

FROM EDITORS TO ALGORITHMS
A Values-Based Approach to Understanding
Story Selection in the Facebook News Feed

Michael A. DeVito

Department of Communication Studies
Northwestern University
Evanston, IL, USA

devitom@u.northwestern.edu

<http://www.mikedeviso.net/>

This is a pre-print version of an article published that has been accepted for publication in *Digital Journalism*. For quotation, please consult the final, published version of the article at <http://dx.doi.org/10.1080/21670811.2016.1178592>.

Please cite as:

DeVito, M. A. (2016) From Editors to Algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism Ahead of print*. doi: 10.1080/21670811.2016.1178592

ACKNOWLEDGEMENTS

The author thanks Eileen Emerson for her research assistance, as well as Kerric Harvey, David Karpf, and Emily Thorson of The George Washington University for their guidance and support during the initial research on this piece. The author also thanks the anonymous reviewers, William Marler, and the members of Aaron Shaw's "Bring Your Own Research" group at Northwestern University for their feedback on the manuscript.

FUNDING

This research was partially supported by a thesis grant from The George Washington University's School of Media and Public Affairs, with which the author was previously affiliated.

FROM EDITORS TO ALGORITHMS

A Values-Based Approach to Understanding Story Selection in the Facebook News Feed

Facebook's News Feed is an emerging, influential force in our personal information flows, especially where news information is concerned. However, as the News Feed's story selection mechanism starts to supplant traditional editorial story selection, we have no window into its story curation process that is parallel to our extensive knowledge of the news values that drive traditional editorial curation. The sensitive, trade-secret nature of the News Feed and its constant updates and modifications make a traditional, computer science based examination of this algorithmic giant difficult, if not impossible. This study takes an alternative approach, using a content analysis of Facebook's own patents, press releases, and Securities and Exchange Commission filings to identify a core set of algorithmic values that drive story selection on the Facebook News Feed. Informed by the principles of material culture analysis, it ranks these values to create a window into Facebook's curation process, and compares and contrasts Facebook's story selection values with traditional news values, examining the possible consequences of one set of values supplanting the other. The study finds a set of nine News Feed values that drive story selection: friend relationships, explicitly expressed user interests, prior user engagement, implicitly expressed user preferences, post age, platform priorities, page relationships, negatively expressed preferences, and content quality. It also finds evidence that friend relationships act as an overall influence on all other story selection values.

KEYWORDS: algorithmic curation; algorithms; critical algorithm studies; content analysis; Facebook; material culture analysis; news information; news values

Facebook was not designed to give us the news. It was designed to connect users and help them share and engage with content (Facebook 2012). As such, the service's heart, the News Feed, carries content from the personal to the professional in many different formats, and does not operate directly in parallel with the story selection process at a traditional news organization. However, regardless of design, the News Feed is now taking an increasingly central role in our information flows as its user base expands and users increasingly rely on it as an integrated part of their daily lives (Duggan, Ellison, Lampe, Lenhart & Madden 2014). In particular, it is becoming a key source of news information, with 41 percent of US adults currently receiving news through Facebook (Matsa & Mitchell 2014) and Facebook moving to encourage news organizations to publish content in a News Feed-native format (Somaiya, Isaac & Goel 2015).

As Facebook assumes the role of news source, it takes on some of the influential gatekeeping and agenda-setting functions that have traditionally been performed by human editors (McCombs & Shaw 1972). In particular, it takes these functions on through the mechanism of story selection. In the traditional editorial process, story selection by human editors is guided by *news values*, which Jerry Palmer (2000, p. 45) defines as "a system of criteria which are used to make decisions about the inclusion and exclusion of material and also – crucially and less obviously – about which aspects of stories to present in the form of news output." In the News Feed, all story selection is conducted not by editors, but by algorithms. This process is not exclusive to news content, but all news content is filtered through it (Sittig &

Zuckerberg 2013). These algorithms operate based on what we will call *algorithmic values*, defined as *a system of criteria which are used to make decisions about the inclusion and exclusion of material and which aspects of said material to present in an algorithmically-driven news feed*.

The gatekeeping and agenda-setting roles of both traditional and now algorithmic news sources play key roles in determining the content and vocabulary of the public conversation (McCombs & Shaw 1972; Scheufele 1999; Scheufele 2000), so we need to understand how these roles are carried out. News values give us a great insight into how the process worked traditionally (O'Neill & Harcup 2009; Palmer 2000), but we have no equivalent picture of algorithmic values. The public and academics alike have little access to the creators and methods of algorithms, which are often deliberately obscured to protect intellectual property (Lievrouw 2012; Pasquale 2011; Tufecki 2015). Compounding this lack of access, the algorithms in question also change rapidly and without notice (Karpf 2012), and are sometimes not fully understood even by their creators (Napoli 2014). Our limited understanding of the Facebook News Feed's selection process in particular is based on outdated, incomplete information (e.g., Bucher 2012) – we do not have a clear picture of what the algorithm is, much less what values it is embedding into its story selection process.

This study uses content analysis informed by material culture analysis techniques to identify a set of algorithmic values that drive story selection in the Facebook News Feed. It examines the comparative roles of news and algorithmic values in story selection, reviews the scholarship on introducing bias and values to algorithms, and details the unique challenges that an algorithmically-driven, operationally opaque medium like Facebook poses to understanding news information flow. By sidestepping the key barriers that arise when seeking a formal representation of the News Feed and instead relying on primary source documents generated by Facebook itself, it uncovers a more lasting, values-based picture of how this key information mediator carries out story selection, and how this might change news information flow.

Background

Not everyone can know everything, and not every source of information can publish every bit of information it comes across. Aside from practical concerns like space constraints, there are also the problems of information overload and information anxiety, where too large a body of information becomes practically unusable, and can even lead to the purposeful avoidance of information as a defense mechanism (Bawden & Robinson 2009). There must be a mechanism for selecting content, and the most commonly-used framework for explaining how editors select which stories to publish out of the myriad possibilities is that of news values (Palmer 2000).

There have been many different formulations of what the current news values are, from the perspective of large-scale, international stories (e.g. Gatlung & Ruge 1965) as well as more everyday story selection (e.g. Gans 1979; Harcup & O'Neill 2001; O'Neill & Harcup 2009). A popular formulation, as recorded by Lee (2009), includes “novelty or oddity, conflict or controversy, interest, importance, impact or consequence, sensationalism, timeliness, and proximity.” These explicit values form the basis of the journalistic “gut feeling” for newsworthiness, with the additional motivator of exclusivity often added (Schultz 2007). These news values have changed over time and with cultural norms, and there is no universal agreement as to what the current set is (O'Neill & Harcup 2009; Palmer 2000). However, the presence of an informed discussion on the topic which involves direct input from editors and

journalists themselves (O'Neill & Harcup 2009) allows, at the very least, an accessible understanding of the shape of the values that drive traditional news.

