The color of corruption coverage in Spain
Agenda setting in a polarized media ecosystem

PhD candidate: Pablo Rey-Mazón

Director: Ismael Peña-López

Committee members: Sasha Costanza-Chock, Victor Sampedro Blanco

Information and Knowledge Society Doctoral Programme.

Universitat Oberta de Catalunya (UOC).

Image: 15 months of corruption scandals stories in Spanish newspapers front pages coded by institution (color) with PageOneX software. January 2013- March 2014.

1. Research subject

Mainstream media determine what is important and what is not by selecting and framing stories. Mass media play therefore a key role in shaping public opinion. Since the seminal article *The agenda-setting function of mass media* was published in 1972, and during the last decade, we have witnessed the emergence and the growth of the social media influence, first with the blogosphere and later with social networking sites.

We use the term social media as "web based services that allow individuals to (1) construct a public or semipublic profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd & Ellison, 2008:211). Social media users, part of the reader-writer culture, contribute to disseminate the news and are also able to participate directly in the creation, selection and framing of the stories to modify the agenda setting traditionally dominated by the mainstream media (Negroponte 1995; Shirky, 2008; Sunstein, 2001). News media remain as key players in the mass media ecosystem but they are no longer alone in the way political reality is shaped. The introduction of the term "media ecology" in the 1960s by theorists including McLuhan and Postman made communication scholars understand media as interconnected to each other and society (Strate, 2004).

The main objective of this study is to weight how social networking sites are changing the agenda-setting role of mainstream media: How is the dialogue between social media and news media? How is it driving attention to certain stories and hiding others? What are their similarities and differences? To address these questions we are going to use as our case study the way corruption scandals are covered and framed in the Spanish media ecosystem.

In the past years, corruption cases in Spain have involved almost every institution in the country. The major political parties, the royal family, the main unions, the supreme court or the national bank, just to mention a few, have been involved in corruption scandals usually brought to light by news media and amplified by social media. Every week, a new investigation is unveiled provoking an increasing the sense of indignation. This situation has pushed up the concern for corruption to the second position of the problems perceived by Spaniards[1] (CIS, 2015).

The wide range of actors involved in corruption scandals and the very polarized partisanship of the Spanish media outlets (Hallin☐ & Mancini, 2004) makes corruption an appropriate field of research to analyze news media bias in Spain (Castromil, 2012) and its influence among its readership. A topic that involves political

---

[1] In the "barometer", of the Centro de Investigaciones Sociológicas (CIS, 2015), that asks about the three more important problems for citizens, corruption and fraud is in the top 3 for the past three years.

parties from different ideologies (right and left, from both sides of the aisle) is particularly useful to analyze the role of social media in the perception and framing of such stories. For that purpose we are going to analyze news media stories and social media messages in Twitter about corruption scandals in Spain. We will obtain accurate metrics to understand why not all the cases have the same visibility or gain the same popularity.

2. Specific details of the proposal

2.1. Justification of its interest

In the field of media studies researchers are addressing the complexity of multiple information flows in order to understand how reality is framed by news and social media (Lotan et al., 2011; Bandari et al., 2012; Graeff et al., 2014). To draw a more accurate picture there is a need to take in account all the actors across the entire media ecosystem, and not only mainstream media. As it happened with the emergence of alternative media, like the surge blogs more than a decade ago (Kelly, 2008), social networking platforms like Facebook and Twitter are now becoming now relevant tools that drive an important amount of the Internet traffic. Social media are also a common way among scholars to measure how different actors —political parties, social movements, news media and society in general— frame and deliver their messages in the new media ecosystem, or *networked public sphere* to use Yochai Benkler's phrase. In this research we are going to analyze Twitter messages as a proxy for the social media ecosystem, front pages of paper newspaper and home pages of online newspapers as a proxy for news media selection of stories.

Communication scholars have long analyzed newspaper front pages, using column-inches, as an important indicator of mass media attention (Ed Herman and Noam Chomsky, 1988). Different studies have found that newspapers set the agenda for other mainstream media such as television networks and radio broadcasts (Boyle, 2001; McCombes, Lopez-Escobar, and Llamas, 2006). Whereas front page analysis is not new, as we shall explain, we propose the use of new methods (PageOneX.com and StoryTracker software) to have better and more accurate metrics that take in account size and visual aspect of the analyzed stories for online news sites and paper newspapers. Through the combined analysis of Twitter and the main printed and online newspapers we are going to be able to understand how they influence the public agenda and to understand the triangle media, politics and public.

Corruption cases in Spain have been "mainly brought to light by newspapers" (Costas-Pérez, Solé-Ollé & Sorribas-Navarro, 2012:1). The analysis of corruption coverage in the main newspapers allows us to measure how news media protect or attack certain parties and institutions by hiding or promoting certain stories, we will be able to obtain accurate measures of their political bias. A Twitter analysis allow us research how social media dialogues and modifies, if it does, the agenda setting role of news media. Through this combined analysis we will be able to update theories about the Spanish partisan media ecosystem (Hallin & Mancini,

2004; Castromil, 2012; Bouza et al., 2013). We are specially interested in studying how the situation evolves in a new political and media environment that is shifting from a bipartisan system to a four players game (Rodon & Hierro, 2016) and where new online news sites and social networking platforms are entering with strength the mass media ecosystem. In a system where people select the newspaper more aligned to their ideology and point their attention to alternative sources of information such as social media: what comes first, the agenda setting and framing by mass media or the general public predisposition?

2.2. State of the art - Literature review

This thesis proposal is grounded in the field of media studies. We are going to use a diverse bibliography to agenda-setting and social and news media. The research also has strong ties with corruption and corruption perception studies that have been conducted in different disciplines such as political science, sociology or economics. We will center this section about the scholarship on measuring news and social media and conclude with an examination of the dialogue between them.

