

Analyse graphique des travaux de recherche autour des sorcières et de la sorcellerie



Table des matières

1	Introduction, objectif et contexte	2
1.1	Origine des données	2
1.2	Sujet et démarche pour l'obtention des données	2
1.3	Plan du présent rapport	3
2	Vue générale sur le corpus	4
2.1	Biais et limites du corpus obtenu	4
2.2	Quelques graphiques	4
3	Réseaux de mot-clefs	8
3.1	Mesures de centralité	8
3.2	Communautés de mots-clefs	9
3.3	Liens avec les disciplines	9
3.4	Conclusion	13
4	Réseaux d'auteur-ices	16
4.1	Centralité	16
4.2	Degré des nœuds	16
4.3	Exploration du lien entre l'auteur-ice et le domaine disciplinaire	19
5	Conclusion générale	23
6	Annexes	24

Table des figures

1	Représentation graphique de la répartition des types de documents	5
2	Évolution du nombre de publications présentes sur HAL pour notre requête par année de publication	6
3	Répartition des disciplines pour les travaux de notre requête	7
4	<i>Closeness</i> et <i>betweenness</i> des mots-clefs	10
5	Occurrence des mots-clefs	11
6	Communautés de mots-clefs	12
7	Centralité des nœuds sur un réseau bipartite mots-clefs/domaines	14
8	Communautés mots-clefs/disciplines	15
9	Réseau d'auteur-ices : <i>Closeness</i> (couleur) et <i>Occurrence</i> (taille)	17
10	Réseau d'auteurs : <i>Degré</i> (couleur) et <i>Occurrence</i> (taille)	18
11	Réseau bipartite Auteurs-Disciplines (couleurs pour le type de noeuds et <i>Centralité Eigenvector</i> pour la taille)	20
12	Communautés disciplinaires des auteurs (<i>Modularity Class</i>) avec le <i>degré</i> pour la taille des nœuds	22
13	Réseau bipartite mots-clefs / disciplines	25

1 Introduction, objectif et contexte

Ce document présente l'étude d'un corpus de textes de recherche autour du domaine des sorcières et de la sorcellerie, grâce à l'analyse de graphes. Le corpus est extrait de la base de données [HAL](#). L'export et la production de graphes est faite en python, et le code est disponible [sur cette page](#). Enfin, nous manipulons les graphes grâce au logiciel Gephi.

1.1 Origine des données

HAL est une archive ouverte, en ligne, « destinée au dépôt et à la diffusion d'articles scientifiques ».

HAL, lancée il y a 20 ans, est aujourd'hui reconnue comme l'archive ouverte nationale en France. Cependant, recherche étant majoritairement internationale, de très nombreuses publications qui y sont archivées ne sont pas rédigées en français (qui ne représentent que 39 % des dépôts).

1.2 Sujet et démarche pour l'obtention des données

Nous avons choisi d'analyser les publications sur HAL où il est question de sorcières, de sorciers, de sorcellerie, *etc.*

Pourquoi ce sujet ?

En apparence décalé et fantaisiste, ce sujet nous a interrogé·e·s pour ces possibles liens avec des sujets contemporains (féminisme, écoféminisme, chamanisme, néo-paganisme, ...) en plus d'être historique. Il nous a donc paru intéressant d'analyser la littérature scientifique à ce sujet en partant des données présentes sur HAL, avec tous les biais que cette méthodologie implique.

Pour travailler sur notre hypothèse de départ, nous nous sommes d'abord concentré·e·s sur l'analyse des mots-clés et des domaines disciplinaires engagés dans ces objets d'études. Nous souhaitons également voir si des auteur·ices se dégageaient de ces résultats, et identifier d'éventuel·les expert·es francophones sur ce sujet.

Périmètre de la requête

Pour répondre à notre sujet, nous avons construit notre requête de départ à partir de la forme tronquée [sorc*](#). Pour une question pratique dans le cadre de cet exercice, nous avons choisi de nous limiter aux résultats en langue française et c'est pourquoi nous avons choisi la racine "sorc", qui nous paraissait correspondre à la grande majorité du champ lexical ciblé. Nous avons en revanche fait le choix de ne mettre aucun périmètre temporel particulier afin d'éventuellement observer une évolution de l'intérêt sur ces questions au cours du temps.

Nous nous sommes immédiatement rendu·e·s compte lors des premiers résultats d'un bruit important dû aux noms des auteur·rices. Pour remédier à cela, nous avons donc restreint la recherche du terme [sorc*](#) aux seuls champs titre, mots-clés et résumé. Cela permettait d'ailleurs d'éviter tout autre bruit venant par exemple des noms de structures.

Information exportées pour l'analyse

Pour tenter de mener à bien ces objectifs, nous avons donc défini les informations dont nous aurions besoin à l'export :

- Titre
- Auteur·rice·s
- Structures
- Date de publication
- Domaines
- Mots-clés
- Type de documents

Requête finale

La requête finale utilise l'API de HAL et se présente donc sous la forme suivante :

```
response = requests.get("http://api.archives-ouvertes.fr/search/?q=title_t:sorc* OR keyword_t:sorc* OR abstract_t:  
→ sorc*&wt=csv&rows=900&fl=title_s,authFullName_s, keyword_s,publicationDateY_i,structName_s,  
→ level0_domain_s, domainAllCode_s, docType_s")}
```

1.3 Plan du présent rapport

Afin de donner une vue globale du corpus de textes, nous présentons d'abord quelques données générales à son sujet, ainsi que les biais que présentent notre étude. En accord avec les questionnements évoqués précédemment, nous avons divisé notre étude en deux grandes parties : la première analyse des réseaux de mots-clés, la seconde des réseaux d'auteur·rices. Dans les deux cas, nous utilisons plusieurs mesures, et deux types de graphes : simples et bipartites.

2 Vue générale sur le corpus

2.1 Biais et limites du corpus obtenu

Avant toute analyse du corpus, il nous semble important de rappeler que certain-e-s critères ou données de départ biaisent les résultats et/ou nos analyses.

Tout d’abord, si HAL est une plateforme adéquate pour ce type d’exercice du fait de ses données ouvertes et manipulables librement, toutes les publications scientifiques n’y sont pas présentes (loin de là). Bien que le dépôt soit de plus en plus encouragé, voire obligatoire depuis quelques années, certain-e-s chercheur·euse-s plus impliqué·e-s ou sensibilisé·e-s à la science ouverte risquent d’être sur-représenté·e-s (puisque leurs collègues pourraient ne jamais avoir déposé sur HAL mais être par ailleurs spécialistes). Les pratiques de chaque chercheur·euse, laboratoire ou communauté disciplinaire peuvent donc avoir un impact sur les données présentes sur HAL, et par conséquent sur notre analyse. C’est également valable sur la temporalité des publications : des publications plus anciennes pourraient ne jamais avoir été déposées sur HAL, qui n’existe que depuis 20 ans.

De la même manière, si nous avons choisi de restreindre notre corpus aux publications francophones, nous sommes bien conscient·e-s que nous passons sous silence une partie, peut-être importante, des publications. Cela représente également un biais non négligeable quant aux données récupérées.

Nous considérons donc à partir de maintenant que toutes les analyses que nous pouvons tirer des graphiques ou cartographies ci-après ont comme postulat de départ les biais ci-dessus. Nous les mettons volontairement de côté pour éviter les répétitions.

2.2 Quelques graphiques

La requête réalisée grâce à l’API fournie par HAL nous a permis de trouver 748 documents ayant rapport avec notre sujet d’étude.

Types de travaux

La FIGURE 1 présente la répartition des types de documents dans le résultat de notre recherche. Les articles de revues sont majoritaires, puisqu’ils composent presque le tiers de notre corpus (32 %). En deuxième position par ordre d’importance, les chapitres d’ouvrages composent 22 % de notre corpus, puis viennent les communications au sein de conférences qui constituent 17 % du corpus. Les 29 % restants correspondent à différents autres types de documents (mémoire, thèses, posters, direction d’ouvrage...) qui sont présents dans des proportions plus minimes.

La composition de notre corpus est donc relativement hétérogène.

Notons cependant que cela correspond plus ou moins à la répartition générale des dépôts dans HAL puisque 54 % du total des dépôts correspondent à des articles dans une revue ; 28 % à des communications dans un congrès et 9 % à des chapitres d’ouvrages. Nous constatons donc tout de même que le pourcentage de chapitres est bien plus important dans notre corpus par rapport au taux total dans HAL. Cela est sans doute dû au fait que ce format de publication est très lié aux disciplines (SHS) rattachées à notre étude.

