Leveraging peer-to-peer influence: measuring diffusion in online media publications

Jan Sodoge
Simona Bisiani

September 2020

Abstract

Rapidly changing modes of communication, news consumption, and emergent media business models are shaping the media landscape in the digital age. Within this paper, we analyze these dynamics by comparing the diffusion of media publications between traditional and emergent media players. We use the theoretical foundations of diffusion theory to apply mixed-influence models to assess the impacts of endogenous and exogenous diffusion based on two case studies. We find no significant differences in the influence of endogenous diffusion between emergent and traditional media organizations. Further, no evidence is found for the hypothesized growing importance of endogenous diffusion for traditional media publications over time. The method and framework presented within this paper hold relevance in future studies to the diffusion of publications by media organizations.

1 Introduction

The mediums through which individuals access the news have changed over time and increased in complexity (Tandoc and Jenkins 2017). From the start of humankind news diffused by word-of-mouth (De Fleur 1987). The printing invention, by enabling newspapers, presents one of the first cases where advancements in communication technology changed the diffusion of the news. Throughout history, the invention of information and communication technologies has impacted how news diffuse within societies in terms of underlying mechanisms, pace, and outreach. Today, spread and reach modes exploited by media organizations heavily revolve around digital technologies. These technologies and connected practices highlight the importance, for news organizations, to leverage peer-to-peer influence. Thereby, interpersonal communication is promoting itself as a strategy for success in widening the audience reach. Within this broader development, recent years have seen the emergence of new companies that have understood the growing relevance and benefits of peer-to-peer diffusion. One of these is Buzzfeed, a highly successful and influential emerging media organization relying on peer-to-peer effects to enlarge its audience reach. Buzzfeed operates through a business model that exploits statistics and data exploration methods that aim to consistently make content go viral. Following the hard/soft news categorization by Reinemann and colleagues (2012), publications by Buzzfeed can be classified as soft and has been described as informal (Wihbey 2019). Simultaneously, traditional media organizations have been subjected to the exercise of quickly adapting to new tools and changing audience preferences, as individuals' ways of digesting content progressively revolve around social media (Wihbey 2019). Additionally competing with emergent media organizations for audience attention, traditional media organizations are now faced with shrinking revenues and readership (Tandoc and Jenkins 2017).

The scribed changes to processes of diffusion and spread in the ever-evolving media landscape requires an empirical assessment as to how emergent and traditional media organizations reach their audiences. In this study, we build on theory from the sociology of diffusion to assess the difference in the diffusion of publications between an emergent media organization and a traditional media organization. We select *Buzzfeed* as the emerging media organization, and *die Zeit Online* as the traditional media organization. Methodologically, we avail of reader's comments to individual publications to approximate mixed model diffusion curves that measure the magnitude and relevance of peer-to-peer diffusion.

Based on the scribed context and motivation, our originating question covers and relates to the changes introduced by the digital age and emergent media organizations on the diffusion processes of media organizations' publications. Within this study, we generate two research questions that derive from the aforementioned developments and will be discussed within the following section. First, we are interested in capturing differences in diffusion between the emergent media players and traditional media organizations. Specifically, how do differences in magnitude effects introduced by peer-to-peer influence across media organizations impact the diffusion of publications? Second, we aim to take a longitudinal perspective to assess whether the relevance and influence of peer-to-peer effects have altered within traditional media.

This research paper is structured as follows. First, we review literature and theory on the diffusion of news and present how emergent players and traditional media organizations differ in diffusion processes and dynamics. Second, we describe our methodology, which is based on the modeling of diffusion processes and equation-based mixed-influence. Consequently, we describe the data collection and manipulation processes. We present findings concerning the discussed hypotheses and discuss these findings with context to the existing literature and our contribution to the field.

2 Background

Researcher interested in how news spread have traditionally availed of insights from the broader field of diffusion of innovations, which emerged in the 1960s (Rogers 1960) and established itself in the 1980s (Mahajan 1985). Diffusion theory provides tools that served disparate theoretical interests, ranging from measuring the speed of spread of one news story across a population (Greenberg 1964; Kanhian and Gale 2003), the role of different channels in the news diffusion (Rogers 2000), the life course and development of news (Rogers 2000; Im et al 2010), the characteristics of the audience (Ihm and Kim 2018), or of the news in itself (e.g. the content type) (Inoue and Kawakami 2004), and the relationship between the news and the audience (e.g. personal relevance) (Kanhian and Gale 2003).

Much of the early research used exceptional events as case studies such as Kennedy's assassination (Greenberg 1964), the 9/11 attacks (Kanhian and Gasle 2003), or Roosevelt's death (Miller 1945). Typically, diffusion patterns of these news reports showed that their spread was dependent on the combination of mass media input and interpersonal communication among individuals, meaning that both newsagents and people played a role in the diffusion of news events. A substantial stream of news diffusion research has delved into how news stories spread among a population due to interpersonal communication, what can be called peer influence, or peer-to-peer effects. Traditionally, these mechanisms have manifested themselves in the context of word-of-mouth communication. Kanhian and Gale (2003) describe some of the aspects that are likely to have traditionally driven word of mouth spread of news. For example, the magnitude of the news event, and a sense of immediacy and importance, have been deemed important drivers of people passing the news on to others. Jeffres and Quarles (1983) instead showed that an individual's location affected his reception mode, where being at work meant hearing the news by word of mouth, whereas being at home meant being likely to be exposed by the mass media directly through print and television. More recent studies have incorporated the internet in the analysis of the diffusion of online content (Im et al 2010). Here, internet-based technologies are described to operate as a disruptor of traditional production, distribution, and diffusion modes (Wihbey 2009).

Contemporary journalism is indeed facing numerous challenges, ranging from credibility issues, political polarization, and difficulty in generating revenue streams (Wihbey 2019). A lot of these challenges can be directly linked to the overturning of the definition of news by technology. Today news can come in different formats, at different levels of authority and objectivity, and with substantially differing purposes. Traditional media organizations are confronted by the entry in the journalistic field of new players, which bring to the table original ways of defining news, generating the content, and attracting viewership (Kovach and Rosenstiel 2014; Wihbey 2019).

