ARTICLE

Deploying Network Analysis in Antitrust Law

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

Abstract. The Digital Economy Unit of the French Competition Agency ("FCA") has developed a tool that identifies citations of the FCA's publications in its other publications and represents the interconnections in a graph. This breaks new ground in computational antitrust research by adding another powerful tool to this field: network analysis. The paper first describes the method and data used to create the graph to derive thought-provoking insights from the graph analysis. Then, it presents a method to identify and measure impactful publications. Finally, it introduces potential future research avenues in the field of computational antitrust.

* **Yann Guthmann** is the Head of the Digital Economy Unit at the French Competition Authority.
** **Adrien Frumence** is Data Scientist and Case Handler at the French Competition Authority.
*** **Camille Hoogterp** is Assistant Case Handler at the French Competition Authority.

I. Introduction

Everyone in the field of competition law knows about the famous decision of the Court of Justice of the European Union ("CJEU") called "Chiquita bananas" (*United Brands*) in which the CJEU, for the first time, gave the definition of a dominant position.[1] This definition is still used and cited in almost every abuse of a dominant position case. This shows how some decisions stand out from the rest and can change how we practice law.

The constant renewal in the practice of law is not only due to case-law turnarounds but also to the fact that law–and especially competition law–follows suit of a fast-paced business world. In these last decades, the digital revolution reshaped our entire world. Hence, the adoption of IT technologies, in particular data science based tools, by government and public agencies is essential for keeping track of business practices and for remaining efficient.

Since its creation in September 2020, one of the main goals of the Digital Economy Unit of the French Competition Authority (hereafter "FCA") has been to develop data science tools for its enforcement of competition law. Combined with its digital expertise, the team provides case handlers with efficient tools and insights to facilitate their investigations, regardless of the nature of the case. On that premise, the Digital Economy Unit has joined forces with the Computational Antitrust project at CodeX to contribute to the field of computational antitrust.[2]

As our fellow colleagues of the Computational Antitrust project, we see antitrust cases as the raw material of computational antitrust. In this paper, we describe and provide a valuable tool, empowered by the exploitation of this material, to improve the efficiency of the work of antitrust agencies in today's fast-evolving environment.

We start from a simple observation: publications of the FCA often cite each other. Identifying and recording these citations manually can be very tedious, but fortunately, the process can be automated. In addition to providing a method of automating this process, we have created a tool to visualize the result. The tool takes the shape of a network graph in which the FCA's publications are represented and connected with each other by their citations. The tool ultimately helps to uncover underlying patterns in publications' citations.

---

[1] Case 27/76, United Brands Company and United Brands Continental BV v. Comm'n, 1978 ECLI:EU:C:1978:22, 65.
[2] Thibault Schrepel, *Computational Antitrust: An Introduction and Research Agenda,* 1 STAN. COMPUT. ANTITRUST 1 (2021).

We address three research questions:

1. Can we identify impactful publications using network analysis?
2. How are the FCA's publications connected to each other and what insights can be derived from the graph analysis?
3. How can the decision-drafting process be improved after uncovering the underlying patterns?

Part I of the paper describes the methodology used to create the network graph, Part II analyzes the graph and derives some first insights, and Part III develops the next steps that can be easily achieved in order to further scale up the work. This paper, therefore, seeks to pave the way for other computational antitrust enthusiasts who wish to contribute to a very fruitful area of research.

II. The methodology

A network is a set of items, i.e., vertices or nodes, with connections between them, i.e., edges.[3] Complex network analysis ("CAN") is a discipline that explores quantitative relationships in networks with non-trivial, irregular structures (e.g., antitrust cases). Against this background, subpart 1 describes how we collected the data, subparts 2 and 3 explain the algorithm applied to create the complex network and deploy it as a web application and what we can derive from it at first glance, and subpart 4 details how we enriched the data.

A. Identifying the citations

The first step is to define the starting dataset from which we will identify and extract the citations. We choose to use the exhaustive database comprising the 350 decisions, 276 opinions and 9 interim measures published by the FCA between 2009 and 2021 publicly available on the FCA's website.[4] These 635 publications amount to a total of 24,244 pages.

