Explaining and predicting click-baitiness and click-bait virality

Article *in* Industrial Management & Data Systems · July 2022

DOI: 10.1108/IMDS-01-2022-0003

2 authors:

Swagato Chatterjee
Queen Mary, University of London

Meghraj Panmand
Indian Institute of Technology Kharagpur

Explaining and predicting click-baitiness and click-bait virality

Structured Abstract

Purpose:

In the age of social media, when publishers are vying for consumer attention, click-baits have become very common. Not only viral websites but also mainstream publishers, such as news channels, who use click-baits for generating traffic. Therefore, click-bait detection and prediction of click-bait virality have become important challenges for social media platforms to keep the platform click-bait free and give better user experience. In this study, we try to explore how the contents of the social media posts and the article can be used to explain and predict social media posts and the virality of a click-bait.

Method:

We have used 17745 tweets from Twitter with 4370 click-baits from top 27 publishers, and applied econometric along with machine learning methods to explain and predict click-baitiness and click-bait virality.

Findings:

We find that language formality, readability, sentiment scores, and proper noun usage of social media posts and various parts of the target article plays differential and important roles in click-baitiness and click-bait virality.

Theoretical Contribution:

The paper contributes towards the literature of dark behavior in social media at large and click-bait prediction and explanation in particular. It focuses on the differential roles of the social media post, the article shared and the source in explaining click-baitiness and click-bait virality via psycho-linguistic framework. The paper also provides explanability to the econometric and machine learning predictive models, thus performing methodological contribution too.

Managerial Implications:

The paper helps social media managers create a mechanism to detect clickbaits and also predict which ones of them can become viral so that corrective measures can be taken.

Originality and Value:

This is one of the first papers which focus on both explaining and predicting click-baitiness and click-bait virality.

Keywords: Click-bait, virality, formality, sentiment, text-mining

1. Introduction

In the age of rampant use of the internet and instant gratification, the consumption of online content has increased exponentially. Social media users, for instance, are possibly the biggest content consumers, consuming all types of content. The fear of missing out (FOMO), online fatigue, and social comparisons have reduced the barriers of consuming and sharing unverified online content, leading to an exponential increase in fake news, hoax, rumour, and click-baits contents (Wessel et. al., 2016; Kim and Dennis, 2019; Talwar et. al., 2019). This has happened due to active participation of some users in such dark social media behaviour. Research on such dark behaviour is still in nascent stage (Talwar et. al., 2019; Quandt et al., 2022).

In this particular study, we focus on click-baits. Click-baits may be defined as web content, produced, and marketed to maximize advertisement revenue by attracting lots of traffic to a website via attracting the reader's attention using stories/headlines/posts that entice them to click (Munger et al., 2020). Importantly, this is often done at the cost of quality, authenticity, and exactitude (López-Sánchez et. al., 2018). One must note that click-baits are different from fake news, rumours and hoax. While the later provide false or unverified information, click-bait gives true information but of lower quality. Therefore, while fake news, rumour and hoax also strive to make the user believe on the content, click-bait will not try to do that. The attraction of traffic by the click-baits is achieved by using provocative and sensational text in social media content or in the title text. Often, the language used in the title is just enough to ensure curiosity, but not enough to get the full information about the topic, creating thereby a curiosity gap, which in turn leads the reader to click the link for further info (Potthast et al., 2016). However, the article that the click-bait refers to usually fails to deliver on the promise of exciting/surprising information. As the curiosity and expectations of the users are not fulfilled by the website content, often the websites pointed by click-bait links are not rated well, as they have higher bounce rates (López-Sánchez et al., 2018). Some examples of click-bait headlines include for instance "20 celebrities who have beaten cancer"; "Cabin Crew Takes Secret Pictures, You Won't Believe the Results"; or "Life Insurance companies hate this new trick" etc. We have chosen click-baits as the domain of study as social media platforms have fairly tackled the other three social media evils such as rumour, hoax and fake news. However, social media platforms are still struggling how to tackle click-baits. Zhang and Clough (2020) report almost 70% of the social media content in WeChat, the most popular social networking site in China, are clickbaits. Such huge number makes the domain very relevant for the researchers.

Detecting the click-baits in a social media and stopping them from being viral is very important for the social media platforms and its users. While click-bait cannot be considered as illegal or dangerous, they are usually frowned upon, as they tend to waste the reader's time and generate discontent (López-Sánchez et al., 2018). In fact, click-baits are disliked by both public institutions and private organizations due to their deceptive nature (Mihaylov et al., 2018). For instance, the European Union has declared that they would work against click-bait articles published in online media with "catchy, provocative, and sensationalist front-page" headlines and without quality content (Orosa et al., 2017). On the other hand, social media platforms like Facebook may face trust issues both for their users and public institutions due to such click-baits. Thus, these platforms (notably Facebook) have begun taking initiatives to reduce the number of click-baits by correctly identifying them and punishing the perpetrators (Babu et al., 2017). The viral nature coupled with low-quality content of click-baits has led these platforms to create algorithms, which can automatically detect and filter out click-baits (López-Sánchez et. al., 2018). Extant literature has talked about a number of natural language processing and machine learning-based methods to detect click-baits automatically (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016; Khater et. al., 2018; Zhang and Clough, 2020). Yet, every new type of click-baits is evolving, suggesting some drawbacks of the existing models. One problem that most of these studies face is that they lack any conceptual model or framework while exploring click-baits as the studies primarily focused on click-bait detection instead of focusing on the impact of click-baits on readers (Zhang et. al., 2020). Studies on the psychological process involved in click-bait response of the users is very limited. While some researchers have focused on emotional response of the click-bait headline (Pengnate, 2019), others have focused on how the impact of clickbaitiness of source derogation and sharing behavior of the users (Mukherjee, 2022). However, none of these studies focused on the drivers of click-baitiness and the psychological process under it. Better understanding of the psychological process will not only help the platform managers detecting click-bait better today, but also help them to mitigate future threats. Social media houses and marketers will keep on trying new techniques and applying new technologies to catch user attention. Combating such strategies will be easier for the platform managers if they can understand the psychological process of click-baits better. However, the psychological process of a social media post being considered as click-bait has not been addressed, although such an exploration is important to better understand the possible drivers of click-baitiness. This creates a gap in the literature. We fill this gap by exploring click-bait drivers which are governed by the psychological process as proposed in this study. Moreover, the click-baitiness of a link does not only depend on the potential of the social media post to attract readers, but also depend on the attractiveness of the article title, which is often displayed when a link is shared, and the article content, which differentiates between an otherwise viral content and a click-bait. Therefore, a platform manager should use models and algorithms which also take in to account such information which explaining click-baitiness and predicting and detecting a click-bait. The attractiveness and information quality of a text can be found from linguistic characteristics (Aladhadh et al., 2018; Zagovora et al., 2018). Therefore, such a linguistic analysis of the article title and content is an obvious addition to the extant literature of click-bait prediction models. However, very few studies have actually used linguistic characteristics of the website-content or the title-text while studying click-bait, thus leaving a lacuna in the literature. We fill this lacuna by answering the following question: "What makes a social media post a click-bait?"

Click-baits attract a lot of traffic as they create inquisitiveness in the mind of the readers. However, very few readers are expected to share the click-bait in social media on their own as sharing click-bait may impact their status in the social networks negatively (Mukherjee et al., 2022). Click-baits are mostly shared by the ad spend by the marketers. Therefore, if the social media channels remain careful about such social media posts which are boosted by marketers, clickbait can be eradicated. However, this is not enough. It has been observed in recent times that many click-baits are becoming viral. Therefore, click-bait detection and explaining the click-baitiness of a social media post is not the only problem for the social media channels, stopping the click-baits to become viral is also another challenge for the platform managers. This begs an important research question: 'what exactly drives the virality of click-baits?' The above question is more pertinent as researchers have seen users performing dark social behavior such as trolling, sharing fake news, sharing click-baits, etc. (Lowry et. al., 2016; Wessel et. al., 2016; Talwar et. al., 2019). Therefore, one needs to know whether and how organic virality of click-bait is possible.

The question of virality become further more important when some legitimate media houses are also using click-bait to increase popularity and traffic. Recently even mainstream news media have begun using click-baits of their news posts in order to attract traffic to their online presence, which in fact, has its own set of challenges (Chen et. al., 2015). As Frampton (2015, p1) in his BBC News article on 'Click-bait journalism' said: "*One perennial frustration for the online reader is the "look at me" headline, which can have negative consequences*." This means, that the headlines which are attraction seeking, but not of quality, may lead to negative evaluations by the readers. Still main-stream media houses are increasing using this technique as their click-bait news posts are also getting viral. This makes the virality prediction and identification of the root cause of virality more important for the platform managers, who want these media houses use the platform and at the same time want the platform clean.

