HAL
open science

Appropriation of Social Sciences and Humanities Literature in the public arena

Lucie Loubère, Fidelia Ibekwe

To cite this version:

Lucie Loubère, Fidelia Ibekwe. Appropriation of Social Sciences and Humanities Literature in the public arena. International Conference on Conceptions of Library and Information Science, Jun 2019, Ljubljana, Slovenia. hal-02499038

HAL Id: hal-02499038

https://hal.science/hal-02499038v1

Submitted on 4 Mar 2020

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.

Appropriation of Social Sciences and Humanities Literature in the public arena

Loubère Lucie
Aix Marseille Univ, Université de Toulon, IMSIC, Marseille, France
lucie.loubere@univ-amu.fr

Ibekwe Fidelia
Aix Marseille Univ, Université de Toulon, IMSIC, Marseille, France.
fidelia.ibekwe-sanjuan@univ-amu.fr

Introduction. *We propose a methodology and tools for exploring the citation contexts of scholarly publications on the web.*

Method*. From the logs to Open Edition pages, a platform that hosts scholarly publications in the Social Sciences and Humanities, and the digital traces left on the web by the visitors, we identified web pages that cited publications from the Open Edition platform and delineated the citation contexts using the Texttiling thematic segmentation library. With Unitex, a Natural Language Processing engine, we identified noun phrases from the citation contexts of the citing web pages and built a co-occurrence graph linking the cited Open Edition publications to the phrases found in their citation contexts on the web.*

Analysis*. We generated a dynamic and interactive graph interface allowing the user to navigate between the cited publications from Open Edition and their contexts of citation.*

Results*. The exploration of the graphs revealed the citation themes, their provenance (citers) and other features such as the purpose of citation, actors and events surrounding the cited publications and the polarity of the citation (positive or negative).*

Conclusion*. This study showed that there was a strong lexical coherence between the cited publications from Open Edition and their appropriation in the public arena.*

1. Introduction

Studying the impact of scholarly publications has long been the goal of the metrics field (biblio/scientometrics) (Filliatreau, 2008; Hicks, Wouters, Waltman, Rijcke and Rafols, 2015; Robinson-Garcia, Sugimoto, Murray, Yegros-Yegros, Larivière and Costas, 2019). However, these fields focus solely on the impact of scientific productions within the academic sphere using mathematical and statistical models that yield quantitative indicators. In terms of the units of analysis, scientometrics and bibliometrics studies tend to focus almost exclusively on (co)-citation counts of authors, journals or keywords. While this enabled the comparison of apparently 'neutral indicators' (citation counts, ratios, etc) or the evolution of citations of scholarly publications, they yield little or no information on the reasons for their citations. Algorithms and visualisations offered by commercial vendors for exploring scholarly content are based on a quantitative analysis of the logs on scholarly publication platforms, hence their output continues the same trend as bibliometrics indicators.

The availability of scholarly publications in online repositories and in open access has increased the chances of dissemination and thus their re-use by the general public. The diversity of the audience liable to access scholarly content has led to the diversification of publication media outside of the usual scholarly publication channels such as journal articles, conference papers or books. We have witnessed a proliferation of means of disseminating scholarly publications via academic blogs, scientific magazines destined to a wider audience. More and more scientific publications and discoveries are now being announced via social media (Twitter in particular) before being taken up by the mainstream media. This has rendered more porous the frontier between and the scientific community and the society. The proliferation of more venues for disseminating scholarly content and their reuse in the public

arena via newspaper articles, citations on websites and on social media (Twitter, Facebook) have led to the emergence of more "social" and less scholarly based metrics. Grouped under the name of 'altmetrics' (O'Neill, 2016), these more recent indicators of scholarly impact are a testimony to the current porosity between the academic world and the society at large. However, "altmetrics" has continued in the same tradition as the older biblio/scientometrics by basing its indicators on numerical trace, i.e., computing the number of likes, posts, downloads, tweets or retweets a scholarly publication gets on the web with the result that neither of these fields provide information on the actual use of the scholarly publications cited nor the reasons for which they were cited.

Our study aims to bridge this gap by proposing a combined approach, integrating symbolic and numerical methods for the exploration of the impact of scholarly publications in the public arena. Our study is cast in the framework of the *Open Knowledge Appropriation* project (OKA) funded by the Initiative of Excellence of Aix-Marseille University in France. The central research questions posed by this project is to determine how and to what extent scholarly publications hosted on the Open Edition platforms (https://www.openedition.org/?lang=en) are appropriated by the general public, i.e. what traces of these publications are found on the web, in what form and for what purposes? Open Edition comprises four platforms hosting books, journals, blogs, scientific events and news in the Social Sciences and Humanities (SSH). Some of its content is open access while others require a subscription.