Algorithmic Curation

Ultimately, editors use news values to determine what is “relevant” for their audiences (Gans 1979; Lee 2009); a critical combination of news values gets a story labeled relevant (Harcup & O'Neill 2001) and once that story is labeled relevant, its selection indirectly influences audience attention (Lee 2009). “Relevance,” however, is a tricky construct – it is highly subjective in the absence of defined criteria, and there is no independent definition of what is relevant to a particular audience (Gillespie 2014). Today, algorithms are likewise using supplied criteria to determine what is “relevant” to their audiences and worth knowing (Gillespie 2014; Tufecki 2015).

As we have seen, the definition of information relevance by an entity outside the self is not a new feature of media that started with algorithms. Traditional mass media certainly does not dictate public opinion (Bennett & Iyengar 2008), but it does play a key role in defining what issues are relevant enough to be included in the public conversation (McCombs & Shaw 1972; Scheufele 2000) and the terms in which those issues should be discussed (Scheufele 1999).

Throughout the 20th century, the press has performed these public relevance functions in a semi-transparent manner (Bimber 2003). As noted above, though the vast majority of consumers are not editors, editorial methods and values are well-documented (e.g., Gans 1979; O'Neill & Harcup 2009), tracked over time (e.g., Michelstein & Boczkowski 2009) and understandable by the public. We understand human editors' values, and their flaws, and generally look upon their exercise of these public relevance functions with a critical eye (Fingerhut 2016; Gallup 2015).

Increasingly, these functions are being transferred to mediating algorithms (Gillespie 2014; Pariser 2011). These algorithms represent a media structure, just as the old journalistic regime did (Napoli 2014), and they accordingly have a new set of biases. However, unlike with the traditional press, these biases are not generally recognized, as algorithms are assumed, but not verified, to be impartial (Bozdog 2013; Pariser 2011). Algorithms have “a technologically-inflected promise of mechanical neutrality” according to Gillespie (2014, 181); in the popular imagination, they are unbiased, especially as compared to the much-maligned biases of human journalists and editors (e.g., Vallone, Ross & Lepper 1985). The algorithm, in this view, is finally the solution for the journalistic bias problem.

Algorithmic Values

Algorithms are by no means perfect; they have biases just as surely as human editors do. Structurally, they have trouble with rapid shifts in tastes and situations as well as serendipitous leaps (Churchill 2013), and they rely primarily on correlation, not deep comprehension of information (Andrejevic 2013). They are also affected by what Friedman and Nissenbaum (1996) call “technical bias” which comes from the limitations of technology itself, including the technical limitations of databases they draw from, available storage and processing power, and the influence of any errors in their code (Friedman & Nissenbaum 1996; Gillespie 2014; Introna & Nissenbaum 2000).

Technical issues aside, algorithms also share a key bias with flesh-and-blood editors, what Friedman & Nissenbaum (1996) call “preexisting bias” which comes from individual or

societal input into the design of a system. This can be conscious or unconscious, and just as individual and societal biases form the basis for news values (Schultz 2007), this forms the basis of algorithmic values.

Algorithms necessarily have value decisions embedded in their design, which come into play whenever approximation or interpretation are required, whenever human constructs must be operationalized, and whenever there is a choice of more than one way to solve a particular problem (Bozdag 2013; Friedman & Nissenbaum 1996; Kraemer, van Overveld & Peterson 2011). For example, if Facebook is going to prioritize posts from close friends in the News Feed, engineers must decide on the criteria that defines a “close friendship” versus an acquaintance. Engineers must turn this nuanced human construct with many different interpretations into a single, operational definition, and in doing so embed their values as to what a “close friend” is into the algorithm itself.

These implicit value judgments start long before a single line of code is written. Facebook is a business, so the product must ultimately serve business interests; in this case, the business requires micro-targeting data to sell to advertisers, so the News Feed product needs to value (and therefore prompt) continued user input, direct interaction between users, and increased friend connections in order to identify preexisting product preferences (Andrejevic 2013; Bucher 2012; Churchill 2013; Domingos 2005). Designers must also make the potential decision to prioritize the platform’s own content (Bozdag 2013) or, in a Facebook-specific case, the content of cooperative partners through programs like Instant Articles (Somaiya et al. 2015). Finally, designers must account for regulatory concerns such as Terms of Service compliance policing and copyright, legal, and governmental requests (Bozdag 2013).

As the design process moves past business needs, engineers make values-based decisions with deep impacts on the algorithm’s eventual output. Engineers must choose exactly what variables they are evaluating and working with (Steiner 2012) and what data sources the algorithm will draw on (Gillespie 2014). Inclusion and exclusion criteria here form a layer of value judgments as to what is relevant enough to even be included in a database (Introna & Nissenbaum 2000); not being indexed effectively places content outside an algorithm’s “vision,” as if the content does not even exist.

Additionally, engineers must decide at this point who defines “relevance” in a relational sense. On the one hand, engineers can decide to make popularity the key value, prioritizing the majority’s opinion and, potentially, causing majority dominance over what should be localized topics (Gillespie 2014). On the other hand, individual preference can be made the key value, basing decisions on the user’s own past actions (read as preference), and potentially opening up the possibility of feedback loops and filter bubbles (Pariser 2011).

This values-based decision-making is endemic to algorithmic systems. Nicholas Diakopoulos (2014) has identified points where this happens in each of the major functions of algorithms: prioritization, classification, association, and filtering. Value judgments in these areas are clearly visible in Facebook’s own technical documentation of the News Feed (e.g., Kao, Schafer & Watzman 2014; Sittig & Zuckerberg 2013; Zuckerberg, Bosworth, Cox, Sanghvi & Cahill 2012).

With the Facebook News Feed mediating such a considerable amount of our information flow, it is clear the values guiding that mediation need examination, leading to our research questions:

- **RQ1: What are the core algorithmic values of the Facebook News Feed?**

- **RQ2: How are those algorithmic values ranked in order to select stories for the Facebook News Feed?**
- **RQ3: How do these algorithmic values differ from traditional news values?**

Previous Approaches

The question of what drives the Facebook News Feed is not a new one; academics and marketers alike puzzle over this question. While marketers have a constantly changing stream of folk wisdom and best practices that inform their understanding of the News Feed (search for “facebook news feed marketing” for over 180 million examples), the understanding of the News Feed’s decision-making process in the academic world is weak, at best.

Research on Facebook in general is scattered and fragmented, taking place across many different academic disciplines, with a focus on how people behave on Facebook and how to use Facebook as a research tool, and not how Facebook itself behaves (Caers et al 2013; Wilson, Gosling & Graham 2012). With that lack of study on Facebook itself comes a major gap in the literature: our understanding of the News Feed stops with EdgeRank, the formula that guided the News Feed over five years ago. That’s a problem, because, as Facebook has declared multiple times, EdgeRank is dead (McGee 2013).

Most treatments of the Facebook News Feed lead back to two trade articles that explain EdgeRank based on a PowerPoint presentation given by News Feed engineers Ruchi Sanghvi and Ari Steinberg at the 2010 f8 developer conference (Kincaid 2010; Taylor 2011). With Sanghvi’s presence, it is fair to assume the presentation was authoritative – she is listed on most of the relevant patents.

