



HAL
open science

Les réseaux comme outils d'analyse en bibliométrie. Un cas d'application : les réseaux d'auteurs

Eric Boutin, Philippe Dumas, Hervé Rostaing, Luc Quoniam

► To cite this version:

Eric Boutin, Philippe Dumas, Hervé Rostaing, Luc Quoniam. Les réseaux comme outils d'analyse en bibliométrie. Un cas d'application : les réseaux d'auteurs. 1996. sic_00826955v1

HAL Id: sic_00826955

https://archivesic.ccsd.cnrs.fr/sic_00826955v1

Preprint submitted on 28 May 2013 (v1), last revised 16 Jun 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES RESEAUX COMME OUTILS D'ANALYSE EN BIBLIOMETRIE. UN CAS D'APPLICATION : LES RESEAUX D'AUTEURS

INTRODUCTION:

L'objectif de cette validation est d'appliquer l'approche réseau au domaine de la bibliométrie. Cette validation a fait l'objet d'un article écrit par Boutin et alii et paru en Mars 1996 dans les cahiers de la documentation. La notion de réseaux est bien connue des techniciens de la bibliométrie. L'utilité de ce concept, qui met en évidence les structures des relations existant entre différentes entités, a déjà été présentée dans de nombreux travaux. Ces réseaux peuvent tour à tour représenter l'activité de la co-citation (SMALL, 1973), les réseaux socio-techniques (CALLON, 1993), la structure thématique d'un domaine (DOU et alii, 1989) ou les collaborations scientifiques (PETERS et VAN RAAN, 1991)...

Le point de départ de ces analyses bibliométriques est un ensemble de notices bibliographiques téléchargées à la suite de l'interrogation d'une banque de données. Cette masse d'information possède la caractéristique d'être structurée en champs homogènes.

Un expert effectuant une lecture de cet ensemble de notices peut dégager les relations entre unités bibliographiques ce qui lui permet d'identifier la structure de l'ensemble étudié. Il prend pour point d'ancrage un ou plusieurs champs descripteurs de l'ensemble des notices considérées. Il constitue des dyades, des associations transitives pour arriver finalement à un réseau de relations.

Toutefois, quand le nombre de notices à examiner augmente, la lecture séquentielle de cette information s'avère fastidieuse et l'expert a du mal à se forger une idée objective et complète de la réalité qu'il cherche à appréhender. La construction automatique du tissu de relations entre les différentes unités bibliographiques devient alors indispensable à l'analyse d'une grande masse de données.

Cette validation présente une application de cette approche à la construction de réseaux d'auteurs.

L'exemple présenté ci dessous se propose de visualiser le réseau des auteurs travaillant dans le domaine de la bibliométrie au niveau mondial, et ceci après la consultation de la banque de données Pascal*.

La collecte des références bibliographiques a été effectuée à partir du CDROM Pascal 1984-1994 avec comme équation logique : « bibliometr? or scientometr? or informetr? ». Le résultat de cette collecte est la constitution d'un corpus bibliographique des 1191 références.

Le premier travail a consisté à isoler le champ auteur de chaque référence bibliographique, puis à faire l'inventaire des relations entre auteurs pour finalement présenter ces relations sous forme matricielle.

Plusieurs types de réseaux qui sont présentés de façon successive à travers les trois paragraphes suivants qui s'attacheront respectivement à représenter le réseau global, des réseaux issus de zooms et des réseaux dynamiques.

La construction du réseau global.

Sur le réseau de départ de la *Figure 1*, ne sont représentés, par souci de clarté que les auteurs qui ont une fréquence de publication supérieure ou égale à 3. Chaque auteur est symbolisé par un sommet du réseau. Chaque sommet est représenté par une étiquette qui contient le nom de l'auteur et le nombre d'articles qu'il a publié. Un lien n'est visualisé qu'à partir du moment où les deux auteurs correspondant ont publié ensemble au moins deux articles.

* Banque de données produite par l'INSIT. Elle a l'avantage d'être une source d'information multidisciplinaire et internationale pour les domaines des sciences exactes.

