HAL
open science

Chasing data in the Intermediation Era: Economy and Security at stakes

Aurélien Faravelon, Stéphane Frénot, Stéphane Grumbach

To cite this version:

Aurélien Faravelon, Stéphane Frénot, Stéphane Grumbach. Chasing data in the Intermediation Era: Economy and Security at stakes. IEEE Security and Privacy Magazine, 2016, Economics of Cybersecurity, Part 2, 14 (3), pp.22-31. 10.1109/MSP.2016.50 . hal-01107365v2

HAL Id: hal-01107365

https://inria.hal.science/hal-01107365v2

Submitted on 11 Nov 2015

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.

Chasing data in the Intermediation Era
Economy and Security at stakes

Aurélien Faravelon[1], Stéphane Frénot[1] and Stéphane Grumbach *[2]

[1]University of Lyon
[2]INRIA

Abstract

Intermediation is the action to match two types of actors (users, clients, services, etc.) in a world with incomplete information, where the matching would have been difficult without intermediaries. We show the increasing role of on-line intermediation platforms in the economy, and their growing responsibility for ensuring global security of people and society. Our contribution is twofold. At a theoretical level, we better define algorithmic intermediation by establishing a hierarchy of intermediation actors, based on their degree of abstraction from specific services. Then, surveying data from web analytics tools such as Alexa.com, we show that there are only few important intermediation platforms, headquartered in few countries. These results underline the strategic importance of intermediation actors in terms of security and that the more abstract they are, the more influence they reach.

Keywords: Big data, economics, intermediation

1 Introduction

The growth at an unprecedented pace of the production and exchange of data in digital form has lead to the rapid development of new industries. Their services disrupt the old economic models and raise complex social issues, ranging from the protection of individuals and their privacy, to threats to the society as a whole and its security. New equilibria are taking shape between corporations of the digital age that spread their activities worldwide across borders on the one hand, and states that have legal responsibilities and legitimacy over restricted and well-defined territories on the other hand.

In this paper, we show the dominant role of on-line intermediation platforms, and their increasing responsibility for ensuring global security of people and society. Intermediation players are of great importance for their growing impact on the world economy, their fantastic promises changing our lives and our organisations, as well as for the profound disruption they impose on traditional actors in both the private and the public sectors.

Intermediation dates from the pre-digital era. The term comes from the financial sector, and designates the capacity of an institution - such as a bank - to match the funds of depositors with the needs of potential borrowers. The absence of complete information leads to the existence of intermediaries, able to make the match, that single actors would not achieve, while making profit out of successful matches.

Intermediaries connect distinct groups of users, such as borrowers and lenders, following multi-sided economic models, with different pricing techniques. Intermediaries are moreover in a position to create new services such as, very importantly in the banking sector, the mutualisation of risks.

The advent of new technologies for data processing has completely changed the intermediation business. The complexity of financial intermediation, for in-

*Electronic addresses: `firstname.lastname@inria.fr`

stance, has grown since the 1980s, and has lead to the handling of high frequency data flows, and very abstract financial products. The growth of data in the last decade has created avenues for intermediation in an unbounded number of sectors. The first online intermediation service of importance has been developed in the late 1990s. It is the search engine, which intermediates between users and the knowledge they seek, and has revolutionised our access to knowledge. Since then, a powerful industry has emerged with brand new services, such as social networks, which enjoy a remarkable growth, and reaches also sectors historically dominated by local actors.

Data fuels intermediation corporations the way crude oil fuels the traditional industry. Leaders in these sectors are now rivalling with the oil industry as the top capitalisations. On the web, the leading corporations are all in the intermediation business.

Intermediation corporations are not all alike though. Our first contribution is to establish a hierarchy of intermediation actors, divided into four levels, based on their degree of abstraction. *Production actors* are involved in the making of the goods or services they offer, such as media corporations. Then, *distribution actors*, as Amazon or Netflix, essentially provide products produced by others. *Sectorial intermediation* actors, such as LinkedIn or eBay, target specific economic sectors. Finally, *intermediation platforms* offer essential services to support intermediation as well as other services provided by third players. These categories are not exclusive and most corporations have their activities at one or two levels.

We then show that intermediation corporations are ineludible, thanks to their central position in the web. Usually, research focuses on the structure of hypertext links [9, 8] to show that some pages are central. Being so is crucial for business as well as influence. Various types of analysis of the centrality of positions have been made, such as for specific top-level domains for instance [7]. It has also been applied to the study of specific sectors such as hotel sites [13] and the analysis of trackers - i.e., sites which track users on other sites - in order to assess their income [6]. In contrast, we analyse the web from the perspective of visits and show that intermediation actors are central.