This presentation was the first look into News Feed since its launch four years prior, and Kincaid writes that it was clear to developers that there had been a massive evolution from launch to presentation day. What was revealed was a two-element system, featuring objects (content) and edges (relational interactions – tags, comments, etc.). Edges had three components at the time: affinity between the viewer and the item creator, type of edge (with comments being more important than likes), and time decay. The formula was relatively simple: each of these elements multiplied together for each edge and then all added together to get the EdgeRank. The higher the EdgeRank, the more likely to be in the feed.

The best translation of all this information into academic terms is Tania Bucher’s 2012 article on algorithmic power. According to Bucher, affinity is a measure of the creator/viewer relationship based on an interaction score that tracks wall posts, likes, private messages, etc. Weight is a measure of overall popularity or importance, based on many factors, but especially comments. She also notes that content type and the traffic in the viewer’s own network have an effect on the EdgeRanks. She points out that the algorithm makes assumptions about friendships; not all friendships are weighted equally, and the measure of closeness was, as of 2011, based entirely on interactions. Engagement ranking in general, she writes, is largely based on user effort; chatting with a friend requires more than just commenting, which requires more than liking, which requires more than viewing, and so a hierarchy of weights is produced. However, in 2012, this information was already at least two years outdated, and the chorus of “EdgeRank is dead” was on the rise, with ethnographic research based on interviews with engineers pointing to a much more complex system (Steiner 2012).

Research Challenges

Bucher's paper essentially marks the limits of academic knowledge of how News Feed story selection works ends, and an updated view of how this complex algorithm functions is not easy to come by due to two major factors: the impermanence of Internet technologies and the black-boxed nature of most influential algorithms.

The first challenge is the transience of any understanding of the News Feed we may glean; the web is by nature transient, rapidly changing at a rate that regularly outpaces the research process (Karpf 2012). For algorithms in particular, this is an ongoing, iterative process between engineers and users where new tweaks drive waves of user feedback and market changes, and user feedback and market changes drive new versions (Mager 2012); there are, effectively, no locked, finished algorithms.

The bigger challenge is the black-boxed nature of major algorithms, where the inner workings of a major system are purposely obscured from public view. This is a common practice to protect trade secrets and prevent malicious hacking and gaming of the system (Diakopoulos 2014; Lievrouw 2012; Pasquale 2011). As a result, it is rare for us to get any insight into key values-based processes like variable definition (Steiner 2012). The News Feed itself is heavily protected, and is a complex enough system that even those within Facebook and within the News Feed team may have no clear picture of how it works on an individual basis (Napoli 2014). From a traditional research standpoint, the News Feed is inaccessible.

A rigid examination of Facebook, designed to come up with the modern equivalent of the EdgeRank formula, would be derailed by both of these problems; the black box prevents looking at all the necessary, interlocking factors affecting the algorithm (Hughes 1987), and the constantly-changing nature of Facebook would quickly render any formal description via equations obsolete. A more lasting approach should look at what endures beyond changeable code, and what is on display instead of hidden: the values that drive the decisions inside the black box.

Methods

To access the values that drive the Facebook News Feed, this study turns to public text artifacts. Content analysis is used to infer the value that Facebook places on particular concepts in the News Feed based on frequency of mention (Krippendorff 1989). Context is given to these frequencies by the deployment of material culture analysis, the practice of using artifacts to identify the beliefs (including values) of a given group during a given time period (Prown 1982). It uses a combination of quantitative and stylistic/substantive analysis, both of which are employed here.

Sources

In material culture analysis, there is no chance for a dialogic interpretation of text, as the subjects/creators of artifacts are often dead (Hodder 2000). Facebook is not forthcoming about the inner workings of the News Feed, limiting the possibility of dialogue about the News Feed artifact. To address this, this study evaluates artifacts associated with the News Feed, specifically those directly generated by Facebook.

The "artifacts" in this case are the documents that Facebook itself has created that deal with the News Feed, in the form of four publicly available sources: Facebook's Newsroom blog,

Facebook's Notes blog, Facebook's patent filings, and Facebook's Securities and Exchange Commission filing from its initial public offering. These four groups of content can be read, respectively, as the way Facebook says the News Feed works and the way (on paper, at least) that the News Feed does actually work.

In this type of analysis, purpose and audience need consideration (Hodder 2000). Additionally, the fact that those writing for a public audience will potentially express different, strategically sanitized values than those writing for a utilitarian purpose (Prown 1982) must be accounted for. As such, we will pause here to examine the context of each source. All four are commissioned, signed texts that went through an editing process, but the level of editing and intended audience varies, and there is a mix of general public-facing and primarily technical, utilitarian documentation.

Newsroom. Posts on the Facebook Newsroom blog are fully vetted, highly edited press releases from Facebook's main press site. Their audience is an official audience (journalists, regulators, etc.) and the language is somewhat formal. They are meant to inform, but through the lens of the best interests of the company. They are currently the primary source of news from within Facebook, and should represent the current state of affairs.

Notes. The Notes are blogs written by Facebook employees in a less formal style, often sharing technical updates and accomplishments as well as site policy and feature updates. They are written less formally, suggesting an audience of not just the press but also everyday users. They have a personal dimension, and are often signed by individual engineers or officers. They are less edited than Newsroom posts; at times, they read like personal letters from executives like Mark Zuckerberg. They were more prominent in the early days of the service, and are no longer posted; the information they once conveyed is now contained in the Newsroom, and as such some of the values represented may be older. They are distinct from the current notes posted by Facebook's own internal research and data science teams; those posts are not considered here.

Patents. Patents are the legal documents defining the technology behind Facebook. They are written for a highly technical audience composed primarily of experts and regulators, and as such are very formal. They are also extremely detailed, as they need to give enough detail to protect the company against intellectual property theft. They have a legal requirement to be truthful, unlike the Notes or Newsroom posts, but are also opaque to the average user. They are slightly behind the curve in terms of technical details, as there is at least an 18-month lag between feature creation and filing.

SEC Filing. Facebook's initial public offering filing with the SEC is a legal document detailing Facebook's financial situation and offerings. It has a similar legal requirement of truth as the patents, and is also similarly opaque to everyday users. It is directed at investors and regulators, and reads as half technical document, half sales pitch. It was created as a snapshot of Facebook on the day it went public, but contains many forward-looking statements.

Data Collection

Articles to be analyzed were scraped from the above sources via scripts which pulled content via links from the following listings:

- Newsroom: A search of the entire Facebook newsroom blog for any posts containing a case-insensitive variation of the term "news feed," including the sometimes-used "NewsFeed" (n = 108).

- Patents: A Google Patents search for any patents filed with the US Patent Office with Facebook, Inc. as the assignee and the case-insensitive words “news feed” or “newsfeed” anywhere within the document (n = 92).
- Notes: A search of the Facebook Notes page for the term “feed” (n = 132).
- The SEC filing was downloaded directly, as it is a single document.

Each script pulled the headline and full text of the article, along with metadata including date and tags. These were then run through a processing script, readying them for quantitative analysis and qualitative coding. All basic text processing and analysis was completed using R (R Core Team 2014). Quantitative analysis was done using the tm text mining package (Feinerer, Hornik & Meyer 2008). Aggregation of qualitative coding consisted of simple additive aggregation of the code sheets, which were then broken down by document source.

