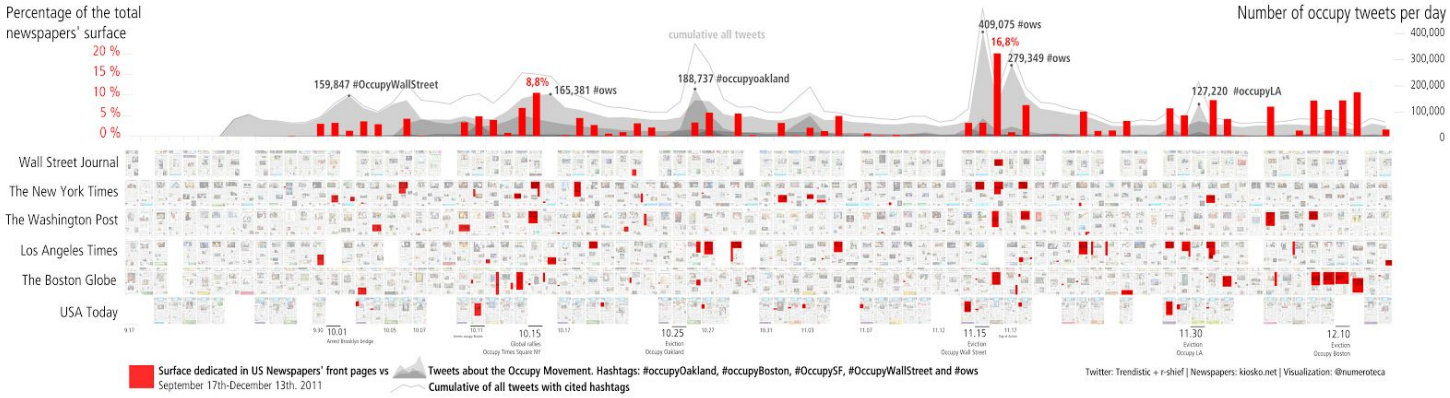
PageOneX: New Approaches to Newspaper Front Page Analysis

Abstract

PageOneX is a Free/Libre and Open Source Software tool designed to aid the coding, analysis, and visualization of newspaper front pages. Communication scholars have long analyzed newspaper front pages, using front page column-inches as an important indicator of mass media attention. In the past, this 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 automates some of these steps, simplifies others, and makes it possible for teams of investigators to conduct shared front page analysis online. We review scholarship in this area, describe our approach in-depth, and provide concrete examples of analysis conducted using PageOneX.

Introduction

Figure 1: PageOneX analysis of Occupy Wall Street newspaper front pages vs. tweets



Source: PageOneX:
http://numeroteca.org/2011/12/17/surface-newspapers-front-pages-vs-twitter-dec13th-occupy-ows

In September of 2011, the Occupy movement exploded into public consciousness, spreading rapidly from the initial encampment at Zuccotti Park to more than 950 sites across the United States and around the world, in more than 80 countries (Kidd, 2014). Tens of thousands of protesters used a wide range of highly mediated tactics to draw attention to growing economic inequality: from the "People's Microphone" to Twitter, from print newspapers such as the Occupied Wall St. Journal to the GlobalRevolution.tv livestream, and from traditional press releases to crowdfunded, professionally produced television ads ([Author, 2012]; Kidd, 2014; Kavada, 2015). Occupy was a cross-platform mobilization process; it took place, as Paulo Gerbaudo notes, both in 'tweets and the streets' (Gerbaudo, 2012). Occupy was a also a cross-platform, participatory, action-linked, and community-accountable form of media mobilization, described as *transmedia organizing* by [Author], who argues that this combination of characteristics is the most effective social movement strategy within the new, converged media ecology ([Author, 2014]). Although public discourse

about the movement emphasized its extensive use of social media, Occupy was also covered extensively by mainstream print and broadcast news, as demonstrated in Figure 1 (above). This visualization, a comparison of newspaper front page stories vs. tweet volume about Occupy between September 17th and December 13th, 2011, was generated using a method that would later be systematized in PageOneX, a Free/Libre and Open Source Software (F/LOSS) tool for media analysis that we present in the pages that follow. The visualization displays both the percentage of newspaper front page surface area dedicated to the Occupy movement (in red) and the number of tweets from various related hashtags (#ows, #occupywallstreet, #occupyLA, #occupyOakland, and #occupyBoston, in grey shaded areas).[1] On any given day in this time period, Occupy was more likely than not to be on the front page of at least one national newspaper (New York Times, Washington Post, Los Angeles Times, Boston Globe, Wall St Journal, and USA Today).[2] The ultimate impacts of the Occupy movement are still being debated;[3] what is not contested is that Occupy received sustained attention across the media ecology. Just how to track and understand attention in an increasingly complex media ecology is a matter of great interest.

Communication scholars have long used column-inches of print newspaper coverage as an important indicator of mass media attention. 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. This laborious and time-consuming process was followed by calculation, analysis, and sometimes data visualization. Our approach automates some of these steps, and dramatically simplifies others. Widespread access to the broadband internet also makes it far easier for distributed teams of investigators to work with

[1] Tweets per day are based on an archive of 12 million Occupy related tweets, released by r-shief.org in the context of the #occupydata hackathon. For more information, see http://occupyresearch.net.
[2] Comparing newspaper coverage with the twitter data reveals that peaks in Tweet volume of Occupy hashtags tend to precede peaks in newspaper front page coverage, although this is not always the case. Future research might explore the predictive power of tweet volume in relationship to next-day front page coverage, and vice versa, as well as explore the significance of moments of convergence and divergence between social media and newspaper attention.
[3] See Costanza-Chock, Schweidler, and DeTar, 2013: http://bit.ly/ows-rise-fall-rise.

common datasets that are hosted online, and recent advances in javascript libraries facilitate the creation of user-friendly graphical user interfaces (GUIs) for content coding tasks. In this paper, we present PageOneX: a Free/Libre and Open Source Software tool designed to simplify the coding, analysis, and visualization of front page print newspaper coverage of news stories and media events. We provide a brief review of scholarship in this area, describe our approach in-depth, give concrete examples of analysis conducted using PageOneX, and conclude with a description of our next steps and an invitation to communication scholars to use the platform and participate in the future development of the tool.

Newspaper front pages: still important in the broader media ecology

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. Many print newspapers have transformed into cross-platform news organizations, launched online editions, developed increased capacity for multimedia reporting, adapted to the 24-hour news cycle, and reduced their number of printed pages or the frequency of print publication. In some cases, newspapers have abandoned their printed editions entirely (The Economist, 2011). However, while print newspaper circulation is, for the most part, declining in the United States, massive and newly literate publics in Brazil, India, China, and other recently industrialized nations are driving an unprecedented rise in total global newspaper circulation (Ibid.) Fears about the supposed death of the printed newspaper are thus, at least in transnational perspective, and at least for the moment, quite parochial.

Not only is newspaper reach growing globally, but even in countries where print circulation is declining, newspapers 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 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). This effect does not seem to be limited to the United States; for example, Qian (2009) found that Chinese newspapers largely set the agenda for Chinese blogs. 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.

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).

Manufacturing Consent and measuring in Column-Inches

While media anthropologists have employed qualitative methods to explore the dynamics that shape the production of the newspaper's front page, the most widely-read and cited study of mass media attention is *Manufacturing Consent*, by Ed Herman and Noam Chomsky (1988). Herman and Chomsky employ 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 (Herman and Chomsky, 1988). For instance, they use all of these measures to compare a series of paired examples of New York Times (NYT) coverage of murders of religious victims in US client states vs. US enemy states:

Figure 2: Mass-Media Coverage of Worthy and Unworthy Victims

	NEW YORK TIMES								TIME and NEWSWEEK				CBS NEWS			
	Articles[1]		Column inches		Front-page articles		Editorials[1]		Articles[1]		Column inches		No. of news programs[1]		No. of evening news programs	
	No.	% of row 1	No.	% of row 1	No.	% of row 1	No.	% of row 1	No.	% of row 1	No.	% of row 1	No.	% of row 1	No.	% of row 1
Victims																
1. Jerzy Popieluszko, murdered on Oct. 19, 1984	78	(100)	1183.0	(100)	10	(100)	3	(100)	16	(100)	313.0	(100)	46	(100)	23	(100)
2. 72 religious victims in Latin America, 1964–78[2]	8	(10.3)	117.5	(9.9)	1	(10)	—	—	—	—	16	(5.1)	—[3]	—	—	—
3. 23 religious, murdered in Guatemala Jan. 1980–Feb. 1985[4]	**7**	**(9.0)**	**66.5**	**(5.6)**	**—**	**—**	**—**	**—**	**2**	**(12.5)**	**34.0**	**(10.9)**	**2**	**(4.3)**	**2**	**(8.7)**
4. Oscar Romero, murdered Mar. 18, 1980	16	(20.5)	219.0	(18.5)	4	(40)	—	—	3	(18.8)	86.5	(27.6)	13	(28.3)	4	(17.4)
5. 4 U.S. religious women, murdered in El Salvador, Dec. 2, 1980	26	(33.3)	201.5	(17.0)	3	(30)	—	—	5	(31.2)	111.0	(35.5)	22	(47.8)	10	(43.5)
6. Total of lines 2–5	57	(73.1)	604.5	(51.1)	8	(80)	—	—	10	(62.5)	247.5	(79.1)	37	(80.4)	16	(69.6)

