Is the Buzz on? – A Buzz Detection System for Viral Posts in Social Media

Abstract

Today, online social networks (OSNs) constitute a major part of our lives and have, to a large extent, replaced traditional media for direct communication, as well as information dissemination and gathering. In the vast amount of posts that get published in OSNs each day, some posts do not draw any attention while others catch on, become viral, and develop as so-called buzzes. Buzzes are defined through their characteristics of immediacy, unexpectedness, and intensity. The early detection of buzzes is of vital importance for companies, public figures, institutions, or political parties – e.g., for the pricing of profitable advertising placement or the development of an appropriate social media strategy. While previous researchers developed systems for detecting trending topics, mainly characterised by their intensity, this is the first study to implement a buzz detection system (BDS). Based on almost 120,000 manually classified Facebook posts, we estimated and trained models for the BDS by applying various classification techniques. Our results highlight that, among other predictors, the number of previously passive users who then engage in the buzz post, as well as the number of likes given to the comments, are important. Evaluating the BDS over a five-month evaluation period, we found that these two classifiers perform best and detected over 97% of the buzzes.

Keywords: Buzz Detection, Viral Phenomena, Buzzes, Decision Support, Online Social Networks, Social Media

1. Motivation

Online social networks (OSNs) are a vital platform for participation in social and public life. Every day, OSN users on Facebook, Twitter, or Instagram publish a vast amount of posts. The majority of these do not find much attention, while a few exceptional posts catch on and become viral. Two different types of viral phenomena exist: (1) “trending topics” that fulfil the properties of immediacy and intensity (such as discussions about football matches or the deaths of famous persons) and often re-occur over time (such as lotteries or the annual *Movember* fundraising campaign) (e.g., Asur et al. 2011; Lau, Collier, and Baldwin 2012; Zubiaga et al. 2015), as well as (2) “buzzes” that fulfil the properties of immediacy, unexpectedness, and intensity (Lesot et al. 2012; Deusser et al. 2018). The main difference between trending topics and buzzes is that trending topics do not necessarily fulfil the characteristic of unexpectedness (cf. Section 2). In this study, we concentrate on buzzes: Although they can have significant economic, social, and political consequences, they are not yet as well-studied as trending topics.

Buzzes generally take one of three forms: firestorms, lovestorms and hot topics (Jansen 2019). Companies that are confronted with online firestorms (Pfeffer, Zorbach, and Carley 2014), for instance, can suffer major economic consequences. For example, in 2013, the Italian food and pasta company *Barilla* experienced such a firestorm after *Barilla*'s chairman offended homosexual people in a radio interview. Thousands of social media users retaliated against the company and advocated for boycotting *Barilla* (Facebook 2013a, 2013b). However, buzzes can also have positive impacts. As virality plays a crucial role in advertising (e.g., Hayes, King, and Ramirez 2016; Moldovan, Steinhart, and Lehmann 2019), practitioners may use buzzes for the following purposes: (1) the pricing of profitable advertising placement (Berger and Milkman 2012), (2) integrating with their marketing strategy, like *Delta Airlines* did in their on-board security video (YouTube 2015), or (3) as a launching pad for an improvised marketing intervention (Borah et al. 2019), i.e., real-time marketing communication, such as *Oreo*'s prompt reaction to the power outage during the 2013 *Super Bowl*.

As these examples demonstrate, there is great value in being able to detect buzzes early. Thus, the objective of this study is to develop a concept for buzz detection and its prototypical implementation, while also sharing insights on factors that are important for identifying buzzes early in their evolution.

We collected metadata from public Facebook pages, manually classified almost 120,000 posts, estimated a logistic regression (LR) model, and trained different established classifiers – AdaBoost, support vector machine, and random forests – in order to create our buzz detection system (BDS). The results of the LR yield that lagging variables (cf. Section 3.2.2), such as the comments' content length and their likes, as well as deviation variables (cf. Section 3.2.2), such as the number of previously passive users who then engage in the buzz post and the number of page owner comments, are important factors that help to identify buzzes early in their evolution. Testing the BDS over a five-month evaluation period, we find that an LR with a threshold of 0.01 (i.e., "LR 1%") and SVM performed best. These two algorithms jointly detected over 97% of the buzzes. Comparing these results to the ones of an existing model that comes closest to ours and predicts viral online content, we find that the existing model performs better than chance, but does not outperform our proposed model.

So far, researchers have only developed detection systems for trending topics. To the best of our knowledge, no study to date has tried to implement a detection system for buzzes. Thus, this study makes a significant contribution to research on viral online content. Our practical contribution lies in providing a BDS that could actually be used for brand and social media management.

We structure the remainder of this paper as follows: Section 2 elaborates on the theoretical background behind buzzes and their detection. Section 3 extensively describes the architecture of the BDS. We present the evaluation of the BDS in Section 4, followed by a discussion in Section 5. Section 6 concludes with a summary of the study's findings and limitations, and offers a glimpse into possible avenues for future research.

2. Theoretical Background

This study builds on three different approaches that contribute to the detection and prediction of viral online content. First of all, we need to gather information that generally makes content go viral. Secondly, it is important to our study to know which variables constitute a model for viral content prediction and detection. Thirdly, the major objective of this study is to actually implement a detection system for buzzes. Thus, it is important to build on the insights of researchers who have already developed detection systems for trending topics.

In the following, we introduce these three approaches on viral online content in more detail: (1) Researchers concentrated on identifying and understanding variables that drive information diffusion in social networks to explain why some online content gets more attention than others (e.g., Stieglitz and Dang-Xuan 2013; Heimbach et al. 2015; Heimbach and Hinz 2016; Akpinar and Berger 2017; Quesenberry and Coolsen 2019). (2) Researchers developed models that allow for the detection of viral contents (e.g., Szabo and Huberman 2010; Tsur and Rappoport 2012; Deusser et al. 2018; Tellis et al. 2019; Rietveld et al. 2020). (3) A few researchers actually built systems to detect trending topics (Mathioudakis and Koudas 2010; Cvijijk and Michahelles 2011; Lau, Collier, and Baldwin 2012). From the first two approaches, we can conclude that content-, user-, and engagement-related variables play an important role in identifying drivers of viral online content; we will include these variables in the model that underlies the BDS (cf. Section 3.2.2). As the third approach is most relevant to our work, we elaborate on existing detection systems in more detail to derive further avenues for improvement, which we will address in the present study. There are three studies in which researchers proposed systems for the real-time detection of viral online content (Mathioudakis and Koudas 2010; Cvijijk and Michahelles 2011; Lau, Collier, and Baldwin 2012). However, Mathioudakis and

Koudas (2010), Cvijijk and Michaehelles (2011), Lau, Collier, and Baldwin (2012) concentrated on trending topics, whereas we are interested in buzzes. Thus, they did not explicitly consider the unexpectedness characteristic of viral content. To explain trending topics, the authors gave examples, like the deaths of Michael Jackson and Amy Winehouse, a terrorist attack, as well as a National Basketball Association match. These examples are not highly unexpected, as we can already anticipate that news reporting and public discussions will accompany the death of a famous person, a terrorist attack or a major sports event. Our study focuses on unanticipated and unexpected cases (cf. the buzz definition in Section 3.2.1) – a choice that constitutes its importance and uniqueness.

The unexpectedness characteristic is also the reason why we chose to use manually classified data (buzz vs. non-buzz) as our dependent variable, rather than rely on the number of likes, comments, or shares. After all, those factors would not necessarily have detected posts that fulfil all three buzz characteristics, i.e., immediacy, unexpectedness, and intensity. The aforementioned examples usually obtain lots of reactions, represented by an extraordinary amount of likes, comments, or shares in OSNs. However, they do not represent buzzes and would therefore have been misclassified by a model solely estimated and trained on engagement variables like likes, comments, and shares.

To clarify this point, consider a post about a lottery: When a large number of users can participate, we anticipate that this post will receive an extraordinary number of likes, comments or shares, but this is not unexpected and therefore does not create a buzz according to our definition. Facebook live sessions are another example: Users can view a Facebook page owner who is live on Facebook and comment on the live session post in real time in order to ask questions or share reactions. Such live sessions usually result in an extraordinary number of comments, which, again, is not very surprising. Thus, using the number of comments as a dependent variable, our model would detect many posts that do not fulfil the buzz criteria. By instead using manually classified data to delineate posts as buzzes and non-buzzes, we can estimate and train models that help us to detect buzzes based on posts that are immediate, unexpected, and intense.

Nevertheless, we examined the three previous studies in which the researchers developed systems to detect trending topics in order to analyse their approaches and derive relevant system features. Table 1 summarises the results of these analyses. Because of our focus on buzz detection, we did not seek to cluster several posts to one trend; thus, we did not solely rely on content-related variables, like “bursty” keywords, but instead took into account other indicators. Following Mathioudakis and Koudas (2010), our system allows for user interactions – e.g., to select Facebook pages for buzz monitoring (for the full scope of user interactions, see Section 3.3). Similarly to Cvijijk and Michahelles (2011) and Lau, Collier, and Baldwin (2012), we will test our system in dedicated evaluation periods and compare it to the performance of an existing model that predicts viral online content.

