

# *Modéliser la lecture*

*Perspectives d'analyse quantitative de  
la réception scientifique à la suite de  
l'étude Usage Alpha.*

Cette recherche a reçu le soutien du Programme d'investissement ISTEX  
(ANR-10-IDEX-0004-02)

*Workshop Appropriation du  
savoir ouvert  
6-8 novembre 2019.*



# Plan

Introduction méthodologique ?

1. Les lecteurs inattendus
2. Les lecteurs attendus
3. Les disciplines d'Open Edition
4. Perspectives.

[https://analytics.huma-num.fr/OpenEditionLab/umberto\\_oe/](https://analytics.huma-num.fr/OpenEditionLab/umberto_oe/)



# Modélisation et humanités numériques

# Contexte scientifique



## Transactional log analysis: Online catalogs

J. Tolle. *Proceedings of the 6th annual international ACM SIGIF retrieval*, page 147-160. ACM, (1983)

### Links and resources

**BibTeX key:** [tolle1983transactional](#)

**search on:**

## Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction

Maristella Agosti · Franco Crivellari ·  
Giorgio Maria Di Nunzio

Received: 8 January 2010 / Accepted: 21 June 2011 / Published online: 14 July 2011  
© The Author(s) 2011

**Abstract** In the last decade, the importance of analyzing information management systems logs has grown, because log data constitute a relevant aspect in evaluating the quality of such systems. A review of 10 years of research on log analysis is presented in this paper. About 50 papers and posters from five major conferences and about 30 related journal papers have been selected to trace the history of the state-of-the-art in this field. The paper presents an overview of two main themes: Web search engine log analysis and Digital Library System log analysis. The problem of the analysis of different sources of log data and the distribution of data are investigated.

L'analyse quantitative des logs est un champ de recherche ancien (1ère étude de l'OCLC dès... 1983) qui ne prend vraiment de l'ampleur que depuis très peu de temps

# Contexte scientifique



**PIWIK**  
Open Analytics Platform

- Learn more
- Community
- Help
- Marketplace
- Developers
- Hosting
- About
- Blog

## Liberating Analytics

Piwik is the leading open-source analytics platform that gives you more than just powerful analytics:

- Free open-source software
- 100% data ownership
- User privacy protection
- User-centric insights
- Customisable and extensible

De nouveaux systèmes d'analyse des logs permettent d'obtenir des données fiables (sans les robots...)

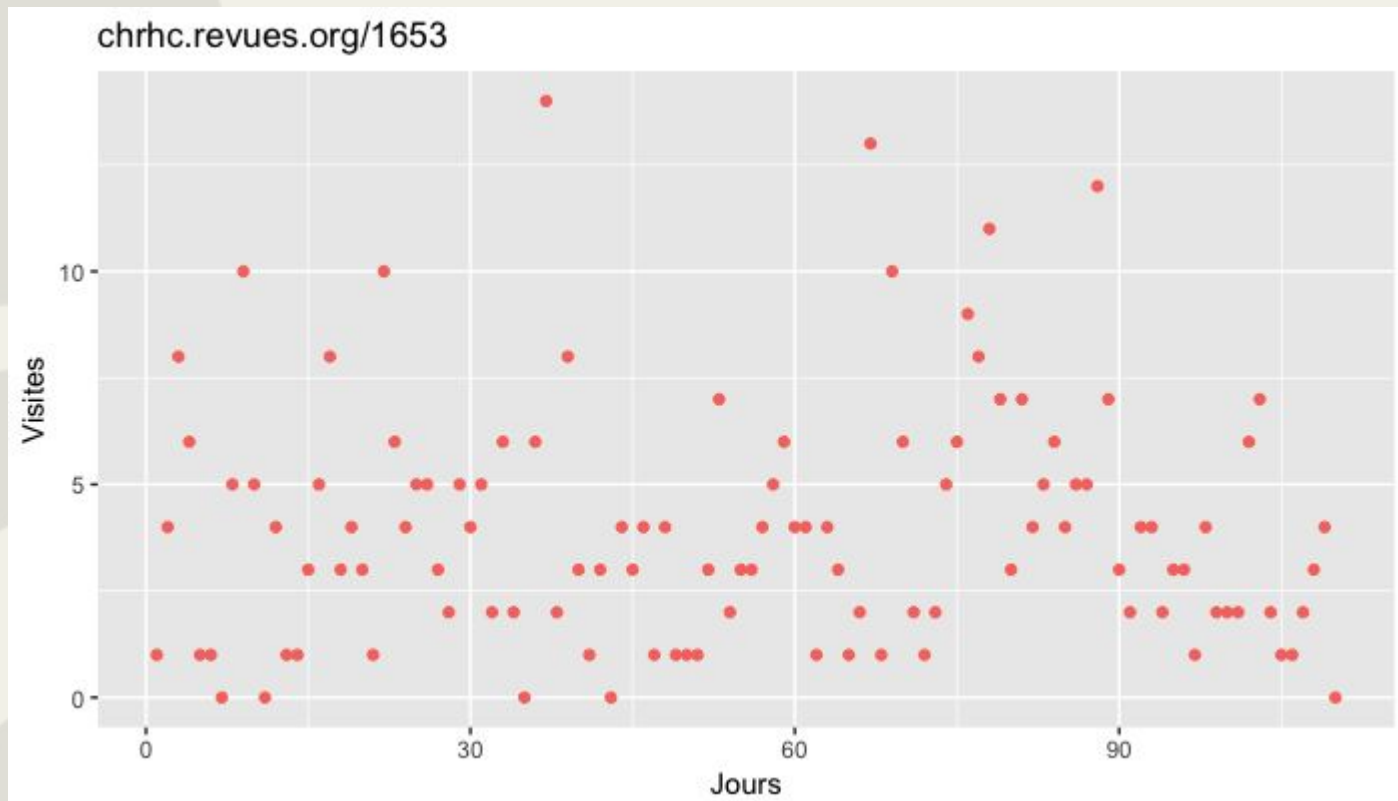
# Corpus et méthodologies

## Un projet de *Little Big Data*

- **Exploration de l'ensemble des logs** de janvier 2017 à juillet 2019, soit **160 millions de sessions**. C'est un élargissement significatif du projet initial qui prévoyait des échantillons sur une durée de quelques jours.
- **Mise en visibilité de nouveaux objets** : l'utilisation des logs complet rend possible le repérage du lecteur inattendu et la réalisation de segmentations fines.
- **Contraintes logistiques** : sans être considérable, le volume des données étudiées **met à l'épreuve les outils d'analyse usuels** et elles sont en évolution constante. Les temps de traitement sont substantiellement allongés.
- **Données complexes à modéliser** : des modèles inhabituels sont nécessaires pour rendre compte de la diversité des phénomènes.

# Modélisation et humanités numériques

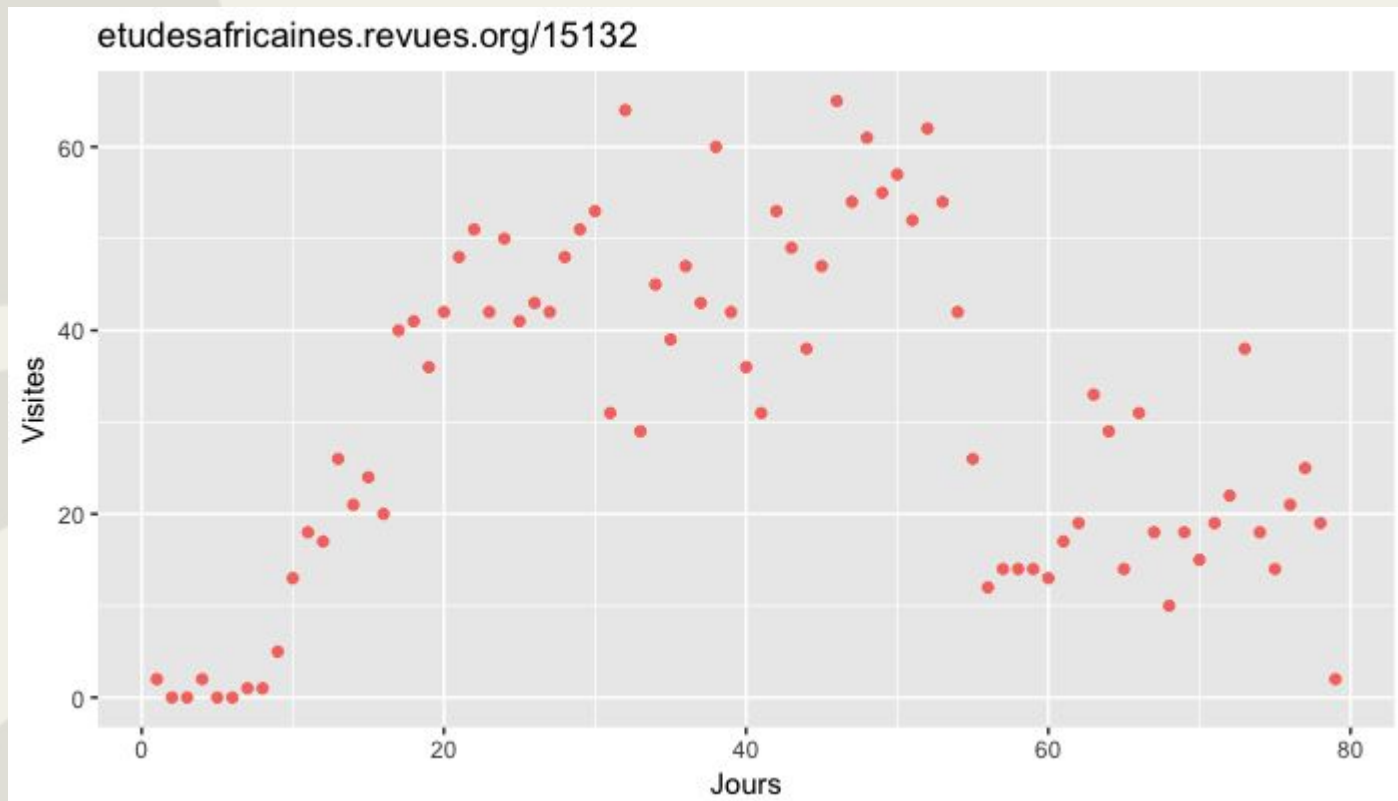
## Le paradoxe du siège d'avion



Chaque article a sa propre vie...

# *Modélisation et humanités numériques*

## Le paradoxe du siège d'avion

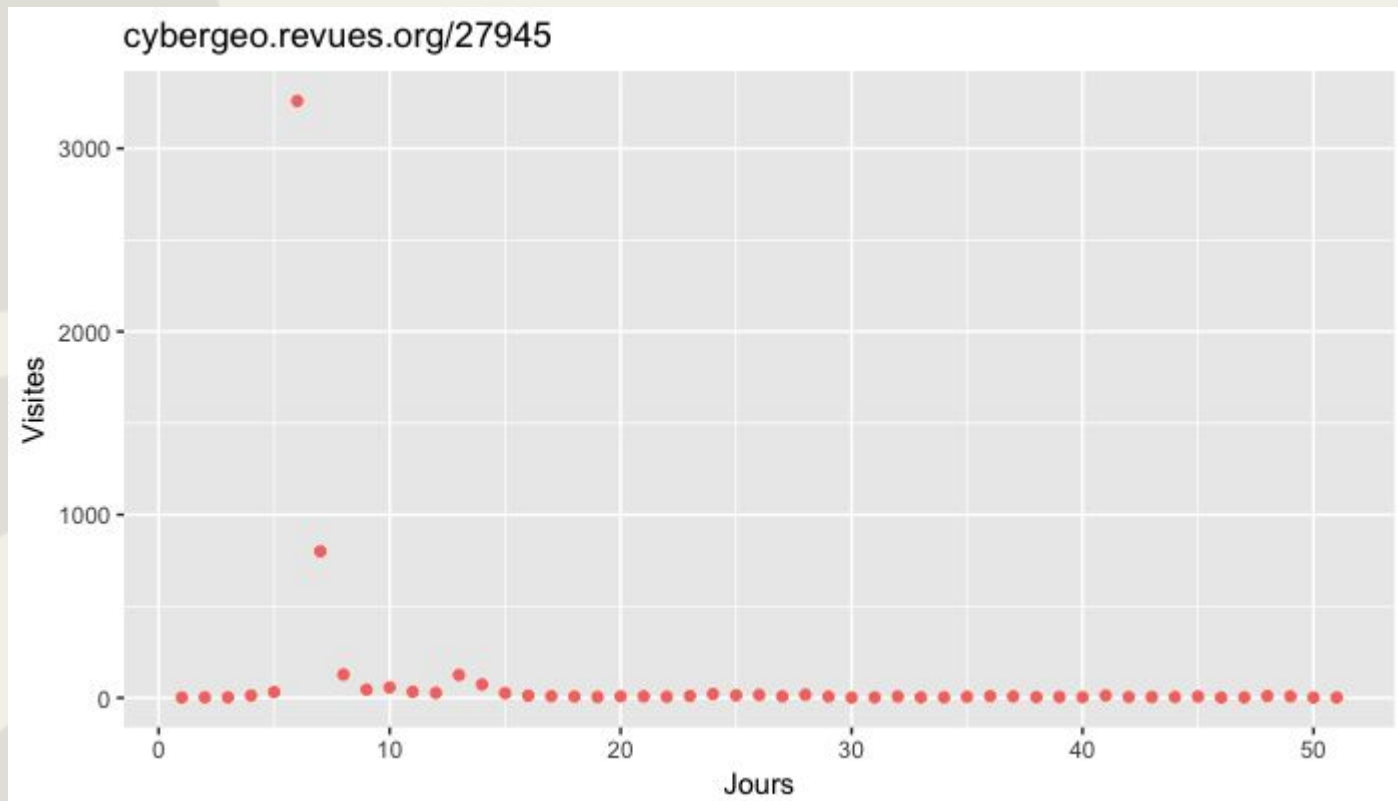


Chaque article a sa propre vie...



# Modélisation et humanités numériques

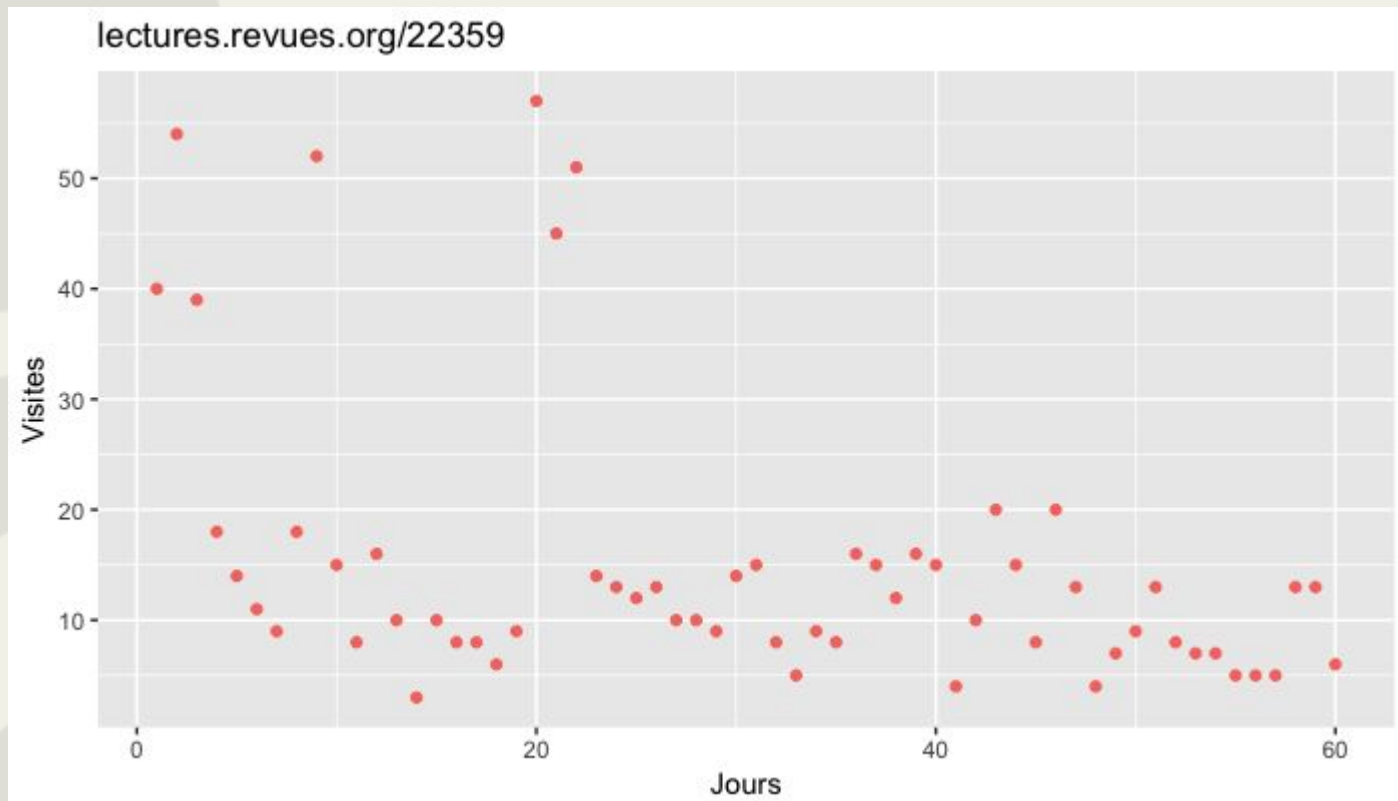
## Le paradoxe du siège d'avion



Chaque article a sa propre vie...

# *Modélisation et humanités numériques*

## Le paradoxe du siège d'avion

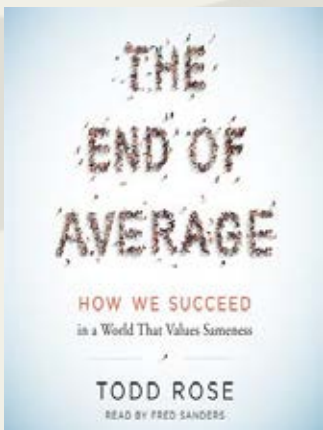


Chaque article a sa propre vie...

# *Modélisation et humanités numériques*

## Le paradoxe du siège d'avion

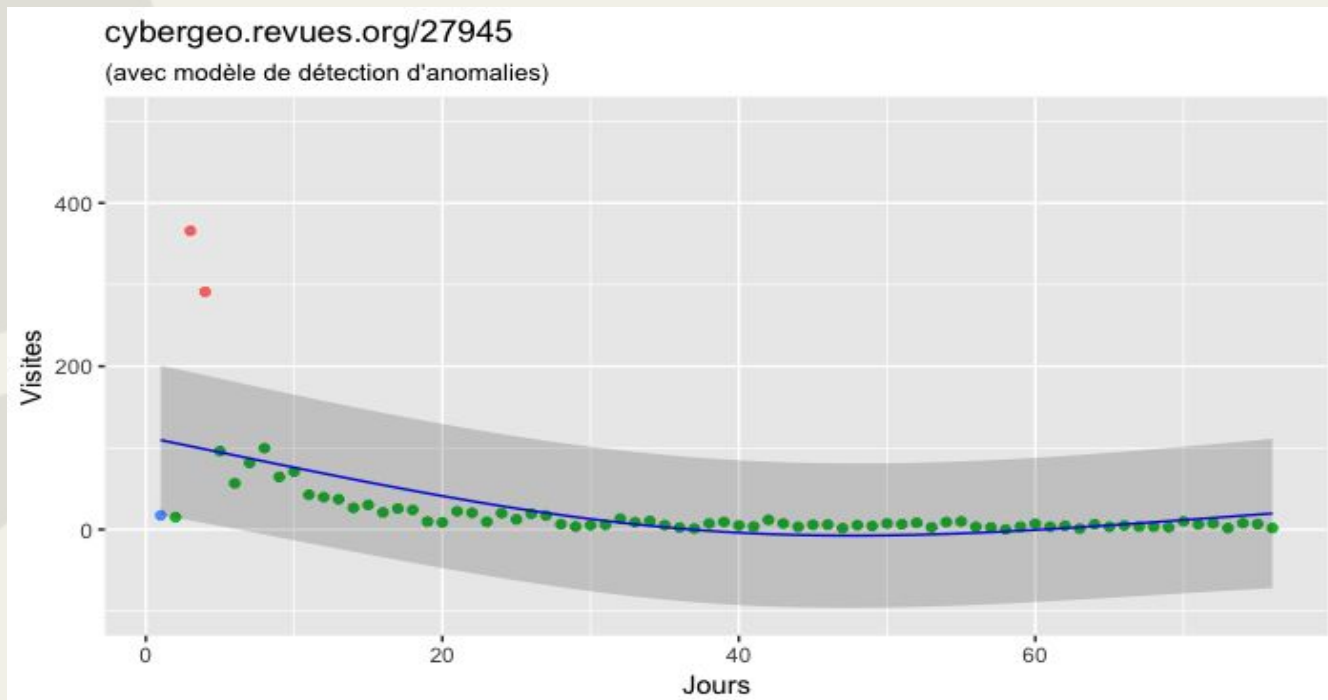
Un problème identifié par l'armée de l'air américaine en 1945 : le siège d'avion moyen ne convient à personne.



Cockpit of Concorde. CC BY SA - Christian Kath

# Modélisation et humanités numériques

## Classer l'inclassable



Paradoxalement les cas-limites, voire les « ruptures » dans le modèle peuvent former des observations plus intéressantes que les classifications correctes.

# Modélisation et humanités numériques

## Classer l'inclassable

Un exemple d'intertextualité particulièrement intense entre roman-feuilleton et sport, le récit d'un combat d'escrime très... romancé (31 janvier 1922 : roman-feuilleton 30%, sport 35%)

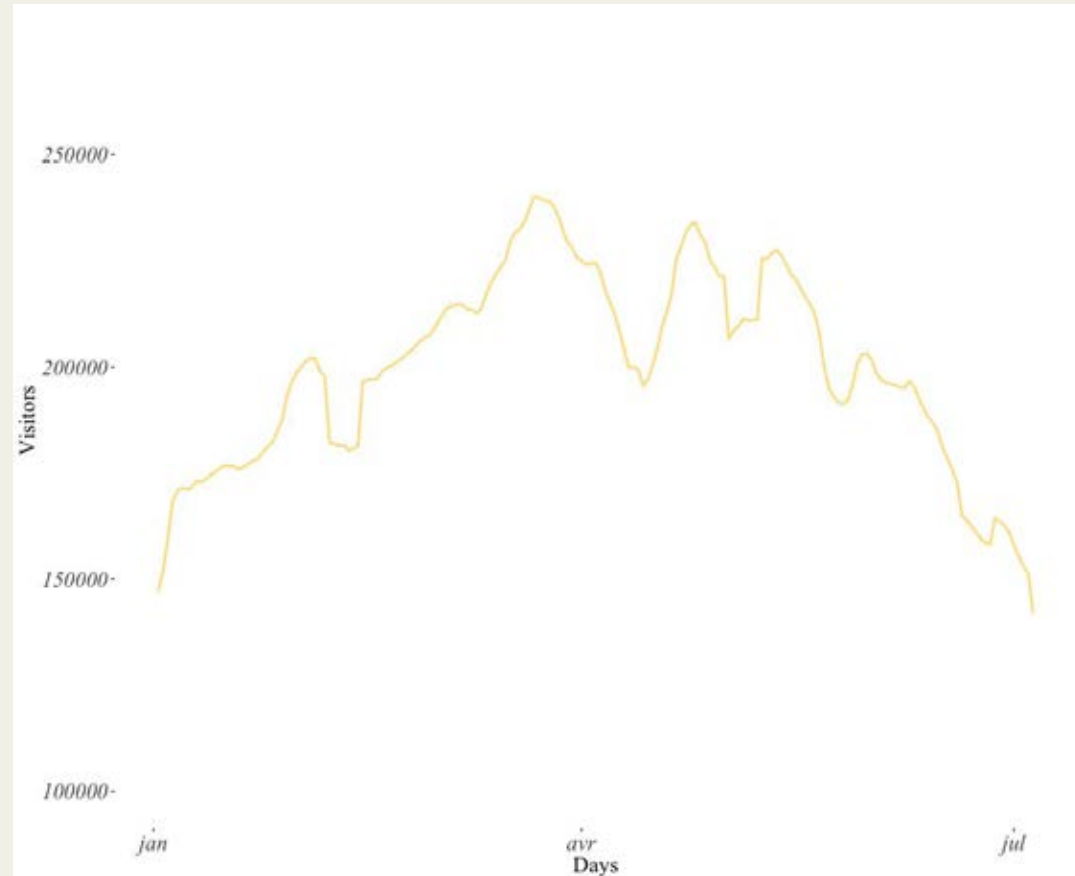


Paradoxalement les cas-limites, voire les « ruptures » dans le modèle peuvent former des observations plus intéressantes que les classifications correctes.

