Traitement de l'information : analyse des données classiques versus analyse réseau. Un cas d'application : la bibliométrie

Eric Boutin, Luc Quoniam, Hervé Rostaing, Philippe Dumas

TRAITEMENT DE L'INFORMATION: ANALYSE DE DONNEES CLASSIQUES VERSUS ANALYSE DE RESEAU. UN CAS D'APPLICATION: LA BIBLIOMETRIE.

BOUTIN E*., QUONIAM L**., ROSTAING H**., DUMAS Ph*.

* Centre de Recherche Le Pont, IUT, TC, BP 132 Université de Toulon 83957 La Garde email boutin@univ-tln.fr
** CRRM, Fac. St Jérôme 13397 Marseille cedex 20.

L'analyse en terme de réseau a fait l'objet d'un certain nombre de recherches, dans le domaine des sciences humaines et sociales. Les principaux résultats sont présentés dans l'ouvrage de Wasserman (1994).

L'objet de ce travail est de porter à la connaissance de la communauté académique une pratique originale de traitement de l'information qui utilise un graphe appelé réseau pour représenter une information complexe de façon synthétique. Cette technique est tout d'abord positionnée par rapport aux « analyses de données classiques » et fait ensuite l'objet d'une application dans le cas particulier du traitement de données bibliométriques.

Les données bibliométriques possèdent trois caractéristiques principales qui fondent leur spécificité:

- L'information à analyser est massive. La masse d'information à traiter suggère l'utilisation d'outils qui permettent le traitement automatique de cette information.
- L'information à analyser est structurée autour de champs homogènes: champ auteur, mots-clés, résumé... Cette caractéristique la distingue des techniques d'analyse de l'information en texte intégral qui raisonnent directement sur une information brute.
- L'information à analyser est obtenue à la suite de l'interrogation de banques de données internationales. La nature de cette information est très variée puisqu'il peut s'agir aussi bien d'une information scientifique, technique, technico-économique... Cette troisième caractéristique permet d'établir la spécificité de ce type d'information par rapport à l'information obtenue à la suite d'enquêtes.

Dans ce contexte particulier de l'analyse bibliométrique, la problématique générale du traitement de l'information est la suivante. A la suite de l'interrogation d'une banque de données, l'analyseur collecte une masse d'information brute prenant la forme d'un ensemble de notices. A ce stade, l'information est « inerte ». Pourtant, cette information colletée est le fruit d'une dynamique humaine qui sous tend une dynamique d'idées. Ces documents sont en effet le résultat d'un processus interactionniste, coopératif voir concurrentiel entre des acteurs qui ont développé ces idées et publié leurs résultats. Or ce processus échappe en grande partie à l'individu qui collecte des documents. Seul l'expert du domaine, qui a participé ou observé cette dynamique, a le recul nécessaire pour dégager de l'ensemble des idées générales. Il serait intéressant de faire revivre cette dynamique, d'activer la dimension humaine sous-jacente en visualisant la structure du phénomène. Cette structure est une nébuleuse qui peut être analysée sous plusieurs plans en interaction comme par exemple: le plan des acteurs, celui des idées et celui du temps.

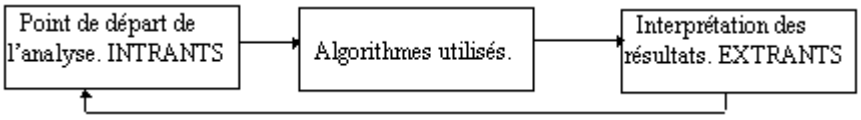
L'approche de la représentation de l'information sous forme de réseau se prête particulièrement à ce genre de finalité même si d'autres techniques d'analyse de données peuvent être utilisées. Ces autres techniques d'analyses qui débouchent sur des représentations cartographiques d'un ensemble d'information seront désignées par la suite sous le terme générique « d'analyses de données classiques ». Le corps du document reprendra trois étapes permettant de mettre en évidence les caractéristiques spécifiques de l'approche réseau par rapport à l'approche « analyse de données

classique ». Cette comparaison de méthodes prolonge la réflexion de Tijssen (1989). Nous conclurons en montrant en quoi les deux démarches sont complémentaires.

Pour faciliter la clarté de l'exposé, nous proposons de suivre un même cas d'application tout au long de cette présentation. Ainsi, nous sommes-nous intéressés à un corpus de 1191 références télédéchargées de la base de données Pascal. Ce corpus traite de la bibliométrie. Il a été obtenu à la suite de l'équation logique « Bibliometric? or scientometric? or informetric? ». Ce choix a été guidé par deux motifs principaux. Le premier est que les techniques bibliométriques nous étant familières, nous pouvons jouer le rôle à la fois d'analyseurs et d'experts capables de valider les résultats obtenus. D'autre part, les résultats de l'étude ne sauraient être préjudiciables à un intérêt privé comme cela aurait pu l'être si nous avions analysé les brevets d'une entreprise dans une problématique de veille concurrentielle.