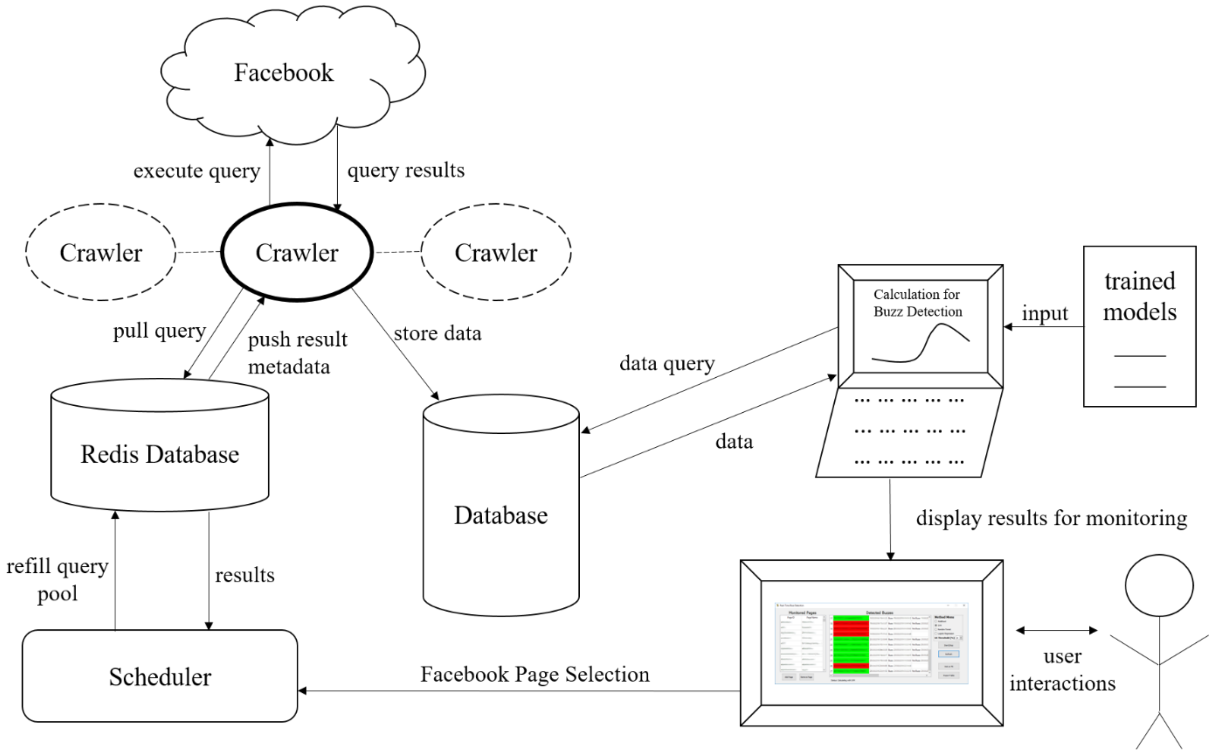
Table 1 Analysis of the three most relevant studies and their implications for our study

System Characteristics	Mathioudakis & Koudas (2010)	Cvijijk & Michahelles (2011)	Lau, Collier, and Baldwin (2012)	Our Study
Purpose	*TwitterMonitor*: Trend Detection	Trending Topic Monitoring on Facebook	Trending Topic Identification on Twitter	Buzz Detection on Facebook
Approach	1. identify "bursty" keywords 2. group "bursty" keywords into trends 3. extract additional trend information	1. topic identification 2. cluster detection	• Latent Dirichlet Allocation (LDA) topic modelling • cluster detection	• 199,910 manually classified posts • estimate logistic regression/train different classifiers based on content-, user-, and engagement-related variables
User Interaction	• to submit own trend description • to sort trends by different criteria	no user interaction	no user interaction	among others: • to select Facebook pages for buzz monitoring • to directly visit detected buzzes on Facebook • to export results for further analyses
System Evaluation	no system evaluation	• 10 experiments with 1,000 posts each from different time intervals • performance metrics reported on cluster-level	• 1-month evaluation period on Twitter • trending topics reported • no performance metrics reported	• 2 evaluation periods (3 months + 2 months) with more than 210,000 posts • manual classification • performance metrics reported on post-level • comparison to an existing model that predicts viral online content

3. Architecture of the Buzz Detection System (BDS)

In the following, we describe the architecture of our BDS. Figure 1 offers an overview of the architecture and summarises its components, which we will detail in Section 3.1 to Section 3.3. We first concentrate on the backend of the BDS and describe the processes of data collection and data storage. Second, we explain the buzz detection process by referring to the models we trained, and then move to introduce the frontend of the BDS.

Figure 1 Architecture of the buzz detection system (BDS)



3.1 Data Collection and Storage

The backend of the BDS fulfils the tasks of collecting selected Facebook pages and storing these data, which we elaborate in the following.

In order to collect metadata from public Facebook pages, we used the official Facebook Graph API. We distributed the data collection over a number of independent crawlers. Each of these crawlers acted as a registered user of a Facebook App, which was specifically created for the purpose of this research, and used their own API access token. The crawlers were coordinated using a “Redis” database. “Redis” is an in-memory database, meaning it operates mainly on memory rather than disk storage, giving it vastly increased performance for use cases such as ours. After completing their assigned request, each crawler sent the returned data to the SQL server and a summary of the result to the scheduling server. This design allowed each crawler to be completely unaware of other crawlers and enabled the seamless addition and removal of crawlers.

Each request was issued towards a specific Facebook page or post. If the target was a page (post), the crawler retrieved all posts (comments) as well as the current page like count since the page (post) was last requested. In order to maintain an up-to-date view of each page and post, we had to regularly query them. As activity varies greatly among the monitored pages and posts – some posts receive thousands of comments per hour while others are never commented on – the scheduling server prioritised the request with the highest expected number of updates. This prioritisation depends on both the length of time from when the entry was last queried as well as the type and relevance (or activity rate) of each entry. To prevent posts with little activity from being entirely ignored, and to ensure that new posts are discovered quickly, we added two empirically chosen restrictions to this priority-based querying: At least every 20^{th} query had to pertain to a page and 10% of the total requests had to be dedicated to keeping all posts updated regularly. So the entire database had to be queried at least once in the timeframe that it could be completely queried if only 10% of the queries were available. For example, if requests are processed at a rate of 100 requests per minute, and if there are a total of 1,000 entries, then each entry must be queried at least once every 100 minutes.

In the described way, we repeatedly queried a set of 1,021 public Facebook pages from August 18^{th}, 2016 until April 2^{nd}, 2018. The collected pages belonged to famous persons like politicians, musicians, and athletes, but also to companies, public institutions, political parties, fire and police departments, and relatively unknown people from all over the world. Due to the rarity of buzzes, and to be sure that a buzz occurred on a given page, we selected 76 of the pages through daily checks on news aggregation websites reporting on buzzes, such as news.feed-reader.net. The other 945 pages were selected randomly by applying a three-level randomisation process in order to avoid any selection bias. The randomisation process took place on the website where Facebook lists all existing public Facebook pages. On the first level, we randomly chose a letter indicating the first letter of a Facebook page's name in the A to Z collection provided by Facebook. For the second level, we randomly chose a set of the letter intervals displayed by Facebook on the following page. On the third level, we randomly selected pages displayed on the following page listing the public Facebook pages in order to monitor them with the help of the BDS. In total, we gathered data from 537,467 posts, including 78.85 million comments.

We stored all data required to calculate our variables in a PostgreSQL database. The data comprise: page information, like page ID, page likes, and page name; post information, like post ID, content, timestamp, likes, and comments; and comment information, like comment ID, commenter ID, content, and timestamp.

3.2 Trained Models for Buzz Detection

As we are dealing with a binary classification problem, we employed LR and more advanced machine learning (ML) techniques (i.e., AdaBoost, support vector machine, and random forest). We include all four approaches in the BDS. LR is useful for interpreting results and can be applied directly in order to derive recommendations for human decision-makers, while the ML techniques yield superior classification results. In order to train the models for the BDS, we used an additional data set. In the following, we first introduce the training data set and the variables that we used when applying the four classification approaches. Afterwards, we separately introduce the four approaches and describe the results of the trained models.

3.2.1 Training Data Set

For the training data collection, we also used the official Facebook Graph API. We collected metadata from the 76 selected public Facebook pages (cf. Section 3.1) in the period between the page's existence and the end of August 2016. In order to ensure enough time for buzz activity to develop, we cleaned the data by separately including only posts that were older than two weeks. Finally, our data set consisted of 119,910 posts. First, these posts were manually classified as buzzes and non-buzzes by four coders based on the definition of buzzes as the coding criteria (Deusser et al. 2018: p 1444):

"*A Buzz is a specific post, behaviour, or topic which initially spreads via social media and that suddenly draws surprising as well as extraordinary attention leading to many views. In this respect, a post is a text, a video, and/or a picture. It can lead to certain online reactions such as liking, sharing, commenting on, emulating, and/or offline reactions such as participating in events or purchasing a certain item. In most cases this exceptional attention lasts for a few days or weeks. In rare cases it lasts for a few months. At the beginning, in most cases there is no news agency involvement. However, later on single media outlets might report on the [buzz].*"

All four coders classified more than one fourth of the data in order to check their reliability. At least two coders classified the remaining posts. In order to verify the intercoder reliability, we relied on the tetrachoric correlation coefficient – a measure specifically designed for binary data (Carroll 1961; Divgi 1979). Table 2 summarises the results.

Table 2 Overview of the intercoder reliability measures between all pairs of coders

Coder 1 & 2	Coder 1 & 3	Coder 1 & 4
n = 32,730 r = .9320*	n = 89,218 r = .9172**	n = 32,730 r = .8786*
Coder 2 & 3	Coder 2 & 4	Coder 3 & 4
n = 45,697 r = .8664*	n = 32,730 r = .8885*	n = 50,483 r = .6638*
Note: n = number of posts both coders classified; r = tetrachoric correlation score; * p < .05; ** p < .01		

The coefficient's value of 0 (1) means no (perfect) agreement. Generally, researchers regard ratings with values above 0.7 as strong associations. In order to prevent any misclassification, the authors scrutinised the cases of disagreement and decided whether a post was a buzz or not. After this classification process, our data set consisted of 201 buzzes.

3.2.2 Operationalisation

The dependent variable, "buzz", is dichotomous: It is marked as 1 if it represents a buzz based on the definition (cf. Section 3.2.1) or 0 otherwise.