Source: Herman & Chomsky[4], 1988, p. 185

Herman and Chomsky argue that the ideological bias of the US ruling class is reflected in the far more extensive coverage of murders that take place in 'enemy states,' compared with relative inattention to similar murders when they take place in 'client states,' as measured by column inches, number of stories, number of editorials, number of front page articles, and other quantitative indicators.[5] In another article that employs a similar methodology, Herman and Brodhead (1986) compare NYT coverage of two airline shootdowns, the 1983 Soviet Union downing of Korean flight KAL 007, and the 1973 Israeli downing of a Libyan civilian airliner. They find the ratio of column-inch space in the two cases to be 5.7/1. The authors argue that this skewed column-inch

[4] The media coverage for an 18 month period from the time of the first report of the victim's disappearance or murder.
[5] These data can also be used to explore the relationship between front page articles and the total number of articles in the newspaper about a certain topic. According to Chomsky and Herman: Number of articles total / Number of articles in front page (ratio): 1. 78/10 (7.8/1). Jerzy Popieluszko; 2. 8/1 (8/1). 72 religious victims; 3. 7/0 (-). 23 religious victims; 4. 16/4 (4/1). Oscar Romero; 5. 26/3 (8.6/1). 4 US religious victims; 6. 57/8 (7.1/1). Average 2-5. The ratios vary between 4 and 8, but seem very consistent between the different cases, usually around 8. Wolfe, Boydstun & Baumgartner (2009) also study the relationship between front page articles and total articles in depth.

ratio reflects the shaping of news coverage by the NYT to fit the ideological bent of the Reagan administration: while both events are tragic, the attack on the Korean flight receives 5.7 times as much coverage as the attack on the Libyan flight, because the Soviet Union is thought of by newspaper editors as the 'enemy' of the United States while Israel is thought of as a 'friend.'

While column-inch measurement gained great visibility in the 1980s, largely due to the popularity of *Manufacturing Consent,* it was not then a new approach. 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 in the New York Age, Pittsburgh Courier, and Chicago Defender by content types 'social betterment,' 'neutral,' and 'antisocial.' Yuell then measured each type of material in column inches,[6] and analyzed the percent of the total content by category:

Figure 3: Type of Material in three Black newspapers

Paper	Type of Material	Per cent of Amount
New York Age	Social Betterment	65.69
Pittsburgh Courier	Social Betterment	46.20
Chicago Defender	Social Betterment	37.11
New York Age	Neutral	22.97
Pittsburgh Courier	Neutral	33.30
Chicago Defender	Neutral	37.57
New York Age	Antisocial	11.34
Pittsburgh Courier	Antisocial	20.50
Chicago Defender	Antisocial	25.32

Source: Yeuell, 1928: 96

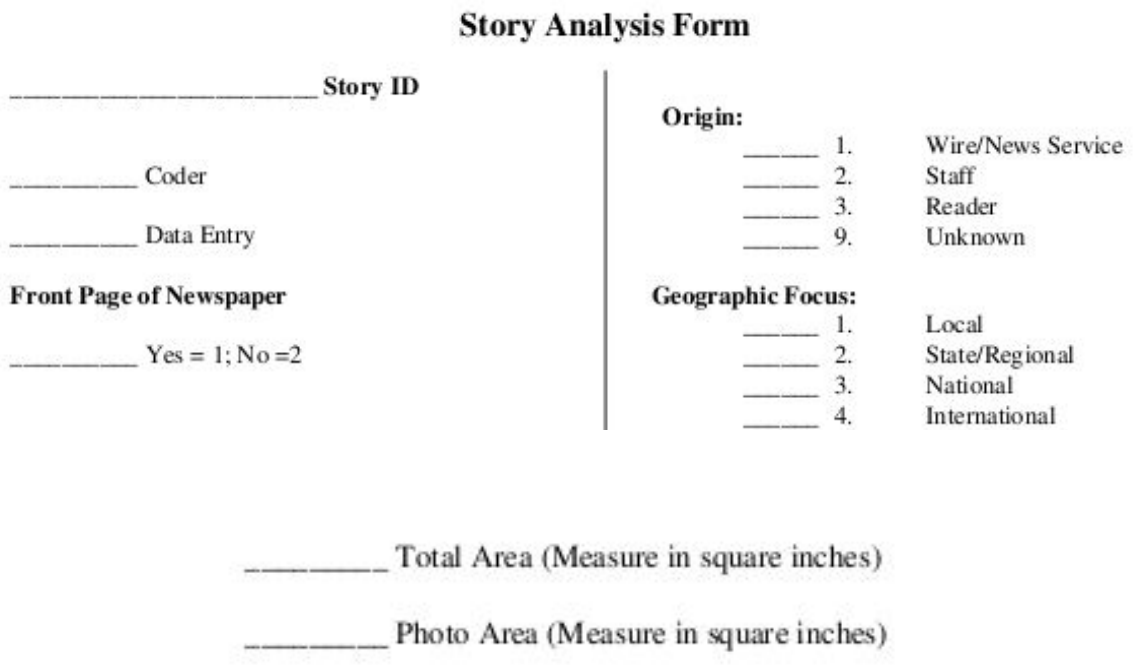
Yeuell did not provide the sum total of column-inches, but rather, the percentage of total coverage occupied by each type of news. This approach made comparison possible across different newspaper sizes and formats, and we have incorporated this insight into the PageOneX tool.

[6] Yeuell: "The method used in analysis was the relative amount of space devoted to different types of news matter. Or, to be more specific, the unit of measure is the column inch. A column inch is an item of news one inch long and the width of a column." p.96.

Researchers employing similar methods have, over time, developed key metrics including number of articles, number of column lines, number of editorials, number of front-page articles, and column-inches. In the 1920s, education scholar H. L. Donovan used quantitative content analysis of the Chicago Daily Tribune to explore the question 'what do adults read?' He found that crime was by far the most frequent subject, followed by accidents, then government and politics, as measured by column inches and column lines (Donovan, 1925). In *An Analysis of the Use of Column Inches in School Papers* (1926), the same author measured column inches, then calculated and visualized the percentage of coverage dedicated to various topics in school newspapers. In 1930, Belmont Mercer Farley analyzed 41,000 column inches of newspaper coverage about public schools, in order to develop recommendations about how to best communicate the need for the public school system to a skeptical public (Farley, 1930). Measurements of column-inches were also used early on to provide an indicator of attention, and to quantify the news impact of events.[7] 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. For example, Lynch and Peer (2002) developed a step-by-step guide to newspaper content analysis that includes an overview of area measurement:

[7] For example, Clark (1930) describes analysis of the number of articles, column inches, editorials, and photographs devoted to coverage of the Iowa meeting of the American Association for the Advancement of Science (AAAS): “In advance of the meeting there were published 1,710 articles (including 30 editorials and 36 photographs) totaling 10,998 column inches. In addition, advance notice of the meeting was given from 15 radio stations. During the meeting there were published 7,436 articles (including 33 editorials and 88 photographs) totaling 74,781 column inches. The press of the State of Iowa alone therefore published in connection with the meeting 9,146 articles (including 63 editorials and 124 photographs) totaling 85,779 column inches. In the Des Moines papers the meeting was noticed by 15 eight-column headlines.” (Clark, 1930:218).

Figure 4: Story Analysis Form



Story Analysis Form

_____ **Story ID**

_____ Coder

_____ Data Entry

Front Page of Newspaper

_____ Yes = 1; No =2

Origin:

_____ 1. Wire/News Service
_____ 2. Staff
_____ 3. Reader
_____ 9. Unknown

Geographic Focus:

_____ 1. Local
_____ 2. State/Regional
_____ 3. National
_____ 4. International

_____ Total Area (Measure in square inches)

_____ Photo Area (Measure in square inches)

Source: Lynch & Peer, 2002. Analyzing newspaper content. A how to guide.