Our specific research question is to study the different uses people make of scholarly contents outside of *bona fide* scholarly communication channels such as journals, articles, conferences, books, and what we can learn about how scholarly literature circulates and is re-appropriated in the public arena. Our focus is therefore not on counting citations received by a given scholarly publication nor on counting the amount of logs or traffic to those publications, rather we are interested in what the citing documents on web pages (the citers) actually say about the cited documents. For this, we need to consider their citation contexts. Taking into account the context in citation studies has always been a bottleneck issue because it requires knowledge intensive approaches amongst which Natural Language Processing (NLP) is an obvious choice.
Our study there departs from traditional biblio/scientometrics in that we design a methodology and tools to study the uses of scholarly literature in a non-academic setting based on the citation contexts. Our research therefore deals with scholarly communication and its link with scholarly literacy in the society.

2. Methodology

Our methodology combines symbolic and numerical approaches. The symbolic approach draws on Natural Language Processing (NLP) methods and tools as well as on discourse analysis while the numerical approach deploys a top-down hierarchical clustering algorithm coupled with graph generation and information visualisation methods and tools. The sections below describe in more details the steps needed to go from corpus collection to graph exploration showing the citation contexts.

2.1 Corpus building

The starting point of our study are logs of web pages to any of the four platforms hosted by Open Edition, namely:

- Open Edition Books (https://books.openedition.org/)

- Open Edition Journals (https://www.openedition.org/catalogue-journals)
- Open Edition research blogs (https://www.openedition.org/catalogue-notebooks)
- Calenda: announcements of scientific events (https://calenda.org/)

Open Edition conserves the log data specific to its visited documents, the users, and the provenance of the latter. Figure 1 below is an extract of these logs. It shows the identifiers of the session, the visitor's ID, the time of connection, its duration, the URL from where the visit came (the referrer) and the page consulted on Open Edition.

idvisit	idvisitor	visit_last_action_time	visit_first_action_time	referer_url	name
177909	1.51571811112E+19	2017-01-01 00:00:14	2017-01-01 00:00:00	https://www.google.fr/	rha.revues.org/6885
177909	1.51571811112E+19	2017-01-01 00:00:14	2017-01-01 00:00:00	https://www.google.fr/	rha.revues.org/7255
177914	1.48617546873E+19	2017-01-01 00:00:25	2017-01-01 00:00:15	https://www.google.it/	romatevere.hypotheses.org/297

Figure 1: Examples of logs to pages hosted on Open Edition.org.

We collected logs for the period between 01/01/2017 to 01/01/2018.

2.2 Identification of visiting web pages

Using the log collection tool (Matomo, https://matomo.org/), we eliminated from our corpus logs from robots performing automated visits, logs from search engines because they only provide the queries entered by a visitor to get to Open Edition pages but not the precise context of citation of an Open Edition document. We also excluded password protected sites such as social media (Facebook, Snapchat, ...), intranets of academic or corporate entities and sites that do not allow access to the specific pages containing the link to an Open Edition publication. Since the application of the New European Data Protection Regulation, some servers no longer specify the complete URL of the referrer page. Thus, visits coming from Wikipedia only provide the generic link 'http://wikipedia.fr' regardless of the actual Wikipedia page from which a visit was made. Hence, we could not exploit citation contexts coming from Wikipedia. After removal of these domain names, we extracted the textual content of the remaining visiting web pages and identified the position of the link to Open Edition. This gave rise to 33, 667 citing web pages for the period considered.
The corpus we collected is heterogeneous on several levels. The first level of heterogeneity lies in the diversity of themes present on Open Edition platforms. The platforms cover the whole spectrum of social sciences humanities. Publications on neuropsychology sit alongside those on literature, philosophy, history, sociology or law. The second level of heterogeneity concerns the diversity of re-appropriation media (the citing document types). Amongst web pages citing Open Edition publications, are discussion forums, blogs of researchers, government sites, associations, the mainstream media and the layman/woman. The disciplinary diversity mentioned above is linked to the thematic diversities of each journal / blog. A sociology journal can deal with a multitude of subjects which themselves can be of interest to several fields. This wealth of content, far from being a limitation, constitutes in itself a first level of exploration of our corpus. In the sections below, we will describe in more details the approaches and methods we deployed to analyse and explore the citation contexts of Open Edition publications on the web.

