

Analyse de l'interdisciplinarité dans les réseaux d'acteurs scientifiques

Michel de Moura, UVSQ

Henri Bretel, Université Paris-Saclay

1. Constitution du corpus
2. Méthodologie de traitement
3. Visualisation des réseaux
4. Discussion et prochaines étapes

1. La constitution du corpus en bref

Objectif :

- Etablir le corpus des publications relevant du DIM, c'est-à-dire :
 - ✓ En sciences du patrimoine et des matériaux anciens
 - ✓ dont au moins un des auteurs est issu d'une institution francilienne
 - ✓ sans limitation de date

Sources :

- Google Scholar, HAL, Scopus, Web of Science

Outil de collecte :

- Zotero

Résultats :

- 5188 publications collectées et vérifiées
- 4113 validées pour traitement

1. Corpus : étapes de constitution

Etape 0 : Collection HAL du DIM MAP

- Il existait une collection HAL du DIM MAP, à curation manuelle
- Elle est incomplète mais sert d'exemple pour une première exploration

Etape 1 : Liste d'auteurs du DIM PAMIR

- Nous utilisons la liste des 955 scientifiques faisant partie d'une unité de recherche participant au DIM
- Celle-ci est réduite à 110 auteurs dont les activités concernent exclusivement les sujets du DIM (critères : unités ou personnes repérées manuellement par des experts dans la liste)
- Extraction des publications depuis Google Scholar, Web of Science, Scopus, HAL (méthodologies diverses)

Etape 2 : mots-clés et institutions

- Nous récupérons les mots-clés associés au corpus issu de la première étape (critères : fréquence et pertinence)
- Nous dressons la liste des institutions concernées par le DIM (les tutelles franciliennes des unités utilisées en étape 1)
- Nous réitérons nos requêtes sur les bases de données en les élargissant grâce à ces deux critères

Première approche : publications de la collection DIM MAP

Cette étape est séparée du reste du processus. Elle a permis de commencer le travail de définition d'un corpus, mais n'intervient plus ensuite.

```
url='http://api.archives-ouvertes.fr/search/DIM-MAP/?q=*:*&indent=true&rows=100&fl=docId,doiId_s,keyword_s,domain_s'
resp=requests.get(url).json()
```

```
docs=resp['response']['docs']
df=pd.DataFrame(docs)
df.to_csv('DIM-MAP_resp.csv', index=False)
```

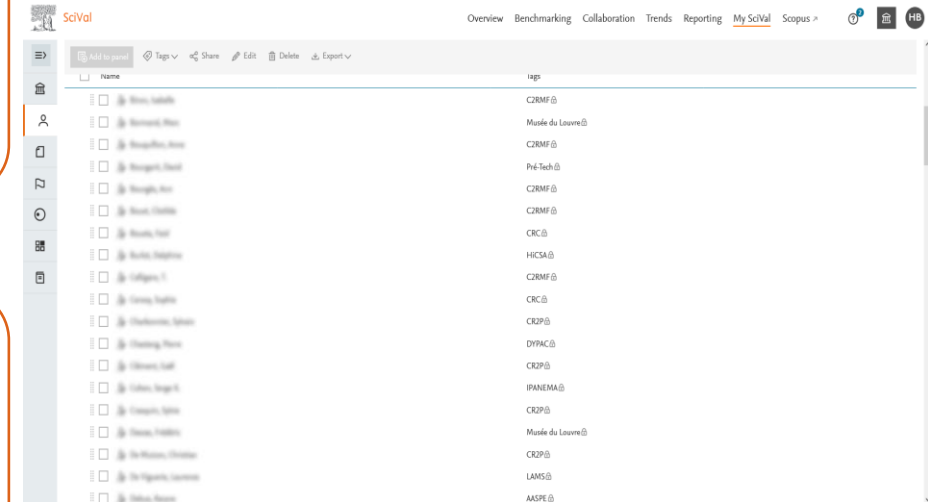
Facettes sur les mots-clés

```
kuds=df['keyword_s'].dropna()
```

```
kuds_dict={}
for kwlist in kuds:
    for kw in kwlist:
        try:
            kuds_dict[kw]+=1
        except KeyError:
            kuds_dict[kw]=1
```

```
kuds_df=pd.DataFrame(kuds_dict.values(), index=kuds_dict.keys(), columns=['nbdocs']).sort_values('nbdocs', ascending=False)
```

