Opening the black box of perceived quality: Predicting endorsement on a blog site

Sotirakou, C.; Trilling, D.; Germanakos, P.; Mourlas, C.

DOI
10.1145/3350546.3352553

Publication date
2019

Document Version
Final published version

Published in
Proceedings: 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2019)

Opening the black box of perceived quality: Predicting endorsement on a blog site

Catherine Sotirakou
Faculty of Communication and Media Studies, National & Kapodistrian University of Athens
Athens, Greece
katerinasot@media.uoa.gr

Damian Trilling
Amsterdam School of Communication Research, University of Amsterdam
Amsterdam, The Netherlands
d.c.trilling@uva.nl

Panagiotis Germanakos
User Experience ICD, Product Engineering, SAP SE
Walldorf, Germany
panagiotis.germanakos@sap.com

Constantinos Mourlas
Faculty of Communication and Media Studies, National & Kapodistrian University of Athens
Athens, Greece
mourlas@media.uoa.gr

ABSTRACT

Uncovering their readers' perceptions is of key importance for every news media organization to find methods to improve the quality of their product. It has the potential to facilitate journalists' work in attracting attention and gaining a loyal audience. Discovering which elements of a news story influence readers' perceptions has been a cross-disciplinary research goal for the past years, because it can play a crucial role in news dissemination and consumption in the digital age. Drawing upon literature in the various areas such as journalism, psychology, computer science, and AI, this paper proposes a machine learning approach that explores three dimensions of article features that can help predicting the online behavior of the reader. Results show that how the story is written, the topic, and certain aspects of the author's online reputation can affect reader endorsements and the perceived quality of an article.

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**; • **Applied computing** → **Document management and text processing**.

KEYWORDS

Natural Language Processing, News, Computational Journalism, Computer-assisted Content Analysis, Machine Learning

ACM Reference Format:

Catherine Sotirakou, Damian Trilling, Panagiotis Germanakos, and Constantinos Mourlas. 2019. Opening the black box of perceived quality: Predicting endorsement on a blog site. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI '19), October 14–17, 2019, Thessaloniki, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3350546.3352553

WI '19, October 14–17, 2019, Thessaloniki, Greece

ACM ISBN 978-1-4503-6934-3/19/10…$15.00
https://doi.org/10.1145/3350546.3352553

1 INTRODUCTION

Likes, shares, and comments can determine how many people have a chance to see social media content, because they serve as an input for the filtering algorithms that decide the ordering of the posts in every news feed. Communication scientists have started to investigate the so-called "shareworthiness" of a news story: features of an online news article (such as its novelty, tone, the actors involved, etc.) that help predicting the number of shares an article will receive [9, 36, 37]. Yet, such content features are not the only features to consider to understand what explains engagement with online news and information. Accordingly, Orellana-Rodriguez and Keane [28] distinguished in their survey on news dissemination on Twitter between user features, content features, and context features. Context features, for instance, include temporal features and location features, while user features relate to the features of the account from which the original information was published.

In the current work, we consider these three feature dimensions and we extend them by additionally proposing "style" as a category within content features and a set of new features that best express the three dimensions. We apply this framework to a new data set extracted from Medium.com, a well-known blog platform. We explore which feature category explains best how many so-called "claps" (a form of likes) an article receives, and how this differs between different types of articles.

2 RELATED WORK

We can broadly distinguish three types of user engagement metrics: clicks, likes, and shares. Clicks are a sign that a news article caught the user's attention but do not provide much actual feedback. However, when a user consciously decides to hit "like" or "share", that reflects a more personal attachment to and connection with the content and can serve as a valuable indication of the perceived quality of the article. After all, people also use these actions to construct a personal image, because what they share reflects on who they are or want to be [36]. Medium.com, the blog platform we are studying, features an interesting combination of liking and sharing: It allows the user to "clap" for an article, which shows the author that their story was liked; but the number of claps (it is possible to clap up to 50 times for the same story) is also used to determine which stories one's followers will see – thus, it can be seen as an indirect way of sharing as well. It is reasonable to assume that people do not want to be seen as recommending articles that are not good. Given that the claps on Medium.com are public and result not only in feedback to the author, but also in recommendations, we can argue that claps can be used as a proxy to measure the perceived quality of an article. Perceived quality is highly dependent on the consumer's judgment [27]. Accordingly, Aaker defined perceived product quality as "the customer's perception of the overall quality or superiority of the product or service with respect to its intended purpose, relative to alternatives" [1]. We are interested in the perceived quality of online news articles. We follow a very broad definition of news here and follow Bogart [4], who defined news as the "useful current information that is not necessarily generally important, but is of considerable importance to some people". As decisions within news organizations are increasingly driven by metrics of clicks and user engagement [21, 35], it is of crucial importance to find out how to reconcile high quality of content with perceived quality as expressed by user engagement.