2.3 Delineating citation contexts

The first step is to delineate the citation contexts. To this end, we extracted all the texts on the 33, 667 citing pages. However, we are not concerned with the whole article on every citing page. In order to reduce the window to the immediate citation context to an Open Edition document, we turned to thematic segmentation approaches and after some literature reviews,

we chose the Textilling library (Hearst, 1997). This tool segments a text (discourse) into thematic segments based on their lexical cohesion. Since we were only interested in capturing just the citation context surrounding the link to an Open Edition page, we only needed to identify and extract this local context. For the pages where the citation to an Open Edition document contains little or no surrounding text (isolated links), we performed a manual selection of what could constitute a citation context. To analyse the topical content of the citation contexts, we identified the noun phrases contained therein using Unitex, an open access NLP package developed by Sébatien Paumier at the University of Marne-la-Vallée (https://unitexgramlab.org/fr). Unitex is based on Max Silberztein 1993's NLP engine called Intex® (http://www.nyu.edu/pages/linguistics/intex/). Unitex performs several NLP tasks such as morphological analysis and tagging, identification of textual concordances or regular expressions and corpus alignment (in different languages). Its graphic interface enables the user to draw the sequence of morpho-lexical categories s/he wishes to extract from a corpus as finite state automata which are then launched on the corpus. The results are highlighted in the corpus. Figure 2 is an example of a very simple automaton that identifies all text sequences in which a noun is followed directly by an adjective.

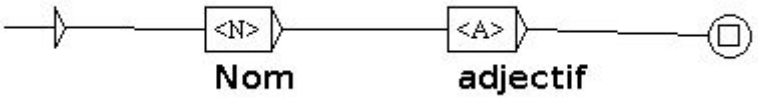

Figure 2: A simple finite state automaton that identifies sequences of nouns followed by adjectives in a text.

We wrote several finite state automata to identify phrases comprising at least 2 sequences of one or more noun(s), article(s), adverb(s) and adjective(s) which are searched for recursively. The automaton Figure 3 below illustrates this.

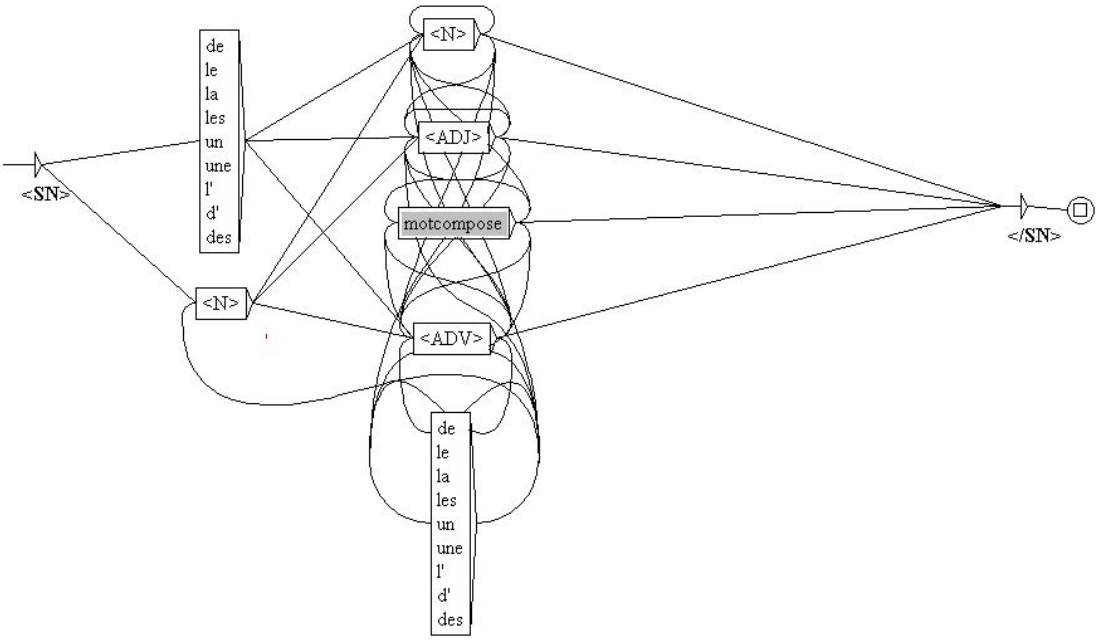

Figure 3: A finite state automaton that identifies noun phrases contained in the citation contexts to Open Edition documents.

This finite state automaton identified noun phrases that give an indication of what the citing pages said about the cited documents on Open Edition. Figure 4 is an example of noun

phrases thus identified which served as input for the next phases of clustering, graph generation and visualisation.

```
ent (des livres, <SN>des thèses</SN>, des articles de recherche universi
res, des thèses, <SN>des articles de recherche universitaire</SN>).{S} E
nt illustrées en <SN>adoptant un point de vue</SN> décalé, et surtout di
 Proche-Orient : <SN>la destruction d</SN>'un patrimoine mondial _ [Parl
la destruction d'<SN>un patrimoine mondial</SN> _ [Parlons peu parlons S
nce {S}Accueil > <SN>La Science</SN> en images > Science ça tourne ! > A
ence en images > <SN>Science ça tourne</SN> ! > Avides de Recherche {S}T
ourne ! > Avides <SN>de Recherche</SN> {S}Terrorisme et guerre au Proche
 Proche-Orient : <SN>la destruction d</SN>'un patrimoine mondial mercred
la destruction d'<SN>un patrimoine mondial mercredi</SN> 9 novembre 2016
 9 novembre 2016 <SN>par Rédaction</SN> de Parlons peu parlons Science {
lusieurs années, <SN>le terrorisme</SN> et la guerre traversent plusieur
le terrorisme et <SN>la guerre</SN> traversent plusieurs régions du mond
```

Figure 4: Examples of noun phrases identified from the citation contexts.

