MEDIA AND PUBLIC INTEREST IN THE ERA OF WEB ANALYTICS: A CASE STUDY OF TWO SPANISH LEADING NEWSPAPERS

Medios e interés público en la era de la analítica web: un caso de estudio de dos de los principales diarios españoles

Santiago Justel-Vázquez, Josep-Lluís Micó-Sanz and Guillem Sánchez-Marín

Abstract
The labels “The most viewed”, “The most shared”, and “The most commented” are examples of the logic from web analytics tools in an information ecosystem where the audience has acquired an enormous role because it is now possible to quantify more precisely than ever before. In this context, the present investigation focuses on how the digital media agenda is currently being constructed in relation to the increasing importance given to pageview statistics. The implications of this new scenario for our society as a whole will be analyzed. This research takes the form of a case study of two of Spain’s leading newspapers, Lavanguardia.com and Elperiodico.com, using content analysis and in-depth interviews with working professionals. The aim of the research is to attempt to answer the following questions: How relevant is public interest news? How much importance is accorded to anecdotal and viral content?

Keywords
Journalism; Agenda; Democracy; Internet; Digital media; Web analytics; Audiences; Public interest.
Respuesta a la siguiente pregunta: ¿qué peso tiene el interés público en las noticias y el contenido viral o anecdótico que se comparte en las páginas vistas. Se analizan las implicaciones de este nuevo escenario en nuestras sociedades democráticas. Esta investigación toma la forma de un caso de estudio sobre dos de los principales diarios españoles, El País y El Mundo, recurriendo al análisis de contenido y a entrevistas en profundidad con profesionales en activo. El objetivo es dar respuesta a la siguiente pregunta: ¿qué peso tiene el interés público en las noticias y el contenido viral o anecdótico que se propaga en la web?

Palabras clave
Periodismo; Agenda; Democracia; Internet; Medios digitales; Analítica web; Audiencias; Interés público.

Resumen
Las etiquetas “Lo más visto”, “Lo más compartido” y “Lo más comentado” son ejemplos de la lógica que las herramientas de analítica web imponen a los medios digitales en un ecosistema informativo en el que la audiencia, que hoy se puede cuantificar más al detalle que nunca, ha adquirido un protagonismo enorme. En este contexto, el presente artículo se centra en estudiar cómo se construye hoy la agenda de los medios en relación con la creciente importancia dada a las estadísticas cuantificadas de manera más detallada. Se analizan las implicaciones de este nuevo escenario en nuestras sociedades democráticas. Esta investigación toma la forma de un caso de estudio sobre dos de los principales diarios españoles, El País y El Mundo, recurriendo al análisis de contenido y a entrevistas en profundidad con profesionales en activo. El objetivo es dar respuesta a la siguiente pregunta: ¿qué peso tiene el interés público en las noticias y el contenido viral o anecdótico que se propaga en la web?

1. Introduction
The seeds of this research were planted back in the beginning of 2010. It arose from the new presence on newspaper websites of tables or sections making reference to “The most viewed”, “The most commented”, or “The most shared”, and of labels such as “Popular” and “Trend”. Such labels are an example of the essence and logic that the internet imposes in online media.

What is the digital media agenda? How are the statistics obtained from web analysis influencing the exercise of journalism, which at the same time finds itself being conditioned by the economic crisis and the transformation of its organizational model? When importance is accorded to the internet imposes online media.

1.1. Literature review
The scope of this article and the tension that exists between the commercial and professional poles of the media is not new to journalism (Sparrow, 1999, p. 74). The press first had to achieve enough freedom to be independent of political power, the close monitoring of which is an important role assigned to newspapers since the beginning by democratic theory (Jefferson, 1984; De-Tocqueville, 2000). This was followed by the progressive institutionalization and integration of the increasingly large mass media companies within the economic structure of society. This posed a threat that was not foreseen from a theoretical perspective: the commercial aspect of newspapers conditioning the work of journalists. This phenomenon raised concerns starting early in the twentieth century and has remained a part of the debate about the media in contemporary democracies. Concern about this issue became the basis for socially responsible media theory (Siebert; Peterson; Schramm, 1963).

