UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

MEASUREMENTS AND ANALYSIS OF INDIVIDUAL AND COLLECTIVE ADVERTISING

Autor: Juan Miguel Carrascosa Amigo
Director: Dr. Rubén Cuevas Rumín

DEPARTAMENTO DE INGENIERÍA TELEMÁTICA

Leganés, 27 de Abril de 2016

TESIS DOCTORAL

MEASUREMENTS AND ANALYSIS OF INDIVIDUAL AND COLLECTIVE ADVERTISING

Autor: Juan Miguel Carrascosa Amigo
Director: Dr. Rubén Cuevas Rumín

Firma del tribunal calificador:

Firma:

Presidente:

Vocal:

Secretario:

Calificación:

Leganés, 27 de Abril de 2016

"Marketing is becoming a battle bases more on information than on sales power."

Philip Kotler

Resumen

El problema fundamental del ecosistema de publicidad online es su falta de transparencia, evitando que los usuarios y los anunciantes comprendan adecuadamente su funcionamiento. Esta falta de transparencia puede poner en peligro el crecimiento sostenible de este mercado tan rentable. La comunidad investigadora ha comprendido este problema y recientemente ha comenzado a desarrollar metodologías y herramientas para dar a conocer los aspectos más relevantes de la publicidad online. Los resultados de esta tesis tienen como objetivo contribuir a este esfuerzo colectivo.

En esta tesis se definen dos tipos diferentes de publicidad dentro del contexto de la publicidad online: la publicidad individual y la publicidad colectiva.

(i) En el contexto de esta tesis, hemos definido la publicidad individual como la publicidad que se dirige directamente a un usuario con unas características concretas. Las personas navegan a través de millones y millones de sitios web que conforman Internet. Cada uno de nuestros pasos en Internet, deja una huella única que nos define como un individuo único (por ejemplo, nuestro país, sexo, edad, intereses, gustos, etc.) y que puede ser potencialmente utilizada por el ecosistema de publicidad.

(ii) Por otro lado, la publicidad colectiva se refiere a un grupo de usuarios con características comunes. Por ejemplo, una misma zona geográfica (región, ciudad, país, etc.) o en el mismo rango de edad. Debido a la naturaleza de este tipo de publicidad, el estudio se ha llevado a cabo en una red social concreta, Twitter, y específicamente, hemos analizado la funcionalidad conocida como Trending Topics, al ser una herramienta válida para ser utilizada en los contextos de marketing y publicidad.

En esta tesis se presentan dos metodologías independientes relacionadas con los tipos de publicidad definidas anteriormente. Estas metodologías permitirán dar a conocer aspectos fundamentales de las prácticas utilizadas en el ecosistema de publicidad.

Uno de los problemas clave de la publicidad individual es la privacidad. Como internautas, los sitios que visitamos, los clics que hacemos o los correos electrónicos que enviamos, nos definen a la perfección de cara al ecosistema de publicidad. Por tanto, esta gran cantidad de información hace posible, no solo campañas personalizadas para cada usuario, si no que también incluye un gran riesgo para nuestra privacidad y el uso

de nuestros datos personales. Por tanto, la primera contribución de esta tesis es definir una metodología para analizar el uso actual del Online Behavioural Advertising (OBA) que ha incrementado su importancia como método para aumentar la eficacia de la publicidad online, pero que también tiene un impacto negativo sobre la privacidad del usuario. OBA funciona mediante la asociación de tags o etiquetas a los usuarios en función de su actividad online para luego usar estas etiquetas para realizar campañas dirigidas a ellos. El aumento de esta técnica se ha visto acompañado por una mayor preocupación sobre la privacidad de los usuarios por parte de los investigadores, los reguladores y la prensa. Esta tesis presenta una nueva metodología para medir y comprender OBA en el mercado de la publicidad online. Esta metodología se basa en la creación de personas artificiales con intereses y gustos distintos como "la cocina", "cine", "deportes de motor", etc. A su vez, hemos desarrollado un sistema de medición automático, escalable que permite realizar pruebas con múltiples configuraciones. Como resultado cabe mencionar que hemos observado que OBA es una práctica frecuente y que los intereses mejor valorados por los anunciantes son utilizados con mayor intensidad en las campañas dirigidas.

Por otro lado, en el contexto de la publicidad colectiva, faltan metodologías que permitan cuantificar la eficacia de las actuales herramientas publicitarias colectivas y también realizar estudios comparativos con la eficiencia de los canales de publicidad colectiva tradicionales (anuncios en televisión, radio, periódicos...). En esta tesis se aborda este problema y se propone una metodología para caracterizar la visibilidad de los Trending Topics en varios países y ciudades mediante el uso de métricas basadas en el tiempo de exposición de los Trending Topics y la penetración de Twitter en esas zonas geográficas. Además, comparamos en varios países la visibilidad que proporcionan los Trending Topics y los canales de publicidad tradicionales (por ejemplo, los anuncios en periódicos o en la radio). Este estudio confirma que los Trending Topics ofrecen una visibilidad comparable a los canales de publicidad tradicionales en los países en los que hemos realizado el estudio comparativo. Además, la publicidad colectiva, debido a su naturaleza, permite que la información se difunda fácilmente entre los diferentes usuarios de una red social a través de los comentarios. Esta característica convierte la publicidad colectiva en una pieza de información con un tratamiento diferente a la publicidad individual y, por lo tanto, se requiere un estudio adicional de la propagación de esta información. Con este fin, se ha realizado un estudio sobre los patrones existentes entre países siguiendo un modelo líder-seguidor sobre un conjunto de datos de miles de Trending Topics en decenas de países y cientos de ciudades.

Las conclusiones extraídas de esta tesis han permitido una mejor comprensión de la utilización de la publicidad online en la web, proporcionando una mayor transparencia al ecosistema de publicidad y cuantificando el impacto de las nuevas herramientas publicitarias colectivas.

Abstract

The fundamental problem of the online advertising ecosystem is its lack of transparency, which avoid people or advertisers to properly understand its functionality. This lack of transparency may jeopardize the sustainable growth of this profitable market. The research community has understood this message and has recently started to develop methodologies and tools to unveil relevant aspects of online advertising. The results of this thesis aim to contribute to this collective effort.

We define two different types of advertising, individual advertising and collective advertising, both present in the context of online advertising.

(i) In the context of this thesis, we define individual advertising as the advertising that is targeted directly to a single user. People, as Internet users, browse over the millions and millions of websites that make up the Internet. Each of their single steps, leaves a unique footprint that defines us as a single online individual (e.g. country, sex, age, interests, tastes, etc.) who can be potentially targeted by the advertising ecosystem.

(ii) Collective advertising refers to a group of users with common characteristics. For example, same geographical area (i.e. region, city, country, etc.) or same age range. Due to the nature of this type of advertising, the study has been conducted in an Online Social Network (OSN), Twitter, and specifically we have analyzed the functionality known as Trending Topics due to being a powerful tool used in marketing and advertising contexts.

This thesis presents two independent methodologies in these types of advertising, which help to unveil fundamental aspects of the practices in the advertising ecosystem.

One of the key problems of individual advertising is privacy. As Internet users, the sites we visite, the clicks we do or the emails we send, define us perfectly facing the advertising ecosystem. This wealth of information allows customized campaigns for each user but also includes a huge risk to our privacy and the use of our personal data. Then, the first contribution of this thesis is to define a methodology to analyze the current use of Online Behavioural *targeted* Advertising (OBA) that has risen in prominence as a method to increase the effectiveness of online advertising but also has a negative impact on user privacy. OBA operates by associating tags or labels to users based on their online activity and then using these labels to target them. This rise has been accompanied

by privacy concerns from researchers, regulators and the press. This thesis presents a novel methodology for measuring and understanding OBA in the online advertising market. We rely on training artificial online personas representing behavioural traits like "cooking", "movies", "motor sports", etc. and build a measurement system that is automated, scalable and supports testing of multiple configurations. We observe that OBA is a frequent practice and notice that categories valued more by advertisers are more intensely targeted.

In the context of collective advertising, there exists a lack of methodologies to quantify the efficiency of online collective advertising tools and comparing it with the efficiency of traditional collective advertising channels (e.g., ads in TV, radio, newspapers). In this thesis we address this problem and propose the first methodology to characterize the visibility of Trending Topics in several countries and cities by using metrics that rely on the exposure time of Trending Topics and the penetration of Twitter, we compare the visibility provided by Trending Topics and traditional advertisement channels such as newspapers' ads or radio-stations' commercials for several countries. The study confirms that Trending Topics offer a comparable visibility to the aforementioned traditional advertisement channels in those countries where we have conducted our comparison study. Also, collective advertising, given its nature, allows the information to be spread among different users of the social network via comments or shares. This feature converts the collective advertising in a piece of information with a different treatment comparing to the individual advertising and, therefore, it also requires a study of this spread of information. To this end, we have conducted a study on existing patterns across countries following a leader-follower model over a dataset of thousands of Trending Topics from dozens of countries and hundreds of cities.

The conclusions drawn from this thesis have allowed a better understanding of the use of online advertising on the web, providing more transparency to the advertising ecosystem and quantifying the impact of new collective advertising tools.

Agradecimientos

A mi familia, por su apoyo y cariño incondicional.

A mi padre, por ser mi referente en este largo camino.

A mi madre, por todo su amor.

A mi hermano, por su optimismo y fuerza de voluntad.

A Olga, mi compañera de viaje, por estar siempre cuando lo he necesitado.

También agradecer a la Universidad Carlos III de Madrid, al Departamento de Ingeniería Telemática y, en especial, al grupo de investigación NETCOM, por ayudarme a crecer personal y profesionalmente durante estos últimos años. Me llevo muchas experiencias, grandes compañeros y mejores amigos. En especial gracias a Roberto, Ángel, Gordillo, David, Lisardo, Raquel, Rafa, Grego, Vero, ... sin vosotros, este doctorado no hubiera sido lo mismo.

A Rubén, mi tutor y amigo. Gracias por ofrecerme esta oportunidad y creer en mí todos estos años.

Gracias a todos por estar en los buenos y en los malos momentos y, sobretodo, por ayudarme a superar este reto.

A todos vosotros, GRACIAS.

Contents

List of Figures

List of Tables

Abbreviations

API	**A**pplication **P**rogramming **I**nterface
BAiLP	**B**ehavioural **A**dvertising in **L**anding **P**ages
CPA	**C**ost **P**er **A**cquisition
CPC	**C**ost **P**er **C**lick
CPM	**C**ost **P**er **M**ille
DNT	**D**o **N**ot **T**rack
EM	**E**xpectation **M**aximization
FNR	**F**alse **N**egative **R**ate
FPR	**F**alse **P**ositive **R**ate
GDP	**G**ross **D**omestic **P**roduct
GRP	**G**ross **R**ating **P**oint
HtV	**H**igh **t**emporal **V**isibility
IQR	**I**nter **Q**uartile **R**ange
LtV	**L**ow **t**emporal **V**isibility
MtV	**M**edium **t**emporal **V**isibility
NV	**N**et **V**isibility
OBA	**O**nline **B**ehavioural **A**dvertising
OSN	**O**nline **S**ocial **N**etwork
PV	**P**otential **V**isibility
PoV	**P**otential **o**nline **V**isibility
QoS	**Q**uality **o**f **S**ervice
SIR	**S**usceptible **I**nfected **R**ecovered
TT	**T**rending **T**opic
TTK	**T**argeted **T**raining **K**eywords
WW-TT	**W**orld **W**ide **T**rending **T**opic

Chapter 1

Introduction

Advertising has been around since there exists a need to communicate the existence of products for marketing. One of the most common forms of advertising has been the oral expression but thanks to advances like the printing press by Gutenberg, the industrial revolution, the radio or the television; advertising has been constantly evolving to this day. Nowadays, we could say that the traditional media includes, mainly, newspapers, radio and television. But the explosion of the Internet, as a new player in the advertising ecosystem, has enlarged and changed the rules of the marketing game.

Interactive Advertising Bureau US (IAB US) publishes every year a report including the worth of revenue generated by the Online Advertising. The headline of the latest annual report (2015) was the following one: "US Internet Ad Revenues Reach Record-Breaking $49.5 Billion in 2014, a 16% Increase Over Landmark 2013 Numbers, Marking Fifth Year in a Row of Double-Digit Growth for the Industry" [1]. Similar figures emerge from the European report where IAB Europe mentions that the online ad industry accounts for 1.4 million jobs and €100 billion of Gross Value Added (GVA[1]) [2].

In addition to the economical revenue, it is important to remark the relevance of Online Advertising for those companies in the forefront of innovation such as Facebook and Google which are two of the largest Internet companies by revenue and market capitalization [3]. For these companies, Online Advertising accounts for the vast majority of theirs revenue [4, 5] and their online advertising services (i.e. Google AdWords and Facebook Ads) have evolved into the main source of revenue for these companies.

These previous numbers highlight the impressive growth currently taking place in Online Advertising and the impact it produces on the global economy. But in spite of the current importance that the Online Marketing and Online Advertising has in our global

[1] A standard measure of the contribution to the overall economy similar to GDP.

economy and our lives, we are facing a completely opaque ecosystem with a huge lack of transparency. This lack of transparency along with the ability of online advertising to exploit personal information from users via targeted advertising accentuates the privacy concerns and raises questions and doubts about how companies make use of the personal information they collect. In addition, new services appear over the Internet that are useful for mass marketing campaigns. One of these services are the Trending Topics list. A functionality from Twitter where advertisers have the possibility of "promoting" content during a period of time and (or) for a specific group (e.g. geographical area). This new functionality extends the concept of collective marketing known so far for its use in traditional mass media and increases the collective advertising on the Internet.

Consequently, despite the advantages that the Internet offers in terms of marketing, the complexity and dynamism of current advertising ecosystem requires greater transparency and better understanding of the implications of the use of personal information in targeted advertising. Also there is no knowledge about the impact of the new methods used in collective marketing and comparing their potential of marketing with the traditional one.

In this context, we lack the required methodologies that can help us to better understand specific aspects of the individual and collective advertising. Indeed, building such a methodologies is a very challenging problem since: 1) they required the collection of large amounts of real data from the individual or collective advertising service under analysis; 2) they required to build a measurement system that is automated and scalable and 3) the obtained data may be noisy due to the great level of complexity and dynamisms of these advertising services.

This thesis fills this gap by presenting two independent methodologies addressing fundamental questions in the context of individual and collective advertising. On the one hand, we focus on understanding privacy concerns due to the lack of transparency about personal information from users in the context of the individual advertising. In this scenario, individual advertising is represented by display ads in the web. On the other hand, we present the first methodology able to quantify the potential of marketing in new forms of collective advertising emerging along with Online Social Networks. In particular, we define the first methodology and metrics, which allow to compare the potential of marketing in Trending Topics (an online collective advertising tool) vs marketing campaigns in newspapers or radio stations (traditional collective advertising tools).

The presented methodologies settle the basis and fundaments for the development of others of its kind by the research community contributing to the collective effort to make the online advertising ecosystem more transparent.

Traditional Media	Online Advertising (Internet)	
Newspaper, Radio, TV	**Individual Advertising**	**Collective Advertising**
Broadcasting	Targeted/Personalization	Online Broadcasting
Static content	Dinamic Content	Dinamic Content
No Privacy Concerns	Privacy Concerns	No Privacy Concerns
GRP, CPM...	CPC, CPM...	CPC, CPM...

TABLE 1.1: Summary of main features of advertising in traditional media and online.

Following sections provide a more detailed motivation on the aspects mentioned above showing a more precise vision of the topic and allowing the reader a greater understanding of the following chapters.

1.1 Traditional Media and Online Advertising

The Internet, as a *new* channel of advertising shares common features with the traditional mass media. On the one hand, both present the basic principle of marketing: being able to draw the attention of as many people as possible (i.e. broadcasting). On the other hand, they have the ability to focus on specific groups. One example of these specific groups could be people from the same geographical zone, ads shown in a local newspaper or ads shown in its online version for the traditional media or the online advertising respectively.

In contrast, the dinamism and flexibility that the Internet offers allow to deliver ads with richer content for the users. This flexibility and the big amount of data along the Internet have allowed the emergence of new methods in marketing online such as personalization or targeted advertising, improving their Quality of Service (QoS) while increasing the effectiveness of marketing and, obviously, the benefits of publishers and advertisers.

This new feature of *personalization* in the online advertising is the main difference with traditional media. Taking this in mind the online advertising could be split into two groups: invidiual advertising and collective advertising. Both share the main characteristics from online advertising but shown some differences in terms of scope. For better clarity and understanding, Table 1.1 shows a summary of the main features of individual and collective advertising. It is worth noting that, although this thesis does not pretend to make a comparative analysis of advertising in traditional media and online advertising, we consider relevant to highlight common characteristics and major differences between both to put in context and motivate this thesis. So, information about traditional mass media is also shown in the table.

From Table 1.1, it is important to emphasize the following key aspects:

- Individual advertising is a new type of advertising where personal information becomes more important. **Targeted ads and personalization** are new approaches in marketing online. As Internet users, the *sites* we visite, the *clicks* we do, the *likes* we share, or the *emails* we send, define us perfectly facing the advertising ecosystem. This wealth of information allows customized campaigns for each user but also includes a huge risk to our privacy and the use of our personal data. Therefore, this personalization involves **privacy concerns** about personal data, being part of this data, related to sensitive topics such as our sexual orientation or health. Individual advertising is a new type of marketing where personal information takes importance
- We can observe that the main differences between the traditional media and the collective advertising are minimal in general terms but exhibit important connotations. For example, although both rely on **broadcasting**, traditional media depends mostly on oral communication to spread it around the community while collective advertising, due to the network connectivity that the Internet offers, take advantages of shares, likes or comments from users to spread it. This means that despite having one common goal, their achievement is completely different. Online advertising exploits the ease of the Internet to spread faster than traditional media.
- Related to the preceding item, a clear example is the effect of online advertising in specific environments such as **Online Social Networks** (e.g. Facebook or Twitter). In these environments, the dissemination of any piece of information (e.g. photos, news, ads, etc.) is vastly superior thanks to the ease of sharing and **content dissemination** via likes, shares or comments among users of the social network.

Then, based on these previous key aspects, the analysis and measurements to be conducted in this thesis can be summarize in three main bullets: (i) to understand the use of *targeted* advertising and provide greater understanding of the use of personal data and its privacy concerns as part of the individual advertising analysis; (ii) analyze and understand the visibility offered by online social networks from the collective advertising perspective and (iii) provide a greater understanding of existing information flows on online social networks for specific pieces of information.

Therefore, as we mentioned in previous section, before going into detail on individual advertising (Chapter 3) and later, collective advertising (Chapter 4) and how it spreads

in OSNs (Chapter 5), the next three sections provide a more detailed motivation on the three bullets mentioned above showing a more precise vision of the topic and allowing the reader a greater understanding of the following chapters.

1.2 Privacy Concerns in Online Advertising

Business models around *personal information*, that include monetizing personal information via Internet advertising and e-commerce [6], are behind most free Web services. Information about consumers browsing for products and services is collected, e.g., using tracking cookies, for the purpose of developing tailored advertising and e-marketing offerings (coupons, promotions, recommendations, etc.). While this can be beneficial for driving web innovation, companies, and consumers alike, it also raises several concerns around its privacy implications. There is a fine line between what consumers value and would like to use, and what they consider to be overly intrusive. Crossing this line can induce users to employ blocking software for cookies and advertisements [7–10], or lead to strict regulatory interventions. Indeed, this economics around personal information has all the characteristics of a "Tragedy of the Commons" (see Hardin [11]) in which consumer privacy and trust towards the web and its business models is a shared *commons* that can be over-harvested to the point of destruction.

The discussion about privacy red-lines has just started[2] and is not expected to conclude any time soon. Still, certain tactics, have already gained a taboo status from consumers and regulators, e.g., price discrimination in e-commerce [6, 12–14]. Online Behavioural *targeted* Advertising (OBA, Section 3.3) on sensitive categories like sexual orientation, health, political beliefs etc. [15], or tricks to evade privacy protection mechanisms, like Do-Not-Track signals, are additional tactics that border the tolerance of most users and regulators. The first contribution of this thesis is to build a reliable methodology for detecting, quantifying and characterizing OBA in display advertising on the Web and then use it to check for controversial practices. This methodology is widely explained in Chapter 3.

1.3 Online Social Networks

Online Social Networks (OSNs) in general and Twitter in particular have changed the way in which people communicate, but also have a significant impact on the public image

[2]FTC released in May 2014 a report entitled "Data Brokers – A Call for Transparency and Accountability", whereas the same year the US Senate passed the "The Data Broker Accountability and Transparency Act (DATA Act)".

of celebrities or politicians and are being used by important companies with marketing and/or advertisement purposes [16]. In particular, Twitter has its own business web page[3] and marketing on Twitter has become a business itself [17, 18]. Twitter offers a functionality, that among other uses, is of high relevance in this context named *Trending Topics* (TTs) which are officially described as: "*the hottest emerging topics (or the "most breaking" breaking news), rather than the most popular ones*" [19]. As acknowledged by experts in the field of marketing, surprise is one of the most powerful marketing tools [20]. TTs hold by definition this surprise component and marketing experts have been exploiting it. For instance, TV and radio-station shows have started to announce *hashtags*[4] so that all tweets regarding the show can be aggregated using a hashtag which eventually may become Trending Topic. If that happens it is reported as a big success. Trending Topics have been also used with marketing purposes in politics. For instance, in the last public debate for the Spanish presidency in 2011, one of the candidates became TT as a result of an orchestrated operation by his party supporters. This was used as an unequivocal proof by his party and by several media that he had won the debate [21]. In addition, some social movements such as the "occupy" movements augmented their visibility among the population after becoming TT [22]. Furthermore, the commercial interest of Trending Topics for companies is reflected by the *Promoted Trending Topics* service offered by Twitter[5]. These are a special type of TTs that can be purchased in slots of 24 hours for around $200K [23]. This service is regularly used by companies in the context of advertisement and marketing campaigns.
Finally, another symptom of the relevance of TTs is the recent movement made by Facebook to implement its own Trending Topics service that is currently available for users in United States [24].

However, to the best of our knowledge, this (seemingly) common idea that TTs are a useful tool in marketing contexts is not supported by any scientific or technical work. We believe that a solid scientific basis is required to allow experts in different disciplines to make informed decisions regarding the actual impact that TTs may have in marketing, advertisement, and related contexts. This thesis constitutes a first effort in that direction in which we perform a thorough analysis of the actual *visibility* provided by TTs in Chapter 4.

[3] https://business.twitter.com/marketing-twitter
[4] A hashtag is a special type of word that starts by the symbol #. It is a common practice that people tweeting about a common topic use a common hashtag to identify it.
[5] https://business.twitter.com/products/promoted-trends

1.4 Online Social Networks: Information Flow

Since the existence of online social media, citizens around the world use it to communicate beyond mass media blackouts. For example, IRC channels served as a way for individuals to report news in 1991 during the media blocks in the Soviet Union coup de etat and in the Gulf War[6]. The growth of social media use in developed societies allowed individuals to take one step further, organizing actions and spreading relevant information around their environment. One example of such emergence of coordination away from mass media are the actions of the *Anonymous* group [25], which in 2008 organized demonstrations and produced reports against the Church of Scientology. More recently, the widespread adoption of social media around the world has triggered events that were reported by individuals beyond media blockages, including actions of social movements like the Spanish *Indignados* [26], the Gezi protests in Turkey [27], and the revolutions during the Arab Spring [28].

While social media are clearly relevant for news and culture, there are still many open questions about the potential, role, limitations, and biases of online social media. Social media overcome some limitations of traditional mass media that are commonly attributed as sources of biases. First, the cost to set up an information channel in social media is negligible, allowing individual users to become information channels themselves. This overcomes the ownership barrier of traditional media [29] and potentially weaken biases related to information centralization. Second, social media have the potential of a very broad and deep coverage of all kinds of content, allowing any information to be found, reported, and eventually attract collective attention. But such potential might not necessarily be realized, in particular when information overloads and misinformation spread due to the communication brevity and informality of many social media platforms. In addition, social media are not free of the influence of other biases that can limit their transparency and coverage. For example, a major part of the funding in social media comes from advertising strategies, in which the product is the attention of users and not the reported content [30]. Furthermore, social media are not isolated communities, and traditional mass media biases can potentially resonate in each social medium. A recent example of selective reporting in both mass and social media is the reaction to the Charlie Hebdo shootings in January 2015, which received an extremely large attention share in comparison to similar attacks against freedom of speech in former Yugoslavia or the Middle East [31].

[6] http://www.ibiblio.org/pub/academic/communications/logs/

1.5 Thesis Overview

The reminder of this document is organized as follow. Chapter 2 presents the relevant literature related to this thesis. Chapter 3 details our methodology to evaluate individual advertising, in terms of measuring Online Behavioural targeted Advertising. Moreover, we describe the development and evaluation of a system that we developed for implementing our methodology for measuring OBA and the results obtained. We devote Chapter 4 to the analysis of collective advertising, describing our large scale measurement methodology to collect information for thousands of Trending Topics over a period of several months and the metrics and datasets used. Additionally, we present a methodology to characterize the visibility of Local TTs in a country or city. Chapter 5 is devoted to complement the previous analysis of the collective advertising by doing and analysis of the information flow inside this specific type of advertising. To finish this document the conclusion obtained and some future research lines are presented in Chapter 6.

1.6 Contributions

1.6.1 Publications

This thesis covers contributions from the following literature:

- Carrascosa, J. M., Mikians, J., Cuevas, R., Erramilli, V., & Laoutaris, N. I Always Feel Like Somebody's Watching Me. Measuring Online Behavioural Advertising. In Proceedings of the 11th ACM International on Conference on emerging Networking Experiments and Technologies (CoNEXT 2015). ACM.
- Carrascosa, J. M., Cuevas, R., González, R., Azcorra, A., & García, D. Quantifying the Economic and Cultural Biases of Social Media through Trending Topics. PloS ONE 2015.
- Carrascosa, J. M., González, R., Cuevas, R., & Azcorra, A. Are trending topics useful for marketing?: visibility of trending topics vs traditional advertisement. In Proceedings of the first ACM Conference on Online Social Networks (COSN 2013). ACM.

1.6.2 Software and Tools

Following we show a list of tools implemented during the development of this thesis breaking them into primary tools (i.e. included and explained in the next chapters) or side tools.

List of primary tools:

- Twitter Trending Topics Monitoring Tool: Tool that allows to collect the top list of Trending Topics over 250 different locations (countries and cities) every 5 minutes.
- Semantic Classification Tool: This tool leverages the information from DBpedia[7] to classify different names, terms, words or expressions into meaningful semantic categories.
- Understanding Behavioural Ads: Complex tool developed to collect ads from several webs to provide a better understanding of advertising on the web focusing on Online Behavioural Advertising. It also provides a categorization module of webs based on several sources.

List of side tools:

- Marketing Online Tool: Tool divided in 2 modules to explore how online companies make use of our personal information in different services: (1) analysis of Google and Yahoo services (e.g. mail, maps, search, video). (2) Analysis of Google Campaigns and the cost-per-click (CPC) associated to keywords or interests.
- Twitter Monitoring Tool: Flexible tool that allows measuring multiple aspects of Twitter: selected users' information, tweets associated to keywords, etc.
- Facebook Monitoring Tool: Divided in 2 modules: (1) collect all the information associated to a set of users: friends, likes, posts, general info, etc. and (2) collect all the activity from a Facebook page: users, posts, reactions to these posts, etc.

1.6.3 Datasets

In this section, the main datasets used in this thesis are listed and briefly described.

- Individual Advertising Dataset: This dataset is composed by all the information (websites, ads, cookies, etc.) collected during the experiments of the 51 artificial

[7] http://dbpedia.org/About

personas and 21 sensitive personas defined in Chapter 3. See more details about this dataset in Section 3.4.2.

- Websites Categorization: This dataset includes the categorization for every website and ad from the previous dataset with 3 different sources: Cyren, Google Ad Words and McAfee. See more details about this dataset in Section 3.4.2.
- Trending Topics from Twitter: Complete dataset including the top list of Trending Topics for 62 countries and 215 cities every 5 minutes for different periods of time from 2011 to 2014. See more details about this dataset in Section 4.3.4 and Section 5.3.1.
- World Bank Dataset: This dataset includes relevant information related to demographic, economic, and cultural factors for each country in the previous dataset extracted from the World Bank Database. See more details about this dataset in Section 5.3.2.

Chapter 2

State of the Art

The measurements and characterization of the Internet applications have attracted the attention of the research community in the last years. In this chapter we focus on the recent literature and measurement studies on Online Advertising and Online Social Networks. Furthermore, we highlight the main literature related to information propagation, specifically, in social media.

2.1 Measurements studies on Online Advertising

This thesis is related to recent literature in the areas of measurement driven studies on targeting/personalization in online services such as search [32, 33] and e-commerce [6, 12, 13]. More specifically, in the context of advertising the seminal work by Guha et al. [34] presents the challenges of measuring targeted advertising, including high levels of noise due to ad-churn, network effects like load-balancing, and timing effects. The methodology describes in Section 3.4 considers these challenges. Another early work by Lorolova et al. [35] presents results using microtargeting to expose privacy leakage on Facebook.