Intermediation platforms are revolutionising many economic sectors such as transportation or the hotel business [15]. As they build new services and gather large number of users, they disrupt legacy economic models.

Moreover, as platforms acquire more power, they might acquire dominant positions and weigh on local policies for instance [14]. Sometimes, they conflict with local legislations. The relationship between corporations controlling the new services and states is of increasing importance. Geographical territories, where laws apply in areas limited by boundaries, conflict with "digital territories" which ignore most of these frontiers. Such conflict strongly imparts data security.

The dependency of most countries on foreign platforms raises many issues related to trade, sovereignty, security, or even values. Global issues of data flows and data storage are at the heart of today economic, political and business agendas. Edward Snowden's revelations starting in june 2013 have made the debate on these technical issues accessible to a large audience. Consumers regularly protest against the lack of privacy protection when states pass new laws and corporations adjust their storage strategy according to data policies, left aside financial or fiscal arguments which are beyond the scope of this paper, to install their data centres.

We consider the domains on which platforms and states are either cooperating or conflicting. Conflicts concern in particular labor laws, taxation, money, privacy rules as well as security. Conflicts have led to trials, new laws, as well as interruption of service of platforms over a territory, as was recently the case for Google News in Spain. We then consider the common grounds on which platforms and states are collaborating, which include economic as well as security issues.

The paper is organised as follows. In the next section, we present the methodology of our approach. In Section 3, we define more precisely the concept of intermediation and present a hierarchy of actors based on their abstraction level from specific services. In Section 4, we analyse the relationship between intermediation platforms and countries. Sections 5 and 6 are devoted to respectively the most powerful plat-

forms and countries, demonstrating the extreme concentration of the intermediation power both in the industry and over the geography. Finally, in Section 7, we consider the role of platforms as disruptive operators - especially in the field of security and then we study their relations with states.

2 Methodology

We use data extracted from Alexa.com and Trafficestimate.com[1].

Our sample contains 30 countries, which include the Top 25 countries of the world based on their 2013 Gross Domestic Product (GDP)[2]. Additionally, we have included countries, which figure in the Top 5 GDP in their respective continents, but did not make it to the Top 25 global GDP list, to ensure a sufficient representation of each region of the world. The final list of countries we study contains China (CN), India (IN), Japan (JP), South Korea (KR), Taiwan (TW), Thailand (TH), Indonesia (ID), Australia (AU), United States (US), Brazil (BR), Mexico (MX), Argentina (AR), Colombia (CO), Venezuela (VE), Canada (CA), United Kingdom (GB), Germany (DE), Italy (IT), France (FR), Poland (PL) Spain (ES), Netherlands (NL), Russia (RU), South Africa (ZA), Egypt (EG), Algeria (DZ), Nigeria (NG), Turkey (TR), Iran (IR), Saudi Arabia (SA).

For each country, we consider its Top 25 sites, according to their Alexa rank. Ranks are based on the traffic data provided by users in Alexa's global toolbar panel over a rolling three months period. A site's rank is based on a combined measure of *Unique Visitors* (the number of unique Alexa users who visit a site on a given day) and *Pageviews* (the total number of Alexa user URL requests for a site).

The restriction to the Top 25 sites is rather meaningful. Indeed, according to our computations from Alexa's data, the Top 25 sites represent about half of the traffic of the Top 500 sites in average in most countries we study. The Top 25 sites of each country therefore amounts to a significant part of the data flows and give a good estimate of the global picture. The overall number of sites in the Top 25 in the 30 countries we study, amount to 419 sites. They can be seen as the most influential worldwide.

Throughout our study, we define a *platform* as a set of services - such as mail or social networks. A platform belongs to a *corporation* - such as Google and Facebook - which can possess several platforms. A platform may be distributed over several *sites* - such as google.com, google.fr, etc. Throughout our study, we group the metrics of the different sites belonging to the same platform.

We use the number of visits as an indication of a platform's importance. Indeed, visits testify for users and the possibility for a platform to retrieve data from its users and, in return, develop new services or enhance existing ones or benefits from phenomena such as the network effects to grow bigger. Google Translate, social networks and services such as carpooling websites, for instance, all owe their rise to the data they collect. As they grow bigger, platforms acquire more and more disruption power.

