

“Sex, Scandals, and Celebrities”?

Exploring the Determinants of Popularity in Online News

ANGÈLE CHRISTIN¹

Fernand Braudel IFR Postdoctoral Fellow
Fondation de la Maison des Sciences de l'Homme
Centre Maurice Halbwachs (ENS-EHESS)
Data & Society Research Institute
angele.christin@gmail.com



The internet is transforming journalism in many ways. The circulation of newspapers and magazines has declined dramatically, as have advertising revenues, in part due to online competition. The development of news websites has been accompanied by many innovations in journalistic writing and formats, such as more casual writing styles, shorter blog posts, interactions with online readers, participatory journalism, and multimedia productions that weave together text, images, videos, and hyperlinks (Singer et al., 2011; Dagiral and Parasie, 2010; Anderson, 2013; Usher, 2014; Boczkowski, 2010, 2005; Deuze, 2007).

Less visible than these changes, but perhaps more profound in terms of differentiating print and online news, is the increasing importance of internet metrics in web newsrooms: web journalists and editors are now confronted with a constant flux of quantitative data about the online success of their articles (Anderson, 2011; Usher, 2013; Christin, 2014; Napoli, 2011; MacGregor, 2007). A growing number of analytics software programs (Google Analytics, Omniture, Chartbeat, Visual Revenue, etc.) provide detailed and often real-time statistics about the behavior of online readers. These metrics typically include the number of visits, “likes” on Facebook, “tweets” on Twitter, comments, and number of returning users. They also track the average time spent per reader on the page, the scrolling patterns

Pour citer cet article

Référence électronique

Angèle Christin, « “Sex, Scandals, and Celebrities”? Exploring the Determinants of Popularity in Online News », *Sur le journalisme, About journalism, Sobre jornalismo* [En ligne], Vol 4, n°2 - 2015, mis en ligne le 15 novembre 2015.

URL : <http://surlejournalisme.com/rev>

of readers, and additional information about how readers accessed the articles.

In most web newsrooms, journalists and editors have open access to these analytics programs, which are usually installed on every computer in the office. Journalists often observe and compare the number of page views for their articles, make predictions about which of their pieces will be popular online, and express surprise or disappointment when the actual numbers do not match their expectations. Journalists now have a distinct representation of their readers based on direct quantitative feedback. This “*quantified audience*” (Anderson, 2011) is a major change compared to the “*imagined audience*” of print newspapers, where journalists used to have rather abstract representations of their audience and did not know which articles were the most popular among their readers (Gans, 1979; Schlesinger, 1978; Darnton, 1975).

Based on this quantitative information, how do journalists make sense of their readers’ preferences? Reactions are mixed. In online news, like in many other cases of cultural production, “*all hits are flukes*” (Bielby and Bielby, 1994): journalists claim that it is impossible to predict which articles will “*go viral*.” However, journalists also draw on their observations of traffic numbers to make statements about the kinds of articles that tend to be popular online. Specifically, journalists assert that readers usually click more on stories involving sex, celebrities, crime, and practical advice pieces. In contrast, prestigious sections covering international news, politics, and the arts are said to perform poorly in terms of page views.

Yet it is an open question whether web journalists and editors are correct in their representations of their readers’ preferences. In this article, I examine the determinants of online popularity by drawing on mixed methods, using qualitative interviews and ethnographic fieldwork to complement a statistical analysis of the content published on a news website. With the assistance of a computer scientist, I collected all of the articles posted between May 2007 and September 2012 (13,159 articles) on a stand-alone French news site that I call *LaPlace.com* (the name of the publication has been changed), where I interviewed 34 journalists and conducted ethnographic fieldwork between 2011 and 2013.²

The data set gathered on *LaPlace*’s website features many variables of interest for each article: the number of visits, “tweets” on Twitter, “likes” on Facebook, comments left by readers, the headline and key-words of each article, the time of publication, number of words, and number of quotes. It

also includes several variables about the authors (gender, employment status, and Twitter presence). These variables allow me to examine the determinants of online popularity over time at the French website and compare these findings with the journalists’ representations as gathered in my interviews and observations.

This research design is innovative in several ways. First, the statistical analysis takes advantage of the new opportunities created by the internet and the rise of “Big Data” available online (Boyd and Crawford, 2012; Lewis, Zamith, and Hermida, 2013). Second, the comparison of the qualitative material (e.g., the journalists’ representations of their readers’ preferences) and the statistical findings (the determinants of online popularity) affords a more nuanced reading of the ethnographic data (Gramain and Weber, 2001). If web journalists’ representations diverge from the actual preferences of their readers, what does it indicate about the professional culture of news sites and the worldview of web journalists?

This article proceeds as follows: first, I introduce *LaPlace*, the French publication under consideration, and describe the representations of journalists and editors about internet traffic. Second, I introduce the data set and the variables used in the statistical analysis. Third, I summarize my quantitative findings and add important qualifiers to the representations put forward by *LaPlace*’s journalists. Lastly, I discuss the implication of these results for online readership and newsroom cultures.

HOW *LAPLACE*’S JOURNALISTS UNDERSTAND THEIR READERS’ PREFERENCES

This section explores how journalists and editors at *LaPlace* understand traffic numbers and interpret their readers’ behavior. I describe two central ideas prevalent among web journalists: the “sex, scandals, and celebrities” hypothesis and the idea that “*commenters are trolls*.” First, however, I introduce the publication under consideration.

***LaPlace* between participation and politics**

LaPlace was created in 2007 in Paris by several journalists formerly employed by a prestigious, left-leaning national newspaper. The founders provided most of the initial capital but relied from the beginning on several external investors.

LaPlace’s editorial line developed over the years around two central pillars, a participatory approach

and a left-leaning political stance. *LaPlace*'s participatory approach was described by its founders as providing "information with several voices": "the journalists, the experts, and the internet users." According to the website's FAQ page, "*LaPlace relies on the collaborative work of three voices: the journalistic staff, which consists of experienced journalists and young reporters (...); a circle of specialists, experts, and eyewitnesses who shed light on the news; and internet users themselves, who participate in the life of LaPlace with their comments but also with their articles, links, photos, and videos.*" Editors describe blogs and "testimonies" ("témoignages," first-person pieces written by internet users) as an important part of the editorial project. In accord with its liberal and left-leaning orientation, the website became well known for several scoops implicating the former president of France, the conservative Nicolas Sarkozy, and his party the Union pour un Mouvement Populaire (UMP). The articles published on *LaPlace* typically rely on an informal, witty writing style, with headlines featuring puns or provocative formulations.

LaPlace can still be described as a "niche" website, in the sense that it does not cover all the "beats" and subjects that a generalist website such as *LeMonde.fr* or *LeFigaro.fr* might provide. For example, *LaPlace* does not cover local news, the weather, lifestyle, or travel in a systematic way. Instead, it primarily covers politics, news, culture, technology, and social issues. Though the website's editors emphasize the news-oriented aspect of the publication, it is safe to say that *LaPlace* is closer to a web magazine than an online newspaper, even if this distinction has become less relevant on the web compared to the print world (Schudson, 2011).

In 2011, the website had two million euros of revenue, mostly from advertising, but it was still not profitable. The founders thus decided to sell their shares to a larger media group. The media group and the founders of *LaPlace* issued a statement reaffirming the editorial independence of the website. In 2013, *LaPlace* had 25 employees, including 15 journalists, and attracted slightly more than two million unique visitors each month.

I first studied the news production process at *LaPlace* using ethnographic methods: regular observations of the office were complemented by semi-structured interviews with current or past employees (Beaud and Weber, 2010). I spent between one and three days per week at the newsroom between December 2011 and March 2012, and again between March and June 2013, as part of a larger study of six web newsrooms in Paris and New York. During my days of observation, I sat at the different desks with

the journalists, shadowed them during their daily work activities, and attended editorial meetings. In parallel, I conducted a total of 34 recorded interviews with staff journalists, editors, freelancers, and bloggers.³ In addition, I followed the online activity of *LaPlace.com* intensively, reading the articles, exploring the archives of the website, and analyzing the daily exchanges between journalists on Twitter. *LaPlace*'s managers also gave me access to one of the software programs providing real-time statistics on the traffic of the website.

Sex, scandals, and celebrities

What kinds of articles are likely to attract large numbers of readers? *LaPlace*'s journalists and editors have strong views on this question: they all agree that most articles about sex and celebrities are particularly popular among their readers, whereas pieces on world news, politics, and culture do not attract as many readers, as indicated by the following quotes from my interviews.

