



HAL
open science

Scandales sanitaires et espace médiatique numérique

Dominique Cardon, Jean-Philippe Cointet

► **To cite this version:**

Dominique Cardon, Jean-Philippe Cointet. Scandales sanitaires et espace médiatique numérique. Sciences po Paris. 2019, 25 p. hal-03933171

HAL Id: hal-03933171

<https://hal.science/hal-03933171>

Submitted on 10 Jan 2023

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.

Scandales sanitaires et espace médiatique numérique - Projet exploratoire ANSES - médialab

Dominique Cardon, Jean-Philippe Cointet
médialab, Sciences Po

Introduction	2
Structure de l'écosystème médiatique français	3
Première stratégie: Les sujets connexes à l'ANSES dans l'écosystème médiatique	6
Structure de la couverture médiatique des dossiers	10
Deuxième stratégie: l'espace santé	14
Cartographie détaillée du dossier glyphosate	16
Bibliographie	19
Annexes	20

Introduction

Ce projet exploratoire avec l'ANSES s'appuie sur l'infrastructure de captation de données médiatiques numériques que le médialab a mis en place en partenariat avec le Berkman Center. Ce dispositif est ici déployé pour documenter et analyser la structure des débats médiatiques qui se cristallisent autour d'un certain nombre de dossiers qui concernent directement l'ANSES. L'objectif est d'accompagner l'Agence dans la compréhension des trajectoires médiatiques qui rendent tel ou tel sujet problématique.

Le médialab a lancé depuis le printemps dernier une veille systématique sur le web par près de 400 médias français. Cette veille consiste en l'indexation systématique de l'ensemble des contenus publiés par les principaux médias d'information en France. Notre liste de sources inclut les grands quotidiens nationaux, une grande partie de la presse régionale et locale, de nombreux magazines d'actualité générale ou spécialisée (notamment sur des questions de consommation, de santé ou d'environnement), des médias purement numériques, des blogs d'opinion influents, etc.

La cartographie des sources construite dans ce corpus de médias de l'espace public permet d'en dégager une structure définissant l'appartenance des différents médias à différentes zones de l'espace public (centre et hyper-centre, contre-espace informationnel, médias politisés, magazines et PQR). Cette structure permet de rendre compte des milieux privilégiés de déploiement des controverses sanitaires au sein de l'espace public numérique. Cette méthode cartographique permet notamment d'être attentif à l'autorité des sources, à leur visibilité et et au déplacement des controverses. La première partie du rapport résume les propriétés structurelle de l'écosystème médiatique français sur lequel nous fondons les analyses ultérieures

Notre étude se déploie ensuite en deux parties. La première partie consiste à répertorier les sujets liés aux questions de santé publique qui circulent dans l'espace médiatique française. Pour construire un tel inventaire, nous avons mis en place deux stratégies complémentaires. La première stratégie consiste à répertorier la présence médiatique et numérique des dossiers de l'ANSES à partir d'une liste d'article mentionnant l'Agence dans l'ensemble de la presse française. La seconde stratégie consiste à établir la liste des sujets les plus visibles et les plus partagés en concentrant notre attention sur des médias spécialisées sur les thématiques de santé et de bien-être. Une fois la liste de dossiers établis, on s'efforce d'en proposer une typologie en fonction des espaces de l'espace public numérique qu'ils occupent. Enfin, on s'attachera dans une troisième étude à rendre compte plus finement de la topologie des débats autour du glyphosate.

Structure de l'écosystème médiatique français

L'essentielle de nos analyses s'appuient sur la caractérisation structurelle de l'espace médiatique numérique français. Cette architecture est construite par les relations d'*autorité* qu'entretiennent les différents sites web, que l'on peut lier à l'intégration de chaque média au sein du système de reconnaissance produit par les pratiques éditoriales des journalistes. C'est ce que montre la cartographie de plus de 14000 liens hypertextes échangés entre les articles de notre corpus. Lorsqu'un article du site A envoie un lien vers le site B il s'agit bien d'un signe de reconnaissance donné par le journaliste du site A à un confrère du site B. En revanche cela ne signifie pas que B reconnaît A : le lien hypertexte est directionnel et la reconnaissance est asymétrique. Si l'on regroupe les sites qui partagent une distribution voisine de médias qui les citent et qu'ils citent, on obtient les différentes catégories que visualise la roue ci-dessous (fig 1)

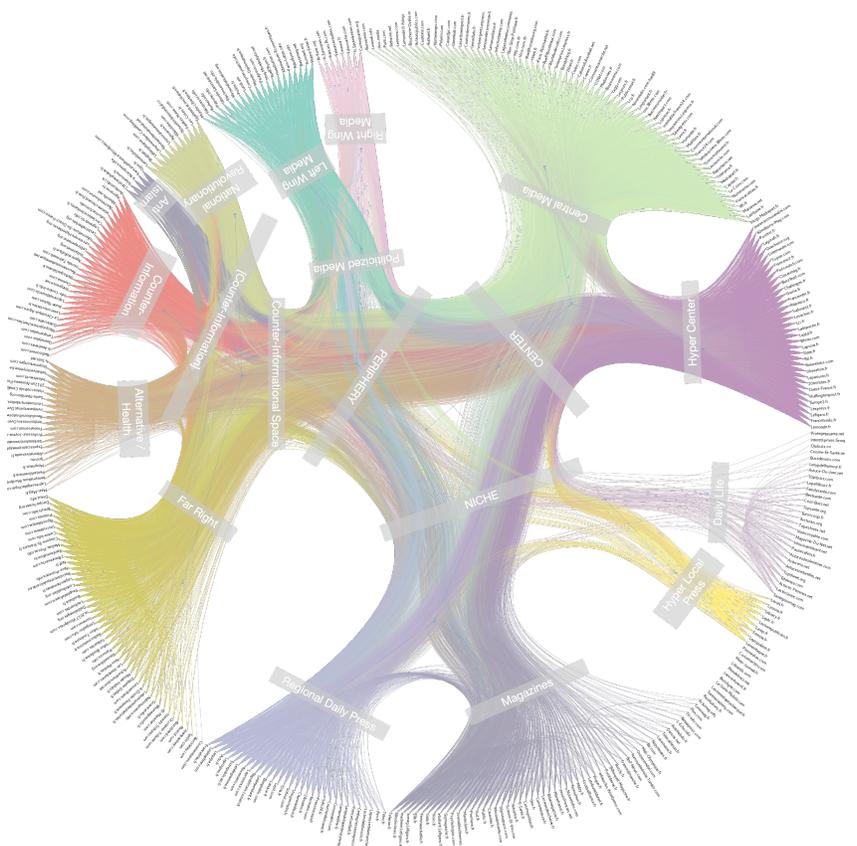


Fig. 1. La roue des médias organisée en fonction de la distribution et de la direction des liens hypertextes entre articles, blocs de couleurs obtenus grâce à la méthode SBM (Peixoto, 2014). [Version haute définition.](#)