With the advent of the internet and in the context of an economic crisis which is especially acute in the media, the commercial pressure seems even greater on a sector that has witnessed the gradual disappearance of its traditional business model and has yet failed to encounter any alternatives. Thus it becomes crucial to consider aspects such as the essence of the media (Bettig; Hall, 2002; Hoskins; McFadyen; Finn, 2004) and the intersection between the press, the market, and democracy (McChesney; Wood; Foster, 1998; McChesney, 1999, 2000; Baker, 2002; 2015; Almiron, 2010; Fenton, 2014).

The current research draws on the works of authors such as Lipmann (1995; 1997) and McCombs (2004) in addressing many of the issues related with the composition of the media agenda and more specifically with the way in which this agenda becomes the public agenda.

Newspapers, television, and news websites provide the public with clues about the relevance accorded to different themes.

“Over time, the issues emphasized in news reports become the issues regarded as most important among the public. The agenda of the news media becomes, to a considerable degree, the agenda of the public” (McCombs, 2004, p. 2).

By accepting that the public agenda is influenced to a greater or lesser degree by the media agenda, the journalistic selection process is given the responsibility to demonstrate concern for that which is relevant to society. If communication creates a community (Dewey, 1966), the health of the latter depends on the quality of the former.

Immersing oneself in the realm of this agenda invites one to also enter the complex realm of the processes by which public opinion is formed (Childs, 1962; Habermas, 1990; Noelle-Neumann, 1986) and to review the criteria for newsworthiness (Warren, 1959; Tuchman, 1978).

A media agenda dominated by banal themes would result in an impoverished public debate that is also dominated by banal themes. This research takes as its starting point the concern arising from such a scenario, in which the combina-

Register for free at https://www.scipedia.com to download the version without the watermark
tion of the media’s economic problems and the possibilities offered by web analytics displace public interest and favor the primacy of commercial criteria. The public interest (Croteau; Hoynes, 2001; Downs, 1962; McQuail, 1998) is one of the concepts on which this research rests and which emanates from the political condition of the person when they form part of a community.

No one questions the usefulness of web analytics for business, but its application to journalism, because of journalism's direct connection with the public interest, is more complex. In the field of marketing the objectives of web analytics are clear: the numbers, data, statistics, and analysis provide a sound basis for decision making (Kaus-hik, 2010, p. 11). However, in the field of journalism the concept becomes “muddy and contentious” (Tandoc; Thomas, 2015, p. 244), as it has the ability to tip the balance in the profession between its public service role and its role as a medium for advertising. It does not seem realistic to hope that the media will turn their backs on the reality and logic that prevail on the internet if they want to ensure their survival. But on the other hand, due to its social function the newspaper business cannot be equated to any other.

This study aims to outline the composition of the digital media. Additionally, it aims to measure the contribution of the media to the quality of the public debate, as they are one of our societies’ main discussion forums.

Recent publications focusing on the disruption brought about by the arrival of web analytics in newsrooms and the implications of this (Tandoc; Thomas, 2014; Tandoc, 2014) with regard to new ways of consuming information (Olmstead; Mitchell; Rosenstiel, 2011), the divergence between the choices made by journalists and those made by the audience (Boczkowski; Peer, 2011), SEO tools and indicators (García-Carrero et al., 2016), or on media activity in social networking sites (Lasorsa; Lewis; Holton, 2013; Holcomb; Gross; Mitchell, 2011; Oriella PR Network, 2012; Broersma; Graham, 2013; Doval, 2014) have confirmed the need for the research carried out over recent years into this very current issue.

There are studies that corroborate the approach proposed in the current article, such as that conducted by Curran et al. (2009), which was guided by the hypothesis that a media system based on the market, offering softer news rather than hard news impedes the public from exercising their right to be informed.

2. Methodology

The current research employs case study methodology in order to contribute to an emerging area of study. It focuses on the online edition of the two main Spanish newspapers based in Barcelona: lavanguardia.com and elperiodico.com.