Measuring news media

Newspaper front pages are still important in the broader media ecology[2]

Even as the media ecology has been rapidly transformed by the continued globalization of the cultural industries (Hesmondhalgh, 2007), the spread of broadband internet and mobile phones, the rise in read/write digital media literacies (Anderson, Glaisyer, Smith, and Rothfeld, 2011), and the steady growth of social media, daily print newspapers continue to be a key mechanism for organizing both public and elite attention.

In countries where print circulation is declining, newspapers still have an impact beyond direct readership, since they greatly influence attention across the media ecology. For example, Golan (2006) and Reese & Danielian (1989) examined what they call the intermedia agenda setting relationships between newspapers, magazines, and TV, and found that mass media outlets tend to provide similar amounts of coverage of stories. They also demonstrated the key agenda-setting role of the New York Times, specifically, for TV news. Boyle (2001) found that newspapers set the agenda for television networks, while other researchers found that newspaper coverage heavily influences both television and radio broadcast news agendas (McCombes, Lopez-Escobar, and Llamas, 2006). Breed (1955) demonstrated that large metropolitan newspapers set the agenda of smaller newspapers, while White (1949) and later Snider (1967) found that news story selection by the editor of a small-town Midwestern newspaper in the US was strongly shaped by story availability and categorization in a wire service, a finding later supported by McCombs and Shaw (1976). Newspaper coverage also plays an important agenda-setting role for blogs and social media (McCombs and Shaw, 2006).

---

[2] Part of this section was also written for Costanza-Chock and Rey-Mazón (2016).

Newspaper front pages, while declining in the total share of media attention they command, thus not only remain relevant but continue to function as critically important spaces in the broader media ecology. The front page of the print edition of newspapers has long been, and still remains, both a powerful indicator of the importance of a given story, and a key mechanism for driving attention across all media platforms. At least in Spain, front pages of daily newspapers are directly displayed and commented in TV and radio news programs as an indicator of what other media find relevant.
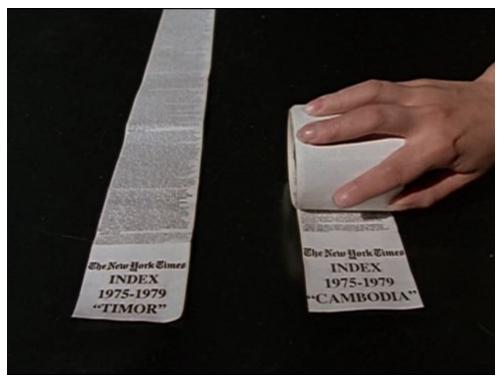
Accordingly, communication scholars have researched newspaper front pages for decades, and continue to do so. Many media scholars have studied the production of the newspaper front page. Newsroom anthropologists have closely explored the news production process, including the social, cultural, political, economic, and interpersonal forces that shape editorial decisionmaking and front-page prioritization (Forrest, 1934; Gans, 1979; Broder, 2000). Given the influence of print newspaper coverage on the overall news agenda, newspaper editors, unsurprisingly, spend a great deal of time and energy deciding what stories to place on the front pages of their print editions (Sumpter, 2000). The front page is also a highly contested political space. For example, Álvarez-Peralta (2010, 2011) argues that front page analysis provides a more direct lense on a newspaper's editorial line than the newspaper's editorials. He notes that editorials are carefully constructed, public-facing documents that frame the opinions of the newspaper editors for a critical reading public, whereas analysis of the content and framing of front page stories provides access to the paper's 'hidden' editorial stance (Álvarez Peralta, 2010).

From column inches to percentage

While column-inch measurement gained great visibility in the 1980s, largely due to the popularity of *Manufacturing Consent* (Herman and Chomsky, 1988), it was not then a new approach. In the past, researchers taking this approach needed to obtain physical copies of newspapers, measure column-inches by hand (usually with a ruler), and manually input measurements into a spreadsheet or database.

Herman and Chomsky employed quantitative measures including column-inch comparison, number of articles, number of front page articles, and number of editorials, as empirical support for their widely cited (if much attacked) 'propaganda model' of the news system. Measurements of column-inches were also used early on to provide an indicator of attention, and to quantify the news impact of events. The techniques developed by early mass communication scholars for newspaper column-inch analysis evolved over time, such as through the incorporation of systematized coding instructions, training, and measures of intercoder reliability, and these techniques eventually became part of standard curricula in communication studies. In one of the earliest studies to use this method, Yeuell (1928) performed a content analysis of the three most popular Black newspapers, coding articles by content types.' Yeuell then measured each type of material in column inches, and analyzed the percent of the total content by category. He did not provide the sum total of column-inches,

but rather, the percentage of total coverage occupied by each type of news. This approach made comparison possible across different newspaper sizes and formats.

Measuring genocide coverage of friend and enemy states in column-inches. Source: Manufacturing Consent: Noam Chomsky and the Media (1992).

More recently, a number of communication scholars have developed studies that employ metrics of the number of articles found on newspaper front pages. For example, Boydstun (2013) created a database that categorizes every front page article (1998-2006) from the New York Times, according to a codebook developed by the Policy Agendas Project (policyagendas.org). The dataset enables scholars to analyze front page topic attention over time, and to compare front page coverage to full-paper coverage. Wolfe, Boydstun, and Baumgartner (2009) analyzed Boydstun's data and found that front page coverage is more likely to repeat topics that were covered on the previous day, compared to full-paper coverage (a characteristic they describe as 'friction.') They also found some important differences between the topics that dominate the front page and those that occupy full paper coverage, although overall they note a strong positive correlation (r=0.78) between front page coverage and full paper coverage of any given topic (Wolf, Boydstun, & Baumgartner, 2009). Following the same methodology, Chaqués-Bonafont, Palau, and Baumgartner (2014) analyzed 13 years of front pages of the two main Spanish daily newspapers to create a database to study media bias and the impact of media coverage of corruption on Spanish public opinion.