To capture the features of the economics of intermediation, we design metrics, based on the flow of visits to the top actors of the Web. Our goal is to reveal that the higher the abstraction of intermediation activities, the larger the traffic a platform attracts is. These metrics allow to experimentally establish that the importance of intermediation systems follows a power law, with a very rapid decrease of their influence.

Firstly, we investigate the importance of platforms. To do so, we analyse the number of visits to our sample of websites in order to determine the most visited platforms. We also study the relations between platforms by analyzing the upstream to a site - i.e. the visits to a site X from another site Y - and the downstream from a site - i.e. the visits from a site X to another site Y. Eventually, we define the international influence of a platform as the number of countries for which it belongs to the Top 25.

Then, we measure the influence of a country over another. We consider that a country X influences another country Y, if a corporation, headquartered in X, has platforms in the Top 25 of Y. For each country of our sample, we compute the size of its "influence

[1] alexa.com,trafficestimate.com, data were collected on 19 december 2014, we explicitly state when otherwise.

[2] https://www.cia.gov/library/

zone", i.e. the number of countries influenced by a country.

3 The rise of intermediation

Intermediation allows new services that only intermediaries can deploy. Banking, is an example of such services. The use of algorithms to carry intermediation in potentially all economic sectors is new. *Algorithmic intermediation* is being generalised with the explosion of data and the advent of complex data analytics. Its role in the digital economy is already remarkable. It applies to a large spectrum of sectors of activity, with distinct economic models.

Our aim is to understand why intermediaries are so important today. To do so, we build a model of intermediation activities. We propose to distinguish between different types of intermediation according to their relations to the products or services they deal with. These relations strongly influence a platforms's relations with its users.

An *"abstract"* service does not focus on a specific usage. For instance, a social network is an abstract service as it offers a wide set of functionalities not restricted to an economic sector, and it allows to built other functionalities on top of it using its API. In contrast, an online shop provides a limited amount of functionalities.

We distinguish four levels of intermediation activities, ranked according to their degree of abstraction in ascending order (Figure 1). Levels may overlap: a company may develop services in several categories of activities.

At the bottom of the hierarchy, the PRODUCTION level encompasses industries that produce goods or services and sell them online to their customers, essentially with a very restricted form of intermediation between their services and their customers. The press constitutes a good example of such industries, with a direct relationship to their readers.

One level above, DISTRIBUTION corporations commercialise goods produced by others. Netflix is a good example of that level, giving mostly access to cultural products they do not produce but distribute to their customers.

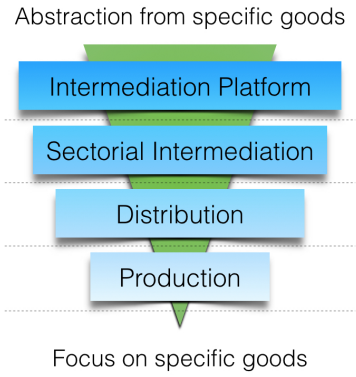

Figure 1: Levels of abstraction of intermediation

The next level, SECTORIAL INTERMEDIATION, includes corporations which provide online services which allow their users to connect with specific goods or services. The search engine belongs to that category together with actors, such as Blogspot or LinkedIn and more generally online dating sites or job sites.

Eventually, the highest level, INTERMEDIATION PLATFORM, is constituted by corporations that offer an ecosystem on top of which others can build and distribute their services. Facebook and Google, for instance, are the most prominent platforms. At this level, corporations offer a sort of global operating system disconnected from physical supports, that allows the development of unbounded types of activities.

Categories may overlap. Amazon, for instance is mostly in the distribution category but operates as well as sectorial intermediation, while Netflix mostly distributes digital contents but also produces some. For PRODUCTION systems, the intermediation activity might be shallow. Transport operators can be seen as intermediating between drivers and clients for instance. But what is of interest, is that although the intermediation might be shallow, there is a possibility of *desintermediation* of the activity, which is what is going on in the transport sector in particular. Companies may challenge traditional transport business models - which focus on production and distribution activities - by performing intermediation. Carpooling

platforms are examples of such challengers.

Corporations at all levels harvest data, which is of strategic importance for their activity. Production systems might get a good knowledge of their users - such as their purchase history or specific tastes. Yet, very sophisticated *recommendation* algorithms, as those successfully deployed by Netflix or Amazon, only appear at the distribution level. For Sectorial intermediation systems, the main challenge is to maintain direct connection with their users, while ensuring the gatekeeping to the services they propose with no middleman. As they do so, they collect a large amount of information from their users. Eventually, intermediation platforms also ensure a direct link to their users, and try to become a universal gate to all services. For that purpose, they facilitate the development of services developed by others, by offering APIs. They thus collect data about their users and about the traffic generated by applications built on top of their services.

