



Cartographie de la recherche scientifique

Abdou Beukeu

► To cite this version:

| Abdou Beukeu. Cartographie de la recherche scientifique . 2016. hal-01357006v3

HAL Id: hal-01357006

<https://hal.science/hal-01357006v3>

Preprint submitted on 17 Oct 2016

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.

Cartographie de la recherche scientifique :
Cas des Universités Cheikh Anta Diop de Dakar (UCAD) et Gaston
Berger (UGB) de Saint-Louis

Abdou Beukeu SOW

Doctorant en Science de l'Information et de la Communication
Université Lille 3, membre du Groupe d' Études et de Recherche Interdisciplinaire en
Information et Communication (GERiiCO)

Résumé

Cet article se propose de faire une cartographie de la recherche des deux plus grandes universités publiques du Sénégal en s'appuyant sur la méthode des citations bibliographiques : les co-citations et les co-signatures et sur un outil d'aide à l'analyse et à la visualisation de contenu : Gephi¹. L'objectif est de montrer comment les techniques de visualisation de l'information peuvent fournir une meilleure compréhension de la recherche au Sénégal et faciliter les prises de décisions sur les questions liées à l'Information Scientifique et Technique.

Mots-clés : Cartographie de la recherche, visualisation de l'information, bibliométrie, Université Cheikh Anta Diop de Dakar, UCAD, Université Gaston Berger, UGB

Abstract

This article aims to make a research mapping of the two largest public universities in Senega. The work is based on the bibliographic citation method: co-citation and co-signatures on the analysis support tool and visualization of content: Gephi. The aim is to show how information visualization techniques can provide a better understanding of research in Senegal and facilitate decision-making on issues related to the Scientific and Technical Information.Technique.

Keywords: Science mapping, visualization, Bibliometrics, Université Cheikh Anta Diop de Dakar, UCAD, Université Gaston Berger, UGB

¹ Logiciel de cartographie open source

Introduction

Les chercheurs sénégalais publient à leur propre rythme, sans contraintes majeures. Aucune politique ne les impose à publier périodiquement. Malgré cela, il est important de signaler que le Sénégal dispose d'un potentiel documentaire riche (qualité)² en termes d'Information Scientifique et Technique (I.S.T), qui cependant reste difficile à localiser et à accéder. Cette situation contribue largement au classement des universités sénégalaises.

Malgré les efforts considérables du gouvernement sénégalais traduits par le renforcement du budget consacré à l'enseignement supérieur et la création de nouvelles universités, la politique en termes de valorisation de la recherche manque encore de pertinence. Les conséquences sont visibles. Entre 2013 et 2015, l'université Cheikh Anta Diop qui faisait la fierté des universités francophones est passée de la 13^e place à la 39^e place dans le classement des universités africaines, détrônée par l'université de Ouagadougou³. Créé depuis 1977, le Centre National de Documentation Scientifique et Technologies (CNDST) est chargé de mener la politique nationale en termes d'information à caractère scientifique. Sa mission⁴ est : de coordonner l'action des organismes constituant le réseau national d'information et de documentation scientifique et technique ; de promouvoir toute action d'intérêt commun susceptible de renforcer les moyens d'information et de documentation scientifique et technique.

Les prises de décisions sur les questions liées à l'I.S.T doivent reposer sur des analyses bibliométriques préalables. Il nous semble dès lors opportun de bien évaluer la visibilité scientifique des universités et chercheurs sénégalais pour mieux définir les stratégies et politiques en termes de valorisation de la recherche au Sénégal.

² Par qualité, nous entendons l'IST produite sur le Sénégal et qui répond le plus aux besoins (réalités politiques économiques et culturelles) des Sénégalais.