Code permettant de récupérer les mots-clés associés aux publications de la collection HAL liée à l'ancien DIM (MAP)



name	tags
Marie Perle	C2RMF @
Musée du Louvre	Musée du Louvre @
C2RMF	C2RMF @
Pré-Tech	Pré-Tech @
C2RMF	C2RMF @
C2RMF	C2RMF @
CRC	CRC @
HICSA	HICSA @
C2RMF	C2RMF @
CRC	CRC @
CR2P	CR2P @
DIPAC	DIPAC @
CR2P	CR2P @
IPANEMA	IPANEMA @
CR2P	CR2P @
Musée du Louvre	Musée du Louvre @
CR2P	CR2P @
LAMS	LAMS @
AASPE	AASPE @

Exemple de la représentation dans SciVal d'auteurs du DIM validés pour le corpus de l'étape 1

1. Corpus : validation et finalisation

- A partir des mots-clés récupérés depuis le premier corpus et de l'ensemble des institutions participant au DIM, nous rassemblons sur Zotero 5188 publications issues de diverses sources.

Titre	Créateur
Narrow Aperture Detection Electrodes ICR Cell with Quadrupolar Ion Detection for FT-ICR MS at the Cyclotron Frequency	Nagornov et al.
Natural risks and water management in Delphi	Perrier et al.
Nature concepts et enjeux des référentiels isotopiques pour décrire les systèmes d'élevage du passé	Balasse et al.
Naviguer dans la toile : imagerie 3D de textiles minéralisés du site de Nausharo (première moitié du IIIe millénaire av. J.-C.)	Li et al.
ND:YAG vs ER:YAG: A comparative study of laser varnish removal on easel paintings	Lopez et al.
Neanderthal-like traits visible in the internal structure of non-supranuchal fossae of some recent Homo sapiens: The problem of their identification in hominins and phylogenetic implications	Nowaczewska et al.
Neanderthals paleoenvironment in Western Provence: The contribution of Les Auzières 2 (Méthamis, Vaucluse, France)	Marchal et al.
Neanderthal environment in Liguria during the upper pleistocene. Pollen analysis from caverna delle fate (Finale Liguria, Italy)	Karatsori et al.
Neanderthal selective hunting of reindeer? The case study of Abri du Maras (south-eastern France)	Daujeard et al.
Negahi be Eghtesad-e zisti Tappeh Hessar dar dore ye III B ; Natayej motale'at Bastanshenasi hafari ye 1995 (Aperçu sur l'économie de subsistance de Tappeh Hessar III B ; Résultats des études archéozoologiques d...)	Mashkour
Neogene and Quaternary fossil remains of beaked whales (Cetacea, Odontoceti, Ziphiidae) from deep-sea deposits off Crozet and Kerguelen islands, Southern Ocean	Lambert et al.
Neogene palaeoenvironments of the Namib Desert: a brief synthesis	Senut et Ségalen
Neolithic sheep birth distribution: Results from Nova Nadezhda (sixth millennium BC, Bulgaria) and a reassessment of European data with a new modern reference set including upper and lower molars	Balasse et al.
Ness of Brodgar's seaweed-eating sheep: adaptations to an island archipelago environment on Neolithic Orkney, Scotland	Blanz et al.
Networks and Monumentality in the Pacific: Proceedings of the XVIII UISPP World Congress (4-9 June 2018, Paris, France) Volume 7 Session XXXVIII	Hermann et al.
Neue Erkenntnisse zum Arnegundegrab. Ergebnisse der Metallanalysen und der Untersuchungen organischer Überreste aus Sarkophag 49 der Basilika von Saint-Denis	Buchet et al.
Neue Identifizierungsstrategie eisenzeitlicher Korallen anhand optischer und Raman-spektroskopischer Charakteristiken	Fürst et al.
Neurocranial development of the coelacanth and the evolution of the sarcopterygian head	Dutel et al.
Neutral theory: From complex population history to natural selection and sociocultural phenomena in human populations	Austerlitz et Heyer
Neutron imaging investigation of fossil woods: non-destructive characterization of microstructure and detection of in situ changes as occurring in museum cabinets	Odin et al.
New AGLAE - Un outil d'exception au service de la connaissance des biens culturels	Pacheco
New approaches for investigating paintings by ion beam techniques	Beck et al.
New archeointensity data from French Early Medieval pottery production (6th-10th century AD). Tracing 1500 years of geomagnetic field intensity variations in Western Europe	Genevey et al.
New archeointensity data from Novgorod (North-Western Russia) between c. 1100 and 1700 AD. Implications for the European intensity secular variation	Salnaia et al.
New Bioarchaeological Evidence and Radiocarbon Dates from the Lambayeque/Sicán Culture Camelids from the El Brujo Complex (Northern Coast of Peru): Implications for Funerary and Herd Management Prac...	Santana-Sagredo et al.
New chronological data (ESR and ESR/U-series) for the earliest Acheulian sites of north-western Europe	Voinchet et al.
New Clues on the Origin of the Inca Mirror at the Museum National d'Histoire Naturelle in Paris	Calligaro et al.
New constraints on geomagnetic field intensity variations in the Balkans during the Early Byzantine period from ceramics unearthed at Thasos and Delphi, Greece	Genevey et al.
New Constraints on the Metallurgical Activity at the Morimond Abbey (Haute-Marne) from the Archeomagnetic Analysis of two Excavated Brick Fireplaces	Rouzeau et al.