Intermediation platforms are the main players on the web as they attract hundreds of millions of users and thus collect the largest amounts of data and meta-data. The Top 25 global sites according to Alexa correspond to 22 distinct platforms. 7 platforms mainly have activities of the intermediation platform level, and 14 whose activities mainly belong to the sectorial intermediation level. Only one has activities which mainly belong to the distribution category, Amazon.

The absence of corporations mostly dedicated to production activities from the Top 25 may be explained by the dependency of their activities on a specific geographic territory. Intermediation corporations are over-represented because they offer so many functionalities that users are likely to need them. Indeed, orders of magnitude separate each level of abstraction of the intermediation in terms of visits. Corporations mainly involved in DISTRIBUTION only attract 1,1 billion montly visits month when SECTORIAL INTERMEDIATION corporations attract 9,4 billions visits and INTERMEDIATION PLATFORMS 15,1 billions of visits. Once more, as intermediation platforms offer a large range of services which can be used for pretty much anything, they are able to attract most visits. Thus, tey also capture most data on the web.

4 Platforms and Countries

Large intermediation systems play a fundamental role in the economy. They offer essential services, much like electricity or water supply, which are used by people as well as corporations to support their own services. Unlike essential utilities, such as energy, telecom, or transportation, which are strongly regulated in all countries by local authorities to ensure their fair distribution over the territory and in the whole population, the new services are provided by multinational corporations and are only starting to get the attention of the legislators.

In the words of Joe Nye, cyber-power is "a set of resources, that relate to the creation, control and communication of electronic and computer-based information - infrastructures, network, software and human skills" [11]. No doubt that platforms which manipulate data and meta-data and possess infrastructures hold cyber-power. As such, they play a central role in cyber-security. We understand cybers-ecurity both as security at a national and international level and security from the viewpoint of individual which may try to protect their data for privacy reasons for instance. The relationship between countries and platforms is of outermost importance when considering issues such as privacy and security. Indeed, both issues are related to the legal framework implemented in each country, while platforms operate on a global basis. Hosting major digital corporations is thus a key factor in the power of a country. The American Patriot Act, for instance, demands that American platforms provide the US government with data no matter where platforms store data.

The rather uneven distribution of platforms in the world leads to an unbalanced geopolitical situation. Figure 2 shows the difference of context in the US, China, Russia, France and the United Kingdom (UK). For each country, the proportion of national actors from the country's Top 25 sites mostly belonging to a specific category (such as Intermediation platform, distribution, etc.) operating on its territory is shown.

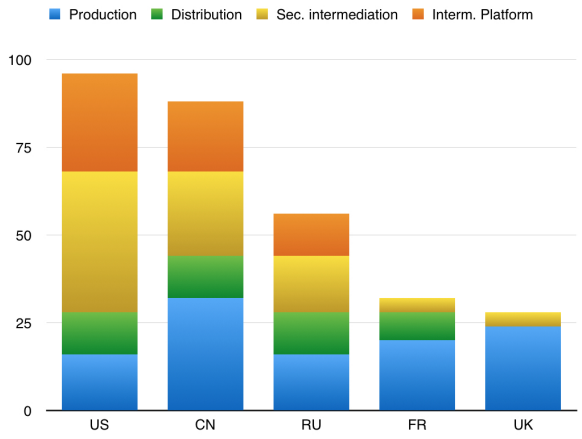

Figure 2: Proportion of national corporations at each intermediation level of the 5 permanent members of the UN Council

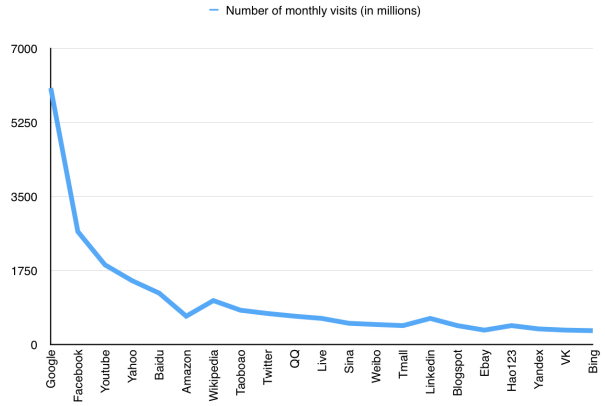

Figure 3: Global traffic of top world corporations

The Production category appears in all national Top 25 list when it was missing from the global Top 25 list. Users, indeed, access online newspapers and buy goods from national retails operators. However, only three countries possess intermediation platforms and, in the UK and France the sectorial intermediation corporations are mostly dedicated to goods selling. The US host most of the sites which their citizens visit, the same stands for China. Russia still hosts most of the corporations which attract its citizens but in a smaller proportion. In contrast, most sites which operate in France and the UK depend on by foreign countries.