# Modélisation et humanités numériques

## La permanence du cycle

Un phénomène qui se constate régulièrement dans les logs : les pratiques des lectures sont « cycliques », en accord avec des récurrences hebdomadaires / saisonnières

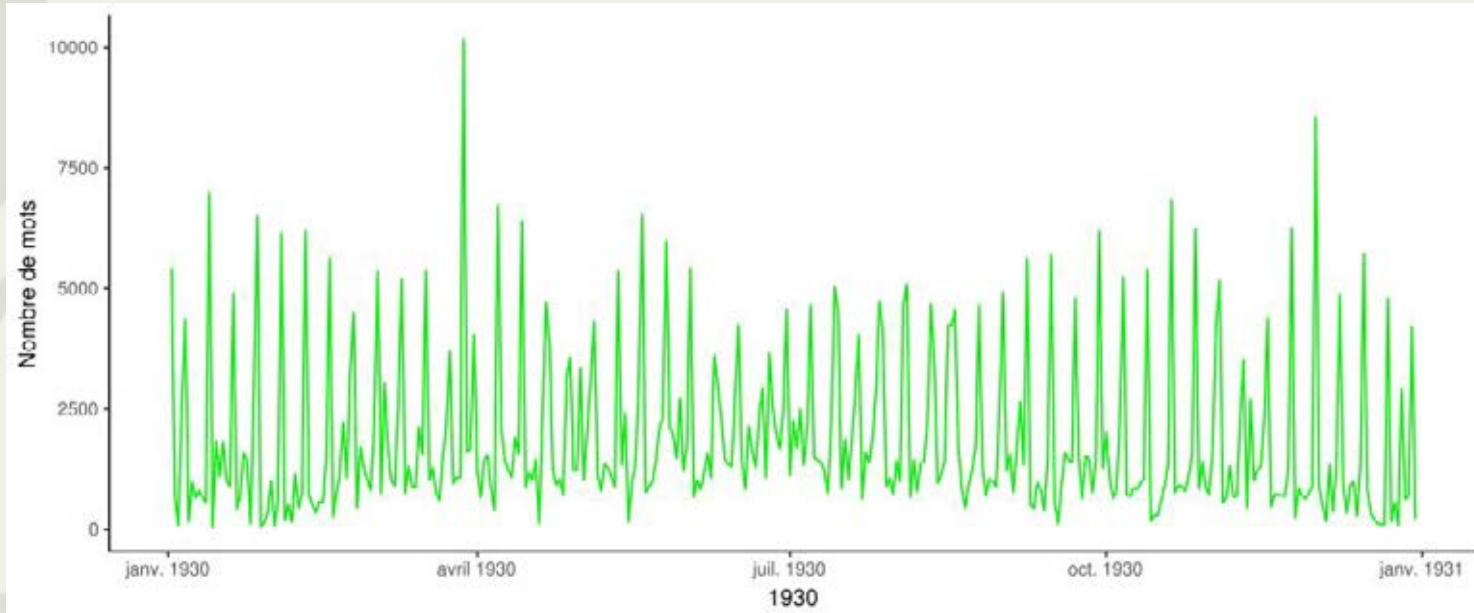


Visiteurs par jours (après correction)

# *Modélisation et humanités numériques*

## La permanence du cycle

Et c'est aussi une tendance que l'on perçoit en classant automatiquement le sport dans la presse des années 1920.



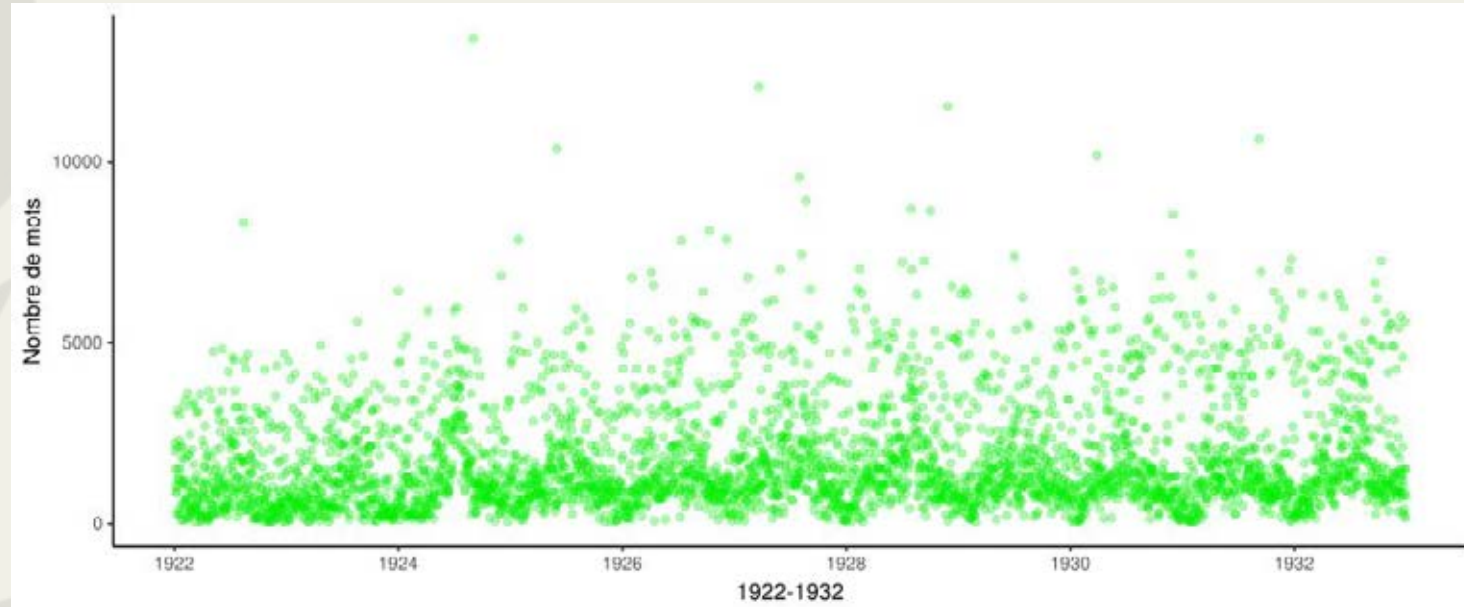
Nombre de mots consacrés au sport dans Le Matin en 1930 (variations hebdomadaires)



# *Modélisation et humanités numériques*

## La permanence du cycle

Et c'est aussi une tendance que l'on perçoit en classant automatiquement le sport dans la presse des années 1920.



Nombre de mots consacrés au sport dans Le Matin de 1922 à 1932 (variations saisonnières)





# 1. Les lecteurs inattendus

*De l'identification de cas particulier à la découverte d'un phénomène global*

# 1. Lecteurs inattendus

## L'émergence du lecteur inattendu

OPEN ACCESS AND INSTITUTIONAL REPOSITORIES, SCHOLARLY PUBLISHING

### The unexpected reader

🕒 NOVEMBER 15, 2011 👤 KEVIN SMITH, J.D. 💬 4 COMMENTS

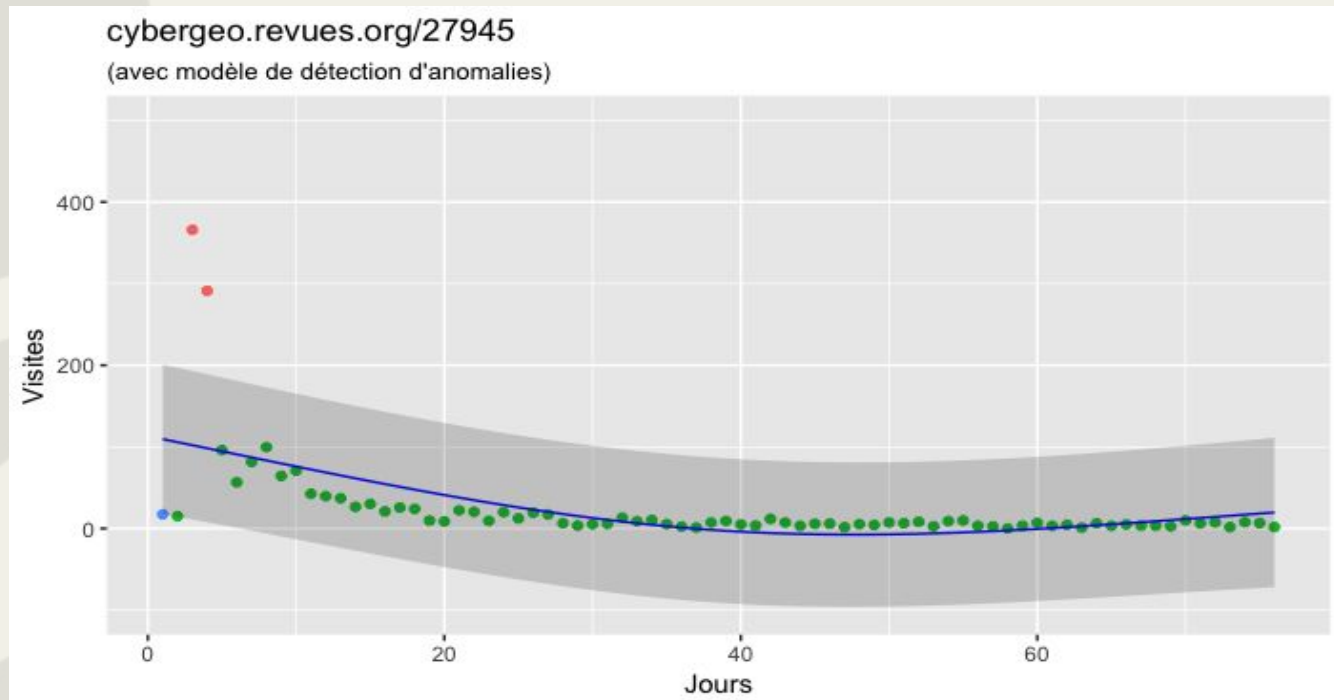
I have just returned from the Berlin 9 Conference on Open Access, which was held in Washington, D.C. at the lovely conference center facilities of the Howard Hughes Medical Institute. It was a fascinating meeting, and quite different in tone from the one I attended last year in Beijing.

In its opening paragraph, this [Chronicle of Higher Education report](#) on the conference captures the fundamental difference. This year the conference was much more clearly focused on the *impact* of open access on research; rather than talking about how open access will be accomplished, the discussion assumed that open access is inevitable and instead emphasized the differences that the evolution to open will make.

Une nouvelle forme de lecture rendue possible par le libre accès : la diffusion n'est plus uniquement restreinte aux pairs

# 1. Lecteurs inattendus

## Modéliser le lecteur inattendu



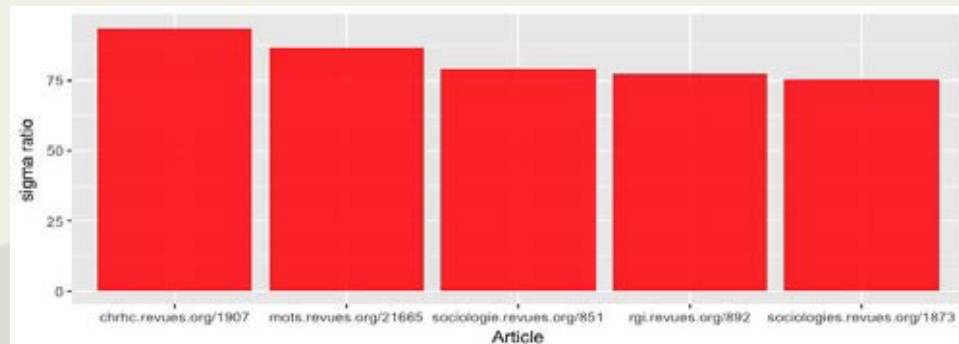
Le modèle représente les évolutions *structurelles* de la fréquentation

Ce qui en déborde est “inattendu”

# 1. Lecteurs inattendus

## Modéliser le lecteur inattendu

```
> anomaly_detection@url_prediction_anomalies %>% arrange(desc(fit_sigma_ratio)) %>% select(-lwr, -upr)
# A tibble: 32,087 x 5
  url diff_days log_value_anoms fit fit_sigma_ratio
  <chr> <dbl> <dbl> <dbl> <dbl>
1 chrhc.revues.org/1907 82 1748.3 40.590581 93.69053
2 mots.revues.org/21665 50 197.9 8.253494 86.53669
3 sociologie.revues.org/851 24 144.5 7.647771 79.31815
4 rgi.revues.org/892 97 168.5 8.926589 77.46158
5 sociologies.revues.org/1873 34 265.8 13.085899 75.29770
6 confins.revues.org/11729 6 2615.8 272.741627 74.65389
7 insaniyat.revues.org/5292 83 226.0 11.101971 73.66207
8 confins.revues.org/11701 10 183.5 15.243524 73.32456
9 com.revues.org/5838 36 294.4 19.330700 72.79988
10 terrain.revues.org/4263 31 98.0 6.078802 72.54528
```



Notre modèle permet de repérer les anomalies les plus extrêmes sur l'ensemble de revue.org

# 1. Lecteurs inattendus

## *Commémorer*

### Une commémoration

Hassan Fathy, construire *avec* ou *pour* le peuple?

Thierry Paquot

p. 15-25



Résumé | Index | Plan | Texte | Notes | Citation | Auteur

#### Résumé

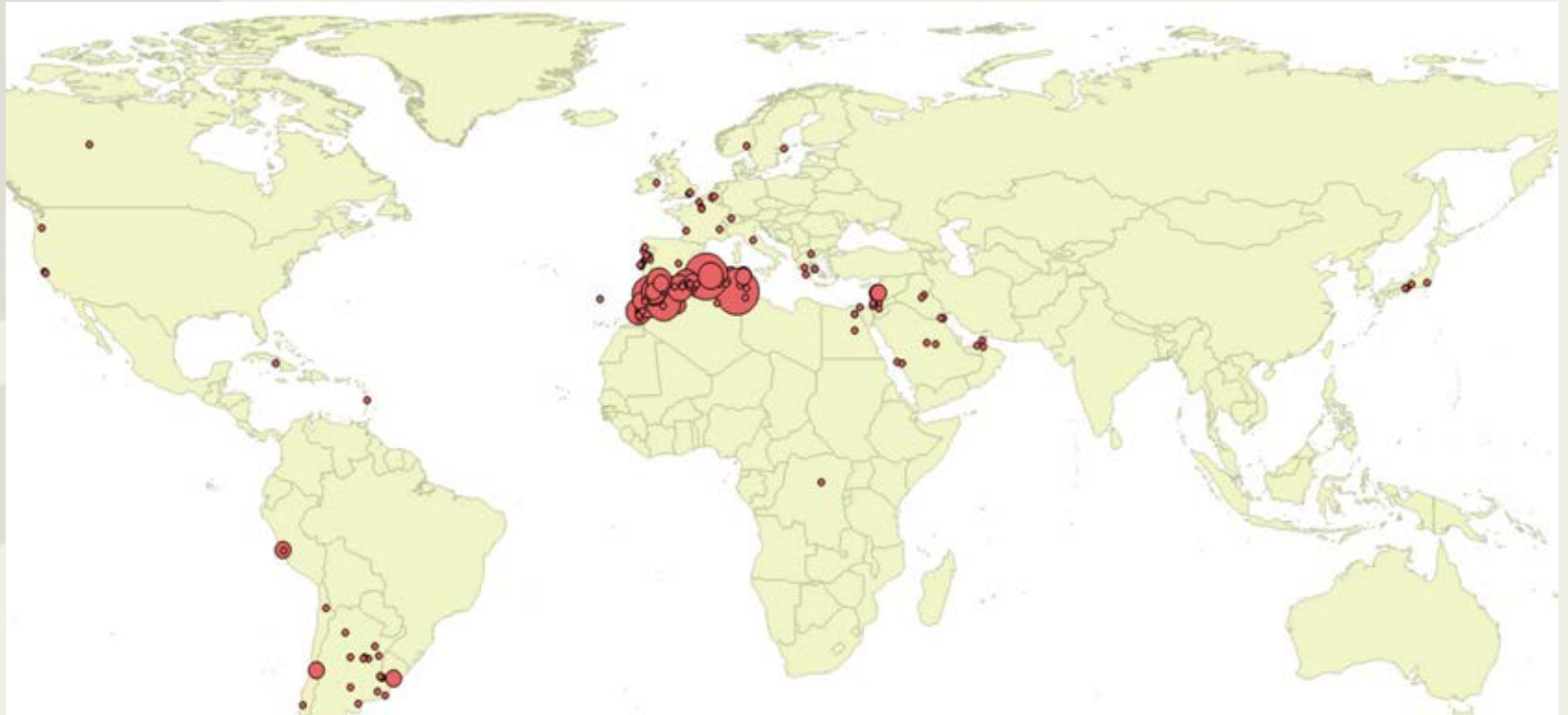
Au lendemain de la Seconde Guerre mondiale, l'architecte égyptien Hassan Fathy, en se lançant dans l'aventure architecturale du village de Gournah, propose des constructions économes et expérimente de nouveaux procédés constructifs. Militant de l'auto-construction et d'une esthétique de qualité, il associe les peuples à la construction de leurs maisons et exploite les traditions locales. La réception de son œuvre pose la question d'une architecture pour le peuple et d'une rupture douce avec le passé.

<http://chrhc.revues.org/1907>

2000 visites en une journée pour un vieil article qui n'en a que 3 ou 4 habituellement. La cause : un doodle de Google consacré à l'architecte égyptien Hassan Fathy

# 1. Lecteurs inattendus

## *Commémorer*



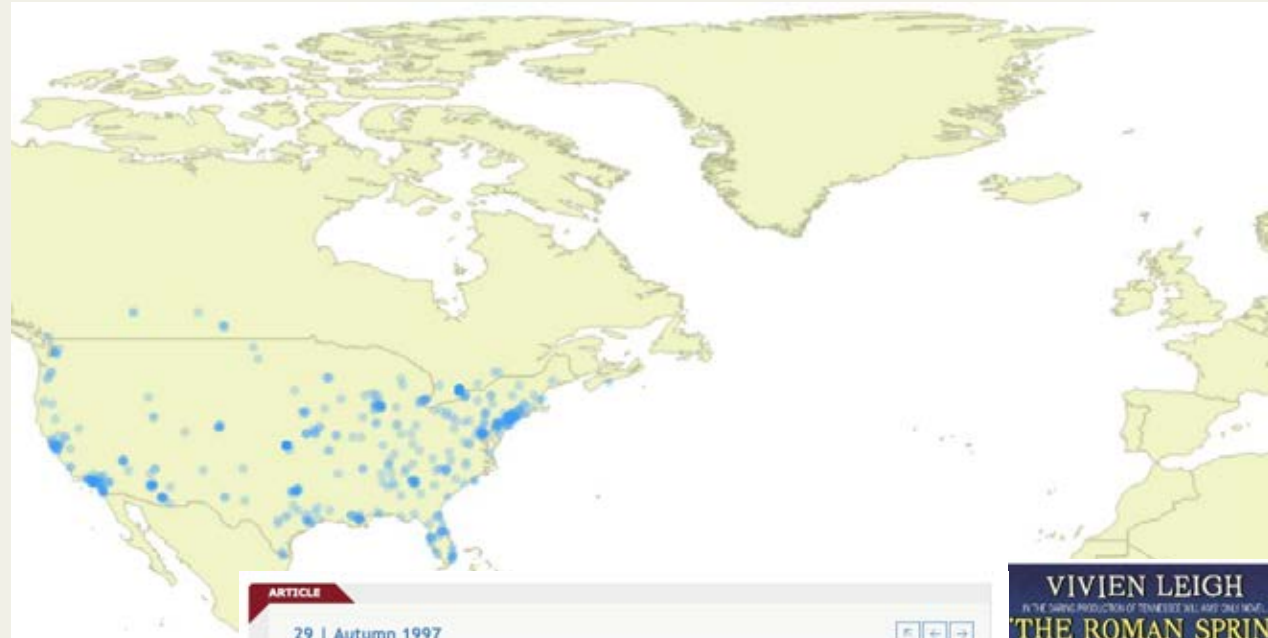
Les coordonnées des visites suggèrent que le doodle a surtout été activé dans les pays du Maghreb



# 1. Lecteurs inattendus

## *Prolonger un film...*

Hassan Fathy est un cas exceptionnel mais il y a des anomalies beaucoup plus fréquentes : les articles sur des films ont plein de visites... après leur diffusion à la télévision.



ARTICLE

29 | Autumn 1997

Other places, other selves

**The kindness of strangers : savage rescue in Tennessee**  
Williams' *The roman spring of mrs. Stone*

Theoharis Theoharis Constantine

Résumé | Texte | Citation | Auteur

RÉSUMÉ

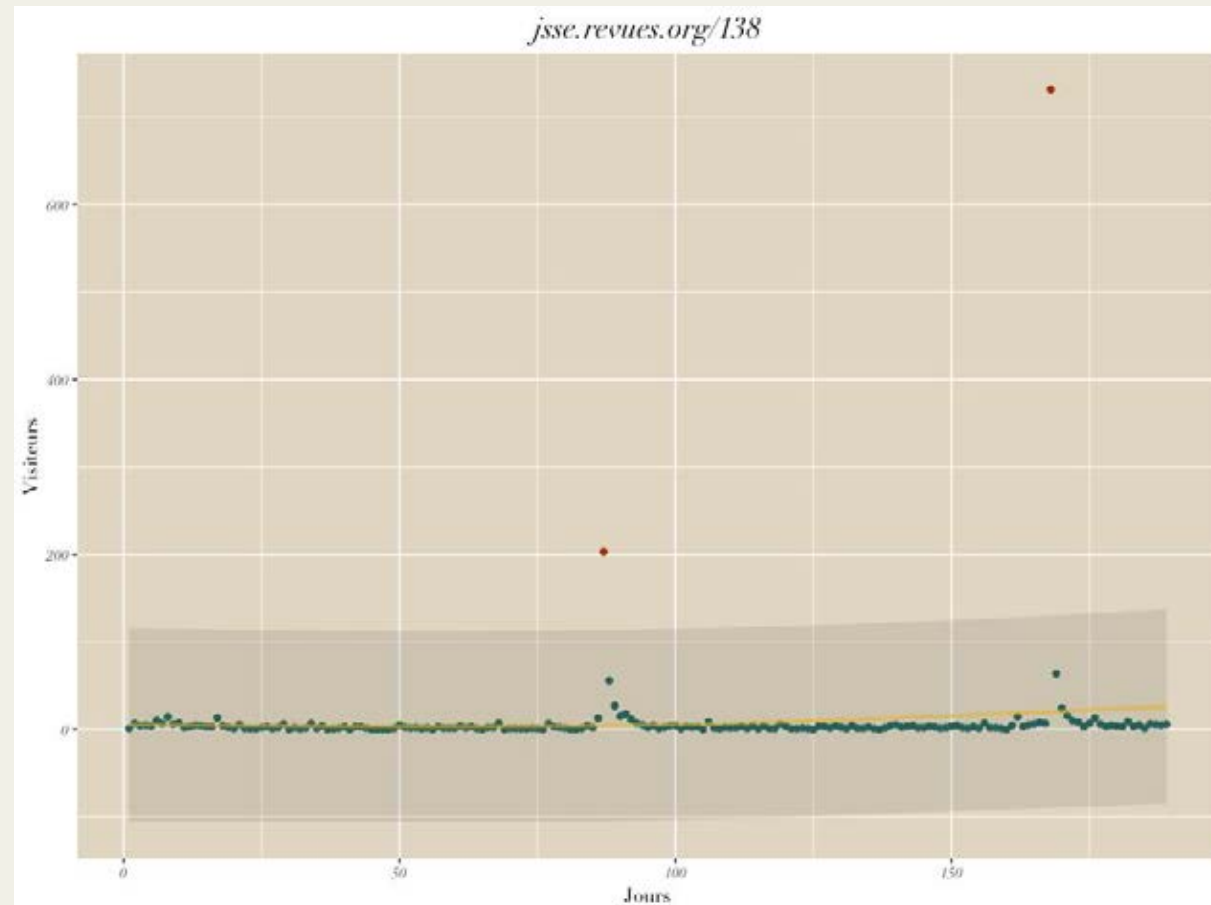
Dans la longue nouvelle "Le Printemps romain de Mrs Stone", Williams présente l'éveil tardif à la vie d'une actrice américaine qui vit loin de chez elle. A travers des stratégies narratives qui engagent l'ambiguïté, des modes de construction de personnages fondés sur le paradoxe, l'auteur fait traverser à son héroïne deux amours sacrificielles : en passant d'un ami entretenu attiré à un jeune garçon qui se prostitue dans la rue, Mrs Stone laisse derrière elle les conventions d'une vie élégamment décadente pour pénétrer dans le violent univers du paganisme que le titre de la nouvelle semble suggérer. Sa délivrance est, certes, crue, teintée de barbare. Cette lecture psychanalytique de l'identité sexuelle du personnage est fondée sur la conception classique de la représentation du désir.