Di Tella and Franceschelli (2009) measured the percentage of newspaper front pages dedicated to corruption scandals in the four main newspapers in Argentina during the period 1998–2007, using a methodology that they describe as follows: "For each day in our sample period, and for each newspaper, a research assistant measured the area covered by any front page article that dealt with any corruption scandal that involved

members of the current national administration (e.g., the president or the ministers) and then divided it by the total area of the front page. Our approach involves two steps. In the first step, we use content analysis to select reports involving corruption scandals of the government. In the second step, we simply measure the area occupied by this scandal on the front page (see Puglisi and Snyder 2008 for a discussion)" (Di Tella & Franceschelli, 2009: 8). In this way, they demonstrated a negative correlation between corruption coverage and government advertising (more government advertising correlated with less corruption coverage), a finding consistent with independent reports of provincial officials withholding advertising as a punishment for papers that provided critical coverage (Ibid.).

As we have seen, front page newspaper analysis has long been an important method for communication scholars. Although newspapers now compete with a broader array of platforms than ever before for audience attention, and the print editions of many papers are struggling to find their footing in a digitally converged media ecology, newspaper front pages remain key spaces of communicative power. They continue to shape attention and contribute to setting the news agenda across multiple media platforms. Quantitative analysis of the contents of newspaper front pages therefore retains analytical force. As we later explain, we are going to use PageOneX, a software for front page analysis that we have developed.

Research about online news sites, and particularly of their home pages, do not have, obviously, such a long history (Li, 2013; Salaverría & Sancho, 2007) and it has also to face the rapidly changing environment of digital technologies. In 2014, an internal report on digital innovation at the The New York Times stated that visits to the homepage had fallen 50 percent in the period 2011-2013. This information made some journalists certify the death of the homepage or at least a change in the way homepages were traditionally conceived[3] (Kirkland, 2014). In the same period, the page views remained fairly steady at The New York Times website, so the loss of visits to the homepage was related to the way visitors accessed their site, which might indicate a growth of visits via social media.

Therefore, it seems that homepages in online news sites are less influential than before the emergence of Facebook and Twitter[4], but they are still relevant in the way they drive attention. We expect to see in online newspapers a similar selection of content to the one noted by Álvarez-Peralta (2010, 2011) for printed newspapers and be able analyze the 'hidden' editorial stance of the medium. When studying homepages other variables are going to come into play, like the 'on the fly' decisions by web editors to change the news in the homepage according web statistics (low or high traffic) that news stories receive. In the methodology section

---

[3] The website still had 80 million monthly visits, nevertheless. Kirkland (2014) states "That's why so many news organizations are desperate to find ways to keep readers on site after they've visited an article page. The Los Angeles Times brings its homepage to you, transitioning to a homepage or section page at the bottom of many stories. Quartz doesn't have a homepage — just a scroll of stories. TIME.com has an infinite scroll of stories populated with top news. And Advance newspapers have experimented with retooling homepages as streams of stories".

[4] There ather other community based social news networking service such as Reddit, and meneame.net for Spain, that drive also social media traffic to online news media.

we explain that we are going to use astpages.org and Storytracker software to archive and analyze home pages stories.

Measuring social media: Twitter

Since its creation in 2006, the a microblogging platform Twitter has become one of the most popular social media websites where users exchange information. Its openness by default and the possibility to get the data[5] and directly measure the way information propagates has also attracted the attention of social and communication researchers (Ye & Wu, 2010; Xin et al., 2011). As several scholars have noted there are problems that are limiting the utility of this type of research, for example "the apparent ease with which tweets may be aggregated belies the difficulty of designing a reliable, reproducible data collection", the way "Twitter itself is a dynamic system subject to near constant change", or that "Twitter use is not evenly distributed among Internet users" (Driscoll & Walker, 2014:1746). We are addressing these problems in our research and are going to use a consistent terminology (Bruns & Burgess, 2011a, 2011 b, 2012; Howison, Wiggins, & Crowston, 2011, Congosto, 2016) to enable comparison across different studies.

Driscoll and Walker (2014) also studied and compared the three publicly accessible API: the Search API, REST API, and Streaming API with the real time stream of tweets, provided by the Gnip PowerTrack commercial service (Walker, Hemsley, Eckert, Mason, & Nahon, 2013). The results showed that the Streaming API is suitable for longitudinal data collection, and not so much for high traffic events, due to the 50 tweets per second limit of this method. For massive and short events González-Bailón et al. (2012) also found out that Search API results are not a random sample of overall Twitter activity. As we discuss in the methodology section, we are going to use the software t-hoarder for Twitter data acquisition.
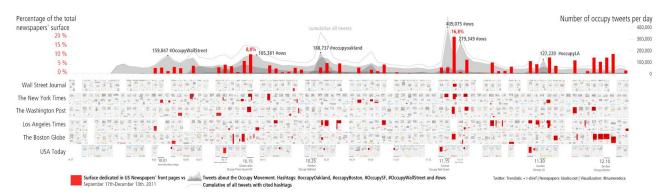
Social and news media dialogue

The interaction between social networking sites and digital news media together with the introduction of new digital technologies has interested communication scholars regarding how these new technologies have directly affected journalistic practices (Deuze, 2003, 2005; Lasorsa et al, 2012; Lewis, Kaufhold, & Lasorsa, 2010; Phillips et al., 2009; Phillips, 2010).