Other recent studies have focused on economics of display advertising [36], characterizing mobile advertising [37] and helping users to get control of their personal data and traffic in mobile networks [38], designing large scale targeting platforms [39] or investigating the effectiveness of behavioural targeting [40]. Moreover, Lécuyer et al. [41] developed a service-agnostic tool to establish correlation between input data (e.g., users actions) and resulting personalized output (e.g., ads). The solution is based on the application of the differential correlation principle on the input and output of several shadow accounts that generate a differentially distinct set of inputs.

This work is different in focus to this previous literature since we are primarily concerned with Online Behavioural Advertising (OBA) in display advertising, with the intention of understanding the collection and use of sensitive personal information at a large scale. To the best of our knowledge, only a couple of previous works analyze the presence of OBA advertising using a measurement driven methodology. Liu et al. [42] study behavioural advertisement using complex end-user profiles with hundreds of interests (instead of personas with specific interests) generated from an AOL dataset including users online search history. The extracted profiles from passive measurements are rather complex (capturing multiple interests and types), and are thus, rather inappropriate for establishing causality between specific end-user interests and the observed ads. Our approach is active rather than passive, and thus allows us to derive an exact profile of the interest that we want to capture. Furthermore, the authors collapse all types of targeted advertising (demographic, geographic and OBA), excepting re-targeting, whereas we focus on OBA due to its higher sensitivity from a privacy perspective. Barford et al. [43] present a large-scale characterisation study of the advertisement landscape. As part of this study the authors look at different aspects such as the new ads arrival rate, the popularity of advertisers, the importance of websites or the distribution of the number of ads and advertisers per website. The authors examine OBA very briefly. They trained personas but as they acknowledge their created profiles present a significant contamination including unrelated interests to the persona. Our methodology carefully addresses this issue. Moreover, these previous works check only a small point of the entire spectrum of definitions, metrics, sources, filters, etc. For instance, they rely on Google ads services to build their methodologies which reduces the generality of their results. This work has taken a much broader look on OBA including both the methodology, the results, and the derived conclusions. Finally, to the best of our knowledge, ours is the first work reporting results about the performance of the used methodology, the extent to which OBA is used in different geographical regions and the utilization of DNT across the web.

2.2 Online Social Networks

The measurements and characterization of Online Social Networks (OSNs) have attracted the attention of the research community in the last years. Below we highlight the main literature that is related to one of the main Twitter functionality about the advertising on this OSN: Trending Topics. Specifically, we focus on the literature about the measurements and analysis of Trending Topics and also the semantic classification of them.

Measurement and Analysis of Trending Topics: [44] performed the most exhaustive characterization of Twitter so far. As part of this study the authors briefly analyze Trending Topics using coarse temporal metrics and quantitative metrics to classify Trending Topics in few externally defined (i.e., artificial) categories. Furthermore, [45] use quantitative metrics to analyze the formation, persistence and decay phases of Trending Topics. Both works rely on quantitative metrics that, as shown by [46], may lead to unreliable results due to the best effort nature of Twitter APIs. In addition, [47] have analyzed the propagation of City-Level Trending Topics among cities in US. Finally, [48] studied the differences between the tagging pattern in Twitter and other OSN systems. The authors present the phenomenon of the Twitter *micro-meme*: emergent topics for which a tag is created, used widely for a few days and then disappears. Although these papers provide initial valuable results, they focus on specific aspects of Trending Topics different to the one addressed in this thesis, i.e., the characterization of the visibility offered by TTs in different countries and the impact of these in the advertising ecosystem.

Semantic classification of Trending Topics: [49] use a dataset formed by around 800 Trending Topics and classify them into 18 different categories using a text- and a network-based methodologies that achieve an accuracy of 65% and 70%, respectively. Furthermore [50] assign 15 different properties to Trending Topics (including some unreliable quantitative properties) to classify them into 4 classes using a similar text-based methodology as the one used in [49]. They validate their technique using a training and a test sets with 600 and 436 Trending Topics, respectively. In this case they report an accuracy of 78.4%.

2.3 Information propagation: social media

The selection and content of centralized media can be affected by *subjective* and/or *systemic* biases. Subjective biases operate at the level of the individual information of the reporter, during the evaluation of the informativeness in the context of current events [51]. Since online media content is collectively curated by large groups of Internet users, other subjective biases can emerge from shared values [52], information overloads [53], and cultural preferences [54]. Systemic biases operate at a mesoscopic level, creating patterns that cannot be observed at the level of a news piece or a journalist, but can be observed at larger scales when sufficient content is analyzed [29]. In the context of international news, there is no evidence that can attribute these biases to supranational power structures [55], but incentive mechanisms can bias mass media through economic and social forces [29, 56]. Examples of empirically tested presence of these systemic

biases relate them to the increase of reporting with Gross Domestic Product (GDP) of the country where news originate [57], and its decrease with geographical distance [58] and political stability [59]. This theory is still applicable to online social media as part of a larger medium that also includes traditional mass media.

With respect to social media, previous works focused either on individual biases in news sharing [60–62], or on social and geographical factors or content sharing relevant to viral marketing and information technologies [63–65]. Within this context of individual behaviour, the concept of algorithmic biases in personalization [66] add a complementary view to our collective analysis on systemic and subjective biases. Algorithmic biases are often related to the filter bubble [67], a phenomenon that creates echo chambers [68] that build on already existing individual confirmatory biases [69]. While such concept can be linked to polarization phenomena [70], we focus on testing the economic and cultural factors hypothesized by the theories of media biases. Our approach covers a wider spectrum of online social media, in which not only breaking news play a role, but also popular content like gossip, cultural trends, and fashion define the information that is considered relevant and consumed across societies.

Chapter 3

Individual Advertising

3.1 Motivation

Online Behavioural *targeted* Advertising (OBA) has risen in prominence as a method to increase the effectiveness of online advertising. OBA operates by associating tags or labels to users based on their online activity and then using these labels to target them. This rise has been accompanied by privacy concerns from researchers, regulators and the press. In this chapter, we present a novel methodology for measuring and understanding OBA in the online advertising market. We rely on training artificial online *personas* representing behavioural traits like 'cooking', 'movies', 'motor sports', etc. and build a measurement system that is automated, scalable and supports testing of multiple configurations. We observe that OBA is a frequent practice and notice that categories valued more by advertisers are more intensely targeted. In addition, we provide evidences showing that the advertising market targets sensitive topics (e.g, religion or health) despite the existence of regulation that bans such practices. We also compare the volume of OBA advertising for our personas in two different geographical locations (US and Spain) and see little geographic bias in terms of intensity of OBA targeting. Finally, we check for targeting with do-not-track (DNT) enabled and discovered that DNT is not yet enforced in the web.

3.2 Introduction

The technologies and the ecosystem for delivering targeted advertising is truly mind boggling, involving different types of entities, including Aggregators, Data Brokers, Ad Exchanges, Ad Networks, etc., that might conduct a series of complex online auctions

to select the advertisement that a user gets to see upon landing on a webpage (see [71] for a tutorial and survey of relevant technologies). Furthermore, targeting can be driven by other aspects e.g., location, gender, age group, that have nothing to do with specific behavioural traits that users deem as sensitive in terms of privacy, or it can be due to "re-targetting" [72–74] from previously visited sites. Distinguishing between the different types of advertising is a major challenge towards developing a robust detection technique for OBA. On a yet deeper level, behaviours, interests/types, and relevant metrics have no obvious or unique definition that can be used for practical detection. It is non-trivial to unearth the relative importance of different interests or characteristics that can be used for targeting purposes or even define them. Last, even if definitional issues were resolved, how would one obtain the necessary datasets and automate the process of detecting OBA at scale?

One of the main contribution of this thesis is the development of an extensive methodology for detecting and characterizing OBA at scale, which allows us to answer essential questions: (*i*) How frequently is OBA used in online advertising?; (*ii*) Does OBA target users differently based on their profiles?; (*iii*) Is OBA applied to sensitive topics?; (*iv*) Is OBA more pronounced in certain geographic regions compared to others?; (*v*) Do privacy configurations, such as Do-Not-Track, have any impact on OBA?.

Our methodology addresses all above challenges by 1) employing various *filters* to distinguish interest-based targeting from other forms of advertising, 2) examining several alternative *metrics* to quantify the extent of OBA advertising, 3) relying on multiple independent *sources* to draw keywords and tags for the purpose of defining different interest types and searching for OBA around them, 4) allowing different geographical and privacy configurations. This work combines all the above to present a much more complete methodology for OBA detection compared to very limited work existing in the area that has focused on particular special cases over the spectrum of alternatives that we consider (see Section 2 for related work).

A second contribution of this work is the implementation and experimental application of our methodology. We have conducted extensive experiments for 72 interest-based personas (e.g., 'motorcycles', 'cooking', 'movies', 'poetry' or 'dating') including typical privacy-sensitive profiles (e.g., 'AIDS & HIV', 'left-wing politics' or 'hinduism'), involving 3 tagging sources and 3 different filters. For each experiment we run 310 requests (on average) to 5 different context free "test" websites to gather more than 3.5M ads. Having conducted more than 2.9K experiments combining alternative interest definitions, geographical locations, privacy configurations, metrics, filters and sources of keywords to characterize OBA, we observe the following:

(1) OBA is a common practice, 88% of the analyzed personas get targeted ads associated to all the keywords that define their behavioural trait. Moreover, half of the analyzed personas receive between 26-62% of ads associated to OBA.

(2) The level of OBA attracted by different personas shows a strong correlation (0.4) with the value of those personas in the online advertising market (estimated by the CPC suggested bid for each persona).

(3) We provide strong evidences that show that the online advertising market targets behavioural traits associated to sensitive topics related to health, politics or sexual orientation. Such tracking is illegal in several countries [75]. Specifically, 10 to 40% of the ads shown to half of the 21 personas configured with a sensitive behavioural trait correspond to OBA ads.

(4) We repeat our experiments in both US and Spain and do not observe any significant geographical bias in the utilization of OBA. Indeed, the median difference in the fraction of observed OBA ads by the considered personas in US and Spain is 2.5%.

(5) We repeat our experiments by having first set the Do-Not-Track (DNT) flag on our browser and do not observe any remarkable difference in the amount of OBA received with and without DNT enabled. This lead us to conclude that support for DNT has *not yet* been implemented by most ad networks and sites.

Our intention with this work is to pave the way for developing a robust and scalable methodology and supporting toolsets for detecting interest-based targeting. By doing so we hope to improve the transparency around this important issue and protect the advertising ecosystem from the aforementioned Tragedy of the Commons.

The rest of this chapter is organized as follows: In Section 3.3 we describe what Online Behavioural Advertising is. Section 3.4 describes the proposed methodology to unveil and measure the representativeness of OBA. Using this methodology we have implemented a real system that is described and evaluated in Section 3.5. Section 3.6 presents the results of the conducted experiments. Finally, Section 3.7 summarizes the main results.

3.3 Online Behavioural Advertising (OBA)

Online Behavioural *targeted* Advertising (OBA) is the practice in online advertising wherein information about the interests of web users is incorporated in tailoring ads. This information is usually collected over time by aggregators or ad-networks while users browse the web. This information can include the publishers/webpages a user browses as well as information on activity on each page (time spent, clicks, interactions, etc.).

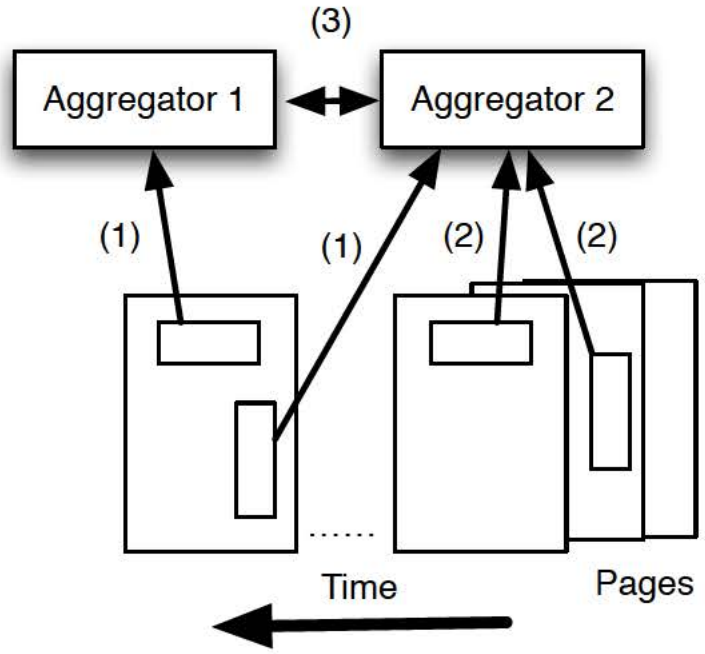


FIGURE 3.1: High level description of how OBA can happen: User browses multiple webpages over time, each page has ads and aggregators present on them. When the user is on the current page (1), the aggregators present on that page can either be new or aggregators that were present on previous pages (2). Hence Aggregator 2 can leverage on past information to show a tailored ad, while Aggregator 1 can either show a run-of-network (RoN) ad or get information from Aggregator 2 (3) to show a tailored ad.

Based on the overall activity of the users, profiles can be built and these profiles can be used to increase the effectiveness of ads, leading to higher click-through rates and in turn, higher revenues for the publisher, the aggregator and eventually the advertiser by making a sale. We note that such targeting is referred to as network based targeting in the advertising literature.

In Figure 3.1, we provide a very high-level overview of how OBA can happen, and information gleaned by browsing can be used. Assume user Alice has no privacy protection mechanisms enabled in her browser. As she is visiting multiple publishers (e.g., websites), her activity is being tracked by multiple aggregators that are present on each publisher, using any of the available methods for tracking users [36, 76]. When Alice visits a publisher, aggregators (aggregator 2) present on that publisher could have already tracked her across the web and based on what information they have about her, they can target her accordingly. Another scenario can be when an aggregator (aggregator 1) is present on the current publisher where Alice is but was not present on previous publishers. In this case, the aggregator can either show a run-of-network ad (un-tailored) or obtain information about Alice from other aggregators and/or data sellers to show tailored ads. Indeed the full ecosystem consisting of aggregators, data sellers,

ad-optimizers, ad-agencies etc. is notoriously complex[1] [71], however for the purposes of this work, we represent all entities handling data other than the user and the publishers, either collecting or selling data, as aggregators.

Other types of (less privacy intrusive) targeted advertising techniques include: (*i*) *Geographical Targeted Ads* are shown to a user based on its geographical location; (*ii*) *Demographic Targeted Ads* are shown to users based on their demographic profile (age, sex, etc) that is estimated by aggregators in different manners, for instance, through the user's browsing history [43]; (*iii*) *Re-targeting Ads* present to the user recently visited websites, e.g., a user, that has checked a hotel in website A, receives an ad of that hotel when visiting website B few hours latter. Finally, a user can be exposed to *Contextual Ads* when visiting a website. These ads are related to the theme of the visited website rather than the user's profile and thus we consider them as non-targeted ads.

3.4 Methodology to measure OBA

In this section we describe our methodology to unveil the presence (or absence) of OBA advertising as well as to estimate its frequency and intensity compared to more traditional forms of online advertising.

3.4.1 Rationale and Challenges

Our goal is to uncover causal links between users exhibiting a certain behavioural trait and the display ads shown to them. Notice that we *do not* claim or attempt to reverse engineer the complex series of online auctions taking place in real time. We merely try to detect whether there is any correlation between the advertisements displayed to a user and his past browsing behaviour.

We create artificial *personas* that present a very narrow web browsing behaviour that corresponds to a very specific interest (or theme), e.g., 'motor sports' or 'cooking & recipes'. We train each persona by visiting carefully selected websites that match its interest and by doing so invite data aggregators and trackers to classify our persona accordingly. We refer to the visited websites as *training* webpages. For instance, the training set for the 'motor sports' persona would be formed by specific motor sports webpages. Therefore, two first challenges for our methodology are which personas to examine and how to select training webpages for them that lead to a minimal *profile*

[1] http://www.displayadtech.com/the_display_advertising_technology_landscape#/the-display-landscape

contamination [43]. By contamination, we are referring to the association of tags and labels not related to the main theme of the persona.

Once the personas and the training webpages have been properly selected, we need to retrieve the ads that these personas obtain upon visiting carefully selected *control* webpages that meet the following criteria: (*i*) include a sufficient number of display ads, and (*ii*) have a neutral or very well defined context that makes it easy to detect context based advertisements and filter them out to keep only those that could be due to OBA. We use weather related webpages for this purpose.

The ads shown to a persona in the control pages lead to websites that we refer to as *landing* webpages. Therefore if the theme of the landing webpages for a persona has a large overlap with the theme of its training websites we can conclude that this persona frequently receives OBA ads. To automate and scale the estimation of the topical overlap between training and landing pages, we rely on online tagging services (e.g., Google AdWords, Cyren, etc) that categorize webpages based on keywords. We use them to tag each training and landing webpage associated to a persona and compute the existing overlapping. Note that we decided to use several online tagging services or *sources* to remove the dependency on a single advertising platform (a limitation of previous works like in [42, 43]).

As indicated before OBA can co-exist with several other types of advertisement on the same page, including: re-targeting ads, contextual ads, geographically targeted ads. Our goal is to define a flexible methodology able to detect and measure OBA in the presence of such ads. Thus, the fourth challenge that our methodology faces is to define filters for detecting and removing these other types of ads.

The final challenge for our methodology is to define meaningful, simple, and easy to understand and measure metrics to quantify OBA using keywords from the training and landing pages of a persona.

3.4.2 Details of the Methodology

- Selection of Personas & Training Pages: The selection of personas with a very specific behavioural trait, and thus a reduced profile contamination, is a key aspect of our methodology. To achieve this in an automated and systematic manner we leverage the Google Ad Words' hierarchical category system, which includes more than 2000 categories that correspond to specific personas (i.e., behavioural traits) used by Google and its partners to offer OBA ads. For each one of these personas, Ad Words provides a list of related websites (between 150 and 330 webs). We use these websites as training pages for

the corresponding persona. Specifically, we consider the 240 personas of level 2 from the Google Ad Words' category system and apply the following three-steps filtering process to collect keywords for each persona while catering to avoid profile contamination:

- *Step 1:* For a given persona p, we collect the keywords assigned by Ad Words to every one of its related websites and keep in our dataset only those websites that have p's category among their keywords. For instance, for p = 'motor sports', we only keep those related websites categorized by Ad Words with the keyword 'motor sports'. After applying this step 202 personas remain in our dataset.

- *Step 2:* For each persona p, we visit each website selected during *Step 1*, using a clean, in terms of configuration, browser, i.e., without cookies, previous web-browsing history, or ads preferences profile. Then, we check the categories from the Google Ad Words system added to the ads preference profile[2] of our browser after visiting those websites. We only keep those related websites, which add to the ads preference profile exclusively p's category or p's category plus a second related one. For instance, for p = 'motor sports', we only keep a related website in our dataset if it includes to the ads preferences profile the category 'motor sport' or the category 'motor sports' plus a second one such as 'cars' or 'motorbikes'. After applying this step 104 personas remain in our dataset.

- *Step 3:* Our final dataset consists of 51 personas that are left with at least 10 training pages each after steps 1 and 2. Note that by having at least 10 training pages we intend to capture enough diversity in the visited sites and expose our personas to a number of trackers that would approximate what a real user would find. Indeed, the considered personas are exposed to 15-40 trackers whereas the examination of the browsers of 5 volunteers revealed that they were exposed to 18-27 trackers[3].

In addition to the above systematically collected personas, we have also selected manually 21 *sensitive personas* related to topics that, for instance, privacy regulation in Europe do not allow to track or process (e.g., health, ethnicity, sexuality, religion or politics) [75]. Interestingly the categories of our *sensitive personas* do not appear in the public Google Ad Words' Hierarchical Category System, however when querying for them in Ad Words we obtain a similar output as for any other persona. We apply the same steps as above with a single difference in *Step 2*, where we keep only websites that do not add any category in the ads preference profile of our browser. By doing so, we ensure that our *sensitive personas* are not being associated with any additional behavioural trait.

[2]Google's ads preferences profile represents the behavioural trait inferred by Google for a browser, based on the previously visited sites. It includes one or more categories from the Ad Words category system.

[3] We have used the tracker detection tool provided by the EDAA [77] to obtain these results.

The final list of 51 regular and 21 sensitive personas can be checked in Figure 3.3 and Figure 3.4, respectively.

- Selection of Control Pages: As indicated in the methodology's rationale we need a set of pages that are popular, have ads shown on them and yet have low number of easily identifiable tags associated with them and thus do not contaminate the profile of our personas. We used five popular weather pages[4] as control pages since they fulfil all previous requirements.

- Visiting Training and Control Pages to obtain ads: Once we have selected the set of training and control pages for a persona, we visit them with the following strategy (see Section 3.5.1 for details). We start with a fresh install, and select randomly a page from the pool of training+control pages to visit with the interval between different page visits drawn from an exponential distribution with mean 3 mins[5]. By doing so, on the one hand, we regularly visit the training pages so that we allow trackers and aggregators present in those pages to classify our persona with a very specific interest according to our deliberately narrow browsing behaviour. On the other hand, the regular visits to control pages allow us to collect the ads shown to our persona to latter study whether they are driven by OBA. An alternative strategy would be to visit first the training pages several times to get our persona profiled by aggregators and visit only control pages. We avoided this strategy because visiting consecutively multiple weather sites fooled the data aggregators into believing that our browser was a "weather" persona.

- Tagging Training and Landing Pages: In order to be able to detect systematically correlations between training and landing pages we need to first identify the keywords that best describe each webpage in our dataset. For this purpose, we use 3 different sources: Cyren[79], Google Ad Words[80] and McAfee[81]. Each source has its own labeling system: Google Ad Words labels web-pages using a hierarchical category system with up to 10 levels and tag categories with 1 to 8 keywords. Cyren and McAfee provide a flat tagging system consisting of 60-100 categories and label web-pages with at most 3 keywords. Note that by utilising multiple sources we try to increase the robustness of our methodology and limit as much as possible its dependency to the idiosyncrasies of particular labeling systems. Finally, it is worth noting that the coverage of the considered tagging services is very high for our set of training and landing pages. In particular Google, McAfee and Cyren were able to tag 100%, 99.0% and 95.5% of the training pages and 100%, 97.2% and 93.3% of the landing pages, respectively.

[4] http://www.accuweather.com,http://www.localconditions.com, http://www.wunderground.com, http://www.myforecast.com, http://www.weatherbase.com

[5] This distribution is selected to emulate a human-being generated inter-arrival time between visits according to recent measurement studies [78].

- Training Set Keywords: To achieve the aforementioned robustness against the particularities of individual classification systems, we filter the keywords assigned to a page by keeping only those that are assigned to the page by more than one of our 3 sources. The idea is to quantify OBA based on keywords that several of our sources agree upon for a specific page relevant to the trained persona. Assume we have a training webpage W tagged with the set of keywords K_1 to K_3 for each one of the 3 sources above. Our goal is to select a keyword k within K_i ($i \in [1,3]$) only if it accurately defines W for our purpose. To do this, we leverage the Leacock-Chodorow similarity [82] ($S(k,l)$) to capture how similar two word senses (keywords) $k \in K_i$ and $l \in K_j$ ($j \in [1,3]$ & $j \neq i$) are. Note that two keywords are considered similar if their Leacock-Chodorow similarity is higher than a given configurable threshold, T, that ranges between 0 (any two keywords would be considered similar) and 3.62 (only two identical keywords -exact match- would be considered similar). We compute the similarity of k belonging to a given source with all the training keywords belonging to other sources and consider k an accurate keyword only if it presents a $S(k,l) > T$ with keywords of at least N other sources. Note that N is also a configurable parameter that allows us to define a more or less strict condition to consider a given training keyword in a given source.

- Filtering different types of ads: To complete our methodology we describe next the filters used in order to identify and progressively remove landing pages associated with non-OBA ads:

- *Retargeting Ads Filter (F_r)*: In our experiment a retargeting ad in a control page should point to either a training or a control page previously visited by the persona. Since, we keep record of the previous webpages visited by a persona, identifying and removing retargeting ads from our landing set is trivial.

- *Static and Contextual Ads Filter ($F_{s\&c}$)*: We have created a profile that after visiting each webpage removes all cookies and potential tracking information such that each visit to a website emulates the visit of a user with empty past browsing history. We refer to this persona as *clean profile*. By definition, when visiting a control webpage the clean profile cannot receive any type of targeted behavioural ad and thus all ads shown to this profile correspond to either static ads (ads pushed by an advertiser into the website) or contextual ads (ads related to the theme of the webpage). Hence, to eliminate a majority of the landing pages derived from static and contextual ads for a persona, we remove all the common landing pages between this persona and the clean profile.

- *Demographic and Geographical Targeted Ads Filter ($F_{d\&g}$)*: We launch the experiments for all our personas from the same /24 IP prefix and therefore it is likely that several of them receive the same ad when geographical targeting is used. Moreover, we have

computed the Leacock-Chodorow similarity between the categories of each pair of personas in our dataset to determine how close their interests are. To filter demographic and geographical ads we proceed as follows: for a persona p that has received an ad A, we select the set of other personas receiving this same ad ($O(A) = [p_{1,A}, p_{2,A}, ...]$) and compute the Leacock-Chodorow similarity between p and $p_{i,A} \in O(A)$. If the similarity between p and at least one of these personas is lower than a given threshold T', we consider that the ad has been shown to personas with a significantly different behavioural trait and thus it cannot be the result of OBA. Instead, it is likely due to geographical or demographic targeting practices.

- Measuring the presence and representativeness of OBA: We measure the volume of OBA for a given persona p by computing the overlapping between the keywords of the training and landing pages for p. Note that we consider that a training keyword and a landing keyword overlap if they are an exact match. In particular, we use two complementary metrics that measure different aspects of the overlapping between the keywords of training and landing pages. However, let us first introduce some definitions used in our metrics: *(i)* We define the set of unique keywords associated with the training pages for a persona p on source s as $K_{T_{ps}}$; *(ii)* We define the set of unique keywords associated with the landing pages of ads shown to a persona p on source s on control pages as $K_{L_{ps}}$; *(iii)* Finally we define the set of unique keywords associated to a single webpage W on source s as K_{W_s}. Note that the set of keywords associated to a web-page remains constant for a given source regardless the persona. Using these definitions we define our metrics as follows:

Targeted Training Keywords (TTK): This metric computes the fraction of keywords from the training pages that have been targeted and thus appear in the set of landing pages for a persona p and a source s. It is formally expressed as follows:

$$TTK(p, s) = \frac{|K_{T_{ps}} \wedge K_{L_{ps}}|}{|K_{T_{ps}}|} \in [0, 1] \tag{3.1}$$

In essence, TTK measures whether p is exposed to OBA or not. In particular, a high value of TTK indicates that most of the keywords defining the behavioural trait of p (i.e., training keywords) have been targeted during the experiment.

Behavioural Advertising in Landing Pages (BAiLP): This metric captures the fraction of ads whose landing pages are tagged with at least one keyword from the set of training pages for a persona p and a source s. In other words, it represents the fraction of received ads by p that are likely associated to OBA. BAiLP is formally expressed as follows:

$$BAiLP(p,s) = \frac{\sum_{i=1}^{L_{ps}} f(K_{W_s^i}) \cdot ntimes}{L_{ps}} \in [0,1]$$
$$where\ f(K_{W_s^i}) = \begin{cases} 1 & if \quad (K_{T_{ps}} \wedge K_{W_s^i}) \geq 1 \\ 0 & if \quad (K_{T_{ps}} \wedge K_{W_s^i}) = 0 \end{cases} \tag{3.2}$$

Note that *ntimes* represents the number of times an ad has been shown to p and L_{ps} is defined as the set of landing pages for a persona p and source s.

In summary, TTK measures if OBA is happening and how intensely (what percentage of the training keywords are targeted) whereas BAiLP captures what percentage of the overall advertising a persona receives is due to OBA (under different filters).

3.5 Automated System to measure OBA

In this section we describe the development and evaluation of a system that we developed for implementing our previously described methodology for measuring OBA.

3.5.1 System implementation and setup

A primary design objective of our measurement system was to be fully automated, without a need for man-in-the-loop, in any of its steps. The reason for this is that we wanted to be able to check arbitrary numbers of personas and websites, instead of just a handful. Towards this end, we used a lightweight, headless browser PhantomJS ver. 1.9 (http://phantomjs.org/) as our base as we can automate collection and handling of content as well as configure different user-agents. We wrote a wrapper around PhantomJS that we call PhantomCurl that handles the logic related to collection and pre-processing of data. Our control server was setup in Madrid, Spain. The experiments were run from Spain and United States. In the case of US we used a transparent proxy with sufficient bandwidth capacity to forward all our requests. We used a user-agent[6] corresponding to Chrome ver. 26, Windows 7. Our default setup has no privacy protections enabled for personas, but for the clean profile, we enable privacy protection and delete cookies after visiting each web-site. A second configuration set-up enables the Do Not Track[7] [83] for all our personas. For each persona configuration (no-DNT and DNT) and location (ES and US) we run, in parallel, our system 4 times in slots of 8-hours in a window of

[6] We have repeated some experiments using different user-agents without noticing major differences in the obtained results.