# 1. Lecteurs inattendus

## *Prolonger un film...*

Hassan Fathy est un cas exceptionnel mais il y a des anomalies beaucoup plus fréquentes : les articles sur des films ont plein de visites... après leur diffusion à la télévision.





# 1. Lecteurs inattendus

## *De la science à l'anecdote à la science*

De nombreux épisodes de fréquentation inattendu mettent en évidence le rôle de nouveaux intermédiaires entre publications scientifiques et lecteurs inattendus

The Recipes Project

ABOUT CREDITS RECIPES PROJECT FLASHBACK THEMATIC SERIES ADVICE FOR CONTRIBUTORS

Food, Magic, Art, Science, and Medicine

RECENT POSTS

Indigo or no indigo?

Strike notice 4: feeding a strike

Strike Notice 3: International women's day

Strike

Med

reli

CAT

Se

EARLY MODERN, EXPERIMENTATION, FOOD AND DRINK, GERMANY, INGREDIENTS, MATERIAL HISTORY, POSTS, PRINTED BOOKS, RESEARCH AND WRITING, STUDENTS

BEER SOUP: THE BREAKFAST OF EARLY MODERN RULERS

30/05/2013 M. TAYLOR-POLESKEY 4 COMMENTS

leskey

Prince Friedrich Wilhelm, the Elector of Prussia began each morning with a beer soup. He then fell asleep and attended to the day's business until

les 200 meilleurs commentaires en afficher 500

triés par: apprécié

[-] michaelrohansmith 1094 points il y a 10 mois

Boil all the water you use, for drinking and washing.

perma-lien embed enregistrer

[-] Arsnicthegreat 338 points il y a 10 mois

That's still good advice in a lot of places.

perma-lien embed enregistrer parent

# 1. Lecteurs inattendus

## *De la science à l'anecdote à la science*

De nombreux épisodes de fréquentation inattendu mettent en évidence le rôle de nouveaux intermédiaires entre publications scientifiques et lecteurs inattendus

The screenshot shows a web page with a breadcrumb trail: Accueil > High-Tech / Science > Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie. Navigation buttons for 'TOP PRÉCÉDENT' and 'TOP SUIVANT' are visible. The article title is 'Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie'. The author is Pierre Galouise, dated 4/04/2017, with 699 shares. The category is High-Tech / Science. A sidebar on the right contains an 'Index' with categories like Auteurs, Mots-clés, Chronologique, and Thématique. Below the index is a 'Derniers numéros' section with a list of issues from 2002 to 2014. The main content area on the right shows the start of an article titled 'Les chemins de fer et l'heure légale' by Lucien Baillaud, with a 'Plan' section listing various topics related to time zones and standard time.

Accueil > High-Tech / Science > Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie

TOP PRÉCÉDENT  
10 explications possibles derrièr ...

TOP SUIVANT  
10 villes des Etats-Unis à voir a ...

### Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie

Par Pierre Galouise  
le 4/04/2017

699 PARTAGES

Catégorie : High-Tech / Science

Entre quelques **phénomènes météorologiques** au fil de l'Histoire, on a parfois l'impression que le top a principalement été inspiré par l'ouvrage **« la mesure du temps »** qui nous livre une liste de faits et c'est DINGO.

#### Index

- Auteurs
- Mots-clés
- Chronologique
- Thématique

#### Derniers numéros

- 24-25 | 2002

#### Numéros en texte intégral

- 45 | 2014
- 44 | 2013
- 42-43 | 2012
- 41 | 2010
- 40 | 2009
- 39 | 2008
- 38 | 2008
- 36-37 | 2007

### Les chemins de fer et l'heure légale

Lucien Baillaud

Plan | Texte | Notes | Citation | Auteur

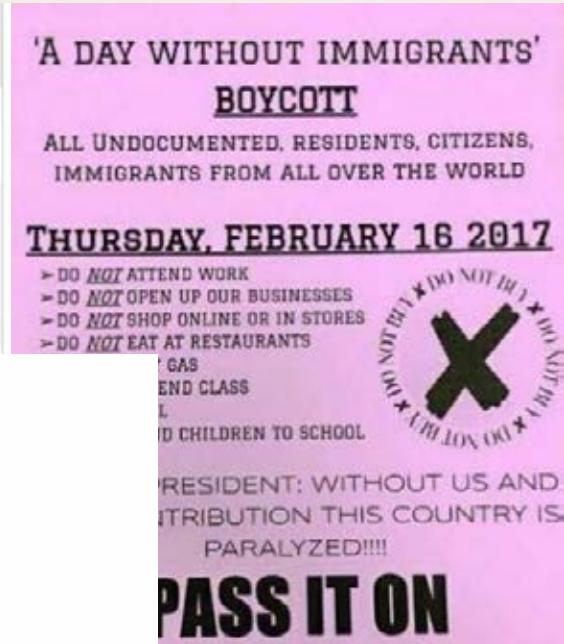
#### Plan

- Les inconvénients des heures locales
- Vers l'unification des heures : les heures nationales
- L'heure de Rouen
- Le système américain
- La conférence de Washington
- Vers la généralisation des fuseaux horaires
- 24 heures ? ou 12 et 12 ?
- La décimalisation du temps et les problèmes de minutes
- Conclusion

# 1. Lecteurs inattendus

## *Conserver la mémoire*

Certains épisodes de lecteurs inattendus illustrent la fonction “mémorielle” de la recherche scientifique : ici une analyse du “Day Without Immigrants” de 2009 redécouverte en 2017.



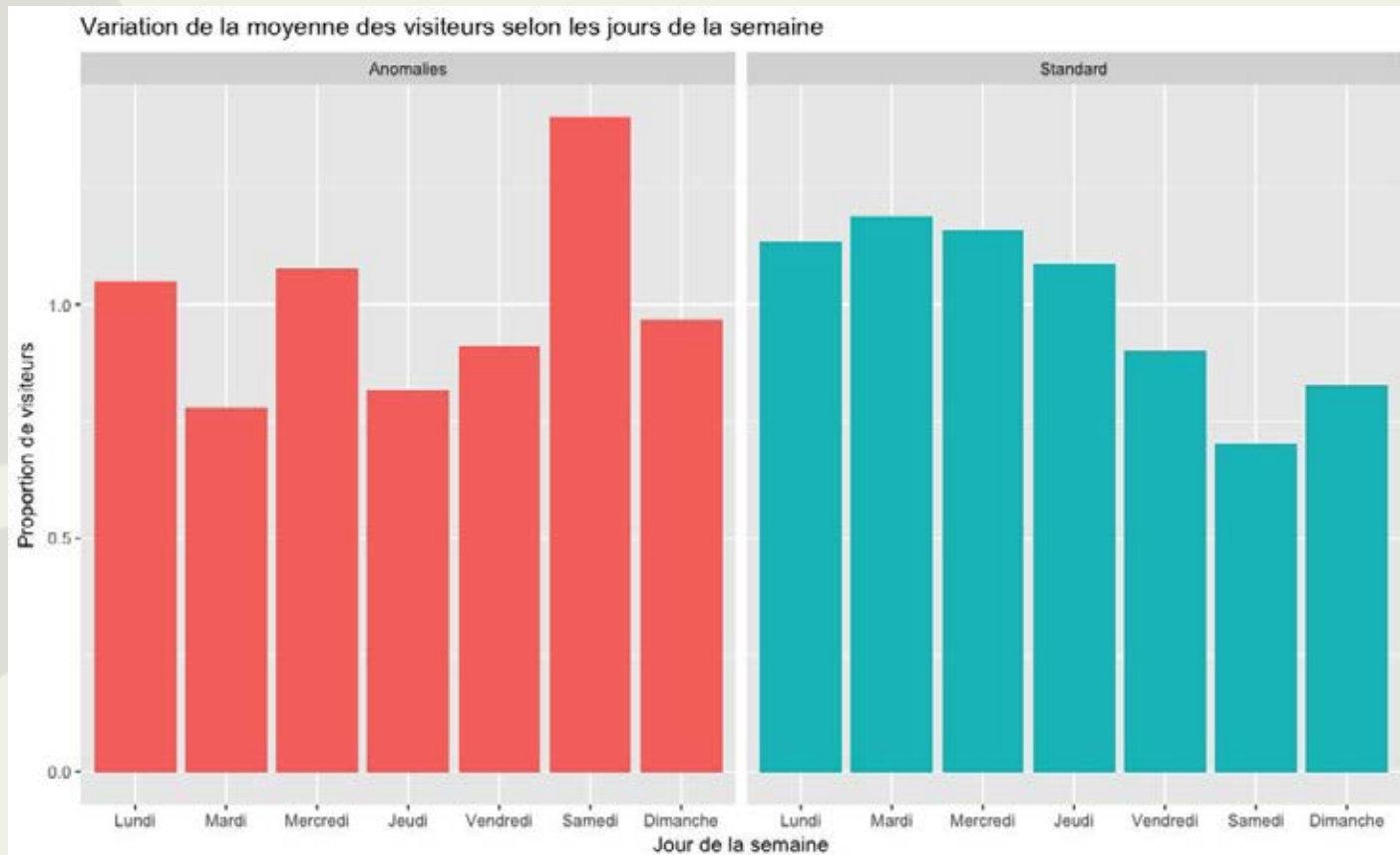
# 1. Lecteurs inattendus

Au-delà des cas particuliers ici mis en lumière, la lecture inattendue est une pratique courante pour les publications en libre accès. Au-delà de leur réception par les pairs, les articles peuvent :

- Préserver et documenter une mémoire collective
- Prolonger la découverte d'une production culturelle (film, livre, etc.)
- Recontextualiser la vulgarisation scientifique.

# 1. Lecteurs inattendus

## *Vers une typologie de l'inattendu ?*

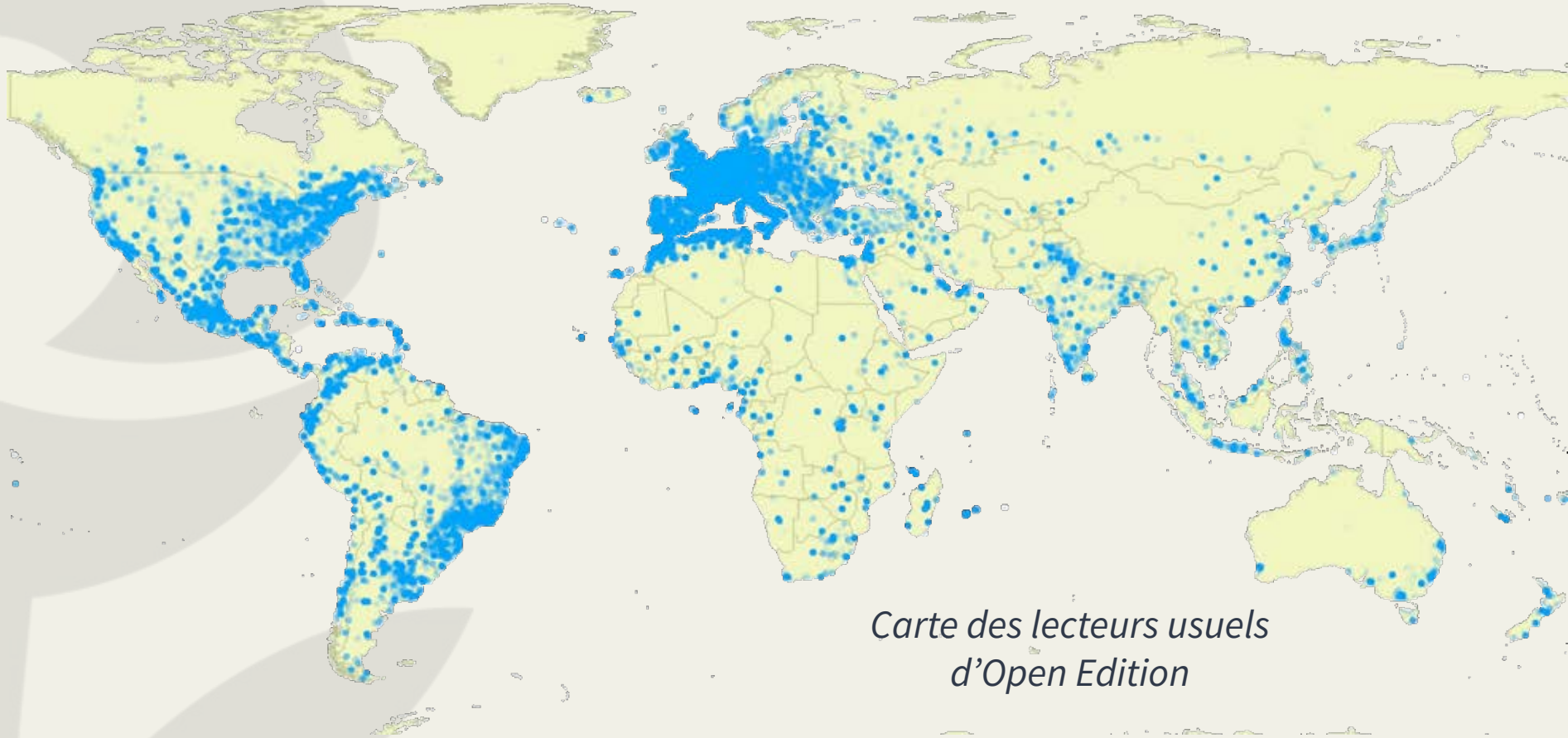


Le lecteur inattendu est un lecteur comme les autres mais il vient surtout... le samedi.



# 1. Lecteurs inattendus

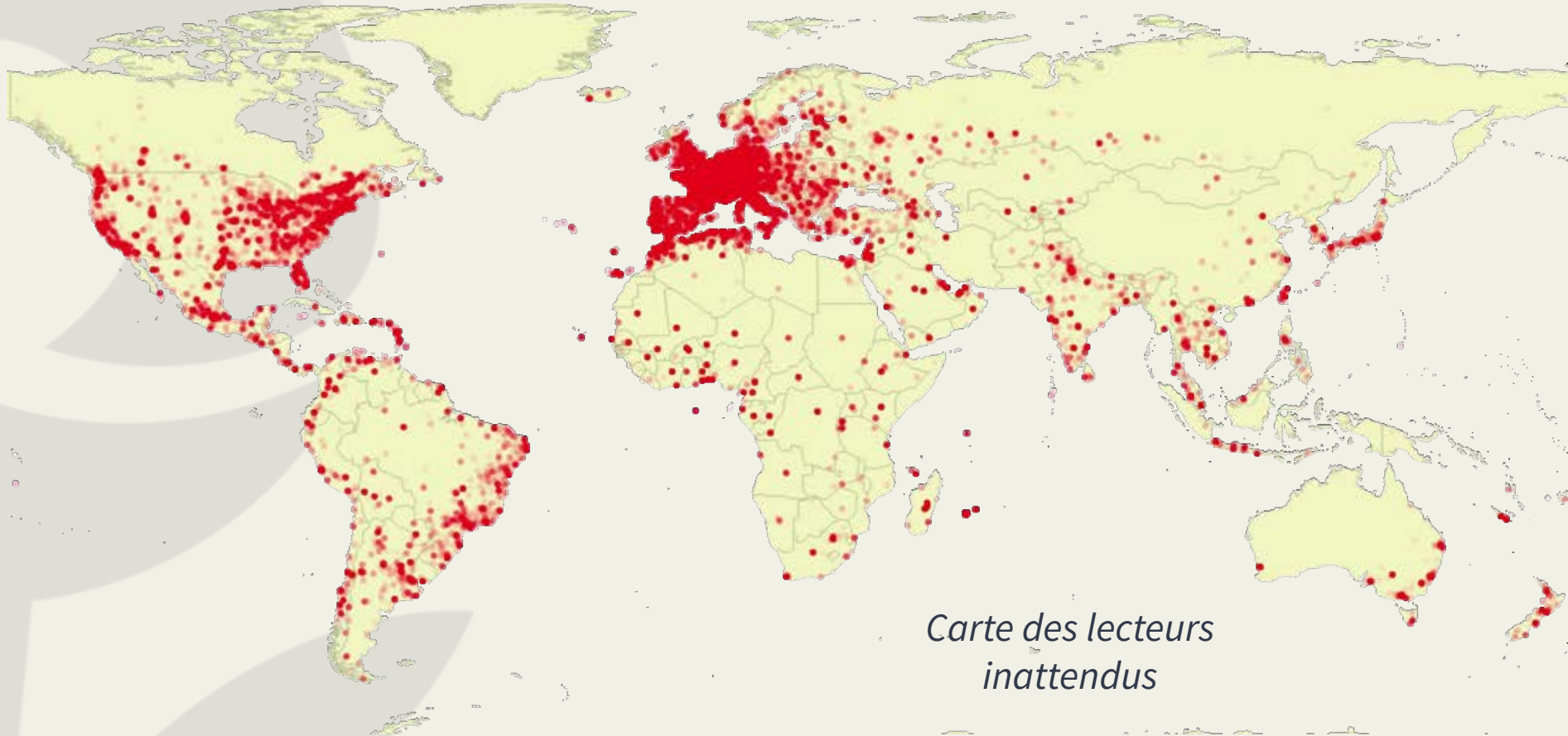
*Vers une typologie de l'inattendu ?*



Le lecteur inattendu est... un lecteur comme les autres. Il vient de partout.

# 1. Lecteurs inattendus

*Vers une typologie de l'inattendu ?*

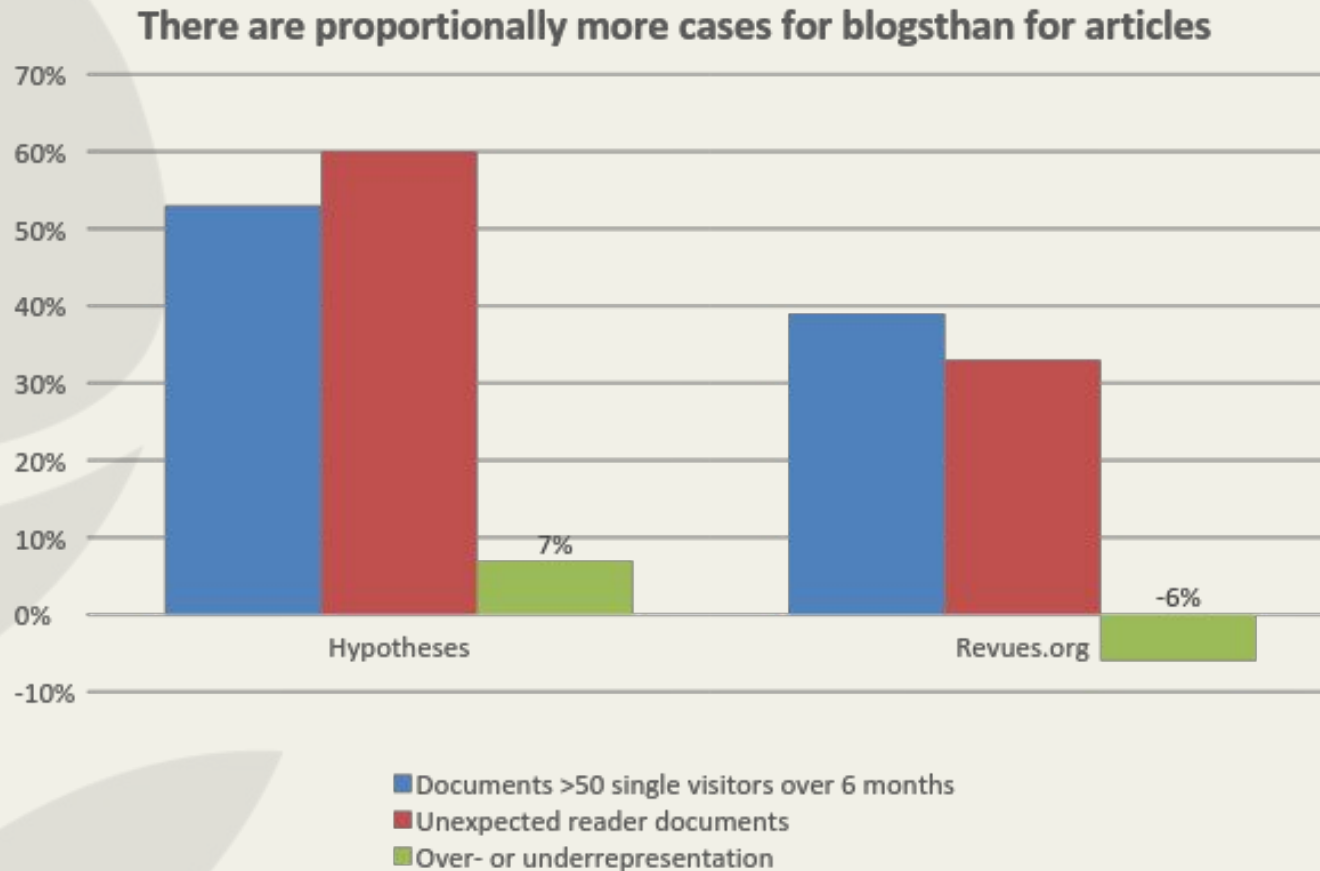


*Carte des lecteurs  
inattendus*

Le lecteur inattendu est... un lecteur comme les autres. Il vient de partout.

# 1. Lecteurs inattendus

## *Vers une typologie de l'inattendu ?*

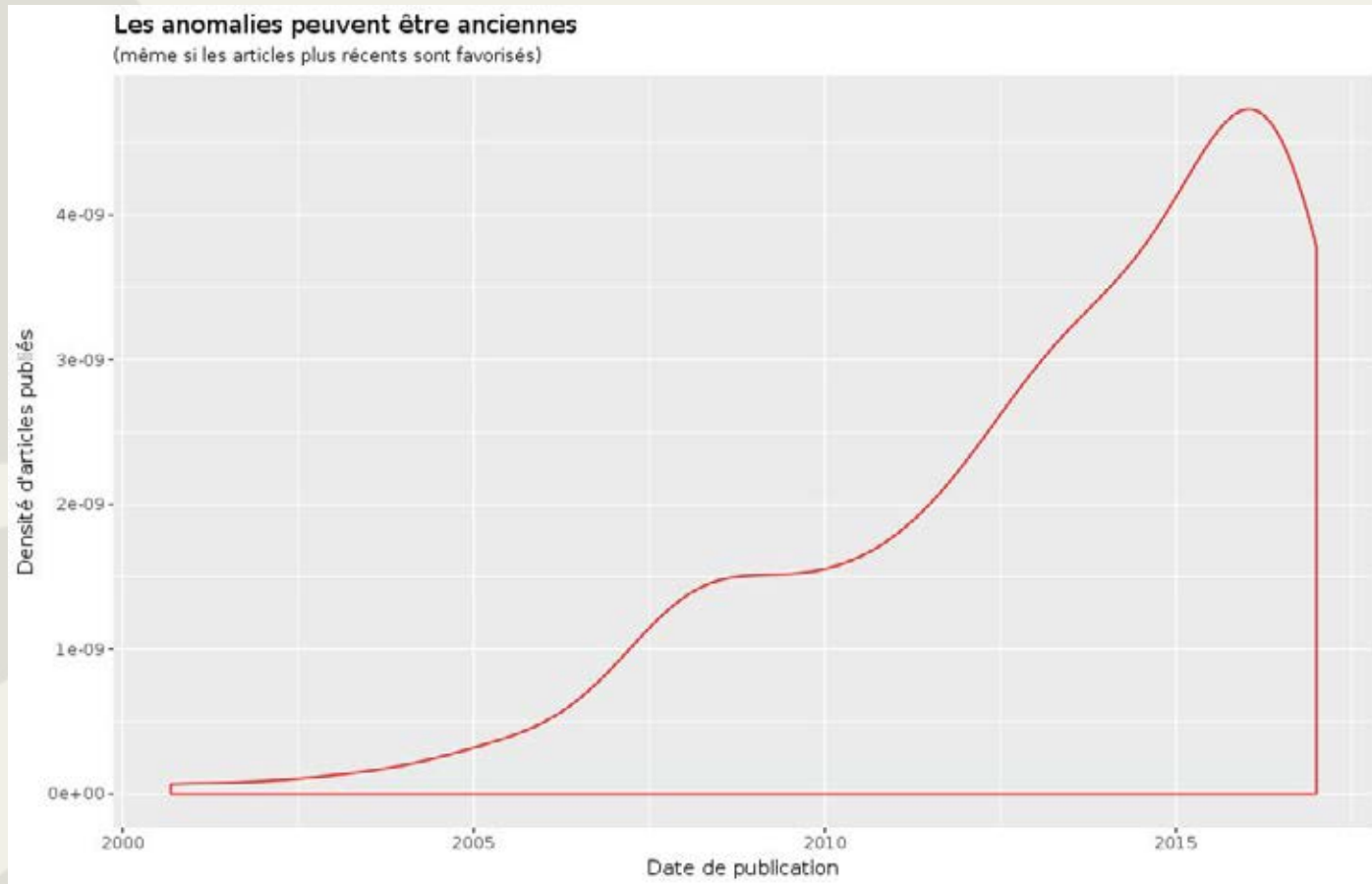


Le lecteur inattendu apprécie la forme blog sans dédaigner le format classique de l'articles de revues scientifiques



# 1. Lecteurs inattendus

## *Vers une typologie de l'inattendu ?*



Les lecteurs inattendus réactualisent fréquemment des publications anciennes.