Before the hype of the networked social media, such as Facebook and Twitter, when blogs were the an important alternative way of social communication online, Kelly (2008) studied whether the blogosphere and other web-native media were to replace mainstream media as primary sources of information: "the blogosphere is becoming as important as the front page of the paper for landing eyeballs on a journalist's

---

[5] Driscoll & Walker (2014) p. 1746: "Among the many sites where big social data is collected, Twitter is particularly compelling because of its perceived accessibility. In comparison to Facebook, which is largely closed-off to the academic community, or a high-bandwidth site like YouTube, tweets are small in size, public by default, numerous, and topically diverse. With little more than a laptop, an Internet connection, and a few lines of scripting code, researchers can aggregate several million tweets in a short period of time using widely-available, low-cost tools".

article. There is a cycle of attention between blogs and the MSM [Mainstream Media], in which the MSM uses the blogosphere (...), and the blogosphere channels attention back to the MSM" (Kelly, 2008:2). In the same way a decade ago scholars studied the complementary function of mainstream media and the blogosphere media studies are focusing now in the rise of social and the way it impacts the mainstream media (Newman, 2009, 2011).

Surface coverage in newspapers front pages vs. Twitter about Occupy from September 17th through December 13th, 2011.The tweets per day come from the data released by r-shief.org. Pablo Rey Mazón. Link to image.

Twitter and Facebook were in origin networking communication tools where users were supposed to publish personal status updates. With the time users started sharing all types of information and rebroadcasting content from the web in the form of URL. In 2009, Twitter exemplified this evolution when they changed the prominent question in their website "what are you doing?" for "what's happening?". Later on these social networking platforms like Twitter enabled the embed and publication of multimedia content (images, videos) and became "a central role in the way stories are sourced, broken and distributed" (Newman, 2011:6).

Online news media, aware of the potential of these tools and the high Internet traffic that they command have also engaged in the dissemination of their news stories through these new channels of communication. Furthermore, tools like Facebook Instant Articles allows news media to publish their own articles natively on the Facebook platform. Twitter has enabled the possibility to publish video natively, like the Facebook Live application. Therefore, the current analysis of social media necessarily includes news media outlets as they are also users and actors of social networking sites. Both ecosystems can no longer be studied separately. The use of multiple online news sources with tools like Media cloud "an open news platform that enables public access to full-length historical news texts, which enables unprecedented access to the online news landscape" (Chuang et al., 2014) or the use of multiple sources of online and off–line information across complex media ecosystems (Graeff et al., 2014).

2.3. Questions and objectives

Questions

General question: How are social networking sites changing the agenda-setting role of traditional media?

Secondary questions:

How is the dialogue between social media and news media?

How are news and social media driving attention to certain stories and hiding others?

How are new media using social networking platforms?

Which is the proportion of the tweets that contain URL that point to online news media sites?

Objectives

The main objective of this study is to answer how are social networking sites changing the agenda-setting role of traditional media.

This proposal studies corruption coverage in news and social media in Spain because we want:

- to update theories about the Spanish partisan media ecosystem --now populated with new digital outlets and social media-- and understand how media outlets are reacting to the new wave of corruption scandals. We are specially interested to see how the situation evolves in a political situation that is changing and new online news sites are entering with strength the media ecosystem.

- to find out if (and how) corruption coverage influences/modifies public opinion in order to help readers understand how persuasive is the mass media ecosystem[6]. In a system where people select the newspaper more aligned to their ideology and point their attention to alternative sources of information setting such as social media: what comes first, the agenda setting and framing by mass media or the general public predisposition?

Secondary research questions, attending to specific context

Current events related to this research can provide interesting situations and context that can help to test hypothesis and develop theory based through their study:

- In recent years the directors of the main newspapers (*El País*, *El Mundo* and *La Vanguardia*) have changed: how was the coverage of political corruption evolved before and after? How does corruption coverage influences or is influenced by the change of directors?

- It is expected that during election periods the amount of coverage about political parties increases: which effect does it have on corruption coverage? Should we expect an increase or decrease of coverage? How does the partisanship of the Spanish media ecosystem relates to this? How is Twitter influenced by election periods? And for the case of catalan newspapers and corruption scandals?

---

[6] Castromil (2012) sets a similar question for the Spanish media ecosystem "Polarized pluralism" media system type as enunciated by Hallin and Mancini (2004).

2.4. Working hypothesis

At this stage this research considers three working hypothesis:

H1: Social media does not have a relevant/strong influence in the way agenda setting role is defined by the news media.

H2: News and social media dialogue and share influence in the way they bring attention to stories and they set the agenda.

H3: For certain stories and topics social media has bigger influence than news media on the definition of the public agenda.

Temptative conclusions

Whereas Twitter as a social media is able to command attention to certain stories, news media still have the gatekeeping role to define what is relevant. In the event that an organized or unorganized group of Twitter users is able to bring to light certain stories, it still needs the coverage of news media to amplify the story, reach more Twitter users (and a larger audience) and influence what is newsworthy in the broad media ecology. However, this gatekeeping role by the news media is not as immovable as before, as the barriers the borders between journalism, citizen journalism and social media are blurred and new independent media more closed to the social are entering into play.

2.5. Research methodology

The unit of analysis of this research are social media messages in Twitter and stories in front pages of daily newspaper and stories in home pages of online news sites. We want to measure the importance given by the news media to corruption scandals and compare it to the discussion in social media so that we can infer similarities and differences.

Period of analysis

Given the availability of the data, we are going to analyze daily front pages of newspapers from 2010[7] to 2017. These dataset will serve as a background indicator for the longer period.

Daily homepages of online news sites and tweets will be gathered in the period from 2016 (second half) to 2017.