Philippe (editor-at-large): "*If we only cared about the number of clicks, we would only write about celebrities. Celebrities... it always works. (...) But if we only do Syria, Africa, and... say, the socialist party's plan on housing policy, we won't get any readers. Syria is a good example... People are being butchered every single day and no one cares.*"

Marina (editor-in-chief): "*Our readers couldn't care less about Syria. A headline on Syria will only get 4,000, 5,000 clicks, which is not a lot. (...) So we have to check what's working on Twitter, Chartbeat, Facebook, and if we see that traffic is all flat, we post an article, a scandal, something about sex, or something about corruption... because we know that we need to hook the readers in...*"

Anne (staffer): "*Well, we know that articles about sex and 'testimonies' will get a lot of clicks. Sometimes it's more complicated... Some articles, though, it's easy to predict... For example the story about the cops who posted comments on Facebook about the teenagers that they were frisking, well, in the headline you have 'cops', 'Facebook', and 'arrests,' obviously it's clickable. The words just catch the eye... But if you have an article about senators, embezzlement, and dual mandates... It's probably not going to work as well.*"

Agnès (staffer): "*In the culture section, one has to become humble. When you write about*

graphic novels, you know that you won't get a lot of page views."

These representations were confirmed during my days of observation: journalists kept commenting on the articles and made predictions about which ones were likely to be "one-million click articles" or "flops." This was particularly clear during editorial meetings: journalists often jokingly threatened to write articles featuring "clickbait" headlines about sex or to include pictures of cute kittens in the first paragraph, both options being described as sure ways to attract more traffic.

Several ideas emerge from these observations and interviews. First, journalists and editors at *LaPlace* seem to agree that articles about sex, scandals, celebrities, and pictures of animals are likely to be popular online. Journalists and editors also claim that articles on world news and culture do not attract many readers. They often use a cynical tone when describing the preferences of their readers, not unlike the journalists studied by Gans in the 1960s, who lamented the fact that their audience did not have the same news-reading habits as they did (Gans, 1979).

Articles about "scandals" form another category that allegedly attracts many readers on *LaPlace*. According to journalists, "scandals" are actions by public figures that can be regarded as legally or morally questionable. Some examples of notable incidents of this kind during the period under consideration included Cecilia Sarkozy (the former wife of President Sarkozy) casting a blank vote in the 2007 elections or Dominique Strauss-Kahn's arrest in New York. This category of "scandals" also broadly includes the corruption cases mentioned by Marina. All these stories are likely to be successful online, according to *LaPlace*'s editors and journalists.

In addition, staffers assert that reader-submitted first-person narratives, or "testimonies" (*témoignages*), also tend to be popular. "Testimonies" is a format which is specific to *LaPlace*: readers send stories to the website about their life or a specific experience that they would like to share with a larger audience. Testimonies are similar to blogs posts in the sense that in both cases the authors are not professional journalists. Yet there is an important difference between the two forms of writing: bloggers repeatedly post on their blogs and tend to specialize on a given topic, whereas testimonies are usually a one-time collaboration.

Last, journalists often explain that the format of articles also matters: short articles or posts are more likely to attract a larger audience, whereas

longer articles tend to be less popular. Of course, online readers do not know whether the article is going to be short or long before they click on the headline. Yet, according to *LaPlace*'s journalists, the higher popularity of short articles might be due to the growing importance of social media and mobile audiences in web traffic. Readers "share," "like," or "tweet" about articles on Facebook and Twitter, thus sending new groups of readers to the articles, who in turn are also more likely to share it with their friends. This process of online referral on social media is said to work better for short, "fun" pieces than for long, analytic articles.

Distinguishing between Facebook users, Twitter users, and commenters

More generally, *LaPlace*'s journalists establish strong distinctions between the categories of readers, most significantly between Facebook users who "like" articles, Twitter users who "tweet" about articles, and online readers who write comments at the bottom of articles. According to the journalists, these different categories of readers are attracted by different kinds of articles. Journalists claim that commenters are a very specific group of readers. Gael – one of *LaPlace*'s editors who is also the community manager – gives us his understanding of the distinction between different categories of readers:

Gael (managing editor): *"The profiles of the readers are becoming more and more diverse. We have the internet users who read our articles on Facebook... It's a particular profile. They read our paper through their friends' links, or they like our page on Facebook, but they never go to the website directly... They're younger, 25 or 30 years old, whereas our average reader is more like 30-35 years old. But our returning readers, who are usually the ones who post comments all the time, every day or almost every day, they hate Facebook because of its use of private data, because of its ideology... These readers are very important for our identity, they've been there from the start. Then there are some readers, many of them actually, who go on the main page and maybe check out a couple of articles, but don't read the comments, and they don't post anything."*

It is possible to distinguish between three broad categories of readers. The first group is the mass of online readers who access the website's content through Facebook. These readers usually browse through a couple of articles, do not post comments, and sometimes share the articles with their friends and followers. Readers on Facebook are said to be

younger, less interested in political stories, and more attracted by “fun,” “weird,” or unusual stories with pictures that they can “like” and “share” with their friends. The second group consists of the readers who access the articles and comment on the content on Twitter. These readers are said to be more educated on average than on Facebook; they are more interested in stories about the media, politics, and the internet. Journalists often say that most of the people who “tweet” about their articles are either other journalists or people who are related to the journalistic sphere broadly considered. The third and final group is the commenters, about whom *LaPlace*’s journalists have complex feelings. On the one hand, they emphasize the value of the feedback and information provided by these regular readers. On the other hand, they often criticize the behavior of a small group of commenters, pointing out that their frequent comments are negative and offensive:

Louise (staffer): *“The commenters... It’s a gold mine, all the information they’re giving us... But we don’t really use it. Also, some lunatics are posting insults every day! They’re insulting the people we interview and they’re threatening us... It’s not super interesting (laughs).”*

Alexandre (staffer): *“The comments are usually annoying... There is a lot of trolling involved, people who come to mess around, who are very negative... It’s depressing. Anyway, I usually don’t read the comments. When they ask actual questions, I answer.”*

Gael (assistant-editor): *“The commenters... They’re very critical, very politicized, often offensive... It’s true that they often drag the discussion down...”*

Web journalists often rely on the metaphor of “trolling” to characterize regular commenters. In the internet repertoire, a “troll” is someone who creates discord by posting inflammatory messages and upsetting people online. For example, a recurring commenter on the website took the pseudonym “Gutter” (“*Egoût*”): most of his/her comments harshly criticize the political and editorial line of the website, in addition to insulting each journalist on a daily basis. According to the staffers, commenters are particularly likely to post comments about topics that are political or controversial (such as racism, sexism, religion, the Middle East, and immigration). This adds nuance to the idea that articles about politics are not popular on *LaPlace*’s website. It indicates that the concept of “popularity” needs to be further unpacked in order to take into account the different modes of access and expression of online readers.

Before turning to the presentation of the data set, it is important to mention that the representations of *LaPlace*’s journalists are consistent with the existing academic literature on readership and interactivity in online news, in two main ways. First, as recent research indicates, the popularity of “sex, scandals, and celebrities” appears to be a constant in online news, regardless of the sites and countries involved. In their book *The News Gap* (2013), Boczkowski and Mitchelstein draw on a content analysis of about 40,000 articles in eight online publications in four different countries and find that readers prefer stories about sex, celebrities, crime, and practical matters – what they call “*non-public interest stories*.” Second, scholars have repeatedly shown that, despite web journalists’ emphasis on breaking the unidirectional flow of information from the newsroom to the readers, newsroom routines are “sticky”: web journalists still frequently express their indifference and sometimes their resistance to including lay internet users in the news-making process (Boczkowski, 2005; Domingo, 2008). In this perspective, readers’ participation would operate as a “myth” detached from actual practices in most web newsrooms (but see Nip, 2006; Braun and Gillespie, 2011; Lewis, 2012; Canu and Datchary, 2010).

These two features seem to apply to the case of *LaPlace*, yet a note of caution is needed before comparing *LaPlace* to other news sites (especially to generalist legacy media sites), since the French site has a highly specific editorial line and a “niche” position in the French market for online news. For this reason, our statistical findings should be compared with caution to the results provided by existing studies of online readership and participation.