Exemple de publications triées au sein de la bibliothèque Zotero ayant servi à la finalisation du corpus

- Des experts du domaine valident ensuite manuellement l'appartenance au champ du DIM de ces publications.
- Le corpus conservé au terme de ce processus (4113 publications) n'est pas exhaustif mais il est considéré comme pertinent et suffisamment représentatif pour lancer l'étude.

2. Méthodologie de traitement

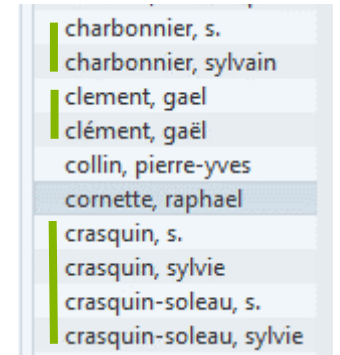
1. Nous utilisons les collaborations entre auteurs comme matériau de base du réseau
2. Nous devons identifier des affiliations des auteurs
 1. Source du lien auteur-affiliation : HAL (permet de traiter plus précisément les affiliations anciennes ou à fusionner)
 2. Méthode : d'abord par l'affiliation la plus courante de chaque auteur, puis par l'affiliation la plus récente.

hal_ids	last_affil
[385997, 12510364]	Histoire naturelle de l'Homme préhistorique
[11275339, 444479]	IRCER - Axe 2 : procédés plasmas et lasers
[1132033, 11213404, 11464170]	Culture et Environnements, Préhistoire, Antiquité, Moyen-Age
[765269, 495488, 1135501]	Laboratoire de Biologie et Ecophysiologie
[12025183]	Delft University of Technology
[234815]	Biogéosciences [UMR 6282] [Dijon]
[12544617, 12318179, 12385611, 12395839, 688183, 1487170, 559915, 310894]	Préhistoire et Technologie
[12300050, 12142821, 11876586, 241661, 12616621]	Institut de minéralogie, de physique des matériaux et de cosmochimie

Extrait de tableau de données comparant les identifiants de structures HAL possibles trouvé pour des auteurs et les noms des structures retenues comme les dernières affiliations utilisées

2. Méthodologie : collaborations entre auteurs

- Chaque auteur doit être repéré de manière unique : certains profils sont en doublons et supprimés manuellement, mais nous devons réattribuer les publications pour obtenir un résultat correct.
- Chaque auteur ne peut être affilié qu'à une seule structure : l'affiliation multiple pourrait causer des « collaborations » biaisées entre les structures successives d'affiliation des auteurs.
- Aucune limite n'est fixée au nombre d'auteurs par publication : le domaine n'est pas propice aux publications les plus démesurées (plus de 1000 auteurs) mais le corpus comporte une petite proportion de collaborations de grande taille (plus de 100 auteurs).