A similar study focused on any other media reality would make a good subject for future research. The research employed qualitative techniques (Taylor; Bogdan, 1998) including interviews, a participant and non-participant observation period, and an exploratory web analysis. The combination of different techniques “opens up enormous opportunities for mutual advantages in each of three major phases-design, data collection, and analysis” (Sieber, 1973, p. 1337) and results in an investigation that is both deeper and clearer.

Several previous investigations such as that carried out by Parasie and Dagiral (2013) based on a combination of interviews and content analysis; or those by Domingo, Masip, and Mico-Sanz (2008); or Singer (2004) employing methodological triangulation have served as a found-
2.1. Interviews

Semi-structured in depth interviews (Berger, 2000; Wimmer; Dominick, 2011) were developed from questionnaires that included some specific questions related to respondents’ posts and also took into account further issues that arose from the answers obtained. The interview questions concern, among other issues, the criteria of newsworthiness, the changes in work routines introduced as a result of the internet, the perception of and interaction with the audience, business models in the digital environment, and the role of the press at the current time.

A total of 10 interviews were conducted (in November and December, 2015) with directors of the newspapers that form the focus of this study; middle management (section heads and copy editors); editors; trainee editors; and other specific job profiles, such as the head of web analytics, in an attempt to cover practically all the profiles present in a newsroom.

2.2. Content analysis

The work of several authors who employ these methods in the field of communications (Berger, 2000; Wrench et al., 2008; Wimmer; Dominick, 2011; Krippendorff, 2012) and some studies focusing on online media (Jones, 1998; Templar, 2001) were taken as a reference in the planning and execution of the content analysis of the present investigation.

Three modes of analysis were established for each of the points in time in which screen grabs were taken:

- The contents of the “Most viewed” list
- The news items present in the first scroll or home screen
- The quantification and classification of all information from the web.

The objective was to observe the composition and structure of the digital editions of the selected newspapers paying particular attention to aspects such as the degree of public interest of the news content. This was carried out by means of investigating both the distribution and hierarchy of the news items, the choice of front cover topics and contrasting them with the “Most viewed” ranking.

3. Results

The most valuable results are obtained from the comparison of the results of the three analyses, enriched in the next section with quotes obtained from the interviews.

The tables 1 and 2 (at the end of the document) demonstrate the most significant results.

3.1. Classification categories

Three categories were created to classify the news items with respect to their degree of relationship with the public interest. Drawing from a pilot exploratory analysis (sample zero), a scale of 1 to 3 was established, with 1 indicating the highest link with the public interest and 3 the lowest.

The criteria included to establish a classification method for this investigation were as follows; newsworthiness (Warren, 1975; Tuchman, 1983), the classical distinction inherited from Anglo-Saxon journalism of hard news and soft news and also the definitions of public interest and the public sphere:

Category 1. Hard news

Category 2. Intermediate news: This category includes contents that are not essential to the functioning of a democratic society but which may include many incidents or events considered equally relevant to the general public or to a specific public. They represent the news items that are neither considered hard news nor tabloid or anecdotal news.

Category 3. Sensational or anecdotal news: It includes...
news items or videos that proliferate on the internet and spread virally on social networking sites with no connection with the public interest and whose function is to entertain or elicit emotions.

### 3.2. The relevant and the anecdotal

Category 3 content (tables 1 and 2) represents between 5% and 9.35%, depending on the newspaper and the sample, of all the content present on the cover page, while they represent between 15.6% and 28.4% of the “Most viewed” list. These results demonstrate that they have a higher success rate than corresponds to their numerical share, if it were the case that all content types had the same rate of success.

These news items are not found in either of the print editions of these newspapers and are clearly aimed at attracting traffic, which they do. Furthermore, they do so without requiring a special presence on the cover page. Generally, newspapers do not highlight this content at the top of their cover pages. Their presence on the home page measured between 0% and 3% of the content, except on one occasion (11% of the first sample for Elperiodico.com). The print newspapers do publish anecdotal or lurid tabloid news, but such items often occupy discrete positions in the middle or toward the end of the second column.