We select the 2009-2021 time period for two reasons. First, the FCA has existed in its current form since mid-2008. Hence, the year 2009 is the first complete year when the FCA operated with its current governance structure. Second, older publications are often rendered obsolete by recent developments. For the sake of simplicity, we only focus on three types of publications (decisions, interim measures, and opinions) and we leave aside all merger control publications which amounted to 2,746 publications during the same period.[5]

The next step is to identify and extract citations in the dataset. This step is facilitated by how the FCA numbers its publications, always following this pattern:

---

[3] Mark EJ Newman, *The Structure and Function of Complex Networks,* 45 SIAM REV. 167 (2003).

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

[5] Autorité de la concurrence, *Liste des décisions de contrôle des concentrations*, https://www.autoritedelaconcurrence.fr/fr/liste-de-controle-des-concentrations.

Type of publication	Pattern
Decision	YY-**D**-NN
Opinion	YY-**A**-NN
Interim measure	YY-**MC**-NN

Table I: The numbering of FCA publications

Where YY represents the last two digits of the year of publication and N represents a number which is incremented for each kind of publication. For example, a decision would be named 14-D-13 (the 13th decision of 2014), an opinion would be named 15-A-18 (the 18th opinion of 2015) and an interim measure would be named 19-MC-01 (the 1st interim measure of 2019).

More broadly, the FCA has 22 types of publications that all follow the same general pattern: 2 digits - 1 to 5 letters - 2 to 3 digits. Consequently, we developed a program that identifies these patterns and ran it through the 20,000+ pages of the dataset. In so doing, we were able to identify 14,036 citations from our 635 publications starting point.

The first finding is that 40% of the citations are references to confidential business information decisions that are quoted only once by the antitrust case publication they refer to. Against this finding, we discarded all the citations that are not a decision, an opinion or an interim measure. The citation dataset represents 8,375 extracted citations and is composed of 662 decisions, 341 opinions and 40 interim measures for a total of 1,043 publications from 1987 to 2021.

B. The complex network graph

Two of the most important network concepts are items and the connections between them. Items are known as nodes and connections as edges. Here, the construction of the complex network both programmatically (using NetworkX)[6] and interactively (using Dash cystoscape)[7] first requires making each publication a node and representing the relationships (citations) between them as edges. The relationships between entities are naturally directed (asymmetric): a 2021 publication can quote a 2014 publication, but the opposite is not possible. While nodes or edges describe the structural properties of networks, we also add non-structural properties as attributes of the network. These are mainly meta-data for each publication, the structure extraction of the citation, and the web-audience score. Finally, we select geometric positions for each node according to the layout algorithm.

---

[6] The Python library provides a collection of functions for constructing, measuring, and drawing complex networks. NetworkX, https://networkx.org/.
[7] Dash Cytoscape is a Python graph visualization component for creating easily customizable, high-performance, interactive, and web-based networks. Plotly, https://dash.plotly.com/cytoscape

To do so, we choose the algorithm of Fruchterman and Reingold to draw the complex graph by force-directed-placement.[8] Using this algorithm, all nodes repel each other, thus respecting the principle of magnets. The further apart the nodes are, the less they repel each other. Edges act as a spring between two nodes. At each pass of the algorithm (i.e., 5,000 epochs), the sum of the forces is applied to each of the nodes. These nodes are moved until a stable state is found.

Also known as spring embedders, such algorithms calculate the layout of a graph using only information contained within the structure of the graph itself, rather than relying on domain-specific knowledge (i.e., antitrust law).
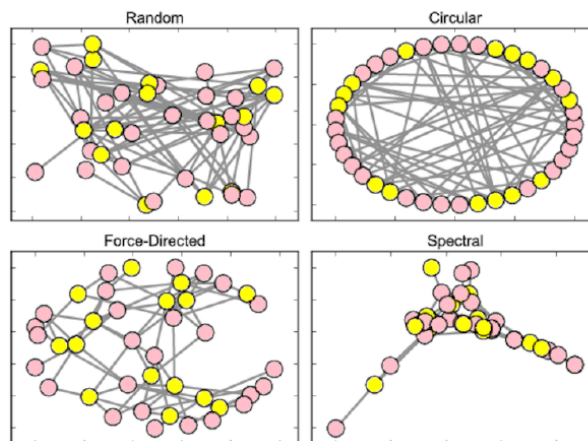