A screenshot of a software interface for deduplicating author names. It shows a list of names in a light blue box with a vertical scrollbar on the left. The names are: charbonnier, s.; charbonnier, sylvain; clement, gael; clément, gaël; collin, pierre-yves; cornette, raphael; crasquin, s.; crasquin, sylvie; crasquin-soleau, s.; crasquin-soleau, sylvie. The entry 'cornette, raphael' is highlighted in a darker blue, indicating it is selected for deduplication.

charbonnier, s.
charbonnier, sylvain
clement, gael
clément, gaël
collin, pierre-yves
cornette, raphael
crasquin, s.
crasquin, sylvie
crasquin-soleau, s.
crasquin-soleau, sylvie

Exemple de liste d'auteurs à dédoublonner (liste sur VosViewer)

3. Visualisation des réseaux

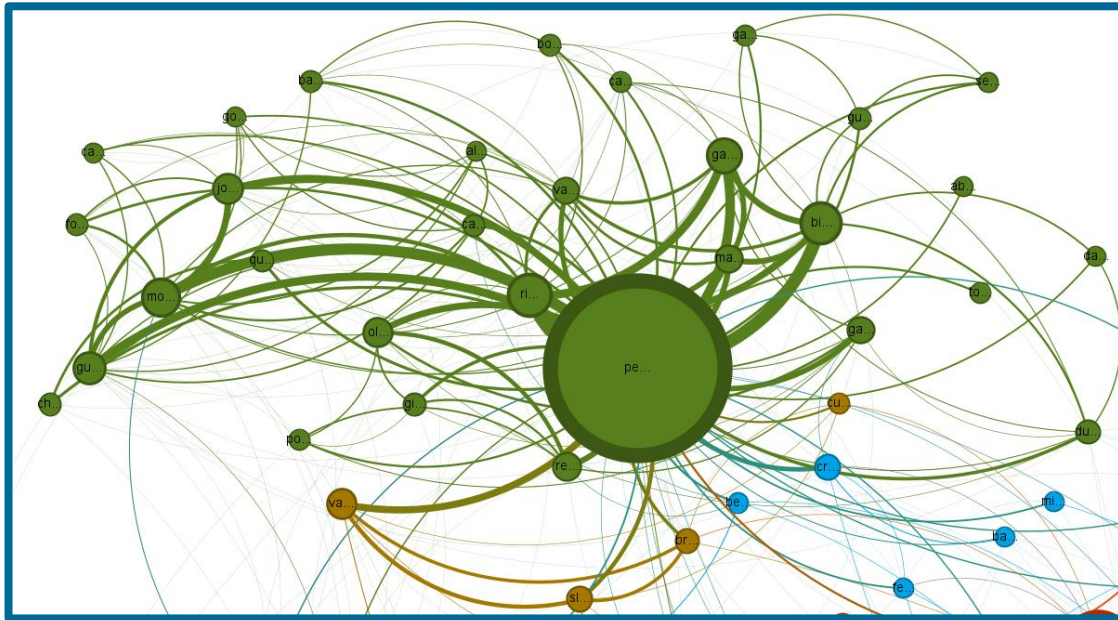


Réseau des collaborations entre auteurs dans le corpus du DIM PAMIR
Taille des nœuds : nombre de publications dans le corpus. Taille des liens : nombre de publications communes
Couleurs : clusters obtenus par application de l'algorithme de modularité de Gephi, résolution 1
Spatialisation : Force Atlas 2

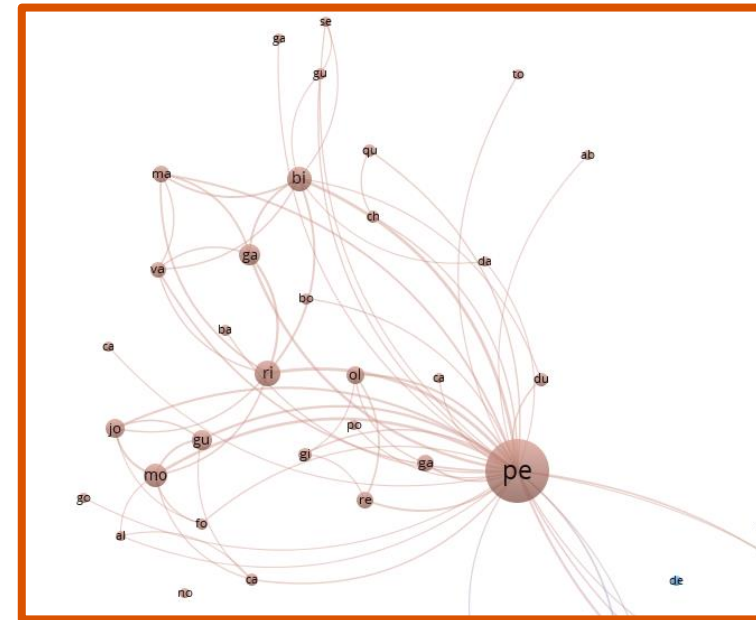
3. Visualisation des réseaux : choix du logiciel

