Seeing Is Not Always Believing

An Exploratory Study of Clickbait in WeChat

Zhang, Wenping; Du, Wei; Bian, Yiyang; Peng, Chih-Hung; Jiang, Qiqi

Document Version
Accepted author manuscript

Published in:
Internet Research

DOI:
10.1108/INTR-09-2019-0373

Publication date:
2020

License
Unspecified

Citation for published version (APA):
Zhang, W., Du, W., Bian, Y., Peng, C.-H., & Jiang, Q. (2020). Seeing Is Not Always Believing: An Exploratory Study of Clickbait in WeChat. *Internet Research*, *30*(3), 1043-1058. https://doi.org/10.1108/INTR-09-2019-0373

Seeing Is Not Always Believing: An Exploratory Study of Clickbait in WeChat

Abstract

Purpose: The purpose of this study is to unpack the antecedents and consequences of clickbait prevalence in online media at two different levels, namely, 1) Headline-level: what characteristics of clickbait headlines attract user clicks and 2) Publisher-level: what happens to publishers who create clickbait on a prolonged basis.

Design/methodology/approach: To test the proposed conjectures, the authors collected longitudinal data in collaboration with a leading company that operates more than 500 WeChat official accounts in China. This study proposed a text mining framework to extract and quantify clickbait' rhetorical features (i.e., hyperbole, insinuation, puzzle, and visual rhetoric). Econometric analysis was employed for empirical validation.

Findings: The findings revealed that 1) hyperbole, insinuation, and visual rhetoric entice users to click the baited headlines, 2) there is an inverted U-shaped relationship between the number of clickbait headlines posted by a publisher and its visit traffic, and 3) this non-linear relationship is moderated by the publisher's age.

Research limitations/implications: This research contributes to current literature on clickbait detection and clickbait consequences. Future studies can design more sophisticated methods for extracting rhetorical characteristics and implement in different languages.

Practical implications: The findings could aid online media publishers to design attractive headlines and develop clickbait strategies to avoid user churn, and help managers enact appropriate regulations and policies to control clickbait prevalence.

Originality/value: The authors propose a novel text mining framework to quantify rhetoric embedded in clickbait. This study empirically investigates antecedents and consequences of clickbait prevalence through an exploratory study of WeChat in China.

Keywords: Clickbait, Rhetoric, Visit traffic, Online publisher, WeChat

Article classification: Research paper

Introduction

Clickbait, defined as “(On the Internet) content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page” in the Oxford English Dictionary, prevails in the Internet, especially in online media, recently. This prevalence is partly due to the massive information available online; hence, to attract users’ attention, publishers often create exaggerated and eye-catching headlines to entice users to click to find out more. Examples of clickbait headlines are “You’ll never believe…”, “What every mother needs to know…”, and “The ten things you’ve been done wrong in exercises”. Publishers who resort to clickbait will receive more page views or clicks, which is vital given the importance of visit traffic on the Internet. However, the end result is that users are underwhelmed by the content that they have been tricked into viewing and may lose trust in that publisher.

WeChat, as a leading social media platform in China, is also prone to the clickbait problem. As the main marketing channel in WeChat, WeChat Official Accounts (WOAs) post articles to attract users to subscribe. To stand out, WOA publishers often craft catchy titles and attention-grabbing descriptions to increase visitor traffic through which publishers could gain revenue from either advertisers or user payments. This trick indeed works in the short term. However, constantly crafting clickbait creates mistrust among subscribers and in turn results in a dilemma for keeping existing subscribers and attracting newcomers for WOA. For articles, if clicks can be easily obtained by clickbait headlines, editors may put little effort in improving quality of articles to attract reads. The clickbait prevalence in WeChat may lead to the phenomenon of Gresham’s Law (Rowland and Marz, 1982), i.e., bad money drives out good. Platform reputation will be finally ruined with huge number of low-quality articles with clickbait headlines.

Due to its deceptive nature, clickbait is widely criticized by both public institutions and companies (Mihaylov *et al.*, 2018). For instance, 28 European Union (EU) member countries found that online media tended to publish articles with “catchy, provocative, and sensationalist front-page” headlines in lieu of delivering quality content, and the EU declared the need to alter such tendency (Orosa, 2017). Moreover, Facebook announced to undertake initiatives to significantly reduce clickbait by intensifying the punishment (Arun *et al*., 2017). Despite the implementation of various regulations for reducing clickbait, it remains prevalent (Chakraborty *et al.*, 2016; Rony *et al.*, 2017). The ineffectiveness in controlling clickbait might result from the insufficient understanding of its benefits for publishers.

The information-gap theory of curiosity explained why users intend to click the baited headlines (Loewenstein, 1994). Elaborately, clickbait stimulates curiosity when a user perceives a gap between his/her knowledge and attention. Such gap generates a “feeling of deprivation labeled curiosity” and motivates the user to acquire the missing information to alleviate the “feeling of deprivation”. Prior studies argued that such baited stimuli could be developed in terms of using rhetorical strategies. For instance, Flower and Hayes (1980) found that rhetoric could construct mystical stimuli to attract the reader’s attention; a similar conclusion was reached by McQuarrie and Mick (1996), who found rhetorical strategies to be effective in attracting consumer attention. In marketing studies, rhetoric has been widely accepted as an effective strategy for designing appealing advertising messages (Phillips and McQuarrie, 2002). It can be inferred that users are prone to clicking rhetorically featured clickbait. As rhetoric is widely used in clickbait headlines, this study innovatively resorts to rhetoric identification to detect clickbait in the present research.

Differing from users whose attraction to clickbait is mainly driven by the feeling of deprivation labeled curiosity, publishers strategically create clickbait for monetization (Cook, 2016; Rochlin, 2017). In particular, a publisher monetizes its website in terms of increasing site traffic by using clickbait (Rochlin, 2017). For example, Farhi (2007) found that online publishers significantly relied on web advertising revenues from site traffic to secure their survival. Therefore, opportunistically employing clickbait to attract a large volume of visits is an effective and efficient approach for increasing publishers' revenues. Besides, with the increased number of visitors, the publisher can more easily acquire potential subscribers. Schlosser *et al.* (2006) stated that converting visitors to users provided great benefits and thus, should be regarded as the highest priority task in digital strategy management. Wang *et al.* (2005) found that subscription fees or subscriber payoffs significantly contributed to the income of online publishers. In this regard, clickbait, as an effective approach for user acquisition, is widely popularized by online publishers (Lombardi, 2017).

Despite widespread prevalence of clickbait in Internet and online media, not many studies have investigated such phenomenon to date. Several extant studies have confounded the concept of clickbait and fake news though they are completely different in essence. In particular, clickbait is targeted for attracting users to click and read whereas fake news is created to disseminate fictitious information for malicious purposes (Bakir and McStay, 2018; Lazer *et al.*, 2018; Rubin *et al.*, 2019). Notably, differing from fake news that contain little factual basis, the content under clickbait is simply valueless. This study's literature review suggested that current researches on clickbait could be classified into two streams, i.e., algorithms or application designs for clickbait detection and consequences of clickbait. The former investigated the design and evaluation of different clickbait detection methods to target clickbait prior to its prevalence. The latter stream of research mainly focused on the impact of the prevalence of clickbait. Such post-hoc evaluation did not provide tangible initiatives or suggestions for managing the deluge of clickbait in online media. The antecedents and consequences of clickbait prevalence remain unclear.