[7] Do Not Track is a technology and policy proposal that enables users to opt out of tracking by websites they do not visit (e.g., analytics services, ad networks, etc).

3 days so that all personas are exposed to the same status of the advertising market. These time slots generate 310 visits per persona (on average) to the control pages that based on the the results in [43] suffices to obtain the majority of distinct ads received by a persona in the considered period of time. To process the data associated to each persona, configuration and geographical location we use 3 sources to tag the training and landing pages (Google, McAfee and Cyren), 3 different combinations of filters (F_r; F_r and $F_{s\&c}$; F_r, $F_{s\&c}$ and $F_{d\&g}$) and 2 metrics (TTK and BAiLP). Furthermore, we use different values of T and N (for the selection of training keywords) and T' (for filtering out demographic and geographic targeted ads). Overall our analysis covers more than 2.9K points in the spectrum of interest definitions, metrics, sources, filters, geographical locations, privacy configurations, etc.

Before discussing the obtained results (Section 3.6), in the next subsection we evaluate the performance of our system to identify OBA ads using standard metrics such as accuracy, false positive ratio, false negative ratio, etc. It is worth mentioning that, to the best of our knowledge, previous measurement works in the detection of OBA [42, 43] do not perform a similar evaluation of their proposed methodologies.

3.5.2 System Performance to identify OBA ads

To validate our system we need to generate a ground truth dataset to compare against. We used humans for a subjective validation of correlation between training and landing pages as done also by previous works [41, 49, 50, 84]. To that end, two independent panelists subjectively classified each one of the landing pages associated to few[8] randomly selected personas as OBA or non-OBA. Note that the classification of these two panelists was different in 6-12% of the cases for the different personas. For these few cases a third person performed the subjective classification to break the tie.

For each ad, we compare the classification done by our tool as OBA vs. non-OBA with the ground truth and compute widely adopted metrics used to evaluate the performance of detection systems: Recall (or Hit Ratio), Accuracy, False Positive Rate (FPR) and False Negative Rate (FNR). Table 3.1 shows the max and min value of these metrics across the analyzed personas for our three sources (McAfee, Google and Cyren). We observe that, in general, our system is able to reliably identify OBA ads for all sources. Indeed, it shows Accuracy and Recall values over 94% as well as a FNR smaller than 4.5%. Finally, the FPR stay lower than 10% for all the analyzed personas excepting for the 'Yard & Patio' persona where the FPR increases up to 25%.

[8]Note that the manual classification process required our panelists to carefully evaluate around 300-400 landing pages per persona. Then, it was infeasible to perform it for every persona.

	Recall	Accuracy	FPR	FNR
McAfee	99.1/95.6%	99.0/94.2%	25.5/3.8%	4.4/0.1%
Google	99.1/95.7%	99.0/94.3%	25.7/3.8%	4.3/0.1%
Cyren	98.7/95.5%	98.6/94.1%	25.4/4,27%	4.5/1,3%

TABLE 3.1: Max and Min values of Recall, Accuracy, FPR and FNR of our automated methodology to identify OBA ads for our three sources (McAfee, Google and Cyren) for the analyzed personas. Max and Min values correspond to 'Bycicles Accesories' and 'Yard & Patio' personas, respectively.

Training Pages	
http://poolpricer.com	http://levelgroundpool.com
http://whirlpool-zu-hause.de	http://poolforum.se
http://eauplaisir.com	http://photopiscine.net
http://a-pool.czm	http://allas.fi
http://seaglasspools.com	http://piscineinfoservice.com

TABLE 3.2: Sample of the training webpages for 'Swimming Pools & Spas' persona

3.6 Measuring OBA

In this section we present the results obtained with our measurement system for the purpose of answering the following essential questions regarding OBA (*i*) How frequently is OBA used in online advertising?; (*ii*) Does OBA target users differently based on their profiles?; (*iii*) Is OBA applied to sensitive topics?; (*iv*) Is OBA more pronounced in certain geographic regions compared with others?; (*v*) Does Do-Not-Track have any impact on OBA?

We will start by analysing a concrete example and try to help the reader follow along the different steps of our methodology. After that we will present holistic results from a large set of experiemtns.

3.6.1 Specific case: Swimming Pools & Spas and Google

Let us consider a persona, 'Swimming Pools & Spas', and a source, 'Google', to present the results obtained in each step of our methodology for this specific case. Table 3.2 shows the set of training webpages for the 'Swimming Pools & Spas' persona. One can observe by the name of the webpages their direct relation to the 'Swimming Pools & Spas' persona.

We train this persona as described in Section 3.4.2. Then, in the post-processing phase we tag the training and landing webpages using our 3 sources. To describe the process in this subsection we refer to the results obtained for 'Google'. Table 3.3 shows the 6 keywords that Google assigns to the training websites in the first column. However, this initial set of keywords may present some contamination including keywords unrelated to the 'Swimming Pools & Spas' persona. Hence, we compute the semantic similarity

Training Keywords	Filtered Training Keywords
Gems & Jewellery	—
Gyms & Health Clubs	—
Outdoor Toys & Play Equipment	Outdoor Toys & Play Equipment
Security Products & Services	Security Products & Services
Surf & Swim	Surf & Swim
Swimming Pools & Spas	Swimming Pools & Spas

TABLE 3.3: Keywords associated to the training websites for 'Swimming Pools & Spas' using Google as source before and after applying the semantic overlapping filtering with N = 2 and T = 2.5

	TTK	BAiLP
F_r	1	0.17
$F_{s\&c}$	1	0.75
$F_{d\&g}$	1	0.97

TABLE 3.4: TTK and BAiLP values after applying each filter for the 'Swimming Pools & Spas' persona and 'Google' source.

between these keywords and the keywords assigned by other sources to the training webpages with N = 2 and T = 2.5. This technique eliminates 2 keywords and leaves a final set of 4 training keywords shown in the second column of Table 3.3.

Let us now focus on the landing webpages. Our experiments provide a total of 381 unique landing webpages after filter F_r. Then, we pass each of these webpages for $F_{s\&c}$ and $F_{d\&g}$ filters sequentially. Each filter eliminates 226 and 128 of the initial landing pages, respectively. This indicates that contextual ads (eliminated by $F_{s\&c}$) are the more frequent type of ads. After applying each filter we compute the value of the two defined metrics (TTK and BAiLP) using the resultant set of landing pages and its associated keywords and show them in Table 3.4. The results suggest a high presence of OBA ads. Indeed, the obtained TTK values indicate that 100% of the training keywords are targeted and thus they appear among the landing keywords. Moreover, the BAiLP shows that, depending on the specific applied filter, between 17 and 97% of received ads by our 'Swimming Pools & Spas' persona are associated to landing pages tagged with keyword from the training set and thus are likely to be associated to OBA advertising. Note that in the extreme case where no filters are applied, BAiLP represents the percentage of all ads shown that are suspected to be targeted (17% for 'Swimming Pools & Spas' persona). In the other extreme, when all filters are applied, BAiLP shows the same percentage after having removed all advertisements that can be attributed to one of the known categories described in Section 3.3 (97% for 'Swimming Pools & Spas' persona).

Finally, Table 3.5 shows the Top 10 landing pages associated to a larger number of ads shown during our experiment. We observe that the three most frequent landing pages, that amount to most of the ads shown to our persona, are related to Swimming Pools, pointing clearly to OBA.

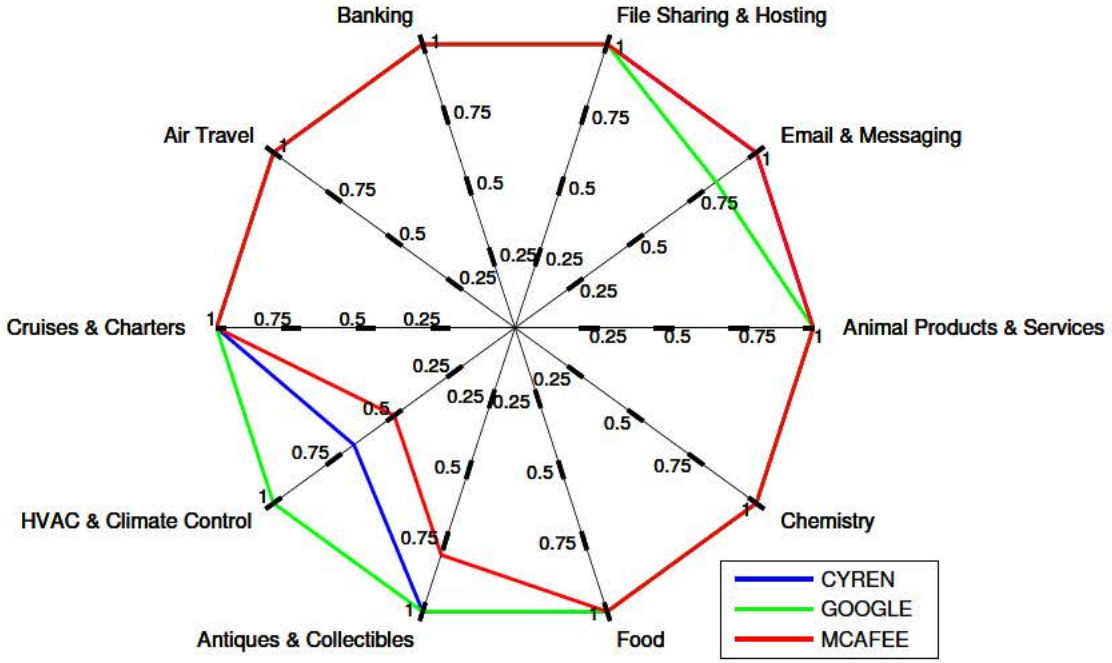
Landing Webpages	Num. ads
www.abrisud.co.uk	1195
www.endlesspools.com	106
www.samsclub.com	16
www.paradisepoolsms.com	8
www.habitissimo.es	8
www.abrisud.es	6
ww.atrium-kobylisy.cz	6
www.piscines-caron.com	5
athomerecreation.net	4
www.saunahouse.cz	4

TABLE 3.5: Top 10 list of landing webpages and the number of times their associated ads were shown to our 'Swimming Pools & Spas' persona.

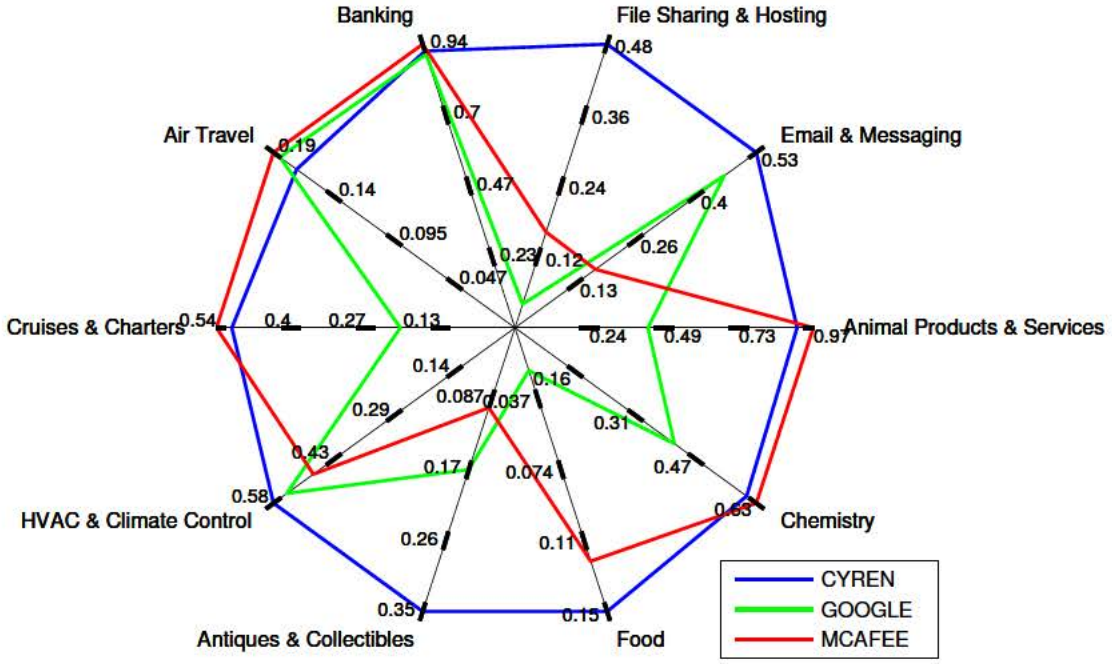
3.6.2 How frequent is OBA?

Let us start analyzing the results obtained with our methodology for each independent source. For this purpose, Figure 3.2 presents the values of TTK and BAiLP for 10 selected personas in a radar chart. In particular, these results correspond to experiments run from Spain, with DNT disabled and all filters (F_r, $F_{s\&c}$, $F_{d\&g}$) activated. First, TTK shows its maximum value (i.e., 1) in 9 of the studied personas for Google and Cyren and in 8 personas for McAfee. Moreover, TTK is not lower than 0.5 in any case. This result indicates that regardless of the source used to tag websites, typically all the training keywords of a persona are targeted and then appear in its set of landing keywords. Second, we observe a much higher heterogeneity for BAiLP across the different sources. In particular, Cyren seems to consistently offer a high value of BAiLP in comparison with the other sources whereas McAfee offers the highest (lowest) BAiLP for 5 (3) of the considered personas and shows a remarkable agreement (BAiLP difference < 0.05) with Cyren in half of the considered personas. Google is the most restrictive source offering the lowest BAiLP for 6 personas. In addition it only shows close agreement with Cyren and McAfee for two personas ('Air Travel' and 'Banking'). These results are due to the higher granularity offered by Google compared to Cyren and McAfee that makes more difficult finding matches between training a landing keywords for that source. If we now compare the BAiLP across the selected personas, we observe that for 27 of the 30 considered cases BAiLP ranges between 0.10 and 0.94 regardless of the source. This indicates that 10-94% of the received ads by these personas are associated to landing pages tagged with training keywords and then, they are likely to be the result of OBA.

These preliminary results suggest an important presence of OBA in online advertising. In order to confirm this observation and understand how representative OBA is, we have computed the values of our two metrics, TTK and BAiLP, for every combination of persona, source, set of active filters and setting N = 2, T = T' = 2.5 in our dataset. Again these experiments are run from Spain and with DNT disabled. In total 4 runs of 459 independent experiments were conducted. Figure 3.3 shows the average and



(a) TTK



(b) BAiLP

FIGURE 3.2: TTK and BAiLP for 10 personas and all sources for N = 2, T = T' = 2.5 and all filters activated (F_r, $F_{s\&c}$, $F_{d\&g}$)

standard deviation values of TTK and BAiLP for the 51 considered personas, sorted from higher to lower average BAiLP value. Our results confirm a high presence of OBA ads. The obtained average TTK values indicate that for 88% of our personas all training keywords are targeted and appear among the landing keywords (for the other 12% at least 66% of training keywords match their correspondent landing keywords). This high overlapping shows unequivocally the existence of OBA. However to more accurately quantify its representativeness we rely on our BAiLP metric, which demonstrates that half of our personas are exposed (on average) to 26-63% of ads linked to landing pages tagged with keywords from the persona training set. Since the overlap is consistently high, independently of the source, filters used, etc., we conclude that these ads are likely the result of OBA.

3.6.3 Are some personas more targeted than others?

Figure 3.3 shows a *clear variability in the representativeness of OBA for different personas*. Indeed, the distribution of the average BAiLP values across our personas presents a median value equal to 0.23 with an interquartile range of 0.25 and a max/min value of 0.63/0.02. This observation invites the following question: *"Why are some personas targeted more intensely than others?"*. Our hypothesis is that the level of OBA received by a persona depends on its economic value for the online advertising market. To validate this hypothesis we leverage the AdWords keyword planner tool[9], which enables us to obtain the suggested Cost per Click (CPC) bids for each of our personas.[10] The bid value is a good indication of the relative economic value of each persona. Then, we compute the spearman and pearson correlation between the BAiLP and the suggested CPC for each persona in our dataset. Note that to properly compute the correlation, we eliminate outlier samples based on the suggested CPC.[11] The obtained spearman and pearson correlations are 0.44 and 0.40 (with p-values of 0.004 and 0.007), respectively. These results validate our hypothesis since *we can observe a marked correlation between the level of received OBA (BAiLP) and the value of the persona for the online advertising market (suggested CPC bid).*

[9] https://adwords.google.com/KeywordPlanner

[10] Specifically, we use the keyword defining the interest of each persona to obtain its suggested CPC bid.

[11] We use a standard outlier detection mechanism, which considers a sample as an outlier if it is higher (smaller) than Q3+1.5*IQR (Q1-1.5*IQR) being Q1, Q3 and IQR the first quartile, the third quartile and the interquartile range, respectively.

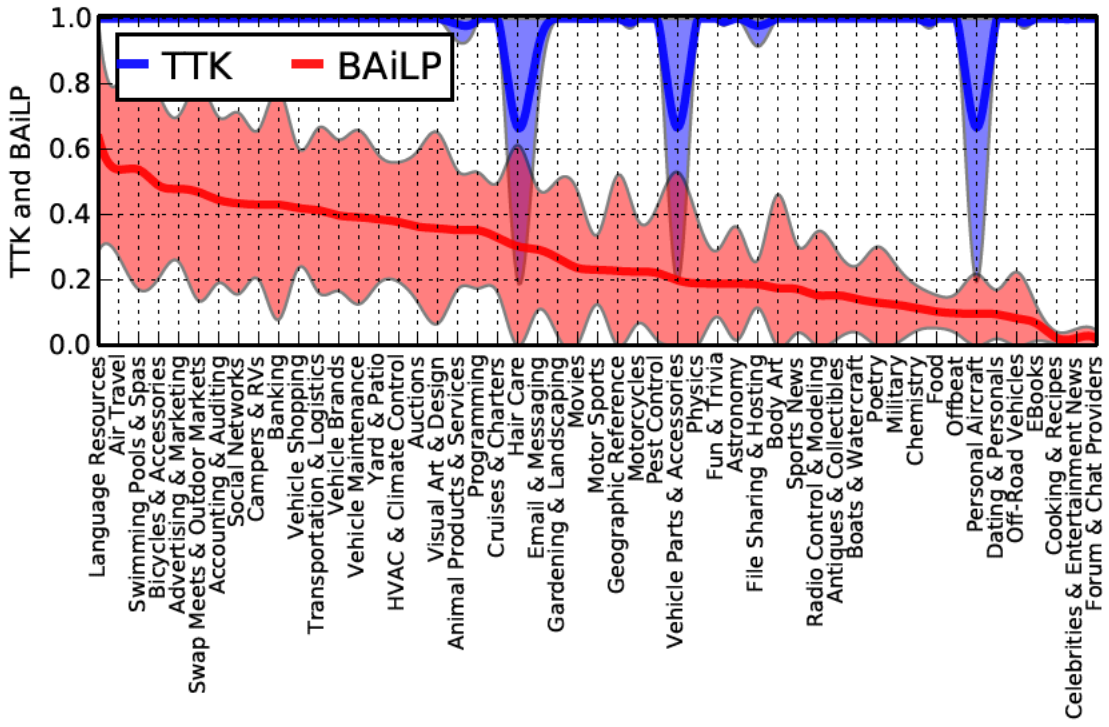


FIGURE 3.3: Average and standard deviation of TTK and BAiLP for each regular persona in our dataset sorted from higher to lower average BAiLP.

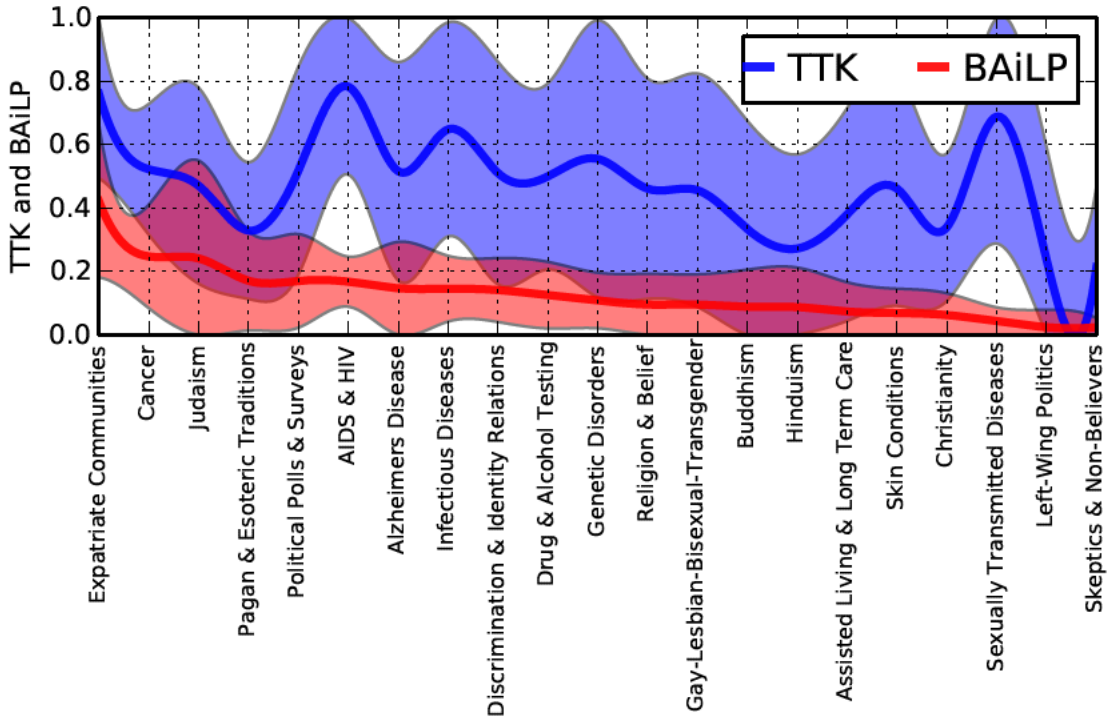


FIGURE 3.4: Average and standard deviation of TTK and BAiLP for each sensitive persona in our dataset sorted from higher to lower average BAiLP.

3.6.4 Is OBA applied to sensitive topics?

The sensitive personas in our dataset present behavioural traits associated to sensitive topics including health, religion, and politics. Tracking these topics is illegal (at least) in Europe. To check if this is being respected by the online advertising market, we repeat the experiment described in the previous subsection for all our 21 sensitive personas, setting the geographical location in Spain. In this case, we run 4 repetitions of 189 independent experiments.

Figure 3.4 shows the average and standard deviation values of TTK and BAiLP for each sensitive persona, sorted again from higher to lower average BAiLP value. One would expect to find values of TTK and BAiLP close to zero indicating that sensitive personas are not subjected to OBA. Instead, our results reveal that despite the lower values compared to the personas of Figure 3.3, the median value of average TKK is 0.47 indicating that for half of the sensitive personas at least 47% of the keywords defining their behavioural trait remain targeted. Moreover, BAiLP results show that 10-40% of the ads received by half of our sensitive personas are associated to OBA. In summary, *we have provided solid evidence that sensitive topics are tracked and used for online behavioural targeting despite the existence of regulation against such practices.*

3.6.5 Geographical bias of OBA

In order to search for possible geographical bias of OBA, we have run the 459 independent experiments described in Subsection 3.6.2 using a transparent proxy configured in US so that visited websites see our measurement traffic coming from a US IP address. For each persona, we have computed the average BAiLP across all combinations of sources and filters for the experiments run in Spain vs. US and calculated the BAiLP difference. Figure 3.5 shows the distribution of average BAiLP difference for the 51 considered personas in the form of a boxplot. Note that a positive (negative) difference indicates a major presence of OBA ads in Spain (US). The BAiLP differences are restricted to less than 10 percentage points across all cases with an insignificant bias (median of BAiLP difference = 2.5%) towards a major presence of OBA ads in Spain than in US. Hence, we conclude that *there is not a remarkable geographical bias in the application of OBA*. Note that we have repeated the experiment with our other metric, TTK, obtaining similar conclusions.

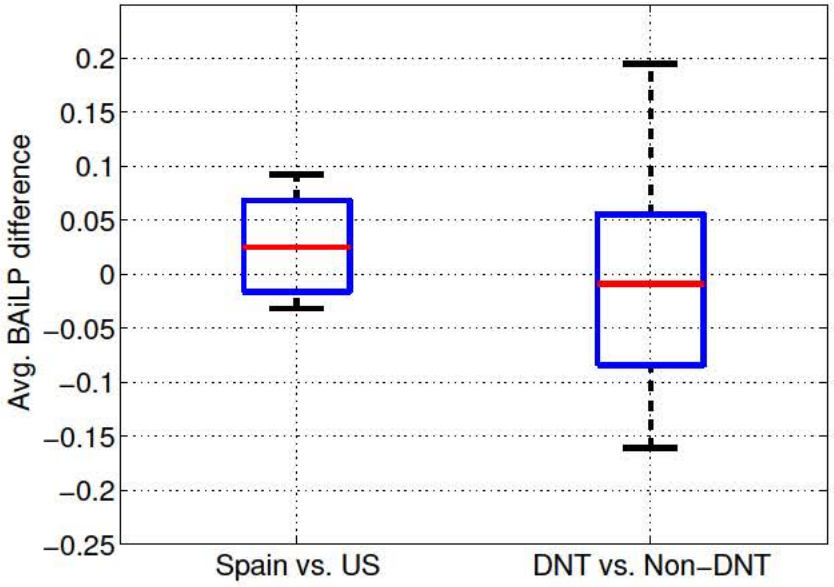


FIGURE 3.5: Distribution of the average BAiLP difference for the 51 regular personas in our dataset for the cases: Spain vs. US (left) and DNT vs. non-DNT (right)

3.6.6 Impact of Do-Not-Track in OBA

Following the same procedure as in the previous subsection, we have computed the average BAiLP difference when DNT is activated from when DNT is not for each one of the 51 regular personas in our dataset, fixing in both cases the geographical location to Spain. Figure 3.5 depicts the distribution of the average BAiLP difference where, a positive (negative) difference indicates a major presence of OBA ads with the DNT activated (deactivated). The median of the distribution is $\sim$0, indicating that half of the personas attract more OBA ads either with DNT activated or not. Moreover, the IQR reveals that half of the personas present a relatively small BAiLP difference ($\leq$ 8 percentage points). Therefore, *the results provide strong evidences that DNT is barely enforced in Internet and thus its impact in OBA is negligible.* Again, we have repeated this experiment with TTK obtaining similar conclusions.

3.7 Discussion

This chapter presents a methodology to identify and quantify the presence of OBA in online advertising. We have implemented the methodology into a scalable system and run experiments covering a large part of the entire spectrum of definitions, metrics, sources, filters, etc that allows us to derive conclusions whose generality is guaranteed. In particular, our results reveal that OBA is a technique commonly used in online advertising. Moreover, our analysis using more than 50 trained personas suggests that the volume of OBA ads received by a user varies depending on the economical value associated

to the behaviour/interests of the user. More importantly, our experiments reveal that the online advertising market targets behavioural traits associated to sensitive topics (health, politics or sexuality) despite the existing legislation against it, for instance, in Europe. Finally, our analysis indicates that there is no significant geographical bias in the application of OBA and that do-not-track seems to not be enforced by publishers and aggregators and thus it does not affect OBA. These essential findings pave a solid ground to continue the research in this area and improve our still vague knowledge on the intrinsic aspects of the online advertising ecosystem.

3.8 Other Applications

In this section we describe a side application of the methodology explained above as an example of the applicability of the methodology in other scenarios. In fact, an adapted version of this methodology has been used to validate the impact of a new service Web Identity Translation (WIT) on the adverstising ecosystem.

Background: Web Identity Translation (WIT) [85] is a service focus on the needs of privacy users and the needs of the advertisers to drive Online Behavioural Advertising (OBA). WIT is a proxy that intervenes between users and the advertising ecosystem to protect users identity without impacting the advertising ecosystem. We could briefly defined the main functionality of this proxy as tracking cookies from the browser and subtitutes them with private cookies to hide the identity of the users and to avoid re-identification.

Goal: Our objective in this section is to show that the ads seen by an original and a WIT-ed profile are "similar". This is a challenging objective since even if one replays the exact web-stream many, if not most, ads will be different for a variety of reasons that impact on both auctioned and non-auctioned ad campaigns. Therefore a simplistic one-for-an comparison of ads does not make much sense since it is not what we expect to see with or without WIT. What really matters is if the proportion of ads that are targeted remain qualitatively similar, i.e., if a user that was searching for a car kept receiving car related ads (albeit for different brands or dealers) after having been WIT-ed.

Therefore, the main goal is to answer the following essential questions: (i) Does the advertising ecosystem still be aware of the users' interests after WIT intervention?; (ii) Are still the users getting similar ads, in terms of topics, related to their interests after WIT intervention?.