- Nous avons le choix entre Gephi et VosViewer, Les différences entre les deux sont les suivantes :
 - VosViewer a une prise en main plus simple,
 - Gephi a l'avantage d'une plus grande flexibilité grâce aux extensions dont il dispose
 - Les modèles de spatialisation proposés sont plus variés dans Gephi
- Nous décidons d'utiliser Gephi pour les comparaisons entre clusters de collaboration et clusters « théoriques » de laboratoires

Gephi



VosViewer

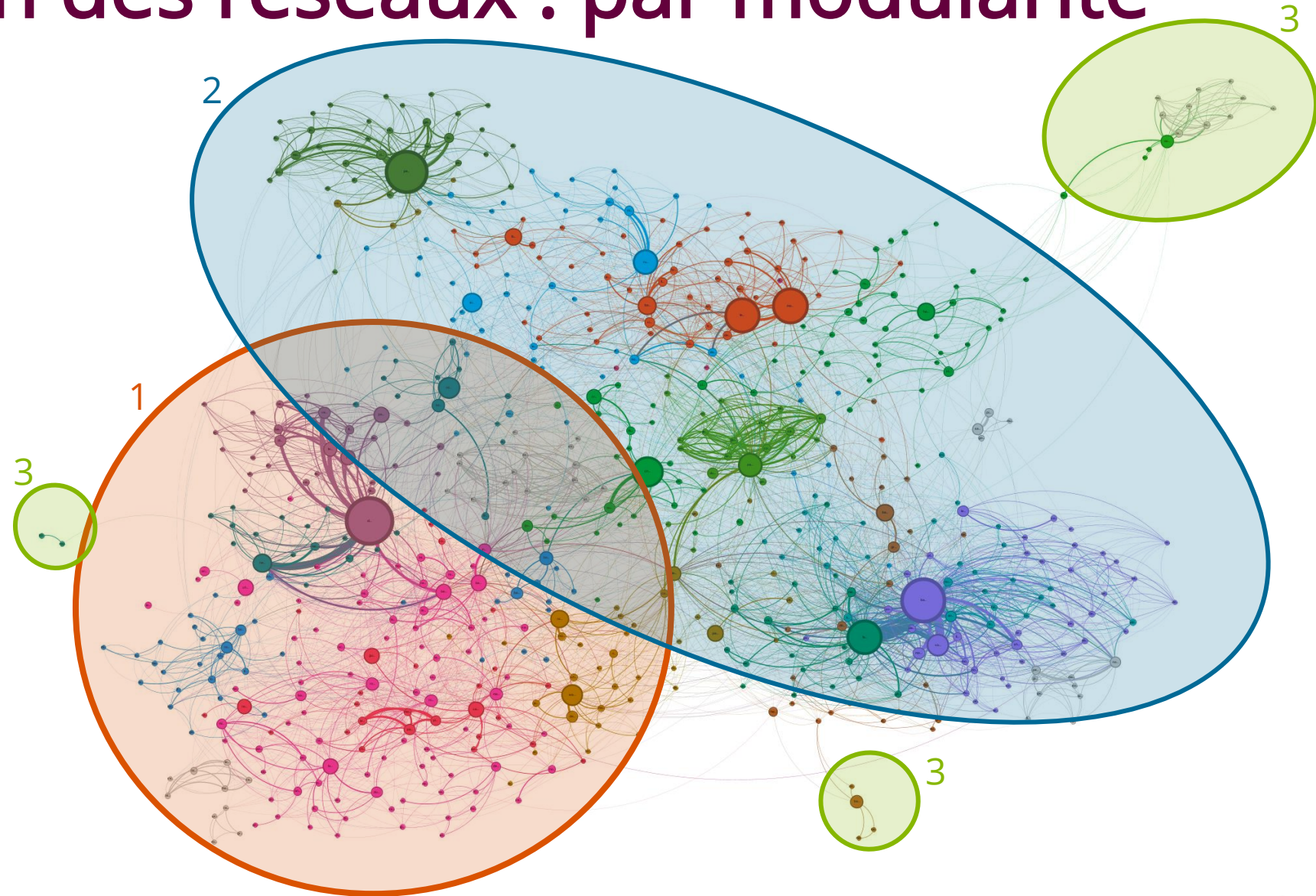


Comparaison entre les visualisations par défaut de Gephi (cadre bleu) et VosViewer (cadre orange) pour le même groupe de chercheur analysé sur les mêmes données

