

Apports de l'approche bibliométrique et de l'outil Cortext Manager à la revue de littérature scientifique. Une application à la prise en compte de l'environnement en Sciences Régionales (1999-2019)

Béatrice Davi, Stéphanie Truchet-Aznar, Olivier Aznar, Francis Aubert

▶ **To cite this version:**

Béatrice Davi, Stéphanie Truchet-Aznar, Olivier Aznar, Francis Aubert. Apports de l'approche bibliométrique et de l'outil Cortext Manager à la revue de littérature scientifique. Une application à la prise en compte de l'environnement en Sciences Régionales (1999-2019). NOV'AE, 2022. hal-03885367

HAL Id: hal-03885367
https://hal.science/hal-03885367

Submitted on 5 Dec 2022

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.

Apports de l'approche bibliométrique et de l'outil Cortext Manager à la revue de littérature scientifique

Une application à la prise en compte de l'environnement en sciences régionales (1999-2019)

Béatrice DAVI[1]
Stéphanie TRUCHET-AZNAR[2]
Olivier AZNAR[3]
Francis AUBERT[4]

Correspondance
beatrice.davi@vetagro-sup.fr

Résumé.

Cet article présente une démarche bibliométrique originale ainsi que certaines des fonctionnalités d'un outil d'analyse bibliométrique, la plate-forme Cortext Manager, à travers une application portant sur la prise en compte des enjeux environnementaux dans les articles publiés en sciences régionales entre 1999 et 2019. La première étape de l'analyse a consisté à décrire, à partir d'indicateurs bibliométriques classiques, l'ensemble du corpus de 1 145 articles extraits de *Web of Science*. Dans un deuxième temps, la réalisation d'une analyse des cooccurrences entre les mots-clés a permis de distinguer 4 grands groupes thématiques. Enfin, de nouvelles analyses bibliométriques, couplées avec une analyse qualitative des articles les plus cités, ont permis de caractériser chaque groupe thématique du point de vue des documents phares, des principaux thèmes abordés et des travaux séminaux. Des analyses bibliométriques complémentaires, menées sur chacun de ces groupes et croisées avec les analyses qualitatives sur les principaux articles des groupes, ont permis d'avoir une vision plus précise du contenu des articles et de dégager les principales questions scientifiques abordées ainsi que les groupes de travaux séminaux cités.

Mots-clés

Bibliométrie, Cortext Manager, analyse de cooccurrences, méthodologie, environnement, Sciences régionales.

1 Bibliothèque agronomique, VetAgro Sup, 63370, Lempdes, France.
2 UMR Territoires, UCA, AgroParis Tech, INRAE, VetAgro Sup, 63178, Aubière, France.
3 UMR Territoires, UCA, AgroParis Tech, INRAE, VetAgro Sup, 63370, Lempdes, France.
4 UMR Cesaer, Institut Agro de Dijon, 21079 Dijon Cedex, France.

Contributions of the bibliometric approach and the Cortext Manager tool to the review of scientific literature

An application to the analysis of the consideration of the environment in regional science litterature (1999-2019)

Béatrice DAVI[1]
Stéphanie TRUCHET-AZNAR[2]
Olivier AZNAR[3]
Francis AUBERT[4]

Correspondence
beatrice.davi@vetagro-sup.fr

Abstract

This article presents an original bibliometric approach and certain functionalities of a bibliometric analysis tool, the Cortext Manager platform, via an application focused on the analysis of the integration of environmental issues in articles published on regional science between 1999 and 2019. The first step of the analysis consists in describing, on the basis of classical bibliometric indicators, the entire corpus of 1 145 articles extracted from the *Web of Science*. Then, an analysis of keywords cooccurrences is carried out to distinguish 4 main thematic groups. Lastly, new bibliometric analyses, coupled with a qualitative analysis of articles cited the most frequently, makes it possible to characterize each thematic group from the angle of outstanding documents, the main themes dealt with, and seminal works. Additional bibliometric analyses carried out on each of these groups and crossed with the qualitative analyses on the main articles of the groups provide a more precise view of the content of the articles, and identifying the major scientific issues broached and the groups of seminal works cited.

Keywords

Bibliometrics, Cortext Manager, analysis of cooccurrences, methodology, environment, regional science.

1 Bibliothèque agronomique, VetAgro Sup, 63370, Lempdes, France.
2 UMR Territoires, UCA, AgroParis Tech, INRAE, VetAgro Sup, 63178, Aubière, France.
3 UMR Territoires, UCA, AgroParis Tech, INRAE, VetAgro Sup, 63370, Lempdes, France.
4 UMR Cesaer, Institut Agro de Dijon, 21079 Dijon Cedex, France.

Introduction

Apparue au début du XXe siècle, la bibliométrie consiste en l'application de méthodes statistiques ou mathématiques sur des ensembles de références bibliographiques, dans le but d'aider à leur comparaison et à leur compréhension (Rostaing, 1996). À partir des années 1960, elle se développe fortement suivant deux approches : une approche sociologique et une approche évaluative. La première, impulsée notamment par Derek Price (1963) et souvent qualifiée de Science de la Science, mobilise la bibliométrie pour analyser les mécanismes sociologiques qui sous-tendent l'activité scientifique. Dans le sillon de la création de l'*Institute for Scientific Information* (ISI) par Eugène Garfield en 1960, la seconde s'appuie plus particulièrement sur l'analyse des citations dans un but d'évaluation de l'activité scientifique (Garfield et Merton, 1979). Depuis les années 1990, le recours à la bibliométrie s'étend à un objectif plus large d'analyse de la science, favorisé par la diffusion de techniques et d'outils d'analyse. Ainsi, dans des domaines variés, les travaux de recherche mobilisent de manière croissante l'analyse bibliométrique pour réaliser des revues de littérature. L'objectif est alors de rendre compte d'ensembles de travaux publiés dans une revue (Baumont et Girard, 2019), portant sur une approche particulière (Surinach *et al.*, 2003), ou encore abordant une notion spécifique (Tancoigne *et al.*, 2014). De manière classique, l'utilisation d'indicateurs uni-variés construits à partir des informations bibliographiques d'un corpus de publications permet d'analyser l'évolution du nombre de publications et de citations, d'examiner leur répartition entre revues ou encore d'identifier les principaux auteurs. Mais l'analyse bibliométrique permet également d'identifier des liens au sein d'un corpus de publications à partir de l'analyse des cooccurrences de différents types d'items, comme les mots-clés ou les auteurs cités. Ces analyses de cooccurrences ouvrent de nombreuses possibilités. Ainsi, deux types d'analyse des cooccurrences peuvent être menés sur la base des références bibliographiques d'un corpus de publications : le couplage bibliographique, qui permet d'identifier les publications qui se réfèrent aux mêmes travaux (Kessler, 1963), et l'analyse des co-citations, qui permet d'identifier les références citées conjointement par plusieurs publications du corpus (Marshakova, 1973 ; Small, 1973). Les analyses de cooccurrences s'appuient sur différents outils d'analyse et de visualisation de données. Ces programmes permettent, notamment, de représenter les liens identifiés à travers des cartes relationnelles et de caractériser les items en fonction de leur répartition et de leur position dans ces cartes. Appliquée à l'analyse textuelle, les cooccurrences permettent de repérer les termes associés entre eux et ainsi d'identifier les grands champs thématiques présents dans un corpus de références bibliographiques (Bordignon, 2021).

Face aux interrogations croissantes concernant les dynamiques des territoires dans un contexte de changement climatique, nous avons souhaité faire un état des lieux de la manière dont les enjeux environnementaux étaient appréhendés dans le domaine des sciences régionales. Notre démarche consistait à analyser un ensemble d'articles traitant à la fois d'environnement et de développement régional et publiés dans des revues en sciences régionales sur une période de vingt ans (1999-2019). Une première étape de recherche bibliographique nous a permis d'isoler 1 145 articles. Compte tenu de ce nombre, nous avons choisi de privilégier l'approche bibliométrique qui permet d'appréhender un corpus conséquent de publications qu'une revue de littérature narrative ne peut aborder. L'objectif de notre analyse était de dresser un bilan précis du traitement des enjeux environnementaux dans les travaux en sciences régionales et, plus précisément, de dégager, au sein de ce corpus, des sous-ensembles d'articles cohérents du point de vue des thématiques et des travaux de référence. Plusieurs outils d'analyse bibliométrique permettent de réaliser des analyses de cooccurrences d'items et ainsi de distinguer des groupes thématiques, d'auteurs ou encore de travaux cités au sein du corpus. Compte tenu de notre objectif, nous avons choisi l'outil Cortext Manager[1], développé par l'Unité Mixte de Recherche LISIS (CNRS - ESIEE - INRAE - Université Gustave Eiffel) car il permet, à la suite d'une analyse de cooccurrences, d'affecter les publications du corpus initial à un groupe (et un seul) parmi les groupes générés (Breucker *et al.*, 2016). Grâce à cette fonctionnalité, nous avons pu mettre en place une méthode d'analyse bibliométrique en trois temps. La première étape a consisté en une analyse bibliométrique portant sur l'ensemble du corpus d'articles à partir d'indicateurs classiques (ex : évolution du nombre d'articles et de citations, répartition du corpus selon les revues). Dans une deuxième étape, nous avons réalisé une analyse des cooccurrences de mots-clés, qui nous a permis de distinguer quatre grands groupes thématiques intitulés *« Sprawl & Land use »*, *« Clusters & Innovation »*, *« Politics & Governance »*, *« Amenities & Migration »*. Une fois les articles du corpus affectés à l'un de ces quatre groupes thématiques, nous avons pu réaliser de nouvelles analyses de cooccurrences portant sur les mots-clés, les expressions

1 Pour en savoir plus : voir le site de Cortext Manager (https://www.cortext.net/projects/cortext-manager/) et la documentation sur cet outil (https://docs.cortext.net/).

et les références citées (analyse des co-citations) pour chacun des groupes. Ces différentes analyses bibliométriques, couplées à une analyse qualitative portant sur un nombre limité d'articles, nous ont permis de mieux caractériser les articles au sein des groupes thématiques. Nous avons pu ainsi décrire ces groupes du point de vue des champs disciplinaires et des types de travaux (ex : empiriques, conceptuels), identifier les principaux courants auxquels ils se réfèrent (ex : écoles de pensée), distinguer les documents phares et dégager des pistes de réflexion sur la manière d'appréhender l'environnement et les thèmes absents dans les articles de notre corpus.