# 1. Lecteurs inattendus

## *L'application Umberto*

Umberto présente deux écrans successifs.

- le premier écran permet de sélectionner l'épisode de lecteurs inattendus que l'on souhaite explorer.
- le second écran présente les informations disponibles sur l'épisode de lecteur inattendu retenu.

# 1. Lecteurs inattendus

## *L'application Umberto*

Détecteur de lecteurs inattendus

https://rstudio.labocleo.org/p/4122/

**OpenEdition**

Bienvenue dans le détecteur de lecteurs inattendus !

Plateformes :

- Hypothèses
- Revues.org
- OpenEdition Books
- Calenda

Publications :

Choisir un ou plusieurs espaces de publications :

Période :

Choisir la période concernée

01/01/2017 à 13/11/2017

Intensité minimale

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Choisissez l'épisode de lecteurs inattendus que vous voulez explorer

Afficher 10 éléments

Rechercher :

	Plateforme	Publication	Document	URL	Intensité	Date
1	Revue.org	Entrélaçs	Matthew Barney, artiste/auteur : pour une hybridation des genres	<a href="http://entrelacs.revues.org/343">http://entrelacs.revues.org/343</a>	97	04/10/2017
2	Revue.org	Études de communication	Technologies de l'information et de la communication et modernisation des services publics. Quelques remarques et repères pour une exploration critique	<a href="http://edc.revues.org/1135">http://edc.revues.org/1135</a>	95	25/09/2017
3	Revue.org	Genre, sexualité et société	Virginité des filles et rapports sociaux de sexe dans quelques récits d'écrivaines marocaines contemporaines	<a href="http://gss.revues.org/1413">http://gss.revues.org/1413</a>	93	22/04/2017
4	Revue.org	Cahiers d'histoire. Revue d'histoire critique	Hassan Fathy, construire avec ou pour le peuple?	<a href="http://chrhc.revues.org/1907">http://chrhc.revues.org/1907</a>	92	24/03/2017
5	Revue.org	Sociologie	Communauté et société	<a href="http://sociologie.revues.org/2574">http://sociologie.revues.org/2574</a>	91	23/07/2017
6	Hypotheses.org	Archivalia	Open Access Woche (Tag 2)	<a href="http://archivalia.hypotheses.org/68021">http://archivalia.hypotheses.org/68021</a>	91	25/10/2017
7	Hypotheses.org	Zilsel	Réponse à la réponse de Pascal Ragouet	<a href="http://zilsel.hypotheses.org/2928">http://zilsel.hypotheses.org/2928</a>	90	23/10/2017
8	Revue.org	Insaniyat / إنسانيات	2002-1963) من الثورات الزراعية إلى الإصلاحات البورقوية (العلامة في الجزائر)	<a href="http://insaniyat.revues.org/7027">http://insaniyat.revues.org/7027</a>	87	30/06/2017
9	Revue.org	Cahiers de la Méditerranée	Les Alpes-Maritimes et la crise franco-monégasque de 1962	<a href="http://cdim.revues.org/3023">http://cdim.revues.org/3023</a>	87	22/08/2017
10	Hypotheses.org	Le blog du collectif RENOIR	Offre de stage "Sauvegarde du patrimoine de l'observatoire du Pic du Midi"	<a href="http://renoir.hypotheses.org/1946">http://renoir.hypotheses.org/1946</a>	86	07/10/2017

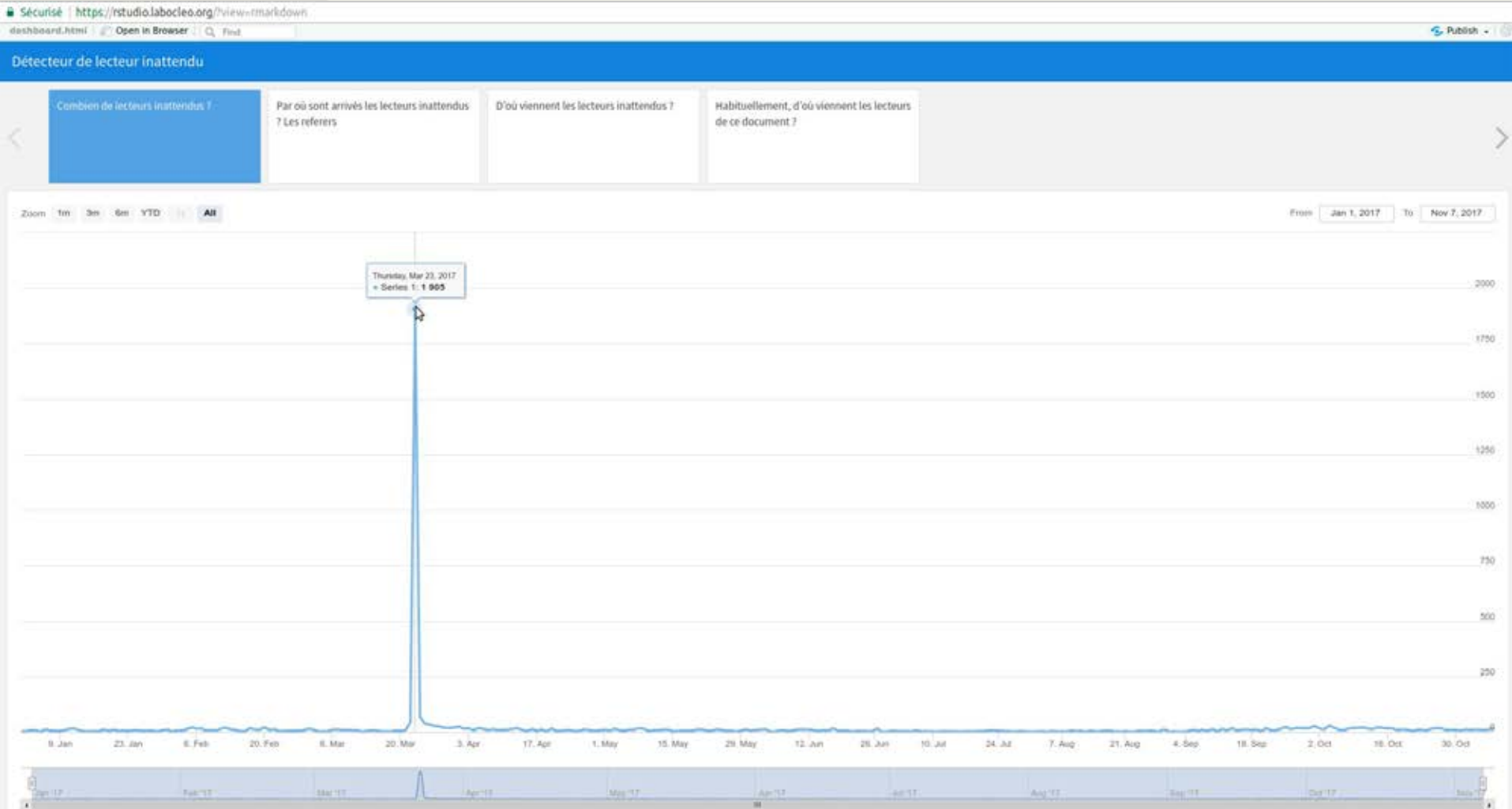
Affichage de l'élément 1 à 10 sur 479 éléments

Précédent 1 2 3 4 5 ... 48 Suivant

Explorez !

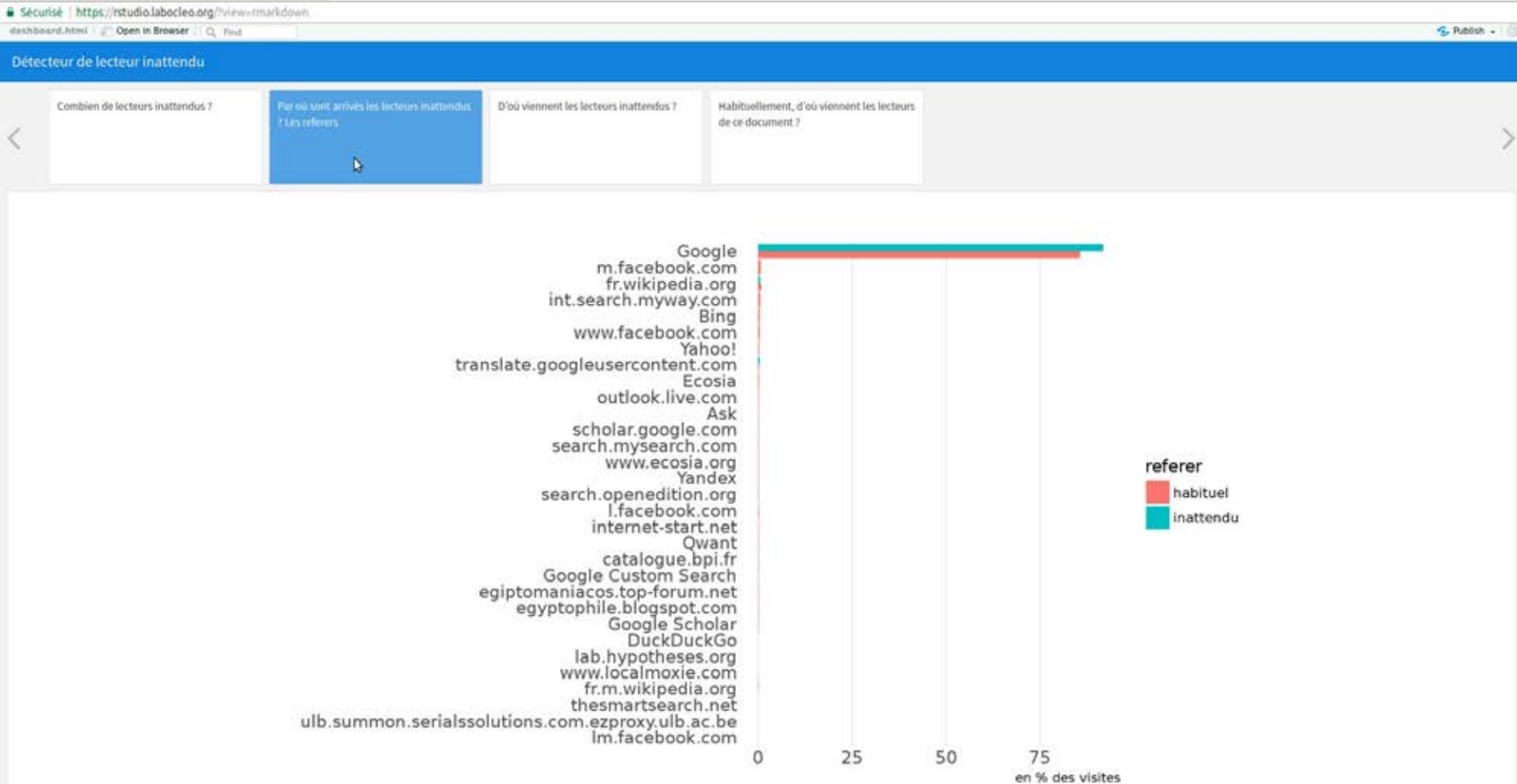
# 1. Lecteurs inattendus

## *L'application Umberto*



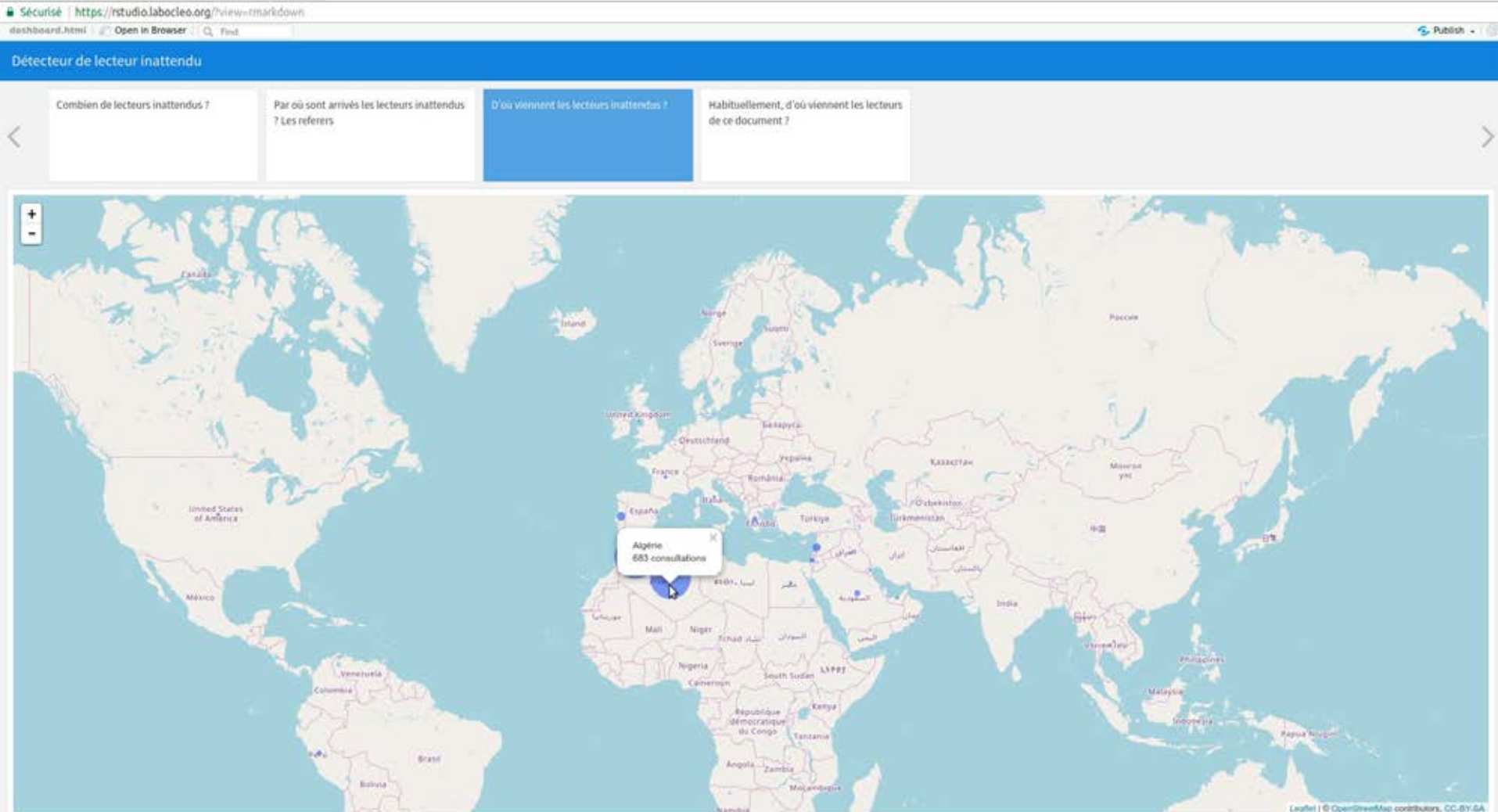
# 1. Lecteurs inattendus

## *L'application Umberto*



# 1. Lecteurs inattendus

## *L'application Umberto*





## 2. Les lecteurs institutionnels

*Des pratiques usuelles du lecteur  
attendu.*

## 2. Résultats obtenus

### *Observation des usages*

#### Une autre ressource : les IPs collectives

Compilation de  
2514 adresses IPs  
de 1287  
institutions.

```
# A tibble: 2,514 × 2
  ip                institution
  <chr>             <chr>
1 81.80.82.0        A2C
2 195.220.17.0     Abbaye aux dames
3 193.52.26.0      ABES
4 193.52.69.0      ABES
5 194.199.47.0     Académie d'Amiens
6 195.221.155.0    Académie d'Amiens
7 195.221.156.0    Académie d'Amiens
8 194.199.46.0     Académie d'Amiens
9 195.83.242.0     Académie d'Aix-Marseille
10 195.83.253.0     Académie d'Aix-Marseille
# ... with 2,504 more rows
# i Use `print(n = ...)` to see more rows
```

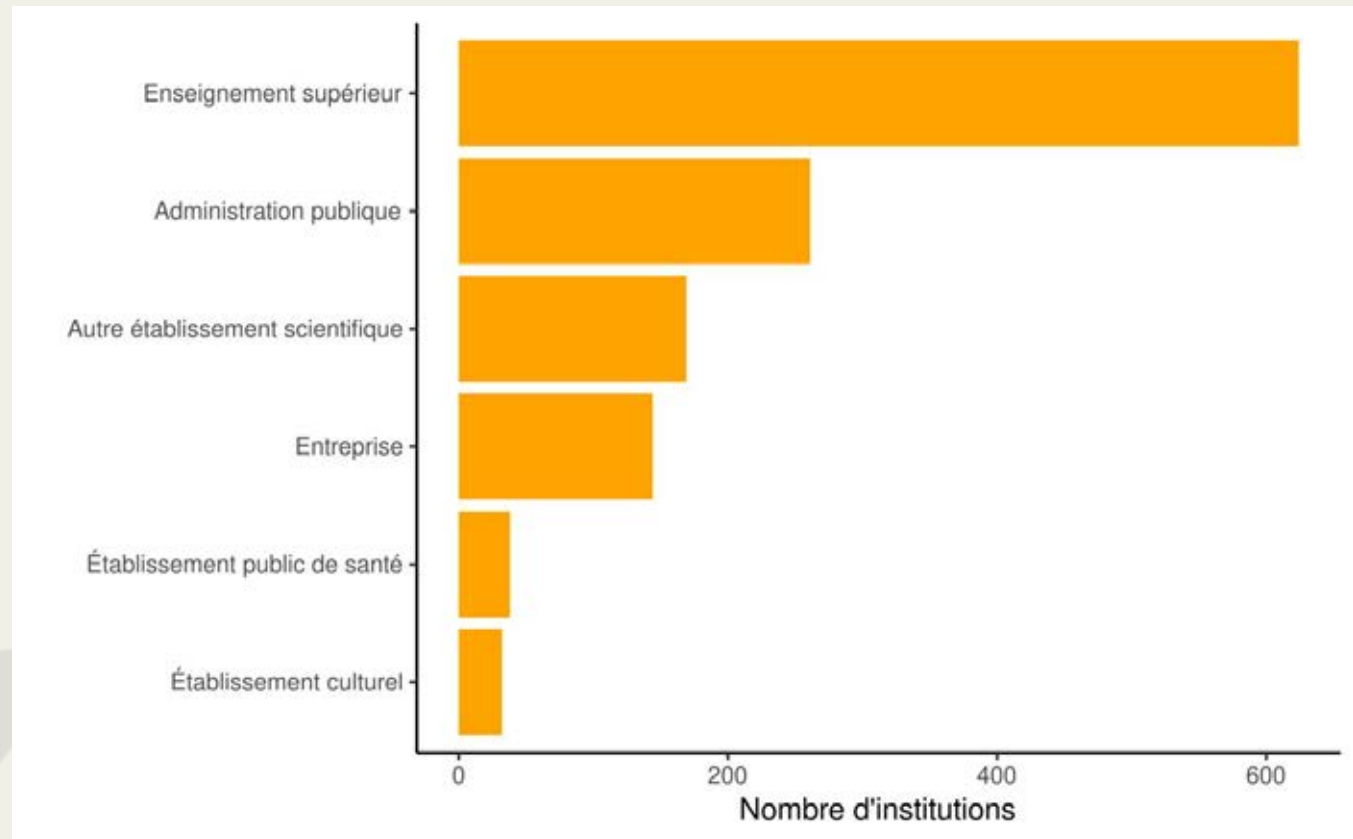


## 2. Résultats obtenus

### *Observation des usages*

#### Une autre ressource : les IPs collectives

Compilation de  
2514 adresses IPs  
de 1287  
institutions.

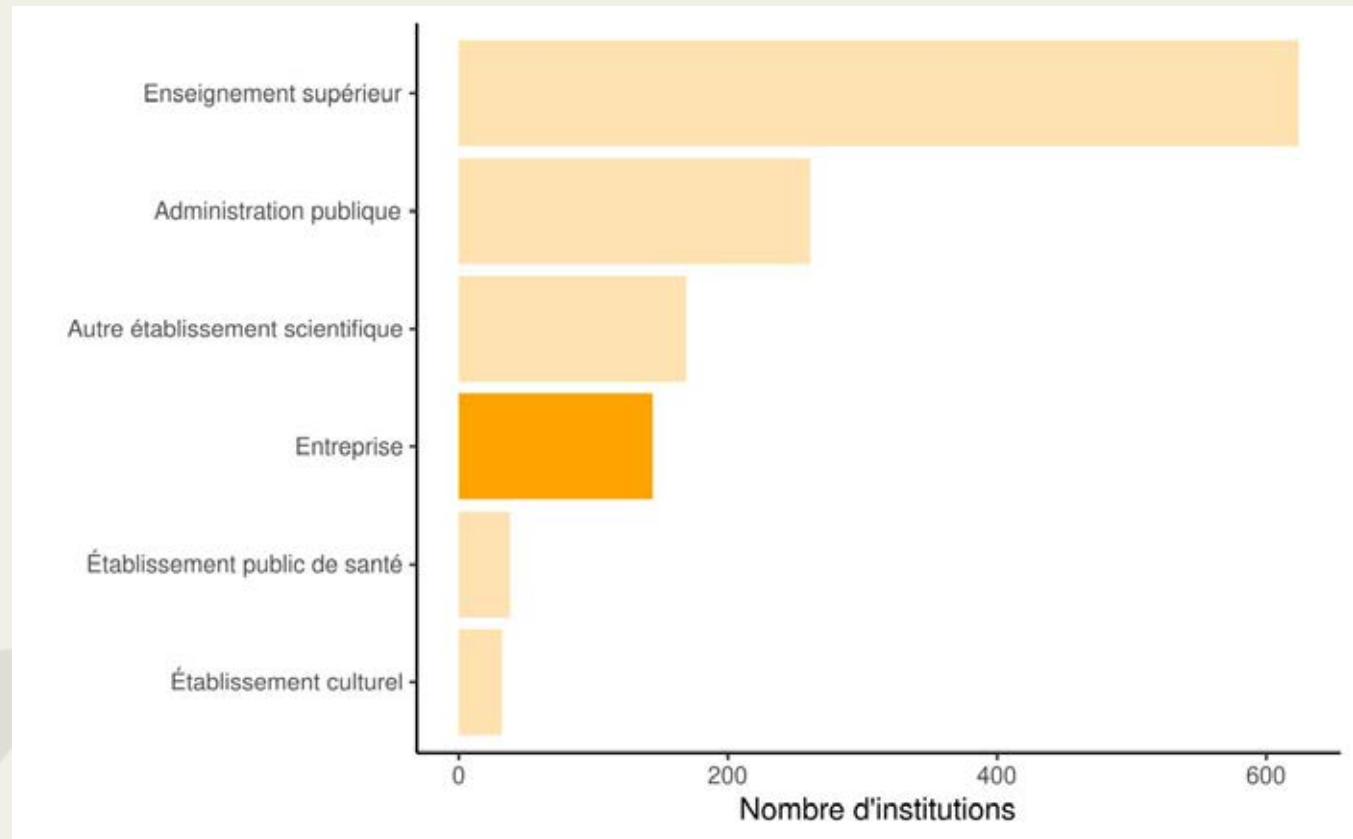


## 2. Résultats obtenus

### *Observation des usages*

#### Une autre ressource : les IPs collectives


Compilation de  
2514 adresses IPs  
de 1287  
institutions.