3. Visualisation des réseaux : par modularité

On repère dans le corpus deux clusters fortement interconnectés :

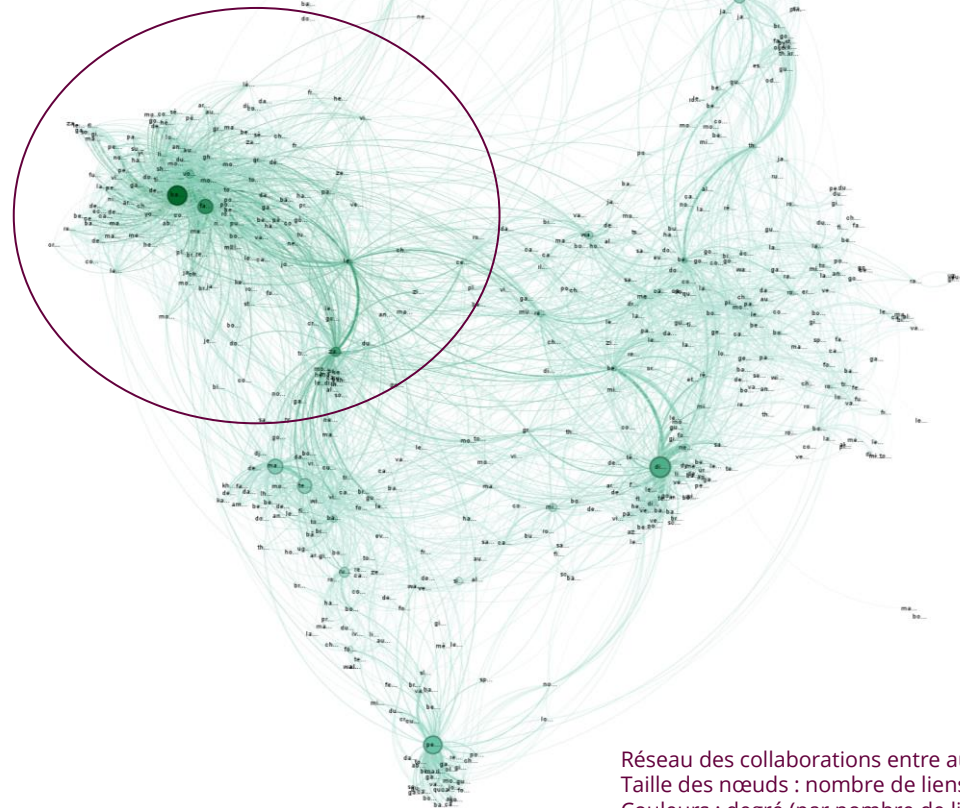
1. le **premier** cluster est large et constitué de nombreux chercheurs ayant produit des volumes similaires de publications, à l'exception d'un auteur plus prolifique, bien que non central au sein du cluster (Di.)
2. le **second** cluster est doté de plusieurs sous-clusters (à ce stade, nous ne cherchons pas à savoir s'ils sont thématiques ou institutionnels), chacun centrés sur une ou deux figures d'auteurs prolifiques (Pe., Ru., Ma., Te., Ch., Ba1.)
3. Chaque cluster comporte également une ou plusieurs **équipes** moins interconnectées au reste, en particulier pour le premier l'équipe autour de Ba2. et pour le second celle de Mu. A ce stade, nous ne cherchons pas à expliquer ces liens.



