



ARTICLE

(Translated from the original English version)

## Analyse de réseau en droit de la concurrence

Yann Guthmann,\* Adrien Frumence,\*\* & Camille Hoogterp\*\*\*

**Abstract.** Le service de l'économie numérique de l'Autorité française de la concurrence (« ADLC ») a développé un outil capable d'identifier les citations des publications de l'ADLC dans ses autres publications et de représenter ces interconnexions dans un graphe. L'intégration d'un outil puissant dans, l'analyse de réseau, constitue une innovation puissante dans le domaine de recherche sur le « computational antitrust ». L'article décrit d'abord la méthode et les données utilisées pour créer le graphe afin de tirer de l'analyse du graphe des enseignements intéressants. Il présente ensuite une méthode permettant d'identifier et de mesurer les publications ayant un fort impact. Enfin, il présente les futures pistes de recherche potentielles dans le domaine du « computational antitrust ».

\* **Yann Guthmann** est le chef du service de l'économie numérique de l'Autorité française de la concurrence.

\*\* **Adrien Frumence** est Data Scientist et rapporteur à l'Autorité française de la concurrence.

\*\*\* **Camille Hoogterp** est assistante rapporteure à l'Autorité française de la concurrence.

## I. Introduction

Tous ceux qui travaillent dans le domaine du droit de la concurrence connaissent le célèbre arrêt de la Cour de justice de l'Union européenne (« CJUE ») intitulé « Bananes Chiquita » (*United Brands*), dans lequel la CJUE définit, pour la première fois, une position dominante.<sup>1</sup> Cette définition est encore utilisée et citée dans presque tous les cas d'abus de position dominante. Cela montre comment certaines décisions peuvent sortir du lot et changer la façon dont nous pratiquons le droit.

Le renouvellement constant de la pratique du droit n'est pas seulement dû aux revirements de jurisprudence, mais aussi au fait que le droit (notamment le droit de la concurrence) suit le rythme effréné du monde des affaires. Au cours des dernières décennies, la révolution numérique a remodelé notre monde entier. Par conséquent, l'adoption de technologies informatiques, en particulier d'outils basés sur la science des données, par les administrations et les organismes publics est essentielle pour suivre les pratiques commerciales et rester efficace.

Depuis sa création en septembre 2020, l'un des principaux objectifs du service de l'économie numérique de l'Autorité française de la concurrence (ci-après « ADLC ») est de développer des outils de *data science* pour son application du droit de la concurrence. Outre son expertise numérique, l'équipe fournit aux rapporteurs des outils et des informations efficaces pour faciliter leurs enquêtes, quelle que soit la nature de l'affaire. Sur cette base, le service de l'économie numérique s'est associé au projet « Computational Antitrust » de CodeX pour apporter sa contribution dans le domaine.<sup>2</sup>

Comme nos collègues du projet « Computational Antitrust », nous considérons les affaires de concurrence comme la matière première du « computational antitrust ». Dans cet article, nous décrivons et fournissons un outil précieux, rendu possible par l'exploitation de cette matière brute, pour améliorer l'efficacité du travail des autorités de la concurrence dans l'environnement actuel en constante évolution.

Nous partons d'un constat simple : les publications de l'ADLC se citent souvent les unes les autres. Identifier et enregistrer ces citations manuellement peut être très fastidieux, mais heureusement, le processus peut être automatisé. En plus de fournir une méthode pour automatiser ce processus, nous avons créé un outil pour visualiser le résultat. L'outil prend la forme d'un graphe de réseau dans lequel les publications de l'ADLC sont représentées et reliées entre elles par leurs citations. L'outil permet finalement de découvrir des tendances sous-jacentes dans les citations des publications.

---

<sup>1</sup> Affaire 27/76, *United Brands Company et United Brands Continentaal BV contre Commission des Communautés européennes*, 1978 ECLI:EU:C:1978:22, 65.

<sup>2</sup> Thibault Schrepel, *Computational Antitrust: An Introduction and Research Agenda*, 1 STAN. COMPUT. ANTITRUST 1 (2021).

Nous abordons trois questions de recherche :

1. Peut-on identifier les publications ayant un fort impact en utilisant l'analyse de réseau ?
2. Comment les publications de l'ADLC sont-elles reliées les unes aux autres et quels enseignements peut-on tirer de l'analyse des graphes ?
3. Comment améliorer le processus d'élaboration des décisions après avoir découvert les modèles sous-jacents ?

La partie I de l'article décrit la méthodologie utilisée pour créer le graphe de réseau, la partie II analyse le graphe et en tire quelques premières conclusions, et la partie III développe les prochaines étapes qui peuvent être facilement réalisées afin d'étendre le travail. Cet article vise donc à ouvrir la voie à d'autres passionnés du « computationnel antitrust » qui souhaitent contribuer à un domaine de recherche très fructueux.

## II. La méthodologie

Un réseau est un ensemble d'éléments, c'est-à-dire de sommets ou de nœuds, avec des connexions entre eux, c'est-à-dire des arêtes.<sup>3</sup> L'analyse des réseaux complexes est une discipline qui explore les relations quantitatives dans les réseaux présentant des structures non triviales et irrégulières (par exemple, les affaires de concurrence). Dans ce contexte, la sous-partie 1 décrit la manière dont nous avons collecté les données, les sous-parties 2 et 3 expliquent l'algorithme appliqué pour créer le réseau complexe et le déployer en tant qu'application web et ce que nous pouvons en tirer à première vue, et la sous-partie 4 détaille la manière dont nous avons enrichi les données.

### A. Identifier les citations

La première étape consiste à définir l'ensemble de données de départ à partir duquel nous allons identifier et extraire les citations. Nous avons choisi d'utiliser la base de données exhaustive comprenant les 350 décisions, 276 avis et 9 mesures conservatoires publiés par l'ADLC entre 2009 et 2021 et accessibles au public sur le site web de l'Autorité.<sup>4</sup> Ces 635 publications représentent un total de 24 244 pages.

Nous avons choisi la période 2009-2021 pour deux raisons. Premièrement, l'ADLC existe sous sa forme actuelle depuis la mi-2008. Ainsi, l'année 2009 est la première année complète où l'ADLC a fonctionné avec sa structure de gouvernance actuelle. Deuxièmement, les publications plus anciennes sont souvent rendues obsolètes par les développements récents. Pour des raisons de simplicité, nous ne nous concentrons que sur trois types de publications (décisions, mesures conservatoires et avis) et nous laissons de côté toutes les publications relatives au contrôle de concentration, qui se sont élevées à 2 746 publications au cours de la même période.<sup>5</sup>

---

<sup>3</sup> Mark EJ Newman, *The Structure and Function of Complex Networks*, 45 SIAM REV. 167 (2003).