Herein, we have given a theoretical explanation of the drivers of clickbaitiness and click-bait virality, which no other prior work on click-baits have provided. Moreover, we have used computational linguistics and sentiment mining techniques for click-bait detection and click-bait virality prediction. We find that both the formality and sentiment scores of various parts of the website content, along with the post title can enrich conceptual framework and predictive validity for both click-baitiness and click-bait virality. We have also found that the posts targeted towards a person/object/place etc. and readability of the posts can increase the social media engagement with a click-bait post. The study contributes towards the literature of click-bait detection, click-bait virality, dark-behavior of social-media-users, application of linguistics in social media research, etc. The study also helps in creating better click-bait detection engines and understanding the user behavior so that corrective measures can be taken.

In the following sections, we cover extant literature to lay the foundation for our research. Next, we run an empirical study on click-bait detection and click-bait virality prediction, followed by a discussion of the results, theoretical contribution, managerial implications, and future scope of research.

2. Literature Review

2.1. Click-bait Detection Models

Click-baits are not fake news or spam, which have been extensively studied in the last two decades. Click-baits, unlike fake news or spam, are websites with authentic but poor-quality information (López-Sánchez et. al., 2018). Therefore, click-bait detection has also been separately addressed by extant literature (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016; López-Sánchez et. al., 2018). Click-baits are ambiguous, exaggerated, inflammatory, bait-and-switch, teasing, often with wrong formatting and graphic (Biyani et. al., 2016). Chakraborty et. al. (2016) compared click-baits and non-click-baits in terms of sentence structure and language use. However, other researchers have used the website content associated with the click-bait link and used features such as term-frequencies (unigrams and bigrams), binary features encoding the presence of exclamation marks, numbers, and superlative adverbs, text sentiment, headline formality etc. (Biyani et al., 2016; Potthast et. al., 2016; Khater et. al., 2018; Pujahari and Sisodia, 2019). Pandey et al. (2018) have used swarm intelligence algorithm and found better predictive performance, although they used similar set of variables as done by past researchers. Zhang et al. (2020) majorly focused on click-bait headlines only, vis a vis their effect on traffic generation. Zhang and Clough (2020) created an automated click-bait detection technique using data from Chinese media where they used usage of certain words in the clickbaits. Rajapaksha (2020), in his doctoral dissertation, used multimodal fusion and transfer learning methods.

Our study is different from all of the studies above in a number of areas. We have used linguistic characteristics of various parts of the website content, and social media posts separately to create features, which have not been explored in extant literature. Moreover, extant literature did not focus on the conceptual model of ‘click-baitiness’, i.e., what makes a click-bait post. Extant literature has also not focused on how a click-bait becomes viral. Such knowledge is important for click-bait news items or authentic click-baits, as virality is also an important outcome that tabloid journals are looking for (Chen et. al., 2015). Our study fills this gap. The overall research questions of our study are:

RQ1: How can we find whether a social media post is click-bait or not?

RQ2: How can we predict the virality of a click-bait?

RQ3: How the conceptual framework of clickbaitiness and click-bait virality can be developed?

2.2. Social Media Post Virality

Extant literature has focused on social media post virality extensively. Some of the factors that are found to be crucial for social media post virality includes the content (Porter and Golan, 2006), social network structure (Weng et. al., 2013), seeding strategies, and source characteristics (Han et. al., 2020). In terms of the content, the attractiveness (Porter and Golan, 2006), informativeness, and emotional aspect of content (Tellis et. al., 2019) lead to virality. In terms of the social network, the source, structure, and size of the network and how the message is propagated through a network may have its influence on virality (Weng et al., 2013). Even in the context of dark content, researchers have explored the antecedents of sharing of misinformation (Tornberg, 2018), fake news (Talwar et al., 2019; 2020), low-quality information (Qiu, 2017). However, no exploration has been done on how a click-bait can become viral and what are the antecedents of such an event. Although, Mukherjee et al. (2022) explored the linkage between clickbaitiness and sharing of social media posts, they found that click-baits are less likely to become viral on their own as the users may understand manipulative intent of the click-baits and will derogate the source. Therefore, their finding suggests that sharing of dark content will not happen in the context of click-bait, which is in contrast to past literature. Thus, our study tries to fill the gap by contributing to the literature of click-bait virality and to the overall literature of social media virality.

3. Theoretical Framework

3.1. Stimulus-Organism-Response Model

We have used the stimulus-organism-response (S-O-R) model as the theoretical framework. The 'S-O-R framework' was first proposed by Mehrabian and Russell (1974), and was later modified by Jacoby (2002). According to this framework, environmental stimuli provoke cognitive and emotional conditions of an individual, leading to behavioral responses (Donovan and Rositer, 1982). Such a framework has been used in consumer behavior, computer and website experience, advertising research, etc. (Mollen and Wilson, 2010; Rose et. al., 2012; Eroglu et. al., 2003). In the social media context, SOR model has been used in brand engagement in brand communities (Kamboj et al., 2018), purchase decision based on online reviews (Bigne et al., 2020) etc. We have applied the 'S-O-R framework' in the domain of consumer behavior.

Stimulus means the influencing factor, which arouses an individual (Eroglu et. al., 2001). In the context of click-bait, the attractive and curiosity generating nature of a website article's heading and the text may be considered as the stimuli. Such text in the headings increases the reader's curiosity for further probing, leading to information clutter in the users' minds. This can be further explained by the information-gap theory of curiosity (Loewenstein, 1994). As per this theory, the gap between an individual's knowledge and attention leads to stimulation of a 'feeling of deprivation' or curiosity.

'Organism' is the cognitive and affective condition of the users resulting from the stimuli, which will lead to their response (Loureiro and Ribeiro, 2011). While affective condition alludes to the feelings and emotions generated in a user's mind, the cognitive condition refers to mental states related to the acquisition, processing, retention, and retrieval of information (Eroglu et. al., 2001). In the case of click-baits, the affective condition of curiosity in the mind of the user triggers emotions such as anticipation and interest. After reading the full article of click-bait, the user cognitively processes the information promise from the text post and the article text. Such matching of anticipated and obtained information go through cognitive processing. If the match is high, the emotions generated would be positive (such as admiration, trust, surprise, amazement, joy, etc.); on the other hand, if the match is low, the emotions generated would be negative (such as anger, rage, disgust, annoyance, boredom, etc.).

Lastly 'response' refers to the outcomes of the above-mentioned mechanism in terms of user behavior. For instance, before the emotions of anticipation and interest generated from curiosity, it leads to the behavioral response of clicking the click-bait link and reading the article. Moreover, after reading the article, positive emotions lead to sharing and liking the post, while negative emotions lead to blocking or unfollowing the social media page, which shared the click-baits. We use the above psychological framework to explain our hypotheses in the following sections.

3.2. Formality of Language

We have used formality language as an important construct in our framework. Heylighen and Dewaele (1999) coined the definition and measure of formality in language, which has then been used in computational linguistic quite extensively. The theoretical definition of formality has been given based on two major aspects, i.e., context-dependence and fuzziness. Understanding a language does need an understanding of the background and the context. Formal language tries to ensure that the recipient would understand the message without any knowledge of the context or background or without any assumptions. Such avoidance of context-dependence makes formal language direct and unambiguous. On the other hand, fuzziness refers to a situation when the meaning of the expression is not unambiguously determined, even if the context is provided. For example, if a person says ‘The room is hot’, it is difficult to understand what is the exact temperature. Formal styles of language not only avoid context-dependent expressions, but they also tend to avoid fuzzy expressions. Formal speakers of the language tend to choose the least fuzzy expression while communicating. Meanwhile, expressions in themselves could be both fuzzy and context-dependent. For example, a ‘big’ house means something different in the context of a metropolitan city in comparison to the context of a remote village. Both these types of ambiguity need additional information to be resolved, but in case of context dependence, such information is readily available, whereas, in case of fuzziness, additional efforts are needed to collect more information to resolve this ambiguity. Similarly, there is also expression which is characterized by a high degree of fuzziness and a low degree of context-dependence. An example of this could be the evasive answers given by politicians, which actually minimize the amount of precise information. Formality, on the other hand, maybe defined as a linear combination of context independence; it is the inverse of fuzziness (called precision). An orthogonal dimension of expressivity is also defined, which is a linear combination of context-dependence and precision.