Toute méthode d'analyse de données peut s'exprimer de façon simple en répondant aux questions suivantes:
- Quelles sont les données qui sont «injectées » dans l'analyse?
- Quels sont les algorithmes qui permettent de traiter ces données?
- Quels sont les résultats de l'analyse et quelles interprétations peut-on en faire?

Le système récursif évoqué ci dessous suggère un processus d'aller retour entre inputs et outputs en fonction des outputs obtenus.
Seule la partie interprétation des résultats nous intéresse véritablement ici. Toutefois, pour être capable d'interpréter correctement les résultats obtenus, il est nécessaire de s'intéresser aux deux premières étapes.



I- Point de départ de l'analyse:

a- Un point de départ commun:

L'analyse en terme de réseau, comme l'analyse de donnée classique, a pour point de départ un ou plusieurs champs d'un ensemble de notices bibliographiques. Nous allons exposer la démarche en s'intéressant au « plan des idées » en analysant le champs « Descripteur Français » d'un ensemble de notices. Ce champ représente les concepts majeurs abordés dans les documents.

Nous allons illustrer la démarche et raisonner sur les 5 premiers articles de la base de travail. Les articles sont décrits par un ensemble de 12 mots-clés.

DF - Loi de probabilité; Loi Lotka; Loi Bradford; Etude comparative
DF - Loi Lotka; concentration; Indice de Gini
DF - Entropie; théorie information; Classification
DF - Entropie; théorie Shannon
DF - classification; représentation graphique; réseau

tableau 1: descripteurs français des 5 premières notices de la base

Cette information peut être présentée sans perte de sens sous forme de matrice binaire. Cette matrice retrace la présence ou l'absence de chaque mot-clé pour chacune des références étudiées.

	Article 1	Article 2	Article 3	Article 4	Article 5
Entropie	0	0	1	1	0
Classification	0	0	1	0	1
Loi lotka	1	1	0	0	0
Théorie Information	0	0	1	0	0
Théorie Shannon	0	0	0	1	0
Réseau	0	0	0	0	1
Représentation graphique	0	0	0	0	1
Concentration	0	1	0	0	0
Loi Probabilité	1	0	0	0	0
Loi Bradford	1	0	0	0	0
Indice Gini	0	1	0	0	0
Etude comparative	1	0	0	0	0

tableau 2: matrice binaire associée

b- La divergence des méthodes:

Une fois la matrice binaire construite, l'objectif est de mesurer l'intensité du lien entre les variables prises deux à deux. Ceci va se traduire par la construction de matrices carrées symétriques. De telles matrices peuvent être construites automatiquement à l'aide du logiciel Dataview développé au CRRM par Rostaing (1993). C'est à ce niveau que les analyses de données classiques et l'analyse de réseau divergent.

- Les analyses de données classiques mesurent l'intensité du lien en ayant recours à des **indices d'association**. Ces indices statistiques sont nombreux. Nous raisonnerons ici sur l'indice de Jaccard, souvent utilisé, et illustrerons le principe en mesurant le lien entre les deux premiers mots-clés au sens de cet indice. L'état des relations entre ces deux mots-clés peut se résumer par le tableau élémentaire suivant:

		« Entropie »	
		Présence	Absence
« Classification »	Présence	N_A	N_B
	Absence	N_C	N_D

N_A désigne le nombre d'articles dans lesquels ces deux mots-clés sont co-présents. Ici, il sont co-présents dans l'article 3 donc une fois: Ici, $N_A=1$

N_B désigne le nombre d'articles qui comportent « classification » comme descripteur mais qui ne comportent pas « Entropie ». Ici, $N_B=1$.

En suivant le même principe on trouve $N_C=1$ et $N_D=2$

L'indice de Jaccard apprécie l'association entre les deux mots-clés par l'expression: $\frac{N_A}{N_A + N_B + N_C}$ ce qui donne un indice de 0.33 dans l'exemple considéré.