Cette cartographie donne une représentation cohérente des différents types de médias rassemblés en fonction des critères de légitimité mis en oeuvre au sein de la profession journalistique. Trois univers apparaissent dans cet espace :

- un *Cœur* qui rassemble les médias de l'*Hyper-centre* (principaux journaux, radios et télévisions) et du *Centre* de l'espace médiatique (grands quotidiens régionaux, magazines importants et médias plus spécialisés mais dominants) ;
- une *Périphérie* qui rassemble deux groupes différents : d'une part des *Médias partisans de droite* (*Causeur, Valeurs actuelles, Fdesouche...*) et de *gauche* (*Monde diplomatique, Alternative économique, Acrimed...*), et d'autre part la galaxie des médias de *Contre-informations* que l'on peut classer en sous-ensembles (*extrême-droite, anti-islam, santé alternative, nationaux-révolutionnaire...*) ;
- des *Niches* spécialisées qui rassemblent *Magazines, PQR*, et d'autres types de médias plus éloignés de l'information nationale d'actualité.

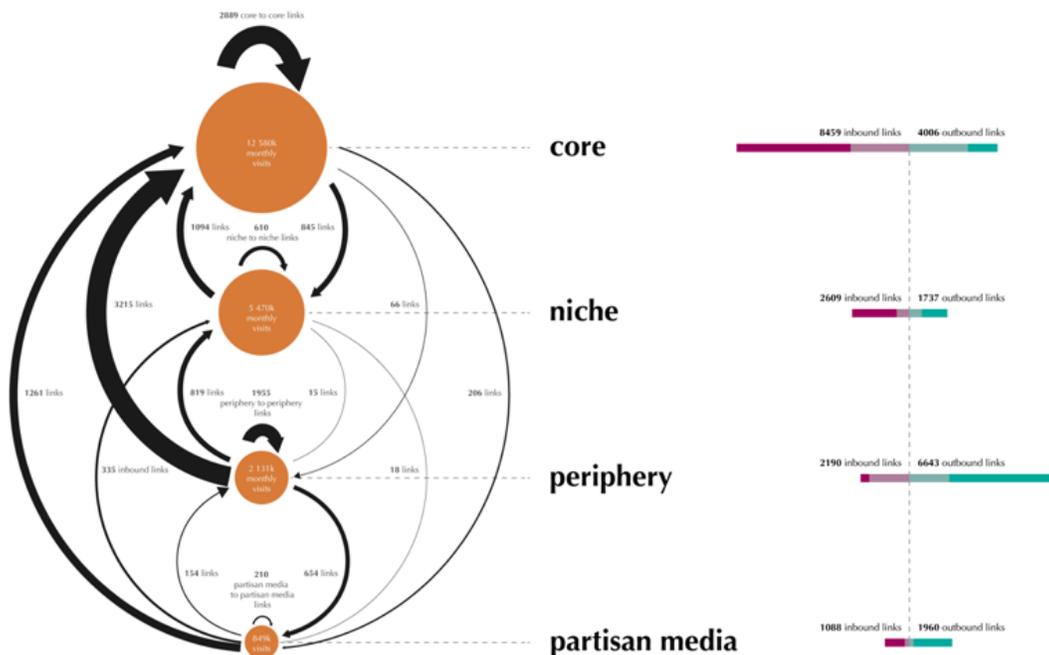


Fig 2. Diagramme résumant l'asymétrie des relations d'autorité (volume de liens hypertexte) entre classes de média (les médias "politisés" ont été isolés dans cette visualisation)

La seule distribution des liens hypertextes entre les articles des médias de notre corpus permet de construire une classification très cohérente de l'espace journalistique français. Il est frappant de constater que cette classification automatique est très proche de celle

effectuée à la main par les Décodeurs du journal *Le Monde* pour désigner la fiabilité des sources d'information selon les critères de la profession. On peut donc en conclure que le lien hypertexte entre les articles des journaux constitue toujours une très bonne approximation de l'autorité des différentes publications.

La structure inférée est en réalité hiérarchique et permet de distinguer 13 blocs distincts au niveau le plus granulaire. Nous avons retravaillé à la marge cette typologie pour intégrer une catégorie santé qui intègre aussi bien des sites de santé spécialisé qui font autorité ("Santé institutionnelle") que des sites santé et bien être relevant d'après nos calculs et selon les experts de l'ANSES de l'espace contre-informationnel ("santé alternative"). Une troisième catégorie intermédiaire regroupe 4 sites supplémentaires (cf Tableau 1). Dans les différentes parties de notre analyse, nous nous référerons tantôt aux catégories les plus fines de notre hiérarchie (médiat de droite, de gauche, Santé institutionnelle, etc.), tantôt des catégories plus grossières. Ces catégories nous serviront par la suite à rendre compte de la couverture différentielle des différents dossiers dans l'espace public médiatique. Le tableau A1 en annexe fournit une liste exhaustive des 390 médias de la roue et de leur catégorie d'appartenance au niveau le plus granulaire.

Sites de Santé	Catégorisation
Allodocteurs.fr	Santé AHS
E-Santé.fr	Santé AHS
Morandinisante.com	Santé AHS
Psychologies.com	Santé AHS
Santeacademy.com	Santé Alternative
Electrosensible.info	Santé Alternative
Alternativesante.fr	Santé Alternative
artdevivresain.Over-Blog.com	Santé Alternative
Initiativecitoyenne.be	Santé Alternative
Ipsn.eu	Santé Alternative
Professeur-Joyeux.com	Santé Alternative
Santé-Nutrition.org	Santé Alternative
Santenatureinnovation.com	Santé institutionnelle
Bestsanté.com	Santé institutionnelle
Familysanté.com	Santé institutionnelle

Protegetasante.net	Santé institutionnelle
Santeplusmag.com	Santé institutionnelle
Topsante.org	Santé institutionnelle

Tableau 1: Sites de santé et catégories associées

Première stratégie: Les sujets connexes à l'ANSES dans l'écosystème médiatique

Notre première stratégie est la suivante. L'ANSES est le point de départ de notre requête. On a construit un corpus de près de 1500 articles dont le contenu mentionne au moins une fois l'acronyme de l'Agence parmi l'intégralité de l'espace médiatique couvert par notre périmètre de veille (plus d'un million d'articles sur un an).