<sup>4</sup> Autorité de la concurrence, *Liste des avis et décisions*, <https://www.autoritedelaconcurrence.fr/fr/liste-des-decisions-et-avis>.

<sup>5</sup> Autorité de la concurrence, *Liste des décisions de contrôle des concentrations*, <https://www.autoritedelaconcurrence.fr/fr/liste-de-contrôle-des-concentrations>.

Type de publication	Modèle
Décision	AA-D-NN
Avis	AA-A-NN
Mesure conservatoire	AA-MC-NN

L'étape suivante consiste à identifier et à extraire les citations dans l'ensemble de données. Cette étape est facilitée par la façon dont l'ADLC numérote ses publications, en suivant toujours ce modèle :

**Tableau I :** La numérotation des publications de l'ADLC

Où AA représente les deux derniers chiffres de l'année de publication et N représente un nombre qui est incrémenté pour chaque type de publication. Par exemple, une décision serait nommée 14-D-13 (la 13e décision de 2014), un avis serait nommé 15-A-18 (le 18e avis de 2015) et une mesure conservatoire serait nommée 19-MC-01 (la 1ère mesure conservatoire de 2019).

Plus généralement, l'ADLC compte 22 types de publications qui suivent toutes le même modèle général : 2 chiffres - 1 à 5 lettres - 2 à 3 chiffres. Par conséquent, nous avons développé un programme qui identifie ces modèles et l'avons fait tourner à travers les plus de 20 000 pages de l'ensemble de données. Ce faisant, nous avons pu identifier 14 036 citations à partir de notre point de départ de 635 publications.

La première constatation est que 40 % des citations sont des références à des décisions relatives à des informations commerciales confidentielles qui ne sont citées qu'une seule fois par la publication de l'affaire de concurrence à laquelle elles se réfèrent. Face à ce constat, nous avons écarté toutes les citations qui ne sont pas une décision, un avis ou une mesure conservatoire. La base de données de citations représente 8375 citations extraites et se compose de 662 décisions, 341 avis et 40 mesures conservatoires pour un total de 1043 publications de 1987 à 2021.

## B. Le graphe de réseau complexe

Deux des concepts de réseau les plus importants sont les éléments et les connexions entre eux. Les éléments sont appelés des nœuds et les connexions des arêtes. Ici, la construction du réseau complexe à la fois programmatically (en utilisant NetworkX)<sup>6</sup> et interactivement (en utilisant Dash Cytoscape)<sup>7</sup> nécessite d'abord de faire de chaque publication un nœud et de représenter les relations

<sup>6</sup> La bibliothèque Python fournit une collection de fonctions pour construire, mesurer et dessiner des réseaux complexes. NetworkX, <https://networkx.org/>.

<sup>7</sup> Dash Cytoscape est un composant Python de visualisation de graphes permettant de créer des réseaux facilement personnalisables, performants, interactifs et basés sur le web. Plotly, <https://dash.plotly.com/cytoscape>

(citations) entre elles comme des arêtes. Les relations entre les entités sont naturellement dirigées (asymétriques) : une publication de 2021 peut citer une publication de 2014, mais l'inverse n'est pas possible. Alors que les nœuds ou les arêtes décrivent les propriétés structurelles des réseaux, nous ajoutons également des propriétés non structurelles en tant qu'attributs du réseau. Il s'agit principalement des métadonnées de chaque publication, de l'extraction de la structure de la citation et du score d'audience sur le web. Enfin, nous sélectionnons les positions géométriques de chaque nœud selon l'algorithme de mise en page.

Pour ce faire, nous choisissons l'algorithme de Fruchterman et Reingold pour dessiner le graphe complexe basé sur les forces.<sup>8</sup> Avec cet algorithme, tous les nœuds se repoussent, respectant ainsi le principe des aimants. Plus les nœuds sont éloignés les uns des autres, moins ils se repoussent. Les arêtes agissent comme un ressort entre deux nœuds. À chaque passage de l'algorithme (c'est-à-dire 5 000 périodes), la somme des forces est appliquée à chacun des nœuds. Ces nœuds sont déplacés jusqu'à ce qu'un état stable soit trouvé.

Également appelés « spring embedders », ces algorithmes calculent la disposition d'un graphe en utilisant uniquement les informations contenues dans la structure du graphe lui-même, plutôt que de s'appuyer sur des connaissances spécifiques à un domaine (par exemple, le droit de la concurrence).

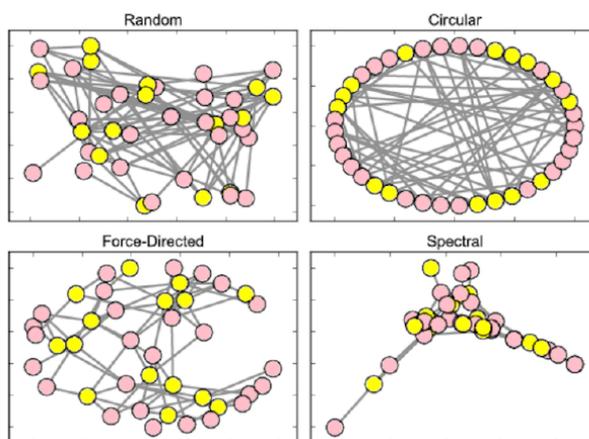


Figure I : Aperçu des algorithmes de mise en page populaires pour le dessin de réseaux complexes<sup>9</sup>

Par conséquent, dans notre réseau complexe, les publications sont classées en fonction de leurs similitudes. Cela signifie que même les publications qui ne se citent pas directement les unes les autres peuvent être fortement connectées et très étroitement positionnées. Selon cette mesure, deux publications dont les citations et les références se chevauchent fortement sont supposées avoir plus de chances de couvrir un sujet connexe.

<sup>8</sup> Thomas M. J. Fruchterman & Edward M. Reingold, *Graph Drawing by Force-Directed Placement*, 21 SOFTWARE, PRAC., & EXPERIENCE 1129, 1131-38 (1991).

<sup>9</sup> Dmitry Zinoviev, COMPLEX NETWORK ANALYSIS IN PYTHON : RECOGNIZE - CONSTRUCT - VISUALIZE - ANALYZE - INTERPRET, 58-59 (Adaobi Obi Tulton ed., 2018) (affichant les mises en page de ces algorithmes et fournissant des explications de base sur leur fonctionnement).

