



The changing public sphere on Twitter: Network structure, elites and topics of the #righttobeforgotten

new media & society

1–20

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1461444816651409

nms.sagepub.com



Shuzhe Yang

Goethe University Frankfurt, Germany

Anabel Quan-Haase

Western University, Canada

Kai Rannenberg

Goethe University Frankfurt, Germany

Abstract

Since the ruling of the European Court of Justice, the right to be forgotten has provided more informational self-determination to users, whilst raising new questions around Google's role as arbiter of online content and the power to rewrite history. We investigated the debate that unfolded on Twitter around the #righttobeforgotten through social network analysis. The results revealed that latent topics, namely Google's role as authority, alternated in popularity with rising and fading flare topics. The public sphere, or *Öffentlichkeit*, that we observed resembles the traditional one, with elite players such as news portals, experts and corporations participating, but it also differs significantly in terms of the underlying mechanisms and means of information diffusion. Experts are critical to comment, relay and make sense of information. We discuss the implications for theories of the public sphere and examine why social media do not serve as a democratising tool for ordinary citizens.

Keywords

European Union, Google, right to be forgotten, social network analysis, Twitter

Corresponding author:

Shuzhe Yang, Chair of Mobile Business & Multilateral Security, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany.

Email: shuzhe.yang@m-chair.de

Introduction

Every day millions of individuals worldwide produce and upload user-generated content in the form of comments, likes, images and posts. For most users, the creation of content and the engagement with others' content are a fun and rewarding routine that is integrated into their everyday rhythms and practices (Hogan and Quan-Haase, 2010). Third parties also generate copious amounts of data on users such as automatic credit ratings and location data. While users are often oblivious to how and where personal information is stored, the data can have significant consequences for an individual's reputation and career, as they may have uploaded unflattering or compromising images or content (Yang and Albers, 2013).

The European Union (EU) has recognised the potential threat of data disclosure to an individual's reputation and the need to provide users with the right to informational self-determination, that is, the right of the individual to decide about what personal information should be public and under what circumstances (Boguslaw and Westin, 1968). In our hyperconnected society (Quan-Haase and Wellman, 2005), this right is frequently violated because information spreads easily and quickly at little cost (Yang, 2015). As a countermeasure, the European Court of Justice (ECJ, 2014) introduced a new ruling on 13 May 2014, the right to be forgotten (RTBF), which allows citizens to request personal information to be removed from a search engine's result page.¹ This decision is transformative in terms of how data ownership and stewardship are understood, and as such it has received much scholarly attention (e.g. Gilbert, 2015; Shoor, 2014). However, there is a gap in the literature in terms of understanding how citizens, the media and key stakeholders discuss, interpret and make sense of the ruling. Understanding how this debate takes place in the public sphere is particularly relevant for future developments of the RTBF, as public opinion is one central mechanism for gaining knowledge about the influence of legislation on people's digital lives and can serve as a catalyst for future revisions to the law.

This article focuses on the debate that took place on Twitter a year after the RTBF ruling. Much debate has focussed on how social media open new opportunities for reviving and revitalising the public sphere, not as a single, one-dimensional space, but rather as 'a more complex system of distinct and diverse, yet interconnected and overlapping, publics [...] which represent different topics and approaches to mediated communication' (Bruns and Highfield, 2016: 124). We will examine how Twitter could function as part of this alternative form of public sphere in which information, ideas and debates can diffuse fast and in an unfettered way (Colleoni et al., 2014). The study gathered real-time data from Twitter users to learn about attitudes and opinions surrounding the debate and key topics that emerged. We aim to reveal public concerns and provide knowledge that could help lawmakers to improve the legal framework to empower citizens and to help companies in their implementation of the RTBF. The following three research questions guide the study:

1. What are the central topics and sub-topics discussed by Twitter users employing the #rtbf? How do central topics and sub-topics relate to one another?
2. Who are the key players in the RTBF debate?

3. What can we learn about the nature of the public sphere on Twitter from analysing tweets related to the RTBF?

Related work

Mayer-Schönberger's (2009) work has sparked an intense debate on the potential harm of compromising data on an individual's reputation and the need for informational self-determination. In his view, data need to have an expiration date – once the data have reached their shelf life, they should be deleted automatically. In this context, the RTBF's role is to enable users to control their personal history to achieve so-called social forgiveness (e.g. Korenhof et al., 2014; Mayer-Schönberger, 2009) and allow them to reinvent their digital persona (Novotny and Spiekermann, 2014).

The European Commission (2012) integrated part of Mayer-Schönberger's ideas into its 2012 draft of the General Data Protection Regulation (GDPR) to aid individuals in enforcing their informational self-determination. The European Commission did not integrate an expiration date for data, but it did give individuals the right to request that their personal information be deleted. While the GDPR is based on Mayer-Schönberger's ideas, it is quite different in that it is a 'control right' meant to give users control over their personal data and further extend this right to third parties such as Google. The RTBF applies not only to search engine providers, but also to 'other intermediaries [...] whenever the conditions established in the ruling are met' (Art. 29 WP, 2014: 8). So far, the definition of 'other intermediaries' remains unclear and unspecified (Gilbert, 2015). The EU institutions published the GDPR on 4 May 2016, explicitly mentioning the RTBF, which in the GDPR is combined with the right to erasure (European Parliament, 2016). In the GDPR, the EU institutions decided not only to integrate the RTBF as per the previous ECoJ decision but also to extend its scope by allowing individuals and companies to request the erasure of their personal data from all websites and other storage locations (European Parliament, 2016, Article 17, 2). The GDPR necessitates data processors and controllers to 'take reasonable steps, including technical measures, to inform [other] controllers which are processing the data, that the data subject has requested the erasure by such controllers of any links to, or copies or replications of that personal data' (European Parliament, 2016, Article 17, 2). This article focuses on the RTBF decision as outlined by the ECoJ, as this is as of today the only legally binding ruling. The ECoJ's RTBF decision and the RTBF in the GDPR have similar aims. Therefore, we follow the EU's understanding of the RTBF, which sees it as a person's right to 'obtain from the controller the erasure of personal data concerning him or her without undue delay' (European Parliament, 2016, Article 17, 1).