It doesn’t seem that quality digital newspapers are currently covering important issues in the form of entertainment, but the custom of offering their audience relevant issues surrounded by banalities might create the same effect.

### 3.3. Professional judgment and the public interest

Of the total, the contents of category 1 and 2 represent an average of 95% and 90.65% respectively of the total content of the cover pages in the first and second samples taken from Lavanguardia.com, the percentages that provide a picture of their presence in the “Most viewed” list are 71.5% and 77.6% respectively. Thus it could be said they are underrepresented in this list. In the case of Elperiodico.com the combined sum of the contents of categories 1 and 2 present on all cover pages represent an average of 91.5% and 93% of the total, while the importance accorded to them in the list of most popular content is 82% and 84.5% respectively. They are therefore underrepresented, as in the case of Lavanguardia.com, although to a lesser extent.

### 3.3. Professional judgment and the public interest

Interestingly, the third column of tables 1 and 2 (that of the percentage represented by each category on the home page) can be interpreted as being the reflection of the judgment of the editor in chief or the chief of digital content who organizes the most relevant content, and the second column, that of the percentage represented by each cate-
The perception of journalists that the anecdotal and banal succeed more easily than the relevant may condition the way in which certain themes are selected in the future. Category 3 themes work more easily and also generally require much less time and effort to produce. One of the editors interviewed complained that an in-depth report, drawn up after consulting various experts was less likely to succeed than an item that simply contains a striking image or viral video, with such content frequently getting a huge number of page views. An item showing the latest suggestive photo of a celebrity or a funny clip of animals can reap many of the page views received by a section in any one day; they are a good source of traffic and can be edited in a few minutes, with hardly any investment of resources.

4. Discussion

In the editorial offices of the newsrooms no one doubts that quality can reap many page views, but they affirm that it can sometimes be disheartening to see the “Most viewed” list. According to some of the professionals interviewed, various newspapers have opted to modify the way the most viewed list is composed in order to filter the content of some sections; in this way preventing anecdotal content that the newspaper considers to be incompatible with its brand slipping in to the rankings. Another newspaper plans to directly delete such content. As previously mentioned, newspapers are a bit ashamed of this content, but are unwilling to give such content up as they are a rich source of traffic.

4.1. The banal and brand prestige

Due to its dynamic nature, the internet requires a degree of continuous learning. Digital journalism has its own distinguishing features and characteristics, widely studied during the last decade (Masip et al., 2010). But “the ‘know how’ of online journalism is a task that gets updated on daily experience” (García-Avilés, 2014, p. 268).

In this way, the media tends to speak of itself as making many mistakes: “not everything is done well”, “the balance to which we aspire is not always achieved”, or “sometimes content is published which should not have been given space”.

The need for income forces all kinds of formulas to be tested. Such formulas include the publication of content that has gone viral on social networking sites, or items of questionable newsworthiness, or even veracity, but which are exciting or strange. As demonstrated in the analysis of the total content of the cover page this content represents between 5 and 10% of the news items of the newspapers studied, leading to the question: Does the publication of such content result in an erosion of the prestige of the title’s brand? In the words of the sub-director of one of the newspapers analyzed, it is an issue that “worries and occupies the media”.

A long-term strategy of infotainment distances interested viewers and destroys the authority of the news media that opts for it in order to publish serious news (Kovach; Rosenstiel, 2007, p. 194). It doesn’t seem that quality digital newspapers are currently covering important issues in the form of entertainment, but the custom of offering their audience relevant issues surrounded by banalities might create the same effect.

The anecdotal creates traffic; about 30% of the most viewed news items are about the media itself and up to 100% in some specific samples are Category 3 items and therefore bring in money through advertising. Newspapers are reluc-
tant to give up this income. Yet the continuous adaptation required by the internet has awakened the fear that such a mixture of banal content might undermine the prestige of the brand.

Mixing the banal with the relevant is a new cocktail for the traditional media. The best formula to mitigate possible adverse effects should be explored. The creation of specific and differentiated channels to house such content would provide an alternative that avoids damaging the brand with the publication of anecdotal or viral content. A second alternative coming from newsrooms is to “do the anecdotal well”, although it is argued that failure to do so is also, in many cases, “a problem of resources”. In another journalistic setting with a higher budget, or more time, the results would be better.