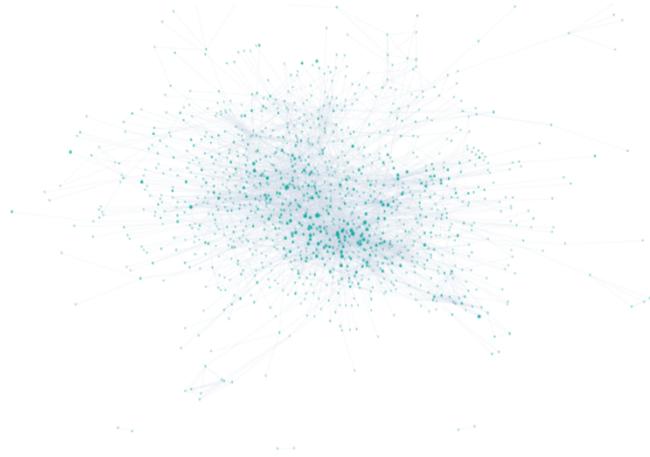


Figure II : Vue d'ensemble du réseau complexe<sup>10</sup>

### C. Ce que l'on voit à première vue

À première vue, le graphe est difficile à lire, étant donné ses 1 173 nœuds et ses 4 015 arêtes. Comme le graphe n'est pas entièrement connecté, nous le partitionnons en sous-ensembles.<sup>11</sup> Plus précisément, le graphe possède huit composantes connectées. Mais il existe une disparité entre ces composantes : l'une d'entre elles contient 1 157 publications/points, représentant 98,7 % de tous les nœuds (appelée la composante connectée géante).<sup>12</sup> Dans l'ensemble, presque tous les documents sont connectés, et seuls quelques-uns d'entre eux sont isolés. Une curiosité intéressante est le fait que 16 documents sont semi-isolés, ce qui signifie qu'ils ne sont pas connectés au sous-ensemble principal du graphe mais créent sept sous-ensembles très petits (composés pour la plupart de deux/trois nœuds).

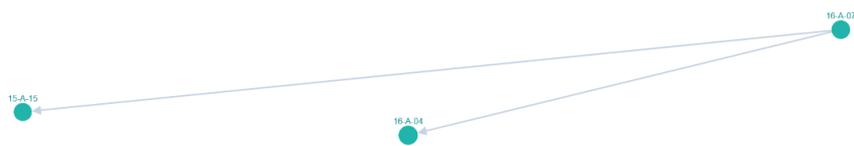


Figure III : Un exemple de sous-graphe composé de trois nœuds qui ne sont pas connectés au réseau complexe principal

Il faut garder à l'esprit que notre réseau complexe n'est pas un arbre de citations. Cela a déjà été fait auparavant.<sup>13</sup> Dans notre graphe de co-citations, les

<sup>10</sup> Le réseau est disponible à l'adresse <https://sen-codex.dev/>, ou dans la section « Chiffres clés » du site web de l'AFC.

<sup>11</sup> Un réseau convexe est un graphe dans lequel on peut relier, directement ou indirectement, tout nœud à tout autre nœud du graphe par une chaîne d'arêtes.

<sup>12</sup> Mark Newman, Albert-László Barabási, & Duncan J. Watts, *Chapter Three: Empirical Studies*, in *THE STRUCTURE AND DYNAMICS OF NETWORKS* 167, 170 (Mark Newman, Albert-László Barabási, & Duncan J. Watts eds., 2006) (définissant diverses formes de composantes connectées géantes).

<sup>13</sup> Voir Citation Tree, <https://www.citationtree.org>.

publications sont classées en fonction de leurs similarités.<sup>14</sup> Cela signifie que même les publications qui ne se citent pas directement peuvent être fortement connectées et positionnées à proximité les unes des autres dans le graphe.



Figure IV : Les publications liées aux télécommunications sont mises en évidence par des carrés rouges - le graphe montre des liens forts et cachés entre elles.

Sur le graphe ci-dessus, on peut voir les positions proches de la décision 20-D-01 relative à une pratique mise en œuvre dans le secteur de la diffusion de la télévision numérique terrestre et de l'avis 12-A-13 relatif à l'analyse des marchés de gros en amont des services de diffusion terrestre de programmes audiovisuels. La décision 20-D-01 ne cite pas l'avis 12-A-13, mais les deux sont néanmoins fortement connectés et positionnés à proximité l'un de l'autre dans le réseau complexe. La méthode décrite ci-dessus pourrait améliorer la précision des recommandations de publication.

#### D. Création du livre de code

L'extraction de l'ensemble des données de citation est une première étape nécessaire. Après avoir cherché une méthodologie pour améliorer la qualité des données affichées par le graphique (en ajoutant le contexte de chaque citation), nous avons choisi d'utiliser la structure des publications, c'est-à-dire les titres et sous-titres. Connaître les parties de la publication dans lesquelles se trouvent les citations donne l'importance de la citation et donc, l'importance de la publication dans la pratique décisionnelle.

Nous rassemblons toutes les publications de l'ADLC en format PDF. Ce format peut être décrit comme semi-structuré car il ne conserve qu'une partie de la structure du document original, qui est, le plus souvent à l'ADLC, un document Word. En pratique, on peut facilement retrouver le texte et sa police ainsi que la position des différents éléments du texte. Cependant, le document perd des

<sup>14</sup> Kevin. W. Boyack & Richard. Klavans, *Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately?*, II J. AM. SOC'Y FOR INFO. SCI. & TECH. 2389, 2391, 2394-95 (2010) (explication de l'analyse des co-citations).

informations sur l'organisation des paragraphes, les styles, la mise en page du document et les notes de bas de page.

Afin d'identifier la structure de la publication, nous nous appuyons sur une combinaison d'heuristiques spécifiques aux publications de l'ADLC et qui ont varié dans le temps. Par exemple, la plupart des titres sont dans une police plus grande que le texte principal ou en gras. En outre, les titres peuvent commencer par des nombres (ou des chiffres romains), des lettres minuscules simples ou des lettres majuscules suivies de parenthèses, par exemple b). Nous remarquons qu'aucune heuristique ne peut détecter tous les titres sans faux positifs, mais leur combinaison fournit une précision suffisante. Malheureusement, les heuristiques ne sont pas généralisables aux publications d'autres autorités de la concurrence ou même aux publications plus anciennes de l'ADLC qui suivaient un format différent (c'est-à-dire avant 2015). Ces différences peuvent être dues à une évolution des modèles de mots utilisés ou du processus de rédaction. Nous avons donc mené le processus de détection sur les décisions, identifiées par un modèle « AA-D-NN », publiées de 2015 à 2021. L'ensemble de données ainsi obtenu contient 127 décisions et 1 688 citations, une fois que les faux positifs (environ 3 % du total) ont été éliminés. La principale difficulté de détection a été rencontrée dans les décisions post-2018 : de nombreuses citations sont reléguées en note de bas de page, il a donc fallu développer une heuristique pour détecter ces cas afin d'attribuer les mentions en note de bas de page à la partie dans laquelle se trouve la référence, et non à la note de bas de page elle-même.