³ Ces données sont obtenues à partir des résultats de *Ranking Web of university* qui est le plus grand classement académique des établissements d'enseignement supérieur en fonction de leur production, de leur présence et de leur impact sur le web

⁴ Cf : décret n 77 272 du 6 avril 1977 (Sénégal)

Cartographie d'une recherche ou visualisation de l'information

La capacité de l'homme à visualiser l'information joue un rôle majeur dans ses processus cognitifs. L'usage des méthodes graphiques permet à celui-ci de mieux appréhender des notions qui lui étaient peu observables auparavant. La visualisation de l'information est une représentation graphique de l'information dont l'objectif est de fournir une compréhension du contenu de l'information, de produire des représentations visuelles à partir de données brutes. Selon Tamara Munzner[1], la visualisation de l'information n'est pas un champ scientifique et technique comme le définit Polanco[2]. Pour Munzner, c'est plutôt un phénomène en réaction au fait de l'explosion des données qui de surcroît progresse plus rapidement que le taux de croissance de la puissance informatique. Si les interprétations sur le concept divergent, tous les auteurs s'accordent au moins pour dire que la représentation visuelle fait appel à des ressorts cognitifs qui fonctionnent plus rapidement que ceux d'ordre d'activation de la compréhension comme le raisonnement mathématique, physique. La visualisation de l'information est une manière de construire du sens à partir de données brutes pour faire apparaître de phénomènes jusqu'alors imperceptibles. Le succès de cette approche réside dans le fait que « les humains sont fortement habitués aux images et aux informations visuelles. Une représentation visuelle peut communiquer certains types d'information beaucoup plus rapidement et efficacement qu'une autre méthode »[3]. Au-delà de l'intérêt de manipuler des données brutes qui n'en disent pas beaucoup, la visualisation de l'information offre aux lecteurs une facilité de compréhension. Dans son travail de recherche, Pierre Humberte[4] pense que cette approche est aussi une réponse au problème lié à la sous-exploitation des systèmes de recherche d'information (SRI). Citant Ihadjadene (1998) et Favier et al.[5], il montre que les SRI sont peu exploités et que la visualisation est un recours pour pallier ce phénomène. Des recherches menées par Ihadjadene sur l'usage des moteurs de recherche, des bibliothèques numériques et des www-Opacs prouvent que les « *ressources du système sont sous-utilisées et que les outils mis à disposition de l'utilisateur final pour explorer le nombre élevé de réponses sont insuffisants et inadaptées* ». De leur côté, Favier, et al. montrent dans leurs travaux sur le moteur recherche AltaVista, que « *85 % des usagers se contentent des dix premiers résultats fournis sur la première page et 78 % des requêtes ne sont pas modifiées dans le but de les améliorer* ». Ces recherches indiquent que l'identification et l'interprétation des informations sont de plus en plus hors de portée des capacités humaines. Au-delà de leurs apports en bibliométrie, les techniques de visualisation de l'information, en plein essor, méritent l'attention des acteurs en science de l'information et de la communication. Depuis quelques années, la visualisation de l'information est devenue une pratique qui intéresse les Sciences de l'Information et de la Communication, cependant les

méthodes qui y sont appliquées notamment dans l'organisation des connaissances remontent un peu plus en arrière.

Méthodologie

L'objectif est de faire une cartographie de la recherche des universités Cheikh Anta Diop de Dakar (UCAD) et Gaston Berger (UGB) de Saint-Louis. Ceci, en s'appuyant sur la méthode des citations bibliographiques : les co-citations et les co-signatures et sur un outil d'aide à l'analyse et à la visualisation de contenu : Gephi.

Les analyses bibliométriques reposent traditionnellement sur deux types d'indicateurs : les indicateurs d'activité ou « univariés » et les indicateurs relationnels[6]. Les premiers servent essentiellement à comparer entre eux, les éléments d'un ensemble de références bibliographiques. Il s'agit dans ce système de mesure de dénombrer l'élément bibliographique étudié. On peut identifier deux procédures avec ce système : le comptage des références (mesure de la productivité) et le comptage des citations reçues par les références (mesure de l'impact). Le second type indicateur cherche à suivre l'évolution des multiples interactions qui se développent au sein de la recherche. Ces indicateurs peuvent être regroupés en deux grandes familles : celle de la première génération et celle de la seconde génération. Contrairement aux indicateurs de la seconde génération, ceux de la première n'entrent pas dans le contenu des documents analysés.