Heylighen and Dewaele (1999) have developed a measure of formality score, which has further been used in categorizing different types of the corpus. While developing the measure, words have been divided into two classes, context-dependent expressions, and context-independent expressions. Context-dependent words are mostly adverbs, pronouns, and interjections, while context-independent words are usually nouns, adjectives, and prepositions. To measure formality, frequencies of formal categories are added, and the frequencies of deictic categories are subtracted, post which the score is normalized to 100. The following is the formula for formality score:

F = (noun frequency + adjective freq. + preposition freq. + article freq. – pronoun freq.– verb freq. – adverb freq. – interjection freq. + 100)/2

The frequencies refer to the percentages of the number of words belonging to that specific category with respect to the total number of words. Extant literature has used the above-mentioned formality score for measuring textual formality (Lahiri et. al., 2011; Pavlick and Tetreault, 2016; Hyland and Jiang, 2017), including in the context of microblogs such as tweets (Sen et al., 2015). We have also used the above definition and measure of formality score to find its relationship with click-bait and its virality. In the context of click-bait detection, formality scores have been used by Biyani et al. (2016) and Pujahari and Sisodia (2019). However, these studies have used the formality score only in the context of the article and not explored the effect of formality scores in different parts of the clickbait (social media post, article headline, article content etc.). Moreover, none of them have given the theoretical underpinning on why formality of the text can lead to clickbaitiness. Additionally, an exploration of the effect of such a formality score of clickbait virality is also missing. We try to fill these gaps.

3.3. Formality and Click-baitiness

We have seen how the S-O-R framework explains how the attractive nature of click-bait social media posts creates curiosity, leading to emotions such as interest and anticipation, prompting thereby the reader to click the link and reading the article. This mechanism is triggered only when the social media post can generate enough curiosity in the mind of the reader, which leads to a further probe of information (Zhang et. al., 2020). Therefore, a social media post needs to be such that the information is not holistic in nature (Zhang et. al., 2020). Such fuzzy nature is the hallmark condition of click-baits. Therefore, click-baits often try to create interests for a subject by signaling something unexpected or unknown within that context, which effectively triggers curiosity. Based on the same, one could assume that both fuzziness and context-dependence are major characteristics of click-bait media posts, which lead to consumer response behavior following the S-O-R framework. Therefore, as both fuzziness and context-dependence are negatively related to the formality of the language, we posit:

H1: Formality of the language of the social media post is negatively related to the click-baitiness of the post

The paper tries to explore click-baitiness and not the click-through-rate of a link. The word 'bait' is associated with false hope. Therefore, click-bait is not only related to the 'probability of click' but also with the degree of false hope or 'bait'. A post becomes clickbaity only when the article associated with the post does not keep the promise of information quality generated by the post. Before that, the post is just an attractive post, but not a clickbait. Therefore, the article level information must be used while exploring the clickbaitiness of a post, based on the following characteristics of clickbait that the article related to clickbait have lower information quality.

Click-baits tend to have higher bounce rates, as the content quality does not keep its promise generated by the social media post. On the other hand, and most importantly, websites, which use click-bait for traffic generation, tend to make a lot of money from advertisements posted on their pages (López-Sánchez et. al., 2018; Zhang et. al., 2020). Thus, higher traffic to the websites naturally leads to a higher impression of advertisements, which in turn generates revenue. Nevertheless, the higher bounce rate is a detrimental factor for revenue generation, because bounced traffic is not considered while calculating the impressions. Therefore, one of the major challenges for click-baits include the ability to reduce bounce rates. Extant literature suggested that user attention is an important driver of the reduction of bounce rate (Sng, 2017). In order to minimize the bounce rate, click-bait articles keep the reader's tension, curiosity, and anticipation alive to such an extent, whereby the reader is almost glued to the website and doesn't feel the need to go off it. Thus, the longer a reader stays on the website, the higher are the chances of it not being counted among bounced traffic. Looking at this scenario from a linguistic viewpoint, one notes that the language used in such cases is highly informal, context dependent and fuzzy thereby. This suggests that the articles which have higher formality are expected to be less click-baity. Thus, we propose:

H2: Formality of the language of the article title is negatively related to the click-baitiness of the post

H3: Formality of the language of the article is negatively related to the click-baitiness of the post

Figure 1 gives the theoretical framework cum process diagram of consumer reactions to click-baits. As can be understood from the process diagram, the perceived click-baitiness depends on the perceived disconfirmation between the information quality of the post and information expectation of the post. As the informality leads to increased information expectation by creating an information gap, informality widens such disconfirmation and increases click-baitiness.

<Figure 1 here>

3.4. Article and Title Formality and Click-bait Virality

Click-bait virality will depend on the social media sharing and engagement of the users with the social media post. Such sharing and engagement can be conducted by two types of users: one who clicks the link and reads the article and one who does not. In this section we explain the engagement and sharing behavior of the users who are of the first type. In the next section, we discuss about the second type.

In the first type, as the users read the article, the text and the title of the article will be more salient during the sharing behavior. The influence of the article title and text on the click-bait virality can be explained using the expectancy disconfirmation theory. According to this theory, individuals create an expectation based on signals received from external and internal sources (Oliver, 1980). The overall evaluation by the individual will depend on the disconfirmation of the perception from the expectation (Oliver, 1997). Such evaluation, if positive, leads to positive attitude formation and behavioral outcome. Expectancy disconfirmation theory has been extensively used in marketing and information science domain, in the context of customer satisfaction, user engagement etc. (Liao et. al., 2016; Qazi et. al., 2017).

In the context of clickbaits, the higher is the information content of the text and the title of the article, the lower is the disconfirmation between the information expected and the information received. As formality of the language used will reduce ambiguity of the information, formality will also increase the quality and richness of the information received which will reduce the disconfirmation. Such lower gap will induce positive attitude and emotions in the mind of the users leading to increased sharing behavior of the users. Therefore, we posit:

H4: Formality of the language of the click-bait article title is positively related to the virality of the click-bait post

H5: Formality of the language of the article positively related to the virality of the click-bait post

3.5. Click-baitiness and Click-bait Virality

Motivations of engaging in dark behavior in social media have been an interesting area in academia (Lowry et. al., 2016). Extant literature has suggested a number of reasons why users engage with dark posts in social media, such as fake news, rumors, and click-baits (Wessel et. al., 2016; Talwar et. al., 2019). The motivations behind sharing the fake news and click-baits can often be common. Self-disclosure, gossip seeking behavior, social comparison, entertainment, pass-time, socialization, fear of missing out (FOMO), and social fatigue can lead to user engagement with fake news and dark content in social media (Talwar et. al., 2019; Apuke and Omar, 2020). Moreover, instantaneous sharing of content for awareness generation and altruism is strongly related with dark content sharing, while fact-checking reduces such content sharing behavior (Apuke and Omar, 2020; Talwar et. al., 2020). Thus, one can argue that click-bait sharing is a low cognitive process governed by instantaneous reactions, social fatigue, and FOMO (Islam et al., 2020). Thus, the informality of the social media post and the overall clickbaitiness of the social media post drive the sharing behavior of such posts. The above relationship is expected to be more prominent for such users who have not identified the post to be a click-bait or did not click the post yet, as one who clicks the link and realize the article to be a clickbait may show lesser dark behaviour than one who does not. Thus, we posit:

H6: Formality of the language of the click-bait social media post is negatively related to the virality of the click-bait post

H7: Click-baitiness of a click-bait post is positively related to the virality of the click-bait post

3.5. Other Relevant Variables

3.5.1. Readability

Readability can be characterized as the ease of comprehending a written text (Klare, 1963). Readability has been extensively used in the domains of education, literacy, and psycholinguistics (Chall & Dale, 1995; DuBay, 2004). Readability is measured mainly based on word complexity or word length or the number of syllables (e.g., Flesch, 1948; Gunning, 1952; McLaughlin, 1969) and syntactic complexity or average sentence length (Chall & Dale, 1995; DuBay, 2004). Texts which have longer sentences and unfamiliar words cannot be read fluently, while fluency of text is something the readers enjoy (Alter & Oppenheimer, 2009). As enjoyment, attraction and engagement are inter-related (Berger & Milkman, 2012), fluent social media posts get higher readership and engagement (Pancer et. al., 2019). The fluency has a positive impact on veracity, confidence and many other judgements (Reber et al., 2004; Winkielman et al., 2003). The metacognitive experience resulting from fluency leads to positive affect in the mind of the readers. The readers attach the content as the attribution of the positive affect (Kohli et al., 2005; Moreau et al., 2001). Such fluency-affect relationship leads to higher social media engagement of the users (Pancer et. al., 2019). The same is expected to be observed in the context of click-bait too where the goal is to increase consumption and engagement with the click-bait. Following the above discussion, we expect that readability will improve click-baitiness and click-bait virality.