The motivation of this study is to investigate what headline features attracts reads and the consequences of clickbait prevalence to publishers. More specifically, this study attempts to answer the questions at two different levels through an empirical study of WeChat in China, namely, 1) Headline-level: what characteristics of clickbait headlines attract user clicks and 2) Publisher-level: what happens to publishers if they create clickbait on a prolonged basis. Answering these questions is not only vital to understand the antecedents and consequences of clickbait prevalence but also conducive to implementing regulations or policies to reduce future clickbait creation.

The reminder of this study is organized as follows. The next section proposes this study's research conjectures based on existing research and reviews the two research streams, i.e., clickbait detection methods and consequences of clickbait prevalence. Next, the third section describes the research setting and methodology, i.e., text mining and econometric analysis, to test the proposed conjectures. Following this, two empirical studies are conducted to test the conjectures at the user level and publisher level, respectively. Finally, the findings are discussed and conclusions presented in the last section.

Research Conjectures

Although clickbait has prevailed in online media for decades, few studies discuss clickbait in the information systems discipline. It has been noted that clickbait is different from fake news (Chen *et al.*, 2015). Thus, earlier studies which investigated fake news were not included in this section. This study classifies clickbait research into two main streams: the detection of clickbait and the consequences of clickbait prevalence. The former was about the design and evaluation of clickbait detection methods while the latter investigated the various consequences, particularly the negative impacts, resulting from the deluge of clickbait. More details are presented below.

Clickbait Detection

Clickbait detection method aims to identify clickbait so as to control clickbait prevalence (Daoud and El-Seoud, 2019; Gairola *et al.*, 2017; Shu *et al.*, 2018; Sisodia, 2019). The earliest research on clickbait detection was in 2016, namely the study by Potthast *et al.* (2016), which employed a dataset from Twitter to construct the first clickbait corpus containing 2,992 tweets from which 767 pieces of clickbait were extracted. In particular, the authors designed a detection approach that built upon 215 semantic features and incorporated the application of a series of simple random forest classifiers to achieve 0.76 precision and 0.76 recall, respectively. Precision (positive predictive value) is the fraction of relevant instances among the retrieved instances. Recall (sensitivity) is the fraction of relevant instances that have been retrieved over the total volume of relevant instances. Inspired by this attempt, Cao and Le (2017) expanded the approach to 331 semantic features and applied a more sophisticated algorithm, namely the random forest regression, to target potential clickbait on social media. This approach outperformed its predecessor and achieved 0.82 accuracy and 0.61 F1-score, respectively. Accuracy is the faction of the correctly retrieved instances. F1-score is the overall performance score and refers to the balance (trade-off) between precision and recall. A higher F1-score indicates a better overall performance. Numerous studies employing similar designs were proposed with improved performance. Table 1 summarizes the representative studies of clickbait detection approaches in chronological order.

Table 1. A Brief Review of Clickbait Detection Approaches (Algorithmic Design)

Building upon the algorithmic design of clickbait detection, certain studies developed applications of such detection (Rubin *et al.*, 2019). For instance, Chakraborty *et al.* (2016) designed a browser extension that provided a warning on potential clickbait by comparing the feature distributions between trained clickbait and non-clickbait datasets; Rony *et al.* (2017) designed and developed an automatic bot and integrated it into a web browser to help users avoid clicking baited headlines on social media.

State-of-art clickbait detection methods either employ classifiers with huge amounts of hand-crafted lexical and syntactic headline features or leverage end-to-end deep learning methods that are capable of automatically inducing implicit features. Both types of methods focused on improving detection performance with "black box" algorithms but failed in providing maneuverable conclusions for platform managers with hundreds of headline features or implicit features (e.g., what features are useful in detecting clickbait and what clickbait features attract users).

To the best of the authors' knowledge, few clickbait detection studies consider features from readers' perspective, i.e., what clickbait features attract readers. Zheng *et al.* (2017) argued that behavioral traits were helpful to design clickbait detection and included these as semantically textual features in their application design. To validate this hypothesis, the authors employed gradient boosting decision tree by considering behavioral traits, and outperformed conventional methods. López-Sánchez *et al.* (2017) argued that clickbait perception was contingent upon users and applied deep learning and metric learning to improve the adaptability of the clickbait detection model. In addition, literature on communication and media research unveiled that rhetorical language is more attractive to users (Benoit and Smythe, 2003; Scaraboto *et al.*, 2012). Practically, rhetorics could be widely found to frame the headlines to gain readers' attention in both online and traditional media. Therefore, this study aims to leverage rhetorical features for clickbait detection and unveil the role of rhetorical features of headlines in attracting users, and proposes the first conjecture:

Research Conjecture 1: *The extent of rhetorics embedded in a clickbait is positively associated with the number of baited users.*

Consequences of Clickbait Prevalence

Besides clickbait detection approaches, another stream of studies discussed the consequences or impact of this prevalence. Intuitively, the primary objective of creating clickbait was to increase click likelihood. Publishers leveraged clickbait to attract users' attentions to entice clicking through intriguing headlines (Anand *et al.*, 2017; Potthast *et al.*, 2016). Moreover, publishers employed the social sharing functionality to promote clickbait on social media (Rubin, 2017). Given that the headlines were prominently displayed in the shared newsfeeds on social media, users' friends would be enticed to click on such clickbait and be redirected to the publisher's site. Considering that visit traffic served as a primary antecedent to publishers' income, clickbait that helped increase visit traffic would not be sorely resisted by the publishers.

However, users receiving massive clickbait may develop negative perceptions of the publisher(s) concerned. Accumulatively, users were very likely to abandon a publisher that constantly created clickbait. For instance, Beleslin *et al.* (2017) asserted that using clickbait to attract users was a risky strategy because negative attitudes towards clickbait could be progressively cultivated over time. Scacco and Muddiman (2016) comparatively studied factual headlines and clickbait. They reached similar conclusions that users presented negative attitudes and reactions when they were perennially exposed to clickbait. Previous studies argued that user behaviors could be viewed as a manifestation of their attitudes (Ajzen, 1985). As a result, users refused to receive posts from the publishers resulting in user churn. This study found that although using clickbait initially helped publishers accumulate visit traffic, such traffic would drop if clickbait continued to be created over time. Thus, an inverted U-shaped relationship was postulated between the number of clickbait headlines (created by a publisher) and visit traffic (to the focal publisher).