Après une présentation du principe général de l'analyse bibliométrique, cet article présente notre démarche d'analyse bibliométrique appliquée à la prise en compte de l'environnement dans les articles publiés en sciences régionales entre 1999 et 2019, et réalisée à l'aide de l'outil Cortext Manager.

L'approche bibliométrique : l'analyse des cooccurrences de mots-clés et des co-citations

Principe des cooccurrences et terminologie

Dans une approche bibliométrique, l'analyse des cooccurrences permet d'analyser les liens existant au sein d'un corpus de publications. Le principe des cooccurrences consiste à identifier deux à deux tous les liens entre les modalités d'une variable catégorielle. Il peut s'agir des mots-clés mais aussi des auteurs des publications, des termes présents dans du texte libre ou des références citées par les publications du corpus. Dans le cas des mots-clés, toutes les cooccurrences entre mots-clés sont recensées pour chaque publication du corpus, puis dénombrées sur l'ensemble des publications du corpus et, enfin, projetées dans une matrice de contingence, qui sert de base de calcul aux algorithmes pour la construction du graphe ou carte relationnelle (Figure 1).

Sur cette carte relationnelle apparaissent les différents éléments qui composent le graphe :

- Les **items** ou **noeuds** (ici les mots-clés) sont représentés par des cercles. Plus l'item a de poids (ex : nombre d'occurrences) dans le corpus de publications, plus le cercle qui le représente est grand. Dans notre exemple, un mot-clé sera représenté par un cercle d'autant plus volumineux qu'il sera présent dans notre corpus de publications.
- Les **liens** sont représentés par les traits reliant les cercles. Un lien est une connexion ou une relation entre deux items. D'un point de vue visuel, plus le lien entre les items est fort, plus les cercles sont proches sur la carte. Dans notre exemple, la coprésence ou cooccurrence de deux mots-clés dans une publication sera représentée par un trait entre ces deux mots-clés. Et plus une association entre deux mots-clés sera présente dans notre corpus, plus le trait qui relie les deux cercles représentant ces mots-clés sera court.
- Un **réseau** est constitué par l'ensemble des items (ou nœuds) et des liens projetés sur la carte. Il peut être composé de plusieurs clusters ou groupe d'items.
- Un **cluster**, ou un groupes d'items, est une sous-partie d'un réseau, caractérisée par un ensemble de nœuds définis statistiquement comme proches les uns des autres. La proximité entre des nœuds peut être identifiée, par exemple, par une mesure de similarité. Ces différents groupes sont représentés par les couleurs des cercles et des traits qui les relient.

Analyse des cooccurrences de mots-clés

L'analyse des cooccurrences de mots-clés au sein d'un corpus de publications vise, d'une part, à identifier les principaux concepts introduits et les principaux thèmes abordés et, d'autre part, à caractériser les liens entre eux et leur position les uns par rapport aux autres. Cette méthode repose sur l'hypothèse implicite que certains mots-clés présents dans les publications reflètent l'argumentation scientifique des auteurs et que, quand une paire de mots-clés est utilisée pour indexer un grand nombre d'articles, ces mots représentent une forte association entre les problèmes ou les concepts auxquels ils se réfèrent (Rostaing, 1996).

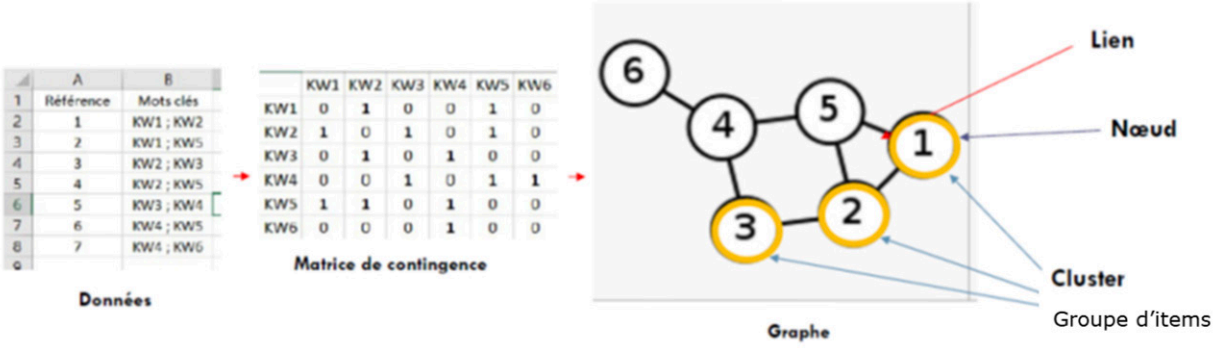


Figure 1. Principe des cooccurrences

Encadré 1. Comparaison de deux outils libres d'analyse bibliométrique

Au cours de notre analyse bibliométrique, nous avons testé et comparé deux outils en accès libre : VOSviewer et Cortext Manager (cf. ci-après).

VOSviewer

Vosviewer, développé par l'Université de Leiden aux Pays-Bas, permet de réaliser, à partir de fichiers bibliographiques issus de *Web Of Science* (WOS), de *Scopus* et de *PubMed* ou à partir de données au format RIS, des cartes de réseaux de cooccurrences. Ces cartes peuvent porter sur les mots-clés, les auteurs, les collaborations internationales. L'outil permet également de réaliser des analyses sur les citations, telles que l'analyse des co-citations (Small, 1973) ou du couplage bibliographique (Kessler, 1963). Il permet également de réaliser des analyses sur des données textuelles, comme le titre ou le résumé (Leiden University's Centre for Science and Technology Studies, 2020). Cet outil présente plusieurs avantages. D'abord, son utilisation est simple et intuitive. Ensuite, en modulant certains paramètres de l'analyse des cooccurrences de mots-clés, il est possible de créer une arborescence thématique des termes utilisés dans les publications et ainsi d'identifier des groupes, puis des sous-groupes thématiques créés lors de la répartition automatique dans les clusters.

Logiciel en java de l'Université de Leiden

- Cooccurrences mots clés
- Coauteurs, collaborations internationales
- Citations
- Couplage bibliographique
- Cooccurrences données textuelles

Points forts : interface conviviale et prise en main rapide
Téléchargement : https://www.vosviewer.com/

CORTEXT

La plateforme **Cortext Manager** a été créée et développée par l'Unité Mixte de Recherche LISIS (CNRS-ESIEE-INRAE-Université Gustave Eiffel). Comme VOSviewer, elle permet de réaliser des cartes relationnelles issues des principales bases bibliographiques. Sa prise en main, moins intuitive, nécessite un peu plus de temps d'apprentissage. Mais elle présente plusieurs avantages. Elle permet d'utiliser des scripts spécifiques pour le comptage, l'évolution historique des données ou l'analyse lexicale. Elle permet également d'enrichir la base de données relatives au corpus avec des variables externes ou des variables que l'on a créées au cours des analyses Cortext, ce qui ouvre un large éventail de possibilités. Enfin, l'outil Cortext permet d'affecter les publications du corpus à un groupe (et un seul) parmi les groupes générés par les cooccurrences de mots-clés ou par les co-citations.
Notre choix final s'est porté sur l'outil Cortext Manager qui intègre cette possibilité d'affecter les publications à un groupe thématique, fonctionnalité précieuse compte-tenu des objectifs de notre analyse.

Plate-forme collaborative en ligne développée par l'unité LISIS de l'INRAE

- Données quantitatives: comptages (nombre / parts de publications selon différents critères) et mesures
- Données qualitatives : cartes réseaux basées sur l'analyse de cooccurrences d'items (clusters, heatmap, historical map…)

Points forts : nombreuses fonctionnalités d'analyse lexicale, possibilité d'enrichir le corpus avec de nouvelles variables, projection des publications dans les clusters.

Accès Cortext Manager v2
Documentation: https://docs.cortext.net

Passerelles

Les deux outils permettent d'exporter les jeux de données sous différents formats : cvs, xls, gefx, pdf, jpg, png, svg…

Analyse des co-citations

L'analyse des co-citations s'appuie, quant à elle, sur la section bibliographique des publications du corpus étudié (Marshakova, 1973 ; Small, 1973). Plus précisément, elle consiste à comparer les références bibliographiques des publications du corpus et à identifier les récurrences par une mesure de co-citations. Celle-ci correspond au nombre de fois où deux références bibliographiques apparaissent simultanément dans l'ensemble des publications du corpus (Figure 2).
L'analyse des co-citations repose sur l'idée que les groupes de documents cités par les publications du corpus constituent les « souches de littérature-coeur » (Rostaing, 1996). Une fois que la structuration des références bibliographiques a été identifiée par l'analyse des co-citations, il est ensuite possible de connaitre les articles du corpus qui citent ces différents groupes de références bibliographiques. Ainsi, l'analyse des co-citations permet d'identifier différents groupes de références « souches » au sein des documents cités par les articles du corpus mais aussi d'identifier, au sein du corpus, les articles qui citent les mêmes groupes « souches » et les mêmes travaux dans la littérature.