Ce réseau fait apparaître un certain nombre d’auteurs isolés compte tenu des restrictions émises précédemment. Toutefois, on peut identifier un certain nombre de groupes correspondant aux différentes équipes de recherche dans le domaine. L’ensemble est appelé métastructure du réseau. Il faut noter que la proximité entre groupes n’a aucune signification ni justification, si ce n’est l’esthétisme de la représentation globale. Pour une matrice de relations, il existe une quasi-infinité de réseaux esthétiquement équivalents qui permettent de la représenter graphiquement. Le choix de l’un ou de l’autre s’effectuera de façon arbitraire.

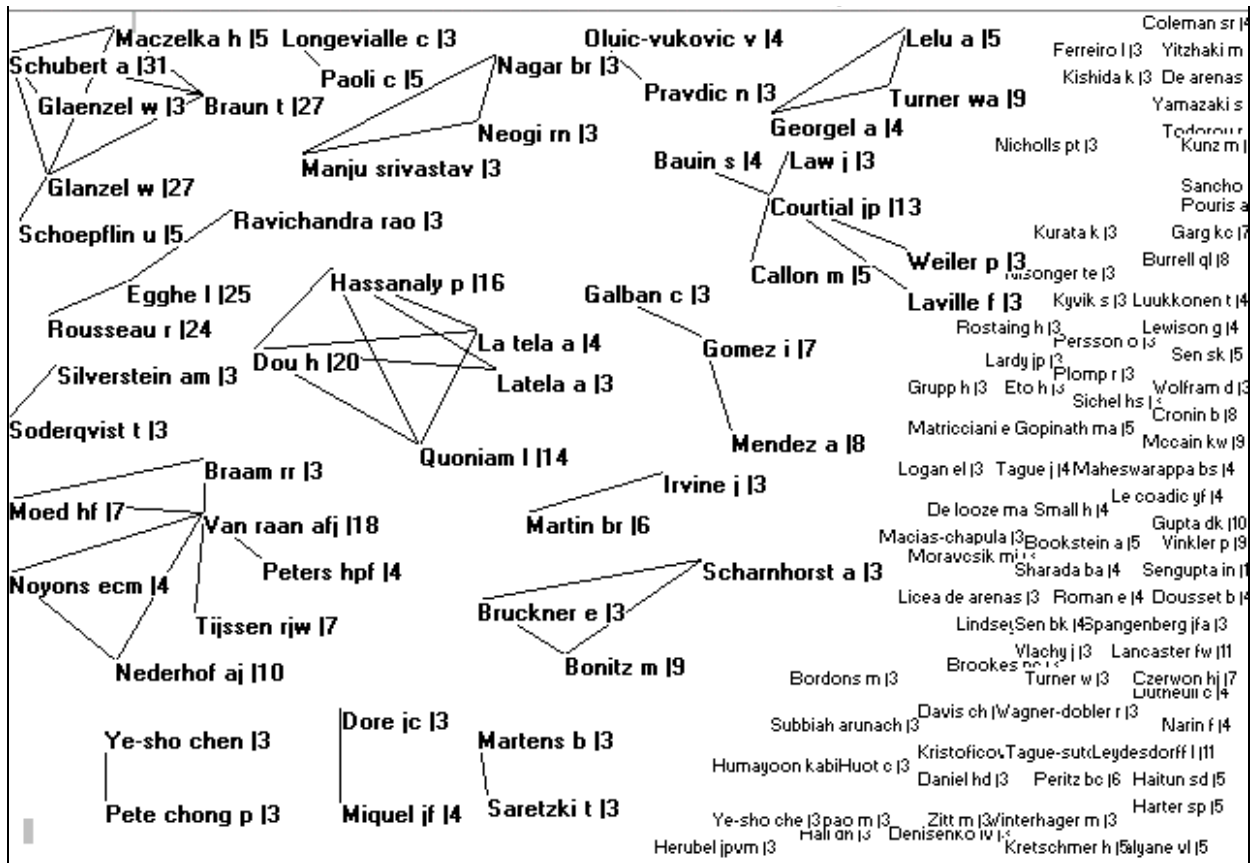


Figure 1 : Construction du réseau d'auteurs.