Although finding a news story interesting is a matter of personal taste, it is possible to create a taxonomy of so-called news values [8, 12]. This approach suggests that certain attributes of news can influence journalists' and audience news selection [7, 16] as well as engagement with digital news stories [36, 39]. Yet, there is no full consensus yet on which features have the largest impact on the so-called "shareworthiness" [36]. On the one hand, Trilling et al. find geographical proximity, involving Western countries, conflict, human interest, negativity, positivity, and exclusiveness to increase engagement, with slight differences between Facebook and Twitter engagement [36]. Valenzuela et al. confirm the findings for exclusiveness and negativity, and added – amongst other features – deviance and relevance to the list [37]. Garcia-Perdomo et al., in contrast, cannot confirm an influence of proximity, but do confirm an influence of conflict and human interest as well as deviance (here called unusualness) [9]. They also find impact and prominence to be relevant predictors. In contrast to [37], they also found some effect of usefulness. Kilgo et al. concluded that the influence of such news values was less important than emotional appeals; however, their conclusions are based on very small cell sizes [18]. Relatedly, other work investigated how news values differ between sensationalist and non-sensationalist content [17].

3 MODEL & FEATURE EXTRACTION

Our primary concern was to create an inclusive model to measure and quantify the value of perceived quality across various article categories. We structured the backbone of our model using the overarching dimensions proposed by Orellana-Rodriguez and Keane [28], namely (i) user, (ii) content, (iii) context.

3.1 User features

Author's Popularity: Having a large number of followers may result in a great number of votes in a small period of time for a very small piece of text, while new authors have to work hard and produce long pieces of writing to achieve the same result [15].

Another way for us to determine an author's popularity is to look for the number of people they follow, as this can also contribute to build reputation and gain more votes, since users leverage networking communities by befriending other users and ask for votes in return [29].

Author's tone: Authors are also characterized by their tone of speech. Self-referential or self-reflexive messages are frequently used to attract the readers' attention [25], while human-related stories are more likely to get positive feedback [22] and personal pronouns are an indicator of the existence of people in the story. We investigate the use of personal pronouns (I, you, we, etc.), reflexive pronouns (myself, yourself, etc.), and possessive pronouns (mine, yours, etc).

Subjectivity: Social media drive journalists towards more subjective ways of presenting information [20]. We use the Python library TextBlob to measure subjectivity in textual data.

Modality: Grammatical modality is signaled by grammatical moods that express a speaker's general intentions and is implemented grammatically through three moods namely indicative, imperative and subjunctive. To capture the degree of certainty of an author, we used pattern.en [5], which measures modality as a value between -1.0 and $+1.0$, where values $> +0.5$ represent certainty.

3.2 Content features

We focus on content features that we expect to influence the reader. We slightly extend the model of Orellana-Rodriguez and Keane [28], by separating it into structure and styles features.

3.2.1 Structure. Article Length: We used the so-called "reading time" provided by Medium.com.

Images: The use of salient pictures is a way to attract audience attention and achieve link clicks (or other follow-up actions like donating money, see [23]). Additionally, maintaining a good balance of text and images is an important factor for capturing attention and is used extensively for marketing campaigns. Consequently, to study images we use two different features: the total number of pictures in the article and the image-to-text ratio.

Bullet points: Listicles, meaning a mixture of 'list' and 'article' are very popular types of articles and according to research, they tend to be highly shareable [26]. We therefore looked for the existence of bullet points.

Links: We searched for external links in the article, because a news story that contains links may relate to more verified sources, complementary information, and may present different angles. Furthermore, providing extra reading material can be a sign of extensive reportorial effort.