As Wihbey (2019) explains, the study of online news diffusion requires a meaningful comprehension of what he calls "social facts" (Wihbey 2019: 215), defined as the package of (i) the content of the news and (ii) the additional information that accompanies the news and signals its approval or popularity. Oftentimes digital platforms provide, alongside media content, some indicators of content popularity. This could be, for example, an appreciation metric, like the number of likes the item has received, or an exposure measure, showing the number of times or users that have consumed the media item. The role of this additional information in shaping the diffusion of the content can be relevant to assess, as these "social cues" (Wihbey 2019: 215), alongside information about the media, are increasingly exploited by filtering algorithms and recommendations systems that highly determine the news people will digest (Wihbey 2019). An important insight here is that social cues can also serve as tools for people to give sense to a publication. A cue regarding high sharing of the publication might, for example, instigate an individual to do the same, or it simply might validate that news for the reader. Social media sharing operates as word of mouth in the online realm, and what makes people share or pass on news content has not yet been fully explored. One important element to consider when analyzing social media sharing is the role the social media platform in itself plays in influencing that process. Broadly, there are three potential ways in which social media can interfere in the news-sharing process. First, social media are often equipped with internal algorithms that tailor news suggestions to each individual user. Second, they allow individuals to re-share media content, critically increasing the role each individual has in exposing content to new potential audiences. Third, social media come with individualized network structures that can directly impact the diffusion of publications (Kim et al 2013).

Another important insight from the previous argument is that it is no longer sufficient for news media organizations to rely on their audience-base to survive. Crucially, knowledge of how people reach content is an essential component of the evolution process traditional media organizations are undergoing due to the rise of digital technologies. Nowadays, even the most established news organizations are seeing their audiences reducing at a steady pace (Tandoc 2018). However, emerging companies come with innovative business models, that demonstrate this does not need to be the faith for news media organizations. Companies that shine for their ease in dragging audiences onto its platform are *Mashable, Upworthy, and Buzzfeed* (Wihbey 2019). Buzzfeed, as described by Ellis (2014), is a news and entertainment website that uses a combination of user-generated content, traditional reporting, and online aggregation methods to generate stories. The website uses insights from social media to predict content virality and choose where to distribute such online content, and 75 % of its traffic comes from social media, mostly through Facebook (Saba 2014). Virality embodies the phenomenon of what Watts and Peretti (2007: 1) have called "the ultimate free lunch", the launch of content that generates attention through multiple acts of sharing from one user to the next, similarly to that of an epidemic. A non-peer-reviewed paper by Watts and Peretti (2007) studied the phenomenon of online virality. The authors point out strong levels of unpredictability as an important characteristic of the virality of journalistic work. The unpredictability of virality in the diffusion of journalistic work is significant because of several, interrelated factors: structural parameters of the social network in which a publication diffuses, individual user-preferences, personal thresholds, and the sharing behavior within different temporal phases of the diffusion. Based on the work by Centola (2010) some patterns can be identified on the micro-level. Particularly, unlike viral diseases, where one single exposure to a person with the virus is sufficient for contagion to happen, virality in a social network often requires multiple exposures to some media content for an individual to choose to pass it on (in what is known as complex contagions) (Centola 2010).

The design and strategy of Buzzfeed aim at leveraging virality by the means of social contagion while assuming that it cannot be predicted. A focal point was set on the "social process of passing" (Wihbey 2019:94) to create this contagion process. Knowing virality is exceptionally hard to predict, Peretti was described by Watts as seeking to instead "trigger social contagion in a probabilistic way, playing the odds and learning what works from the data" (Wihbey 2019:94). Wihbey (2019) describes Buzzfeed as obsessed with peer-to-peer influence, knowing social cues and social facts are critical for content success as they directly inform on what individuals are interested in at any given time, all while providing insights on the type of conversation the content generates.

The importance of companies like Buzzfeed cannot be dismissed because they can heavily impact the formation of public knowledge, the definition of what is news, and the future for traditional media players. Thus it is relevant to assess to what extent emerging media organizations and traditional news organizations differ in terms of diffusion given their different degree in focus on peer-to-peer effects. Based on reviewed strands of literature, we come to the following hypothesis.

H1: *Effects of peer-to-peer influence are more significant for an emerging media organization compared to traditional media.*

A question that arises from the previous reflections is: how much are the old players adapting to the modus operandi of the new players, and how much are the new players adhering to the core values of journalism as established by the cornerstones in the field? Paradoxically, emergent media organizations started appearing in times of economic hardships faced by traditional media (Tandoc and Jenkins 2017). Specifically, outlining the respective challenges traditional media face, Lievrouw (2002) describes the necessary adoption of contemporary technological media platforms and organizations. Wolf and Schnauber (2014) discuss these adoptions as crucial in the context of an increasing user-behavior of news consumption to mobile phones. Bechmann (2012) describes the responses of traditional media organizations in expanding to multiple platforms to publish content, particularly to those enabled by digital technologies and the internet. These processes are reflected within the importance media organizations assign to their appearance and establishment of themselves on these platforms (Wolf and Hohlfeld 2012). Similarly, social media networks can be discussed as a new platform traditional media are keen to embed in (Ripollés 2012). In this process

of adopting and establishing within new digital platforms, Wihbey (2019) outlines the orientation of many traditional media organizations on the practices of emergent media players like Buzzfeed. E.g. metrics and tools developed by *Buzzfeed* that measure the level of peer-to-peer sharing are adopted by different traditional media organizations (Wihbey 2019). Therefore, as a second hypothesis, we expect traditional media to have increased values of peer-to-peer-based diffusion over time, following the models presented by emergent media organizations.
H2: *Growing social media reach for traditional media organizations has led to an increase in the relevance of endogenous diffusion over time.*

3 Methodology

Peer-to-peer based diffusion reflects as relevant in creating virality effects and high levels of reach for emergent news media organizations. To study and compare the diffusion processes and the role of peer-to-peer dynamics across emergent and traditional media, this research is methodologically embedded in the sociological theory of diffusion. Here, diffusion describes a process by which an innovation is spread in a social system over time through particular flows of social interaction and communication (Rogers 1962). On a theoretical level, research differentiates endogenous and exogenous diffusion as distinct underlying processes or forces driving diffusion (Mahajan 1985). *Exogenous diffusion* describes a diffusion process where an external source (e.g. a media organization) imposes a centralized source of influence on the individuals to adopt an innovation. Within this process of external influence, each individual is equally likely to adopt the innovation. *Endogenous diffusion* is based on internal processes within the social system, among the individuals, that include social interaction e.g. passing of information about an innovation. Social contagion mechanisms are discussed in the literature as relevant in underlying endogenous diffusion processes (Mahajan 1985, Centola 2010, Wihbey 2019). While the previous descriptions of diffusion processes exclusively either driven by endogenous or exogenous dynamics are highly stylized and theory-based, most processes of diffusion analyzed in empirical research show an interplay of both endogenous and exogenous diffusion dynamics. Within this research, we conceptualize innovation processes on the macro-level by applying mathematical diffusion models. This group of models represents diffusion as a function that maps a time point in a diffusion process to the number of individuals who adopted an innovation. Research developed multiple equation-based diffusion models with different focal points that are used both to forecast diffusion dynamics and to analytically describe underlying dynamics driving certain kinds of diffusion processes.