H8: Readability of the (a) social media post, (b) article title and (c) article content will be positively related to click-baitiness

H9: Readability of the (a) social media post, (b) article title and (c) article content will be positively related to clickbait virality

3.5.2. Content with names and social mentions

Extant literature suggests that social mentions have a significant impact on the virality of a social media post like Twitter posts (Pramanik et. al., 2016; Bao et. al., 2018). Mentions in the micro-blogging website increase readership and probability to engagement (Bao et. al., 2018). While social mentions are basically names and hence proper nouns, usage of proper nouns are expected to increase click-bait virality. Moreover, often contents on celebrities, a special location, or some brand attract more attention to the users which can be used by the click-bait publishers to attract traffic. Therefore, such posts have higher chances of going viral too. Hence, we hypothesize:

H10: Usage of proper noun in (a) social media post and (b) article title of a click-bait increases the click-bait virality

Figure 2 gives the theoretical model for click-baitiness and click-bait virality.

<Figure 2>

4. Empirical Study

4.1. Data

We have used the data provided by Potthast et. al. (2018) in their Click-bait Detection Challenge 2017. The data consist of every tweet of 27 top-most retweeted English-language news publishers published from December 1, 2016, to April 30, 2017. The dataset had tweet text, media attachments, and Twitter metadata.

To ensure proper representativeness, stratified sampling has been done from a corpus of 459541 tweets, based on various publishers and various days of the months, as discussed by Potthast et. al. (2018), yielding a set of 38517 tweets and archived articles. Amazon Mechanical Turk (AMT) has been used to annotate such tweets by five different workers in terms of click-bait strength using a 4-point scale (1=not click-baiting and 4=click-baiting). Quality testing has been done on the annotation to ensure credibility. The mode of four annotators' response has been used for classifying a tweet into a click-baiting category (1 to 4) and the fifth annotator's responses have been used to break the ties. The researchers received a Fleiss' κ of 0.35 which can be considered as a "fair" agreement (Landis and Koch, 1977; Potthast et. al., 2018). A median split of the mean scores has been done to get the click-bait class (click-bait or non-click-bait). This may create bias as, in real life, the click-baits are much lesser in count in comparison to non-clickbaits. However, this method will balance the dataset. The validity check of the categorization was done by comparing our categorization with categorization done by experts (Potthast et. al., 2018). Out of these tweets, we randomly chose 17745 tweets, which were presented before the researchers as training data. As the click-bait challenge, 2017 is closed now, we could not access the testing data and hence had to use only the training data representative of the full dataset of our model.

The final dataset contains 17,745 data points out of which 13,375 are non-click-baits, while the remaining 4,370 are click-baits. We only considered the text part of the dataset, and did not consider media attachments; nevertheless, we checked if there were any media attached or not. The dataset contained the text of the post, the post timestamp, the article title and content, post-media availability, mean score of click-baitiness given by the annotators, and the click-bait class. We web-scraped the number of followers of the publishers, the total number of retweets, shares, and likes of the twitter posts on 1st February 2020.

4.2. Text Mining and Feature Generation

At first, we pre-processed all text data; this includes converting them to lower case, removing punctuations and numbers. However, we did not remove the stop words, as these are required for formality score calculation. Then, we divided the article contents for each social media post into two equal parts to calculate the formality scores for these sections. Further, we calculated the formality scores based on the formula stated above. Formality scores were calculated for the twitter post, the target article's heading, and the two parts of the target article's text. For Parts-of-speech tagging, the Spacy library of Python was used.

We used the text length and sentiment scores of the twitter post, the target article's heading and the target article's text as covariates. Sentiment scores were calculated using the Vader library in Python (Hutto and Gilbert, 2014). Sentiment scores were included in the model to test for possible positivity bias of the users which may lead to click-baitiness and click-bait virality (Spottswood and Hancock, 2016). Post-media availability, a binary variable, indicating whether a media (picture or video) was attached to the social media post, was also used as a covariate. Post-media attracts more attention than text. Therefore, the presence of post-media can increase click-baitiness. Other covariate that was included in the virality model was the number of followers of the publisher. The number of followers is expected to impact the virality of any social media post (Han et. al., 2020). We used the number of proper nouns to check if clickbaits which are targeted towards a person or a place increase click-baitiness. This is expected as users who are followers of the person or the place will get attracted by such clickbaits. We have also included the readability scores of the twitter post, article title, and article posts in the model as covariates. We measured readability using the Gunning Fog Index (Gunning, 1969). This index tries to estimate the number of years of formal language training is required to understand the text. A higher FOG Index means lower readability.

4.3. Empirical Validation of Conceptual Model

To validate the conceptual models, we ran linear regression on click-baitiness, virality, and logistic regression of click-bait class. Virality has been measured as log(1+number of retweets + number of likes + the number of comments). Virality does not have any specific measure and researchers have used their own measures to quantify virality (Han et. al., 2020). Our measure quantifies the overall social engagement of the users with the posts which acts as the proxy of virality. However, we have also performed modeling for individual aspects of virality (likes, retweets, and replies). We have used log(1+number of likes), log(1+number of retweets), and log(1+number of comments) as the independent variables. The independent variables considered include the formality scores, text length, and sentiment scores of the twitter post, the target article's heading, and the two parts of the target article's text along with media availability. We had to drop the text length of the third quarter of the click-baiting article and social media availability for multicollinearity issues. We scaled all the independent variables before using them in regression. The descriptive of the rest of the variables are given in Table 1. Moreover, the regression results are posted in Table 2 and Table 3. We first ran the regression on the covariates only, and then included the formality scores in the model to check whether the goodness of fit increased. Herein, the multicollinearity check suggests that the VIF scores are below 4, suggesting thereby that no significant multicollinearity exists in the models. Notably, the models also show fitment in improvement when the formality scores are included over and above the control variables.

<Tables 1, 2 and 3 here>

After controlling for sentiment, text count, number of followers of the publisher, readability, and the usage of proper nouns, we found that the formality of the twitter post and article title is negatively related with click-baitiness ($\beta_{twitter\text{-}post}$=-0.061, $p<0.001$, $\beta_{article\text{-}title}$=-0.017, $p<0.001$) and click-bait class ($\beta_{twitter\text{-}post}$=-0.491, $p<0.001$, $\beta_{article\text{-}title}$=-0.144, $p<0.001$). This supports H1 and H2. Further, we also found that the language formality of the article is negatively related to click-baitiness (β=-0.026, $p<0.001$) and click-bait class (β=-0.180, $p<0.001$). This supports H3. Further, we also found that FOG Score of the twitter post and the article are negatively related to click-baitiness ($\beta_{twitter\text{-}post}$=-0.028, $p<0.001$, $\beta_{article}$=-0.010, $p<0.001$) and click-bait class ($\beta_{twitter\text{-}post}$=-0.249, $p<0.001$, $\beta_{article}$=-0.127, $p<0.001$), thus partially supporting H8. We also found that the sentiment of the twitter post and the article are positively related with click-baitiness ($\beta_{twitter\text{-}post}$=-0.073, $p<0.001$; $\beta_{article}$=0.01, $p<0.001$) and click-bait class ($\beta_{twitter\text{-}post}$=0.184, $p<0.001$, $\beta_{article}$=0.114, $p<0.001$).

While exploring virality we found that the language formality of the title of the click-baiting article is positively related to virality (β=0.05, $p<0.05$), which supports H4. However, the above result is visible only in the case of replies (β=0.102, $p<0.001$) and not in other components of virality such as retweets (β=0.057, NS) and likes (β=0.044, NS). Additionally, we found that the formality of the article text is negatively related with virality (β=0.136, $p<0.001$), retweets (β=0.056, $p<0.05$), likes (β=0.171, $p<0.001$), and replies (β==0.060, $p<0.01$), thus supporting H5. Moreover, the formality of the twitter post is negatively related with the virality of the click-baits (β=-0.159, $p<0.001$), retweets (β=-0.143, $p<0.001$), likes (β=-0.155, $p<0.001$), and replies (β==-0.189, $p<0.001$), supporting H6. We also found that click-baitiness is positively related to click-bait virality (β=0.046, $p<0.001$), likes (β=0.067, $p<0.01$), and replies (β==0.059, $p<0.01$), supporting H7.

We also found that the FOG score of the twitter post ($\beta_{virality}$=-0.082, p<0.001; $\beta_{retweets}$=-0.069, p<0.05; β_{likes}=-0.081, p<0.001; $\beta_{replies}$=-0.092, p<0.001), article content ($\beta_{virality}$=-0.125, p<0.001; $\beta_{retweets}$=-0.073, p<0.01; β_{likes}=-0.145, p<0.001) and article title ($\beta_{virality}$=-0.15, p<0.001; $\beta_{retweets}$=-0.149, p<0.001; β_{likes}=-0.144, p<0.001; $\beta_{replies}$=-0.147, p<0.001) are negatively related to click-bait virality. This supports H9.