2.4 Building graphs of co-cited contexts and of cited documents

The next step consisted in building a co-occurrence graph of noun phrases occurring in the citation contexts to Open Edition publications. An edge is drawn between two noun phrases if they co-occurred in the same citation contexts. These graphs of similarity thus relate nodes (the cited Open Edition page and the noun phrases found in the citation contexts) by edges (the number of co-occurrence of the same phrases in the same contexts citations). The first time two noun phrases are found in the same citation context, the value of the edge is set to 1. As more citation contexts are explored, this value is incremented if the same two noun phrases are found in the same citation contexts.
Table 1 hereafter shows some of the noun phrases (in blue) identified in the citation context of an Open Edition publication.

referrer url	http://www.certop.cnrs.fr/La-parole-aux-chercheurs-Vincent
cited Open Edition url	sms.hypotheses.org/6419
Delineated citation context	L'histoire et la sociologie des sciences et des techniques ont marqué depuis quelques années un intérêt de plus en plus net pour les « équipements » et leur rôle dans la conduite et l'organisation de la recherche. On peut définir ceux-ci comme des infrastructures matérielles (par exemple un avion destiné à étudier l'atmosphère et les dispositifs de mesure associés) et immatérielles (les logiciels permettant de piloter ces dispositifs ou d'analyser les données qu'ils recueillent), et comme un ensemble d'instruments et de personnels mis à disposition des chercheurs pour réaliser des expériences. Lire la suite sur Mondes Sociaux (LabEx sms)

Table 1: Example of an Open Edition blog cited on an external webpage with its citation context.

The graph generation algorithm used here is the Leuven method (Blondel, Guillaume, Lambiotte, and Lefebvre, 2008) available in the Gephi package (Bastian, Heymann, and Jacomy, 2009). It proceeds in two phases: first, the algorithm creates a cluster for every node, then each node is absorbed into the cluster closest to it. If this structure allows a gain of modularity, it is kept, if not, it is cancelled and the process is reiterated until it converges or it

reaches optimal modularity, i.e., no change improves it. The second phase reproduces the same pattern by considering the communities built previously as distinct nodes, the edges connecting the nodes are calculated as the sum of the weights of the links between the individuals constituting the two communities. These groupings therefore depend on the cited page but also on the co-presence of two segments (noun phrases) in the same citation context.

2.5 The interactive graph navigation interface

The graph thus generated becomes a navigable interface only in its dynamic export. We used the sigma.js (http://sigmajs.org/) library developed by Alexis Jacomy for the Oxford Internet Institute which allows the generation of an interactive graph on websites. It also allows us to enrich this visualisation with additional elements such as links to the citation contexts and the citing web pages. For the sake of legibility, we assigned colours to the clusters formed. The final navigation tool is divided into three parts: a search component, the main graph and an information component (see Figure 5 below).

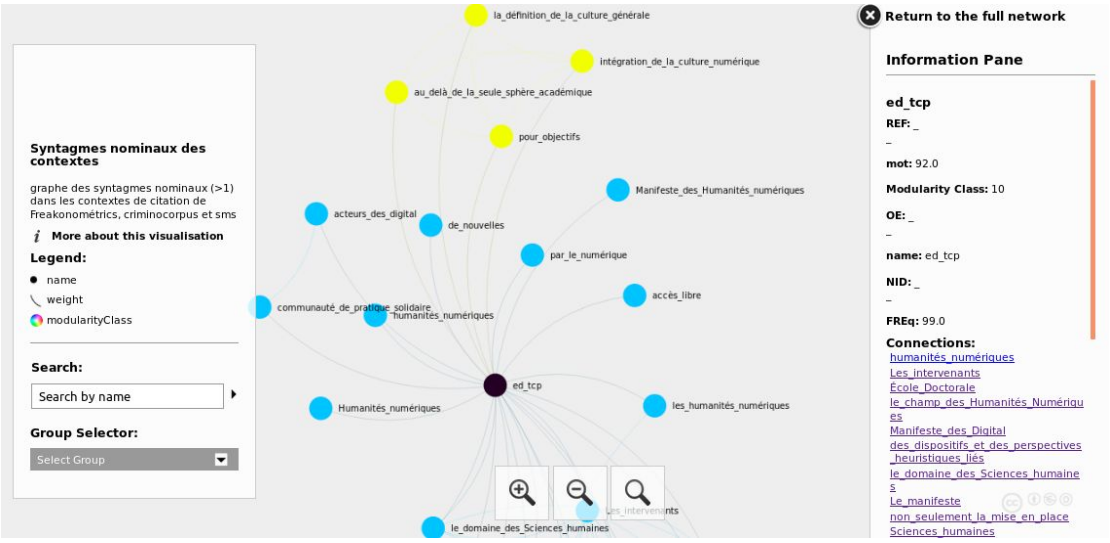

Figure 5: The navigation interface showing cited documents on Open Edition and their contexts of citation.

