An Evaluation of Retrospective Facebook Content Collection from Political Facebook Pages

Marko Bachl¹

¹ University of Hohenheim, Department of Communication

Abstract

The content on Facebook pages is obviously subject to changes. However, there is often a considerable time lag between the sampling period and the collection date in studies on public Facebook communication. The present evaluation study aims to assess whether—and if so, how—problematic this common practice is in terms of unavailable content. I investigated whether a sample of content items, which were posted to 408 German political Facebook pages in March 2017, could be retrospectively collected in June 2017, September 2017, and March 2018. 27% of the 132,068 test items were no longer accessible after 12 months. The deletion of complete pages was rare, but some prominent pages were not available retrospectively. Posts by the pages themselves were far more likely to be available compared to content, which was posted by private accounts to the pages. There were only small differences between the political parties, but there was substantial variation between the distinct pages. I conclude with a discussion of implications for single research projects (applying timely data collection whenever possible) and for communication research and related disciplines (considering institutionalized data collection efforts).

Keywords: Dynamic content, Data collection, Content analysis, Archive, Facebook

Word count: XXX

Prepared for the conference proceedings of the 2017 annual meeting of the DGPK Methods section, “Dynamische Prozesse der öffentlichen Kommunikation: Methodische Herausforderungen”, edited by Stefan Geiß, Christian Schemer, Teresa K. Naab, Christina Peter, and Philipp Müller. I want to thank Sebastian Stier for insights into the social media data collection at GESIS and Teresa K. Naab and Stefan Geiß for comments.

Correspondence concerning this article should be addressed to Marko Bachl, University of Hohenheim, Department of Communication. E-mail: marko.bachl@uni-hohenheim.de

An Evaluation of Retrospective Content Collection on Facebook

Content from the Internet is today an established research object in communication research (Skalski, Neuendorf, & Cajigas, 2017). Studies on user-generated online content also have become increasingly popular (Naab & Sehl, 2017). One important challenge in online content analysis is the initial content collection, because Internet content is inherently dynamic: It can change over time or even disappear completely. McMillan (2000) aptly titled her early review of online content analyses “the microscope and the moving target” (p. 80). 17 years later, Naab and Sehl (2017) still concluded that “several of the challenges posed by the continuously changing nature of the content are not fully met” (p. 1256). In this report, I look at the retrospective collectability of political content on one relevant online platform. Facebook is the online social network site with the greatest reach in most countries and it is considered an important platform for political communication. Scholars investigated, for example, how political parties communicate on Facebook (Arzheimer, 2015; Magin, Podschuweit, Haßler, & Russmann, 2016; Stier et al., 2018; Stier, Posch, Bleier, & Strohmaier, 2017), how users interact with posts from media organizations’ pages (Schmidt et al., 2017), or which (media) sources were referred to by parties and their followers on political pages (Bachl, 2018).

Lacking a systematic review of the divergences between the sampling period and time of data collection in the literature, I will provide a comparison in some selected studies. This serves to illustrate the practical relevance of the problem: How much time passes between the original publication of the Facebook content and its collection? In an own study, I collected posts, comments, and replies, which were posted on AfD-related Facebook pages during the year 2016, in December 2016 (Bachl, 2018). Arzheimer (2015) was interested in all posts by the party and the users on the main AfD page since its inception in March 2013. The posts were retrieved in July 2014. The sampling period of Stier et al. (2017) was December 2014 to August 2016, all content was collected at the end of the sampling period. Schmidt et al. (2017) proceeded similarly for their sampling period of six years (January 2010 to December 2015). Magin et al. (2016) investigated content from political party pages in the four weeks before the 2013 German and Austrian national elections. The content collection took place in November 2015. The first “Social Media Monitoring” project of GESIS relied on a purely retrospective collection in November 2013 for the sample period from January 2009 to October 2013 (Kaczmirek et al., 2014). The second edition of the project for the Bundestag election 2017 again initially planned to use data from a retrospective collection. Luckily, they conducted an additional collection in which the data for each post was downloaded eight days after its creation for a shorter sampling period (August, 15th, to September, 30th, 2017). A comparison of both data sets revealed that the retrospectively collected data set missed a non-negligible amount of content compared to the rolling collection. The authors consequently decided to publish the latter data set (Stier et al., 2018).