De proche en proche, il est possible de construire la matrice de Jaccard qui traduit l'ensemble des associations entre mots-clés. Celle ci est présentée Tableau 3.

	entropie	classification	lo lotka	théorie information	théorie shannon	réseau	représentation graphique	concentration	loi de probabilité	loi bradford	indice de gini	etude comparative
entropie	1,00	0,33	0,00	0,50	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00
classification	0,33	1,00	0,00	0,50	0,00	0,50	0,50	0,00	0,00	0,00	0,00	0,00
loi lotka	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,50	0,50	0,50	0,50	0,50
théorie information	0,50	0,50	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
théorie shannon	0,50	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
réseau	0,00	0,50	0,00	0,00	0,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00
représentation graphique	0,00	0,50	0,00	0,00	0,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00
concentration	0,00	0,00	0,50	0,00	0,00	0,00	0,00	1,00	0,00	0,00	1,00	0,00
loi de probabilité	0,00	0,00	0,50	0,00	0,00	0,00	0,00	0,00	1,00	1,00	0,00	1,00
loi bradford	0,00	0,00	0,50	0,00	0,00	0,00	0,00	0,00	1,00	1,00	0,00	1,00
indice de gini	0,00	0,00	0,50	0,00	0,00	0,00	0,00	1,00	0,00	0,00	1,00	0,00
etude comparative	0,00	0,00	0,50	0,00	0,00	0,00	0,00	0,00	1,00	1,00	0,00	1,00

Tableau 3: Matrice de Jaccard associée aux 5 premières références

Le recours à ce type d'indice peut être rapidement discuté:
Ces indices permettent de mesurer l'intensité de la relation entre deux mots-clés en ne prenant pas seulement en considération l'importance du lien entre lesdits mots-clés mais aussi le poids statistique de chacun d'eux. Ainsi, une association unique de deux mots-clés au sein d'un corpus n'aura pas le même sens si ces deux mots-clés sont rares ou s'ils sont très fréquents.
De plus, l'indice de Jaccard respecte le principe de métrique qui est à la base de toute analyse de données classique.

Toutefois, le recours à ce genre d'indices d'association soulève un certain nombre de problèmes. D'une part, la multiplicité des indices potentiels fait qu'il est difficile de maîtriser la logique à l'oeuvre: pour pouvoir interpréter les résultats obtenus à partir d'un indice particulier, une solide expérience empirique est recommandée.

D'autre part et surtout, le recours à de telles matrices de distance dénature le lien réel entre les mots-clés. Un biais est introduit en ce sens que ce n'est plus l'information brute de départ qui est manipulée mais un « construit » mathématique. Ainsi, à partir de la matrice présentée Tableau 3, on pourra dire, que les deux termes « théorie de shannon » et « entropie » sont plus proches au sens de l'indice de Jaccard que les termes « entropie » et « classification ». On parvient à positionner ces trois mots-clés relativement les uns par rapport aux autres mais pas dans l'absolu. Il y a donc une perte d'information entre matrice binaire et matrice de distance.

- Dans l'approche en terme de réseaux, on privilégie au contraire une démarche plus pragmatique en raisonnant tout le long de la chaîne de traitement, sur l'information brute initiale contenue dans la matrice binaire. L'information brute est traduite sous forme de deux matrices carrées symétriques complémentaires.

La première exprime, pour chaque couple de mots-clés le nombre de références qu'ils décrivent conjointement. En reprenant l'exemple ci dessus, on obtient la matrice présentée tableau 4.

	entropie	classification	lo lotka	théorie information	théorie shannon	réseau	représentation graphique	concentration	loi de probabilité	loi bradford	indice de gini	etude comparative
entropie	2	1	0	1	1	0	0	0	0	0	0	0
classification	1	2	0	1	0	1	1	0	0	0	0	0
loi lotka	0	0	2	0	0	0	0	1	1	1	1	1
théorie information	1	1	0	1	0	0	0	0	0	0	0	0
théorie shannon	1	0	0	0	1	0	0	0	0	0	0	0
réseau	0	1	0	0	0	1	1	0	0	0	0	0
représentation graphique	0	1	0	0	0	1	1	0	0	0	0	0
concentration	0	0	1	0	0	0	0	1	0	0	1	0
loi de probabilité	0	0	1	0	0	0	0	0	1	1	0	1
loi bradford	0	0	1	0	0	0	0	0	1	1	0	1
indice de gini	0	0	1	0	0	0	0	1	0	0	1	0
etude comparative	0	0	1	0	0	0	0	0	1	1	0	1

Tableau 4: Matrice des mots-clés.

Les termes « Entropie » et « Classification » décrivent ensemble 1 article.
Les valeurs de cette matrice sont beaucoup plus facilement interprétables que celles contenues dans la matrice de Jaccard présentée Tableau 3. Le même poids est affecté à chaque lien quelle que soit la fréquence de chacun des mots-clés de l'association.
Néanmoins, le passage de la matrice binaire à la matrice du tableau 4 se traduit par une perte d'information. En effet, la matrice du tableau 4 fournit la fréquence de la co-présence de deux termes mais on ne sait plus dans quels articles ces deux termes sont co-présents.
Pour cette raison, on peut introduire une seconde matrice symétrique, appelée matrice de Condorcet qui indique pour chaque couple d'articles le nombre de mots-clés en commun qui les décrivent. La matrice associée à ce genre de préoccupation est présentée Tableau 5.

