Social Science Computer Review
1-18

DOI: 10.1177/0894439317690337
journals.sagepub.com/home/ssc

Exploring Issues in a Networked Public Sphere: Combining Hyperlink Network Analysis and Topic Modeling

Daniel Maier[1], Annie Waldherr[2], Peter Miltner[1], Patrick Jähnichen[3], and Barbara Pfetsch[1]

Abstract

We propose a methodological approach to analyze the content of hyperlink networks which represent networked public spheres on the Internet. Using the case of the food safety movement in the United States, we demonstrate how to generate a hyperlink network with the web crawling tool Issue Crawler and merge it with the results of a probabilistic topic model of the network's content. Combining hyperlink networks and content analysis allows us to interpret such a network in its entirety and with regard to the mobilizing potentials of specific sub-issues of the movement. We focus on two specific sub-issues in the food safety network, *genetically modified food* and *food control*, in order to trace the involved websites and their interlinking structures, respectively.

Keywords

social network analysis, topic model, food safety, social movement, collective action, hyperlink

For communication researchers, the Internet has not only brought up a variety of compelling research questions but also put pressure on scholars to develop and apply methods appropriate for investigating these questions. When turning to analyze web communication, online public spheres, and social media networks, scholars need to adjust traditional methods such as content analysis or surveys and also develop entirely new approaches. Accordingly, social network analysis and automated content analysis are probably the most promising techniques for communication research. During the last few years, tremendous efforts have been put into further developing these techniques and applying them to web research. Scholars have also combined both approaches to study networks and content in conjunction (Bennett, Foot, & Xenos, 2011; Elgesem, Steskal, & Diakopoulos, 2015;

[1] Freie Universität Berlin, Berlin, Germany
[2] Westfälische Wilhelms-Universität Münster, Münster, Germany
[3] Humboldt Universität zu Berlin, Berlin, Germany

Corresponding Author:
Daniel Maier, Freie Universität Berlin, Garystr. 55, Berlin 14195, Germany.
Email: maier@zedat.fu-berlin.de

Himelboim, McCreery, & Smith, 2013; Kim, 2012; Tateo, 2005; Tremayne, Zheng, Lee, & Jeong, 2006).

Our study contributes to the ongoing challenge of combining content and network analysis. From a methodological perspective, we develop and test an explorative approach, which allows for an insight into the semantic dimensions of hyperlinked networks. Specifically, we use web crawling methods to generate an issue network and combine the network data with the inductive computational text mining approach of topic modeling. This combination allows us to trace not only social connections among actors online but also to identify what messages they put forth.

Our empirical case refers to online communication of the food safety movement in the United States. The broad issue of food safety comprises a variety of sub-issues such as genetically modified food (GM food) and food control. Due to different ethical and political constellations, these sub-issues are supposed to bear differently on mobilizing potentials among nongovernmental organizations (NGOs) and other civil society actors. We aim to scrutinize those sub-issue-dependent networks with our methodological approach. By both identifying the linkages between actors and discovering which sub-issues they jointly refer to, we get a full picture of the multiple dimensions and layers of the debate in this community.

From a theoretical point of view, we draw on the concept of issue networks. An issue network is a set of websites that all treat a common issue and are interconnected through hyperlinks (Marres & Rogers, 2005). Issue networks match nicely with the theoretical concept of the networked public sphere in that they constitute sets of actors with a topical relationship indicated by hyperlinks (Marres & Rogers, 2005). Bennett, Lang, and Segerberg (2015) consider such networks as representing "issue publics" or public spheres, respectively. Accordingly, we analyze an issue network in order to empirically determine a thematically centered public sphere.

In the first section of this article, we introduce the concept of networked public spheres as the theoretical framework of our study. We also explicate why this concept is appropriate for studying the potential of online mobilization by civil society actors. The subsequent sections are devoted to describing the data that were collected and processed by web crawling and topic modeling. We start the presentation of findings by giving an overview of the identified network and the topics of food safety therein. Eventually, we analyze two selected topics, food control and GM food, in detail, in order to draw conclusions on the mobilization potentials of the online issue networks of the selected sub-issues. We also discuss limitations and future implications of the study.

Theoretical Background

Our case study of online issue networks is based on theoretical arguments about the web as a networked public sphere consisting of multiple interconnected public communication spaces (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2015; Friedland, Hove, & Rojas, 2006). In this perspective, we empirically investigate the networked public sphere along the two dimensions of *content synchronization* and *structural connectivity* of debates. We argue that the content and structure of an issue network are indicative of how to study how mobilizing actors work toward an integration of the networked public sphere.

According to the majority of scholarly definitions, a public sphere is constituted by nonprivate communication within an openly accessible (metaphorical) space (Castells, 2008, p. 90). This space is a precondition for the circulation of information and the discourse about public affairs (Dahlgren, 2005; Papacharissi, 2010). With regard to the political system, the public sphere fulfills the function of an intermediate communication system between political decision makers and citizens, which can be described with respect to the standing actors and their connections and arguments. Now, several

scholars argue that the Internet fuels the structural diversity of this intermediate connection. It offers more public spaces with easier access than ever before but also more personalized communication environments (Bennett & Manheim, 2006; Pariser, 2011). These venues differ in their communication modes, institutional settings, size, and reach: from small chat rooms or fora of individuals to large networks of organized actors who communicate about issues on the web—so-called issue networks (Marres & Rogers, 2005). Instead of delimiting or isolating these various communication platforms, the web interconnects these otherwise distinct communication spaces. The varying modes and genres of communication result in an integrated "networked public sphere" (Benkler et al., 2015; Friedland et al., 2006).

The concept of a public sphere is defined by its societal and political functions, which can be met only when public communication is sufficiently integrated. In the scholarly discussion, two measurable criteria for integration are stressed. The first criterion is the extent to which the same issues are discussed simultaneously from similar perspectives in various subspaces of a public sphere (Eder & Kantner, 2000). The second criterion focuses on the communicative relations and interactions between speakers (Adam, 2008). Kleinen-von Königslöw (2010, pp. 49–64) summarizes these criteria as the *similarity* and *interconnectedness* of debates. Engrafting these criteria into the concept of a networked public sphere, we use the term content synchronization for similarity and structural connectivity for interconnectedness.

Among all participants of public debate, for civil society actors, these criteria are of particular importance. They seek integrated public communication to improve opportunities to build up mobilized alliances (Gamson & Wolfsfeld, 1993; Koopmans, 2004). Scholars emphasize that the networked character of the Internet-based venues of communication offers an opportunity structure for citizens and civil society actors to actively participate in political debate (Dahlgren, 2005). Civil society actors, which were rather marginalized in the mass media's public sphere (Gamson & Wolfsfeld, 1993), can now gain visibility and strengthen the salience of their positions via online communication (Pfetsch, Adam, & Bennett, 2013). Hence, the Internet effectively changes and improves the discursive opportunity structure for social movements and civil society actors (Cammaerts, 2012).