**DATA AND METHOD: A STATISTICAL ANALYSIS
OF LAPLACE’S CONTENT, 2007-2012**

In collaboration with a computer scientist, I devised a web crawl (an automated computer program) in order to gather all the articles published on *LaPlace*’s main page between May 2007 (when the website was created) and September 2012. This process gathered a total of 13,159 articles. The crawling process collected all the articles published in the central column of the website (nicknamed “the river” [*le fleuve*] by the journalists). Focusing on this section of the website automatically excluded two types of articles from the analysis. First, articles that were never published on the homepage are not included in our sample. During my observations, I observed very few instances where the editors vetoed publishing an article on the homepage. These cases typically involved blog posts written by blog-

gers or internet users (rather than articles written by staff members or freelance journalists) which were deemed to be of poor quality and unfit to be posted on the homepage. Second, the sample does not include the short rewrites of news wires and external web content that are published in the section called the “lookout,” a column located on the left side of the homepage. In the brief posts published in this section, staffers ‘signal’ articles published on other websites and provide a short paragraph summarizing the information. Articles that elaborate more lengthy analyses of external content are published in the “river” section of the website and thus are included in the sample.

Main variables and descriptive statistics

For each of the 13,159 articles published in the “river,” we collected several variables of interest. First, we gathered information on the characteristics of the articles themselves: the time of publication (minute, hour, day, month, and year); the number of words, quotes, images, and hyperlinks present in the text; and the “tags” of the article (key-words describing the topic and format of the article). In addition, drawing on the information provided on the website, we gathered variables about the articles’ online popularity: the number of visitors (number of people who clicked on the article), number of times the article was mentioned on Twitter, the number of times the article was “liked” on Facebook, and the number of comments left by readers. Finally, we included information about the author(s) of the article, such as the author’s name, the number of authors, the author’s status (staffer, freelancer, or blogger), and the gender of the author.

These last two variables were documented using a two-step process. First, the computer scientist extracted the information about the gender of the author and his or her status based on the biographical note provided for each writer on the website. These biographical notes are created by the editors and include the name of the authors, their gender, and their occupational status, e.g., whether they are “journalists” or not. Thus, the original categorization of the author as a journalist is decided by *LaPlace*’s staff members. Yet this first coding process left us with many missing variables. To remedy this, I manually checked and recoded the “employment status” and “gender” variables by searching online for the names of each of the 2,169 authors included in the data set. I used the following criteria in order to code an author as a “journalist” when the information is missing. To be categorized as journalists, authors either have to define themselves as journalists on Twitter or on their personal websites, or (when I could not find a Twitter account or personal website) must

have worked for at least one news organization. I created an additional category among the “journalists” by distinguishing the “staffers” affiliated with *LaPlace* – people working in the newsroom, including tenured employees (“*Contrats à durée indéterminée*”), untenured employees (“*Contrats à durée déterminée*”), and interns (“*stagiaires*”) – from the “freelancers” (“*pigistes*”) who write for *LaPlace* on a contractual basis.

Table 1. Descriptive statistics about the characteristics of the articles

Descriptive statistics		N	%
Articles published	Total	13,159	100
Day:	Weekday	9,970	76
	Weekend	3,189	34
Time of day:	Morning	2,910	22
	Afternoon	6,062	46
	Evening	3,835	29
	Night	352	3
Number of authors:	Total	2,169	100
	Non-journalists	1,243	57
	Freelance journalists	857	40
	Staffers	49	3
Number of articles written by:	Non-journalists	3,088	23
	Freelance journalists	4,435	33
	Staff journalists	5,397	44
Number of articles written by:	Males	7,299	55
	Females	4,715	36
	N/A (news organizations)	1,145	9
Number of words	Mean	1,008	–
	Minimum	1	–
	Maximum	14,930	–
Number of quotes	Mean	3	–
	Minimum	0	–
	Maximum	65	–

Table 1 provides descriptive statistics about the articles and their authors. Most articles are published during weekdays (76% between Monday and Friday) rather than weekends. *LaPlace* publishes almost half of its articles during the afternoon (46% of the articles are published between 12pm and 6pm) and to a lesser extent in the evening and the morning; very few articles are published at night. 2,169 authors had published at least one article on *LaPlace* as of September 2011. Among these 2,169 authors, 57% are non-journalists, 40% are freelance journalists, and only 3% are staffers affiliated with *LaPlace*, which confirms the idea that the content published on the website comes from multiple sources. More than half of the authors are non-journalists. However, this percentage drops when one looks at the proportion of articles written by each of these categories of authors: staffers provide 44% of the content published on the website, against 33% for freelancers and only 23% for non-journalists. The gender balance is decidedly in favor of men: 55% of the articles published were written by men, against 36% by women. The other 9% of the articles come from partner websites, such as think tanks and NGOs, rather than individuals. Finally, articles published on *LaPlace* are on average 1,008 words long (the longest article has 14,930 words – longer than some academic articles) and feature an average of 3 quotes (the maximum number is 65 quotes for an article).

Table 2. Descriptive statistics about the measurements of online success

	Mean	Median	Min	Max
Number of visits	35,738	24,311	39	4,351,786
Number of «likes» on Facebook	330	71	0	128,445
Number of «tweets» on Twitter	37	13	0	2,679
Number of comments	147	104	0	2,310

Table 2 provides additional descriptive statistics on the main variables measuring online success. The median number of visits for *LaPlace*'s articles is 24,311 and the average number of visits is 35,738. This difference between the median and the mean is due to the presence of outliers: articles that are "hits" and attract very high numbers of visitors. For example, the most popular article ever published on *LaPlace* was viewed by more than 4 million people (we will come back to this

article later). The median number of comments for *LaPlace*'s articles is 104; the median number of "likes" is 71 and the median number of "tweets" is 13. These lower numbers of "likes" and "tweets" can be explained in two ways. First, Facebook and Twitter only became popular in recent years in France, around 2008-2009 for Facebook and 2011-2012 for Twitter. Second, sharing an article with one's friends on Twitter or Facebook reveals a deeper engagement with the content of the article compared to simply reading an article or even posting an anonymous comment on the website.

Table 3 describes the evolution of the main variables between 2007 and 2012. Before comparing numbers across years, one should remember that *LaPlace* was created in May 2007 (thus 2007 only has 8 months of data) and that the data collection ends in September 2012 (thus the 2012 only includes 9 months of data). The table documents a rapid increase in the amount of content posted online over the last five years. Between 2007 and 2012, the number of articles published per year on the website has been multiplied by 3.2. Most of this increase is due to staffers: the number of articles published by staffers has been multiplied by 5.8, whereas the number of articles published by non-journalists has (only) doubled over the period. These trends should obviously be understood in the light of *LaPlace*'s expansion: the staff grew from 10 to 15 full-time staff writers between 2007 and 2012. It is intriguing to note that articles became longer over time, from 768 words on average in 2007 to 1033 words in 2012. The average number of quotes per article also increased.

The evolution of the different measurements of success reveals a mixed picture. The median number of visits grew quickly between 2007 and 2008; it remained stable between 2008 and 2010. Over the past two years, however, there has been a decline in the number of visits per article (-32% between 2010 and 2012). Likewise, the median number of comments has been stable between 2008 and 2011, before decreasing between 2011 and 2012 from 124 to 74. In contrast, the average number of "likes" and "tweets" per article constantly increased over the period, which is logical given the expansion of the two social sites between 2007 and 2012.

'Tags' and topics of the articles

In the final part of the data section, I introduce several additional variables for identifying each article's topic. As mentioned above, the web crawl collected all the tags attached to the articles. Copy editors – not journalists – usually decide which tags

Table 3. Main variables, 2007-2012

Variables	2007	2008	2009	2010	2011	2012
Number of articles published online	1035	922	1827	2826	3263	3268
Articles written by staffers	280	254	676	1066	1496	1625
Articles written by non-journalists	332	251	461	693	697	654
Median number of words per article	694	991	876	880	994	960
Median number of quotes per article	0	2	2	2	2	2
Median number of links per article	3	3	5	4	5	4
Median number of visits	12032	28091	28754	28147	25175	19150
Median number of likes	0	2	19	101	220	106
Median number of tweets	0	0	0	1	29	51
Median number of comments	57	126	112	124	124	74

to attach to any given article. Tags are drawn from a pre-existing list in order to convey essential information about the article and improve its visibility on search engines. When a new topic emerges that cannot be described using existing tags, copy editors may create a new tag. The number of tags varies between 0 and 29, with an average of 5 tags per article in the sample. As of October 2012, a total of 3,217 different tags have been used. For example, an article about the Olympic Games at Sochi, entitled “The cost of the Olympic Games at Sochi? A farce!,” had the following tags: “Olympic Games Sochi 2014,” “Russia,” “Corruption,” “Putin,” “ski,” “Winter Sports,” and “Caucasus.”