Mayer-Schönberger's (2009) arguments and the GDPR draft have received support among the research community. Castellano (2012), in his analysis of existing data protection regulations in the EU, concluded that updated laws were needed in the context of Web 2.0. In this environment, the traditional means of control of content, via gatekeepers such as news agencies, broadcast corporations and standards of journalism, cease to exist. Ausloos (2012) analysed the benefits and disadvantages of the RTBF for individuals and proposed a version of the RTBF that strikes a balance between individuals' right to informational self-determination and freedom of speech for the public's interest.

Shoor (2014) investigated non-legislative approaches to privacy and data protection and concluded that the RTBF is probably the most suitable solution because other technical solutions are hard to implement due to the decentralised nature of the Internet. Ambrose (2012) suggested adding a time component to information, a similar argument to Mayer-Schönberger's (2009), and thereby introduced the notion of a data lifecycle. Other researchers have been more critical of the RTBF and have suggested examining the risks it presents to society, instead of focussing only on its benefits. Rosen (2012) argued that the RTBF can be a double-edged sword: either an effective tool that empowers users and helps them to enforce their right to informational self-determination or the single biggest threat to the Internet, as it has the potential to undermine freedom of speech. Scholarship has expanded the debate to legal frameworks in other countries (e.g. Japan), and similar concerns to Rosen's have been voiced (e.g. Gilbert, 2015). Overall, the literature suggests that debate needs to continue around the potential benefits of the RTBF as well as its possible detrimental effects on society, as the RTBF could shape the future of the Web.

Twitter activity around #righttobeforgotten

From its early inception, the Internet has been exalted for its potential to foster political debate and serves as an alternative and more inclusive public sphere (Brundidge, 2010). The public sphere is a critical forum for open debates, as it constitutes 'a constellation of communicative spaces in society that permit the circulation of information, ideas, debates – ideally in an unfettered manner – and also the formation of political will (i.e. public opinion)' (Dahlgren, 2005: 148). In the context of political discussion, social media platforms are seen as particularly well suited for fostering open debate due to their high level of interactivity, fast diffusion of information and capability for facilitating discussions in real time without geographical constraints (Colleoni et al., 2014). Among social media platforms, according to Colleoni et al. (2014), Twitter is perceived in the domain of political communication as a central space for the exchange of information and the debate of current topics because of four unique characteristics. First, all tweets are visible to everyone by default, except when a tweet or an account has been set to private. Second, hashtags, as a key feature of Twitter, facilitate discussions around specific topics without explicitly creating bounded groups. This feature also enables users to follow a debate anonymously without being noticed, that is, users can lurk. Third, Twitter's retweet function allows information to be diffused quickly and effortlessly. Finally, and important in the context of political communication, Twitter, in comparison to other sites like Facebook, does not require reciprocal relationships among users, allowing for the easy emergence of central nodes of information diffusion.

To address the three research questions guiding this study, we collected real-time data from Twitter over a 2-month period. Anderson (2008) has claimed that 'with enough data, the numbers speak for themselves. [...] There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?' Despite this, we did not rely solely on big data for our analysis, we also employed the following three strategies to supplement our data: we examined individual Twitter accounts, we drew on mainstream news reporting to gain an understanding of the social and informational contexts of the debate

and we looked for background information in the data that could provide context for the findings and trends. Our approach, thus, relied on large amounts of data, but was not blind to the nuances available in these data and their meaning. In particular, we followed Monroe et al. (2015) and recognised the value of big data analytics in the following areas:

1. Helping scientists to design better experiments for testing hypotheses that could not be examined without large amounts of data;
2. Making more precise comparisons between two or more sub-populations by providing fine-grained samples;
3. Identifying behaviour that could not be observed before by uncovering behaviours that are otherwise actively hidden by individuals or organisations (e.g. censorship).

The study relied on Netlytic² to collect data from Twitter. Netlytic is a cloud-based research tool designed to collect, manage and analyse data from social media platforms. Its key strengths are that it allows scholars to explore emerging themes and trends and facilitates the examination of interactions among Twitter accounts as they take place in real time (Gruzd et al., forthcoming). In Netlytic, researchers can control what data are collected by adjusting and modifying keywords and hashtags as needed. It also provides flexibility for data cleaning by filtering irrelevant or duplicate data and includes features that allow scholars to drill down and examine individual records and investigate how a particular record was used to create the various data visualisations (Gruzd et al., forthcoming). Hence, investigations of data can occur at various scales, which is important for sensemaking.

As discussed above, on 13 May 2014, the ECoJ ruled against Google in the RTBF case. Our data collection started about a year later and took place from 20 May to 18 July 2015. As a result of the late start date, the initial Twitter exchanges following the RTBF decision were not included – exactly 12 months and 7 days of data are missing. Future research can attempt to purchase this dataset from such providers as Gnip and expand the present investigation. While the initial tweets are absent, past research by Yang (2015, 2016) suggests that a time lag can be beneficial to capture engagement by a broader public instead of limiting the investigation to the initial reaction. Hence, two important considerations related to the time frame of data collection, and which help put our findings in context, are as follows:

1. As the data collection was started 1 year after the ECoJ's decision, we are looking at the debate once it has been established and consolidated, rather than at the initial flurry of activity. This means that opinions are more thought-through and issues have been considered from different points of view.
2. Furthermore, due to the elapsed time, we hoped that information regarding the decision had reached a broader audience and that this audience would have had the chance to consider the options available to search engine providers as to how to practically implement the RTBF ruling and the consequences for users.

The end date of data collection was determined by volume and the fact that the discussion had started to dwindle. After the peak of the discussion was reached in the fourth week with 9954 tweets, only 1099 tweets were gathered in the last week.

Our initial data collection focussed on tweets using the keyword 'right to be forgotten' and the hashtag #rtbf. After 10 days, we examined the preliminary data and determined that two additional hashtags – #2beforgotten and #righttobeforgotten – were often utilised as well. We added these to retrieve as comprehensive a dataset as possible. Twitter hashtags evolve organically as users interact with one another, and hence multiple hashtags can emerge around a single topic (Conover et al., 2011). When we manually scrutinised the data for tweets unrelated to the RTBF discussion, we realised that the hashtag #rtbf collected a lot of irrelevant tweets because it was used not only for discussing the RTBF ruling but also in relation to a Belgian television company that broadcasted the Eurovision Song Contest during the same time frame. Consequently, we excluded tweets from the data collection which included any of the following hashtags or keywords, which were frequently used simultaneously with #rtbf: #eurovision, #eurovisionrtbf, #eurovision2015 and 'eurovision'. Tweets in many different languages were collected, including German, Russian and Korean, with English being the most prevalent language.