The data collection methodology is structured as follows:
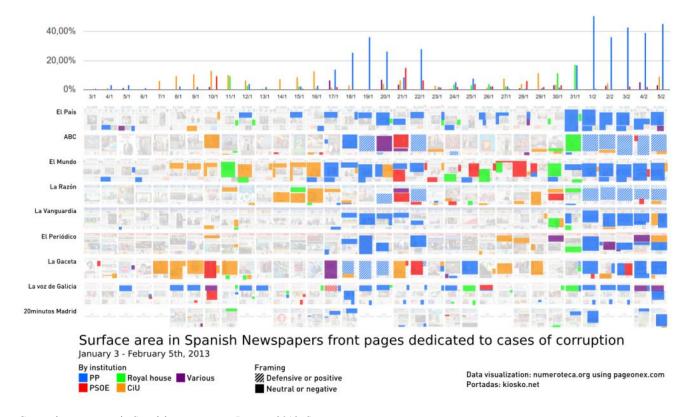
News media stories

For news media we will quantify corruption coverage by measuring: the number of stories; the size, by using the percentage of surface dedicated to the topic; the "color", which institution was related to corruption; and

---

[7] PageOneX.com uses front pages from the Kiosko.net online database.

the position. Informed by other studies and literature in the field of corruption coverage and media studies (Rivero & Fernández-Vázquez, 2011; Baumgartner & Chaqués Bonafont, 2015) we will also study the framing of corruption (Costas-Pérez, Solé-Ollé & Sorribas-Navarro, 2012), whether the coverage is neutral or negative or defensive and giving a positive view.

For research validity each front page and story should be coded for at least two coders to be able to measure the reliability of the data.

We will use PageOneX.com to analyze front pages of paper newspapers and StoryTracker to analyze home sites of online news sites. We will compare our data collection with other existing datasets regarding corruption coverage, such as the analysis of front pages conducted by the Spanish Policy Agendas project (Chaqués-Bonafont, Palau & Baumgartner, 2014) in the two largest newspapers in Spain: *El País* and *El Mundo*; and the analysis by the working group "Agenda y Voto" from the Centro de Investigaciones Sociológicas (CIS).

Corruption coverage in Spanish newspapers. January 2013. Source: numeroteca.org

Printed newspaper front pages

To conduct printed newspaper front page analysis we will use PageOneX.com, an open source software tool designed to aid the coding, analysis, and visualization of newspaper front pages that we have developed (Costanza-Chock and Rey-Mazón, 2016). In the past, measuring news stories involved obtaining physical copies of newspapers, measurement by hand, and manual input of measurements into a spreadsheet or

database, followed by calculation, analysis, and visualization. PageOneX.com automates some of these steps, simplifies others, and makes it possible for teams of investigators to conduct shared front page analysis online.

Newspapers: *El País, El Mundo, ABC, La Razón, La Vanguardia* and *El Periódico*.

Online news sites homepages

We will use similar metrics to printed newspapers for the home page of online newspapers. According to the Estudio General de Medios (2005) survey in Spain daily printed newspapers are losing audience, their penetration is now at 28.5% while in 2008 was 42.1%. In the same period, Internet penetration has risen from 29.9% to 66.7% and 6 among the 10 most visited sites are news websites. In the past years new online media are earning popularity in Spain, such as eldiario.es, elconfidencial.es or elespanol.com. Though they do not have the same prominence as printed and traditional newspapers, they are gaining an important audience, and have been responsible for discovering several important corruption scandals.

To study online media we will use the open source software StoryTracker (http://storytracker.pastpages.org) developed by Ben Welsh to track stories in news sites homepages. PastPages.org, the open database of homepages has started archiving the main Spanish online news sites for this study.

News online sites: all the online versions of printed newspapers together with *eldiario.es, elespanol.com, publico.es, 20minutos.es* and *elconfidencial.com*.

Social media messages: Twitter

We are going to analyze social media messages in Twitter with the t-hoarder platform software[8] developed by Mari Luz Congosto at the Universidad Carlos III. T-hoarder "captures messages posted by Twitter users, analyses and visualises the results, allowing the detection of the most viral moments and the most prominent users" (Congosto, 2016:11). We will also use her Twitter data set (available at http://t-hoarder.com) about certain corruption cases and tweets (Congosto, 2016). We are going to quantify conversations about different corruption scandals (search by words and related hashtag) by the number of messages, retweets, number of users and the topic they are referred to. We will also analyze user networks and the dissemination of tweets by the mainstream media in Twitter.

T-hoarder uses Searching API streaming to collect the tweets. Once acquired the data are cleaned of false positive before its processing[9].

---

[8] Code available https://github.com/congosto/t-hoarder
[9] We will use the methodology described by Congosto at https://github.com/congosto/t-hoarder/blob/master/README.md.

Public opinion and context

To provide related information to our stories data set and in order to measure impact in public opinion we are using two different approaches, existing public opinion surveys and focus groups related to corruption perception.

- Public opinion surveys: according to Villoria & Jimenez (2012) there are more than fifteen global and regional surveys about corruption (UNDP, 2008). In Spain we are going to use the monthly national survey, "the barometer", of the Centro de Investigaciones Sociológicas (CIS), that asks about the three more important problems for citizens, where corruption and fraud are since 2013 in the top three[10], or the surveys by the Observatorio Imagen de España by the Real Instituto Elcano. Palau & Davesa (2013) compared corruption coverage and public perception of corruption.
- Focus groups to provide context to understand how people "read" front pages, how they judge about scandals along personal ideologies. We want to contrast data from media coverage with direct perceptions of readers. Participants will be selected to have diverse affinity to political parties and different ideologies. We will use the newspapers front pages of the week as a starting point for a conversation about corruption.

A series of semistructured interviews will be conducted with editors, journalists and web managers of the online news sites under analysis to get insights about the process of configuration of the home pages.