Retrospectively collected content may be incomplete compared to the population of content, which was present during the sampling period. Both the administrators of the public pages and the individual users can decide to modify or delete the content at any time and for many reasons. The complete deletion of prominent pages is the clearest example. Hannelore Kraft, the former minister-president of North Rhine-Westphalia, and

Frank-Walter Steinmeier, the current Bundespräsident of Germany, recently deleted their public profiles.¹ Their decisions made it impossible to retrospectively collect all relevant content for a comprehensive study on the use of Facebook during the 2017 election campaign in North Rhine-Westphalia or for a study about changes in the Bundespräsident's online communication over the course of his career. Beyond such obvious problems, the deletion of many single content items across pages and users can similarly interfere with the validity of retrospectively collected data. The consequences may be particularly problematic if the missingness occurred not at random. The assumption, that specific content (e.g., hate speech, arguments against a party on its own page, communication mishaps) is more likely to be deleted, seems highly plausible. The results would then not only be less precise, but also biased. With this in mind, I attempt to answer two straightforward research questions:

1. Which proportion of the content items on selected political Facebook pages can be retrospectively collected after three to four, six to seven, and eleven to twelve months?
2. How do the odds to retrospectively collect a content item differ between content types, parties, and pages?

The first question simply wants to know which shares of the original content would still have been available if the collection was conducted later in time. The exact dates of the three test collections are not especially meaningful, as they were chosen for pragmatic reasons (conference submission, conference presentation, writing of this manuscript). They are, however, sufficiently typical to provide useful points of reference for comparison with other studies. The second question differentiates the messages based on their content types and based on the pages and the parties, to which the pages belong. With regard to content types, I distinguish posts by the pages themselves from posts by individual accounts. Comments and replies were treated as separate content types. This distinction provides points of reference for studies which focus on the messages of political communicators (posts from pages) and for studies which are interested in user-generated content (user posts, comments, replies) on political Facebook pages. In addition, the hierarchical structure of the content items on Facebook must be considered when thinking about patterns of content deletion. Each post is published on one page, each comment refers to one post, and each reply refers to one comment. If, for example, a post is deleted, all comments to this post and their replies also vanish. It can therefore be expected that content items, which are lower in the hierarchy, are more likely to be missing because of the deletion of higher-order items. The comparison of parties and pages is indicative to gauge how much studies on certain political parties or studies which focus on specific pages are affected by retrospectively non-available content items.

It should be noted that the present study does not try to explain *why* some content items cannot be retrieved after a given time period. The reasons why any given item is missing is practically impossible to infer from publicly available data. An unavailable content item could have been deleted by the creator, by the page's administrators, or on behalf of Facebook. Similarly, a user with all of his or her content items could have been banned from one page or from the whole platform. He or she could also have decided to change

¹Steinmeier now uses a new profile: <https://www.facebook.com/Bundespraesident.Steinmeier/>.

the privacy settings, or delete or deactivate his or her whole account. The simple goal of the present study is to provide estimates of the magnitudes of losses which have to be expected when collecting content from German political Facebook pages retrospectively. Any interpretations beyond this descriptive aim are only informed speculation. A survey of page administrators or Facebook users would be necessary for a thorough investigation of the underlying causes.

Methods

The present analysis is based on a larger data collection effort in the run-up to the 2017 Bundestag election. All posts, comments, and replies on 422 political Facebook pages were collected every 6 hours since late February 2017 via the Facebook Graph API.² The pages were sampled to represent the most important pages of the major political parties in Germany, and, additionally, of the PEGIDA movement.³ Pages of the AfD were oversampled because of a research interest in populist-right online communication. For this evaluation study, I selected a stratified subsample of all content items which were published in March 2017. A maximum of 25 posts a) from each page itself and b) from users to each page were randomly selected. Then a maximum of 50 comments to each post and, in turn, a maximum of 50 replies to each comment were randomly selected. The procedure resulted in a sample of 408 pages (the other pages did not post or receive any content in March 2017), 8,032 posts from pages (i.e., by an administrator of the page in the name of the page), 1,013 posts from users, 69,182 comments, and 53,841 replies (overall 132,068 content items). Each content item on Facebook has a unique alphanumeric identifier, which was used to check whether it could still be retrieved in three test runs on June 12th, 2017, September 13th, 2017, and March 14th, 2018.