The engagement of civil society organizations, such as NGOs, nonprofit organizations, citizen initiatives, or social movement organizations in the public debate, is crucial for political deliberation. In this regard, Habermas (2006) argues that "[a]ssociational networks of civil society . . . translate the strain of pending social problems and conflicting demands for social issues into political issues" (p. 417). In order to successfully mobilize their cause, civil society organizations are interested in building coalitions with other actors; they seek to increase awareness for their issues and thereby raise the commitment of a critical amount of citizens (Gerhards & Rucht, 1992). Thus, an online issue network that constitutes a well-integrated public sphere in terms of the abovementioned criteria (content synchronization and structural connectivity) may boost the discursive opportunities and visibility of movement actors (Koopmans, 2004) and thereby increase the overall mobilization potential for a specific cause.

We therefore assume that a high degree of content synchronization and structural connectivity in an online issue network induced by civil society can be regarded as a high potential to mobilize attention and support for an issue and thus as an indicator for contention in society. For these reasons, one should expect that civil society actors are strategically interested in and actively working toward the integration of online networks and debates (Castells, 2008).

Against this background, we aim to explore the mobilizing potential of online issue networks in two case studies on food politics. We analytically apply the abovementioned criteria and demonstrate how to measure the degree of network integration.

An investigation of issue networks needs to be informed about the nature of hyperlinks. Hyperlinks are essential structural elements of online communication that enable actors to associatively refer to the communications of one another (Park, 2003). Rogers (2010) differentiates between cordial links, which connect websites of affiliate organizations, from aspirational links and critical links. Aspirational links are set by minor organizations or groups to connect to more established and powerful ones, while critical links are set to refer to actors which are criticized in a given context. According to Rogers, cordial links are the most frequently used and most significant category of hyperlinks. These hyperlinks maintain the collectively shared identity of movement groups in that they are directed to actors who share a congruent framing of political issues (Ackland & O'Neil, 2011; Diani, 2003). Moreover, receiving such links is a form of endorsement that may cause an increase in site traffic and subsequent forms of support (Ackland & O'Neil, 2011, p. 180). Well-integrated central actors enjoy higher visibility and influence than marginalized actors (Rogers, 2002).

However, the nature of hyperlinks is highly context dependent. The only common denominator is that they can always be considered to represent strategic acts of communication (Jackson, 1997; Shumate, 2012). We therefore interpret hyperlinks to indicate the degree to which an actor is integrated in a debate about an issue (Koopmans & Zimmermann, 2010, p. 175).

In this regard, issue networks publicly render the configurations of actors around a common issue and give them the possibility of political articulation (Marres, 2006). Issues are contentious by definition (see next section). However, it depends on a study's focus whether its aim is to reconstruct different positional camps *within* a single network (Meraz, 2012) or to map out only one camp of more or less like-minded websites (Bennett, Lang, & Segerberg, 2015). Our approach is the latter.

Method

Definitions and Operationalization

In order to investigate the network's degree of integration and thus infer statements about its mobilizing potential, we examine the two dimensions of content synchronization and structural connectivity of issue networks. The structural connectivity dimension is pretty straightforward, as it can be operationalized using descriptive measures from social network analysis, such as the network's density, the average nodal degree, and the clustering coefficient. The content dimension, in contrast, is less straightforward. One must clearly define what kind of content shall be investigated. Communication theory roughly provides three basic concepts: *topics*, *issues*, and *frames*. To avoid terminological confusion, we clarify these concepts and discuss how they are connected.

From the perspective of communication theory, topics are a basic precondition for interactional human communication and can be regarded as general categories that help us in structuring the complexity of reality and serve as points of reference for meaningful communication (Luhmann, 1971). In research on contentious politics, the term issue is not only more common than the term topic, but it also denotes that a matter is under debate and people can take sides. Issues are contentious by definition, "with individuals and groups taking opposing positions" (Miller & Riechert, 2001, p. 108). The contentious character of an issue denotes a basic conflictive perspective (or a superordinate contentious frame) on a topic. Finally, interpretive frames are less general categories that can be regarded as topical attachments that employ a more specific perspective (Eilders, 2000). Based on Entman's (1993) frame definition, Miller and Riechert (2001) argue that (in political communication) frames become "manifest in the choice and range of terms that provide the context in which issues are interpreted and discussed" (p. 109). They acknowledge that "the words are

Table 1. Google Search Terms and Seed URLs.

Search Terms	Seed URLs
Food safety, safe + food, food scandal, genetically modified foods, food + consumer protection, food + consumers, food + risk, food safety + campaign, food + labelling, food safety + control	http://www.centerforfoodsafety.org/ http://www.cspinet.org/foodsafety/ http://www.foodandwaterwatch.org/food/ http://www.organicconsumers.org/foodsafety.cfm http://notinmyfood.org/newsroom http://barfblog.foodsafety.ksu.edu/barfblog http://www.greenpeace.org/international/en/campaigns/agriculture/ http://www.pewhealth.org/topics/food-safety-327507

indicative of perspectives, or points of view, by which issues and events can be discussed and interpreted" (Miller & Riechert, 2001, p. 114).

Since we use topic modeling as part of our methodology, we also need to be clear what the term topic in *topic modeling* actually means. Jacobi, van Atteveldt, and Welbers (2016) agree that a topic model's resulting topics can in some cases indeed be interpreted in theoretical terms as issues or frames. "However, what exactly topics represent . . . is ultimately an empirical question" (Jacobi, van Atteveldt, & Welbers, 2016, p. 91), and it should not be disregarded that topic is actually just the term used for a latent variable that captures the "abstract notion" of a topic (Blei, Ng, & Jordan, 2003, p. 995). For communication research, this abstract notion needs to be interpreted in theoretical terms to make use of it (Jacobi et al., 2016, p. 90). In fact, Gelman and Shalizi (2013) rightly emphasize that basically every Bayesian analysis, such as topic modeling, needs a thorough interpretation by experts in their respective domains.

As intended by the design of the study, the data support that the topical categories that result from the applied modeling procedure should be interpreted as issues or sub-issues of food safety, respectively. As a result, we use the term sub-issue in the Findings section of the article when we refer to the topical categories of the topic model. However, in a technical context, the reference to the issue concept appears inappropriate. The term topic is used instead throughout to refer to the technical concept.

Gathering Hyperlink Network Data and Website Content

The data were gathered in a two-step procedure. In a first step, the network data were collected, that is, information about the interlinking structure of a set of websites. In a second step, we harvested the content of the webpages in the respective websites of the network.