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 (http://www.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 newspapers 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. Most recently, quantitative approaches to measuring and visualizing news attention flows are experiencing a new boost in visibility, for a variety of reasons. Massively increased information flows, greater availability of datasets, and the steady increase in the power and sophistication of Natural Language Processing, Machine Learning (NLP/ML), and

tools for 'Big Data' analysis and visualization have combined to generate a surge of interest in this area.[8] At the same time, widespread access to the broadband Internet provides new possibilities for geographically distributed teams of researchers to collaborate on content analysis. We felt that the time was ripe to build new tools for front page analysis, rooted in an understanding of the history of the subfield and based on a user-centred design process. In the following section, we describe our approach to front page newspaper analysis and data visualization, made publicly available via the PageOneX platform.

PageOneX

Informed by the existing scholarship on front page newspaper analysis, and based on a data visualization process developed by [Author], we developed a Free/Libre and Open Source Software (F/LOSS) tool for newspaper front page analysis and visualization called PageOneX. The software, initially created to run locally on the end user's computer, is now available as a free hosted service at http://pageonex.com, and the code and documentation are available at https://github.com/numeroteca/pageonex/wiki. PageOneX is designed to facilitate the relatively rapid creation of 'small multiple' data visualizations of front page newspaper coverage. 'Small multiple', a term popularized by Edward Tufte (2001), is a kind of data visualization that affords the visual comparison of multiple series of data: series are displayed separately and placed next to each other. In the case of PageOneX, the data displayed include small images of the newspaper front pages themselves, as well as the surface area on the newspapers' front pages that is dedicated to a specific news story (a 'thread') that runs chronologically along the x-axis (time). Coders use a point-and-click web interface to manually select the portion of front pages dedicated to that thread. The selected areas are then displayed as shaded or colored transparencies on top of the original newspaper front pages. When viewed at sufficiently close range (zoomed in to a story), the text of

[8] For examples, see http://blog.pageonex.com/references/#visualizations.

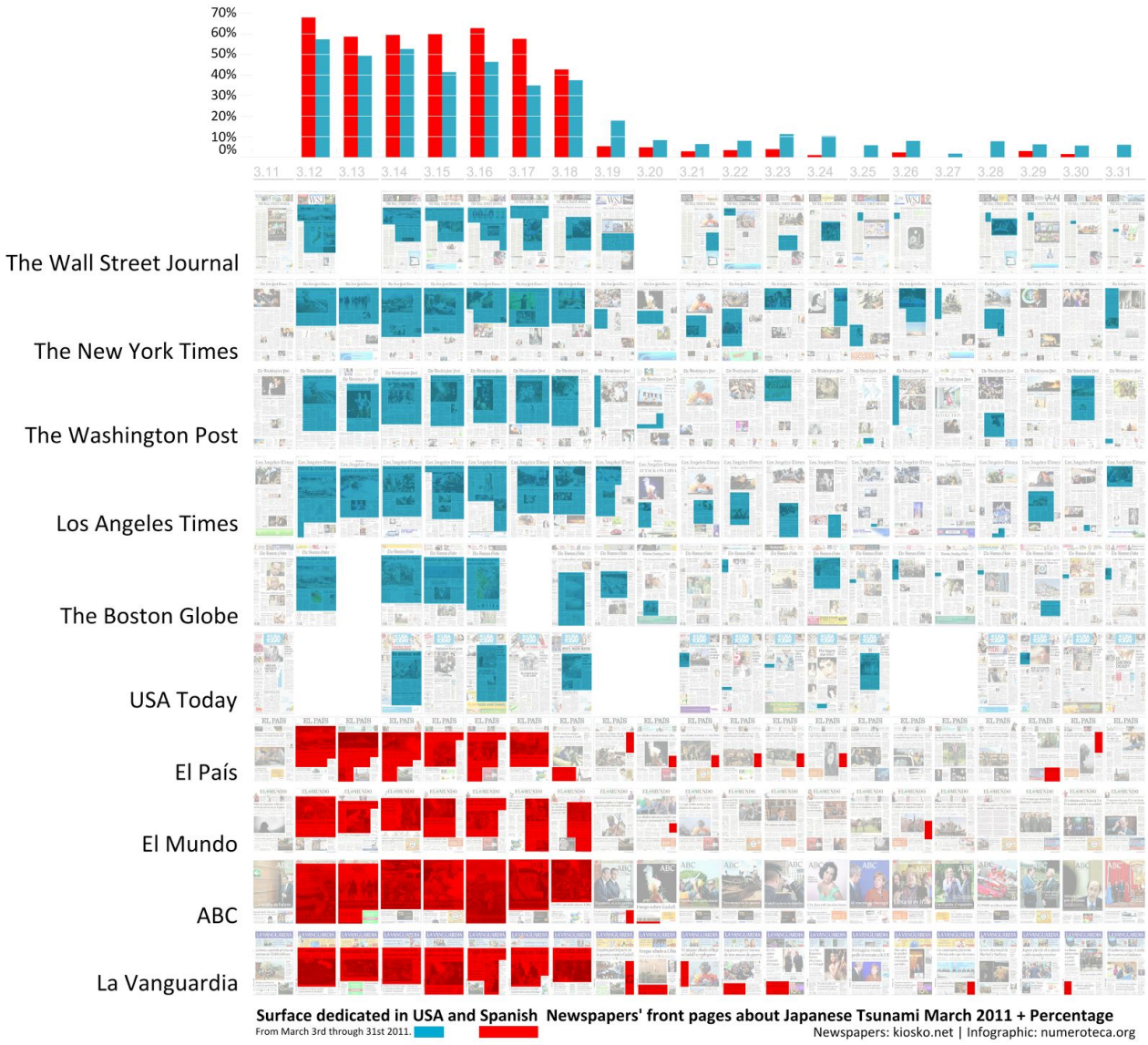
the original story can still be easily read. However, this form of graphical display of information is most compelling at a larger scale: small thumbnails of the newspaper front pages are arrayed chronologically from left to right, and the viewer is able to see at a glance the evolution of coverage, or the lack of coverage, of a particular news thread over time, as well as compare front page coverage between different newspapers.

Our approach allows the viewer to have a sense of the overall front page coverage trend, while retaining the ability to zoom in at any point to the raw data: images, headlines, and article text. [9] We also automatically measure the surface area of highlighted (coded) text, relative to the size of the newspaper front page, and produce a bar chart of the percentage of front page coverage over time. To make this measurement comparable between different newspapers and countries[10] we use the percentage of total surface area, rather than the real area in square inches or millimeters (as did Yeuell, as early as 1928, and similar to Tella & Franceschelli, in 2009). For example, we used this procedure to compare coverage of the 2011 Japanese Tsunami between broadsheet-sized US newspapers and berliner-sized Spanish newspapers:

[9] For an example of zooming in to raw data, and a process description of building the tool, see http://numeroteca.org/2012/06/06/pageonex-ready-steady-go.

[10] The area of newspaper front pages changes depending on their size and format. For example, there are 'tabloids,' 'berliners,' 'broadsheets,' and other standard sizes, usually related to a type of newspaper or to a region or country. For more see http://www.papersizes.org/newspaper-sizes.htm.

Figure 5: Spanish vs. US newspaper coverage of the 2011 Japanese Tsunami



Source: Numeroteca.org: http://numeroteca.org/?p=1340.

This visualization displays small multiples of four Spanish (red) and five US (blue) newspapers' front page coverage of the 2011 Japanese Tsunami, from March 3rd (left) through March 31 (right). Along the top of the visualization, we use a bar chart to display the percentage of total front page surface area dedicated to the topic.[11] The basic PageOneX visualization displays both the data itself (front pages and highlighted surfaces), as well as a visualization of the measurement in a bar or line

[11] When an image of a individual front page is not available, the graphic display of the thread contains a gap. The percentage of surface area is always calculated from the total surface area of the available front pages.

chart. Newspaper front pages provide one surface area datum per day, so a line connecting 2 surface area data points (from two consecutive days) produces a false sense of continuity.[12] Line charts are useful, however, to compare more than 2 news threads, as we shall see in the examples from the Arab Spring below.

Technical background

PageOneX.com is a Free/Libre and Open Source platform (licensed under the GNU Affero GPL v3) for newspaper front page analysis and visualization. Initially, [Author 2] developed this type of data visualization through a 'manual' process: he downloaded images of newspaper front pages from the web, then used a vector graphics program to draw rectangular transparencies on top of them in order to highlight certain stories. Next, he developed a script using the programming tool Processing (http://processing.org) in order to automatically a) download newspaper front pages from the web service kiosko.net, and b) build a Scalable Vector Graphics[13] file with the array of front page images, to be coded by manually drawing and arranging colored transparencies. In a third stage of development, [Author 2] worked with Google Summer of Code intern Ahmd Refat, then with developers Rahul Bhargava and Edward Platt at the MIT Center for Civic Media, to build PageOneX.com, a hosted web platform built in Ruby on Rails that allows anyone to use the PageOneX tool. PageOneX.com was developed to provide a ready-to-use front page analysis tool for anyone with a connection to the Internet. It was designed to lower the barrier to software-based front page newspaper analysis by avoiding the technical problems of installation, and to make the raw data and data visualizations publicly available. The platform semi-automates the process of newspaper selection, download, thread coding, and data visualization.