Coding

Code building was a multi-step process that continued throughout the study. Codes were primarily inductive, grounded directly in the actual texts to be analyzed (Glaser & Strauss 1967). This was achieved through an iterative coding process that started out with quantitative text analysis through word counts and keyword in context analysis (Ryan & Bernard 2000). Deductive information based on observation of Facebook through regular use was mixed in to the initial coding process in order to follow a grounded theory approach where deductive codes provide a theoretical framework and inductive codes provide a direct connection to the data (Berg & Lune 2012).

The codebook instructed the coder to evaluate each document for what is said and written about the News Feed in terms of the importance of various inputs to the feed. “Importance” is defined here as influence over the behavior of the News Feed. In particular, close attention was paid to repetition of a certain input, as well as formal or informal rankings or relationships between inputs. The coder recorded this information in an Excel spreadsheet with quasi-binary variables for each of the distinct inputs identified in the literature and the quantitative text analysis; a 0 indicated the input was not important according to the document, and a 1 indicated that an input was important. Additionally, a 2 could be recorded to indicate the most important inputs in a situation where a document mentioned or ranked multiple inputs. The coder could also flag documents that detailed changes to the News Feed, as well as documents that specifically needed further review in the substantive close reading phase.

The initial codebook was used by an undergraduate research assistant to code each document, and in the grounded theory tradition the coder was able to suggest new codes and theoretical notes to be integrated into further theory (Glaser & Strauss 1967; Ryan & Bernard 2000). Several new codes were added during open coding. The coder also had the option of marking an article as irrelevant, as the wide search parameters above were sure to turn up some false positives. Use of this flag resulted in a final n of 186 (72 Newsroom posts, 52 patents, 61 Notes blogs, and the SEC filing). A larger n would be preferable, but this study was focused on Facebook’s own output, which was limited; it does succeed in looking at the entire population of official Facebook statements on the News Feed up to April 2015.

As a check of intercoder reliability, I re-coded a portion of the undergraduate research assistant’s coding; intercoder agreement was at 91.58 percent with a Cohen’s kappa of 0.857, indicating a reliable coding strategy. The results were then aggregated based on the number of times each value was coded as important across the document set. Finally, I performed a close reading of the more technical documents (the patents and SEC filing) in light of their original

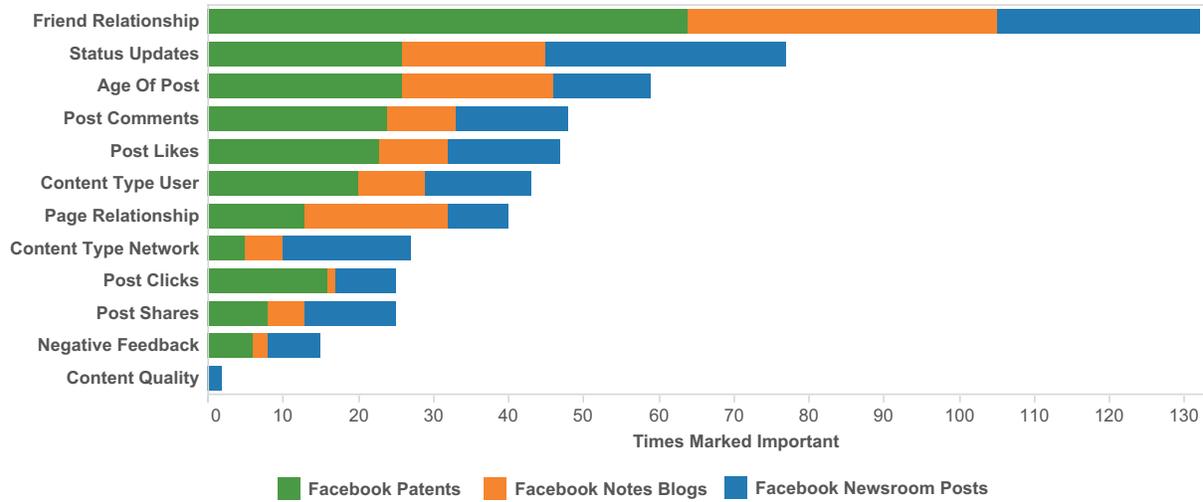


Figure 1: Aggregate importance of Facebook News Feed input factors

contexts, paying special attention to any documents flagged by the coder as needing further review, detailing changes, or containing comparative rankings of News Feed inputs.

Results

The aggregate results of coding the Newsroom posts, Notes blogs, and patent filings (Figure 1) reveal key temporal differences between the document sets. This is likely directly related to when the documents were produced. As Figure 2 illustrates, the volume of documents generated by each source has changed considerably over the past decade, with Notes blogs dominating Facebook’s public output in the early days of the service, peaking in 2009, and Newsroom posts steadily gaining prominence from 2008 onwards. Based on this change over time, we can conclude that the patents and Notes blogs follow an earlier model of the News Feed, closer to the vision of the algorithm that EdgeRank embodied; this tells us where the

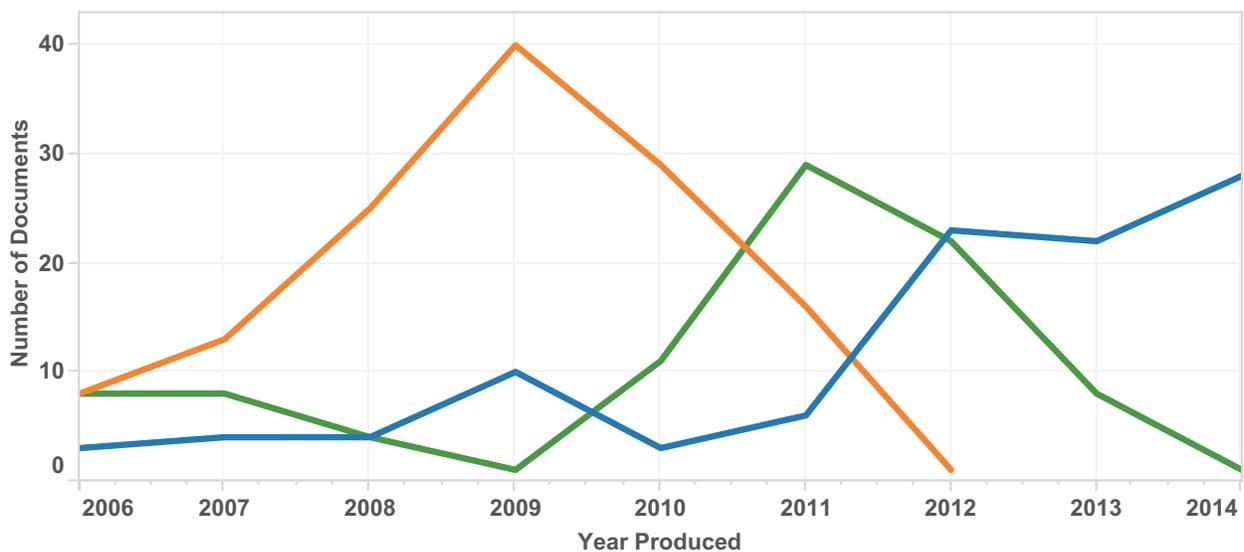


Figure 2: Document volume per source over the period 2006-2014

network has been. The Newsroom results tell us where the network is going and where it may be right now. Based on this difference between the Newsroom results and the rest, it is fair to say that Facebook’s values are in the process of changing. This is unsurprising, and is again backed up by the data; overall, 93.5 percent of the posts analyzed documented an update of some kind to the News Feed – change is endemic to Facebook. The results below are best taken as a snapshot at this point in time, per the tenets of material culture analysis.

