HAL open science

Fouille de publications scientifiques pour une analyse bibliométrique de l'activité de recherche sur la déforestation

Nathalie Neptune, Josiane Mothe, Julius Akinyemi

To cite this version:

Nathalie Neptune, Josiane Mothe, Julius Akinyemi. Fouille de publications scientifiques pour une analyse bibliométrique de l'activité de recherche sur la déforestation. Brigitte Grau (LIMSI-CNRS); Haïfa Zargayouna (LIPN, UMR CNRS, Université Paris 13). 2ème Rencontre entre chercheurs en Intelligence Artificielle (IA) et en Recherche d'Information (RI) (Journée commune AFIA - ARIA 2017), Dec 2017, Paris, France. , 2017. hal-03120276

HAL Id: hal-03120276

https://hal.science/hal-03120276v1

Submitted on 31 Jan 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.

Journée commune AFIA - ARIA
12 décembre 2017
Paris, France

Fouille de publications scientifiques pour une analyse bibliométrique de l'activité de recherche sur la déforestation

Nathalie Neptune*, Josiane Mothe*, Julius Akinyemi**

*Institut de Recherche en Informatique de Toulouse
**Entrepreneur-In-Residence, MIT Media Lab/Founder-CEO UWINCorp Inc.

Introduction

La déforestation est un phénomène très répandu qui touche des portions de territoires assez importantes surtout dans les régions tropicales. La télédétection permet aux chercheurs de suivre et d'analyser l'évolution spatiotemporel de ce phénomène.

En utilisant la fouille de texte et de méta-données sur les publications scientifiques sur le thème de la déforestation, nous visons à identifier où a lieu la production scientifique sur la déforestation et comment les chercheurs sont connectés entre eux.[1] Grâce à des analyse de réseau, il est possible de mettre en évidence les tendances en termes de collaboration entre les auteurs. [2] L'analyse des mots-clés permet d'identifier les sites touchés par la déforestation auxquels les chercheurs s'intéressent le plus, les forêts tropicales et l'Amazonie, de même que des sujets connexes ayant rapport à l'environnement et à la santé.

Méthodologie

- Analyses géographiques et temporelles
- Fréquences, co-occurrence, groupements, graphes, diagrammes de dispersion
- Croisements avec les données de l'Organisation des Nations Unis pour l'alimentation et l'agriculture.

Résultats

Auteurs, Pays et Domaines

- 16 136 publications de Web of Science Core Collection
- 31 772 auteurs
- 149 pays et territoires

Pays	Publications
USA	10288
Brésil	4152
Royaume Uni	2579
Allemagne	1539
Chine	1487

Tableau des pays ayant le plus de publications

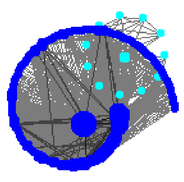
Auteurs	Publications
Fearnside, PM	97
Laurance, WF	73
Houghton, RA	63
Lambin, EF	58
Shimabukuro, YE	57

Tableau des auteurs ayant le plus de publications

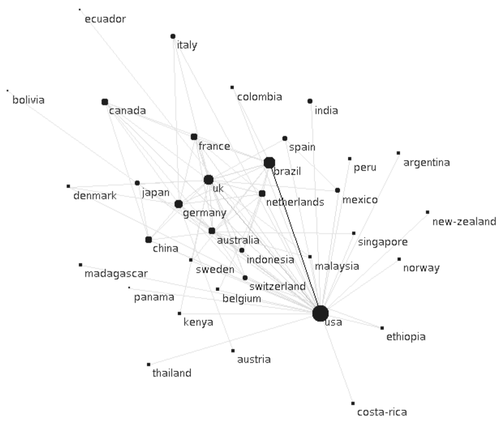
Categories	Publications
Environmental Sciences	3890
Ecology	2812
Geosciences, Multidisciplinary	1571
Environmental Studies	1491
Biodiversity Conservation	1298

Tableau des categories (champ WC du Web of Science) comptant le plus de publications

Réseau des catégories liées à la médecine tropicale et à la parasitologie



Les deux principaux regroupements de catégories: Sciences environnementales et santé (médecine tropicale, parasitologie, santé du travail)

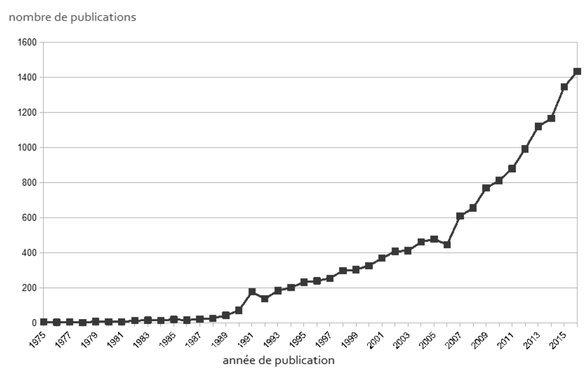


Réseau des pays ayant collaboré avec les États-Unis

Pays et régions géographiques d'intérêt

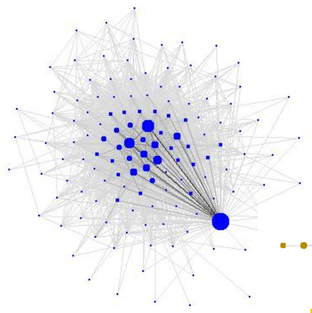
Brazil	Australia	Amazonia	Mediterranean basin
USA	Thailand	Africa	Indian Peninsula
Mexico	Cameroon	Europe	Sumatra
Indonesia	Argentina	America	Borneo
Madagascar	Congo	Rondonia	Sub-sahara
Costa Rica	Malaysia	Mato Grosso	