Mahajan (1985) describes a basic equation-based diffusion model (see Equation 1) that models the rate of diffusion at time t. The rate of diffusion is modeled as a function of the number of individuals who at time t did not adopt the innovation. To assess the number of individuals who have not adopted an innovation, the model assumes a defined number of individuals that represent the potential audience to adopt an innovation ($\bar{N}$). The audience thereby defines the maximum of the function towards which the diffusion function converges. The characteristic dynamics of a diffusion process in this model are determined by the *coefficient of diffusion*. The specification of the diffusion coefficient determines the involved mechanisms of endogenous and exogenous diffusion.

$$\frac{dN(t)}{dt} = g(t) * (\bar{N} - N(t)) \tag{1}$$

Exogenous diffusion models integrate a constant influence rate of diffusion. These dynamics are captured by a constant factor of a within the diffusion model that interacts with the group of individuals who have yet not adopted an innovation. At the time t the share of individuals to adopt an innovation is identical while the pool of potential adopters decreases over time (Rossman, 2014). A diffusion curve of an exogenous diffusion model displays high rates of diffusion that decline constantly over time as fewer individuals adopt the innovation in each time step.

$$\frac{dN(t)}{dt} = a * (\bar{N} - N(t)) \tag{2}$$

Equation 2 describes the rate of diffusion as based on the interaction of individuals who adopted an innovation at time t and the share of individuals who did not adopt the innovation at time t. Thus, this multiplication is discussed to reflect the social interaction within a group of individuals. The interaction is modeled by a *index of imitation*, here b that describes the strength of the interaction on the adoption of the innovation. When modeled, cumulative endogenous diffusion takes the shape of an S-curve. In the initial phases of the introduction of innovation (or the generation of a rumor, as well as the creation of content online) the adoption starts slow, but as soon as it reaches the maximum rate of diffusion it will exponentially increase, to slow down again once it has reached most of the population (Rossman 2014).

$$\frac{dN(t)}{dt} = b * N(t) * (\bar{N} - N(t)) \tag{3}$$

The mixed-influence model refers to a combination of both models of endogenous and exogenous diffusion dynamics (see Eq. 4). Mathematically, the mixed-influence model can be rephrased as an addition of the diffusion rates of both endogenous and exogenous influences.

$$\frac{dN(t)}{dt} = (a + b * (N(t)) * (\bar{N} - N(t)) \tag{4}$$

Mixed-influence models can be used to assess the magnitude of endogenous and exogenous diffusion processes as they reflect parameters for both endogenous and exogenous diffusion (Mahajan 1985). In this study, we apply a mixed-influence model to assess these for the diffusion dynamics of media organization publications. As a mixed-influence model, we apply the generalized Bass-model

(Bass 1969) which models diffusion processes based on coefficients reflecting the strengths of each endogenous and exogenous influence processes. Bass's model is based on the assumption that two forces influence an individual to adopt an innovation that corresponds to the discussed dynamics of endogenous and exogenous diffusion: behavioral tendencies to imitate other individuals and ambitions for innovating behavior. These dynamics behaviors correspond to endogenous and exogenous diffusion where the coefficient of imitation reflects endogenous dynamics and the coefficient of innovation reflects exogenous dynamics. We use the ratio of the coefficients of imitation and innovation q/p (also referred to as bass-ratio in the following) to assess the influence and relevance of endogenous diffusion compared to exogenous diffusion. Ratios of higher numerical value reflect an increasing relevance of endogenous dynamics to the diffusion of an innovation. The ratio is discussed with literature as a relevant parameter with respect to modeling diffusion curves and analyzing the relevance of both diffusion mechanisms (Bemmaor and Janghyuk 2002, Ntwoku and Meso 2017). We apply the *diffusion* package in the R programming language to estimate Bass-models based on the observed diffusion curves of journalistic publications.

4 Data

We collect data from two media organizations to represent an emergent media organization and a traditional news organization. The latter is represented by the German *die Zeit Online* which is published as a print newspaper and an online organization. This organization was selected both as it represents a traditional media, historically based on a well-established printed newspaper, and as we can avail of their web archive of publications, constituting the most extensive database of publications and precisely marked comments we can access. Both display a comment section directly below the publications. In this research, we exploit data from this section to generate diffusion curves based on the number of comments and their distribution over time.

We use the distribution of comments over time as an approximation of the diffusion dynamics of the publication. It serves to approximate the diffusion of a publication where the ground truth is represented by the unique views (unique views of unique users) a publication receives over time. While we recognize that the number of views of a publication is more a precise measure of diffusion, there is to the best of our knowledge no media organization offering consistent data of publication views publicly. The total number of comments can be anticipated as highly influenced by the engagement or content of the publication itself. Thus, comparing publications by the number of comments they receive in the context of diffusion represents a biased measure. Nevertheless, in this research design, we aim to compare the dynamics of the diffusion curves (endogenous/exogenous characteristics) that are independent of the number of comments a publication receives.

In this research, characteristics of diffusion curves of publications are compared by the means of fitting the Bass-model equation on the observed distribution of comments over time. The resulting coefficients of the Bass-model are independent of the total number of comments but depend only on the development and distribution of these over time. In this context, we standardize the range of comments over time to an interval of [0,1] to reflect on the diffusion process. In comparing publications solely by the characteristics of resulting diffusion curves, it is a central assumption to our research that the share of views corresponds proportionally to the number of comments made on a publication on the level of a single publication. This implicates the share of readers who comment is hypothesized to remain constant over time as a share of those who read the publication. For example, when fewer people read a publication in the later stages of the diffusion process, proportionally fewer people will comment on the publication. We find support for this assumption by research that highlights the ability to successfully predict views by social interactions with the publication such as comments (Castillo, Carlos, et al 2014; Keneshloo, Yaser, et al 2016). In using comments to create diffusion curves, we aim to account for biases introduced by the engagement rate of a publication by counting only comments by unique users and thus ignore second or further comments by the same user. Also, we delete threads of comments which result in discussions and are likely influenced by the engagement of a publication. These threads or sub-comments are relevant features in both media organizations that we sample data from. Further strategies we integrate to control for engagement influences are highlighted in the following sections.