Readability: Readability testing is a way to automatically check the clarity of writing. We use the Flesch-Kincaid Grade Level [19], which has been widely applied to journalistic and political communication research [32, 38].

Novelty: If an article covers new, exclusive information that is not present elsewhere, it may be more interesting for readers and enhance engagement [36]. We used a three-day sliding window and used the normalized (to account for title length) number of nouns (which we determined using spacy [13]) that a headline shared with all concatenated headlines in the window to calculate a popularity score. A low popularity score can be interpreted as an indicator of high novelty [36].

3.2.2 Style. *Beautiful phrasing:* High-quality articles are often written using beautiful language [2, 22], meaning more unusual phrasing and creative words that lead to more positive feedback [34]. We used Term Frequency-Inverse Document Frequency (tf·idf) as a proxy, because the use of rarer words can be seen as an indicator for the use of beautiful language.

Sentiment: Both positivity and negativity have been shown to influence news sharing [34, 36, 37], with some psycho-physiological experiments showing that negative news evokes stronger and more sustained reactions than does positive news [33]. We used Vader [14], which calculates the overall polarity of the text along with the sentiment: positive, negative or neutral. Other studies have focused on evidence that online content that causes awe or anger and anxiety emotions to the users increases the chances of becoming viral [3, 10]. For the emotions extraction, we used the NRC suite of lexica, that is based on Plutchik's model, that suggests the existence of eight basic emotions: joy, sadness, anger, fear, trust, surprise, disgust, and anticipation [24]. For the measurement, we computed the counts of emotion words, each normalized by the total number of article words. Except for polarity and emotions, we also calculated the overall emotional density expressed in the text.

Title polarity: The headline of an article is a significant aspect of every story as modern news readers and social media users scan the headlines before they decide to click [6, 31]. To measure polarity expressed in the headline, again, we used Vader [14].

3.3 Context features

Genre: The topic of the article reflects certain characteristics of its nature and it has been greatly investigated in previous work [2, 22].

Temporal: We have articles from only one year, so we only used hour and type of day as temporal features.

Cultural distance: Previous research [36] found that geographical proximity and stories that mention Western countries increase social media engagement. For our study, we extracted references to geolocations from the texts using the Mordecai system [11], which extracts toponyms and returns their coordinates.

4 METHOD & DATASET

We collected data from Medium.com and retrieved articles from the categories News, Technology, Health, Business, Sports, and Lifestyle, as well as their number of "claps", which we use as a proxy measure for perceived quality. Our data set consists of articles published between September 2017 to September 2018 and were collected automatically in January 2019. The total amount of articles before cleaning was 247, 071 and had a large distribution of claps ranging from 0 to 292K claps. All articles were processed for stop words, nonstandard words and characters removal, stemming, and tokenization. In addition, we removed articles not written in English, NaN values, and HTML code from the text. This left us with 200, 710 (Table 1) cases for inclusion in our models.

Articles with claps above the 99.8th percentile were considered outliers and were removed. Furthermore, we normalized the features and checked for feature correlation. In addition, we conducted a Principal Component Analysis (PCA) to search for redundant features. However, the analysis showed that we could not meaningfully reduce our large set of components into a smaller one without losing a lot of variance.

Table 1: Descriptive statistics

Category	No. Articles	Mean (claps)	Median (claps)	75th percentile (claps)
News	17762	150	0	22
Technology	30883	236	2	59
Business	35768	65	0	14
Health	49072	39	0	3
Sports	12664	19	0	5
Lifestyle	54561	347	22	121

5 DATA ANALYSIS & FINDINGS

We separated articles with low and high claps, and used the scikit-learn [30] implementation of tree-based machine learning approaches (decision tree, random forest, xgboost) for this binary classification task. Tree models can be more interpretable than other complex models and provide more descriptive explanations of the effect of the features.

5.1 Phase A

We randomly sampled our original data set and created a new one that consisted of 60K articles, 10K from each category, (News, Technology, Health, Business, Sports, Lifestyle). From this new dataset, we randomly selected 14K articles from the lowest and 14K articles from the highest quartile of claps.