In total 30,894 relevant tweets produced by 18,959 user accounts were gathered in the specified time frame that provided insight into how the discussion unfolded and how various groups and individuals engaged with the topic. Looking at tweets over time was critical to the aim of this study, as it allowed us to observe the development of trends in the corpus, the emergence of new sub-topics and how the community evolved (boyd and Crawford, 2012). Additionally, the amount of collected data from Twitter provided a broader data foundation than other data collection methods, such as surveys, experiments or interviews. A key advantage of drawing data from Twitter was the large and diverse sample size encompassing perspectives, opinions and interactions of individuals from different countries, cultural backgrounds, socioeconomics and, most importantly, for this specific study, legal frameworks (boyd and Crawford, 2012). In contrast to other means of data collection, where individuals are aware of being part of a study, data gathered from Twitter provide insight into individuals' perspectives and opinions on topics, in this case the RTBF, in a more naturalistic setting; at least one that is not influenced by reactivity to the study (Ansolabehere and Hersh, 2012). To protect the privacy of users, we anonymised the data.

Using social network analysis to analyse Twitter data

We employed social network analysis (SNA) to analyse and visualise the Twitter data. SNA is not only an important theoretical lens but also a methodological approach for investigating connections and the flow of information in social media, as it generates rich and complex data as users discuss, debate and exchange their opinions in real time (Quan-Haase and McCay-Peet, forthcoming). SNA was ideal for this study because it allowed us to identify central actors as well as examine the overall structural features of the network. We used the following two types of open-source software: OpenRefine³ and Gephi.⁴ OpenRefine was used to process and prepare data for further analysis in Gephi (e.g. extract mentions and hashtags into separate columns, create a node and an edge table). We used Gephi for visualising the results and its integrated features for calculating network structural measures.

Table 1. Timeline of important news published before and after data collection.

Date and news	Discussed topics
4 July 2014: Legal experts discuss whether Google violated EU Data Protection Directive. (source: <i>The Register</i>)	<ul style="list-style-type: none"> • Google's effort to comply with the RTBF. • Google's role as decision maker with regard to 'privacy protection for the individual vs. public interest'.
13 May 2015: <i>WSJ digits</i> report that France wants Google to expand the scope of the RTBF. (source: <i>WSJ digit</i>)	<ul style="list-style-type: none"> • Whether Google's 'privacy-protection vs. public interest' decisions are correct. • Reviewing what decision has been reached 1 year after the RTBF decision. • Reviewing the current situation. • What to do in the future to enhance the RTBF.
21 September 2015: French regulators reject Google's RTBF appeal. (source: <i>The Guardian</i>)	<ul style="list-style-type: none"> • Google's appeal against the expansion of the RTBF to global scale has been rejected by Commission Nationale de l'Informatique et des Libertés (CNIL). • CNIL will likely begin to apply sanctions against Google.
4 May 2016: EU institutions published the GDPR, which integrates the RTBF in Article 17. (source: European Parliament, 2016)	<ul style="list-style-type: none"> • Extension of the scope of the RTBF to remove duplicates and copies of data. • Extension of the RTBF to a global scale.

EU: European Union; RTBF: right to be forgotten; *WSJ*: *The Wall Street Journal*; GDPR: General Data Protection Regulation.

Central keywords

To provide some context for interpreting the Twitter data, Table 1 lists some of the key pieces of news published prior to and following the data collection period. The British technology news website *The Register* reported on 4 July 2014 that legal experts were discussing Google's role as sole decision maker with regards to whether information requested for deletion falls under an individual's right to privacy or if the information is considered of public interest. On 13 May 2015, *Wall Street Journal Digits* reported on France's intention to request that Google extend globally the RTBF's scope.

The impact of the two news reports on the RTBF debate on Twitter is visible in the word cloud depicted in Figure 1.⁵ Size reflects the keyword's ranking, with bigger words occurring more often in the dataset. The number in the right upper corner shows the absolute frequency of each keyword.

The keywords 'France', 'requests' and, in particular, 'Google' were the most frequently used, with 4905, 5087 and 20,064 mentions, respectively. There were three reasons why 'Google' and #Google were so frequently mentioned in tweets related to the RTBF. First, due to Google's predominance, it is seen as a synonym for conducting web searches, reflected in the verb 'to google'. Second, the RTBF decision was specifically directed at Google because it is the number one search engine in Europe and has the largest market share worldwide (Statista, 2015). Google's share in Europe is 90% in comparison to Bing and Yahoo with 2.7% and 2.3%, respectively (Rosoff, 2014). Finally, the RTBF could impact people's access to information and thus threaten freedom of speech

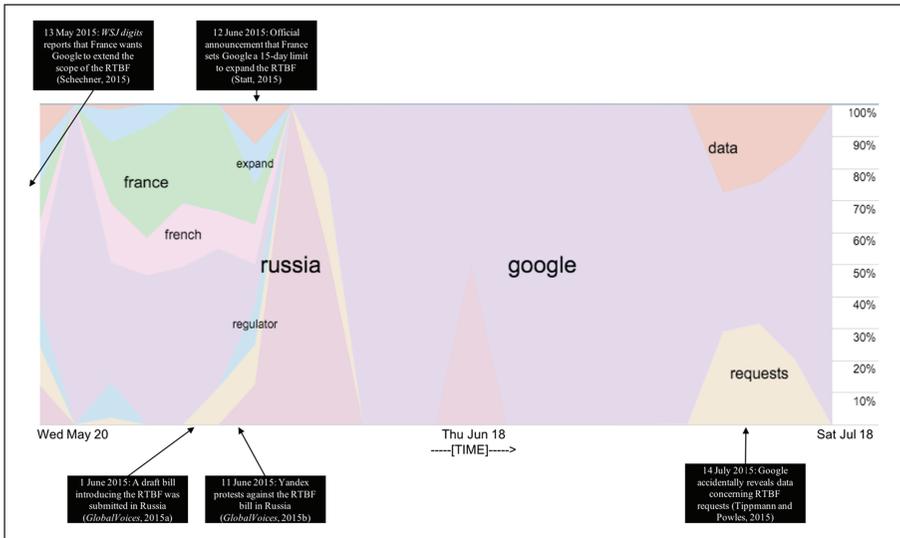


Figure 2. Thematic relevance and news.