Two shortcomings of the procedure should be noted right away: First, the original data collection already missed content which was posted *and* deleted in the 6-hour windows between two collection runs. This is relevant for research questions on content which is presumably deleted quickly by the pages' moderators, such as hate speech and trolling. It may also preclude items which were published by error or immediately received unfavorable responses and were therefore deleted quickly. Research interests which inherently include the risk of content deletion will most likely require an even closer tracking of new content.

Second, I did not systematically evaluate technical inconsistencies of the data retrieval procedure. In each evaluation period, all content items were checked several times and the results differed very slightly in each run. About 0.5% to 1.5% of all items had divergent results in one run compared to the next one about 10 minutes later. A cursory review of the inconsistencies did not suggest any systematic losses, but the technical reliability of the procedure surely remains a topic that deserves further attention. I decided to use the results from the first run in each test period for the subsequent analyses. About 1% of the items in the study showed highly implausible results which are most likely due to (technical) errors in data retrieval.⁴ 186 content items were found only in the third test, but neither in the

²<https://developers.facebook.com/docs/graph-api>

³For sake of simplicity, I refer to the organizations behind the Facebook pages as "parties". This also includes the PEGIDA movement and its pages, although PEGIDA is of course not a political party.

⁴Another explanation would be that the items come from users who deactivated and then again reactivated their Facebook accounts. While a small amount of divergences could be caused by such user actions, I doubt

Table 1

Pages without collectable content items after March 2017

Page	Identified in
Frank-Walter Steinmeier	June 2017
Hannelore Kraft	June 2017
BadenWürttemberg freiheitlich-patriotisch-traditionsbewusst *	September 2017
Junge Alternative Lüneburg	September 2017
Alexander Dobrindt, MdB **	March 2018
Alternative für Deutschland - Ortsgruppe Rostock	March 2018
Alternative für Deutschland - Unterstützer Kreisverband Ennepe-Ruhr	March 2018
Alternative für Deutschland AfD - Ostholstein	March 2018
Die Linke.SDS Berlin	March 2018
Michael Fuchs	March 2018
Pegida NRW	March 2018

Note. * All content items were deleted, page still exists. | ** Only userposts to page in March 2017. All content items were deleted, page still exists.

first nor in the second. 163 items were found only in the second test. 636 items were not found in the first, but in both later tests. 545 items were missing only in the second test.

A public repository on the Open Science Framework documents some accompanying materials to this article: <https://osf.io/an9u5/>. It contains a minimal data set, which is sufficient to reproduce the statistical analyses presented herein. However, I cannot make the full data set publicly available because of privacy and data protection concerns and the Facebook terms and services. The variable “id” in the public data set can be used to retrieve the content items which were in the initial sample for this evaluation study. A continuation of the study over a longer period is made possible in principle with this information. The OSF page also contains an *R* (R Core Team, 2017) script, which documents data collection and preparation. *RFacebook* (Barbera, Piccirilli, Geisler, & van Atteveldt, 2017) was used for data collection. Data management and preparation were mainly conducted with *dplyr* (Wickham & Francois, 2016) and *tidyr* (Wickham, 2017). The original manuscript was typeset with *papaja* (Aust & Barth, 2017) and contains all code necessary to reproduce the statistical analyses. They were conducted with *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *arm* (Gelman & Su, 2016) and the plots were created with *ggplot2* (Wickham, 2009), *ggstance* (Henry, Wickham, & Chang, 2016), and *egg* (Auguie, 2017).