Moreover, given that clickbait largely spread in social media through Word-of-Mouth, e.g., reposting or clicking the "like" button (Fulgoni and Lipsman, 2017), older publishers with often more subscribers could exert higher influence on increasing visit traffic by creating clickbait. More specifically, more people from their larger user base could be influenced and enticed to click on the baited headlines. Nonetheless, by continuing to create clickbait, these older publishers would simultaneously risk higher chances of losing existing subscribers. Existing users may get annoyed with the clickbait and choose to discontinue their subscription to the concerned publisher. On the basis of the foregoing, the second conjecture on clickbait from publishers' perspective is developed:

Research Conjecture 2: *The number of clickbait headlines posted by an older publisher has a stronger inverted U-shaped relationship with its visit traffic than for younger publishers.*

Data and Methodology

This research focuses on WOAs that created clickbait headlines to entice users to click and read their articles. WOAs appear either in the "Subscribed Articles" section or in the "Chat" section of WeChat, and are brought to the top of the subscriber's message interface upon sending notifications. Subscribers can use the conversation interface of an account to read past articles. Users can also send articles that they find interesting to friends or share articles in the “Moments” section. Thus, articles may be potentially disseminated to millions of users through such “friend” ties. The number of user visits, through which companies can monetize WOAs by charging users directly or earning advertisement income, is vital to WOAs. Given that users get bored with overwhelming and useless subscription articles, publishers craft catchy titles and attention-grabbing descriptions, i.e., clickbait, in order to stand out and draw user traffic.

The data for this study was collected from a leading digital media company (hereinafter referred to as *New-Media* to preserve the company’s identity) in China. New-Media operates more than 500 WOAs in China with more than 10 million subscribers. Each account has its own designated editor and publishes articles in its own niche such as “Fitness”, “Lifestyle”, and “Photography.” All articles from these WOAs could be shared within the social media for free reading. The bottom of each article presents a series of sponsored advertisements, which contribute to the key revenues of New-Media.

Driven by the information-gap theory of curiosity, this study first designed and developed four rhetorical features, namely hyperbole, insinuation, puzzle, and visual rhetoric, to detect clickbait headlines. The details of these measurements are provided in Study 1. Corresponding values of the four rhetorical features of article headlines were computed respectively, and the headlines with positive values in any rhetorical dimension, regardless of the numerical values, were labeled as clickbait. In addition, this study recruited three student assistants to manually check the validity of the results. In particular, 1,000 headlines and related articles were randomly selected. The headlines and related content were sequentially presented to the assistants who were requested to vote on whether the headline was a clickbait or not. The results (97.8%) confirmed the validity and accuracy of the measurement. In particular, the headlines that were not labeled as clickbait were assigned null values in any rhetorical dimension.

This study conducted two empirical studies to separately validate the proposed two conjectures. Study 1 collected data of published articles and aimed to investigate the relationship between the extent of rhetorics in clickbait and the number of baited users. Study 2 collected data of publishers and aimed to validate: 1) the quadratic relationship between the number of clickbait and the publisher's visit traffic and 2) the moderating role of the publisher's age.

The Study 1 sample contains 7,481 articles with clickbait headlines published from July 1st, 2017 to December 31st, 2017. The life cycle of an article was estimated to be 5-7 days from its published date to its "death" (when it is forgotten by users). Therefore, after consulting some editors of publisher accounts and the managers from New-Media, this study set a one-week lag for collecting article data comprising the numbers of visits, "likes" and "shares" of each published article. In addition, article headlines were crawled to analyze the extent of rhetorics embedded in clickbait.

In Study 2, the data on publisher accounts managed by New-Media was extracted for a publisher-level investigation into the change in the site's visitor traffic resulting from peer influence or existing user churn. To precisely observe the fluctuation in visit traffic, the fundamental user base volume is imperative. In other words, clickbait posted by publisher accounts with only a few subscribers may not exert significant influence on the publisher's visit traffic. After consulting with practitioners from New-Media, this study focused on the publisher accounts having at least 50,000 subscribers, which resulted in a sample comprising 202 publisher accounts. These accounts were created for targeting 25 distinctive niches. Given that these accounts did not post articles every day, this study created a panel as a month-level observation. In this regard, a 6-month observation period might be insufficient. In the end, this study comprised 2,424 observations.

Results

Study 1: Headline-level Investigation

This study employed the number of unique visits of each article seven days after its initial publication as the dependent variable (i.e., the number of baited users). To remove the clicks created by bots, this study set a threshold, τ, referring to the interval between the timestamp of hitting the headline and the timestamp of landing on the article page to filter out fake clicks. In particular, this study removed clicks whose τ values were fewer than 5 seconds. The dependent variable was denoted as RDS_i.

For the independent variable, i.e., the extent of rhetorics, this study designed and developed a series of text mining methods to extract different rhetorical characteristics of articles' clickbait headlines. By referring to prior studies of clickbait and linguistic research (e.g., Anand *et al.*, 2017; Biyani *et al.*, 2016; Deighton, 1985; Hart and Daughton, 2015; Vatz, 1973; Zhang *et al.*, 2018), this study semantically and syntactically featured four prominent rhetorical characteristics, namely hyperbole, insinuation, puzzle, and visual rhetoric, and computationally measured the four rhetorics with the proposed framework. The four rhetorical features were verified as the most representative ones widely used in clickbait after consulting practitioners from New-Media.

Hyperbole refers to the use of exaggeration as rhetoric. Overstatement or over-expression is widely found in hyperbole rhetoric. Hyperbolic terms refer to the terms frequently used in overstatements. Most widely used hyperbolic terms include exaggeration words (e.g., shock, and astonish) and extremely large or small quantifiers (e.g., million, billion, and second). “*Make millions of dollars within 10 days!*” and “These statistics may shock you” are representative hyperbole headlines including hyperbolic terms. This study manually constructed a hyperbole lexicon from clickbait headlines in the sample to detect the hyperbolic terms in the headlines. The hyperbole strength of a headline could be measured by:

$$\text{Hyperbole\_Strength} = \frac{No.of\ hyperbolic\ terms\ in\ the\ headline}{Total\ no.of\ terms\ in\ the\ headline} \quad (1)$$

Insinuation refers to an indirect (and usually malicious) implication. To do so, editors usually adopt tempting or metaphorical words or expressions. Insinuation terms usually have connotative meanings besides the obvious/direct semantic meaning. These terms will lead to rich imaginations. Common examples include wordplay (e.g., polyseme and homonym), punch line (e.g., “Wash hair”, which refers to a well-known scandal of a popular star), etc. Example headlines with insinuation rhetorics are “*It had to be chew!*” and “*TIGER PUTS BALLS IN WRONG PLACE AGAIN*”. Similar to hyperbole, it can be measured by:

$$\text{Insinuation\_Strength} = \frac{No.of\ insinuation\ terms\ in\ the\ headline}{Total\ no.of\ terms\ in\ the\ headline} \quad (2)$$