	Article 1	Article 2	Article 3	Article 4	Article 5
Article 1	4	1	0	0	0
Article 2	1	3	0	0	0
Article 3	0	0	3	1	1
Article 4	0	0	1	2	0
Article 5	0	0	1	0	3

Tableau 5: matrice des relations entre articles.

Les articles 1 et 2 ont un mot-clé en commun.

Les matrices présentées tableau 4 et 5 préservent l'intégralité de l'information contenue dans la matrice binaire. Mathématiquement l'utilisation de l'une ou de l'autre se traduit par la résolution du problème primal ou dual. Dans les faits, la matrice mots-clés mots-clés sera utilisée pour faire ressortir la structure des idées contenues dans les documents tandis que la matrice article-article permettra de regrouper les articles décrits par des mots-clés identiques et correspond plus à une problématique de classement automatique de documents.
La démarche de l'analyse réseau est séduisante en ce sens qu'elle conserve la totalité de l'information contenue dans la matrice binaire d'origine.

II- Les algorithmes utilisés:

l'objectif n'est pas de rentrer dans le détail du fonctionnement des algorithmes mais d'en présenter le principe général pour être à même d'interpréter les résultats de chaque analyse. Le lecteur intéressé pourra se rapporter à un article antérieur de Boutin (1995-1).
La représentation de l'information contenue dans les matrices 3, 4, 5 se heurte au même écueil. Chacune de ces matrices pourrait être représentée par un polyèdre à n sommets dans un espace à n-1 dimensions, n représentant la taille de la matrice considérée. Toutefois, une telle représentation est d'une part impraticable et d'autre part illisible car on ne peut concevoir de représentation claire qu'en 2 voire 3 dimensions.
A ce stade, chaque technique d'analyse de données est particulière. L'analyse du « cadrage multidimensionnel des données » (MDS) par exemple a pour objectif d'arriver à trouver par un processus itératif les coordonnées des différents sommets dans le plan telles que les distances entre ces sommets soient les plus proches possibles des distances calculées dans une matrice d'association du type de celle présentée tableau 3. Les techniques d'analyse d'inertie (AFC, ACP, AFCM) se traduisent par la recherche des dimensions qui rendent le mieux compte de la dispersion du nuage.
L'analyse en terme de réseau cherche à construire un graphe particulier appelé réseau qui satisfasse certaines contraintes esthétiques. Un réseau est composé de sommets représentant dans notre cas les mots-clés et d'arcs déterminant l'intensité du lien entre les sommets pris deux à deux. La figure 1 reprend l'exemple du réseau de relations entre les douze mots-clés présentés ci dessus.
Le réseau figurant sur la partie gauche est plus lisible que celui de droite. Cette lisibilité supérieure est appréciée à travers une **fonction d'esthétisme** qu'il s'agit de maximiser. Un réseau sera d'autant plus esthétique que:
- le nombre d'intersection entre deux sommets est limité.
- Le nombre d'intersection entre les arêtes est limité.
- Le nombre d'intersection arêtes-sommets est limité.

Certains types d'algorithmes permettent de faire converger le réseau vers une solution esthétiquement supérieure: il en va ainsi des algorithmes génétiques dont on trouvera une présentation dans le rapport de Aubry (1993) ou de la technique du recuit simulé qui fait l'objet d'une application par Davidson (1989) dans le domaine du tracé automatique de graphe.