Rationale and Challenges: Our goal is to demonstrate that the use of WIT as a method of privacy protection does not have a relevant impact over the advertising

ecosystem. It means that if a user, due to his interests, is shown ads in a specific topic then, after using WIT, the topic will persist. Notice that we do not attempt to uncover correlation between users' web history and ads shown to him. We merely try to detect whether users' interests have been affected (i.e. different perception by the advertising ecosystem) by using WIT.

To achieve this goal, we face several challenges. Firstly, one of the requirements of the system is the feasibility of operating as a real system. Therefore, we need to get a dataset of real users and their web history for the validation. Furthermore, the length associated to these web history (i.e. number of visited webpages) is a key aspect of the validation process. Our hypothesis is that users with identical web history and browsing patterns will tend to exhibit same interests, and then same type of ads, as long as an enough browsing time has elapsed (i.e. in the *long* term) and/or enough webpages with high relevance for the advertisers have been visited[12]. The second challenge is related to obtaining relevant ads shown to the analysed user. Most websites contain several static ads or ads related to the content of the website (i.e. contextual ads), and these ads are shown without distinction to most of the users. Therefore, we need to address how to filter these types of ads in order to differentiate the basis profile of a user from its differentiator profile. Finally, how to decide if two profiles are *similar* among them.

Details of the Methodology: In summary, the main differences are the following:
- Selection of Users/Profiles: real web history instead of artificial users. We run 5 instances of each user.
- Visiting Pattern: sequential navigation along the whole web history instead of randomly.
- Control Pages (i.e. set of pages to collect ads): we injected 3 control pages every 10 pages in the web history.
- Filtering different types of ads: we only filter static and contextual ads (i.e. ads shown to the clean profile).
- We compared (always in terms of keywords) the original stream with the WIT-ed stream (over all repetitions) and then again the original stream and the clean profile that had an empty browsing history.

Results: The validation is conducted using a dataset composed by 21 real web history. It is worth noting that the duration of these web history may range from a few days to months and, therefore, replicate the same temporal pattern is completely unworkable. Therefore, we have set an interval time between websites equal to 60 seconds.

[12]Looking to the advertising ecosystem, it seems obvious to think that visits to Online Social Networks (e.g. Facebook or Twitter) do not provide the same type of information about your interests as if you visit, for example, several car insurance webpages

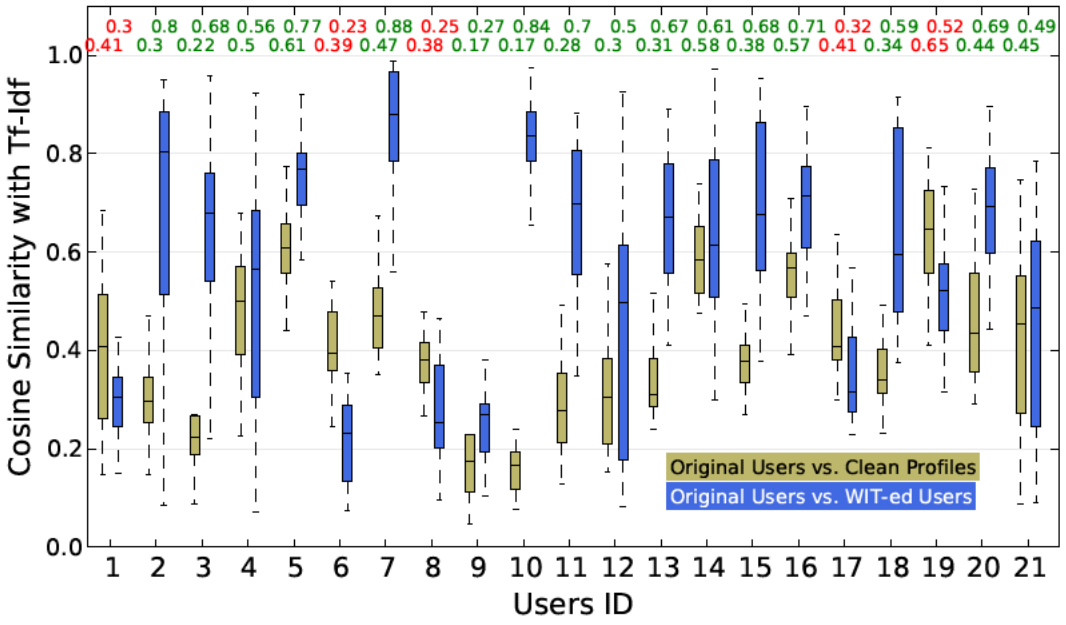


FIGURE 3.6: Comparison of the distribution of the cosine similarity for the 21 *real users* with the users from the *control* set and the *intervene* set respectively. Median values for each pair of boxplots are shown at the top of the figure.

In Figure 3.6 we have a pair of boxplots associated to each user. On the one hand, the boxplot on the left (i.e. light color) shows the distribution of the cosine similarity between the 5 instances of the user with the 5 instances of the control set. On the other hand, on the right (i.e. dark color) shows the distribution of the cosine similarity with their intervened version. Furthermore, at the top of the figure median values are shown where green values implies that, in median, the set of instances of the *Original Users* are more similar to the instances of the *WIT-ed Users* than to the *Clean Profiles*. Values equals to 1 mean that users are identical in terms of the topics seen in the set of ads while values closer to 0 reflect that users have a completely opposing interests.

Figure 3.6 shows that for 16 of the 21 users the targeting similarity between their original and WIT-ed streams is more similar than the corresponding targeting similarity between their original stream and that seen by a clean profile. This result leaves us confident that WIT does not degrade significantly the ability of the advertising ecosystem to perform effective onine targeted advertising.

Chapter 4

Collective Advertising

4.1 Motivation

Trending Topics seem to be a powerful tool to be used in marketing and advertisement contexts, however there is not any rigorous analysis that demonstrates this. In this chapter we present a first effort in this direction. We use a dataset including more than 800K Trending Topics from 35 countries and 215 cities collected over a period of 3 months as basis to characterize the visibility offered by Trending Topics. Furthermore, by using metrics that rely on the exposure time of Trending Topics and the penetration of Twitter, we compare the visibility provided by Trending Topics and traditional advertisement channels such as newspapers' ads or radio-stations' commercials for several countries. Our study confirms that Trending Topics offer a comparable visibility to the aforementioned traditional advertisement channels in those countries where we have conducted our comparison study. In addition, using a standard metric, the Cost Per Mille, we show that marketing campaigns based on Trending Topics seem to be more cost effective than those based on newspapers' ads. Hence, we conclude that Trending Topics can be useful in marketing and advertisement contexts in an efficient manner.

4.2 Introduction

In particular, we study the visibility of World Wide Trending Topics (WW-TT), but more interestingly from the point of view of marketing[1] we analyze the visibility provided by the Trending Topics from 35 different countries (Country-TT) and the Trending Topics of the main cities of these countries (City-TT). Toward this end, we first define

[1]Marketing experts are interested on studying different regional markets.

and implement a high resolution measurement methodology that leverages the Twitter API to collect the list of TTs with a resolution of dozens of seconds. Using this methodology we have collected 3 WW-TTs datasets between Sep 2011 and May 2013 that all together include more than 80K TTs. Using these datasets we demonstrate that the resolution provided by our methodology enables the detection of any change in the visibility of TTs. Identifying these changes is of high importance in the aforementioned marketing or advertisement contexts.
Furthermore, we use the same methodology to collect a dataset including more than 110K Country-TTs from 35 countries and more than 700K City-TTs over a period of 3 months in 2013. We use this dataset to compare the visibility offered by TTs across these countries and cities. In order to perform a complete comparison we define three metrics. The first one helps us to compare the *net-visibility* (i.e., the actual time of exposure) of TTs whereas the other two metrics named *potential-visibility* and *potential-online-visibility* take into account the penetration of Twitter among the population and the population with Internet access in a country (or a city), respectively. These metrics give an insight on the fraction of the population (or "online population") that the Local TTs are able to reach in a country (or a city). In addition, we use the aforementioned metrics to compare the visibility offered by TTs and traditional advertisement channels such as newspapers' ads and radio-stations' commercials for several countries with rather different demographics and cultural backgrounds. In addition, we use an standard metric such as the Cost Per Mille (CPM) to compare the cost effectiveness of advertising campaigns based in Trending Topics and newspapers' ads.

To complete our analysis, we analyze the variability offered by TTs visibility within a country and a city. In particular, for 3 selected countries (Ireland, New Zealand and UK) we present a deep analysis of their TTs visibility: (*i*) using a novel and efficient methodology we classify the TTs of a country in different semantic categories and study which categories are more likely to become TT and which ones offer higher visibility periods; (*ii*) we study whether TTs visibility follows a diurnal pattern as Internet traffic [86] and many other online services do. Finally, we present a study of the variability of TTs visibility for the cities of 4 countries (US, UK, BR and JP) in our dataset and make a comparison of the visibility offered by the TTs of the Top 15 cities in Twitter.

In summary, the main contributions of this chapter are twofold: First, a measurement methodology that allows to monitor the visibility of TTs and its evolution over time. Second, a methodology to properly characterize the visibility of TTs within a country that permits to perform meaningful comparative analyses with other countries or with traditional advertisement channels. The utilization of these methodologies led to the following insights:

- Our results show that the median visibility of TTs is higher than that offered by radio-stations' commercials and newspapers' ads in 4 and 9 out of 10 studied countries, respectively. Hence, we conclude that (at least for the studied countries) TTs can be considered a useful tool in marketing and advertisement contexts.

- However, there is a strong variability on the visibility that TTs offer in different countries and also across Trending Topics within a country (i.e. City-level). In addition, the penetration of traditional media and TTs varies substantially accross countries. Therefore, we cannot generalize the previous conclusion for all the TTs in every country.

- The CPM of promoted Trending Topics in US is 2 order of magnitude smaller than that of a major newspaper such as the Wall Street Journal. This indicates that marketing campaigns based on Trending Topics are very cost efficient.

- Our detailed examination of few countries reveals that "Hashtags" present a higher visibility than other "non-hashtaged" TTs related to "Sport Events" or "Celebrities". Furthermore, the exposure time of TTs presents a clear diurnal pattern for most of the studied countries. Specifically, TTs provide longer visibility periods during night hours when fewer users are connected.

- City-Level Trending Topics offer two order of magnitude more visibility than Country-level Trending Topics for the analyzed countries.

The rest of the chapter is organized as follows: Section 4.3 describes our measurement methodology and our datasets. Section 4.4 details our methodology to evaluate the visibility of TTs within a country while we devote Section 4.5 to put in context our analysis doing a comparison with traditional advertisement channels. Section 4.6 compares the cost-efficiency of advertisement campaigns based on promoted Trends and newspapers ads in US. Section 4.7 dissects the visibility of TTs within a country from a semantic perspective and a city-level analysis whereas Section 4.8 characterizes the visibility of City-Level Trending Topics. Finally, in Section 4.9 we discuss the main conclusions of this chapter.

4.3 Measurement Methodology, Metrics and Datasets

In this section we describe our large scale measurement methodology to collect information for thousands of Trending Topics over a period of several months. Additionally, we define temporal metrics to be used in the rest of the chapter. We also discuss the basic

filtering techniques applied to produce meaningful datasets and finally we summarize the datasets used to conduct our analysis.

4.3.1 Measurement Methodology

Twitter provides different APIs to access the information available in the system[2]. In our methodology we leverage two of these APIs, namely the REST and Streaming APIs. We query the REST API to obtain the list of 10 TTs at a given instant and for a given location (e.g., a country or city). Since the maximum number of queries allowed by Twitter to the REST API is 150 per hour, we are able to collect the list of TTs every 24 seconds for a given location. This guarantees a fine grain time resolution in the sampling of the Trending Topics list. Furthermore, we query the Streaming API to retrieve the tweets associated to a given Trending Topic. The Streaming API offers a best effort service in which the system provides as many tweets as it can (depending on the load) including the term (i.e., Trending Topic) requested in the query. In particular, our tool uses the Streaming API to collect tweets associated to the 20 most recent TTs at any moment.
Using multiple instances of our tool we are able to collect data from World Wide (WW) Trending Topics as well as Local Trending Topics from 35 different countries and 215 cities in parallel[3].

4.3.2 Temporal Metrics

The visibility of a TT is basically defined by the time that it is shown to users that we refer to as *exposure time*. We use the following meaningful metrics to capture the temporal characteristics of TTs:

- *number of active periods*, this metric counts the number of times that a given topic has become TT. We refer to each one of those active periods as an *instance*.

- *total active time*, this metric captures the total time a topic has been TT across one or multiple active periods, i.e., the total exposure time.

- *age*, this metric measures the total time between the first instant and the last instant a topic is a TT across one or multiple active periods.

To clarify these concepts, let us consider the following simple example: a topic that has been Trending Topic on Jan 1^{st} 2013 between 9 AM and 9:30 AM, on Jan 1^{st} between

[2]https://dev.twitter.com/docs
[3]This was the number of countries and cities offering the list of Trending Topics at the moment of the data collection campaign.

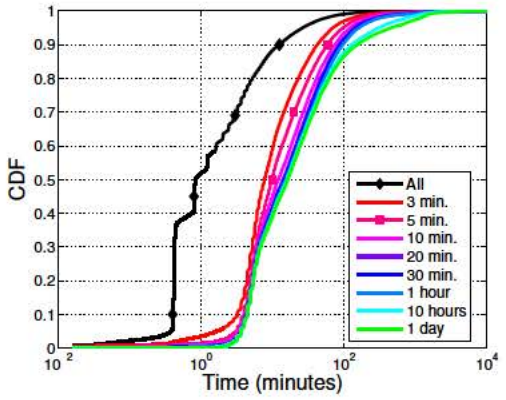


FIGURE 4.1: *Active time* of WW-TT instances without oscillations filtering (*All*) and with oscilations filtering (filtering times from 3 min to 1 day).

6 PM and 6:20 PM and on Jan 2^{nd} between 8:50 AM and 9 AM. Then, the *number of active periods* for this Trending Topic is 3 (or in other words this TT has 3 associated instances), the *age* is 24 hours (from 9 AM Jan 1^{st} to 9 AM Jan 2^{nd}) and the *total active time* is 60 min (30, 20 and 10 minutes in the first, second and third active periods, respectively).

Previous studies have considered the volume of tweets [44, 45] to analyze TTs using the Search or the Streaming API. Although, this metric does not capture the visibility of TTs as well as those presented above, it could be an interesting complementary metric for our study. Unfortunately, as demonstrated in [46], the volume of tweets obtained from the Streaming API is not a reliable metric. In particular, that study shows that due to the best effort nature of the Streaming API in those peak hours where the number of tweets associated to a topic is higher the API provides the lower number of tweets[4]. In short, using the volume of tweets as a metric may lead to wrong results and thus we do not use it for our analysis.

4.3.3 Data Filtering

As described before, our methodology allows to gather the list of the 10 TTs for a given location (e.g., WW or Local TTs for a country or city) every 24 seconds. Unexpectedly, there is a high variability in the composition of this list in a time scale of few minutes (or even seconds). We conjecture that this high variability is due to those topics that are ranked by Twitter Trending Topics selection algorithm around the 10^{th} position that enter and leave the Top 10 list frequently. The curve labeled as *"All"* in Figure 4.1 shows the distribution of the *active time* for each WW-TT instance in our dataset. We observe that half of the instances present an *active time* lower than 1 minute. Therefore, the

[4]Note that this observation also applies to the search API since it provides a subset of the tweets provided by the Streaming API [87].

Trending Topics selection algorithm works in intervals of seconds. Note that previous works considered that the list of TTs was updated in intervals of 5 minutes [44] or 20 minutes [45].

This real time selection of TTs produces a phenomenon that we refer to as *oscillations*. This occurs when a topic enters and leaves the Trending Topic list several times in a short period of time (e.g., a few minutes). However, *oscillations* are unlikely to be observed by users since neither the web interface of Twitter nor Twitter API-based applications refresh the Trending Topic information as frequently as our measurement tool. Therefore, in order to better approximate the user experience we would like to process the collected data in order to filter these short-term *oscillations*. For this purpose, we consider that a topic that presents one or more *oscillations* within a period of X minutes has been a Trending Topic during the whole X minutes period. Figure 4.1 shows the CDF of the *active time* of single instances of TTs after applying the described technique for X = 3, 5, 10, 20, 30, 60, 600 and 1440 minutes. The result suggests that a value of X = 5 min suffices to eliminate most of the short-term *oscillations* (i.e., those in the order of seconds or few minutes) and do not merge those long-term *oscillations* (i.e., those in the order of tens of minutes). Therefore, we filter out the *oscillations* using this value. We have repeated the experiments described along the chapter with other values of X (3 and 7 minutes) obtaining similar results.

4.3.4 Datasets

Using the measurement methodology and data filtering technique described in this section we collected the following datasets:

WW-TT: This dataset is formed by 3 traces including all the WW-TTs in 3 different periods of approximately 3 months each.

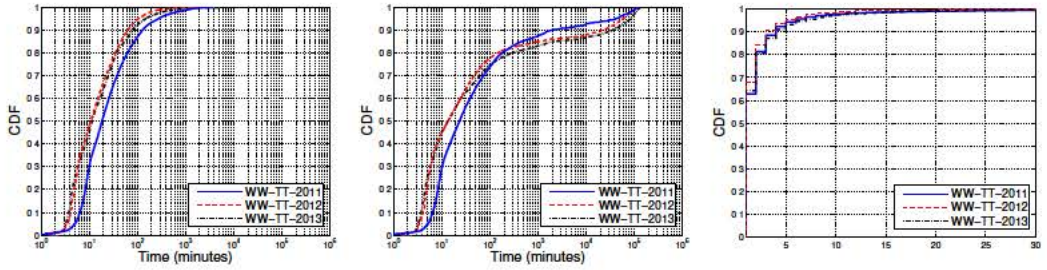
Country-TT: This dataset was collected in parallel to our most recent WW-TT trace. It includes the Local Trending Topics for 35 countries over a period of 3 months.

City-TT: This dataset was also collected in parallel to our most recent WW-TT trace. It includes the Local Trending Topics for 215 cities spread across 30 countries over a period of 3 months.

The specific dates of data collection along with the number of TTs included in each trace are shown in Table 4.1.

TABLE 4.1: **Basic statistics of Datasets.**

	Period	TT Instances	Unique TTs
WW-TT-2011	09/07/2011 - 11/30/2011	31251	13964
WW-TT-2012	12/01/2011 - 02/25/2012	80856	43985
WW-TT-2013	02/20/2013 - 05/20/2013	67221	29326
Country-TT-2013	02/20/2013 - 05/20/2013	713012	112196
City-TT-2013	02/20/2013 - 05/20/2013	6572036	123050



(a) Total active time (b) Age (c) Number of active periods

FIGURE 4.2: CDF of the temporal metrics of TTs within our WW-TT datasets.

4.3.5 Accuracy of the measurement methodology

The final goal of our measurement methodology is to accurately collect the visibility offered by TTs at any moment, expressed through the previously defined temporal metrics. Hence, the proposed methodology should be able to discover any change on the visibility of TTs.

Figure 4.2 presents the distribution of the number of active periods, total active time and age across TTs within our three Worldwide datasets. We observe that TTs within WW-TT-2012 and WW-TT-2013 show a similar visibility that is significantly different from that shown by TTs within WW-TT-2011. In particular, Figure 4.2(a) reveals that the median value for the total active time halves, from 20 to 10 minutes, between WW-TT-2011 and WW-TT-2012 and then remains stable in WW-TT-2013. This result suggests that the TTs selection algorithm was modified to severely reduce the visibility of TTs in December 2011, most likely during the large system upgrade process carried out by Twitter on that month [88]. However, to the best of the authors knowledge, this modification on the Trending Topics selection mechanism was not publicly announced by Twitter despite the implications that it might have.

In order to corroborate the previous observation, we have calculated the distribution of the total active time for each individual month in our Worldwide datasets but December 2011 (for being the month where the modification took place) and performed a Kolmogorov-Smirnov test [89] for each pair of distributions. The obtained results show that the distributions of Sep'11, Oct'11 and Nov'11 are similar between them and so are

the distributions of Jan'12, Feb'12 and those from 2013. Specifically, the parameter K of the test varies between 0.06 and 0.15 in all cases. However, when we compare any of the first three months to any of the other months the Kolmogorov-Smirnov test concludes that the distributions are significantly different, in particular, K varies between 0.27 and 0.32.

Moreover, Figure 4.2(b) shows the distribution of TTs age for our three WW-TT datasets. Again, we observe that the distribution for this metric is similar for WW-TT-2012 and WW-TT-2013 and different from WW-TT-2011. This confirms the reported change in TTs visibility. In particular, the modification in the TTs selection algorithm in Dec 2011 yielded around 80% of TTs (i.e., those that have one or two close active periods) to present a lower age in our WW-TT-2012 and WW-TT-2013 than in the WW-TT-2011 dataset. However, this trend is reversed for the 20% TTs presenting a longer Age (i.e., those with several associated active periods). This suggests that the TTs selection algorithm implemented since Dec 2011, in addition to shorten the active time of TTs instances, also requires that the period of time with a relative reduced volume of tweets for a topic to become TT again to be longer. It is worth to mention that we have performed equivalent Kolmogorov-Smirnov tests for this metric as for the total active time obtaining similar results.

Finally, Figure 4.2(c) shows the distribution of the number of active periods (or instances) for our WW-TT datasets. The results indicate that this distribution is similar for the three datasets. Then, the modification of TTs selection algorithm in Dec 2011 has not affected the ability of TTs to achieve this status multiple times, however as noted before the time between TTs instances has increased.

In summary, the results presented in this subsection confirm that the proposed measurement methodology is capable of accurately capture the visibility associated to TTs as well as identifying any change it may suffer along time.

4.4 Methodology to characterize the visibility of Local TTs

In this section we present a methodology to characterize the visibility of Local TTs in a country or city and compare it with that offered by TTs in other countries or cities. For this purpose we define three meaningful metrics named *net-visibility*, *potential-visibility* and *potential-online-visibility*.

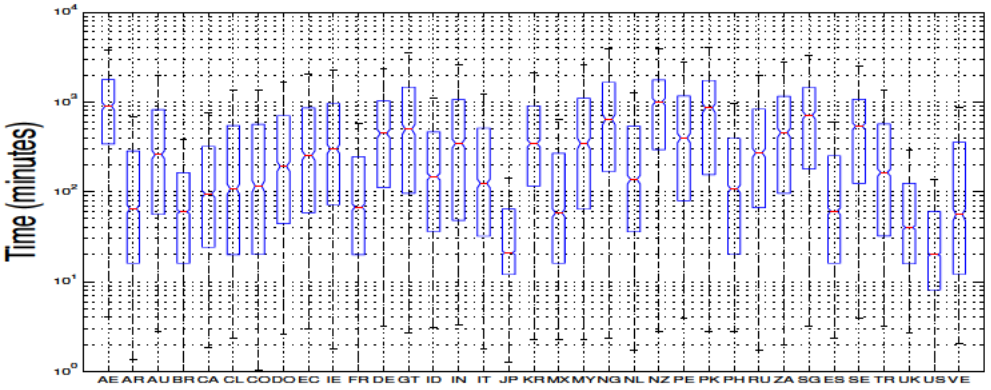


FIGURE 4.3: Distribution of Total Active Time for the TTs in each one of the 35 countries of our Country-TT-2013 dataset.

4.4.1 A first look at TTs visibility within a country

Let us use the temporal metrics defined in Section 4.3.2 to make a first comparison of the visibility granted by TTs across different countries.

Figure 4.3 shows the distribution of the *total active time* for TTs in each one of the 35 countries of our Country-TT-2013 dataset. Each distribution is represented in the form of a boxplot where the box shows the 25, 50 and 75 percentiles of the distribution and the whiskers indicate the 5 and 95 percentiles, respectively. Note that any boxplot used in the rest of this chapter presents this same information unless otherwise stated.
We observe that there is an important variability in the *total active time* for the TTs within a country. We will address this issue in Section 4.7. Of more interest for this section is the significant difference among the distribution of *total active time* for different countries[5]. In particular, the median value of the *total active time* varies around 2 order of magnitude between 20 min in US and 1000 min in New Zealand (NZ). This observation suggests the presence of well differentiated groups of countries with respect to the visibility provided by TTs.
In order to find these groups we leverage standard clustering techniques. Specifically, we use the following 9 input variables to our clustering algorithm: 25, 50 and 75 percentiles of the *total active time*, the *age* and the *number of active periods* for the TTs of a given country. We use the EM clustering algorithm since it provides as output the optimum number of clusters[6]. This clustering process results in 3 distinct clusters[7]. Figure 4.4 shows the distribution of the median value of the three temporal metrics (*total active*

[5]Note that we also observe a significant variability for the *age* and the *number of active periods* across countries but we do not present the results due to space limitations.

[6]The EM algorithm follows a cross-validation approach to find the optimum number of clusters [90]. Furthermore, we double-check the correlation between variables to eliminate redundant information in the clustering process.

[7]We have repeated the clustering exercise using EM and different number of seeds and for all cases we always obtain the same optimum number of clusters.

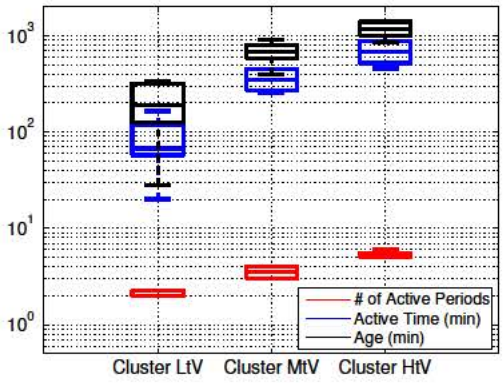


FIGURE 4.4: Summary of the distribution of temporal metrics for the HtV, MtV and LtV clusters.

time, *age* and *number of active periods*) for the countries within each cluster in the form of boxplots. We can observe that the clustering algorithm produced meaningful results since the clusters are perfectly separated and thus represent three different groups that we refer to as: *High temporal Visibility* (HtV), *Medium temporal Visibility* (MtV) and *Low temporal Visibility* (LtV). In particular, the median values for the *total active time* of the HtV, MtV and LtV groups are 700, 350 and 70 min, respectively.

Note that the temporal metrics and, specifically, the *total active time* of a TT captures the *net-visibility* associated to that TT. This is the total time that the TT is visible (or exposed). In the next subsection we develop further the concept of *net-visibility*.

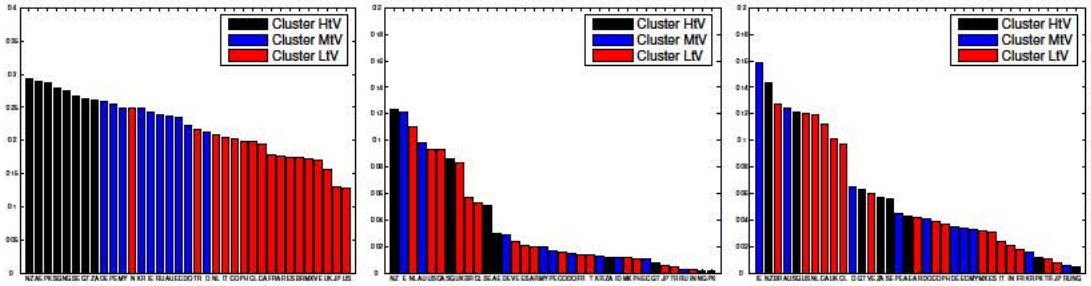
4.4.2 Net-Visibility

We define a normalized version of the *total active time* to represent the *net-visibility* associated to a TT. We refer to this metric as *net-visibility* (NV) and express it as follows:

$$NV = \frac{log(total\ active\ time)}{log(max(total\ active\ time))}\ \alpha \in [0,1] \tag{4.1}$$

where the *max(total active time)* is the duration of our measurement period that is the maximum active time that a TT may have in our dataset. Moreover, the list of TTs shares the bandwidth of the medium (e.g., PC or tablet screen) with other elements like the timeline or the recommendation of users to follow. Then, it is likely that some users do not pay attention to the Trending Topics while browsing through the Twitter interface. The aim of the parameter α in the previous expression is capturing this behaviour.

This phenomenon has been well studied in the area of online advertisement where it



(a) Net-Visibility (b) Potential-Visibility (c) Potential-Online-Visibility

FIGURE 4.5: Trending Topics' visibility metrics for the 35 countries in our Country-TT dataset.

is refereed to as *banner blindness* [91, 92]. In a recent study, [93] analyze the banner blindness among users who browse a web with exploratory purposes, i.e., not looking for a specific piece of information. This browsing behaviour represents well the typical browsing pattern of Twitter users. The authors quantify the banner blindness through a normalized metric of the *recognition* that captures whether a user reminds or not one (or more) banner(s) that was (were) shown during the browsing session. The value of this metric ranges between 0 (no recognition) and 1 (full recognition). The obtained results indicate that the average recognition for users performing an explorative browsing is 0.51. Given the similarity between the described scenario and ours, we will consider a value of $\alpha = 0.51$ along the chapter.