We used the randomForest package from the Python library scikit-learn [30] to create a classifier of our articles and to measure the importance of the predictor variables, which is calculated during the construction of the decision trees by checking the raise of prediction error when data for that variable is permuted while all others are left unaltered. Based on its F1 score, we decided on using the XGBoost algorithm with an F1-Score of 80%.

Figure 1) presents the importance of the variables of the selected dimensions. The dimension of User is the first with Authors' popularity (Followers & Followees), Modality and Tone, second is the Content dimension and specifically Structure (Article Length, Images, Readability) and Style (Sentiment).

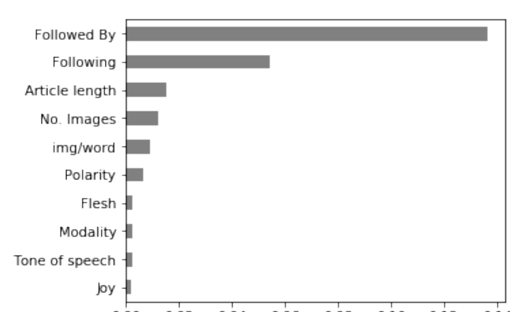

Figure 1: Features importance score for high claps bucket

We observe that the social network of an author, along with length, pictures and polarity are the most important characteristics that influence the perceived quality of an article. Interestingly, the results reported in Figure 1 indicate high accuracy and give an idea about the predictive power of the features, which will be used as the baseline for Phase B.

5.2 Phase B

To check whether some features are more important depending on the genre, we created different data sets for every category and the same process was applied to split the articles into two classes of the same proportion of texts, one bucket with the 25% higher clapped texts and the low-clapped bucket with the ones that got 0 claps. Despite the fact that all articles were taken from a single source, differences exist on the distribution of the claps in every category as shown in Figure 2.

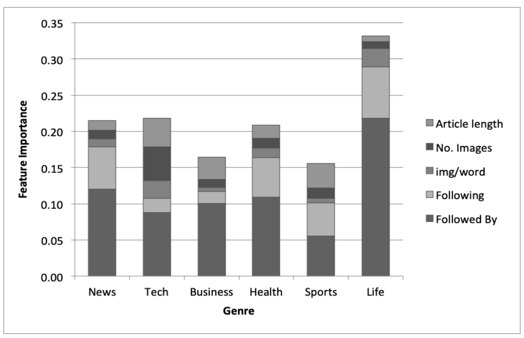

Figure 2: The importance measures of Top-5 features in 6 different article categories

6 DISCUSSION OF THE RESULTS

First, the most important feature is the number of followers an author has, which is the only feature that remains stable in all data sets. This is somewhat expected because the more followers the author has, the more claps their articles will receive. Similarly, the number of people that the author follows influences the perceived quality of their article, meaning that the social network of the author has a positive effect. Second, the length of an article is associated with perceived quality. However, we cannot interpret from the results if longer articles are correlated with better metrics or the opposite is true in this case. Third, images seem to be of great significance in terms of positive feedback. Both the number of pictures and image-to-text ratio are among the top five important features in most data sets. This is a logical consequence since a picture communicates a great deal of information and can make the text more interesting and easy to understand. Fourth, sentiment has an impact on the crowd's endorsements, since the polarity of a text, which translates to the sum of positive and negative scores, appears to be strongly correlated to perceived quality. Moreover, positivity, negativity, neutrality, joy, trust, and anticipation also show up as less important features.

Our findings suggest some differences depending on the category. Specifically, the existence of external links in the body of the article appears to be somewhat important in the case of News and Lifestyle, while beautiful phrasing is fairly relevant to Health, Business and Lifestyle categories. Meanwhile, readability (Flesh score) is considerably associated with perceived quality when it comes to Business articles. Notably, subjectivity and tone are associated with high claps only in Business and Technology respectively. Likewise, the presence of bullet points in the text is quite significant in Technology and Lifestyle, whereas grammatical modality is slightly correlated to Sports and News topics. Surprisingly, the only temporal feature that ended up having a small degree of significance is the hour, which came up in Health and Lifestyle data sets.

The above results show that perceived quality could not be quantified using a unified value across articles' types but rather should be treated as a situation-specific factor.