[12] Bar charts give a more accurate view of the surface area devoted to the story, whereas line charts run the risk of suggesting a visual interpolation of discrete data. For an illustration, see an analysis of Spanish newspaper coverage of the 2010 Haitian earthquake, using line charts, at *http://numeroteca.org/2011/11/20/prensa-terremoto-haiti*.
[13] Scalable Vector Graphics (SVG) is a family of specifications of an XML-based file format for describing two-dimensional vector graphics, both static and dynamic.

PageOneX.com allows users to create their own threads, collaborate with others, and browse threads created by others. To open a thread, the user fills in a form and provides a title and time frame, then selects the newspapers that she wants to study and creates the topics (categories) that she wants to code. The user also provides a description of the categories. Once the form is submitted, the tool automatically downloads the newspaper front page images that match the time frame and the selected newspapers (the images are downloaded from web platform http://kiosko.net).[14] Once the thread is created, the user begins coding. The interface presents one front page image at a time: the user can point, click, and drag multiple rectangles to select and code areas on the page. She categorizes each rectangle using one of the previously created topics. Alternatively, the user can tag the front page image as "nothing to code," to indicate that the image has been reviewed, but contains no article related to the study topics. The user iterates through this process to code the front pages of all newspapers in the selected date range. There is also a 'display' view, or dynamic visualization of the coding that has taken place so far. The display view shows the array of thumbnail images (every row is one newspaper, and every column is one day) with the highlighted areas in color. Clicking on any thumbnail displays a full size image of the newspaper front page, which makes it possible to read the newspaper content. Images that have already been coded (either by drawing a rectangle or with the "nothing to code" button) are visually differentiated from uncoded images, to help the user quickly see which pages must still be coded. Additionally, above the matrix of images, there is a dynamically generated bar chart that displays the percentage of total front page surface area dedicated to each topic, for each day. Additionally, the platform supports collaboration among teams of media researchers, who can take part in a shared coding process. PageOneX makes it possible to crowd-source the coding process, an approach that is increasingly viable since front page newspaper stories are major media events that attract attention

[14] Another site that provides images of current newspaper front pages is the Newseum. However, the Newseum's archive of front pages is currently stored locally on DVDs, so it is not accessible via the web. We hope to collaborate with the Newseum in the future on a searchable, codable archive. See http://newseum.org/todaysfrontpages/flash.

from diverse scholars, activists, and researchers who may not already know one another, but have a shared interest in tracking attention to a particular story. Users may contribute to existing projects by joining a thread, or they may generate their own news threads to follow.

PageOneX has a number of advantages over previous approaches to coding newspaper front pages: it avoids the need to gather physical copies of the newspapers; web-based coding and measurement is much faster than manual measurement with a ruler; it is possible to easily work in teams distributed across multiple locations; and the reader or viewer of the visualization has immediate, interactive access to the source data, not only to the researcher's tables and visualizations. The PageOneX team has also developed clear, comprehensive documentation, both for end-users who plan to use the software to analyze particular stories, and for developers who would like to install, customize, or extend the functionality of the system. As the user community of PageOneX grows, we are working to ensure that our development process is as open and transparent as possible. Our roadmap, detailing our timeline for new features, issue tracker for bug reports and feature requests, and code base, for developers who would like to modify, improve, or extend our code, can all be found at PageOneX.com.

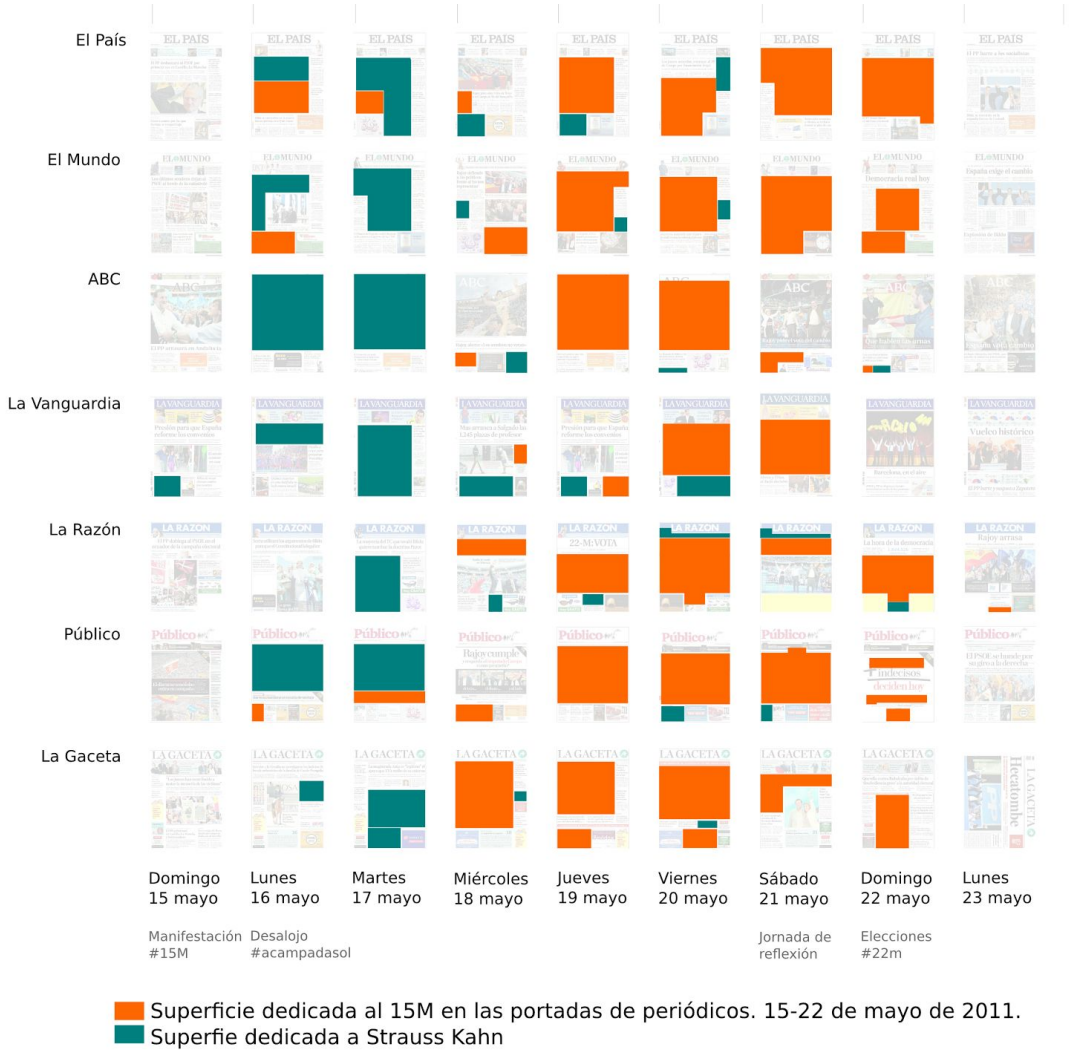
Examples of analysis with PageOneX

In this section, we describe a series of examples of media research that uses PageOneX to analyze and visualize front page news attention in a range of situations. We explore the use of PageOneX to compare ‘battles for attention’ between co-occurring news stories; to compare the length of the news cycle between asynchronous stories; to compare front page newspaper attention with attention on Twitter; and to compare multiple newspapers’ coverage of the same story.

A. Battles for attention between co-occurring stories

Communication scholars have used various approaches to study news agenda competition between issues (McCombs and Shaw, 2006). One use of PageOneX is as a tool to examine how unrelated news stories collide on the surface of newspapers. For example, we examined coverage of the detention of Dominique Strauss-Kahn, former manager of the International Monetary Fund, following allegations that he had sexually assaulted a hotel employee, alongside coverage of the *Indignados* movement in Spain:

Figure 7: Front Page Coverage of Strauss Kahn detention vs *Indignados* movement in 7 Spanish newspapers, May 15th-23rd, 2011.