Puzzle refers to inquiry, questioning, or interrogation. From a linguistic perspective, puzzles could be created in two ways: keeping pronoun(s) in headlines and creating question-based headlines (e.g., echo questions or interrogative sentences). For the former, pronoun strength could be measured in similar ways as the previous two measurements. An example of this rhetorical clickbait can be “*This kid opens a present. You won’t believe what happens when they see what’s inside!*”. For the latter, a typical example is “*Can you solve this ancient riddle? Most people failed!*”. To measure the strength of the latter form of puzzle, this study firstly defined a set of figurative forms (e.g., interrogation, rhetorical questions, and elliptical sentences), denoted by S. Next, this study defined a set of figurative forms of each headline, denoted by E. The following formula was used to depict the extent of a puzzle in the question-based headlines:

$$Figurativeness\_Strength = \begin{cases} 1 & if\ S \cap E = \emptyset \\ 0 & otherwise \end{cases} \quad (3)$$

The overall measurement of the extent of puzzle in each headline could be written as the weighted sum of the figurativeness strength and pronoun strength:

$$Puzzle\_Strength = \alpha \cdot Figurativeness\_Strength + (1 - \alpha) \cdot \frac{No. of\ pronomial\ terms\ in\ headline}{Total\ no. of\ terms\ in\ the\ headline} \quad (4)$$

where α is a smoothing factor, e.g., α is set to 0.5 if both types of puzzles contribute equally in constructing the puzzle.

<u>Visual Rhetoric</u> refers to communicative visual images. Three types of visual rhetoric are widely used in clickbait: symbols, digits, and pictures. This study manually constructed a symbol set Sym. The set Sym' was used to denote the symbols in each headline. Previous studies indicated that headlines with more symbols could attract more attention (Cao and Ye 2009; Cui *et al.*, 2012); therefore, the symbol strength can be measured as $symbol\_strength = |Sym'|/|Sym|$. The extent of digits could be measured in a similar manner as the measurement of hyperbole or insinuation. Given that there was at most one picture inserted as a thumbnail under each headline, this study employed a binary value, where 1 denoted the inclusion of a picture and 0 denoted the lack of a picture. Eventually, the visual rhetoric of each headline can be measured by:

$$Visual\_Strength = \beta \cdot symbol\_strength + \gamma \cdot digits\_strength + \delta \cdot picture\_strength \quad (5)$$

where β, γ, and δ are smoothing factors and satisfy $\beta + \gamma + \delta = 1$ (e.g., $\beta = \gamma = \delta = 1/3$).

Collectively, this study used HYP_i, INS_i, PUZ_i, and VIS_i to represent the extent of rhetorics of Article $\underline{i}$ in terms of the aforementioned four types of rhetorical characteristics, respectively. Moreover, this study included a rich set of control variables including headline length (HDL_i), number of likes (LKN_i), number of shares (SRN_i), and whether this article was a lead article (ILP_i) in the issue. The headline length (HDL_i) was computed by the number of words. Next, publishers selected certain articles as lead articles, setting them at the top of their social media page occasionally; this study used a binary value ILP_i to indicate whether an article was a lead article (if yes, $ILP_i = 1$) or not. It is to be noted that the topic of each article was not included. Although each publisher had its own niche, the topics overlapped among their published articles. For example, the topic of the article, "Diet Recipe" from the publisher "Fitness Girl" might have overlapped with the articles from another publisher named "Healthy Life" In this regard, controlling the publishers' demographics might result in contradictions. The variable definitions and descriptive statistics are presented in Table 2 below.

Table 2. Summary of Variables Used in Study 1 (7,481 observations)

Correlations among the studied variables are reported in Table 3; a majority of the bivariate correlations were below the recommended 0.70 threshold level. Only two pairs of variables had slightly higher correlation values. To rule out collinearity concerns, this study calculated the variance inflation factors (VIF) of each variable. The maximum estimated value of VIF is 6.53, which is lower than the recommended threshold of 10.0 (Cohen *et al.*, 2003).

Table 3. Correlation of Variables Used in Study 1

Based on these variables, the model could be described as follows:

$$f(RDS_i) = \beta_0 + \beta_1 HYP_i + \beta_2 INS_i + \beta_3 PUZ_i + \beta_4 VIS_i + \beta_5 HDL_i + \beta_6 LKN_i + \beta_7 SRN_i + \beta_8 ILP_i + \varepsilon_i \quad (6)$$

Given that the dependent variable (RDS_i) was a typical counting variable, this study employed both Poisson regression and Negative binomial regression to estimate the coefficients. The results are reported in Table 4 below.

Table 4. Model Estimation Results of Study 1

The results show that the negative binomial regression model has lower AIC and BIC values. It indicates that negative binomial regression model fits the data better than the Poisson regression model. One of the plausible reasons is that Poisson regression makes a strong assumption that the variance is equal to the mean (Schilling and Phelps, 2007). The dependent variable was over-dispersed, which resulted in its variance exceeding its mean value. This dispersed data followed a gamma distribution; therefore, the negative binomial distribution could achieve better estimation results (Hilbe, 2011).

The estimated results revealed that the extent of rhetorics, especially the hyperbole, insinuation, and visual rhetoric, enacted positive influences on the number of clicks of each article, whereas creating puzzles in the headlines had no significant contribution to increasing the clicks of articles.

To this end, the study’s first research conjecture has been well tested. The exploratory results reveal that clickbait characterized by hyperbolic, insinuating, and visual rhetoric contribute to the increase in the number of received clicks. Thus, setting rules or regulations to restrain the excessive use of such rhetoric in headlines could help reduce clickbait prevalence. On the other hand, editors who create headlines based on facts can choose to use hyperbole, insinuation, and visual rhetoric appropriately to increase their readership.

Study 2: Publisher-level Investigation

In this study, the dependent variable denoted by AUN_{jt} was the aggregated number of monthly unique visitors to articles published by each publisher account in the last month. This measurement of visit traffic differed from the number of visits or number of unique visits widely employed in web analytics (Vellingiri *et al.*, 2015). The present study employs the number of unique visitors (with relatively smaller values) instead of the number of unique visits to precisely quantify the user retention ability of publisher accounts.

The key predictors in Study 2 include (1) $CFQ_{j,t-1}$: the number of clickbait published by each publisher account in the last month and (2) $AGE_{j,t-1}$: the age (in terms of number of months) of each publisher account since its establishment. Headlines were identified as clickbait by applying the measurement approach for extracting rhetorical features.

This study also included a set of control variables: the average number of visits to each article of each publisher account ($ARS_{j,t}$), number of articles published by each publisher account ($ATN_{j,t}$), number of newly added subscribers of each publisher account ($NRU_{j,t}$), total number of subscribers of each publisher account ($PTU_{j,t}$), and categorical niche of each publisher account ($PDM_{j,t}$). The details of variable definitions and descriptive statistics are presented in Table 5 below.

Table 5. Summary of Variables Used in Study 2 (2,424 observations)

The correlations among these variables are presented in Table 6 below. A majority of the bivariate correlations were below 0.70. Moreover, none of the variation inflation factors were higher than 10.0.