Cette stratégie est limitée au sens où elle ne permet pas d'embrasser l'intégralité des conversations médiatiques sur les questions de santé discutées en ligne. Pour autant, elle est intéressante car elle permet de repérer l'ensemble des dossiers dans lesquels l'ANSES est directement interpellée. On note figure 13 qu'en dépit de fortes fluctuations dues à certains événements remarquables (notamment la sortie du rapport de l'ANSES demandant la fermeture des cabines de bronzage le 10 octobre qui a généré près de 50 articles) l'activité médiatique autour de l'agence est quasiment continue. On remarque également que l'écho des articles sur les réseaux sociaux et sur Facebook en particulier est parfois disproportionné par rapport à la couverture médiatique. Ainsi l'article "*Alerte sur les antipuces bourrés d'insecticides neurotoxiques*" et publié dans Libération le 7 avril 2018 dans lequel l'ANSES est interpellée a été très largement partagé sur Facebook (39286 partages). Le mois suivant, l'alerte au slime et notamment l'article "*Brûlures, eczéma : le Slime, la pâte fabriquée par les enfants et les ados, est un jouet dangereux*" publié par actu.fr génère un nouveau pic d'activité sur Facebook.

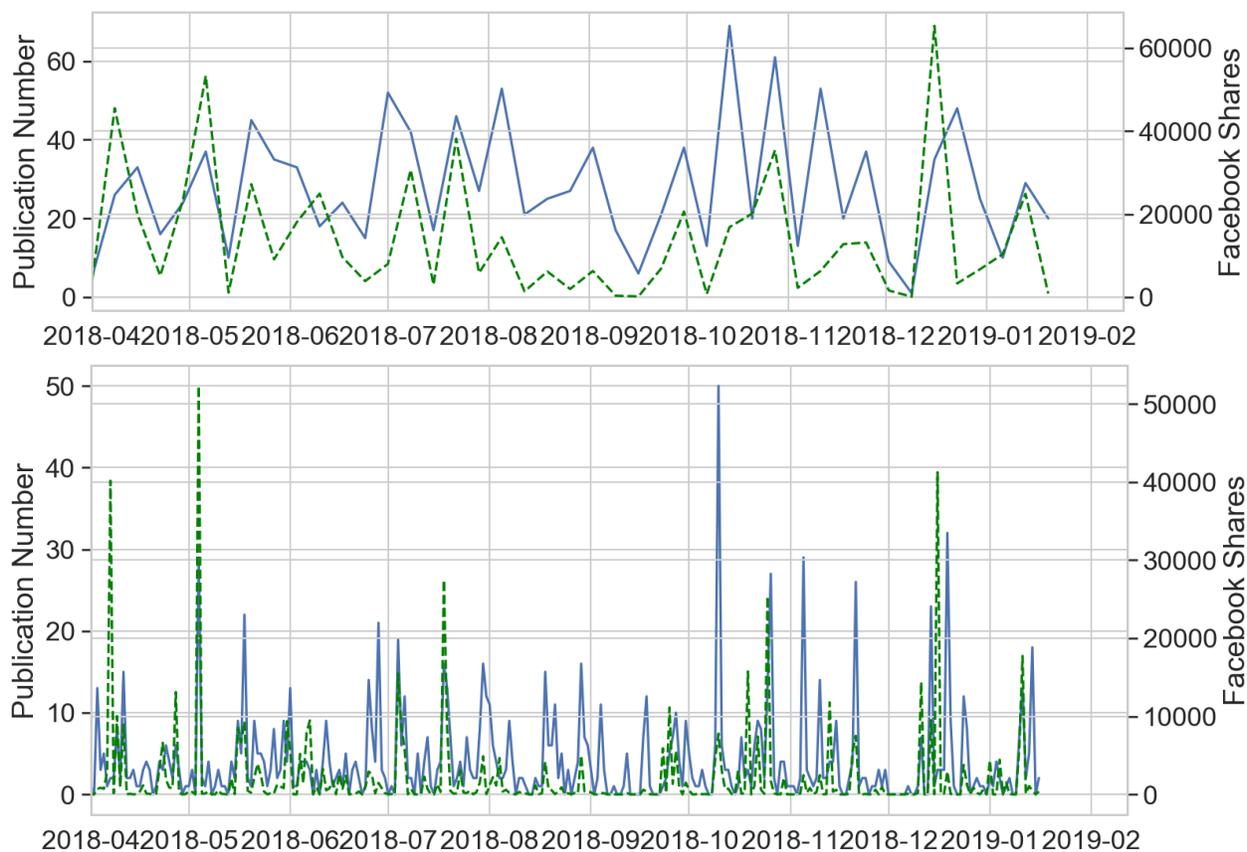


Fig 3. En bleu, évolution hebdomadaire du nombre d'articles mentionnant l'ANSES. En vert nombre cumulé de partages des articles correspondants sur Facebook. Les scores sont calculés à l'échelle de la journée (en bas) ou de la semaine (en haut)

Une fois ce corpus de documents constitué, nous avons fait appel à un modèle thématique (Blei et al, 2003). Les modèles thématiques (communément appelé "topic model" dans la littérature) infèrent automatiquement les thématiques qui composent les documents individuels d'un ensemble de documents (Cointet et Parasie, 2018). La méthode LDA (Analyse Latente de Dirichlet en français) requiert de pré-définir le nombre de sujets ou de thématiques. Dans notre cas, après avoir testé d'autre paramètres, nous avons privilégié ce choix d'une quarantaine de sujets.