Constitution du corpus

Traditionnellement, les études sur la mesure de la science se font à partir de bases bibliographiques comme le Web of Science, Scopus, Google Scholar, etc. Parmi ces multiples bases, un choix s'impose pour avoir un corpus homogène et assez représentatif. Le WoS est une base de données trop sélective. Elle privilégie l'anglais par rapport au français. Étant donné que la plupart des productions scientifiques sénégalaises est en français, il n'est pas étonnant de remarquer l'absence des chercheurs sénégalais dans les revues indexées par le WoS. En plus de cela, le WoS reste peu accessible pour les pays africains à cause de son coût très élevé. D'un point de vue scientifique, Google scholar ne respecte pas certaines règles de publication. L'un de ses inconvénients pour une analyse c'est d'être trop général. Lancé en fin 2004, il inventorie des articles approuvés ou non par des comités de lecture (peer-reviewed), des thèses de type universitaire, citations ou encore des livres scientifiques. En choisissant ces deux bases on risque alors d'avoir peu de publications sénégalaises (silence) ou trop (bruit), ce qui pourrait biaiser nos analyses. C'est donc pour ces raisons évoquées que nous avons choisi de constituer nos corpus à partir de Scopus dont le Sénégal s'est récemment doté.

Dans un premier temps nous voulions nous intéresser à l'ensemble des publications scientifiques sénégalaises, ce qui nous donnerait un corpus trop grand pour faire une analyse propre. Après avoir essayé plusieurs requêtes, nous avons pu constater en nous limitant aux deux plus grandes et premières universités du Sénégal (Cheikh Anta Diop et Gaston Berger), une forte présence de la médecine, des mathématiques et de l'agriculture. Partant de là, nous avons résumé notre analyse sur ces trois domaines et sur la période 2011 à 2015.

L'importation des données dans Gephi

Les premières cartographies ont été réalisées à l'aide d'outils tels que Pajek (sorti en 1996). Ces outils de cartographie jouent un rôle important dans la visualisation de l'information. Parmi cette gamme se trouve Gephi. Utilisé par beaucoup d'institutions scientifiques en France⁵ comme aux États-Unis, le logiciel permet de réaliser des représentations de corpus d'informations bien structurées. Le logiciel présente des fonctionnalités qui offrent la possibilité d'observer les données sur différents plans, de réaliser des filtres en fonction des éléments statistiques ou bien encore d'identifier des clusters et de faire des focus (vue à grande échelle) sur eux. Grâce à ses multiples fonctionnalités, Gephi offre la possibilité d'analyser des données en regroupant les éléments observés en « clusters » (thématique, période, groupe de recherche).

Résultats et discussions

C'est sur la base de deux requêtes utilisant les affiliations des universités comme mots-clés que nous avons interrogé la base Scopus. Les résultats des deux requêtes montrent la prédominance de certaines disciplines par rapport à d'autres. À l'université Cheikh Anta Diop comme à l'université Gaston Berger, la médecine, les mathématiques et l'agriculture sont de loin les domaines de recherche les plus représentatifs dans la période qui nous intéresse. (Voir tableau 1 et 2).

Sur tous les graphes obtenus (voir figures), la largeur des liens indique la proximité d'un auteur par rapport à un autre ou à un domaine de recherche donnée. Un auteur qui publie plus dans un domaine X sera plus proche du nœud représentant ce domaine X. Un auteur qui publie plus avec un auteur A sera très proche de celui-ci. Cette distance est dite Chemin géodésique en analyse de réseau. Dans l'exemple de la figure 1, on peut constater la position de l'auteur "Sene. M" qui partage un lien avec URF sciences appliquées et la section mathématique. Cela montre

⁵ En France : INRA (Institut National de la Recherche Agronomique), CNRS (Centre national de la recherche scientifique), INIST (Institut de l'information scientifique et technique), OST (Observatoire des sciences et techniques) Aux États unis : NASA (National Aeronautics and Space Administration), NIGMS (National Institute of General Med Science), NIAID (National Institute of Allergy and Inf Disease), NICHD (Nat. Inst of Child Health Hum Dev).