Both *Buzzfeed* and *die Zeit Online* show differences in the design and display of comments and timestamps in their comment sections. Thus, two different approaches for data collection are necessarily applied that are discussed in the following [1].

For *die Zeit Online* we access the online archive to collect information on comments using a headless web-crawler. The archive enables us to collect data on publications between 2010 (i.e. the first year where we observe comments) and 2019. Here, the timestamps displayed for each comment are static and consist of year, month, day, hour, and minute. Thereby, crawling these timestamps allows generating precise diffusion curves for each publication using the following procedure. We define t_0 as the date of publication of the respective publication. Timestamps are scraped for all comments of the publication. Both dates of publication and timestamps are converted to the *ISOdatetime* format in R that represents the timestamp as a numeric value. This allows picturing the diffusion process on a numeric interval and calculating time differences between timestamps. We create a cumulative diffusion curve by sorting the timestamps of comments, mapping them using a cumulative count on the interval of the *ISOdatetime* values.

Buzzfeed displays the timestamps of comments dynamically i.e. the label of a comment may change from "One hour ago" to "One day ago" if a longer time passes by. Thereby, using the single comments timestamps to generate diffusion

[1] Code for crawler and data analysis is available at github.com/jansodoge/journalism_diffusion_processes

curves would have created inaccuracies as timestamps of older comments are inaccurate. Thus, it was necessary to track the development of the number of comments using an on-line approach for crawling. Practically, we construct a web-crawler that counts the number of comments on each publication every hour for a time frame of 60 consecutive hours in total. Here, diffusion curves are generated as we map, for each publication, the number of comments made each hour on a time interval. t_0 is defined as the first time a publication was captured on the*Buzzfeed* website.

Several measures are taken to standardize and thus make publication comparable between both organizations. First, timestamps are standardized to an identical t_0 (i.e. the time of publication). Also, the number of comments over time is standardized to an interval of $[0, 1]$. For *die Zeit Online* we observe a large majority of comments published within the first 60 hours after publication (see Fig 1). To enable greater comparability concerning *H1* we decided to reduce the comments for this time frame of 60 hours that corresponds to the data selected for *Buzzfeed*. An exploratory analysis showed that the excluded comments at *die Zeit Online* reflected stages in the diffusion where the diffusion curves slowly converged to the audience. Thereby, we reason no loss of information in reducing the time frame here.

On the resulting data for both media organizations, we select publications with (i) at least an increase of 20 comments within the observed time frame and (ii) with less than 20 comments at the earliest point we observe these. The second criteria specifically apply to data generated from *Buzzfeed* where the crawler captures publications that are already published for multiple hours and evolved through diffusion parts. The first criterion (i) is applied to select publications with a minimum number of observations that serve sufficient to fit the Bass-model on the data.

In particular for data derived from *die Zeit Online* the observations (i.e. timestamps of comments) are unequally distributed over time. For diffusion curves based on comment timestamps, we observe a lower density of comments in a defined time interval where the diffusion function converges. This can be observed specifically in data from *die Zeit* (see Fig. 1b) as it results from fewer comments made in these parts when the rates of diffusion of respective decrease significantly. This unequal distribution of comments thus relates to the dynamics of diffusion curves (fewer comments implicate fewer views at later stages in the diffusion process of a publication). This phenomenon is less salient for *Buzzfeed* where the number of comments is assessed roughly in time steps of one hour (not precisely given technical-aspects in the crawling process). We apply linear interpolation to derive discrete, fixed time steps in diffusion curves. This allows a standardized and comparable format for diffusion curves with respect to time t (t_0, t_1, t_{n-1}, t_n).

For the final dataset, this results in 46 publications that match the criteria and are used within this research from *Buzzfeed*. The mean of the maximum number of comments reached for these publications is 81.3 while the minimum is around 11.2. These publications on average diffuse from 11.2 to 81.3 comments over time. Among these publications, five are identified with the *quiz* section on

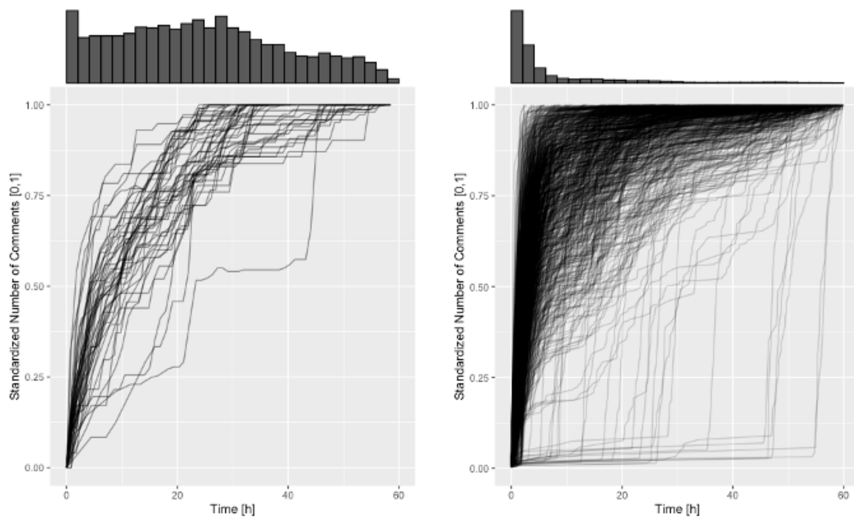

Figure 1: Interpolated diffusion curves for *Buzzfeed* (left), *Die Zeit Online* (right). Histograms reflect density of points of measurements.

Table 1: Dataset structure and variables

Publication ID	Imitation	Innovation	Ratio imitation/innovation	Media organization
url_1	p_1	q_1	p_1/q_1	Buzzfeed
...	...	...	...	Buzzfeed
url_{46}	p_{46}	q_{46}	p_{46}/q_{46}	Buzzfeed
url_{47}	p_{47}	q_{47}	p_{47}/q_{47}	die Zeit
...	...	...	...	die Zeit
url_{831}	p_{831}	q_{831}	p_{831}/q_{831}	die Zeit

Buzzfeed, and 11 are identified with the *TV and movie* section. While we also crawled data on the *shopping* section, we did not find any publications matching our criteria. For *die Zeit* we obtain diffusion curves for 795 publications for 2011 (117), 2012 (71), 2014 (94), 2015 (176), 2017 (59), 2018 (278). The respective years were chosen as they reflect three-time intervals within the past decade with equal time steps separating them.