## 2. Résultats obtenus

### *Observation des usages*

## Une autre ressource : les IPs collectives

 **OpenEdition**  
Umberto, le détecteur de lecteurs d'Open Edition

Les visites inattendues | **Les visiteurs institutionnels** | Les sources des visites inattendues

Search:

Show **10** entries

**Pays :**  
Veuillez choisir le pays d'origine de l'IP  
Tous les pays

**Type d'institution :**  
Veuillez choisir le type d'institution  
Tous types

Le détecteur de lecteurs institutionnels a été appliqué à l'ensemble des logs d'Open Edition de janvier 2017 à juillet 2019. Les données peuvent être téléchargées depuis le lien suivant :  
[Télécharger les données complètes](#)

Identifiant	Institution	Affiliation	Type	Domaine	Pays	Visites
1	3e œil productions	3e œil productions	entreprise	cinéma	France	1387
2	A2C	A2C	entreprise	immobilier	France	2460
3	Aachen Universität	Aachen Universität	enseignement supérieur	université étrangère	Allemagne	2642
4	Aalborg Universitet	Aalborg Universitet	enseignement supérieur	université étrangère	Danemark	1145
5	Abbaye aux dames	Abbaye aux dames	institution religieuse	association	France	6
6	Abeille Musique	Abeille Musique	entreprise	musique	France	1350
7	ABES	ABES	autre établissement scientifique	EPA	France	4242
8	Académie d'Aix-Marseille	Académie d'Aix-Marseille	administration publique	éducation	France	3511
9	Académie d'Amiens	Académie d'Amiens	administration publique	éducation	France	44877
10	Académie de Besançon	Académie de Besançon	administration publique	éducation	France	1211

Showing 1 to 10 of 1,299 entries

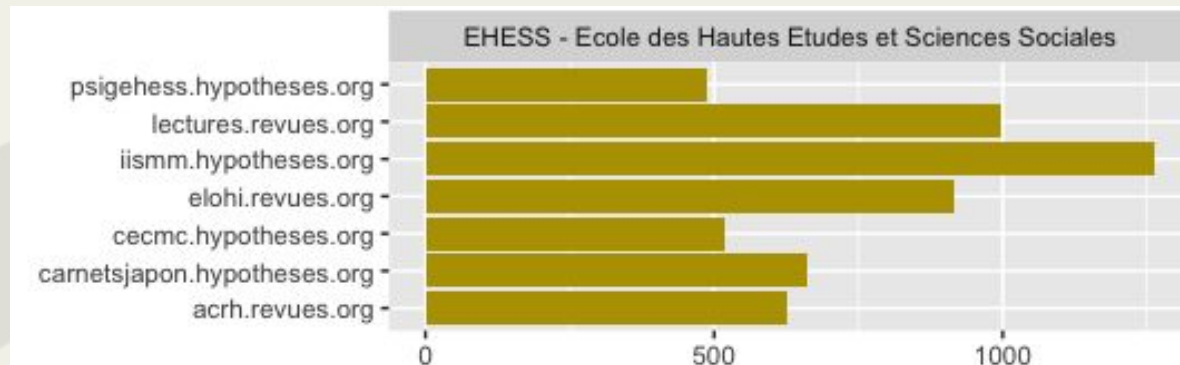
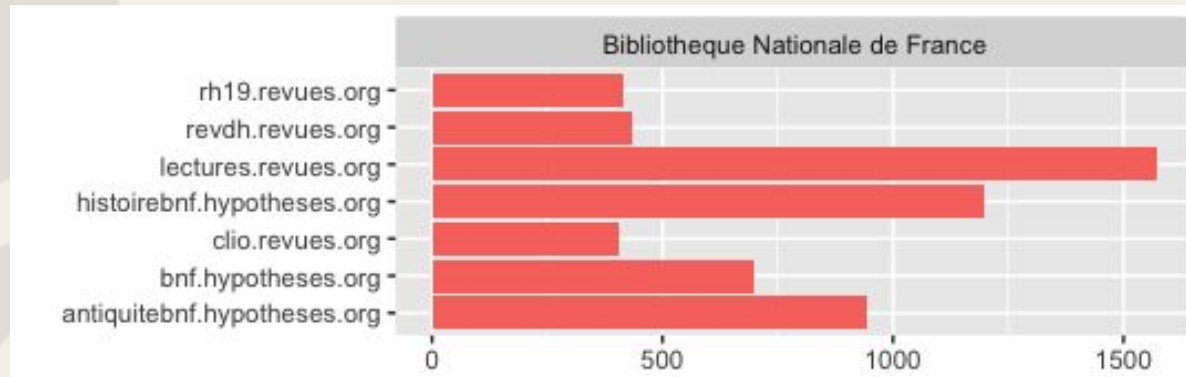
Previous **1** 2 3 4 5 ... 130 Next

## 2. Résultats obtenus

### *Observation des usages*

## Une autre figure de lecteur : le lecteur “institutionnel”

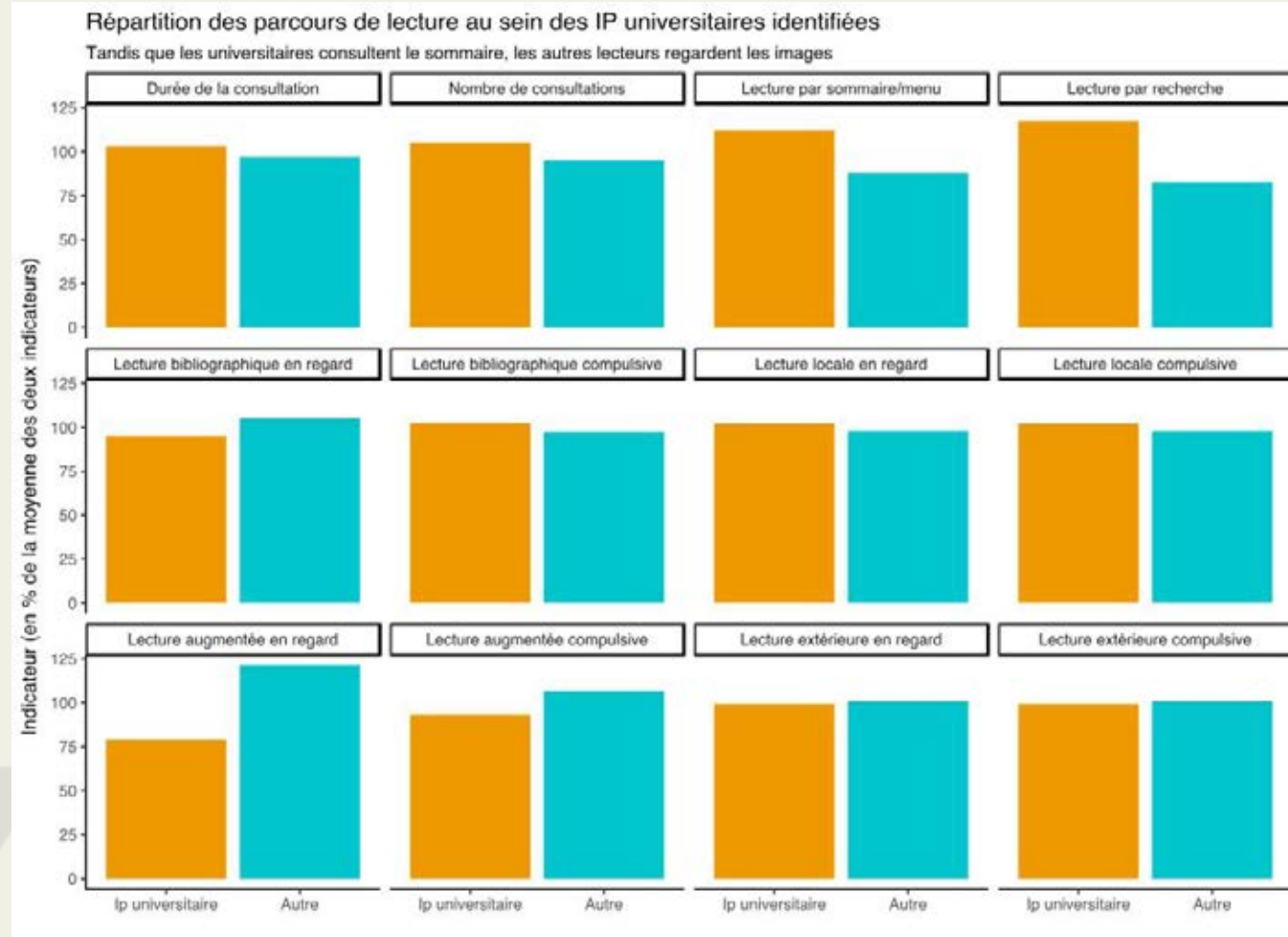
Les IPs universitaires se caractérisent par un mode de consultation “communautaires” qui favorise les publications de leurs institutions



## 2. Résultats obtenus

### *Observation des usages*

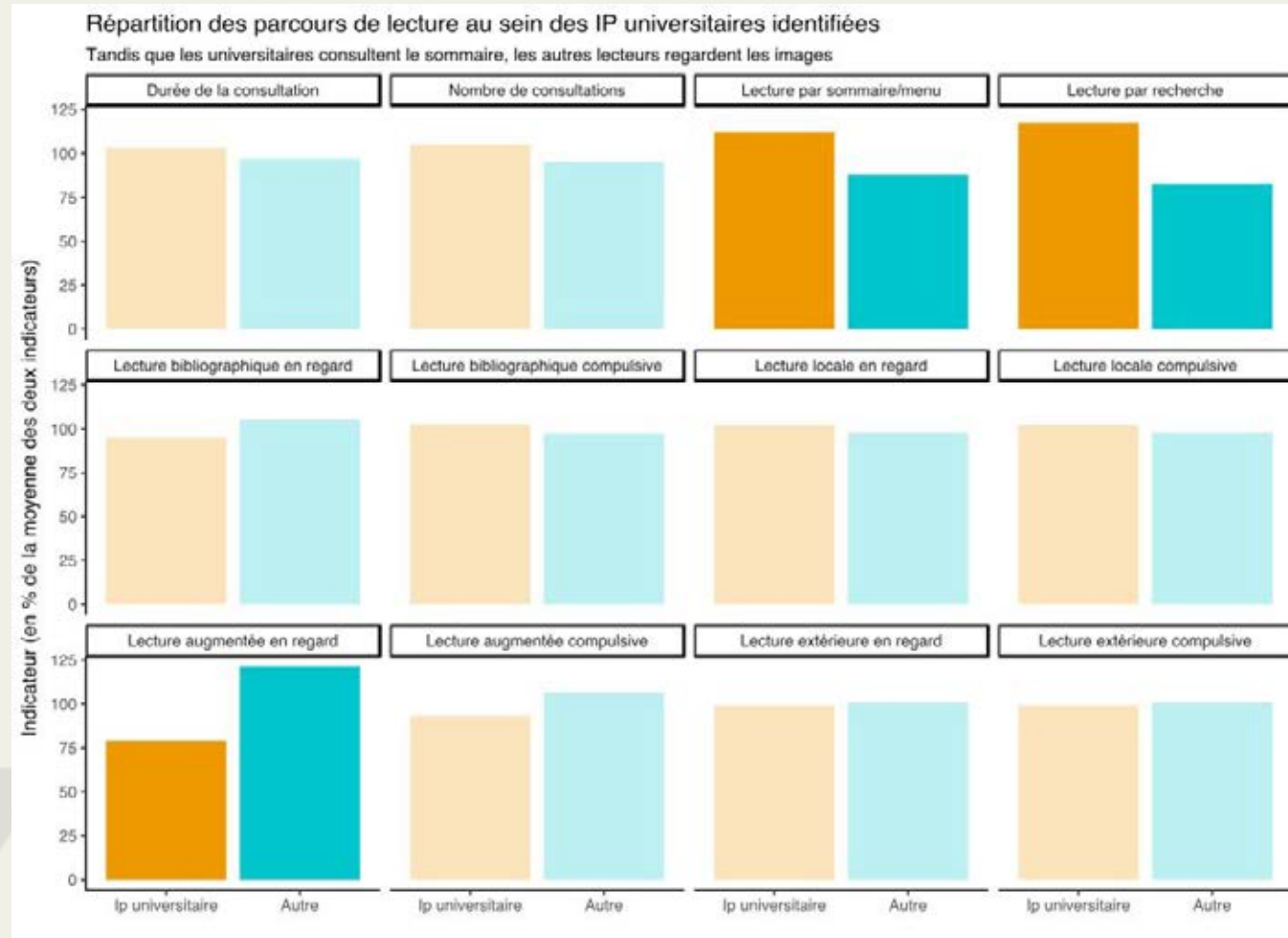
Les pratiques de lecture varient en fonction du profil du lecteur : les IPs universitaire lisent le sommaire, les autres IPs regardent les images



## 2. Résultats obtenus

### *Observation des usages*

Les pratiques de lecture varient en fonction du profil du lecteur : les IPs universitaire lisent le sommaire, les autres IPs regardent les images



## 2. Résultats obtenus

### *Observation des usages*

## Un lecteur institutionnel inattendu : les banques

Parmi les organisations hors recherche, les banques sont très présentes avec plusieurs milliers de sessions ouvertes par la BNP, le Crédit Mutuel, le Crédit Agricole...

585	Institut National de Physique Nucléaire et de Physique des Particules	CNRS	autre établissement scientifique	CNRS	France	308465
132	Centre de Calcul el-Khawarizmi	Centre de Calcul el-Khawarizmi	autre établissement scientifique	autre établissement scientifique	Tunise	210249
78	Bibliothèque Nationale de France	Bibliothèque Nationale de France	établissement culturel	bibliothèque	France	173474
1192	Université Paris-8	Université Paris-8	enseignement supérieur	université française	France	126223
1057	Université de Bordeaux	Université de Bordeaux	enseignement supérieur	université française	France	126035
1085	Université de Lausanne	Université de Lausanne	enseignement supérieur	université étrangère	Suisse	125431
402	École Normale Supérieure de Lyon	École Normale Supérieure	enseignement supérieur	EPSCP	France	123863
1182	Université Paris-1	Université Paris-1	enseignement supérieur	université française	France	116868
342	Direction interministérielle du numérique et du système d'information et de communication de l'État	Direction interministérielle du numérique et du système d'information et de communication de l'État	administration publique	administration nationale	France	116103
1122	Université de Strasbourg	Université de Strasbourg	enseignement supérieur	université française	France	111008
442	État de Vaud	État de Vaud	administration publique	administration nationale	Suisse	110388
292	Crédit Mutuel	Crédit Mutuel	entreprise	banque	France	106174



## 2. Résultats obtenus

### *Observation des usages*

## Un lecteur institutionnel inattendu : les banques

Parmi les organisations hors recherche, les banques sont très présentes avec plusieurs milliers de sessions ouvertes par la BNP, le Crédit Mutuel, le Crédit Agricole...

292	Crédit Mutuel	Crédit Mutuel	entreprise	banque	France	106174
81	BNP Paribas	BNP Paribas	entreprise	banque	France	20299
291	Crédit Agricole	Crédit Agricole	entreprise	banque	France	18004
97	Caisse nationale de l'assurance maladie des travailleurs salariés	Caisse nationale de l'assurance maladie des travailleurs salariés	établissement public	banque	France	12082
95	Caisse d'épargne	Caisse d'épargne	entreprise	banque	France	10002
96	Caisse des Dépôts	Caisse des Dépôts	entreprise	banque	France	4503
98	Caisses Desjardins	Caisses Desjardins	entreprise	banque	Canada	4137
527	HSBC	HSBC	entreprise	banque	France	2546
83	BNP Paribas - Montreuil	BNP Paribas	entreprise	banque	France	1668
893	Société Générale	Société Générale	entreprise	banque	France	1583
82	BNP Paribas Fortis	BNP Paribas Fortis	entreprise	banque	Belgique	1482
68	Banque Populaire	Banque Populaire	entreprise	banque	France	1391
763	Natixis	Natixis	entreprise	banque	France	1196

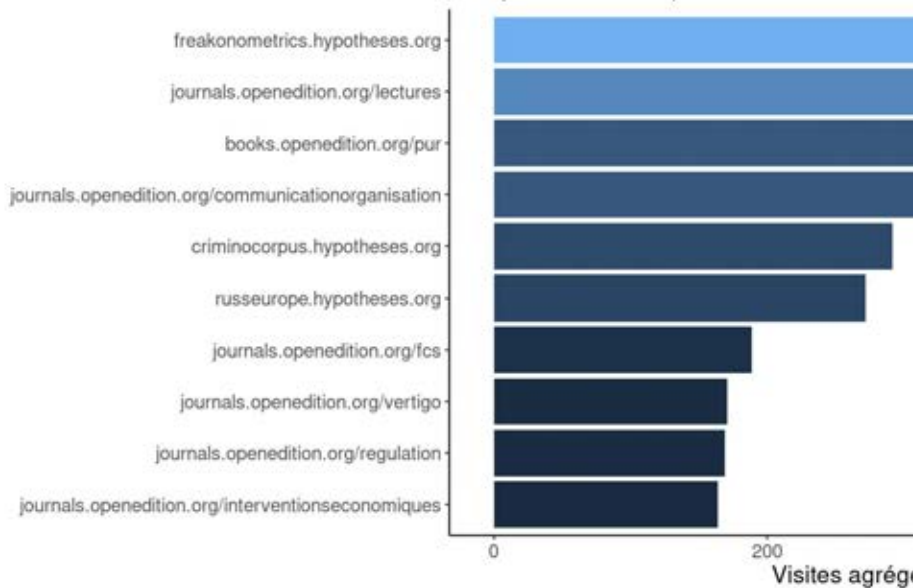
## 2. Résultats obtenus

### *Observation des usages*

## Un lecteur institutionnel inattendu : les banques

Parmi les organisations hors recherche, les banques sont très présentes avec plusieurs milliers de sessions ouvertes par la BNP, le Crédit Mutuel, le Crédit Agricole...

Fréquentation en provenance de BNP Paribas



18 | 2E SEMESTRE / AUTUMN 2015

Contestations monétaires. Une économie politique de la monnaie

Contestations monétaires. Une économie politique de la monnaie

**Les nouvelles monnaies numériques : au-delà de la dématérialisation de la monnaie et de la contestation des banques**

*New Digital Currencies: Beyond Monetary Dematerialisation and the Contestation of Banks*

Alain Laurent et Virginie Monvoisin

p. 1-24

<https://doi.org/10.4000/regulation.11524>

Résumé | Index | Plan | Texte | Bibliographie | Documents annexes | Notes | Illustrations | Citation | Auteurs

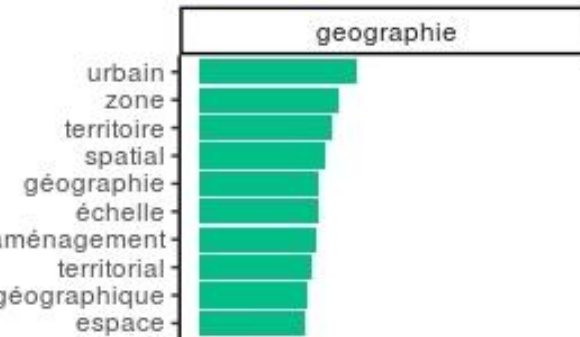
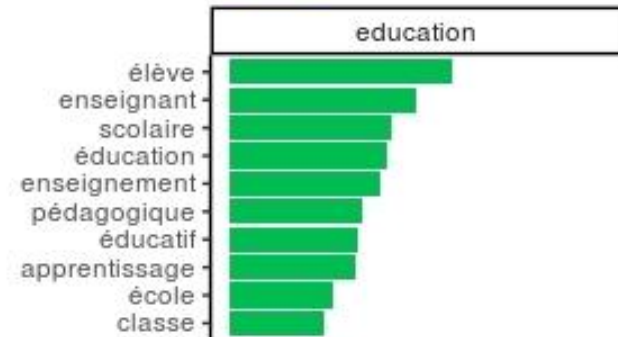
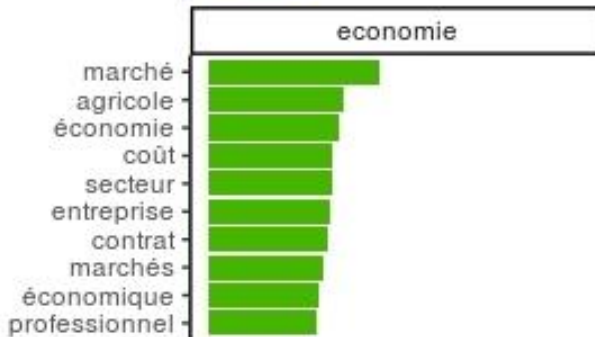
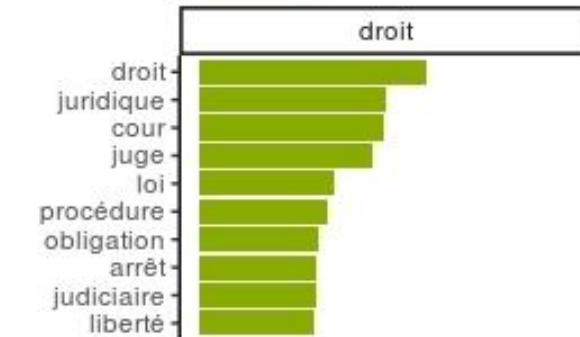
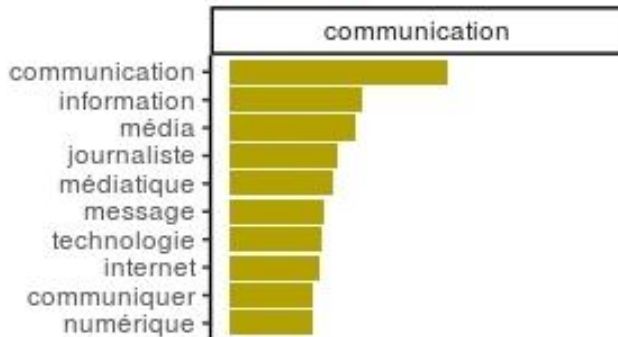
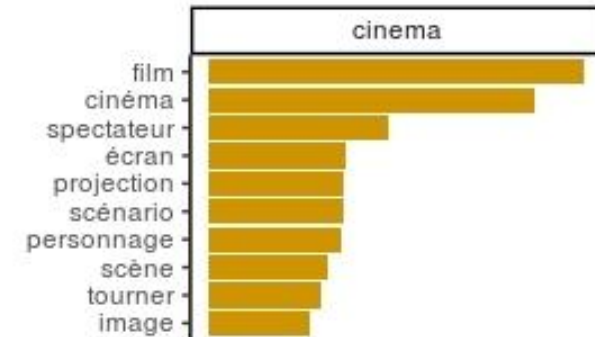
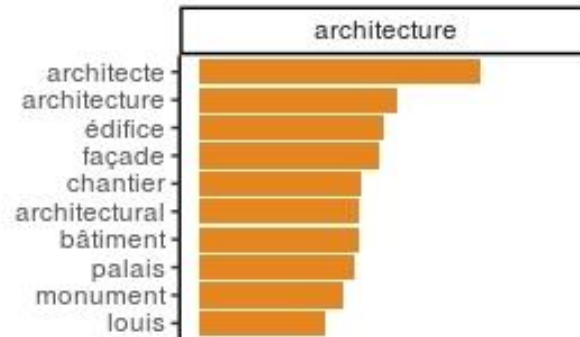
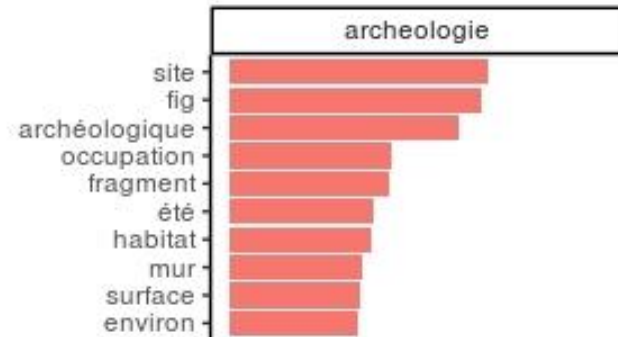