Évolution de la production dans le temps

La FIGURE 2 présente la quantité de documents correspondant à notre requête en fonction des années de publication. On observe une croissance exponentielle entre 1980 et 2015. La quantité d’articles baisse légèrement entre 2015 et 2020. La dernière barre du graphique n’apporte que peu d’informations, puisqu’elle correspond à la période allant de 2020 à 2025. Ce n’est donc pas surprenant qu’elle soit bien inférieure aux précédentes à l’année 2021.

Une telle évolution peut s’expliquer par deux facteurs, dont il est difficile d’évaluer l’importance respective avec ces seules informations :

- HAL n’existait pas avant 2001 (on peut donc penser que la plupart des travaux d’avant cette date n’y ont pas été déposés), et on peut supposer que beaucoup de travaux datant d’avant les années 1990 n’ont pas été numérisés, donc encore moins archivés sur HAL.

Rappelons par ailleurs la mise en application de la loi pour une république numérique. Votée en 2016, cette

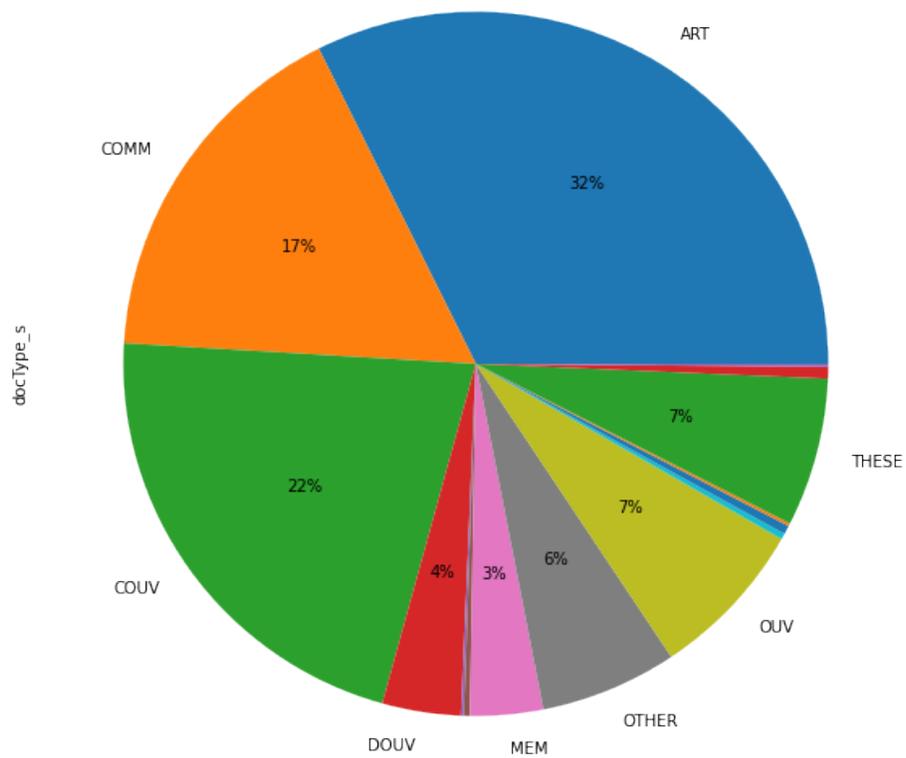


FIGURE 1 – Représentation graphique de la répartition des types de documents. COMM correspond à des communication dans des conférences, ART à des articles de revues, OUV à des chapitres d’ouvrages, DOUV à des direction d’ouvrages et MEM à des mémoires.

loi oblige la mise à disposition gratuite sur un e archive ouverte des résultats issus de recherches financées par des fonds publics. Cette loi a donc induit une augmentation des dépôts des publications depuis 5 ans maintenant et peut ainsi expliquer en partie le nombre de publications important de ces dernières années.

- Un vrai développement de l'intérêt et des travaux de recherche autour des sorcières et de la sorcellerie depuis les années 2000, comme on le voit dans la culture plus populaire (littérature jeunesse avec des phénomènes mondiaux comme *Harry Potter* à la fin des années 1990, ou le développement de ces thématiques dans les réseaux féministes et sur les réseaux sociaux depuis quelques années), avec des intérêts associés à la sorcellerie, en passant par le renouveau des mouvements néo-païens²

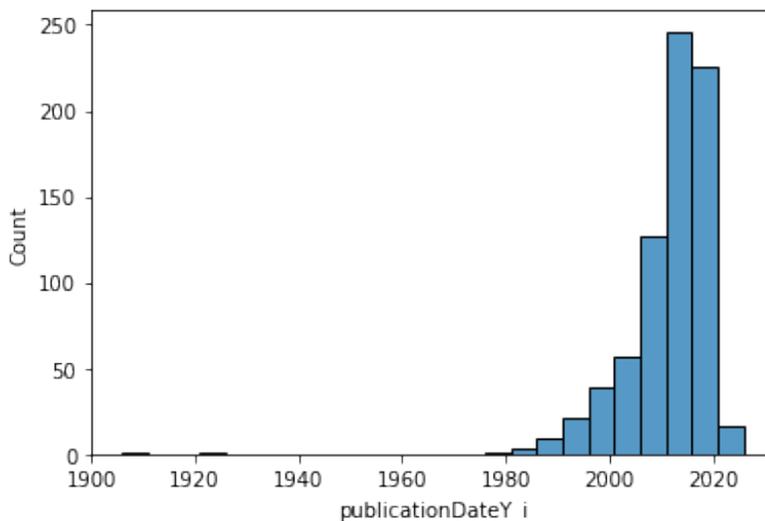


FIGURE 2 – Évolution du nombre de publications présentes sur HAL pour notre requête par année de publication. Une barre du graphique représente 5 années.

Domaines de recherche concernés

Enfin, la FIGURE 3 nous permet d'avoir un aperçu de la répartition des travaux par discipline. Sur HAL, à chaque dépôt peut être associé un ou des domaines et un ou plusieurs sous domaines disciplinaires. Ce graphique nous montre qu'un total de 95 disciplines sont représentées, le plus souvent en proportion assez minime : 26 disciplines seulement ont une fréquence ≥ 3 .

Les disciplines les plus représentées sont, sans surprise, liées aux sciences humaines : histoire (shs.hist), littérature (shs.litt), anthropologie sociale et ethnologie (shs.anthro-se) et sociologie (shs.socio). En lien avec notre hypothèse de départ, notons que la discipline des études de genre (shs.genre) représente 10 travaux, c'est-à-dire un peu plus de 1,3 % du corpus.

De manière générale, ce graphique confirme l'intuition que notre requête était très vaste.

2. Voir [Néopaganisme](#) sur Wikipédia. Consultation le 21/11/2021.

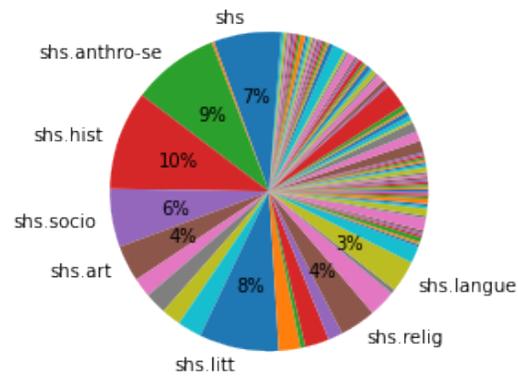


FIGURE 3 – Répartition des disciplines pour les travaux de notre requête. Plusieurs disciplines peuvent être indiquées par document. Par souci de lisibilité, nous n'affichons pas les proportions ni les labels pour les disciplines représentées à moins de 3 %. SHS signifie « Science de l'homme et société ».

3 Réseaux de mot-clefs

Nous avons commencé par étudier les liens entre les différents mots-clefs liés aux documents renvoyé par notre requête. L'idée générale était de dégager des grandes thématiques, possiblement interdisciplinaires, sous-tendant les travaux sur la sorcellerie et les sorcières. Nous partions notamment avec l'idée que le thème de la sorcellerie est de plus en plus fréquent dans des thématiques féministes, et souhaitions voir si cette hypothèse pouvait être confirmée ou infirmée.