Aggregate results reveal that friend relationships are by far the most important factor in the Facebook News Feed, followed by status updates. When the results are ranked by source (see Table 1), status updates trump friend relationships in the Facebook Newsroom posts. This is not a true mismatch, as both are clearly the algorithm’s strongest values, and may represent a technical reality of friends as the key factor, and a public relations reality of representing the News Feed as totally responsive to one’s own input. As such, the top two algorithmic values of the News Feed appear to be **friend relationship** associated with a piece of content and **explicitly expressed user interests** determined through text mining of status updates.

The age of a post is third in the aggregate ranking, and a point of interest, as this contradicts a number of accounts in the trade press (e.g., McGee 2013). This may be a function of the older data points in the patents and Notes where post age is ranked third and second, respectively, while in the more recently updated Newsroom posts it is ranked seventh. As such, **post age** is clearly an algorithmic value, but likely ranks lower than aggregate results suggest.

Engagement behavior comes next, with comments on posts and likes on posts ranking third and fourth, respectively, in aggregate. A third type of engagement behavior, shares, also appears to be gaining importance based on more recent documents. Taken together, the importance the documents put on these three inputs suggests that **prior user engagement** on posts is a key part of Facebook’s value structure.

The News Feed also values input that it extrapolates from user behavior. User content-type preferences as well as user clicks are the sixth and ninth ranked inputs respectively, so are less important values, but values nonetheless, and will be grouped as **implicitly expressed user preferences**.

The remaining values are difficult to accurately rank, as their representation varies across sources. **Page relationship**, the linkage of a user to the page of origin for a post, is prominent in the Notes blogs but middling in the patents and minor in the Newsroom posts, suggesting it may have been recently devalued. **Platform priorities**, as represented by network-preferenced

Facebook Newsroom	Facebook Patents	Facebook Notes Blogs
Status Updates (32)	Friend Relationship (64)	Friend Relationship (41)
Friend Relationship (27)	Status Updates (26)	Age Of Post (20)
Content Type Network (17)	Age Of Post (26)	Status Updates (19)
Post Likes (15)	Post Comments (24)	Page Relationship (19)
Post Comments (15)	Post Likes (23)	Post Likes (9)
Content Type User (14)	Content Type User (20)	Post Comments (9)
Age Of Post (13)	Post Clicks (16)	Content Type User (9)
Post Shares (12)	Page Relationship (13)	Post Shares (5)
Post Clicks (8)	Post Shares (8)	Content Type Network (5)
Page Relationship (8)	Negative Feedback (6)	Negative Feedback (2)
Negative Feedback (7)	Content Type Network (5)	Post Clicks (1)
Content Quality (2)	Content Quality (0)	Content Quality (0)

Table 1: Facebook News Feed input factors ranked by importance.

content types, is minor in the Notes blogs and patents, but ranks third overall within the Newsroom results; it is only recently that Facebook has started privileging video at a feed level (Somaiya et al. 2015), and so this can be taken as more prominent in the more recent versions of the feed. A similar trend can be seen for **negatively expressed preferences** in the form of negative feedback (e.g., the “see fewer post like this” tool). Finally, the Newsroom results are the only ones to mention **content quality**, a factor that is highly discussed as of late in the tech press as a major focus area for Facebook (Kacholia 2013). Clearly, all of these are algorithmic values, but values in flux. Erring on the side of building a model of algorithmic values that will endure for future versions of the algorithm, we will temporarily rank them in order of importance with an emphasis on the most recent updates.

SEC Filing

As is standard in a grounded content analysis, the full-text coding was followed up by close reading and analysis of key documents (Ryan & Bernard 2000), starting with Facebook’s SEC filing from its public offering in 2012. Though not included in the aggregate analysis, the SEC filing is a look into Facebook’s financial priorities and how they explicitly represent themselves as a business.

Overall, this document provides additional support for friend relationships as Facebook’s number one algorithmic value, as sharing and the commerce opportunities that come with it are filtered through (and enabled by) this lens of social connections. News Feed appears to be at the core of this, judging by its description in the products section:

“News Feed. The Facebook News Feed is the core feature of a user’s homepage and is a regularly updating list of stories from friends, Pages, and other entities to which a user is connected on Facebook. It includes posts, photos, event updates, group memberships, app updates, and other activities. Each user’s News Feed is personalized based on his or her interests and the sharing activity of the user’s friends. Stories in a user’s News Feed are prioritized based on several factors, including how many friends have Liked or Commented on a certain piece of content, who posted the content, and what type of content it is. News Feed is a key component of our social distribution capability.”

(Facebook 2012, 82)

The key theme of the entire sales pitch contained within the document is Facebook’s strength at social connectivity, and the business possibilities it opens up. The opening summary explicitly lays out social connectivity as Facebook’s central feature:

“People use Facebook to stay connected with their friends and family, to discover what is going on in the world around them, and to share and express what matters to them to the people they care about.” (Facebook 2012, 1)

The pitch to application developers and advertisers that follows builds on this, focusing on the ability to use social connections, and the data they generate, for money-making activity. In fact, most of the section that discusses advertising products is about contextual advertising based on the actions of friends, such as check-ins at a business or liking a brand page or post.

Early on in the document, Facebook lays out a list of their priorities in “creating value” for users (Facebook 2012, 2), developers (3), and advertisers (3) that follows a similar pattern. Most of Facebook’s priorities regarding users have to do with facilitating increasing numbers of monetizable social connections and providing more data to the platform. Their priorities regarding developers focus on using friend links to distribute products and personal data to customize experiences. Their priorities regarding advertisers center around using user data for

targeting and social relationships for reach. This all lends additional support to explicitly expressed user interests and implicitly expressed user preferences as algorithmic values.

Interestingly, one of the top risk factors listed in the summary risk protection section of the filing is privacy laws (5). This, too, tells us something about Facebook's values: with strict data protection and privacy laws, the data mining and social connections of Facebook would be at serious risk, and potentially inoperable. This would effectively kill Facebook's business. As such, despite references to user privacy in the Newsroom and Notes blogs, we will not include privacy here as an algorithmic value.