Une fois l'information extraite, nous obtenons la liste des citations avec la partie de la décision dans laquelle se trouve chaque citation. Nous nous limitons au titre et au sous-titre pour analyser la structure : bien que la hiérarchie du plan dépasse régulièrement 6 niveaux, les informations les plus pertinentes se situent aux deux premiers niveaux. Pour faciliter la synthèse, nous avons choisi de classer ces titres et sous-titres dans un nombre limité de catégories. Ce travail est effectué manuellement, mais une interface spécifique a été développée pour le faciliter. De plus, cette catégorisation est effectuée par une personne maîtrisant les concepts du droit de la concurrence et le raisonnement classique des décisions.

Afin d'avoir une idée d'étiquettes qui seraient à la fois exhaustives et exclusives, nous nous sommes inspirés d'un outil interne de l'ADLC (nommé Galileo) qui filtre certaines décisions de l'ADLC par titres principaux. L'étiquetage manuel a permis d'identifier 33 catégories au total - 4 catégories et 29 sous-catégories, réparties comme suit :

Catégorie/sous-catégorie	Nombre de citations	Pourcentage
<b>Discussion</b>	1,386	82 %
Pratiques	660	39 %
Sanctions pécuniaires	272	16 %
Marchés pertinents	143	8 %
Procédure devant l'AFC	97	6 %

Imputabilité	72	4 %
Champ de compétence de l'AFC	67	4 %
Procédures autres que la procédure de contestation	29	2 %
Engagements	17	1 %
Mesures conservatoires	15	1 %
Injonctions	7	0 %
Droit applicable	4	0 %
Applicabilité du droit de l'Union européenne	3	0 %
<b>Constatactions</b>	<b>175</b>	<b>10 %</b>
Secteur	83	5 %
Procédure	43	3 %
Contexte	16	1 %
Entreprises concernées	16	1 %
Pratiques dénoncées	15	1 %
Engagements	2	0 %
<b>Mesures correctives</b>	<b>76</b>	<b>5 %</b>
Montant de base	44	3 %
Sanctions pécuniaires	10	1 %
Montant final	8	0 %
Injonctions	7	0 %
Circonstances propres à l'entreprise	5	0 %
Ajustements finaux	2	0 %
<b>Mise en œuvre de la procédure d'engagements</b>	<b>51</b>	<b>3 %</b>
Évaluation préliminaire	21	1 %
Pratiques	15	1 %
Marchés pertinents	9	1 %
Engagements	4	0 %
Applicabilité du droit de l'Union européenne	2	0 %
<b>Total</b>	<b>1,688</b>	<b>100 %</b>

Tableau II : Livre de code

Le nombre et le pourcentage de citations sont basés sur leur catégorie et sous-catégorie identifiées par les parties de la décision où elles se trouvent.

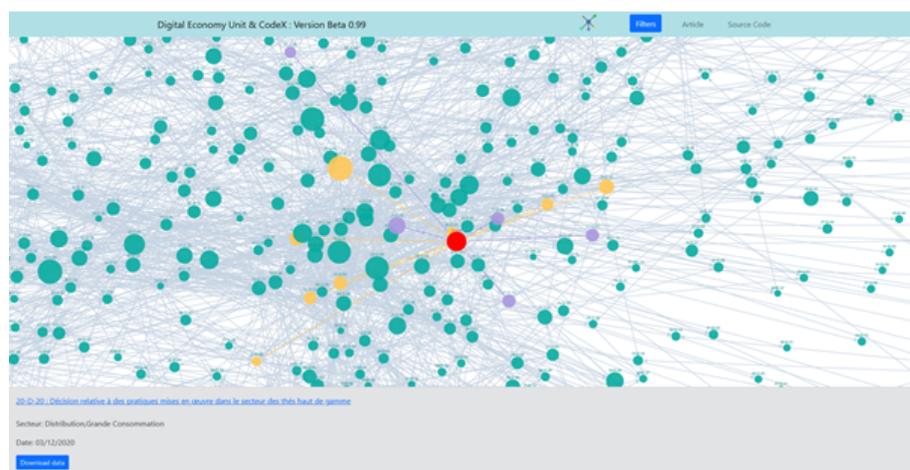
Comme nous l'avons intuitivement prédit, une écrasante majorité de citations (82 %) se trouve dans la catégorie « évaluation ». Une autre constatation immédiate est que les décisions d'engagement ne suivent pas la même structure que les autres. Nous avons donc créé une catégorie spécifique « Mise en œuvre de la procédure d'engagement ». D'autres autorités de la concurrence peuvent facilement réutiliser cette méthodologie pour identifier la structure de leurs décisions et l'harmoniser. Ces mêmes autorités pourraient également effectuer le travail que nous avons fait sur les citations, c'est-à-dire identifier les parties dans lesquelles se trouvent la plupart des citations et faire quelques déductions. Ces travaux peuvent également modifier la manière dont les autorités de la concurrence rédigent leurs décisions ou, du moins, leur faire prendre conscience de certains schémas dans le processus de rédaction des décisions.

### III. Conclusions préliminaires

Le graphe et l'ensemble de données nous permettent de tirer quelques enseignements préliminaires. La sous-partie 1 explique comment on peut utiliser l'outil graphique. La sous-partie 2 tire des enseignements statistiques globaux de l'ensemble de données utilisé pour créer l'outil et de celui créé par l'outil. La sous-partie 3 présente les raisons pour lesquelles les publications les plus citées arrivent en tête du classement.

#### A. Utilisation de l'outil graphique interactif de l'ADLC

Cet article présente le premier outil graphique interactif en réseau utilisé pour explorer les publications de l'ADLC.<sup>15</sup>



**Figure V :** Exemple d'une publication mise en avant (20-D-20 relative à des pratiques mises en œuvre dans le secteur des thés haut de gamme) et de ses interconnexions.

<sup>15</sup> Le réseau est disponible à l'adresse <https://sen-codex.dev/>, ou dans la section « Chiffres clés » du site web de l'AFC.