Dans les réseaux présentés dans cette section, nous avons éliminé tous les mots-clefs n'apparaissant qu'une seule fois, pour plus de lisibilité. Par ailleurs, il nous semble important de signaler que seuls quatorze mots-clefs dépassent les 3 occurrences, et cinq les 5 occurrences³, ce qui rend considérablement plus compliquée (et, de fait, moins intéressante...) l'analyse des graphes. Par ailleurs, 209 travaux, soit presque 28 % du corpus étudié, n'ont aucun mot-clef.

Enfin, nous attirons l'attention sur le fait que certains mots-clefs auraient pu être regroupés : par exemple « sorcière », « sorcières » et « Sorcières – Dans la littérature », « Moyen Age » et « Moyen Âge », « Littérature jeunesse » et « Littérature de jeunesse », *etc.* Cela aurait nécessité un traitement du texte supplémentaire, que nous avons choisi de ne pas effectuer par manque de temps et de compétences, et parce que cela ne nous semblait pas être l'objectif de ce travail.

Quelques chiffres généraux sur le graphe

Avant de plonger dans les représentations, voici quelques chiffres sur le graphe de mots-clefs que nous avons obtenu, qui permet de tirer quelques premières conclusions :

- **Densité** : 0,014 – Il y a peu de liens entre les mots-clefs. On a en effet beaucoup de mots-clefs isolés, et rares sont ceux qui ont un degré supérieur à 3 (*cf.* FIGURE 5).
- **Diamètre** : 4 – Notre graphe ne semble donc pas être un petit monde.

3.1 Mesures de centralité

Le réseau présenté en FIGURE 4 nous permet de souligner plusieurs choses :

- D'abord, la majorité des mots-clefs ont le même niveau de centralité, à peu de choses près. « Sorcellerie » est, sans surprise, extrêmement central puisqu'il présente la *betweeness* la plus élevée. Le mot-clef « Magie » est un peu moins central, mais reste important. Cela n'est pas surprenant non plus, puisqu'il semble compliqué d'évoquer la sorcellerie sans évoquer la magie, et ce quel que soit le champ disciplinaire. Ces deux mots-clefs sont d'ailleurs ceux qui ont le degré le plus élevé (*cf.* FIGURE 5), c'est-à-dire qui ont le plus de liens avec d'autres nœuds. Ces observations ne font finalement que confirmer que notre corpus traite bien de sorcellerie.

Le mot-clef « Femmes » est également relativement central, dans une moindre mesure. Là encore, c'est relativement facile à interpréter : on parle plus souvent de sorcières que de sorciers. Mais cela confirme tout de même notre intuition que le sujet est relié à des questions de genre. Cependant, cela ne donne pas d'indications quant aux liens avec des thématiques féministes. Une interprétation incluant les domaines de recherche, ou une analyse textuelle des résumés des travaux du corpus, pourraient peut-être compléter ce point.

Enfin, ce qui est plus étonnant est plutôt que tous les autres mots-clefs aient une *betweeness* égale ou quasi-égale : « Violences » par exemple n'est pas plus un « pont » entre les nœuds que quelque chose d'aussi spécifique que « PEUPLIER », au contraire de ce à quoi on aurait pu s'attendre. Ce résultat est probablement lié au fait que beaucoup de travaux du corpus n'aient pas de mots-clefs ; et que beaucoup de ces mots-clefs aient une occurrence très faible d'où ce résultat très diffus.

- Les résultats sont moins surprenants lorsqu'on se penche sur la *closeness* des nœuds, qui nous donne une indication sur la distance moyenne entre un nœud et les autres. Les nœuds les plus foncés sur le graphe (donc ceux qui sont le plus proche de tous les autres) sont aussi les plus généraux comme « Sorcellerie ». D'autres rejoignent directement les disciplines majoritaires mises en évidence FIGURE 3 : « Fiction » ou « Histoire des femmes », « ontologie sorcellaire » par exemple. On peut supposer que ces termes, s'ils sont relativement proches des autres mots-clefs du corpus, nous donnent des indications sur les grandes thématiques de ce

3. Sorcellerie, magie, démonologie, sorcières, et sorcières – dans la littérature

dernier : on y retrouve donc l’histoire et la littérature, mais aussi des questions liées au droit (« Procédure pénale », « Justice ») et des thématiques de genre (« Histoire des femmes », « Femmes », « Idéal féminin »). Certains nœuds centraux (au sens de la *closeness*) sont toutefois plus difficiles à interpréter, car ils semblent concerner des sujets très précis (« Sorcellerie bocaine » par exemple) – là encore, ce résultat est possiblement imputable à la proportion de travaux n’ayant pas ou peu de mots-clefs.

3.2 Communautés de mots-clefs

La FIGURE 6 nous permet de mettre en avant les communautés de mots-clefs. A nouveau, il est important de garder à l’esprit que la plupart des mots-clefs ont une très faible occurrence, ce qui est important dans notre interprétation : il est très probable qu’une communauté ne soit « constituée » en fait qu’un ou deux articles.

Le graphique nous permet de repérer huit communautés de mots-clefs. Certaines d’entre elles semblent référer à un sujet de recherche très précis, comme par exemple les nœuds de couleur vieux rose reliant « Roman arthurien tardif » et « Perceforest » (qui est « le nom d’une composition en prose anonyme écrite en français vers 1340⁴ » qui « relie la tradition des romans d’Alexandre à celle des grands romans arthuriens.⁵ »). D’autres communautés, comme celle en vert⁶, qui ne paraissent pas évidentes au premier abord, relèvent d’un document précis⁷. Ainsi, le faible nombre d’occurrences de la plupart des mots-clefs nous permet seulement de repérer des documents solitaires qui regroupent quelques mots-clefs spécifiques.

3.3 Liens avec les disciplines

Comme nous l’avons vu à la sous-section précédente, les communautés de mots-clefs sont assez divisées. Par ailleurs, l’étude de la centralité des nœuds semblait faire ressortir des disciplines comme le droit.

Afin d’approfondir cela, nous avons tenté de faire des réseaux bipartites entre les mots-clefs et les disciplines, en reprenant le constat que notre corpus était très vaste et comprenait de nombreux domaines de recherche (cf. FIGURE 3⁸). Pour plus de lisibilité sur ces réseaux, nous ne faisons pas apparaître les mots-clefs et disciplines ayant moins de 2 occurrences.

Premièrement, la FIGURE 7 nous permet d’affiner notre compréhension de la centralité du mot-clef « sorcellerie » : Gephi nous permet de voir qu’il est relié à toutes les disciplines majoritaires (sociologie, histoire, psychologie, littérature et anthropologie notamment). Il est aussi directement lié à des disciplines minoritaires. Cela n’est pas surprenant car il s’agit du mot clef le plus générique associé à notre recherche. Mais cela nous permet de constater que c’est bien une thématique qui fait le pont entre plusieurs disciplines des SHS.

La FIGURE 8 nous permet ensuite d’expliquer le peu de liens directs entre mots-clefs que nous avons observé à la section précédente. On voit en effet des communautés très fortes, réparties « autour » de chacune des disciplines. Par exemple, la communauté en vert que nous avons décrite sur la figure 6, reliant *idéal féminin*, *justice*, et *traitement médiatique*, se retrouve sur la FIGURE 8 dans la même communauté (en vert), qui s’organise autour des nœuds « shs.litt » et « shs ». On voit également des communautés rassembler des mots-clefs qui n’étaient intégrés dans aucun groupe dans la FIGURE 6, avec la création de liens avec des disciplines. Par exemple, « Hittite⁹ », « Procès » et « Religion » sont regroupés dans la communauté en rose sur la FIGURE 8, qui réunit les disciplines histoire et religion. En revanche, la communauté de mot-clefs que l’on pensait associée au droit précédemment ne se retrouve pas du tout dans le groupement qui s’organise autour de cette discipline, mais s’intègre plutôt à la communauté « histoire » – probablement des travaux autour des procès historiques, par exemple.

De manière générale, on constate que nos différentes communautés de mots-clefs repérées à la section précédente ont tendance à s’organiser autour de disciplines particulières. L’ajout des disciplines sur notre graphe permet par

4. D’après l’article [Perceforest](#) sur Wikipédia. Consultation le 14/11/2021.

5. *Ibid.*

6. Reliant *Idéal féminin*, *Justice*, *Fiction* et *Traitement médiatique*

7. En effet, [une simple recherche avec ces 4 mots-clefs sur HAL](#) ne renvoie qu’un seul document.

8. La FIGURE 13, en annexe, permet de voir sous une autre représentation la domination de certaines disciplines : l’histoire, l’anthropologie sociale et ethnologie et la littérature notamment.