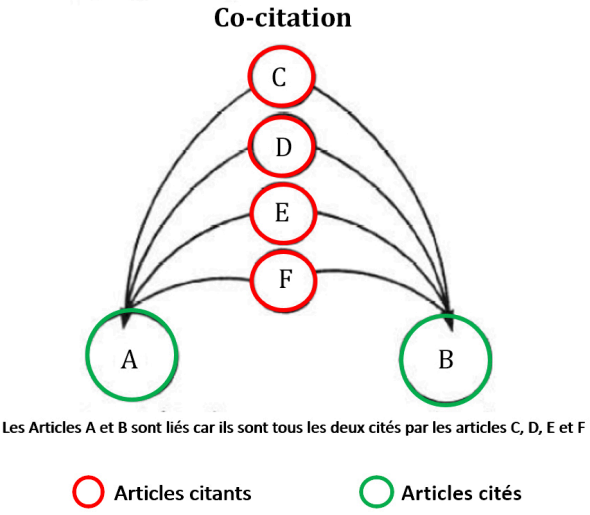


Figure 2. Principe des co-citations d'après Garfield (2001)

Un exemple d'application : la prise en compte de l'environnement dans les articles en sciences régionales (1999-2019)

Une étape liminaire essentielle : la constitution du corpus d'articles

Avant de mener l'analyse bibliométrique, la première étape a consisté en la collecte des articles à traiter (Figure 3).

Pour identifier les travaux en sciences régionales sur lesquels allait porter notre analyse, nous avons appliqué la méthodologie proposée par Surinach *et al*. (2003) qui ont réalisé un état des lieux des travaux publiés dans les revues en sciences régionales entre 1990 et 2000. Ainsi, nous avons établi une première liste de revues en sciences régionales en croisant les titres inclus dans la base Econlit des revues en économie avec ceux répertoriés dans une liste de catégories WOS[2] relevant des sciences régionales. Ensuite, nous avons sélectionné, dans la liste de revues obtenue, les revues dont les titres et/ou les *Aims & Scope* font explicitement référence aux notions de région, de croissance régionale ou de développement régional. Nous avons alors abouti à une liste de 18 titres de revues (Tableau 1).
Nous avons ensuite extrait certains articles de ces revues, en utilisant les fonctionnalités de *Web Of Science* (WOS). Le

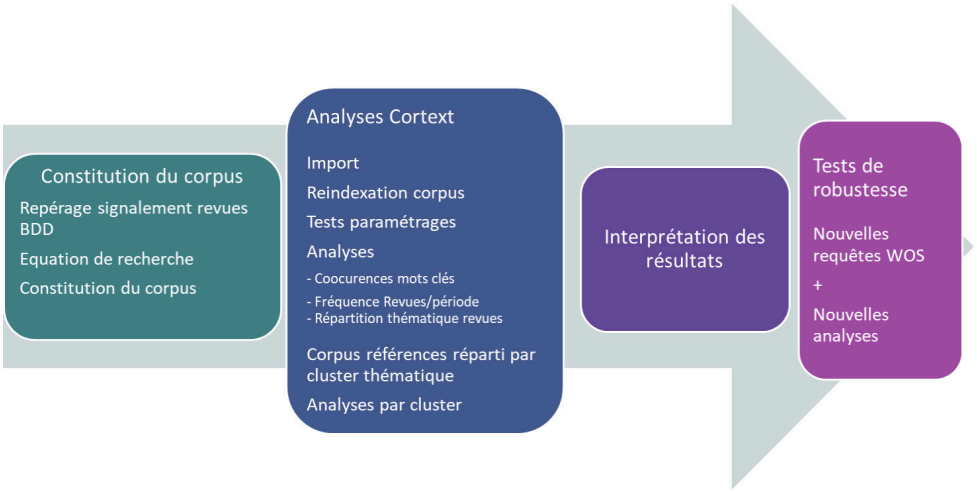


Figure 3. Présentation des étapes de l'analyse bibliométrique

2 Considérant que la valence Economie était apporté par la base Econlit, nous avons pour notre part retenu seulement trois catégories WOS différentes : Geography, Regional & Urban Planning et Urban studies.

choix de cette base bibliographique a découlé de nos besoins pour mener l'analyse bibliométrique. En effet, pour la réalisation de l'analyse de co-citations, nous devions être en mesure d'exporter les données citationnelles liées aux publications. Or, celles-ci sont disponibles uniquement dans les bases WOS et *Scopus*. Nous avons arrêté notre choix sur le WOS car, pour les 18 revues sélectionnées, le nombre de publications liées était supérieur et les données citationnelles mieux harmonisées.

Parmi l'ensemble des articles publiés dans ces 18 revues, nous avons cherché à identifier les articles abordant à la fois l'environnement et le développement régional ou la croissance régionale. Pour cela, nous avons élaboré une requête de sélection des articles en deux temps. La première requête, assez large, ciblait les termes liés à l'environnement couplés avec les termes liés au développement régional, avec une recherche portant sur le champ topic (titre, résumé et mots-clés) du WOS.

Tableau 1. Répartition des articles du corpus entre les 18 titres de revues

Titre	Nbre d'articles	%
Annals of regional science	62	5,4 %
Cambridge journal of regions economy and society	26	2,3 %
Economic geography	28	2,5 %
Environment and planning A	243	21,2 %
Growth and change	53	4,6 %
International journal of urban and regional research	83	7,3 %
International regional science review	30	2,6%
Journal of economic geography	32	2,8 %
Journal of geographical systems	20	1,8 %
Journal of regional science	45	3,9%
Journal of urban affairs	43	3,8 %
Journal of urban economics	25	2,2%
Papers in regional science	41	3,6 %
Real estate economics	6	0,5 %
Regional science and urban economics	38	3,3 %
Regional studies	142	12,4 %

À l'issue d'un balayage aléatoire des résultats, nous avons constaté que les termes « *nature* » et « *environment* », polysémiques, renvoyaient beaucoup de bruit, c'est-à-dire que très peu d'articles les employaient (notamment dans le résumé) dans le sens « écologique » du terme. Pour confirmer cette première impression, nous avons utilisé le script Terms extractions de Cortext Manager. Celui-ci nous a permis de visualiser l'ensemble des expressions intégrant ces deux termes pivots et le nombre d'occurrences associées dans le corpus. Nous avons ensuite remanié notre équation de manière à écarter les termes inadéquats.
Nous les avons remplacés par « *natural* » et « *environmental* », et les avons réintégrés en recherche sur les mots-clés uniquement. D'autres termes ont également été exclus de la recherche, comme « *natural experiment* » ou « *built environment* » (Figure 4). Cette phase d'aller-retour entre les résultats de la requête et l'élaboration de la requête était essentielle pour réduire autant que faire se peut la présence d'articles hors sujet.

```
(TS=("region*" OR "growth" OR "development" OR "location*") AND TS=("ecolog*" OR
"environmental" OR "sustainab*")) OR((TS=("natural" NOT "natural experiment") OR (AK=
"nature" OR KP= "nature") OR TS=("nature" NEAR "conservation")) AND TS=("region*" OR
"growth" OR "development" OR "location*")) OR (((AK="environment" OR KP="environment")
NOT (AK="built environment" OR KP="built environment")) AND TS=("region*" OR "growth" OR
"development" OR "location*"))
```

Figure 4. Formulation finale de la requête d'extraction du corpus

Enfin, nous avons limité notre recherche aux articles originaux et exclu volontairement les documents de type revues de littérature (*reviews*) pour ne pas fausser le résultat des analyses. En effet, le nombre de citations est généralement plus élevé pour ce type de document, et les bibliographies plus conséquentes peuvent perturber l'interprétation des analyses de co-citations ou de couplage bibliographique, en mettant davantage l'accent sur les travaux les plus anciens au détriment des thèmes émergents. Néanmoins, une analyse indépendante des revues de littérature pourrait apporter un éclairage utile pour les travaux futurs.
La période couverte était 1999-2019. Au final, notre corpus était constitué de 1 145 articles. Une fois notre corpus de publications constitué, nous avons exporté nos publications au format *plain text* en sélectionnant le filtre *Full record and cited references*. Les données bibliographiques compressées ont ensuite été importées dans le logiciel Cortext Manager au format ISI.