Results

Before turning to the research questions, we take a look at the pages whose complete March 2017 content was no longer collectable after one year (Table 1). This concerned 11 of the 408 pages. The deletion of the content on the four pages of individual politicians was presumably motivated by major changes in their careers. Hannelore Kraft resigned from that they occur that frequently.

all public and party offices after the 2017 NRW election. Frank-Walter Steinmeier was elected Bundespräsident in March 2017. Michael Fuchs did not stand for re-election and left the Bundestag after the 2017 election. All three pages were no longer publicly available. The page of Alexander Dobrindt seems to have been reactivated after March 2017. No own content was published during the sampling period of the evaluation study, only some comments by other users were found. They were all removed in between the September 2017 and the March 2018 evaluation runs, maybe in an attempt to “start fresh” with new communication activities after the transition from federal minister to head of the CSU parliamentary group. All four examples suggest (albeit with a tiny sample size of politicians) that the deactivation or a major overhaul of the public Facebook pages of politicians shortly after important career changes can pose a threat to the completeness of retrospectively collected data. The deletion of whole pages did not matter much quantitatively, but there is a real chance that any crucial page’s content will no longer be available.

The disappearance of most of the pages from political organizations concerned smaller, local chapters. They may be less important for many typical research questions. However, the inability to retrospectively collect the content on the Pegida-related “BadenWürttemberg freiheitlich-patriotisch-traditionsbewusst” page uncovered a surprising—at least to me—pattern of deletions. The page *regularly* removes all older posts every few months. The prevalence of this phenomenon beyond the present sample and the underlying motivation have to remain unclear for now. However, it is obvious how such a systematic revision of a page’s publication history may invalidate the data from a retrospective collection even if the page itself remains available.

Research question 1 asked about the shares of content items which could be collected retrospectively. Overall, 17% of the 132,068 test items were no longer accessible after three to four months, 23% after six to seven months, and 27% after eleven to twelve months. A more detailed picture of the temporal volatility of Facebook content has to take the path dependencies due to the hierarchal organization of the content items into account.

Figure 1 presents these dependencies organized by content type. The white areas show the shares of items in a category, which could be accessed in the evaluation run. The areas in different shades of gray indicate the shares of items categorized by the hierarchically highest item which was also missing. The first column in each panel clearly indicates that posts by the pages themselves were most stable, with only 5% of inaccessible posts even after about one year. The political organizations in this sample mostly seem to have refrained from removing their own content. User-generated content was far more volatile. This especially holds for posts by users to the topmost level of a page. Almost one fifth of them could no longer be collected after three to four months. About one year after publication, over one third was unavailable. Comments and replies were also quite volatile, with one fifth or one third deletions after one year, respectively. The inaccessibility of higher-order content elements, mainly of comments, was an important reason for the disappearance of replies. About 20% of all replies could no longer be accessed after one year because the respective comment had been removed.

Research question 2 asked about differences between content types, parties, and pages. Figure 2 summarizes the results for content types and parties. The first panel confirms the descriptive results from Figure 1. Posts by the pages themselves were far more stable than user-generated content. Userposts and replies had the lowest odds of being

Figure 1. Shares of accessible and deleted content items

Notes. The percentages report the shares of items in a category, which could be accessed in the evaluation run.

collected retrospectively. The differences between the parties and PEGIDA were rather small compared to the differences between content types or between pages. Content items on FDP-related pages could be retrospectively retrieved somewhat more often, the content on AFD-related pages was somewhat more likely to vanish. The pattern of differences did not change substantially over the course of the three evaluation periods. Only the deletion rates on PEGIDA pages have grown slightly stronger compared to the overall trend.

Figure 3 presents the differences between the pages. The upper panel shows how the odds to retrospectively collect a content item vary between all pages. The considerable variation is comparable in magnitude to the differences between the content types. The distribution is skewed somewhat to the left. Overall, the rates of successful retrospective collection were high on most pages, but some pages lost substantial parts of their content. The lower panel presents the 15 pages with the strongest absolute deviations from the grand mean of successful retrospective collection in the last evaluation run after one year (with exception of the pages in Table 1 which had no retrospectively collectable content). The highly diverse set of pages mirrors the limited differences between the parties (Figure 2, lower panel). Both the positive and negative tails of the distribution include pages from different parties as well as pages from prominent politicians and small, local party chapters.