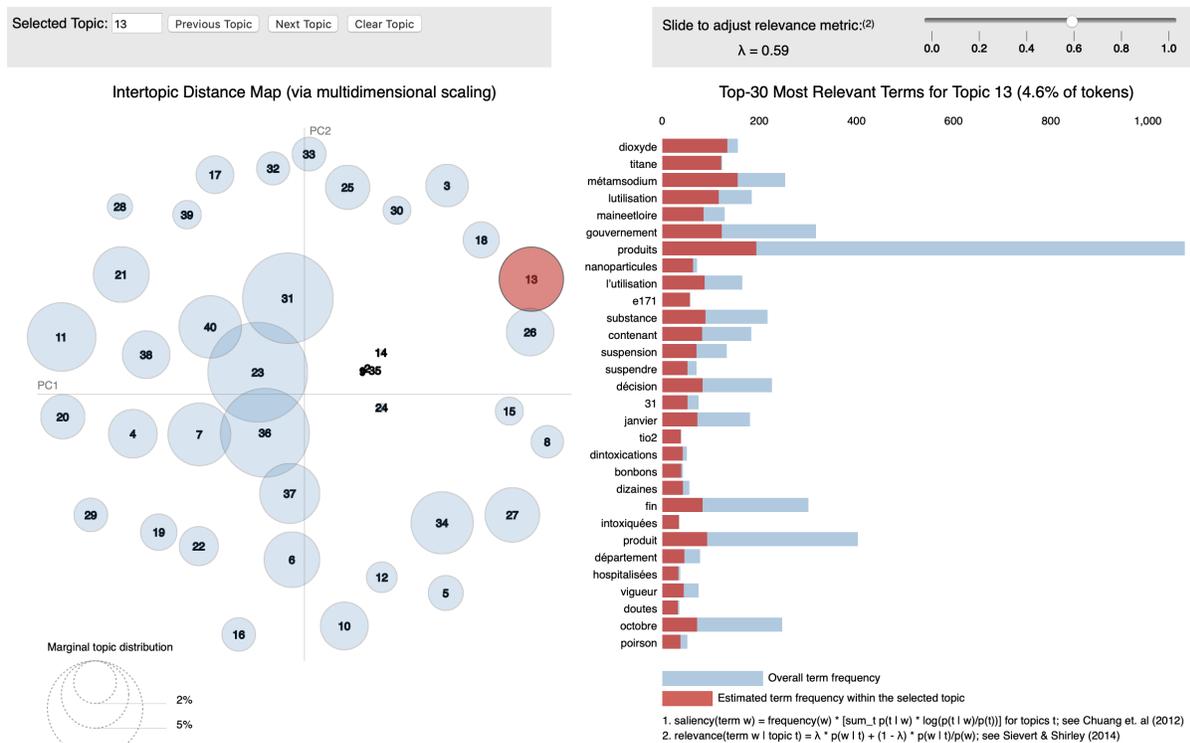


Fig 4. Visualisation du modèle thématique à 40 sujets calculé par l'algorithme LDA. Le sujet 13 traite du problème des nano-particules dans notre alimentation comme l'oxyde de titane et le colorant E171

Mais que cachent ces 40 thématiques ? Nous nous sommes largement servis de l'interface pyLDavis (Sievert and Shilley, 2014) pour explorer efficacement les différentes thématiques qui ressortent de l'analyse. La figure 4 ci-dessus donne une idée du fonctionnement de l'interface qui peut être directement interrogée [en ligne](#). Sans rentrer dans les détails, chaque thématique est définie par un ensemble de termes dont la probabilité de présence dans les documents associés à ladite thématique est forte. Nous nous servons de la distributions des termes au sein de chaque thématique mais aussi des articles les plus fortement associés à chaque thématique pour interpréter et étiqueter chaque sujet. A l'issue de ce travail, nous avons repéré une vingtaine de sujets distincts¹:

- glyphosate
- slime
- chlordécone
- protections intimes féminines
- propolis
- cabines uv
- métamsodium

¹ Certaines thématiques nous sont apparus redondantes. On a ainsi trouvé plusieurs sujets autour du glyphosate selon que la presse ne s'émeuve des risques qu'il présente pour l'alimentation, pour la santé des abeilles, etc. Certaines sujets renvoient également à des styles d'écriture plutôt qu'à des thématiques en tant que telles.

- pesticides
- tiques
- nutriscore
- vêtements neufs
- terrains synthétiques
- mélatonine
- maladies professionnelles
- pollution de l'air
- abeilles
- barquettes
- aloe vera
- EHS/électrosensibles
- titane

Ces sujets ont rythmé l'actualité de l'agence, soit parce qu'elle a récemment publié des rapports au sujet des substances ou technologies concernées (mélatonine, cabines UV, slime, nutriscore, etc.) ou parce qu'elle est concernée historiquement par des dossiers au long cours toujours sensibles (chlordécone, électrosensible, abeilles, etc.).

Par suite, nous avons construit une série de requêtes permettant de délimiter des corpus thématiques spécifiques à chaque dossier. Seuls les requêtes permettant de récupérer au moins 100 articles ont été conservées, ce qui résulte en 14 dossiers au total. A ce stade, la référence à l'ANSES n'est plus requise pour figurer dans les différents corpus.

Dossier	Requête	Taille du corpus
glyphosate	(risque AND (sante OR sanitaire)) AND glyphosate	2134
linky	(risque AND (sante OR sanitaire)) AND linky	647
slime	(risque AND (sante OR sanitaire)) AND SLIME	203
chlordécone	(risque AND (sante OR sanitaire)) AND chlordécone	253
protections intimes féminines	((risque* OR chimi*) AND (sant* OR sanit*)) AND ("protection* intime*" OR tampon* OR "serviette* hyg*")	961
métamsodium	(risque AND (sante OR sanitaire)) AND métamsodium	128
pesticides	(risque AND (sante OR sanitaire)) AND pesticide*	3706
tiques	(risque AND (sante OR sanitaire)) AND tique*	
abeille	((risque AND (sante OR sanitaire)) OR moralité) AND abeille	1007
barquettes	(risque AND (sante OR sanitaire OR contaminant)) AND barquette*	153

EHS	(risque AND (sante OR sanitaire)) AND (EHS OR électrosensib*)	378
titane	("dioxyde de titane" OR TiO2 OR E171 or 77891) AND (risqu* OR sant* OR sanit*)	237
nano	(risque AND (sante OR sanitaire)) AND nano*	728

Tableau 2: Liste des dossiers dont la requête regroupe plus de 100 articles dans le corpus global

Structure de la couverture médiatique des dossiers

Etant donnés les 14 dossiers qui viennent d'être listés, on interroge dans cette partie leur cinétique médiatique (dans la presse et sur Facebook) mais aussi la nature de la configuration des espaces médiatiques qu'ils occupent. Notre hypothèse, en criblant un grand nombre de dossiers de manière systématique est d'interroger l'existence de motifs qui associent à certains types de risques (chimiques vs physiques, nationaux vs locaux, alimentaires ou non) un certain type de réponse médiatique (magazine vs presse quotidienne, presse régionale ou nationale, droite ou gauche, presse spécialisée santé ou généraliste, etc.)