Figure I: Overview of popular layout algorithms for complex network drawing[9]

As a result, in our complex network, publications are arranged according to their similarities. That means that even publications that do not directly cite each other can be strongly connected and very closely positioned. According to this measure, two publications with highly overlapping citations and references are presumed to have a higher chance of covering a related subject matter.
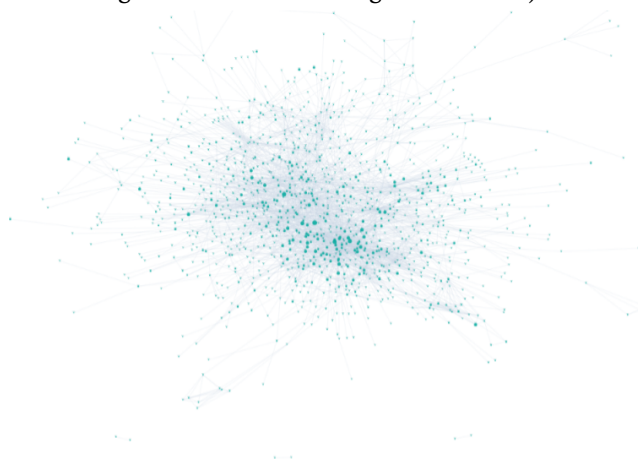

Figure II: Overview of the complex network[10]

---

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

[9] Dmitry Zinoviev, COMPLEX NETWORK ANALYSIS IN PYTHON: RECOGNIZE – CONSTRUCT – VISUALIZE – ANALYZE – INTERPRET, 58-59 (Adaobi Obi Tulton ed., 2018) (displaying the layouts for these algorithms and providing basic explanations of how they work).

[10] The network is available at https://sen-codex.dev/, or in the "Key Figures" section of the FCA's website.

C. What we see at first glance

At first glance, the graph is hard to read, given its 1,173 nodes and 4,015 edges. Because the graph is not fully connected, we partition the graph into subsets.[11] More specifically, the graph possesses eight connected components. But there is a disparity between these components: one contains 1,157 publications/points, representing 98.7% of all nodes (called the giant connected component).[12] Overall, almost all documents are connected, and just a few of them are isolated. One interesting curiosity is the fact that 16 documents are semi-isolated, meaning that they are not connected to the main graph subset but create seven very small subsets (mostly composed of two/three nodes).
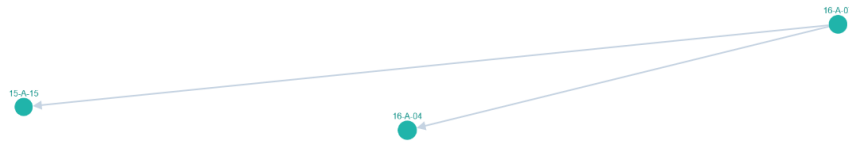

Figure III: An example of a subgraph composed of three nodes that are not connected to the main complex network

One should keep in mind that our complex network is not a citation tree. Those have been done before.[13] In our co-citation graph, publications are arranged according to their similarities.[14] That means that even publications that do not directly cite each other can be strongly connected and positioned close to each other in the graph.

---

[11] A convex network is a graph in which one can connect, directly or indirectly, any node to any other node of the graph by a chain of edges.

[12] 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) (defining various forms of giant connected components).

[13] *See* Citation Tree, https://www.citationtree.org.

[14] Kevin. W. Boyack & Richard. Klavans, *Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately?*, 11 J. AM. SOC'Y FOR INFO. SCI. & TECH. 2389, 2391, 2394-95 (2010) (explaining co-citation analysis).

Figure IV: Publications related to telecommunications are highlighted in red squares–the graph shows hidden strong connections between them

On the graph above, one can see the close positions of Decision No. 20-D-01 related to a practice implemented in the digital terrestrial television broadcasting sector and Opinion No. 12-A-13 related to the analysis of the upstream wholesale markets for terrestrial broadcasting services of audiovisual programs. Decision No. 20-D-01 does not cite Opinion No. 12-A-13, but both are still strongly connected and positioned close to each other in the complex network. The method outlined above could improve the accuracy of publication recommendations.