This unbalanced distribution of platforms creates new opportunities for surveillance and security enforcement that we investigate in Section 7.

5 Platforms of influence

We now focus on the relations between platforms and countries. We hypothesize that only a few platforms strongly influence the web and that a small set of platforms reach most countries.

Indeed, when we order the web traffic to the platforms in descending order according to their global rank on Alexa, we observe a power law (Figure 3)[3]. There are only few influent platforms: Google comes first, followed by Facebook. The Top 25 platforms attract most of the visits and, most likely, most of the data. They are thus major economic powers. For instance, in 2013, Amazon was larger than the next dozen Internet retailers combined[4].

Yet, there is a strong imbalance between the platforms. Google receives twice as much traffic as any other platform. This observation echoes the domination of Google over the global Top 25 sites.

The international influence also follows a power law. Only 9 platforms belong to 50%, or more, of the top 25 lists we study. From the 10th to the 50th sites, there are only "local" platforms which only influence a small set of countries (4).

Interestingly, the international influence of platforms does not exactly mimic their rank. For instance, Baidu is ranked 5 worldwide on Alexa. However, it only belongs to 4 Top 25 national lists. This situation may be explained by the rank's mode of computation (on Alexa.com, a rank is a combination of average daily visitors and page views over the past three months) and the corporation's international development strategy.

[3]For this Figure, data was extracted on January 6, 2015. We have

[4]http://www.forbes.com/

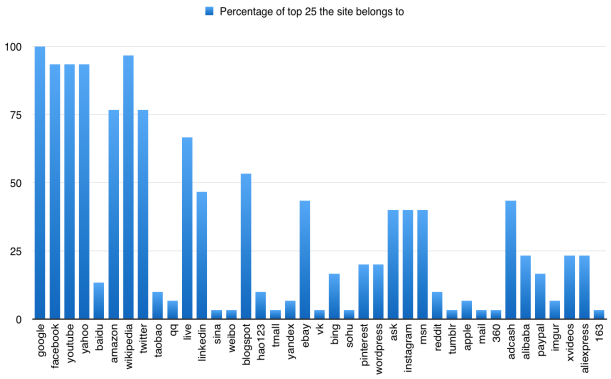

Figure 4: Top 50 sites international influence

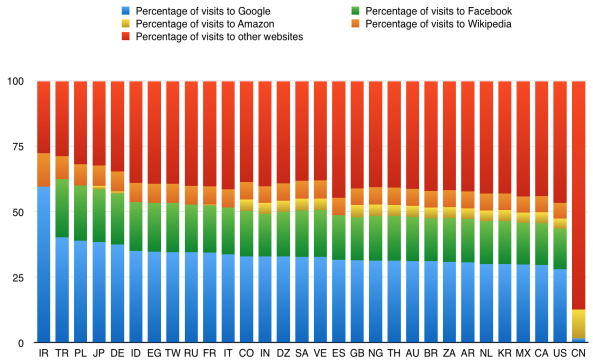

Figure 5: Influence of dominant actors

In all countries but China, Google gets more than a quarter of the traffic dedicated to the Top 25 sites. Google is especially strong in Europe when it is an American corporation. This situation may reflect the absence of alternative search engine in Europe for instance.

At any rate, Figure 4 highlights Google's dominant position. This place partly comes from the global use of Google services. However, not only does Google influence all the countries we study, it also attracts a large part of the traffic to the Top 25 sites of each country (Figure 5). Eventually, local policies may not hinder the development of platforms. Google is officially absent from China but Google.co.hk appears in the Chinese Top 25 sites.

It is customary in Europe to discuss the Influence of the "GAFA", namely Google, Amazon, Facebook and Apple on the digital economy. Our study casts another light on their power. Apple is no major web actor. Yet, Google oversteps the influence of other platforms with a comparable rank. Indeed, Facebook only has half of Google's influence. Amazon does not belong to all the Top 25 lists we study and Wikipedia only has less than 20 percents of Google influence. If we admit that the Top 25 world sites represent half the traffic to the Top 500 sites, Google attracts 10 to 12% of the traffic to the Top 500 world sites.