Table 3 Descriptions of the independent variables and their descriptive statistics

Variable Category	Variable Reference	Variable	Variable Description	Mean	SD	Min	Max
Lagging Variables	content-related	*firstLevel-ContentLength*	average length of first-level comments' (comments on post) content at time *t*	64.4 (46.35)	102.47 (91.78)	0 (0)	11812 (7517)
		secondLevel-ContentLength	average length of second-level comments' (comments on comments) content at time *t*	39.21 (31.99)	89.45 (64.34)	0 (0)	14632 (7487)
	user-related	*repeatUsers*	$\frac{\text{number of users having commented on a post more than once at time } t}{\text{number of comments post has received at time } t}$	.074 (.139)	.119 (.344)	0 (0)	1 (52)
	engagement-related	*commentLikes*	average number of likes the comments to a given post have received at time *t*	1.02 (.452)	1.44 (.929)	0 (0)	39.33 (160)
Deviation Variables	user-related	*newUsers*	$\frac{\text{number of users having commented on a page for the first time at time } t}{\text{number of all users having commented on the post at time } t}$	.332 (.006)	.289 (.042)	0 (0)	1 (1)
		pageOwner-Comments	$\frac{\text{number of page owner comments on a given post at time } t}{\text{average number of page owner comments on the page owner's page}}$	.128 (.627)	.692 (6.39)	0 (0)	75.93 (1451.25)
	engagement-related	*comments*	$\frac{\text{number of comments on a given post at time } t}{\text{average number of comments on posts on the given page}}$	.832 (.736)	5.81 (3.49)	0 (0)	1005.95 (675.62)
Note: Values in parentheses refer to the statistics of the whole data set from August 2016 until April 2018							

Given the available data, we derived the independent variables from theory. Meire, Ballings, and Van den Poel (2016) distinguished between a post's leading and lagging information. Leading information represents the data being available at a certain Facebook profile or page before the posting of a specific post, like the total number of posts on a page. Lagging information refers to the data being available after a specific post was published, such as the number of comments and likes the post received. As lagging information is only partly useful for early detection systems, we follow Meire, Ballings, and Van den Poel (2016) and calculate deviation variables, which are a combination of leading and lagging information.

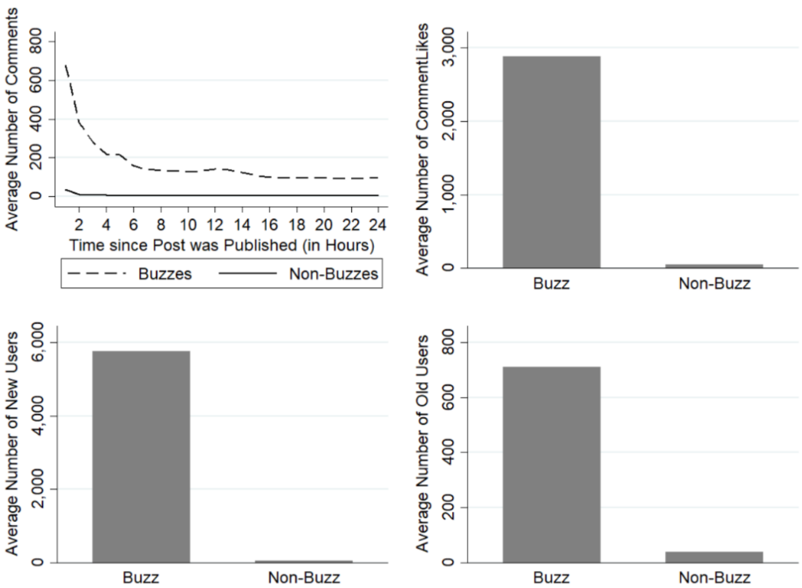
Table 3 summarises the descriptions of the variables, as well as the descriptive statistics of both the training dataset and the full dataset (encompassing all the data from August 18th, 2016 until April 2nd, 2018) used to evaluate the BDS.

Lagging variables. Two lagging variables refer to the content of the comments on a given post at time *t*. Following various researchers (e.g., Tsur and Rappoport 2012; Meire et al. 2016), we accounted for the average length of both the content of first-level comments (comments on the post), i.e., *firstLevelContentLength*, and the average length of the content of second-level comments (comments on comments), i.e., *secondLevelContentLength*. By including *repeatUsers* in our model, we considered a user-related variable, in line with prior researchers (e.g., Zubiaga et al., 2015; Deusser et al. 2018). This variable signifies the number of users who have commented on a given post or its comments more than once, divided by the number of total comments that the post has received at time *t*. We follow Meire, Ballings, and Van den Poel (2016) and Deusser et al. (2018) by addressing users' engagement on a post by including *commentsLikes* which signifies the average number of likes the comments to a given post have received at time *t*. According to Deusser et al. (2018), this is one variable that helps to detect the unexpectedness characteristic of buzzes. Given the example of lotteries: lotteries may receive a high number of comments as this represents the way of how users can participate in a lottery. As stated earlier, lotteries do not represent buzzes as it is not very surprising that many people participate and thus the post may become viral. Usually, lotteries' comments do not receive many likes. Thus, by including the *commentsLikes* variable, we are able to reflect a certain unexpectedness.

Deviation variables. Following Meire, Ballings, and Van den Poel (2016), we calculated deviations by combining the lagging information (drawn from a post on a certain Facebook page at a certain point of time *t*) with the leading information from the same page. One derivation variable reflects the number of users commenting on a given post (e.g., Zubiaga et al. 2015; Deusser et al. 2018). We took into account the variable *newUsers* by distinguishing between old and new users commenting on a given post. Old users are those who had commented on a given Facebook page before. New users are those who commented on a certain page for the first time. *NewUsers* stands for the percentage of new users commenting on a given post. A further user-related variable is *pageOwnerComments*, which captures whether page owners have commented more often on a given post at a time *t* than on average on their page. *PageOwnerComments* refers to the number of page owner's comments to a given post at time *t* divided by the average number of page owner's comments on their respective Facebook page (Deusser et al. 2018). According to Deusser et al. (2018), this is a further variable that helps to detect the unexpectedness characteristic of buzzes. Given the example of Facebook live sessions, the number of page owner comments is usually pretty low, as they answer the questions, users pose by writing comments, live in the video. As described earlier, live sessions do not fulfil all buzz characteristics. Thus, *pageOwnerComments* is a good indicator, since page owners usually comment on a given post themselves if something unforeseen or unexpected happens and is discussed in the comments. A further deviation variable is *comments* (e.g., Szabo and Huberman 2010; Feroz Khan and Vong 2014; Deusser et al. 2018), which is the number of comments a post received at time *t* divided by the average number of comments posts received on that page. This variable indicates if a post has received more comments at a time *t* than the average for that specific page.

Based on the training data set, Figure 2 offers a first insight into the differences between buzzes and non-buzzes according to (1) their temporal courses in terms of the average number of comments they receive, (2) the number of likes their comments receive on average, (3) the average number of previously inactive users commenting on the page for the first time (new users), and (4) the average number of users who have already been active by commenting on the page before (old users).

Figure 2 Differences between buzzes and non-buzzes



3.2.3 Logistic Regression (LR)

The LR estimates the likelihood of occurrence of the outcome variable depending on different variables. The following equation describes this regression approach:

$$z = \alpha + \beta_1 * firstLevelContentLength + \beta_2 * secondLevelContentLength + \beta_3 * repeatUsers + \beta_4 * commentsLikes + \beta_5 * likesShares + \beta_6 * newUsers + \beta_7 * pageOwnerComments + \beta_8 * comments$$

where α is the Y intercept and the βs are the regression coefficients. α and β are estimated by the maximum likelihood method.

Modelling the likelihood of occurrence is not based on a linear regression approach, but on a logistic function which is as follows:

$$\pi(\mathrm{Y}) = \frac{e^z}{1 + e^z}$$

where π denotes the probability of the outcome variable Y (which represents a buzz) and e represents Euler's number.

We first checked the correlation among the independent variables and found that none of the correlations exceeded critical values (see Table A1 in the Appendix for details). In the following, we report the results of our LR models.

Table 4 Summary of the results of the estimated LR models predicting buzzes

	Model 1	Model 2	Model 3	Model 4	Model 5
firstLevelContentLength	1.0004** (.00007)	1.0003** (.00007)	1.0003** (.00008)	1.0003* (.0002)	1.0005** (.0001)
secondLevelContentLength	1.0004** (.0002)	1.0004** (.0001)	1.0003** (.0001)	1.0006** (.0002)	1.0005** (.0001)
repeatUsers		9.54** (1.58)	9.76** (1.86)	6.05* (.924)	4.77** (1.36)
commentLikes			1.22** (.024)	1.25** (.036)	1.24** (.032)
newUsers				465.57** (160.73)	181.82** (77.19)
pageOwnerComments				1.13** (.027)	1.09** (.019)
comments					1.09** (.014)
Constant	.002** (.0001)	.001** (.0001)	.001** (.0001)	.00002** (.00001)	.00003** (.00001)
Pseudo R^2	.004	.014	.035	.2237	.4487
Wald χ^2	39.18	202.48	223.32	483.51	340.48
BIC	2363.27	2351.84	2312.52	1894.82	1380.48
N	94,927	94,927	94,927	94,927	94,927
Note: Odds ratio and robust standard errors in parentheses; * p < .05, ** *p* < .01					

Table 4 covers all the models we estimated in order to explain the dependent variable *buzz*. Please note that we only included those variables in our models that yielded highly significant results ($p < .01$), as Lin, Lucas, and Shmueli (2013) suggested for large samples. In this way, we ensured the best performances for our BDS.

Model 1 contains the content-related lagging variables *firstLevelContentLength* and *secondLevelContentLength*, which both showed a significant positive influence ($p < .01$) on the outcome variable *buzz*. Thus, the higher the average length of the content in both first- and second-level comments, the more likely the considered post is a buzz.

In Model 2, we added the user-related lagging variable *repeatUsers*, which had a significant positive influence ($p < .01$) on *buzz* as well. This finding indicates that, at time *t*, a post is more likely to be a buzz if more unique users have commented on a given post more than once. Model 3 added the engagement-related lagging variable *commentLikes*. It showed a significant positive influence ($p < .01$) on *buzz*, meaning that the more likes the comments of a given post have received on average at time *t*, the higher the post's likelihood of being a buzz. In Model 4, we added the user-related deviation variables. Both *newUsers* and *pageOwnerComments* have a high positive influence ($p < .01$) on the outcome variable *buzz*. Thus, the more users that comment on a post on a specific page for the first time, as well as the more that page owners commented on posts on their own page, the more likely it is that the given post is a buzz. Model 5, our final model, added the engagement-related deviation variable *comments*, which has a high positive influence ($p < .01$) on *buzz*. The more comments a post receives, the more likely it is that this post displays buzz characteristics.