In order to test the idea that sex, celebrities, and scandals are more successful in terms of visits than world news, culture, and politics, I defined five categories of articles based on the aggregation of relevant tags about “politics,” “world news,” “culture,” “sex,” and “celebrities and scandals.” Instead of aiming for exhaustive categories – which seems hardly possible given that there are more than 3,000 tags – I gathered a critical mass of articles for each category by focusing first on the tags that received many occurrences in the data set.

Several points should be mentioned before turning to the empirical findings. First, the list of the different categories of articles proposed here might appear heterogeneous: some categories have a strong cohesion and are highly institutionalized in news organizations (for instance the “culture,” “politics,” and “world news” categories) whereas others appear more disparate or subjective (“sex” and “scandals”). The main purpose of these categories is to test *LaPlace* journalists’ representa-

tions about the kinds of articles that are successful online. Therefore, this design follows the journalists’ own classification system instead of an analytic one.

Second, it should be noted that the different tags and categories defined here are not mutually exclusive. For example, an article about Dominique Strauss-Kahn’s arrest in New York will be categorized as belonging to the “scandal,” “politics,” and “sex” categories simultaneously. The highest correlation is the one between “politics” and “scandals”: 184 articles in the sample were “tagged” both in the “politics” and in the “scandals” category. The Pearson’s correlation coefficients for the different tags under consideration reveal that the most highly correlated categories are “scandals” and “politics” ($r = 0.0537$).

Finally, I draw on several additional variables in order to test the relative popularity of user-generated content. First, I use the variable on the employment status of the author as a first approximation of participatory content by looking at the online success of articles written by “non-journalists.” Indeed, content published by “non-journalists” can be of three types on *LaPlace*: blog posts by one of the regular bloggers; “testimonies;” or op-eds. All three formats count as user-generated content. Second, I draw on two detailed variables to distinguish between “testimonies” and “blogs.” I created a category called “testimonies” including all the articles published with the tag “testimonies.” This variable includes 276 articles. I also defined a category for “blogs.” Since there is no specific tag for blogs, I gathered the names of all the bloggers on the dedicated page of the website

Constructing the categories Tags and methodology

The “*world news*” category includes all the articles that have one of the following tags: “world” (972 occurrences), “United States” (582 occurrences), “Europe” (335 occurrences), “Arab Spring” (314 occurrences), “China” (227 occurrences), “Israel” (199 occurrences), “Tunisia” (199 occurrences), “Palestine” (140 occurrences), “Maghreb” (136 occurrences), “Russia” (129 occurrences), “Syria” (111 occurrences), and “Iran” (89 occurrences). Put together, these different tags gathered a total of 2,443 articles.

The “*politics*” category includes all the articles with the following tags: “*LaPlace* politics” (1,280 occurrences), “Nicolas Sarkozy” (former president, 1,241 occurrences), “UMP” (conservative party, 555 occurrences), “Presidential election 2012” (522 occurrences), “PS” (socialist party, 417 occurrences), “François Hollande” (current president, 353 occurrences), “FN” (far-right party, 253 occurrences), “Front de gauche” (far-left party, 69 occurrences), “Modem” (center-right party, 65 occurrences), “Presidential election 2007” (56 occurrences), and “PC” (communist party, 46 occurrences). A total of 2,896 articles are included in the “politics” category.

The “*culture*” category includes the tags “films” (395 occurrences), “culture” (344 occurrences), “music” (306 occurrences), “books” (301 occurrences), “photos” (131 occurrences), “arts” (113 occurrences), “movies” (81 occurrences), “festi-

vals” (57 occurrences), and “paintings” (37 occurrences). A total of 1,389 articles are counted in the “Culture” category.

I created a “sex” category by aggregating the following tags: “sex” (352 occurrences), “rapes” (75 occurrences), “prostitution” (70 occurrences), “gender questions” (59 occurrences), “LaPlace_69” (53 occurrences) (a section devoted to sex-related articles), and “intimate life” (12 occurrences). A total of 489 articles belong to the “sex” category.

Last, for the “*scandals and celebrities*” category, I included tags on affairs, corruption, and fraud, but also specific names of politicians and celebrities associated with contentious affairs: “censorship” (134 occurrences); “affairs” (e.g., “scandal” in French) (97 occurrences); “corruption” (84 occurrences); “Liliane Bettencourt” (80 occurrences) (heiress of L’Oréal involved in the transfer of secret funds to the conservative party); “Dominique Strauss-Kahn and the Sofitel Affair” (44 occurrences); “tax haven” (35 occurrences); “celebrities” (35 occurrences); “fraud” (34 occurrences); “Carla Bruni-Sarkozy” (33 occurrences) (model, singer, and wife of the former president Nicolas Sarkozy); “Cecilia Attias” (19 occurrences) (former wife of Nicolas Sarkozy); “conspiracies” (18 occurrences); and “Valérie Trierweiler” (11 occurrences) (journalist and partner of François Hollande at the time). The “scandal” category includes 566 articles.

that introduces the 73 blogs. 39 bloggers were active between 2007 and 2012. I then collected all the articles that they published during the period – a total of 526 articles between 2007 and 2012.

ANALYZING THE ONLINE CONTENT OF *LAPLACE*

What are the determinants of online success?

Are *LaPlace*’s journalists correct when they claim that sex, scandals, and celebrity-related articles are particularly popular online whereas world news, politics, and culture articles attract lower numbers of visits? Table 4 introduces a negative binomial regression that explores this question. In the following regression models, the unit of analysis is the article and the dependent variable is the number of visits. Poisson regressions are usually the most appropriate for count data, but the dependent variable here is characterized by a strong over-dispersion. Hence, I use a negative binomial regression instead (Land,

McCall and Nagin, 1996). As a robustness check, I ran the same models without the outliers (excluding the top percentile of articles that had received the highest number of visits) and found similar results.

Table 4. Regression model: *Number of visits and the determinants of success*

Variables	Model 1	Model 2	Model 3
“Sex” tag	0.91*** (0.038)	0.75*** (0.040)	0.70*** (0.040)
“Scandal” tag	0.19*** (0.043)	0.08* (0.041)	0.08* (0.041)
“World news” tag	-0.11*** (0.019)	-0.17*** (0.018)	-0.17*** (0.019)
“Politics” tag	0.12*** (0.018)	0.11*** (0.018)	0.08*** (0.019)

Variables	Model 1	Model 2	Model 3
“Culture” tag	-0.20*** (0.024)	-0.16*** (0.023)	-0.13*** (0.023)
Interaction term (Politics*Scandal)	0.12 (0.076)	0.21** (0.074)	0.18* (0.072)
Female author		0.08*** (0.015)	0.08*** (0.019)
Twitter account		0.03 (0.020)	0.03 (0.022)
Number of words (1 st quartile)		0.00 (0.021)	-0.03 (0.022)
Number of words (4 th quartile)		0.25*** (0.020)	0.23*** (0.020)
Number of quotes (1 st quartile)		-0.02 (0.021)	-0.04 (0.021)
Number of quotes (4 th quartile)		0.09*** (0.022)	0.09*** (0.022)
Staffer			0.11*** (0.023)
Non-journalist			0.06** (0.023)
Testimony			0.61*** (0.050)
Blog post			-0.27*** (0.035)
Constant	10.43*** (0.010)	10.61*** (0.041)	10.51*** (0.044)
Observations	13,152	12,008	11,957
Ll	-150175	-136085	-135231
Pseudo R-Squared	0.01	0.01	0.01

Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05. Several control variables (year, month, day, hours of publication, dummies for staffers, number of hyperlinks) are not reported in this table.

Model 1 explores the association between the topic of the articles and the number of visits that they attract; it confirms the representations of *LaPlace*'s staffers, with one important exception. As emphasized by the French journalists, articles about sex, scandals, and celebrities are more likely to receive a high number of visits compared to other articles: the coefficients for the “sex” and “scandals” tags are positive and highly significant. Specifically, the differences in logs of expected counts for online visits are expected to be 0.91 units higher for articles about sex than for other articles. Articles about sex are much more popular than articles about scandals: the coefficient for the “sex” tag is more than four times higher than the coefficient for the “scandal” tag. In contrast, the coefficients for the “world news” and “culture” tags are both negative and highly significant. Yet, surprisingly, articles about politics – which were predicted to be unpopular – turn out to be significantly more successful than other types of articles on *LaPlace*'s website: the coefficient for the “politics” tag is positive (0.12) and significant at the 0.001 level. Since tags overlap, this positive coefficient may be due to the fact that many articles categorized in “politics” are also about “scandals.” Thus, I added an interaction term to control for this phenomenon, but the coefficient for politics remains positive even with the interaction term. It should be noted that the amount of variance explained is low, with a pseudo R-Squared of 0.01.