Table 6. Correlation of Variables Used in Study 2

To test the second research conjecture, i.e., the quadratic relationship and the moderation effect, the following econometric model is built for coefficient estimation:

$$f(AUN_{jt}) = \gamma_0 + \gamma_1 CFQ_{j,t-1} + \gamma_2 CFQ_{j,t-1}^2 + \gamma_3 AGE_{j,t-1} + \gamma_4 AGE_{j,t-1} * CFQ_{j,t-1} + \gamma_5 AGE_{j,t-1} * CFQ_{j,t-1}^2 + \gamma_6 ARS_{j,t} + \gamma_7 ATN_{j,t} + \gamma_8 NRU_{j,t} + \gamma_9 PTU_{j,t} + \gamma_{10} PDM_{j,t} + \xi_{i,t-1} \quad (7)$$

Similar to Study 1, the Poisson regression and negative binomial regression are employed, respectively. In addition, the Hausman test is applied to determine whether the fixed-effect model outperformed the random-effect model (Borenstein *et al.*, 2010). The null hypothesis was rejected. Thus, the fixed effect model is chosen to fit the data. The estimated results are presented in Table 7 below.

Table 7. Model Estimation Results of Study 2 (Fixed-effect model)

The estimated coefficients in Table 7 provide evidence to support the second research conjecture. Both models revealed the curvilinear relationship between the number of clickbait created by a publisher and its visit traffic. The coefficient of the linear term $CFQ_{j,t-1}$is positive, whereas that of $CFQ_{j,t-1}^2$ is negative. Although clickbait created by a publisher help attract visitors, the user would get bored if the publisher constantly created clickbait. The positive and significant relationship between $AGE_{j,t-1}$ and AUN_{jt} shows that the publisher's age could help increase visit traffic. Older publishers tend to have a larger user base with time and thus, more visit traffic. The negative and significant coefficients for $AGE_{j,t-1} * CFQ_{j,t-1}$ and $AGE_{j,t-1} * CFQ_{j,t-1}^2$ provide evidence of the negative impact of the age of the publisher on the non-linear relationship. The excessive clickbait created by older publishers could result in a negative impression of a larger existing user base. Hence, for older publishers, the greater the number of clickbait created, the greater is the drop of visit traffic.

Discussion and Conclusion

While the move to online media consumption was supposed to enable people to access information more effectively, instead, people are somewhat overwhelmed by the deluge of information made available. The prevalence of non-factual content in online media made it more difficult for people to retrieve their desired information. The deluge of fake news or malicious rumors on the internet has been effectively reduced through various initiatives, including legislation or relatively mature technology-based solutions. Residing in the grey area between fake news and factual reports, clickbait has not been thoroughly studied so far. While relatively lean, the research topics of the literature could be classified into two streams: design and evaluation of clickbait detection methods (e.g., Biyani *et al.*, 2016; Rony *et al.*, 2017) and the various consequences of clickbait prevalence (e.g., Beleslin *et al.*, 2017; Scacco and Muddiman, 2016). However, these studies did not adequately address why clickbait prevail on the Internet. Disclosing the antecedents driving clickbait prevalence ought to facilitate the implementation of appropriate policies and regulations to effectively reduce its quantity and popularity. The findings from this study contributed to bridge this gap in the literature.

As an exploratory study, two intriguing research conjectures have been proposed to account for antecedents and consequence of clickbait prevalence. In particular, this study asserted that 1) clickbait is characterized by rhetoric, which entices users to click; and 2) there is a quadratic (inverted U-shaped) relationship between the number of clickbait headlines posted by a publisher and the publisher’s visit traffic, and such relationship is moderated by the publisher’s platform age. The first conjecture involved unpacking the prevalence of clickbait from user perspective and the latter was proposed to outline the consequential impact of such prevalence on the publishers. To test these two conjectures, this study collected longitudinal data in collaboration with a leading digital media company in China, and applied sophisticated analytical frameworks, i.e., a series of self-designed and developed text mining methods and econometric analysis, for empirical validation. This study obtained three important findings.

First, besides verifying that rhetorical characteristics indeed enticed users to click baited headlines, this study obtained an in-depth understanding by revealing the significant role of hyperbole, insinuation, and visual rhetoric in enticing users to click these headlines. Interestingly, the puzzle, which was widely used for attracting attention in online advertisements (Alwitt, 2002; Bizzozero *et al.*, 2016; Fazio *et al.*, 1992; Jiang *et al.*, 2012), was not significantly associated with the user decision on clicking. From the regulator's perspective, this finding is instrumental in proactively examining potentially baited headlines by targeting these rhetorical features. This finding could also benefit editors who create headlines that reflect facts. Such editors could leverage these three identified rhetorical features to promote articles by attracting more attention and readership.

Second, the findings revealed an interesting phenomenon. Although using clickbait could be an incentive to temporarily increase visit traffic, continuing with such a strategy could lead to user churn over time. Understanding this consequence could discourage publishers from increasingly creating clickbait over time because of its impending negative outcome. This could effectively restrain the growth of clickbait in the online media. Indeed, this could also prevent publishers from falling into a dilemmatic bottleneck in user growth.

Last, but not least, this study found that older publishers were prone to outperform in terms of visit traffic growth. However, it was also important to point out that user churn might also impend more easily and promptly. Remarkably, compared with newly established publishers, the older publishers had a higher influence in spreading clickbait. Moreover, according to the paradigm of organizational ecology (Hannan and Freeman, 1993), older organizations tended to employ less aggressive but moderate strategies. Thus, by knowing the risk of the double-edged sword resulting from creating clickbait, the influential and older publishers were unwilling to create clickbait impulsively, which also indirectly restricted the popularity of clickbait in the Internet.

This research also contributes to current literature on clickbait detection and clickbait consequences. On one hand, our findings advance knowledge on the effective use of rhetoric identification to detect clickbait online. State-of-art clickbait detection methods employed either huge amounts of lexical and syntactic headline features or implicit embedding features, which cause difficulty to apply the complex features to real-world clickbait detection (Daoud and El-Seoud, 2019; Gairola *et al.*, 2017; Shu *et al.*, 2018; Sisodia, 2019). In addition, the complex feature design and black box machine learning techniques fail to provide practical suggestions for platform managers in controlling clickbait prevalence. Our work unveils the rhetoric used in clickbait headlines and proposed a clickbait detection framework with four prominent rhetorical dimensions, namely hyperbole, insinuation, puzzle, and visual rhetoric, and achieves satisfactory performance. The simple and practical feature extraction is helpful for managers to analyze which rhetorical dimension attracts clicks most.