Our final model yielded a Pseudo R^2 of 44.87%. We obtained the largest growth in R^2 when adding the deviation variables. In order to assess the robustness of our model, we applied the following checks: First of all, building our model step by step, we could already see that our results were stable and that our variables were consistently valid. Second, as we are dealing with rare case events yielding from the imbalance between buzzes and non-buzzes, we recognise that the standard LR applied may be insufficient. Thus, we repeated the model estimation using the "relogit" function, as it can explicitly deal with such rare cases (King and Zeng 2001). Our obtained results were substantially the same, which further indicates our model's robustness. When controlling for both the day of the week and the month of the year, we effectively found the same results. The lack of significance among the control variables further supports the stability of our model.

3.2.4 Machine Learning Techniques

We applied AdaBoost, support vector machine, and random forest in the BDS, as they provided the most robust results (cf. Deusser et al. 2018) and are among the best classifiers (Fernández-Delgado et al. 2014). We optimised the hyper-parameters of all our approaches using traditional grid search – a process of determining parameters through exhaustive search in a predetermined interval of possible hyper-parameter values. We assigned a bias towards recall during this grid search, as avoiding false negatives is of higher priority than avoiding false positives due to the rarity of buzzes. In the following, we present each technique and describe its implementation relative to the BDS.

AdaBoost (AB). AB is a ML technique that incorporates a weighting mechanism that focuses on misclassified observations from the previous iterations (Freund and Shapire 1996). It sequentially re-weighs the training data by giving more weight to the incorrectly classified observations and less weight to the correctly classified observations. Thus, AB focuses on the observations that are hard to classify, whereby the final model represents a linear combination of all the previous models (Hastie et al. 2005). We implemented AB by using the “sklearn.ensemble.AdaBoostClassifier” package in Python. There are two parameters to set: (1) the learning rate and (2) the number of estimators. We found 0.8 to be the optimal learning rate and 60 estimators to be the optimal number of estimators.

Support Vector Machine (SVM). Vapnik (1995, 1998) introduced SVM, which is based on statistical learning theory. As in our case of binary classification, SVMs aim to find a linear optimal hyperplane that maximises the margin of separation between positive and negative observations. Similarly, quadratic optimisation problems only concentrate on support vectors that represent the data points closest to the optimal hyperplane. But the data must be linearly separable, which is not often the case in practice. Thus, a kernel function is applied in order to transform the input space into a higher dimensional feature space and allow for a linear separation. We implemented SVM by using the “sklearn.svm.SVC” package in Python and found that the polynomial kernel is optimal in our case.

Random Forests (RF). According to Fernández-Delgado et al. (2014), RF is the best multi-purpose classification technique. RF grows an ensemble of trees by aggregating all tree predictions based on majority voting (Breiman 2001). Thus, it overcomes limited robustness due to its high instability, coupled with the suboptimal performance of decision trees (Dudoit, Fridlyand, and Speed 2002; Hastie et al. 2005). Furthermore, RF does not overfit (Breiman 2001), provides variable importance (Sandri and Zuccolotto 2006), and only needs two parameters to be set (Larivière and Van den Poel 2005; Duda, Hart, and

Stork 2012): (1) the number of trees to be grown and (2) the number of randomly selected predictors at each node of each tree. We implemented RF by using the “sklearn.ensemble.RandomForestsClassifier” package in Python. Following Breiman (2001), we set the number of trees to 60 and the number of predictors as the square root of the total number of variables, i.e., three because the model includes seven explanatory variables.

3.2.5 Evaluation of Models on Test Set

We applied the final models to an additional, independent test set consisting of almost 24,000 posts and 43 buzzes. Table 5 and Table 6 summarise the results.

Table 5 Evaluation of the final LR model on the test set

		π = 0.01		π = 0.05		π = 0.2	
		Actual		Actual		Actual	
		Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz
Prediction	Buzz	27	215	20	28	11	10
	No Buzz	16	23,725	23	23,912	32	23,930
Accuracy		99.04 %		99.79 %		99.82 %	
Precision		11.16 %		41.67 %		52.38 %	
Recall		62.79 %		46.51 %		25.58 %	
F1		18.95 %		43.96 %		34.37 %	

Table 6 Evaluation of the final machine learning technique models on the test set

		AdaBoost		SVM		Random Forests	
		Actual		Actual		Actual	
		Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz
Prediction	Buzz	20	13	9	18	15	5
	No Buzz	23	23,927	34	23,922	28	23,935
Accuracy		99.85 %		99.78 %		99.86 %	
Precision		60.60 %		33.33 %		75.00 %	
Recall		46.51 %		20.93 %		34.88 %	
F1		52.63 %		25.71 %		47.62 %	

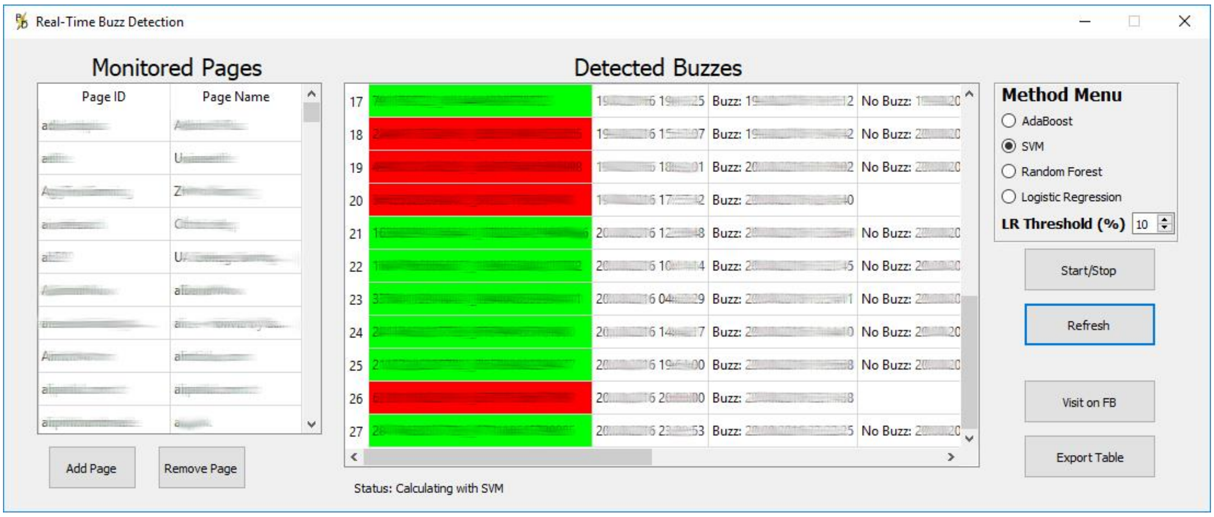
Following Deusser et al. (2018), we evaluated the LR model with different thresholds, namely π = 0.01, π = 0.05, and π = 0.2. On the one hand, a rather small π of 0.01 yields a higher recall (63%) at a lower precision (11%). A large π-value (0.2) yields a precision of over 50% and a recall of under 50%. A threshold of an intermediate value, such as 0.05, leads to more balanced results. Consequently, we can state that the threshold value π depicts the trade-off between identifying buzzes correctly, on the one hand, and a high number of false positive classifications, on the other hand. Regarding the ML techniques, we find that SVM yields a precision of 33.33% and a recall of 20.93%. AB and RF performed better on the test data set: The number of misclassified posts is relatively low for both AB (precision of 60.60% and recall of 46.51%) and RF (precision of 75.00% and recall of 34.88%). Considering the higher importance of avoiding false negatives due to the rarity of buzzes, “LR 1%”, “LR 5%”, AB, and RF all performed fairly well on the test data set.

3.3 Frontend

The frontend consists of a *pyQt 5.0 UI* and contains a page table, a buzz table, a method menu, and various buttons, as depicted in Figure 3.

Due to privacy reasons, we blurred the names of the Facebook pages and their respective post IDs. The page table on the left lists all the pages that are currently monitored by the BDS. The buttons “Add Page” and “Remove Page” allow one to add further Facebook pages or to remove selected pages from the list. The buzz table in the middle lists all detected buzzes and keeps track of posts that were once detected as buzzes, but stopped fulfilling the buzz criteria as time went on. Therefore, it lists the post’s identification number, its creation time, the timestamp when it was first discovered as a buzz and the corresponding probability when the LR is chosen from the method menu. In the following columns, the buzz table respectively displays the timestamp for when a post is no longer detected as a buzz and the timestamp if it fulfils the buzz criteria again. Additionally, the rows are colour-coded: red refers to a buzz and green to non-buzz. Especially for the LR, a green-red scale shows a post’s buzz probability. The status line informs users about the current state of the BDS – whether it is “calculating” and with which specific method, or whether it is “idle” and waiting for users’ interactions.

Figure 3 Screenshot of the functional buzz detection system



In the method menu on the right, users can choose from different methods (i.e., AB, SVM, RF, or LR) with various threshold options. In Section 4, we elaborate on these different methods. The “Start/Stop” button allows the user to start and stop the BDS. The “Refresh” button allows the BDS to continue buzz detection after users have chosen a different method. The “Visit on FB” button enables users to directly visit selected Facebook posts from the buzz table. The “Export Table” button allows users to export the current state of the buzz table as a CSV file in order to conduct further analyses and calculations.