9. Les Hittites sont un peuple ayant vécu en Anatolie au II^{ème} millénaire avant JC., qui semblaient pratiquer des rituels de type magique.

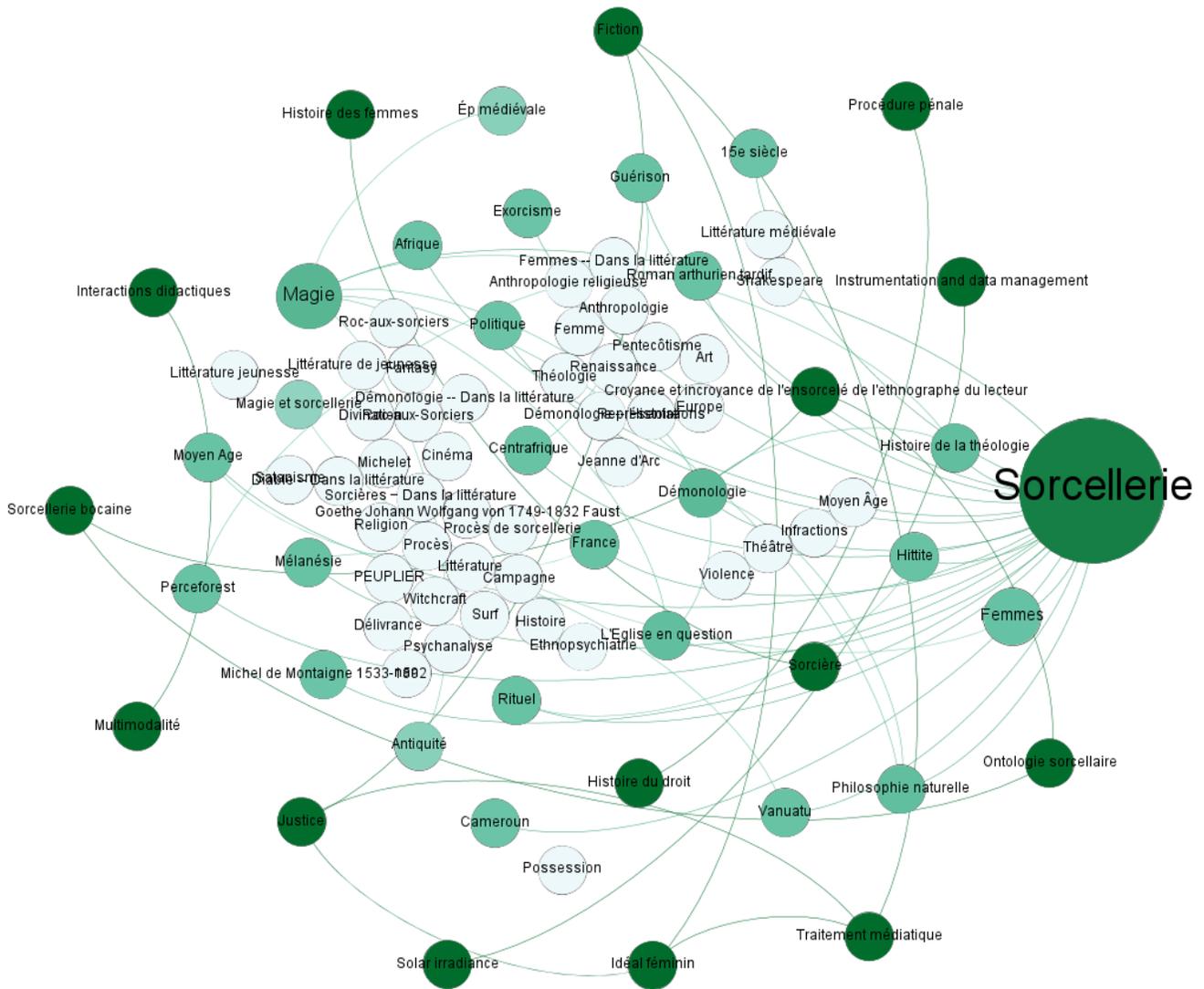


FIGURE 4 – Réseau des mots clefs représentant la *closeness* (intensité de couleur) et la *betweneess* (taille) des différents nœuds.

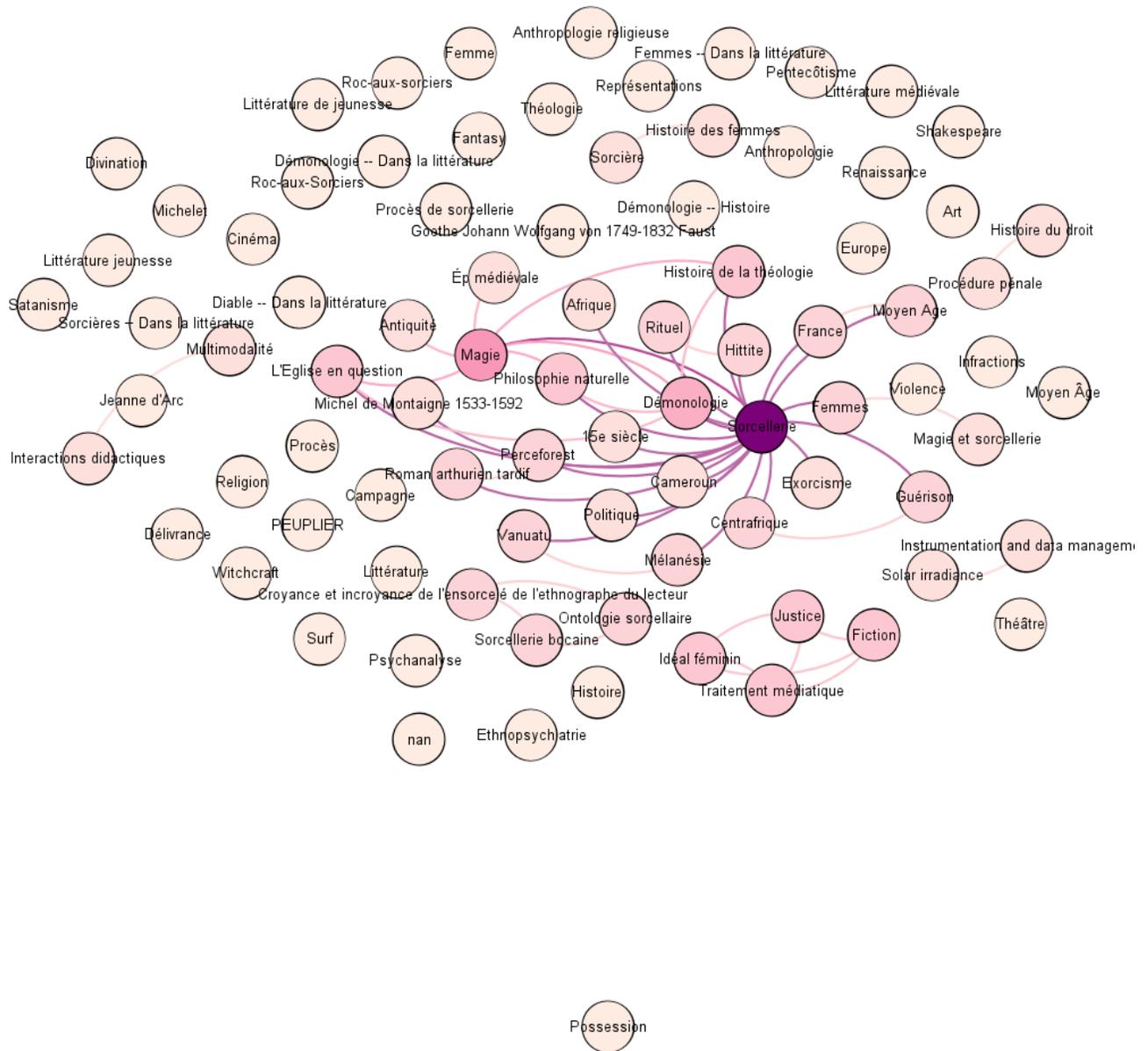


FIGURE 5 – Représentation de l’occurrence des mots-clés. Les mots-clés les plus sombres sont ceux qui ont le degré le plus élevé. « Sorcellerie » a un degré de 22, « Magie » de 7, « Démonologie » de 5. A noter que ce sont aussi les 3 termes avec le plus d’occurrences (respectivement 76, 36 et 29).

ailleurs de visualiser de nouvelles communautés, d'en préciser ou d'en expliquer certaines.