D. Creating the coding book

Extracting the citation dataset is a necessary first step. After searching for a methodology to improve the quality of data displayed by the graph—adding the context of each citation—we chose to use the structure of the publications, *i.e.*, the titles and subtitles. Knowing the parts of the publication in which the citations are found gives the importance of the citation and therefore, the importance of the publication in the decision-making practice.

We gather all the FCA's publications in PDF format. This format can be described as semi-structured because it retains only part of the structure of the original document–which is, most often at the FCA, a Word document. In practice, one can easily retrieve the text and its font as well as the position of the different text elements. However, the document loses information about paragraph organization, styles, document layout and footnotes.

In order to identify the structure of the publication, we rely on a combination of heuristics that are specific to the FCA's publications and have varied over time. For example, most titles are in a larger font than the main text or in bold. In addition, titles may begin with numbers (or Roman numerals), single lowercase letters, or uppercase letters followed by parentheses, e.g., b). We notice that no heuristics can detect all titles without false positives, but their combination provides sufficient accuracy. Unfortunately, the heuristics are not generalizable to publications of other competition agencies or even to older publications of the FCA that followed a different format (i.e., before 2015). These differences may be due to an evolution

of the Word models used or the drafting process. Therefore, we conducted the detection process on decisions, identified by a "YY-D-NN" pattern, issued from 2015 to 2021. The dataset obtained this way contains 127 decisions and 1,688 citations, once false positives (around 3% of the total) have been removed. The main detection difficulty was encountered in post-2018 decisions: many citations are relegated to footnotes, so it was necessary to develop a heuristic to detect these cases in order to attribute the footnoted mentions to the part in which the reference is located, and not the footnote itself.

Once the information is extracted, we obtain the list of citations with the part of the decision in which each citation is found. We limit ourselves to the title and subtitle to analyze the structure: although the hierarchy of the plan regularly exceeds 6 levels, the most relevant information is located at the first two levels. To facilitate the synthesis, we choose to categorize these titles and subtitles in a limited number of categories. This work is done manually, but a specific interface has been developed to ease it. Moreover, this categorization is carried out by a person mastering the concepts of competition law and the classical reasoning of decisions.

In order to get an idea of labels that would be both exhaustive and exclusive, we got inspired by an internal tool of the FCA (named Galileo) that filters some FCA's decisions by main titles. The hand labeling resulted in the identification of 33 categories in total–4 categories and 29 subcategories, distributed as follows:

Category/subcategory	Number of citations	Percentage
Assessment	1,386	82%
Practices	660	39%
Fines	272	16%
Relevant markets	143	8%
Procedure before the FCA	97	6%
Imputability	72	4%
Jurisdiction of the FCA	67	4%
"Non-challenge" procedure	29	2%
Commitments	17	1%
Interim measures	15	1%
Injunctions	7	0%
Applicable law	4	0%
Applicability of EU law	3	0%
Findings	175	10%
Sector	83	5%

Procedure	43	3%
Context	16	1%
Companies involved	16	1%
Reported practices	15	1%
Commitments	2	0%
Remedies	76	5%
Base amount	44	3%
Fines	10	1%
Final amount	8	0%
Injunctions	7	0%
Company-specific circumstances	5	0%
Final adjustments	2	0%
Implementation of commitments procedure	51	3%
Preliminary assessment	21	1%
Practices	15	1%
Relevant markets	9	1%
Commitments	4	0%
Applicability of EU law	2	0%
Total	1,688	100%

Table II: Coding book
The number and percentage of citations is based on their category and subcategory identified by the parts of the decision where they are found.

As we intuitively predicted, an overwhelming majority of citations (82%) are in the "assessment" category. Another immediate insight is that commitment decisions do not follow the same structure as the other ones. Therefore, we have created a specific category of "Implementation of commitments procedure." Other competition agencies can easily reuse this methodology to identify the structure of their decisions and harmonize it. These same agencies could also perform the work we did on citations, i.e., identify the parts in which most citations are located and make some deductions. This work can also change the way competition agencies draft decisions or at least render them aware of certain patterns in the decision-drafting process.

III. Preliminary takeaways

The graph and the dataset enable us to derive some preliminary insights. Subpart 1 explains how one can handle the graph tool. Subpart 2 derives overall statistical insights from the dataset used to create the tool and the one created by the tool. Subpart 3 introduces the reasons why the most cited publications top the ranking.