Les analyses bibliométriques simples menées sur l'ensemble du corpus d'articles

L'analyse bibliométrique, réalisée sur les 1 145 articles de notre corpus, comprend plusieurs étapes. Dans un premier temps, nous avons réalisé une analyse globale de l'ensemble du corpus avec des indicateurs uni-variés classiques du type de ceux générés par la fonction *Analyse results* du WOS : répartition du corpus selon les titres de revue ou encore selon les pays d'affiliation des auteurs. Le nombre moyen de citations a été extrait à partir de la fonction *Citation report* du WOS.
L'analyse temporelle de l'ensemble du corpus montre que le nombre d'articles fluctue d'une année sur l'autre. Néanmoins, il tend à augmenter au cours de la période, passant de 25 articles par an sur les cinq premières années de la période à environ 80 articles par an sur les cinq dernières années (Figure 5). De même, le nombre moyen de citations par article déjà publié augmente fortement au cours de la période, reflétant ainsi une expansion de ce champ de recherche.
Le script *Cortext Epic Epoch* nous a également permis d'analyser l'évolution historique de la répartition du corpus entre les 18 titres de revues. Dans ce script, une chronologie des fréquences est construite pour chaque titre de revue, sui-

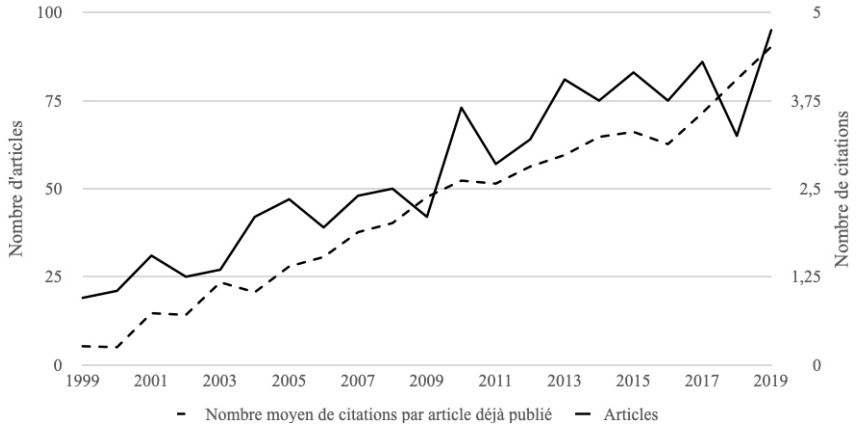


Figure 5. Évolution du nombre d'articles et de citations (1999-2019)

vant une périodicité définie par l'utilisateur. Un graphique du type *bump graph* nous a ensuite permis de visualiser l'évolution dans le temps de cette fréquence et du classement des titres de revue (Figure 7). Sur ce graphique, on observe ainsi que la répartition des articles de notre corpus est relativement stable pour les revues qui éditent le plus d'articles sur l'environnement tout au long de la période d'analyse (*Environment and Planning A, Urban Studies* et *Regional Studies*), mais qu'elle fluctue beaucoup plus pour les autres titres de revues.

Encadré 2. Méthodes d'analyse utilisées via la plateforme Cortext Manager

La plateforme Cortext Manager propose de nombreuses fonctionnalités (Figure 6). Pour notre étude, nous n'avons utilisé qu'une partie de ces scripts pour réaliser différents types de tâches :

- Les scripts *Corpus explorer, Data curation* et *Query* : pour visualiser des données ; fusionner, renommer ou supprimer des champs ; effectuer des requêtes et des extractions des variables à différents formats.
- Les scripts *Term extractions, Corpus list* et *Corpus term indexer* : pour extraire des termes afin de constituer, modifier (ou importer) des lexiques ; indexer des champs textuels (titre, résumé) ou des listes de termes (Keywords).
- Le script *Network mapping* : pour construire des réseaux de cooccurrences (mots-clés, expressions et co-citations) ; visualiser selon des critères statistiques les termes proches ; les représenter graphiquement.
- Les scripts *Epic Epoch* et *Contingency matrix* : pour représenter l'évolution d'une variable sur une période choisie ou représenter la corrélation entre deux variables.

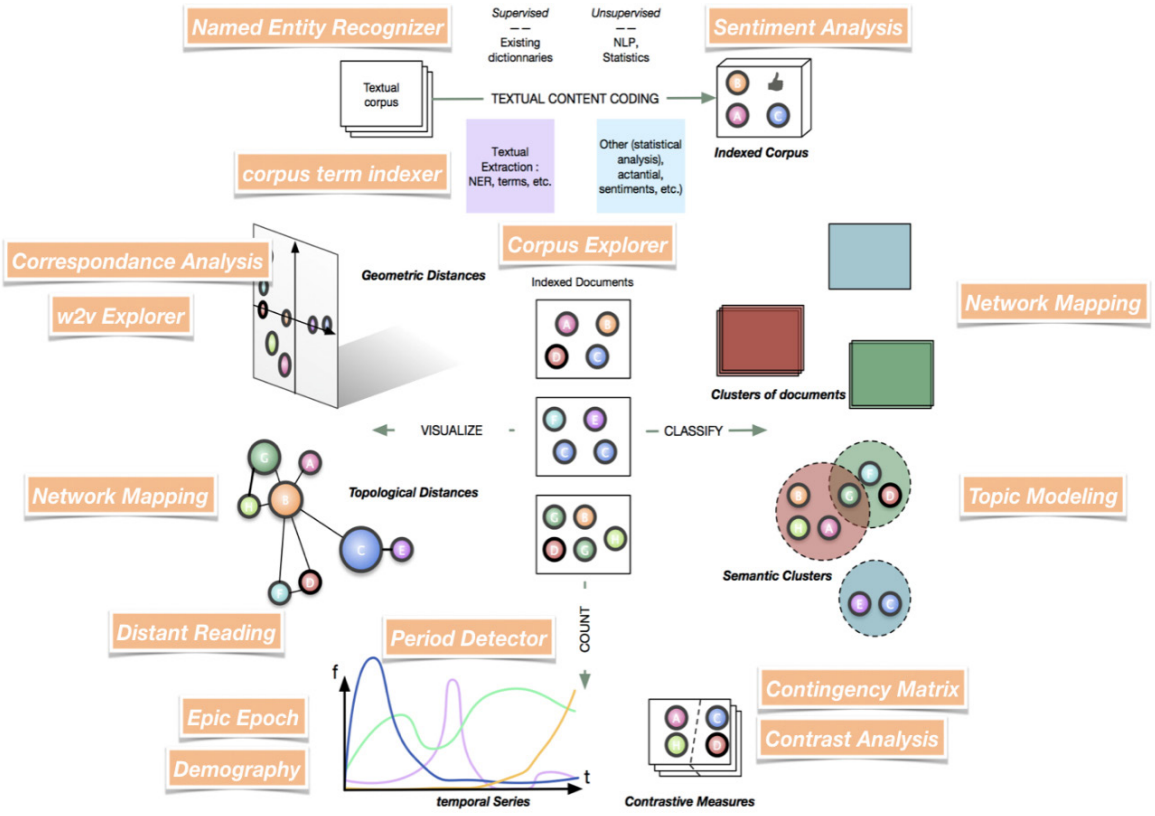
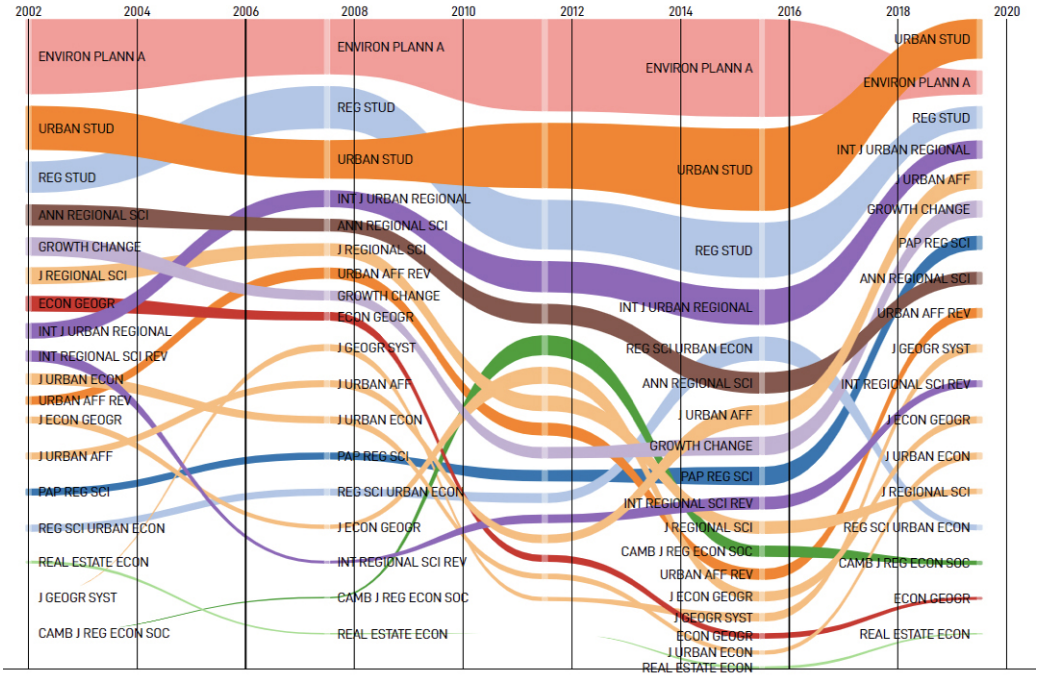


Figure 6. Principales méthodes d'analyses proposées dans Cortext Manager (Source : https://docs.cortext.net/)



Paramétrage: Script Epic Epoch
Size of the hierarchy: 20
Number of time slices: 5
Time slices distribution : homogeneous

Figure 7. Évolution de la répartition des 18 titres de revues du corpus

L'identification de groupes thématiques au sein du corpus d'articles

L'analyse des cooccurrences de mots-clés

Compte tenu de la taille conséquente de notre corpus, nous avons fait le choix de réaliser, dans une deuxième étape, une analyse des cooccurrences de mots-clés[3] afin de distinguer, au sein du corpus, quelques grandes thématiques. Nous avons basé cette analyse sur les « Keywords auteur » et les « Keywords Plus ». Les « Keywords auteur » correspondent aux mots-clés originaux, fournis par les auteurs. Les « Keywords Plus » sont des mots-clés générés automatiquement par un algorithme informatique de WOS qui correspondent à des mots ou des phrases qui apparaissent fréquemment dans les titres des références bibliographiques citées. Alors que les « Keywords auteur » sont plus complets pour représenter le contenu d'un article, les « Keywords Plus » sont particulièrement adaptés à l'analyse bibliométrique pour étudier la structure des connaissances des domaines scientifiques (Zhang *et al.*, 2016).