Source: Netlytic.

of mentions of the words quickly dropped. At this point, the attention focussed on the keyword ‘Russia’, which dominated the discussion for a few days after a new bill that introduced the RTBF in that country (GlobalVoices, 2015a). The RTBF as introduced by the ECoJ prohibits public figures such as politicians from removing information deemed of public interest, whereas the Russian ruling extends its scope to allow this (Shaftan, 2015).

On Twitter, the Russian ruling sparked much debate. Discussions focussed around how public figures could abuse the law by removing compromising reports to enhance their online reputation. In this case, the RTBF would function as a mechanism to restrict the right to free speech instead of its originally intended purpose (e.g. Rosen, 2012). Within the ECoJ (2014) ruling, public figures cannot remove information that is deemed of public relevance, but can ask for information to be removed that applies to their private lives. However, it is hard to find a balance between freedom of speech and an individual’s right to privacy because there is no clear line as to where public life ends and private life starts. Russia’s interpretation and version of the RTBF circumvented this problem by extending the scope to all individuals in society including public figures (Doyle, 2015). The ruling can potentially also empower politicians by allowing them to increase their informational self-determination beyond the private sphere.

Looking at a snapshot of the RTBF debate, Figure 3 shows a rapid increase in posts around 13 June 2015 resulting from the news regarding the Russian government’s interpretation of the RTBF (GlobalVoices, 2015b). The radically different interpretations of the RTBF decision have the potential to polarise future debates on attaining the right balance between freedom of speech and privacy rights, particularly in relation to its application to public figures such as politicians, celebrities or athletes.

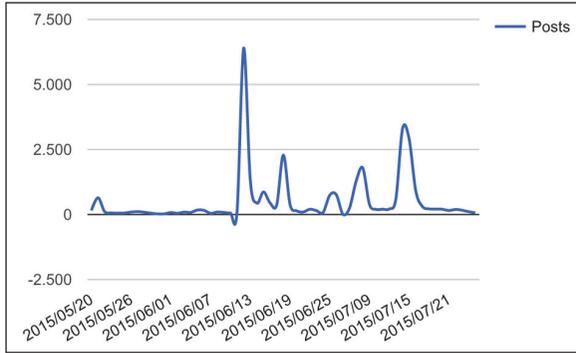


Figure 3. Peak around the Russian right to be forgotten (RTBF) debate.
Source: Netlytic.

Figure 2 shows that the keywords ‘Google’ and ‘Russia’ were inversely related, that is, when ‘Russia’ emerged as a central keyword, the debate no longer included the keyword ‘Google’ and vice versa. This could potentially be related to the fact that Google’s market share and popularity is low in Russia, where other search engine providers, such as Yandex, dominate the search engine market (Clayton, 2013). Therefore, concerns about how Google operates and how it will implement the RTBF ruling are less relevant to the discussion surrounding the RTBF in Russia.

Influencers in #righttobeforgotten

How influence occurs on social media is a central topic. Influencers can be defined in terms of their centrality in a network – major influencers are referred to as stars (Heider, 1958) or opinion leaders (Merton, 1968) – or in terms of their ability to move information between otherwise disconnected networks; major players in this category are referred to as gatekeepers (Rogers, 1962), cosmopolitans (Merton, 1968) or bridging actors (Burt, 1992). We first look at central individuals in terms of volume of contributions and network centrality and then discuss those players who control the flow of information and help the network remain interconnected.

Figure 4 shows that the top 10 posters comprised nine user accounts representing corporations or news outlets. Only one personal account (@Account_A) was among the top 10; this user was very engaged and posted 33% of all posts. From the user’s account description, it was evident that the user worked in the communications industry as the head of communications of a non-profit organisation. Postings may have been motivated both by the user’s employment and personal interest.

The news account @Account_K was the most frequently mentioned account, with over 450 mentions (see Figure 5). In contrast to the top 10 posters, the top 10 accounts by mentions contained three personal accounts:

1. A British Broadcasting Corporation (BBC) North America Technology reporter (Account_M),

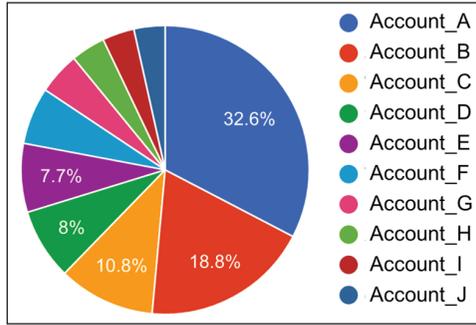


Figure 4. Top 10 accounts by posts in the right to be forgotten (IRTBF) network. Source: Netlytic.

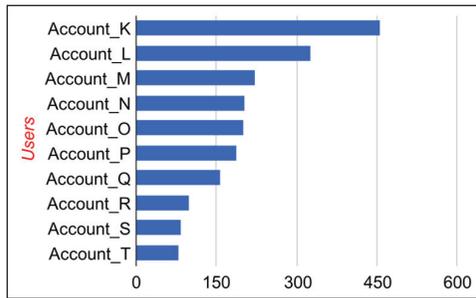


Figure 5. Top 10 accounts by mentions in the right to be forgotten (RTBF) network. Source: Netlytic.

2. A lawyer and technology researcher from Cambridge University (*Account_Q*) and
3. A law lecturer, who focuses on privacy and human rights (*Account_S*).

The frequent mention was due to each user’s knowledge and prominence in the areas of privacy and law. While all three were not among the top posters, they regularly posted tweets with their opinions, links to resources and other relevant posts. News portals focussed on tech-related topics, such as *@Account_K*, were often mentioned in tweets, and their content was frequently retweeted because they regularly posted updates and comments relevant to the operations of search engines. *@Account_L* (Google’s company account) was also frequently mentioned in tweets because the RTBF decision affected its operations the most (ECoJ, 2014), but *@Account_L* did actually not post content nor participate in the debate.

Network structure of #righttobeforgotten

The network structure indicated a sparsely knit and decentralised network consisting of 9192 separate clusters, which is a large number of clusters for a network consisting of

Table 2. Centrality measurements of the top five bridging actors.