# 3. Les disciplines d'Open Edition

*Classer automatiquement les articles  
pour documenter les visites*

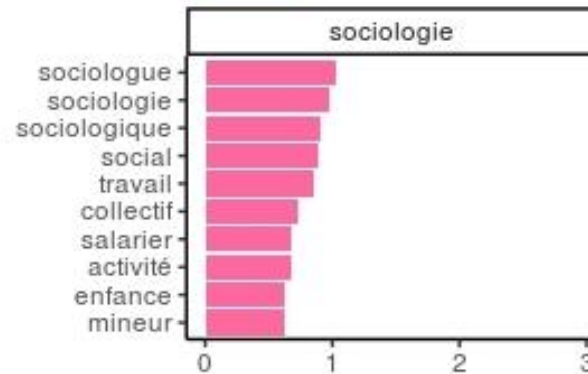
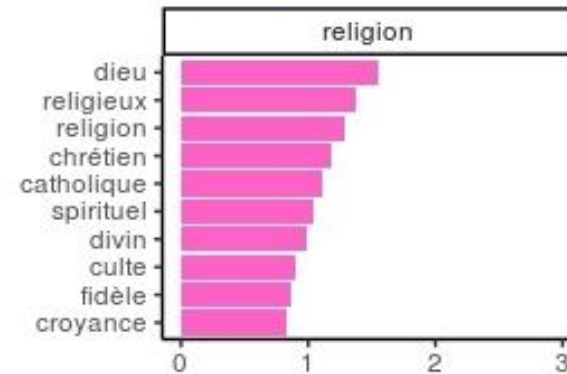
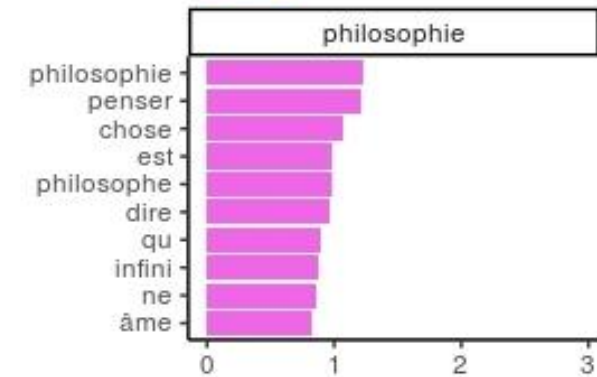
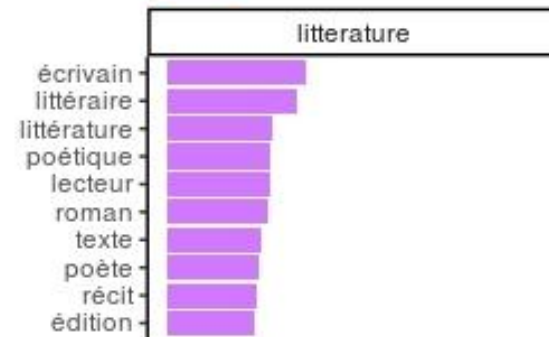
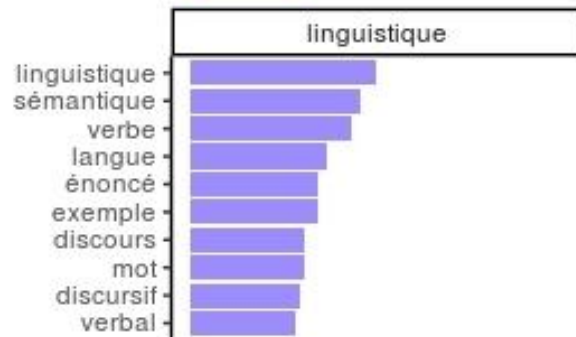
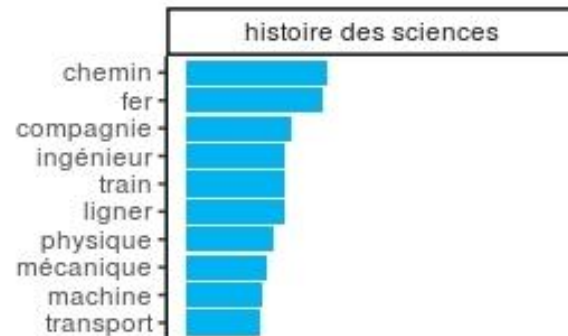
# 3. Les disciplines d'Open Edition

## *Un modèle pour 18 disciplines*



# 3. Les disciplines d'Open Edition

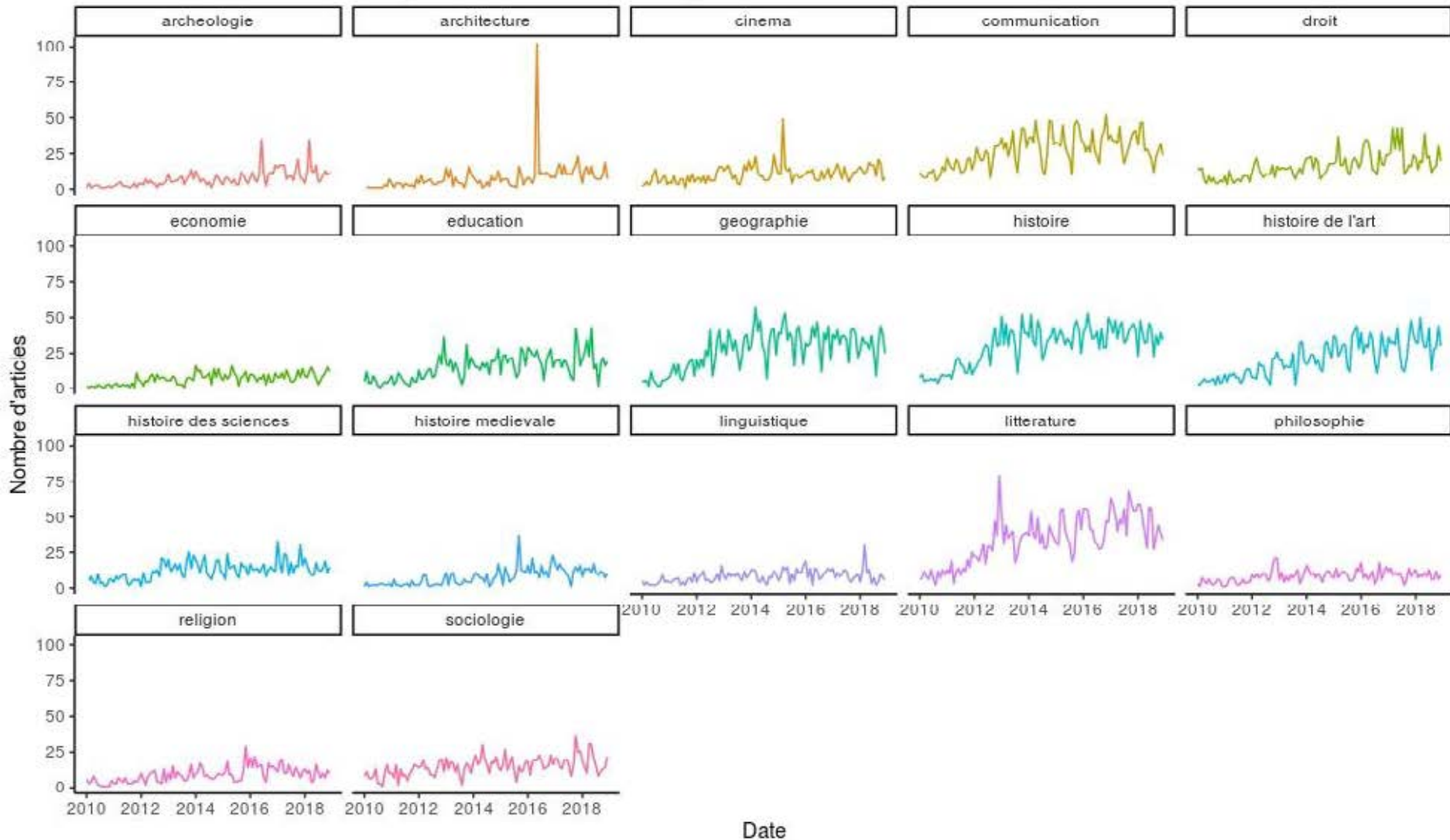
## *Un modèle pour 18 disciplines*



# 3. Les disciplines d'Open Edition

## *Les grandes tendances*

Classification des articles de plus de 1000 mots sur Hypotheses

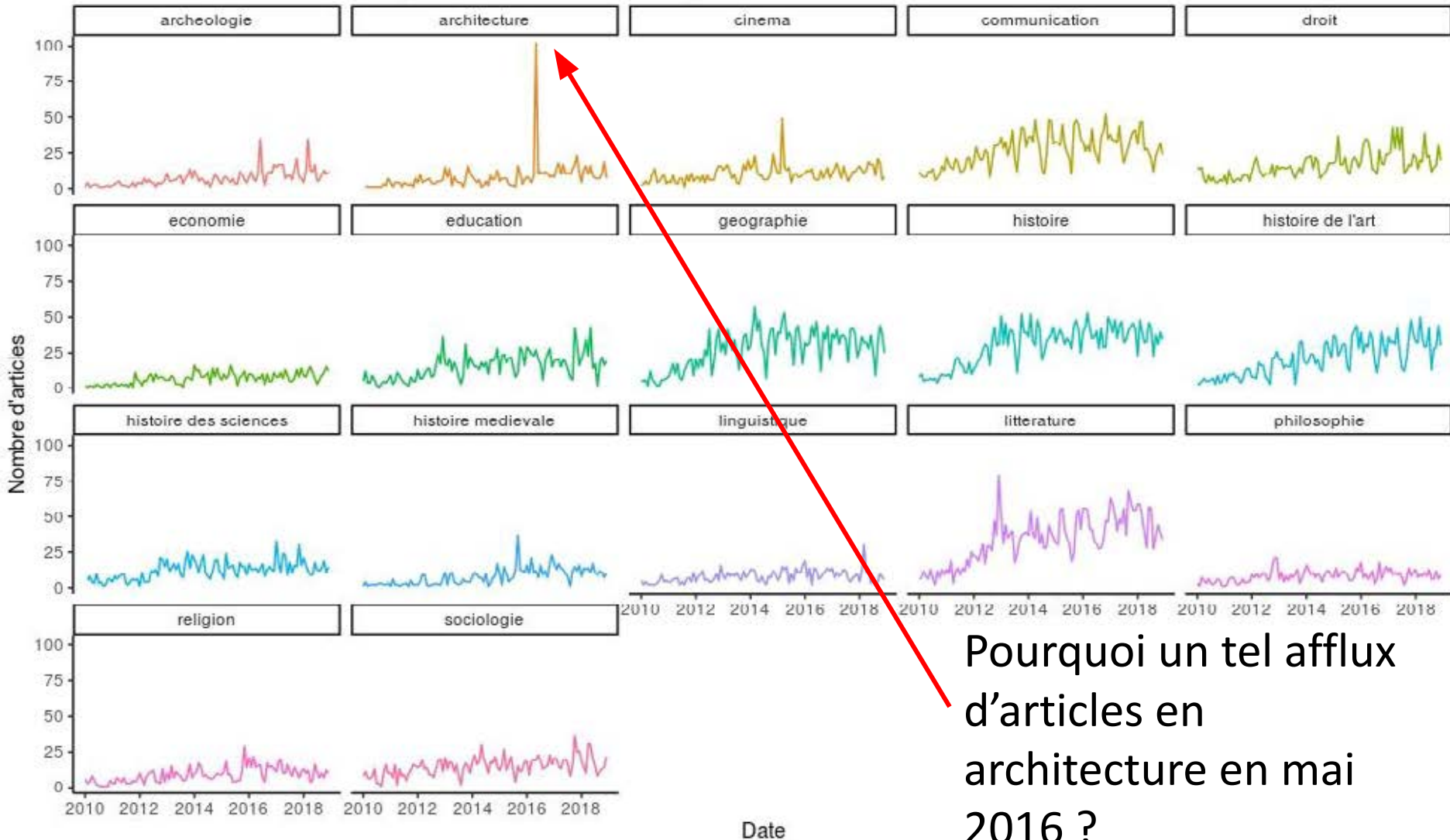




# 3. Les disciplines d'Open Edition

## *Les grandes tendances*

Classification des articles de plus de 1000 mots sur Hypotheses



Pourquoi un tel afflux d'articles en architecture en mai 2016 ?



# 3. Les disciplines d'Open Edition

## *Les grandes tendances*

Après vérification cela correspond à la publication de nombreuses fiches du Centre d'archives d'architecture du XXe siècle (par exemple <https://gtc.hypotheses.org/1108>).

**Guides des sources** Comité d'histoire du ministère de la Culture et de la Communication

À propos Histoire de l'enseignement de l'architecture Grands Travaux culturels

CENTRE D'ARCHIVES D'ARCHITECTURE DU XXE SIÈCLE / FONDS HENSA20

### Fonds Gérard THURNAUER

PAR MPEPIN · PUBLIÉ 19/05/2016 · MIS À JOUR 22/10/2018

Référence	378 Ifa, THUGE
Niveau de description	Fonds d'archives
Intitulé	THURNAUER Gérard (1926-2014)
Dates extrêmes	1957 ; 1980-2007
Nom des principaux producteurs	Gérard Thurnauer
Importance matérielle	16 ml de dossiers, 2 tiroirs de meuble à plans (grands albums), 2,5 ml de tubes de plans roulés, une maquette.
Langue des documents	français
Historique de la conservation	<a href="#">Centre d'archives d'architecture du XXe siècle</a>
Localisation physique	Centre d'archives d'architecture du XXe siècle 127, rue de Tolbiac 75013 Paris



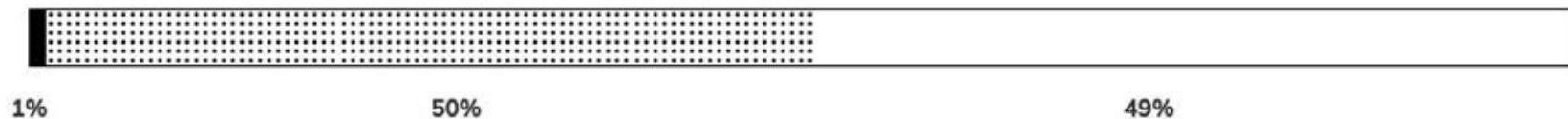
# 4. Perspectives

*Des modalités citationnelles aux  
parcours dirigés...*

### 3. Citations en contexte

## *La nouvelle donne : ouvrir les citations*

How many citations are open today?



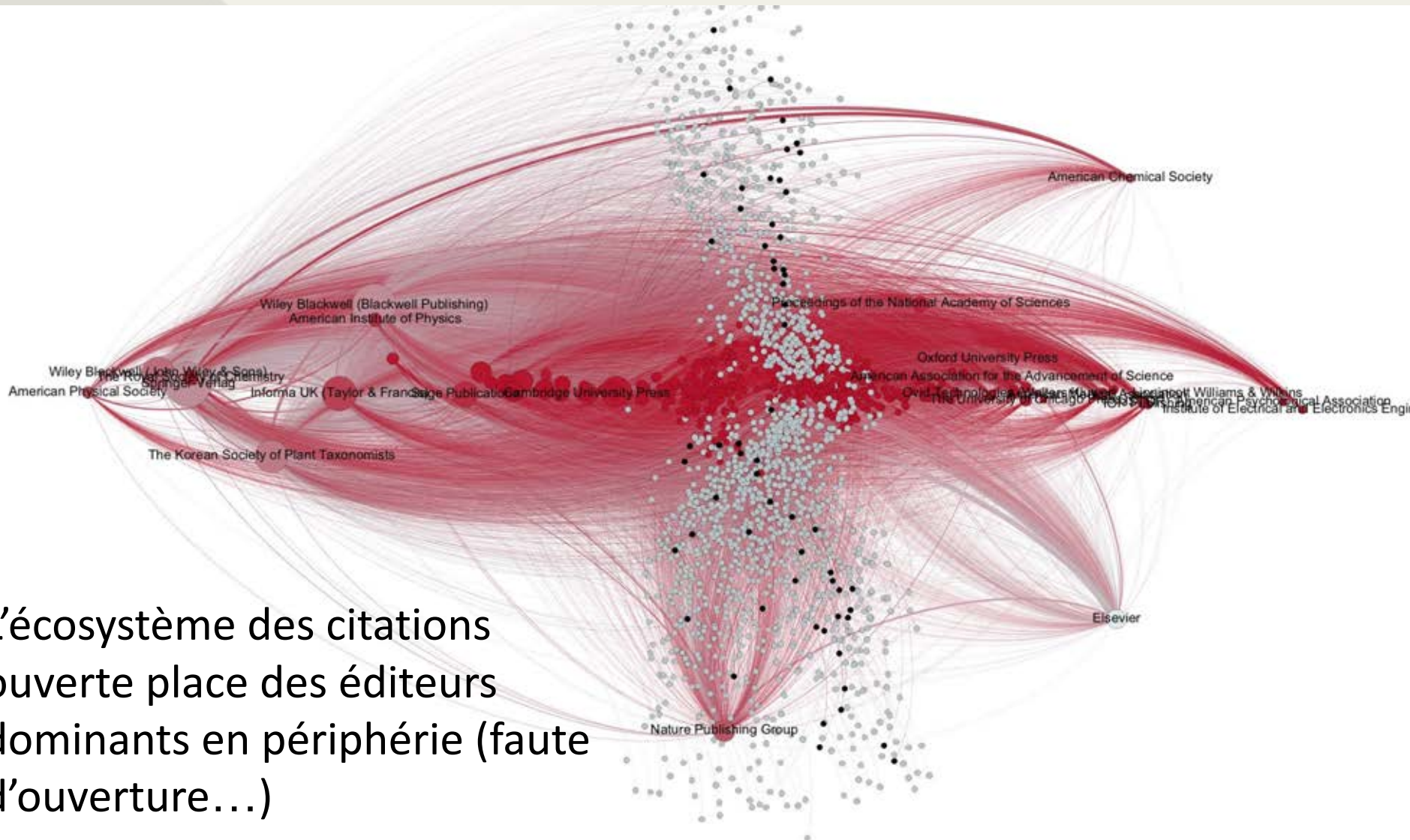
As of January 2018, the fraction of publications with open references has grown from 1% to more than 50% out of 38 million articles with references deposited with Crossref.

We encourage all other scholarly publishers to follow the example of these trail-blazing publishers by making their reference metadata publicly available. Please contact Crossref Support ([support@crossref.org](mailto:support@crossref.org)) for more information, or to let them know that you are ready to open up your reference metadata now. See also our list of responses to [frequently asked questions](#).

Depuis tout juste un an les mesures de la réception scientifique s'ouvrent massivement : il y a aujourd'hui 50% de citations ouvertes sur Crossref.

### 3. Citations en contexte

## *La nouvelle donne : ouvrir les citations*

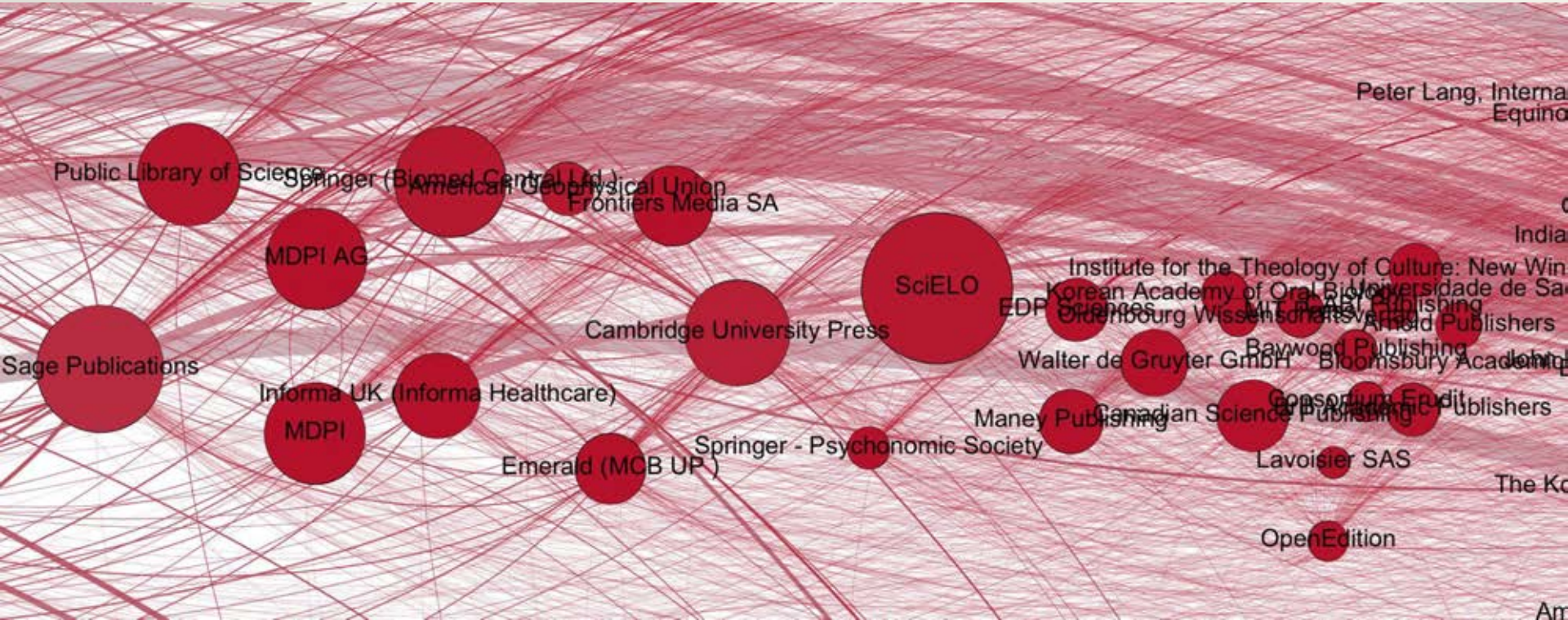


L'écosystème des citations ouverte place des éditeurs dominants en périphérie (faute d'ouverture...)



### 3. Citations en contexte

## *La nouvelle donne : ouvrir les citations*



Et les acteurs ouverts (dont Open Edition) au milieu).

### 3. Citations en contexte

## *La nouvelle donne : ouvrir les citations*



Andrew Piper  
@\_akpiper

Abonné

En réponse à @scott\_bot @TimHitchcock et 2 autres

yes though often the humanities are under-represented. we're trying to change that by doing a citation project that focuses specifically on the humanities. the data annotation is massively laborious though.

Traduire le Tweet

11:32 - 31 oct. 2018

4 J'aime



OpenCitations a retweeté



OpenCitations @opencitations · 26 oct.

The #OpenCitations Corpus has ~14M citations thanks to the #EXCITE Project -> it now contains also bibliographic information from #SocialSciences papers, that can be queried and accessed by using our #SPARQL endpoint, REST API, and free-text search tool [opencitations.net](http://opencitations.net)

Traduire le Tweet

As of October 26, 2018, the OCC has ingested the references from 326743 citing bibliographic resources and contains information about 13964148 citation links to 7565367 cited resources.

Steffen Staab, Philipp Mayr, Silvio Peroni et David Shotton



32



31



Le passage à un paradigme ouvert relance la question de la représentativité des sciences sociales dans les index de citations...