4. Evaluation of the Buzz Detection System

In order to evaluate the BDS, we used two different evaluation periods and compared the results to the ones of an existing model that predicts viral online content. The first evaluation period comprises two months from 1st August, 2017 until 30th September, 2017, while the second one covers three months from 1st November, 2017 until 31st January, 2018. Across both evaluation periods, 213,228 posts were published with a total of almost 23.5 million comments from 873 different Facebook pages. We purposefully chose two different evaluation periods in order to evaluate the BDS’ performance in a period outside of a festive season and in a period encompassing several festive days, such as Christmas and New Year. In this way, we could see how the BDS reacted to all the Christmas and New Year’s greetings on these days. Such posts usually satisfy two of the three buzz characteristics – immediacy and intensity – but do not fulfil the characteristic of unexpectedness. Therefore, we performed a more extensive evaluation in this second period, especially with regard to a more exhaustive manual classification, as well as a more fine-grained evaluation of the different classification techniques and their performance.

Ideally, at least two coders would have classified all 213,228 posts in order to tell whether these posts were buzzes or non-buzzes. Potentially, every post published before the evaluation period could be – and some actually were – detected as buzzes within the evaluation period. Thus, we would have had to manually classify all the previous posts from the monitored pages as buzzes and non-buzzes in order to correctly perform our evaluation – an undertaking that is not practical in terms of time, labour and cost. Therefore, in order to evaluate the BDS’ performance, we limited the number of posts to classify, which we will describe in the following two sections. These sections also include all relevant details for each classification period.

4.1 Evaluation Period 1

In the first evaluation phase, 57,612 posts were published with a total of almost 10 million comments on 813 different Facebook pages that we randomly collected for monitoring (cf. Section 3.1 on random data collection process). As described previously, we limited the number of posts to manually classify. In this first period, we classified all the posts that were detected as buzzes by any classification technique (n = 510). By manually classifying these 510 posts, we confirmed 120 posts to be buzzes. In Table 7, we summarise the confusion matrices of each classification technique. Within both evaluation periods, we chose 0.01, 0.05, 0.1, and 0.2 as the thresholds for the LRs. As can be seen, the higher the threshold, the better the accuracy. However, as buzzes represent rare cases, we are more interested in avoiding false negatives than avoiding false positives. Thus, we concentrated on the recall in order to evaluate the performance of the different classification techniques. We obtained the best results for “LR 1%”, “LR 5%”, and SVM – 59.17%, 42.50%, and 40.00%, respectively – when referring to the posts that were correctly classified as buzzes at the end of the evaluation period and that were once correctly classified as buzzes during the evaluation period. We also see that a higher threshold for the logistic regressions accompanies lower recall and higher precision. Regarding RF, we find that the classification technique performed poorly: Only one buzz was detected within the evaluation period. An accuracy of almost 80% might seem quite high, but there is actually no recall, which renders this classification technique inferior in the BDS. Similarly, AB only classified one buzz correctly. Still, 35 posts were correctly identified as buzzes during the evaluation period, but classified as non-buzzes by the end of the period, which leads to a recall of 30% of detected buzzes at a specific point of time within the evaluation period. We discuss the potential reasons for this observation in Section 5. This evaluation is already a useful assessment of the system’s performance and the results – especially based on a logistic regression – look promising. However, this evaluation does not allow us to test how the system would react to truly expected events that go viral. Moreover, we would like to extend the number of human classifiers to ensure that the results are not driven by one subjective decision. We will address these shortcomings with the second evaluation.

4.2 Evaluation Period 2

In the second evaluation phase, 155,616 posts were published with a total of almost 13.5 million comments on 821 different Facebook pages that we randomly collected for monitoring (cf. Section 3.1 on manual data collection process). As described previously, we limited the number of posts to manually classify. In order to evaluate the BDS’ performance in this second evaluation period, we limited the number of posts by the

Table 7 Confusion matrices of each classification technique for the 510 posts detected within the additional evaluation period

		LR π = 0.01		LR π = 0.05		LR π = 0.1		LR π = 0.2		SVM		AdaBoost		Random Forests	
		Actual		Actual		Actual		Actual		Actual		Actual		Actual	
		Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz
Prediction	Buzz	33	67	20	38	14	33	12	27	27	123	1	0	1	0
	No Buzz	38* 49**	124* 199**	31* 69**	64* 288**	30* 76**	57* 300**	24* 84**	50* 313**	21* 72**	103* 164**	35* 84**	33* 357**	0* 119**	0* 389**
Accuracy		69.80 %		72.94 %		72.75 %		73.53 %		57.65 %		76.67 %		76.47 %	
Precision		33.00 % (51.45 %)		34.48 % (57.30 %)		29.79 % (57.14 %)		30.77 % (57.14 %)		18.00 % (28.07 %)		100 % (100 %)		100 % (100 %)	
Recall		27.50 % (59.17 %)		16.67 % (42.50 %)		11.67 % (36.67 %)		10.00 % (30.00 %)		22.50 % (40.00 %)		0.83 % (30.00 %)		0.83 % (0.83 %)	
F1		32.70 % (55.04 %)		22.47 % (48.80 %)		16.77 % (44.67 %)		15.09 % (39.34 %)		20.00 % (32.99 %)		1.65 % (46.15 %)		1.65 % (1.65 %)	

* = post classified as buzz at some point of time within the evaluation period, but classified as non-buzz by the end of the evaluation period;
** = post was never detected as a buzz by the classification technique and is therefore classified as a non-buzz
in parentheses () recall, precision, and F1 calculated by including the posts that were once correctly classified as buzzes (*)

following: we classified (1) all the posts that were detected as buzzes by any classification technique (n = 468); (2) 100 randomly selected posts (20 posts per SVM, LR threshold = 1%. LR threshold = 5%, LR threshold = 10%, LR threshold = 20% because these were the methods with a high number of non-buzzes) that were classified as buzzes at some time during the evaluation period, but were classified as non-buzzes by the end of it; (3) 100 randomly selected posts that were entirely undetected by either of the classification techniques. In this evaluation, three coders classified a total of 668 posts using the buzz definition as the coding criteria (cf. Section 3.2.1). Table 8 summarises the intercoder reliability measures. This approach ensures a certain level of objectivity.

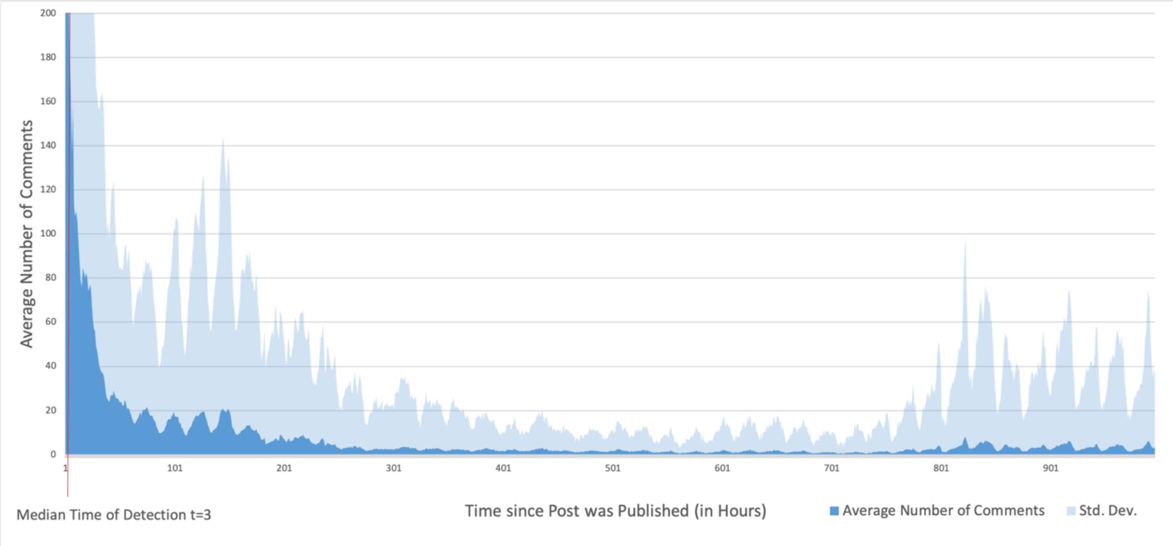
Table 8 Overview of the intercoder reliability measures between all pairs of coders

Coder 1 & 2	Coder 1 & 3	Coder 2 & 3
n = 668 r = .9308*	n = 668 r = .9475*	n = 668 r = .8573*
Note: n = number of posts both coders classified; r = tetrachoric correlation score; * p < .01		

Among the 668 classified posts, the coders identified 145 buzzes. Thereby, 102 (= 70.34%) were actually detected as buzzes by either of the classification techniques, while 39 buzzes (= 26.70%) were correctly detected as buzzes by either of the classifications at some point of time during the evaluation period, but classified as non-buzzes by the end of the evaluation. Four buzzes (= 2.76%) were not detected at all by the classification techniques. In total, 141 out of 145 buzzes were listed as buzzes in the “Buzz Table”, indicating that the BDS would have shown users 97.24% of the buzzes by combining all the classification techniques. During this second evaluation period, Christmas- or New-Year-related posts were not detected by either of the classification techniques, meaning that the BDS successfully passed this quality test.

Figure 4 shows the average number of comments given to the detected buzzes per hour since the buzz post was published. The figure also depicts the standard deviations of all the buzzes’ comments within each hour. The vertical line at t = 3 hours represents the median time to detect a buzz. After this time the average of about 27.7% of all comments had been posted. Thus, in theory, about 72.3% comments could be moderated because the system has issued a warning.

Figure 4 Average number of comments given to the detected buzzes and their standard deviations per hour over time



In Table 9, we present the overlapping and non-overlapping classification results from the different techniques. The “LR 1%” correctly classified nine buzzes that had not been detected as buzzes by any other technique. The SVM classified 34 buzzes correctly, while one buzz was correctly classified, and then de-classified, by the end of the evaluation period. These 35 posts had not been detected as buzzes by any other classification technique. All other posts were detected by at least two classification techniques.