The number of proper nouns used in the article title is positively related with virality (β=0.295, p<0.001), retweets (β=0.232, p<0.001), likes (β=0.324, p<0.001) and replies (β==0.237, p<0.001). Proper noun usage in social media post is positively related with overall virality (β=0.051, p<0.05) and replies (β=0.095, p<0.001). This supports H10.

Further, we found that the sentiment of the article ($\beta_{virality}$=-0.062, p<0.01; $\beta_{retweets}$=-0.093, p<0.001; $\beta_{replies}$=-0.114, p<0.001) and title ($\beta_{replies}$=-0.097, p<0.001) are negatively related to click-bait virality. On the other hand, the sentiment of the twitter post in positively related with virality ($\beta_{virality}$=0.07, p<0.01; β_{likes}=0.089, p<0.01). The number of followers of the twitter publisher is also positively related with virality ($\beta_{virality}$=0.484, p<0.001; $\beta_{retweets}$=0.460, p<0.001; β_{likes}=0.489, p<0.001; $\beta_{replies}$=0.428, p<0.001).

Longer post ($\beta_{virality}$=-0.181, p<0.001; $\beta_{retweets}$=-0.154, p<0.001; β_{likes}=-0.172, p<0.001; $\beta_{replies}$=-0.260, p<0.001) and longer title (β=-0.166, p<0.001; $\beta_{retweets}$=-0.126, p<0.001; β_{likes}=-0.192, p<0.001; $\beta_{replies}$=-0.107, p<0.001) is negatively related to virality. The theoretical and managerial implications of all the results are given in the discussion section.

4.4. Predictive Models

We explored both regression and machine learning models for click-bait detection and virality prediction. For click-bait detection, the click-bait mean score has been used as the dependent variable. This is done to provide a comparative performance of our data with that of the submissions of Click-bait Challenge 2018, from where the data has originally been taken. For virality prediction, we used the virality score as per the conceptual models discussed above. For both cases, we used regression-based techniques instead of classification, as the dependent variable is continuous in nature. Importantly, the dependent variables include the formality score, readability score, sentiment score, and the word-count of the twitter post, the target article's heading, and the two parts of the target article's text along with the availability of post-media. We scaled all the features using MinMax scaler, and randomly divided the data into training, validation, and testing data on a 50:25:25 ratio. Further, on the training data, we have run the following regression and machine learning models: linear regression, lasso regression, elastic net regression, decision tree, K-nearest neighbor (KNN), gradient boosting, random forest, and support vector machine. We have performed 3-fold cross-validation on the training data. Next, we have done the hyperparameter tuning of the best model in the validation data. Lastly, we used the tuned model to predict in the testing data.

For clickbaitiness, we find gradient boosting, which gives the best result in the training data, as listed in Table 4. The MSE of test data was found to be 0.045. Further, the tuned model gives MSE score on 0.037 is the validation data and 0.043 is the testing data. Taking 0.5 as the cut-off score for click-baitiness to predict a twitter post to be a click-bait, we get to see the performance of classification, as listed in Table 5. The results are good when compared to the results of other works done of Click-bait Challenge 2018 data (Potthast et. al., 2018). The average MSE score of the submissions of the challenge was 0.0775 and the median score was 0.0445.

We also successfully predicted click-bait virality, and the results of prediction accuracy are listed in Table 4. The random forest model is found to be the best performing model with the lowest MSE score in the training data (0.872). Later we hyper tuned the same with the validation data and found MSE to be 0.75 in validation-data and 0.92 in testing data. While we cannot claim that our predictive model outperformed all other available models, it did a decent job comparison to available models and, at the same time, we came up with a model which gives explainable predictive solutions.

<Tables 4 and 5 here>

5. Discussion

In this study, we found how the linguistic nature of the text of the social media post and target article can be used to provide a conceptual framework and predictive validity to click-baitiness and click-bait virality. We found that the formality of the language of the social media post and the article title are negatively related to the click-baitiness of the post. This can be explained by the S-O-R framework and information-gap theory which suggests that the attractive nature of click-bait social media posts creates curiosity which in turn leads to emotions such as interest and anticipation leading to clicking the link and reading the article (Mehrabian and Russell, 1974; Donovan and Rositer, 1982; Loewenstein, 1994; Eroglu et. al., 2001; Jacoby, 2002; Loureiro and Ribeiro, 2011). Such curiosity generation is associated with fuzziness and context-dependence of the language of the social media post and the title of the article, hence the above relationship. We also found that the language informality of the article makes it more click-baity as such language practice generates optimal curiosity which reduces bounce rate but at the same time does not provide quality information or does not reduce information-gap (Sng, 2017).

Next, we have also explained how click-baits become viral. We have also found that clickbaitiness itself leads to click-bait virality, which is in line with the motivation of sharing and engaging with dark content in social media such as fake news, rumors, and click-baits (Talwar et. al., 2019; Talwar et al., 2020; Islam, 2020). We also found that the informality of the text of the social media post leads to click-bait virality as an interest provoking social media post makes a sharer more popular in the social media space. On the other hand, the formality of the article title and text increases the sharing behavior as such formality removes ambiguity and makes the article information rich.

Another variable related to computational linguistic which contributes to click-baitiness and click-bait virality is the readability of the post, title, and content. Readability, in general, improves click-baitiness and click-bait virality. This is in line with our expectation as readability or fluency improves reading enjoyment which increases attraction towards and engagement with social media posts (Pancer et. al., 2019).

We also found the positive sentiment in the text of the leads to clickbaitiness. This may be because of positivity bias among the users who want to consume positive content more (Spottswood and Hancock, 2016). "Clickbaitiness" does intend the user to consume the content although there is low informational value. The positive sentiment in the social media post will attract the user towards the article link. The positive sentiment in the actual article will ensure the users consumer the content and do not bounce from the link. This will ensure higher ad revenue. However, we found the sentiment of the text of article title and article content is negatively related to virality. This may be because sharing of misinformation or dark content is related to lower cognitive effort, entertainment seeking, self-promoting behavior. Negative content is more in sync with such dark behavior as supported by extant literature (Stieglitz and Dang-Xuan, 2013; Wang et al., 2020). Therefore, the result seems in sync with extant knowledge.

Often longer informal article title creates more curiosity while longer social media post makes the social media post itself unattractive. Therefore, while longer article title is positively related to click-baitiness, longer social media post is not. This can also be because we have sourced our data from twitter which is a platform of microblogging and users are habituated to read shorter posts (Zhao and Rosson, 2009). After controlling for clickbaitiness and text formality scores, longer article titles and social media posts are negatively related to virality, may be because long click-baity text creates higher anticipate information and hence higher dissatisfaction leading to lower social engagement.

We found more followers of the publisher will lead to the increased virality of the click-bait which is self-explanatory. However, we also found that higher usage of proper nouns in article title can lead to higher virality. Proper nouns are associated with names of persons, brands, and places; often clickbaity articles targeting a person, a brand or a place leads to higher emotional responses which lead to virality. Moreover, proper nouns can also be related to social mentions in twitter posts which leads to virality (Pramanik et. al., 2016; Bao et. al., 2018).

We also found good predictive power of our model in comparison to existing models of click-bait detection. We also found that the regression models are at par with standalone machine learning models, while predictive click-baits and their virality. The ensemble models like random forest and boosting model like gradient boosting works better than both regression and standalone machine learning models.

5.1. Theoretical Contribution

The paper has a number of theoretical contributions. Firstly, this is one of the first papers, which conceptually explores the click-baitiness of social media posts and their virality. It also provides the predictive validity of the conceptual models. Extant literature has focused on click-bait detection (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016, Zhang et al., 2020) and the impact of click-bait in business and public opinion (Munger et al., 2020; Zhang et. al., 2020). However, providing theoretical reasoning of a predictive model becomes important to properly tackle clickbait problems. Moreover, with the advent of usage of click-bait by the mainstream media and news publishers, the organic virality of click-baits also becomes a challenge; and this hadn't been studied thus far in extant literature. In our study, we explore this gap. Thus, our paper contributes to overall click-bait literature.

Secondly, extant literature of click-bait detection has focused on machine learning tools and post characteristics (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016, Zhang et al., 2020). A few studies have focused on linguistic aspects but remained limited to the social media post and/or article headline (Biyani et. al., 2016; Khater et. al., 2018; Zhang et al., 2020). However, we have majorly explored the linguistic characteristics of not only the social media posts but also the article itself. We have also explored language formality of the article text, which has not been used in the context of click-bait detection before. Thus, this represents our work's significant methodological contribution in click-bait detection literature (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016, Zhang et al., 2020). It also contributes to the stream of literature that explores the effect of linguistic characteristics in social media posts (Berger & Milkman, 2012; Pancer et. al., 2019).