Figure 2. Differences in retrospective accessibility between content types and parties

Notes. The figure shows point estimates and intervals of ± 2 standard errors. Zero is the grand mean log(odds) of successful retrospective collection in each evaluation run (June 2017: 2.76 or, converted to a probability, .94; September 2017: 2.31 or .91; March 2018: 1.88 or .87; note that the grand means are not identical with the simple shares from above). Positive values indicate above-average odds, negative values below-average odds. The model includes content type, parties, and pages. The results for pages are shown in Figure 3.

$n_{\text{observations}} = 121,864$; $n_{\text{pages}} = 408$; $n_{\text{parties}} = 8$; $n_{\text{content types}} = 4$.

There seems to be no obvious pattern to distinguish between categories of pages with highly stable or more volatile content. The patterns of differences were quite similar for all three evaluation runs. Among the notable pages in the bottom panel, only the accessibility of content items on Andrea Nahles' page showed a substantial change between the June and the September 2017 evaluation runs compared to the overall trend.

Discussion

In sum, the evaluation study detected substantial volatility in the content which was posted on a large set of German political Facebook pages in March 2017. The deletion of complete pages did not matter much quantitatively. However, the absence of some pages (Hannelore Kraft, Frank-Walter Steinmeier) from the data set would have had devastating consequences for common research projects in political communication. Posts, which were published by the pages themselves, were far more likely to be available retrospectively than user-generated content (userposts, comments, replies). There were only small differences between the political parties, but there was substantial variation between the distinct pages. None of the findings are overly surprising. It is, after all, well-known that content on the Internet and especially user-generated content are subject to changes. However, the common knowledge

Figure 3. Differences in retrospective accessibility between pages

Notes. The upper panel shows the density of the distributions of the point estimates of all pages. The lower panel shows the point estimates and intervals of ± 2 standard errors for the 15 most extreme pages (with the exception of the pages in Table 1) in March 2018. The names of some pages were abbreviated: Mttlstnds-undWrtschftsdCDU/CSU: Mittelstands- und Wirtschaftsvereinigung der CDU / CSU; AltrntvfrDtschlnd,KrsvrbdDsbr: Alternative für Deutschland, Kreisverband Duisburg; JngAltrntvfrDtschlnd-Bzkrsvrbns: Junge Alternative für Deutschland Bezirksverband Südbaden; Zero is the grand mean log(odds) of successful retrospective collection in each evaluation run (June 2017: 2.76 or, converted to a probability, .94; September 2017: 2.31 or .91; March 2018: 1.88 or .87; note that the grand means are not identical with the simple shares from above). Positive values indicate above-average odds, negative values below-average odds. The model includes content type, parties, and pages. The results for content types and parties are shown in Figure 2.

$n_{\text{observations}} = 121,864$; $n_{\text{pages}} = 408$; $n_{\text{parties}} = 8$; $n_{\text{content types}} = 4$.

seems to have only limited consequences for the empirical practice of Facebook studies. The empirical evidence presented in this report will hopefully inspire critical reflection and methodological improvements toward less *ad hoc* data collection practices.

The results have implications for the interpretation of studies which employed retrospective content collection, for researchers who plan to conduct studies on Facebook content, and for communication research and related fields who are interested in social media content. Past Facebook studies, like the ones listed in the introduction, obviously cannot change their data collection procedure. Similarly, researchers, who wish to conduct a study on public Facebook communication during a past sampling period, have no viable alternative to collecting the content retrospectively. In these instances, the implications of the present evaluation study depend on the research interest. If a study is primarily interested in the content, which was published by the pages themselves (i.e., their posts), retrospective content collection seems to be an acceptable approach. However, the prominent examples of completely unavailable pages reminded us that not every research question on parties' and politicians' Facebook communication activities can be answered with retrospectively collected data. Plausibility checks of the data sets are certainly in order. Checking for deleted data in the absence of a full data set for comparison is, of course, mostly educated guesswork. For political pages, contrasting the collected data with a timeline of politically relevant events and with a list of known relevant politicians is one possible starting point. If the Facebook page of a major politician is missing from the sample or if party or politician pages do not contain posts during important election campaigns, then skepticism is certainly in order. The untested assumption of complete data is not prudent, even for posts by the pages.