Source: Numeroteca: "Front page surface dedicated to 15M (Indignados movement) and

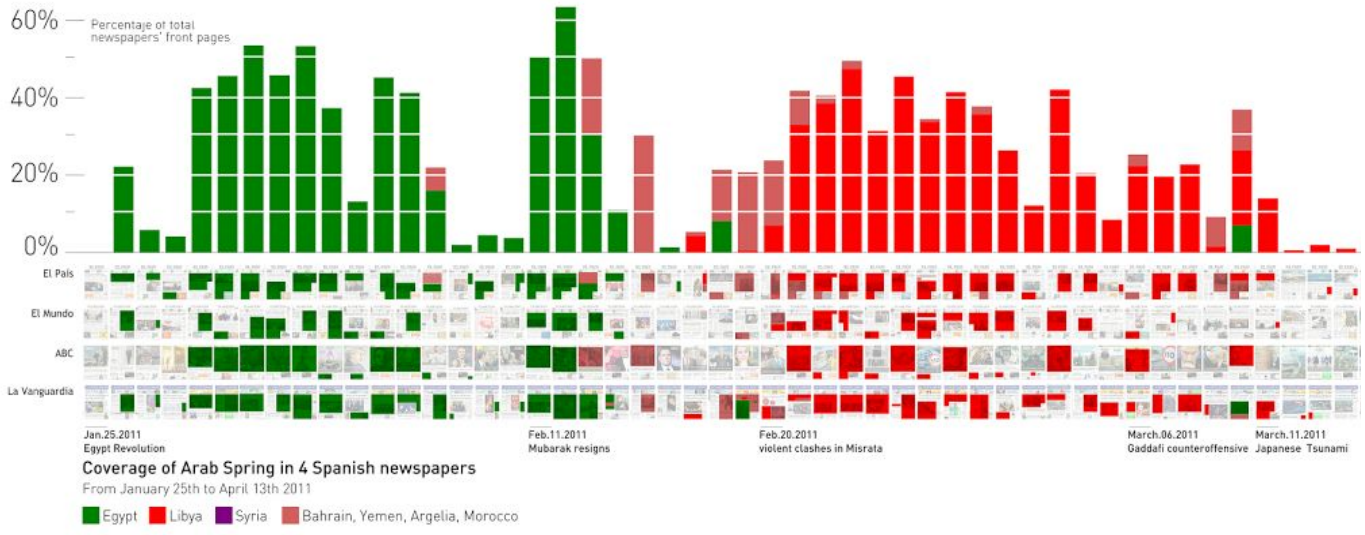
Strauss-Kahn"

http://numeroteca.org/2011/05/26/superficie-dedicada-15m-en-portadas-y-strauss-kahn/

We can see at a glance that coverage of the *Indignados* movement (#15M) was quite limited the day after the 15th of May march, and even after the police shut down the #acampadelsol encampment on the 16th. It was not until a large march on Thursday, May 19th, followed by the Spanish government's prohibition against protests, that the story of the movement was catapulted onto the front page (for further analysis of the #15M movement and its coverage in newspapers vs. social media, see section D, below). Meanwhile, the detention of Strauss-Kahn was driven almost entirely off the front page by the new unfolding story in the squares and streets of Spain.

In another example, an analysis of Spanish newspaper coverage of the Arab Spring, we can observe how closely related news stories 'fight' for a position on page one. The figure below is a visualization of coverage of the 2011 revolutions in Egypt, Syria and Lybia, made by [Author 2] for the Institute for the Study of Conflict and Humanitarian Action (*Instituto de Estudios sobre Conflictos y Acción Humanitaria*, IECAH):

Figure 8: Front Page Coverage of the Arab Spring



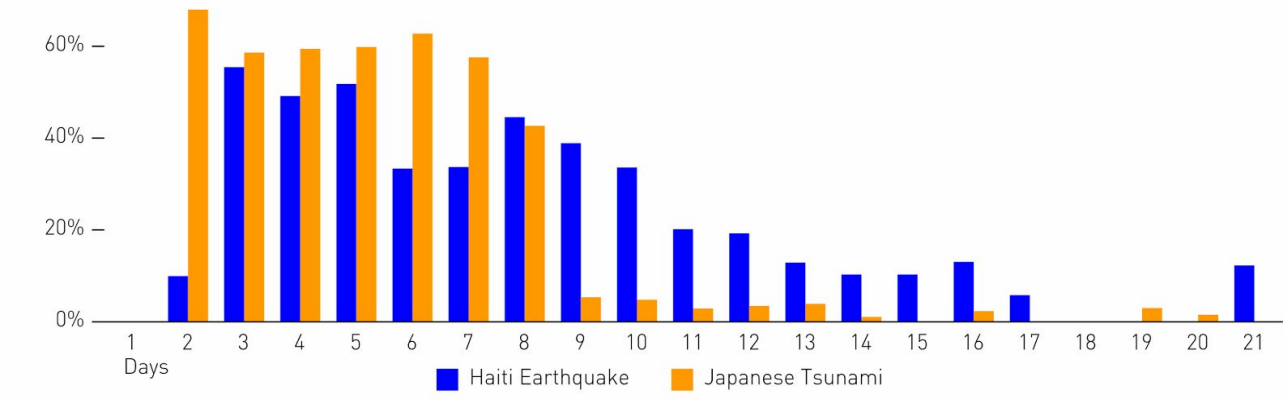
Source: "Newspaper Front Page Analysis: How Do They Tell The Story?" http://numeroteca.org/2013/06/18/newspaper-front-page-analysis-how-do-they-tell-the-story/arab-spring-spanish-newspapers-bars/

The bar chart shows the percentage of front page surface area dedicated to the 'Arab Spring' revolutions in Egypt, Lybia, Syria and other countries, in four Spanish newspapers. The original data and the coded front pages are visible as small multiples below. This visualization illustrates the periodicity of newspaper attention, which focused on one 'revolution' at a time: even as the stories took up between 20%-60% of the front pages of the Spanish press, coverage shifted from one country to the next in a one- to two-week cycle. There was rarely a simultaneous juxtaposition of news from multiple revolutionary situations; the visualization corroborates Boydstun's similar 2013 analysis of newspaper coverage of the Arab Spring. In these examples, PageOneX was used to compare the coverage of unrelated stories that unfold simultaneously, as well as to examine related stories that battle for attention on the front page.

B. Comparing the length of the news cycle for asynchronous stories

PageOneX also allows us to compare asynchronous stories (stories that unfold at different moments in time). In the example below, we compare coverage of the Japanese Tsunami (2011) and the Haitian Earthquake (2010) in four Spanish newspapers. This approach permits us to observe the news cycle for major events; in this case, to compare how long each crisis lasted in the Spanish press:

Figure 9: 2010 Haitian earthquake vs. 2011 Japanese Tsunami in four Spanish newspapers



Source: Numeroteca: http://numeroteca.org/2013/06/18/newspaper-front-page-analysis-how-do-they-tell-the-story/haiti-vs-japan/

In this visualization of coverage of the Haitian earthquake (day 1 = January 12th, 2010) and the Japanese Tsunami (day 1 = March 11th, 2011), we can easily compare the length of time that each story received front page coverage. It is possible to observe a rapid drop-off in coverage of the Japanese crisis after day 8 (03.18.2011), as well as the one- to two-week nature of front page attention in the case of these major events.[15] Again, this observation is consistent with findings by Boydstun (2013), who studied both volume and length in days of coverage of these same events, and also compared them to the far greater (and longer) coverage of Hurricane Katrina.[16]

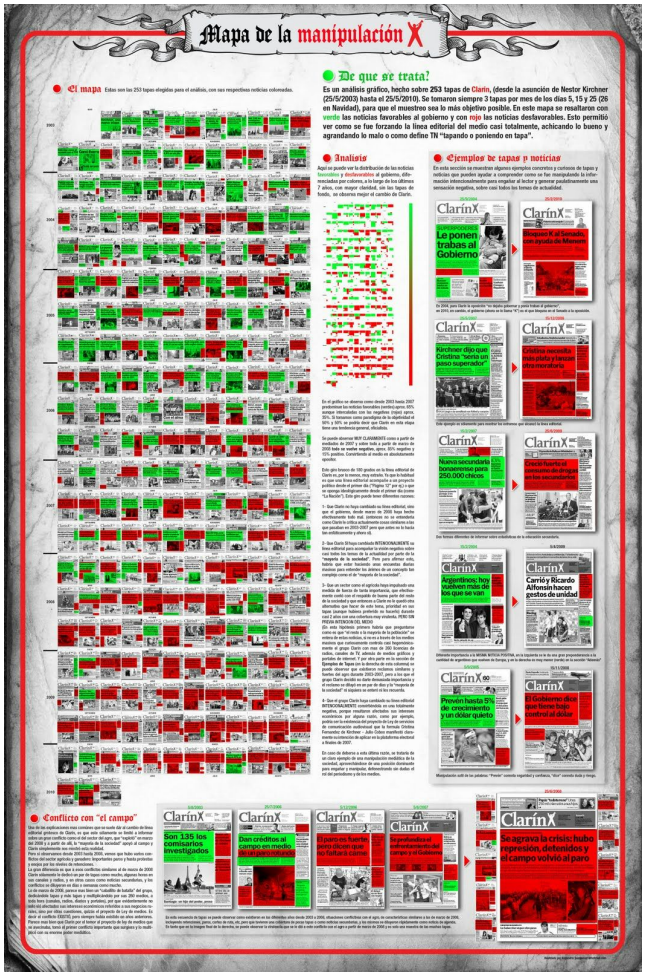
C. Qualitative analysis of news content

The sheer amount of coverage as a proportion of newspaper front pages is an important indicator of attention over time, but it tells us little about framing or other crucial aspects of mass media attention to a story. PageOneX can also be used to code and visualize front page newspaper content according to multiple user-created categories. Coders assign articles within a given thread to one of any number of categories, and each category appears as a separate color. Articles can be marked as positive, negative, or neutral in tone, by content type (politics, sports, technology, and so