Third, we have explained click-baitiness and click-bait virality from a theoretical point of view. Most of the extant literature has not explored the psychological framework behind click-bait consumption and social engagement (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016; Potthast et. al., 2016, Munger et al., 2020). Others have focused on information gap theory to study consumer response to clickbaits (Zhang and Clough, 2020; Zhang et al., 2020). We used the S-O-R framework and information gap theory to explain how users react to click-bait social media posts, and how the post and article characteristics go on to impact the psychological framework of the users. Thus, we also contribute to both click-bait literature (Chen et. al., 2015; Biyani et. al., 2016; Chakraborty et. al., 2016;

Potthast et. al., 2016, Zhang et al., 2020) and the application of the S-O-R framework (Mehrabian and Russell, 1974; Donovan and Rositer, 1982; Jacoby, 2002; Eroglu et. al., 2003; Mollen and Wilson, 2010; Rose et. al., 2012) and Information gap theory (López-Sánchez et. al., 2018; Zhang et. al., 2020).

One interesting finding of our study is to find the link between clickbaitiness and click-bait virality. We also found that the drivers of click-baitiness i.e., the informality of the text of the social media post and the article are positively related to clickbait virality. This makes an important contribution towards the social engagement literature (Zadeh and Sharda, 2014), with special focus on the motivations of the users behind sharing and engaging with dark content in social media such as fake news, rumors, and click-baits (Wessel et. al., 2016; Talwar et. al., 2019; Talwar et al., 2020; Islam, 2020). This also provides the validity of dark behavior of social media users based on large scale data. It also contrasts with the recent findings of Mukherjee et al. (2022) who suggests that clickbaits will not lead to higher sharing.

The paper also contributes towards the stream of literature which focuses on text readability and its outcome. Extant literature has found that readability improves engagement in various domains such as education, literacy, psycholinguistics, social media (Klare, 1963; Chall & Dale, 1995; Pancer et. al., 2019; DuBay, 2004). Our paper strengthens the above argument and show readability has its applications in the dark social media posts too. Further exploration can also be done in the context of rumour, hoax, fake news, fake reviews etc.

Our result also contributes towards the extant literature focusing on social media content and user behavior (Stieglitz and Dang-Xuan, 2013; Spottswood and Hancock, 2016; Wang et al., 2020). We have shown contrasting effect of text sentiment on attracting, consuming and sharing social media posts. While positivity biased in seen during getting attracted to and consuming a social media post and the article link, even if it is a click-bait, the sharing behavior of click-bait are more associated with negative sentiment. The above strengthens the positivity bias in social media consumption hypothesis (Spottswood and Hancock, 2016). However, the negativity bias during sharing gives new insights into the sentiment-dark-behavior relationship (Stieglitz and Dang-Xuan, 2013; Wang et al., 2020).

5.2. Managerial Contribution

The paper has a number of managerial contributions. Firstly, the paper provides a new click-bait detection model using the linguistic characteristics of the article. Our click-bait detection is source independent and can identify click-baits even from mainstream media. Therefore, it can be used to regulate social media platforms more efficiently and in an unbiased way. Moreover, social media platforms have developed algorithms that can already identify headlines that are deceptive. The findings of this work where language and formality are used will enrich such models and make it more powerful in identifying dark contents. The platform managers can create algorithms which can automatically find the readability, formality, sentiment etc. scores from the text of the social media post and the articles shared. These scores will improve the clickbait-detection models used by the platform managers.

Secondly, we have not only predicted possible click-baits, we have also predicted their virality. Therefore, a midlevel click-baity post with very high potential virality can be blocked or regulated using our models. Our virality predictions could also be used by publishing companies to predict whether they need to promote or demote a click-bait post. A platform manager should create a joint metric which is dependent on both click-baitiness and predicted virality, instead of relying only one of the two metrics. Such metric can be used to rate the potential threats from a social media post in terms of click-baitiness. Platform managers should try to minimize such threats.

Thirdly, our paper looks to explain the psychological framework followed by users while reacting to social media posts and engaging with a click-bait. Knowledge of such a psychological process would certainly help the publishers, along with the platform managers give better information and service to the users. Explainable prediction models are becoming more or more important every day in the context of ethical AI, which our study ensures. Moreover, recent changes in social media platforms, such as regulatory requirements etc., have made social media companies more socially responsible and users more aware. In this light, this paper will give direction to social media companies in training their users to detect click-baits easily based on some basic rules. Social media companies can also flag potential click-baits, thus helping the users.

Fourthly, we found that the informality of the text of the social media post leads to click-bait virality while the formality of the article title and text increases the sharing behavior. Therefore, the results indicate that the formality of the social media post and that of the actual article should be treated differently by the social media platforms while detecting a click-bait. This is a very important finding in the context of click-bait. Platform managers can create programs which can collect the text of the social media post along with the title-text and content-text of the article which is shared in the post. They can also write an algorithm following our paper or similar papers which will use the formality scores of these texts from three different sources (social media post, article title, article text) to predict click-baitiness and detect click-baits.

4.3. Limitations and Future Scope

While we have tried to explore the psychological process of click-baits, we have worked on the above based on macro-level data. An individual user-level analysis is warranted to further explore the nuances of consumer psychology involved in the process. Future researchers can focus on the same. We could collect the channel level follower data only during the data collection process, while a better measure would have been historical data of followers when the post was published. With the lack of such data, we used the current followers as a proxy. This is a limitation of our study. We have also used current virality measures. Though it is also a limitation, click-bait topics by major publishers are mostly contextual and contemporary to the time of sharing. Therefore, we do not expect a huge change in virality scores away from the publishing date. This makes the current virality score not so different from virality scores close to the publishing date. We also did not have the exact amount spent on the social media posts, and therefore could not capture the effect of ad spend on virality; this again, is a limitation of our study. The dataset only contains data on US news publishers and not content publishers. Future research should focus on applying linguistic characteristics in click-bait detection and virality prediction for social media posts from content publishers too. Future researchers can also test the applicability of these results in emotion-baits which is an important and upcoming issue in social media. Finally, marketers should also think the long-term effect of click-bait as a marketing strategy. Although click-baits can increase visibility and attract attention of the users, the image of the publisher or the brand may take a hit. Further research should focus on the long-term effect of click-baiting strategy on brand image.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235.

Aladhadh, S., Zhang, X., & Sanderson, M. (2018). Location impact on source and linguistic features for information credibility of social media. *Online Information Review*. 43(1), 89-112

Apuke, O. D., & Omar, B. (2020). Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 101475.

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

Bao, P., Shen, H. W., Huang, J., & Chen, H. (2018). Mention effect in information diffusion on a micro-blogging network. *PloS one*, *13*(3).

Bigne, E., Chatzipanagiotou, K., & Ruiz, C. (2020). Pictorial content, sequence of conflicting online reviews and consumer decision-making: The stimulus-organism-response model revisited. *Journal of Business Research*, 115, 403-416.

Biyani, P., Tsioutsiouliklis, K., & Blackmer, J. (2016). "8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 9-16). IEEE.

Chall, J. S., & Dale, E. (1995). Readability revisited: The new Dale-Chall readability formula. Cambridge, MA: Brookline Books.

Chen, Yimin, Niall J. Conroy, and Victoria L. Rubin. (2015) "Misleading online content: recognizing click-bait as" false news"." *Proceedings of the 2015 ACM on workshop on multimodal deception detection*.

Donovan, R. J., & Rossiter, J. R. (2002). Store atmosphere: an environmental phychology approach. *Retailing: critical concepts. 3, 2. Retail practices and operations*, *2*(1), 77.

DuBay, W. H. (2004). The principles of readability. Impact Information: Costa Mesa, CA.

Eroglu, S. A., Machleit, K. A., & Davis, L. M. (2003). Empirical testing of a model of online store atmospherics and shopper responses. *Psychology & marketing*, *20*(2), 139-150.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–233.

Frampton. B (2015), Clickbait: The changing face of online journalism, *BBC News,* Retrieved from: https://www.bbc.com/news/uk-wales-34213693 on 1st May, 2020

Gunning, R. (1952). The technique of clear writing. New York: McGraw-Hill.

Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication*, 6(2), 3-13.

Han, Y., Lappas, T., & Sabnis, G. (2020). The Importance of Interactions Between Content Characteristics and Creator Characteristics for Studying Virality in Social Media. *Information Systems Research*, Forthcoming.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal?. *English for Specific Purposes*, *45*, 40-51.

Islam, A. N., Laato, S., Talukder, S., & Sutinen, E. (2020). Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. *Technological Forecasting and Social Change*, *159*, 120201.

Jacoby, J. (2002). Stimulus-organism-response reconsidered: an evolutionary step in modeling (consumer) behavior. *Journal of consumer psychology*, *12*(1), 51-57.