Note that the *net-visibility* for a country is computed as the median of the *net-visibility* of all TTs of that country. We have computed the *net-visibility* for the 35 countries included in our Country-TT-2013 dataset. Figure 4.5(a) presents a ranking of countries based on their *net-visibility* (from highest to lowest). The results indicate that, as expected, countries within the HtV class present the highest *net-visibility*.
Although *net-visibility* is definitely an interesting metric, it does not properly characterize the actual potential visibility offered by a TT since it does not take into account the penetration of Twitter in a country. For instance, the actual visibility granted by TTs in a country with 10K Twitter users and a *net-visibility* of 0.9 may be lower than in a country with 100M Twitter users and a *net-visibility* of 0.1. In the latter case the TTs would be visible for a shorter period of time but are (potentially) exposed to a much larger number of users.

4.4.3 Potential-Visibility & Potential-Online Visibility

To properly characterize the potential visibility offered by a TT we need to consider both the *net-visibility* and the penetration of Twitter in the country. Toward this end,

we have defined a normalized metric that considers these two aspects. We refer to this metric as *potential-visibility* (PV) and it is expressed as follows:

$$PV = NV \frac{\#Twitter\ users}{country\ population} \in [0, 1]. \tag{4.2}$$

where, the fraction term represents the penetration of Twitter in a country. In particular, the #Twitter users is calculated as the % of registered Twitter users in a country (as obtained from our previous work [94]) multiplied by the most recent value of overall registered users reported by Twitter (554M)[8]. Furthermore, the population of each country is obtained from the World Bank statistics [95]. The *potential-visibility* for a country is computed as the median of the *potential-visibility* for the TTs of that country.

We have defined a second valuable metric, the *potential-online-visibility* (PoV). This is a normalized metric that considers the penetration of Twitter among the Internet users of a country rather than among the whole country population. The number of Internet users for a country is also obtained from the Mundial Bank statistics. The expression for the PoV for a TT is the following:

$$PoV = NV \frac{\#Twitter\ users}{\#Internet\ users} \in [0, 1]. \tag{4.3}$$

Differently from the *potential-visibility*, that characterizes the capacity of TTs to reach the population of a country, this metric captures the capacity of TTs to reach the Internet users of that country. Then, a person or company interested on having online presence would be more interested in this second metric[9]. Furthermore, it is worth noting that by definition the *potential-online-visibility* $\geq$ *potential-visibility* and the equality happens only if the Internet penetration in a country is 100% (i.e., all the citizens from a country have Internet access). Again, the *potential-online-visibility* for a country can be computed as the median of the *potential-online-visibility* for its TTs.

Figures 4.5(b) and 4.5(c) present the sorted list (from highest to lowest) of the 35 studied countries based on their *potential-visibility* and *potential-online-visibility*, respectively. These figures allow to easily identify those countries in which TTs have potential to reach a larger portion of the population (Figure 4.5(b)) and/or the online population (Figure 4.5(c)). We believe that these metrics are of high interest to evaluate the usefulness of TTs in marketing and advertisement contexts.

[8] http://www.statisticbrain.com/twitter-statistics/

[9] Note that in many cases advertisement campaigns have a specific target audience. Our metrics can be adapted to those cases. In particular, we would need to change the penetration value considering the estimated number of Twitter users belonging to the target audience and the size of the target audience in the numerator and denominator, respectively.

As we guessed, the *potential-visibility* (and the *potential-online-visibility*) depicts a quite different picture than the *net-visibility*. For instance, Ireland (IE) that is ranked 14^{th} based on the *net-visibility* occupies the 2^{nd} position based on the *potential-visibility* (1^{st} based on the *potential-online-visibility*). This occurs because despite IE has a medium *net-visibility*, it shows a high Twitter penetration and thus the potential of TTs to reach a higher portion of the population is higher than in most of other countries. We observe the opposite effect for Nigeria (NG) that has the 5^{th} highest *net-visibility*, but due to the low penetration of Twitter in the country, it shows the 2^{nd} lowest *potential-visibility* (the lowest *potential-online-visibility*).

Finally, it is worth to mention that we observe slight variations between the ranking of *potential-visibility* and *potential-online-visibility* metrics for most of the countries. This variability is dictated by the different penetration of Internet in different countries.

4.5 Trending Topics vs. Traditional Advertisement Channels

In this section we first introduce the most common metric used to measure the visibility of ads in traditional media and discuss why it is not appropriate to assess the visibility of Trending Topics. Afterwards, we leverage the methodology and metrics described in the previous section to make a comparison of the potential visibility offered by TTs and ads in traditional media.

4.5.1 Background on assessment of visibility in Traditional Advertisement Channels

There is a standard metric used to measure the visibility achieved by ads in traditional media (e.g., radio-stations, TV channels or newspapers). This metric is named Gross Rating Point (GRP) [96, 97] and is expressed as follows:

$$GRP = frequency \cdot reach \tag{4.4}$$

Where the *reach* and the *frequency* are defined as:

- The *reach* is the ratio between the number of individuals within the target audience (e.g., men over 50) that use the specific media (e.g., a specific radio-station or TV channel) and the total number of individuals within the target audience.

- The *frequency* is the ratio between the number of views (listenings) of an ad and the number of people who viewed (listened to) that ad. In other words, it indicates the average number of views (listenings) of an ad per user.

On the one hand, the *reach* used in the GRP is exactly the same metric as the *penetration* we use to compute our PV. On the other hand, marketing companies rely on the information provided by audiometers to compute the *frequency* for ads in TV-channels or radio-stations. These are devices installed in houses that monitor the watching (listening) activity of TV (radio-station) users. In the case of newspapers this metric is estimated based on the *Readership*. This is, the number of daily readers of a newspaper. Unfortunately, the frequency is a metric rather difficult to measure for alternative advertisement channels such as Trending Topics. Indeed, there is a controversial debate regarding the suitability of GRP for advertisement in online media [98, 99].
Our PV metric considers the time of exposure of an ad, that is an objective metric (similarly to the frequency), but it can be accurately measured for both traditional advertisement channels (e.g., radio-stations' commercials or newspapers' ads) and alternative ads channels such as TTs. Hence, our PV (contrary to GRP) allows comparing the visibility of traditional and new types of advertisement channels.

4.5.2 Visibility of Trending Topics vs. Newspapers' ads and Radio-stations' commercials

In this subsection we apply the metrics defined in Section 4.4 to traditional advertisement channels such as newspapers' ads and radio-stations' commercials and compare their visibility to that offered by TTs for 10 selected countries: Canada (CA), Colombia (CO), Ireland (IE), France (FR), Germany (DE), Guatemala (GT), New Zealand (NZ), Spain (ES), United Kingdom (UK) and United States (US).

Let us focus first on newspapers' ads. We consider full-page ads for our analysis and thus α is equal to 1 because the ad uses all the bandwidth of the medium. For comparison purposes we assume that an ad appears in a newspaper every day over a period equivalent to the duration of our Country-TT-2013 dataset (90 days). Finally, [100] report that the average time that readers dedicate to an ad in newspapers is 17.26 seconds. In particular, their results are obtained from an experimental study in which they use eye-tracking techniques on a population of slightly more than 3600 users. Using these values we can estimate the average total active time associated to newspapers' ads that would be equal to 17.26 (sec/day)* 90(days) = 25 min and 53 sec. Moreover, the information regarding newpapers' readership is typically available. In particular we have collected that information for some of the most popular newspapers in the countries

under consideration. The described data allows us to estimate the *net-visibility* and the *potential-visibility*[10] for popular newspapers of the studied countries.

Now, we consider the example of radio-stations' commercials. Again, α is 1 because radio stations' commercials use all the bandwidth of the medium. We consider the traditional duration of radio-stations' commercials of 60 seconds for our analysis. Note that slots of 15 or 30 seconds are typically offered by radio-stations as well [101, 102]. Furthermore, radio-stations' advertisement campaigns vary between few weeks and few months depending on their goal. Then, for comparison purposes we consider the duration of our dataset (90 days) that is included in this range. Finally, the advertiser has to define a schedule for the ad. This is, the number of used slots per day and time-frames associated to those slots (morning, afternoon, evening or night). To this end, advertisement companies indicate that an ad should be listened at least 3 or 4 times by a person in order to be sure that he/she got the message [103, 104]. Hence, they use this reference value to define the most suitable schedule for each specific campaign. In our case, we consider an aggressive campaign in which the ad is played three times in every time-frame (12 times a day) so that the probability of people listening to it 4 times is high.

We can use the previous data to estimate the *total active time* associated to a radio-station's commercials as 60 (sec/commercial) *12 (commercials/day)*90 (days) = 1080 minutes. Furthermore, the audience of some of the most popular radio-stations in the considered countries is publicly available. Hence, with the described data we can compute our visibility metrics for those radio-stations.

The computed *net-visibility* for radio-stations' commercials and newspapers' ads is 0.5927 and 0.2760, respectively. Comparing these results with the median *net-visibility* of TTs for the 35 countries shown in Figure 4.5(a) we observe that radio-stations' commercials present a significantly higher *net-visibility* that TTs in all the 35 countries. Furthermore, TTs offer a slightly higher *net-visibility* than newspapers' ads in only 3 countries: New Zealand (NZ), Arab Emirates (AE) and Pakistan (PK). Hence, we conclude that ads in traditional media enjoy longer exposure times than Trending Topics. However, as indicated in Section 4.4 the *potential-visibility* is a more accurate metric since it takes into account the penetration of the specific media in the country. Figure 4.6 shows the *potential-visibility* associated to popular radio-stations' commercials and newspapers' ads as well as the median *potential-visibility* of TTs for the 10 considered countries. We observe that the *potential-visibility* depicts a different picture than the *net-visibility* due to the different penetration of Twitter, newspapers and radio-stations

[10]The *potential-online-visibility* does not make sense in this case since we are not considering online media.

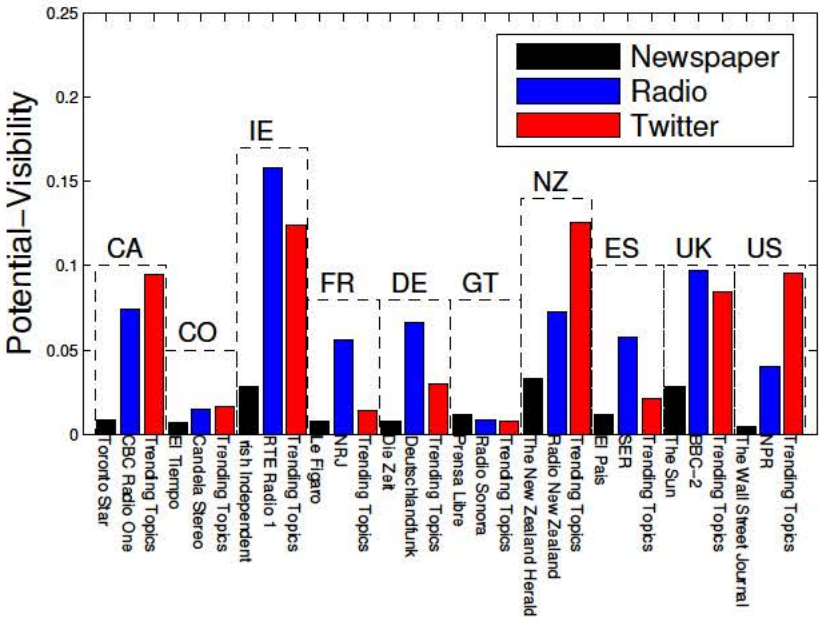


FIGURE 4.6: Potential-visibility for TTs, radio-stations' commercials and newspapers' ads for the 10 considered countries. The x-axis presents the names of the studied media for each advertisement channel and country.

in these countries. In particular, radio-stations' commercials, Trending Topics and newspapers' ads show the highest *potential-visibility* in 5 countries (IE, FR, DE, ES and UK), 4 countries (CA, CO, NZ and US) and 1 country (GT), respectively. Moreover, in all countries, excepting Guatemala, Trending Topics show a higher *potential-visibility* than newspapers' ads. These results, indicate that despite having a lower exposure time, the higher penetration of Twitter compared to traditional media makes that Trending Topics have a higher potential visibility than radio-stations' commercials in several countries and newspapers' ads in almost every considered country.

Therefore, we conclude that Trending Topics offer a visibility comparable to other traditional ad channels for the analyzed countries. This confirms that Trending Topics are a useful tool for marketing and advertisement purposes.
However, several considerations should be taken into account with respect to our results. First, Trending Topics should not be considered as a substitute to traditional ads channels. Instead, they should be considered a complementary tool in advertisement and marketing contexts. In particular, in traditional ad channels the advertiser buys several slots and it has the certainty that its product would be shown to the audience during those slots. However, the same advertiser may lunch a marketing campaign in Twitter but it has not the guarantee that its product will become Trending Topic. In fact, the definition of strategies to help companies to generate Trending Topics is still an unsolved

matter and requires further research[11]. Second, our analysis has been conducted under certain assumptions. For instance, we have only considered popular radio-stations and newspapers in each country with a higher penetration than the average newspapers or radio-stations in those countries. Furthermore we have considered values that represent realistic advertisement campaigns in newspapers and radio-stations, but other type of campaigns are possible and may lead to different visibility results. Finally, some other subtle aspects such as how the ad support (e.g., audio vs text vs images) affects the attention of the user have not been considered.

4.6 Estimating the economical value of getting a Trending Topic

In the previous section we have analyzed the visibility offered by TTs showing that it is similar to that offered by other traditional media. In this section we would like to extend our previous discussion providing some economic figures. In particular, we analyze and compare the cost effectiveness of an advertisement campaign based on newspapers ads and promoted services in Twitter. For this purpose we rely on a well-known metric in the advertisement industry, the Cost Per Mille (CPM). Furthermore, the conducted analysis helps to better understand what is the value associated to a regular Trending Topic. Before presenting the obtained results we briefly describe the different promoted products of Twitter and explain why we have selected the CPM in front of other existing metrics.

4.6.1 Promoted Products on Twitter

In April 2010 Twitter launched the first phase of its *Promoted Tweets* platform going a step further in Twitter ads [105]. Around six months later *Promoted Accounts* and *Promoted Trends* were added to the suite of Promoted Products [106]. These are paid-products which are purchased by advertisers to amplify their visibility and promote their brands, products and services in Twitter.

In this work we are interested in the *Promoted Trends*. A Promoted Trend, as defined by Twitter [107], is similar to a regular Trending Topic that appears on top of the list of regular TTs. Promoted Trends are purchased in slots of 24 hours and are shown to Twitter users in a given geographical area, typically a country, during these 24 hours. The cost of a promoted Trend varies depending on the geographical location, for instance

[11] Note that companies have the option of purchasing Promoted Trending Topics that follow a similar business model (pay-per-slot) as traditional ad channels.

Promoted Trends are cheaper in Latin America than in US. Specifically, the cost of a 24 hours Promoted Trend has significantly increased from $80K in 2010 to $200K in 2013 [108] in US. Furthermore, the cost also varies for special events. For instance, world wide Promoted Trends are being sold for $600K during the 2014 Soccer World Cup [109]. Finally, there is a difference worth noting between Promoted and Regular TTs. Promoted Trends are shown to all Twitter users within the targeted country C whereas the regular TTs of C are visible to those Twitter users that have selected to see the list of TTs of C. For instance, US Promoted Trends are shown to all Twitter users that log in Twitter from an US IP address, whereas US regular Trending Topics are shown to those Twitter users (regardless if they are in US or not) that select to see the list of US Trending Topics in their account.

The previous information gives us a first rough indication of the value of a regular Trending Topic. For instance, a regular TT in US lasting 24 hours would be valued in ∼$200K. Note that this is a ball park estimation due to the highlighted differences between Promoted and regular TTs.

4.6.2 Metrics to measure the Cost Effectiveness of Advertisement Campaigns

The Cost Per Mille (CPM) is a standard metric to assess and compare the cost effectiveness of advertising campaigns in different media [97]. It is expressed as follows:

$$CPM = \frac{Advertising\ Cost}{Generated\ Impressions} \tag{4.5}$$

where *Generated Impressions* represent the amount (in thousands) of, for example, the circulation (the views) of an ad (Promoted Trend) in newspapers (Twitter). Note that our assumption is that whenever a user logs in their Twitter account one impression is generated with probability equal to the value of α (0.51) introduced in Section 4.4.2. Hence, the total number of daily generated impressions of a Promoted Trend in a country C is computed as:

$$Generated\ Impressions_{TW}(C) = \alpha \cdot \#Daily\ Time\ Line\ Views_{TW}(C) \tag{4.6}$$

where $\#Daily\ Time\ Line\ Views_{TW}(C)$ represents the daily number of timeline views made by active users in C.

There exist other metrics used to assess the cost effectiveness of online advertising campaigns such as the *Cost-Per-Click* (CPC) used, for instance, in banner-based ads or

Cost-Per-Acquisition (CPA) which is mainly used in affiliate marketing. These metrics rely on external information. In the case of the CPC, we need to know whether the ad was clicked and thus resulted in a visit to the webpage of the advertiser. Whereas to compute the CPA we need to know if the click on the ad ended up into a purchase in the advertiser webpage. Contrary to the CPM, these metrics do not work for ads in traditional media. Then, we leverage the CPM in the rest of the section to conduct our comparative study.

4.6.3 CPM of Promoted Trends vs Newspapers' ads

We perform our comparison study for US since it is the country for which we have accurate values for the cost of a Promoted Trend ($200K). Therefore using the definition of CPM and $Generated\ Impressions_{TW}(C)$ defined above[12], we obtain that the CPM for Promoted Trends in US is $0.63. This value is very close to the average CPM ($0.56) reported for Social networking sites [111].

In order to compare the cost effectiveness of Promoted Trends with traditional media, we leverage the CPM reported by the Wall Street Journal (WSJ) [112], a major US newspaper. In particular, the CPM of a full page ad of the WSJ with an associated cost of $150K is $28.81.

The obtained results suggest that the cost effectiveness of Promoted Trends is two order of magnitude larger than the one of ads in major newspapers. This result is reasonable because ads in newspapers involve physical material, qualified personnel and additional resources. Furthermore, as indicated in Section 4.5, our discussion does not take into account some relevant, but hard to measure, aspects such as how different ad supports (e.g., text vs image) influences the attention of the user.

Finally, the conducted analysis helps us to provide a ball park estimation of the economical value of a regular Trending Topic in terms of advertising campaign costs. The obtained results suggest that regular TTs are expected to obtain lower CPM than ads in traditional media (i.e., newspapers) and thus are more cost efficient.

[12]Note that to compute the daily number of timeline views we use the data reported by Twitter about the average number of timeline views and the daily number of active US users in [110].

4.7 Analysis of the variability of TTs visibility within a country

As Figure 4.3 revealed, there exists a notable difference, over an order of magnitude, in the *total active time* across Local TTs in a country. Hence, distinct TTs within a country enjoy rather different visibility. In this section we dig into this difference. In particular, we conduct the following analyses: *(i)* we present a methodology whose aim is to unveil which type of TTs are likely to provide a higher visibility and *(ii)* we study whether the visibility offered by TTs at different times of the day presents an identifiable daily pattern. Due to space limitation, we present the obtained results for three selected countries. Specifically, we have chosen one country from each one of the temporal-visibility groups defined in Section 4.4 to guarantee the diversity in our selection: New Zeland (NZ) from the HtV group, Ireland (IE) from the MtV group and UK from the LtV group. Note that we will refer to results for other countries when warranted.

Before going into those analyses, we would like to briefly extend our comparison between TTs and traditional advertisement channels. In the previous section we have used the median value of the different visibility metrics of TTs within a country to perform the comparison study. However, due to the high variability of TTs visibility in a country, we would like to present more statistically meaningful results. To this end, we have computed the percentage of TTs that present a higher *potential-visibility* than radio-stations' commercials and newspapers' ads for each one of the 10 countries analyzed in Section 4.5. Figure 4.7 shows the obtained results. First, at least 85% TTs present a higher *potential-visibility* than newspapers' ads in all countries but Guatemala in which due to the high penetration of the considered newspaper only the top 1% most visible Trending Topics would achieve a higher visibility than ads in that newspaper. Second, in the case of radio-stations' commercials we observe a high variability in the results. For instance, in FR, GE and ES the visibility of commercials in the considered radio-stations' is higher than for any TT whereas in US we observe the opposite effect, 99% TTs enjoy more visibility than commercials in the considered radio-station. This variability is dictated by the interplay of the penetration of different media as well as the associated *net-visibility*.

In summary, these results confirm the conclusion from Section 4.5: Trending Topics offer a visibility comparable to traditional ad channels and then they are useful as a tool in marketing and advertisement contexts. However, the high variability observed in the visibility of TTs across (and within) countries requires to conduct an individual analysis for each specific case to obtain accurate results.

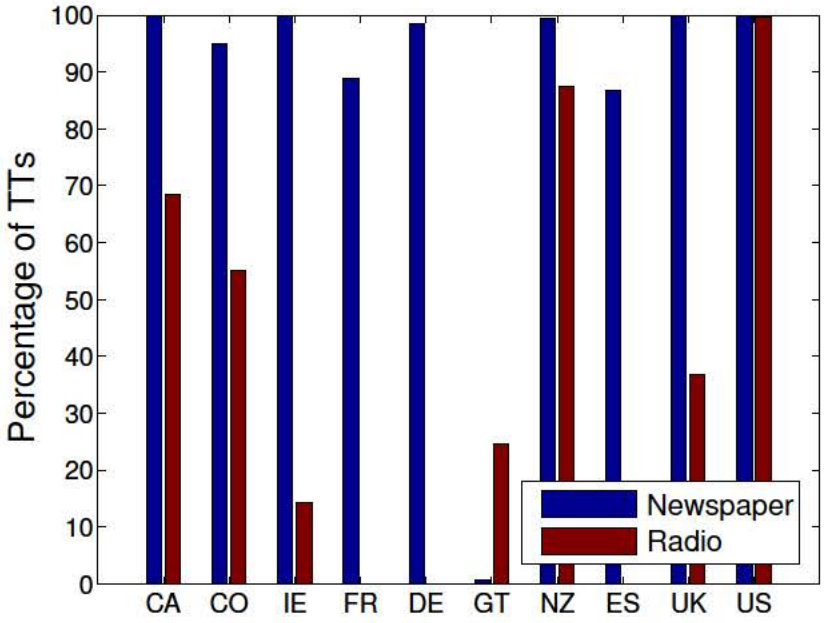


FIGURE 4.7: Percentage of TTs with higher *potential-visibility* than newspapers' ads and radio-stations' commercials for the 10 considered countries.

Finally, we would like to highlight that in order to study the variability of TTs visibility within a country we use the *total-active-time* in the rest of the section. Note that *net-visibility* is a normalized version of this metric and Twitter penetration, used to compute the *potential-visibility*, is the same for all TTs within a country. Then, results derived with the *total active time* and these other metrics are equivalent.

4.7.1 Visibility of different semantic classes of TTs

In this subsection we first define an efficient methodology to group TTs by their semantic meaning into different semantic classes. Then we apply this methodology to the Local TTs of the selected countries. Finally, we compute the distribution of the *total active time* for the TTs within each semantic class so that we can report what types of TTs offer higher visibility in each country.

Methodology

Our tool uses the following sources in order to assign a specific TT to a semantic category:

- *DBpedia*[13] is a sub-project of Wikipedia that aims to create an ontology to classify different names, terms, words and expressions available in Wikipedia pages. In particular,

[13]http://dbpedia.org/About

TABLE 4.2: List of Semantic classes and categories. For each category we also indicate the source as follows: *DBp* for DBpedia, *IMDb* for IMDb and *Self* for Self-defined categories.

CLASS	CATEGORY	EXAMPLE
Hashtags	— **(Self)**	#FirstQuestionsAsked
Sports-Related	Athlete (DBp)	Andrew Sheridan
	Sport Events (Self)	Chelsea - Liverpool
	Other Sport Issues (DBp)	Conmebol
	College Coach (DBp)	Mike McQueary
Feelings & Emotions	Positive (Self)	Happy Birthday Britney
	Negative (Self)	RIP Perez Hilton
Places &Buildings	Architectural Structure (DBp)	US Capitol
	Administrative Region (DBp)	Northern Iowa
	Feature (DBp)	La Cartuja
	City (DBp)	Amsterdam
	Populated place (DBp)	Cannes
Celebrities	Agent (DBp)	Hugh Grant
	Office holder (DBp)	Bill Clinton
	Politician (DBp)	John McCain
	Artist (DBp)	Freddie Mercury
	Famous person (IMDb)	Christine Reyes
	Writer (DBp)	Edgar Allen Poe
Entertainment	Album (DBp)	Love After War
	Movies (IMDb)	Finding NEMO
	Book (DBp)	Geek Love
	Work (DBp)	Reservoir Dogs
	Character (IMDb)	Batman & Robin
	Video games (DBp)	Death Race
	Film (DBp)	Celda 211
	Single (DBp)	Bad Romance
	TV show (DBp)	American Idol
Companies	Organization (DBp)	RTVE, Spanair
	Privately held company (DBp)	Twitter
	Public company (DBp)	Jackson Hewitt
Others	Diseases (DBp)	HIV
	First name (DBp)	Danielle
	Wide-body aircraft (DBp)	Boeing 767
Unclassified	—	Take Facebook Down

it provides a hierarchical ontology that currently covers 359 semantic categories that are described by one or more properties from a pool of 1775.

- *IMDb*[14] is a popular database including information related to a large number of entertainment resources such as movies, TV shows, actors/actresses, etc. Contrary to DBpedia, IMDb does not provide a structured classification for the stored resources.

[14]http://www.imdb.com

TABLE 4.3: Distribution of Local TTs from UK, IE and NZ across the defined semantic classes.

	NZ	IE	UK
Hashtags	53,72%	47,31%	39,13%
Sports-Related	3,46%	5,96%	10,80%
Feeling & Emotions	0,80%	0,84%	0,76%
Places & Buildings	3,86%	7,39%	4,58%
Celebrities	7,45%	9,79%	14,19%
Entertainment	8,64%	7,58%	8,85%
Companies	3,59%	3,95%	3,06%
Others	13,70%	11,67%	8,99%
Unclassified	4,79%	5,51%	9,65%

- *Self-defined categories*: Manual inspection of TTs reveals some common semantic categories that although easily identifiable for a human being are not recognized by either DBpedia or IMDb. In particular, we identify two of these categories: (*i*) *Sport Events*, our manual inspection reveals that TTs are commonly used to reflect events related to different sport games, such as the score of football games. Examples of this are TTs such as 'Arsenal 1-2 Manchester United' or 'Gol de Benzema'. (*ii*) *Feelings/Emotions*, our manual inspection also suggests that TTs are used to express emotions, feelings, preferences, greetings, etc. Therefore it is common to find TTs including words such as 'Happy', 'Love' or 'Hate'. Examples of these TTs are 'Happy Birthday Andy Carroll' or 'We Love Hunger Games'. Therefore, our tool classifies those TTs that include one (or more) emotion-related word(s) and neither DBpedia nor IMDb are able to classify in the *Feelings/Emotions* class.
Moreover, we consider *Hashtags* as a separate category. As indicated in the Introduction hashtags are a special functionality of Twitter that is widely used and thus understanding whether they offer a higher/lower visibility than "non-hashtaged" topics is of high interest for commercial and advertisement purposes.

The large number of potential output categories provided by DBpedia and the lack of structure of IMBb would make infeasible to conduct a meaningful analysis of the semantic context of TTs using their provided results. To address this issue, we have performed a careful merging process in which we group semantic categories obtained from DBpedia, IMDb and our self-defined categories into a handful set of semantic *classes* that permits us to present a meaningful discussion. Note that for this process we have used as reference the 18 classes defined in [49]. Indeed, the 18 classes defined in [49] can be easily merged into the 9 classes resulting from our process (with the exception of hashtags). We have decided to define a smaller number of classes because using 18 classes results in few of them being scarcely populated.

Table 4.2 lists the defined semantic classes and, for each class, presents the most important categories along with its original source (i.e., DBpedia, IMDb or self-defined categories). In particular, we use the following preference order in our semantic classification process for a given TT: we first try to classify it using DBpedia in a semantic category and class. If DBpedia fails we use IMDb and in case it also fails we use our Self-defined categories. Those topics that are not classified after these three steps are added to the *Unclassified* class. Finally, our manual inspection of the TTs within the Unclassified class reveals that most of these topics correspond to complex sentences similar to some hashtags but without the initial '#'. Some examples are: 'Tomorrow is Friday', 'Bieber Fever Is Incurable', 'Ian Is Our Pride', 'M or P' and 'Lin is 6'. It can be noticed that some of them are difficult to be semantically classified even for a human being without the required context knowledge (e.g., 'M or P').