Avant d'initier l'analyse des cooccurrences, il a été nécessaire de procéder à une étape d'examen et de nettoyage des mots-clés, puis de ré-indexation du corpus. Au cours de cette étape, nous avons fusionné les termes « Keyword auteur » et « Keyword plus » sans doublon avec le script *Data curation*. Les formes singulier/pluriel, avec ou sans tirets ont été harmonisées. Les mots-clés géographiques ont été retirés et les codes de classification JEL remplacés par les mots-clés correspondants. Le tableau de correspondances constitué a été enregistré au format csv (*Dictionnary of equivalent string*) puis utilisé pour réindexer le corpus avec le script *Corpus list indexer*. Le script *List builder* a permis de créer la nouvelle liste constituée de plus de 3 800 mots-clés, utilisée ensuite pour l'analyse des coocurrences réalisée avec le script *Network mapping*.

Ce dernier script permet de configurer et de personnaliser les analyses de cooccurrences en fonction du type de données et du résultat attendu. Les principaux paramètres modulables sont les suivants :

- l'algorithme de mesure de proximité (*raw, chi2, cosine, distributional…*), qui est adapté au type de réseau, homogène ou hétérogène. On parle de réseau hétérogène lorsque les liens entre les nœuds relient deux champs dif-

3 La taille conséquente du corpus de publications rend difficile la réalisation d'une analyse des cooccurrences de termes ou d'expressions dans du texte libre (titres et résumés) sur l'ensemble des publications. En effet, il est indispensable de réaliser, avant l'analyse des cooccurrences, une analyse lexicale approfondie de manière à regrouper les expressions similaires, supprimer les éléments non pertinents, etc. Sur un corpus de cette taille, cette étape aurait été trop fastidieuse si elle avait porté sur les expressions plutôt que sur les mots-clés.

férents (auteurs/KW, par exemple). Pour les réseaux homogènes analysant les cooccurrences Keywords/Keywords, on utilise l'algorithme *distributional* ;

- l'algorithme de détection de communauté : nous avons utilisé *Louvain resolution*, préconisé par le support de la Plateforme Cortext Manager pour sa stabilité ;
- l'algorithme de classement qui permet de faire varier la taille des nœuds en fonction d'une valeur numérique. Par exemple, la taille de chaque nœud d'un groupe de co-auteurs peut être proportionnelle au nombre total de publications d'un auteur ou au nombre total de cooccurrences dans le corpus ;
- le filtrage des nœuds et des liens permet d'augmenter ou de réduire le nombre de groupes et/ou de restreindre l'analyse ou la représentation graphique à une partie seulement du corpus. Il est notamment possible d'indiquer un seuil, un nombre de *top nodes* correspondant au nombre d'items représentés par un nombre décroissant d'occurrences.

L'analyse de cooccurrences de mots-clés, réalisée sur les mots-clés apparaissant au moins 5 fois dans l'ensemble des articles du corpus, nous a permis d'identifier quatre groupes thématiques intitulés[4] *« Politics & Governance », « Amenities & Migration », « Cluster & Innovation »* et *« Sprawl & Land use »*. Une carte relationnelle permet de visualiser leur composition ainsi que leur position relative (Figure 8[5]). Une première analyse de cette carte et des fichiers .csv liés nous a permis d'étudier les mots-clés qui contribuent à ces différents groupes thématiques et leur importance (nombre d'occurrences) et ainsi de commencer à caractériser et comparer le contenu des groupes thématiques (Tableau 2).

Une fois déterminée la structure du corpus en groupes thématiques, Cortext Manager permet d'affecter les articles du corpus à ces groupes, par comparaison entre les mots-clés de chaque article, et la composition en termes de mots-clés de chacun des groupes. Un article peut être assigné à aucun, un ou plusieurs groupes à la fois. Une nouvelle table est générée dans la base de données et peut être utilisée comme un nouveau champ pour une analyse ultérieure. L'option « Attribuer un cluster unique à chaque enregistrement (meilleure correspondance) » de Cortext Manager permet, en outre, d'assigner un groupe unique à chaque publication. Dans ce cas, chaque article est associé au groupe le plus proche en termes de mots-clés, sur la base d'une mesure de similarité. Ce poids, permettant de mesurer la relation entre un document et un groupe, est calculé en additionnant les *Term Frequency-Inverse Document Frequency* (TF-IDF[6]) de chaque mot-clé contribuant au groupe et également présent dans les mots-clés de l'article pondérés par une mesure de l'importance du mot-clé dans le groupe. Le groupe le plus similaire à un article donné est alors sélectionné, pour autant que cette similarité soit supérieure à un seuil prédéterminé, ce qui signifie que certains articles du corpus peuvent encore rester non attribués (INRAE/LISIS Unit, 2020a).

Nous avons réalisé une série de tests de paramétrage, qui nous ont permis entre autres de tester la robustesse des groupes (ou clusters) constitués. Les paramétrages faisant varier le nombre de groupes identifiés au sein du réseau, le choix des paramètres s'est fait de manière itérative au moyen d'un arbitrage entre le nombre de groupes et le nombre d'articles rattachés à un de ces groupes. Nous avons fixé les paramètres de manière à obtenir un nombre de groupes relativement réduit, ce qui permettait de dégager des grandes thématiques, tout en conservant le maximum d'articles du corpus classés dans l'un des groupes (ou à réduire le nombre d'articles orphelins).

Cette fonctionnalité de projection des publications dans les groupes nous a permis de répartir les articles de notre corpus entre les quatre groupes thématiques identifiés par l'analyse des cooccurrences de mots-clés. Au final, sur les 1 145 articles du corpus initial, l'analyse des cooccurrences de mots-clés et la fonction de projection des publications nous a permis de classer 1 085 articles entre les quatre groupes thématiques (Tableau 2).

On observe que les groupes *« Politics & governance »* et *« Amenities & Migration »*, auxquels contribuent respectivement 104 et 101 mots-clés, sont ceux auxquels se rattachent le plus grand nombre d'articles. En effet, 443 articles sont rattachés au premier groupe et 308 au second, ce qui représente respectivement 41 % et 28 % des articles classés.

La caractérisation d'ensemble des groupes thématiques

La fonctionnalité de projection des publications, présente

4 Les intitulés des groupes correspondent à la cooccurrence de mots-clés la plus fréquente au sein du groupe.

5 Cette figure correspond à la carte réalisée à partir du fichier GEFX issu de Cortext et modifié avec le logiciel Gephi. Pour des raisons de lisibilité, la carte relationnelle présentée ici ne fait apparaître que les 100 premiers mots-clés (paramètre top nodes) et les relations entre eux. Ces modifications ne concernent pas la topologie du réseau (la taille relative des nœuds, l'épaisseur relative des liens et les distances relatives sont identiques à celles générées par Cortext).

6 Le TF-IDF est une méthode de pondération utilisée dans la fouille de textes. Cette mesure statistique permet d'évaluer l'importance d'un terme contenu dans une publication, comparativement à son importance dans le corpus.

Paramétrage: Script Network mapping
100 top nodes
Weight : frequency
Detection algorithme : Louvain resolution
Proximity Threshold: '0.1'
Proximity measure : distributional

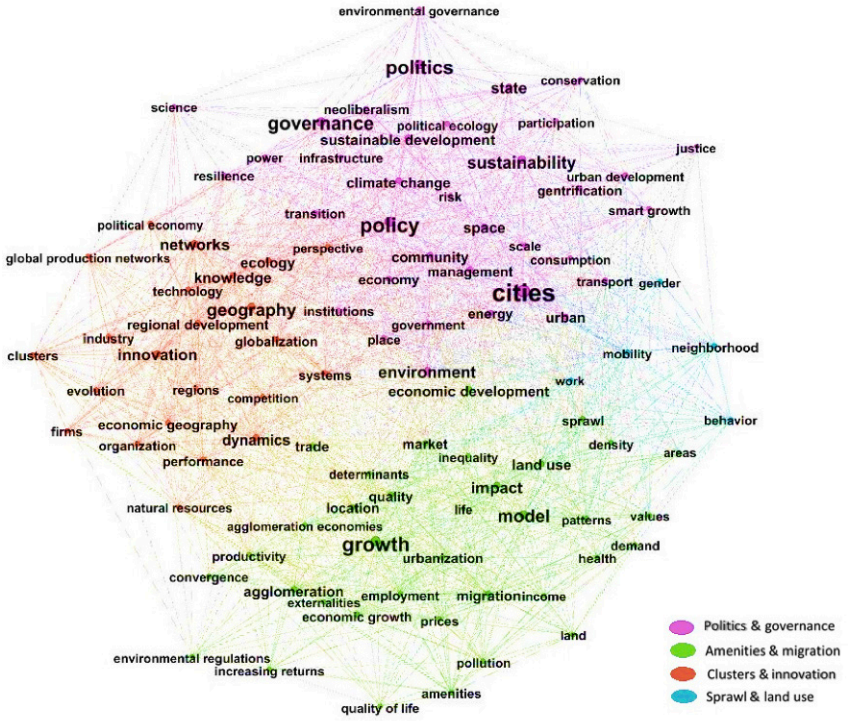


Figure 8. Principales cooccurrences Keywords Auteur et Keyword plus

Tableau 2. Composition des quatre groupes thématiques

	Politics & Governance	Amenities & Migration	Clusters & Innovation	Sprawl & Land use	Ensemble
Nombre de mots-clés	104	101	79	55	339
10 principaux mots-clés (nombre d'occurrences)	Cities Policy Politics Governance Sustainability Environment State Space Economic development Climate change	Growth Model Impact Agglomeration Migration Market Location Urbanization Trade Quality	Geography Innovation Networks Knowledge Dynamics Ecology Economic geography Regional development Clusters Performance	Land use Sprawl Mobility Density Health Neighborhood Behavior Transport Inequality Patterns	
Nombre d'articles classés	443 (41 %)	308 (28 %)	208 (19 %)	126 (12 %)	1 085 (100 %)