Patents

The original core of the News Feed is described in US patent 8,583,690 (Sittig & Zuckerberg 2013). It has been updated several times since its initial filing in 2005, but in the abstract, it makes the value the News Feed started with clear:

*“A system, method, and computer program for generating a social timeline is provided. A plurality of data items associated with at least one relationship between users associated with a social network is received, each data item having an associated time. **The data items are ordered according to at least one relationship.** A social timeline is generated according to the ordered data items.”* (Sittig & Zuckerberg 2013, emphasis mine)

The patent goes on to detail how this socially-based timeline module uses relationship data to generate personalized stories. Other data is incorporated to generate personalized stories, but friends always come first. This principle is well-supported by later patents; all member actions are seen by the system through the lens of friends who might do the same, friends of friends who might do the same, etc., according to patent 7,827,208 (Bosworth, Cox, Sanghvi, Ramakrishnan & D'angelo 2008), which deals with generating relationship-based personalized content. The business reasons behind this are clear and consistent – patent 8,655,987 (Schoen et al. 2014) deals with the creation of sponsored stories, a key source of Facebook's advertising revenue, and it explicitly establishes that sponsored posts are based on your social graph. They are looking for a “match in activity stream” that's based on what your friends are associated with; no friend data, no sponsored stories – and no revenue.

The focus on relationships is made even more clear by a patent filed a year after the original, right as the News Feed began to take off. US patent 8,171,128 (Zuckerberg et al. 2012), on creating a feed of media content based on member interactions, explicitly lays out that media content is selected based on the user's connections with other users. If there are no friend connections, the News Feed as described simply can not function. It also makes plain that user relationships determine the ranking/order of content as well, with additional input from a monitoring module, which looks at the user's reaction to the content the algorithm has selected.

All of this can be read as a clear confirmation that friend relationships are not only a core algorithmic value but one that moderates and informs all the other values. However, the patents also complicate this value, making clear that friend relationships are not simple links. Facebook is far beyond binary relationships; patent 8,402,094 (Bosworth & Cox 2013) establishes that scaled affinity scores are created between individual users and other elements in the system, including other users, pages, keywords, applications, and other content. Interestingly, a passage within this patent suggests that current relationships, as they stand now, may be the precise form of friend relationship in play when friendships act as a moderating lens. In Facebook's example, a current girlfriend would rank higher than an ex girlfriend, despite the fact that there is much more built-up affinity between the user and an ex girlfriend than a brand new girlfriend – unless

the user's activity log still shows them obsessing over the ex regularly. It appears friend relationships are actually that fine-grained.

Additionally, explicit user interests and implicit user preferences are deeply tied into multiple algorithmic systems, including the News Feed, according to the patents. Facebook is actively tracking our "movements" through the system and assigning relationships based on any interaction with any element in the system, according to US patent 7,699,123 (Zuckerberg et al. 2010). Facebook is even reading into the individual words we use to harvest useful data about us. Patent 8,824,764 (Nowak & Eckles 2014) documents how linguistic data can be used to infer personality characteristics, which in turn is used for "targeting, ranking, selecting versions of products, and various other purposes." This, along with similar evidence from the SEC filing above, suggests that both of these values should be highly ranked among the News Feed's algorithmic values.

Discussion

We can now formulate a model of the News Feed's algorithmic values that parallels traditional news values in both structure (an easily-parseable list) and purpose (story selection). The nine algorithmic values we have identified, in descending order of influence over the News Feed, are friend relationships, explicitly expressed user interests, prior user engagement, implicitly expressed user preferences, post age, platform priorities, page relationships, negatively expressed preferences, and content quality. Evidence from the close readings has shifted the order from the initial formulation at the top of the results section, but the clear primacy of friend relationships continues to inform the entire structure. In fact, the patents in particular have established that friend relationships are a guiding value that mediates the application of all the other values.

This is the first of two key differences between news values, the old drivers of information relevance, and algorithmic values, their potential replacement. The standard news values listed by Lee (2009) act in concert and tension with one another; no single one is dominant, and it is a combination of these values that ultimately results in story selection (Harcup & O'Neill 2001). A combination of algorithmic values is still required for story selection, but friend relationships eclipse and moderate all these values; a news judgment, for example, can leave out any of the news values, but Facebook can never leave out friend relationships. Even the underlying journalistic drive to scoop the competition (Schultz 2007) is simply an extra motivator, not an active influence over the expression of the other values. Friend relationships, in contrast, explicitly moderate how the other algorithmic values will be expressed, even down to the level of what text is displayed to users.

The second major difference between news and algorithmic values is the core concerns they can be boiled down to. Shoemaker & Cohen (2005) argue that all news values are ultimately either about deviance (e.g., abnormality, breaking the status quo) or social significance (e.g., impact on society). The underlying concerns of the News Feed's algorithmic values are fundamentally different. All nine of the identified algorithmic values can be reduced down to personal significance, e.g., impact to self, interests, and friends. This is a radical departure from traditional news values, and brings longstanding concerns about the role of personalization in selecting news content to the forefront.

Many theorists (e.g., Bozdag 2013; Gillespie, 2014; Pariser 2011; Sunstien 2001) have warned of the potential for personalization to form filter bubbles and feedback loops that reduce exposure to counterattitudinal information, cutting off the debates and exchanges of ideas that

are central to the operation of a democracy. Empirical work has since confirmed that this sort of enforced selective exposure leads directly to the sort of polarization (Stroud 2010) that has been concerning academics for the past 15 years. A platform that delivers news through a filter that is heavily weighted towards the personally-focused algorithmic values identified here may bake this potential for polarizing personalization directly into its design. On Facebook in particular, this could be compounded by the role we have identified for friend connections as the overarching lens through which other values are focused. Friends on social media tend to be homophilic and ideologically clustered (Gaines & Mondak 2009; Kushin & Kitchner 2009); this potentially places the lens through which the News Feed algorithm filters all other values firmly within your personal bubble to begin with.

Of course, this algorithmically-driven personalization of news is not exclusive to Facebook, or to social media platforms. Algorithmically-driven systems have become thoroughly embedded in all areas of journalism, to the extent that Primo and Zago (2015) assert that the act of journalism itself is now best viewed as not exclusive to humans, but rather an interaction between human journalists and technological tools. This extends to editorial decision making and content curation within traditional news organizations, where algorithmically-driven personalization is a popular strategy for attempting to maintain relevance with users (Thurman & Schifferes 2012). Facebook, however, stands to have a wider influence over curation than any one news organization, with its large, diverse user base, and increased efforts to act as a publication platform through expanding Instant Articles (Alpert 2016).

In addition to all of this, data collection appears to be a core value of the Facebook platform overall. It is not explicitly tied into the News Feed's story selection, but, based on the close readings above, appears to be a motivating concern. Facebook does appear to be explicitly designed to value disclosure (Andrejevic 2013; Light & McGrath 2010) and many of the identified algorithmic values either serve the purpose of prompting the disclosure of additional data or directly rely on it for proper operation. This leaves us with a combination of a business and a personal concern as the core of Facebook's algorithmic values, in direct contrast to the combination of a novelty and a societal concern that drive news values.

Clearly, Facebook's algorithmic values are very different, in both content and underlying structure, from traditional news values. As one replaces the other, it is essential to be mindful of this and continue to research how this impacts information flow and its social and civic consequences. This includes applying the methods used to verify and understand journalism in the 20th century, especially in-depth ethnographic research on the internal decision-making culture and process of Facebook. It is also essential to call for organizations like Facebook to publicly reflect on their growing influence over content curation. Facebook has shown an institutional desire to explore how its systems affect information flow and polarization (e.g., Bakshy, Messing & Adamic 2015), but has yet to directly address their expanded role as a key news information provider.