A. Using FCA's interactive graph tool

This article introduces the first interactive network graph tool used to explore the FCA's publications.[15]
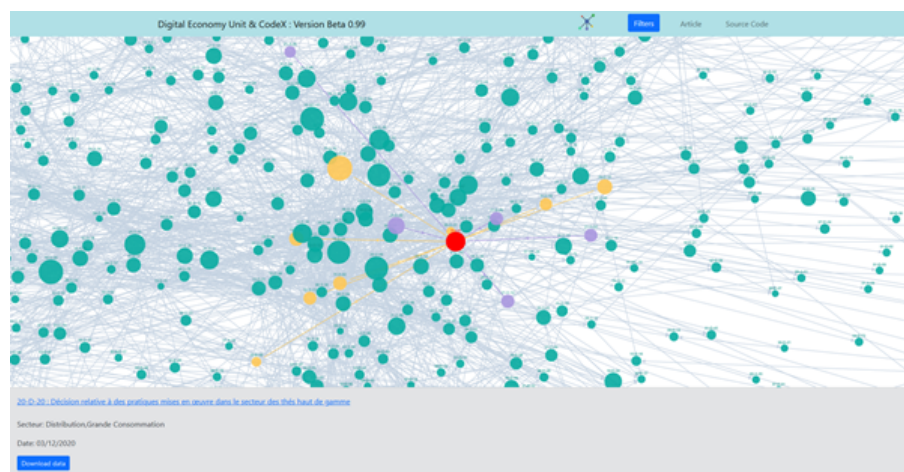

Figure V: Example of a highlighted publication (20-D-20 regarding practices implemented in the high-end tea sector) and its interconnections

The tool is designed to outline important publications. Each node (circle) in the network represents a publication of the FCA. Selected publications are highlighted in red. Interconnections are highlighted in yellow when they link to older publications and purple when they link to newer publications that cite the one selected.

On the bottom panel, users find the metadata related to the selected publication and can download the interconnections data (publications cited/citing the selected one). On the top-left panel, users have access to different filters:

- Highlighting a specific publication by its name (e.g., "20-D-20");
- Highlighting all the publications related to a sector by its name (e.g., "Digital");
- Defining a time period on the network graph (e.g., "2018-2021" means that only citations found in publications ranging from 2018 to 2021 in the starting dataset will be displayed).

---

[15] The network is available at https://sen-codex.dev/, or in the "Key Figures" section of the FCA's website.

With our layout algorithm, similar publications (based on their interconnections, even indirectly connected) are clustered together and are connected by yellow or purple edges. Popular publications (frequently cited and with an important web audience) are represented by bigger circles (nodes).

With that in mind, identifying an impactful publication of the FCA requires identifying a large node at the center of a big cluster. More specifically, users can:

- Get a visual overview of French antitrust cases: filtering a publication or a sector will build a complex network graph of connected publications;
- Create a bibliography for research projects: start with a reference or a sector and use the tool to fill in the gaps and find closely related publications;
- Identify all key publications: users can visualize important French antitrust publications based on their impact on newer publications (number of citations) and the importance of web audience;
- Discover the most relevant prior and derivative works: use the tool to find important prior publications in their field of interest.

Year	Nb of Publications	Nb of citations	Citations by publication	Nb of pages	Nb of pages by document	Citations per pages	Average age of citations[16]
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

---

[16] Measured in years since publication.

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			

Table III: Statistics regarding the citations found in the publications from the starting dataset

B. Global findings

The dataset we used to create the graph, and the one it gives in return, make a convenient material for a statistical analysis. The citations between the FCA's publications yield two types of insights in respect of the way one looks at citations.

First, one can look at the distribution of citations among the publications. Doing so gives perspective about how the FCA drafts its publications and how it changed over time. One can, for instance, analyze the evolution of the number of citations year by year. Even though there is no particular trend throughout the years, we observe that the last two years have seen a noticeable increase in the number of citations by publication. We also notice a sharp increase in the number of pages by document through the years. While a typical publication was around 30 pages long in 2009-2011, its length has since doubled. Interestingly, the ratio of citations to pages seems quite stable during 2009-2021, and so one could assume that the increase in the number of citations is the result of the increase in the length of the FCA's publications the last few years. Furthermore, one can derive the age of citations.[17] Here, we also notice an increase throughout the years. Around 2010, citations were 5 years old on average, while after 2020 they are more than 8 years old.