Consequently, the Bass-model is fitted onto the generated diffusion curves of each publication. The resulting dataset contains observations representing a publication characterized by the coefficients for innovation (exogenous diffusion), imitation (endogenous diffusion), and the ratio of both. For publications from *die Zeit Online* two additional variables are introduced: a boolean that captures whether the publication was published during night-time or day-time; a numeric variable that counts the maximum number of comments observed for the respective publication. These will be integrated and discussed within the context of the second hypothesis.

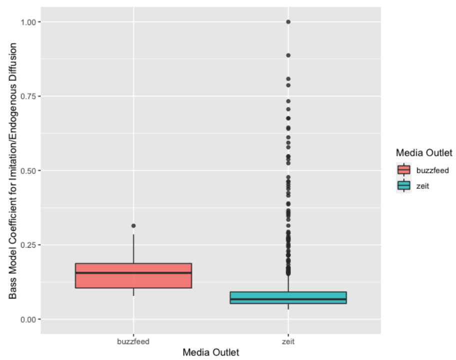

Figure 2: Coefficients of imitation: comparison between *Buzzfeed* and *die Zeit Online*

5 Results

5.1 Hypothesis 1

To investigate *H1* we compare the coefficients representing the levels of imitation (p) of the modelled diffusion curves between Buzzfeed and *die Zeit Online*. The distribution of imitation coefficients is displayed for both media organizations.

In direct comparison, we observe different distributions between both media organizations. For *die Zeit Online* we observe a significant number of extreme values (outliers) within the fourth quantile, which create a skewed distribution. For *Buzzfeed* the observations display lower levels of standard deviation and thus only one outlier outside of the whiskers. Within these differing characteristics of the imitation parameters distribution, comparing the median of both groups shows higher imitation coefficients for Buzzfeed. While the mean is less vulnerable for extreme values, using the mean to compare both samples shows higher imitation coefficients for *die Zeit Online*. The latter finding can be related to the differing distributions between both samples. Given the different results using these measures of distribution, we apply an unpaired two-sample t-test to assess whether differences between both samples are significant. Both samples fulfill necessary conditions for this test i.e. independent samples, normal distribution and similar variances. The test shows no significant differences between both samples ($p = 0.6756$). Based on this we can thus provide no evidence for our first hypothesis and reject it.

5.2 Hypothesis 2

To test the section hypothesis we apply a longitudinal perspective on the modelled diffusion coefficients for publications published by *die Zeit Online*. We apply the ratio of the coefficient of imitation and the coefficient of innovation (bass-ratio) to track the relative important of imitation (i.e. endogenous dif-

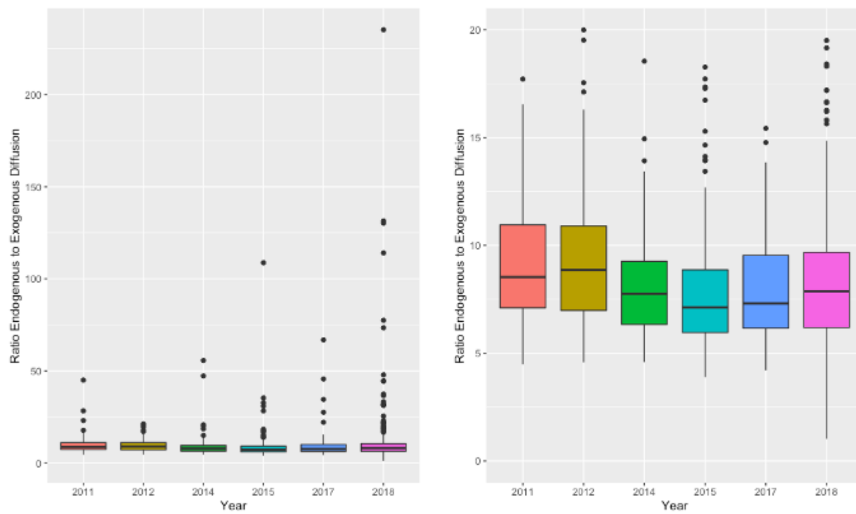

Figure 3: Longitudinal development of endogenous to exogenous diffusion ratio for *die Zeit Online*: overall distribution (left), zoomed in on box-plot centers (right)

fusion). A visual investigation of the historic developments is displayed within Figure 3. As we discussed in the context for *H1* the overall distribution of coefficients of imitation reflects here similarly in the ratio of imitation and innovation with respect to the role of extreme values. In the context of the ratio we consider, extreme values reflect publications that have significantly higher contributions by endogenous diffusion. We observe the role of extreme values to shape the respective year's sample distribution to increase in later years. This is reflected in an increasing standard deviation between 2011 and 2018. Similar to the observations made on the development of mean and median, we find the mean to be influenced by the extreme values. Measured by the mean we observe an increase in the importance of endogenous diffusion over time. Using the median, we do not find these trends reflect. There is no clear trend to be recognized over time in the median of the bass-ratio we apply as a measure.

In a consequent step, we construct a linear regression model to assess the impact of historic development on the ratio of endogenous and exogenous diffusion. Thus, the ratio is modelled as the dependent variable which and the temporal dimension is reflected within the independent variables. We model the temporal dimension within three categories. While we possess data on diffusion curves for years 2011,2012,2014,2015,2017,2018, we sort these into three distinct categories: early- (2011-2012), mid- (2014-2015) and late-years (2017-2018). Given the gaps in data collection for two years, we decide to use this scale instead of modelling *years* as a continuous variable.

We construct three models. A first model (1) uses only the three categories of years (using the early-years as baseline category). A second model (2) in-

Table 2: Measures of centrality and variability for endogenous to exogenous diffusion ratio for *die Zeit Online* by year

Year	Median	Mean	SD	Number of publications
2011	8.7	10.01	4.95	117
2012	8.87	9.91	3.99	71
2014	7.81	9.28	7	94
2015	7.18	8.97	8.81	176
2017	7.52	10.62	10.26	59
2018	8.15	12.29	19.92	278

tegrates the maximum number of comments as a control variable. Including maximum number of comments aims at controlling for influences of engagement i.e. whether publications with more comments diffuse differently than those with less. Third, we introduce a dummy-variable to control for whether the publication was published during day-time (8-22) or night-time. This serves to control for publications that were published during the night-time hours and therefore do not display exogenous diffusion dynamics captured by an initially steep diffusion curve. In each model, early-years are used as the baseline to compare to mid- and late-years.