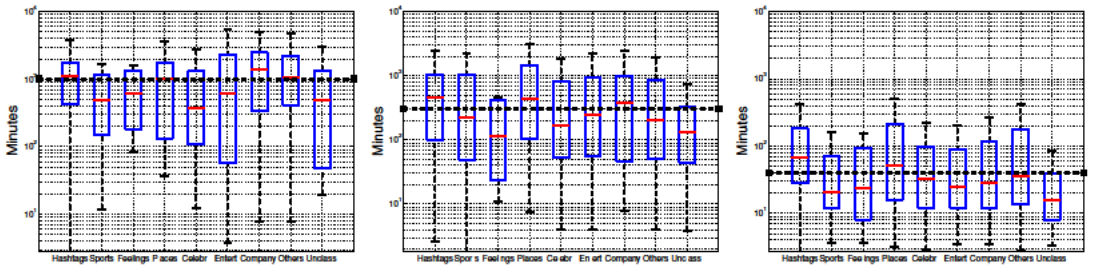
Performance Evaluation

We have used the described methodology to classify the TTs included in our datasets. Table 4.3 summarizes the percentage of TTs that have been classified as well as those that our tool is unable to classify for each analyzed country (Unclassified). The results suggest that our tool is fairly efficient since it is able to automatically classify more than 90% of the TTs in the worst considered case (UK).

However, the effectiveness of a classification tool is not measured by the percentage of resources that it is able to classify but the percentage that it is able to classify correctly. In particular, we define two types of errors for our classification tool: (*i*) *false positives* are those TTs that our tool assigns to a wrong class and (*ii*) *false negatives* are those TTs that our tool was unable to classify but a human being would be able to classify in any of the defined semantic classes.

The detection of false positives and negatives needs to be done manually. Note that this is a common practice used in previous works [49, 50]. Conducting such an experiment for all the TTs from our dataset is a very tedious and time consuming task. Therefore, we have selected a random set of 1000 TTs and three different persons[15] have manually detected the false positives and negatives for this subset of TTs. Note that the differences between the classification done by these three persons over the same random set varies less than 1%. This suggests that the error introduced by human beings is negligible and thus the result of the manual classification can be considered a good approximation to the ground truth. In addition, sampling introduces an error in the proportion of Trending Topics per category used during the validation with respect to the actual

[15] These three persons were not connected to our research project to guarantee the objectivity.



(a) Active Time NZ (b) Active Time IE (c) Active Time UK

FIGURE 4.8: Distribution of the *active time* across TTs within each semantic classes for NZ, IE and UK (the horizontal dashed line shows the median active time of all the Local-TTs of the correspondent country).

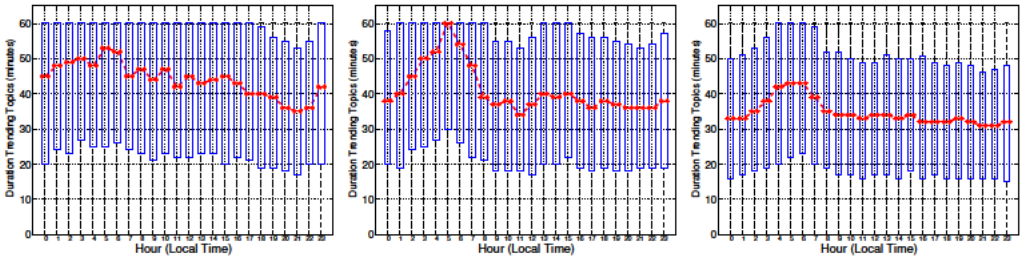
proportions. This error can be computed using a hypothesis test for a proportion [113]. This is a well-known tool widely used to compute confidence intervals for the results of surveys. In particular, in our case in which we use a sample of 1K TTs, the error introduced by sampling in the proportion of Trending Topics in any class is $\leq 3.1\%$ (with 95% confidence) for any size (i.e., number of Trending Topics) of the dataset. This suggests that: first, the obtained results are reasonably accurate and, second, the used methodology scales well since manually inspecting a sample of 1000 TTs (that as we have demonstrated is doable for a human being) suffices to not incur in high errors in the considered proportions for different classes.

Our detection experiment reveals that, one the one hand, 41% of the unclassified TTs are false negatives. Since the *Unclassified* class represents less than 10% of our TTs, we conclude that overall only around 4% of the TTs corresponds to false negatives. On the other hand, false positives are also infrequent and represent only 5% of the inspected TTs. In a nutshell, these results indicate that our semantic classification tool is quite accurate and its automatic process is able to classify more than 91% of the TTs correctly.

Visibility of TTs across semantic classes

Figure 4.8 depicts the distribution of the *total-active-time* for every semantic class of the three analyzed countries in the form of boxplot. In addition, we plot a horizontal dashed line that indicates the median *total-active-time* for all TTs in the country for reference.

First of all we observe a high variability among the visibility offered by different TTs within each class. Despite this variability, we still can derive useful observations. For instance, "Hashtags" and "Places" are the only two cateogries whose median *total active time* is above the median value of the country, for all three countries. Interestingly, this result along with results in Table 4.3 suggest that adding a # in front of the term to be advertised seems to increase the chances to become TT and to enjoy a longer



(a) Active Time NZ (b) Active Time IE (c) Active Time UK

FIGURE 4.9: Distribution of the *total active time* of TTs instances within each one of the 24-hour slots of a day.

active time. Surprisingly, categories such as "Sport" and "Celebrities" that attract a fair amount of attention from Mass media do not appear among those offering higher visibility. This may indicate that Twitter users do not get excited about these topics for long time. Finally, we observe differences across countries that indicate that each *national market* shows preferences for different types of topics. For instance, TTs related to companies present the highest visibility in NZ whereas TTs in this category show a rather low active time in UK. Furthermore, TTs related to "Sports" present a quite low visibility in NZ and UK but not in IE.

4.7.2 Daily Pattern of Trending Topics Visibility

Internet traffic as well as most on-line services present a daily usage pattern bound to the daily schedule of their users [86]. In this subsection we focus on understanding whether the visibility offered by Local TTs presents an identifiable daily pattern. For this purpose, we divide a day in its 24 one-hour slots[16] and for each slot we calculate the distribution of the *active time* for the TT instances present in that slot. Note that the maximum *active time* that a TT instance can have in a slot is 60 minutes.

Figure 4.9 shows the obtained results for UK, IE and NZ. The x-axis shows the 24 time slots described in the previous paragraph and the y-axis shows the distribution of the *active time* of the TTs present in each time slot in the form of boxplot. Note that the time slots represent local time for each country. We observe that there is a marked daily pattern in the distribution of the *active time* for the different hour-slots. In fact, for every country we can see the presence of few slots where TTs tend to have a higher *active time*. Specifically, these slots correspond to the night (sleeping) hours in which a lower activity of Twitter users helps TTs to remain visible longer time. However, the higher *net-visibility* enjoyed in those hours does not really lead to a higher *potential-visibility*

[16]Slot 0 includes information for the 60 minutes between 12AM and 1AM, slot 1 includes information for the 60 minutes between 1AM and 2AM and so on.

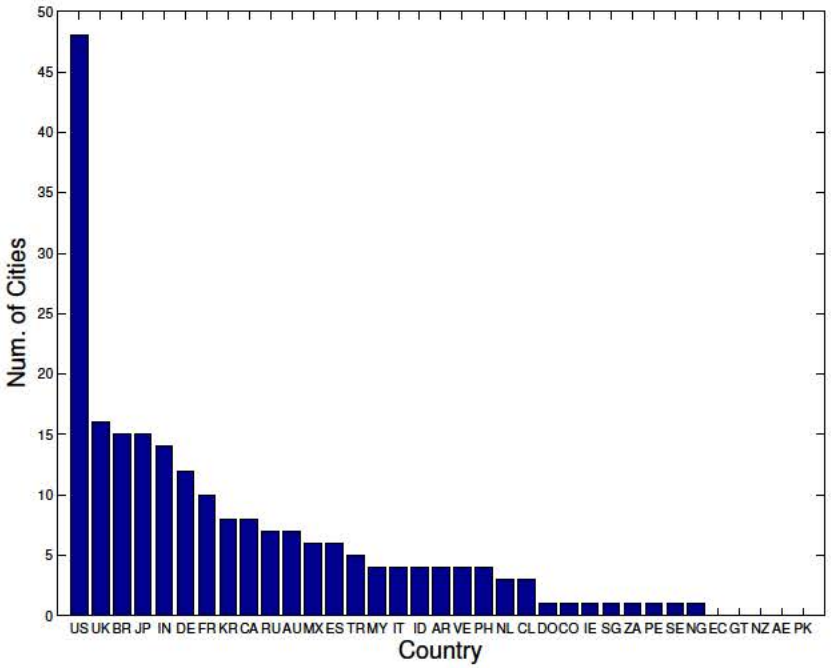


FIGURE 4.10: Number of cities associated to each country in our dataset.

since the number of users connected to Twitter at those hours is likely to be significantly smaller than in the morning, afternoon or evening. In summary, most of the countries show the previously reported daily pattern, with few exceptions such as Japan, US and some Latin-American countries (e.g., Colombia or Venezuela), in which we observe a flatter shape. Thus, the difficulty of getting a TT in these countries is independent of the time of the day. Finally, we have separately studied the daily-pattern for week days and weekends for every country without noticing major differences.

4.8 Analysis of City-Level Trending Topics

To complete the thorough analysis of the visibility offered by Trending Topics within a specific country discussed above, in this section we give one step further to analyze the visibility of city-level TTs. We first characterize the city-level TTs for the different countries in our dataset and afterwards we discuss their visibility aspects.

4.8.1 Characterization of City-Level Trending Topics

At the time of the data collection procedure Twitter offered City-Level Trending Topics for 215 cities across 30 countries. Figure 4.10 shows the distribution of number cities offering City-Level TTs across the 35 countries in our dataset. We observe that the distribution is not homogeneous. Indeed, there are 5 (GT, PK, NZ, EC and AE) and 8

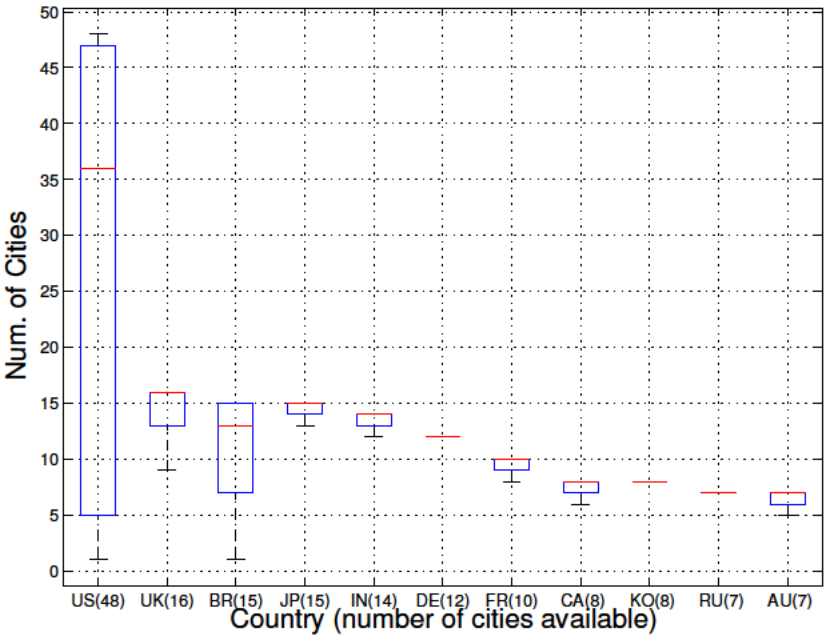


FIGURE 4.11: Distribution of the number of cities that each City-Level TT reaches for the studied countries. Number of available cities for each country indicated in brackets in the x-axis.

(DO, CO, IE, SG, ZA, PE, SE and NG) countries with none or just one associated city, respectively. The countries with a larger number of associated cities are US (48), UK(16), BR (15) and JP (15). This heterogeneity seems to be associated to the variability of the penetration of Twitter in different countries. In order to offer the TT functionality a minimum volume of active users in a city is required. Then, those countries showing a larger number of Twitter users seem to have a larger number of cities with the City-Level TTs feature enabled.

Next we aim to analyze the homogeneity of the Trending Topics appearing across different cities within a country. For this purpose, for every City-Level TT in a country we compute the number of cities in which it appeared. Figure 4.11 shows the distribution of the number of cities associated to each City-Level TT for the 11 countries with at least 7 cities in our dataset[17] in the form of boxplot. We observe that for all countries the median of the distribution is higher than $N/2$ (being N the number of cities associated to the country). This means that for every country 50% of the City-Level TTs appeared at least in half of the cities. Furthermore, for all countries, excepting US, UK and BR, at least 75% of City-Level TTs have appeared in N or N-1 cities (e.g., in JP that has 15 associated cities, 75% of the City-Level TTs have appeared in 14 or 15 cities). Indeed, for few countries (DE, KO and RU) every City-Level TT appears in all cities.

[17]We have performed the same analysis for the rest of countries obtaining similar results.

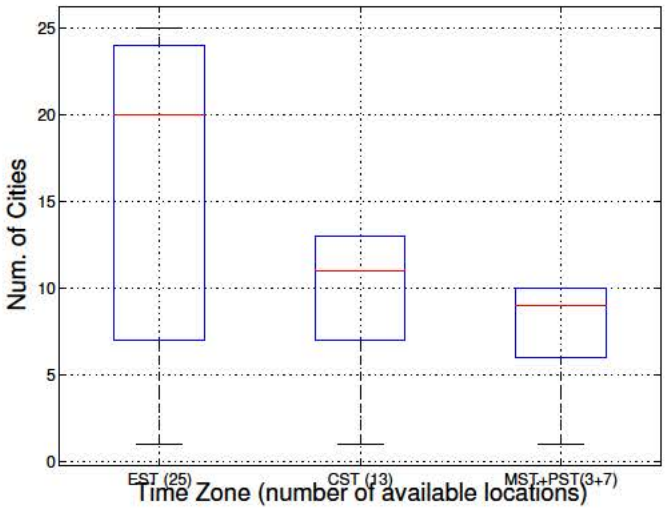
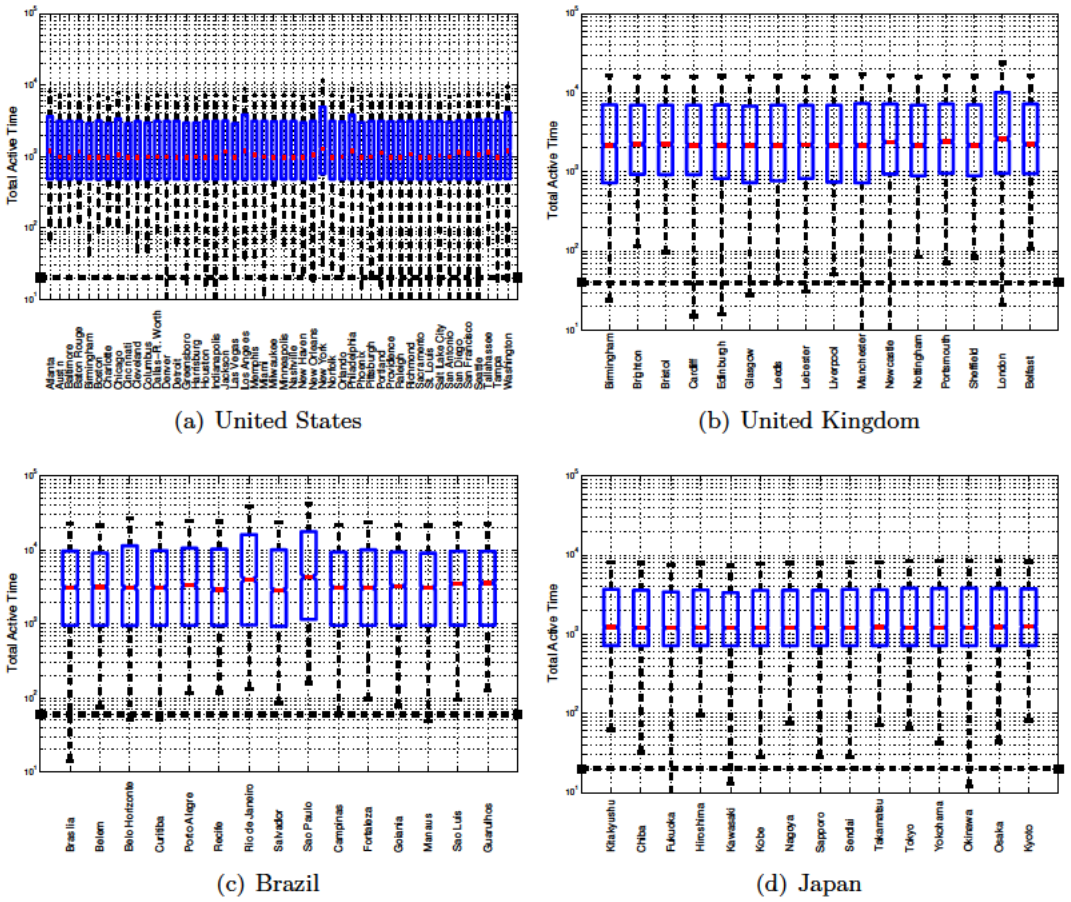


FIGURE 4.12: Distribution of the number of cities that each City-Level TT reaches for the US cities in the EST, CST and MST+PST times zones. Number of available cities for each time zone indicated in brackets in the x-axis.

If we now focus on the 3 exceptions, on the one hand, UK presents a similar behaviour to the one just described since 75% of its City-Level TTs appear in at least 13 of the 16 associated cities. On the other hand, BR and US present a complete different behaviour. The Inter-Quartile Range (IQR) varies between 5 and 47 cities in US and between 7 and 15 cities in Brazil. We conjecture that the fact that Brazil and US are large countries with a large number of Twitter active users lead them to present the reported different behaviour. We further analyze this phenomenon for US since it presents the most distinct behaviour. To this end, we have divided the cities of US into three groups based on their time zone: Eastern Standard Time (EST) including 25 cities, Central Standard Time (CST) including 13 cities, and Mountain Standard Time and Pacific Standard Time (MST-PST) that includes 10 cities. Figure 4.12 shows the distribution of the number of cities in which a City-Level TT appears for each one of the defined groups. The results indicate that even dividing US into its 3 main geographical areas, the IQR is significantly larger than in any other studied country but Brazil. This result suggests that the size of the country is not as determinant as the level of activity in the heterogeneity of the City-level TTs sharing across cities.

In summary, the obtained results indicate that a large fraction of City-Level TTs are shared among (almost) every city in a country, excepting for Brazil and US. Thus, it seems that national events, rather than local events, attract the attention of Twitter users in a country.



(a) United States (b) United Kingdom

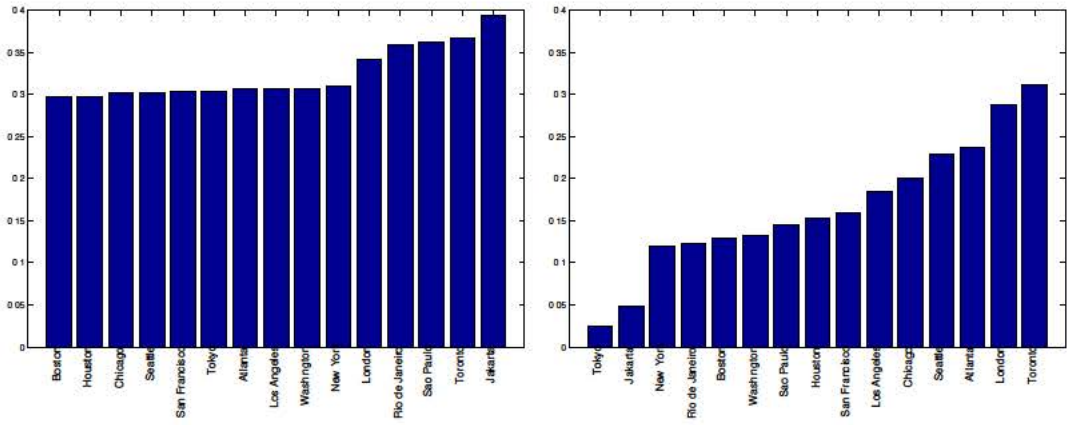
(c) Brazil (d) Japan

FIGURE 4.13: Distribution of *Total Active Time* for the City-Level TTs for each city in US, UK, BR and JP (the horizontal dashed line shows the median Total Active Time associated to the correspondent Country-Level TTs).

4.8.2 Analysis of the Visibility of City-Level Trending Topics

In order to analyze the visibility associated to the City-level TTs in a country, we have computed the Total Active Time for the Trending Topics of each city for the four countries with the largest number of cities in our dataset (US, UK, BR and JP). Figure 4.13 presents the distribution of Total Active Time for the Trending Topics of each city in the studied countries. Note that the distribution is shown using a boxplot format and the y axis is in log scale. Furthermore, we depict the median Total Active Time associated to the Country-level TTs with an horizontal dashed line for each country.

The obtained results highlight the following interesting findings: (*i*) For every analyzed country, the median of the Total Active Time of Country-Level TTs is roughly 2 order of magnitude smaller than the median Total Active Time for the TTs of any city. This result is expected since the competition to become Country-level TT is much higher than at the City-Level (all City-Level TTs compete among them to become Country-level TT). This leads to a higher dynamicity in the composition of the Country-Level TTs



(a) Top Cities: Net-Visibility (b) Top Cities: Potential-Visibility

FIGURE 4.14: Visibility metrics for the City-Level TTs of the Top 15 cities in Twitter.

list and thus a lower Total Active Time. From a marketing/advertisement perspective this has important implications since a marketing/advertisement campaign focused on a specific city can achieve a high visibility through City-Level Trending Topics; (*ii*) The distribution of Total Active Time remains constant across the cities of a specific countries with the exception of few cities in US, UK and BR. Unexpectedly, we observe that these exceptions correspond to the largest metropolitan area cities [114] of each country: New York, Los Angeles, Philadelphia and Washington in US, London in UK and, Rio de Janeiro and Sao Paulo in BR. Hence, larger cities get higher Active Time than those with smaller populations. Again this result seems to have clear implications in the field of marketing and advertisement.

To conclude our analysis we compare the the visibility of City-level TTs for the 15 cities with a larger number of users in Twitter obtained from [115]. To this end we leverage the methodology described in Section 4.4 and compute the *net-visibility* and *potential-visibility* for these cities. Figures 4.14(a) and 4.14(b) show the *net-visibility* and *potential-visibility* for each one of the studied cities, respectively. Note that the cities are sorted from left to right from the lowest to the highest value. As occurred at the Country-level TTs, the *net-visibility* and the *potential-visibility* depict very different stories due to the different penetration of Twitter in the different studied cities. If we focus on the *net-visibility*, we observe that all the considered US cities offer the lowest values along with Tokyo whereas Jakarta and Toronto offer the highest values. However, when we factor in the Twitter penetration in the city through the *potential-visibility*, the US cities spread across the whole Ranking. Furthermore, the low penetration of Twitter in Jakarta makes it to drop to the penultimate position in the *potential-visibility* ranking. Finally, it is worth noting that Toronto pass from the Top 2 to the Top 1 position. This

indicates that the visibility offered by TTs in Toronto make them a valuable marketing tool in this city.

4.9 Discussion

Despite Trending Topics are a well-know feature regularly exploited in the context of marketing and advertisement, we still stand on preliminary ground in terms of understanding this tool. In this chapter we characterize the visibility of Trending Topics across 35 countries. In particular, we present a measurement methodology along with a methodology to thoroughly analyze the visibility of Trending Topics that we believe can be of high value for experts of different disciplines in marketing and advertisement contexts. The results obtained applying these methodologies indicate that, in general, Trending Topics present a comparable visibility to other traditional advertisement channels and thus they can be considered a useful tool in marketing and advertisement contexts. However, the high variability on the visibility offered by Trending Topics across (and within) countries suggests that we should apply the described methodology to obtain accurate results for each specific case.

Chapter 5

Collective Advertising: Information Flow

5.1 Motivation

Online social media has recently irrupted as the last major venue for the propagation of news and cultural content, competing with traditional mass media and allowing citizens to access new sources of information. In this chapter, we study collectively filtered news and popular content in Twitter, known as Trending Topics (TTs), to quantify the extent to which they show similar biases known for mass media. We use two datasets collected in 2013 and 2014, including more than 300.000 TTs from 62 countries. The existing patterns of leader-follower relationships among countries reveal systemic biases known for mass media: Countries concentrate their attention to small groups of other countries, generating a pattern of centralization in which TTs follow the gradient of wealth across countries. At the same time, we find subjective biases within language communities linked to the cultural similarity of countries, in which countries with closer cultures and shared languages tend to follow each others' TTs. Moreover, using a novel methodology based on the Google News service, we study the influence of mass media in TTs for four countries. We find that roughly half of the TTs in Twitter overlap with news reported by mass media, and that the rest of TTs are more likely to spread internationally within Twitter. Our results confirm that online social media have the power to independently spread content beyond mass media, but at the same time social media content follows economic incentives and is subject to cultural factors and language barriers.

5.2 Introduction

In a formal sense, we analyze collective information filters, which contain selected information of temporal relevance that gains special visibility in a centralized communication channel [116]. While the datasources for mass media are defined by newspapers, magazines, and television programs, such content in online social media appears in link sharing communities, like Reddit [117], and aggregation mechanisms in microblogging platforms. In this chapter, we focus on Twitter Trending Topics (TTs), which serve as a global filtering mechanism for Twitter users to select which content is relevant and when. This way, TTs serve as a centralized channel of communication from many to many, serving the purpose of central media within Twitter due to the public interest, temporal component, and global reach of TTs through the Twitter interface. We present our study of media biases in TTs in two large-scale datasets from 2013 and 2014, detecting temporal patterns of TT appearance that reveal the network of media influence across countries. Based on that network, we perform a confirmatory analysis to test systemic and subjective biases in social media, and an exploratory analysis on the relation between TTs and mass media.

5.3 Materials and Methods

5.3.1 Data on Trending Topics

We gathered a dataset of Local TTs in Twitter: Phrases or words that appear in Twitter at a much higher frequency than the rest in a certain country during a short period of time. Local TTs serve as crowdsourced aggregators of recent topics of high relevance and popularity, filtering the activity of a country through a centralized channel of temporally relevant information. Among the information provided by its REST API, Twitter provides the list of 10 Local TTs at query time in a given location (e.g., a country). We retrieved the list of TTs at all the available countries in an automated fashion, gathering the list of Local TTs for each country every 5 minutes, which is the interval used by Twitter to update the list of TTs [118].

We executed our retrieval software in two study periods: February 20th to May 20th, 2013 and April 14th to July 4th, 2014. During each study period, we collected information on all available countries with Local TTs provided by the API, which amounts to 35 countries in the first period and 62 countries[1] in the second one (including the 35

[1]Countries in Twitter are identified by their ISO-3166 codes. Note that we treat the US territory of Puerto Rico as a country to be consistent with the Twitter API.

previous ones). The first study period produced a dataset of more than 112.000 Local TTs and the second period a set of more than 188.000 TTs. We refer to these datasets as TT-2013 and TT-2014, respectively, and are publicly available at [119]. Finally, it is worth noting that our data collection methodology is compliant with Twitter's terms of Use and Service[2]. Moreover, since we are able to collect all TTs for every available country at a rate equal to the TTs' update interval configured by Twitter, we conclude that there is no bias in the obtained results due to the data collection methodology.

5.3.2 Data on demographic, economic, and cultural factors

For each of the 62 countries, we collect a dataset to quantify variables that are subject to induced biases in the appearance of TTs. We measure the wealth of a country through its GDP using purchasing power parity rates as reported by the World Bank statistics of 2010[3]. We also obtain the most recent directed migration statistics (from 2000) between each pair of countries from the World Bank database. For each country, we set its timezone as the UTC offset in winter time of the largest city as reported by Wikipedia. Furthermore, we also extract from Wikipedia a list of official and national languages in each country.

We quantify the culture of each country c using the four principal cultural dimensions of Hofstede's model [52]: power distance p_c, individualism i_c, masculinity m_c, and uncertainty avoidance u_c. Our intention applying Hofstede's dimensions is not to explore the role of any particular aspect of culture, but to quantify the distance between cultures as an aggregate of the differences in the shared values of two societies. While some countries are not included in Hofstede's dataset, 31 countries from TT-2013 and 43 countries from TT-2014 are present.

5.3.3 Detecting mass media influence in Twitter

We developed a method to determine if a TT appearing in country c at a given date d is related to content reported in traditional mass media (e.g., a newspaper). If a TT is related to at least one mass media item in the country c during a time window of $d \pm N$ days, we consider the TT as *External*, and *Internal* to Twitter if it did not appear in mass media. We use Google News to detect if a TT is reported in the mass media of the country, querying for certain terms related to the TT for the interval of N days before and after the TT emerged. We apply this method to the TTs of four countries (US, ES,

[2] https://twitter.com/tos https://dev.twitter.com/overview/terms/agreement-and-policy

[3] Data on purchasing power parity from Argentina during the study period is not available in the World Bank.

CA and GB) collected over a period of time of 1 month between April and May, 2014. Furthermore, we consider a time window of 5 days ($N = 2$) around the appearance of the TT.

For a given TT, our method is divided into a *pre-processing phase* to translate the TT into an appropriate format to query the Google News service, and a *search phase* in which the Google News service seeks for news including the words forming our TT.

The **pre-processing** phase is divided into two steps. First, (when required) we transform the TT in a set of meaningful words. For instance a TT "#BarackObamaInNewYork" would be transformed into "Barack Obama In New York" (the # is removed and the words forming the TT are properly separated). Second, we filter all the words contained in the lexicon of the top 1000 most frequent words of the language of the country, constructing a set of terms to query without common words that can easily produce false positives. If all the words in the TT are included in the most frequent lexicon, that TT is automatically filtered out in this phase.