Second, one can also look at which documents are quoted and how often to apprehend the "importance" of FCA's publications.

---

[17] For example, if a 2015 quotes a publication from 2010, the age of the citation is 5.
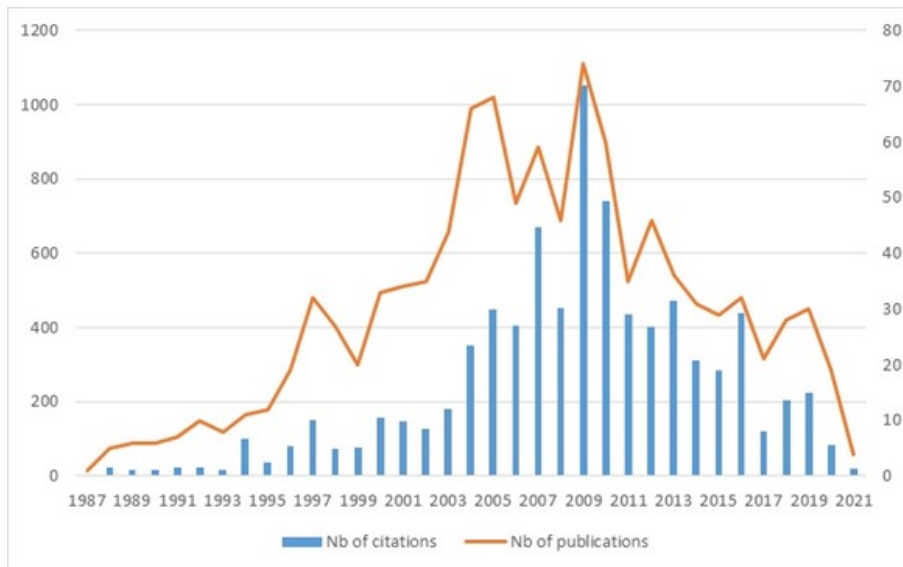
Figure VI: Citations by the year of the publications they refer to

When considering the number of citations and documents quoted by year, we notice the shape of a bell curve centered on the year 2009. This can be explained as such: for the part to the right, the most recent publications have less documents that can quote them, for the part to the left, the older publications lose relevance as they age.

In order to try to grasp the relative importance of each year, one can look at the ratio of the number of quotations against the number of documents. The higher this ratio, the higher the impact. On average, this ratio is equal to 8.03, meaning that each publication identified by our method is quoted 8 times on average. 2009 is the best year with a ratio of 14.23 and just behind are 2016 with 13.75 and 2013 with 13.14. This could suggest that these years were of particular importance for the FCA.

The data also confirms the assumption that on average, the older a publication gets, the less frequently it is quoted. Between 1987 and 2006, the ratio of quotations to the number of publications only amounts to 5. There is one caveat, though, that one should keep in mind regarding the findings above: the present analysis applies to publications that were flagged by our program, but there are also publications of the FCA that are (alas) never quoted.

For the years 2009 to 2021, we can compare the citations dataset with the starting dataset and see what percentage of the FCA's publications are later quoted by other publications.

Year of document	Nb of Documents	Nb of documents quoted	Ratio of documents quoted
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%

Table IV: Percentages of the citations of publications from years 2009 to 2021

On average, 70% of the publications of the FCA are quoted later on. Not surprisingly, only 10% of the 2021 publications are quoted, but the percentage already reaches 71% for 2019 publications. This could suggest that the cases the FCA investigates usually involve the same sectors or practices and thus the agency can rely on recent cases. We also notice that 2016 and 2017 are below average compared to the standard of the FCA, as less than 60% of the publications are quoted later on. 2016 seems to be a "hit or miss" year, either the publications are often quoted (13.75 on average for those that are quoted, as shown before), or they are not quoted at all.

One should keep in mind that the measure of importance presented above has its flaws. First, the importance of a publication is only measured by the fact that other FCA publications subsequently quote it, so some equally important but otherwise distinguishable publications may be missed. Second, the time factor has a major influence: the older the decision, the more likely it is to be quoted, so the tool can miss on recent decisions. Thinking about a way to better assess the criterion of importance could significantly improve the analysis. Nonetheless, as part 2.3 argues, the method used in our study can provide valuable insights.