Model 2 elaborates on Model 1 by adding important control variables. First, I include the gender of the author (female=1) and whether the author has a Twitter account (Twitter=1).⁴ I also incorporate the number of words and number of quotes for each article, for which I created quartiles. The third quartile is omitted from the regression and Table 4 only reports the lowest and highest quartiles (e.g., the 25% articles featuring respectively the lowest and highest numbers of words and quotes). Last, I added controls for the hour and date of publication (not reported in the table). In Model 2, the sign and statistical significance of the coefficients for the different tags remain highly stable compared to Model 1. In addition, articles written by women are more likely to receive a higher number of visits than articles written by male authors. A possible explanation for this finding is that women are more likely than men to write on “soft” news such as social issues and lifestyle (Neveu, 2000; Damian-Gaillard, Frisque, and Saitta, 2012), which might not be captured by the existing tags. Having an account on Twitter is positively but non-significantly correlated with the number of visits. Last, articles that feature a high number of words and quotes are correlated with high numbers of visits: the coef-

ficients are positive and highly significant for the two top quartiles.

Model 3 keeps the control variables and adds several additional variables about the authors. First, I rely on a broad categorization of the occupational status of the article’s author: is he or she a staffer, a freelance journalist, or a non-journalist? “Freelance journalist” is the omitted category in the model. Model 3 documents that articles written both by staffers and non-journalists are overall significantly more successful than articles written by freelance journalists. The last two variables of Model 3 unpack the category of “user-generated content” by distinguishing between “blog posts” and “testimonies.” Interestingly, the coefficients are significant but of opposite signs. The coefficient for testimonies is positive, high (0.61), and significant at the 0.001 level, thus indicating that testimonies are very popular on the website. In contrast, the coefficient for blog posts is negative and also significant at the 0.001 level, which means that blog posts tend to be less successful online than other types of articles.⁵

Different profiles of readers: visits, “likes,” “tweets,” and comments

Whereas the previous model only considered the number of visits as an indicator of online popularity, this section compares the determinants of ‘success’ when different measurements are taken into account: not only number of page views, but also “likes” on Facebook, “tweets” on Twitter, and comments on the website. Table 5 provides several negative binomial regression models. The number of visits, likes, tweets, and comments are the dependent variables for Model 1, 2, 3, and 4 respectively.

Table 5 shows that the determinants of success on *LaPlace* are not strikingly different when success is measured by the number of comments compared to other indicators such as visits, likes, and tweets. Overall, coefficients are highly similar in Models 1, 2, 3, and 4. Articles about sex or politics perform better in terms of visits, likes, tweets, and comments compared to articles about world news and culture. In addition, articles written by non-journalists are more successful than articles written by freelancers; blogs posts are not very popular, whereas testimonies attract many visits, likes, and tweets.

The only counterintuitive difference emerging between Models 1, 2, 3 and Model 4 is that articles about “scandals” are relatively unpopular in terms of comments: the coefficient is negative (-0.13) and significant at the 0.01 level. In addition, the coefficient for “politics” is higher in Model 4 than in Model 1, 2, and 3, thus indicating

that articles about politics attract more comments than visits, “likes,” and “tweets.”

Table 5. Regression model: visits, likes, tweets, and comments

Variables	Model 1 Visits	Model 2 Likes	Model 3 Tweets	Model 4 Com- ments
“Sex” tag	0.70*** (0.040)	0.66*** (0.088)	0.44*** (0.079)	0.30*** (0.045)
“Scandal” tag	0.08* (0.041)	0.26** (0.089)	0.25** (0.081)	-0.13** (0.046)
“World news” tag	-0.17*** (0.019)	0.00 (0.040)	-0.07 (0.037)	-0.14*** (0.021)
“Politics” tag	0.08*** (0.019)	0.10* (0.041)	0.18*** (0.038)	0.30*** (0.021)
“Culture” tag	-0.13*** (0.023)	-0.04 (0.049)	-0.23*** (0.046)	-0.39*** (0.026)
“Politics” “Scandal”	0.18* (0.072)	-0.05 (0.157)	0.58*** (0.147)	0.09 (0.081)
Number of words (1 st quartile)	-0.03 (0.022)	-0.01 (0.047)	-0.17*** (0.045)	-0.16*** (0.024)
Number of words (4 th quartile)	0.23*** (0.020)	0.36*** (0.043)	0.20*** (0.040)	0.09*** (0.022)
Constant	10.51*** (0.044)	3.07*** (0.094)	-0.44*** (0.094)	4.77*** (0.049)
Observations	11,957	11,963	11,963	11,959
Ll	-135231	-70721	-43376	-70360
Pseudo R-squared	0.01	0.03	0.08	0.02

Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05. Several control variables (year, month, day, time of day, dummies for staffers, gender of the author, whether the author is a staff writer or not, Twitter account, number of quotes, and number of hyperlinks) are not reported in this table.

In order to add nuance to the findings of this statistical model, I turn to a more qualitative analysis of the “top ten” articles that attracted the highest number of visits, likes, tweets, and comments since the creation of *LaPlace*. Table 6 relies on the following color code: dark color cells are for “hard news” articles; medium color cells are for user-generated articles; light color cells are for “sex and gender” articles; and black underlined headlines indicate articles that appear several times in the table.

The most successful article ever published on *LaPlace* (4,351,786 visits) is entitled “Video: a French-Tunisian professor murdered during a protest in Tunisia.” An internet user sent to *LaPlace* a video of the murder that marked the unravelling of the Arab Spring in January 2011. Marina, then managing editor of the website, embedded the video in the page and wrote an article summarizing the information; this piece of breaking news was followed by many media outlets around the world. The second most visited article (3,282,176 visits) also relates to the beginning of the Arab Spring. Written by a freelance journalist from Tunisia and published in December 2010, it describes how the suicide of an unemployed man led to arbitrary repression and censorship by the Tunisian authorities. The third most successful article (1,587,103 visits) was written by a staffer and published in October 2010; it is a political article with a witty and sarcastic tone describing the internal wars of the conservative UMP party (it compares several politicians to the Three Musketeers and the “Dalton brothers” of the Lucky Luke comic books series).

On Facebook, the most popular article (128,445 likes) is entitled “Dear Neighbors: Best and Worst Neighbor Stories. Pictures” In this webpage, which is frequently updated by the staffers, internet users are invited to send pictures of the funniest notes that their neighbors have left in the communal parts of the buildings. *LaPlace*’s journalists then publish the pictures and sometimes add a comment. The second most popular article on Facebook (49,822 likes) is a politics piece entitled “600 good reasons not to vote for Sarkozy.” First written by a freelance journalist, this article has been frequently updated and completed. It lists some of the most controversial quotes and political measures taken by Nicolas Sarkozy during his presidency. The third most successful article on Facebook is a testimony entitled “A high school teacher: I faked a Wikipedia account to trick my students.” Out of the ten most popular articles on Facebook, four are about sex and gender.

On Twitter, the most popular article is once again “600 good reasons not to vote for Sarkozy” (2,679 tweets). The second most successful article is entitled “Get lost, asshole! The story of the man in-

sulted by Sarkozy” (2,087 tweets). This is a fictional piece imagining the life of the (unidentified) man who was insulted by Sarkozy during a public meeting. The third most popular story is “The 46 first steps of a Twitter addict: Where are you?” It was originally written in English and published by a U.S. blogger on his website.

Finally, the most popular article in terms of comments is entitled “Heated debates at the Knesset: ‘Traitors, go to Gaza!’” It is a short “lookout” piece which summarizes current political debates in Israel (2,310 comments). The second most popular article is entitled “The hearing of Dominique Strauss-Kahn postponed because of DNA testing.” It was written by Phillipe, the editor-in-chief (1,800 comments). The third most popular article was written by a staffer and is entitled “True and false conspiracy theories after 9/11.”

Table 6 both confirms and adds nuance to some of the findings provided by the regression models. The articles that are popular in terms of “likes” and “tweet” appear to be fairly similar: testimonies score well on Facebook as well as on Twitter and three articles appear both in the Twitter and Facebook columns (“Dear Neighbors,” “600 reasons not to vote for Sarkozy,” and “She makes fun of Morano”).