On the other hand, our findings provide new insights regarding the impact of clickbait on publishers. Very few attempts relating to clickbait outcomes indicate the negative impact of clickbait on publishers, such as threatening publishers' credibility, reducing the information quality, or turning off readers (Molyneux and Coddington, 2019; Roelofs and Gallien, 2017; Zannettou *et al.*, 2018). We unveil the interplay among the clickbait, the visit traffic of publishers, and the characteristics of the publishers. Elaborately, the identified inverted U-shape relationship among clickbait and visit traffic enriches the intuitive understanding of the negative impact of clickbait. The results further reveal that the experience of publishers determines the power of the effect. Thus, this research provides the grained evidence of the clickbait impact and enriches the clickbait consequence literature. As an exploratory study, this work presents various caveats, which serve as suggestions for future research. First, this study only extracted four types of rhetorical features in the headlines. Although practitioners have concurred with the generalizability of these four as the most representative features, more sophisticated methods for extracting more rhetorical characteristics are suggested. Second, to further understand user psychological rationales, this study encourages future works to conduct a psychometric analysis using different methods, e.g., surveys, lab and field experiments, and even NeuroIS. Third, the causality issues are not perfectly addressed in this study though the lagged variables have been included in the model. Further experimentation can be considered to rule out endogeneity. Last but not least, the text mining framework was designed and implemented to process the Chinese language. Future studies can implement the approach using different languages, which can further improve the precision and effectiveness of the findings.

Acknowledgements

This paper has been selected as the 27th European Conference on Information Systems (ECIS 2019) fast-track paper (Zhang, W., Jiang, Q. and Peng, C. H., “UNFOLDING THE CLICKBAIT: A SIREN'S CALL IN THE ATTENTION ECONOMY.”, in The 27th European Conference on Information Systems) for possible publication in *Internet Research* journal.

References

Ajzen, I. (1985), "From intentions to actions: a theory of planned behavior", in Kuhl J., Beckmann J. (Eds.), *Action control*, Springer, Berlin, Heidelberg, pp. 11-39.

Alwitt, L.F. (2002), "Suspense and Advertising Responses", Journal of Consumer Psychology, Vol. 12 No. 1, pp. 35-49.

Anand, A., Chakraborty, T. and Park, N. (2017), "We used neural networks to detect clickbaits: You won't believe what happened next!", in *European Conference on Information Retrieval*, Springer, Cham, pp. 541-547.

Badu, A., Liu, A. and Zhang, J. (2017), "New Updates to Reduce Clickbait Headlines", available at: https://about.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/ (accessed 17 May 2017).

Bakir, V. and McStay, A. (2018), "Fake news and the economy of emotions: Problems, causes, solutions", *Digital Journalism*, Vol. 6 No. 2, pp. 154-175.

Beleslin, I., Njegovan, B.R. and Vukadinović, M.S. (2017), "Clickbait titles: risky formula for attracting readers and advertisers", in *XVII International Scientific Conference on Industrial Systems*, Novi Sad, Serbia, pp. 364-369.

Benoit, W.L. and Smythe, M.J. (2003), "Rhetorical theory as message reception: a cognitive response approach to rhetorical theory and criticism", *Communication Studies*, Vol. 54 No. 1, pp. 96-114.

Biyani, P., Tsioutsiouliklis, K. and Blackmer, J. (2016), "8 amazing secrets for getting more clicks: Detecting clickbaits in news streams using article informality", in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI, Phoenix, Arizona, USA, pp. 94-100.

Bizzozero, P., Flepp, R. and Franck, E. (2016), "The importance of suspense and surprise in entertainment demand: evidence from Wimbledon", *Journal of Economic Behavior & Organization*, Vol. 130, pp. 47-63.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2010), "A basic introduction to fixed-effect and random-effects models for meta-analysis", *Research Synthesis Methods*, Vol. 1 No. 2, pp. 97-111.

Cao, X. and Le, T. (2017), "Machine learning based detection of clickbait posts in social media", available at: https://arxiv.org/pdf/1710.01977.pdf (accessed 1 April 2018).

Cao, Z. and Ye, J. (2009), "Attention savings and emoticons usage in BBS", in Fourth International Conference on Computer Sciences and Convergence Information Technology in Seoul, South Korea, pp. 416-419.

Chakraborty, A., Paranjape, B., Kakarla, S. and Ganguly, N. (2016), "Stop clickbait: detecting and preventing clickbaits in online news media", in *2016 IEEE/ACM International Conference Advances in Social Networks Analysis and Mining, IEEE, San Francisco, USA*, pp. 9-16.

Chen, Y., Conroy, N.J. and Rubin, V.L. (2015), "Misleading online content: recognizing clickbait as false news", in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, ACM, Seattle, USA*, pp. 15-19.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2003), "Applied Multiple regression/Correlation Analysis for the Behavioral Sciences", Routledge, London.

Cook, C. (2016), "Money under fire: the ethics of revenue generation for oppositional news outlets", *The International Journal of Communication Ethics*, Vol. 13. No. 2/3, pp. 66-80.

Cui, A., Zhang, M., Liu, Y., Ma, S. and Zhang, K. (2012), "Discover breaking events with popular hashtags in Twitter", in *Proceedings of the 21st ACM international conference on Information and knowledge management in Maui, USA*, ACM, pp. 1794-1798.

Daoud, D.M. and El-Seoud, M.S.A (2019), "An effective approach for clickbait detection based on supervised machine learning technique", *International Journal of Online Engineering*, Vol. 15 No. 3, pp. 21-32.

Deighton, J. (1985), "Rhetorical strategies in advertising", *North American Advances in Consumer Research*, Vol. 12, pp. 432-436.

Farhi, P. (2007), "Salvation? The embattled newspaper business is betting heavily on web advertising revenue to secure its survival. But that wager is hardly a sure thing", *American Journalism Review*, Vol. 29 No. 6, pp. 18-24.

Fazio, R.H., Herr, P.M. and Powell, M.C. (1992), "On the development and strength of category–brand associations in memory: The case of mystery ads", *Journal of Consumer Psychology*, Vol. 1 No. 1, pp. 1-13.

