Content Virality on Online Social Networks: Empirical Evidence from Twitter, Facebook, and Google+ on German News Websites

Irina Heimbach
TU Darmstadt
Dept. of Economics and Law
heimbach@emarkets.tu-darmstadt.de

Benjamin Schiller
TU Dresden
Faculty of Computer Science
benjamin.schiller1@tu-dresden.de

Thorsten Strufe
TU Dresden
Faculty of Computer Science
thorsten.strufe@tu-dresden.de

Oliver Hinz
TU Darmstadt
Dept. of Economics and Law
hinz@emarkets.tu-darmstadt.de

ABSTRACT

The virality of content describes its likelihood to be shared with peers. In this work, we investigate how content characteristics impact the sharing likelihood of news articles on Twitter, Facebook, and Google+. We examine a random sample of 4,278 articles from the most popular news websites in Germany categorized by human classifiers and text mining tools. Our analysis reveals commonalities and subtle differences between the three networks indicating different sharing patterns of their users.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics; G.3 [**Probability and Statistics**]: Correlation and regression analysis

General Terms

Measurement; Economics; Human Factors

Keywords

Online social networks; Content virality; News articles

1. INTRODUCTION

The digitization and emergence of Internet-based services altered human communication in many respects. While we had to cut out or copy a newspaper article in the past to share it with others, we can easily send a link to an interesting article or a funny video to our friends, colleagues, and relatives via email or share them on *online social networks* (OSNs) like *Facebook*, *Twitter*, and *Google+* nowadays. Some content attracts high attention and becomes *viral*, exhibiting an almost contagious behavior [16, 5]. We define *content virality* as some feature of the content that enhances its likelihood to be shared in different communication channels, distinct from *content popularity* which is commonly measured by the total number of accesses to the content. For a discussion of alternative measures of virality, we refer to [16]. As argued by [16], virality is strongly related „to the content being spread, rather than to the influencers who spread it".

The phenomenon of online content sharing belongs to the general domain of diffusion processes, where the characteristics of the diffusing object - be it a product, an idea, a piece of information, or behavior - play an important role. While the impact of characteristics of physical products are well studied in this context, e.g., relative advantage and complexity as key factors of adoption [38], the characteristics of digital products are insufficiently investigated. The studies by Berger and Milkman [7] and Guerini et al. [16] constitute laudable exceptions and investigate the content characteristics in a structured manner. The study by Berger and Milkman [7] finds that positive and emotionally written New York Times articles are shared more frequently. Furthermore, in their sample, articles which evoke strong emotions like anger and awe are shared more often. Although they controlled for several factors and their findings are robust and confirmed in experiments and a field study, the research on drivers of content virality warrants further examinations. The study by Guerini et al. [16] proposes prediction models for different alternative measures of virality using support vector machines.

Studies by Szabo and Huberman [45], Berger and Iyengar [6], Aral and Walker [3], and Schulze et al. [42] investigate diffusion outcomes in context of different communication channels and design sharing mechanisms. Their findings suggest an interrelation between content characteristics and the characteristics of the channel, a relation that has not been studied so far.

Therefore, we intend to study the interplay between the shared content and the use of different channels. In this

HT'15, September 1–4, 2015, Guzelyurt, TRNC, Cyprus.
© 2015 ACM. ISBN 978-1-4503-3395-5/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2700171.2791032.

work, we investigate how content characteristics impact the sharing likelihood of news articles on the OSNs Twitter, Facebook, and Google+. We examine a random sample of 4,278 articles from the most popular news websites in Germany categorized by human classifiers and text mining tools.

Understanding the reasons why some content becomes popular is of high relevance for theory and business practice, as the media industry is challenged by the development of Internet-based services and has to face the transition of social life into the digital environment. First, more and more readers substitute printed magazines and newspapers with online content which is currently often still free of charge. But as more and more media companies like the New York Times [1] shift to freemium business models, knowledge about the drivers of content virality could be useful for sophisticated pricing strategies for online content [31]. Second, the authors might craft purposefully content that goes viral. Third, if the authors are not willing to adapt or even purposefully write content with the aim that it becomes popular and viral in different social media, the results of our research can provide then at least useful suggestions in selecting the appropriate channel for the message. For example if the content provider knows that Facebook users are more interested in funny and entertaining stories than in well-investigated profound polarizing articles, then it might focus its social media strategy on other, more suitable, OSNs than Facebook. The problem of selecting appropriate communication channels is becoming more relevant, as more instant messengers, OSNs and content aggregators emerge offering their own sharing plugins, e.g., like the messenger WhatsApp in 2014. Due to limited space on the website and especially for mobile ones, the content providers will be not able to place all social plugins around a content and will be forced - sooner or later - to deal with the problem of selecting the most appropriate communication channels for spreading their content online. Finally, understanding and knowing on which OSN a content will be shared more can allow content providers to match ads to the audience of that OSN [31].