The left pane enables the user to search for all or parts of a noun phrase or to select a cluster by its colour. The central pane contains the interactive graph which can be navigated, nodes' labels are shown when the mouse hovers over them. A click on a node zooms the graph to this node and its neighbours and displays contextual information on the right pane with the links to the noun phrases in the citation contexts as well as direct connections with other phrases and the cited document.

3. Results

We will illustrate the results of our corpus processing stages described above by exploring the citation contexts of three blogs hosted on Open Edition:

- ThatCamp Paris (ed_tcp)
- Monde sociaux (ed_sms)
- Criminocorpus (ed_criminocorpus)

The graph presented in 5 and accessible here makes it possible to visualise the phrases present in the contexts of citations of these blogs. In the results below and for reasons of legibility, we added a root "ed_" to the names of these blogs. Figure 6 hereafter shows the immediate

citation context of these three blogs with their names highlighted as prominent nodes and around them, the noun phrases extracted from their citation contexts on the citing web pages.

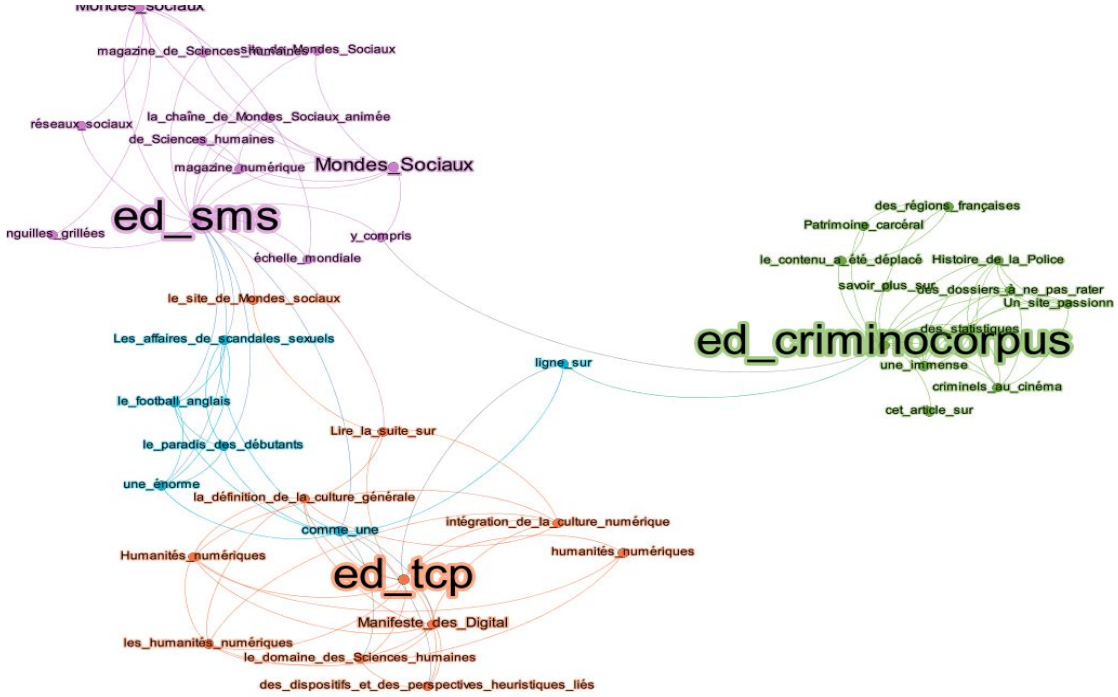

Figure 6: Overall graph showing the citation contexts of three blogs: '*ThatCamp Paris (ed_tcp), Monde Sociaux (ed_sms), Criminocorpus.*'

3.1 Citation Contexts of the blog 'ThatCamp Paris'

This blog was launched in 2009 for the preparation of the first non-conference on Digital Humanities. The cited article is a manifesto aimed at fostering events that promoted informal discussions and collaborations between stakeholders of Digital Humanities. This blog offered a launching pad for gathering ideas, participations and proposals of interested persons to organise various events such as workshops, datasprint etc. Although the last article of the blog was dated 2015, visits to this blog continues and we were able to identify 38 citing web pages between January 2017 to January 2018. The phrases extracted in citation contexts to this blog are coloured in the orange zone in Figure 6. A further splitting of this cluster brought several themes to the fore (see Figure 7 here below).