For all the regression models we find no integrated coefficient with a p-value lower than 0.05 which would be considered statistically significant. Therefore, we phrase p-values lower than 0.1 as *weakly significant* within the following analysis to point out differences in significance of coefficients. For the first model (1) we the group of late years to display positive, weakly significant effect on the ratio of endogenous to exogenous growth. This can be interpreted as a higher importance of endogenous diffusion to news within the late years compared to the early years in our sample. For the mid-years we find a non-significant, negative effect. Within the second model (2) the weakly significant effect of the late-year category remains with a similar magnitude. The introduce control variable of maximum number of comments does not show up significant. The third model that introduces the control for day-time shows the late-year category as non-significant while the introduced control variable is negative and weakly significant. The negative coefficient can be interpreted as that publications published during night-time reduce the importance of endogenous diffusion. The effect of the introduction of this variable on the influence of the late-years variable can be interpreted as that the differences in bass-ratio between publications are rather explained by the time of publication than the years. Here then, a mediating effect is likely in place. When excluding the publications published during the night the coefficient of *Late-years* becomes insignificant and drops to 1.53.

Table 3: Linear Regression Model Analysis, Hypothesis 2

	Dependent variable:		
	Ratio endogenous to exogenous diffusion		
	(1)	(2)	(3)
Late-years (2017,2018)	2.027*	2.041*	1.748
	(1.203)	(1.202)	(1.213)
Mid-years (2014,2015)	−0.890	−0.830	−0.847
	(1.255)	(1.255)	(1.254)
max. comments		−0.011	−0.012
		(0.007)	(0.007)
Day-time			−3.591*
			(2.129)
Constant	9.971***	11.069***	14.695***
	(0.964)	(1.215)	(2.468)
Observations	795	795	795
R^2	0.010	0.012	0.016
Adjusted R^2	0.007	0.009	0.011
Residual Std. Error	13.215 (df = 792)	13.205 (df = 791)	13.189 (df = 790)
F Statistic	3.872** (df = 2; 792)	3.317** (df = 3; 791)	3.205** (df = 4; 790)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Share of significant ($p < 0.1$) coefficients, sample of 80 % of dataset, 10,000 simulations

	(1)	(2)	(3)
Late Years	.33	.35	.08
Mid Years	.002	.002	.002
Max. comments	NA	.23	.26
Day-time	NA	NA	.5
Intercept	1	1	1

5.3 Robustness Analysis

We conduct a robustness analysis of the presented results of the regression model to assess the insecurities and variability introduced by the extreme values within the samples representing diffusion at *die Zeit Online*. We run the regression models based on randomly sampled (without replacement) subsets of 80% of the original dataset for 10,000 runs. We evaluate the share of simulations where coefficients are significant at $p < 0.1$ (see Appendix A). For the simple model (1) that applies only the year-category as a predictor, we find the late-years category significant for 33% of simulations. When integrating the maximum number of comments as a control variable, the share slightly increases to 35% while the max. comments coefficient is significant in 23% of simulations. When additionally controlling for day-time, this variable is significant for 50% of simulations while the share of late-years category drops to 8%. The results present evidence of the instability of the results. The low probabilities for the late-year category (max. 35% in (1) and (2)) highlight the instability of the results is highly dependent on the extreme values.

6 Discussion

This research aims to expand on literature of news diffusion by exploring how publications from different media organizations spread. The media landscape has seen the recent entry of new companies which have brought with themselves novel methods of reaching audiences and generate revenues. Critically, companies like Buzzfeed are described as exploitative of peer-to-peer effects, whereby they rely on digital forms of word of mouth to gain audiences. Their modus operandi highlights the importance, in the news diffusion process within the digital context, of "the people formerly known as the audience" (Rosen 2012: 14). The inherent complexity emerging from the interplay of audiences, social media networks, and media organizations online has brought researchers to explore various aspect of how these impact news diffusion processes. While some focus on the diffusion of a single news event (Rogers 2000), some on the characteristics of the audience in relation to passing content on (Kanhial and Gale 2003), and others on the intrinsic features of the news in itself (Inoue and Kawakami 2004), empirical studies comparing diffusion processes among diverse media organizations are yet to be pursued. In this research we contribute to this gap by comparing diffusion curves of an emerging media player, *Buzzfeed*, and a traditional news organization, *die Zeit Online*. Similarly, there have not been any attempts, to the best of our knowledge, to measure the evolution of diffusion patterns within a media organization over time. Thus, we incorporate in this research a longitudinal measurement of diffusion processes at a traditional media organization. The motivation behind the temporal dimension arises from the fact that media organizations are increasingly embedded in social media, as a way to reach audiences.

We empirically tested two hypotheses which connect with contemporary research on diffusion of news in a changing media ecosystem. We find no evidence for our first expectation and thus reject the hypothesis of significant differences in the contribution of endogenous diffusion between an emergent and a traditional media organizations. Using a t-test to compare samples of both types of media organizations, no significant differences in the coefficients of imitation were found. While the medians between both types differed, the skewed distribution of imitation coefficients per publication for the traditional media organization led to rejection of significant differences. The second hypothesis investigated a longitudinal perspective on traditional media organizations where literature suggested an increasing importance of endogenous diffusion patterns. Using a linear regression model we find a weakly significant effect of increasing importance of endogenous diffusion in a simple model using only the longitudinal dimension to model the ratio of endogenous to exogenous diffusion. Though when controlling for whether a publication was published during day-time or night-time, we find that the effect of years becomes insignificant while the control variable becomes significant. Likely then, publications published during night-time were modelled as driven largely by endogenous diffusion as comments within the first hours did not appear given to a missing audience that delayed the diffusion process until the morning hours. Thus, a mediating effect of the time of publication in the investigated causal relationship can be assessed. Where the distribution of the diffusion coefficients for *die Zeit Online* is skewed with significant extreme values, the robustness analysis stressed the instability introduced thereby to our results.