# 3. Citations en contexte

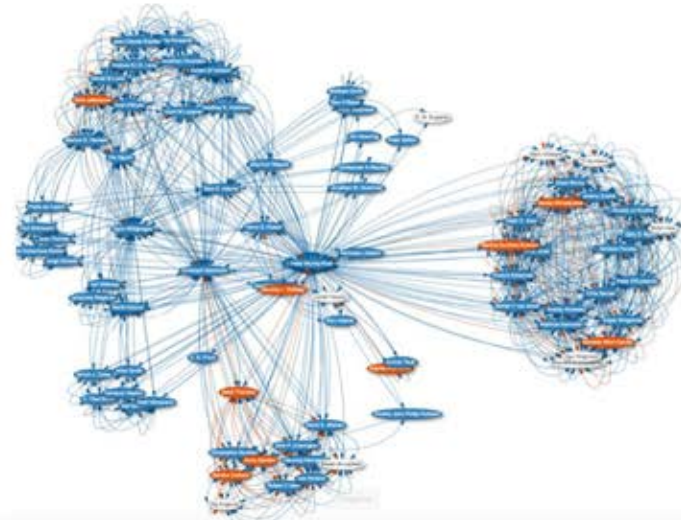
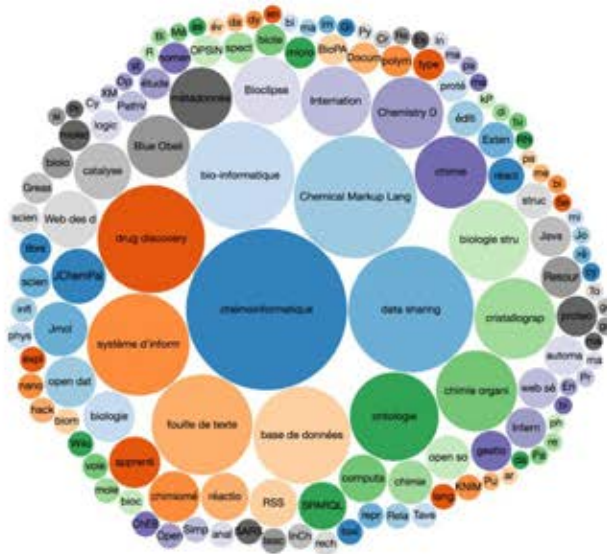
## *La nouvelle donne : ouvrir les citations*

### Citing authors

Authors that cite the author (excluding self citations).

Show 10 entries

Count	Citing author	Orcid
53	<a href="#">Egon Willighagen</a>	0000-0001-7542-0286
31	<a href="#">Christoph Steinbeck</a>	0000-0001-6986-0814
28	<a href="#">Ola Spjuth</a>	0000-0002-8083-2864
25	<a href="#">Jeremy G. Frey</a>	0000-0003-0842-4302
18	<a href="#">Rajarshi Guha</a>	0000-0001-7403-8819
15	<a href="#">Marcus D. Hanwell</a>	0000-0002-5851-5272
13	<a href="#">Jonathan Alvarsson</a>	0000-0002-8682-7208
13	<a href="#">Andreas Bender</a>	0000-0002-8683-7546
12	<a href="#">Michel Dumontier</a>	0000-0003-4727-9435
11	<a href="#">Chris T. Evelo</a>	0000-0002-5301-3142



...et rouvre un débat ancien sur l'usage des données de citation : faire de l'évaluation quantitative à courte vue ou... projeter les graphes de la connaissance.



### 3. Citations en contexte

#### *Garder un lien avec le texte*

1. Paying homage to pioneers
2. Giving credit for related work (homage to peers)
3. Identifying methodology, equipment, etc.
4. Providing background reading
5. Correcting one's own work
6. Correcting the work of others
7. Criticizing previous work

*Can Citation Indexing be automated?* (1964)

8. Substantiating claims
9. Alerting to forthcoming work
10. Providing leads to poorly disseminated, poorly indexed, or uncited work
11. Authenticating data and classes of fact—physical constants, etc.
12. Identifying original publications in which an idea or concept was discussed.
13. Identifying original publication or other work describing an eponymic concept or term as, e.g., Hodgkin's Disease, Pareto's Law, Friedel-Crafts Reaction, etc.
14. Disclaiming work or ideas of others (negative claims)
15. Disputing priority claims of others (negative homage)

Dès 1964, Eugene Garfield souligne les limites de l'extraction brutes de citation qui met sur le même plan des pratiques radicalement distinctes. Il identifie quinze modalités citationnelles.

# 3. Citations en contexte

## *Garder un lien avec le texte*

Table 2 Examples of eve words for type C

Although,	Though,	,although
However,	however, their	however, the
but the	but it	But they
In spite of	Instead of	But instead

Table 3: Examples of cue words for type B

based mainly on	bans	is based on
the basic	used in	uses of
used by	to use a	can use
that can	We can	We use

Nanba & Okumura (1999)

*Toward Multi-paper  
summarization*

Table 9. Content Analysis of Citations (kinds of uses of previous research) in Science Studies, Vol. 1-4 (1971-1974)

Content category	N	%
1. Cited source is mentioned in the introduction or discussion as part of the history and state of the art of the research question under investigation	135	5.8
2. Cited source is the specific point of departure for the research question investigated	14	0.6
3. Cited source contains the concepts, definitions, interpretations used (and pertaining to the discipline of the citing article) <sup>a</sup>	25	1.1
4. Cited source contains the data (pertaining to the discipline of the citing article) which are used sporadically in the citing text	56	2.4
5. Cited source contains the data (pertaining to the discipline of the citing article) which are used for comparative purposes, in tables and statistics	123	5.3
6. Cited source contains data and material (from other disciplines than citing article) which is used sporadically in the citing text, in tables or statistics	—	—
7. Cited source contains the method used	32	1.4
8. Cited source substantiates a statement or assumption, or points to further information	1842	80.0
9. Cited source is positively evaluated	55	2.4
10. Cited source is negatively evaluated	9	0.4
11. Results of citing article prove, verify, substantiate the data or interpretation of cited source	7	0.3
12. Results of citing article disprove, put into question the data as interpretation of cited source	10	0.4
13. Results of citing article furnish a new interpretation/explanation of the data of the cited source	1	0.0

Spiegel-Rösing (1977),

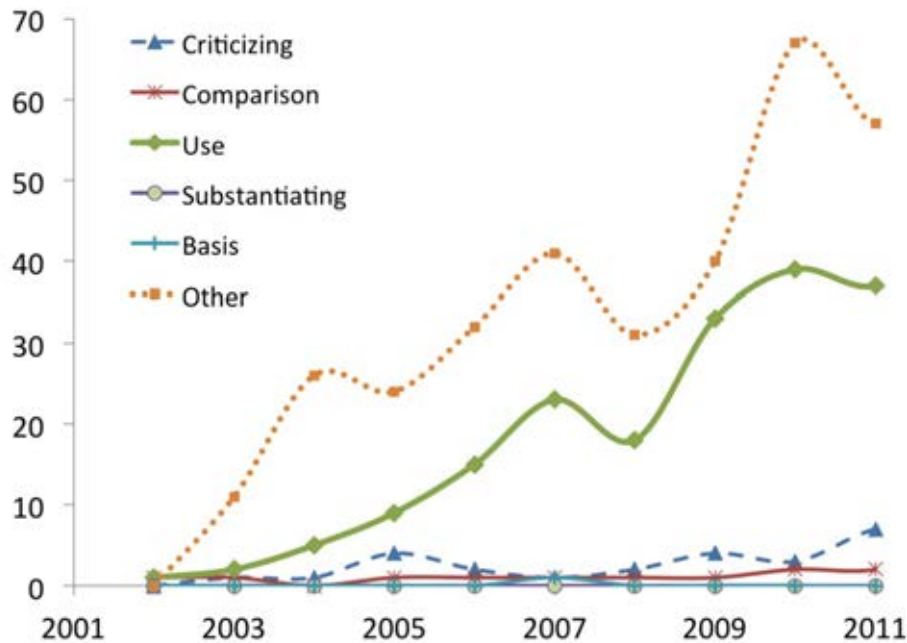
*Bibliometric and content analysis*

Après Garfield, les typologies se succèdent mais les implémentations concrètes sont limités à de petits corpus

# 3. Citations en contexte

## *Garder un lien avec le texte*

*NLP Driven Citation Analysis for Scientometrics*



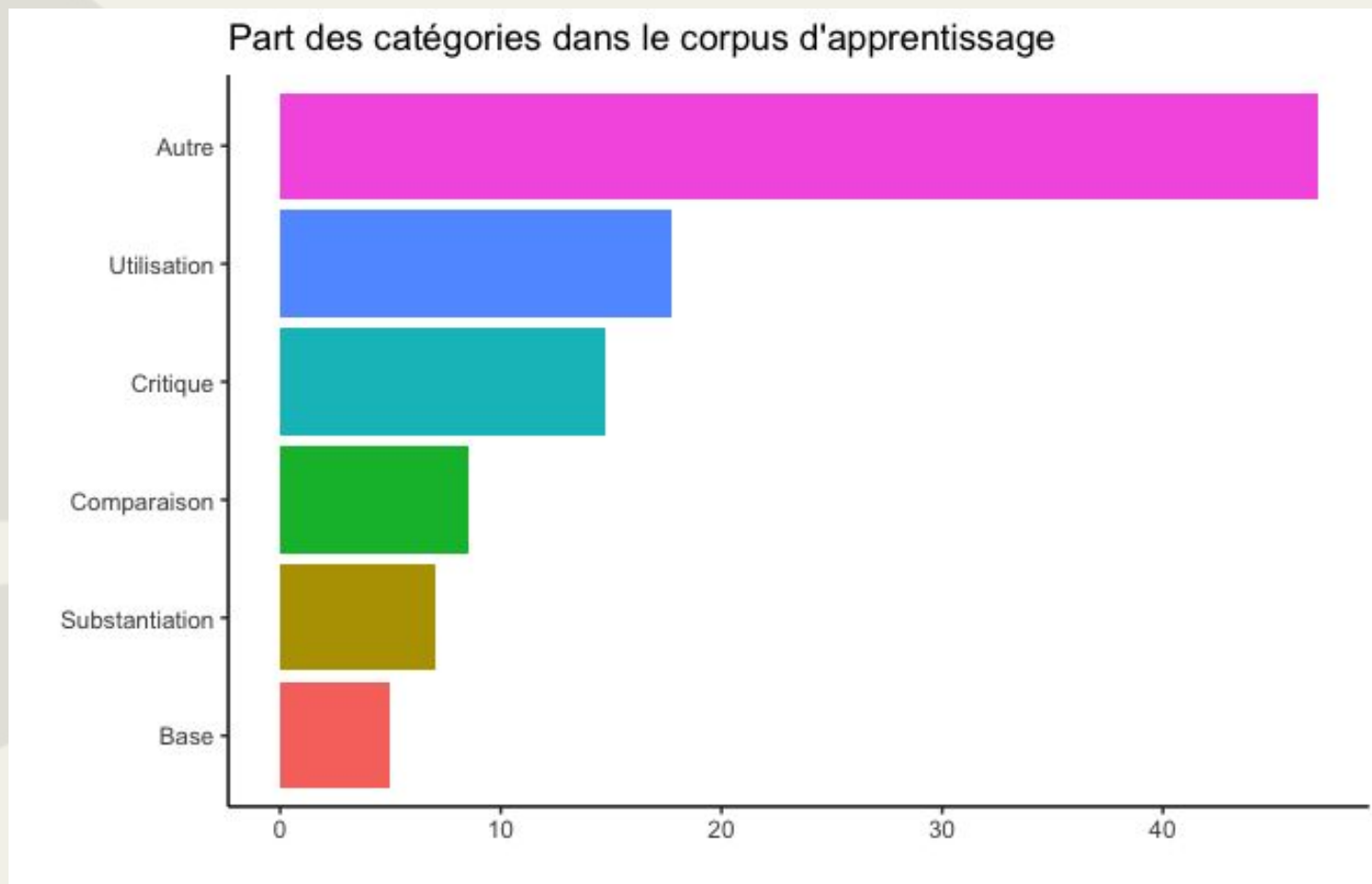
11

L'essor récent des méthodes de *text mining* ouvre la voie à des projets plus ambitieux, comme celui de Jha et al. (2016)

Category	Description	Example
Criticizing	Criticism can be positive or negative. A citing sentence is classified as "criticizing" when it mentions the weakness/strengths of the cited approach, negatively/positively criticizes the cited approach, negatively/positively evaluates the cited source.	Chiang (2005) introduced a constituent feature to reward phrases that match a syntactic tree but did not yield significant improvement.
Comparison	A citing sentence is classified as "comparison" when it compares or contrasts the work in the cited paper to the author's work. It overlaps with the first category when the citing sentence says one approach is not as good as the other approach. In this case we use the first category.	Our approach permits an alternative to minimum error-rate training (MERT; Och, 2003);
Use	A citing sentence is classified as "use" when the citing paper uses the method, idea or tool of the cited paper.	We perform the MERT training (Och, 2003) to tune the optimal feature weights on the development set.
Substantiating	A citing sentence is classified as "substantiating" when the results, claims of the citing work substantiate, verify the cited paper and support each other.	It was found to produce automated scores, which strongly correlate with human judgements about translation fluency (Papineni et al. , 2002).
Basis	A citing sentence is classified as "basis" when the author uses the cited work as starting point or motivation and extends on the cited work.	Our model is derived from the hidden-markov model for word alignment (Vogel et al., 1996; Och and Ney, 2000).
Neutral	A citing sentence is classified as "neutral" when it is a neutral description of the cited work or if it doesn't come under any of the above categories.	The solutions of these problems depend heavily on the quality of the word alignment (Och and Ney, 2000).

### 3. Citations en contexte

## *Garder un lien avec le texte*



La proposition de Jha et al. s'appuie sur une classification supervisée des phrases. Les résultats sont plutôt probants sauf que... 47% des formes de citation ne correspondent pas à la typologie (autres)

# 3. Citations en contexte

## *Rhétoriques des liens sur Hypotheses*

Un premier essai sur *Hypotheses* :

- Constitution d'un **corpus aléatoire de 10000 billets** en français.
- Création des **données lexicales avec *Spacy*** (formes syntaxiques, entités nommées, dépendances, découpage des phrases...)
- **Préservation de l'emplacement des hyperliens** avec les idx. Les références (sous forme de notes de bas de page) sont aussi sauvegardées mais ici l'établissement d'un lien est moins trivial.



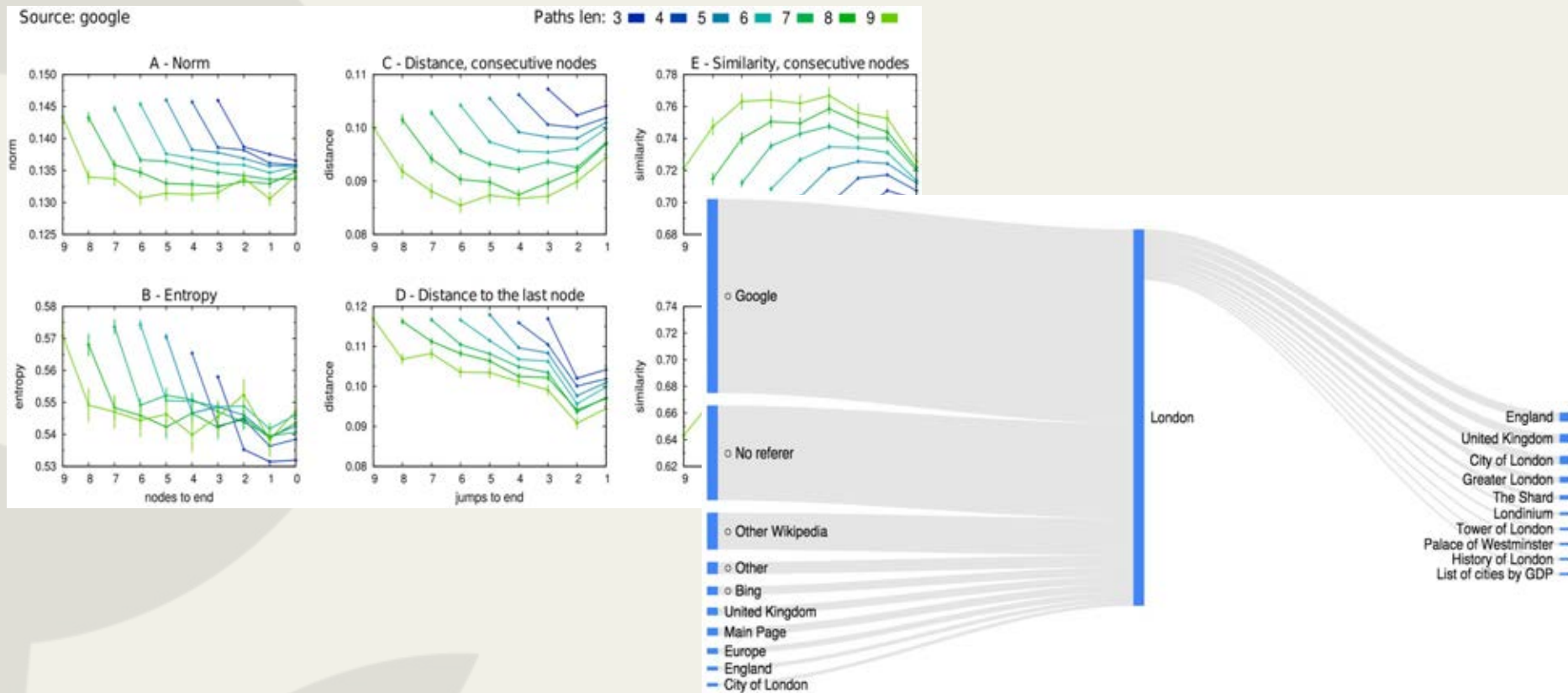
### 3. Citations en contexte

## *Rhétoriques des liens sur Hypotheses*

title	idx	text	lemma	pos	tag	dep	ent
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	104	Oran	Oran	PROPN	PROPN_Gend...	nmod	LOC
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	109	porte	porter	VERB	VERB_Mood=...	ROOT	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	115	désormais	désormais	ADV	ADV__	advmod	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	125	son	son	DET	DET_Number...	nmod:po...	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	129	nom	nom	NOUN	NOUN_Gende...	obj	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	132	,	,	PUNCT	PUNCT__	punct	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	134	à	à	ADP	ADP__	case	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	136	l'	l'	DET	DET_Definite...	det	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	138	initiative	initiative	NOUN	NOUN_Gende...	obl	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	149	de	de	ADP	ADP__	case	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	152	la	le	DET	DET_Definite...	det	NA
<a href="http://arcs.hypotheses.org/1108">http://arcs.hypotheses.org/1108</a>	155	ministre	ministre	NOUN	NOUN_Gende...	nmod	NA

Aperçu des données lexicales de Spacy. Il serait possible des données de ce type sur l'ensemble des corpus d'Open Edition en laissant tourner le scripts pendant plusieurs jours (environ 2 heures pour 10 000 textes).

# 3. Perspective *Contextualisation de la réception*



Le projet Usage Alpha s'inscrit dans un contexte scientifique où la réception est mesurée pour elle-même et non dans une perspective d'évaluation.





# Conclusion

*Ce n'est que le début...*

# Conclusion

- Pour prolonger la réflexion, plusieurs pistes d'étude :
  - Repérer les similarités thématiques dans les parcours
  - Identifier les modalités de la citation à partir des termes qui accompagnent un lien
  - Affiner la typologie des parcours
- Exploitation des résultats dans le cadre du projet Appropriations financés par la fondation A\*MIDEX

# Communication

## Plusieurs communications :

Pierre-Carl Langlais, *Modéliser la lecture - Une approche quantitative de la consultation des revues scientifiques en accès ouvert*, Séminaire DH, EHESS, Juin 2017

<https://drive.google.com/file/d/0B775SZpk-FplbFlzd28xSlJjSXM/view?usp=sharing>

Marin Dacos et al., “The unexpected reader”, *DH2017 Conference keynote* – 8th August 2017

<https://dh2017.adho.org/program/plenieres/?lang=fr>

Présentation de cas emblématiques d’usage inattendu :

<http://lab.hypotheses.org/category/etude-sur-les-usages-des-lecteurs>

Pierre - Carl Langlais, Retours d’expérience : Evolution de la communication scientifique. Journée « Innométriques 2017 », CNRS.

[https://drive.google.com/file/d/1jP-tAM\\_JniE3cPkagx-15S5mQAAdY8sw/view](https://drive.google.com/file/d/1jP-tAM_JniE3cPkagx-15S5mQAAdY8sw/view)

# Communication

## Communications envisagées :

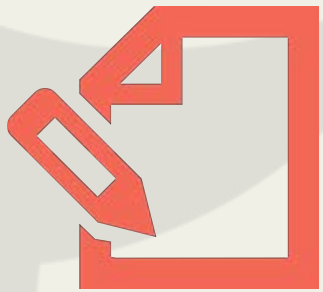
- publication sur le carnet d'OpenEdition Lab (<http://lab.hypotheses.org/>) de deux billets portant sur le prototype de détecteur de lecteurs inattendus et sur les analyses d'usages observés sur la plateforme OpenEdition
- réponse à deux appels à contribution :
  - LIBER's 2018 Annual Conference (<http://libereurope.eu/blog/2017/11/24/liber-2018-call-papers/>)
  - The 22nd International Conference on Electronic Publishing (EIPub 2018) (<http://epress.utsc.utoronto.ca/elpub2018/call-for-papers/>)
- rédaction d'un article scientifique en français et en anglais

## 2. Cartographier les parcours

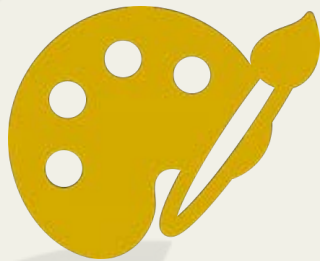
### *Observation des usages*

#### Classifier les parcours

Un préalable : l'étiquetage des formes éditoriales. Les parcours d'Open Edition s'étendent en effet sur plusieurs objets distincts