In order to retrieve the network data, we used the web crawling tool Issue Crawler.[1] Crawling tools such as Issue Crawler take advantage of the network characteristic of the web in that they automatically collect and follow hyperlinks embedded in the source code of webpages. We applied a snowball procedure, which according to Waldherr, Maier, Miltner, and Günther (2016) is the most intuitional, inclusive, and nonrestrictive method to capturing the interlinking structure of an a priori unknown assemblage of websites. This approach is also in line with the inductive identification of issue-specific hyperlink networks proposed by Adam, Häussler, Schmid-Petri, and Reber (2016, p. 235) which requires no definitive knowledge about network boundaries; boundary specification in this view remains an analytical problem.[2]

The reconstruction of our hyperlink network began with the definition of seed URLs, that is, the starting points of the crawling procedure. These starting points have to be defined carefully, as they determine the overall structure of the resulting network. We conducted Google searches, a literature review, and gathered expert opinions in order to choose the most relevant websites of civil society

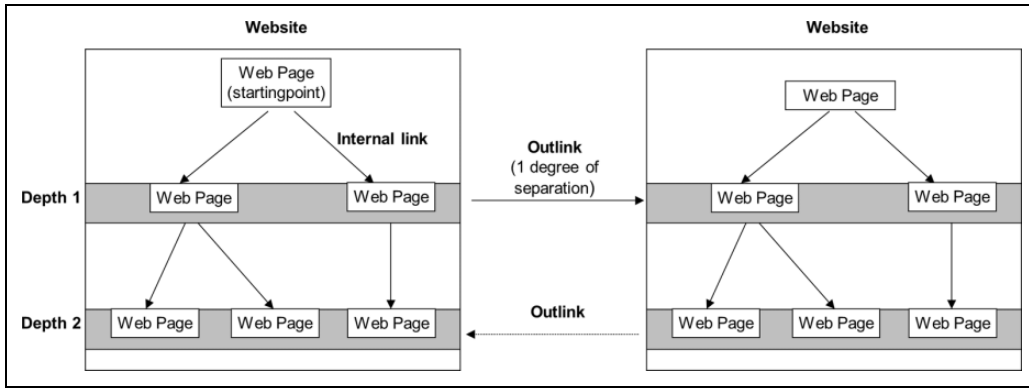

Figure 1. Logic of the crawling procedure.

actors engaging in food safety in the United States (see Table 1). This resulted in a list of 68 possible websites, which were checked for availability, up-to-dateness and most importantly—since we wanted to collect an issue network—whether food safety was an important issue. The finally chosen seed URLs are listed in Table 1.

In the case of a snowball crawling technique, the crawling algorithm fetches the seed URLs and follows every embedded hyperlink to the subsequent pages. After including these referred pages into the crawled pool of pages, the procedure recursively repeats itself. The crawling tool archives the history of visited webpages and hyperlink paths.

In order to prevent the network from becoming too big to analyze, we determined specific values for the two relevant parameters, *crawling depth* and *degree of separation*, which restrict the inclusion of webpages into the network. Crawling depth refers to the vertical dimension of the crawling process (see Figure 1) and restricts the algorithm to following internal links within a website up to a certain depth. We set the crawling depth to a value of 2. In contrast, the degree of separation refers to the horizontal crawling dimension and affects the hyperlinks between websites. We set the degree of separation to a value of 1, which means that the maximum distance of any website in the network to a seed site is a hyperlink path length of 1.

The crawling procedure results in a list of webpages, including the information about the website they belong to (nodelist) and how these websites are interlinked with each other (edgelist). In other words, the result of the crawling procedure is a directed hyperlink network.[3]

In order to download and archive the content data of the network's websites,[4] we used the free software tool Wget (version 1.13.4).[5] The outcome of this downloading process is a large collection of files from which we extracted the HTML files for further analysis.

The hyperlink network was gathered in November 2014 and comprises 17,881 webpages that belong to 3,755 different websites. These 3,755 websites refer to one another with 8,148 hyperlinks.[6] We were able to download and archive 11,845 of the webpages,[7] that is, 66% of the original network's pages. This loss of pages also affects the according network: the downloaded webpages belong to 2,211 websites (59%), with 5,886 (62%) hyperlinks connecting them. Every subsequent analysis refers to this reduced hyperlink network (2,211 websites; 11,845 pages; 5,886 hyperlinks).

Topic Modeling

The set of the collected 11,845 HTML files forms the basis of the text corpus on which we applied the topic modeling procedure. The HTML files were read in using an HTML parser that extracted the

plain text from the body of each file.[8] The topic model was estimated for the resulting corpus of text documents.

A topic model is a Bayesian hierarchical probabilistic model. It defines an artificial process for generating documents, describing how the actually observable data (the words in the documents) get into their places. In the most basic kind of topic model, latent Dirichlet allocation (LDA; Blei et al., 2003), this process is controlled by two latent factors, the topics and the proportions of topics in documents. The first factor, a topic, is formally defined as a probability distribution over the words in the corpus. The set of words with the highest probability in individual topics is assumed to describe them thematically. The second factor, the documents' topic proportions, is again a set of probability distributions (one for each document) defined over the assumed number of topics. Every topic is attributed a probability to appear in a document, and the probabilities of all topics for a single document sum up to 1. Simply put, in a topic model, the individual words that we see in a document are generated by first finding a topic through the documents' distribution over topics and then finding words from the chosen topic. Both choices are random draws from their respective distributions.

In our particular study, we use a nonparametric reinterpretation of the LDA model, the hierarchical Dirichlet process (HDP) topic model (Teh, Jordan, Beal, & Blei, 2006), to circumvent the problem of defining a certain number of topics a priori (Griffiths & Steyvers, 2004). The number of topics is instead inferred from the data. The only choice left is that of an abstract granularity parameter β ($\beta > 0$), which indirectly influences the number of the resulting topics. The lower the value of β, the higher the granularity of the model and the more topics we expect. For calculation, an HDP Java-based software implementation created by one of the authors was used.[9]

Two candidate topic models with different granularity parameters ($\beta = .1$ and $\beta = .5$) were calculated.[10] We solely focus on the model with the lower granularity level ($\beta = .5$). It resulted in an estimated optimal number of 53 topics, which we were able to interpret significantly better than the topics from the other model. This evaluation was based on the word distributions of the topics. Since the topics with the smallest corpus proportion are almost always of poor quality (Mimno, Wallach, Talley, Leenders, & McCallum, 2011, p. 262), we considered 23 topics to be negligible because they neither reflect coherent topical concepts nor did they make up for more than 1% of the modeled corpus.

As described above, the topic model's underlying text documents were extracted from HTML files using an HTML body extraction method. Although this is a common way to distill the usable content from HTML files with previously unknown structures (Günther & Scharkow, 2014), the method is prone to deliver results that deviate from what we would manually select as usable textual content. Hence, so-called boilerplate content, including website navigation, link lists and ads, and so on, potentially also becomes part of the topic model. Boilerplate topics are common phenomena in topic models (Mimno & Blei, 2011). Although boilerplate topics have no substantive meaning, their emergence sharpens the other meaningful topics "by segregating boilerplate terms in a distinct location" (DiMaggio, Nag, & Blei, 2013, p. 586).

For the interpretation of the remaining 30 topics, we used lists of the topics' 30 most likely terms. Additionally, we read 10 documents with the highest topic proportions for each topic to validate topical coherence.[11] Moreover, we calculated the topics' concentrations across the network using the Hirschman–Herfindahl Index (HHI) in order to evaluate how many actors actually contribute to a certain topic. The HHI can take values ranging from 1 (maximum concentration) to $\frac{1}{N}$, where N corresponds to the number of sites, which means, the topic is equally distributed among all sites of the network.[12]

The HHI turned out to be a useful measure for the evaluation of the topics. Almost every topic that we interpreted to be a boilerplate topic features a high HHI value (HHI > .2). Taking the topics' top words and the HHI value into account, we found 15 topics to represent meaningful coherent concepts, while the remaining 15 topics were considered boilerplate or noninterpretable. Therefore, we focus on these 15 remaining interpretable topics.