d'une part que la section mathématiques est rattachée à l'URF sciences appliquées et d'autre part que "Sene. M" signe sous les deux institutions qu'il représente. Ce qui les donnerait plus de visibilité. La position et la taille du nœud représentant "Seck S.M" dans le sous-réseau « URF santé » montrent que celui-ci occupe une place stratégique au sein de l'URF. La taille de son nœud indique en effet qu'il entretient avec les autres auteurs et le nombre de copublications. La taille des nœuds est proportionnelle au nombre de co-publications de chaque auteur. Sur ce même réseau des co-auteurs UGB, il est aisé de remarquer le sous-réseau ISRA placé en périphérique. La position d'un sous-réseau ou d'un nœud en périphérique indique un éloignement qui peut traduire l'émergence d'un domaine de recherche, une faible productivité, ou sa non-appartenance à une communauté. Le cas du sous-réseau ISRA peut alors faire croire que des chercheurs de l'UGB collaborent avec des chercheurs de l'ISRA. Contrairement au réseau UGB, ce que nous pouvons d'emblée constater avec le réseau UCAD c'est que malgré la présence de certains domaines comme les mathématiques, l'agriculture, la chimie, etc., les sous-domaines de la médecine sont les plus présents. C'est dire, que la faculté de médecine est à l'UCAD est plus visible que les autres dans le monde. La forte demande d'admission des étudiants étrangers confirme ce fait. La figure 3 représente le réseau de co-publication dans la faculté de médecine. Elle montre pour chaque sous-discipline identifiée, les auteurs dont les travaux sont plus visibles. La taille de nœuds est proportionnelle au nombre d'articles co-publiés pour les auteurs et au nombre total d'articles pour un sous-domaine donné. Cela signifie que les auteurs comme "Ndiaye M." ; "Ndiaye D." "Mboup S." ; "Faye A." ; "Faye O." occupent une place stratégique dans leur domaine respectif.

La figure 3 représente le réseau de co-auteurs de la faculté de médecine de l'université Cheikh Anta Diop. Elle fait apparaître les départements les plus productifs et les auteurs qui s'y rattachent. La proximité entre les nœuds représentant les auteurs et ceux représentant les disciplines montrent aisément la spécialité de chaque chercheur. Autrement dit, plus un auteur est proche d'un domaine, plus il est certain d'appartenir à ce domaine. Il peut cependant arriver qu'un auteur ait autant des publications dans deux domaines différents.

Nous pouvons dire à partir de cette figure que la « Parasitologie », l'« Odontologie » et la « Biologie » sont les domaines où l'université Cheikh Anta Diop publie le plus. De la même manière les chercheurs comme « Ndiaye M », « Ndiaye D » et « Mboup S » sont très influents dans leur domaine, sans vouloir chercher à quoi cette influence est due.

Les résultats obtenus de ces graphes montrent bien comment on peut se servir des techniques de visualisation pour observer l'état de la recherche dans un pays.