Table 10 summarises the confusion matrices of each classification technique and supports the findings from the first evaluation period. As in the first evaluation period, we chose 0.01, 0.05, 0.1, and 0.2 as the thresholds for the LRs. As can be seen, a higher threshold accords with better accuracy. When focusing on the recall, we obtained the best results for “LR 1%” and SVM – 42.07 % and 38.62 %, respectively – when referring to posts that were still classified as buzzes at the end of the evaluation period. A higher LR threshold is accompanied by higher precision and lower recall. Regarding RF, and similarly to the first evaluation period, we found that this classification technique performed poorly: It detected no buzzes within this second evaluation period. Only one post was correctly classified as a buzz during the evaluation period, but it was classified as a non-buzz by the end of the evaluation period. An accuracy of almost 80% might seem quite high, but the lack of any recall renders this classification technique inferior in the BDS. AB also performed poorly, as it only correctly classified two buzz posts. Six posts were correctly identified as buzzes during the evaluation period, but classified as non-buzzes by the end of the period. Like RF, AB RF, AB featured high accuracy, but very low recall. We discuss the potential

Table 9 Overview on overlapping and non-overlapping classification results of the 141 posts by the different classification techniques

		Others		LR 5%		LR 5% & LR 10%		LR5% – 20%		LR5% – 20% & SVM		LR5% – 20% & AdaBoost		
		1	not classified (.)	1	0	1	0	1	0	1	0	1	0	
LR 1%	1	0	9	4	1	1	0	18	0	6	1	1	0	
	0	0	0	0	2	0	4	0	13	0	19	0	0	
		9		*7*		*5*		*31*		*26*		*1*		*79*

		Others		LR 1%		LR 1% & LR 5%		LR1% – 10%		LR1% – 20%		AdaBoost		
		1	not classified (.)	1	0	1	0	1	0	1	0	1	0	
SVM	1	0	34	2	1	1	1	3	1	(6)	2	1	1	
	0	0	1	2	0	0	0	0	0	1	(19)	0	0	
		35		*5*		*2*		*4*		*3*		*2*		*51*

		LR 1% & LR 5%	
		1	0
LR 10%	1	(1)	0
	0	5	(4)
		5	

		LR1% – 10%	
		1	0
LR 20 %	1	(18)	0
	0	1	(13)
		1	

		LR1% – 20%		LR1% – 20% & SVM	
		1	0	1	0
AdaBoost	1	(1)	0	0	0
	0	2	0	2	0
		2		*2*	

		LR1% – 20% & SVM		
		1	0	
AdaBoost & Rand.F.	1	0	0	
	0	1	0	
		1		*11*

Table 10 Confusion matrices of each classification technique for the 668 considered posts within the evaluation period

		LR π = 0.01		LR π = 0.05		LR π = 0.1		LR π = 0.2		SVM		AdaBoost		Random Forests	
		Actual		Actual		Actual		Actual		Actual		Actual		Actual	
		Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz	Buzz	No Buzz
Prediction	Buzz	61	213	46	108	36	89	31	73	56	184	2	0	0	0
	No Buzz	43* 41**	74* 236**	44* 55**	91* 324**	45* 64**	83* 351**	36* 78**	68* 382**	24* 65**	80* 259**	6* 137**	0* 523**	1* 144**	0*523**
Accuracy		55.54 %		69.01 %		70.36 %		72.01 %		59.13 %		78.59 %		78.29 %	
Precision		22.26 % (32.81 %)		29.87 % (45.45 %)		28.80 % (47.65 %)		29.81 % (47.86 %)		23.33 % (30.30 %)		100 % (100 %)		–	
Recall		42.07 % (71.72 %)		31.72 % (62.07 %)		24.83 % (55.86 %)		21.38 % (46.21 %)		38.62 % (55.17 %)		1.38 % (5.52 %)		0 %	
F1		29.11 % (45.02 %)		30.77 % (52.48 %)		26.67 % (51.43 %)		24.90 % (47.02 %)		29.09 % (39.12 %)		2.72 % (10.46 %)		–	

* = post classified as buzz at some point of time within the evaluation period, but classified as non-buzz by the end of the evaluation period;
** = post was never detected as a buzz by the classification technique and is therefore classified as a non-buzz
in parentheses () recall, precision, and F1 calculated by including the posts that were once correctly classified as buzzes (*)

reasons for this observation in Section 5.

As this second evaluation period was one month longer than the first one, and encompassed a major festive season, we performed a more extensive analysis on the time each classification took. In order to evaluate the different classification techniques' speed in detecting a buzz, we would have had to monitor each post manually in order to note the moment the post really became a buzz. In the majority of the cases, the classification techniques detected posts as buzzes within a few seconds, minutes, or hours. However, published posts sometimes receive little attention until a few days, weeks, or even months later. Of course, manually monitoring each single post in order to evaluate the speed of each classification technique is not an efficient practice. Thus, we decided to report on the elapsed time between a post's publication and each technique's timestamp for detecting the buzz. First, we report the amount of time each classification technique took on average to identify the true positives. Table 11 summarises these results.

Table 11 Average time each classification technique took to classify a post as a buzz

	All Buzzes	Buzzes classified in < 100 days	Buzzes classified in < 14 days	Buzzes classified in < 1 day
LR 1%	n = 61 48 days, 08:30:07	n = 50 5 days, 23:53:49	n = 45 2 days, 06:52:56	n = 22 06:18:52
LR 5%	n = 46 54 days, 05:26:32	n = 36 7 days, 00:06:44	n = 31 2 days, 03:09:51	n = 16 07:40:32
LR 10%	n = 36 67 days, 08:33:09	n = 27 6 days, 14:49:08	n = 23 1 day, 19:10:22	n = 14 08:33:43
LR 20%	n = 31 55 days, 21:52:53	n = 25 7 days, 04:30:38	n = 21 2 days, 16:35:31	n = 11 07:31:44
AdaBoost	n = 2 2 days, 05:07:48	n = 2 2 days, 05:07:48	n = 2 2 days, 05:07:48	n = 1 02:17:28
SVM	n = 56 30 days, 17:53:42	n = 50 3 days, 19:53:27	n = 45 1 day, 17:42:35	n = 30 06:31:03

In order to account for outliers (the buzz posts that took an exceptional amount of time for our classification techniques to detect), we report on not only the average times for all classified buzzes, but also buzzes that were detected in less than 100 days, in less than 14 days, and in less than one day. As can be seen, on average the SVM is generally the fastest at detecting buzzes. With an increasing threshold, the LRs naturally take longer to identify a post as a buzz. If AB detects a buzz, it usually does so quite quickly. As RF failed to detect a single buzz within the evaluation period, we cannot report any results for this classification technique.

Next, we focused on the posts that at least two techniques detected as buzzes and compared the speed of the classification techniques. Thereby, we also considered posts that were classified as non-buzzes by the end of the evaluation period, although they had been identified as buzzes previously. For this reason, we utilised the timestamp of a post's first detection as a buzz. Table 12 summarises the results. Here, we also removed outliers: posts for which either SVM or the LRs with their various thresholds took an exceptionally long period of time (on average over 5,000 hours) to discover the post as a buzz compared to the respective other classification techniques. Besides, we performed additional analyses on these outliers and found that the majority of these posts stemmed from very large Facebook pages with a number of page likes above the 95% percentile of all the pages we monitored. Additionally, we examined those posts where at least one method detected the post as a buzz at a certain point of time during the evaluation period, but classified it as a non-buzz by the end of the evaluation period. We found that, in most cases, the classification techniques swung between classifying a post as a buzz and a non-buzz within seconds – namely, whenever a new comment was added to the post.

Table 12 Average times for LR 1%, AdaBoost (AB), SVM are faster than the respective other techniques

faster than	LR 1%	LR 5%	LR 10%	LR20%	AB	RF	SVM
LR 1%	–	n = 50 6 days, 17:03:01	n = 45 11 days, 02:19:45	n = 37 12 days, 09:13:40	n = 2 0 days, 01:08:53	–	n = 16 23 days, 10:17:54
LR 1% w/o outliers	–	n = 49 6 days, 20:14:51	n = 44 11 days, 08:11:02	n = 36 12 days, 17:00:20	n = 2 0 days, 01:08:53	–	n = 15 2 days, 09:19:55
AB	n = 2 1 day, 10:08:21	n = 2 3 days, 05:12:45	n = 2 4 days, 22:54:40	n = 2 9 days, 05:16:14	–	n = 1 205 days, 08:58:08	n = 1 3 days, 20:42:52
SVM	n = 13 57 days, 17:11:29	n = 6 56 days, 02:50:44	n = 5 67 days, 17:00:13	n = 4 16 days, 06:08:51	n = 3 2 days, 01:19:02	–	–
SVM w/o outliers	n = 9 1 day, 08:11:00	n = 4 0 days, 07:44:49	n = 3 1 day, 00:43:02	n = 3 3 days, 11:50:23	n = 3 2 days, 01:14:02	–	–

4.3 Model Comparison

In order to further evaluate the buzz detection system's performance, we compare our proposed model to a model that comes closest to ours predicting the virality of news articles based on online sharing behaviour. Heimbach and Hinz (2018) postulated this model building upon the findings by Berger and Milkman (2012). We used Model 2 in Table 7 (Heimbach and Hinz 2018) to re-run our buzz detection system on the data sets of the two discussed evaluation periods. In evaluation period 1, the considered model detected 25,537 posts as buzzes and in evaluation period 2, there were 63,291 posts detected. Due to two reasons we decided to take the top 1% of the detected buzzes in these two different periods and manually classified 887 posts: (1) the high number of posts and its impracticality in terms of time, labour and cost to manually classify these posts, and (2) the fact that our thorough manual classification of the training data set of almost 120,000 posts proved that buzzes rarely happen and the number of detected buzzes by the new model was exceptionally high. In the first evaluation period, seven out of 255 detected posts were actually classified as buzzes (i.e., 2.75%). In the second evaluation period, 28 out of 604 detected posts were buzzes (i.e., 4.64%). These results show that the model chosen for this comparison is definitely working better than chance. However, the high number of false positives is due to the fact that many posts were detected as buzzes that did not fulfil the buzz characteristic of "intensity" as there were no or hardly any reactions to the posts in forms of likes, shares, and comments. It was also obvious that in the second evaluation period the model detected many posts as buzzes that represented Christmas wishes so that the above average reactions to these posts were not very surprising and the model did not pass the test as described earlier in this section. These findings underline that more sophisticated models are necessary in order to detect the unexpectedness characteristic of buzzes as outlined in Section 3.2.2 and as incorporated in our model for buzz detection.