[15] Closer examination of the source data shows that Tsunami coverage was almost entirely replaced by coverage of the UN's authorization of war in Libya.
[16] See http://psfaculty.ucdavis.edu/boydstun/Supplementary_Information_for_Making_the_News_files/Figure2.1.pdf.

on), by author gender, or by any categorization schema that the researcher creates. This specific approach (color coding front page article content by category) has been employed by others in the past. For example, the following front page visualization of the Argentinian newspaper Clarin, by an artist identified only as 'Alejandro,' shows articles that were favorable (green) or unfavorable (red) to the Kirchner government over an 8 year time period:

Figure 11: Positive and negative news about the Argentine government over 8 years



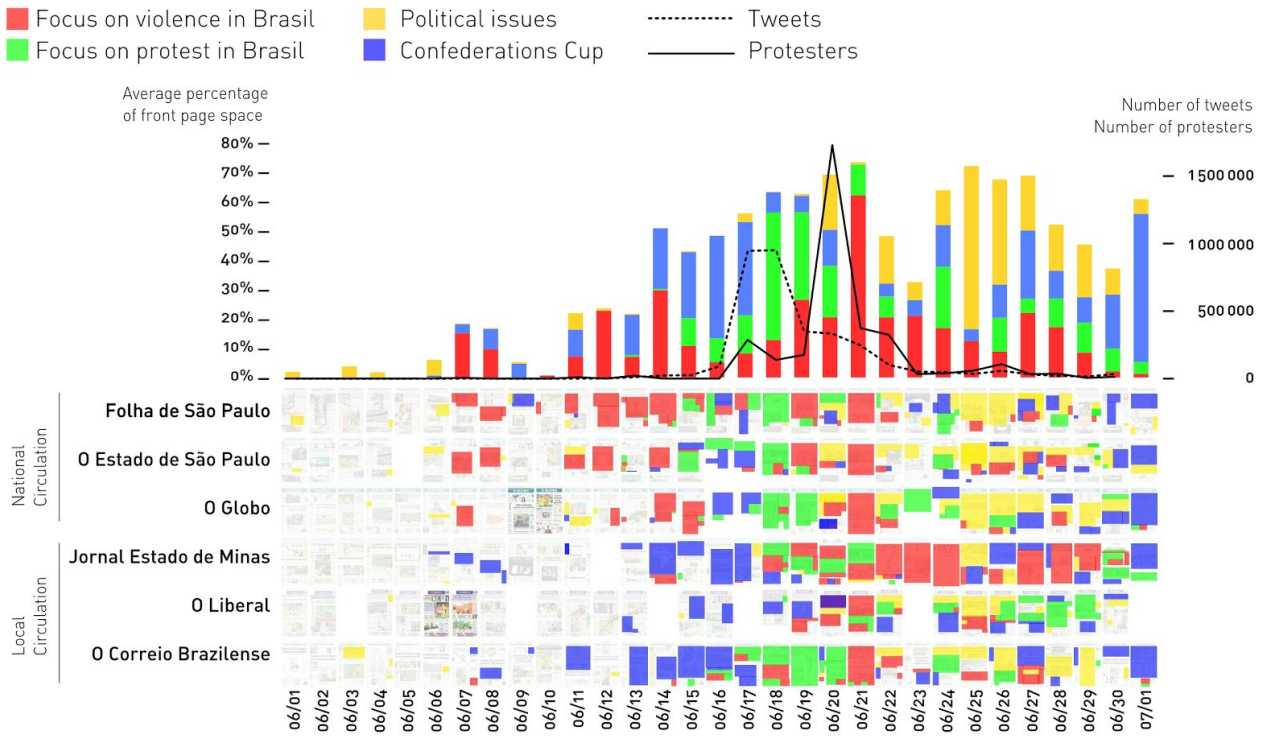
Source: Alejandro, "Graphic to uncover manipulation," zavekainfografica blog: http://zavekainfografica.blogspot.com/2011/05/grafico-para-descubrir-manipulaciones.html.

This visualization of 253 front pages (3 front pages were selected from each month) reveals the

steady shift of front page coverage from mostly positive, during the first half of the study period, to mostly negative, near the end of the Kirchner government.

In another example of frame analysis using PageOneX, Debora Leal researched front page coverage of widespread Brazilian protests in June 2013:

Figure 13: Front page coverage of Brazilian protests: "violence" vs. "protest" frames



Source: Leal, 2013: http://blog.pageonex.com/2013/08/21/protesters-or-vandals-how-the-traditional-media-approached-brazilian-protests-in-june-2013/

These protests were sparked by an increase in transportation fees, then quickly spread across the country and broadened in scope to include a wide range of political and economic demands. Leal coded the front pages of six newspapers, three national (O Estadão, A Folha de São Paulo, and O Globo) and three local (Jornal Estado de Minas, from the state of Minas Gerais, O Liberal, in Para, and O Correio Braziliense, in Brasilia). She created four categories (violence, protest, confederation cup, and political issues raised by protesters), and used these to code the main framing of the

protests in each news article. Leal draws several conclusions from her analysis. First, she argues that before June 13th, most protest coverage focused on violence; after multiple journalists were injured and arrested during the protests in São Paulo, the Brazilian mass media began to broaden their frame to include protesters' demands. Second, she notes that "After the big march on June 20th, the protests dominated the newspapers' front pages. The focus on violence was much more prevalent than the focus on the protest itself (green)." However, as she goes on to describe, coverage after June 22nd shifted: from that date forward, much of the coverage discussed policy reforms (yellow), while violence (red) or protest (green) frames tapered off in frequency. Ultimately, Dilma Rousseff, president of Brazil, responded to the protests with significant proposals for political reform (Leal, 2013).

D. Comparison with other media sources

One of the most compelling applications of PageOneX is to compare front page newspaper coverage with coverage in other media platforms. As we discussed in the introduction, communication scholars have examined the flow of news across multiple platforms, an area of research sometimes referred to as *news diffusion* or *intermedia agenda setting*, or in the context of political protest, *transmedia mobilization*. Qian (2009) examined the intermedia agenda setting effects between the New York Times, Chinese Newspapers, and Chinese blogs during the Beijing Olympics, and found that the New York Times set the agenda for Chinese Newspapers, while Chinese Newspapers set the agenda for Chinese blogs. Lopez-Escobar et. al. (1998) set a high bar for intermedia agenda-setting research, with a cross-lagged correlation analysis of newspapers, TV news, and political advertising during the 1995 Spanish elections. They found that political advertising in newspapers had an influence on both TV and newspaper news agendas, and that TV news influenced TV political advertising. In this section, we summarize work by media researchers who have used PageOneX to compare front page newspaper coverage of mass protests with social