dans l'outil Cortext Manager et qui permet de les affecter à un groupe thématique et un seul, ouvre de nombreuses possibilités d'analyse. Elle nous a permis par exemple d'identifier les titres des revues qui contribuent le plus, à travers les articles qu'elles ont publiés, à chacun des groupes thématiques. Le script *Contingency Matrix* de Cortext Manager permet de visualiser la distribution conjointe selon deux champs des publications du corpus à travers une matrice de contingence. Celle-ci montre le degré de corrélation entre deux variables : les cellules rouges sont les plus corrélées, les cellules bleues sont corrélées négativement et les blanches ne présentent aucune corrélation. Appliqué à l'analyse de la relation entre nos groupes thématiques et les titres de revues, ce script nous a permis de visualiser rapidement la dépendance statistique entre ces deux caractéristiques (Figure 9).

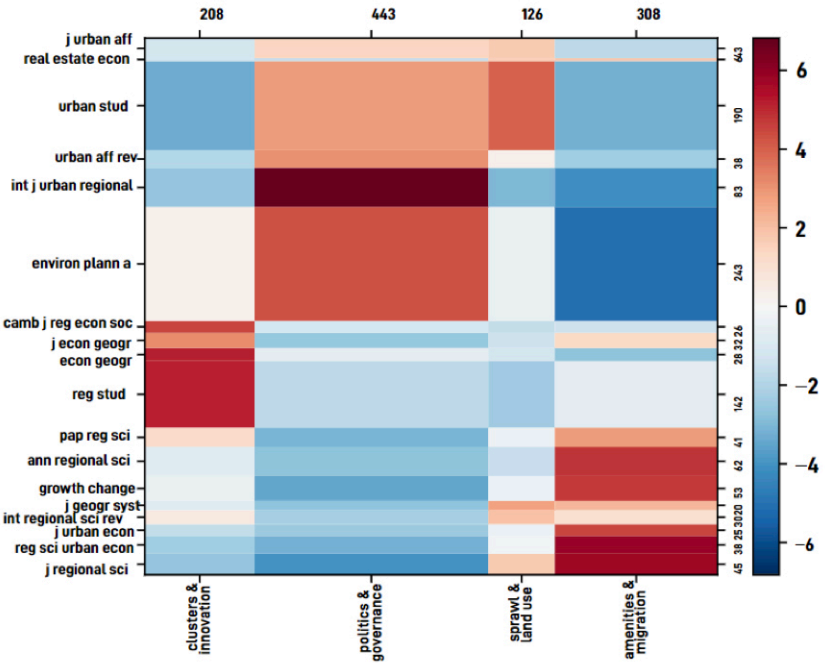
Ainsi, on voit que les travaux rattachés au groupe thématique *« Politics & Governance »* sont davantage publiés dans les revues *International Journal of Urban and Regional Research, Environment & Planning A, Urban Studies* et *Urban Affairs Review* alors que les articles du groupe *« Amenities & Migration »* sont davantage publiés dans les revues *Journal of Regional Science*, *Regional Science and Urban Economics*, *Annals of Regional Science, Growth Change* et *Journal of Urban Economy*. En outre, les revues *Regional Studies, Economic Geography, Cambridge Journal of Regions Economy and Society* et *Journal of Economic Geography* sont surreprésentées dans les travaux du groupe *« Clusters & Innovation »* alors que ce sont les revues *Urban Studies* et *Journal of Geographical Systems* qui le sont pour les travaux du groupe *« Sprawl & Land use »*.

Tests de robustesse

Au cours de l'analyse de l'ensemble du corpus, nous avons réalisé plusieurs tests de robustesse, en sus des tests de paramétrage de l'analyse des cooccurrences de mots-clés. Les tests de robustesse ont consisté notamment à modifier la requête initiale lancée dans le WOS et à analyser les changements induits sur la classification en groupes thématiques. Nous avons réalisé deux types de modification de la requête.

Afin de tester la pertinence de notre requête par rapport à notre questionnement, nous avons, en premier lieu, élargi notre requête et observé comment la classification évoluait si on retenait, parmi les articles publiés par les 18 revues sur la même période, tous ceux qui abordaient l'environnement et non uniquement ceux qui abordaient à la fois l'environnement et le développement régional ou la croissance régionale. Dans ce cas, la requête a été modifiée de manière à ne conserver que la partie « environnement » de la requête, en supprimant les mots-clés de la partie « développement régional » (Figure 10).

Cette modification ajoute 853 nouvelles publications. À travers un balayage rapide de ces articles, nous avons pu voir que la nouvelle version de la requête amenait des publications trop éloignées de notre questionnement initial (ex : environnement et santé). Ceci s'explique par le fait que certaines des revues sélectionnées, à l'image de la revue *Environment & Planning A,* publient des travaux en sciences



Paramétrage: Script Contingency matrix 20 top nodes

Figure 9. Matrice de contingence entre les groupes thématiques et les titres de revues

```
(TS=("ecolog*" OR "environmental" OR "sustainab*")) OR (TS=("natural" NOT "natural experiment")) OR (AK= "nature" OR KP= "nature") OR (TS=("nature" NEAR "conservation")) OR ((AK="environment" OR KP="environment") NOT(AK="built environment" OR KP="built environment"))
```

Figure 10. Formulation de la requête excluant les termes liés à l'économie régionale

régionales mais pas exclusivement (ex : économie institutionnelle, économie du développement). Nous avons donc fait le choix de conserver les deux parties de la requête, la partie « environnement » et la partie « développement régional », pour rester cohérent avec notre questionnement initial. Le deuxième type de modification de la requête visait à analyser la sensibilité de notre classification à l'intégration de nouveaux termes liés à l'environnement. Nous avons ainsi vérifié que l'ajout de termes tels que « pollution » ou « biodiversité » à la requête de base ne modifiait le corpus de publications que de façon marginale (Figure 11). La requête intégrant le terme « pollution » en Topic renvoie 120 publications, dont 69 qui étaient déjà présentes dans le corpus et 51 articles supplémentaires. L'analyse des cooccurrences de mots-clés réalisées avec le même paramétrage mais avec ces 51 nouveaux articles aboutit à l'identification de cinq groupes thématiques. Parmi eux, quatre groupes sont très proches des groupes initiaux et le groupe supplémentaire ne comprend que trois publications et trois mots-clés associés. De la même manière, la requête intégrant le terme *« biodiversity »* en Topic amène 24 articles au total, dont 16 qui figuraient déjà dans le corpus et uniquement 8 articles supplémentaires.

```
(TS=("region*" OR "growth" OR "development" OR "location*") AND TS=("ecolog*" OR "environmental" OR "sustainab*" OR "pollution")) OR((TS=("natural" NOT "natural experiment") OR (AK= "nature" OR KP= "nature") OR TS=("nature" NEAR "conservation")) AND TS=("region*" OR "growth" OR "development" OR "location*")) OR (((AK="environment" OR KP="environment") NOT (AK="built environment" OR KP="built environment")) AND TS=("region*" OR "growth" OR "development" OR "location*"))
```

Figure 11. Formulation de la requête initiale incluant le terme « pollution »

Les analyses bibliométriques menées sur chacun des groupes thématiques : l'exemple du groupe « Cluster & Innovation »

Afin d'affiner encore notre vision du corpus d'articles, dans une troisième étape, nous avons mené des analyses spécifiques sur chacun des groupes thématiques identifiés précédemment. Nous avons réalisé, à partir du logiciel Cortext Manager, une analyse des co-citations ainsi que de nouvelles analyses de cooccurrences de mots-clés et de cooccurrences d'expressions. Nous avons croisé ces analyses avec un dernier niveau d'analyse plus qualitative. Nous présentons ici quelques-uns des résultats obtenus à partir de l'exemple du groupe thématique *« Cluster & Innovation »*.

Analyse des Co-citations

Outre l'analyse des cooccurrences de mots-clés ou d'expressions, le script *Network mapping* permet également d'identifier des réseaux de co-citations et de les visualiser à travers des cartes historiques (Figure 12[7]). Sur ces figures en forme de radar, les références bibliographiques sont positionnées en fonction de leur date de publication.
Dans le cas du groupe thématique *« Cluster & Innovation »*, l'analyse des co-citations nous a permis de distinguer quatre groupes de travaux séminaux. La représentation sous forme de radar met en évidence le fait que le groupe qui intègre les souches de littérature les plus anciennes (en l'occurrence Jacobs (1969) et Holling (1973)) perdure au cours du temps. Elle permet également d'observer rapidement que les trois autres groupes de travaux souches sont constitués principalement par des travaux publiés entre 1990 et le début des années 2000.