5. Discussion

In the following, we discuss our theoretical contributions and the managerial implications.

Contribution to Theory. The BDS is the first detection system for buzzes – one type of viral phenomena that have not been analysed extensively yet. While most researchers have focused on trending topics (e.g., Mathioudakis and Koudas 2010; Cvijijk and Michahelles 2011; Lau, Collier, and Baldwin 2012), we recognised the fundamental economic and societal importance of buzzes and the need for their early detection in the mere flood of information in OSNs. We believe that incorporating the unexpectedness characteristic of buzzes makes this study's whole research aim unique and crucial. Without accounting for the unexpectedness of posts, we would only discover postings as buzzes that can easily be anticipated – for example, lotteries, Facebook live sessions, and well wishes on holidays and birthdays. Such episodes are not necessarily of interest to companies, public figures, institutions, or political parties. Thus, our central contribution is actually detecting postings that fulfil all three buzz characteristics: immediacy, unexpectedness, and intensity.

By leveraging insights from previous studies about identifying buzzes early in their evolution, we developed the first system for detecting buzzes. Evaluating the whole system over a 5-

month evaluation period, we found that the theoretically derived variables for buzz detection, as well as the estimated and trained models that housed them, perform exceptionally well in practice. Applying different classification techniques, we found that by combining the results of SVM and “LR 1%”, it is possible to detect almost 100% of buzzes. Based on our results, we can contribute the following insights for any future research: SVM singularly and correctly detected over 24% of buzzes, even those had not been detected by all other classification techniques. The posts in question were mostly characterised by a high engagement rate, i.e., they obtained most of the likes, shares, or comments on the respective page. Yet, this does not always hold true in the case of very large Facebook pages with more than 14 million page likes. Moreover, SVM generally detected buzzes faster than the respective other classification techniques. “LR 1%” is the second-best classification technique in terms of detecting buzzes that other techniques missed (6.21%) and in terms of speed. As with posts detected only by SVM, these were posts with a high engagement rate except for those stemming from Facebook pages with more than 10 million page likes. Unexpectedly, and in contrast to Deusser et al. (2018), AB and RF performed poorly. This may be due to the fact that these techniques usually operate with many more variables. We restricted the variables to ones that performed very well within the LR model and to those that were highly significant in our large sample, as suggested by Lin, Lucas, and Shmueli (2013). However, when we ran RF on the whole data set in order to examine its performance a bit closer, the BDS detected 23 posts as buzzes. The BDS classified 14 of the 23 as buzzes and the remaining nine as non-buzzes. Having three coders manually classify the 14 posts confirmed that these posts were correctly classified as buzzes. Eight of the nine other posts were also manually classified as buzzes. Thus, at a particular moment while classifying these posts, RF had already correctly classified the posts as buzzes. In one case, the post was correctly classified as a non-buzz in the end. Therefore, we can conclude that if RF detects a buzz, then it is very accurate. Regarding the speed, it took RF the following times to detect the buzzes: for all buzzes (n = 14): 20 days, 01:31:01, for all buzzes detected in less than 100 days (n = 13): 7 days, 02:40:08, for all buzzes detected in less than 14 days (n = 12): 4 days, 11:07:53, and for all buzzes detected in less than 1 day (n = 1): 00:59:53. As with the other classification techniques, there were solitary buzzes detected after a long delay. Other than that, the speed is comparable to the other classification techniques.

Regarding the detection speed in general, we discovered that not every post goes viral and becomes a buzz directly after its publication. In some cases, posts lie dormant until a sudden interest transforms them into buzzes. Thus, the detection speed has to be interpreted very carefully because a rather long detection time does not necessarily mean something negative. It could merely be a result of an unknown catalyst. However, it is almost impossible to identify this trigger and ascertain the speed with which the classification technique correctly detects a buzz. Hence, we resorted to using the publication date of a post, even though the act of publishing is not necessarily the trigger for a buzz. However, this might be overruled if several other classification techniques detect a buzz much earlier than one single technique struggling to detect it at all.

We also discovered that for a few posts, long detection times could be a result of larger Facebook pages. With pages that have over 1.6 million likes, the high activity could be making it more difficult for the BDS to detect buzzes. In terms of the comments, the posts themselves were also active for a relatively long time and mostly detected as buzzes when the last comments were written on the post. We have to analyse this interrelation in future research.

Regarding the quick oscillation between the detection of a buzz and classifying it as a non-buzz, we analysed the respective posts. Although it is difficult to explain the decisions of the classification techniques, we expect that writing a new comment to a post may have exceeded some threshold for classifying this post as a buzz – even though at the timestamp of a comment in the following second, the post could already be classified as a non-buzz again. Taking into account all these aspects, this study valuably contributes to and extends the existing literature on viral online content, both in general and in terms of their detection systems.

Managerial Implications. Practitioners, such as companies, public figures, institutions, or political parties, can use the BDS for various reasons – the most general one being the detection of buzzes amidst the flood of information in OSNs. Based on our results, the most relevant indicators for buzzes are user-related variables (*newUsers* and *repeatUsers*) as well as reactions to comments on the post (*commentLikes*). Regarding the number of users involved in buzzes, it is important to notice that buzzes have quite a large coverage. If posts attract new users, they are more likely to become a buzz, which may then create a cycle where the buzz attracts even more new users. This may be amplified further by news agencies that attract additional users who would commonly be passive participants (Reynolds 2002; Roberts, Wanta, and Dzwo 2002). Consequently, a positive feedback loop with new users can lead to buzzes. A further relevant indicator for buzzes is a high number of users that repeatedly interact with each other through commenting or replying to comments, thereby leading to vivid discussions. According to Holt (2004), most participation in discussions arises when people debate different viewpoints. This is in line with the buzz type “hot topics” where people debate controversial topics (Jansen 2019) and it explains the high number of users who repeatedly comment on the same post or reply to the same post comments. Furthermore, if users initiate lots of reactions within the discussions in the form of likes to the post’s comments, the post is more likely to become a buzz. This underlines the fact, in line with previous research (Lesot et al. 2012; Deusser et al. 2018), that buzzes are characterised by intense activity.

Practitioners can also make use of different existing buzz types (Jansen 2019) in order to develop a social media strategy that is appropriate to each specific type: (1) Some buzzes may carry joyful and positive messages that enjoy great popularity. Reacting to such buzzes – also known as lovestorms (Ács and Pagh 2017) – or incorporating them into social media or marketing strategies can be of considerable benefit for companies, public figures, and political parties. For example, *Delta Airlines* included several lovestorms in their on-board security video (YouTube 2015). (2) Buzzes can also carry harmful and negative messages that can cause massive waves of outrage. These buzzes – also known as firestorms (Pfeffer, Zorbach, and Carley 2014) – could encompass companies' newly released products, the actions of public figures, or statements from political parties. A timely response to the firestorm by the concerned person or company may prevent any further reputation or image loss. For example, the food company *Barilla* became the target of such a firestorm in 2013 when *Barilla's* chairman offended homosexuals in a radio interview. *Barilla* reacted to this public outrage by posting two videos on Facebook that were meant as an apology (Facebook 2013a; Facebook 2013b). (3) Other buzzes may represent so-called hot topics that include discussions, e.g. about refugees, obesity, or gender differences. Some parties may want their names to appear in the ongoing discussion and will act as thought-leaders by commenting on hot topics to assert their positions and persuade others to their side (Rooderkerk and Pauwels 2016). For example, a German consumer organisation involved in investigating and comparing goods and services triggered a hot topic discussion by publishing a report on potential carcinogenic substances in well-known cosmetic products (Facebook 2015). Among the general discussion on this topic, some users commented by referring to their carcinogenic-free products or posting links to their blogs.

Furthermore, practitioners can use the displayed buzzes for profitable advertising placement – as a post enters buzz territory, advertising platforms can charge higher prices to place adverts next to these kinds of posts. Moreover, they can analyse buzzes from other Facebook pages in order to bolster the chance of their own posts becoming buzzes. In this analysis, they can incorporate Berger's (2013) six STEPPS (social currency, triggers, emotion, public, practical value, stories) for contagious products and ideas. For example, social media managers can use the results of the BDS to analyse the buzzes that their competitors experienced or generated, and further examine the triggers behind a certain post becoming a buzz, including its topic and the story others created around it. They can even exploit detected buzzes as the starting point for engaging in improvised marketing intervention (Borah et al. 2019), which is the creation and execution of a real-time marketing communication in response to an external event. The BDS is also relevant for practitioners who want to monitor the number of buzzes generated by their own social media strategies. Practitioners may be able to determine the impact of their buzzes in terms of the number of followers gained or even the firm revenue generated. By incorporating the six STEPPS, they can also compare non-buzzes and buzzes in order to infer the marketing strategies that lead to the most impactful posts.

6. Conclusion

OSNs play a major role in today's world. Every day, the virtual world is inundated with posts, but only some catch on and become viral phenomena. Previous research has already intensively studied trending topics (Mathioudakis and Koudas 2010; Cvijijk and Michahelles 2011; Lau, Collier, and Baldwin 2012), but so-called buzzes have yet to be extensively analysed (Deusser et al. 2018). Buzzes fulfil the criteria of immediacy, unexpectedness, and intensity, which distinguishes them from well-studied trending topics. Their detection can be of great importance, e.g. for determining the social media strategy of companies, public figures, institutions, or political parties. By collecting metadata from public Facebook pages and manually classifying almost 120,000 posts, we estimated an LR model and trained different classifiers (i.e., AB, SVM, and RF) in order to build the first BDS. The results of the LR offer initial insights into important factors that aid buzz identification early in their evolution. These are lagging variables, such as the comments' length and likes, as well as derivation variables, like the number of new users and the number of page owner comments. Evaluating the BDS over a five-month period, we found that SVM and "LR 1%" performed best, together detecting over 97% of the buzzes.