Studies on party-user interactions or user-generated content, which applied retrospective content collection, should certainly be taken with a grain of salt. The reported results strongly suggest that substantial shares of userposts, comments, and replies on political Facebook pages vanish from the platform. The present study did not try to identify the (likely) causes for the deletion of user-generated content. It can only be assumed that a combination of active moderation by the pages' administrators and actions taken by the users were responsible for the lower odds of retrospective availability. Whatever the reasons might be—the results clearly show that the user-generated content which could be retrieved after one year is an incomplete and most likely biased representation of the original content.

The most straightforward consequence for future research projects is to use more timely and repeated data collection procedures. It is my impression that researchers routinely apply the data collection approaches from traditional content analysis to the “moving target” (McMillan, 2000, p. 80) of online and social media communication. Relatively easy access via APIs and user-friendly tools like *Facepager* (Jünger & Keyling, 2018) convey the look-and-feel of traditional media archives such as Nexis or WISO.⁵ This has lured some researchers into collecting data from Facebook like from curated archives, where retrospective data collection is the inherent purpose of the service. If one wants to pick an analogy from established research methods, it makes a whole lot more sense to think about the data collection plan of a Facebook study like the one of a longitudinal survey. Obviously, nobody would believe that asking about respondents' attitudes from last year is a sensible research design decision.

⁵<https://www.nexis.com> and <https://www.wiso-net.de>.

If one “asks” a political Facebook page via the API about its content from a year ago, the answer is likely to be similarly unreliable. A timely data collection procedure does not pose a major challenge for many typical research projects in political communication. The most important pages are often known in advance, as are the sampling periods of interest, which are usually defined by election dates or other long-term expectable events. Timely data collection should be considered the standard approach if these conditions hold.⁶

Finally, the uncertainty about the retrospective accessibility of public Facebook communication highlights the need to think about institutional initiatives to reliably archive online and social media content of public interest and to make it available for non-commercial research purposes. One obvious solution would be a closer cooperation with the platforms themselves, in this case, Facebook. The recent revelations surrounding the data collection practices of Cambridge Analytica, which were at least in part made possible by a researcher who passed on data collected for scientific purposes,⁷ have led to mixed developments in this regard. On the one hand, recent updates of the Graph API have already seriously limited the ability of independent researchers to collect publicly accessible data from Facebook.⁸ On the other hand, Facebook agreed to a research partnership, in which previously unattainable data sets from within Facebook can be analyzed by independent scholars.⁹ It remains to be seen whether the data sets from this initiative will be sufficiently robust to changes of the original content. The description of the first data set mentions that the data from deleted users cannot not be included—even if the data set is only available in aggregated form and does not contain information about individual users. Another possibility would be a coordinated data collection effort—if the scope of public data access remains sufficient for such an endeavor. The “Social Media Monitoring” initiative by GESIS (Kaczmirek et al., 2014; Stier et al., 2018) is an interesting step in this direction. Yet the project is also confronted with many legal and ethical challenges beyond the technical implementation, ranging from data and privacy protection to copyright and the platforms’ terms and conditions. Some aggregated results can be retrieved via <http://mediamonitoring.gesis.org/>, but the distribution of the raw or at least more detailed data is not possible at the moment.

It is beyond the scope of the present evaluation study to outline an institutional solution to the challenges of changing online and social media content. Yet the empirical results make it clear that such initiatives would certainly be appreciated by many researchers. At present, the best practice is an own timely data collection effort and I strongly encourage its application whenever possible. But not every research group has the resources to implement and operate an adequate infrastructure. Because of this, and also because we do not always know in advance which content we would like to collect, some Facebook (and more generally, social media) studies will continue to rely on retrospective content collection.

⁶Another practical reason to favor timely over retrospective data collection procedures in Facebook studies is a bug in the Facebook Graph API, which to the best of my knowledge is still unresolved at the time of this writing: <https://developers.facebook.com/bugs/1838195226492053/>. In short, the collection of past posts from a page will sometimes be incomplete. The bug occurs mostly for pages with high activity. Although this issue is not related to the empirical results presented in this report, it should be mentioned here, because it makes large-scale retrospective collections at least infeasible, if not impossible.