Le résultat de notre analyse est compilé pour l'intégralité des dossiers ainsi que le corpus "ANSES" dans la figure ci-dessous (figure 5)

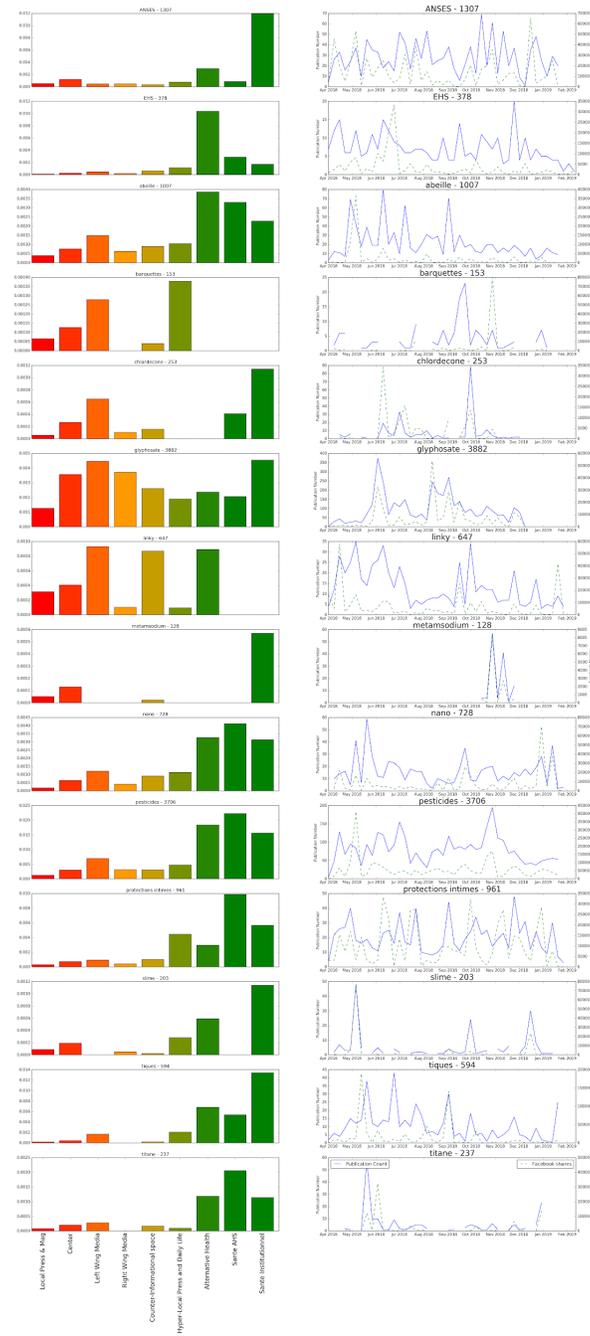


Fig. 5: A droite, le profil d'occurrences du nombre d'articles propres à chaque dossier et de leur partage sur Facebook ; à gauche, l'histogramme correspond à la fraction des articles de chaque bloc de médias qui sont capturés dans chaque corpus. De gauche à droite: PQR, Centre, Médias de gauche, Médias de droite, contre-espace informationnel, presse départementale, santé alternative, intermédiaire ou institutionnelle.

Logiquement, on observe la prééminence des sites de santé qu'ils soient institutionnels ou alternatifs dans la quasi-intégralité des cas. Seuls les dossiers barquette et glyphosate font exception. Mais avant de donner trop d'importance aux sites santé, il faut bien préciser que les histogrammes de la figure 3 montrent la fraction des articles du bloc de média considéré qui mentionnent chaque dossier. Le compte brut d'articles est donc normalisé par le nombre total d'articles que *mediacloud* a indexé dans chaque bloc. Hors, les blocs santé sont de loin les moins peuplés. La sur-représentation relative de la plupart des sujets dans ces groupes de médias ne se traduit donc pas par une sur-représentation absolue. La figure A1 en annexe montre d'ailleurs bien que les sites de blocs Coeur et PQR sont bien systématiquement les principaux pourvoyeurs d'articles liés à chaque sujet.

Intéressons nous maintenant aux motifs plus fins qui se dégagent dans la distribution des différents espaces médiatiques qu'occupent les différents dossiers si on fait abstraction des médias de santé institutionnel et de le bloc AHS (la santé alternative étant alors rapatriée dans le bloc contre-informationnel) (voir Fig 6). On voit alors émerger un certain nombre de profils typiques

- Les dossiers fortement couverts par la presse très locale: protections intimes, slimes, tiques. On retrouve des objets de consommation courants ou des menaces qui sont très territorialisées (les tiques en forêt)
- Les dossiers donnant lieu à une couverture "politisée" qu'elle soit particulièrement de gauche (abeille, chlordécone, linky, tiques, titane, nano, pesticides) ou des deux bords (glyphosate)
- Les dossiers qui relativement aux autres blocs agitent fortement la sphère contre-informationnelle (linky, EHS, titane, nano, abeilles pesticides dans une moindre mesure) qui recoupent les précédents à l'exception du dossier des tiques.