As a result, we yielded information about both the structural composition of the hyperlink network and the topic composition of its content. We concatenated the two data sets into a single comprehensive data source by defining the topic website proportions as node attributes of the network's websites.

Findings

The hyperlink network as empirical manifestation of the online food safety issue network is composed of 2,211 websites with 5,886 hyperlinks connecting them. The density of the graph is rather low (graph density < .001) due to the high amount of nodes. The nodes' average degree is 2.3, which means that, on average, each of the websites features 2.3 connections to other websites (in-links plus out-links). This, of course, is modestly informative, given that the graph's in- and out-degree distributions are massively skewed and approximately follow power-law distributions (Barabási & Albert, 1999), that is, many websites rarely receive and set links, while only a few receive and set many links. What is most important for evaluating structural connectivity is the network's clustering coefficient. As proposed by Watts and Strogatz (1998), the average local clustering coefficient measures the "cliquishness" (p. 401) or connectivity of a network. Given a node i, the local clustering coefficient $C(i)$ is equal to the proportion of i's *effectively* connected neighbors versus i's *potentially* connected neighbors (see also Newman, 2010, pp. 198–200). The average clustering coefficient $\bar{C}(i)$ is then defined as the average value over the local clustering coefficients for each node i in the network. The $\bar{C}(i)$ of an online issue network can thus be interpreted as the average strength of the communicative connectivity among websites. The issue network's $\bar{C}(i) = .175$, that is, the average probability that a randomly chosen website (in our network) forms a triadic configuration with two of its direct neighbors is 17.5%. We can use this value as baseline indicator for the public spheres structural connectivity and compare it with sub-issue specific networks.

Reconsidering our aim to assess the sub-issue spectrum of the collected issue network, we conclude from our data that the hyperlink network contains a diverse range of sub-issues in the contemporary food safety debate in the United States (Table 2). The debate incorporates sub-issues relating to sustainability in agriculture in general (Sub-issue 16), the production of GM food (Sub-issue 10), or consumers' action against antibiotics in animal feed (Sub-issue 18). These sub-issues are, of course, associated with consequential problems of food control (Sub-issue 7), such as outbreaks of infectious disease by means of contaminated food. All these effects threaten human health (Sub-issue 22) either because bacteria or viruses cause infections (Sub-issue 17) or because unhealthy ingredients (e.g., too much sugar) in highly processed food cause obesity (Sub-issue 33). Civil society actors who back up their arguments with scientific evidence seem to clamor for governmental regulation of the food industry (Sub-issue 25), educational approaches for solving existing problems, and public health research programs (Sub-issue 20). Only few sub-issues are concentrated (HHI > .2) among small sets of websites (e.g., Sub-issues 14, 18, and 19), whereas most sub-issues of food safety are widely dispersed across the network.

In order to examine exemplary online public spheres and detect mobilization potentials in U.S. food politics, we focus on two sub-issues: food control (Sub-issue 7) and GM food (Sub-issue 10). Both sub-issues can be considered key problems of the food safety debate (Anderson, 2000). Also, empirically, both make up a similar share of the modeled content (GM food: 3.5%, food control: 4.2%) and show similar dispersion indicators (GM food: HHI = .022, food control: HHI = .041). In public perception, the GM food sub-issue is connected to ethical concerns about diverse aspects of the use of biotechnology (Knight, 2009). Environmental and antiglobalization activist organizations publicly talk about these ethical concerns. While U.S. federal regulation of

Table 2. Description of Topics/Sub-issues in the Topic Model.

Sub-issue	Share (%)	Hirschman–Herfindahl Index	Label	Top 15 Words
20	9.0	.021	Public health research	Health, public, policy, research, education, information, program, resources, national, news, programs, events, members, microbiology, center
25	4.7	.011	Governmental regulation	State, public, bill, government, states, federal, trade, health, house, congress, campaign, court, senate, rights, action
7	4.3	.041	Food control	Food, safety, products, meat, beef, animal, news, chicken, poultry, salmonella, industry, health, product, recalls, foods
18	4.1	.342	Consumer action	Food, antibiotics, consumers, share, meat, CSPI, labeling, click, arsenic, consumer, safety, union, trader, newsroom, posted
16	3.5	.044	Sustainable agriculture	Food, farm, local, farmers, coffee, produce, home, fresh, policy, community, market, canning, farming, news, hunger
10	3.5	.022	Genetically modified food	Food, organic, crops, genetically, farmers, pesticides, pesticide, engineered, Monsanto, agriculture, seed, corn, foods, modified, seeds
5	3.5	.015	Research	Study, health, research, risk, disease, studies, article, data, levels, found, journal, human, published, exposure, science
3	3.3	.022	Climate change and energy	Climate, energy, read, global, change, report, environmental, environment, carbon, power, emissions, world, water, warming, natural
17	3.0	.122	Infectious foodborne disease	Health, outbreak, case, salmonella, infections, information, reported, disease, cases, infection, count, page, illness, persons, outbreaks
9	2.9	.331	Drinking water and fracking	Water, food, fracking, service, public, watch, works, quality, board, bottled, resources, Moines, environmental, local, radiation
22	2.7	.062	Health and disease	Cancer, health, ebola, disease, care, skin, news, brain, natural, Mercola, medical, article, heart, food, breast
33	2.3	.048	Unhealthy diet	Food, foods, products, sugar, nutrition, trans, Coca Cola, diet, salt, health, healthy, coke, product, ingredients, drinks
19	1.9	.217	Mercury in fish	Fish, retrieved, salmon, shutterstock, page, seafood, species, mercury, Wikipedia, search, stock, terms, create, wild, images
36	1.8	.102	Animal and dairy products	Milk, farm, U.S. Department of Agriculture, dairy, animal, cattle, organic, beef, farmers, cows, agriculture, animals, livestock, disease, producers
14	1.1	.225	Organic clothing	Organic, cotton, clothing, green, natural, made, fiber, fibers, products, chemical, sustainable, chemicals, fashion, health, skin

Note. For lack of space, only the top 15 words are represented in the table. CSPI = Center for Science in the Public Interest.

GM foods is relatively loose, there are many policies for the control and inspection of food products to prevent cases of food contamination with viral, bacterial, or other pathogens (Herring, 2015).

As we have chosen an explorative approach for this study, we did not formulate strong hypotheses about the nature of the issue networks and how they might differ with respect to their mobilization potentials. However, we expected to find a highly integrated networked public sphere among NGOs for the GM food sub-issue and less so for the food control sub-issue. We expect that because of the missing regulations and the strong ethical dimensions of the GM issue in the United States. In addition, we suppose that there will be concern about GM food by many environmental and anti-globalization activists alike.