Our efforts deviate from previous detection systems for viral online content in five key ways: We (1) focus on the detection of buzzes and thereby consider the unexpectedness characteristic, (2) manually classify posts to form the foundation of the BDS, (3) take into account prominent variables identified in previous research on viral online content, (4) account for user interaction, and (5) evaluate the BDS over a five-month evaluation period. Thus, we theoretically contribute new insights upon which further studies can build. We have provided a BDS that is intended for practical application while also suggesting recommendations for practitioners, e.g. to combine the results of SVM and "LR 1%" in order to obtain the best results.

Of course, this study features several limitations that represent avenues for future research. As we could not classify all posts as buzzes or non-buzzes, there is the chance that the BDS missed some buzzes. Likewise, AB and RF performed poorly because of the small number of variables used for classification. Thus, future studies could incorporate more variables and train more advanced models. Granted, this might require more data from Facebook that are not readily available. Future studies could also focus on optimising the BDS for larger Facebook pages (in terms of page likes) so that buzzes on these pages are detected earlier. In terms of detection speed, future research may also examine variables that help to detect buzzes in an even faster time. For example, the number of comments given in the first two hours after the publication of a post could be a promising predictor worth analysing. To generalise our findings, similar studies should be conducted on other social networks, such as Twitter or Instagram.

Acknowledgements

First of all, the authors would like to thank the “Deutsche Forschungsgemeinschaft (DFG)” for their funding without whom this research would not have been possible. Secondly, we are very thankful for the valuable feedback by Wendy Moe, editor of the *Journal of Interactive Marketing*, University of Maryland, and the two anonymous reviewers. Further, we would like to express our deep gratitude to Sebastian Pohl who helped us to develop the buzz detection system. Thanks to his relentless work, we were able to implement our idea and make this publication possible.

References

Ács A and Pagh SQ (2017) “Lovestorm” How a Local Community Reclaims Positive Public Perception by Using Social Media. In: *4th European Conference on Social Media*, pp. 1-9.

Akpinar E and Berger J (2017) Valuable Virality. *Journal of Marketing Research* 54(2): 318-330.

Asur S, Huberman BA, Szabo G and Wang GC (2011) Trends in Social Media: Persistence and Decay. In: *5th International Conference on Weblogs and Social Media*, pp. 1-8.

Berger J (2013) *Contagious: Why things catch on*. Simon and Schuster.

Berger J and Milkman KL (2012) What makes Online Content Viral?. *Journal of Marketing Research* 49(2): 192-205.

Borah A, Banerjee S, Lin YT, Jain A, Eisingerich A (2019) Improvised Marketing Interventions in Social Media. *Journal of Marketing* ISSN 0022-2429 (In Press).

Breiman L (2001) Random Forests. *Machine Learning* 45(1): 5-32.

Carroll JB (1961) The Nature of the Data, or how to Choose a Correlation Coefficient. *Psychometrika* 26(4): 347-372.

Cvijikj IP and Michahelles F (2011) Monitoring Trends on Facebook. In: *9th Conference on Dependable, Autonomic and Secure Computing*, pp. 895-902.

Deusser C, Jansen N, Reubold J, Schiller B, Hinz O and Strufe T (2018) Buzz in Social Media: Detection of Short-lived Viral Phenomena. In: *International World Wide Web Conference*, pp. 1443-1449.

Divgi DR (1979) Calculation of the Tetrachoric Correlation Coefficient. *Psychometrika* 44(2): 169-172.

Duda RO, Hart PE and Stork DG (2012) *Pattern Classification*. John Wiley & Sons.

Dudoit S, Fridlyand J and Speed TP (2002) Comparison of Discrimination Methods for the Classification of Tumors using Gene Expression Data. *Journal of the American Statistical Association* 97(457): 77-87.

Facebook (2013a) Barilla Apology. Available at: https://tinyurl.com/y3ckkwao (accessed 15 July 2019).

Facebook (2013b) Guido Barilla Apology. Available at: https://tinyurl.com/y5wld3al (accessed 15 July 2019).

Facebook (2015) Stiftung Warentest. Available at: https://tinyurl.com/y5n8uoey (accessed 15 July 2019).

Feroz Khan G and Vong S (2014) Virality over YouTube: An Empirical Analysis. *Internet Research* 24(5): 629-647.

Fernández-Delgado M, Cernadas E, Barro S and Amorim D (2014) Do we need Hundreds of Classifiers to solve Real World Classification Problems?. *The Journal of Machine Learning Research* 15(1): 3133-3181.

Freund Y and Schapire RE (1996) Experiments with a new Boosting Algorithm. In: *International Conference on Machine Learning*, pp. 148-156.

Hastie T, Tibshirani R, Friedman J, Hastie T, Friedman J and Tibshirani R (2009) *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Second Ed., Springer.

Hayes JL, King KW, Ramirez Jr A (2016) Brands, Friends, & Viral Advertising: A Social Exchange Perspective on the Ad Referral Processes. *Journal of Interactive Marketing* 36: 31-45.

Heimbach I and Hinz O (2016) The Impact of Content Sentiment and Emotionality on Content Virality. *International Journal of Research in Marketing* 33(3): 695-701.

Heimbach I and Hinz O (2018). The Impact of Sharing Mechanism Design on Content Sharing in Online Social Networks. *Information Systems Research* 29(3): 592-611.

Heimbach I, Schiller B, Strufe T and Hinz O (2015) Content Virality on Online Social Networks: Empirical Evidence from Twitter, Facebook, and Google+ on German News Websites. In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pp. 39-47.

Jansen N (2019) The Fiery, the Lovely, and the Hot – Analsyis of Online Viral Phenomena in Social Media. In: *27th European Conference on Information Systems*, pp. 1-16.

King G and Zeng L (2001) Logistic Regression in Rare Events Data. *Political Analysis* 9(2): 137-163.

Larivière B and Van den Poel D (2005) Predicting Customer Retention and Profitability by using Random Forests and Regression Forests Techniques. *Expert Systems with Applications* 29(2): 472-484.

Lau JH, Collier N and Baldwin T (2012) On-line Trend Analysis with Topic Models:# Twitter Trends Detection Topic Model Online. In: *Proceedings of International Conference on Computational Linguistics*, pp. 1519-1534.

Lesot MJ, Nel F, Delavallade T, Capet P and Bouchon-Meunier B (2012) Two Methods for Internet Buzz Detection Exploiting the Citation Graph. In: *International Conference on Fuzzy Systems*, pp. 1-8.

Lin M, Lucas Jr. HC and Shmueli G (2013) Research Commentary – Too Big to Fail: Large Samples and the p-Value Problem. *Information Systems Research* 24(4): 906-917.

Mathioudakis M and Koudas N (2010) Twittermonitor: Trend Detection over the Twitter Stream. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 1155-1158.

Meire M, Ballings M and Van den Poel D (2016) The Added Value of Auxiliary Data in Sentiment Analysis of Facebook Posts. *Decision Support Systems* 89: 98-112.

Moldovan S, Steinhart Y, and Lehmann DR (2019) Propagators, Creativity, and Informativeness: What Helps Ads Go Viral. *Journal of Interactive Marketing* 47: 102-114.

Pfeffer J, Zorbach T and Carley KM (2014) Understanding Online Firestorms: Negative Word-of-Mouth Dynamics in Social Media Networks. *Journal of Marketing Communications* 20(1-2): 117-128.

Quesenberry KA and Coolsen MK (2019) Drama Goes Viral: Effects of Story Development on Shares and Views of Online Advertising Videos. *Journal of Interactive Marketing* 48: 1-16.

Reynolds A (2002) News Influence Pictures of the World. *Media Effects: Advances in Theory and Research* 10(1): 1-16.

Roberts M, Wanta W and Dzwo T.-H. D. (2002) Agenda Setting and Issue Salience Online. *Communication Research* 29(4): 452-465.

Rietveld R, van Dolen W, Mazloom M and Worring M (2020) What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram. *Journal of Interactive Marketing* 49: 20-53.

Rooderkerk RP and Pauwels KH (2016) No Comment?! The Drivers of Reactions to Online Posts in Professional Groups. *Journal of Interactive Marketing* 35: 1-15.

Sandri M and Zuccolotto P (2006) Variable Selection Using Random Forests. In: Zani PS, Cerioli PA, Riani PM and Vichi PM (eds) *Data Analysis, Classification and the Forward Search. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin/Heidelberg, pp. 263-270.

Stieglitz S and 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.

Szabo G and Huberman BA (2008) Predicting the Popularity of Online Content. arXiv e-prints (2008). Available at SSRN 1295610.

Tellis GJ, MacInnis DJ, Tirunillai S and 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.

Tsur O and Rappoport A (2012) What's in a Hashtag?: Content Based Prediction of the Spread of Ideas in Microblogging Communities. In: *Proceedings of the 5th International Conference on Web Search and Data Mining*, pp. 643-652.

Twitter (2013) Oreo Cookie – Power out? No problem. Available at: https://tinyurl.com/y9gd45b6 (accessed 01 May 2020).

Vapnik V (1995) The Nature of Statistical Learning Theory. Springer, New York.

Vapnik V (1998) Statistical Learning Theory. Wiley, New York.

YouTube (2015) Delta's New Amazing Safety Video. Available at: https://tinyurl.com/y45xftz2 (accessed 15 July 2019).

Zubiaga A, Spina D, Martinez R and Fresno V (2015) Real-time Classification of Twitter Trends. *Journal of the Association for Information Science and Technology* 66(3): 462-473.

Appendix

Table A1 Correlation matrix of the independent variables

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	*firstLevelContentLength*	1.000						
(2)	*secondLevelContentLength*	.182*	1.000					
(3)	*repeatUsers*	.237*	.275*	1.000				
(4)	*commentLikes*	.153*	.236*	.181*	1.000			
(5)	*newUsers*	.095*	.071*	.090*	.121*	1.000		
(6)	*pageOwnerComments*	.056*	.072*	.192*	.054*	.053*	1.000	
(7)	*comments*	.021*	.055*	.056*	.065*	.098*	.152*	1.000
Note: * p < .01								