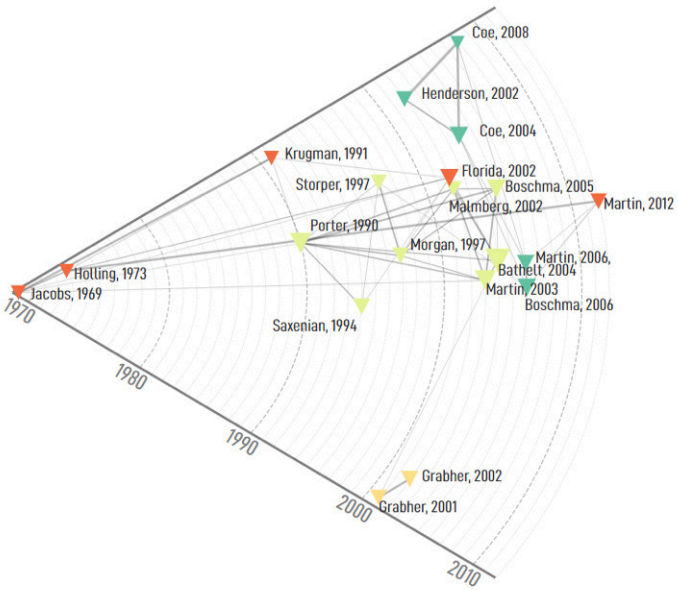
À nouveau, la fonctionnalité de projection des publications nous a permis de rattacher les articles qui contribuent au groupe thématique *« Cluster & Innovation »* à l'un de ces quatre groupes de travaux séminaux et ainsi de regrouper les articles du groupe par communauté de pensée ou type d'approche scientifique.

Analyse des cooccurrences d'expressions du titre et du résumé

Sur le même principe, nous avons réalisé, pour chacun des quatre groupes thématiques, une analyse des cooccurrences d'expressions[8] basée sur les titres et résumés des publications du corpus. L'analyse réalisée sur le groupe thématique *« Cluster & Innovation »* nous a permis d'identifier 58 expressions

7 Nous avons testé différentes analyses de co-citations, avec différents seuils pour le nombre de références bibliographiques à prendre en compte. Ce nombre de 20 a été finalement retenu à l'issue d'un arbitrage entre, d'une part, le nombre d'articles associés aux quatre groupes de travaux séminaux identifiés (en l'occurrence 113 articles parmi les 208 que compte le groupe *« Clusters & Innovation »*) et, d'autre part, la lisibilité de la carte et l'intérêt analytique des groupes constitués.

8 L'analyse de la cooccurrence de termes peut être réalisée sur la base des termes simples (monograms) ET des multi-termes (n-grams) ou des multi-termes seuls. Nous avons fait le choix de nous limiter aux multi-termes car l'intégration de termes simples nécessite un travail supplémentaire pour écarter les termes simples non informatifs (INRAE UMR LISIS, 2020b). Or, l'absence des termes simples peut fausser l'analyse des cooccurrences et les profils thématiques identifiés. Dans notre cas, ce risque est modéré, car l'analyse des cooccurrences d'expressions est une analyse complémentaire, qui vient à la suite d'une analyse fine des mots-clés (et donc des termes simples les plus importants) présents dans le groupe thématique.



Paramétrage: Script Network mapping
20 top nodes
Historical map : true
Weight : cooc-sum
Detection algorithme : automatique (Louvain)
Proximity Threshold : automatique
Proximity measure : distributional

Figure 12. Co-citations des 20 références les plus citées du groupe « *Clusters & Innovation* »

différentes réparties entre 9 sous-groupes thématiques. Nous avons pu visualiser ces différents sous-groupes d'expression et les positionner les uns par rapport aux autres grâce à une nouvelle carte relationnelle (Figure 13).
La fonctionnalité de projection des articles nous a permis d'associer ceux qui contribuent au groupe thématique *« Cluster & Innovation »* à un (et un seul) de ces 9 sous-groupes thématiques. On observe ainsi que, parmi les sous-thèmes identifiés au sein de ce groupe thématique, ceux auxquels contribuent le plus d'articles de notre corpus portent sur l'écologie industrielle et le développement local, sur les districts industriels et l'évolution des clusters, ainsi que sur les filières et l'industrie alimentaire.

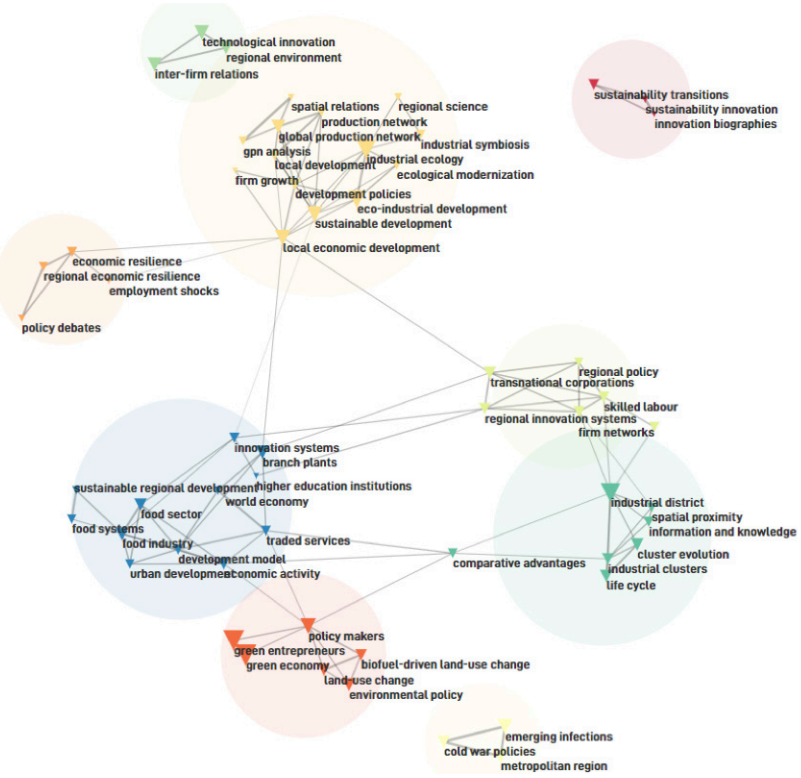
Si elle a permis d'améliorer encore un peu plus l'appréhension de notre corpus initial, l'analyse des cooccurrences d'expressions présentait des limites. Ainsi, elle ne permettait pas d'avoir une vision précise du contenu des articles et de dégager une structuration hiérarchique des questions scientifiques abordées. C'est la raison pour laquelle nous avons fait le choix de croiser cette analyse avec une analyse qualitative, réalisée sur une sélection d'articles au sein du groupe thématique.

Analyse qualitative et grille de lecture

Le découpage de notre corpus de 1 145 publications en quatre groupes thématiques nous a permis d'obtenir des ensembles d'articles cohérents et de tailles plus modestes. Nous avons pu dès lors croiser ces analyses automatiques, réalisées grâce à la plateforme Cortext Manager, avec une analyse qualitative plus fine. Cette dernière étape reposait sur une grille de lecture conçue préalablement. Elle a consisté, d'abord, en un balayage de l'ensemble des articles des quatre groupes thématiques, portant plus précisément sur les titres, mots-clés et résumés. Cette première lecture rapide visait à identifier, autant que faire se peut, la thématique principale abordée par chacun des articles du groupe thématique ainsi que le type de méthodologie dominant. Afin de renseigner de manière plus complète notre grille de lecture, nous avons procédé, ensuite, à une lecture plus approfondie des 25 % les plus citées[9] parmi les publications de chaque groupe thématique. Ceci nous a permis d'avoir une idée plus précise des principaux champs disciplinaires, des questions de recherche, des méthodes d'analyse ou encore des conceptions de l'environnement sous-jacentes.

En ce qui concerne le groupe thématique *« Cluster & Innovation »*, cette phase d'analyse qualitative nous a permis d'identifier cinq sous-thèmes différents et de construire une carte mentale des différentes questions de recherche abordées (Figure 14). La comparaison de cette carte mentale et de la carte obtenue par l'analyse des cooccurrences d'expressions montre des similitudes. Ainsi, la carte des cooccurrences d'expression donne, en première approche, une assez bonne idée des sous-thèmes abordés (ex : clusters et districts industriels, dynamiques d'innovation, écologie industrielle).

9 Afin de ne pas survaloriser les publications les plus anciennes, nous avons classé les articles selon le nombre moyen de citations par an, et non selon le nombre total de citations sur la période.