L’analyse de ce réseau peut s’effectuer en répondant à quelques questions (DEGRENNÉ et FORSE, 1994) :

- Au sein de chaque groupe, l’organisation entre auteurs est-elle une organisation plutôt centralisée ou partagée ? Ce type de réflexion permet de rendre compte de la politique des laboratoires en matière de publications scientifiques.
- Les groupes formés correspondent-ils à des centres de recherche identifiables ? Ce type d’interrogation permet de mettre en évidence des collègues invisibles d’auteurs qui, sans faire partie du même laboratoire, n’en publient pas moins ensemble.

Réseaux obtenus à la suite de zooms :

Toutefois, ce réseau très général ne représente pas la réalité dans toute sa complexité. Celle-ci ne peut être appréhendée que lorsque l’on effectue un zoom sur une certaine portion du réseau. C’est ce qui est suggéré par la Figure 2. Ce zoom s’effectue en ajoutant certains critères pendant la construction du réseau de relations. Ce réseau visualise en effet sans restriction de fréquences les

relations de copublication entre les auteurs qui ont publié en langue française. Sur ce réseau, les auteurs qui ont plus d'une publication sont représentés en utilisant une taille de caractères plus grande. On retrouve la structure par groupe de la figure 4 mais celle-ci est moins apparente.

On voit poindre dans cette *Figure 2* le rôle de pivots joué par certains auteurs. Ces pivots appelés aussi isthmes sont représentés dans un rectangle encadré. Si on enlève ces auteurs, on disconnecte à chaque fois le groupe en plusieurs sous-groupes. Ces auteurs particuliers se situent à la frontière de plusieurs sous-groupes de recherche. Ils sont les points de passage obligé pour rapprocher les recherches des sous-groupes. On peut définir de la même façon la notion de xisthme. Dans la *Figure 2*, Paoli, Dionne, La Tela constituent un 3-isthme. Il faut enlever en effet simultanément ces trois auteurs pour disconnecter le groupe en deux sous groupes.