2.6. Means and resources

From a technical perspective we have the support of the MIT Center for Civic Media for the PageOneX servers maintenance and software upgrades. We need to add a new feature to PageOneX to allow the use of multiple taxonomies[11] (https://github.com/numeroteca/pageonex/issues/82)

We have the support of Ben Welsh for the Spanish online news sites home page archive at pastpages.org and for the home page analysis software StoryTracker.
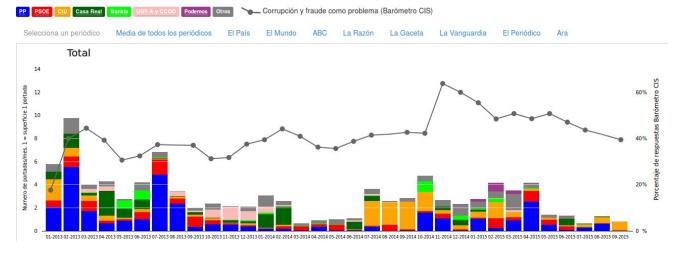
Mariluz Congosto provides her Twitter datasets collected for her own PhD dissertation (Congosto,2016) and support for the software t-hoarder.

I also have my own shared server with the Montera34 community where I host the blog of this phd (http://numeroteca.org/tag/phd) and data that I am collecting.

---

[10] In Castromil (2012) it is noted that public opinion (CIS) it is not always aligned to what mainstream media consider relevant: "Baste comprobar que, en numerosas ocasiones, las prioridades de la ciudadanía, medidas en las preguntas de agenda pública y personal que recogen los barómetros mensuales que realiza el Centro de Investigaciones Sociológicas (CIS), no coinciden con lo que los medios de comunicación consideran prioritario".

[11] We need to classify corruption stories by institution and also by frame (positive versus neutral or negative).

Superficie dedicada en portada a casos de corrupción. 2013-2015

As a starting point Ive been gathering the data for this research at the *The Color of corruption* public database (numeroteca.org/colorcorrupcion, Rey-Mazon, 2015). It is possible to access and visualize the percentage of front page, institution involved in the corruption scandals in front pages stories in the main Spanish newspapers since January 2013. This data have already been used in Rodón & hierro (2016) and in the *2015 Anuari dels Silencis Mediàtics* (Year book of media silence) edited by media.cat and the critical journalism Ramon Barnils Group (grupbarnils.cat). The recent use of these data demonstrates an increasing demand for this kind of datasets and research in the academic and journalism disciplines

The access to the *Jornadas de Periodismo de datos* and to the community *Periodismo de datos* at the Medialab Prado in Madrid provides easy access to key actors (editors, journalists and web managers) in the ain online news sites.

2.7. Schedule

	2016	2017	2018	2019
Literature review	x x x			
Data gathering	x x x			
Data analysis	x x x	x x		
Writing			x x	x

Literature review involves more than reading, it includes meta-analysis, proof-of-method related to dissertation idea and submit dissertation research plan proposal for approval.

Data gathering and data analysis are simultaneous processes, as the software we plan to use serve for data gathering, analysis and visualization. Data gathering will involve learning of software for home page scraping (StoryTracker) and Twitter analysis with T-hoarder and setting up the different software environments.

References

Álvarez Peralta, M. (2010). *La crisis en portada: Análisis de la construcción de la actual crisis económica en el discurso del diario El País*. D.E.A. dissertation, Department of Journalism III, Complutense University of Madrid, Spain.

Álvarez-Peralta, M. (2011). Popularization of the Scientific Knowledge about the Economic Crisis in the Spanish Mass Media: Trends, Inaccuracies and Omissions in the Documentation Process of Press Discourse on Economics. *The International Journal of Science in Society*, *2*, 203-218.

Anderson, C. W., Glaisyer, T., Smith, J., & Rothfeld, M. (2011). Shaping 21st century journalism. Leveraging a Teaching Hospital Model in journalism education. *New America Foundation*.

Bandari, Roja, Sitaram Asur, and Bernardo A. Huberman. "The pulse of news in social media: Forecasting popularity." *arXiv preprint arXiv:1202.0332*(2012). http://arxiv.org/pdf/1202.0332.pdf

Baumgartner, F. R., & Chaqués Bonafont, L. (2015). All News is Bad News: Newspaper Coverage of Political Parties in Spain. *Political Communication*, *32*(2), 268-291.

Bouza, Fermín *et al* (2013): "Génesis y desarrollo de los encuadres periodísticos en un sistema mediático polarizado: El caso español (2004-2008)" en Capdevila, Arantxa *et al* (eds): *Estudios de Comunicación Política. Libro del año 2012*. Madrid. Tecnos. http://www.tecnos.es/ficha.php?id=3261574

Boydstun, Amber E. *How policy issues become front-page news*. ProQuest, 2008. https://etda.libraries.psu.edu/paper/8802/4082

Boydstun, Amber E. *Making the news: Politics, the media, and agenda setting*. University of Chicago Press, 2013.

Boyle, T. P. (2001). Intermedia agenda setting in the 1996 presidential election. *Journalism & Mass Communication Quarterly* 78, no. 1: 26-44.

Broder, D. S. (2000). Behind the front page. Simon and Schuster.

Bruns, A., & Burgess, J. E. (2011). New methodologies for researching news discussion on Twitter.ç

Bruns, A., & Burgess, J. E. (2011, August). The use of Twitter hashtags in the formation of ad hoc publics. In *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*.

Bruns, A., & Burgess, J. (2012). Notes towards the scientific study of public communication on Twitter. *Science and the Internet*, 159-169.

