media/rep/

Repositorium für die Medienwissenschaft

NECSUS
European Journal of Media Studies

DFG Deutsche Forschungsgemeinschaft

Philipps Universität Marburg

NECSUS

Social infomediation of news on Twitter
A French case study

Nikos Smyrnaios and Bernhard Rieder

Abstract

Social infomediation is an emerging phenomenon that sees growing numbers of Internet users share and comment on news items on Facebook and Twitter. This study analyses a large sample of French-speaking Twitter users over a period of two months. First, we study some general characteristics of our sample's usage of Twitter, such as timescale, productivity, hashtag, and URL distribution. We then compare the French online media agenda to the most shared and discussed news items in our sample in order to highlight similarities and differences. Our findings show that even though they depend on mainstream media coverage, Twitter user preferences often push political and technological stories that have been overlooked or even ignored to the forefront.

Keywords: Twitter, infomediation, news, France, diversity

The rising power of online social networks over the last few years has brought about the phenomenon of 'social dissemination' of online information. If services such as Facebook and Twitter have initially been used as support for everyday sociability,[1] their usage has rapidly gone beyond mere interpersonal communication and become an integral part of the electronic public sphere. Such an entanglement between mass media and interpersonal information dissemination had already surfaced in the works of Lazarsfeld as early as the 1940s. Nevertheless, the rise of modern social media has altered the depth and scale of the phenomenon. Today, tens of millions of individuals worldwide share information in real time and voice their opinions publicly or semi-publicly on news related to political, social, cultural, and economic issues. Rather than through frictionless dissemination, such exchanges on Twitter occur within a 'conversational

ecosystem' where various levels of discourse intertwine and information feeds are continuous.[2] Far from your typical interpersonal discussions or mass media broadcasts, the use of Twitter is characterised by 'context collapse', in which communication goals and potential audiences overlap.[3] At that point, a complex communication scheme emerges in which scope and geometry are variable.

In this paper[4] we argue that, indeed, one must relinquish the notion of an 'average' Twitter user. Twitter gives rise to a wide range of usage modes, levels of intensity and visibility, and types of shared and consumed content. This diversity co-exists with concentration tendencies that often manifest themselves in distribution that follows specific power laws: a restricted number of users, content, sources, and topics dominate the others but still leave room for a genuine plurality that can be discovered by diving below the surface. Finally, Twitter is not sealed off from other platforms – quite the contrary: it is inextricably linked to other so-called Web 2.0 services, as well as to the information and media sphere in general.

In order to embrace this complex set of trends that characterise social networking sites (or SNS) in relation to journalism, we rely on the concept of the *social infomediation of news*.[5] The notion of infomediation has been successively used in information science,[6] economics,[7] and management[8] in order to explain the new forms of mediation that take place in digital networks between suppliers and consumers of information. In the area of online news one can observe the growing intervention of firms that originate from the 'logical layer' of the Internet, made of software and services.[9] Web service providers and software firms progressively establish themselves as inevitable passage points between producers and editors of content and also Internet users, including the many transversal ties in this increasingly complex field of actors. These infomediaries play an increasing role in the distribution of online information by making supply meet demand in a variety of ways.

Social infomediation involves online social platforms that meet the three requirements for being a SNS[10] as well as communities or groups of Internet users who share and comment on news content. The particular technical and economic characteristics of each platform dictate a number of usage constraints. This results in a specific type of infomediation scheme which requires an 'activation' of sorts (e.g. clicking a share button) and a specific appropriation on the part of its users that is conditioned by various factors, such as socio-economic and cultural characteristics, objectives, and usage contexts.[11]

The result of this triangular interaction between content, sharing platforms, and users constitutes a process of *social infomediation*. In this case, it is not simply a computer algorithm that is used to match a heterogeneous and plethoric online information offer with a diverse demand from various users, as is the case with Google News for instance. Rather, it is the interactions that take place between individual users through proprietary and structuring platforms. The use of SNS such as Twitter and Facebook to share news links is intermediated by these platforms in infrastructural terms, but also via specific functionalities and somewhat obscure rule sets. Facebook's EdgeRank, an algorithm that governs the level of visibility of information being circulated on its platform, is a good example of this.[12] Technical logic, often optimised for commercial gain, and rich social dynamics intertwine to produce complex communication spaces that take a growing part in the way Internet users perceive the world and participate (or not) in debates of public character.

The relationship between 'traditional' producers of news and these socio-technical platforms is a particularly relevant issue in this context. By comparing the topics most covered by various news websites with those most discussed and shared in SNS, one can investigate possible convergences or divergences between these two domains in order to gain further insight into a series of transformations that cannot be subsumed by terms of dubitable analytical value, such as *democratisation* or *decentralisation*. To that end, we have carried out an analysis of news-related exchanges on Twitter and a comparison with the dominant news agenda of the French web, which is clearly understudied compared to English-language content. However, before providing an overview of our methodology we need to further specify the context and thrust of our investigation.

Infomediation on Twitter

Since at least the June 2009 Iranian presidential elections and the repression that followed, Twitter has become not only an information medium of choice for news but also a forum for political mobilisation, as seen during the Arab revolutions and the WikiLeaks controversy.[13] The omnipresence of individuals and organisations that are specifically interested in the news and/or are politically engaged (media, news reporters, bloggers, political activists, NGOs, etc.) makes Twitter an invaluable tool to the networked production and dissemination of news described by Gilad Lotan and colleagues.[14] Alfred Hermida further developed this assessment by advancing

the concept of 'ambient journalism', which includes 'real-time, networked digital technologies as awareness systems that offer diverse means to collect, communicate, share and display news and information in the periphery of a user's awareness'.[15] According to Hermida, Twitter is now at the core of this ambient journalism, where citizens and reporters co-produce information fragments that, once they are aggregated by the media, make up the whole of the news.

Beyond their professional use by journalists, social media also influence the way the general public gains access to news content. Indeed, a growing number of Internet users now get their news by way of links shared using this type of service. According to the State of the Media report,[16] nearly one in ten American Internet users say they regularly follow news-topic recommendations made on Facebook and Twitter.[17] A survey carried out in France shortly before the 2012 presidential election showed that 15% of respondents had turned to these two services for information pertaining to the election campaign, although the exact proportion for each service is not known.[18] Nevertheless, according to the same survey, politically-active Internet users are three times more likely to have a Twitter account than the average population of Internet users (31% and 10% respectively).