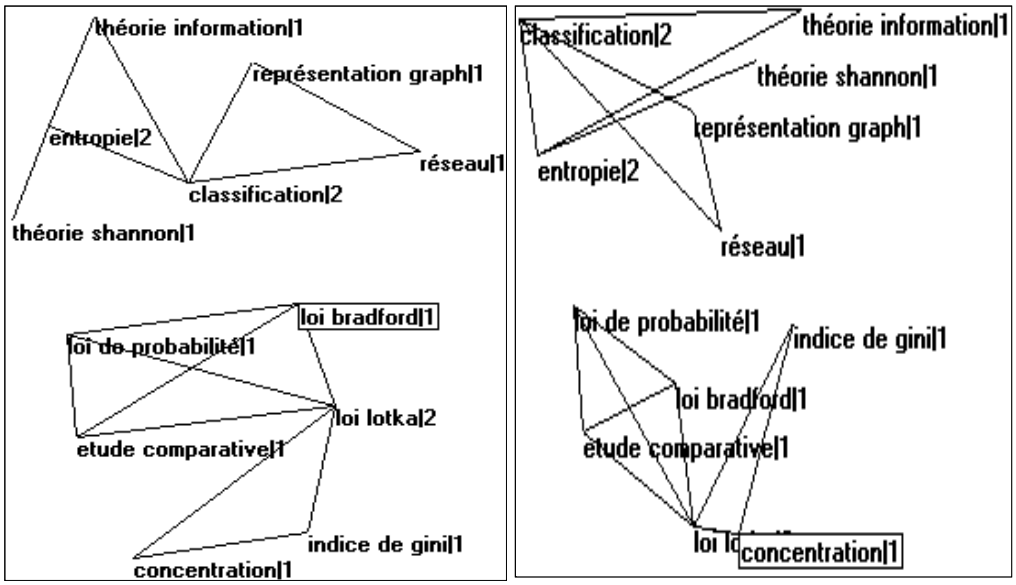


Figure 1: Deux exemples de réseaux de relation entre douze mots-clés.

III- Les résultats et leur interprétation:

Nous avons choisi de construire le réseau des mots-clés associé aux 1191 références de la base de travail. L'information contenue dans le champ mot-clé est extraite puis analysée automatiquement à travers le module de multidimentionnal scaling (MDS) d'un logiciel commercial[1]. Parallèlement, le même jeu de données est exploité à travers le logiciel Matrisme[2] qui permet la construction automatique de réseaux. (Figure 3).

Une comparaison des résultats est établie sur la base de la qualité et de la facilité d'interprétation des résultats obtenus.