Account	In-degree	Out-degree	Betweenness score
@Account_Q	174	44	69,805.89
@Account_V	26	10	40,999.06
@Account_X	35	21	39,618.94
@Account_U	9	25	31,546.89
@Account_A	9	448	22,874.54

18,959 nodes. The average cluster coefficient was .04, which means that a few nodes already formed a cluster (Cherven, 2015). That is, each cluster is rather small in relation to the overall network size. The modularity of the network was .876, which means strong interconnections among nodes within a cluster but sparse interconnectivity to other clusters (Newman, 2006). There were few connections among nodes, with the network density value being .0000377179 (Coleman and Moré, 1983). To better visualise and examine the topological features of the network, we relied on the ‘giant component’ filter. This filter allowed us to focus on large connected clusters by fading out smaller clusters (Cherven, 2015) and resulted in 102 clusters, each with many bridging nodes facilitating the flow of information.

Betweenness centrality measures how often a node is on the shortest path between two nodes and thereby indicates how often a node connects with otherwise disconnected clusters (Cherven, 2015). Five nodes serve as bridges with high betweenness scores connecting various clusters (for their individual betweenness scores, see Table 2). One of the bridges, @Account_X, is run by the Stanford Center on Democracy, Development and the Rule of Law. The other four accounts are all personal accounts: @Account_Q is a lawyer and technology researcher at Cambridge University, @Account_V is the personal account of Organization for Security and Co-operation in Europe (OSCE) representative on Freedom of the Media, @Account_U is a senior lecturer at the School of Computing at the University of Kent and @Account_A is an Italian cyber-utopian. Figure 6 illustrates the clusters and functions of bridging actors. The size of the account labels represents the size of the betweenness score of the node.

Table 2 depicts the betweenness scores, in- and out-degree centrality for the top five bridging actors. Taking into account that 18,959 nodes comprised the overall network of tweets exchanged in the time frame, the influence of the five bridging actors cannot be ignored. Moreover, their level of engagement was also visible in that they were exchanging information not only with multiple clusters but also with each other.

@Account_Q and @Account_X have the largest in-degree values with 174 and 35, respectively. @Account_A and @Account_Q have the largest out-degree values with 448 and 44, respectively. Considering that the average degree centrality in our sample is .639, it is apparent how influential and significant these top five bridging actors are in the network. We can conclude that @Account_A is an information hub, with high out-degree value, that distributes information about the RTBF, while @Account_Q can be considered a core bridging actor that has the greatest influence on the RTBF Twitter network by connecting otherwise isolated clusters. The betweenness scores (see Table 2) show that

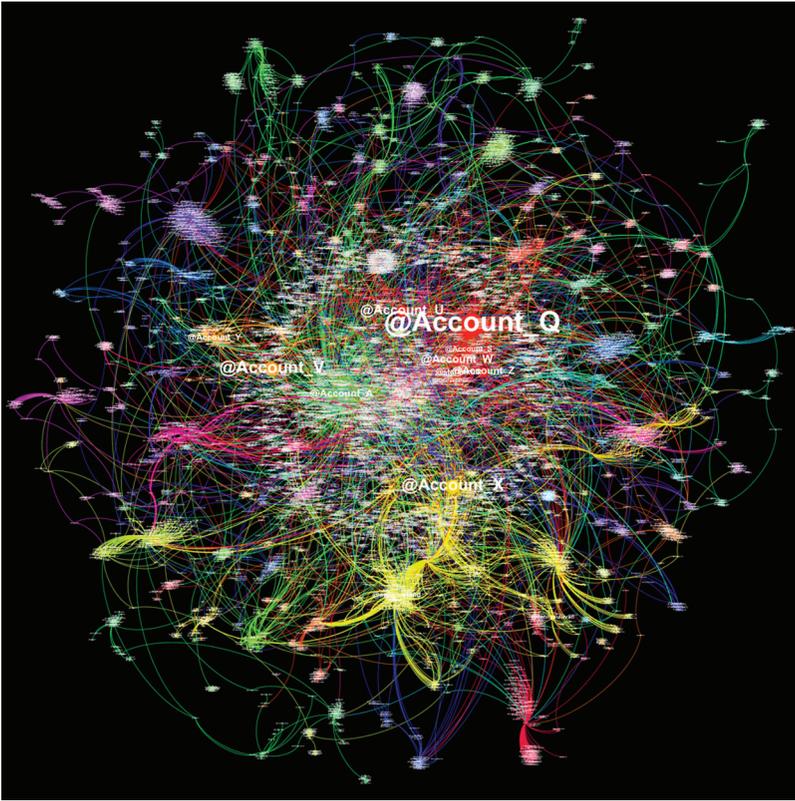


Figure 6. Clusters and bridges in the right to be forgotten (RTBF) network.
Source: Gephi.

the top five bridging actors appear at least 22,874 times on the shortest path between two nodes, whereas *Account_Q* as the most influencing node appears 69,805 times on the shortest path. Considering that the total number of shortest paths in the network is 155,571, the influence of the bridging actors becomes apparent. The function of information hubs and bridging actors is critical for keeping the discussion alive and the network interconnected.

Discussion

Dynamics of latent and flare topics

Research question (RQ) 1 examined the various topics discussed on Twitter concerning the RTBF. One central topic spanned the entire period and prevailed throughout the debate, namely Google's role in the RTBF. We refer to this as a *latent topic*, as it was pervasive and dominant. Identifying the latent topic was critical, as the focus on a single topic kept the discussion alive and functioned as a recurring theme in the narrative. While

it is true that alternative search engines are less central due to their low market share, they are still key players in many regions of the world. This demonstrates how Google's monopoly plays out with regards to debates on Twitter; the brand becomes accountable for its practices, as the entire debate has been funnelled towards its operations.

Several sub-topics emerged and flooded the Twitter stream for 3–5 days at a time. We refer to these topics as *flare topics* because of their sudden and intense appearance followed by a quick fading from the discourse. These flare topics were often linked to news originating in the mainstream media and would spark heated, topic-specific debates. They would quickly reach a peak and then slowly flatten and completely vanish; this is when the latent topic would return into focus and sustain the debate.

We also observed that these flare topics could completely shift the focus of the discussion. For example, when Russia became a key topic, the previously ongoing discussion around Google and France came to a sudden halt. This suggests that discussions evolving around the RTBF were guided primarily by one dominant latent topic, but that they received new impetus for further discussion from multiple flare topics. Furthermore, it seems that there is currently no flare topic that can substitute for the focus on Google and take the position of a latent topic, although the RTBF decision also involved other search engines and equally affected their operations. These findings could be used to anticipate future developments concerning the RTBF ruling in the Twittersphere.