Limitations

This study is not the final word on Facebook's algorithmic values; we know that this examination is limited by the transient nature of Internet technology (Karpf 2012) and will need to be updated periodically. In particular, the News Feed's values must be reexamined in the light of emerging trends like the increased emphasis on content quality and platform-established content priorities as well as the push to have news organizations post in a Facebook-native

format (Somaiya et al. 2015). It is possible that, within the space of a few years, these values may become much more prominent in the News Feed as primary drivers of story selection, especially for news information. Indeed, recent developments, such as moves to prioritize a new type of live video service introduced by Facebook (Facebook 2016) and expand the prominence and footprint of Instant Articles (Alpert 2016), suggest that this transition is underway.

Additionally, this study still suffers from some of the limitations of the fact that the News Feed exists in a black box (Pasquale 2011). The outside confirmation that material culture analysis usually seeks is limited here to the close readings, as direct quantitative manipulation of the News Feed is extremely difficult, and actively policed by Facebook. Future work could potentially expand on this by adding additional sources, such as press articles with interviews from Facebook staff and research co-authored by Facebook engineers.

Conclusion

Facebook's own patents make clear that a simple equation for describing the News Feed will not be forthcoming; US patent 8,768,863 (Gubin, Kao, Vickry & Maykov 2014) clearly establishes that the modern News Feed is not just a weighted formula with thousands of inputs, but rather a constantly updated, personalized machine learning model, which changes and updates its outputs based on your behavior, the behavior of people you are connected with, and the behavior of the affinity and personality-based sub-group of users the system judges you to belong to. Facebook's formula, to the extent that it actually exists, changes every day. However, the values that drive it do not; they change, but much more slowly, and in a much more human-readable way. This study has identified the set of algorithmic values that drive the News Feed, how they compare to one another, and how they compare to the traditional drivers of our news information flow. With this information, and a commitment to reexamine these values periodically, we can begin to understand and investigate this influential algorithm in a lasting, cross-disciplinary manner, and perhaps understand other giants in algorithmic curation (e.g., Google Search and Google News) using the same values-based approach.

REFERENCES

- Alpert, Lukas I. 2016. "Facebook Will Make Instant Articles Available to All Publishers." *The Wall Street Journal*, February 17. <http://www.wsj.com/articles/facebook-opens-up-instant-articles-to-all-publishers-1455732001>.
- Andrejevic, Mark. 2013. *Infoglut: How too much information is changing the way we think and know*. New York, NY: Routledge.
- Bakshy, Eytan, Solomon Messing, and Lada Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348 (6239): 1130-2.
- Bawden, David, and Lyn Robinson. 2009. "The dark side of information: overload, anxiety and other paradoxes and pathologies." *Journal of Information Science* 35 (2): 180-91.
- Bennett, W Lance, and Shanto Iyengar. 2008. "A new era of minimal effects? The changing foundations of political communication." *Journal of Communication* 58 (4): 707-31.
- Berg, Bruce L., and Howard Lune. 2012. *Qualitative research methods for the social sciences*. 8th ed. Boston, MA: Pearson.
- Bimber, Bruce. 2003. *Information and American democracy: Technology in the evolution of political power*. Cambridge, UK: Cambridge University Press.
- Bosworth, Andrew, and Chris Cox. 2013. Predicting user responses to invitations in a social

- networking system based on keywords in user profiles. US 8145584 B2.
- Bosworth, Andrew G., Chris Cox, Ruchi Sanghvi, Thyagarajapuram S. Ramakrishnan, and Adam D'Angelo. 2008. Systems and methods for generating dynamic relationship-based content personalized for members of a web-based social network. US 20080040370 A1.
- Bozdag, Engin. 2013. "Bias in algorithmic filtering and personalization." *Ethics and Information Technology* 15 (3): 209-227.
- Bucher, Taina. 2012. "Want to be on the top? Algorithmic power and the threat of invisibility on Facebook." *New Media & Society* 14 (7): 1164-1180.
- Caers, Ralf, Tim De Feyter, Marijke De Couck, Talia Stough, Claudia Vigna, and Cind Du Bois. 2013. "Facebook: A literature review." *New Media & Society* 15 (6): 982-1002.
- Churchill, Elizabeth F. 2013. "Putting the person back into personalization." *interactions* 20 (5): 12-5.
- Diakopoulos, Nicholas. 2015. "Algorithmic Accountability: Journalistic investigation of computational power structures." *Digital Journalism* 3 (3): 398-415.
- Domingos, Pedro. 2005. "Mining social networks for viral marketing." *IEEE Intelligent Systems* 20 (1): 80-2.
- Duggan, Maeve, Nicole B. Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. 2014. "Social Media Update 2014." Washington DC: Pew Research Center. <http://www.pewinternet.org/2015/01/09/social-media-update-2014/>
- Facebook. 2012. "Form S-1 Registration Statement." <https://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm>.
- . 2016. "News Feed FYI: Taking into Account Live Video When Ranking Feed." <http://newsroom.fb.com/news/2016/03/news-feed-fyi-taking-into-account-live-video-when-ranking-feed/>.
- Feinerer, Ingo, Kurt Hornik, and David Meyer. 2008. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25 (5): 1-54.
- Fingerhut, Hannah. 2016. "Millennials' views of news media, religious organizations grow more negative." <http://www.pewresearch.org/fact-tank/2016/01/04/millennials-views-of-news-media-religious-organizations-grow-more-negative/>.
- Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in computer systems." *ACM Transactions on Information Systems (TOIS)* 14 (3): 330-47.
- Gaines, Brian J., and Jeffery J. Mondak. 2009. "Typing together? Clustering of ideological types in online social networks." *Journal of Information Technology & Politics* 6 (3-4): 216-231.
- Gallup. 2015. "Americans' Trust in Media Remains at Historical Low." <http://www.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx>
- Galtung, Johan, and Mari Holmboe Ruge. 1965. "The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian newspapers." *Journal of Peace Research* 2 (1): 64-90.
- Gans, Herbert J. 1979. *Deciding What's News: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. Evanston, IL:Northwestern University Press.
- Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Pablo Boczkowski and Kirsten Foot, 167-193. Cambridge, MA:MIT Press.
- Glaser, Barney, and Anselm Strauss. 1967. "The discovery grounded theory: strategies for