Flower, L. and Hayes, J.R. (1980), "The cognition of discovery: defining a rhetorical problem", *College Composition and Communication*, Vol. 31 No. 1, pp. 21-32.
Fulgoni, G.M. and Lipsman, A. (2017), "The downside of digital word of mouth and the pursuit of media quality: how social sharing is disrupting digital advertising models and metrics", *Journal of Advertising Research*, Vol. 57 No. 2, pp. 127-131.
Gairola, S., Lal, Y.K., Kumar, V. and Khattar, D. (2017), "A neural clickbait detection engine", available at: https://pdfs.semanticscholar.org/96f9/e55ff59ec574ce500a9c499124709db404da.pdf (accessed 1 May 2018).
Hannan, M.T. and Freeman, J. (1993), *Organizational Ecology*, Harvard University Press, Cambridge, CB.
Hart, R.P. and Daughton, S.M. (2015), *Modern Rhetorical Criticism*, Routledge, New York, NY.
Hilbe, J.M. (2011), *Negative Binomial Regression*, Cambridge University Press, Cambridge, CB.
Jiang, Q., Tan, C.H. and Wei, K.K. (2012), "Cross-website navigation behavior and purchase commitment: a pluralistic field research", in *16th Pacific Asia Conference on Information Systems in Vietnam, Ho Chi Minh City,* Paper 193.
Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J. and Zittrain, J.L. (2018), "The science of fake news", *Science*, Vol. 359 No. 6380, pp. 1094-1096.
Loewenstein, G. (1994), "The Psychology of curiosity: a review and reinterpretation", *Psychological Bulletin*, Vol. 116 No. 1, pp. 75-98.
Lombardi, C. (2017), "Competition and the public interest in the digital market for information", Discussion Paper, *Institute for European Integration, Europa-Kolleg Hamburg,* No. 1/17.
López-Sánchez, D., Herrero, J.R., Arrieta, A.G. and Corchado, J.M. (2017), "Hybridizing metric learning and case-based reasoning for adaptable clickbait detection", *Applied Intelligence*, pp. 1-16.
Mihaylov, T., Mihaylova, T., Nakov, P., Màrquez, L., Georgiev, G.D. and Koychev, I.K. (2018), "The dark side of news community forums: Opinion manipulation trolls", *Internet Research*, Vol. 28 No. 5, pp. 1292-1312.
McQuarrie, E.F. and Mick, D.G. (1996), "Figures of Rhetoric in Advertising Language", *Journal of Consumer Research*, Vol. 22 No. 4, pp. 424-438.
Molyneux, L. and Coddington, M. (2019), "Aggregation, Clickbait and Their Effect on Perceptions of Journalistic Credibility and Quality", *Journalism Practice*, pp. 1-18.
Orosa, B.G., Santorun, S.G. and García, X.L. (2017), "Use of clickbait in the online news media of the 28 EU member countries", *Revista Latina de Comunicación Social*, Vol. 72, pp. 1261-1277.
Phillips, B.J. and McQuarrie, E.F. (2002), "The development, change, and transformation of rhetorical style in magazine advertisements 1954–1999", *Journal of Advertising*, Vol. 31 No. 4, pp. 1-13.
Rochlin, N. (2017), "Fake News: Belief in Post-truth", *Library Hi Tech*, Vol. 35 No. 3, pp. 386-392.
Roelofs, P. and Gallien, M. (2017), "Clickbait and impact: how academia has been hacked", *Impact of Social Sciences Blog*, available at: http://eprints.lse.ac.uk/85023/1/impactofsocialsciences-2017-09-19-clickbait-and-impact-how-academia.pdf (accessed 5 April 2018).
Rowland, C. K. and Marz, R. (1982), "Gresham's Law: The regulatory analogy", *Review of Policy Research*, Vol. 1 No. 3, pp. 572-580.
Potthast, M., Köpsel, S., Stein, B. and Hagen, M. (2016), "Clickbait Detection", in *European Conference on Information Retrieval,* Springer, Cham, pp. 810-817.
Rony, M.M.U., Hassan, N. and Yousuf, M. (2017), "BaitBuster: a clickbait identification framework", in *The Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI, pp. 8216-8217.
Rubin, V.L. (2017), "Deception detection and rumor debunking for social media", in Sloan, L. and Quan-Haase, A. (Eds.), *The SAGE Handbook of Social Media Research Methods*, pp. 342-364.
Rubin, V., Brogly, C., Conroy, N., Chen, Y., Cornwell, S.E. and Asubiaro, T.V. (2019), "A news verification browser for the detection of clickbait, satire, and falsified news", *The Journal of Open Source Software*, Vol. 4 No. 35, pp. 1-3.
Scacco, J. M. and Muddiman, A. (2016), "Investigating the influence of "clickbait" news headlines", *Engaging News Project Report*, available at: https://mediaengagement.org/research/clickbait-headlines/ (accessed August 2016).
Scaraboto, D., Rossi, C.A.V. and Costa, D. (2012), "How consumers persuade each other: rhetorical strategies of interpersonal influence in online communities", *BAR-Brazilian Administration Review*, Vol. 9 No. 3, pp. 246-267.

Schilling, M.A. and Phelps, C.C. (2007), “Interfirm collaboration networks: the impact of large-scale network structure on firm innovation”, *Management Science*, Vol. 53 No. 7, pp. 1113-1126.
Schlosser, A.E., White, T.B. and Lloyd, S.M. (2006), “Converting web site visitors into buyers: how web site investment increases consumer trusting beliefs and online purchase intentions”, *Journal of Marketing*, Vol. 70 No. 2, pp. 133-148.
Shu, K., Wang, S., Le, T., Lee, D. and Liu, H. (2018), “Deep headline generation for clickbait detection”, in *2018 IEEE International Conference on Data Mining (ICDM)*, IEEE, pp. 467-476.
Sisodia, D.S. (2019), “Ensemble learning approach for clickbait detection using article headline features”, *Informing Science: The International Journal of an Emerging Transdiscipline*, Vol. 22, pp. 31-44.
Vatz, R.E. (1973), “The myth of the rhetorical situation”, *Philosophy & Rhetoric*, Vol. 6. No. 3, pp. 154-161.
Vellingiri, J., Kaliraj, S., Satheeshkumar, S. and Parthiban, T. (2015), “A novel approach for user navigation pattern discovery and analysis for web usage mining”, Journal of Computer Science, Vol. 11 No. 2, pp. 372-382.
Wang, C.L., Ye, L.R., Zhang, Y. and Nguyen, D.D. (2005), “Subscription to fee-based online services: What makes consumer pay for online content?”, *Journal of Electronic Commerce Research*, Vol. 6 No. 4, pp. 304-311.
Zannettou, S., Chatzis, S., Papadamou, K. and Sirivianos, M. (2018), “The good, the bad and the bait: Detecting and characterizing clickbait on YouTube”, in *2018 IEEE Security and Privacy Workshops, IEEE*. pp. 63-69.
Zhang, W., Kang, L., Jiang, Q. and Pei, L. (2018), “From buzz to bucks: The impact of social media opinions on the locus of innovation”, *Electronic Commerce Research and Applications*, Vol. 30, pp. 125-137.
Zheng, H.T., Yao, X., Jiang, Y., Xia, S.T. and Xiao, X. (2017), “Boost clickbait detection based on user behavior analysis”, in *Web and Big Data. APWeb-WAIM 2017, Lecture Notes in Computer Science, Springer, Cham,* Vol. 10367, pp. 73-80.
Zhou, Y. (2017), “Clickbait detection in tweets using self-attentive network”, available at: https://arxiv.org/pdf/1710.05364.pdf (accessed 1 May 2018).