The remainder of this paper is structured as follows: First, we review the existing research on information diffusion and content characteristics as a driver of such diffusion in Section 2. Then, we present our data set, our estimation model, variables, and the estimation approach in Section 3. In Section 4, we present and discuss the results of our analysis. Finally, in Section 5, we summarize our work and conclude with implications for research and business practice.

2. RELATED WORK

Previous research on content sharing and on the related fields of word of mouth (WOM) and information diffusion has a long research history. Previous research has shown the importance of WOM for increased sales [15, 9], for the adoption and discovery of new products and ideas [40, 11, 14, 13, 32, 5], and for other economic processes like on bidding behavior in online auctions [22]. Further, previous research tries to identify influential people in social networks [1, 20, 26], investigates how network structures impact social contagion [4, 30], and tries to determine how this information can be used to improve viral marketing strategies [27, 37, 47, 21]. Other research investigates temporal and topological patterns of information diffusion [33, 45, 50].

[1] available on `http://bit.ly/15xvO6J`

Individuals decide every time they receive a viral message, a piece of content, or some other information whether and to whom they forward it [36, 43]. Newer studies address the motives to transmit content [36, 19, 25, 43, 23, 6] and the characteristics of product or content [16, 7, 10, 8, 42]. Previous studies show that motives to share content are related to altruism [36, 23, 35] as many people want to help others. This becomes also evident as content with high practical utility is shared rather often [7].

Content characteristics for their part are the least studied component of the social communication. Phelps et al. [36] find that the most forwarded emails contain jokes and chain letters but they do not further differentiate between content characteristics. In contrast, the study by Berger and Milkman [7] classifies New York Times articles on several dimensions and provides deep insights into the main drivers of content virality. An important content factor is its sentiment. The findings on its impact however are ambiguous [17, 7, 2]. Mazzarol et al. [35] report that the received word of mouth is both positive and negative but always with rather extreme values. Rosen and Tesser [39] state that people are reluctant to send negative WOM preventing building negative attitude towards their personalities. Berger and Milkman [7] find that positive content is shared more often. Contradictory to Berger and Milkman [7], Hansen et al. [17] analyze three samples from Twitter and find that negative news-Tweets are more likely to be re-tweeted but the popularity of non-news-Tweets relates to a positive sentiment of the content. Angelis et al. [2] find in a series of laboratory experiments that people tend to generate positive WOM about their own experiences but to transmit negative news about the experiences of others.

Content sentiment is tightly related to the content's emotionality. Dobele et al. [12] analyze real viral marketing campaigns and show that the success of a campaign is highly related to the emotionality of its messages. According to Berger and Milkman [7] content evoking high-arousal emotions like awe and anger activate people to share it with their peers. Content related to deactivating emotions like sadness is shared less frequently [7].

In his seminal work, Rogers [38] has already highlighted the importance of the communication channel in disseminating a product or a new idea. Studies by Szabo and Huberman [45], Berger and Iyengar [6], Aral and Walker [3] and Schulze et al. [42] find differences in diffusion processes between different communication channels or the different design of sharing mechanisms within a communication channel. An unpublished study by Berger and Iyengar [6] finds that discontinuous communication channels (email or text posts) give the conversational partners an opportunity to select the most interesting topic or brand in contrast to continuous channels (phone and face-to-face) over which people discuss the topic which just crosses their mind. As digital content is often shared via OSNs using social plugins, the characteristics of each OSN might have an effect on the sharing probability. Aral and Walker [3] investigate how the sharing mechanism (broadcast vs. personalized referrals) impacts the reach of a Facebook application. The very recent study by Schulze et al. [42] evidences the existence of a mismatch between the product characteristics to be promoted over an OSN and the users' perceptions of the network. Szabo and Huberman [45] find that while content shared on Digg.com saturates very quickly, YouTube videos keep receiving attention for a long time. The findings of these three studies let suppose that content characteristics and the characteristics of an OSN could interact, a question that has not been studied yet. Further, the differences in use of OSNs might reflect their different user groups. Therefore, it is important to know who the users of the different OSNs are and how they can be characterized by content they share over the respective OSN.