We have focused so far on the overall visits to platforms. Yet, the distribution of the "upstream" also shows the importance of intermediation platforms. Users surf between them and use them as hubs to other sites and platforms. The platforms thus control the streams of visits and may collect data out of it, through cookies or social tools for instance. As they do so, they may develop new innovative services. The study of the upstream also shows another side of the relationships between countries: locally hosted sites may belong to a country's Top 25 but the stream of visits is still centered around intermediation - and mostly foreign - platforms. This conclusion stands in our sample even though the proportion of foreign websites in the top 25 differs from on country to another. In the US or China, for instance, the upstream is centered around intermediation platforms but most sites are nationale ones. In the rest of the sample, the top 25 is still centered around intermediation but pretty much all intermediation corporations are foreign and most of the time American. France is a good example of such countries.

6 Countries of influence

We have analyzed the extreme concentration of activity over a few intermediation platforms. The same phenomenon happens at the level of geographical territories. Out of the 30 countries we considered, only 11 host an influential platform (Figure 6). The US hosts most of the influential corporations, while the

other countries considered, apart from China, host one influential corporation.

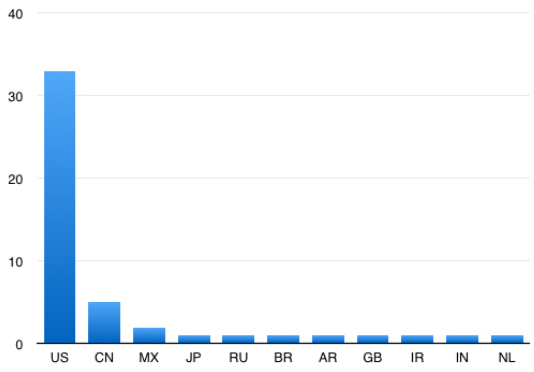

Figure 6: Number of influential platforms by country

By hosting corporations' headquarters and data centres, countries influence other countries, reached by the services of these corporations: the services influence local usages and practices and capture data submitted to the influential country's privacy and security frameworks. In our sample, US platforms reach all the the countries we studied Figure 7). Chinese influential platforms reach 8 countries, while Brazilian ones do so in 5 countries, Argentinian appear ones in 3 countries, British and Mexican ones in 2 countries. The influential sites of the other countries only reach one country but the country where they are headquartered.

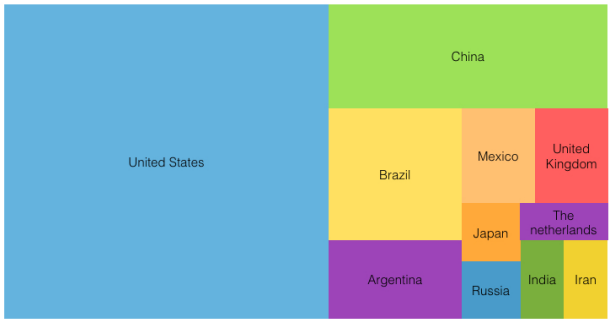

Figure 7: Size of influence zone by countries

We identify three categories of countries. "*Global powers*" influence most countries, as they host a large number of influential platforms. Only the US fits into this category. "*Regional powers*" reach a limited number of countries, generally in a regional or linguistic area. The 10 subsequent countries of Figure 7 belong to this category. Eventually, the 19 other countries we have studied do not host any influential platform, they are "*influenced countries*".

To capture the difference between these categories, let us keep in mind that all Top 25 global platforms are American and consider a particular country, France, an "influenced country", representative of the situation of most European countries.

Most of the platforms belonging to the French top 25 are foreign. The traffic figures are even more impressive, with only 22% of the national traffic on national platforms. Foreign platforms thus capture 78% of the traffic to the Top 25 platforms in France. Out of the 30 countries we considered, most follow a pattern similar to France (Figure 8). Most of the traffic of all countries goes to US sites, about a third to national sites, and a tiny portion to sites of third countries.

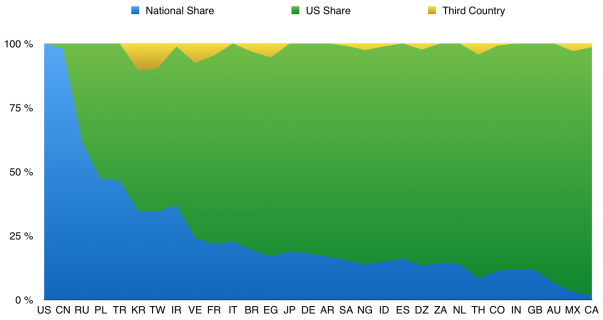

Figure 8: Ratio of sites in the Top 25 of each country headquartered in the US, nationally, or in a third country

7 Platforms as disruptive operators

Platforms alter existing services and create new ones by introducing new intermediaries. For instance, e-commerce platforms are intermediaries between customers and sellers. In the publishing sector, platforms challenge existing actors such as publishers by *disintermediating* them, while becoming intermediaries between service providers - such as editors - and customers. Thus, intermediation can also be seen as the ability to play the role of intermediary and, by doing so, create or alter economic sectors.