In the **search** phase we access the Google News service with the following information: (*i*) we query for the keyword(s) produced by the pre-processing phase; (*ii*) a filter on a specific media outlet depending on the country as explained below; and (*iii*) a time window to perform the search. Specifically, we configure the search of Google News to only consider news pieces that were produced within the designated period, and to look for the keyword(s) in headline and the body of the articles. Therefore, if the result of the *search* phase includes at least one piece of news, we classify the associated TT as External (i.e., reported in Twitter and mass media), and categorize the TT as Internal (i.e., exclusively reported in Twitter) otherwise. Moreover, we also record the earliest date of the news returned by Google News for each External TT. We acknowledge that the proposed methodology may produce some miss-classifications. For instance, a TT #BarackObamaInNewYork would be classified as external if a news piece with a headline *'Barack Obama's administration passes a new Education bill'* appears in the considered time interval ($d \pm N$ days). Despite both events are different, both of them present the common words 'Barack Obama'. To the best of the authors knowledge there is not an available ground truth dataset that would allow to validate our methodology. In the absence of such ground truth, we provide an initial validation of our method through a qualitative pull-out [27] using a random set of 1000 TTs from Spain. An independent rater (a student not taking any other part on this research) was instructed to indicate whether the semantic content of the word(s) of a TT describes current news or not (e.g., in the case of including exclusively very common words). Therefore, those TTs classified by the rater as not news should be filtered out by our algorithm in the pre-processing phase. The result of this pull-out is that 96% of the considered TTs are equally classified

by the rater and our method. A χ^2 test on this result shows that the 95% confidence interval of the accuracy of our method is $[0.945, 0.971]$, and thus it is safe to assume that our method accurately filters out those not newsworthy TTs during the pre-processing phase.

We have applied the described method to the TTs of four countries (US, ES, CA and GB) collected over a period of time of 1 month between April and May, 2014. This includes 5297, 2598, 2385 and 4598 TTs from US, ES, CA and GB respectively. When accessing the Google News service, we focus on the online versions of three large news papers of each country[4]. Note that considering all indexed media by Google News in a country would generate significant noise (e.g., blogs that report the list of Trending Topics) in our results. Instead, the online version of main newspapers provide a broad coverage of different type of news (breaking news, politics, sports, society, science, etc) and at the same time are quite dynamic venues that rapidly report any major event happening. Therefore, they are an appropriate venue to identify the most relevant news in a country. Finally, we filtered news reporting for a time window of 5 days around the appearance of the TT, and repeated the same detection technique for a window of 7 days, obtaining very similar results due to the breaking nature of news.

5.4 Results

5.4.1 Leader-Follower structures

The temporal sequence of appearance of Local TTs allows us to analyze the structure of leader-follower relationships among countries in Twitter. This type of relationship constitutes a media bias in which the temporal patterns of news and popular content are a manifestation of an alignment of incentives, rather than a causal relationship. Herman and Chomsky detail this kind of bias in their Propaganda model [29], operationalizing it as a *sourcing filter* in which some sources are overlooked, distorting the information presented to the public. Thus, leader-follower relationships can appear without a hidden power that manipulates media outlets in different countries; they can be the product of a set of shared interests that create ordered patterns where content originates and where it is consumed afterwards.

We detect leader-follower relationships among countries through the ordering of the appearance of Local TTs. If such relationship exists, there will be a tendency for Local

[4]New York Times, Washington Post and USA Today for US. El Pais, El Mundo and ABC for ES. The Globe And Mail, National Post and Vancouver Sun for CA. Telegraph.co.uk, The Guardian and The Independent for GB.

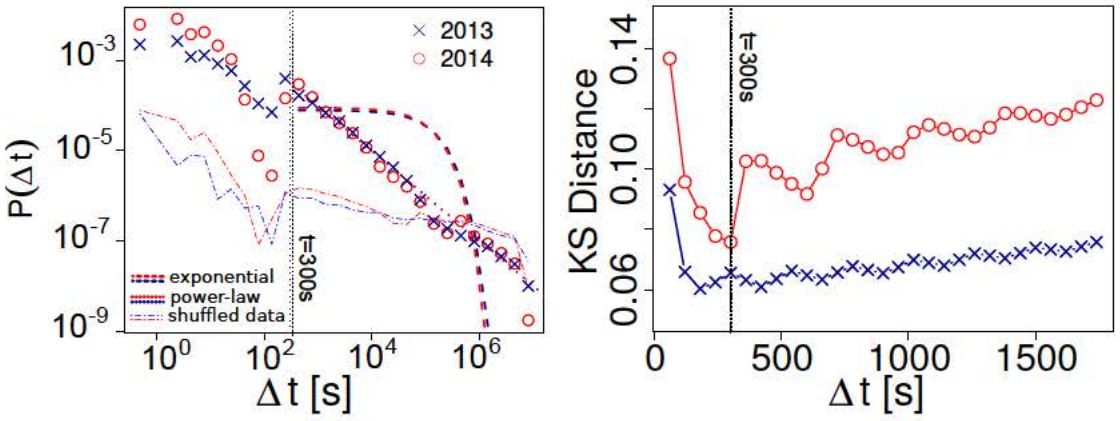


FIGURE 5.1: **Analysis of time intervals between the appearance of TT.** Left: distribution of time intervals, dashed lines show power-law, exponential fits, and distributions for shuffled datasets. Right: KS Distance between the empirical distribution and a power-law fit for ranging values of the minimum Δt of the fit.

year	N	xmin	α	KS	λ	R	p-value
2013	152938	300	$1.000047(10^{-6})$	0.066	$7.9 \cdot 10{-}6$	198.2	0.0
2014	279247	300	$1.0041(10^{-5})$	0.076	$9 \cdot 10{-}6$	293.03	0.0

TABLE 5.1: **Summary of inter-event distribution fits.** Statistics of power law and exponential fits for inter-event distributtions in TT.

TTs to appear in the *leader* country before they emerge in the *follower* country. Thus, we take into account in our analysis the events in which a Local TT appears in country C_i at time t_i, and later in country C_j at time $t_j > t_i$. To test if these temporal sequences are not the product of independent events unrelated to leader-follower relationships, we apply the model of priority processes and bursty patterns in queue theory [120, 121]. If Local TTs appear in pairs of countries by chance, as the result of independent phenomena, the time intervals between the appearances Δt will follow an exponential distribution $P(\Delta t) \sim \lambda e^{-\lambda \Delta t}$ as the result of decoupled Poisson processes. On the other hand, if the appearances of Local TTs are correlated in time, the time sequence will make the distribution of Δt to follow a power-law $P(\Delta t) \sim \Delta t^{-\alpha}$ [121]. The exponent α of this power-law is characteristic of which kind of dynamics produce the correlation. The case of $\alpha = 2.5$ corresponds to exogenously triggered dynamics in which external events produce the TTs [122]; $\alpha = 1.5$ indicates an endogenous process with leader-follower relationships with infinite queues; and $\alpha = 1$ top the same process but with finite queues [121]. This kind of temporal patterns are known to appear in communication processes, including the correspondence of Einstein and Darwin [120], e-mail communication [121], chatroom interaction [123], and Twitter dialogues [124].

The left panel of Figure 5.1 shows the distributions of times between the appearance of TTs in different countries $P(\Delta t)$, for both TT-2013 and TT-2014 datasets. To test the alternative hypotheses of the existence or nonexistence of correlations between TTs

appearances, we fit power-law and exponential distributions through maximum likelihood and the Kolmogorov-Smirnov criterion [125] (see more details in Table 5.1). The resulting theoretical distributions are plotted in Figure 5.1, suggesting that a power-law fit is better than the exponential alternative. Log likelihood ratio tests [126] between the exponential and power-law models give significant estimates of 198.22 (TT-2013) and 293.03 (TT-2014), providing very strong support to reject the independent events hypothesis, in favor of the hypothesis that the appearance of TTs follows a correlated process. We further test this conclusion by shuffling the TTs coocurrence data, permuting the timestamps of appearance of each TT in leader and follower countries. This way, we randomize the pairs, but keep the empirical properties of the distributions of TT appearance in each country. The probability of finding a short interval between the appearance of TTs in this ancillary dataset are much lower than in the empirical distribution, as shown in the left panel of Figure 5.1. Furthermore power-law fits fail in comparison to exponential fits, allowing us to conclude that the empirical shape of $P(\Delta t)$ is far from random.

The exponent of the fits are $\alpha_{TT-2013} = 1.00005 \pm 10^{-6}$ and $\alpha_{TT-2014} = 1.004 \pm 10^{-5}$, indicating that the process of TTs appearance is not due to exogenous factors but depends on leader-follower relationships with finite queues [121]. This also indicates that countries have strong limitations in their capacity to adopt TTs in comparison to how many are generated in the rest of the world. However, note that these correlations are not observable at all timescales, as the power-law distributions are fit above a minimum value of Δt. The right panel of Figure 5.1 shows the Kolmogorov-Smirnov estimate (KS) for a set of minimum values, revealing that the minimum for TT-2013 and TT-2014 are 3 and 5 minutes, respectively. This means that correlations at a timescale smaller than 3 minutes cannot be observed in any of our datasets, and that between 3 and 5 minutes the results are inconsistent. Thus, we take 5 minutes as a minimum criterion to deduce the manifestation of a leader-follower relation. This coincides with our sampling frequency, allowing us to remove from the data those cooccurrences for which we do not have sufficient evidence to consider them product of a leader-follower process.

We quantify the tendency for country C_f to follow country C_l through the amount of TTs that appeared in C_f at least 5 minutes after they appeared in C_l. When applied to every pair of countries in our dataset, these counts define a weighted, directed network in which nodes are countries and links have weights corresponding to the number of TTs that appeared in the leader-follower relationship. We refer to this graph, shown in Figure 5.3 for TT-2013 and TT-2014, as the International Structure of TTs. Finally, we refer to the countries in which a given TT appears in the first Δt (i.e., 5 minutes) after the surge of a TT as *sources* since based on our model they cannot be followers of any other country for this TT.

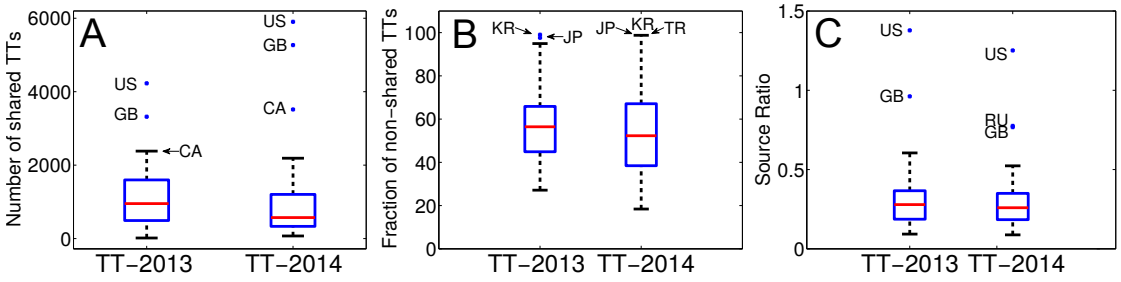


FIGURE 5.2: **Heterogeneity in TT sharing behaviour**. Boxplots of the number of shared TTs (A), the fraction of non-shared TTs (B) and the source ratio (C) for the countries in both TT-2013 and TT-2014 datasets. Red lines show median, boxes 25 and 75 percentiles, bars represent 1.5 times the interquartile range, and outliers are marked.

5.4.2 Heterogeneity in sharing behaviour

For each country in our TT-2013 and TT-2014 datasets, we computed the total number of associated TTs as well as its break down into those that remain local within the country (i.e., are not shared with other countries) and those shared with at least another country. These metrics allow us to understand how much individual countries share TTs internationally. In addition, a critical aspect to define the importance of a country is its capacity to generate TTs that are afterwards consumed by others. According to our leader-follower model, this happens when a country is a source for a TT that is afterwards consumed by others. To characterize the bias of countries towards generating or consuming TTs we compute the *Source Ratio* (SR) as the ratio between the number of TTs in which the country acts as a source and the number of TTs in which it is not a source but a consumer.

Figure 5.2 presents the distribution of the following three metrics across the countries represented in TT-2013 and TT-2014 in the form of a boxplot: (*i*) total number of shared TTs, (*ii*) fraction of non-shared TTs and (*iii*) SR. Overall, in both datasets we observe a significant heterogeneity across the three metrics. For instance, in the case of TT-2014, the interquartile range and the max-min difference for each metric are: (*i*) 871 and 5832, for the number of shared TTs, (*ii*) 28.69% and 80.37%, for the fraction of non-shared TTs, and (*iii*) 0.17 and 1.25, for the SR. Note that, as Figure 5.2 shows, the results are similar for TT-2013.

As seen before, the TTs of a country can be divided in those that stay local in the country (non-shared TTs) versus those that are shared. Furthermore, among the shared TTs, each country might act as a source if it is among the first set of countries in which the TT appeared, thus leading some following countries and contributing to the corresponding links in the network of Figure 5.3.

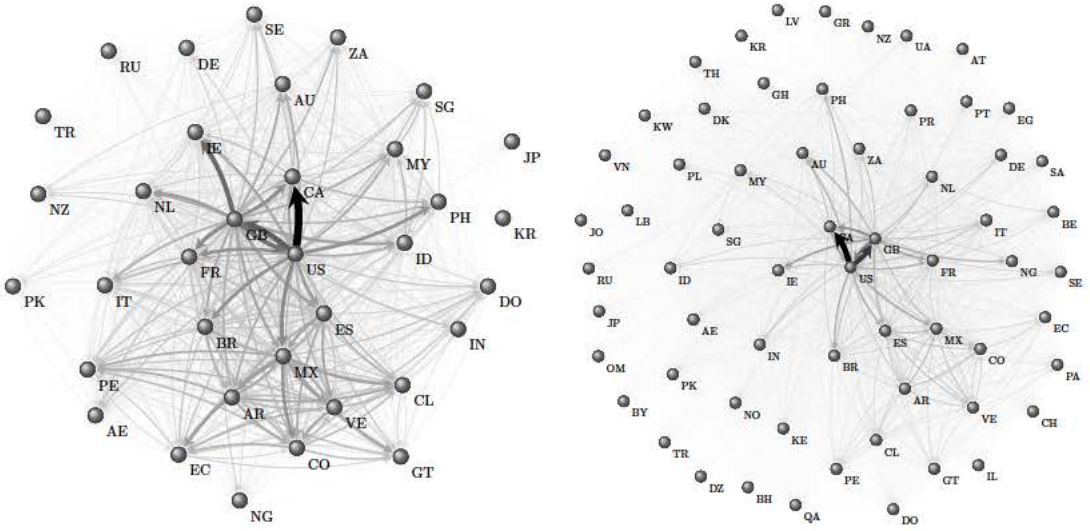


FIGURE 5.3: **Visual Representation of the International structure of TTs.** TT-2013 is shown on the left and TT-2014 on the right. The size and darkness of the links are proportional to their weights.

The descriptive statistics on the amounts and ratios of shared TTs and on the frequency at which countries act as sources show two kinds of outliers. First, some countries are practically isolated, for example Japan, Korea, and Turkey have more than 95% non-shared TTs, while the median is below 60%. Second, some countries have extremely central roles, such as the United States and Great Britain, sharing 67% and 50% more TTs and acting as source in two to three times more tweets than the third largest contributor (Canada).

The above observations of heterogeneity are consistent with a centralization pattern, which we empirically test to assess if countries pay a significantly higher attention to small groups of other countries. We measure the concentration associated to both leading and following activity at the individual country level, through the analysis of the distribution of in and out weights of the links in the international structure of TTs shown in Figure 5.3. We quantify the centralization of attention from and to a country C_i as the Gini coefficient of the weights of outgoing and incoming links, respectively. The Gini Coefficient measures inequality, varying from 0 (complete equality) to 1 (complete inequality). In our case, a high Gini coefficient is an indicator of centralization in either the attention a country pays or in the audience that follows the TTs of a country. As a result, for each country in each dataset, we produced two values of *leading* Gini and *following* Gini, each one measuring the local centralization around the country in either in-degree or out-degree.

To assess the statistical significance of the Gini coefficients, we compare the empirical

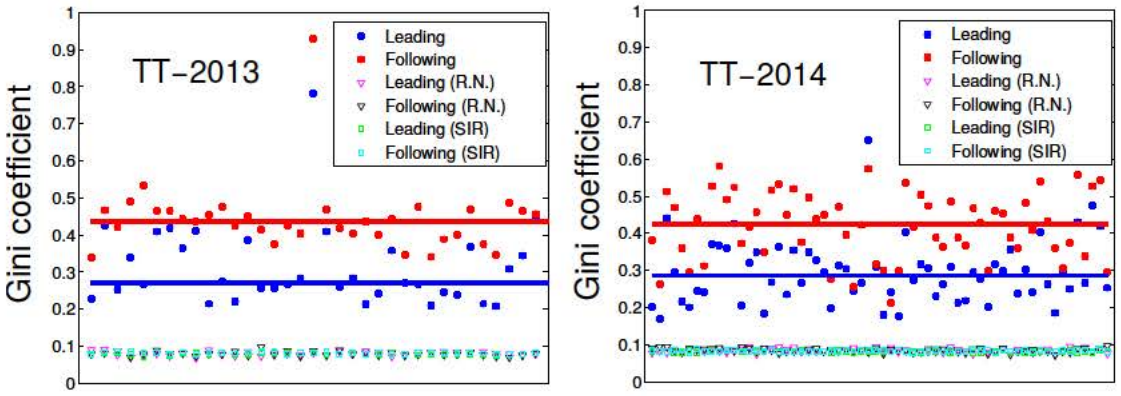


FIGURE 5.4: ***Leading* and *Following* Gini coefficients.** Each Gini coefficient is calculated for the networks of leading-following relations in each dataset and mean values for ten simulations of random networks and SIR processes. Horizontal lines show means over all countries.

estimates with two null models: an Erdos-Renyi random network (RN) [127] and a Susceptible-Infected- Recovered (SIR) process [128]. For the RN case, we produced ten random networks for each dataset, maintaining the amount of nodes, links, and total weight of the original graph. For the SIR case, we simulated ten times a simplified version of the process in the network [129] using the parameter $\beta = 4/N$, which fits the density of our empirical networks. For each simulation of RN and SIR we computed the leading and following Gini of each node in the same manner as we did for the empirical network.

Figure 5.4 shows the leading and following Gini coefficients for each of the countries in TT-2013 and TT-2014, and the averages of the corresponding nodes in the null models. The dispersion of the Gini values in the RN and SIR ensembles are very low, and cannot be appreciated at the scale of the picture. Horizontal lines show the means across nodes for each case, illustrating that the mean leading (following) Gini coefficient is 3.4 (5) times larger in the empirical graph than in the random graph ensemble for TT-2014 and 3.4 (5.5) for TT-2013. This shows an important level of centralization in both the attention attracted by countries from others and the attention countries dedicate to others. Moreover, the attention dedicated to others (following Gini) is more concentrated in fewer countries than the attention received (leading Gini). These results reveal that the network of leader-follower relationships does not only display heterogeneity in overall TTs production and source levels, but also is highly centralized at the local level in both attention and leadership aspects.

5.4.3 Systemic and subjective biases of leader-follower relations

We empirically test the coexistence of subjective and systemic biases in the international coverage of TTs at the link level, using our TTs datasets. We follow a regression model similar to the ones previously applied to the analysis of mass media [58, 59], including geographic, demographic, economic, and cultural aspects of the involved countries.

Subjective biases would manifest as a stronger leader-follower relationship measured as the amount of TT_{xy} (i.e., TTs in which country y follows country x), when countries x and y have similar cultures. In this analysis, clearly the dependent variable is TT_{xy}, which is hypothesized to depend on other factors of the countries involved. We measure the euclidean distance in cultural space as $C_{xy} = \sqrt{\sum_{v \in D} (v_x - v_y)^2}$, where D is the set of cultural dimensions in Hofstede's model and v_x is the value of dimension v for country x as explained in the Data and Methods Section. If the propaganda model holds, systemic biases will increase the leadership of countries with high GDP towards countries with lower GDP, and thus the strength of leader-follower relationships will depend on the GDP difference $G_{xy} = GDP_x - GDP_y$. Furthermore, systemic biases are hypothesized to reach beyond linguistic barriers, in comparison to subjective biases that should depend on shared languages. We quantify linguistic barriers through a binary variable L_{xy} that takes the value 1 if two countries share an official or national language, and 0 otherwise.

We test the existence of subjective and systemic biases controlling for the effect of timezones in our analysis. The fact that some countries start the day before others is subject to have an effect in leader-follower relationships, creating an additional bias in which leadership relationships tend to go from east to west. We include the time difference between the largest cities of each country T_{xy}, subtracting the timezone of y from the timezone of x, and converting it to a 12 hour basis. In addition, demographic factors also have the potential to influence the international coverage of TTs, where migration patterns might change the relevance of TTs depending on the migration from leader to follower m_{xy} and from follower to leader m_{yx}.

We fit a linear regression model with an interaction effect on language as:

$$\begin{aligned} TT_{xy} &= I + a_e m_{xy} + b_e m_{yx} + c_e T_{xy} + d_e G_{xy} + e_e C_{xy} \\ &+ L_{xy} \left(f + a_i m_{xy} + b_i m_{yx} + c_i T_{xy} + d_i G_{xy} + e_i C_{xy} \right) \end{aligned} \tag{5.1}$$

The above model is a combination of two linear models, one that holds for all pairs of countries regardless of their common languages, in which biases are *external* to language barriers and are quantified by a_e, b_e, c_e, d_e, e_e. The second part of the model holds only

TABLE 5.2: **Regression results of TT model for TT-2013 and TT-2014**

	2013	2014
I	$\mathbf{-0.17}$ (0.04)$^{\cdots}$	$\mathbf{-0.13}$ (0.02)$^{\cdots}$
a_e	0.21 (0.11)	0.16 (0.06)$^{\cdot}$
b_e	0.00 (0.10)	**0.17** (0.06)$^{\cdot\cdot}$
c_e	−0.04 (0.03)	−0.04 (0.02)
d_e	**0.39** (0.05)$^{\cdots}$	**0.24** (0.03)$^{\cdots}$
e_e	−0.03 (0.04)	−0.06 (0.02)$^{\cdot}$
f	**0.53** (0.07)$^{\cdots}$	**0.45** (0.05)$^{\cdots}$
a_i	−0.14 (0.12)	−0.04 (0.07)
b_i	0.15 (0.10)	0.03 (0.06)
c_i	**0.20** (0.07)$^{\cdot\cdot}$	0.05 (0.05)
d_i	−0.15 (0.07)$^{\cdot}$	−0.06 (0.05)
e_i	$\mathbf{-0.40}$ (0.07)$^{\cdots}$	$\mathbf{-0.38}$ (0.05)$^{\cdots}$
R^2	0.26	0.22
Adj. R^2	0.25	0.22
Num. obs.	870	1722

$^{\cdots}p < 0.001$, $^{\cdot\cdot}p < 0.01$, $^{\cdot}p < 0.05$

when countries have some language in common ($L_{xy} = 1$), and represents the additional biases *internal* to language barriers a_i, b_i, c_i, d_i, e_i. The intercept of the second part of the model (f) is a constant factor that is only present when $L_{xy} = 1$, and quantifies the *height* of the language barrier in comparison to the constant intercept I. Thus, f measures the bias in TTs that can be attributed only to the fact that two countries share a language. To be able to compare datasets, we renormalized all variables and filtered out countries without values in Hofstede's dataset, analyzing a total of 870 pairs from TT-2013 and 1722 pairs from TT-2014.

Table 5.2 shows the regression results for the above model on the TT-2013 and TT-2014 datasets. The estimate of d_e is positive and significant for both datasets, showing the existence of a systemic bias that depends on the economy regardless of language barriers. In line with our theoretical argumentation, this suggests that TTs follow the gradient of wealth: TTs created in rich and big countries are more likely to appear in follower countries with less wealth and power. This bias has been also reported for the coverage of traditional news over mass media [130, 131]. We do not find evidence of an additional effect of this GDP bias internally to language communities (the estimates of d_i are much smaller than d_e or not significant), indicating that the systemic bias is indeed global and not influenced by language. The size effect of the GDP difference is relatively large, being the strongest effect present in the case when countries do not share a language. Migration effects do not provide consistent results, only showing some positive effect for the migration from follower to leader in TT-2014. We attribute this result to the larger

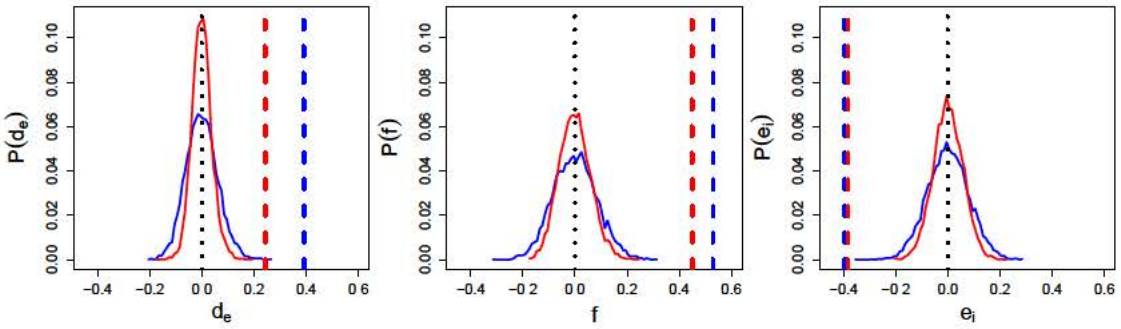


FIGURE 5.5: **Permutation tests of regression results.** Distributions of estimates of d_e, f, and e_i for 10.000 permutations of TT_{xy} for TT-2013 (blue) and TT-2014 (red). Vertical lines show the point estimates of OLS regression.

coverage of TT-2014, which includes a broader range of countries and reveals the effect that immigrants in a host country make newsworthy the TTs from their home country.

The estimates of e_e and e_i reported in Table 5.2 show that cultural distance has a negative weight on the strength of leader-follower relationships, in particular within language barriers. The estimate of the external cultural bias is close to 0 for both datasets only significant in TT-2014, revealing an almost negligible influence across cultures that do not speak any common language. The estimate of the influence of language (f) is highly positive and significant, indicating that TTs are much more likely to follow leader-follower relationships of countries that share at least a language.

These results portray the international structure of TTs within and across language barriers. Across languages, money speaks and TTs follow GDP, also influenced to some extent by migration patterns. Within language communities, the bias of GDP is also present, but it is attenuated by stronger relationships between countries closer in cultural space. It is worth noticing that the role of time zones is not significant across languages and only significant within languages in the smaller TT-2013 dataset. The inclusion of T_{xy} in the model makes our estimates of the rest of biases robust to the effect of timezones. The positive estimate of c_i in TT-2013 portrays the phenomenon that TTs go from east to west when countries share a language, but this is not observable in the larger dataset.

We verify the significant and sizable values of d_e, f, and e_i through permutation tests (see Figure 5.5), concluding that our results are robust to empirical correlations in the link structure of the network of TTs. Furthermore, the model has certain prediction power as measured through the adjusted R^2 values of 0.26 and 0.23, i.e., around 25% of the variance of the weight of leader-follower relationships can be explained by our model of systemic and subjective biases.

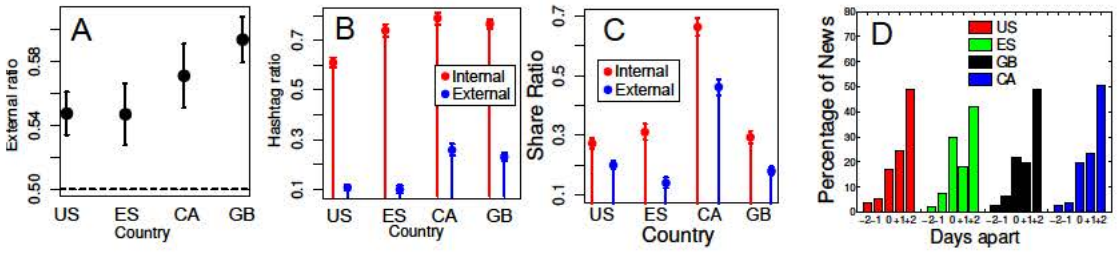


FIGURE 5.6: **Properties of internal and external TTs.** A) Ratio of External TTs; B) Ratio of External and Internal TTs that are hashtags; C) Conditional Probability of Internal or External TTs being shared; D) Percentage of news appearing in principal newspapers with some positive or negative delay with respect to Twitter TTs. All figures show the results of the analysis for US, ES, CA and GB. Error bars indicate the 95% confidence intervals from χ^2 tests.