Article



Fichier  
additionnel



Source



Sommaire

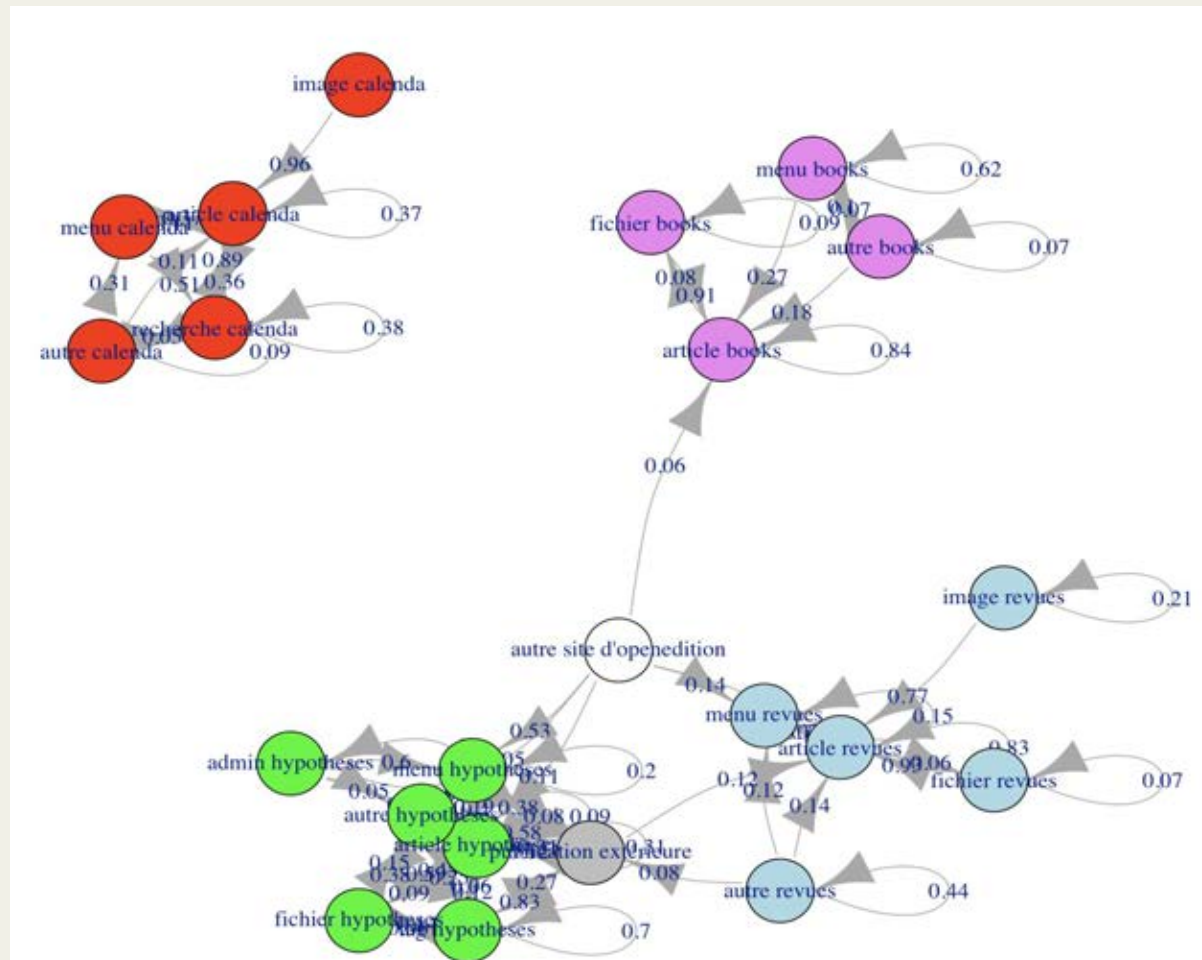


Recherche

## 2. Cartographier les parcours *Observation des usages*

### Classifier les parcours

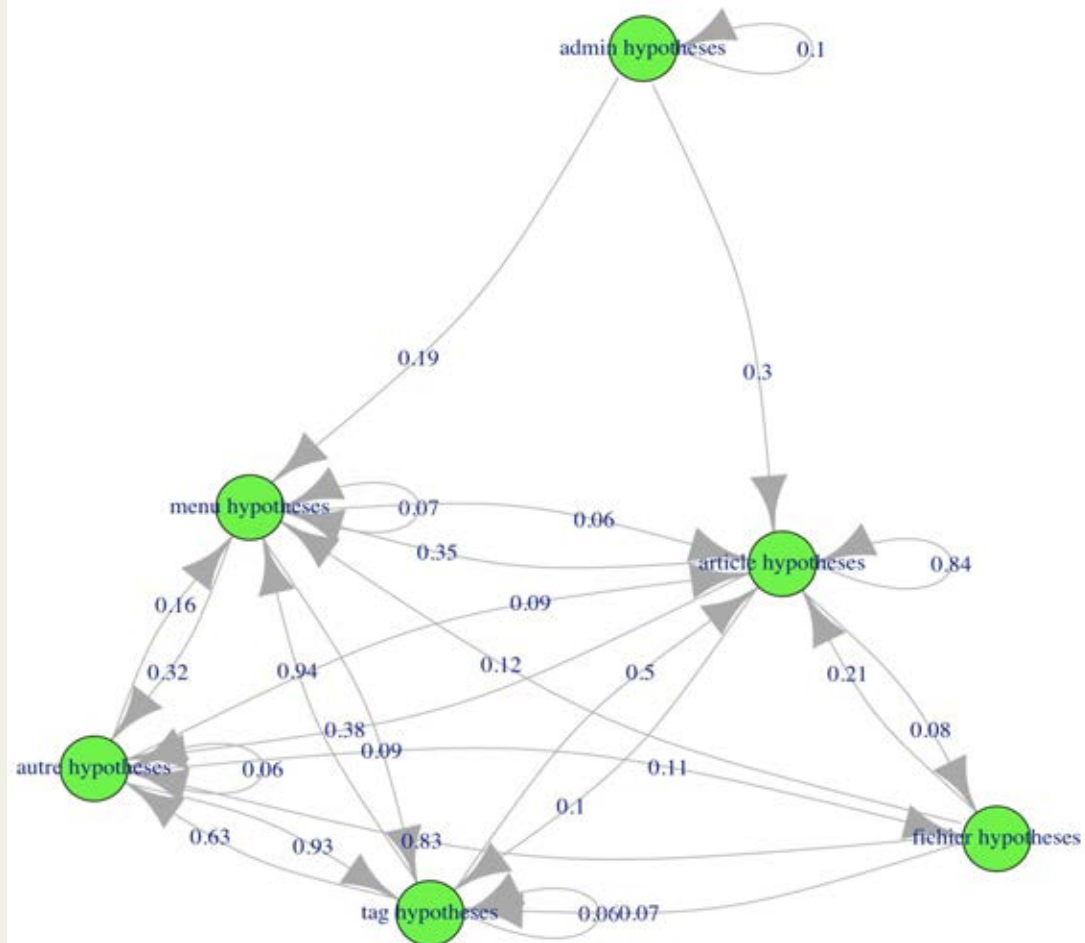
Les formes éditoriales ne s'enchaînent pas de la même manière : il existe des parcours privilégiés (par exemple du sommaire vers l'articles). Le réseau de parcours est réalisé à partir de 200 000 sessions prises au hasard parmi les sessions avec +5 visites



## 2. Cartographier les parcours *Observation des usages*

### Classifier les parcours

Zoom du réseau de parcours sur sa partie la plus touffue : les carnets de recherche d'hypothèses. À la différence de Calenda ou de Books les interactions possibles sont démultipliées.

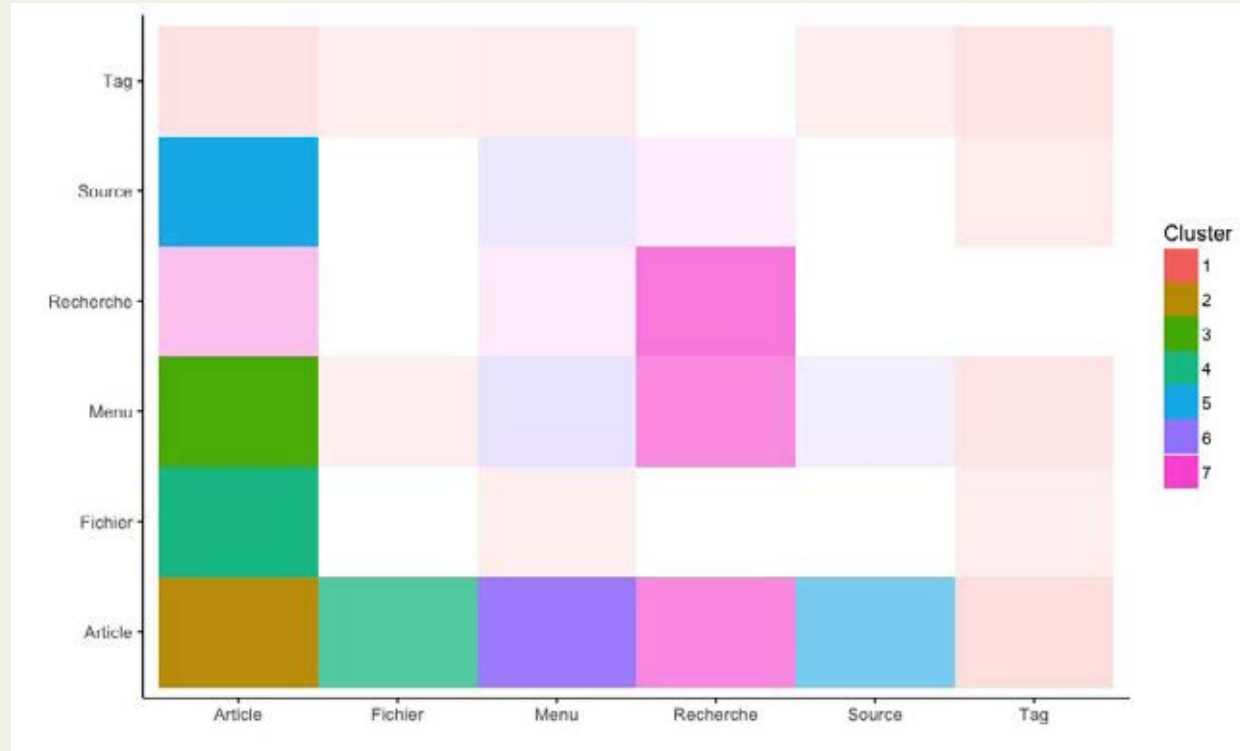




## 2. Cartographier les parcours *Observation des usages*

### Classifier les parcours

Certains enchaînements de formes sont liés : ils correspondent potentiellement à un certain *type* de lecture régulier sur Open Edition.

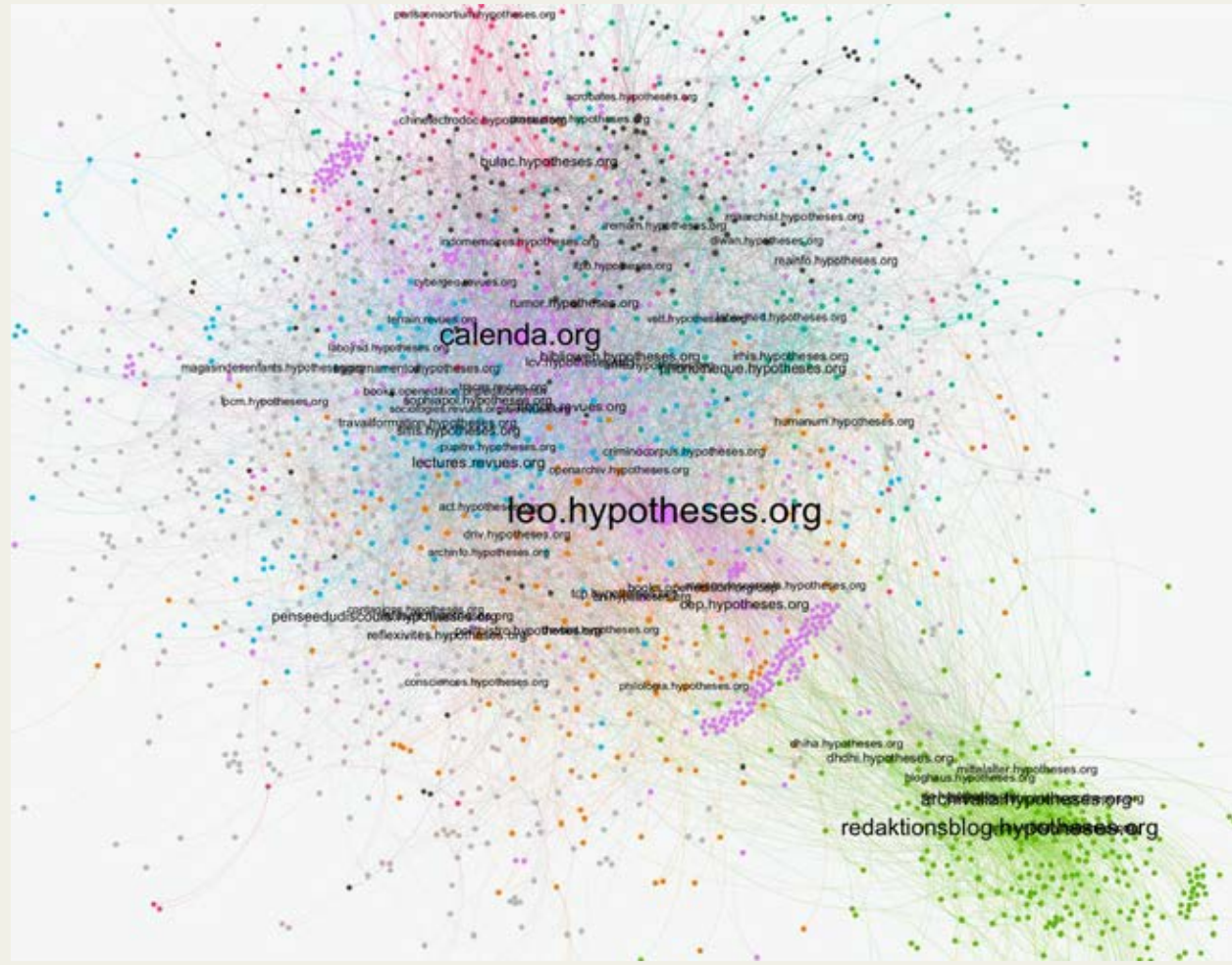


## 2. Résultats obtenus

### *Observation des usages*

#### Classifier les parcours

L'extraction des 800 000 liens présents dans le texte complet des publications d'Open Edition permet d'affiner la classification en distinguant les parcours *anticipés* dans le dispositif lui-même.



## 2. Cartographier les parcours

### *Des consultations en réseau*

#### Une taxinomie des formes de lecture

La classification des parcours repose sur l'identification automatisée de 6 formes de lecture appliquée à

- **La lecture « augmentée »** : ouverture d'images et, dans une moindre mesure, de fichiers intégrés à une publication.
- **La lecture locale** : ouverture de liens présents dans un article conduisant vers un autre article d'une même publication
- **La lecture bibliographique** : ouverture de liens présents dans un publication conduisant vers des publications extérieures
- **La lecture extérieure** : consultation de publications extérieures à OpenEdition qui ne sont pas liées dans la publication.
- **La lecture par sommaire/menu** : utilisation régulière de la page d'accueil de la revue / du carnet au cours de la session.
- **La lecture par recherche** : utilisation régulière des moteurs de recherche au cours de la session.

## 2. Cartographier les parcours

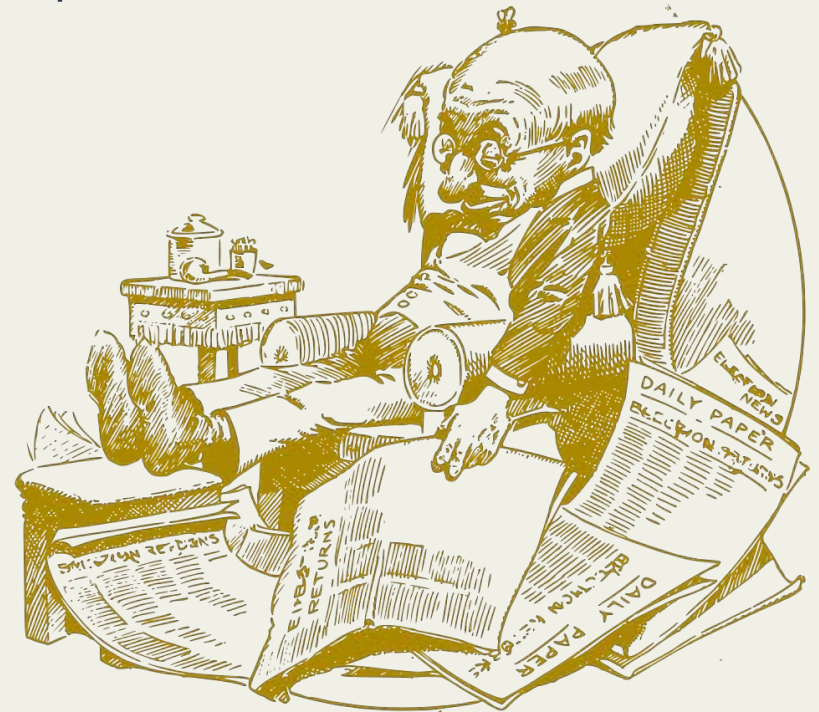
### *Des consultations en réseau*

#### Une taxinomie des formes de lecture

Certaines formes de lectures (augmentée, locale, bibliographique et extérieure) sont subdivisées en deux variantes : en regard et compulsive.



*Lecture en regard*



*Lecture compulsive*



## 2. Cartographier les parcours *Des consultations en réseau*

### Une taxinomie des formes de lecture

Un exemple de lecture en regard : plusieurs dizaines d'aller-retour entre un carnet de recherche sur l'héraldique et un ouvrage de 1861 (!) numérisé sur Gallica sur les armoiries de la noblesse européenne.

**Heraldica**  
nova  
MEDIEVAL AND EARLY MODERN HERALDRY FROM THE PERSPECTIVE OF CULTURAL HISTORY

ABOUT | DISOURCES & TOOLS | FOLLOW

About the Blog | How to Contribute | English | Deutsch | Français | Credits | ISSN: 2199-0816 | Login for Contributors

### Databases & Materials

BY HERALDICA NOWA · PUBLISHED 01/01/2014 · UPDATED 14/06/2017

These databases and dictionaries will help to identify a depicted coat of arms as well as a biased one.

#### Databases

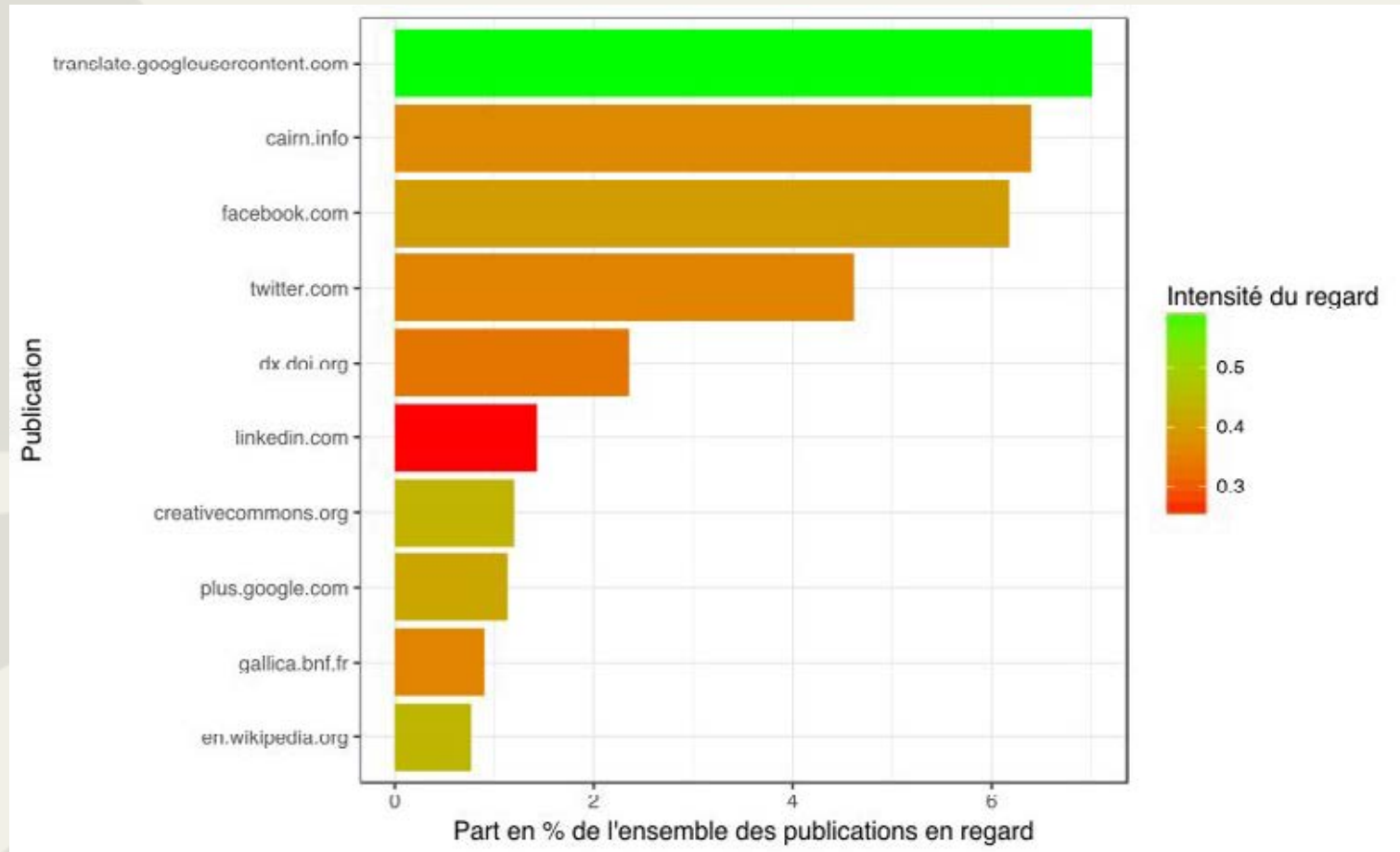
##### Steen Clemmensen, Medieval Armorial

This website is dedicated to armorials. Clemmensen has made an elaborate database of armorials and coats of arms in armorials, which can be downloaded from the website, containing more than 14,000 individual coats of arms. This database enables you, among other things, to identify coats of arms and find out which family bore which coat of arms in the Middle Ages. Furthermore, it presents quality editions of armorials by the author which can be consulted for free.

For more information on this database: <http://heraldica.hypotheses.org/673>.  
Weblink: <http://www.armorial.dk/>



## 2. Cartographier les parcours *Des consultations en réseau*



Les indicateurs de formes de lecture font émerger des pratiques insoupçonnées :  
Google Traduction est fréquemment utilisé pour une lecture en “aller-retour”

## 2. Cartographier les parcours *Des consultations en réseau*

### Une taxinomie des formes de lecture

Un cas de lecture compulsive : ouverture de plusieurs dizaines de lien présents dans un article sur des artistes coréens contemporains.

Art Paris Art Fair 2017 : Artistes coréens, jusqu'au 2 avril

PAR ARTSDELACOREE - 31/03/2017



**Art Paris Art Fair**

Grand Palais

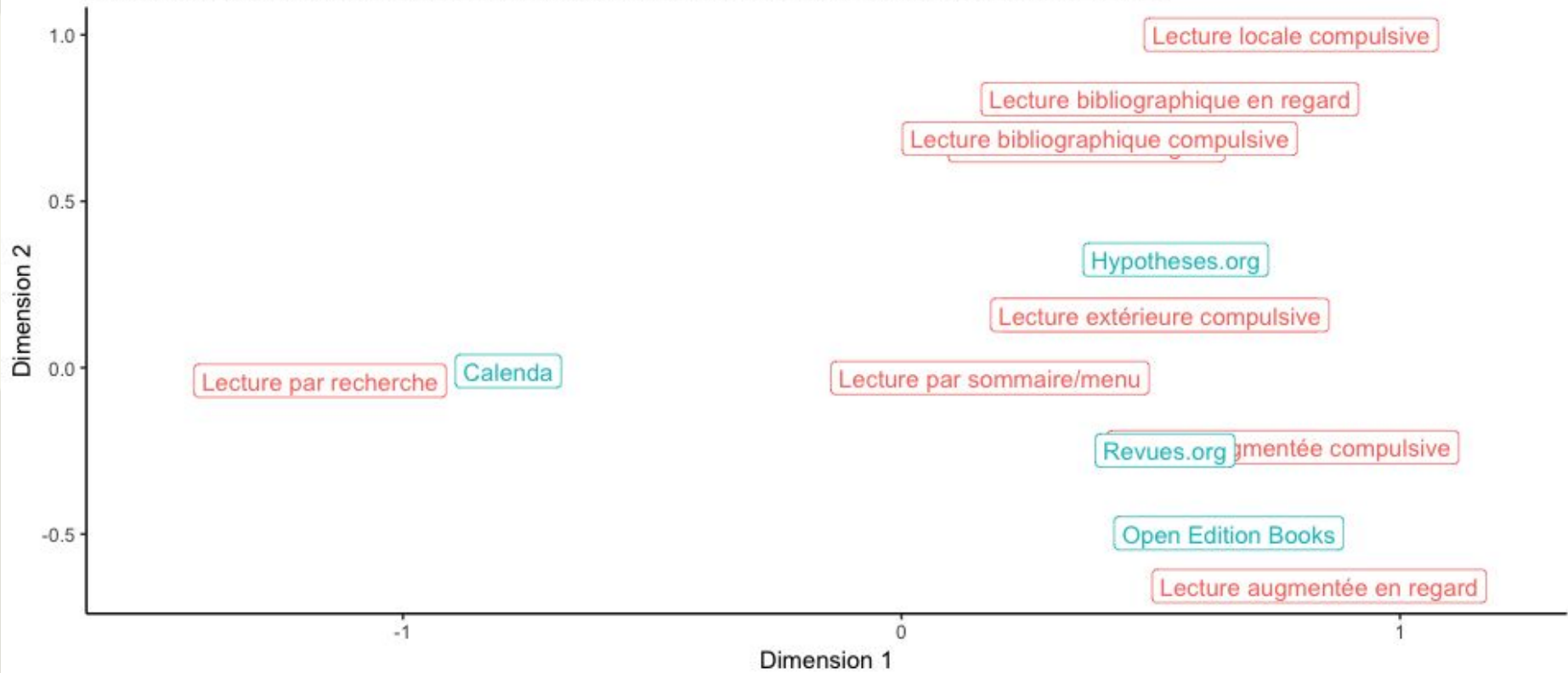


## 2. Résultats obtenus

### *Observation des usages*

#### Analyse de correspondances des parcours d'Open Edition

On lit les références d'Hypotheses, les documents additionnels de Revues et Books et on recherche sur Calenda



Les formes de lecture sont conditionnées par le dispositif éditorial : on lit des sources sur hypotheses ; on regarde des fichiers additionnels sur Revues ; on recherche sur Calenda.