“A further option for the future, beyond a rethinking of the media model, would be to use market research and web analytics to apply professional standards with greater judgement instead of using it to replace them.”

The poor economic situation the majority of the media finds itself in seems to be one of the major conditions for a strong commitment to quality. Directors claim to be clear enough to go with the latter. On a day to day basis, with the pressure on time and ever-increasing multiplication of tasks in the precarious context of the profession, there is little room in the newsroom for reflection on the decisions taken, when there are competing web traffic and creating engagement, it is necessary to be intelligent enough to go with the latter. On a day to day basis, with the pressure on time and ever-increasing multiplication of tasks in the precarious context of the profession, there is little room in the newsroom for reflection on the decisions taken. Thus, the influence of the quantitative, of the data about the number of readers for each item and of visitors who pass through the web in a day or in a month is inevitable.

5. Conclusion
It has become common to see the more anecdotal news items coexisting with topics of maximum interest. Society is becoming used to living with this and to reading diagonally through the news in an increasingly fast way. Most news websites have already multiplied cross thresholds that previously qualified as undesirable. The impact on the prestige of the newspaper’s brand is difficult to calculate but the reaction of some readers in the form of comments provides evidence that the brand is indeed being injured.

4.2. The search for balance
Public interest or audience interest?

“Thus the journalist’s task is to find the way to make the significant interesting in each story and to find the right mix of the serious and the less serious that offers an account of the day” (Kovach; Rosenstiel, 2007, p. 189). In reality there is a wide gap between the soft news, with all its human interest content that remains relevant and which is present in a large number of sections, and those news items that are anecdotal and irrelevant. Similarly, there is a full range of intermediate positions between the belief that the audience is everything and the paper turning its back on the numbers. Thus, another word repeated frequently during interviews is balance. But is this balance achieved? “It is aspired to but not always achieved”. The most viewed is weighted against the editorial line, but sometimes the temptation to publish ‘UFO’s’, as anecdotal news or ‘strange events’ as they are called in the Elperiodico.com newsroom, is irresistible as they are a reliable source of traffic. Lavanguardia.com explains that some sections that are important but costly to maintain and unprofitable in terms of page views, such as the International section, are financed by other softer but more successful content.

In his investigation, Tandoc (2014) speaks of the balance between pursuing quick revenue and accomplishing what is expected from a respected news organization, and concludes that the majority of the time the balance achieved is far from ideal and, is often tipped towards the goal of increasing traffic using web analytics to propose stories that reap a large number of clicks.

Many of the phenomena discussed in this article need to be framed within the context of the search for a business model. That the media are prepared to be flexible with regard to their traditional approach in order to give more space to anecdotal or sensational content that easily goes viral can be linked to the need to achieve good economic results that help them keep their heads above water. As a result, the media sticks to what they know, advertising, and despite the fact that they are bringing in much less income than they did with print newspapers, advertising remains their most secure revenue source. Internet advertising is generally measured for the time being in volume of page views. Thus, the influence of the quantitative, of the data about the number of readers for each item and of visitors who pass through the web in a day or in a month is inevitable.

The results of the web analysis serve to put figures on some aspects of these phenomena. Between one and two of every twenty items of content (between 5 and 10%) that are present on the cover pages of the digital media analyzed are of an anecdotal or sensationalistic nature. Their presence in the most viewed list is greater (between 15.6% and 28%). Their weight on social networking sites (the percentage they represent of the total content published by the newspapers on these platforms) has not been measured in this study, but at first glance seems clearly superior. This might prove an interesting subject of research for the future.

Good journalism has always been subsidized (Anderson; Bell; Shiry, 2012). As with the emergence in the first half of the 20th century of social responsibility theory that had journalism starting to demand the independence of the newspapers from the interests of economic and also, of course, of political power, such formulas (McChesney, 1999; Curran, 2002) may also be necessary now. If we accept the benefits of democracy...
despite its imperfections over any other alternative form of government (Dahl, 1998), we must furnish it with the mechanisms to allow it to work as closely as possible to the ideal model. Regulation should always be at the service of the common good. Although it is necessary to raise this issue, it falls beyond the scope of the current research.