Enfin, il pourrait être intéressant de noter les mots-clefs qui font le pont entre deux disciplines, c'est-à-dire qui apparaissent dans la communauté d'une discipline mais ayant un lien avec une autre. Par exemple, les mots-clefs « JK Rowling », l'auteurice de la saga *Harry Potter*, et « Roald Dahl », auteur de nombreux romans pour enfants, sont tous les deux dans la communauté en orange (s'organisant autour des nœuds disciplinaires « shs.edu » (éducation) et « shs.class » (études classiques)), mais sont tout de même reliés aux nœuds « shs.hist » (histoire) et « shs.litt » (littérature), « shs.art » (art et histoire de l'art) et « shs.genre » (études de genre).

3.4 Conclusion

Les études de graphes sur les mots-clefs nous ont donc notamment permis de mettre en valeur la variété des travaux qui touchent au thème des sorcières et de la sorcellerie. Non-seulement les mots-clefs sont extrêmement divers, mais il y a peu de différence de centralité ou de degrés entre la majorité d'entre eux ; ce qui montre qu'une thématique en particulier ne se démarque pas d'une autre. Par ailleurs, la mise en relation des mots-clefs avec les disciplines permet de montrer comment chaque discipline s'approprie ce thème et le traite avec un point de vue différent. Certains mots-clefs font cependant le pont entre plusieurs domaines d'étude.

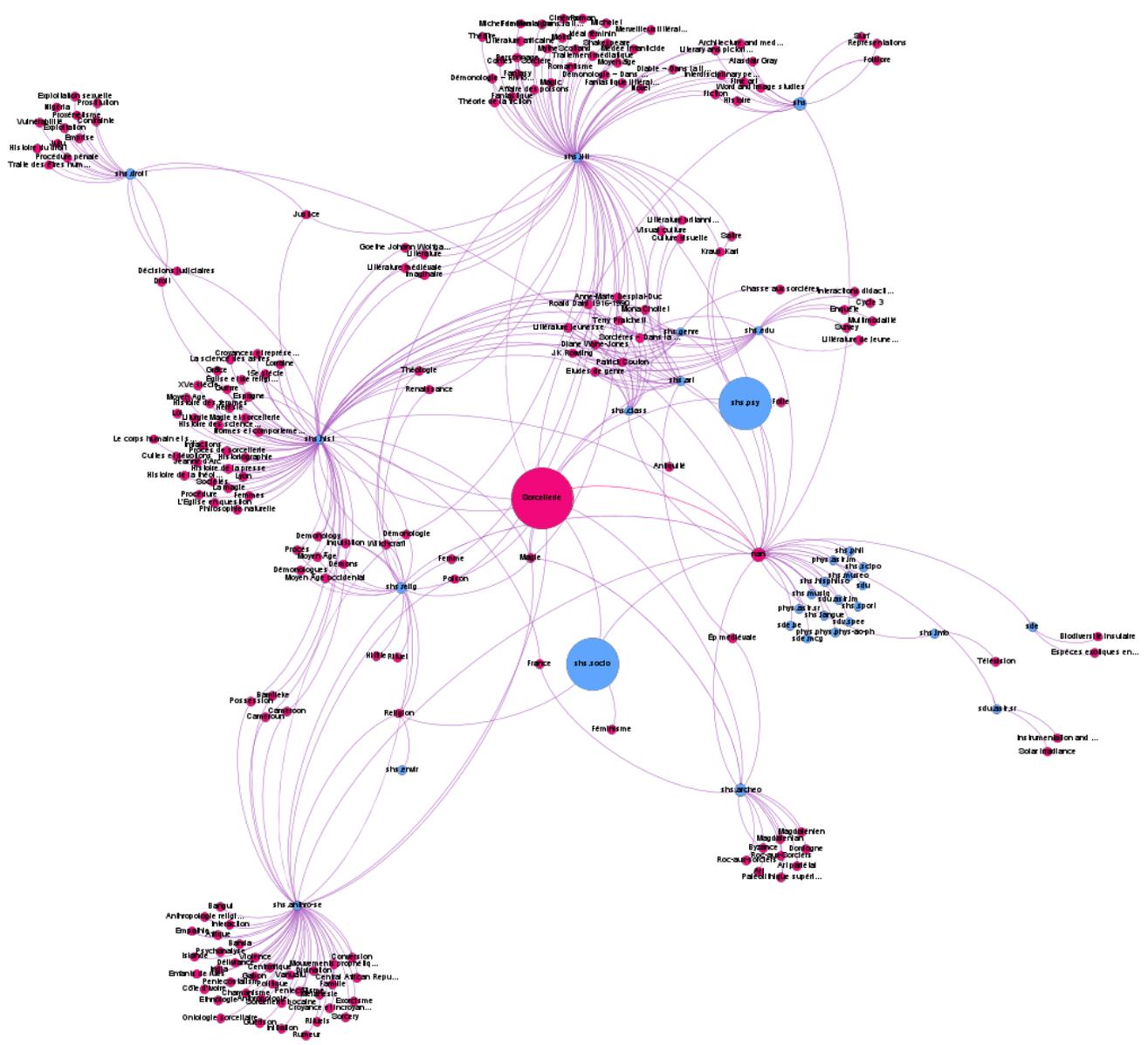


FIGURE 7 – Centralité des nœuds sur un réseau bipartite mots-clés/domaines. Les nœuds en rose représentent les mots-clés, ceux en bleu les disciplines. Le diamètre des nœuds indique leur coefficient de *clustering*.

4 Réseaux d'auteur·rices

Dans notre hypothèse de départ, nous cherchions à savoir si un·e ou des expert·e·s de cet objet d'étude autour de la sorcellerie se démarquai(en)t par rapport aux autres.

4.1 Centralité

Le premier graphe présenté en FIGURE 9 nous montre le réseau des auteur·rice·s de notre corpus. Nous avons choisi comme point de départ ici de montrer la *closeness* par les couleurs. Ainsi, plus la couleur est foncée, plus l'auteur·ice est "central·e" dans le réseau. La taille des nœuds représente quant à elle le nombre d'occurrences (de publication) de l'auteur·rice dans le corpus.

Avant même d'aborder la question de la *closeness*, le résultat peut d'abord sembler décevant : en effet, si nous cherchons à montrer par ce travail des réseaux d'auteur·rice·s sur cet objet d'étude, force est de constater ici qu'il y a peu de collaboration. Le nombre de liens entre les différents nœuds est faible. De fait, on observe la formation de petites communautés de quelques auteur·rices centr·ale·aux qui n'interagissent uniquement qu'avec une petite portion d'autres auteur·rices . Mais les liens entre ces petites communautés sont en revanche inexistantes.

Une majorité d'auteur·rices est donc hors de tout réseau et n'a aucun lien avec les autres. D'ailleurs, constatons par exemple que la chercheuse qui publie le plus, Marianne Closson, avec 16 occurrences, publie seule et sans aucun lien avec le reste des auteur·rice·s listé·e·s ici.

Nous identifions un binôme isolé des autres mais avec une forte *closeness* (Julien Bonhomme et Sandra Fanello). Il·elle·s sont également dans les déposant·e·s les plus important·e·s du corpus. Les auteur·rice·s les plus "centr·ales·aux" du graphe ont quant à elles·eux un nombre d'occurrences très faible (autour de 2) car il est même impossible de distinguer ici leur nom.

On constate enfin la présence de quelques regroupements d'auteur·rice·s composés entre 5 et une petite dizaine de chercheur·euse·s. Bien que ces regroupements ne contiennent pas les auteur·rice·s qui publient le plus, ils occupent néanmoins une place centrale au sein de notre réseau de par leur *closeness* élevée.

Ce graphe qui tente d'analyser la centralité de certains auteurs sur le sujet ne nous permet donc pas de saisir de manière pertinente le fonctionnement d'un réseau sur cet objet d'étude, compte tenu de la faible collaboration qui y règne. Nous avons donc cherché à voir les choses autrement en partant du degré.

4.2 Degré des nœuds

Un second graphe (FIGURE 10), où la *closeness* est remplacée par le degré, permet d'aller un peu plus loin mais d'une autre manière. Ici, plus un nœud est de couleur foncée, plus l'auteur·rice a un fort degré et devrait donc signifier qu'il ou elle collabore davantage avec d'autres personnes sur cette thématique.