[1] Statistica développé par la société STATSOFT

[2] @ Le Pont et CRRM

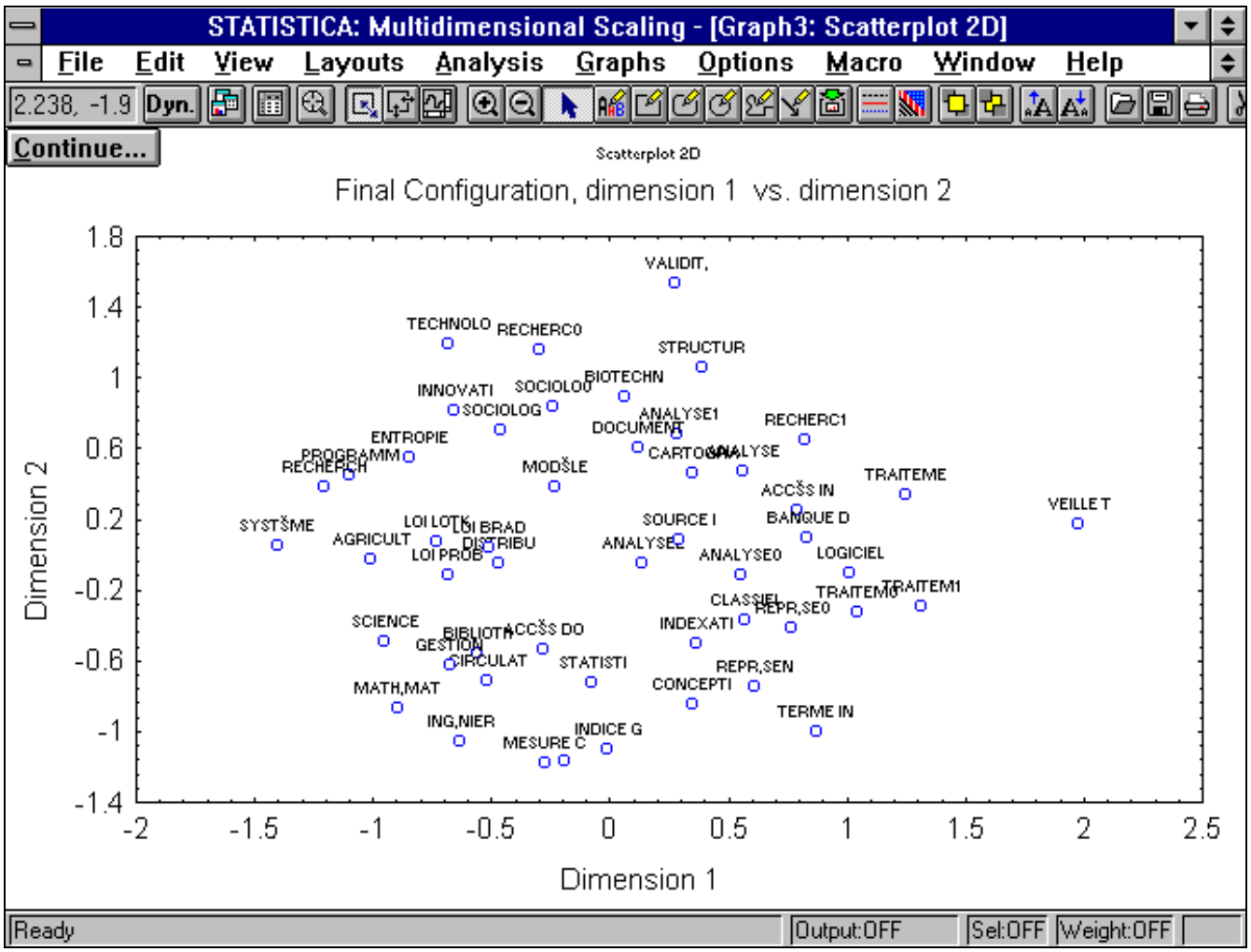


Figure 2: résultat de la MDS

Comme dans la MDS, nous avons choisi, pour ne pas surcharger le réseau présenté figure 3, de représenter tous les mots-clés apparaissant au moins cinq fois dans le corpus. De plus, un lien entre deux mots-clés est visualisé si les deux variables sont utilisées ensemble pour définir au moins trois références.

La valeur située à droite du libellé indique le nombre de fois où le mot-clé apparaît dans le corpus.

L'analyse de la MDS propose une disposition des modalités sur le plan telle que l'intensité du lien entre deux modalités est appréciée par leur distance.

Dans une représentation sous forme de réseau, la coprésence de deux mots-clés dans le même article va se traduire par l'existence d'un lien physique entre les deux sommets en cause. Ainsi, la notion de distance n'a plus l'importance qu'elle avait dans les analyses de données classiques.