Kamboj, S., Sarmah, B., Gupta, S., & Dwivedi, Y. (2018). Examining branding co-creation in brand communities on social media: Applying the paradigm of Stimulus-Organism-Response. *International Journal of Information Management*, 39, 169-185.

Khater, S. R., Al-sahlee, O. H., Daoud, D. M., & El-Seoud, M. S. A. (2018, May). Clickbait Detection. In *Proceedings of the 7th international conference on software and information engineering* (pp. 111-115).

Kim, A., & Dennis, A. R. (2019). SAYS WHO? THE EFFECTS OF PRESENTATION FORMAT AND SOURCE RATING ON FAKE NEWS IN SOCIAL MEDIA. *MIS Quarterly*, 43(3).

Klare, G. R. (1963). The measurement of readability. Ames, IA: University of Iowa Press.

Kohli, C. S., Harich, K. R., & Leuthesser, L. (2005). Creating brand identity: A study of evaluation of new brand names. *Journal of Business Research*, 58, 1506–1515

Lahiri, S., Mitra, P., & Lu, X. (2011). Informality judgment at sentence level and experiments with formality score. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 446-457). Springer, Berlin, Heidelberg.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Liao, G. Y., Huang, H. C., & Teng, C. I. (2016). When does frustration not reduce continuance intention of online gamers? The expectancy disconfirmation perspective. *Journal of Electronic Commerce Research*, *17*(1), 65.

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, 27(4), 962-986.

López-Sánchez, D., Herrero, J. R., Arrieta, A. G., & Corchado, J. M. (2018). Hybridizing metric learning and case-based reasoning for adaptable click-bait detection. *Applied Intelligence*, *48*(9), 2967-2982.

Loewenstein, G. (1994), "The Psychology of curiosity: a review and reinterpretation", Psychological Bulletin, Vol. 116 No. 1, pp. 75-98.

McLaughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of Reading*, 12, 639–646

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

Mihaylov, T., Mihaylova, T., Nakov, P., Màrquez, L., Georgiev, G. D., & Koychev, I. K. (2018). The dark side of news community forums: Opinion manipulation trolls. *Internet Research*.

Mollen, A., & Wilson, H. (2010). Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives. *Journal of business research*, *63*(9-10), 919-925.

Moreau, C. P., Lehmann, D. R., & Markman, A. B. (2001). Entrenched knowledge structures and consumer response to new products. *Journal of Marketing Research*, 38, 14–29

Mukherjee, P., Dutta, S., & De Bruyn, A. (2022). Did clickbait crack the code on virality?. *Journal of the Academy of Marketing Science*, 1-21.

Munger, K., Luca, M., Nagler, J., & Tucker, J. (2020). The (null) effects of clickbait headlines on polarization, trust, and learning. *Public opinion quarterly*. 84(1), 49-73

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 42, 460–469.

Oliver, R. L. (1997). Satisfaction: A behavioral perspective on the consumer. New York: Irwin McGraw-Hill.

Orosa, B.G., Santorun, S.G. and Garcıa, X.L. (2017), "Use of click-bait in the online news media of the 28 EU member countries", *Revista Latina de Comunicacion Social,* Vol. 72, pp. 1261-1277

Pancer, E., Chandler, V., Poole, M., & Noseworthy, T. J. (2019). How readability shapes social media engagement. *Journal of Consumer Psychology*, *29*(2), 262-270.

Pandey, D., Verma, G., & Nagpal, S. (2018, September). Clickbait Detection Using Swarm Intelligence. In *International Symposium on Signal Processing and Intelligent Recognition Systems* (pp. 64-76). Springer, Singapore.

Pavlick, E., & Tetreault, J. (2016). An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics*, *4*, 61-74.

Pengnate, S. F. (2019), Shocking secret you won't believe! Emotional arousal in clickbait headlines: An eye-tracking analysis, *Online Information Review*, 43(7), 1136-1150.

Porter, L. and Golan, G.J. (2006), From subservient chickens to brawny men: a comparison of viral advertising to television advertising, *Journal of Interactive Advertising*, 6(2), 30-38.

Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016). Click-bait detection. In *European Conference on Information Retrieval* (pp. 810-817). Springer, Cham.

Potthast, M., Gollub, T., Hagen, M., & Stein, B. (2018). The click-bait challenge 2017: towards a regression model for click-bait strength. *arXiv preprint arXiv:1812.10847*.'

Pramanik, S., Wang, Q., Danisch, M., Bandi, S., Kumar, A., Guillaume, J. L., & Mitra, B. (2016). On the role of mentions on tweet virality. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 204-213). IEEE.

Pujahari, A., & Sisodia, D. S. (2019). Clickbait detection using multiple categorisation techniques. *Journal of Information Science*, 0165551519871822.

Qazi, A., Tamjidyamcholo, A., Raj, R. G., Hardaker, G., & Standing, C. (2017). Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach. *Computers in Human Behavior*, *75*, 450-460.

Qiu, X., Oliveira, D. F., Shirazi, A. S., Flammini, A., & Menczer, F. (2017). Limited individual attention and online virality of low-quality information. *Nature Human Behaviour*, *1*(7), 0132.

Quandt, T., Klapproth, J., & Frischlich, L. (2022). Dark social media participation and well-being. *Current opinion in psychology*. 45, June, 101284

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8, 364–382

Rose, S., Clark, M., Samouel, P., & Hair, N. (2012). Online customer experience in e-retailing: an empirical model of antecedents and outcomes. *Journal of retailing*, *88*(2), 308-322.

Sen, A., Rudra, K., & Ghosh, S. (2015, January). Extracting situational awareness from microblogs during disaster events. In *2015 7th International Conference on Communication Systems and Networks (COMSNETS)* (pp. 1-6). IEEE.

Sng, Y. F. (2017). Study on Factors Associated With Bounce Rates on Consumer Product Websites. In *Business Analytics: Progress on Applications in Asia Pacific* (pp. 526-546).

Spottswood, E. L., & Hancock, J. T. (2016). The positivity bias and prosocial deception on Facebook. *Computers in Human Behavior*, 65, 252-259.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*, *29*(4), 217-248.

Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019). Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services*, *51*, 72-82.

Talwar, S., Dhir, A., Singh, D., Virk, G. S., & Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*, *57*, 102197.

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. *Journal of Marketing*, *83*(4), 1-20.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one*, *13*(9), e0203958.

Wang, Rui, Yuan He, Jing Xu, and Hongzhong Zhang. "Fake news or bad news? Toward an emotion-driven cognitive dissonance model of misinformation diffusion." *Asian Journal of Communication* (2020): 1-26.

Weng, L., Menczer, F., & Ahn, Y. Y. (2013). Virality prediction and community structure in social networks. *Scientific reports*, *3*, 2522.

Wessel, M., Thies, F., & Benlian, A. (2016). The emergence and effects of fake social information: Evidence from crowdfunding. *Decision Support Systems*, *90*, 75-85.

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). *The hedonic marking of processing fluency: Implications for evaluative judgment*. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zadeh, A. H., & Sharda, R. (2014). Modeling brand post popularity dynamics in online social networks. *Decision Support Systems*, *65*, 59-68.

Zagovora, O., Weller, K., Janosov, M., Wagner, C., & Peters, I. (2018). What increases (social) media attention: Research impact, author prominence or title attractiveness?. *arXiv preprint arXiv:1809.06299.*

Zhang, C., & Clough, P. D. (2020). Investigating clickbait in Chinese social media: A study of WeChat. *Online Social Networks and Media*, 19, 100095.

Zhang, W., Du, W., Bian, Y., Peng, C. H., & Jiang, Q. (2020). Seeing is not always believing: an exploratory study of click-bait in WeChat. *Internet Research*. 30(3), 1043-1058

Zhao, D., & Rosson, M. B. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 243-252).

