autres méthodes couvertes. La méthode comporte toutefois sa part d'inconvénients :

- Ainsi qu'appliquée par Rey et Mattheis (2000), la méthode incorpore énormément de renseignements, dont tous ne sont pas nécessairement pertinents. Par exemple, on peut se questionner sur le choix des auteurs de générer une série de filières, avec l'analyse des composantes principales qui regroupe l'ensemble des secteurs industriels, incluant ceux dont le degré d'association à une filière s'avère très faible (l'un des scénarios des auteurs n'incluant aucun seuil minimal pour les poids factoriels). Nous pensons que cela risque de mettre du « bruit » dans le système et d'avoir un impact négatif sur le décodage des liens importants.
- Il est difficile de savoir, en bout de ligne, pour quelle raison un secteur appartient à une filière : est-ce à cause d'une similitude sur le plan du profil d'achats, ou de celui des ventes?

Tableau 3.7 Synthèse comparatif des avantages et inconvénients des méthodes couvertes

Méthode	Avantages	Inconvénients
<p><b>ACP/AF</b></p> <p>Roepke et autres (1974) Czamanski (1974) Feser et autres (2000) Rey et Mattheis (2000)</p>	<ol style="list-style-type: none"> <li>1. L'ACP/AF identifie les secteurs qui ont des liens complémentaires, c'est-à-dire qui partagent des bassins d'intrants ou des marchés similaires.</li> <li>2. Bien que la source de complémentarité des secteurs groupées ne fasse pas toujours partie d'une filière, il est possible de l'identifier en inspectant la matrice des flux interindustriels.</li> <li>3. Les valeurs propres et les poids factoriels donnent une mesure du degré d'importance des liens des secteurs groupés.</li> <li>4. Les poids factoriels indiquent le degré de corrélation d'un secteur avec <i>chacune</i> des filières identifiées.</li> <li>5. Il existe des logiciels d'analyse statistique sophistiqués pour effectuer ce genre d'analyse.</li> </ol>	<ol style="list-style-type: none"> <li>1. L'ACP/AF est davantage appropriée pour repérer les secteurs complémentaires que pour regrouper ceux liés verticalement. Il faut inspecter la matrice des flux interindustriels <i>a posteriori</i>, une fois les composantes ou les facteurs extraits, pour repérer les secteurs étant à la source des regroupements (sur le plan des achats ou des ventes).</li> <li>2. L'interprétation des résultats peut être difficile si l'on a peu ou trop de composantes/ facteurs extraits.</li> </ol>
<p><b>Analyse de classification</b></p> <p>Caber et all (1991) Blin et Cohen (1977) Abbott et Andrews (1990) Rey et Mattheis (2000)</p>	<ol style="list-style-type: none"> <li>1. La méthode repère les secteurs qui ont des liens complémentaires, c'est-à-dire qui partagent des bassins d'intrants ou des marchés similaires.</li> <li>2. La méthode génère des filières qui sont mutuellement exclusives.</li> <li>3. La présentation des résultats sous la forme d'un dendogramme facilite leur interprétation, puisqu'on peut visualiser le niveau de similitude des secteurs formant une filière.</li> <li>4. Il existe des logiciels d'analyse statistique sophistiqués pour effectuer ce genre d'analyse.</li> </ol>	<ol style="list-style-type: none"> <li>1. La méthode est davantage appropriée pour repérer les secteurs complémentaires que pour regrouper ceux liés verticalement. Il faut inspecter la matrice des flux interindustriels <i>a posteriori</i> pour repérer les secteurs étant à la source des regroupements (au niveau des achats ou des ventes).</li> <li>2. Dans les faits, un secteur industriel est souvent associé à plus d'une filière. Or, la méthode ne relie un secteur qu'à une seule filière; contrairement à l'ACP/AF, on n'obtient pas une mesure de l'importance du lien qu'a un secteur avec chacune des filières identifiées.</li> <li>3. Il semble qu'avec les matrices de coefficients d'achats et de ventes, l'analyse de classification forme peu de filières, souvent très grosses.</li> <li>4. Jusqu'à ce jour, la méthode a été peu utilisée pour le repérage <i>per se</i> de filières industrielles. Nous ne pouvons donc nous référer qu'à un ensemble restreint d'applications.</li> </ol>

<b>Méthode</b>	<b>Avantages</b>	<b>Inconvénients</b>
<p><b>Théorie des graphes</b></p> <p>Campbell (1972) Slater (1977) Hauknes (1999)</p>	<p>1. La méthode lie les secteurs ayant directement des liens entre eux.</p>	<p>1. On perd énormément de renseignements en menant ce type d'analyse.</p> <p>2. La méthode ne produit pas l'équivalent des valeurs propres et des poids factoriels, qui donnent une idée de l'importance du degré d'association des secteurs.</p> <p>3. Le chercheur doit souvent poser un jugement qualitatif sur la composition des filières et déterminer les frontières de celles-ci.</p>
<p><b>Méthode du Maxima</b></p> <p>Roelandt et autres (1999) Peeters et autres (2001)</p>	<p>1. La méthode lie les secteurs ayant directement des liens entre eux, et qui plus est des liens très forts.</p>	<p>1. On perd énormément de renseignements en menant ce type d'analyse.</p> <p>2. La méthode ne produit pas l'équivalent des valeurs propres et des poids factoriels, qui donnent une idée de l'importance du degré d'association des secteurs.</p> <p>3. Le chercheur doit souvent poser un jugement qualitatif sur la composition des filières et déterminer les frontières de celles-ci.</p>
<p><b>Consensus Clustering</b></p> <p>Rey et Mattheis (2000)</p>	<p>1. La méthode permet d'étudier le regroupement des secteurs sous plusieurs angles.</p> <p>2. Il existe des logiciels d'analyse statistique sophistiqués pour effectuer ce genre d'analyse.</p>	<p>1. Ainsi qu'appliquée par Rey et Mattheis (2000), la méthode incorpore énormément de renseignements, dont tous ne sont pas nécessairement pertinents. Cela risque de mettre du « bruit » dans le système.</p> <p>2. Il est difficile de savoir, en bout de ligne, pour quelle raison un secteur appartient à une filière (achats? ventes?)</p>