A further option for the future, beyond a rethinking of the media model, would be to use market research and web analytics to apply professional standards with greater judgment instead of using it to replace them. In this way we would take advantage of information gathered from the audience by bringing the readers, listeners, or viewers closer, as demonstrated in the research referred to by Kovačch and Rosenstiel (2007, p. 236). A shift from page views to other indicators of reader attention (Chartbeat, 2015) would also be a good starting point.

Acknowledgments
This work is part of the project Información política, Twitter y democracia: El periodismo en el entorno de los medios sociales (Plan Nacional de I+D+i. CSO2014-52283-C2-2-P), Spanish Ministry of Economy and Competitiveness.

Table 1. Results of the comparative analysis for Lavanguardia.com

<table>
<thead>
<tr>
<th>Category contents</th>
<th>Percentage of total space on the cover represented by each category</th>
<th>Percentage represented by each category in the “Most viewed” list</th>
<th>Percentage represented by each category on the home page</th>
</tr>
</thead>
<tbody>
<tr>
<td>First sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 contents</td>
<td>50</td>
<td>37</td>
<td>82</td>
</tr>
<tr>
<td>Category 2 contents</td>
<td>45</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>Total of contents 1+2</td>
<td>95</td>
<td>72</td>
<td>100</td>
</tr>
<tr>
<td>Category 3 contents</td>
<td>5</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Second sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 contents</td>
<td>46.35</td>
<td>48</td>
<td>76</td>
</tr>
<tr>
<td>Category 2 contents</td>
<td></td>
<td>30</td>
<td>22</td>
</tr>
<tr>
<td>Total of contents 1+2</td>
<td>90.65</td>
<td>78</td>
<td>98</td>
</tr>
<tr>
<td>Category 3 contents</td>
<td>9.35</td>
<td>22</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. Results of the comparative analysis for Elperiodico.com

<table>
<thead>
<tr>
<th>Category contents</th>
<th>Percentage of total space on the cover represented by each category</th>
<th>Percentage represented by each category in the “Most viewed” list</th>
<th>Percentage represented by each category on the home page</th>
</tr>
</thead>
<tbody>
<tr>
<td>First sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 contents</td>
<td>41.5</td>
<td>39.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Category 2 contents</td>
<td>50.0</td>
<td>43.0</td>
<td>29.0</td>
</tr>
<tr>
<td>Total of contents 1+2</td>
<td>91.5</td>
<td>82.0</td>
<td>89.0</td>
</tr>
<tr>
<td>Category 3 contents</td>
<td>8.5</td>
<td>18.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Second sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 contents</td>
<td>47.0</td>
<td>52.5</td>
<td>72.5</td>
</tr>
<tr>
<td>Category 2 contents</td>
<td>46.0</td>
<td>32.0</td>
<td>24.5</td>
</tr>
<tr>
<td>Total of contents 1+2</td>
<td>93.0</td>
<td>84.5</td>
<td>97.0</td>
</tr>
<tr>
<td>Category 3 contents</td>
<td>7.0</td>
<td>15.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

6. References


http://doi.org/10.1080/1461670X.2011.571825


Masip, Pere; Díaz-Noci, Javier; Domingo, David; Micó-Sanz, Josep-Lluís; Salvadorra, Ramón (2010). “Investigación internacional sobre ciberperiodismo: hipertexto, interactividad,
multimedia y convergencia”. El profesional de la información, v. 19, n. 6, pp. 568-576. https://doi.org/10.3145/epi.2010.nov.02


Digitalización enriquecida de fondo antiguo y patrimonial

Software de gestión para Bibliotecas, Archivos y Museos

R recolector OAI-PMH de metadatos de diversos proveedores

¡Con estándares internacionales para un mundo enlazado y abierto!

DIGIBIS, C/ Alenza, 4. Madrid. Tel.: 914 32 08 88. E-mail: digibis@digibis.com

www.digibis.com