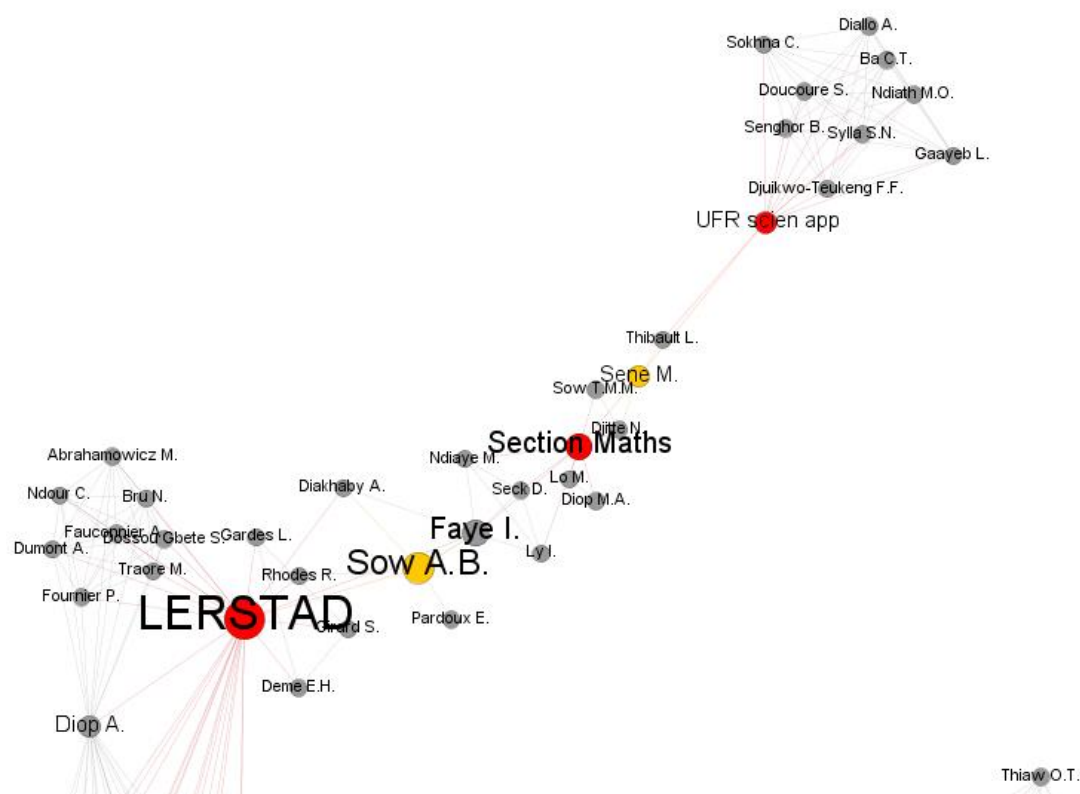


Figure 1 – Position d'un acteur sur le réseau

Nombre de publications	Domaine
311	Médecine
207	Agriculture et sciences biologiques
138	Immunologie et microbiologie
128	Biochimie, Génétique
122	Chimie

Table 1 – Répartition de la production de l'UCAD

Nombre de publications	Domaine
59	Mathématiques
35	Médecine
32	Agriculture et sciences biologiques
17	Informatique
15	Biochimie, Génétique
15	Sciences sociales