As shown by its nice shape, this very homogeneous graph recalls the foundation of Digital Humanities which is strongly embedded in the social sciences and humanities. We also see the presence of a node labelled "digital humanities manifesto", a reference to the collective document from the first ThatCamp. Recurrent phrases in the citation contexts mostly echoed the subject of the blog they were citing, namely:

- *the field of digital humanities,*
- *digital humanities*
- *humanities*
- *the field of humanities*
- *Manifesto of the digital humanities*

Aside from these noun phrases which revealed the topical content of this blog, a more detailed graph also revealed actors active on this topic, be they partners of the *ThatCamp* movement, actors of digital humanities, doctoral schools, universities or scholarly associations. All the actors appear to be united by their citations and reliance on articles published on this blog. The exploration of this wider audience allowed us to identify the announcement for a day's workshop organised by a doctoral school in a French University (http://www.univ-paris3.fr/ed268-2017-2018-conference-du-samedi-a-quoi-bon-les-humanites-numeriques-pour-les-sciences-du-langage--466360.kjsp?RH=1263513068245).

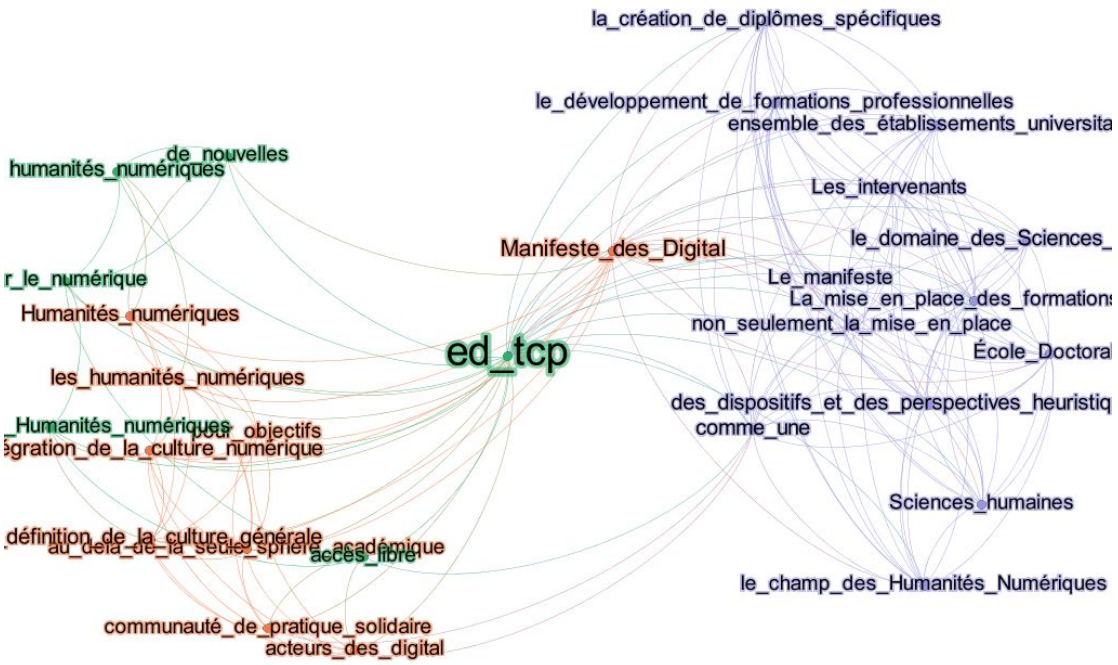

Figure 7: Sub-graph showing the citation contexts of the '*ThatCamp Paris'* blog.

3.2 The citation context of the Criminocorpus blog

The second publication whose citation contexts we studied is managed by CLAMOR (UMR 3726), a CNRS (*Centre National de Recherche Scientifique*) research laboratory as part of an online Francophone scientific publication platform on the history of justice, crimes and punishment. This publication title comprises a museum, a journal and a blog and it is the latter whose external citation contexts we analysed. With a strong publishing activity (several articles per day), it is also the blog that received the most citations on the web (94 citation contexts). The noun phrases extracted from these citation contexts are coloured in green in Figure 6. A more detailed view of this cluster (Figure 8) revealed several topics.
In the citation contexts, we again find a significant proportion of noun phrases echoing the cited blog's content (*criminal phenomenon, criminals in the cinema, statistics'*, *the museum, the first digitally native museum)* but also a recall of the publication platform's other activities with noun phrases such as '*thematic exhibitions, permanent exhibitions, cultural events*'. Phrases such as '*ordinary violence, the criminal controversy, birth of the scientific police*' reflected book reviews that were published by the *Criminocorpus* journal. Further exploration of the citation contexts to this blog uncovered the presence of a rhetorical marker such as "*this article*" in the extracted phrases which is characteristic of scholarly practice in introducing a publication. Other noun phrases conveyed an evaluative dimension such as '*objects constituting relatively rare sources, an exciting site, dossiers not to be missed*', thus showing the adhesion and positive recommendations of the blog's citers.