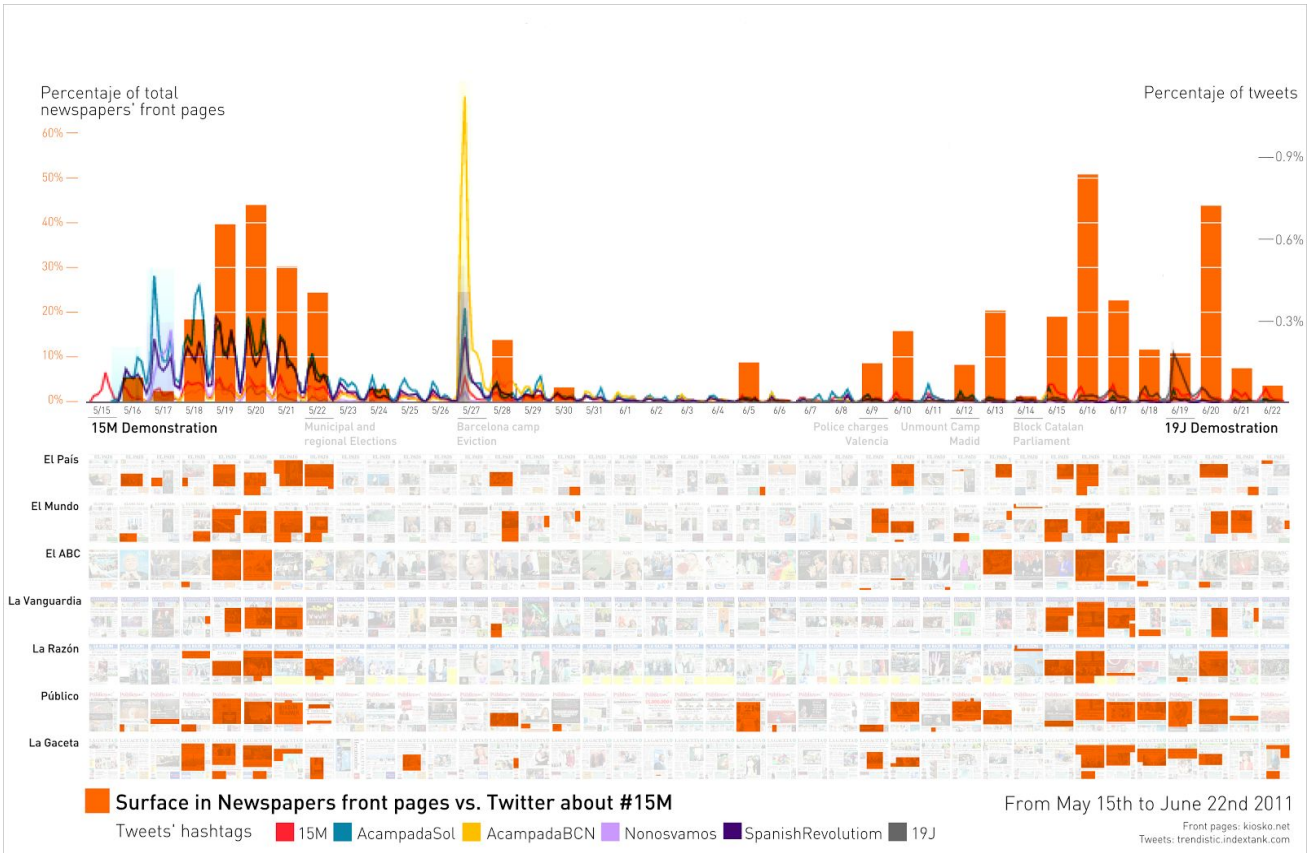
media attention to those same protests.

For example, we compared Tweets[17] to front page newspaper coverage of the Spanish *Indignados* movement. The movement is also known as 15M, based on the date of the first mobilization: the 15th of May, 2011, when about 20,000 people marched in Madrid and 15,000 in Barcelona, among other locations (Castells, 2013). The most popular hashtags used by the movement during its early days were #acampadasol, #15M, and #spanishrevolution, according to research by the Institute for Biocomputation and Physics of Complex Systems (BIFI) at the University of Zaragoza (Borge-Holthoefer, et al., 2011; and see http://15m.bifi.es/index_en.php). We compared a sample of tweets[18] using these hashtags with the front pages from 7 major newspapers in Spain:

Figure 14: 15M movement in Tweets vs. Newspapers

[17] Most cross-platform studies using PageOneX have focused in particular on the relationship between Twitter and newspaper front pages. Researchers frequently use Twitter to study social media information flows, because Tweets are (relatively) open and available for study. Although Facebook is more important to the spread of news, based on its far higher penetration rate, it is a closed network that is more difficult for outside researchers to study.

[18] We did not have access to the 'Firehose' or the full corpus of tweets; instead, we used twitter data from Trendistic, an online Twitter data visualization provider.



Source: Numeroteca, Surface dedicated to the 15M movement in 7 Spanish newspapers. http://numeroteca.org/2011/11/02/surface-dedicated-in-newspapers-front-pages-vs-twitter-about-15m-with-acampadabcn-15m-19j-acampadabcn/

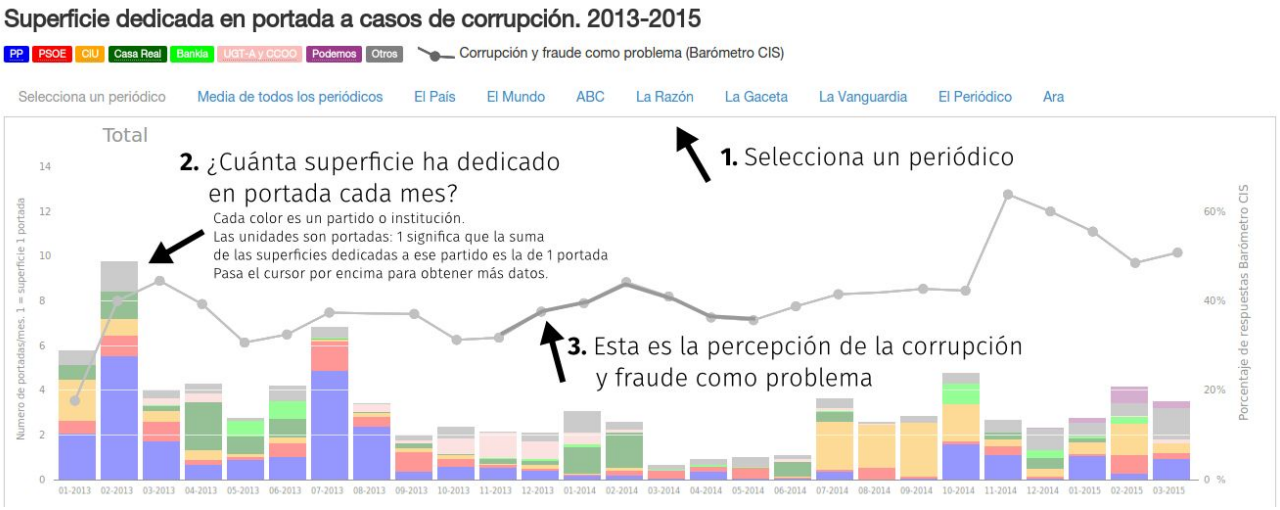
Above the image of coded newspaper front pages is a line chart that displays the volume of 15M tweets that used the movement's top hashtags, from May 15th through June 22nd, 2011. We can see that, in the early stages of the movement, strong twitter activity precedes newspaper coverage by several days. Additionally, the largest spike of twitter activity (on May 27th) is followed by a relatively small amount of of front page coverage. Also interesting is the resurgence of front page coverage near the end of the study period, which does not seem to be matched by a concomitant spike in related tweets, although the most likely explanation is that the movement's most popular hashtags had changed by this point. Our point here is not to draw definitive conclusions about the relationship between tweet volume and front page newspaper coverage, but to demonstrate the possible utility of PageOneX as one tool among many for analysis and visualization of media

attention, including for cross-platform analysis.

E. Comparison of multiple newspapers' coverage of the same story

Finally, PageOneX can also be used to compare coverage of the same ongoing story across multiple newspapers. For example, the ways that amount, duration, framing, and other aspects of coverage may differ between the front pages of various papers can be analyzed, visualized, and seen at a glance. One excellent example of this technique is the study "The Color of Corruption," detailing coverage of endemic corruption across all major Spanish political parties, as reported in the front pages of the largest Spanish newspapers over a period of 30 months:

Figure 15: The Color of Corruption



Source: http://numeroteca.org/colorcorrupcion.

This interactive data browser allows the user to explore the ongoing coverage of political corruption in Spain, as covered by the country's major newspapers (El Pais, El Mundo, ABC, La Razón, La Gaceta, La Vanguardia, El Periódico, and Ara). [Author 2] coded 30 months of these papers' front pages for coverage of political corruption, categorized by the political party that the story indicts. The result is a visualization that reveals not only the volume of total coverage of the scandals, and

volume of coverage in comparison between papers, but also makes visible which papers dedicate more front page coverage to corruption in one political party vs. another. The user can interact with the visualization by selecting a particular newspaper, or by rolling over the color bars to reveal the underlying data.[19] There are many more examples of PageOneX visualizations that use the tool to analyze various aspects of newspaper front page coverage, but in the interests of space we will not discuss them here. We hope that we have provided an interesting overview of some of the possibilities of the tool, and we invite the reader to explore it further at http://pageonex.com.

Conclusions, Limitations, and Next Steps

The PageOneX platform semi-automates a particular method for analyzing and visualizing newspaper front pages. The tool provides a straightforward, user-friendly interface to analyze and visualize news attention over time, as well as across newspapers. It is at the moment a simple tool designed entirely around newspaper front pages, which remain one key component of a much broader and increasingly complex media ecology. In the future, we hope to develop additional functionality to allow comparison between newspaper front pages and other forms of media attention, including the full text of print newspapers, the online editions of newspapers, TV and Radio broadcast transcripts, and various social media platforms.