Cet outil est conçu pour présenter les publications importantes. Chaque nœud (cercle) du réseau représente une publication de l'AFC. Les publications sélectionnées sont surlignées en rouge. Les interconnexions sont surlignées en jaune lorsqu'elles renvoient à des publications plus anciennes et en violet lorsqu'elles renvoient à des publications plus récentes qui citent celle sélectionnée.

Sur le panneau inférieur, les utilisateurs trouvent les métadonnées relatives à la publication sélectionnée et peuvent télécharger les données d'interconnexion (publications citées/citant la publication sélectionnée). Dans le panneau supérieur gauche, les utilisateurs ont accès à différents filtres :

- Mise en évidence d'une publication spécifique par son nom (par exemple, « 20-D-20 ») ;
- Mise en évidence de toutes les publications liées à un secteur par son nom (par exemple, « Numérique ») ;
- Définition d'une période de temps sur le graphe du réseau (par exemple, « 2018-2021 » signifie que seules les citations trouvées dans les publications allant de 2018 à 2021 dans l'ensemble de données de départ seront affichées).

Avec notre algorithme de mise en page, les publications similaires (sur la base de leurs interconnexions, même indirectement connectées) sont regroupées et sont reliées par des arêtes jaunes ou violettes. Les publications populaires (fréquemment citées et ayant une audience importante sur le web) sont représentées par des cercles (nœuds) plus grands.

Dans cette optique, l'identification d'une publication importante de l'ADLC nécessite l'identification d'un grand nœud au centre d'un grand cluster. Plus précisément, les utilisateurs peuvent :

- Obtenir une vue d'ensemble des affaires de concurrence en France : en filtrant une publication ou un secteur, ils obtiendront un graphe complexe de publications connectées ;
- Créer une bibliographie pour leurs projets de recherche : partir d'une référence ou d'un secteur et utiliser l'outil pour combler les lacunes et trouver des publications étroitement liées ;
- Identifier toutes les publications clés : les utilisateurs peuvent visualiser les publications françaises importantes du domaine de la concurrence en fonction de leur impact sur les publications plus récentes (nombre de citations) et de l'importance de l'audience sur le web ;
- Découvrir les travaux antérieurs et dérivés les plus pertinents : utiliser l'outil pour trouver les publications antérieures importantes dans leur domaine d'intérêt.

Année	Nombre de publications	Nombre de citations	Citations par publication	Nombre de pages	Nombre de pages par document	Citations par page	Âge moyen des citations <sup>16</sup>
2009	99	653	6,60	2 275	22,98	0,29	4,88
2010	69	831	12,04	2 172	31,48	0,38	4,50
2011	41	450	10,98	1 369	33,39	0,33	4,68
2012	54	790	14,63	2 255	41,76	0,35	6,12
2013	46	731	15,89	1 841	40,02	0,40	5,02
2014	39	696	17,85	1 695	43,46	0,41	6,51
2015	39	571	14,64	1 795	46,03	0,32	6,50
2016	54	556	10,3	1 927	35,69	0,29	5,11
2017	38	336	8,84	1 036	27,26	0,32	6,13
2018	40	441	11,03	1 267	31,68	0,35	7,61
2019	42	695	16,55	2 141	50,98	0,32	6,87
2020	32	772	24,13	2 198	68,69	0,35	8,80
2021	42	853	20,31	2 273	54,12	0,38	8,30
Total	635	8 375		24 244			

Tableau III : Statistiques concernant les citations trouvées dans les publications de l'ensemble de données de départ

## B. Constatations générales

L'ensemble de données que nous avons utilisé pour créer le graphique, et celui qu'il donne en retour, constituent un matériel pratique pour une analyse statistique. Les citations entre les publications de l'ADLC donnent lieu à deux types d'observations en ce qui concerne la façon dont on considère les citations.

Tout d'abord, on peut examiner la distribution des citations parmi les publications. Cela donne une idée de la manière dont l'ADLC rédige ses publications et de la façon dont cette dernière a évolué au fil du temps. On peut, par exemple, analyser l'évolution du nombre de citations année par année. Même s'il n'y

<sup>16</sup> Mesuré en années depuis la publication.

a pas de tendance particulière au fil des ans, on observe que les deux dernières années ont vu une augmentation notable du nombre de citations par publication. On constate également une forte augmentation du nombre de pages par document au fil des années. Alors qu'une publication type comptait environ 30 pages en 2009-2011, sa longueur a doublé depuis. Il est intéressant de noter que le rapport entre les citations et les pages semble assez stable au cours de la période 2009-2021. On peut donc supposer que l'augmentation du nombre de citations est le résultat de l'augmentation de la longueur des publications de l'ADLC au cours des dernières années. En outre, on peut déduire l'âge des citations.<sup>17</sup> Ici aussi, on constate une augmentation au fil des années. Vers 2010, les citations avaient en moyenne 5 ans, alors qu'après 2020, elles ont plus de 8 ans.

Deuxièmement, on peut aussi regarder quels documents sont cités et à quelle fréquence pour appréhender l'« importance » des publications de l'ADLC.

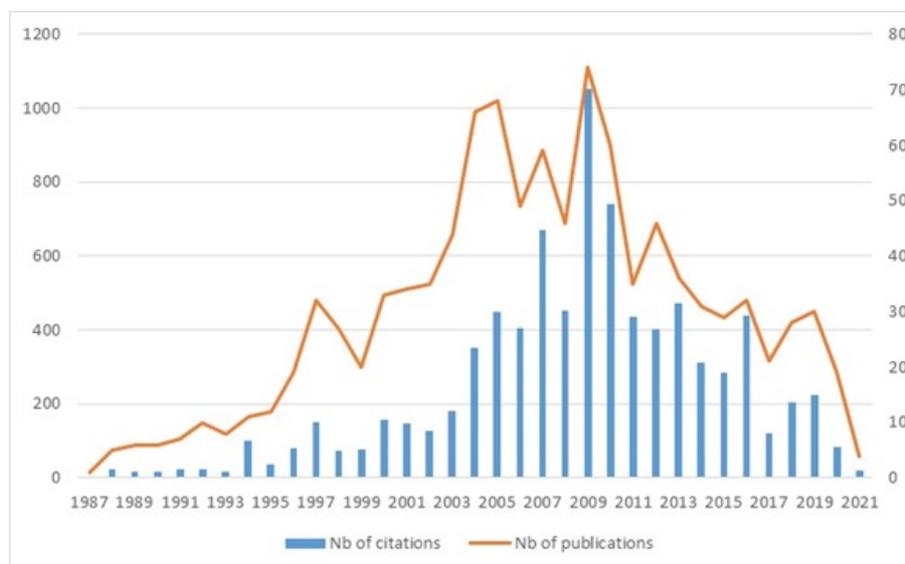


Figure VI : Citations selon l'année des publications auxquelles elles font référence

Si l'on considère le nombre de citations et de documents cités par année, on remarque la forme d'une courbe en cloche centrée sur l'année 2009. Cela peut s'expliquer ainsi : pour la partie à droite, les publications les plus récentes ont moins de documents qui peuvent les citer, pour la partie à gauche, les publications plus anciennes perdent de leur pertinence en vieillissant.