One of Twitter's news features is the major diversity perceived by its users. According to research conducted by the Pew Foundation, Twitter users receive links from a variety of sources – not just friends and relatives, but also individuals and organisations that they have no personal connection to. The result is a feeling among Twitter users (one not necessarily shared by Facebook users) that the information they receive would be difficult to find anywhere else. In other words, given the characteristics of the infomediation it provides for – relative to the technical features of the platform as well as to the sociological aspects of its audience – such a tool seems better equipped than comparable services to take full advantage of the potential diversity of 'weak links'.[19] This particular aspect of Twitter is confirmed by a phenomenon referred to as 'indirect media exposure', observed through the analysis of a large corpus of data.[20] In short, it means being put into contact with news content without having searched for it. This can be seen as a form of *serendipity* – that is, the discovery, posting, and fortuitous reappropriation of information.[21] Through retweets in particular, users of the service can receive information from sources that they do not know and at first demonstrate no interest in. This phenomenon may lead these individuals to be exposed to a broader spectrum of political views and opinions than they are accustomed to.[22] This trend is particularly apparent

when users with opposing views hold discussions,[23] and during major social mobilisations.[24]

Despite the caution that must be exercised in the interpretation of these studies, they still paint a picture of how emerging trends in news access are mainly infomediation-based. Such new practices can potentially translate into an increase in traffic for news websites courtesy of Twitter, and ultimately into economic and journalistic consequences. At the same time, given their diversity and complexity, these phenomena raise socio-political issues regarding how the public media sphere will evolve in the coming years. This article puts forward a (relatively) large-scale empirical study based on an exploratory 'digital methods' approach,[25] so as to develop a better understanding of these difficult-to-grasp phenomena. In particular, we focus on two lines of interrogation: first, we investigate the question of *diversity*, concerning both topics and user practices; second, we study the relationship between Twitter and traditional media outlets by comparing the *topic agenda* between the two domains.

Methodology and area of inquiry

An empirical study of Twitter provides as much opportunities as it does challenges. The various APIs[26] allow for relatively complete access to data and activity from public accounts.[27] Also, Twitter users publish millions of messages every day, and such a large amount of data can easily surpass the capacities of even the most ambitious of research projects. Consequently, given that an exhaustive analysis of all the messages published on Twitter would be very difficult if not impossible, we have elected to base our observations on a subset of data. The chosen sampling method needed to meet the following four objectives: create a sample based on user accounts, rather than on selected themes, so as to assess and compare a wide range of news topics; compile a sample made up primarily of French and French-speaking users; focus on users interested in general and political news; and gather a large enough sample to allow for an in-depth analysis.

To attain these objectives we have opted for a snowball approach transposed to Twitter. We started with an initial core of 496 accounts belonging to French and French-speaking users that were interested in general and political news. The accounts were cherry-picked by three researchers following an exploration phase spanning several days.[28] The core was made up of politicians, activists, bloggers, and professionals from the media and the French-speaking Internet who regularly use Twitter. At the time the initial

core was assembled most of the users enjoyed a certain level of exposure on the platform, notably in terms of mentions and retweets – although this does not inherently imply a large number of followers.[29]

In a second step, we acquired all of their friends and followers thanks to the REST API, which brought our sample to a total of 326,532 (mostly Anglophone) accounts. To allow for a better management of such data using the technical means available to us – but mostly to stay on our Francophone target – we only kept accounts that were linked as friends or followers to at least 10 users of our initial core. Of the resulting 24,351 user accounts, 22,322 were public, non-protected accounts, and 17,359 posted at least one message during the two-month observation period (15 February – 15 April 2011). All analyses were carried out on this last sample using the REST API – without a doubt the most reliable method for accessing Twitter data – making it possible to work on all the tweets posted by these active users throughout the observation period (5,883,659 tweets in total). Tweet-ism[30] and DMI-TCAT[31] were our tools of choice for exploring the corpus, as they provide specific analytical interfaces as well as file-export functions allowing for easy transfer to other software. We mostly used Gephi,[32] a graph-visualisation platform, also the statistics-visualisation software Mondrian.[33] Our method picked up a certain number of celebrity accounts (Barack Obama, Lady Gaga) and spam accounts. Nevertheless, an analysis of the user profiles and the language used to post tweets confirmed a strong French predominance centered on Paris (5,828 users – nearly a third of our sample – explicitly stated 'Paris' in the location field). Our sample also focused on users working in the media or related professions. The bio field of 1,549 accounts (8.9% of the sample) included the word 'journalist'.

The user accounts of our sample therefore featured a few particularities – a Paris location for a third of them and a reference to a journalism-related position for nearly a tenth – corresponding to certain traits characteristic of individuals immersed in general and political news. In that regard, the sample obtained appears to be suited to the objectives of the current research project, which is centered on online news while also constituting a significant segment of the total population of Twitter users in France. It should be recalled that our sample was extracted from roughly 25,000 accounts, including over 17,000 public and active accounts, compared to the total number of Twitter accounts in France which, according to Opin-ionWay[34] and Spintank,[35] ranged from 150,000 to 250,000 at the time our data was collected.

In order to compare our findings about news sharing on Twitter to the dominant agenda of French news websites, we used data from the IPRI[36]

research project. Based on a sample of 209 websites, the researchers of that project operated a quantitative content analysis on 37,569 articles published in March 2011 by French news websites.[37] This large-scale analysis was used to determine the levels of variety (number of topics discussed) and balance (share of each topic within the corpus of articles) of the news agenda on the French web.

General overview of findings

Our corpus of tweets is balanced between those that contain a link, revealing a true intention to share/disseminate information, and those that do not, which are simply comments and conversations (see Figure 1). Nevertheless, it is important to stress that a few of the accounts are actually bots that produce a large amount of link-containing messages comparable to spam. Consequently, it can be assumed that with 'real' users there is a slight predominance of tweets without links, even though this distribution is subject to important variations depending on the topics covered. Compared to a recent finding, based on a one-day exploration of a random sample of Twitter[38] where only 11.7% of tweets were found to contain links, this is still a high proportion.