Nous voyons ici que les deux autrices qui se démarquent sont Camille Bourdier et Sylvie Beyries. Nous les avons également vues sur le graphe de *closeness* (FIGURE 9), bien qu'elles ne soient pas les plus centrales. Notre binôme qui était central dans la FIGURE 9, a ici un degré évidemment plus faible puisqu'elles ne collaborent qu'entre elles deux.

Nous pouvons également constater à partir de ce graphique qu'un autre plus petit réseau de quelques chercheur·euse·s se démarque des autres. Uniquement composé de 5 personnes, ce petit groupe n'est certes pas aussi productif que certains des plus gros·ses auteur·rices de notre réseau, mais il se démarque par l'inter·activité de ses membres. Il·elle·s font en effet parti des auteur·rice·s qui collaborent le plus au sein de notre réseau.

Ce second graphe (FIGURE 10) nous permet également d'observer, comme nous l'avons senti plus haut, que les différents clusters que nous avons décrit précédemment se démarquent beaucoup moins : les plus gros·ses déposant·es collaborent très peu ou pas du tout.

De ces deux premiers graphes, nous concluons donc à une faible collaboration générale entre les auteur·rice·s de ce corpus et sur cet objet d'étude. Une hypothèse pour expliquer ce phénomène pourrait être que ce sujet est pluri·disciplinaire. Ainsi, selon les disciplines des auteur·rice·s (et donc des publications), peut·être pourrions·nous aller

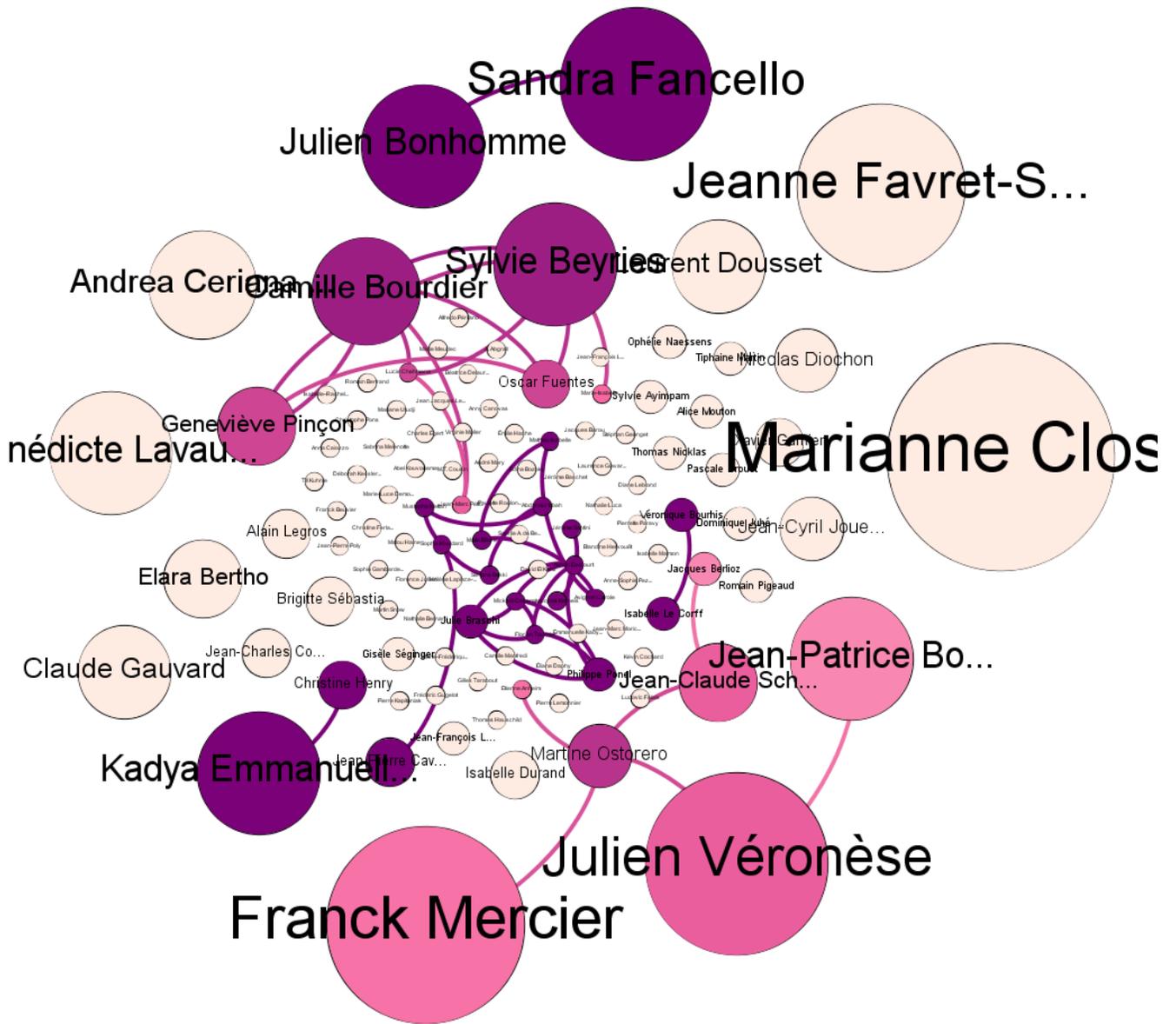


FIGURE 9 – Réseau d’auteur-ices : *Closeness* (couleur) et *Occurrence* (taille)

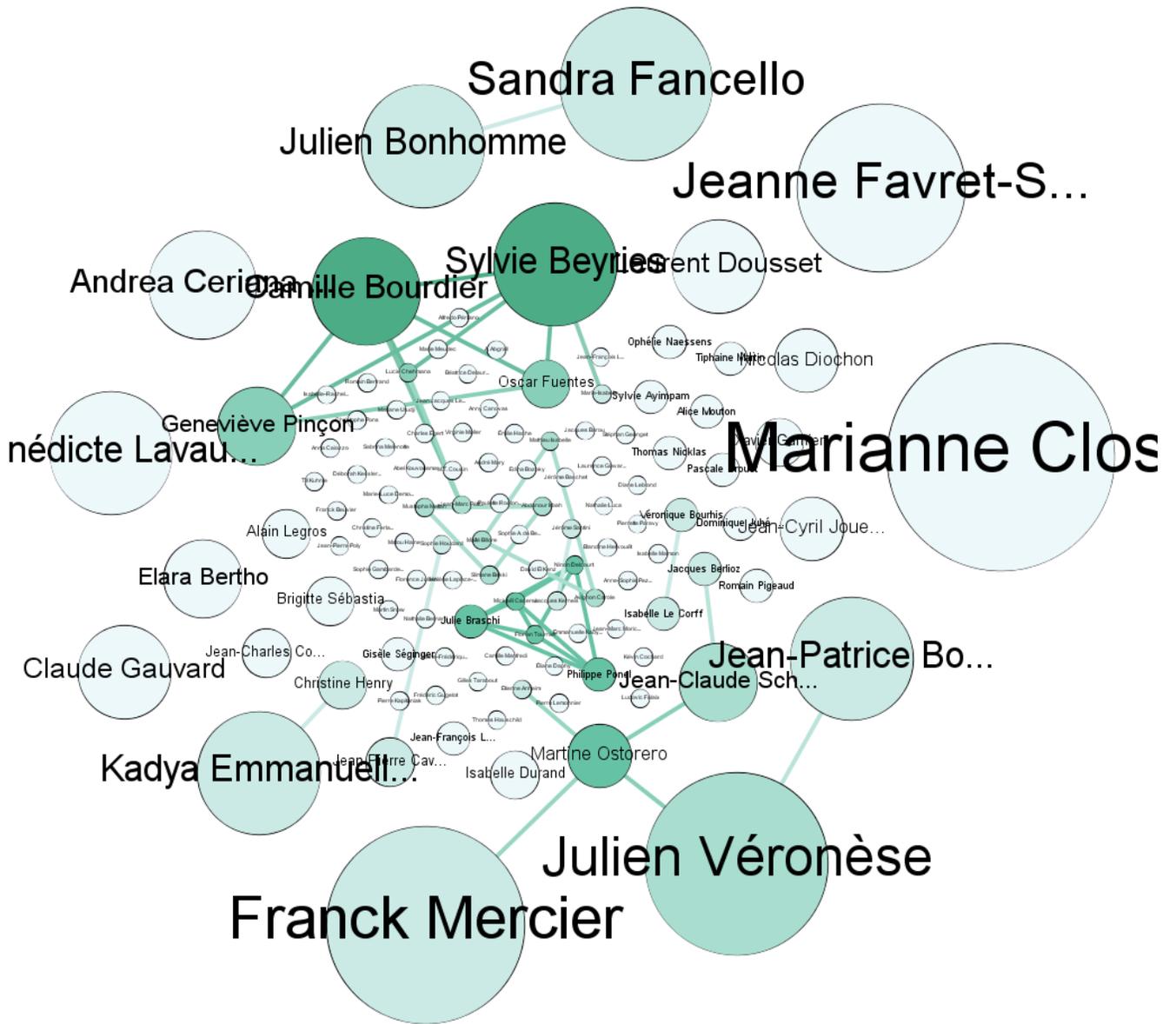


FIGURE 10 – Réseau d’auteurs : *Degré* (couleur) et *Occurrence* (taille)

plus loin dans l'analyse du corpus ?