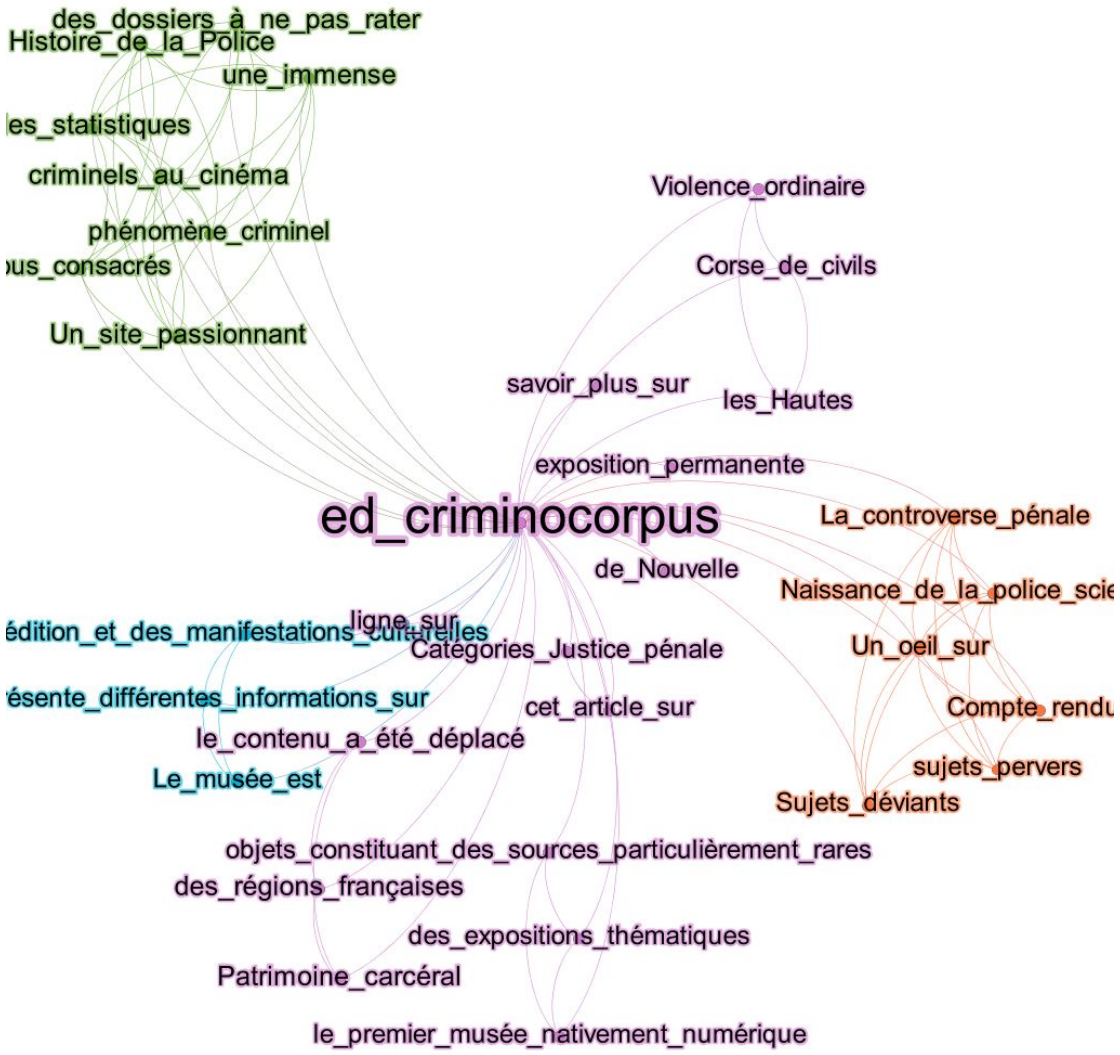

Figure 8: Sub-graph showing the citation contexts of the '*Criminocorpus*' blog.

3.3 Citation contexts of the Mondes Sociaux blog

The third publication whose citation contexts we analysed is hosted by a French Excellence research laboratory called "Labex SMS" for '*Structuration des Mondes Sociaux*' (Structuring Social Worlds). This publication platform is managed by the University of Toulouse and the blog is described as a French-language open access digital magazine promoting the sharing and the circulation of knowledge to a wider audience by writing short articles on previously published content. As a bi-monthly magazine, it yielded 78 exploitable citation contexts found in the purple cluster (*ed_sms*) in the main graph (Figure 6). Further splitting of this cluster revealed a very densely populated structure (see Figure 9 hereafter) that brought to the fore more specific topics.