⁷See, for example, the extensive coverage by The Guardian: <https://www.theguardian.com/news/series/cambridge-analytica-files>.

⁸<https://developers.facebook.com/blog/post/2018/04/04/facebook-api-platform-product-changes/>

⁹<https://socialscience.one/>

The present evaluation study has shown that this approach may be acceptable in some instances (especially if the focus is on professional communicators' publication activities), but will probably lead to substantially and systematically incomplete data sets. At the very least, we should take this into account when evaluating the results of our studies.

References

- Arzheimer, K. (2015). The AfD: Finally a successful right-wing populist eurosceptic party for Germany? *West European Politics*, 38(3), 535–556. doi:[10.1080/01402382.2015.1004230](https://doi.org/10.1080/01402382.2015.1004230)
- Auguie, B. (2017). *Egg: Extensions for 'ggplot2', to align plots, and set panel sizes*. Retrieved from <https://CRAN.R-project.org/package=egg>
- Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bachl, M. (2018). (Alternative) media sources in AfD-centered Facebook discussions. *Studies in Communication and Media*, 7(2), 128–142. doi:[10.5771/2192-4007-2018-2-128](https://doi.org/10.5771/2192-4007-2018-2-128)
- Barbera, P., Piccirilli, M., Geisler, A., & van Atteveldt, W. (2017). *Rfacebook: Access to facebook api via r*. Retrieved from <https://CRAN.R-project.org/package=Rfacebook>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:[10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
- Gelman, A., & Su, Y.-S. (2016). *Arm: Data analysis using regression and multi-level/hierarchical models*. Retrieved from <https://CRAN.R-project.org/package=arm>
- Henry, L., Wickham, H., & Chang, W. (2016). *Ggstance: Horizontal 'ggplot2' components*. Retrieved from <https://CRAN.R-project.org/package=ggstance>
- Jünger, J., & Keyling, T. (2018). *Facepager. An application for generic data retrieval through APIs*. Retrieved from <https://github.com/strohne/Facepager/>
- Kaczmarek, L., Mayr, P., Vatrupu, R., Bleier, A., Blumenberg, M. S., Gummer, T., ... Wolf, C. (2014). Social media monitoring of the campaigns for the 2013 German Bundestag elections on Facebook and Twitter. *GESIS-Working Papers*. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0168-ssao-381955>
- Magin, M., Podschuweit, N., Haßler, J., & Russmann, U. (2016). Campaigning in the fourth age of political communication. A multi-method study on the use of Facebook by German and Austrian parties in the 2013 national election campaigns. *Information, Communication & Society*, 20(11), 1698–1719. doi:[10.1080/1369118X.2016.1254269](https://doi.org/10.1080/1369118X.2016.1254269)
- McMillan, S. J. (2000). The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. *Journalism & Mass Communication Quarterly*, 77(1), 80–98. doi:[10.1177/107769900007700107](https://doi.org/10.1177/107769900007700107)
- Naab, T. K., & Sehl, A. (2017). Studies of user-generated content: A systematic review. *Journalism*, 18(10), 1256–1273. doi:[10.1177/1464884916673557](https://doi.org/10.1177/1464884916673557)
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*, 114(12), 3035–3039. doi:[10.1073/pnas.1617052114](https://doi.org/10.1073/pnas.1617052114)
- Skalski, P. D., Neuendorf, K. A., & Cajigas, J. A. (2017). Content analysis in the interactive media age. In K. A. Neuendorf (Ed.), *The content analysis guidebook* (2nd ed., pp. 201–242). Los Angeles: Sage.
- Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., ... Staab, S. (2018). Systematically monitoring social media: The case of the German federal

- election 2017. *GESIS-Working Papers*. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-56149-4>
- Stier, S., Posch, L., Bleier, A., & Strohmaier, M. (2017). When populists become popular: comparing Facebook use by the right-wing movement Pegida and German political parties. *Information, Communication & Society*, 20(9), 1365–1388. doi:10.1080/1369118X.2017.1328519
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>
- Wickham, H. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>