- qualitative inquiry." Chicago: Aldin.
- Gubin, Max, Wayne Kao, David Vickrey, and Alexey Maykov. 2014. Adaptive ranking of news feed in social networking systems. US 20130031034 A1.
- Harcup, Tony, and Deirdre O'Neill. 2001. "What is news? Galtung and Ruge revisited." *Journalism Studies* 2 (2): 261-80.
- Hodder, Ian. 2000. "The interpretation of documents and material culture." In *Handbook of Qualitative Research*, edited by Norman Denzin and Yvonna Lincoln, 393-405. Thousand Oaks: Sage.
- Hughes, Thomas P. 1987. "The evolution of large technological systems." In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, edited by Wiebe Bijker, Thomas Hughes, and Trevor Pinch, 51-82. Cambridge: MIT Press.
- Introna, Lucas D, and Helen Nissenbaum. 2000. "Shaping the Web: Why the politics of search engines matters." *The Information Society* 16 (3): 169-85.
- Kacholia, Varun. 2013. "News Feed FYI: Showing More High Quality Content." <http://newsroom.fb.com/news/2013/08/news-feed-fyi-showing-more-high-quality-content/>.
- Kao, Wayne, Daniel Schafer, and Joshua Van Dyke Watzman. 2014. Action clustering for news feeds. US2014/0143258A1.
- Karpf, David. 2012. "Social science research methods in Internet time." *Information, Communication & Society* 15 (5): 639-661.
- Kincaid, Jason. 2010. "EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick." <http://techcrunch.com/2010/04/22/facebook-edgerank/>.
- Kraemer, Felicitas, Kees van Overveld, and Martin Peterson. 2011. "Is there an ethics of algorithms?" *Ethics and Information Technology* 13 (3): 251-60.
- Krippendorff, Klaus. 1989. "Content analysis." In *International Encyclopedia of Communication*, edited by E. Barnouw, G. Gerbner, W. Schramm, T.L. Worth and L. Gross, 403-7. New York, NY: Oxford University Press.
- Kushin, Matthew J., and Kelin Kitchener. 2009. "Getting political on social network sites: Exploring online political discourse on Facebook." *First Monday* 14 (11).
- Lee, Jong Hyuk. 2009. "News values, media coverage, and audience attention: An analysis of direct and mediated causal relationships." *Journalism & Mass Communication Quarterly* 86 (1): 175-90.
- Lievrouw, Leah A. 2012. "The Next Decade in Internet Time: Ways ahead for new media studies." *Information, Communication & Society* 15 (5): 616-38.
- Light, Ben, and Kathy McGrath. 2010. "Ethics and social networking sites: a disclosive analysis of Facebook." *Information Technology & People* 23 (4): 290-311.
- Mager, Astrid. 2012. "Algorithmic ideology: How capitalist society shapes search engines." *Information, Communication & Society* 15 (5): 769-87.
- Matsa, Katerina E., and Amy Mitchell. 2014. "8 Key Takeaways about Social Media and News." Washington DC: Pew Research Center. <http://www.journalism.org/2014/03/26/8-key-takeaways-about-social-media-and-news/>
- McCombs, Maxwell E, and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public Opinion Quarterly*: 176-87.
- McGee, Matt. 2013. "EdgeRank is Dead: Facebook's News Feed Algorithm Now Has Close to 100k Weight Factors." *Marketing Land*. <http://marketingland.com/edgerank-is-dead->

- facebook-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908
- Mitchelstein, Eugenia, and Pablo J Boczkowski. 2009. "Between tradition and change: A review of recent research on online news production." *Journalism* 10 (5): 562-86.
- Napoli, Philip M. 2014. "Automated media: An institutional theory perspective on algorithmic media production and consumption." *Communication Theory* 24 (3): 340-60.
- Nowak, Michael, and Dean Eckles. 2014. Determining user personality characteristics from social networking system communications and characteristics. US20140074920 A1.
- O'Neill, Deirdre, and Tony Harcup. 2009. "News values and selectivity." In *The Handbook of Journalism Studies*, edited by K. Wahl-Jorgensen and T. Hanitzsch, 161-74. London, UK:Routledge.
- Palmer, Jerry. 2000. *Spinning into control: News values and source strategies*. New York, NY:Leicester University Press.
- Pariser, Eli. 2011. *The filter bubble: What the Internet is hiding from you*. New York: Penguin.
- Pasquale, Frank A. 2011. "Restoring Transparency to Automated Authority." *Journal on Telecommunications and High Technology Law* 9: 235-256.
- Primo, Alex, and Gabriela Zago. 2015. "Who and What Do Journalism? An Actor-network Perspective." *Digital Journalism* 3 (1): 38-52.
- Prown, Jules David. 1982. "Mind in matter: An introduction to material culture theory and method." *Winterthur Portfolio*: 1-19.
- R Core Team. 2014. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing.
- Ryan, Gery W, and H Russell Bernard. 2000. "Data management and analysis methods." In *Handbook of Qualitative Research*, edited by Norman Denzin and Yvonna Lincoln, 769-802. Thousand Oaks: Sage.
- Scheufele, Dietram A. 1999. "Framing as a theory of media effects." *Journal of Communication* 49 (1): 103-22.
- . 2000. "Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication." *Mass Communication & Society* 3 (2-3): 297-316.
- Schoen, Kent, Ning Li, Robert Kang-Xing Jin, Philip Anastasios Zigoris, Jessica Gronski, Jordan Walke, and Eric Michel Giovanola. 2014. Sponsored-stories-unit creation from organic activity stream. US 8,655,987 B2.
- Schultz, Ida. 2007. "The journalistic gut feeling: Journalistic doxa, news habitus and orthodox news values." *Journalism Practice* 1 (2): 190-207.
- Shoemaker, Pamela J, and Akiba A Cohen. 2005. *News around the world: Content, practitioners, and the public*. New York, NY:Routledge.
- Sittig, Aaron, and Mark Zuckerberg. 2013. Managing information about relationships in a social network via a social timeline. US8583690 B2.
- Somaiya, Ravi, Mike Isaac, and Vindu Goel. 2015. "Facebook May Host News Sites' Content." *The New York Times*. <http://www.nytimes.com/2015/03/24/business/media/facebook-may-host-news-sites-content.html>
- Steiner, Christopher. 2012. *Automate this: How algorithms came to rule our world*. New York: Portfolio/Penguin.
- Stroud, Natalie Jomini. 2010. "Polarization and partisan selective exposure." *Journal of Communication* 60 (3): 556-76.
- Sunstein, Cass. 2001. *Republic.com*. Princeton, NJ:Princeton University Press.
- Taylor, Dan. 2011. "Everything you need to know about Facebook's EdgeRank."

- <http://thenextweb.com/socialmedia/2011/05/09/everything-you-need-to-know-about-facebooks-edgerank/>.
- Thurman, Neil, and Steve Schifferes. 2012. "The future of personalization at news websites: lessons from a longitudinal study." *Journalism Studies* 13 (5-6): 775-90.
- Tufekci, Zeynep. 2015. "Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency." *J. on Telecomm. & High Tech. L.* 13: 203-217.
- Vallone, Robert P, Lee Ross, and Mark R Lepper. 1985. "The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre." *Journal of Personality and Social Psychology* 49 (3): 577-585.
- Wilson, Robert E, Samuel D Gosling, and Lindsay T Graham. 2012. "A review of Facebook research in the social sciences." *Perspectives on Psychological Science* 7 (3): 203-20.
- Zuckerberg, Mark, Andrew Bosworth, Chris Cox, Ruchi Sanghvi, and Matt Cahill. 2012. Communicating a newsfeed of media content based on a member's interactions in a social network environment. US 8,171,128B2.
- Zuckerberg, Mark, Ruchi Sanghvi, Andrew Bosworth, Chris Cox, Aaron Sittig, Chris Hughes, Katie Geminder, and Dan Corson. 2010. Dynamically providing a news feed about a user of a social network. US 7,669,123B2.