As platforms become major intermediaries, they become *basic utilities* while disrupting economic models. For instance, the so-called "sharing economy" disrupts traditional industries and collaboration models [3]. Independent workers have more and more opportunities to work for several platforms such as Amazon Mechanical Turk[5]. This disruption will lead to adjustment in labor laws as the employer/employee relationship evolves and independent workers claim a new status to protect their rights[6].

From a state perspective, platforms possess cyber power as they hold sensitive information and incomparable surveillance power. Indeed, several reports such as [5] show that platforms have become both opponents and partners of governments which try to enforce security through means such as surveillance. We identify three main roles of platforms security wise.

Firstly, platforms may provide useful data or metadata. Several legal frameworks authorize lawful targeted surveillance and most terms of services of platforms include a paragraph which explicitly states that the service will comply with lawful requests. Yet, platforms data may also support mass surveillance or extend a states surveillance power. As the US possesses most top platforms in the world, its reach is thus unprecedented. In particular, meta-data processing allows to retrieve useful sensitive information about people [12]. Some governments thus collaborate with platforms to get their data. For instance, the collaboration between the Chinese government and systems such as QQ is well-known [5].

Secondly, platforms - and especially their commercial interests - may conflict with governmental strategies. In Europe, the data processing and conservations policies of American firms have raised several privacy concerns[7]. In the US, the government has unsuccessfully tried to get data stored abroad by corporations such as Microsoft[8]. Lately, most platforms have decided they would encrypt at least a part of their data to protect them. The US government is strongly opposed to this decision[9]. The Chinese corporation, Xiaomi on the other hand has recently announced it would store its data in the United States out of privacy concerns[10]. As a result, states may implement tools to break in the datasets platforms hold.

Eventually, platform may take up new roles. The "right to be forgotten" [10], for instance, has famously opposed Google and Europe in 2014. A Spanish citizen claimed the right to ask for the delisting of pages which threatened their image. The EU court condemned Google to offer such an opportunity. However, as of today, there is no European actor responsible for the right to be forgotten. On the contrary, Google is in charge on its own and responsible for collecting the delisting claims and processing them.

The relationships between countries and platforms is thus extremely complex, as demonstrated in particular by the Snowden revelations, and it will evolve dramatically in the coming decade. As platforms possess big data and can process them, they are becoming mandatory partners when it comes to cyber power and security. Yet, as we miss a comprehensive international legal framework for instance, civil rights, commercial interests and security imperatives still need to be balanced.

[5] https://www.mturk.com/
[6] http://www.theguardian.com
[7] http://www.europe-v-facebook.org/
[8] http://www.wired.com
[9] http://www.telegraph.co.uk/technology/11139023/US-government-seeks-to-stop-Apple-and-Google-encryption.html
[10] http://www.reuters.com

8 Discussion

We have studied the features of the revolution intermediation introduces in the digital economy. We have shown that it overpowers all other types of activities in terms of importance: it attracts most visits on the web. It is also a concentrated phenomenon: there are only few influent platforms, headquartered in a small number of countries. Our work emphasizes the supremacy of countries such as the United States and China over other areas of the world, such as Europe.

These conclusions lead to a contrasted picture of data security. Influent countries are able to enforce data surveillance and impose their views on data protection. Influent platforms have become both potential partners and opponents in definining of cyber security, be it from an individual point of view (when it comes to privacy for instance) or a state viewpoint.

Our work confirms and extends other inquiries [4, 2]. It also emphasizes the importance of legal frameworks in business strategies: as privacy becomes a competitive advantage, corporations turn to countries with the appropriate security context. Our measures highlight again the importance of the US on the web: they attract most data flows.

No doubt intermediation platforms will set the upcoming political and economic agenda. Several further analyses are needed. First, we have left aside several economic domains of studies, such as tax management in digital activities. We have also ignored several rising types of use, such as mobile applications. Both raise stringent political and economic issues. Uber, for instance, a booming intermediation platform, mostly works on mobile.