Our study is distinct from the stream of research on predicting content popularity from the early users' reactions to content such as, e.g., the number of early votes on Digg.com and number of comments on discussion forums and sites [29, 31, 45]. [29] propose a model for prediction of content popularity on discussion forums dpreview.com and myspace.com using a Cox proportional hazard regression using the number of comments and the number of links in the first hours after publishing. Other studies in this field apply similar research designs. In contrast, we concentrate on inherent content characteristics like content's sentiment and emotionality, topic and which particular emotions it evokes. We also control for author characteristics like writing complexity and author's fame as well as for publishing timing like day-of-the-week and time-of-the-day.

Therefore, we perform a deeper analysis of the drivers of content virality in three different OSNs.

3. DATA, CODING, AND ESTIMATION

Our data set comes from a large-scaled, still ongoing project which collects data of all articles appearing in the most popular German online newspapers and magazines[2] since January 2012 [41]. Using web crawlers, we record an article's title, a link to the full text, the name of the magazine, and the category where it is published. For two weeks after an article's publication, web crawlers visit the web sites every 3 hours and capture the number of Tweets (Twitter), One-ups (Google+), and Shares (Facebook). Here, Tweets refers to the number of Tweets that contain an article's URL. For Facebook, Shares is the sum of Likes, Comments, and Shares that is commonly displayed. In addition, we record whether an article is published on the main page or on subpages only. Given the immense number of articles in our database (454,888 articles published between March 1st and September 30th, 2012), we drew a random sample of 4,278 (about 1%) for this study. Then, we coded articles with respect to several dimensions. Table 1 gives a brief description of the dimensions and the data sources.

Using SentiStrength [46][3], a German dictionary for automated sentiment analysis, we quantified the positivity and emotionality of articles. Positivity is defined as the difference between the shares of positive and negative words in an article [7]. Emotionality is quantified as the total share of positive and negative words [7].

Next, we engaged four human coders to classify articles on further dimensions. The coders were not informed about the research question and were asked to rate articles on the four emotional dimensions (anger, anxiety, awe, and sadness), on the dimensions surprise, practical utility, and interest, and to note the authorship of the articles. We used the coding instructions provided by Berger and Milkman [7][4]. Each article was encoded by one of the coders on a five-point Likert scale [34] according to the extent to which an article evokes certain emotions or might be, e.g., of practical relevance. We trained the coders on a test set of articles to ensure good inter-coder reliabilities of up 0.7 and higher (pairwise Holsti-Index [24]).

Further, we control for the number of pictures and videos included in the article as it may affect content virality. Articles featured with pictures and videos may benefit from increased attractiveness and thus might be shared more on OSNs as shown for user profiles in a business-related OSN [44].

Table 3 gives examples of articles which score highly on different content dimensions. As expected, an article about the new love of German television hostess Michelle Hunziger, is written positively. The disqualification of the German cycling team during the world cup has a high value for anger while a report about child mortality is classified as sad. These examples show that the characteristics we determined for the articles fits well with their respective content.

Author characteristics may also affect content virality. We distinguish the first author's gender since male and female authors might have different writing styles which can impact an article's popularity [7]. Well written articles are more likely to be read, and thus are more likely to be shared. Thus, we control for writing complexity measured by applying the Flesch-Reading-Ease metric, a scale that is so ubiquitous that it is bundled with popular word processing programs and services such as Microsoft Office Word, WordPerfect, and WordPro[5]. We additionally introduce a dummy variable that measures whether the article is based on the reports of news agencies. These articles tend to be very early common knowledge, so that readers believe that everybody is already aware of this information and they therefore refrain from sharing them. We also control for the author's reputation that may also influence an article's virality. We approximated the author's reputation by counting the search hits at the search engine *Bing*[6] for the author's name plus the keyword "author".

Furthermore, we control for position, time [50], and attention competition factors [49]. Most readers start reading online magazines from the home page. So articles which are published there, are more likely to be read and thus to be shared on OSNs. Therefore, we include a variable to describe whether an article appeared on the home page or not. We extracted sections an article was published from the URLs and compress them into 13 distinct categories. These sections are cars, career, society, culture, lifestyle, politics, local, travel, humor, sports, technology, business, and science. Furthermore, there might be time effects on the number of recommendations in OSNs. We control for the time when an article is published. From that, we create 8 dummy variables for 3-hours-periods and seven dummies for the day of the week. Articles which appear at 10:00 p.m. may draw less attention than articles which appear at 10:00 a.m. Similarly, articles which appear on a weekend might have less recommendations as users spend their spare time otherwise than on the Internet. Therefore, we represent the day of the week of an article's publication in 7 dummy variables. We further use a linear time trend that captures the steady growth of each OSN over time. Finally, we account for