Castromil, A. R. (2012). Polarización política y negativismo mediático: similitudes y diferencias en la prensa de derecha y la de izquierda en las elecciones autonómicas y municipales de 2011. *Redmarka: revista académica de marketing aplicado*, (8), 3. http://dialnet.unirioja.es/descarga/articulo/4125393.pdf

Castromil, Antón R. (2013): "Continuidad y cambio en la cobertura de una campaña electoral: De la prensa de referencia a Twitter en las elecciones del 20N" en Cotarelo, Ramón (ed.): *Ciberpolítica. Las nuevas formas de acción y comunicación política*. Valencia. Tirant lo Blanch. http://www.tirant.com/libreria/libro/ciberpolitica--las-nuevas-formas-de-accion-y-comunicacion-politicas-978 8415731085?busqueda=cotarelo&

Castromil, Antón R. (2013): "Negativismo mediático y pluralismo polarizado en la vida política madrileña" en Capdevila, Arantxa *et al* (eds): *Estudios de Comunicación Política. Libro del año 2012*. Madrid. Tecnos. http://www.tecnos.es/ficha.php?id=3261574

Centro de Investigaciones Sociológicas (2015). Barómetro.

Chaqués-Bonafont, L., Palau, A.M. and and Baumgartner, F.R. (2014). Agenda Dynamics in Spain, London: Palgrave. http://www.ub.edu/spanishpolicyagendas/datasetinstruments/

Chuang, J., Fish, S., Larochelle, D., Li, W. P., & Weiss, R. (2014). Large-scale topical analysis of multiple online news sources with Media Cloud. *NewsKDD: Data Science for News Publishing, at KDD*. http://ailab.ijs.si/~blazf/NewsKDD2014/submissions/newskdd2014_submission_4.pdf

Congosto M. T-hoarder dataset. Universidad Carlos III. Departamento de Ingeniería Telemática, Grupo de investigación WebTlab.

Congosto Martínez, M. L. (2016). Caracterización de usuarios y propagación de mensajes en Twitter en el entorno de temas sociales. http://e-archivo.uc3m.es/bitstream/handle/10016/22826/tesis_maria-luz_congosto_martinez_2016.pdf

Costanza-Chock, S., & Rey-Mazón, P. (2016). PageOneX: New Approaches to Newspaper Front Page Analysis. *International Journal of Communication*, *10*, 28. http://ijoc.org/index.php/ijoc/article/viewFile/4442/1652

Costas-Pérez, Elena. "Essays on the Political Economy of Local Corruption." (2014). http://www.tdx.cat/handle/10803/287837

Costas-Pérez, E., Solé-Ollé, A., & Sorribas-Navarro, P. (2012). Corruption scandals, voter information, and accountability. *European Journal of Political Economy*, *28*(4), 469-484.

Deuze, M. (2003). The web and its journalisms: considering the consequences of different types of newsmedia online. *New media & society*, *5*(2), 203-230.

Deuze, M. (2005, 11). What is journalism?: Professional identity and ideology of journalists reconsidered. *Journalism, 6*(4), 442-464. doi:10.1177/1464884905056815

Di Tella, R., & Franceschelli, I. (2009). *Government advertising and media coverage of corruption scandals* (No. w15402). National Bureau of Economic Research. http://www.nber.org/papers/w15402.pdf

Donovan, H. L. (1925). The Content of Ordinary Reading. *The Elementary School Journal*, *25*(5), 370-375.

Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, *8*, 20. http://ijoc.org/index.php/ijoc/article/viewFile/2171/1159

Farley, B. M. (1930). *What to Tell the People about the Public Schools.* In the Journal of Educational Research, 21(1), 52-53.

Forrest, W. (1934). *Behind the Front Page: Stories of Newspaper Stories in the Making*. D. Appleton-Century Company, incorporated.

Fundación Alternativas (2007): Urbanismo y democracia. Alternativas para evitar la corrupción, http://www.falternativas.org/.

Gans, H. J. (1979). *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. Northwestern University Press.

Golan, G. (2006). Inter-media agenda setting and global news coverage: Assessing the influence of the New York Times on three network television evening news programs. *Journalism studies*, *7*(2), 323-333.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2012). Assessing the bias in communication networks sampled from Twitter. arXiv preprint. *arXiv preprint arXiv:1212.1684*. http://arxiv.org/pdf/1212.1684.pdf

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press.

Herman, E. S., & Brodhead, F. (1986). *The rise and fall of the Bulgarian connection*. Sheridan Square Pubns.

Herman, E. S., and Chomsky, N. (1988). *Manufacturing consent: the political economy of the mass media.* New York: Pantheon.

Hermida, A. (2012). Tweets and truth: Journalism as a discipline of collaborative verification. *Journalism Practice*, *6*(5-6), 659-668. http://newmediahistory.pbworks.com/w/file/fetch/71318091/Tweets_and_Truth_preprint.pdf

Hesmondhalgh, D. (2007). The cultural industries. London: Sage.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767.

Jansen, Bernard J., et al. "Micro-blogging as online word of mouth branding." *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009.
https://www.researchgate.net/profile/Jim_Jansen/publication/228551558_The_Commercial_Impact_of_Social_Mediating_Technologies_Micro-blogging_as_Online_Word-of-Mouth_Branding/links/02e7e51f8ad2942c03000000.pdf

Kelly, John. "Pride of place: Mainstream media and the networked public sphere." (2008).
http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Pride%20of%20Place_MR.pdf

Kirkland, K.. 3 takeaways from the 'death of the homepage' and The New York Times innovation report. Poynter.org. 2014.
http://www.poynter.org/2014/3-takeaways-from-the-death-of-the-homepage-and-the-new-york-times-innovation-report/252632/

Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism studies*, *13*(1), 19-36.

Lewis, S. C., Kaufhold, K., & Lasorsa, D. L. (2010, 04). Thinking About Citizen Journalism. *Journalism Practice, 4*(2), 163-179. doi:10.1080/14616700903156919

Li, X. (2013). *Internet newspapers: The making of a mainstream medium*. Routledge.
https://books.google.es/books?hl=en&lr=&id=fVjaAAAAQBAJ&oi=fnd&pg=PA13&ots=rzG_S9WWIk&sig=ahnZ55n6jY0wdluSj4n-pPnRpmo#v=onepage&q&f=false

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., & Pearce, I. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International journal of communication*, *5*, 31.