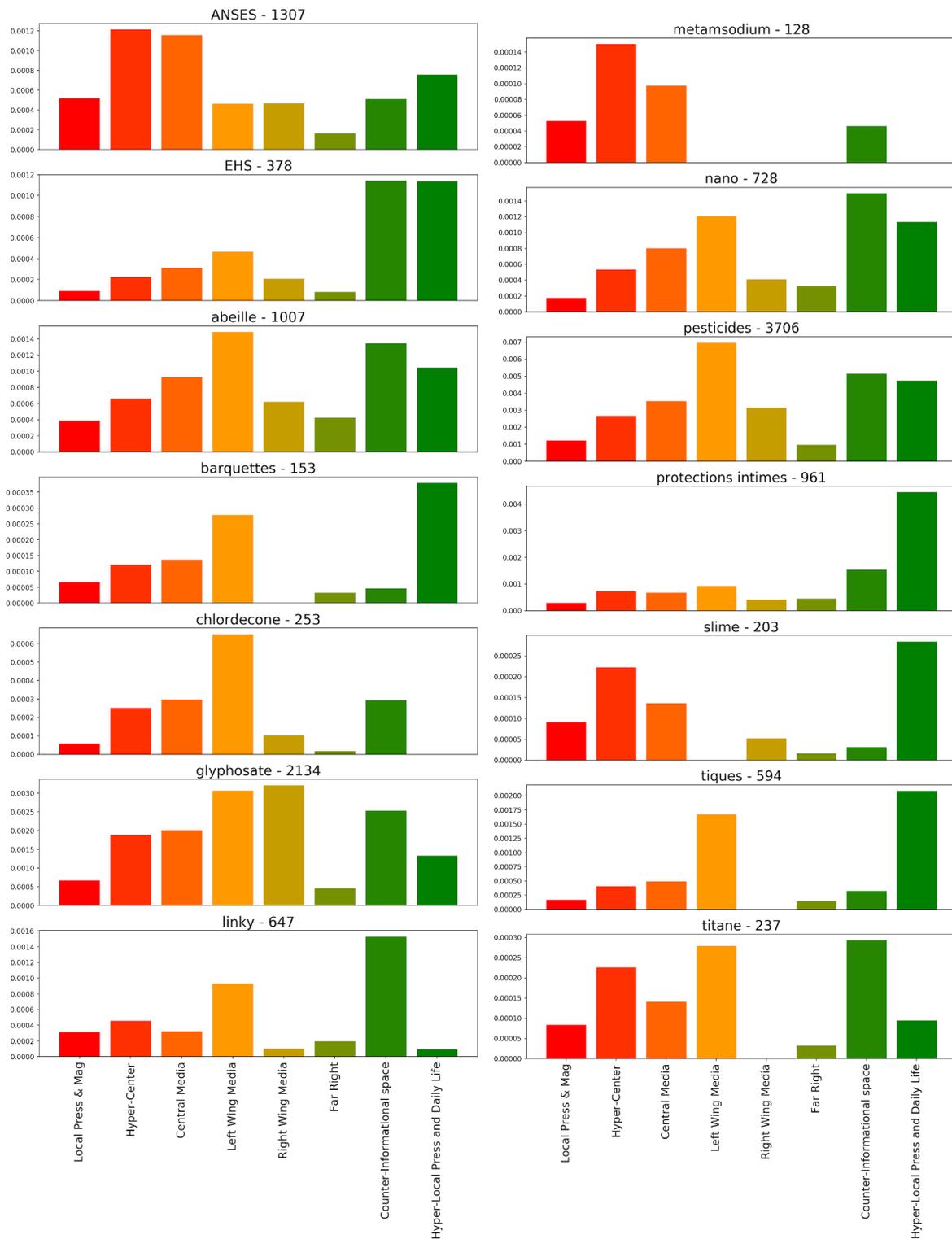


Fig. 6: Distribution des espaces médiatiques de chaque dossier, sous-blocs santé exclus.

Deuxième stratégie: l'espace santé

La deuxième stratégie de délimitation des dossiers de l'ANSES a consisté, à partir de notre périmètre de veille, à extraire les sujets de santé publique discutés fréquemment au sein d'une sous-partie de l'écosystème médiatique. On portera une attention particulière à toute une série de sources relevant du domaine de la santé (au sens large, *i.e.* incluant les questions de bien-être) et proposerons d'en décrire l'agenda.

À partir des articles du corpus, nous avons construit une carte des thématiques en suivant une méthodologie en trois étapes réalisée grâce à la plateforme CorText. La première étape consiste en l'indexation des termes les plus pertinents du corpus. Des outils de traitement de la langue sont appliqués au contenu des articles pour en extraire les groupes nominaux les plus fréquents. Un contrôle manuel a été réalisé sur près de 500 termes (incluant des n-grammes de longueur allant jusqu'à deux comme "huile essentielle" pour éliminer les termes sans intérêt (prénoms, chiffres, termes trop vagues comme "article", etc.). Dans un second temps, un réseau sémantique a été calculé liant les termes les plus similaires selon un calcul statistique qui compare les contextes d'apparition des termes dans l'ensemble du corpus. Les groupes cohésifs de couleur qui composent le réseau (identifiés automatiquement par un algorithme) sont enfin interprétés par nos soins avant de les étiqueter. La cartographie produite par cette analyse permet de distribuer très clairement un ensemble de thématiques qui composent les différents mondes de la Santé dans l'espace médiatique.

Les différents ensemble de terme font apparaître des ensembles thématiques qui permettent d'identifier assez clairement des univers dans le monde de la santé. Sur la carte sémantique des contenus publiés sur les sites de santé, on voit apparaître des thématiques que l'on peut qualifier comme suit :

Vert clair : Alimentation

Bleu : Sport, corps et douleur.

Orange : Famille, psychologie

Jaune : Visage, Peau et cosmétique

Rouge : Recherche et études

Violet : Santé publique, population et risque

Turquoise : Patients et prise en charge

Cette simple identification thématique permet de repérer aisément les différents domaines couverts par les principaux médias du corpus Santé. Si l'on prend les 12 principaux médias du corpus santé, *Psychologies* traitent principalement de la famille....