Table 2 – Répartition de la production de l'UGB

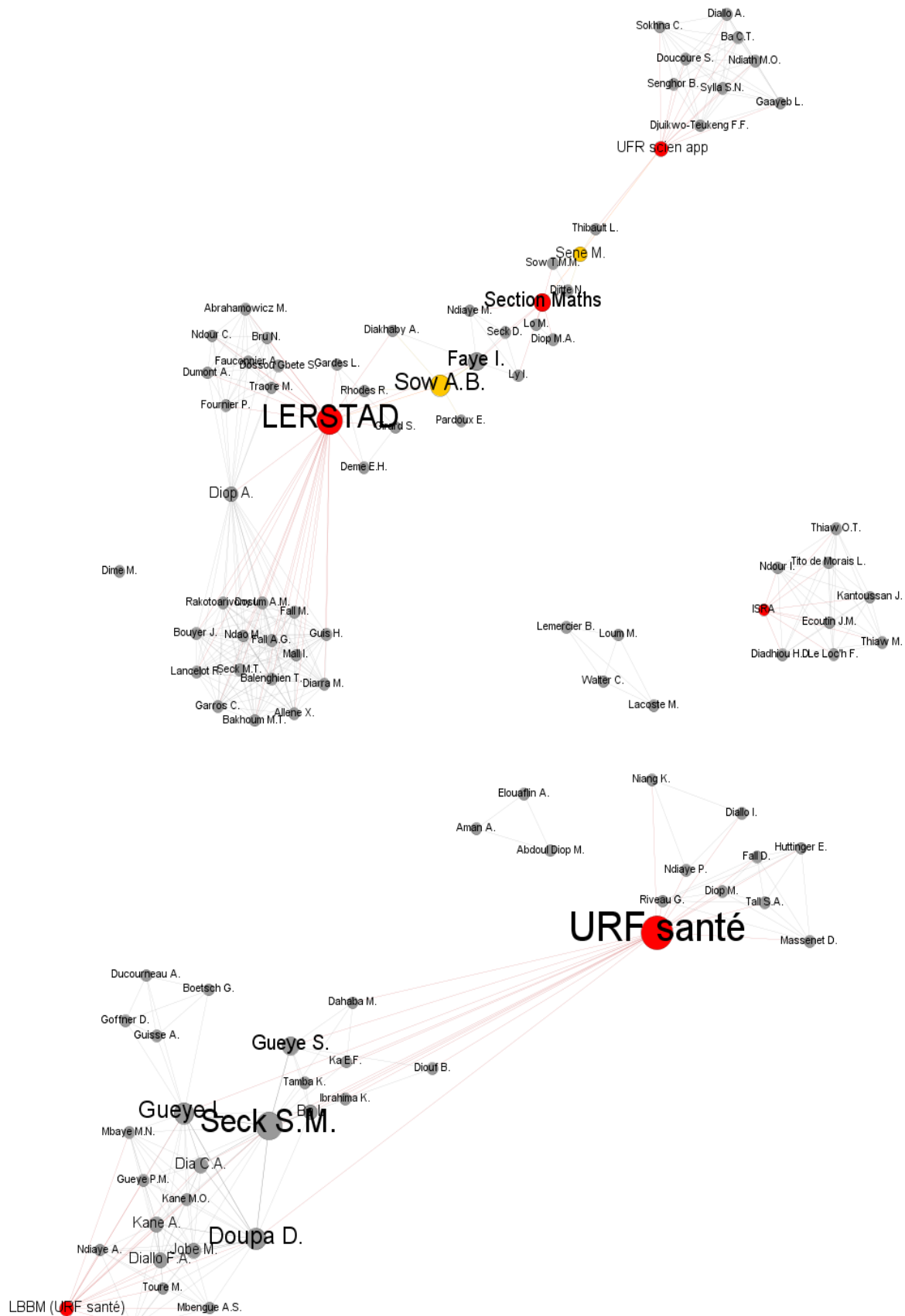


Figure 2– Réseau co-auteurs UGB

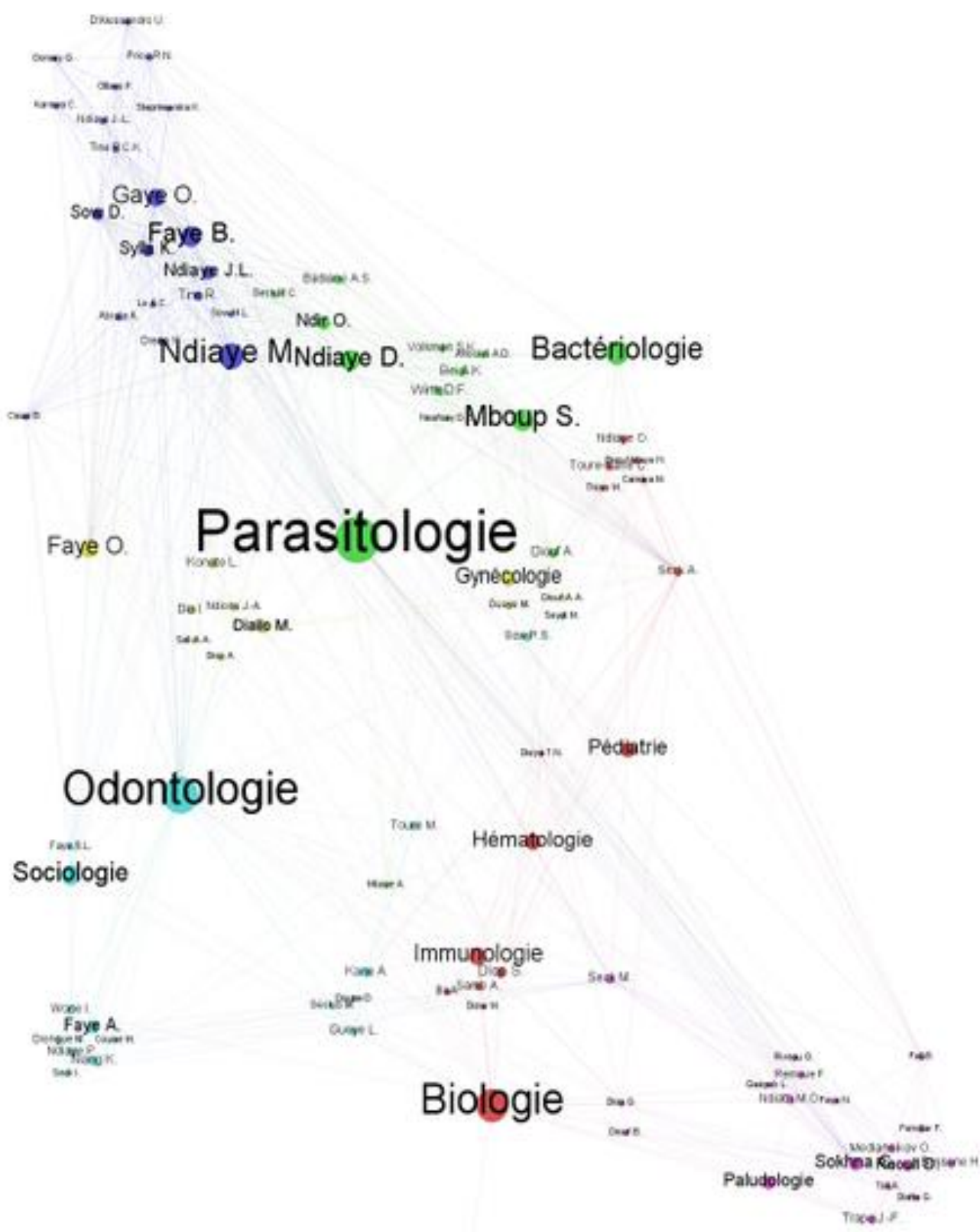


Figure 3— Réseau co-auteurs médecine-UCAD

Conclusion

L'utilité d'une information à caractère scientifique doit être avant tout nationale. Si les méthodes de recherche peuvent être les mêmes dans plusieurs pays, les résultats et les intérêts d'une recherche doivent être spécifiques. Autrement dit, même si la science demeure universelle et l'IST utile quelle que soit son lieu et moment de sa production, la meilleure information pour un pays c'est celle qui prend en compte ses réalités économiques, politique et sociales. À cet effet, il semble primordial de produire des outils d'aide à la décision en commençant par mutualiser toute la production scientifique locale et répertorier les compétences dans tous les domaines de recherche. Cette cartographie de la recherche sénégalaise que nous avons tenté ne peut être complète et bénéfique que si toute la production scientifique est identifiée, localisée et traitée. Une base de données des compétences où seront répertoriées les compétences et productions scientifiques de chacun serait l'outil approprié pour suivre de façon autonome l'évolution de la recherche au Sénégal. Ces analyses bibliométriques sont indispensables pour les décideurs, les