Our attempt to represent a diverse media ecosystem is based on the choice of two case studies: *die Zeit Online* and *Buzzfeed*. With respect to comparing both case studies to assess the influences of endogenous news diffusion, differences between both media organizations, their products and audiences need to be critically discussed. Ihm and Kim (2018) suggest that sharing of news depends on individual-level characteristics relating to desire and considerations of self-presentation. Hence, decisions and behavior of peer-to-peer sharing among individuals has to be considered in the context of active considerations with respect to a formation of social ties and positioning within a social network. Also, Rosengren (1973) highlights socio-demographic variables as relevant in an individual's news sharing behavior. Nevertheless, there are yet no studies we can avail of to integrate connected variables to the specifically selected media organizations covered within this research. Only broader statements in differences between the audiences can be made by that *Buzzfeed* and *die Zeit Online* relate to different audience. While *Buzzfeed* addresses a younger, global, English-speaking audience, *die Zeit Online* addresses a German-speaking, educated audience. Within this research we focus on the role of media organization types and their influences on the diffusion processes. This dimension in the context of new diffusion can be referred to as extrinsic where we refer the cause not within the product that diffuses itself but within its innovator. Instead, on the intrinsic level of news diffusion i.e. the news itself (e.g. topic or content-type, format etc.) literature shows differences in diffusion patterns, too. Kümpel et al. (2015) discuss that content associated with neutral or negative character are less present in peer-to-peer diffusion patterns. As highlighted within section 2, these differences can clearly be stated between the two selected media organizations within this research. Concerning hypothesis 1 where both organizations are compared it is thus relevant to highlight effects resulting intrinsically from the diffusing news publication. For hypothesis 1 we found no significant differences by the means of a t-test which can to a large extent be explained by the extreme values within the sample for *die Zeit Online*. Nevertheless, the differences visualized within the comparison of medians of both samples can thus not exclusively be attributed to the different media organization types but additionally to the character of the diffusing news itself, too. While we highlight the effect of the media organization type in section two, we cannot account for the magnitude of the effects that both the extrinsic and intrinsic forces cause.

To empirically investigate the differences between emergent and traditional media organizations as well as longitudinal developments within traditional media we avail of data on user comments. Comments are used within this research as an approximation to the number of unique views a publication receives. The unique views then present the ground truth for the diffusion of a publication i.e. how many people a publication reaches. This research rests on the described assumption that comments can be viewed as an approximation for the views a publication receives. While we find support in the literature (Castillo, Carlos, et al 2014; Keneshloo, Yaser, et al 2016) based on research that shows how among others the number of comments can predict the clicks of a publication, we cannot tests the assumption as data on views is not publicly available. Future research thus needs to take into account this for further testing of this assumption. Additionally, while research shows that the intrinsic engagement-level of a publication influences the outreach (O'Brien and Heather 2011), there are yet no studies assessing the intrinsic engagement-character of a publication on the level of endogenous vs. exogenous diffusion. Thus, while our approach to measure endogenous and exogenous diffusion levels, the prior findings do not confound our results. Furthermore, we find no significant coefficient when controlling for the maximum number of comments a publication in influencing the bass-ratio in traditional media longitudinally.

Thus, it holds relevant to reflect on limitations to this research introduced by using this approximation metric to assess diffusion. Modeling diffusion processes of publications based on their comments introduces a pre-selection of publications. This is necessary as publications with few comments cannot be precisely modeled using the Bass model. These curves we theorized cannot reflect on diffusion curve dynamics detailed enough for the Bass-model to make a valid claim on the impacts of endogenous or exogenous diffusion. Therefore publications with a range of less than 20 observed comments are dropped from the sample. While this number as a cutoff value is not based on literature, it was arbitrarily selected after a visual inspection of diffusion curves with different numbers of comments. This selection process imposes a critical discussion in the context of endogenous selection bias. This is discussed by Elwert and Winship (2014) as the conditioning influence of collider variables on a link between treatment and outcome. In this study, a potential endogenous selection bias results as the research design integrates control for a collider that though should not have taken place in the research. The findings on the causal effect between treatment and outcome can be biased as the collider mimics the effects of causality that would otherwise not exist. The range of 20 comments as criteria thereby reflects a collider variable in our research design on which basis the conditioning takes place. Analyzing only publications with a sufficient range of comments in the diffusion process is conditioning with the range selection as a collider by which a non-causal association may be introduced to the mechanism we study (Elwert and Winship 2014). Using the unique views of a publication would introduce no similar constrain as diffusion curves are reflected perfectly by the ground truth. Here, we cannot assess the impact of the collider.

The application of the Bass-model introduces several assumptions on the modeled diffusion process (Mahajan 1985). Depending on the specific case study of a diffusion process these assumptions need to be reflected on and compared to the dynamics of observed diffusion and individual-level behavior. In the context of news diffusion among consumers, we critically discuss these assumptions as listed by Mahajan (1985) within the following. First, the mixed-influence model by Bass reflects highly simplifying processes of social interaction in creating endogenous diffusion. As specifically presented in the endogenous diffusion model (Eq. 3), social interaction among individuals that creates endogenous diffusion is based on a hypothetical social system where the interaction between two randomly selected individuals is uniformly distributed. The influence of peer-to-peer diffusion within social networks is a focal point to the mechanisms investigated within this research. This simplifying assumption can thus be criticized where it does not accurately reflect the importance of social networks to this research. Nevertheless, this research does not integrate underlying network structures and remains on a macro-level as micro-level data is not available. Thus, the simplification in this context can be regarded as acceptable as the integration of network dynamics are not reflected in the underlying dataset used and not within the scope of this research. A second assumption of the Bass model relates to the size of the modeled social system in which news diffuse. In the Bass model, the capacity ($\bar{N}$) is a fixed, static parameter. With respect to the diffusion of journalistic publications in a digital space, static limitations of the size of the social system in which these diffuse represents a simplifying assumption. Instead of a static audience (that printed newspapers can be conceptualized as), digital journalism reflects dynamically expanding audiences. These can likewise depend on the underlying network structure as well as the mechanisms driving large social media platforms that distributed content. Third, the Bass-model assumes static features of an innovation i.e. there are no changes to its constituent components. In our research we look at the diffusion of news publications by selecting a range of publications from two media organizations and observing their individual spread over time. However, in reality news stories in an online ecosystem are highly dynamic, meaning they can be updated, rewritten, and evolve, not only by the media agent but also by individuals (Im et al. 2010). This consideration highlights that our research should not be used to gain a deeper understanding of the mechanisms defining the life course of a piece of news or a particular story, as those dynamic elements, as well as intrinsic news characteristics, are entirely left out from the research design. Rather, our interest lies in the quantitative assessment of the difference in diffusion of a single static publication depending on the type of media organization that generated the content. While we have no empirical data on the frequency of these changes, *Buzzfeed* displays a timestamp for the last update on a publication. In most cases, we observe these are additions to a publication or at least impose no significant change in the topic or context of a publication. Further assumptions discussed by Mahajan (1985) are less critical to this research. The binary conceptualization of adopting an innovation (here absorbing online journalism content) is present in our case. Similarly, we can follow the assumption of the Bass-model that an individual can only once adopt an innovation, here absorbing the specific publication.