Timescale and themes
In terms of productivity, we observed that the use of the service is more intense on weekdays and that it decreases on weekends. Another interesting phenomenon is the lower productivity mid-week, notably on Wednesdays, which could be explained by France's work and academic patterns.[39] It is thus possible to infer that Twitter has become an integral part of the daily lives of the users in our sample.
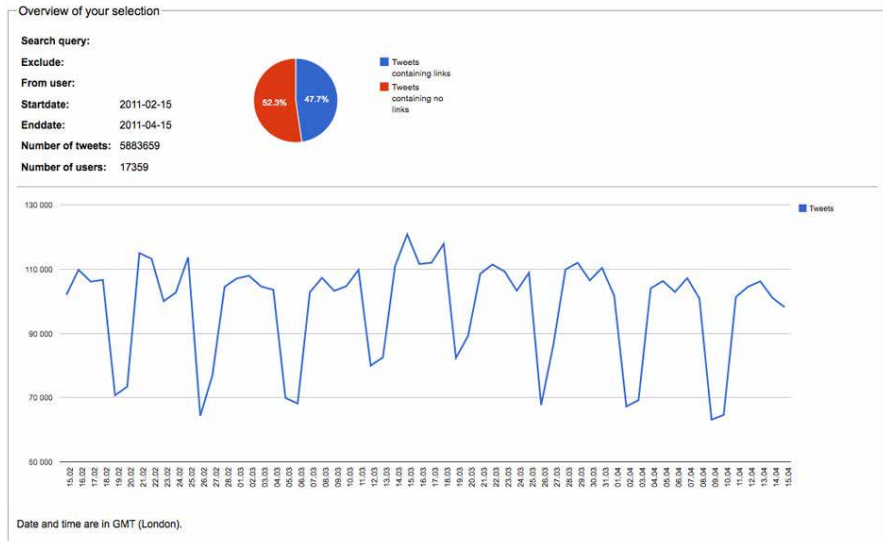
Fig. 1:   *Overview of the DMI-TCAT interface for the observation period (15 Febru-*
      ary – 15 April 2011).

Finally, it is to be noted that the week of 14 March, which was part of the
sample period for the analysis of web articles (7-17 March 2011), is the one
in which the most tweets were produced. This is undoubtedly linked to
the coverage of the tsunami in Japan, which drew considerable user at-
tention to news websites, as shown in Figure 2.[40] When the two curves are
superimposed (Figures 1 and 2) we notice that the day we collected the most
tweets over the two-month period (15 March 2011) also ranked third as the
day with the highest traffic for the sample of websites measured by AT
Internet for that same period. Therefore, it is possible to conclude, as others
before us have,[41] that there is a strong correlation between the intensity of
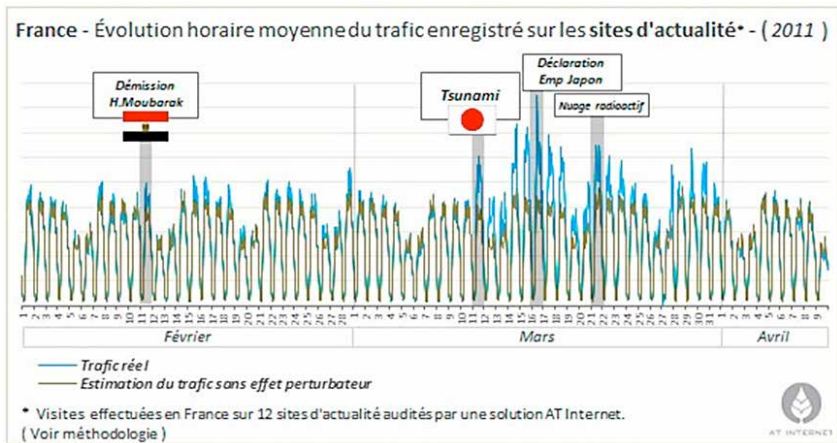the media agenda and the productivity of Twitter users.

France - Évolution horaire moyenne du trafic enregistré sur les **sites d'actualité** - (2011)

Fig. 2:   *Traffic for a sample of 12 news websites*
             (*source: AT Internet*).

The news also manifests itself in the ranking of the most frequently used hashtags,[42] which are shown here as a list (Table 1) and a network graph (Figure 3). Over a third of the tweets collected (37.7%) contained a hashtag. Overall, our corpus of tweets totaled 207,059 unique hashtags. Although hashtags can hardly be interpreted as direct identifiers of subjects and themes, such a large number points to a high diversity as far as message production is concerned.

Table 1:   **Most frequently used hashtags during the observation period (15 February –
             15 April 2011).**

No.	Hashtag	No.	Hashtag	No.	Hashtag
61,746	Ff	13,574	fb	9,323	Kadhafi
34,025	Libya	11,273	fukushima	9,169	Japan
33,254	Libye	10,794	feb17	8,450	Fail
20,526	civ2010	10,368	bm	8,345	Sarkozy
19,322	Cantonales	10,029	tunisie	8,184	Ump
16,812	Japon	9,565	iranelection	8,060	France

The most frequently used hashtag, 'ff' (for 'Follow Friday'), refers to the current Twitter practice of users recommending accounts to follow every Friday. Meanwhile, the other hashtags atop the list concern events surrounding the Arab Spring, the tsunami in Japan, and the French cantonal elections. Furthermore, in the case of international events, English hashtags are used more often, even by French-speaking users. Globally, our sample shows

evidence of deep integration of Twitter usage in everyday life patterns, especially during work time, and a strong correlation of tweet production to variations of the media agenda.
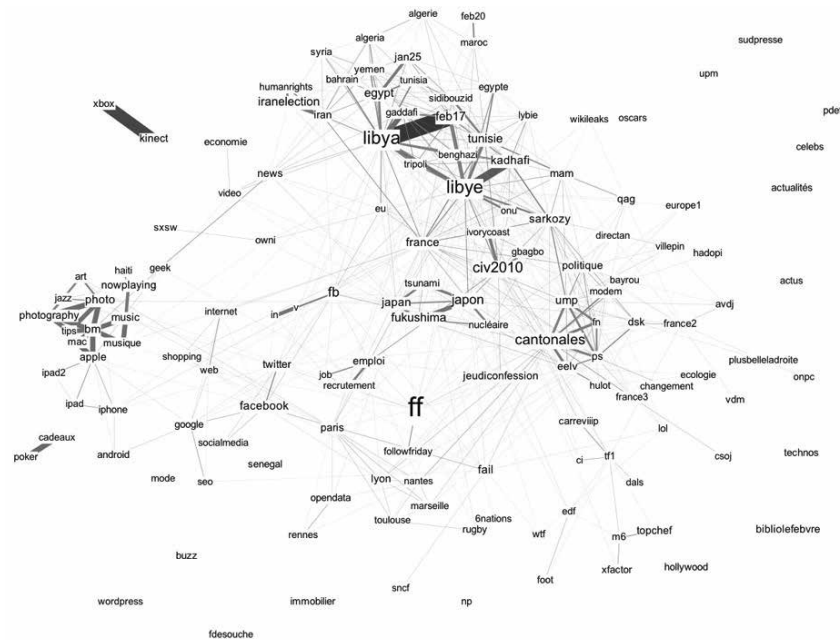

Fig. 3:   Co-occurrence relationship chart of the 144 hashtags used over 2,000 times. The size of a term shows the frequency of the hashtag. A line between two hashtags means that they have been used at least 30 times together in a single tweet (visualisation tool: Gephi).