Current Limitations of PageOneX

The platform currently has significant limitations for systematic media research. For example, the user is currently limited to 500 front pages per thread, although there is a viable workaround.[20] PageOneX does not currently incorporate coder training or automated measures of

[19] See also a PageOneX visualization of 18 months of coverage of Syria http://www.iecah.org/web/visual/syria.htm, where bar chart color indicates the number of newspapers reporting the story on page one on any given day.
[20] We recommend 1 month threads if you are using multiple newspapers. [Author 2] developed a script to download and join all the data from multiple threads and visualize it in a single bar: https://github.com/numeroteca/colorcorrupcion

intercoder reliability. Story selection and categorization can be problematic, and can differ between users if they do not receive proper training and a codebook. In other words, to avoid miscategorization, and for credible, replicable results, the creator of a thread must set clear guidelines for coders and calculate intercoder reliability. Additionally, we would like to implement certain features to improve basic content coding, including multiple categorization of the same article. For example, in the case of the coverage of protests in Brazil, a coder should be able to mark a single article as employing both a 'violence' frame and a 'political issues' frame. Currently, users can create and name as many simultaneous categories as they need, but each area of a page can only be coded as a single category.

In the future, we also imagine extending PageOneX to incorporate tools for cross-platform comparison. We have done this to some extent with Twitter, by using existing third party services to analyze tweet volume over time, then visually comparing the results with newspaper coverage of the same topic. We would like to add statistical analysis of these relationships; for example, is it possible to predict a certain percentage of front page coverage based on tweet volume, or vice versa? Tools to compare front page coverage with video or audio news broadcasts would also be very powerful; for example, we might take advantage of news organization APIs that increasingly provide access to transcripts and/or recordings of TV and radio news to code and visualize the amount of time they devote to a particular thread. Our current limited social media comparison (tweet volume) might be expanded to include mentions and retweets, and to incorporate analytics for additional platforms (in particular, Facebook). Some researchers have already used PageOneX alongside other tools to produce comprehensive cross-platform analysis of media attention; for example, Graeff, Stempeck, and Zuckerman (2014) analyzed the evolution of the Trayvon Martin story by incorporating PageOneX alongside tools including Media Cloud, Google Search query analysis, Change.org petition signatures, bit.ly clicks, and more. Analysis across media platforms may be done either by building new functionality within PageOneX, by making it interoperable with

other F/LOSS projects, and/or by using APIs from other platforms, such as Archive.org's archive of TV news (see http://archive.org/details/tv).

Ideally, certain aspects of the PageOneX coding process that are currently manual would be improved, automated, or semi-automated. Most crucially, the current implementation requires human coders to read and code front page articles based on low-resolution scans available from Kiosko.net.[21] This is severely limiting, primarily since it means that no text-based search of newspaper front page content is currently possible within PageOneX. There are several possible paths to solve this issue: Optical Character Recognition (OCR) could be used to extract searchable text transcripts from each front page image; however, given the current state of OCR and the low resolution of the front page images, this approach would not likely produce usable results. There is, however, promising research into the automatic identification of text blocks and layout structure in PDFs of newspaper front pages, such as the Xed project.[22] Another possibility would be to find an existing database that provides either higher resolution images or PDFs containing the full searchable text of newspaper front pages. A third path would involve making calls from PageOneX to a third party database that contains the full text of newspaper articles, as well as metadata about what page the article appears on. In this way, it should be possible to perform searches for key topic terms within a newspaper or set of newspapers, then return only the images of relevant front pages (those where the search term appears in a story on page 1A) for researchers to code. With access to full text search, the power of the platform could be greatly extended, for example through the use of natural language processing and machine learning (NLP/ML) to automatically locate and code front page articles within a thread or semantic category.

Content analysis, and newspaper content analysis by amount of coverage, has a long history in the field of communication studies. New tools can enable new approaches to this kind of analysis.

[21] Several researchers have also requested the ability to upload their own newspaper images for coding.
[22] See Hadjar, Rigamonti, Lalanne, and Ingold, 2004, for more on Xed.

Some aspects of newspaper content analysis can be automated; some can be distributed or even crowdsourced, and some can be done much more quickly by human coders using web based interfaces. The data produced by such analysis can also be visualized in new ways, and compared against data from other information sources, such as social media. In this paper, we have described previous approaches, explained the approach we have taken with PageOneX, demonstrated the application of the tool to several kinds of content analysis, and suggested areas for further work. We hope that PageOneX becomes a valuable resource for communication scholars, and we invite interested researchers to contact us and to participate in the future development of the platform.

References

[Author 1, 2012]

[Author 1, 2014]

[Author 2, 2013]

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*.

BIFI (2011). *15M Detailed List.* Twitter dataset from the Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza. http://15m.bifi.es/tags_full.php.

Borge-Holthoefer J., Rivero A., García I., Cauhé E., Ferrer A., Ferrer D., et al. (2011). Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study. *PLoS ONE* 6(8): e23883. doi:10.1371/journal.pone.0023883

Boydstun, A. E. (2013). *Making the news: Politics, the media, and agenda setting*. University of Chicago Press. Codebook and dataset available in http://psfaculty.ucdavis.edu/boydstun/Supplementary_Information_for_Making_the_News.html

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

Breed, W. (1955). Newspaper 'opinion leaders' and processes of standardization. *Journalism & Mass Communication Quarterly*, *32*(3), 277-328.

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

Castells, M. (2013). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley & Sons.

Chaqués-Bonafont, L., Palau, A., & Baumgartner, F. (2014). Policy Dynamics in Spain. *Hampshire, UK: Palgrave Macmillan. Forthcoming*.

Clark, A.H. (1930). Press Notice of the Des Moines Meeting in Iowa. Science, New Series, 71(1834), 218.

Costanza-Chock, S., Schweidler, C., and DeTar, C. (2013). "The Rise and Fall (and Rise?) Of Occupy Wall Street: 5 Lessons for Networked Movements." *La Vanguardia Dossier: The Power of the Networks.* http://bit.ly/ows-rise-fall-rise.

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.

Donovan, H. L. (1926). A*n Analysis of the Use of Column Inches in School Papers.* Educational Research Bulletin, 5(14), 292-293.

Economist (2011). The future of news: Back to the coffee house. Retrieved March 10, 2013, from http://www.economist.com/node/18928416.

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.

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

Gerbaudo, P. (2012). *Tweets and the streets: Social media and contemporary activism*. Pluto 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.

Graeff, E., Stempeck, M., and Zuckerman, E. (2014). The battle for 'Trayvon Martin': Mapping a media controversy online and off–line. *First Monday*, 19 (2-3). http://firstmonday.org/ojs/index.php/fm/article/view/4947/3821.

Hadjar, K., Rigamonti, M., Lalanne, D., & Ingold, R. (2004). Xed: a new tool for extracting hidden structures from electronic documents. *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop on*. IEEE.

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.

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

Kavada, A. (2015). Creating the collective: social media, the Occupy Movement and its constitution as a collective actor. *Information, Communication & Society*, (ahead-of-print), 1-15.

Kidd, D. (2014). Social Justice Media: the Case of Occupy. In *Mediaciones: Revista Academica de Comunicación.* Pre-publication Version, October 30, 2014, available at

http://mediaalliance2.live.radicaldesigns.org/downloads/Kidd_Social_Justice_Media-_The_Case_of_Occupy-libre.pdf.

Leal, D. (2013). Protesters or Vandals? How the traditional media approached Brazilian protests in June 2013. PageOneX. http://blog.pageonex.com/2013/08/21/protesters-or-vandals-how-the-traditional-media-approached-brazilian-protests-in-june-2013.

Lopez-Escobar, E., Llamas, J. P., McCombs, M., & Lennon, F. R. (1998). Two levels of agenda setting among advertising and news in the 1995 Spanish elections. *Political Communication*, *15*(2), 225-238.

Lynch, S., & Peer, L. (2002). Analyzing Newspaper Content. *A How-To Guide, Readership Institute, Media Management Center at Northwestern University, Chicago*. http://www.readership.org/content/content_analysis/data/how-to.pdf

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. (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.

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.

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

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

Tella, R. D., & Franceschelli, I. (2009). *Government advertising and media coverage of corruption scandals* (No. w15402). National Bureau of Economic Research.

Tufte, E. R. (2001). *The visual display of quantitative information.* Cheshire, Conn. : Graphics Press.

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

Wolfe, M., Boydstun, A. E., & Baumgartner, F. R. (2009). Comparing the Topics of Front-Page and Full-Paper Stories in the New York Times. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, April 2-5, 2009. Available online at www.unc.edu/~fbaum/papers/Wolfe_Boydstun_Baumgartner_MPSA09.pdf.

Qian, X. (2009). Inter-media agenda-setting effects among Chinese newspapers, Chinese blogs and

the New York Times during coverage of the Beijing 2008 Olympic Games. *Graduate Theses and Dissertations.* Paper 11083. http://lib.dr.iastate.edu/etd/11083.

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