5.4.4 The role of mass media in Twitter

Social media are not isolated from the content generated by mass media, which motivates our analysis of the interaction between both types of media. We apply our method to identify which TTs are associated to mass media content in 4 countries (US, GB, CA and ES), detecting which TTs are Internal (not reported in mass media) or External (also reported in mass media). In the following, we present an explorative study of the properties of TTs in relation to mass media, focusing on three aspects: i) the volume and properties of TTs that are reported in mass media, ii) the extend to which TTs detected in mass media are only local or travel along leader-follower relationships, and iii) the patterns of delay between the appearance of TTs and news reports.

Figure 5.6 A depicts the percentage of External TTs (or External Ratio) for each one of the analyzed countries. This shows that Twitter is not isolated from mass media, since slightly more than half of the TTs overlap with news found through Google News. Despite this important interaction between Twitter and mass media, still roughly between 40-46% TTs are associated to content not reported in mass media. Our data shows that hashtaging is a clear differentiating factor between Internal and External TTs. Figure 5.6 B presents the percentage of Internal and External TTs that are hashtags across the four studied countries. We observe that roughly 60-80% Internal TTs are hashtags whereas these values shrink to 10-25% for External TTs. This suggests that in order to become TT, without the back up of mass media channels, a centralized communication mechanism is required. The hashtaging functionality of Twitter provides such a mechanism, allowing both the collective discussion about the topic and the semiotic creation of a symbol to refer to it [132].

We have analyzed the TTs shared by each pair of the four considered countries. Note that if a TT is Internal in both countries, then it was shared internally within Twitter and it was not reported by traditional media in either country. Moreover, if a shared TT

is External in any of the two countries, the sharing process may have occurred through either mass media or Twitter. Then, to be conservative, we consider that this TT was shared through external media in this case. Based on these considerations, Figure 5.6 C shows the conditional probability that a TT classified as Internal (External) is shared between countries within Twitter (through mass media). We confirm that, for the four analyzed countries, TTs are more likely to be shared internally. Indeed, the probability of being shared internally is 50-100% proportionally higher than externally. In conclusion, the flow of TTs between countries is mainly formed by internal content generated in Twitter rather than news of interest to traditional mass media.

To conclude our analysis, we compare the surging dates of External TTs and their associated news in mass media to check in which communication venue overlapping news are (typically) reported earlier[5]. Specifically, we have computed the difference in number of days between the appearance date of every external TT and its associated piece of news in online newspapers in the four considered countries. Note that the difference ranges between $\pm$ 2 days from the appearance date of the TT since this is the window time that we have configured to query the Google News service. Figure 5.6 D shows the percentage of news that surged two days before, one day before, the same day, one day after and two days after in traditional media than in Twitter. More than 50% of the external TTs are reported earlier in Twitter than in the online edition of main newspapers whereas less than 10% of the news appeared earlier in those newspapers, and the rest appeared the same day in both channels.

5.5 Discussion

We showed how news in social media manifest through Local TTs in Twitter, analyzing two alternative large-scale datasets of TTs in different countries. We validated our analysis of the leader-follower relationships between countries testing the hypothesis of priority processes in queue systems [120, 121], finding a power-law distribution of delay times between the appearance of TTs. This finding conveys knowledge about the dynamics of how TTs travel across countries, in an analogous manner as how power-law degree distributions reveal dynamics of preferential attachment or edge copying mechanisms [133]. Applying the statistical physics of priority processes has potential applications to the analysis of communication dynamics in other online communities, from dialogues to collective reactions.

[5]Note that we are not considering individual tweets but TTs. It seems likely that an individual tweet may capture an event before it is reported in online traditional media. However, when that event becomes a TT it has the entity of a piece of news that has attracted the attention of a large number of people. We are interested in understanding if this happens before or after the event has been reported in mass media.

Leader-follower relationships among countries reveal patterns of heterogeneity in sharing activity as well as concentration of attention, allowing us to test hypotheses inspired in mass media about the role of economic and social factors [29, 56, 57]. We found that content in social media, similarly to mass media, follows the gradient of wealth from rich to poor countries. Our combination of data about TTs with Hofstede's quantification of culture [52] contributes to the wider scientific field of online ethnography [134]. Our results resonate with works about the online manifestation of cultural traits [54, 135, 136], unemployment [137], and economic inequality [138]. Similarly as how our analysis of TTs showed how information in social media can cross international borders, the analysis of digital traces has the potential to explain many other individual and collective aspects of human behaviour.

To analyze the role of mass media in reported TTs, we designed a method to match TTs to news in mass media close to the TT appearance. Using this method, we found that internal TTs are much more likely to manifest around a hashtag, which serves as a symbol to centralize communication in the absence of important mass media channels. This method also allowed us to statistically control for mass media in how TTs are shared across countries, revealing that external TTs are less likely to cross country borders in Twitter than those TTs that were not considered newsworthy by mass media. Further applications of this tool have the potential to enhance the analysis of dynamic collective response patterns in Twitter [132], allowing the measurement of reach and social interaction around news channels.

In this context, Kwak et al. [44] analyzed the overlapping between TTs and mass media news in 2009. First, the authors confirm the condition of breaking news of TTs. Furthermore, they compare news coverage by World Wide Twitter TTs with Hot Topics from Google Trends and headlines of CNN news concluding that CNN was ahead in reporting. Our explorative analysis suggests that this situation has been reversed in the last years, and now Twitter seems to be ahead mass media, at least for the part in which TTs overlap with traditional news channels. Roughly half of TTs overlap with news reported by mass media indicating an important interaction between both venues. Despite this high interaction and the common systemic biases in the international coverage of mass media news and TTs, the major fraction of TTs shared across countries in Twitter correspond to events not reported by major mass media.

We showed how news in social media manifest through Local TTs in Twitter, analyzing two alternative large-scale datasets of TTs in different countries. We validated our analysis of the leader-follower relationships between countries testing the hypothesis of priority processes in queue systems [120, 121], finding a power-law distribution of delay times between the appearance of TTs. This finding conveys knowledge about the

dynamics of how TTs travel across countries, in an analogous manner as how power-law degree distributions reveal dynamics of preferential attachment or edge copying mechanisms [133].

Chapter 6

Conclusions and Future Work

Defining new methodologies to understand fundamental aspects of the Online Advertising has a key importance due to the relevance to the economy and the society. In this thesis, we provide novel methodologies to improve our knowledge on the advertising ecosystem providing transparency and a better understanding of the use of personal information in individual advertising and also, the impact of new collective advertising services (e.g. Trending Topics).

First, we have designed a methodology to understand the use of personal information in individual target advertising by analyzing Online Behavioural Advertising (OBA). Second, we have defined a methodology and representative metrics to measure and analyze the impact and the visibility that new services, such as Trending Topics, offer to the collective advertising. In addition, using the aforomentioned metrics we compare this new online collective services with the traditional mass media. Finally, to complement our analysis of collective advertising we have define a methodology to study the information propagation of marketing pieces, Trending Topics, in Online Social Networks.

More in detail, our analysis of the individual and collective adversing reveal the following contributions and insights:

(i) We present a methodology to identify and quantify the presence of OBA in online advertising. We have implemented the methodology into a scalable system and run experiments covering a large part of the entire spectrum of definitions, metrics, sources, filters, etc that allows us to derive conclusions whose generality is guaranteed. In particular, our results reveal that OBA is a technique commonly used in online advertising. Moreover, our analysis using more than 50 trained personas suggests that the volume of OBA ads received by a user varies depending on the economical value associated to the behaviour/interests of the user. More importantly, our experiments reveal that

the online advertising market targets behavioural traits associated to sensitive topics (health, politics or sexuality) despite the existing legislation against it, for instance, in Europe. Finally, our analysis indicates that there is no significant geographical bias in the application of OBA and that do-not-track seems to not be enforced by publishers and aggregators and thus it does not affect OBA. These essential findings pave a solid ground to continue the research in this area and improve our still vague knowledge on the intrinsic aspects of the online advertising ecosystem.

(ii) Despite Trending Topics are a well-know feature regularly exploited in the context of marketing and advertisement, we still stand on preliminary ground in terms of understanding this tool. In this chapter we characterize the visibility of Trending Topics across 35 countries. In particular, we present a measurement methodology along with a methodology to thoroughly analyze the visibility of Trending Topics that we believe can be of high value for experts of different disciplines in marketing and advertisement contexts. The results obtained applying these methodologies indicate that, in general, Trending Topics present a comparable visibility to other traditional advertisement channels and thus they can be considered a useful tool in marketing and advertisement contexts. However, the high variability on the visibility offered by Trending Topics across (and within) countries suggests that we should apply the described methodology to obtain accurate results for each specific case.

In the case of the information propagation in social networks the most important results obtained are:

(iii) We show how content in social media can break international borders through Twitter TTs, revealing that Twitter is used as an alternative communication channel with respect to mass media. On the other hand, we found significant biases with respect to economic, demographic, and cultural factors. This portrays Twitter as a mixed and multipurpose community, in which content can flow without constraints, but also in which mass media have a strong influence and in which economic and cultural factors bias the flow of content.

Additionally, to analyze the role of mass media in reported TTs, we designed a method to match TTs to news in mass media close to the TT appearance. Using this method, we found that internal TTs are much more likely to manifest around a hashtag, which serves as a symbol to centralize communication in the absence of important mass media channels. This method also allowed us to statistically control for mass media in how TTs are shared across countries, revealing that external TTs are less likely to cross country borders in Twitter than those TTs that were not considered newsworthy by mass media.

As future work we plan to extend the methodology for the analysis of individual advertising allowing more configurations and including new metrics. Also, running experiments with data from real users is also our goal. In addition, we plan to analyze the impact of others collective Online Social Networks (e.g. Facebook) to provide a better understanding of the visibility it offers in terms of advertising and compare it with Twitter and traditional mass media.

References

[1] Interactive Advertising Bureau US. http://www.iab.com/news/u-s-internet-ad-revenues-reach-record-breaking-49-5-billion-in-2014-a-16-increase-over-landmark-2013-numbers-marking-fifth-year-in-a-row-of-double-digit-growth-for-the-industry/, .

[2] Interactive Advertising Bureau Europe. http://www.iabeurope.eu/files/9614/4844/3542/IAB_IHS_Euro_Ad_Macro_FINALpdf.pdf, .

[3] List of largest Internet companies. https://en.wikipedia.org/wiki/List_of_largest_Internet_companies.

[4] Google Annual Report. https://abc.xyz/investor/pdf/20151231_alphabet_10K.pdf.

[5] Facebook Annual Report. https://www.bamsec.com/filing/132680116000043?cik=1326801.

[6] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the Internet. In *Proc. ACM HotNets*, 2012.

[7] Adblock Plus. https://adblockplus.org/.

[8] Disconnect — Online Privacy & Security. https://disconnect.me/.

[9] Ghostery. https://www.ghostery.com/.

[10] Privacy Choice. http://www.privacychoice.org/.

[11] G. Hardin. The tragedy of the commons. *Science*, 1968.

[12] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proc. ACM CoNEXT*, 2013.

[13] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Proc. ACM/USENIX IMC*, 2014.

[14] A Odlyzko. The end of privacy and the seeds of capitalism's destruction. In *Privacy Law Scholars' Conference*.

[15] Latanya Sweeney. Discrimination in online ad delivery. *Queue*, 2013.

[16] Nora G Barnes and Stepahnie Wright. Fortune 500 are bullish on social media: Big companies get excited about google+, instagram, foursquare and pinterest, 2013.

[17] TwitterMarketingAgency.com. Twitter Marketing Agency. http://twittermarketingagency.com/, 2012.

[18] OutSourceGeek.com. Twitter Marketing Company. http://www.outsourcegeek.com/marketing-services/twitter-marketing-service/twitter-marketing-company, 2009.

[19] Twitter. Twitter blog: To Trend or Not to Trend... http://blog.twitter.com/2010/12/to-trend-or-not-to-trend.html, 2010.

[20] Scott Redick. HBR Blog Network. http://blogs.hbr.org/cs/2013/05/surprise_is_still_the_most_powerful.html, 2013.

[21] Alba Garmendia. Twitterholic Politicians. http://litteramedia.wordpress.com/2011/12/13/twitterholic-politicians/, 2011.

[22] Alexandra Hache. Spanish revolution. https://wiki.digitalmethods.net/Dmi/DmiSummer2011SpanishRevolution, 2011.

[23] Peter Kafka. Twitter promoted trends tracked for one month. http://allthingsd.com/20130409/big-media-loves-promoted-trends-twitters-big-dollar-digital-billboards/, 2013.

[24] Cadie Thompson. Facebook friends Twitter-like Trending Topics. https://www.cnbc.com/id/100942887, 2013.

[25] Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *ICWSM*, pages 50–57, 2011.

[26] Javier Borge-Holthoefer, Alejandro Rivero, Iñigo García, Elisa Cauhé, Alfredo Ferrer, Darío Ferrer, David Francos, David Iñiguez, María Pilar Pérez, Gonzalo Ruiz, et al. Structural and Dynamical Patterns on Online Social Networks: the Spanish May 15th Movement as a Case Study. *PloS one*, 6(8):e23883, 2011.

[27] Zeynep Tufekci. The Medium and the Movement: Digital Tools, Social Movement Politics, and the End of the Free Rider Problem. *Policy & Internet*, 6(2):202–208, 2014.

[28] Ethan Zuckerman. New Media, New Civics? *Policy & Internet*, 6(2):151–168, 2014.

[29] Edward S Herman and Noam Chomsky. *Manufacturing Consent: The Political Economy of the Mass Media*. Random House, 2008.

[30] Bernardo A Huberman. Social Computing and the Attention Economy. *Journal of Statistical Physics*, 151(1-2):329–339, 2013.

[31] Noam Chomsky. Paris attacks show hypocrisy of West's outrage, 2015. http://edition.cnn.com/2015/01/19/opinion/charlie-hebdo-noam-chomsky/.

[32] Aniko Hannak, Piotr Sapieżyński, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. Measuring Personalization of Web Search. In *Proc. WWW*, 2013.

[33] Anirban Majumder and Nisheeth Shrivastava. Know Your Personalization: Learning Topic Level Personalization in Online Services. In *Proc. WWW*, 2013.

[34] Saikat Guha, Bin Cheng, and Paul Francis. Challenges in measuring online advertising systems. In *Proc. ACM IMC*, 2010.

[35] Aleksandra Korolova. Privacy violations using microtargeted ads: A case study. In *Proc. IEEE ICDMW*, 2010.

[36] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. Follow the money: understanding economics of online aggregation and advertising. In *Proc. ACM IMC*, 2013.

[37] Narseo Vallina-Rodriguez, Jay Shah, Alessandro Finamore, Yan Grunenberger, Konstantina Papagiannaki, Hamed Haddadi, and Jon Crowcroft. Breaking for commercials: characterizing mobile advertising. In *Proc. ACM IMC*, 2012.

[38] Ashwin Rao, Justine Sherry, Arnaud Legout, Arvind Krishnamurthy, Walid Dabbous, and David Choffnes. Meddle: middleboxes for increased transparency and control of mobile traffic. In *Proc. ACM CoNEXT student workshop*, 2012.

[39] Ye Chen, Dmitry Pavlov, and John F Canny. Large-scale behavioral targeting. In *Proc. ACM SIGKDD*, 2009.

[40] Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. How much can behavioral targeting help online advertising? In *Proc. WWW*, 2009.

[41] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. XRay: Enhancing the Web's Transparency with Differential Correlation. In *USENIX Security*, 2014.

[42] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. Adreveal: improving transparency into online targeted advertising. In *Proc. ACM HotNets*, 2013.

[43] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S Muthukrishnan. Adscape: harvesting and analyzing online display ads. In *Proc. WWW*, 2014.

[44] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media? In *WWW*, 2010.

[45] Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: Persistence and decay. In *ICWSM*, 2011.

[46] Fred Morstatter, Jurgen Pfeffer, Huan Liu, and Kathleen M Carley. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. 2013.

[47] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. Traveling Trends: Social Butterflies or Frequent Fliers. In *COSN*, 2013.

[48] Jeff Huang, Katherine M Thornton, and Efthimis N Efthimiadis. Conversational tagging in twitter. In *21st ACM conference on Hypertext and hypermedia*, 2010.

[49] Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary. Twitter trending topic classification. In *ICDMW*, 2011.

[50] Arkaitz Zubiaga, Damiano Spina, Víctor Fresno, and Raquel Martínez. Classifying trending topics: a typology of conversation triggers on twitter. In *CIKM*, 2011.

[51] Johan Galtung and Mari Holmboe Ruge. The Structure of Foreign News the Presentation of the Congo, Cuba and Cyprus Crises in four Norwegian Newspapers. *Journal of peace research*, 2(1):64–90, 1965.

[52] Geert Hofstede. *Culture's consequences: International differences in work-related values*, volume 5. sage, 1984.

[53] Manuel Gomez-Rodriguez, Krishna P Gummadi, and Bernhard Schölkopf. Quantifying Information Overload in Social Media and its Impact on Social Contagions. In *ICWSM*, pages 170–179, 2014.

[54] David García and Dorian Tanase. Measuring Cultural Dynamics through the Eurovision Song Contest. *Advances in Complex Systems*, 16(08), 2013.

[55] Haoming Denis Wu. Investigating the Determinants of International News Flow A Meta-Analysis. *International Communication Gazette*, 60(6):493–512, 1998.

[56] Einar Östgaard. Factors Influencing the Flow of News. *Journal of Peace Research*, 2(1):39–63, 1965.

[57] Kenichi Ish. Is the US Over-reported in the Japanese Press? Factors Accounting for International News in the Asahi. *International Communication Gazette*, 57(2): 135–144, 1996.

[58] Tsan-Kuo Chang, Pamela J Shoemaker, and Nancy Brendlinger. Determinants of International News Coverage in the US Media. *Communication Research*, 14(4): 396–414, 1987.

[59] Haewoon Kwak and Jisun An. A First Look at Global News Coverage of Disasters by Using the GDELT Dataset. In *Social Informatics*, pages 300–308. Springer, 2014.

[60] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. Social Media News Communities: Gatekeeping, Coverage, and Statement Bias. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1679–1684. ACM, 2013.

[61] Jisun An, Daniele Quercia, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. Sharing Political News: the Balancing Act of Intimacy and Socialization in Selective Exposure. *EPJ Data Science*, 3(1):1–21, 2014.

[62] Jisun An, Daniele Quercia, and Jon Crowcroft. Partisan Sharing: Facebook Evidence and Societal Consequences. In *Proceedings of the second edition of the ACM conference on Online social networks*, pages 13–24. ACM, 2014.

[63] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. Traveling trends: social butterflies or frequent fliers? In *Proceedings of the first ACM conference on Online social networks*, pages 213–222. ACM, 2013.

[64] Krishna Y Kamath, James Caverlee, Kyumin Lee, and Zhiyuan Cheng. Spatio-Temporal Dynamics of Online Memes: a Study of Geo-tagged Tweets. In *Proceedings of the 22nd international conference on World Wide Web*, pages 667–678. International World Wide Web Conferences Steering Committee, 2013.

[65] David Wilkinson and Mike Thelwall. Trending Twitter topics in English: An international comparison. *Journal of the American Society for Information Science and Technology*, 63(8):1631–1646, 2012.

[66] Aniko Hannak, Piotr Sapieżyński, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. Measuring Personalization of Web Search. In *Proceedings of the Twenty-Second International World Wide Web Conference (WWW'13)*, Rio de Janeiro, Brazil, May 2013.

[67] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[68] R Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users1. *Journal of Computer-Mediated Communication*, 14 (2):265–285, 2009.

[69] Eytan Bakshy, Solomon Messing, and Lada Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, page aaa1160, 2015.

[70] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: collective narratives in the age of misinformation. *PloS one*, 10(2):02, 2015.

[71] Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, and Jun Wang. Internet Advertising: An Interplay among Advertisers, Online Publishers, Ad Exchanges and Web Users. *CoRR*, 2012.

[72] Ayman Farahat and Tarun Bhatia. Retargeting related techniques and offerings, April 25 2011. US Patent App. 13/093,498.

[73] Miguel Helft and Tanzina Vega. Retargeting ads follow surfers to other sites. *The New York Times*, 2010.

[74] Anja Lambrecht and Catherine Tucker. When does retargeting work? information specificity in online advertising. *Journal of Marketing Research*, 2013.

[75] European Union Directive 95/46/EC. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

[76] Balachander Krishnamurthy and Craig E. Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proc. WWW*, 2009.

[77] Digital Advertising Alliance Consumer Choice Page. http://www.aboutads.info/choices/.

[78] Ravi Kumar and Andrew Tomkins. A Characterization of Online Browsing Behavior. In *Proc. WWW*, 2010.

[79] Cyren URL Category Check. http://www.cyren.com/url-category-check.html.

[80] Display Planner - Google AdWords. https://adwords.google.com/da/DisplayPlanner/Home.

[81] McAfee. https://www.trustedsource.org/?p=mcafee.

[82] Claudia Leacock, George A Miller, and Martin Chodorow. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, 1998.

[83] Do not track. http://donottrack.us/.

[84] J. Carrascosa, R. Gonzalez, R. Cuevas, and A. Azcorra. Are Trending Topics Useful for Marketing? Visibility of Trending Topics vs Traditional Advertisement. In *COSN*, 2013.

[85] Fotios Papaodyssefs, Costas Iordanou, Jeremy Blackburn, Nikolaos Laoutaris, and Konstantina Papagiannaki. Web Identity Translator: Behavioral Advertising and Identity Privacy with WIT. In *Proc. of the 14th ACM Workshop on Hot Topics in Networks*, 2015.

[86] Kevin Thompson, Gregory J Miller, and Rick Wilder. Wide-area Internet traffic patterns and characteristics. *Network, IEEE*, 1997.

[87] dev.twitter.com. Twitter Rate Limiting. https://dev.twitter.com/docs/rate-limiting, 2013.

[88] Twitter. Twitter Blog: Let's Fly. http://blog.twitter.com/2011/12/lets-fly.html, 2011.

[89] Hubert W Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 1967.

[90] Weka. EM - Weka 3. http://weka.sourceforge.net/doc.dev/weka/clusterers/EM.html, 2013.

[91] Jan Panero Benway. Banner blindness: The irony of attention grabbing on the World Wide Web. In *HFES Annual Meeting*. SAGE Publications, 1998.

[92] Guillaume Hervet, Katherine Guérard, Sébastien Tremblay, and Mohamed Saber Chtourou. Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology*, 2011.

[93] Silvia Heinz, Markus Hug, Carina Nugaeva, and Klaus Opwis. Online ad banners: the effects of goal orientation and content congruence on memory. In *CHI*, 2013.

[94] Roberto Gonzalez, Ruben Cuevas, Angel Cuevas, and Carmen Guerrero. Understanding the locality effect in Twitter: measurement and analysis. *Personal and Ubiquitous Computing*, 2013.

[95] worldbank.org. The World Bank. http://www.worldbank.org/, 2014.

[96] John A Davis. *Measuring Marketing: 110+ Key Metrics Every Marketer Needs*. Wiley. com, 2012.

[97] Paul W Farris, Neil T Bendle, Phillip E Pfeifer, and David J Reibstein. *Marketing metrics: The definitive guide to measuring marketing performance*. Pearson Education, 2010.

[98] Ted Rooke. Are gross rating points really the answer for digital? http://www.imediaconnection.com/content/32278.asp, 2012.

[99] Dave Morgan. Gross Rating Point Metrics Will Be Good for Online Advertising. http://www.mediapost.com/publications/article/155743/, 2011.

[100] Rik Pieters and Michel Wedel. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, pages 36–50, 2004.

[101] Dan O'Day. The perfect length for a radio commercial. http://danoday.com/blog/2010/11/radio-commercial-length/, 2010.

[102] Dan Mccurdy. How Long is a Radio Ad. http://suite101.com/article/how-long-is-a-radio-commercial-a152539, 2009.

[103] AvenueRight.com. Optimum Scheduling for Radio Advertising. http://avenueright.com/entries/71/optimum-scheduling-for-radio-advertising-frequency-is-key, 2010.

[104] MarketingProfs.com. Radio Ad Frequency. http://www.marketingprofs.com/ea/qst_question.asp?qstid=6388, 2005.

[105] Biz Stone. Blog Twitter: Hello World. https://blog.twitter.com/2010/hello-world, 2010.

[106] Twitter. Twitter Blogs: Promoted Promotions. https://blog.twitter.com/2010/promoted-promotions, 2010.

[107] support.Twitter.com. What are Promoted Trends? https://support.twitter.com/groups/58-advertising/topics/244-about-twitter-ads/articles/282142-what-are-promoted-trends, 2013.

[108] Peter Kafka. Twitter Hikes Its Promoted Trend Prices Again, to $200,000 a Day. http://allthingsd.com/20130209/twitter-hikes-its-promoted-trend-prices-again-to-200000-a-day/, 2013.

[109] Cotton Delo. Twitter Is Already Selling World Cup Promoted Trends for 2014. http://adage.com/article/digital/twitter-selling-world-cup-promoted-trends/243018/, 2013.

[110] Amendment. Amendment No.1 to Form S-1. http://www.sec.gov/Archives/edgar/data/1418091/000119312513400028/d564001ds1a.htm, 2013.

[111] Rebecca Liebe. Online News Commands Highest CPM. http://econsultancy.com/blog/6096-online-news-commands-highest-cpm, 2010.

[112] wsjmediakit.com. The Wall Street Journal: General Advertising Rate Card 2014. http://www.wsjmediakit.com/downloads/2014_General_Rate_Card.pdf?140224124240, 2014.

[113] Mario F Triola, William Martin Goodman, Gerry LaBute, Richard Law, and Lisa MacKay. *Elementary statistics*. Pearson/Addison-Wesley, 2006.

[114] Census.gov. Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas. http://www.census.gov/popest/data/metro/totals/2011/, 2011.

[115] Sysomos. Exploring the Use of Twitter Around the World. http://www.sysomos.com/insidetwitter/geography/, 2010.

[116] Clay Shirky. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin, 2008.

[117] Philipp Singer, Fabian Flöck, Clemens Meinhart, Elias Zeitfogel, and Markus Strohmaier. Evolution of Reddit: From the Front Page of the Internet to a Self-referential Community? In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 517–522. International World Wide Web Conferences Steering Committee, 2014.

[118] Trending Topics API Request. https://dev.twitter.com/rest/reference/get/trends/place.

[119] Carrascosa, J. and Cuevas, R. and Gonzalez, R. and Azcorra, A. and Garcia, D. Dataset - Quantifying the Economic and Cultural Biases of Social Media through Trending Topics. http://dx.doi.org/10.6084/m9.figshare.1381869.

[120] Joao Gama Oliveira and Albert-László Barabási. Human Dynamics: Darwin and Einstein Correspondence Patterns. *Nature*, 437(7063):1251–1251, 2005.

[121] Alexei Vázquez, João Gama Oliveira, Zoltán Dezsö, Kwang-Il Goh, Imre Kondor, and Albert-László Barabási. Modeling Bursts and Heavy Tails in Human Dynamics. *Physical Review E*, 73(3):036127, 2006.

[122] Riley Crane, Frank Schweitzer, and Didier Sornette. Power Law Signature of Media Exposure in Human Response Waiting Time Distributions. *Physical Review E*, 81 (5):056101, 2010.

[123] Antonios Garas, David Garcia, Marcin Skowron, and Frank Schweitzer. Emotional Persistence in Online Chatting Communities. *Scientific Reports*, 2, 2012.

[124] David Garcia, Ingmar Weber, and Rama Venkata Kiran Garimella. Gender Asymmetries in Reality and Fiction : The Bechdel Test of Social Media. In *ICWSM*, pages 131–140, 2014.

[125] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-Law Distributions in Empirical Data. *SIAM review*, 51(4):661–703, 2009.

[126] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. Powerlaw: a Python Package for Analysis of Heavy-Tailed Distributions. *PLoS One*, 9(1):e85777, 2014.

[127] P Erdos and A Renyi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[128] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.

[129] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.

[130] Iain Wallace. *The global economic system*. Routledge, 2002.

[131] Immanuel Wallerstein. The World-System after the Cold War. *Journal of Peace Research*, pages 1–6, 1993.

[132] Janette Lehmann, Bruno Gonçalves, José J Ramasco, and Ciro Cattuto. Dynamical Classes of Collective Attention in Twitter. In *Proceedings of the 21st international conference on World Wide Web*, pages 251–260. ACM, 2012.

[133] Michael Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet mathematics*, 1(2):226–251, 2004.

[134] Robert V Kozinets. The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities. *Journal of marketing research*, 39(1):61–72, 2002.

[135] Ruth Garcia-Gavilanes, Daniele Quercia, and Alejandro Jaimes. Cultural Dimensions in Twitter: Time, Individualism and Power. *Proc. of ICWSM*, 13, 2013.

[136] Claudia Wagner, Philipp Singer, and Markus Strohmaier. The Nature and Evolution of Online Food Preferences. *EPJ Data Science*, 3(1):1–22, 2014.

[137] Alejandro Llorente, Manuel Cebrian, Esteban Moro, et al. Social media fingerprints of unemployment. *arXiv preprint arXiv:1411.3140*, 2014.

[138] Gabriel Magno and Ingmar Weber. International gender differences and gaps in online social networks. In *Social Informatics*, pages 121–138. Springer, 2014.