Characteristics of user accounts

A first look at the distribution of user activity within our sample reveals overall data that allows us to establish an order of magnitude. As is often observed on SNS, most statistics from our sample feature distributions that follow a power law: a small number of individuals are responsible for the production of a large percentage of the content (50 user accounts produced 10% of all tweets, and the top 20% produced 80.2%) or have drawn the attention of a large percentage of individuals (50 user accounts are mentioned in 6% of all tweets; cf. Table 2).

Table 2:   Quantitative indicators taken from the sample of user accounts and corpus of
           tweets.

	Top 5	Top 10	Top 20	Top 50	Top 20%
4,584,800 mentions (including retweets) of 266,292 users					
Top users based on mentions	69,764	109,608	169,850	305,556	4,185,469
	(1.5%)	(2.4%)	(3.7%)	(6.67%)	(91,3%)
2,836,902 URLs shared, including 112,862 unique domain names					
URLs owned by top domains	278,099	439,773	617,664	913,783	2,681,515
	(10%)	(15.5%)	(21.8%)	(32.2%)	(94.5%)

A certain number of user accounts dominated our sample, either in terms of activity or popularity. Beyond such a 'star system' logic, there is still room for other user accounts to receive some attention. For instance, Table 2 shows that the five users who are the most mentioned or retweeted represent only 1.5% of all references (which still amounts to nearly 60,000 references over a two-month period), leaving room for less influential users to find themselves at the heart of a debate or discussion on a news topic. However, beyond the top 20% user mentions fall off quickly. Comparatively, the URLs identified in the corpus of tweets which are of particular interest to this study when they point to news websites seem to represent a much more concentrated situation: the five most frequently shared domain names in the tweets total 10% of all the shared URLs (cf. Table 2).

The social dynamics behind these distributions are structured around direct links via mentions (@*user*, thus including retweets). This trend is very strong in our corpus, with 57.7% of the tweets referring to at least one user. At a primary level, we noticed a large diversity of exchanges, with each user referring to an average of 84.4 different individuals in two months. However, this figure dwindles when we consider repeated interactions: 34.5 accounts are mentioned on average at least twice, 10 at least five times, and only 4.4 ten times or more. Brief interactions are extremely common, while stable relationships remain scarce.

The variable geometry of these communication matches on Twitter is assuredly more complex than the quantitative indicators reveal. We thus propose to consider two additional analysis axes that will illustrate possible orientations for differentiating users in the absence (due to uneven distributions) of 'typical' representative units for the whole of the population.
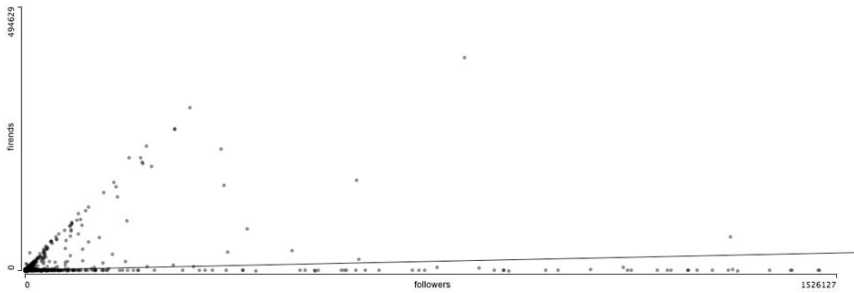
Fig. 4: Scatter plot of the sample based on the number of followers (x) and friends (y). Each point represents an account (visualisation tool: Mondrian).

Figure 4 shows three interesting elements. The least-squares trendline indicates that our sample captured an 'elite', meaning that our users have considerably more followers than they have friends. Also, we can observe the effect of one of the platform's limitations: Twitter makes it technically impossible to follow more than 2,000 accounts unless a user has at least the same number of followers. Finally, two lines stand out on the graph that may establish a broad classification for this elite. The first follows the *x* axis and aggregates 'star' accounts (several followers, few friends) that operate mainly in broadcast mode. The second, at a 45° angle, includes mostly professional users who have built their reputations through an intense Twitter regimen of following, mentioning, tweeting, retweeting, replying, sharing, recommending, and bantering.
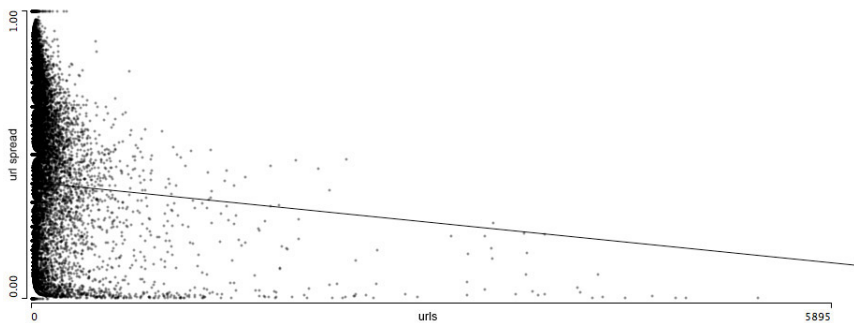

Fig. 5:   Scatter plot of the number of URLs referenced (x) and the URL-spread indicator (y). Each point represents an account (visualisation tool: Mondrian).