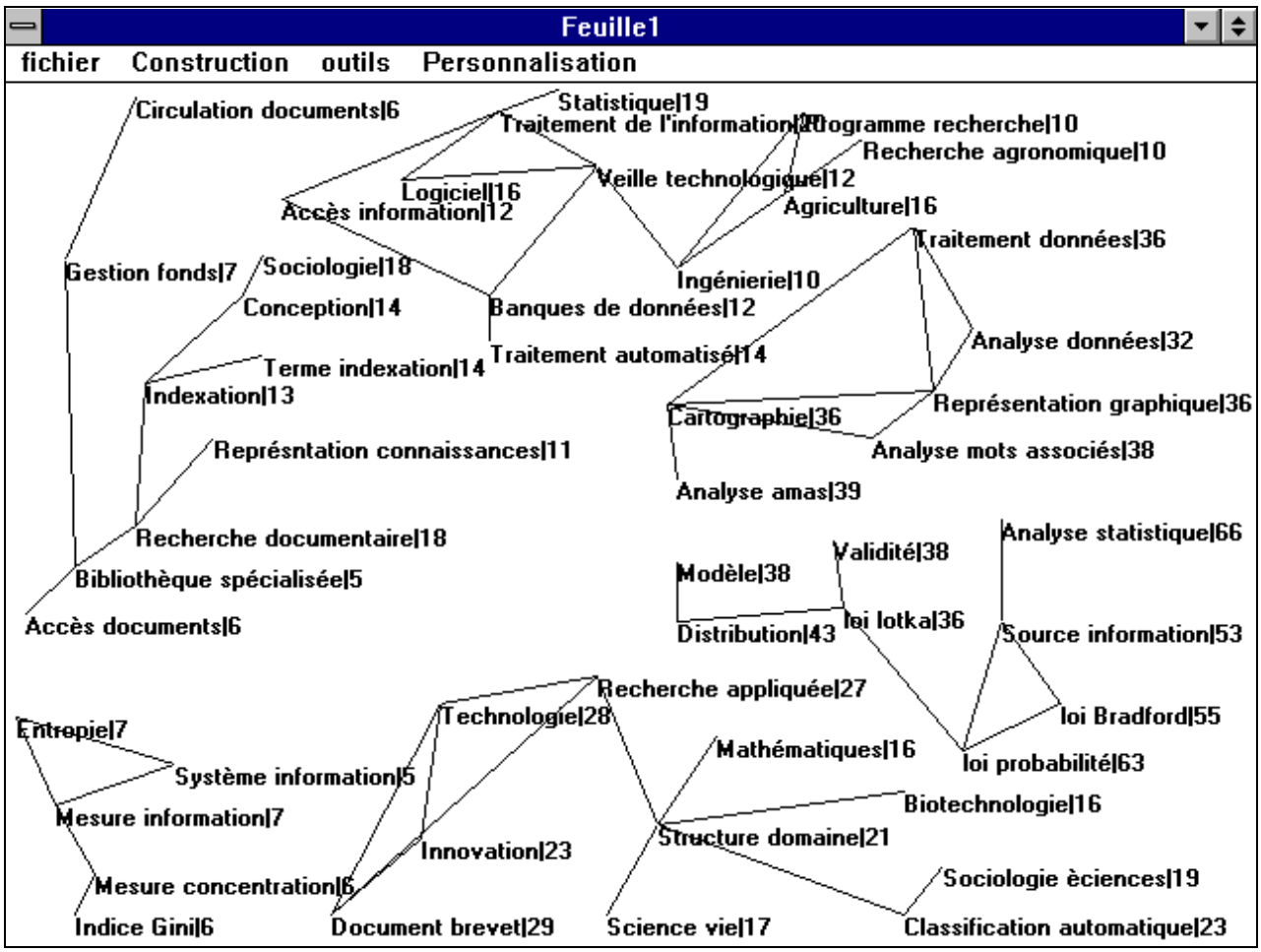


Figure 3: Résultat de l'analyse réseau

Cette représentation sous forme de réseau est originale à plusieurs titres:

- Tout d'abord, l'analyse réseau propose un résultat brut qui peut être affiné par l'analyseur. Cette démarche va consister à éclaircir le réseau en éliminant l' « information triviale » et le « bruit statistique ». L'information triviale se compose des associations de mots-clés les plus fréquentes. Celles-ci constituent le dénominateur commun de l'ensemble considéré. En tant que tel, ce genre d'association apporte une information très générale. Enlever de tels liens du réseau permet de l'éclaircir. L'information qualifiée de bruit statistique est constituée de l'ensemble des associations entre mots-clés qui sont les moins fréquemment obtenues. Ces associations correspondent soit à une information émergente, soit à une information atypique. Le fait de supprimer de telles relations du réseau va accroître sa lisibilité. Des analyses entreprises à partir du concept d'entropie par Lhen (1995) ont permis de dégager des seuils permettant de séparer automatiquement ces types d'informations. Le choix judicieux de ces filtres va permettre de dégager la métastructure du réseau et conduire à une représentation plus visuelle et une interprétation facilitée. En acceptant de perdre une information identifiée, on peut ainsi gagner en signification. Ce processus itératif de modification du résultat de départ ne peut pas être appliqué à une MDS où le résultat, fruit d'un calcul de distance ne peut pas être altéré. Dans le cas présent, les filtres ont consisté à retenir les mots-clés présents au moins 5 fois dans le corpus et ayant une fréquence d'association de plus de 3.

- L'analyse réseau présente l'avantage d'une grande souplesse: en effet, par simple application de filtres spécifiques, l'analyse peut être conduite tant sur l'information émergente que sur l'information dite utile. Celle ci comporte l'ensemble des mots-clés intermédiaires entre les mots-clés qui apportent une information triviale et les mots-clés qui apportent une information qualifiée de « bruit statistique ». De tels filtres ne peuvent pas être appliqués sur une MDS puisque, par définition,

une telle analyse est conduite à partir d'une mesure d'association qui ne correspond pas aux fréquences réelles.

- L'analyse réseau conduit à une compréhension simple de la structure du phénomène étudié. En effet, s'il est difficile, à partir de la MDS présentée figure 2 d'identifier clairement les sous-groupes du graphe, l'analyse en terme de réseau est parfaitement appropriée à ce genre de chose. Le réseau présenté figure 3 fait ressortir 6 groupes de mots-clés « indépendants », selon les filtres appliqués correspondant à six axes majeurs de réflexion dans le domaine: *Axe bibliothèque & gestion documentaire; Veille technologique et ses applications; Construction automatique de cartographies; Lois statistiques; Analyse en terme de classification et brevet; Mesure de l'information*.

Bien davantage, au niveau des sous groupes identifiés eux mêmes, il est possible de faire ressortir le rôle particulier exercé par certains mots-clés, appelés isthmes. Aho (1987) propose un algorithme permettant d'extraire d'un graphe de tels sommets. Dans l'exemple de la figure 8, le mot-clé « veille technologique » est un isthme. Ce mot-clé se trouve au confluent de deux préoccupations: une préoccupation technique (les outils d'analyse de l'information) et une préoccupation ou apparaissent les applications de ce concept dans l'industrie. Ce mot-clé joue un rôle d'intermédiaire entre deux sous groupes.