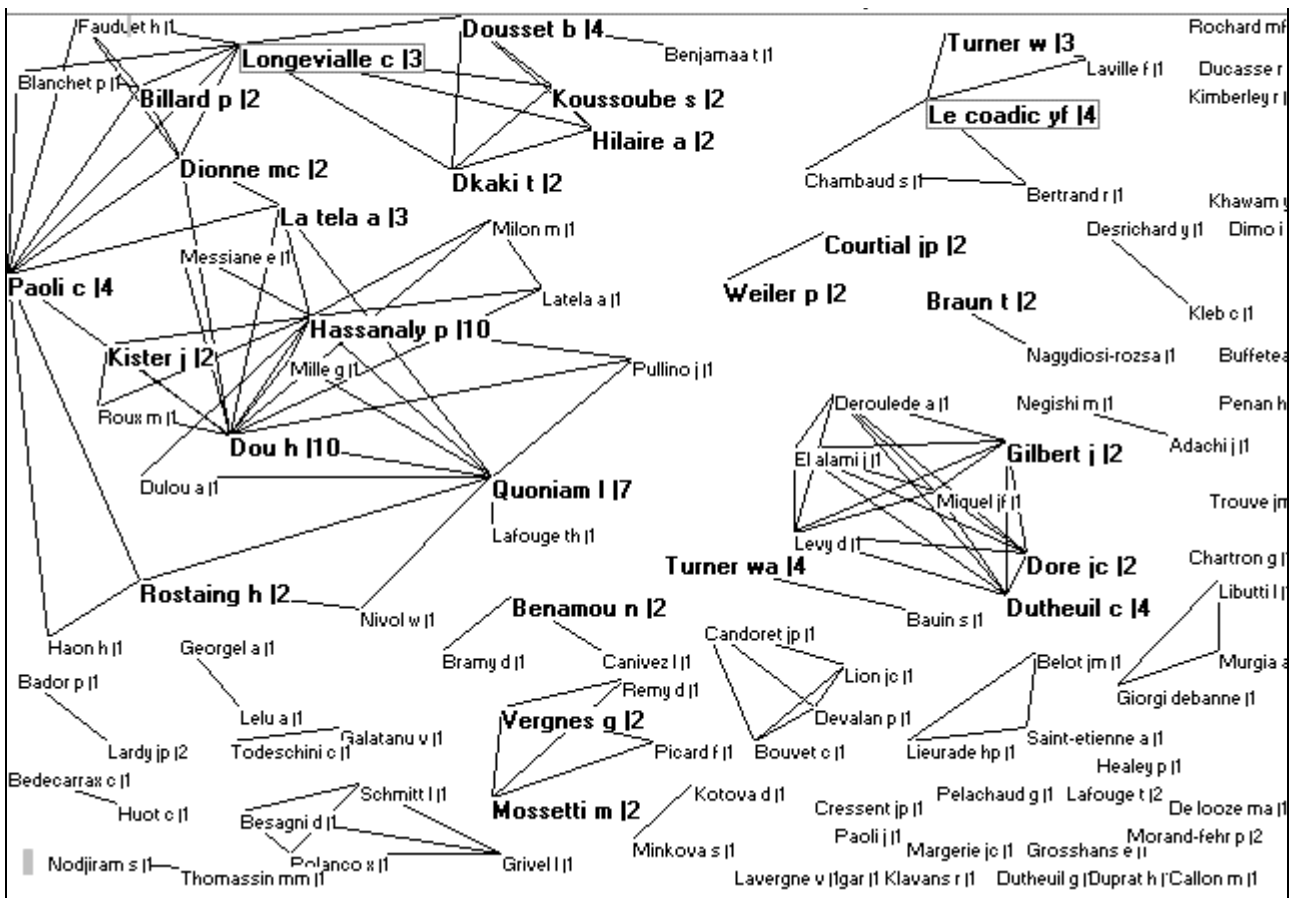


Figure 2 : Réseau des auteurs ayant publié en langue française

Le même principe de « zoom » peut nous conduire à visualiser les relations de copublication des auteurs qui ont publié dans une revue particulière. A titre d'illustration, le réseau de la *Figure 3* représente les relations de copublication des auteurs dans la revue *Scientometrics* spécialisée dans le domaine de la bibliométrie et de la scientométrie. Ce réseau est intéressant en soi mais aussi en comparaison avec le réseau initial présenté *Figure 1*.

Plusieurs questions peuvent là encore servir de piste à l'interprétation :

- La littérature publiée en langue française sur le sujet permet-elle de rendre compte de l'activité de recherche mondiale sur le sujet ? Pour répondre correctement à cette question, il faudrait étudier également les thèmes de recherche à travers l'analyse du champ mots-clés.
- Les équipes publiant en français sur le sujet ont-elles une reconnaissance internationale ?