Figure 5 was built from the URLs appearing in the tweets and uses a derived URL-spread indicator calculated by dividing the number of unique domain names referenced by a user by the number of URLs referenced. An account that only referenced a few domain names receives a low value, whereas a more varied practice in terms of link sharing is rated closer to 1. Two factors stand out here. First, we can see that the majority of accounts adhere to a link-sharing policy that favors diversity by citing several sources. At the same time, 13.65% of accounts are below the 0.1 value, which indicates a referencing practice that focuses on a single domain name. This group of accounts includes essentially commercial players or organisational affiliates – journalists' personal accounts that reference only their own media, for instance. The URL-spread indicator is an excellent means to detect these accounts and allows us to differentiate the various players, who may have very different communication goals.

Our findings show that, globally, while phenomena of concentration and Pareto-like patterns characterise the distribution of indicators such as the number of mentions, retweets, or followers, the diversity of exchanges and encounters that take place within the platform is still remarkably high. This is also true when it comes to content sharing. The desire of many users to vary information sources that they share with others seems to have a stronger impact on the overall production of tweets – at least in our sample – than promotional strategies of organisations and individuals. Domain names constitute an interesting resource for interpretation, and we therefore propose to analyse them more thoroughly.

Most referenced domain names on Twitter

Any platform that imposes such a strong constraint – a tweet cannot exceed 140 characters – operates for the most part through interactions with other services and content available on the web. The act of referencing URLs is therefore a recurrent practice, and it could be argued that referenced content is an integral part of the message itself – at least from the standpoint of the sender.

Domain-name analysis

When it comes to examining the pluralism of the information received, an analysis by domain name is quite revealing. A general overview – scrubbed of three spam accounts – of the most tweeted domain names is therefore instructive:

Table 3:   The 12 most referenced domain names in the tweets (15 February – 15 April 2011).

No. of URLs	Domain Name	No. of URLs	Domain Name
91,527	youtube.com	28,509	twitter.com
49,321	twitpic.com	27,757	yfrog.com
43,629	lemonde.fr	24,819	wikio.fr
38,703	facebook.com	21,480	lefigaro.fr
37,799	paper.li	16,554	dailymotion.com
35,767	foursquare.com	16,506	leparisien.fr

At least two aspects of this table must be highlighted. The overwhelming presence of commercial services dedicated to sharing content generated by amateurs and others (with YouTube at the head of the pack) shows a preponderance towards *multi-platform* usage practices on the part of Internet users. Depending on subjects and common interests, users draw their own informational paths within the ecosystem created by these major platforms, and it would be extremely hard for researchers to follow these paths by digital means. What hides behind a link to YouTube? A video recorded by a demonstrator with a cell phone? A video from a TV network? A semi-professional commentator looking to become the site's next big star? Also, we notice the presence of three traditional media websites, print media in particular (Lemonde.fr, Lefigaro.fr, Leparisien.fr). This stresses the mediaphilic nature of our sample, also the rather successful transfer of popularity by some newspapers to the digital era. The print media still seem to constitute references of choice when it comes to the news, including on Twitter.

Such a strong relationship between Twitter usage and access to news becomes all the more perceptible when we look further into this classification. There are 31 news websites among the 50 most frequently mentioned domain names in the tweets during the 11-day period covered by the IPRI analysis.[43]

Among the media mentioned on Twitter, *Le Monde* clearly pulls ahead, followed by *Le Figaro* and *Le Parisien*, which corresponds roughly to the sites' importance in terms of audience. Indeed, *Le Figaro* and *Le Monde* rank first in the panel of Médiamétrie, and *Le Parisien* takes fifth place for March 2011. In all, 10 of the 20 most mentioned sources on Twitter are part of the 20 most visited news websites. We can see a cyclical pattern here as Twitter users cite popular sources, thereby enhancing the sources' popularity.

The same similarity exists (although to a lesser extent) in terms of a news website's productivity, which is evaluated according to the number of articles published on its main RSS feed. Therefore, *Le Figaro*, *Le Nouvel*

Observateur, *Libération*, *20 Minutes*, and *Le Post* engage in extensive content-dissemination strategies that result in dense traffic and high reference rates on Twitter. However, productivity does not guarantee a high level of referencing in the Twittersphere. Finally, what differentiates news dissemination on Twitter as measured in terms of the global audience of information sites is the presence of technology and foreign-news websites atop the list, most likely due to the events that occurred in Japan during the observation period.

Table 4: Top 20 news websites based on Twitter referencing, web audience, and article production, as of March 2011

(sources: Médiamétrie, Tweetism, IPRI).

Mentions on Twitter March 7–17, 2011	No. of URLs	Web Audience (March 2011)	No. of Page-views	Article Production March 7–17, 2011	No. of Articles
lemonde.fr	8,920	Le Figaro	6.3 mil	Wikio Actualités	2,125
wikio.fr	4,871	Le Monde	6.2 mil	Free Actualités	2,015
lefigaro.fr	4,524	Orange News	5.8 mil	Le Nouvel Observateur	2,012
leparisien.fr	3,723	20minutes.fr	5.5 mil	Voila Actualités	2,009
dailymotion.com	3,434	LeParisien.fr	5.4 mil	Le Journal du Dimanche	1,824
mashable.com	3,228	Nouvelobs.com	5.1 mil	France-Soir	1,820
liberation.fr	2,818	L'Express.fr	4.5 mil	Newspeg Une	1,185
lexpress.fr	2,743	Google News	4.5 mil	Le Figaro	1,166
owni.fr	2,456	TF1 News	4.4 mil	RTL France	1,152
rue89.com	2,402	Yahoo! News	4.4 mil	Google Actualités	1,042
guardian.co.uk	2,397	MSN Actualités	4.2 mil	TV5 Monde	1,008
20minutes.fr	2,393	France Télévisions News	3.5 mil	20 minutes	973
iziva.com	2,364	Libération	3.5 mil	MSN Actualités	919
tempsreel.nouvelobs.com	2,305	Le Point	3.3 mil	Orange Actualités	853
lepost.fr	2,146	Europe 1	3.1 mil	France 3	825
france24.com	2,112	Ouest France	3 mil	France 2	823
lci.tf1.fr	2,049	Dailymotion News	2.7 mil	Le Point	769
scoop.it	2,027	Le Post	2.7 mil	Libération	748
techcrunch.com	1,784	La Dépêche du Midi	2.2 mil	Europe 1	644
journaldugeek.com	1,760	Paperblog	2.1 mil	Le Post	600

The importance of technology sites as well as foreign sites is also noticeable when we look at the most frequently referenced URLs – no longer domain names – in our 7-17 March 2011 sample (the period chosen for the IPRI analysis). Indeed, among the five most frequently mentioned content items, we found pictures from Japan published by three foreign websites – ABC

(314 tweets), Boston.com (274 tweets), and *The New York Times* (200 tweets) – along with two articles from the *Owni.fr* website: one on job creation in the Internet field (272 tweets), and one other on the relationship between French nuclear reactor builder Areva and the Fukushima incident (223 tweets). In fact, the URLs that were referenced over 100 times were either from foreign sources, journalistic pure players, or the *Le Monde* website.