To this point, we have presented insights that can be deducted only by analyzing the edges of the FCA's network. Now, if we combine the network with other data—here, the metadata available on the FCA's website, insightful results quickly emerge.

First, we find that 83% of publications that involved penalties are quoted, whereas only 67% of publications that do not involve penalties are cited. It thus confirms a general assumption that decisions imposing penalties are more impactful. Second, one can filter the impactful publications by sector. On its website, the FCA assigns one or more (or in some rare cases none) sector(s) to its publications. At the time of writing of this paper, we collected the metadata from

our website for all our publications in our starting dataset. Then, we searched the most quoted publication by sector, either on the whole period covered or year by year or over a specific period. As a result, we present the most quoted publications by sector, for three periods of time.

	2009-2012	2013-2016	2017-2021
Retail	09-D-36 (34)	13-D-12 (21)	17-D-20 (9)
Energy / Environment	09-D-14 (22)	13-D-20 (13)	17-D-26 (4)
Digital	10-MC-01 / 10-A-13 (14)	14-D-06 (8)	20-D-04 / 19-D-26 / 18-A-03 (7)
Healthcare	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)
Telecom	09-D-24 (18)	16-D-11 (11)	19-MC-01 (4)
Transport	10-D-39 (30)	13-D-16 (13)	19-D-05 / 17-D-08 (4)

Table V: Most quoted publications by sector and period. The number in brackets refers to the number of citations.

C. Specific insights

Our approach not only provides global insights but also enables us to gain specific insights from specific publications and helps unveil important publications. Regarding the criterion of importance, we chose to conduct a manual analysis on the ten most quoted publications of our dataset (hereinafter: "the top 10").[18]

Decision number	Complete French name	Number of times quoted	Number of publications quoting it
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

---

[18] For example, a publication A can be quoted 10 times by publication X and a publication B can be quoted twice by publication Y and 3 times by publication Z. In this situation, A is quoted by 1 distinct publication and B is quoted by 2 distinct publications and we deemed that B is more important than A.

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
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-cyto-pathologiques	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 |

Table VI: The top 10 most quoted publications

First, it is worth noticing that all these publications are decisions. They rang from 2005 to 2011. Despite its age, the top 10 is still topical: each one of these publications has been quoted in publications issued in 2020 or 2021. We also observe that their length is heterogeneous, ranging from 11 to 163 pages and averaging 77 pages. The sectors are also diversified (construction, services, healthcare, etc.). Finally, these publications are quoted by 31 distinct publications on average (ranging from 25 to 45). Taking all citations into account, they are quoted 53 times on average, ranging from 32 to 112. In order to outline why these publications are cited the most, we conducted a manual paragraph-by-paragraph analysis of each publication. We extracted the number and content of each quoted paragraph and identified which concept these citations refer to. In total, we studied 586 citations.

The easiest decision to explain is No. 10-D-39.[19] This publication is overwhelmingly cited as it lists the criteria considered for granting the "no contest of objections" procedure. [20] These criteria, provided in paragraph 425 and in paragraphs 226 to 228, explain 56% of the 32 citations. Another interesting decision is No. 07-D-41, cited 39 times.[21] Out of these, paragraphs 111 and 81 represent respectively 30% and 15%. By a qualitative analysis of the content of each paragraph, we found that 71% of the citations refer to the conditions under which a professional body is subject to the competition rules. We refined the analysis and found that it could be divided into two further sub-reasons:

Cited publication	Reasons for citing the publication				Most cited paragraphs	
	Major reason for citing the publication	Weight	Sub-reason for citing the publication	Weight	Paragraph number	Weight

| 07-D-41 | Conditions under which a professional body is subject to the competition rules | 71% | Any act issued by a professional body which restricts competition by object or effect may constitute an anti-competitive practice | 34% | 111 | 30% |
| | | | Professional bodies performing a public service mission are not subject to competition law | 20% | 81 | 15% |

Table VII: Weights of the reasons why Decision No. 07-D-41 is cited per paragraph