Key players in the network

RQ2 examined the role of influencers in the RTBF debate. Three types were identified:

1. *Companies*. These represent corporate interests and are affected by the RTBF decision directly (e.g. Google, which is frequently mentioned by others, but does not participate in the debate);
2. *Experts*. These are individuals who are experts (e.g. privacy researchers and lawyers) and contribute with their opinions on and understandings of the decision;
3. *News portals*. This includes varied broadcasters, who are interested in spreading news and engaging a wide audience in an issue of broad public interest (e.g. searchengineland.com).

When looking at actors that serve as bridges between clusters, it was observed that four of five influencers comprised personal accounts belonging to experts. These have the greatest influence on the diffusion of RTBF-related information and their role is also to sustain the network over time. This suggests that the majority of influencers in the network continue to be the gatekeepers and agenda-setting organisations that structured and controlled information in the public sphere in the past.

Twitter as a 21st-century public sphere

RQ3 examined the nature of the public sphere on Twitter. The public sphere that emerged around the RTBF debate on Twitter resembles, on the one hand, the *Öffentlichkeit* as described by Habermas in the 1960s. On the other hand, it differs significantly in terms

of the underlying mechanisms and means of information diffusion. We did not observe Colleoni et al.'s (2014) four characteristics of Twitter in the expected way. Twitter was not able to fulfil its function as a virtual public sphere, as information did not diffuse in an unfettered manner, but rather was controlled by a few key influencers. On Twitter, we continue to see this elite being highly involved and central to the RTBF debate. In particular, news outlets – through editors, newsroom staff and broadcasters – continue to have an agenda-setting function, often being the first to tweet on RTBF news. Their role is critical as they drive the dynamics of networked communication by filtering, making sense of and selecting information and topics of interest to the public (Habermas, 2006; Webster, 2006). The mass media's role in the debate concerning the RTBF is to continuously provide updates and developments.

How agenda setting occurs on Twitter is neither a simple nor a straightforward process, as topics shift quickly and dynamically. We observed that experts directly participate, shape and drive discussions, rather than the mass media serving as an intermediary. In this way, there is much more of a plurality of opinions on Twitter than was ever possible in the traditional public sphere. The position of the traditional mass media has weakened, and social media have emerged as new spaces where experts have the potential to actively participate and steer the discussion (Meraz, 2009).

However, a central question emerges from our study: why are ordinary citizens unengaged in the RTBF debate on Twitter? As mentioned previously, the Internet is often considered as a new public sphere because of its potential to foster political discussions (Brundidge, 2010) and citizen engagement (e.g. Bruns and Highfield, 2016). In the case of the RTBF, however, the results of our study show the opposite, as ordinary citizens do not (actively) participate. There might be several reasons: a space dominated by experts may seem intimidating to ordinary citizens. Twitteracy (a combination of *Twitter* and *literacy*) has been described as an acquired skill (Greenhow and Gleason, 2012), and there may be a barrier to participation in a topic that is highly driven by legal institutions and regulatory bodies. Ordinary citizens may be 'lurkers' or passive information consumers, that is, their participation consists primarily of following debates, retweeting content and observing developments, rather than voicing their opinions. If this is true, it would mean that Twitter's features – in particular, hashtags, which allow users to follow a debate anonymously (Colleoni et al., 2014) – are not always helpful in terms of establishing a virtual public sphere and can actually have the opposite effect. In the case of the RTBF, hashtags do not motivate ordinary citizens to participate; rather, they promote a passive audience. The lack of participation may also be due to ordinary citizens' sense of inexperience or ignorance with regard to the topic of the RTBF (Bishop, 2007). In a space dominated by the elite, ordinary citizens may think that 'they did not need to post [...] and they are being helpful by not posting' (Bishop, 2007: 1889–1890), as they are then not interrupting the discussions moderated by the experts. The lack of engagement may also be related to features specific to Twitter and the sparsely knit network structure it promotes. The high number of cliques and their internal focus may make it difficult for ordinary citizens, who are at the periphery, to penetrate the debate. Ordinary citizens may not be aware of the ongoing debate, as it may take place within contained information bubbles (Pariser, 2011). We describe the Twittersphere as a dynamic space consisting of a multitude of small clusters that function largely independently and in a decentralised fashion.

Our analysis of influencers shows that important players are individuals who are experts in the area. These key players, who could potentially reach out to ordinary citizens and engage them in the debate, have only a limited influence on Twitter, as they have fewer than 3000 followers (except for *@Account_X*, which is a university institution and has 40,000 followers). This limits their role as bridges to the wider public and leads to the debates around the RTBF remaining contained between news outlets, elites and organisations. One way to break this kind of information bubble and engage citizens could be to redesign the personalised recommendation algorithm (Garrett and Resnick, 2011). It could recommend RTBF-related tweets to users who are interested in privacy and data protection. Also, celebrity endorsement could help to promote the relevance of the topic. Such personalised recommendations would not only broaden people's knowledge but also help them engage more meaningfully with current debates that affect them directly.

Study limitations and future work

A total of 30,894 tweets were collected during the 2-month data collection period. Other studies have gathered larger data sets; for example, Conover et al. (2011) scraped over 250,000 relevant tweets within a 6-week period prior to the 2010 US midterm elections. The elapsed time between the RTBF ruling and the start of our data collection, a 1-year gap, may explain the comparatively low numbers. It is also possible that the general public has not as yet joined the debate on Twitter, they may not be fully aware of the RTBF ruling itself, they may not fully comprehend its opaque legal language or they may be unaware of its implications for their own daily web experiences, for example, the possibility of skewed search results (Ambrose, 2012). It is possible that as the debate becomes more widely known among Internet users, greater engagement will occur on Twitter and other social media platforms. Compared to the decision of the ECoJ, the GDPR extends the scope of the RTBF and could spark further debate. Future research could document how the debate shifts and what new sub-topics emerge on Twitter and other platforms. The results of this study could then be useful for anticipating future discussion structures on related topics.