Table 1. A Brief Review of Clickbait Detection Approaches (Algorithmic Design)

Studies	Algorithm(s)	Validation Dataset	Performance
Potthast et al. (2016)	Simple random forest classifiers	2,992 tweets	0.76 precision, 0.76 recall
Biyani et al. (2016)	Gradient boosting decision tree (GBDT)	4,073 webpages	0.749 F-score
Chakraborty et al. (2016)	Support Vector Machines, Decision Trees, and Random Forests	200 news	0.93 accuracy
Cao and Le (2017)	Random forest regression (RFR)	21,000 headlines/titles	0.82 accuracy, 0.61 F1-score
Anand et al. (2017)	A neural network (NN) architecture based on recurrent neural network (RNN)	15,000 news headlines	0.98 accuracy, 0.98 F1-score
Gairola et al. (2017)	RNN	a collection of 19,538 posts	0.65 F1-score
Zhou (2017)	Self-attentive neural network (SANN)	102,045 tweets	0.86 accuracy, 0.68 F1-score
Zheng et al. (2017)	GBDT	Two datasets including 32,037 and 11,193 articles respectively	0.75 precision, 0.81 recall
López-Sánchez et al. (2017)	A case-based reasoning methodology	One dataset contains a total of 32,000 headlines from digital newspapers	0.99 accuracy
Rony et al. (2017)	SoftMax classifier	32,000 news headlines	0.98 accuracy

Table 2. Summary of Variables Used in Study 1 (7,481 observations)

Variable	Notation	Mean	Std. Dev	Min	Max
Number of Clicks of Article *i*	RDS_i	19090.421	18670.623	146	420697
Extent of hyperbole in the headline of Article *i*	HYP_i	0.302	0.256	0	0.796
Extent of insinuation in the headline of Article *i*	INS_i	0.296	0.373	0	0.823

Extent of puzzle in the headline of Article *i*	PUZ_i	0.272	0.236	0	0.639
Extent of visual rhetoric in the headline of Article *i*	VIS_i	0.219	0.282	0	0.701
Length of headline of Article *i*	HDL_i	20.932	6.638	4	64
Number of “Likes” of Article *i*	LKN_i	200.429	285.541	0	4931
Number of “Sharing” of Article *i*	SRN_i	406.214	813.082	0	23852
Whether Article *i* is a lead article or not	ILP_i	*(1965 lead articles)*			

Table 3. Correlation of Variables Used in Study 1

Variables	HYP_i	INS_i	PUZ_i	VIS_i	HDL_i	LKN_i	SRN_i	ILP_i
HYP_i	1							
INS_i	0.150	1						
PUZ_i	0.260	0.337	1					
VIS_i	-0.210	0.113	0.206	1				
HDL_i	-0.186	-0.256	-0.338	0.426	1			
LKN_i	0.416	0.628	0.661	0.503	-0.200	1		
SRN_i	0.476	0.646	0.578	**0.701**	0.377	**0.722**	1	
ILP_i	0.667	0.544	0.482	0.501	-0.338	0.487	0.406	1

Table 4. Model Estimation Results of Study 1

Variable	**Coefficient**			
	Poisson Regression		Negative Binomial Regression	
	Model 1	Model 2	Model 3	Model 4
HYP_i	0.018**	0.018***	0.017*	0.015**
	(0.021)	(0.025)	(0.023)	(0.020)
INS_i	0.023*	0.020***	0.021**	0.028***
	(0.056)	(0.054)	(0.035)	(0.022)
PUZ_i	0.012*	0.009	0.007	0.007
	(0.017)	(0.013)	(0.037)	(0.011)
VIS_i	0.009*	0.012*	0.005*	0.008**
	(0.002)	(0.022)	(0.007)	(0.008)
HDL_i	--	-0.006*	--	-0.004*

		(0.001)		(0.001)
LKN_i	--	0.088***	--	0.074***
		(0.081)		(0.086)
SRN_i	--	0.054***	--	0.044***
		(0.071)		(0.087)
ILP_i	--	0.112*	--	0.089**
		(0.351)		(0.100)
cons	12.936*	13.674**	13.080**	13.013***
	(0.420)	(0.489)	(0.523)	(0.689)
Akaike information criterion (AIC)	5826.7	5541.0	1359.4	1022.5
Bayesian information criterion (BIC)	5891.4	5602.3	1398.8	1051.3

*p-value <=0.1; **p-value <=0.05; ***p-value <=0.01

Table 5. Summary of Variables Used in Study 2 (2,424 observations)

Variable	Notation	Mean	Std. Dev	Min	Max
Visit traffic to publisher *j* at time *t*	AUN_{jt}	376,391.333	254,152.620	142,098	1,008,022
The number of clickbait published by publisher *j* at time *t-1*	$CFQ_{j,t-1}$	70.580	69.216	9	420
Age of publisher *j* at time *t-1* (month)	$AGE_{j,t-1}$	31.320	16.891	7	106
The average number of visits to each article published by publisher *j* at time *t*	$ARS_{j,t}$	14,560.588	13,025.447	94	420,697
The total number of articles published by publisher *j* at time *t*	$ATN_{j,t}$	124.030	159.331	57	686
The number of newly added subscribers of publisher *j* at time *t*	$NRU_{j,t}$	2,196.000	3,609.802	559	36,295
The total number of subscribers of publisher *j* at time *t*	$PTU_{j,t}$	692,033.600	334,535.921	235,239	1,273,382

Categorical niche of publisher *j*	$PDM_{j,t}$	Categorical variables with 25 categories.

Table 6. Correlation of Variables Used in Study 2

Variables	$CFQ_{j,t-1}$	$AGE_{j,t-1}$	$ARS_{j,t}$	$ATN_{j,t}$	$NRU_{j,t}$	$PTU_{j,t}$
$CFQ_{j,t-1}$	1					
$AGE_{j,t-1}$	0.474	1				
$ARS_{j,t}$	0.668	0.550	1			
$ATN_{j,t}$	0.614	0.339	0.470	1		
$NRU_{j,t}$	0.548	-0.441	0.662	0.685	1	
$PTU_{j,t}$	0.440	0.741	0.220	0.643	-0.518	1

Table 7. Model Estimation Results of Study 2 (Fixed-effect model)

Variable	**Coefficient**			
	Poisson Regression		Negative Binomial Regression	
	Model 1	Model 2	Model 3	Model 4
$CFQ_{j,t-1}$	0.085* (0.056)	0.062** (0.041)	0.030** (0.048)	0.026** (0.018)
$CFQ^2_{j,t-1}$	-0.015* (0.009)	-0.014** (0.019)	-0.005** (0.003)	-0.007*** (0.004)
$AGE_{j,t-1}$	0.096* (0.022)	0.103** (0.153)	0.072** (0.120)	0.066** (0.118)
$AGE_{j,t-1} * CFQ_{j,t-1}$	-0.023** (0.014)	-0.017* (0.036)	-0.027** (0.067)	-0.011** (0.024)
$AGE_{j,t-1} * CFQ^2_{j,t-1}$	-0.002* (0.003)	-0.003*** (0.007)	-0.006** (0.015)	-0.009*** (0.011)
$ARS_{j,t}$	--	0.016* (0.033)	--	0.020* (0.089)
$ATN_{j,t}$	--	0.008** (0.016)	--	0.007*** (0.012)
$NRU_{j,t}$	--	0.001* (0.002)	--	0.003 (0.002)
$PTU_{j,t}$	--	0.375** (0.590)	--	0.179** (0.336)

$PDM_{j,t}$	Included but values not presented			
cons	8.726**	8.968***	9.209***	8.483***
	(12.770)	(9.632)	(11.589)	(10.226)
AIC	7504.6	7041.7	3581.2	2783.5
BIC	7566.5	7092.4	3620.7	2839.8

*p-value <=0.1; **p-value <=0.05; ***p-value <=0.01