chercheurs et étudiants nationaux. Dans la même perspective, le champ de la recherche devrait aussi être observé d'un point de vue sociologie puisque la productivité des chercheurs est assujettie à leur motivation qui peut être de plusieurs origines : investissement, reconnaissance des pairs.

Bibliographie

- [1] T. Munzner, "A nested model for visualization design and validation," *IEEE Trans. Vis. Comput.*, 2009.
- [2] X. Polanco, "Aux sources de la scientométrie," *Les Sci. l'information*, 1995.
- [3] M. A. Hearst, "User interfaces and visualization," *Mod. Inf. Retr.*, pp. 257–323, 1999.
- [4] P. Humbert, "Approche informationnelle pour la prise en compte des contextes d'usage dans la conception d'outils de travail collaboratif assisté par ordinateur," 2011.
- [5] L. Favier and M. Ihadjadene, "Vers des Systèmes de Découverte et de Filtrage d'Information Documentaire: Quelle Stratégie Faut-il Mettre en Place?," *Can. Assoc.*, 2000.
- [6] "Callon M., Bastide F., Bauin S. et al. Les mécanismes d'intéressement dans les textes scientifiques, Cah. STS-CNRS, 1984, vol. 4, pp. 88-105. - Google Scholar." [Online]. Available:
<https://scholar.google.fr/scholar?hl=fr&q=Callon+M.%2C+Bastide+F.%2C+Bauin+S.+et+al.+Les+m%C3%A9canismes+d%E2%80%99int%C3%A9ressement+dans+les+textes+scientifiques%2C+Cah.+STS-CNRS%2C+1984%2C+vol.+4%2C+pp.+88-105.&btnG=&lr=>.