Pour essayer de saisir l'importance relative de chaque année, on peut examiner le rapport entre le nombre de citations et le nombre de documents. Plus ce ratio est élevé, plus l'impact est important. En moyenne, ce ratio est égal à 8,03, ce qui signifie que chaque publication identifiée par notre méthode est citée 8 fois en moyenne. 2009 est la meilleure année avec un ratio de 14,23 et juste derrière se trouvent 2016 avec 13,75 et 2013 avec 13,14. Cela pourrait suggérer que ces années ont revêtu une importance particulière pour l'ADLC.

Les données confirment également l'hypothèse selon laquelle, en moyenne, plus une publication est ancienne, moins elle est citée. Entre 1987 et 2006, le rapport

<sup>17</sup> Par exemple, si une publication de 2015 en cite une autre de 2010, l'âge de la citation est de 5 ans.

entre les citations et le nombre de publications ne s'élève qu'à 5. Il faut cependant garder à l'esprit une mise en garde concernant les résultats ci-dessus : la présente analyse s'applique aux publications qui ont été signalées par notre programme, mais il existe également des publications de l'ADLC qui ne sont (hélas) jamais citées.

Pour les années 2009 à 2021, nous pouvons comparer l'ensemble des citations avec l'ensemble des données de départ et voir quel pourcentage des publications de l'AFC est ensuite cité par d'autres publications.

Année du document	Nombre de documents	Nombre de documents cités	Ratio de documents cités
2009	99	74	75 %
2010	69	60	87 %
2011	41	35	85 %
2012	54	46	85 %
2013	46	36	78 %
2014	39	31	79 %
2015	39	29	74 %
2016	54	32	59 %
2017	38	21	55 %
2018	40	28	70 %
2019	42	30	71 %
2020	32	19	59 %
2021	42	4	10 %

Tableau IV : Pourcentages des citations des publications des années 2009 à 2021

En moyenne, 70 % des publications de l'AFC sont citées par la suite. Sans surprise, seules 10 % des publications de 2021 sont citées, mais le pourcentage atteint déjà 71 % pour les publications de 2019. Cela pourrait suggérer que les affaires sur lesquelles l'ADLC enquête concernent généralement les mêmes secteurs ou pratiques et que l'autorité peut donc s'appuyer sur des affaires récentes. On remarque également que 2016 et 2017 sont en dessous de la moyenne par rapport à la norme de l'ADLC, puisque moins de 60 % des publications sont citées par la suite. 2016 semble être l'année du « tout ou rien », soit les publications sont souvent citées (13,75 en moyenne pour celles qui sont citées, comme indiqué précédemment), soit elles ne sont pas citées du tout.

Il faut garder à l'esprit que la mesure de l'importance présentée ci-dessus a ses défauts. Tout d'abord, l'importance d'une publication n'est mesurée que par le fait que d'autres publications de l'ADLC la citent par la suite, de sorte que certaines publications tout aussi importantes mais autrement distinguables peuvent être manquées. Deuxièmement, le facteur temps a une influence majeure : plus la décision est ancienne, plus elle a de chances d'être citée, de sorte que l'outil peut passer à côté des décisions récentes. Réfléchir à un moyen de mieux évaluer le critère d'importance pourrait améliorer sensiblement l'analyse. Néanmoins, comme l'indique la partie 2.3, la méthode utilisée dans notre étude peut fournir des informations précieuses.

Jusqu'à présent, nous avons présenté des idées qui peuvent être déduites en analysant uniquement les arêtes du réseau de l'ADLC. Maintenant, si nous combinons le réseau avec d'autres données (ici, les métadonnées disponibles sur le site web de l'ADLC), des résultats intéressants apparaissent rapidement.

Premièrement, nous constatons que 83 % des publications impliquant des sanctions sont citées, contre seulement 67 % des publications n'impliquant pas de sanctions. Cela confirme ainsi une hypothèse générale selon laquelle les décisions imposant des sanctions ont plus d'impact. Deuxièmement, on peut filtrer les publications ayant un impact par secteur. Sur son site web, l'ADLC attribue un ou plusieurs (ou dans de rares cas aucun) secteur(s) à ses publications. Au moment de la rédaction de ce document, nous avons collecté les métadonnées de notre site web pour toutes nos publications dans notre ensemble de données de départ. Ensuite, nous avons recherché la publication la plus citée par secteur, soit sur l'ensemble de la période couverte, soit année par année, soit sur une période spécifique. En conséquence, nous présentons les publications les plus citées par secteur, pour trois périodes de temps.

	2009-2012	2013-2016	2017-2021
Vente au détail	09-D-36 (34)	13-D-12 (21)	17-D-20 (9)
Énergie / Environnement	09-D-14 (22)	13-D-20 (13)	17-D-26 (4)
Numérique	10-MC-01 / 10-A-13 (14)	14-D-06 (8)	20-D-04 / 19-D-26 / 18-A-03 (7)
Santé	09-D-17 (16)	14-D-06 / 13-A-24 (8)	19-D-01 / 17-A-10 (4)
Services	10-D-13 (27)	16-D-20 (11)	19-MC-01 / 19-D-09 (4)
Télécoms	09-D-24 (18)	16-D-11 (11)	19-MC-01 (4)
Transports	10-D-39 (30)	13-D-16 (13)	19-D-05 / 17-D-08 (4)

Tableau V : Publications les plus citées par secteur et par période. Le nombre entre parenthèses correspond au nombre de citations.