In contrast, the kinds of articles that attract high numbers of comments are different from the more “liked” and “tweeted” stories: they include breaking news (4 out of 10 articles) as well as topics that are controversial in the French context, such as Israeli politics, Dominique Strauss-Kahn’s arrest, 9/11 conspiracy theories, racism, and Catholicism. There is not a single article about sex or gender in the “most commented” list. Only one article (“She makes fun of Morano”) in the most commented articles also appears in the other columns.

Another counterintuitive finding from Table 6 is that the most popular articles in terms of visits are first and foremost “hard news” articles. 4 out of 10 most visited articles are world news pieces about the Arab Spring, which is surprising given the negative coefficient for “world news” in the regression models. One explanation for this result is that, despite the lower overall interest of *LaPlace*’s readers for world news, these specific articles provided breaking news that were followed by many other media outlets, thus reaching a much broader audience than usual.

Hence, even though the regression models did not document striking differences between the determinants of visits, likes, tweets, and the determinants of the number of comments, a more qualita-

Table 6. Ten most successful articles: visits, likes, tweets, and comments (2007-2012)

	Visits	Likes	Tweets	Comments
1	“Video: a French-Tunisian professor murdered during a protest in Tunisia”	“ <u>Dear neighbors: Best and Worst neighbor Stories</u> ”	“ <u>600 good reasons not to vote for Sarkozy</u> ”	“Heated Debates at the Knesset”
2	“Protests in Tunisia after the suicide of an unemployed man”	“ <u>600 good reasons not to vote for Sarkozy</u> ”	“‘Get lost, asshole!’ The story of the man insulted by Sarkozy”	“DSK’s hearing postponed because of DNA tests”
3	“UMP: Christian Jacob tells how he seduced his friend Copé”	“High school teacher, I faked a Wikipedia account to trick my students”	“The 46 first steps of the Twitter addict: where are you?”	“True and false conspiracy theories after 9/11”
4	“Online auctions: I tested the mower scam for you”	“ <u>She makes fun of Morano and is fired for gross misconduct</u> ”	“Twitter: Sarkozy’s team acknowledges interference”	“Blacks, Arabs, and Criminality: Zemmour sentenced”
5	“Egypt: the army deploys its troops in Cairo”	“When I was a child, I wanted to be a baker but I was good at maths”	“What is 3G? An insider’s story to the launching of Free Mobile”	“I am going back to my home country because of racism in France”
6	“Inga Sempé: Design in France is too conservative”	“In the U.S., ridicule is the best way to kill the ‘war on women’”	“ <u>Dear neighbors: Best and Worst neighbor stories</u> ”	“‘Filthy Arab’: the ophthalmologists’ association investigates”
7	“Lou Charmelle, porn actress: ‘there are rules’”	“Eric Fassin: public policies on national identity have built a white France”	“Recent bride, I’m learning about racism for the first time”	“ <u>She makes fun of Morano and is fired for gross misconduct</u> ”
8	“Revolt in Algeria: three deaths and angry youths”	“Fist-fucking in the valley”	“ <u>She makes fun of Morano and is fired for gross misconduct</u> ”	“Integrist Catholics destroy blasphemous ‘Piss Christ’”
9	“ <u>Dear neighbors: Best and Worst neighbor stories</u> ”	“How to write to one’s neighbor when he has loud sex”	“Help senators to refrain from accumulating mandates”	“Video of the high school student hurt in flash-ball”
10	“Strauss-Kahn and his relation to women”	“Scared to get married for life? In favor of untempered love”	“Ms. Figaro names her blog and threatens”	“Toulouse: Mohammed Merah cornered, siege continues”

tive approach indicates that the articles that are the most commented upon differ starkly from the other types of articles that are popular online.

MAKING SENSE OF THE GAP BETWEEN STATISTICAL FINDINGS AND JOURNALISTIC REPRESENTATIONS

Taking advantage of the information available online, this article tests journalistic representations of popularity by examining the actual deter-

minants of success for the articles published on *LaPlace* between 2007 and 2012. The statistical analysis overall confirms the journalists’ idea that articles about sex, scandals, and celebrities are popular, whereas articles about world news and culture fare poorly in terms of traffic. Yet I also document several discrepant findings. First, articles about politics do in fact attract high numbers of visitors on *LaPlace’s* website, even when controlling for the interaction between “politics” and “scandals.” Second, articles written by non-journalists are overall more popular than articles writ-

ten by freelance journalists. Within this category of user-generated content, testimonies are highly successful in terms of visits whereas blogs are unpopular. Third, though web journalists believe that short posts tend to be more popular than long pieces, the statistical analysis indicates the opposite: in fact, longer and more deeply researched articles featuring more quotes attract more visitors, all else being equal. Finally, the analysis provides mixed findings about the distinct preferences of readers who “click,” “like,” “tweet,” and write comments on the articles. The regression model does not reveal any marked difference between the types of articles read by these specific categories of users, but a more qualitative examination of the ten most popular articles in the sample in terms of visits, “likes”, “tweets”, and comments indicates that there are more comments on controversial topics (Israeli-Palestinian conflict, political scandals, racism, etc.) that are not as prominent on Twitter or Facebook. How should we make sense of these findings? In this discussion section, I examine in turn the results regarding political articles, user-generated content, short versus long articles, and online comments.

Political articles and their changing popularity

The online success of political articles constitutes good news for *LaPlace* and for political news sites more generally. It indicates that *LaPlace*'s journalists and editors have found a way to keep their readership interested in politics with a mix of light and funny pieces, “explainers,” interviews, and critical articles. This finding also suggests that it is possible to attract a loyal base of returning readers who come to a given news site because of its engaged liberal political line and are more likely to read political articles than, for example, articles about culture. However, the pessimistic representations of *LaPlace*'s journalists about the low success of politics pieces suggest that several changes in the online competition of the website and the political climate of the country took place over time.

When *LaPlace* was created in 2007, Nicolas Sarkozy had just been elected president; the website soon became a rallying point for left-leaning voters, intellectuals, and politicians. Over time, however, people found other venues providing information on the misdeeds of the conservative presidency. *LaPlace* faced increasing competition as a left-leaning political outlet with the emergence of other stand-alone websites such as *Mediapart* in 2008 and *The Huffington Post* in 2012. *Mediapart* is probably *LaPlace*'s most significant competitor: staffers at *LaPlace* often state that many of their left-leaning readers started following *Mediapart* more assidu-

ously after that website's scoops about the Bettencourt affair in June 2010.

The election of François Hollande in May 2012 was also a mixed blessing for *LaPlace*, since it deprived the journalists of their favorite target: the political right, and more specifically former president Nicolas Sarkozy. The French website tried to adapt to this new political situation by developing its coverage of François Hollande and the Parti Socialiste, but *LaPlace* lost the prominent critical position that was its trademark during the early years of Sarkozy's presidency.

The “newsworthiness” of user-generated content

Second, my statistical analysis shows that blog posts and freelancers' articles are less successful than staffers' articles and testimonies. Yet why are articles written by staff writers and user-generated “testimonies” popular? And what do they have in common? One possible interpretation is to rely on the concept of “newsworthiness” developed in the sociology of news production (Clayman and Reisner, 1998; Shoemaker, 2006).

Staff writers, as professional journalists, are very aware of what is “trending” and “newsworthy” at any given moment of the day or week. Based on this understanding of online traffic, staffers pick topics and write headlines that are designed to be attractive for online readers. Yet freelancers and bloggers are less “in the know”: they do not work in the newsroom, they are less familiar with the editorial line and writing style of the publication, and they are usually less aware of popular topics that fall outside of their specific area of expertise.

Paradoxically, however, internet users who write testimonies might have a better timing and choice of topics than blogger and freelancers. Internet users tend to propose testimonies that are accepted by *LaPlace* when a specific subject matter attracts a lot of media attention. For example, several testimonies were published during the legal debate and political fight about gay marriage in 2013: people shared their experiences of what it would change for gay people to be able to get married or how it felt to be in a gay relationship in this context. Thus, people usually propose testimonies when they feel that the topic is popular, controversial, and “newsworthy.” Readers are interested in these first-hand accounts of highly mediated topics.

Short versus long articles

A third important and counterintuitive finding regards the influence of the length and number of

quotes featured in *LaPlace*'s articles on their online success. Whereas *LaPlace*'s journalists often claim that shorter articles tend to be more popular, the statistical results suggest that longer and more deeply researched articles featuring a higher number of original quotes attract more visitors, all else being equal.