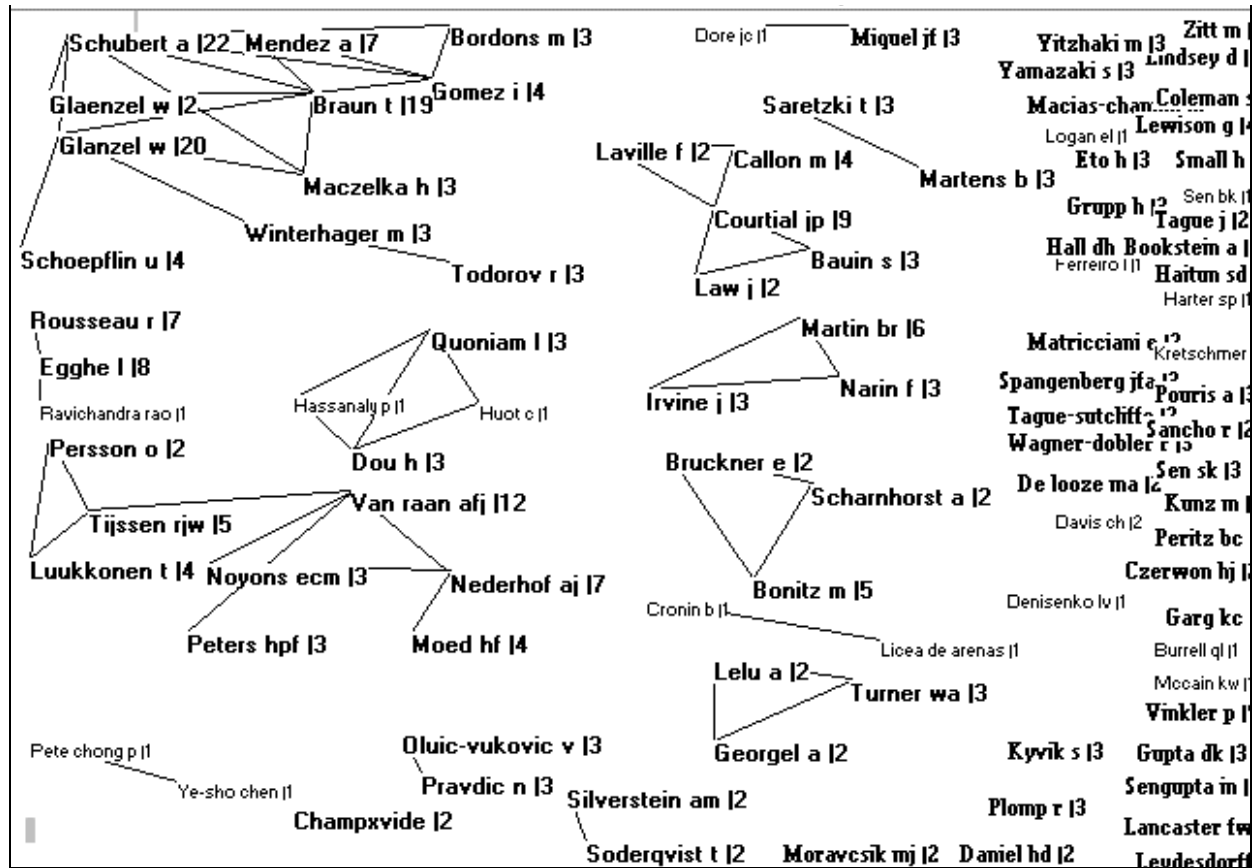


Figure 3: Réseau des auteurs ayant publiés dans la revue Scientometrics

Analyse dynamique du réseau :

Une dernière approche peut être présentée. Celle-ci fait du réseau de la Figure 1 le point d'arrivée d'une évolution dynamique qui a conduit ce réseau à se structurer de cette façon. Pour ce faire, un découpage arbitraire en trois périodes a été obtenu à la suite d'un découpage temporel: 1984-1990, 1990-1992, 1992-1994. Ceci nous permet de représenter trois réseaux. Le réseau correspondant à la période 1984-1990, le réseau correspondant à la période 1984-1992 et le réseau correspondant à l'ensemble de la période étudiée. Ces réseaux sont présentés successivement Figure 4, Figure 5, Figure 6. Les auteurs figurant dans des étiquettes encadrées, sont des auteurs, nouveaux entrant dans le domaine. Ce type d'approche permet de ne pas rester à une vision statique de la réalité mais permet d'intégrer une vision dynamique des choses.

L'analyse de cette dynamique peut être effectuée en répondant à quelques questions :

- Quels sont les pères fondateurs du domaine ?
- Y a-t-il une forte « mortalité » dans ce domaine se traduisant par la disparition de certaines équipes ? Pour répondre parfaitement à cette question, il aurait fallu procéder à des découpages de temps totalement disjoints.
- Y a-t-il une forte « natalité » dans ce domaine se traduisant par l'arrivée de nouvelles équipes ?
- Y a-t-il une tendance à la concentration des groupes ou à la dispersion, à l'isolement ou à la collaboration des chercheurs ?