C'est avec cette idée que nous avons tenté un réseau bipartite entre l'auteur et le domaine disciplinaire.

4.3 Exploration du lien entre l'auteur-riche et le domaine disciplinaire

Tout comme pour l'étude des mots-clefs, le lien avec les disciplines nous a semblé essentiel pour analyser de manière plus pertinente le réseau des auteur-rices compte tenu de la forte pluridisciplinarité que nous sentions autour de ce sujet. A partir de l'import de données dans Gephi, un premier travail a consisté à explorer les différents calculs et mesures, à voir les éventuels intérêts. Nous nous sommes arrêtés sur un premier graphe (FIGURE 11) où les couleurs séparent les auteur-rices d'un côté (en violet) et les domaines disciplinaires d'un autre (en vert clair). La taille de chaque noeud est quant à elle définie à partir de la *Centralité Eigenvector* qui mesure l'importance d'un noeud dans le réseau selon son importance.

Afin de rendre le graphe plus lisible, nous avons travaillé sur le tableau des données au sein de Gephi en ajoutant des colonnes et/ou en transcrivant les codes de chaque domaine en labels significatifs : *shs.litt* est devenu *Littératures*, *shs.hist* est devenu *Histoire*, etc. Ces transcriptions ont été faites à partir de la liste des domaines dans le référentiel de HAL, [AuréHAL](#)¹⁰. Enfin, comme pour les graphes bi-partites précédents, nous n'affichons pas les auteur-rices et les disciplines ayant moins de 2 occurrences dans le corpus.

Que nous apprend ce premier graphe (FIGURE 11) ?

Nous visualisons immédiatement plusieurs grandes disciplines qui se démarquent : *Histoire* en premier avec la *centralité Eigenvector* la plus élevée (1.0) suivi de *Littératures* (0.8), *Anthropologie* (0.5). *Archéologie* et *Religions* arrivent plus bas avec environ 0.2. Notons que les *Études sur le genre* n'arrivent elles que beaucoup plus bas encore avec un score de 0.06, plus bas même que la *Physique*.

Il est d'ailleurs surprenant de voir le groupement a priori inattendu autour de l'*Astrophysique*. Ce groupement devrait d'ailleurs être plus important car deux domaines de HAL porte le même nom mais pas le même code (une évolution dans le référentiel sans doute) et les codes *sdu.astr.im* et *phys.astr.im* désignent tous les deux *Astrophysique - Instrument* ; de même que *sdu.astr.sr* et *phys.astr.sr* désignent *Astrophysique stellaire et solaire*. L'*astrophysique* devrait donc être encore plus importante et deux hypothèses sont ici possibles :

- Une thématique autour de la sorcellerie ou des sorcières intéressent des auteur-rices en astrophysique, ce qui nous semble tout de même extrêmement étonnant.
- Notre requête à partir de la troncature [SORC*](#) a inclus des dépôts de HAL ne concernant pas la sorcellerie, mais un autre terme lié à l'astrophysique. Une rapide requête *via* l'API de HAL¹¹ en rajoutant le bon domaine dans les critères de recherche nous permet d'observer que les articles concernés relèvent en effet du *SORCE* (Solar Radiation and Climate Experiment), une mission satellite sponsorisée par la NASA¹².

Nous voyons pour chaque domaine disciplinaire une auréole d'auteur-riche-s dont chacun-e, au sein de la discipline est plus ou moins à égalité avec ses pairs. La taille de chaque noeud auteur-riche est aussi proportionnelle à la taille de la discipline à laquelle il ou elle se rattache puisque nous sommes basés sur une centralité Eigenvector.

Notons cependant plusieurs auteur-rices, au centre du graphe, qui se démarquent des autres : Tiphaine Martin, Virginie Muller, Florence Jullien, Frédéric Gugelot. Le prochain graphe (FIGURE 12) nous apportera une approche complémentaire. En effet, si nous analysons les liens de ces quatre auteur-rices, nous voyons qu'ils sont chacun-e relié-e à plusieurs disciplines. Tiphaine Marton par exemple, dont la centralité Eigenvector est la plus forte, est reliée à *Histoire*, *Littératures*, *Art et Histoire de l'art*, *Éducation*, *Études classiques* et *Études sur le genre*.

Les auteur-rices qui se démarquaient plus tôt en tant que gros-ses publiant-es/déposant-es (Marianne Closson, Franck Mercier) ou celles et ceux qui collaboraient beaucoup (Julien Bonhomme et Sandra Fancello) sont devenu-es quasiment invisibles au milieu de leurs pairs disciplinaires.