The third example shows that some other top 10 publications are nevertheless more difficult to analyze. Multiple paragraphs of Decision No. 09-D-36 have been quoted for different reasons.[22]

Cited publication	Reasons for citing the publication				Most cited paragraphs	
	Major reason for citing the publication[23]	Weight	Sub-reason for citing the publication	Weight	Paragraph number	Weight
09-D-36	Anti-competitive practices	33%	Criteria for determining the anti-competitive nature of a price discrimination practice	19%	N.A.[24]	10%

		Criteria for determining the anti-competitive nature of exclusivity clauses	8%	233	8%
		Criteria for determining the anti-competitive nature of loyalty and commitment policies	5%	446	7%
	Criteria for determining sanctions	27%	Severity of practices	17%	408	6%
			Reiteration of practices	9%	449	6%
	Definition of relevant markets	15%	Specific product market (mobile telephony in the French Caribbean and Guiana area)	14%	336	5%
					187	5%
					189	5%
					469	5%
					448	3%
					188	3%

Table VII: Weights of the reasons why Decision No. 09-D-36 is cited in FCA's publications per paragraph

Other examples show that top 10 decisions are often cited for different reasons:

Cited publication	Major reason for citing the publication	Weight	Most cited paragraphs	Weight
	Criteria for qualifying a disparagement practice	29%	77	17%

07-D-33[25]	Definition and objective of the reiteration sanction	19%	112	15%
09-D-05[26]	Proportion of the reduction of the fine according to the type of commitments made in the context of a non-challenge procedure	28%	155	35%
	Criteria for assessing the implementation of a compliance program	26%	156, 157, 159, 160	20%

Table IX: Extract of the synthetic table presenting the results of the analysis of two other top 10 publications

As a result, we can draw some general conclusions from the top 10 analysis:

- The method used allows us to explain the main reason(s) why a decision is quoted;
- The method is not optimal: the manual analysis of the paragraphs one by one is very time consuming and involves a strong subjective parameter of human interpretation. This method is therefore not generalizable to all the publications of the graph;
- By contrast, our work on 10 decisions has allowed us to identify the publications that can be considered as the most important in the decision-making process of the FCA. It is however worth noting that we are dealing with the same biases faced in our general analysis (section II.B. above) regarding the criterion chosen to measure the "importance" of a publication;
- Finally, this method represents a new and innovative approach because it enables users to carry out a reversed analysis. One usually starts from a recent publication and goes back to the publication cited. Here, we can start from an "old" publication and view all the younger publications that quote it. Such a method can be reused by practitioners wishing to discover the reasons why a publication is quoted.

Finally, since our manual work cannot be easily scaled or automated, one could consider another method to identify the main reason for citing a publication that is not part of the top 10. First, one would have to identify the relevant publications that quote the source publication by using the dataset created in the Coding Book (section II.D. above), and second, perform the manual work for these publications only. The part in which the publication is cited indeed constitutes a baseline level of information to perform a first sorting.

---

[25] 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
[26] 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.

IV. Next steps

Network analysis applied to the publications of competition authorities is a very promising area of research. The present paper has only outlined the tip of the iceberg of a new area of computational antitrust.

The visualization tool we developed for the sake of this article is available online. We hope many will get to grips with it as it can constitute a helping hand for the antitrust community (case handlers, scholars, lawyers, etc.) to identify at a first glance the interconnections existing between different publications of the FCA. We further hope that this tool will outline the benefits that computational antitrust can bring to competition law.

Last, the FCA's Digital Economy Unit sees several areas of further research:

- One could dive into our dataset (e.g., why a publication is not cited later on, what are the reasons why some trends are observed, etc.);
- One could explore how to better identify recent important publications;
- The network visualization tool could be improved. However, we are not yet aware of the specific parameters that require fine-tuning. Other competition agencies could create their own graphs;
- The detection tool could be improved to identify other kinds of citations, i.e., cases from the European Commission, judgments from other jurisdictions, etc. The detection of European Commission citations could also help interconnecting graphs from different competition agencies (at least in the European Union) thanks to the Commission decisions they cite in common;
- The network analysis could be also done on merger control publications;
- The Digital Economy Unit plans to focus on merger control, with the main objective of identifying the decisions that define new relevant markets.

The Digital Economy Unit will provide its support to researchers interested in these topics.