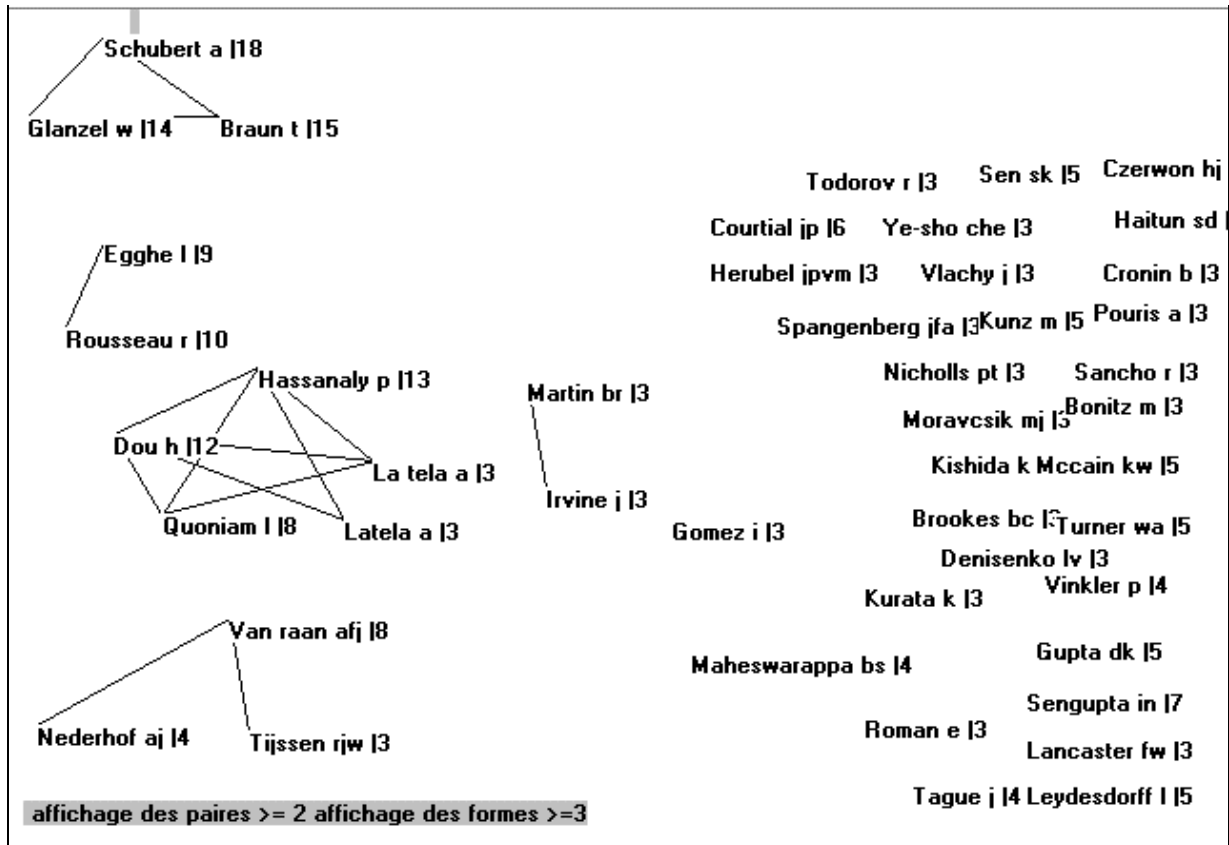


Figure 4 : Réseau d'auteurs correspondant à la période 1984-1990

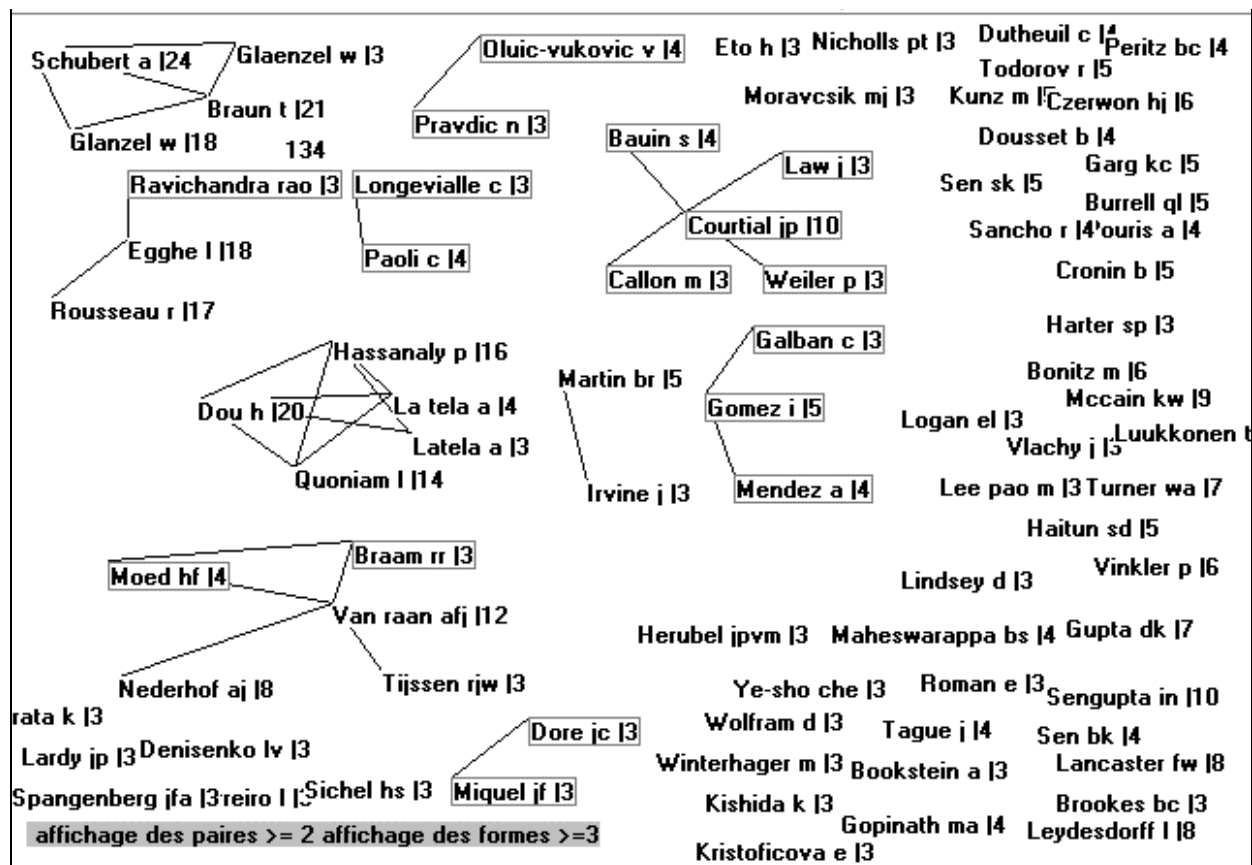


Figure 5 : Réseau d'auteurs correspondant à la période 1984-1992

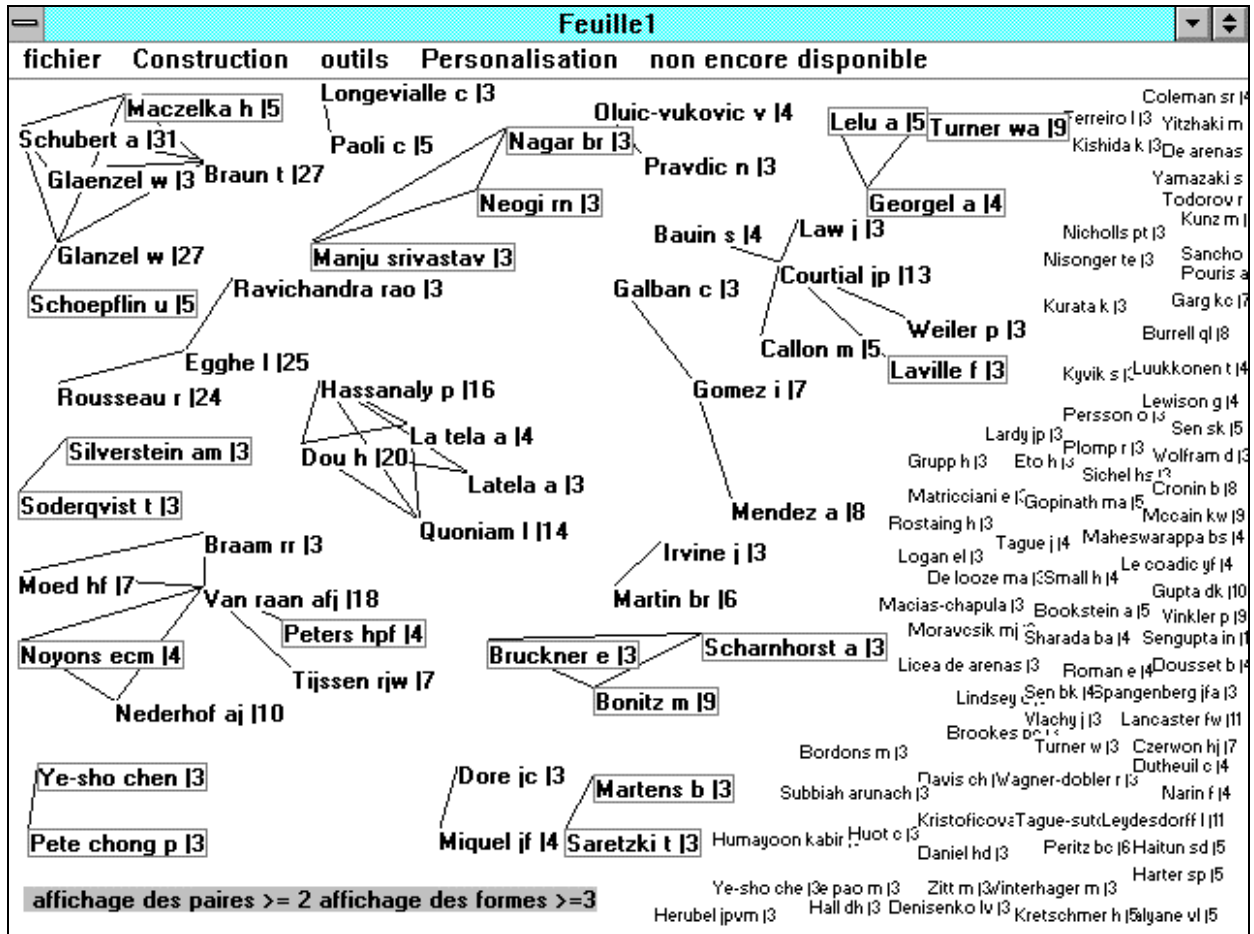


Figure 6 : Réseau d’auteurs correspondant à la période 1984-1994

Conclusion : Eléments de réflexion sur la validité de l’approche réseau :

La construction automatique de réseaux fait apparaître un certain nombre de limites. Si le réseau de relations est construit à partir de références téléchargées de banques de données, deux limites apparaissent suivant que l’on se situe en amont ou en aval de l’algorithme. D’une part, la qualité du réseau est tributaire de la qualité de l’information livrée par le producteur de la banque de données. Le réseau de la Figure 5 exprime bien ce genre de préoccupation. Les auteurs « La Tela » et « Glanzel » sont orthographiés de deux façons différentes dans les références d’origine. Ce problème peut être résolu en amont par un prétraitement