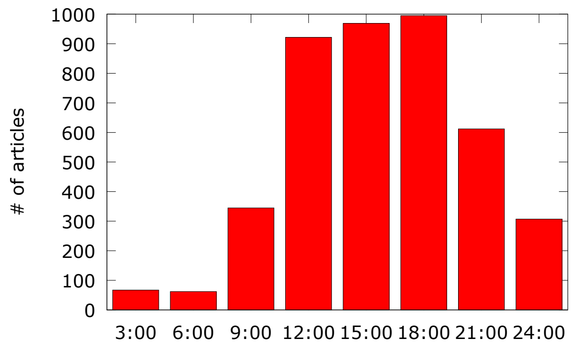
[2]available on `http://www.alexa.com/`

[3]available on `http://bit.ly/1xs2V1r`

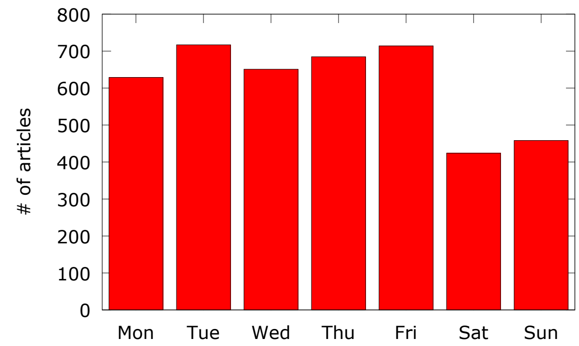
[4]available on `http://bit.ly/1xs35WF`

[5]available on `http://bit.ly/1IMRW9s`

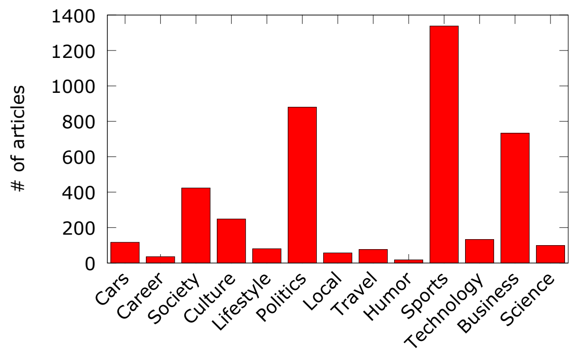
[6]available on `http://www.bing.com`



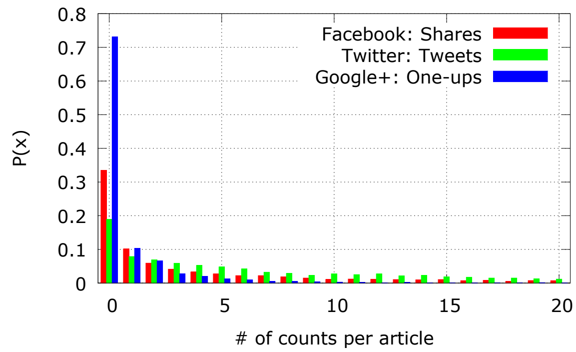
(a) Number of articles per time slot



(b) Number of articles per day of the week



(c) Number of articles per category



(d) Distribution of Share, Tweet and One-up count (for $x \leq 20$)

Figure 1: Statistics for the 4,278 randomly selected articles from German news websites

a magazine's reach and for the number of articles published on the same day. Magazines with higher reach are expected to have more readers who might share articles on OSNs. We completed our data by the moving average monthly reach of newspapers and magazines from the media agency AGOF[7]. As argued by [49], the number of articles is expected to negatively affect the number of recommendations on OSNs due to limited attention of people.

Summary statistics of our dependent variables and other dimensions are given in Table 2. Figure 1d shows the distributions of # of Facebook Shares, Twitter Tweets, and Google+ One-ups, which follow Zipf's law: A huge part of articles gained few recommendations and only few articles were highly shared with peers in the respective OSN. 59.49% of articles are based on reports of news agencies. 11.97% of articles are written by female first authors. 21.30% of articles are published on the start page of the respective news medium. Figures 1c, 1a, and 1b show the distributions of the selected articles along the dimensions article section, time of the day, and day of the week. Our random sample contains a disproportionately high share of sports-related articles, which is caused by the European Championship in Summer 2012. The distributions over the day-of-the-week and time-of-the-day dimensions are plausible.

As our dependent variables take on discrete, non-negative values (i.e. represent count data), we choose a negative binomial regression to analyze our research question. Methods developed for modeling count data (Poisson regression, negative binomial regression, zero inflation, hurdle models etc.) explicitly account for the particular aspects of such data like the preponderance of zeros, small values and the discrete nature of the dependent variables and therefore are better suited than the linear multiple regression approach [28]. We do not apply the Poisson regression model, as it assumes equality of the conditional mean and variance functions (equidispersion). The most common method is then negative binomial model that accounts for overdispersion (variances exceed means) of the data [28].