The URLs analysis in our sample confirms that Twitter is closely articulated with other user-generated content platforms and online media. When it comes to content sharing the traditional media hierarchy is somewhat maintained, with national dailies dominating the top of the list beside YouTube and Facebook. Nevertheless, international sources, independent pure players, and technology websites are also overexposed in our sample. The news hierarchy we found confirms these characteristics.

The hierarchy of news on Twitter

Within the framework of the IPRI project, the level of pluralism of online information was first evaluated based on the news topics covered and the volume of news articles that each topic generated. The news articles, each corresponding to a single URL, were therefore categorised on the basis of the topic they covered. In order to compare the various news topics appearing on Twitter we tracked all the links published by the sample of news websites (38,286 URLs[44]) in the tweets posted by our users. This method was difficult to apply given the various changes URLs undergo when they are shared; however, we put together a series of compensation measures so as to give sufficient validity to our results.[45] We thus found that 15,750 URLs (41% in total) had been referenced in our sample of tweets at least once, and 9,185 (24%) at least twice. Given that a number of media systematically tweet all of their articles, we feel that the second figure is more representative.

Thanks to IPRI's classification of the articles – and therefore that of the URLs – we were able to create a hierarchy of news topics on Twitter. When comparing the news topics that dominated the web's media agenda in France from 7-17 March 2011 to the most referenced subjects on Twitter, a strong similarity can easily be observed. The rankings of the first 10 news topics covered by general and political information websites, as with the Twitter references, showed the omnipresence of major international events, particularly those in Japan and Libya, as well as French political news, such as the former President Jacques Chirac's trial and the poll that showed far-right candidate Marine Le Pen ahead in the first round of the presidential

elections. The only significant difference was the overexposure on Twitter of news related to technology (computer hacking in the French Ministry of Economy) and celebrity gossip (the separation between a former football star and a top model).

Table 5: Top 10 topics covered by news websites and Twitter, 7-17 March 2011.

News Topics on the Web	No. of Articles	News Topics on Twitter	No. of Tweets
Nuclear plant accidents in Japan and related risks	4,147	Nuclear plant accidents in Japan and related risks	11,747
Conflict in Libya	4,131	Violent earthquake and tsunami in Japan	7,821
Violent earthquake and tsunami in Japan	3,101	Conflict in Libya	6,334
Nuclear-energy debate reopened in France and worldwide	1,031	Harris poll showing Marine Le Pen ahead in voter intentions	2,387
Harris poll showing Marine Le Pen ahead in voter intentions	962	Nuclear-energy debate reopened in France and worldwide	2,166
Jacques Chirac's trial	858	Cantonal election campaign	1,283
Renault case	757	Christian Karembeu and Adriana separating	1,031
Conflict in the Ivory Coast	683	Jacques Chirac's trial	1,015
Cantonal election campaign	571	Bercy hacking	888
Bahrain demonstrations and repression	475	Renault case	703

However, the comparison becomes all the more interesting if we reverse the perspective and examine topics that were the most overexposed on Twitter in relation to the number of articles published by news websites. To do so, we calculated a simple ratio by dividing the number of tweets pointing to web articles on specific news topics by the total number of articles. The updated ranking obtained from this ratio unveiled a sort of long-tail effect: isolated articles dealing with subjects that were rarely covered, if at all, by news websites but which were popular with Twitter users. Overall, topics atop the ranking dealt with Internet issues, but they often had a political component in that the subjects addressed were related to censorship, freedom of speech, or the protection of personal information.

The most striking example of this trend is the shutdown of the French version of website ReadWriteWeb, which specialises in technology and web news. Although this was only mentioned in three articles from the IPRI sample of news websites, it was tweeted 444 times by our users. In fact, the ReadWriteWeb announcement had the strongest impact in the French

Twittersphere (793 tweets in total, according to Topsy.com). Far from being a mere piece of information, Fabrice Epelboin's article is a criticism of the American head office's censorship of its French subsidiary, which it had deemed to be too partisan. As such, this issue is rather representative of the topics that ignite passion in the French Twittersphere.

Table 6: Top ten of overexposed news on Twitter, 7-17 March 2011.

News Topic	No. of Articles	No. of Tweets	Tweets/ Article
ReadWriteWeb France shuts down	3	444	148
Total petrol company does not pay taxes in France	1	117	117
Show parodying President Sarkozy's speeches is censored	1	104	104
4chan founder defends his views on Internet anonymity	1	92	92
The history of the Internet through its manifestos	2	162	81
Eviction of Mixbeat from Twitter	1	73	73
Two people in the town Bron brought in for questioning	1	71	71
President Sarkozy's counsel in far-right magazine	2	124	62
Controversy surrounding Bernard-Henri Lévy's reports from Libya for Le Journal du Dimanche	1	52	52
Facebook isn't friends with Chinese activist Michael Anti	1	51	51

Along the same lines, we find the interview with the founder of online forum 4chan regarding anonymity on the web and Facebook's attempt to censor a Chinese activist. Furthermore, SNS are at the heart of several stories that were extremely popular on Twitter, although they were mostly overlooked by news websites. If we consider the other topics related to Internet issues, this amounts to over half of the leading stories on Twitter.

However, this preference for news topics on technology in general and the Internet in particular cohabitates on Twitter with a keen interest in politics, specifically through its more controversial aspects. This can be seen in the success of the story revealing that French oil giant Total does not pay any taxes in France. This single, exclusive article published by *Marianne* magazine was not picked up by any other news websites from the IPRI sample – neither on 16 March, the day it was published, nor on 17 March. Nevertheless, this article was recommended 117 times among our sample of users (202 times in total, according to Topsy). Another theme that clearly stands out in this ranking and that goes in a similar direction is the criticism coming from the French right wing and extreme right wing. This trend seems to indicate that the French Twittersphere leans distinctly

toward the left of the political spectrum – at least based on the results generated from our sample.