The orange zone in Figure 9 (bottom left) gathered phrases such as '*editor of the social world magazine, digital magazine*' that reflected the editorial role of the blog. Other phrases like '*is a digital magazine, humanities, of public, social media*' recalled the purpose of the blog which is to facilitate the circulation of scholarly knowledge to a wider audience. This sub-graph also portrayed the co-existence of scholarly or academic articles alongside articles on scientific mediation and the humanities destined for the general public. The green cluster

(upper right) contains phrases such as '*of humanities sciences, social science magazine, access of humanities'* that position this blog within the humanities as well as phrases reflecting the trust awarded to the blog ('*managed by researchers, recent research work, researcher recognition*). The other sub-clusters echoed the different themes of this blog. Yet other phrases also revealed referencing practices linked to bibliographies.

Figure 9. Sub-graph of showing the citation contexts of the '*Structuration des Mondes Sociaux*' blog.

All in all, our exploration of the citation contexts of these three blogs on web pages shows some regularities in the aspects of the information contained in the cited documents on Open Edition which are re-used on the web. The technical characteristics of the cited documents are expressed in a descriptive way by phrases evoking the numerical formats, the modality of access (open access, free) and the editorial committees. Citing pages also stressed the scientific identity of the cited documents and thus the trust awarded to scholarly documents by members of the public. Another constant was the echoing of the cited documents themes and topics by reusing some of the same phrases found on the cited publications on Open Edition. Finally, in these three examples, readers and citers of the blogs all gave them positive evaluations and recommendations owing mainly to the feeling of trust engendered by their scholarly/scientific nature (the public's awe and respect for science).

Concluding remarks

By focusing on noun phrases as content bearing elements in linguistic utterances, we were able to approach and then highlight the strong lexical proximity between cited scholarly publications on Open Edition platforms and their contexts of re-use and citation in the public space (web pages).

The work presented here is still in progress. Our study has shown how scientific literature from a largely open access hosting platform is re-appropriated and repurposed for various uses in the public arena. It has also shown that it is possible to expand citation studies in a qualitative direction by focusing not only on counting their numbers but on their real impact by studying the contexts of their reuse. We also showed how symbolic (Natural Language Processing) and numerical approaches (clustering, graph generation) can be fruitfully combined to perform a knowledge intensive Human Language Technology task. While the interactivity of the information visualisation interface we built is an asset for exploring complex undirected graphs, it can also constitute a source of bias. A possible bias may lie in the variations in the number of citation contexts for each cited document. In our examples, the three publications (blogs) we used as illustration which yielded 28, 94 and 79 citation contexts respectively. The graph generation process, by its iterative nature, can penalise publications with few citation contexts. A second bias linked to a variation in the lexical density of the citation contexts may occur. For instance, the *Structuration des Mondes Sociaux* blog had fewer citing pages although its citation contexts returned more noun phrases because of a higher lexical density of these contexts. On the other hand, the *Criminocorpus* blog had more citing web pages but whose citation contexts yielded fewer noun phrases. This was because the web pages citing the *Criminocorpus* blog repeated the same sentences from one site to another and also because of the presence of many discussion forums where the immediate citation context contained few sentences. These biases remind us that corpus processing based on quantitative aspects, if they enable us to synthesise and navigate the information in a user-friendly manner, is not sufficient to enable a fine-grained exploration and interpretation of the complexity of the phenomenon under study. In the future, we intend to investigate the relevance of discourse analysis and in particular of rhetorical markers to better understand the purposes for which scholarly publications are being appropriated in the public space.

Acknowledgements
The project leading to this publication received funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a French “Investissements d’Avenir” program.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm*, *8*(2009), 361-362.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Filliatreau, G. (2008). Bibliométrie et évaluation en sciences humaines et sociales : une brève introduction. *Revue dhistoire moderne contemporaine*, *n° 55-4bis*(5), 61-66.

Hearst, M. A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, *23*(1), 33-64.

Hicks, D., Wouters, P., Waltman, L., Rijcke, S. de, & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature Publishing Group*, *520*(7548), 429-431.

Marchand, P., & Ratinaud, P. (2012). L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT*, *2012*, 687-699.

O'Neill, J. (2016). NISO recommended practice: Outputs of the alternative assessment metrics project. *Collaborative Librarianship*, *8*(3), 4.

Robinson-Garcia, N., Sugimoto, C. R., Murray, D., Yegros-Yegros, A., Larivière, V., & Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Informetrics*, *13*(1), 50-63.

Silberztein, M. (1993). *Dictionnaires électroniques et analyse automatique de textes: le système INTEX. Masson.*