Eventually, our metrics should be finer and completed with qualitative measures. We believe that the higher the abstraction level of the platform, the higher the quality of the data harvested. More elaborate measures based on the abstraction level are needed to better quantify and qualify the influence we have diagnosed.

9 Acknowledgments

The authors thank the reviewers for their useful comments as well as Nadeem Anjur and Damien Reimert for their help in developing the tools to collect and analyse the data. This work has been funded by the European project Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities (BYTE).

10 About the authors

Aurélien Faravelon holds a PhD in computer science. His Phd also contains a twist of philosophy as he holds a M.A of philosophy. He is a postdoctoral researcher in the Data at the Core of the Internet (DICE) team at Rhone-Alpes Complex System Institute (IXXI). He currently focuses on data economics and its socio-political consequences.

Stphane Frénot is full professor at Universit de Lyon, works in the CITI Lab at INSA Lyon, and co-manages the INRIA Dice team. His research interests focus on Internet based communication systems. He is specialized in Web software architectures, observes their trends and innovations since 20 years.

Stéphane Grumbach, senior researcher at INRIA, is director of IXXI, the Complex Systems Institute in Lyon. Specialist of data, his research interest focus on the systems handling social data, and their impact on our economies and societies.

References

[1] F. Allen and A. M. Santomero. The theory of financial intermediation. *Journal of Banking & Finance*, 21(11):1461–1485, 1997.

[2] C. Castelluccia, S. Grumbach, and L. Olejnik. Data Harvesting 2.0: from the Visible to the Invisible Web. In *The Twelfth Workshop on the Economics of Information Security*, Washington, DC, United States, June 2013. Allan Friedman.

[3] M. A. Cusumano. How traditional firms must compete in the sharing economy. *Commun. ACM*, 58(1):32–34, Dec. 2014.

[4] M. Falahrastegar, H. Haddadi, S. Uhlig, and R. Mortier. Anatomy of the third-party web tracking ecosystem. *CoRR*, abs/1409.1066, 2014.

[5] A. G. Garcia, C. C. Velasco, E. I. Zamalloa, E. R. Velasco, I. E. Elejabarrieta, J. H. Lotero, J. Mansell, J. J. L. Ibañez, and S. Schuster. Mass surveillance. Technical report, European Parliament, 2014.

[6] P. Gill, V. Erramilli, A. Chaintreau, B. Krishnamurthy, K. Papagiannaki, and P. Rodriguez. Follow the money: understanding economics of online aggregation and advertising. In K. Papagiannaki, P. K. Gummadi, and C. Partridge, editors, *Internet Measurement Conference, IMC'13, Barcelona, Spain, October 23-25, 2013*, pages 141–148. ACM, 2013.

[7] S. A. Hale, T. Yasseri, J. Cowls, E. T. Meyer, R. Schroeder, and H. Margetts. Mapping the UK webspace: Fifteen years of british universities on the web. *CoRR*, abs/1405.2856, 2014.

[8] S. Ihm and V. S. Pai. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 295–312. ACM, 2011.

[9] O. Lehmberg, R. Meusel, and C. Bizer. Graph structure in the web: Aggregated by pay-level domain. In *Proceedings of the 2014 ACM Conference on Web Science*, pages 119–128, New York, NY, USA, 2014. ACM.

[10] A. Mantelero. The eu proposal for a general data protection regulation and the roots of the 'right to be forgotten'. *Computer Law & Security Review*, 29(3):229 – 235, 2013.

[11] J. Nye. *The Regime Complex for Managing Global Cyber Activities*. Harvard Kennedy School, 2014.

[12] T. Pontes, M. A. Vasconcelos, J. M. Almeida, P. Kumaraguru, and V. Almeida. We know where you live: privacy characterization of foursquare behavior. In A. K. Dey, H. Chu, and G. R. Hayes, editors, *The 2012 ACM Conference on Ubiquitous Computing, Ubicomp '12, Pittsburgh, PA, USA, September 5-8, 2012*, pages 898–905. ACM, 2012.

[13] A. Scharlr, K. W. Wöber, and C. Bauer. An integrated approach to measure web site effectiveness in the european hotel industry. *Information Technology & Tourism*, 6(4):257–271, 2003.

[14] A. Themelis. Information and intermediation, abuse of dominance and internet 'neutrality': 'updating' competition policy under the digital single market and the google investigations(?). *European Journal of Law and Technology*, 4(3), 2013.

[15] B. J. Zervas G., Proserpio D. The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Boston U. School of Management Research Paper*, (2013-16), 2013.