Food Control

Focusing on the top words of the food control sub-issue (see Sub-issue 7 in Table 2) immediately leads us to conclude that the sub-issue captures first and foremost the safety of meat products (*food*, *safety*, *products*, *meat*, *beef*, *animal*, and *chicken*). The words *salmonella*, *coli*, *bacteria*, and *outbreak* denote several pathogenic agents of infectious foodborne disease, and all of these terms imply health risks for consumers. In combination with the terms *government*, *standards*, *U.S. Department of Agriculture* (*USDA*), and *industry*, we see that most important regulatory actors, such as the USDA, already play a decisive role in regulating food industry standards or that they at least are called to take action toward such regulation.[13]

Food control is the third most prevalent sub-issue, with a calculated fraction of 4.3% (and an approximated corrected fraction of 8.3%).[14] The sub-issue's rather low HHI value (.041) indicates that it is relatively equally distributed among the network's actors. Figure 2a depicts the network graph, with bigger nodes indicating a greater amount of HTML files contributing to the corpus. The more intense the reporting of a website on the sub-issue is, the more intense is the red color of its node. What we can see is that only a few websites, particularly the starting points of the crawl, are the biggest contributors of a larger amount of documents to the corpus compared to most other websites. The starting points of the crawl also contribute considerably to the focused sub-issue—four of them rank in Positions 3–10 of the top list of the sub-issue's contributors.

Apart from these actors, to which one can ascribe an exceptional position as the leading civil society organizations in the field of food safety, governmental regulatory entities, such as the *Food and Drug Administration (FDA)*, the *Centers for Disease Control and Prevention (CDC)*, and their joint information platform (foodsafety.gov), are the most extensive contributors to the sub-issue. Considering the fact that only a very small fraction of websites in the network (3.6%) is run by governmental entities, this is noteworthy. Finally, the websites of leading news media organizations, such as the *New York Times* and the *Washington Post*, are actively reporting on the sub-issue.

We can state that all of these actors enjoy well-integrated positions in the network,[15] leading to the conclusion that the sub-issue is rather located in the center of the issue network. Moreover, governmental entities apparently perceive it as their duty to regulate. Narrowing down the network to the core websites, that is, the websites that contribute at least 1% of the issue's probability mass, leads to the network depicted in Figure 2b. Only 12 nodes remain in the sub-issue-specific network, which features a high network density (.39) and a much higher average local clustering coefficient (.43) than the whole issue network. The average local clustering coefficient as a measure of the network's cliquishness indicates relatively strong strategic communicative integration (structural connectivity) for the sub-issue of food control. The heterogeneous set of actors including regulatory federal entities refers to one another more intensely with regard to the more specific synchronized sub-issue of food control as compared with the whole issue network.

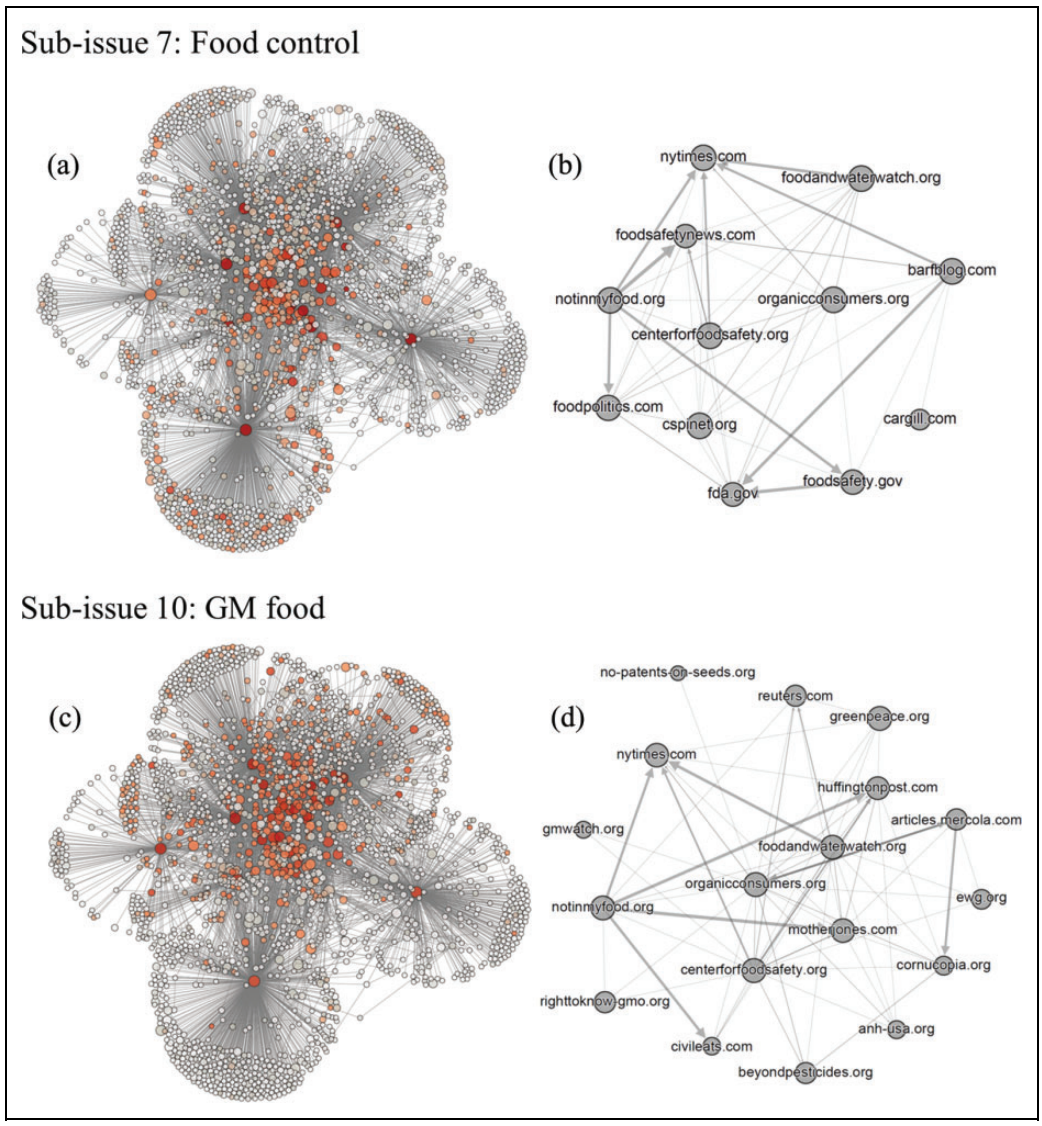

Figure 2. Sub-issue-specific networks and their most important actor constellations. (a) and (c) depict the hyperlink network graphs among 2,211 websites with 5,886 (dichotomized) hyperlinks connecting them. The size of the nodes indicates how many pages they contribute to the topic model. The intensity of the red color of the nodes indicates the amount of files they contribute to the respective sub-issue.[16] (b) and (d) depict the network extracts of the most important websites for each sub-issue. The size of the nodes indicates the number of HTML files that the websites contribute to the respective sub-issue. The graphs were drawn using the *Yifan Hu* algorithm as is implemented in the software package *Gephi* (version 0.8.2 beta).