7 CONCLUSION & FUTURE WORK

We investigated the prediction task of the crowd's endorsement of articles published on the blog platform Medium.com. The best predictor consistently is the number of followers, which is not surprising: if more people see it, more people will clap. Yet, we also found important other predictors that can be interpreted as indicators for the perceived quality of an article. This paper considerably extends a preliminary version of this work [34] by introducing additional features (e.g., from [36]) and including the three dimensions of a news tweet, namely, user, content and context (see [28]). Moreover, we introduce the concept of perceived quality of a news article and we examine which characteristics of a news article predict the reader's positive feedback. To quantify the perceived quality of news articles, the number of "claps" was the dependent variable for our model. Furthermore, in order to define if and to what extent the features contribute in the future online success of an article we used tree-based machine learning algorithms for the prediction tasks. Our findings support what has been showcased already in previous research, that indeed factors related to the author's reputation, along with content-based characteristics of a news article mainly length and images, and sentiment expressed in the text can be predictors of a reader's perceived quality. Results have also shown that an article's type affects the importance of features that reveal its perceived quality differently and should be treated as a situation-specific factor given the nature of each category.

We focused on identifying the role of different feature categories, but not on the nature of their effects. We expect these effects to be non-linear. For example, it is very likely that a feature like the author's popularity is very important to "kick-start" the sharing of an article, but once the article is already popular, this effect may wear off. Next to that, some effects may follow a U-shape (both extremely negative and extremely positive articles may be more likely to be endorsed) or an inverse U-shape (both extremely short (i.e., trivial) and extremely long (i.e., unreadable) articles may not be endorsed. In the current paper, we have provided some guidance on which features may be worth considering, and leave it to future research to disentangle the exact relationship.

ACKNOWLEDGMENTS

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for

Research and Technology (GSRT), under the HFRI PhD Fellowship grant (GA. 14540).