### C. Informations spécifiques

Notre approche ne fournit pas seulement un aperçu global, mais nous permet également de tirer des enseignements spécifiques de publications particulières et de dévoiler des publications importantes. En ce qui concerne le critère d'importance, nous avons choisi d'effectuer une analyse manuelle sur les dix publications les plus citées de notre ensemble de données (ci-après : « le top 10 »).<sup>18</sup>

Décision	Nom	Nombre de fois citée	Nombre de publications qui la citent
07-D-33	Décision n° 07-D-33 du 15 octobre 2007 relative à des pratiques mises en œuvre par la société France Télécom dans le secteur de l'accès à Internet à haut débit	112	45
09-D-36	Décision n° 09-D-36 du 9 décembre 2009 relative à des pratiques mises en œuvre par Orange Caraïbe et France Télécom sur différents marchés de services de communications électroniques dans les départements de la Martinique, de la Guadeloupe et de la Guyane	88	34
09-D-05	Décision n° 09-D-05 du 2 février 2009 relative à des pratiques mises en œuvre dans le secteur du travail temporaire	63	25
07-D-48	Décision n° 07-D-48 du 18 décembre 2007 relative à des pratiques mises en œuvre dans le secteur du déménagement national et international	54	29
10-D-32	Décision n° 10-D-32 du 16 novembre 2010 relative à des pratiques mises en œuvre dans le secteur de la télévision payante	46	25

<sup>18</sup> Par exemple, une publication A peut être citée 10 fois par la publication X et une publication B peut être citée deux fois par la publication Y et 3 fois par la publication Z. Dans cette situation, A est citée par 1 publication distincte et B est citée par 2 publications distinctes et nous avons estimé que B est plus importante que A.

11-D-02	Décision n° 11-D-02 du 26 janvier 2011 relative à des pratiques mises en œuvre dans le secteur de la restauration des monuments historiques	46	35
07-D-41	Décision n° 07-D-41 du 28 novembre 2007 relative à des pratiques s'opposant à la liberté des prix des services proposés aux établissements de santé à l'occasion d'appels d'offres en matière d'examens anatomo-cytopathologiques	39	29
05-D-27	Décision n° 05-D-27 du 15 juin 2005 relative à des pratiques relevées dans le secteur du thon blanc	38	37
09-D-10	Décision n° 09-D-10 du 27 février 2009 relative à des pratiques mises en œuvre dans le secteur du transport maritime entre la Corse et le continent	35	29
10-D-13	Décision n° 10-D-13 du 15 avril 2010 relative à des pratiques mises en œuvre dans le secteur de la manutention pour le transport de conteneurs au port du Havre	33	27
10-D-39	Décision n° 10-D-39 du 22 décembre 2010 relative à des pratiques mises en œuvre dans le secteur de la signalisation routière verticale	32	30

Tableau VI : Les 10 publications les plus citées

Tout d'abord, il convient de noter que toutes ces publications sont des décisions. Elles vont de 2005 à 2011. Malgré l'âge des publications, le top 10 est toujours d'actualité : chacune de ces publications a été citée dans des publications parues en 2020 ou 2021. Nous observons également que leur longueur est hétérogène, allant de 11 à 163 pages, avec une moyenne de 77 pages. Les secteurs sont également diversifiés (construction, services, santé, etc.). Enfin, ces publications sont citées par 31 publications distinctes en moyenne (allant de 25 à 45). En tenant compte de toutes les citations, elles sont citées 53 fois en moyenne, dans une fourchette allant de 32 à 112. Afin de déterminer pourquoi ces publications sont les plus citées, nous avons procédé à une analyse manuelle paragraphe par paragraphe de chaque publication. Nous avons extrait le numéro et le contenu de chaque paragraphe cité et identifié à quel concept ces citations font référence. Au total, nous avons étudié 586 citations.

La décision la plus facile à expliquer est la décision 10-D-39.<sup>19</sup> Cette publication est très majoritairement citée car elle énumère les critères pris en compte pour l'octroi de la procédure de « non-contestation des griefs ». <sup>20</sup> Ces critères, fournis au paragraphe 425 et aux paragraphes 226 à 228, expliquent 56 % des 32 citations. Une autre décision intéressante est la décision 07-D-41, citée 39 fois.<sup>21</sup> Les paragraphes III et 81 représentent respectivement 30 % et 15 % de ces citations. Par une analyse qualitative du contenu de chaque paragraphe, nous avons constaté que 71 % des citations font référence aux conditions dans lesquelles un organisme professionnel est soumis aux règles de concurrence. Nous avons affiné l'analyse et constaté qu'une division en deux autres sous-raisons était possible :

Publication citée	Raisons de la citation de la publication				Paragraphe les plus cités	
	Raison principale de la citation de la publication	Poids	Sous-raison de la citation de la publication	Poids	Numéro de paragraphe	Poids
07-D-41	Conditions dans lesquelles un organisme professionnel est soumis aux règles de concurrence	71 %	Tout acte émanant d'un organisme professionnel qui restreint la concurrence par son objet ou son effet peut constituer une pratique anticoncurrentielle	34 %	III	30 %
			Les organismes professionnels chargés d'une mission de service public ne sont pas soumis au droit de la concurrence	20 %	81	15 %

Tableau VII : Poids des raisons pour lesquelles la décision 07-D-41 est citée par paragraphe

<sup>19</sup> Autorité de la concurrence, Décision n° 10-D-39 du 22 décembre 2010 relative à des pratiques mises en œuvre dans le secteur de la signalisation routière verticale, <https://www.autoritedelaconcurrence.fr/sites/default/files/commitments/10d39.pdf>.

<sup>20</sup> Il s'agit d'une ancienne procédure qui a été remplacée par la procédure de transaction. Plus de détails disponibles sur le site web de l'AFC dans la section « La transaction : une avancée pour les entreprises », <https://www.autoritedelaconcurrence.fr/fr/competence-contentieuse>.

<sup>21</sup> Conseil de la concurrence, Décision n° 07-D-41 du 28 novembre 2007 relative à des pratiques s'opposant à la liberté des prix des services proposés aux établissements de santé à l'occasion d'appels d'offres en matière d'examen anatomo-cyto-pathologiques, <https://www.autoritedelaconcurrence.fr/sites/default/files/commitments/07d41.pdf>.

Le troisième exemple montre que certaines autres publications du top 10 sont néanmoins plus difficiles à analyser. Plusieurs paragraphes de la décision 09-D-36 ont été cités pour différentes raisons.<sup>22</sup>

Publication citée	Raisons de la citation de la publication				Paragraphes les plus cités	
	Raison principale de la citation de la publication <sup>23</sup>	Poids	Sous-raison de la citation de la publication	Poids	Numéro de paragraphe	Poids
09-D-36	Pratiques anticoncurrentielles	33 %	Critères pour déterminer le caractère anticoncurrentiel d'une pratique de discrimination par les prix	19 %	S.O. <sup>24</sup>	10 %
			Critères pour déterminer le caractère anticoncurrentiel des clauses d'exclusivité	8 %	233	8 %
			Critères pour déterminer le caractère anticoncurrentiel des politiques de fidélité et d'engagement	5 %	446	7 %
	Critères de détermination des sanctions	27 %	Gravité des pratiques	17 %	408	6 %
			Réitération des pratiques	9 %	449	6 %
	Définition des marchés pertinents	15 %	Marché de produits spécifique (téléphonie mobile dans la zone des Antilles et de la Guyane)	14 %	336	5 %

<sup>22</sup> Autorité de la concurrence, Décision 09-D-36 du 09 décembre 2009 relative à des pratiques mises en œuvre par Orange Caraïbe et France Télécom sur différents marchés de services de communications électroniques dans les départements de la Martinique, de la Guadeloupe et de la Guyane, <https://www.autoritedelaconcurrence.fr/sites/default/files/commitments//09d36.pdf>.