du champ auteur avant analyse bibliométrique : la normalisation des noms d’auteurs par reformatage. Nous n’avons volontairement pas pratiqué ce traitement pour faire ressortir cette limite. D’autre part, dans l’interprétation des résultats, il peut y avoir confusion entre l’hypothèse et la conclusion. L’outil réseau survalorise en effet le groupe au détriment de l’individu. Si on examine la position sur la Figure 6 des deux auteurs Leydesdorff et Nederhof, on remarque que Nederhof se trouve survalorisé par le fait qu’il appartient à un groupe. Cette limite peut être partiellement contournée si on introduit des tailles d’étiquettes proportionnelles au nombre de publications. L’information fournie par le réseau et les fréquences est plus riche que la seule fréquence qui est le seul critère actuellement reconnu dans les travaux scientifiques sur le sujet. D’autre part, cette approche se caractérise par son caractère réaliste. Il n’y a pas de déformation de la réalité comme c’est parfois le cas dans les méthodes d’analyse de données, ni utilisation de métriques.

Bibliographie :

BOUTIN E, DUMAS P, QUONIAM L, ROSTAING H, DOU H, (1995a) « Génération automatique de réseaux en bibliométrie », Actes du colloque « les systèmes d'informations élaborées », Ile Rousse, 30 Mai - 6 Juin 1995

BOUTIN E, QUONIAM L, ROSTANIG H, DOU H, (1995b) « A new approach to display real co-authorship and co-topicship through network mapping », Acte du colloque « Fifth International Conference on scientometrics & informetrics » Chicago, 7-10 Juin, 1995

DAVIDSON R., HAREL D., (1989), « drawing graphs nicely using simulated annealing », Technical Report CS89-13, Department of Applied Mathematics and Computer Science, The Weizmann Institute, Rehovot, Israel, July 1989

GROVES L.,MICHALEWICZ Z., ELIA P. et JANIKOW C.,(1990) « Genetic algorithms for drawing directed graphs », Methodologies for intelligent Systems, 5, Pages 268-276, 1990

WHITE H., Mc CAIN K., (1989), « bibliometrics », Annual review of information Science and Technology (ARIST), Vol. 24, p. 119-186, 1989.