7 Conclusion

By modeling diffusion curves of news publications we find no statistically significant evidence of higher endogenous diffusion rates for the emerging media organization, that would reflect the peer-to-peer mechanisms described by the literature. Another aspect we explored was whether the increased embedding of traditional news organizations in social networks and focus on peer-to-peer diffusion reflected in an empirically observable increase in endogenous diffusion over time at *die Zeit Online*. Similarly to our first hypothesis, we do not find evidence of increased endogenous diffusion over time for a traditional news organization. Despite inconclusive results, we believe this research is equipped with solid theoretical assumptions, and that it provides a useful framework that can be adopted in future research to further test differences in diffusion processes across media organizations. We encourage future research that studies how different media organizations spread and change over time, as ever-evolving audience preferences and digital tools are substantially reshaping the media ecosystem, with significant implications for journalism. While the presented research yet remains on the macro-level, future research priorities need to focus on the underlying mechanisms on the micro-level. Where we discussed insecurities in our results on the macro-level, an investigation of the micro-level can help to improve an understanding of how differences between emergent and traditional media organizations come about and how consumer behavior impacts the underlying diffusion mechanisms of traditional media over time. Furthermore, the integration of views as a measure of publication diffusion in future research can lead to more empirically significant results and test the application of comments as a proxy.

References

[1] Bass, F. M. (1969). A new product growth for model consumer durables. Management science, 15(5), 215-227.

[2] Bechmann, A. (2012). Towards cross-platform value creation: Four patterns of circulation and control. Information, Communication & Society, 15(6), 888-908.

[3] Bemmaor, A. C., & Lee, J. (2002). The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. Marketing Science, 21(2), 209-220.

[4] Boczkowski, P. J. (2010). News at work: Imitation in an age of information abundance. University of Chicago Press.

[5] O'Brien, H. L. (2011). Exploring user engagement in online news interactions. Proceedings of the American society for information science and technology, 48(1), 1-10.

[6] Castillo, C., El-Haddad, M., Pfeffer, J., & Stempeck, M. (2014, February). Characterizing the life cycle of online news stories using social media reactions. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 211-223).

[7] Centola, D. (2010). The spread of behavior in an online social network experiment. science, 329(5996), 1194-1197.

[8] De Fleur, M. L. (1987). The Growth and Decline of Research on the Diffusion of the News, 1945-1985. Communication Research, 14(1), 109-130.

[9] Ellis, J. (2014). Buzzfeed. Nieman Journalism Lab.

[10] Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual review of sociology, 40, 31-53.

[11] Greenberg, B. S. (1964). Person-to-person communication in the diffusion of news events. Journalism Quarterly, 41(4), 489-494.

[12] Ihm, J., & Kim, E. M. (2018). The hidden side of news diffusion: Understanding online news sharing as an interpersonal behavior. new media & society, 20(11), 4346-4365.

[13] Im, Y. H., Kim, E. M., Kim, K., & Kim, Y. (2011). The emerging mediascape, same old theories? A case study of online news diffusion in Korea. New Media & Society, 13(4), 605-625.

[14] Inoue, Yasuhiro, Kawakami, Yoshiro (2004). Factors influencing tabloid news diffusion: Comparison with hard news.

[15] Jeffres, L. W., & Quarles, R. (1983). A panel study of news diffusion. Journalism Quarterly, 60(4), 722-725.

[16] Kanihan, S. F., & Gale, K. L. (2003). Within 3 hours, 97 percent learn about 9/11 attacks. Newspaper Research Journal, 24(1), 78-91.

[17] Keneshloo, Y., Wang, S., Han, E. H. S., & Ramakrishnan, N. (2016, December). Predicting the shape and peak time of news article views. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 2400-2409). IEEE.

[18] Kim, M., Newth, D., & Christen, P. (2013). Modeling dynamics of diffusion across heterogeneous social networks: News diffusion in social media. Entropy, 15(10), 4215-4242.

[19] Kovach, B., & Rosenstiel, T. (2014). The elements of journalism: What newspeople should know and the public should expect. Three Rivers Press (CA).

[20] Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. Social media+ society, 1(2), 2056305115610141.

[21] Lievrouw, L. A. (2002). Determination and contingency in new media development: Diffusion of innovations and social shaping of technology perspectives. Handbook of new media: Social shaping and consequences of ICTs, 183-199.

[22] Mahajan, V. and Peterson. R. (1985). Models for Innovation Diffusion. Beverly Hills: Sage.

[23] Ntwoku, H., Negash, S., & Meso, P. (2017). ICT adoption in Cameroon SME: application of Bass diffusion model. Information Technology for Development, 23(2), 296-317.

[24] Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2012). Hard and soft news: A review of concepts, operationalizations and key findings. Journalism, 13(2), 221-239.

[25] Ripollés, A. C. (2012). Beyond newspapers: News consumption among young people in the digital era. Comunicar. Media Education Research Journal, 20(2).

[26] Rogers, E. M. (2000). Reflections on news event diffusion research. Journalism & Mass Communication Quarterly, 77(3), 561-576.

[27] Rogers, E. M. (1962). Diffusion of innovations. Glencoe. Free Press.(1976)," New Product Adoption and Diffusion," Journal of Consumer Research, 2, 290-304.

[28] Rosen, J. (2012). The People Formerly Known as the Audience. Teoksessa Michael Mandiberg (toim.) The Social Media Reader.

[29] Rossman, G. (2014). The Diffusion of the Legitimate and the Diffusion of Legitimacy. Sociological Science, 1, 49.

[30] Saba, J. (2014). Beyond cute cats: How Buzzfeed is reinventing itself. Reuters, 23 February.

[31] Tandoc Jr, E. C. (2018). Five ways Buzzfeed is preserving (or transforming) the journalistic field. Journalism, 19(2), 200-216.

[32] Tandoc Jr, E. C., & Jenkins, J. (2017). The Buzzfeedication of journalism? How traditional news organizations are talking about a new entrant to the journalistic field will surprise you!. Journalism, 18(4), 482-500.

[33] Watts, D. J., Peretti, J., & Frumin, M. (2007). Viral marketing for the real world. Harvard Business School Pub.

[34] Wihbey, J. P. (2019). The social fact: News and knowledge in a networked world. Mit Press.

[35] Wolf, C., & Schnauber, A. (2015). News consumption in the mobile era: The role of mobile devices and traditional journalism's content within the user's information repertoire. Digital journalism, 3(5), 759-776.

[36] Wolf, C., & Hohlfeld, R. (2012). Revolution in Journalism?. In Images in mobile communication (pp. 81-99). VS Verlag für Sozialwissenschaften.