Conclusions

Through these various analytical approaches, applied to a specific yet significantly large sample, a more nuanced picture emerges of the complex *social infomediation* that takes place within Twitter. Our study confirms the results of other research on Twitter usage, like deep integration in everyday life, especially during work time, and concentration patterns in the distribution of attention and activity. We also found that content-sharing practices on Twitter tend to privilege mainstream media, particularly national dailies and user-generated platforms. The issues in discussion are also dominated by those present in the agenda of mainstream media. Nevertheless, at the same time, the diversity of exchanges and encounters that take place within the platform remains remarkably high. The spectrum of information and discussions available on Twitter through sharing is also extremely large. Promotional strategies do not seem to surpass individual desire to share and comment on interesting information. Indeed, through our study, Twitter emerges as both an information network and a public sphere.

The convergences and divergences that we observed between the most frequently covered topics by French news websites – measured by the IPRI project's quantitative component – and those that were more present in our Twitter sample lead us to believe that information-sharing practices support the dominating media agenda while also running counter to it at times. Generally, it is possible to conclude that a significant percentage of Twitter users have found a way to push political stories that have been overlooked or even ignored by the mainstream media to the forefront. This in turn increases their level of exposure as well as the diversity of the information found on the platform, and it ignites numerous debates about overlooked political issues.

In complementary research work based on the same data,[46] rather than serving exclusively as a dissemination tool, Twitter appears as a place where 'mass hermeneutics', as described by Lovink,[47] materialises – that is to say, it acts as a device that collects interpretations and comments formulated about an issue (which could be as much a feature article on a political issue as a satirical video uploaded to YouTube). This mass hermeneutics can be linked to the concept of ambient journalism put forth by Hermida,[48] but it emphasises distributed sense-making over 'mere' dissemination.

As a result, although Twitter is often described as a chaotic and fragmented platform, we were able to observe convergence trends among a large number of more or less organised users who contribute to the popularity of certain stories and messages. Despite Twitter's variable communicational geometry, it still allows for specific zones, or rather *feeds*, which are characterised by profound insight and long-standing activism, to develop in parallel with this logic of concentration. These are genuine alternative views of the news and current political issues, as opposed to those expressed in the mainstream media and by traditional power structures. It is just a question of knowing where to find them.

Notes

1. Jouët 2011.
2. Boyd et al. 2010.
3. Marwick & Boyd 2011.
4. This paper is a shortened and adapted translation of Rieder, B. and Smyrnaios, N, 'Pluralisme et infomédiation sociale de l'actualité: le cas de Twitter', Réseaux, Number 176, 2012: pp. 105-139.
5. Smyrnaios 2012.
6. Knauf & David 2004.
7. Hagel & Singer 1999.
8. Águila-Obra et al. 2007.
9. Benkler 2006.
10. According to Boyd & Ellison 2007, an SNS is defined as any online service that allows its users to: 1) create public or semi-public profiles; 2) connect these profiles with a list of users; and 3) browse through their contacts and those of others. The nature of the links and functionalities allowed within the system varies from one SNS to another.
11. Ellison et al. 2011.
12. Bucher 2012.
13. Lotan et al. 2011; Lindgren & Lundström 2011.
14. Lotan et al. op. cit.
15. Hermida 2010.
16. Pew 2012.
17. 'What Facebook and Twitter mean for news', study conducted January 2012 with a representative sample of 3,016 Americans over the age of 18.
18. Online survey by CSA of 1,006 individuals 18 and over conducted 27-29 March 2012.
19. Wu et al. 2011.
20. An et al. 2011.
21. Ertzscheid 2009.
22. Bode et al. 2011.
23. Conover et al. 2011.
24. Poell & Borra 2012.
25. Rogers 2009.

26. In this context, an API (application programming interface) is defined as a means to retrieve structured data from a web service. For an overview of the various APIs made available by Twitter, please see the following link: https://dev.twitter.com/docs.

27. Users can create private accounts, but this practice is relatively rare. According to our analysis, only 1 in 10 accounts is 'protected'.

28. We would like to thank Martin Pasquier for his invaluable assistance in this matter.

29. From the viewpoint of User A, a 'follower' is User B, who follows User A (i.e., has subscribed to that user's tweet feed), and a 'friend' is User C, who is followed by User A.

30. Tweetism is a data collection and analysis platform developed for Twitter by Bernhard Rieder and Raphaël Velt.

31. The Digital Methods Twitter Capture and Analysis Toolset is a data analysis platform developed for Twitter by Erik Borra and Bernhard Rieder as part of the University of Amsterdam's Digital Methods Initiative.

32. http://gephi.org

33. http://stats.math.uni-augsburg.de/mondrian/

34. http://archives.lesechos.fr/archives/2010/LesEchos/20803-127-ECH.htm

35. http://www.spintank.fr/twitter-en-france-etat-des-lieux-chiffres-2011/

36. Our own research was part of the IPRI (Internet, pluralisme et redondance de l'information) research program that was supported by a grant from the Agence nationale de la recherche (ANR-09-JCJC-0125–01b). Involved in the program were several research teams specialising in information and communication (CIM, University of Paris 3, France; ELICO, University of Lyon, France; LERASS, University of Toulouse 3, France; CRAPE, University of Rennes 1, France; GRICIS, UQAM, Montreal, Canada) and computer science (LIRIS, INSA Lyon, France). Internet-user data for March 2011 was acquired from AT Internet and Médiamètrie.

37. Marty et al. 2012.

38. Gerlitz & Rieder 2013.

39. In France a significant number of workers, notably young mothers, do not work on Wednesdays. In many areas Wednesday is also often a day without school.

40. This data on the pluralism consumed confirms the caesura observed during our analysis of the pluralism offered by websites. Before Friday, 11 March, we were dealing with a typical news period, whereas the news became highly charged in the aftermath of the tsunami in Japan.

41. Yang & Leskovec 2011.

42. Hashtags are keywords preceded by the hash symbol that are used to define a tweet or put it into context. An example of a tweet containing a hashtag is as follows: '#Libya: Seïf al-Islam trial in September in Zintan'. Because these keywords are clickable, they can create thematic threads that are easy to browse.

43. Marty et al. op. cit.

44. The number of URLs (38,286) was reduced to 37,569 after the manual identification of a few technical artifacts (see the other article in this series).