3. Visualisation des réseaux : par centralité

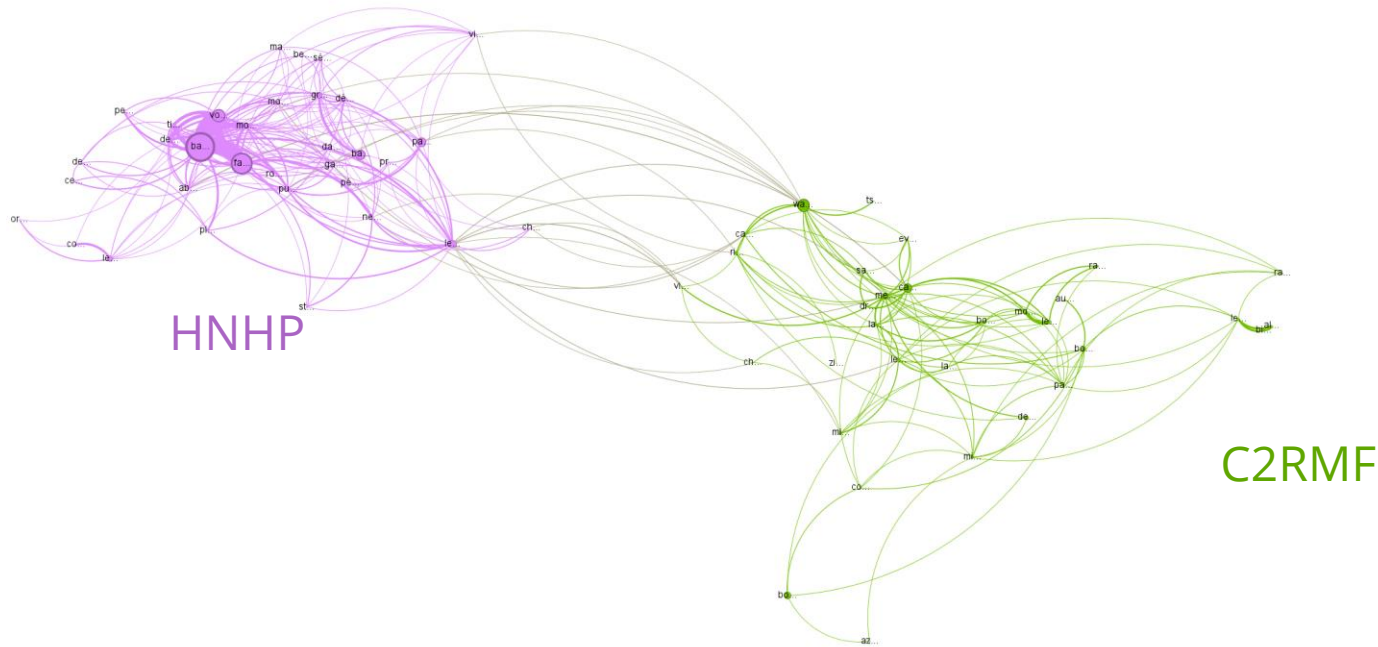
La comparaison avec le graphe obtenu en tenant compte du poids de chaque auteur en nombre de publications dans le corpus permet de mieux faire ressortir les clusters les plus interconnectés : on repère ici davantage les équipes collaborant autour de Ba1, tandis que les autres sont moins fortement interconnectées entre elles.

Cluster « central »



Réseau des collaborations entre auteurs dans le corpus du DIM PAMIR
Taille des nœuds : nombre de liens de collaboration. Taille des liens : nombre de publications communes
Couleurs : degré (par nombre de lien de collaboration)
Spatialisation : Force Atlas 2

3. Visualisation des réseaux : par affiliation



En filtrant les auteurs du graphique en fonction de leur dernière affiliation dans HAL, on obtient une visualisation de la collaboration entre deux structures, dans le corpus du DIM. Ainsi les personnes collaborant entre laboratoires sont plus faciles à repérer, ici Le. ou Wa.

Réseau de collaboration entre auteurs de HNHP et C2RMF dans le corpus du DIM PAMIR
Taille des nœuds : nombre de publications dans le corpus. Taille des liens : nombre de publications communes
Couleurs : Affiliations utilisées en dernier dans HAL par les auteurs
Spatialisation : Force Atlas 2

4. Prochaines étapes

Mise à jour du corpus

- Analyse des biais de conception du premier corpus
- Récupération automatisée de publications

Affinage des analyses

- Correction des biais d'affiliation
- Utilisation d'indicateurs plus performants de l'interdisciplinarité et de collaboration

Automatisation du projet

- Traitements scriptés de la donnée
- Réitération périodique des analyses

université
PARIS-SACLAY

UVSQ 

université PARIS-SACLAY

Questions et échanges

henri.bretel@universite-paris-saclay.fr

michel.de-moura@uvsq.fr