<sup>23</sup> Ou, lorsqu'il est impossible de trouver une raison majeure, la partie de la décision à laquelle appartient le paragraphe cité.

<sup>24</sup> « S.O. » signifie que la citation de la décision n'est pas liée à un paragraphe spécifique.

			187	5 %
			189	5 %
			469	5 %
			448	3 %
			188	3 %

Tableau VII : Poids des raisons pour lesquelles la décision 09-D-36 est citée dans les publications de l'AFC par paragraphe

D'autres exemples montrent que les dix premières décisions sont souvent citées pour des raisons différentes :

Publication citée	Raison principale de la citation de la publication	Poids	Paragraphe les plus cités	Poids
07-D-33 <sup>25</sup>	Critères pour qualifier une pratique de dénigrement	29 %	77	17 %
	Définition et objectif de la sanction de réitération	19 %	112	15 %
09-D-05 <sup>26</sup>	Proportion de la réduction de la sanction pécuniaire en fonction du type d'engagements pris dans le cadre d'une procédure de non-contestation des griefs	28 %	155	35 %
	Critères d'évaluation de la mise en œuvre d'un programme de conformité	26 %	156, 157, 159, 160	20 %

Tableau IX : Extrait du tableau synthétique présentant les résultats de l'analyse de deux autres publications du top 10

Par conséquent, nous pouvons tirer quelques conclusions générales de l'analyse du top 10 :

- La méthode utilisée nous permet d'expliquer la ou les principales raisons pour lesquelles une décision est citée ;
- La méthode n'est pas optimale : l'analyse manuelle des paragraphes un par un prend beaucoup de temps et implique un fort paramètre subjectif d'interprétation humaine. Cette méthode n'est donc pas généralisable à toutes les publications;

<sup>25</sup> Conseil de la concurrence, Décision n° 07-D-33 du 15 octobre 2007 relative à des pratiques mises en œuvre par la société France Télécom dans le secteur de l'accès à Internet à haut débit, <https://www.autoritedelaconcurrence.fr/sites/default/files/commitments//07d33.pdf>

<sup>26</sup> Conseil de la concurrence, Décision n° 09-D-05 du 2 février 2009 relative à des pratiques mises en œuvre dans le secteur du travail temporaire, <https://www.autoritedelaconcurrence.fr/sites/default/files/commitments//09d05.pdf>.

- En revanche, notre travail sur 10 décisions nous a permis d'identifier les publications qui peuvent être considérées comme les plus importantes dans le processus décisionnel de l'DLFC. Il convient toutefois de noter que nous avons affaire aux mêmes biais que ceux rencontrés dans notre analyse générale (section II.B. ci-dessus) concernant le critère choisi pour mesurer l'« importance » d'une publication ;
- Enfin, cette méthode représente une approche nouvelle et innovante car elle permet aux utilisateurs de réaliser une analyse inversée. On part généralement d'une publication récente et on remonte jusqu'à la publication citée. Ici, nous pouvons partir d'une « vieille » publication et voir toutes les publications plus jeunes qui la citent. Une telle méthode peut être réutilisée par les praticiens souhaitant découvrir les raisons pour lesquelles une publication est citée.

Enfin, comme notre travail manuel ne peut pas être facilement étendu ou automatisé, on pourrait envisager une autre méthode pour identifier la principale raison de citer une publication qui ne fait pas partie du top 10. Premièrement, il faudrait identifier les publications pertinentes qui citent la publication source en utilisant l'ensemble de données créé dans le livre de code (section II.D. ci-dessus), et deuxièmement, effectuer le travail manuel pour ces publications seulement. La partie dans laquelle la publication est citée constitue en effet un niveau d'information de base pour effectuer un premier tri.

#### IV. Prochaines étapes

L'analyse de réseau appliquée aux publications des autorités de la concurrence est un domaine de recherche très prometteur. Le présent document n'a fait qu'esquisser la partie émergée de l'iceberg d'un nouveau domaine du « computational antitrust ».

L'outil de visualisation que nous avons développé pour les besoins de cet article est disponible en ligne. Nous espérons de nombreuses personnes s'y intéresseront car il peut constituer une aide pour le milieu de la concurrence (rapporteurs, chercheurs, avocats, etc.) pour identifier au premier coup d'œil les interconnexions existant entre les différentes publications de l'ADLC. Nous espérons également que cet outil donnera un aperçu des avantages que le « computational antitrust » peut apporter au droit de la concurrence.

Enfin, le service de l'économie numérique de l'ADLC voit plusieurs domaines de recherche supplémentaires :

- On pourrait se plonger dans notre base de données (par exemple, pourquoi une publication n'est pas citée par la suite, quelles sont les raisons pour lesquelles certaines tendances sont observées, etc. ;)
- On pourrait étudier comment mieux identifier les publications importantes récentes ;
- L'outil de visualisation du réseau pourrait être amélioré. Cependant, nous ne connaissons pas encore les paramètres spécifiques qui nécessitent un

réglage fin. D'autres autorités de concurrence pourraient créer leurs propres graphiques ;

- L'outil de détection pourrait être amélioré pour identifier d'autres types de citations, c'est-à-dire des affaires de la Commission européenne, des jugements d'autres juridictions, etc. La détection des citations de la Commission européenne pourrait également aider à interconnecter les graphiques de différentes autorités de concurrence (du moins dans l'Union européenne) grâce aux décisions de la Commission qu'elles citent en commun ;
- L'analyse de réseau pourrait également être effectuée sur les publications relatives au contrôle des concentrations ;
- Le service de l'économie numérique prévoit de se concentrer sur le contrôle des concentrations, avec pour principal objectif d'identifier les décisions qui définissent de nouveaux marchés pertinents.

Le service de l'économie numérique apportera son soutien aux chercheurs intéressés par ces sujets.