Nombre de publications par année



Collaboration entre auteurs et pays

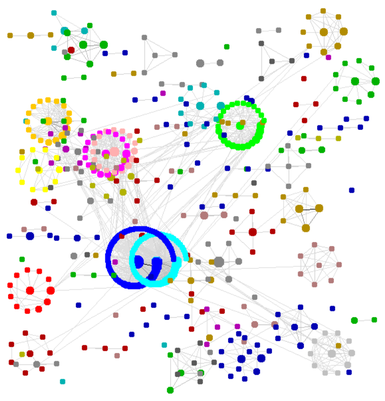
Réseau de collaboration entre les pays



Forte tendance collaborative entre pays et dominance des États-Unis (le plus grand nœud)

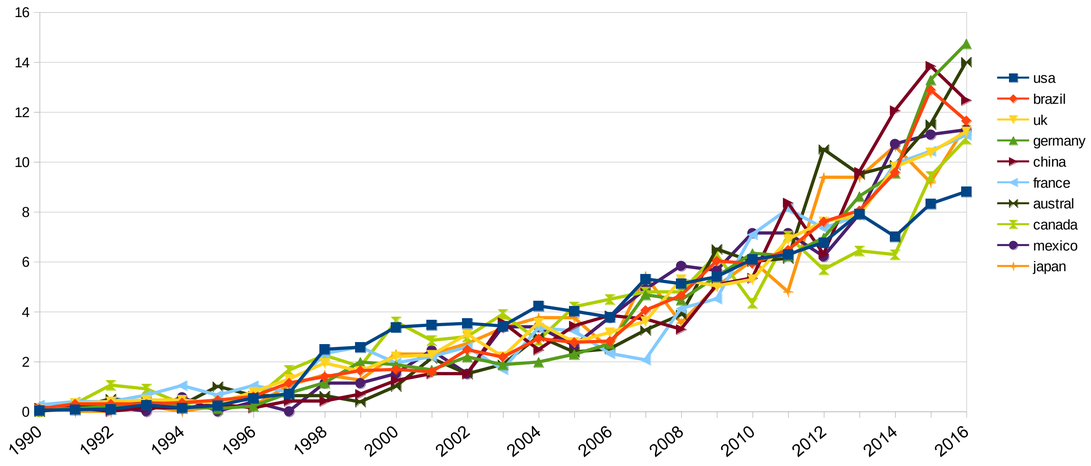
Plusieurs groupements d'auteurs qui collaborent souvent.

Les groupements sont interconnectés.



Réseaux de collaboration entre les auteurs

Évolution de la production de publications par pays



Rapport entre le nombre de publications pour chaque pays sur le nombre total de publications pour le pays, par année

Conclusion et Perspectives

La fouille de texte des données et méta-données sur les publications scientifiques en rapport à la déforestation a permis d'avoir un aperçu de l'évolution de l'activité de recherche sur ce sujet au fil des années ainsi que la collaboration quasi-généralisée entre les pays, et les nombreux réseaux de collaboration entre auteurs. Une augmentation quasi-régulière du nombre de publication est constatée d'une année à l'autre.

Le Brésil ressort comme un acteur important tant par la contribution de ses auteurs que par les mentions qui sont faites du pays dans les publications analysées.

Questions auxquelles les données permettront de répondre

- **Comment les thèmes de recherche évoluent avec le temps?**
- **Quelles sont les institutions les plus productives?**
- **Quelles institutions collaborent entre elles?**
- **Est-ce que les pays les plus affectés par la déforestation contribuent davantage à la production de publications scientifiques?**
- **Est-ce que chaque pays en particulier s'intéresse à des sujets spécifiques?**

Autres analyses

- Extraction d'entités nommées des résumés pour identifier les sujets spécifiques d'intérêt
  - Maladies
  - Plantes
  - Animaux
- Diversité et uniformité des mots-clés
- Types de modèles les plus utilisés

Références

- [1] Borgman, C. L. (1989). Bibliometrics and scholarly communication. Communication Research, 16(5), 583.
- [2] Mothe, J., Chrisment, C., Dkaki, T., Dousset, B., & Karouach, S. (2006). Combining mining and visualization tools to discover the geographic structure of a domain. Computers, Environment and Urban Systems, 30(4), 460-484.

IRIT CNRS - INPT - UPS - UT1 - UTM

Remerciements à la Fondation Schlumberger et à l' École Doctorale 475 Mathématiques, Informatique, Télécommunications de Toulouse et à FabSpace 2.0 (programme Recherche et Développement Horizon 2020 - Convention n° 693210).

UNIVERSITÉ TOULOUSE III PAUL SABATIER
Université de Toulouse