Another limitation of the study is its exclusive focus on Twitter. While the Twittersphere has been deemed an important space for public debate and an extension of more traditional public spheres, studies that exclusively focus on Twitter may have limited generalizability. Discussions in personal circles, social spaces or on other social media platforms such as Facebook are equally valid and important, but were not included in this study. Another limitation comes from our lack of understanding of our sampling frame. Twitter users are not representative of the general population – users have tended to be younger and more tech-savvy – and consequently do not include voices from all strata of society (Quan-Haase and McCay-Peet, forthcoming).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a Social Science and Humanities Research Council of Canada (SSRHC) grant given to A. Quan-Haase (R3603A13) and by the Faculty of Economics and Business Administration at the Goethe University Frankfurt. We would like to thank colleagues at the Faculty of Information and Media Studies, Western University, for their input and particularly to Chandell Gosse and David Brown.

Notes

1. Detailed information: http://ec.europa.eu/justice/data-protection/reform/index_en.htm (accessed 22 November 2015).
2. <https://netlytic.org> (accessed 22 November 2015).
3. <http://openrefine.org> (accessed 22 November 2015).
4. <http://gephi.github.io> (accessed 22 November 2015).
5. Conjunctions were removed (e.g. 'and').

References

- Ambrose ML (2012) It's about time: privacy, information lifecycles, and the right to be forgotten. *Stanford Technology Law Review* 16: 369–422.
- Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete. *Wired*. Available at: http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 3 August 2015).
- Ansolabehere S and Hersh E (2012) Validation: what big data reveal about survey misreporting and the real electorate. *Political Analysis* 20(4): 437–459.
- Art. 29 WP (2014) *Guidelines on the Implementation of the Court of Justice of the European Union Judgment on 'Google Spain and Inc v. agencia española de protección de datos (aepd) and mario costeja gonzález' c-131/12*. Brussels: European Commission.
- Ausloos J (2012) The 'right to be forgotten' – worth remembering? *Computer Law & Security Report* 28(2): 143–152.
- Bishop J (2007) Increasing participation in online communities: a framework for human–computer interaction. *Computers in Human Behavior* 23(4): 1881–1893.
- Boguslaw R and Westin AF (1968) Privacy and freedom. *American Sociological Review* 33(1): 173.
- boyd dm and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Brundidge J (2010) Encountering 'difference' in the contemporary public sphere: the contribution of the internet to the heterogeneity of political discussion networks. *Journal of Communication* 60(4): 680–700.
- Bruns A and Highfield T (2016) Is Habermas on Twitter? Social media and the public sphere. In: Bruns A, Enli G, Skogerbø E, et al. (eds) *The Routledge Companion to Social Media and Politics*. New York: Routledge, pp. 98–130.
- Burt RS (1992) *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
- Castellano PS (2012) The right to be forgotten under European law: a constitutional debate. *Lex Electronica* 16(1): 1.
- Cherven K (2015) *Mastering Gephi Network Visualization*. 1st ed. Birmingham: Packt Publishing.

- Clayton N (2013) Yandex overtakes Bing as world's fourth search engine. *The Wall Street Journal*. Available at: <http://blogs.wsj.com/tech-europe/2013/02/11/yandex-overtakes-bing-as-worlds-fourth-search-engine/> (accessed 24 October 2015).
- Coleman TF and Moré JJ (1983) Estimation of sparse Jacobian matrices and graph coloring blems. *SIAM Journal on Numerical Analysis* 20(1): 187–209.
- Colleoni E, Rozza A and Arvidsson A (2014) Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication* 64(2): 317–332.
- Conover MD, Ratkiewicz J, Francisco M, et al. (2011) Political polarization on Twitter. In: *Proceedings of the fifth international AAAI conference on weblogs and social media (ICWSM)*, Barcelona, Spain, 17–21 July, pp. 89–96. Menlo Park, CA: The AAAI Press.
- Dahlgren P (2005) The internet, public spheres, and political communication: dispersion and deliberation. *Political Communication* 22(2): 147–162.
- Doyle B (2015) Russia's 'right to be forgotten' and China's right to be protected: new privacy and security legislation. *Harvard Journal of Law and Technology Digest*. Available at: <http://jolt.law.harvard.edu/digest/legislation/russias-right-to-be-forgotten-and-chinas-right-to-be-protected-new-privacy-and-security-legislation> (accessed 18 May 2016).
- European Commission (2012) *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*. Brussels: European Commission.
- European Court of Justice (ECoJ) (2014) Case no. c-131/12-Google Spain and Google. Available at: http://curia.europa.eu/juris/document/document_print.jsf?doclang=EN&docid=152065
- European Parliament (2016) *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Brussels, Belgium. Available at: http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf (accessed 18 May 2016).
- Garrett RK and Resnick P (2011) Resisting political fragmentation on the internet. *Daedalus* 140(4): 108–120.
- Gilbert F (2015) The right of erasure or right to be forgotten: what the recent laws, cases, and guidelines mean for global companies. *Journal of Internet Law* 18(8): 1–20.
- GlobalVoices (2015a) How safe are internet search engines from Russian censorship? Available at: <https://advox.globalvoices.org/2015/06/01/how-safe-are-internet-search-engines-from-russian-censorship/> (accessed 17 February 2016).
- GlobalVoices (2015b) Russia moves forward on 'right to be forgotten' bill despite industry protests. Available at: <https://globalvoices.org/2015/06/11/russia-right-to-be-forgotten-industry-protests/> (accessed 26 February 2016).
- Google Inc. (2015) Alphabet code of conduct. Available at: <https://investor.google.com/corporate/code-of-conduct.html> (accessed 31 October 2015).
- Greenhow C and Gleason B (2012) Twitteracy: tweeting as a new literacy practice. *The Educational Forum* 76(4): 464–478.
- Gruzd A, Mai P and Kampen A (forthcoming) A how-to for using Netlytic to collect and analyze social media data: a case study of the Twitter use during the 2014 Euromaidan Revolution in Ukraine. In: Sloan L and Quan-Haase A (eds) *Handbook of Social Media Research Methods*. London: SAGE.
- Habermas J (2006) Political communication in media society: does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory* 16(4): 411–426.