10. Par ailleurs, l'ensemble des codes de domaines est quand à lui trouvable [ici](#)

11. Disponible dans le fichier Google Collab

12. D'après [le site du SORCE](#)

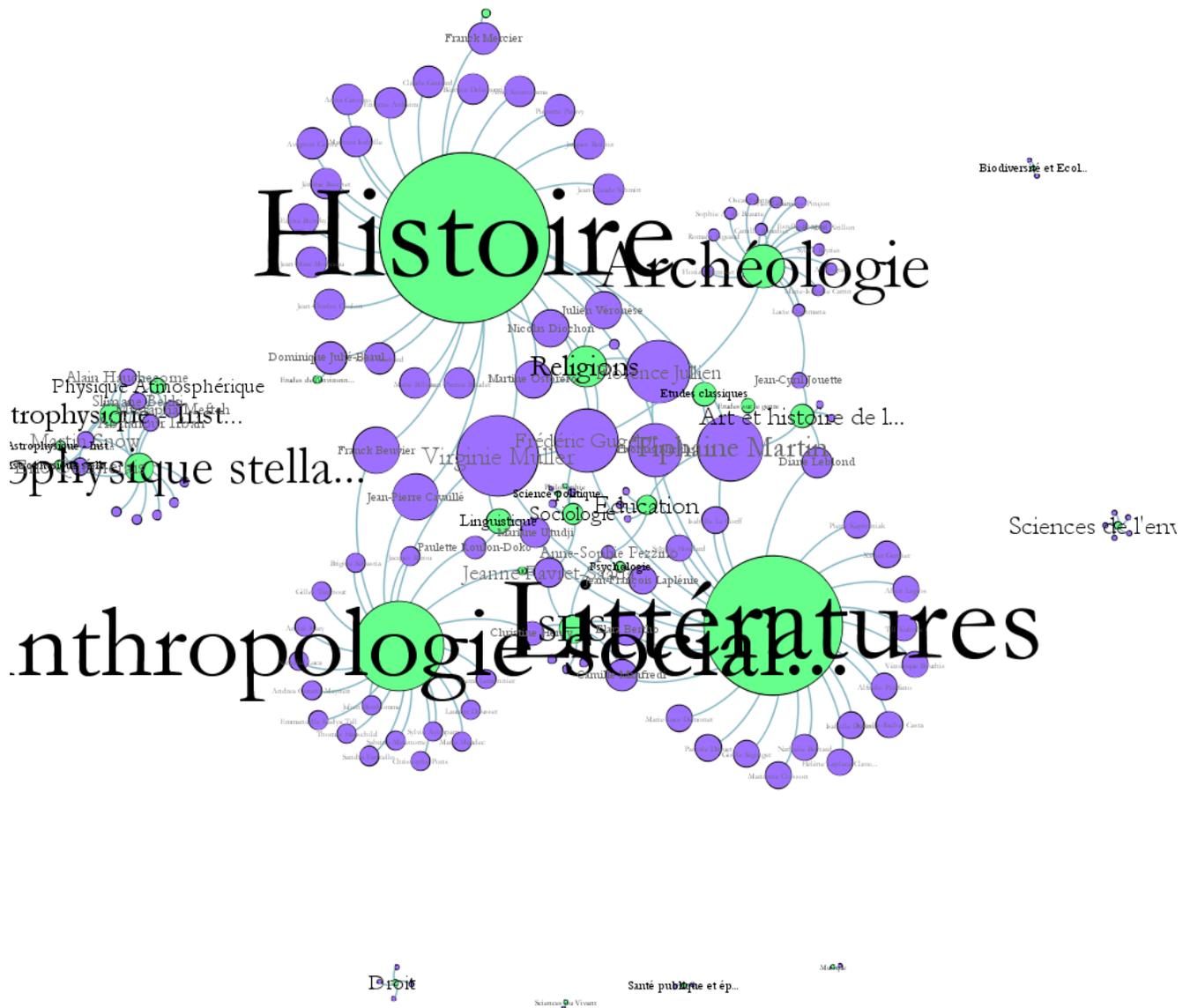


FIGURE 11 – Réseau bipartite Auteurs-Disciplines (couleurs pour le type de noeuds et *Centralité Eigenvector* pour la taille

Que nous apprend le second graphe (FIGURE 12) ?

Ce second graphe (FIGURE 12) nous affiche les mêmes données mais avec cette fois-ci des couleurs se rapportant à la détection de communauté (*Modularity Class*) et une taille des noeuds selon le *degré*. Notons que lors des tests sur Gephi, un certain nombre de mesures n'ont eu aucun résultat : par exemple le coefficient de Clustering à 0.0 pour tous.

Comme précédemment, quatre grosses communautés se détachent : *Histoire, Archéologie, Littératures, Anthropologie*. Nous laissons volontairement de côté l'astrophysique puisque sa présence est hors sujet et notons simplement qu'elle n'a effectivement aucun lien avec le reste des communautés ; tout comme certaines micro-communautés (*Sciences de l'environnement, Droit, Santé publique, Musique,...*)

Une nouvelle fois dans ce graphe, nous constatons la présence des auteur-rices dont les publications font intervenir plusieurs disciplines au centre. Leur degré est d'ailleurs plus important que les autres auteur-rices et même souvent plus important que certaines disciplines.

Cette visualisation permet de voir plus distinctement les "auteur-rices-pont" entre certaines disciplines comme les quatre dont nous parlions juste au dessus. Elle met aussi en évidence tou-tes les auteur-rices dont la centralité *eigenvector* n'est pas forte mais qui sont des "ponts" : par exemple Diane Leblond qui est l'unique lien entre la communauté *Art et histoire de l'art* et la communauté *Littératures* ou encore Dominique Juhé Beaulaton entre l'*Histoire* et les *Études de l'environnement*.

De manière anecdotique, notons que seul l'auteur Franck Mercier, l'un de nos grands déposants dans HAL, est relié au nœud NaN (c'est à dire aux données qui n'ont pas été remplies par le déposant dans HAL). Cela explique peut-être par ailleurs pourquoi sa présence est ici beaucoup moins visible que sur les premiers graphes (FIGURE 9 et FIGURE 10).

Cette exploration du lien entre les auteur-rices et les domaines disciplinaires rattachés aux articles nous a donc bien permis d'affiner notre analyse du corpus en détectant notamment des communautés hors sujet (comme l'astrophysique). Elle nous a également permis d'identifier, comme dans les graphes des mots-clefs, les disciplines les plus fortement représentées. Notons cependant quelques différences dans les résultats puisque dans les FIGURE 8 et FIGURE 6 sur les mots-clefs, on voyait plus distinctement les disciplines Sociologie et Psychologie, beaucoup moins visibles sur les FIGURES 11 et 12. Ce réseau bi-partite auteur-ice-discipline nous permet également de voir l'importance de la pluri, voire de l'interdisciplinarité du sujet ou en tout cas de certain-es auteur-rices.

5 Conclusion générale

Notre travail portait sur la thématique, assez vaste, de sorcière et sorcellerie. Pour étudier la manière dont est traitée celle-ci, nous avons effectué une requête sur HAL en récupérant tous les travaux comportant le terme `sorc*` dans les champs titre, résumé, mots-clés. Notre idée d’origine était de voir si les travaux récupérés avec notre requête faisaient ressortir les thématiques qui nous semblaient de plus en plus discutées, à savoir les questions de genre ou de féminisme. Nous avons également souhaité voir si des auteur·ice·s faisaient figure d’expert·es de cette thématique, au moins en considérant le nombre de publications.

Notre travail présente plusieurs limites. Tout d’abord, nous ne sommes pas à l’abri d’avoir « récolté » des articles ne traitant pas de sorcellerie ou des sorcières (comme on l’a vu avec l’exemple du SORCE à la section 4.3¹³). D’autre part, un traitement textuel des données récupérées, notamment au niveau des mots-clés, était manquant et aurait pu nous fournir des données plus propres. Enfin, se pose la question de la langue avec laquelle nous avons travaillé. Nous n’avons récupéré que des travaux rédigés en français, ce qui masque une grande partie des documents traitant de la question. Par ailleurs, le fait de n’interroger qu’une seule archive ouverte dans le cadre de ce rapport implique que nous n’avons récupéré, dans la grande majorité, des travaux français et récents.

Notre travail a révélé que, si les thématiques de genre apparaissent effectivement (très) légèrement, elles n’y sont pas plus représentées que d’autres, plus liées à l’anthropologie ou à l’histoire, par exemple. Par ailleurs, aucune sous-thématique ne se démarque particulièrement – elles semblent toutes avoir une importance et une centralité équivalente. De manière générale, même si les SHS occupent une place prépondérante dans le corpus, aucune discipline ne semble plus importante qu’une autre. Nous observons un phénomène assez similaire avec les graphes relatifs aux auteur·rices, bien que les disciplines les plus importantes et/ou les plus centrales ne soient pas les mêmes ; ce qu’il est par ailleurs intéressant de relever. L’autre point important qu’on analyse des réseaux d’auteur·rices est le peu de collaboration, ce qui peut sembler contradictoire avec la pluridisciplinarité du sujet. Ainsi, si dans notre corpus, les chercheur·euses collaborent peu ensemble (à quelques exceptions près), les disciplines en revanche se croisent plus souvent. Cette diversité des disciplines influe véritablement sur l’analyse à la fois des relations entre les mots-clés et entre les auteur·rices.

Pour finir, il nous semble important de faire un bilan de cet exercice, nouveau pour nous, d’analyse de données et de réseaux et disons-le, nous avons pris plaisir à analyser et étudier ce corpus de données. Cela a pu être dû au sujet décalé ou aux outils à notre disposition pour chercher et/ou faire ressortir certaines tendances ou incohérences dans nos résultats. Nous avons très souvent eu envie d’aller plus loin en proposant d’autres graphes au fur et à mesure que nous prenions le logiciel en main, d’autres lectures, non seulement pour trouver une logique ce corpus, mais aussi car chaque graphe (résultat d’une hypothèse de départ) nous amenait à nous poser de nouvelles questions et à vouloir aller plus loin dans la manipulation de notre corpus. Nous avons enfin, malgré nos tâtonnements, pu nous familiariser avec les requêtes à l’API de HAL et surtout avec la visualisation via Gephi.

13. En s’amusant un peu avec l’API, on trouve d’autres exemples : `l’un des travaux récoltés` (domaine de l’informatique) traite lui des SoRC, pour *Signatures on randomizable cipher-texts*. `Un autre encore` (domaine des mathématiques) est une thèse qui mentionne une expérimentation sur le texte du *Magicien d’Oz* – thèse qui ne semble pas parler de sorcellerie outre mesure. `Un troisième`, en chimie cette fois, traite du sorbitol et d’acide citrique (*citric acid*), raccourci avec l’acronyme SorCA. D’autres exemples de ce type existent sûrement au sein de notre corpus : cela nous montre globalement l’importance de la précision de la requête de départ, ou du nettoyage des données récupérées. Ces exemples restent tout de même minoritaires, comme on l’a vu dans l’analyse des graphes puisque ni ces disciplines, ni ces mots-clés, ni ces auteur·ices n’apparaissent de manière significative.

6 Annexes