Marraco, M. (2013, January 16). Suiza informa de que Bárcenas llegó a tener 22 millones. *El Mundo*.
Retrieved from http://www.elmundo.es/elmundo/2013/01/16/espana/1358344958.html

McCombes, M., Lopez‑Escobar, E., & Llamas, P. J. (2006). Setting the agenda of attributes in the 1996 Spanish general election. *Journal of Communication* 50, no. 2: 77-92.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, *36*(2), 176-187.

McCombs, M. E., & Shaw, D. L. (1976). Structuring the "unseen environment". *Journal of Communication*, *26*(2), 18-22.

McCombs, M. E., & Shaw, D. L. (2006). The evolution of agenda‑setting research: Twenty‑five years in the marketplace of ideas. *Journal of Communication* 43, no. 2: 58-67.

McCombs, M. E., & Shaw, D. L. (2006). The evolution of agenda‑setting research: Twenty‑five years in the marketplace of ideas. *Journal of Communication* 43, no. 2: 58-67.

Mercado, F., Jiménez, M., Cué C.E., Romero, J.M., (January 31st, 2013). Las cuentas secretas de Bárcenas. *El País*. Retrieved from http://politica.elpais.com/politica/2013/01/30/actualidad/1359583204_085918.html

Negroponte, N. (1996). *Being digital*. Vintage.

Newman, N. (2009). The rise of social media and its impact on mainstream journalism. *Reuters Institute for the Study of Journalism*, *8*(2), 1-5.

Newman, N. (2011). Mainstream media and the distribution of news in the age of social discovery. *Reuters Institute for the Study of Journalism, University of Oxford*.

Palau, A. M., & Davesa, F. (2013). The Impact of Media Coverage of Corruption on Spanish Public Opinion. *Revista Española de Investigaciones Sociológicas*, *144*(1), 97-124. http://reis.cis.es/REIS/PDF/REIS_144_05_ENGLISH1381482849627.pdf

Phillips, A., Singer, J. B., Vlad, T., & Becker, L. B. (2009, 03). Implications of Technological Change For Journalists' Tasks and Skills. Journal of Media Business Studies, 6(1), 61-85.

Phillips, A. (2010, 08). Transparency And The New Ethics Of Journalism. Journalism Practice, 4(3), 373-382.

Philp, M. "Corruption definition and measurement." Measuring corruption (2006): 45-56. http://s3.a

Rey-Mazón, P. (2013). "3 steps to measure the corruption coverage in Spain." *MIT Center for Civic Media Blog*. Web. Accessed October 26, 2015. https://civic.mit.edu/blog/pablo/3-steps-to-measure-the-corruption-coverage-in-spain

Rey-Mazón, P. (2013). Newspaper Front Page Analysis: How Do They Tell the Story?. Oberösterreich, L. A. (2013). *Accountability technologies: tools for asking hard questions*. D. Offenhuber (Ed.). Ambra| V.

Rey-Mazón, P. (2015). The color of corruption. "El Color de la corrupción." *Numeroteca.org*. Accessed October 24, 2015. http://numeroteca.org/colorcorrupcion/

Reese, S. D., & Danielian, L. H. (1989). A closer look at intermedia influences on agenda setting: The cocaine issue of 1986. *Communication campaigns about drugs: Government, media, and the public*, 47-66.

Rivero, G., & Fernández-Vázquez, P. (2011). Las consecuencias electorales de los escándalos de corrupción municipal 2003-2007. *Estudios de Progreso*, *59*(201), 1.

Salaverría, R., & Sancho, F. (2007). Del papel a la Web. Evolución y claves del diseño periodístico en internet. *Diseño periodístico en internet*, 207-239.

Sampedro Blanco, Victor. "Batallas de papel. Medios de comunicación y nuevos movimientos sociales, el caso de la objeción de conciencia." *ZER* 1 (1996): 121-154.

Shirky, C. (2009). *Here comes everybody: How change happens when people come together*. Penguin UK.

Snider, P. B. (1967). "Mr. Gates" revisited: A 1966 version of the 1949 case study. *Journalism Quarterly*, 44, 419-427.

Strate, L. (2004). A Media Ecology. *Communication Research Trends*, *23*(2), 1-48

Sumpter, R. S. (2000). Daily newspaper editors' audience construction routines: A case study. *Critical Studies in Media Communication*, *17*(3), 334-346.

Sunstein, C. R. (2009). *Republic. com 2.0*. Princeton University Press.

UNDP (United Nations Development Program) and Global Integrity: *A Users' Guide to Measuring Corruption*. UNDP Oslo Governance Centre, 2008. lulu.com/content/2523351

Villoria, M., & Jiménez, F. (2012). La corrupción en España (2004-2010): datos, percepción y efectos/Corruption in Spain (2004-2010): Data, Perception and Consequences. *Reis*, 109-134. http://www.reis.cis.es/REIS/PDF/REIS_138_061333356669068.pdf

Walker, Shawn, et al. "SoMe tools for social media research." (2013).

White, D. M. (1950). The gatekeeper: A case study in the selection of news. *Journalism quarterly*, *27*(4), 383-390.

Ye, S., & Wu, S. F. (2010). *Measuring message propagation and social influence on Twitter. com* (pp. 216-231). Springer Berlin Heidelberg. http://ijoc.org/index.php/ijoc/article/viewFile/2171/1159

Yeuell, G. H. (1928). The Negro press as a factor in education. *Journal of Educational Sociology*, *2*(2), 92-98.

Zuckerman, Ethan. *Rewire: Digital cosmopolitans in the age of connection*. 2013.