This constitutes good news for the future of long-form journalism. Yet how can we explain the discrepancy between this result and the idea repeatedly put forward by web journalists that short articles are more successful online? A possible answer is to pay closer attention to the temporal dynamic of online success. When web journalists assess the "success" of article, they usually do it over a short time span, typically over the course of a couple of hours or days. Journalists mostly rely on real-time traffic numbers (such as the ones provided by *Chartbeat*) in order to get a sense of how "well" an article is doing. Yet some articles may attract traffic over a longer period of time, independently of whether they are successful or not during the first days of their online existence.

This is the case for "long-tail" articles. In statistics, a distribution of numbers is said to have a "long tail" when a large number of occurrences are far from the "head" or central part of the distribution. On the web, "long-tail" pieces attract regular traffic over time, either because they become a central point of reference about a given topic that remains interesting over time, or because they attract several waves of referral on online media long after their initial publication, bringing new traffic to the site. Longer and more deeply researched articles are obvious candidates for the "long tail" phenomenon.

Facebook users, Twitter users, commenters, and their specificities

Finally, let me turn to the question of online commenters. From the start, *LaPlace*'s editors and staff writers devoted a great deal of effort to building a strong online community involving regular readers and internet users. Several years later, however, staff writers repeatedly voiced their discouragement. Even though they acknowledge that internet users are a "gold mine" in terms of information and feedback, they also lament the negative impetus and "trolling" nature of most comments.

The data gathered on *LaPlace*'s website only moderately confirms these representations. The regression models showed no striking difference between the determinants of popularity measured in terms of clicks, likes, tweets, and comments. A

more qualitative approach examining the ten most popular articles in the sample revealed interesting nuances between Facebook users, Twitter users, and commenters: the most popular articles on Facebook and Twitter are mostly funny pieces or user-generated testimonies, whereas in terms of number of comments articles about controversial political topics did better.

The absence of statistical results might be due to the nature of the variables used in this analysis. The data set includes a variable about the number of comments posted under each article, but nothing on the content of the online comments (for instance their overall level of offensiveness) or the number of commenters involved (e.g., when there are 200 comments, whether ten commenters posted twenty comments each or two hundred commenters posted one comment each). Further research, both qualitative and quantitative, should continue investigating the dynamics of online commenting communities as one important manifestation of the public sphere in the digital age (Weber, 2014; Robinson, 2010; Diakopoulos and Naaman, 2011).

CONCLUSION

This article confronts journalists' interpretations of traffic numbers with statistical measurement of the determinants of online success. By comparing ethnographic evidence and quantitative data gathered at the French news site *LaPlace*, I explore whether journalists' perception of their "*quantified audience*" (Anderson, 2011) diverges from the actual behavior of their readers. I find that *LaPlace*'s journalists overall have an adequate understanding of traffic numbers, with several notable exceptions regarding political articles, user-generated content, longer articles, and differences between Facebook users, Twitter users, and online commenters. The article proposes several explanations for these counter-intuitive findings having to do with the editorial line and audience of the French website, the changing political context, and the temporal structure of online traffic.

Moving beyond the particularities of the specific news site under consideration, what does this analysis tell us about the professional norms guiding newsrooms and journalists in the digital age? It is important to underline the negative aspects of existing journalistic representations about what drives online traffic. If web journalists believe that they need to publish shorter articles about sex, scandals, and celebrities in order to grow an audience and stay afloat financially, they will orient

news websites in this direction, making it harder for readers to find quality information and long-form journalism online.

Yet the results provided in this analysis also leave room for hope. Indeed, as the specific case of *LaPlace* indicates, political articles as well as long and deeply researched articles can also be popular

over time online. Thus, providing quality information may be in the interest of news organizations as well as that of the public. In order to show this, however, it is essential to give journalists more accurate means to understand the dynamics of online readership.

NOTES

¹ The author would like to thank Paul DiMaggio, Clayton Childress, Gabe Ignatow, Pablo Boczkowski, the participants of the CITASA symposium (University of California, Berkeley) and the WIP workshop (Princeton University), as well as the reviewers and editors for their helpful comments on previous versions of this article. I am very grateful to Yannis Varni, who provided excellent research assistance for the web crawl used in this project.

² The name of the publication and individuals involved have been modified in order to protect its anonymity, in compliance with the recommendations of the Princeton University Institutional Review Board (IRB) for this research protocol (see Bonnet and Robert, 2009; Haggerty, 2004; and Bosk and DeVries, 2004 for a critical assessment of the IRB system).

³ The interviews ranged from 45 minutes to three hours. I transcribed all of the recorded interviews. The translations into English are mine.

⁴ Since the articles are clustered by staffer, the model also includes dummy variables for each of *LaPlace*'s staffers. The sign and significance of coefficients did not vary when the dummy variables were introduced.

⁵ I tried running the model without the "testimonies" variable and only with the "blog" variable: the coefficients remained the same; blogs still attracted fewer numbers of visits compared to other articles.

BIBLIOGRAPHY

- Anderson, C. W., 2013, *Rebuilding the News. Metropolitan Journalism in the Digital Age*, Philadelphia, Temple University Press.
- Anderson, C. W., 2011, "Between Creative and Quantified Audiences: Web Metrics and Changing Patterns of Newswork in Local US Newsrooms," *Journalism*, vol. 12, no5, pp. 550-566.
- Beaud, S., Weber, F., 2010, *Guide de l'enquête de terrain*, Paris, La Découverte.
- Bielby, W. T., Bielby, D. D., 1994, "All Hits are Flukes: Institutionalized Decision Making and the Rhetoric of Network Prime-Time Program Development," *American Journal of Sociology*, vol. 99, no5, pp. 1287-1313.
- Boczkowski, P. J., 2005, *Digitizing the News: Innovation in Online Newspapers*, Cambridge, MIT Press.
- Boczkowski, P. J., 2010, *News at Work: Imitation in an Age of Information Abundance*, Chicago, University Of Chicago Press.
- Boczkowski, P. J., Mitchelstein, E., 2013, *The News Gap. When the Information Preferences of the Media and the Public Diverge*, Cambridge, MIT Press.
- Bonnet, F., Robert, B., 2009, "La régulation éthique de la recherche aux États-Unis: Histoire, état des lieux et enjeux," *Genèses*, vol. 75, no2, pp. 87-108.
- Bosk, C. L., De Vries, R. G., 2004, "Bureaucracies of Mass Deception: Institutional Review Boards and the Ethics of Ethnographic Research," *The ANNALS of the American Academy of Political and Social Science*, vol. 595, no1, pp. 249-263.
- Boyd, D., Crawford, K., 2012, "Critical Questions for Big Data: Provocations for Cultural, Technological, and Scholarly Phenomenon," *Information, Communication, and Society*, vol. 15, no5, pp. 662-679.
- Braun, J., Gillespie, T., 2011, "Hosting the Public Discourse, Hosting thePublic. When online news and social media converge," *Journalism Practice*, vol. 5, no4, pp. 383-398.
- Canu, R., Datchary, C., 2010, "Journalistes et lecteurs-contributeurs sur Mediapart," *Réseaux*, vol. 160, no2, pp. 195-223.
- Christin, A., 2014, "When it comes to chasing clicks, journalists say one thing but feel pressure to do another," *Nieman Lab*, <http://www.niemanlab.org/2014/08/when-it-comes-to-chasing-clicks-journalists-say-one-thing-but-feel-pressure-to-do-another/>
- Clayman, S. E., Reisner, A., 1998, "Gatekeeping in Action: Editorial Conferences and Assessments of Newsworthiness," *American Sociological Review*, vol. 63, no2, pp. 178-199.
- Dagiral, É., Parasie, S., 2010, "Presse en ligne : où en est la recherche?" *Réseaux*, vol. 160-161, no2, pp. 13-42.
- Darnton, R., 1975, "Writing News and Telling Stories," *Daedalus*, vol. 104, no2, pp. 175-94.
- Diakopoulos, N., Naaman, M., 2011, "Towards Quality Discourse in Online News Comments," *Conference on Computer Supported Cooperative Work (CSCW)*, <http://dl.acm.org/citation.cfm?id=1958844>
- Deuze, M., 2007, "What is Multimedia Journalism?" *Journalism Studies*, vol. 5, no2, pp. 139-152.
- Domingo, D., 2008, "Interactivity in the Daily Routines of Online Newsrooms: Dealing with an Uncomfortable Myth," *Journal of Computer-Mediated Communication*, vol. 13, no3, pp. 680-704.
- Gans, H. J., 1979, *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*, Evanston, Northwestern University Press.
- Gramain, A., Weber, F., 2001, "Ethnographie et économétrie: Pour une coopération empirique," *Genèses*, vol. 44, no3, pp. 127-44.
- Haggerty, K. D., 2004, "Ethics Creep: Governing Social Science Research in the Name of Ethics," *Qualitative Sociology*, vol. 27, no4, pp. 391-414.
- Land, K. C., McCall, P. L., Nagin, D. S., 1996, "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models: With Empirical Applications to Criminal Careers Data," *Sociological Methods & Research*, vol. 24, no4, pp. 387-442.
- Lewis, S. C., 2012, "The Tension between Professional Control and Open Participation: Journalism and its Boundaries," *Information, Communication & Society*, vol. 15, no6, pp. 836-866.
- Lewis, S. C., Zamith, R., Hermida, A., 2013, "Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods," *Journal of Broadcasting & Electronic Media*, vol. 57, no1, pp. 34-52
- MacGregor, P., 2007, "Tracking the Online Audience," *Journalism Studies*, vol. 8, no2, pp. 280-298.
- Napoli, P. M., 2011, *Audience Evolution: New Technologies and the Transformation of Media Audiences*, New York, Columbia University Press.
- Neveu, E., 2000, "Le genre du journalisme. Des ambivalences de la féminisation d'une profession," *Politix*, vol. 13, no51, pp. 179-212.
- Nip, J. Y.M., 2006, "Exploring the Second Phase of Public Journalism," *Journalism Studies*, vol. 7, no2, pp. 212-236.
- Robinson, S., 2010, "Traditionalists vs. Convergents. Textual Privilege, Boundary Work, and the Journalist-Audience Relationship in the Commenting Policies of Online News Sites," *Convergence: The International Journal of Research into New Media Technologies*, vol. 16, no1, pp. 125-143.
- Schlesinger, P., 1978, *Putting 'Reality' Together: BBC News*, London, Constable & Co.
- Shoemaker, P. J., 2006, "News and Newsworthiness. A Commentary," *Communications*, vol. 31, no1, pp. 45-64.
- Singer, J. B., Domingo, D., Heinonen, A., Hermida, A., Paulussen, S., Quandt, T., Reich, Z., Vujnovic, M., 2011, *Participatory Journalism: Guarding Open Gates at Online Newspapers*, New York, Wiley-Blackwell.
- Schudson, M., 2011, *The Sociology of News*, New York, W. W. Norton & Company.
- Usher, N., 2014, *Making News at The New York Times*, Ann Arbor, University of Michigan Press.