- Heider F (1958) *The Psychology of Interpersonal Relations*. 1st ed. Hillsdale: Lawrence Erlbaum Associates.
- Hogan B and Quan-Haase A (2010) Persistence and change in social media. *Bulletin of Science, Technology & Society* 30(5): 309–315.
- Korenhof P, Ausloos J, Szekely I, et al. (2014) Timing the right to be forgotten: a study into ‘time’ as a factor in deciding about retention or erasure of data. *SSRN Electronic Journal*.
- Mayer-Schönberger V (2009) *Delete: The Virtue of Forgetting in the Digital Age*. 1st ed. Woodstock: Princeton University Press.
- Meraz S (2009) Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication* 14(3): 682–707.
- Merton RK (1968) *Social Theory and Social Structure*. New York: The Free Press.
- Monroe BL, Pan J, Roberts ME, et al. (2015) No! Formal theory, causal inference, and big data are not contradictory trends in political science. *PS: Political Science & Politics* 48(1): 71–74.
- Newman MEJ (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* 103(23): 8577–8582.
- Novotny A and Spiekermann S (2014) Oblivion on the web: an inquiry of user needs. In: *Proceedings of the 22nd European conference on information systems (ECIS)*, Tel Aviv, Israel, 9–11 June.
- Pariser E (2011) *The Filter Bubble: What the Internet Is Hiding from You*. 1st ed. New York: Penguin Press.
- Quan-Haase A and McCay-Peet L (forthcoming) Social network analysis. In: Bruhn KJ and Craig RT (eds) *International Encyclopedia of Communication Theory and Philosophy*. Cambridge, MA: Willey.
- Quan-Haase A and Wellman B (2005) Hyperconnected net work: computer-mediated community in a high-tech organization. In: Heckscher C and Adler P (eds) *Collaborative Community in Business and Society*. 1st ed. New York: Oxford University Press, pp. 281–333.
- Rogers EM (1962) *Diffusion of Innovations*. 1st ed. New York: The Free Press.
- Rosen J (2012) The right to be forgotten. *Stanford Law Review* 64(88): 5.
- Rosoff M (2014) Here’s how dominant Google is in Europe. *Business Insider*. Available at: <http://www.businessinsider.com/heres-how-dominant-google-is-in-europe-2014-11?IR=T> (accessed 2 March 2016).
- Schechner S (2015) Squabbling over Google’s ‘right to be forgotten’ continues one year on. *The Wall Street Journal*. Available at: <http://blogs.wsj.com/digits/2015/05/13/google-regulators-spar-over-applying-forgotten-right-globally/> (accessed 15 February 2016).
- Shaftan V (2015) Russia signs controversial ‘right to be forgotten’ bill into law. *Data Protection Report*. Available at: <http://www.dataprotectionreport.com/2015/07/russia-signs-controversial-right-to-be-forgotten-bill-into-law/> (accessed 31 October 2015).
- Shoor EA (2014) Narrowing the right to be forgotten: why the European Union needs to amend the proposed data protection regulation. *Brooklyn Journal of International Law* 39(1): 487–519.
- Statista (2015) *Worldwide Market Share of Leading Search Engines from January 2010 to July 2015*. Available at <http://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/> (accessed 18 May 2016).
- Statt N (2015) French regulator says Google must expand ‘right to be forgotten’ to all Google sites. *The Verge*, 21 September. Available at: <http://www.theverge.com/2015/9/21/9365075/french-regulator-google-right-to-be-forgotten> (accessed 18 May 2016).
- Tippmann S and Powles J (2015) Google accidentally reveals data on ‘right to be forgotten’ requests. *The Guardian*, 14 July. Available at: <https://www.theguardian.com/technology/2015/jul/14/google-accidentally-reveals-right-to-be-forgotten-requests> (accessed 18 May 2016).
- Webster F (2006) *Theories of the Information Society*. 3rd ed. New York: Routledge.

- Yang S (2015) Why are people so naïve? Long-term motivation in online reputation management: a grounded theory study. In: *Proceedings of the 21st Americas conference on information systems (AMCIS 2015)*, Fajardo, Puerto Rico, 13–15 August.
- Yang S (2016) Understanding the pain: examining individuals' online reputation management behaviour and its obstacles – a grounded theory. In: *Proceedings of the 49th Hawaii international conference on system sciences (HICSS)*, Kauai, HI, USA, 5–8 January, pp. 3898–3907. New York: IEEE.
- Yang S and Albers A (2013) Overcoming information overload in online reputation management: a systematic literature review. In: *Proceedings of the 21st European conference on information systems (ECIS 2013)*, Utrecht, Netherlands, 5–8 June.

Author biographies

Shuzhe Yang is a research and teaching assistant working at the Chair of Mobile Business and Multilateral Security in the Institute for Business Informatics at the Goethe University Frankfurt, Germany. He received his Master of Science degree in Information Management and Finance, and his PhD degree in Economics from the same university. Shuzhe's research interests include personal online reputation management, privacy on social media, online self-tracking, information retrieval and information system design and architectures. Within these topics, his particular focus is on people's behaviour on and outside social media platforms and the factors that influence people's motivation formation.

Anabel Quan-Haase is an associate professor and holds a joint appointment in the Faculty of Information and Media Studies and the Department of Sociology at the University of Western Ontario. She is the Director of the SocioDigital Lab, and her research interests focus on how people integrate social media into their everyday lives and work settings. Her particular focus is on user engagement and the role of social context in how individuals use and make sense of messages and interactions on social media. Dr Quan-Haase is the author of *Technology and Society: Social Networks, Power, and Inequality* (2015, 2nd ed. with Oxford University Press) and *Information Brokering in the High-Tech Industry* (2009 with Lambert). She is the past president of the Canadian Association for Information Science and a Council Member of the Communication, Information Technologies, and Media Sociology (CITAMS) section of the American Sociological Association.

Kai Rannenberg is a full professor and holds the Deutsche Telekom Chair of Mobile Business & Multilateral Security in the Institute for Business Informatics at the Goethe University Frankfurt, Germany. Before, he was with the System Security Group at Microsoft Research Cambridge, United Kingdom, focussing on Personal Security Devices & Privacy Technologies. Since October 2015, Kai is a Vice President of International Federation for Information Processing (IFIP). From May 2007 till July 2013, he chaired IFIP TC-11 Security and Privacy Protection in Information Processing Systems, after having been its Vice-Chair since 2001. From July 2004 till June 2013, Kai served as the academic expert in the Management Board of the European Network and Information Security Agency (ENISA) and is now a member of ENISA's Permanent Stakeholder Group. Kai's awards include the IFIP Silver Core, the Alcatel Standard Elektrik Lorenz AG (SEL) Foundation Dissertation Award and the Friedrich-August-von-Hayek-Prize of Freiburg University and Deutsche Bank.