GM Food

Regarding Sub-issue 10 (see Table 2), actors communicate about the use of GM/*engineered seeds* in *agriculture*, the labeling of GM *foods*, such as *corn*, and potential adverse effects for flora and fauna *(bees, plants, environment)*. The *issue* also features the names of predominant actors, that is, the industrial enterprise *Monsanto*, which is the market's leader for seeds in agriculture, the *farmers* who

are the potential users of biotechnology, and the *USDA* as the most important federal entity in the politics of GM organisms.[17]

GM food is the fifth most prevalent issue in the U.S. food safety network, with a fraction of 3.5% (corrected fraction: 6.8%). The issue's HHI value is very low (.022), denoting a close to equal contribution to the sub-issue by the network's actors. This also becomes apparent in Figure 2c, where one can see that the issue is more equally distributed among the actors in the center of the network, although its overall fraction is smaller than that of Sub-issue 7. We can state that the debate about GM food is not only restricted to the center core of the network but also reaches beyond to other, less well-integrated areas of the network, which can be located on the upper left side of the graph (connected via greenpeace.int.org).

The actors that maintain the GM food debate (see Figure 2d) can be described as a group of civil society organizations that are either specialized on GM (gmwatch.org, righttoknow-gmo.org, gmo freect.org) or agricultural affairs and food safety (cornucopia.org, ewg.org, centerforfoodsafety.org, foodandwaterwatch.org). These actors are intertwined with both more general news media organizations (such as nytimes.com, huffingtonpost.com) and specialized media sites (like civileats.com or motherjones.com). Interestingly enough, we can observe that the cluster of governmental entities, in contrast to Sub-issue 7, rarely engages in the ongoing debate.

The data reveal that GM food is by no means marginalized in the online network. Instead, we can conclude that the sub-issue itself is not only brought up by specialized organizations but also shares broad attention in the connected civil society community. There is a discernible tendency that the sub-issue is also apparent in different peripheral areas of the graph, which, in comparison to Sub-issue 7, indicates that it is more widely spread across the network.

Again narrowing down the network to the core websites, which were defined as contributing at least 1% of the sub-issue's probability mass (Figure 2d), 18 nodes remain. In this case, the sub-issue-specific network is predominantly based on civil society actors and special-issue news media. Governmentally ruled entities are not part of the condensed network. This sub-issue-specific network features a relatively small density value (.25) compared to the food control network in Figure 2b, which is partly due to the greater amount of websites. What is more interesting is that we can observe that the GM network features a much higher clustering coefficient (.53) than the food control network. In light of these findings, the GM issue features an even stronger communicative integration in terms of structural connectivity than food control.

Conclusion and Discussion

Our study set out to infer from the content and structure of online issue networks of movement organizations and civil society in the U.S. food safety politics about the mobilizing potential of its sub-issues. In the detailed empirical analysis, we combined network analysis and topic modeling and eventually focused on the sub-issues of food control and GM food.

The data for the two sub-issues reveal that both of them feature a relatively high prevalence, which indicates that there is a vivid debate going on in the civil society-induced online public sphere. These debates are located in overlapping sets of websites. However, while food control incorporates many regulatory authorities, such as the FDA and the USDA, these actors are far less well integrated in the GM food debate, although GM food features a slightly higher dispersion according to its HHI value. In the food control debate, we can observe participation of more general civil society organizations, while GM food unifies a more heterogeneous group of civil actors, excluding administrative entities.

In terms of structural connectivity, we conclude that there is a more strongly connected civic coalition in the core network for GM food ($\bar{C}(i) = .53$) than for food control ($\bar{C}(i) = .42$). From the perspective of content synchronization, we realize that food control is a less widely discussed sub-

issues among the diversely oriented civil actors compared to GM food. Furthermore, one could argue that the state actors obviously already claim to take responsibility for controlling food, which makes civil mobilization less pressing. In contrast, our data indicate a lack of state engagement regarding GM food, which at the same time can be considered a more widespread sub-issue among civil organizations. This is probably due to GM food connections to the sustainability debate as well as to the fact that U.S. authorities do not treat GM food differently from other food products. With regard to our theoretical framework, the stronger clustering indicates stronger mobilizing potential for the synchronized sub-issue of GM food. Moreover, for both of the sub-issue-specific networks, we can observe a much stronger cliquishness compared to the whole issue network. This finding indicates that alliance building and structural coherence in issue networks is strongly associated with content synchronization. However, we also observe that the issue network in its entirety is not just a compound of unconnected sub-issue networks. Instead, the issue network is an integrating aggregate, which structurally connects diverse sub-issues of food safety in its core and docks them to related sub-issues in its peripheral regions.

With respect to the methodology of our study, the combined approach of hyperlink network analysis and topic modeling turned out to be a highly instructive approach in the empirical analysis of the structure and content of public debate induced by civil society actors. Topic modeling proved to make a valuable contribution to study issue networks. The calculated model provided reasonable, well interpretable results that could easily be combined with the network data. Furthermore, once specified, the approach allows to easily narrowing down the perspective to specific networks of inductively identified sub-issues. Thus, it offers a valuable alternative to deductive approaches of defining issue networks and dealing with the noisiness of snowball hyperlink networks (Waldherr et al., 2016).

The approach is auspicious, but one also has to be aware of its pitfalls. There is no guideline for how a topic model's topics should be interpreted. Therefore, we categorized the topics identified using categories from communication theory. We also note that the interpretation of the topics is highly context dependent. Our interpretation of the topics as sub-issues might not hold up for other cases. The interpretation of the resulting topics depends on how narrow or broad the selection of the text material in the corpus is defined. Interpreting topics as sub-issues in turn has theoretical implications, as they cannot be separated from their underlying interpretive frames. Instead, sub-issues are compositions of multiple, repeatedly used and partly co-occurring semantic patterns. We therefore conclude that a theoretically guided and well-reasoned decision about a topic model's input data set is a crucial precondition of such analyses. Another challenge of the methodological approach is that there are no standardized guidelines for the validation of topic models.

In future research, some limitations of our study also have to be reassessed. First, the method for extracting usable content of HTML webpages has to be refined in order to reduce boilerplate content. Second, the approach can be used beyond exploration. The knowledge about the topic composition of websites might also be taken as a valuable predictor for the linking structure or vice versa.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article: This publication was created in the context of the Research Unit "Political Communication in the Online World" (1381), subproject 7, which is funded by the Deutsche Forschungsgemeinschaft (DFG,

German Research Foundation). The subproject is also funded by the Swiss National Science Foundation (SNF).