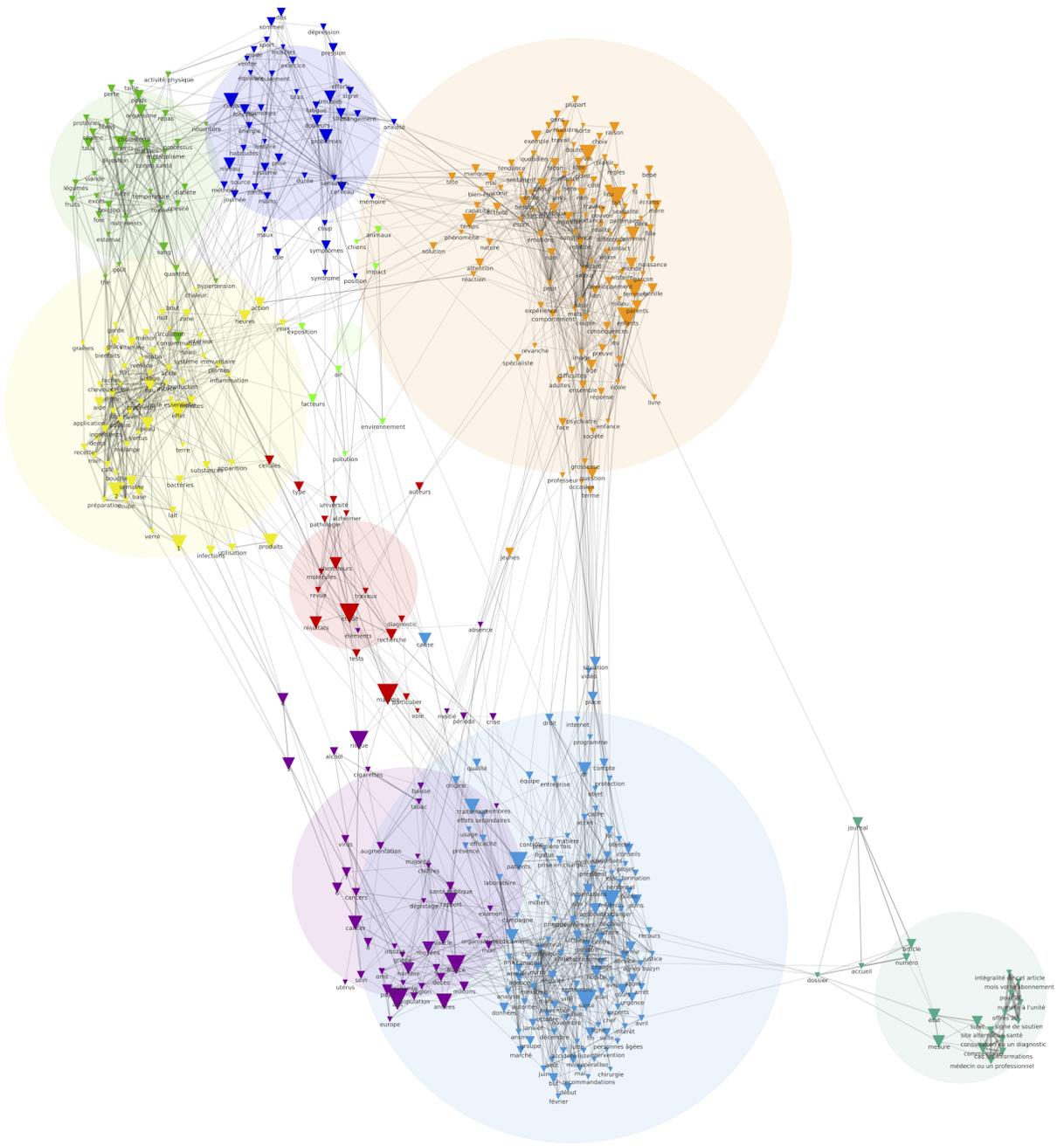


Fig 7: Carte sémantique des contenus publiés parmi les sites santé

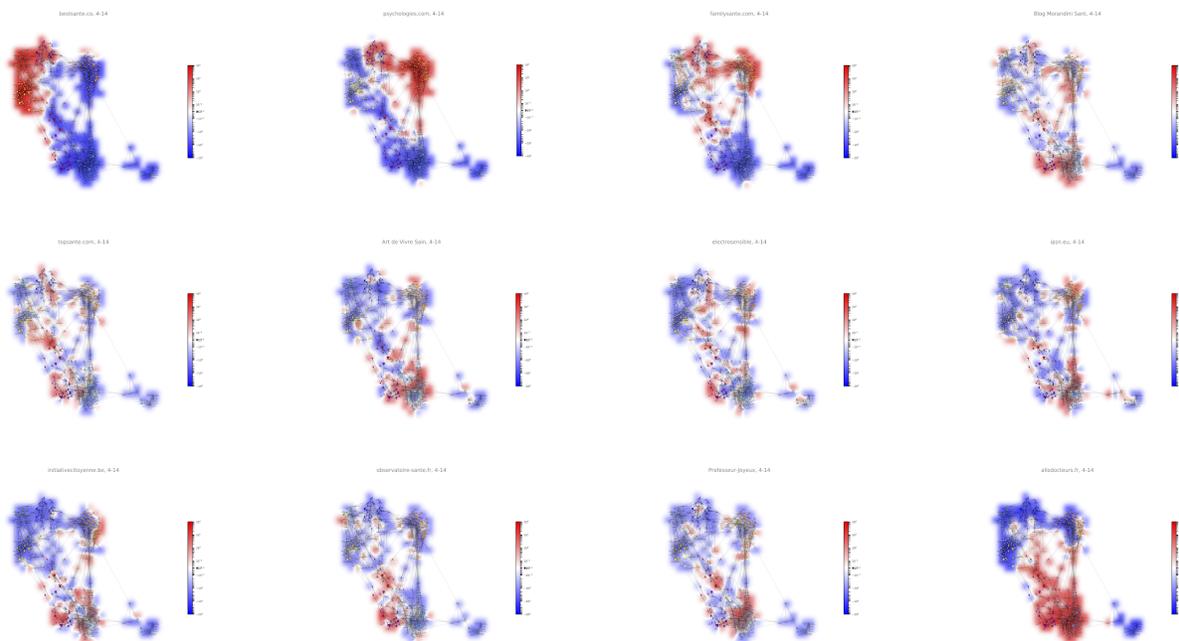


Fig 8 : Carte de densité pour les 12 principaux sites santé de notre corpus

Cartographie détaillée du dossier glyphosate

Nous nous penchons maintenant sur le cas particulier du dossier glyphosate dont nous analysons la structure sémantique qui émerge des 2134 articles qui le composent. Le résultat de la cartographie dont la construction renvoie aux mêmes principes de construction (extraction terminologique, calcul de proximité sémantique entre paires de termes et détection de clusters cohésifs formant autant de thématiques) que la carte précédente. Le paysage médiatique se découpe comme suit (de haut en bas dans le sens horaire):

- [en rouge] Le procès et l'amende record imposée par un tribunal américain à Monsanto suite à la plainte du jardinier Dewayne Johnson, malade d'un cancer, suite à une exposition au Roundup de la marque
- [en vert foncé] Autre récit juridique, celui de Paul François qui a également poursuivi Monsanto pour intoxication notamment devant le tribunal de Lyon
- [en jaune] Le monde de la régulation, de la controverse scientifique et de l'expertise autour des chiffres de dangerosité du glyphosate. On retrouve les grandes institutions internationales mais aussi l'ANSES.
- [en violet] la politique d'occupation des territoires: de la forêt aux exploitations agricoles
- [en orange] les risques liés à la biodiversité