Paramétrage: Script Network Mapping
150 top nodes
Weight : co-occ
Detection algorithme : Louvain
Proximity Threshold: '0.1'
Proximity measure : distributional

Figure 13. Cooccurrences des principales expressions du titre et du résumé du groupe « Clusters and innovation »

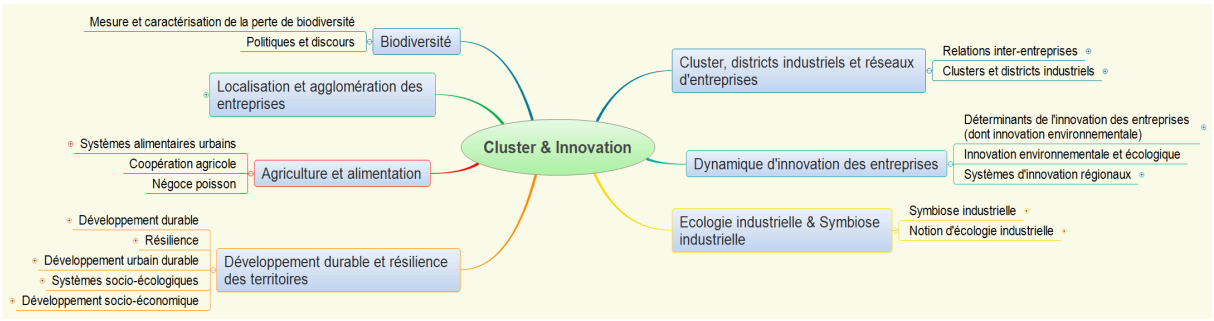


Figure 14. Carte mentale des sous-thèmes abordés par les articles du groupe thématique « Cluster & Innovation »

Néanmoins, si l'on souhaite avoir une vision plus fine des travaux, une analyse qualitative est indispensable. Celle-ci nous amène dans certains cas à regrouper des sous-thèmes qu'une analyse des cooccurrences d'expression distingue, à l'image des sous-thématiques *« economic resilience »* et *« green economy »* que nous avons regroupées dans le thème « Développement durable et résilience des territoires ». À l'inverse, l'analyse qualitative nous conduit à identifier des embranchements et sous-embranchements logiques qu'une analyse des cooccurrences ne permet pas d'identifier directement. Par exemple, les articles portant sur la dynamique d'innovation des entreprises comprennent trois types de travaux : ceux centrés sur les déterminants de l'innovation des entreprises, ceux qui analysent le processus d'innovation environnementale et écologique, ceux qui s'intéressent aux systèmes d'innovation régionaux. Enfin, certains champs (tels que l'approche disciplinaire, la méthodologie ou, dans notre cas, la conception sous-jacente de l'environnement) ne peuvent être identifiés de manière automatique et nécessitent forcément une phase de lecture des publications ou, a minima, des résumés par des experts du domaine scientifique.

Conclusion

L'analyse bibliométrique que nous avons menée visait à analyser un corpus conséquent de publications, constitué de 1 145 articles publiés dans des revues de sciences régionales entre 1999 et 2019. Notre objectif était d'apporter

un éclairage sur la manière dont ces travaux prennent en compte les enjeux environnementaux. Au-delà de l'analyse d'indicateurs uni-variés classiques, l'analyse des cooccurrences de mots-clés nous a permis de distinguer, au sein de cet ensemble de travaux, quatre groupes thématiques différents. Des analyses bibliométriques complémentaires, menées sur chacun de ces groupes et croisées avec des analyses qualitatives sur les principaux articles des groupes, ont permis d'avoir une vision plus précise du contenu des articles et de dégager les principales questions scientifiques abordées ainsi que les groupes de travaux séminaux cités.

Nous avons dû faire face à un certain nombre de difficultés, qui constituent autant de points de vigilance lorsque l'on s'engage dans une analyse bibliométrique. Tout d'abord, le préalable important est de bien formuler les questions en amont et ne pas perdre de vue la problématique initiale ; le risque étant de bifurquer progressivement pour se conformer aux potentialités de l'outil d'analyse. Par ailleurs, sous-estimer l'importance stratégique du travail de constitution du corpus constitue une erreur majeure. Il faut bien prendre en compte sa dimension ainsi que l'importance de la requête initiale qui doit être testée et élaborée en plusieurs temps. Ceci est particulièrement important dans le cas de requêtes larges dans lesquelles certains termes peuvent être polysémiques (« nature » ou « environnement » dans notre cas). En effet, les représentations visuelles peuvent être faussées par l'intégration de publications hors sujet, conduisant ainsi au risque d'une interprétation erronée. De plus, il est essentiel d'évaluer dès le départ les moyens et le temps dont on dispose, car l'investissement nécessaire pour la prise en main de l'outil est non négligeable. Une grande partie de notre travail a consisté à nous former à Cortext Manager. Enfin, il est indispensable d'effectuer des tests de paramétrage. En effet, les résultats peuvent considérablement varier en fonction de la configuration choisie et des types de variables. Il faut donc tester la sensibilité des visualisations pour les adapter aux choix méthodologiques et rester vigilant dans leur l'interprétation.

Ceci étant, l'intérêt de la méthodologie que nous proposons, et qui repose sur certaines fonctionnalités proposées par la plate-forme Cortext Manager, réside en trois principaux points. Elle permet, en premier lieu, d'appréhender pas à pas un corpus de publications très conséquent. Sans cette approche séquentielle, nous aurions renoncé ou nous nous serions limités à une sélection d'articles, les 100 plus cités par exemple. En outre, elle donne, en première lecture, une image plutôt satisfaisante de la répartition du corpus et permet d'appréhender, grâce aux mots-clés, les grandes thématiques abordées. Enfin, l'étape d'analyse intra-groupe permet d'affiner l'analyse thématique initiale et de repérer notamment les travaux séminaux liés. Les limites de cette méthode d'analyse bibliométrique concernent essentiellement l'interprétation des résultats et la nécessité de combiner les visualisations avec d'autres éléments de preuve, liés aux connaissances et jugement intuitif de l'équipe. En effet, les visualisations scientométriques donnent une représentation des points saillants et transcrivent de manière simplifiée et incomplète les données sous-jacentes. Il est donc essentiel de ne pas se limiter aux sorties graphiques et d'analyser les tables de données associées si l'on souhaite avoir une vision à la fois précise et pertinente selon les questions de recherche. Enfin, l'utilisation d'indicateurs bibliométriques ou de cartes relationnelles ne dispense pas de faire une lecture plus fine d'une partie des articles (les plus cités par exemple) pour mieux appréhender le contenu du corpus. ■

Remerciements

Les auteurs remercient Lionel Villard (UMR LISIS) et la plateforme Cortext pour leur appui et leurs réponses aux questions posées concernant l'utilisation de l'outil. Ils remercient également l'équipe éditoriale de NOV'AE et le relecteur anonyme pour leur relecture attentive et leurs suggestions, qui ont permis d'améliorer grandement la qualité de l'article.

Références

Baumont R, Girard A. (2019). Productions Animales dans la production scientifique en élevage et sciences animales. INRA Productions Animales, 32(2):87-94. Disponible sur : https://doi.org/10.20870/productions-animales.2019.32.2.2543. hal-02315856.

Bordignon F (2021). A scientometric review of permafrost research based on textual analysis (1948–2020). Scientometrics, 126(1) : 417-436. Disponible sur : https://doi.org/10.1007/s11192-020-03747-4.

Breucker P, Cointet J, Hannud Abdo A, Orsal G, de Quatrebarbes C, Duong T, Martinez C, Ospina Delgado J P, Medina Zuluaga L D, Gómez Peña D F, Sánchez Castaño T A, Marques da Costa J, Laglil H, Villard L, Barbier M (2016). CorTexT Manager (version v2). [Consulté le 18 mai 2021]. Disponible sur : https://docs.cortext.net.

Garfield E, Merton R K (1979). Citation indexing: Its theory and application in science, technology, and humanities. Vol. 8. New York: Wiley.

Garfield E (2001). From Bibliographic Coupling to Co-Citation Analysis via Algorithmic Historio-Bibliography. A Citationist's Tribute to Belver C. Griffith. Papier présenté à la Drexel University, Philadelphia, PA on November 27, 2001. Disponible sur : http://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf (consulté le 6 mai 2022).

INRAE UMR LISIS (2020a). Network Analysis & Layout. Cortext Manager Documentation [en ligne]. 2020. [Consulté le 24 octobre 2022]. Disponible sur : https://docs.cortext.net/analysis-mapping-heterogeneous-networks/mapping/.

INRAE UMR LISIS (2020b). Terms Extraction. Cortext Manager Documentation [en ligne]. 2020. [Consulté le 24 octobre 2022]. Disponible sur : https://docs.cortext.net/lexical-extraction/.

Kessler, M M (1963). Bibliographic coupling between scientific papers. American Documentation, 14(1):1025. Disponible sur : https://doi.org/10.1002/asi.5090140103.

Leiden University's Centre For Science And Technology Studies (2020). VOSviewer - Visualizing scientific landscapes ». VOSviewer [en ligne]. [Consulté le 12 mars 2020]. Disponible sur : https://www.vosviewer.com/ .

Marshakova, IV (1973). Co-Citation in Scientific Literature: A New Measure of the Relationship Between Publications. Scientific and Technical Information Serial of VINITI, 6:3-8.

Price D (1963). Little Science, big Science. Columbia, New York.

Rostaing H (1996). La bibliométrie et ses techniques [En ligne]. Sciences de la Société (Outils et méthodes). Disponible sur : https://hal.archives-ouvertes.fr/hal-01579948 (consulté le 21 mai 2021).

Small H (1973). Cocitation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for information Science 24(4):265-269. Disponible sur : https://doi.org/10.1002/asi.4630240406.

Surinach J, Duque J. C, Ramos R, Royuela V (2003). Publication patterns in regional and urban analysis: Have topics, techniques and applications changed during the 1990s? Regional Studies 37(4): 351-363. Disponible sur : https://doi.org/10.1080/0034340032000074398.

Tancoigne E, Barbier M, Cointet J.-P, Richard G (2014). Les services écosystémiques dans la littérature scientifique : démarche d'exploration et résultats d'analyse [En ligne]. Rapport Institut National de la Recherche Agronomique. Disponible sur : https://hal.archives-ouvertes.fr/hal-01157253 (consulté le 18 mai 2021).

Zhang J, Yu Q, Zheng F, Long C, Lu Z, Duan Z (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. Journal of the Association for Information Science and Technology 67(4):967-972. Disponible sur : https://doi.org/https://doi.org/10.1002/asi.23437.

Cet article est publié sous la licence Creative Commons (CC BY-SA). https://creativecommons.org/licenses/by-sa/4.0/.

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « NOV'AE », la date de sa publication et son URL.