45. We have created a way to follow up on HTTP 30x redirections, so as to retrieve a portion of the variations produced by content-management systems.

46. Rieder 2012.

47. Lovink 2012.

48. Hermida 2010.

References

An, J., Cha, M., Gummadi, K., and Crowfort, J. 'Media landscape in Twitter: A world of new conventions and political diversity', paper presented at the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain 17-21 July 2011.

Águila-Obra, A. R., Padilla-Melendéz, A., and Serrarols-Tarrés, C. 'Value creation and new intermediaries on Internet: An exploratory analysis of the online news industry and the web content aggregators', *International Journal of Information Management* 27 (2007): 187-199.

Benkler, Y. *The wealth of networks: How social production transforms markets and freedom*. New Haven-London: Yale University Press, 2006.

Bode, L., Hanna, A., Sayre, B., Yang, J., and Shah, D. 'Mapping the political Twitterverse: Finding connections between political elites', Political Networks Paper Archive, Working Paper 59, http://opensiuc.lib.siu.edu/pn_wp/59.

Boyd, D., Gilder, S., and Lotan, G. 'Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter', paper presented at the HICSS-43, IEEE Conference, Kauai, HI, 6 January 2010.

Boyd, D. and Ellison, N. B. 'Social network sites: Definition, history, and scholarship', *Journal of Computer-Mediated Communication*, vol. 13, no. 1, 2007: http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html.

Bucher, T. 'Want to be on the top? Algorithmic power and the threat of invisibility on Facebook', *New Media & Society*, vol. 14, no. 7, 2012: 1164-1180.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. 'Measuring user influence in Twitter: The million follower fallacy', paper presented at the Fourth International Conference on Weblogs and Social Media, Washington, D.C., 23-26 May 2010.

Conover, M.D., Ratkiewicz, J., Fransisco, M., Goncalves, B., Menczer, F., and Flammini, A. 'Political polarization on Twitter', paper presented at the *Proceedings Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 17-21 July 2011.

Ellison, N.B., Steinfield, C., and Lampe, C. 'Connection strategies: Social capital implications of Facebook-enabled communication practices', *New Media & Society*, vol. 13 no. 6, 2011: 873-892.

Ertzscheid, O. 'L'homme est un document comme les autres: du World Wide Web au World Life Web', *Hermès*, no. 53, 2009: 33-40.

Jouët, J. 'Des usages de la télématique aux Internet Studies' in *Communiquer à l'ère numérique: Regards croisés sur la sociologie des usages*, edited by J. Denouel and F. Granjon. Paris: Presses des Mines, 2011, pp. 45-90.

Gerlitz, C. and Rieder, B. 'Mining One Percent of Twitter: Collections, Baselines, Sampling', *M/C Journal*, vol. 16, no. 2, 2012: http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/620.

Hagel J. and Singer, M. *Net worth: Shaping markets when customers make the rules*. Boston: Harvard Business School Press, 1999.

Hermida, A. 'From TV to Twitter: How ambient news became ambient journalism', *M/C Journal*, vol. 13, no. 2, 2010: http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/220.

Knauf, A. and David, A. 'Vers une meilleure caractérisation des rôles et compétences de l'infomédiaire dans le processus d'intelligence économique', paper presented at the *Veille Stratégique, Scientifique et Technologique* – VSST'04 Conference, Toulouse, France, 2004.

Lovink, G. *Networks without a cause: A critique of social media*. Cambridge, Malden: Polity, 2012.

Lotan, G., Graeff, E., Annany, M., Gaffney, D., Pearce, I., and Boyd, D. 'The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions', *International Journal of Communication*, vol. 5, 2011: 1375-1405.

Lindgren, S. and Lundström R. 'Pirate culture and hacktivist mobilization: The cultural and social protocols of #WikiLeaks on Twitter', *New Media & Society*, vol. 13, no. 6, 2011: 999-1018.

Marty E., Rebillad, F., Pouchot S., and Lafouge, T. 'Diversité et concentration de l'information sur le web : Une analyse à grande échelle des sites d'actualité français', *Réseaux*, no. 176, 2012: 27-72.

Marwick, A. E. and Boyd, D. 'I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience', *New Media & Society*, vol. 13, no 1, 2011: 114-133.

Pew Research Center's Project for Excellence in Journalism, 2011, Navigating News Online, http://www.journalism.org/analysis_report/navigating_news_online.

Pew Research Center's Project for Excellence in Journalism, 2012, The State of the News Media, http://stateofthemedia.org.

Poell, T and Borra, E. 'Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto G20 protests', *Journalism*, vol. 13, no. 6, 2012: 695-713.

Rieder, B. 'The refraction chamber: Twitter as sphere and network', *First Monday*, Volume 17, Number 11, 2012: http://firstmonday.org/ojs/index.php/fm/article/view/4199/3359.

Rogers, R. *The end of the virtual: Digital methods*. Amsterdam: Amsterdam University Press, 2009.

Smyrnaios, N. 'How does news infomediation operate online? The examples of Google and Facebook', paper presented at the 10th World Media Economics & Management Conference, Thessaloniki, Greece, 23-27 May 2012.

Wu, S., Hofman, J. M., Mason, W. A., and Watts, D. J. 'Who says what to whom on Twitter', paper presented at the 20th International World Wide Web Conference, ACM, Hyderabad, India, 2011.

Yang, J., and Leskovec, J. 'Patterns of temporal variation in online media', paper presented at the ACM International Conference on Web Search and Data Mining (WSDM), Hong Kong, China, 9-12 February 2011.

About the authors

Dr. Nikos Smyrnaios is a Senior Lecturer at the University of Toulouse and a member of the Laboratory of Applied Research and Studies in Social Sciences (LERASS). His publications focus on the socio-economic stakes and political issues of the Internet. His research fields include online journalism and media, intellectual property & culture in the digital era, as well as social networking sites.

Dr. Bernhard Rieder is Associate Professor of New Media at the University of Amsterdam and a collaborator with the Digital Methods Initiative. His research focuses on the analysis, development, and application of digital research methods, as well as on the history, theory, and politics of software, in particular on the role algorithms play in social processes and in the production of knowledge and culture. He currently participates in the EMAPS project, an EU-funded study of the applications of electronic mapping, led by Prof. Bruno Latour. He is also working on a long-term investigation into the historical and conceptual foundations of information-processing techniques that mine, sort, filter, and connect information on the web.