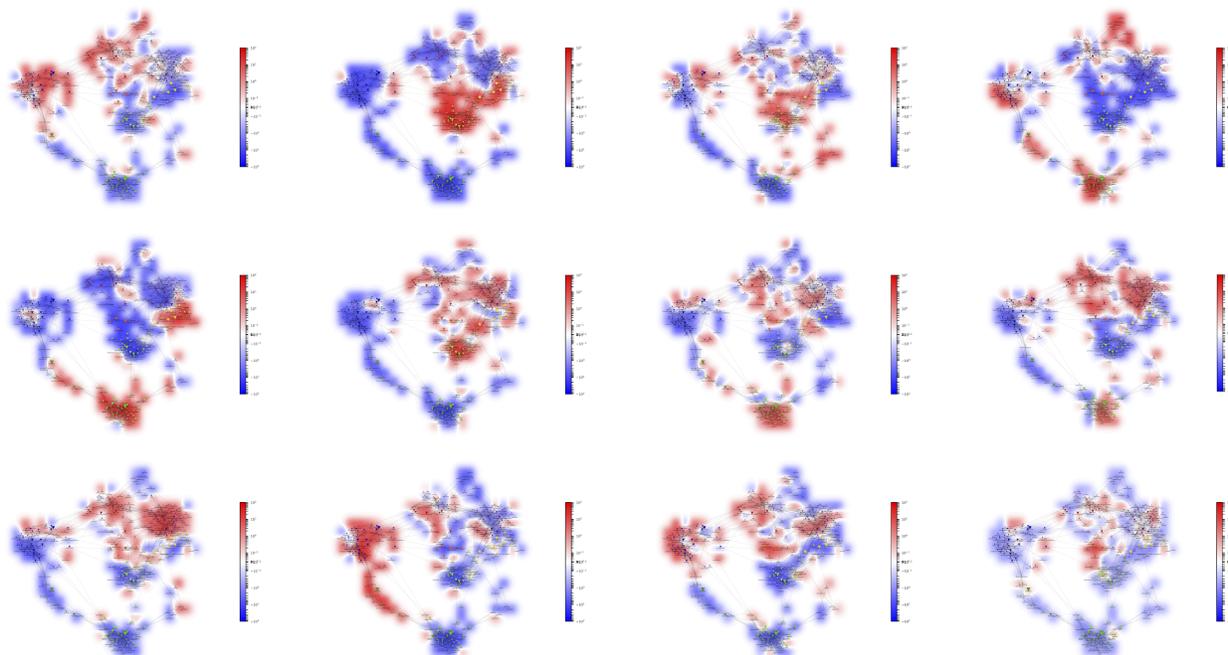


Fig 10 : Un an de discussion sur le glyphosate depuis le mois d'avril 2018 jusqu'à mars 2019. Les "heatmaps" correspondent aux profils d'activité du dossier mois après mois (avril, mai, juin, juillet 2018 sur la première ligne, août, septembre, octobre, novembre sur la deuxième ligne, et décembre, janvier, février et mars en dernière ligne). La condamnation de Monsanto par le tribunal de San Francisco date ainsi d'août 2018. Le vote des députés contre l'amendement proposé par l'élu LREM François-Michel Lambert, qui proposait d'inscrire dans la loi l'interdiction du glyphosate d'ici 2021, à l'occasion de l'examen du projet de loi alimentation et agriculture en septembre.

La figure 8 est intéressante pour rendre compte des différences de traitement médiatique du dossier en fonction des différents blocs de médias. Les médias de droite et de gauche ont ainsi un discours très politisé. C'est aussi le cas des blocs d'extrême droite et de la contre-information (ce dernier se distingue néanmoins par son insistance sur la fusion Monsanto/Bayer). Les déboires juridiques de Monsanto aux Etats-Unis sont préférentiellement traités par les médias mainstream nationaux (Media Centraux, Hyper Centre) tandis que la PQR (mais aussi les sites santé) s'est plus directement intéressée aux questions de régulation et à la question du modèle agricole adéquat.

Annexes

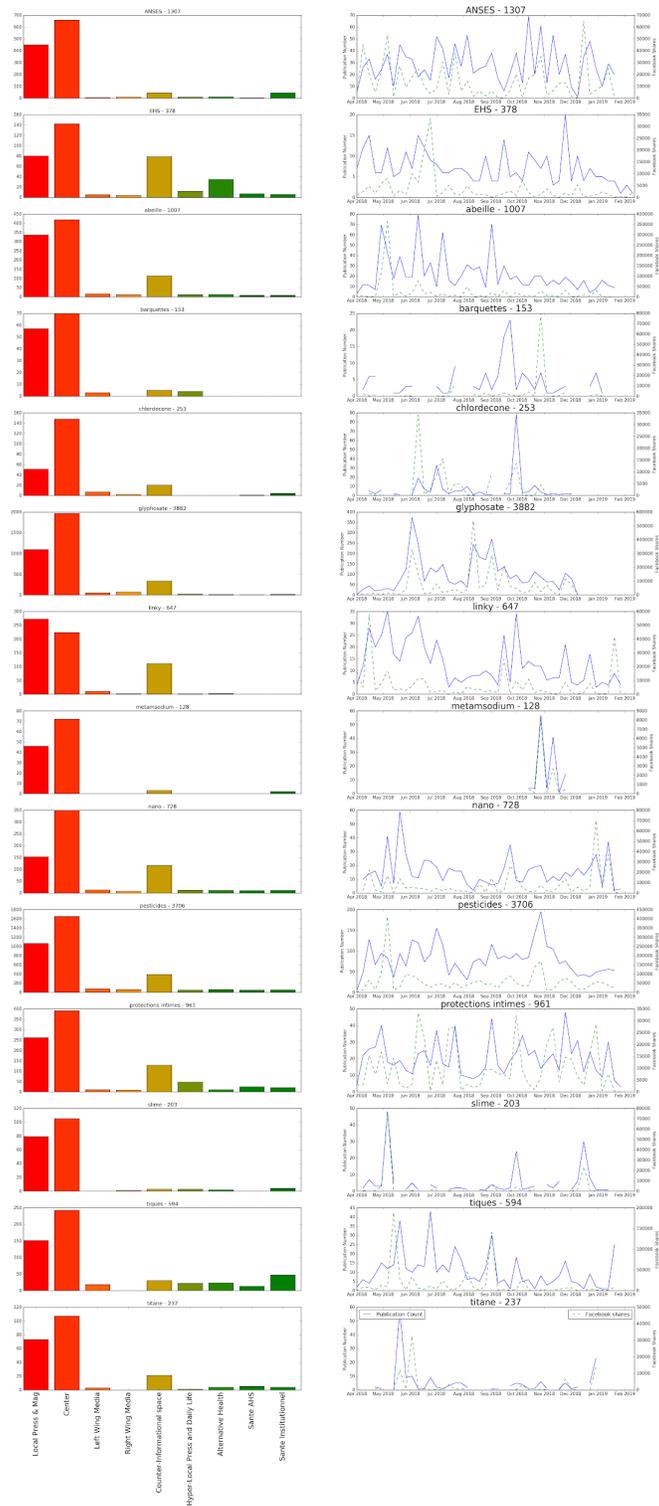


Fig A1. Couverture brute des différents sujets