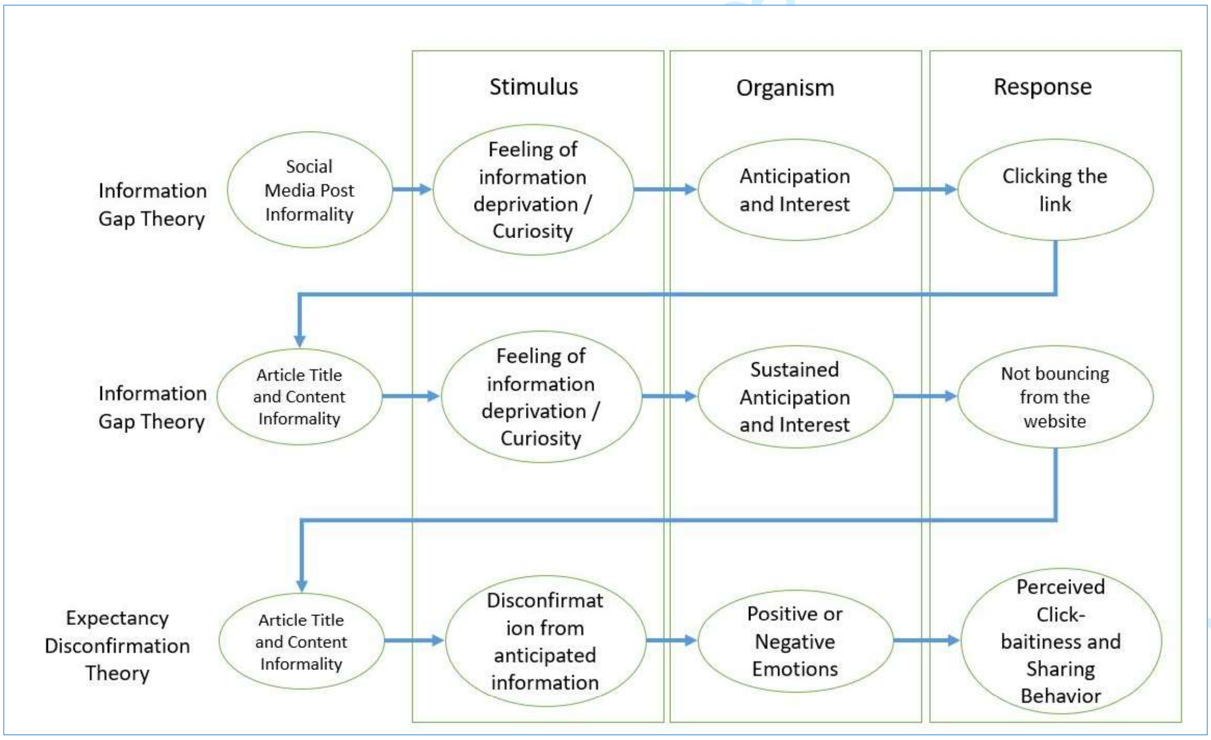

Figure 1: Theoretical Framework cum Process Diagram Explaining User Reaction to Click-baits

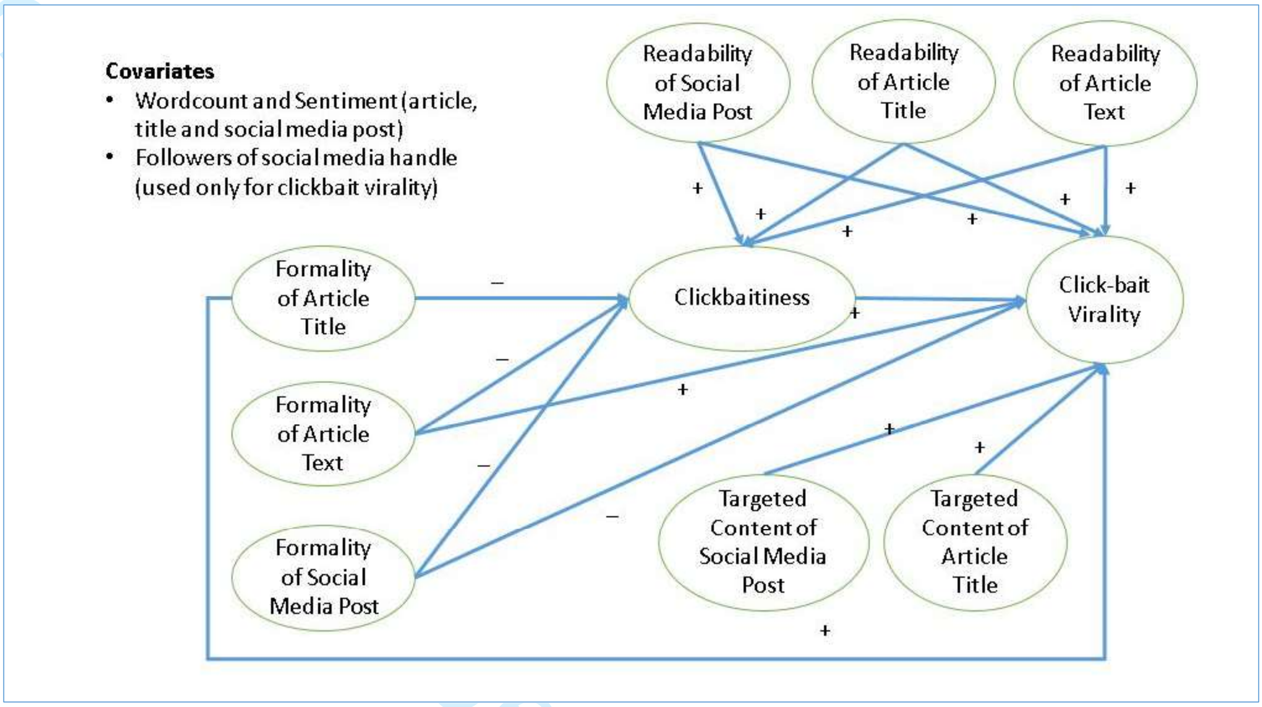

Figure 2: Theoretical model for click-baitiness and click-bait virality

Table 1: Descriptive Statistics of Many Variables in the Dataset (including both click-baits and non-click-baits)

Variables	Mean	SD	MIN	MAX
Retweets Count	104.06	301.57	0	13519
Likes Count	231.24	616.61	0	24109
Replies Count	25.52	66.47	0	2310
Article_Wordcount	630.45	433.1	1	6036
Title_Wordcount	10.35	4.17	1	23
Post_Wordcount	11.68	3.75	1	28
Sentiment_Article	0.33	1.19	-1	1
Sentiment_ArticleTitle	-0.04	0.41	-0.97	1
Sentiment_TwitterPost	-0.03	0.39	-0.96	0.95
Formality_Article	67.84	9.10	0	100
Formality_ArticleTitle	73.73	15.99	0	100
Formality_TwitterPost	70.57	16.73	0	100
Followers of Click-bait Publishers	9318075	11691861	0	47940243
Proper noun count in Click-bait Title	2.85	3.77	0	122
Proper noun count in Click-bait TwitterPost	1.08	1.37	0	15
Virality	4.71	1.63	0	10.33

Table 2: Results of Empirical Validation of Conceptual Models of Click-baitiness and Click-bait Class

Dependent Variables	Click-baitiness	Click-baitiness	Click-bait Class	Click-bait Class
Methods Used	**Linear Regression**	**Linear Regression**	**Logistic Regression**	**Logistic Regression**
AdjR2	0.1161	0.2167		
AIC			18216	17029
(Intercept)	0.326***	0.326***	-1.25***	-1.32***
Article_Wordcount	0.009***	0.009***	0.053***	0.047**
Title Wordcount	0.015***	0.015***	0.235***	0.162***
Post Wordcount	-0.073***	-0.073***	-0.558***	-0.667***
Article_Sentiment	0.010*	0.010***	0.160***	0.114***
Title Sentiment	-0.002	-0.002	-0.012	-0.004
Post Sentiment	0.022***	0.022***	0.174***	0.184***
Post FOG Score	-0.028***	-0.028***	-0.312***	-0.249***
Title FOG Score	0.003	0.002	0.014	0.017
Article FOG Score	-0.010***	-0.010***	-0.216***	-0.127***
Article_Formality		-0.026***		-0.180***
Title Formality		-0.017***		-0.144***
Post Formality		-0.061***		-0.491***

Table 3: Empirical Validation of the Conceptual Model of Click-bait Virality

Dependent Variables	Click-bait Virality	Click-bait Virality	Retweets	Likes	Replies
AdjR2	0.173	0.193	0.108	0.197	0.169
(Intercept)	4.625***	4.623***	3.102 ***	4.157 ***	1.769 ***
Article_Wordcount	-0.004	-0.008	-0.012	-0.010	-0.029
Title Wordcount	-0.172***	-0.166***	-0.126 ***	-0.192 ***	-0.107 ***
Post Wordcount	-0.141***	-0.181***	-0.154 ***	-0.172 ***	-0.260 ***
Article_Sentiment	-0.047*	-0.062**	-0.093 ***	-0.021	-0.114 ***
Title Sentiment	-0.03	-0.018	-0.028	-0.002	-0.097 **
Post Sentiment	0.068*	0.07**	0.058	0.089 **	0.120
Followers	0.47***	0.484***	0.460 ***	0.489 ***	0.428 ***
Proper noun Count in Title	0.289***	0.295***	0.232 ***	0.324 ***	0.237 ***
Proper noun Count in Post	0.051*	0.004	-0.004	0.002	0.095 ***