REFERENCES

[1] David A Aaker. 2009. *Managing brand equity.* simon and schuster.
[2] Ioannis Arapakis, Filipa Peleja, Barla Berkant, and Joao Magalhaes. 2016. Linguistic benchmarks of online news article quality. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1893–1902.
[3] Jonah Berger and Katherine L Milkman. 2012. What makes online content viral? *Journal of marketing research* 49, 2 (2012), 192–205.
[4] Leo Bogart. 1981. *Press and public: Who reads what, when, where, and why in American newspapers.* Lawrence Erlbaum Associates, Hillsdale, NJ.
[5] Tom De Smedt, W Daelemans, and Tom De Smedt. 2012. Pattern for Python. *The Journal of Machine Learning Research* 13 (2012), 2063–2067.
[6] Daniel Dor. 2003. On newspaper headlines as relevance optimizers. *Journal of Pragmatics* 35, 5 (2003), 695–721.
[7] Christiane Eilders. 2006. News factors and news decisions. Theoretical and methodological advances in Germany. *Communications* 31, 1 (2006), 5–24.
[8] Johan Galtung 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 (1965), 64–90.
[9] Víctor García-Perdomo, Ramón Salaverría, Danielle K Kilgo, and Summer Harlow. 2018. To share or not to share: The influence of news values and popular social media content in the United States, Brzil, and Argentinia. *Journalism Studies* 19, 8 (2018), 1180–1201. https://doi.org/10.1080/1461670X.2016.1265896
[10] Rosanna E Guadagno, Daniel M Rempala, Shannon Murphy, and Bradley M Okdie. 2013. What makes a video go viral? An analysis of emotional contagion and Internet memes. *Computers in Human Behavior* 29, 6 (2013), 2312–2319.
[11] Andrew Halterman. 2017. Mordecai: Full Text Geoparsing and Event Geocoding. *The Journal of Open Source Software* 2 (01 2017). https://doi.org/10.21105/joss.00091
[12] Tony Harcup and Deirdre O'neill. 2017. What is news? News values revisited (again). *Journalism studies* 18, 12 (2017), 1470–1488.
[13] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear* (2017).
[14] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.
[15] Yue Jin, Jinghua Huang, and Xinyao Wang. 2017. What Influences Content Popularity? An Empirical Investigation of Voting in Social Q&A Communities.. In *PACIS*. 161.
[16] Hans Mathias Kepplinger. 2008. News values. *The international encyclopedia of communication* (2008).
[17] Danielle K Kilgo, Summer Harlow, Víctor García-Perdomo, and Ramón Salaverría. 2018. A new sensation? An international exploration of sensationalism and social media recommendations in online news publications. *Journalism: Theory, Practice & Criticism* 19 (2018), 1497–1516. Issue 11. https://doi.org/10.1177/1464884916683549
[18] Danielle K. Kilgo, Kyser Lough, and Martin J. Riedl. 2017. Emotional appeals and news values as factors of shareworthiness in Ice Bucket Challenge coverage. *Digital Journalism* (2017). https://doi.org/10.1080/21670811.2017.1387501
[19] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. (1975).
[20] Dominic L Lasorsa, Seth C Lewis, and Avery E Holton. 2012. Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism studies* 13, 1 (2012), 19–36.
[21] Angela M Lee, Seth C Lewis, and Matthew Powers. 2014. Audience clicks and news placement: A study of time-lagged influence in online journalism. *Communication Research* 41, 4 (2014), 505–530.
[22] Annie Louis and Ani Nenkova. 2013. What makes writing great? First experiments on article quality prediction in the science journalism domain. *Transactions of the Association for Computational Linguistics* 1 (2013), 341–352.
[23] Scott R Maier, Paul Slovic, and Marcus Mayorga. 2017. Reader reaction to news of mass suffering: Assessing the influence of story form and emotional response. *Journalism* 18, 8 (2017), 1011–1029.
[24] Malvina Nissim and Viviana Patti. 2017. Semantic aspects in sentiment analysis. In *Sentiment analysis in social networks*. Elsevier, 31–48.
[25] Winfried Nöth. 2007. Self-reference in the media: The semiotic framework. *Self-reference in the Media* (2007), 3–30.
[26] Arika Okrent. 2014. The listicle as literary form. *University of Chicago Magazine* 106, 3 (2014), 52–53.
[27] Peter AM Oude Ophuis and Hans CM Van Trijp. 1995. Perceived quality: a market driven and consumer oriented approach. *Food quality and Preference* 6, 3 (1995), 177–183.
[28] Claudia Orellana-Rodriguez and Mark T Keane. 2018. Attention to news and its dissemination on Twitter: A survey. *Computer Science Review* 29 (2018), 74–94.
[29] Sharoda A Paul, Lichan Hong, and Ed H Chi. 2012. Who is authoritative? understanding reputation mechanisms in quora. *arXiv preprint arXiv:1204.3724* (2012).
[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[31] Joshua M Scacco and Ashley Muddiman. 2016. Investigating the influence of "clickbait" news headlines. *Engaging News Project Report* (2016).
[32] Martijn Schoonvelde, Anna Brosius, Gijs Schumacher, and Bert N Bakker. 2019. Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political speeches. *PloS one* 14, 2 (2019), e0208450.
[33] Stuart Soroka and Stephen McAdams. 2015. News, politics, and negativity. *Political Communication* 32, 1 (2015), 1–22.
[34] Catherine Sotirakou, Panagiotis Germanakos, Andreas Holzinger, and Constantinos Mourlas. 2018. Feedback Matters! Predicting the Appreciation of Online Articles A Data-Driven Approach. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 147–159.
[35] Edson C. Tandoc and Ryan J. Thomas. 2015. The Ethics of Web Analytics: Implications of using audience metrics in news construction. *Digital Journalism* 3, 2 (2015), 243–258. https://doi.org/10.1080/21670811.2014.909122
[36] Damian Trilling, Petro Tolochko, and Björn Burscher. 2017. From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly* 94, 1 (2017), 38–60.
[37] Sebastián Valenzuela, Martina Piña, and Josefina Ramírez. 2017. Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication* 67, 5 (2017), 803–826. https://doi.org/10.1111/jcom.12325
[38] Ben Wasike. 2018. Preaching to the choir? An analysis of newspaper readability vis-a-vis public literacy. *Journalism* 19, 11 (2018), 1570–1587.
[39] Patrick Weber. 2014. Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New media & society* 16, 6 (2014), 941–957.