- L'analyse réseau autorise également la prise en compte d'une analyse dynamique. Lorsqu'on effectue une analyse bibliométrique, celle ci prend place dans le temps. Il peut être intéressant de décomposer la période considérée en tranches qui sont analysées successivement. Le passage d'une période à l'autre se traduit dans l'analyse réseau par la suppression et le rajout de liens à partir d'un réseau où la position des sommets pourra être verrouillée. La dynamique du réseau sur plusieurs périodes est illustrée par Boutin (1995-2). Au contraire une analyse prenant en compte la distance ne permettra pas facilement ce genre de chose puisque c'est la distance entre sommets qui doit rendre compte de l'intensité des relations entre sommets et donc qui varie pour chaque période de temps. Les points se trouvent donc déplacés pour chaque période.

- Chaque réseau peut être modifié par l'utilisateur. En effet puisque la notion de distance n'a pas de sens, il est possible, à partir d'un réseau de départ de « personnaliser » ce réseau en le rendant conforme à sa propre fonction d'esthétisme par une opération de « glisser déplacer » réalisé manuellement grâce à la souris sur les noeuds du réseau. Cette modification du réseau par l'utilisateur ou l'analyseur garantit l'appropriation du réseau et donc la compréhension de la méthode et *in fine* son exploitation.

Conclusion:

Cet exposé a eu pour objectif de positionner une nouvelle technique d'analyse de données par rapport aux techniques existantes. L'analyse réseau occupe une position originale en ce sens qu'elle travaille sur l'information brute et conduit à des réseaux facilement interprétables par un utilisateur néophyte. Néanmoins, il ne faut pas considérer les deux approches comme compétitives mais comme complémentaires.

Il est en effet possible de plaquer sur une analyse de donnée classique les liens correspondant aux fréquences réelles observées. On obtient alors un graphe bénéficiant d'une double grille de lecture non redondante: les variables sont proches les unes les autres en fonction de l'indice d'association retenu et sont liées entre elles en fonction de l'intensité de leur lien. Cette pratique permet de s'affranchir des limites de l'une et de l'autre des méthodes:

- D'une part, il y a désambiguisation du graphe résultat dans la mesure où il est possible de faire ressortir des sous groupes au sein du graphe initial.
- D'autre part, il est possible de mesurer la perte d'information associée au graphe obtenu. En effet, à la différence de l'analyse réseau, les analyses de données classiques fournissent comme sous produit de calcul le pourcentage d'information restitué par la représentation. On dispose ainsi d'un indicateur permettant d'apprécier la pertinence de la représentation.

- Enfin cette analyse mixte permet d'analyser les relations inter-groupes et non plus seulement intra-groupes ce qui n'est pas possible directement dans l'analyse réseau. En effet, dans le réseau de la figure 3, les six axes de recherche du domaine apparaissent comme parfaitement disjoints alors qu'ils sont en réalité imbriqués l'un dans l'autre. En plaquant sur le graphe obtenu par la MDS les relations entre les variables, on peut analyser la proximité existant entre les sous-groupes: On voit par exemple que le sous groupe « Veille technologique » est proche du sous groupe qui s'intéresse aux « représentations cartographiques ». On n'aurait pas pu procéder à ce genre de constat à partir du seul réseau.

Bibliographie

Aubry C, (1993), « Tracé automatique de graphes et réalisation d'une interface graphique sous X Window dans un système d'Analyse Relationnelle des données », *IBM-Centre Européen de Mathématiques Appliquées*, DESS IMOI

Aho A., Hopcraft J. et Ullman J.(1987), *Structure des données et algorithmes*, Interéditions, France.

Boutin E, Quoniam L, Rostaing H. et Dou H. (1995-1), « A New Approach to Display Real Co-authorship and Co-topicship through Network Mapping », *Acte du colloque « Fifth International Conference on Scientometrics and Informetrics »*, Chicago.

Boutin E, Dumas P, Rostaing H, Quoniam L, (1995-2), « Les réseaux comme outils d'analyse en bibliométrie. Un cas d'application: les réseaux d'auteurs. », *Cahier de la Documentation, Association Belge de Documentation*, N°1, 1996, P3-13

Davidson H, Harel D,(1989) « Drawing Graphs nicely using Simulated Annealing », *Department of applied mathematics & Computer Science*, Rehovot, Israel.

Lhen J., Lafouge T., Elskens Y., Quoniam L., Dou H. (1995), « La statistique des lois de Zipf », *Actes du Colloque: « Les systèmes d'information élaborés », Ile Rousse*, Juin 1995.

Rostaing H.,(1993), Veille technologique et bibliométrie: concepts, outils et applications, *Thèse: Aix Marseille III*, 353p.

Tijssen R.J.W., Van Raan A.F.J. (1989) « mapping co-word structure; a comparison of multidimensional scaling and Leximappe » *Scientometrics*, Vol 15, N° 3-4, p 283-295

Wasserman S. Faust K (1994), « Social Network Analysis », USA, University of Cambridge.