Notes

1. For further information on the tool, please visit: http://www.govcom.org/index.html
2. Conducting snowball crawling techniques in order to find issue-related networks typically results in noisy network data, that is, some nodes of the network do not contain content related to the issue under study. This problem results from crawlers' incapability of distinguishing between meaningful strategic hyperlinks and hyperlinks from advertisement banners or other parts of a website (e.g., Adam et al., 2016; Waldherr et al., 2016). However, the application of topic modeling enables us to focus on those actors with meaningful contributions to the food safety issue as shown in the Findings section.
3. What might be confusing at this point is that both the concept of the issue networks and the tool with which its empirical realization (a hyperlink network) was gathered, the Issue Crawler, explicitly refer to the term issue in their names. However, there is by no means a guarantee for the resulting networks to treat common issues (Waldherr et al., 2016). The crawling procedure solely relies on the described rules, without regarding contents.
4. O'Neill, McClain, and Lavoie (2001) argue that we can differentiate between webpages and websites. "A Web site is the collection of all Web pages located at the same top-level . . . URL" (p. 281).
5. Information available at: http://www.gnu.org/software/wget
6. The number of hyperlinks between the webpages is much higher (48,242). After the aggregation from the webpage level to website level, we accounted only for the *presence* (1) or *absence* (0) of hyperlinks between websites.
7. Many websites prohibit an automated access and subsequent download with crawling programs such as Wget. There are further reasons that webpages cannot be downloaded, such as login protections, and so on.
8. English stop words were removed from the text corpus.
9. Software documentation available at: https://bitbucket.org/hotblack_desiato/topic-models
10. The reason for calculating two models with different granularity parameters was to check whether a higher granularity results in a more appropriate model.
11. The proposed interpretation is the result of the document inspection, the topics' word lists and metrics and a discussion in the research team, until consensus was reached.
12. The topic page assignments (as inferred by the model) have to be aggregated to the level of websites, first. Therefore, we summed up all the topic page assignments ($\theta_{i,p}$, i.e., the proportion of topic i located on page p) over all the pages p that belong to a website s, that is, $\{p : p \in s\} : \theta_{i,s} = \sum_{p \in s} \theta_{i,p}$.
13. Some words are not listed in Table 2, but they are among the top 30 words of Sub-issue 7.
14. Regarding the fact that the sum of the interpretable topics of the model accounts for 51.7% of all contents, we calculated an approximate corrected topic fraction by dividing the original topic fraction (4.3%) by the total cumulative fraction of the interpretable topics (51.7%).
15. The visualization algorithm (Yifan Hu), which is based on the Force principle, places mutually connected actors closer to one another, which results in a densely interlinked core of the network (Hu, 2011).
16. Colored versions of Figures 2a and c are contained in the online-version of the article only.
17. Again, not all of the mentioned words are contained in Table 2, but are included in the top 30 words of Sub-issue 10.

References

Ackland, R., & O'Neil, M. (2011). Online collective identity: The case of the environmental movement. *Social Networks*, *33*, 177–190.

Adam, S. (2008). Do mass media portray Europe as a community? German and French debates on EU enlargement and a common constitution. *Javnost—The Public*, *15*, 91–112. doi:10.1080/13183222.2008.11008966

Adam, S., Häussler, T., Schmid-Petri, H., & Reber, U. (2016). Identifying and analyzing hyperlink issue networks. In G. Vowe & P. Henn (Eds.), *Political communication in the online world: Theoretical approaches and research designs* (pp. 233–247). New York, NY: Routledge.

Anderson, W. A. (2000). The future relationship between the media, the food industry and the consumer. *British Medical Bulletin*, *56*, 254–268.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512. doi:10.1126/science.286.5439.509

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2015). Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. *Political Communication*, *32*, 594–624. doi:10.1080/10584609.2014.986349

Bennett, W. L., Foot, K., & Xenos, M. (2011). Narratives and network organization: A comparison of fair trade systems in two nations. *Journal of Communication*, *61*, 219–245. doi:10.1111/j.1460-2466.2011.01538.x

Bennett, W. L., Lang, S., & Segerberg, A. (2015). European issue publics online: The cases of climate change and fair trade. In T. Risse (Ed.), *European public spheres. Politics is back* (pp. 108–137). Cambridge, England: Cambridge University Press.

Bennett, W. L., & Manheim, J. B. (2006). The one-step flow of communication. *The ANNALS of the American Academy of Political and Social Science*, *608*, 213–232. doi:10.1177/0002716206292266

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Cammaerts, B. (2012). Protest logics and the mediation opportunity structure. *European Journal of Communication*, *27*, 117–134. doi:10.1177/0267323112441007

Castells, M. (2008). The new public sphere: Global civil society, communication networks, and global governance. *The ANNALS of the American Academy of Political and Social Science*, *616*, 78–93. doi:10.1177/0002716207311877

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, *22*, 147–162. doi:10.1080/10584600590933160

Diani, M. (2003). Networks and social movements: A research programme. In M. Diani & D. McAdam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 299–319). New York, NY: Oxford University Press. doi:10.1093/0199251789.003.0013

DiMaggio, P., Nag, M., & Blei, D. M. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*, 570–606. doi:10.1016/j.poetic.2013.08.004

Eder, K., & Kantner, C. (2000). Transnationale Resonanzstrukturen in Europa. Eine Kritik der Rede vom Öffentlichkeitsdefizit [Transnational resonance structures in Europe: A critique of the supposed deficit of the public sphere] [Special issue]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *52*, 306–331.

Eilders, C. (2000). Media as political actors? Issue focusing and selective emphasis in the German quality press. *German Politics*, *9*, 181–206. doi:10.1080/09644000008404613

Elgesem, D., Steskal, L., & Diakopoulos, N. (2015). Structure and content of the discourse on climate change in the blogosphere: The big picture. *Environmental Communication*, *9*, 169–188. doi:10.1080/17524032.2014.983536

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*, 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x

Friedland, L. A., Hove, T., & Rojas, H. (2006). The networked public sphere. *Javnost—The Public*, *13*, 5–26. doi:10.1080/13183222.2006.11008922

Gamson, W. A., & Wolfsfeld, G. (1993). Movements and media as interacting systems. *The ANNALS of the American Academy of Political and Social Science*, *528*, 114–125.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*, 8–38. doi:10.1111/j.2044-8317.2011.02037.x

Gerhards, J., & Rucht, D. (1992). Mesomobilization: Organizing and framing in two protest campaigns in West Germany. *American Journal of Sociology*, *98*, 555–595. doi:10.1086/230049

Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*, 5228–5235. doi:10.1073/pnas.0307752101

Günther, E., & Scharkow, M. (2014). Automatisierte Datenbereinigung bei Inhalts- und Linkanalysen von Online-Nachrichten [Automated cleansing of content and link data in online news analyses]. In K. Sommer, M. Wettstein, W. Wirth, & J. Matthes (Eds.), *Automatisierung in der Inhaltsanalyse* [Automation in content analysis] (pp. 111–126). Köln, Germany: Herbert von Halem Verlag.

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*, 411–426. doi:10.1111/j.1468-2885.2006.00280.x

Herring, R. J. (2015). How is food political? Market, state, and knowledge. In R. J. Herring (Ed.), *The Oxford handbook of food, politics, and society*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780195397772.013.35

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication*, *18*, 40–60. doi:10.1111/jcc4.12001

Hu, Y. (2011). Algorithms for visualizing large networks. In U. Naumann & O. Schenk (Eds.), *Combinatorial scientific computing* (pp. 525–549). Boca Raton, FL: CRC Press.