Usher, N., 2013, "Al Jazeera English Online. Understanding Web Metrics and News Production when a Quantified Audience is not a Commodified Audience," *Digital Journalism*, vol. 1, no3, pp. 335-351.

Weber, P., 2014, "Discussion in the Comments Section: Factors influencing Participation and Interactivity in Online Comment," *New Media & Society*, vol. 16, no6, pp. 941-957.

En • What kinds of articles are read most often on news websites? In web newsrooms, journalists now have access to software programs that allow them to track the preferences of their readers in real time. Based on this data, web journalists often claim that “sex, scandals, and celebrities” are the best recipe for attracting readers and “clicks.” Conversely, journalists assert that articles on world news, politics, and culture fare poorly online. Journalists also strongly distinguish between groups of readers, depending on whether they “click,” “like,” “tweet,” or write comments. This paper compares qualitative material and quantitative data to explore whether journalists’ representations of their “quantified audience” diverge from the actual behavior of their readers. Drawing on an original data set composed of 13,159 articles published between 2007 and 2012 on a French news site, I find that articles about sex, scandals, and celebrities indeed attract more readers than articles about world news or culture. Yet articles about politics, long articles, and user-generated contributions are also highly popular. In addition, the preferences of readers vary depending on whether one measures this by the number of “clicks,” “likes,” “tweets,” or comments on the articles. Though regression models do not reveal significant differences, a more qualitative approach indicates that the most popular articles on Facebook and Twitter are humorous pieces and user-generated contributions, whereas controversial political topics attract more comments. I propose several explanations in order to make sense of this gap between the journalists’ representations and the behavior of online readers. These involve the editorial line and audience of the French website under consideration, the changing political context and media coverage of politics in France, and the specific temporal structure of internet traffic for longer articles.

Key-words: web journalism, online audience, journalistic representations, web traffic, mixed methods.

Fr • Quels sont les articles les plus lus sur les sites d’information en ligne? Dans les salles de rédactions web, les journalistes ont désormais accès à des logiciels informatiques leur permettant de détecter les préférences de leurs lecteurs en temps réel. Se basant sur ces données, les journalistes web déclarent souvent que « le sexe, les scandales et les célébrités » sont la meilleure recette pour attirer les lecteurs et leurs « clics ». Inversement, les journalistes affirment que les articles d’information internationale, politique ou culturelle donnent des résultats médiocres en terme de trafic internet. Les journalistes établissent également des distinctions claires entre groupes de lecteurs, selon qu’ils « cliquent », « aiment » (sur Facebook), « tweetent » (sur Twitter), ou écrivent des commentaires sur les articles. S’appuyant sur une double approche quantitative et qualitative, cet article examine si les représentations des journalistes sur leur audience divergent par rapport au comportement réel des lecteurs. S’appuyant sur une base de données composée de 13159 articles publiés entre 2007 et 2012 sur un site d’information français, l’analyse indique que les articles portant sur le sexe, les scandales et les célébrités attirent en effet plus de lecteurs que les articles d’information internationale ou culturelle. Toutefois, les articles portant sur la politique, les articles longs, et les contributions générées par les utilisateurs eux-mêmes sont également populaires. Par ailleurs, les préférences des lecteurs varient selon que les mesures effectuées tiennent compte du nombre de « clics », de « j’aime », de « tweets » ou de nombre de commentaires sur les articles. Bien que les modèles de régression ne révèlent pas de différences significatives, une approche plus qualitative indique que les articles les plus populaires sur Facebook et Twitter sont des articles d’humour et des contenus générés par les utilisateurs, tandis que les sujets politiques controversés génèrent plus de commentaires. Plusieurs explications sont proposées afin de comprendre ce relatif écart entre les représentations des journalistes et le comportement en ligne des lecteurs.

Celles-ci impliquent la ligne éditoriale et l'audience du site français analysé, les évolutions du contexte politique et de la couverture médiatique politique en France, et enfin la spécificité de la structure temporelle du trafic pour les articles longs.

Mots-clés : journalisme en ligne, public, représentations journalistiques, trafic internet, méthodes mixtes.

Pt. Quais são as matérias mais lidas nos sites de notícia? Nas redações de internet, os jornalistas possuem acesso a softwares que lhes permitem detectar as preferências de seus leitores em tempo real. Ao se limitarem a esses dados, os webjornalistas frequentemente declaram que “sexo, escândalo e celebridades” são a melhor receita para atrair leitores e “cliques”. De modo inverso, eles afirmam que suas matérias sobre política, economia e internacional dão resultados medíocres em termos de tráfego na internet. Os jornalistas também estabelecem uma clara distinção entre os grupos de leitores, entre aqueles que “clacam”, “curtem” (no Facebook) e “tuitam” (no Twitter), ou escrevem comentários sobre as matérias. A partir de uma dupla abordagem quantitativa e qualitativa, este artigo examina se as representações dos jornalistas sobre as suas audiências divergem em relação ao comportamento real dos leitores. Utilizando-se de uma base de dados composta por 13.159 matérias publicadas entre 2007 e 2012 em um site de notícias francês, a análise indica que as matérias sobre sexo, escândalos e celebridades de fato atraem mais leitores do que as matérias sobre cultura ou internacional. Todavia, as matérias que tratam de política, as matérias mais longas, e as contribuições geradas pelos próprios usuários também são populares. Além disso, as preferências dos leitores variam quando as mensurações levam em consideração o número de “cliques”, “curtidas”, “tuites” ou de comentários sobre as matérias. Embora os modelos de regressão não tenham revelado diferenças significativas, uma abordagem mais qualitativa indica que as matérias mais populares no Facebook e no Twitter são as que tratam de humor e de conteúdo gerado pelos usuários, enquanto os temas políticos mais controversos geram mais comentários. Várias explicações são propostas com o objetivo de compreender essa relativa distância entre as representações dos jornalistas e o comportamento online dos leitores. Elas incluem a linha editorial e a audiência do site francês analisado, as evoluções do contexto político e da cobertura midiática política na França e, enfim, a especificidade da estrutura tempo tráfego no caso das matérias mais longas.

Palavras-chave: jornalismo online, público, representações jornalísticas, tráfego na internet, métodos mistos.