Jackson, M. H. (1997). Assessing the structure of communication on the World Wide Web. *Journal of Computer-Mediated Communication*, *3*, 00. doi:10.1111/j.1083-6101.1997.tb00063.x

Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*, 89–106. doi:10.1080/21670811.2015.1093271

Kim, J. H. (2012). A hyperlink and semantic network analysis of the triple helix (university–government–industry): The interorganizational communication structure of nanotechnology. *Journal of Computer-Mediated Communication*, *17*, 152–170. doi:10.1111/j.1083-6101.2011.01564.x

Kleinen-von Königslöw, K. (2010). *Die Arenen-Integration nationaler Öffentlichkeiten: Der Fall der wiedervereinten deutschen Öffentlichkeit* [The integration of arenas in national public spheres: The case of the reunified German public sphere]. Wiesbaden, Germany: VS Verlag.

Knight, A. J. (2009). Perceptions, knowledge and ethical concerns with GM foods and the GM process. *Public Understanding of Science*, *18*, 177–188. doi:10.1177/0963662507079375

Koopmans, R. (2004). Movements and media: Selection processes and evolutionary dynamics in the public sphere. *Theory and Society*, *33*, 367–391. doi:10.1023/B: RYSO.0000038603.34963.de

Koopmans, R., & Zimmermann, A. (2010). Transnational political communication on the internet. In R. Koopmans & P. Statham (Eds.), *The making of a European public sphere* (pp. 171–194). Cambridge, MA: Cambridge University Press.

Luhmann, N. (1971). Öffentliche Meinung [Public opinion]. In N. Luhmann (Ed.), *Politische Planung: Aufsätze zur Soziologie von Politik und Verwaltung* [Political planning: Essays on the sociology of politics and administration] (pp. 9–34). Opladen, Germany: Westdeutscher Verlag.

Marres, N. (2006). Net-work is format work: Issue networks and the sites of civil society politics. In J. Dean, J. W. Anderson, & G. Lovink (Eds.), *Reforming politics: Networked communications and global civil society* (pp. 3–18). New York, NY: Routledge.

Marres, N., & Rogers, R. (2005). Recipe for tracing the fate of issues and their publics on the web. In B. Latour & P. Weibel (Eds.), *Making things public: Atmospheres of democracy* (pp. 922–935). Cambridge, MA: MIT Press.

Meraz, S. (2012). The democratic contribution of weakly tied political networks: Moderate political blogs as bridges to heterogeneous information pools. *Social Science Computer Review*, *31*, 191–207. doi:10.1177/0894439312451879

Miller, M. M., & Riechert, B. P. (2001). The spiral of opportunity and frame resonance: Mapping the issue cycle in news and public discourse. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 35–65). Mahwah, NJ: Lawrence Erlbaum.

Mimno, D., & Blei, D. M. (2011). Bayesian checking for topic models. In Association for Computational Linguistics (Ed.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 227–237). Stroudsburg, PA: Association for Computational Linguistics.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In Association for Computational Linguistics (Ed.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Stroudsburg, PA: Association for Computational Linguistics.

Newman, M. (2010). *Networks: An introduction*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199206650.001.0001

O'Neill, E. T., McClain, P. D., & Lavoie, B. F. (2001). A methodology for sampling the World Wide Web. *Journal of Library Administration*, *34*, 279–291. doi:10.1300/J111v34n03_07

Papacharissi, Z. (2010). *A private sphere: Democracy in a digital age*. Cambridge, England: Polity Press.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin Press.

Park, H. W. (2003). Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, *25*, 49–61.

Pfetsch, B., Adam, S., & Bennett, W. L. (2013). The critical linkage between online and offline media: An approach to researching the conditions of issue spill-over. *Javnost—The Public*, *20*, 9–22.

Rogers, R. (2002). Operating issue networks on the web. *Science as Culture*, *11*, 191–213. doi:10.1080/09505430220137243

Rogers, R. (2010). Internet research: The question of method—A keynote address from the YouTube and the 2008 election cycle in the United States Conference. *Journal of Information Technology & Politics*, *7*, 241–260.

Shumate, M. (2012). The evolution of the HIV/AIDS NGO hyperlink network. *Journal of Computer-Mediated Communication*, *17*, 120–134. doi:10.1111/j.1083-6101.2011.01569.x

Tateo, L. (2005). The Italian extreme right on-line network: An exploratory study using an integrated social network analysis and content analysis approach. *Journal of Computer-Mediated Communication*, *10*, 00. doi:10.1111/j.1083-6101.2005.tb00247.x

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, *101*, 1566–1581. doi:10.1198/016214506000000302

Tremayne, M., Zheng, N., Lee, J. K., & Jeong, J. (2006). Issue publics on the web: Applying network theory to the war blogosphere. *Journal of Computer-Mediated Communication*, *12*, 290–310. doi:10.1111/j.1083-6101.2006.00326.x

Waldherr, A., Maier, D., Miltner, P., & Günther, E. (2016). Big data, big noise: The challenge of finding issue networks on the web. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439316643050

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442. doi:10.1038/30918

Author Biographies

Daniel Maier is a research associate and PhD candidate in the Institute for Media and Communication Studies at the Freie Universität Berlin. He graduated in political science from the University of Passau in 2009 and received a master's degree in communication science from the Freie Universität Berlin in 2012. In 2016, he also received a master's degree in public health from Charité Berlin. His research focuses on methodology in communication science, network science, and computational social science. e-mail: maier@zedat.fu-berlin.de

Annie Waldherr is an assistant professor of digitized public spheres in the department of communication at the Westfälische Wilhelms-University in Münster. She received her PhD from the Freie Universität Berlin in 2011 for her dissertation on the dynamics of media attention. In 2005, she graduated in communication science from the University of Hohenheim in Stuttgart. Her research interests include mediated public spheres, political online communication, science and technology discourses, and social simulation. e-mail: annie.waldherr@uni-muenster.de

Peter Miltner is a research associate and PhD candidate in the Institute for Media and Communication Studies at the Freie Universität Berlin. He graduated in communication science from the University of Hohenheim in Stuttgart in 2009 and holds a master's degree in European interdisciplinary studies from the College of Europe in Warsaw (2010). His research interests include political (online) communication and network analysis. e-mail: peter.miltner@fu-berlin.de

Patrick Jähnichen is a postdoctoral researcher at the machine learning group at Humboldt-Universität zu Berlin. He received his PhD from Leipzig University in 2016 for his dissertation on modeling topics dynamically over time. He graduated with an MSc degree in computer science from Leipzig University after having received a bachelor's degree from the University of Cooperative Education in Stuttgart. His main research interests are Bayesian mixture models and their dynamics applied to natural language texts, stochastic processes to steer the dynamics, and statistical inference in these models. e-mail: patrick.jaehnichen@hu-berlin.de

Barbara Pfetsch is a professor of communication theory and media effects research at the Freie Universität Berlin. Her research focuses on comparative analyses of political communication, processes of media agenda building, and political communication on the Internet. She has published several books, including *Political Communication Cultures in Western Europe* (2014), *Comparing Political Communication* (2004), and numerous articles and chapters. e-mail: pfetsch@zedat.fu-berlin.de