The estimation equation looks as follows:

$$Prob(Y = y_i|x_i) = \frac{e^{-\lambda_i} \cdot \lambda_i^{y_i}}{y_i!} \qquad (1)$$

with

$$\lambda_i = e^{x_i' \cdot \beta + \epsilon} \qquad (2)$$

where y_i measures the number of Shares, Tweets, and One-ups in the respective OSN for article i and x_i is the vector describing article i on the different dimensions. ϵ is a gamma distributed error with unity mean and variance α [28]. We used robust standard errors to account for heteroscedasticity in our data set.

4. RESULTS

The results (estimated coefficients, standard errors, and the respective significance levels (p-values) are listed in Table 4. As a reference category in our estimation, we selected an article written by a male first author and published in a

[7]available on `http://www.agof.de/`

Table 3: Exemplary articles which score highly on different dimensions

Variable	Original German Title (and English translation)
Positivity	Neues Glück für Michelle Hunziker: Mit Tomaso auf Wolke sieben *New felicity for Michelle Hunziker: with Tomaso on cloud nine*
Emotionality	Macaulay Culkin: Sein Vater fürchtet um sein Leben! *Macaulay Culkin: His father father fears for his life*
Anger	Radsport WM: Deutsche Teamsprinter disqualifiziert *Cycling worldcup: German team is disqualified*
Anxiety	Nach Fukushima: Japans Regierung erwägt ersten AKW-Neustart *Japan plans to re-start nuclear power plants*
Awe	Wie Guerilla-Gärtner illegal Städte begrünen *How Guerrilla gardeners plant greenery on cities*
Sadness	Bericht von Unicef: Kindersterblichkeit seit 1990 weltweit halbiert *Unicef report: Child mortality halved since 1990*
Surprise	Forderung nach Abschaffung des Paragrafen 173: Grünen-Politiker Ströbele will Inzest erlauben *German politician Ströbele wants to change law to legalize incest*
Practical Utility	Die zehn schönsten Wanderrouten *10 most beautiful hiking routes*
Interest	Chinesischer Jugendlicher: Eine Niere im Tausch für ein iPad *Chinese teenager exchanges kidney for an iPad*
First Author Fame	Ein Witz von Guido Knopp *A joke by Guido Knopp*
Writing complexity	EZB-Mitarbeiter fordern Inflationsschutz für Rente *European central bank employees demand protection of retirement pays from inflation*
Number of Pictures	18 deutsche Filme sind in Cannes am Start *18 German movies are part of Cannes*
Number of Videos	So drücken Sie Ihre Energie-Rechnung *A guide how to lower your energy bill*

culture section on Wednesday between 9:00 p.m. and midnight (Constant term in Table 4).

First, we observe a steady growth of sharing for Google+ ($p < .01$) and a slight decrease for Twitter ($p < .1$) which is captured by our linear time variable representing the publication time. The sharing activities on Facebook seem to stagnate. We also find some periodical patterns: On Google+ and Facebook, articles receive most recommendations on Fridays. For Twitter, there is no difference between Wednesdays (the reference day) and any other day of the week. Facebook users refrain from sharing online content on Sundays. With respect to the time-of-the-day effects there are no significant differences for Google+. Articles have a higher chance to be shared by Facebook users between 6:00 a.m. and 2:00 p.m. ($p < .01$) compared to the reference category published between 9:00 p.m. and midnight. Twitter users tweet less between 12:00 p.m. and 9:00 p.m. ($p < .01$).

With respect to the section where the article is listed, we observe some commonalities and some strong differences between the three OSNs. Articles related to career are likely to be shared on all three OSNs. In contrast, sports-related articles are less likely to be shared. Facebook and Twitter users are less likely to share articles about cars, for Google+ this effect is not significant. Whereas Twitter and Google+ users seem to be interested in business-related articles, Facebook users are less likely to share them with their peers ($p < .1$). Generally, Twitter and Google+ users seem to resemble each other with respect to their recommendation of articles related to politics, technology, and science compared to the reference category (culture). Google+ users also prefer to share lifestyle- and travel-related articles with their peers. Twitter users do not share funny articles with their peers (section humor). With exception of some aforementioned sections, for Facebook, there seem to be no differences in the sharing likelihood of articles in remaining sections. This finding might indicate the high heterogeneity and more equally distributed preferences of Facebook users.

With respect to different emotional and other content dimensions, we found no significant effect for positively written articles and negative effect for emotionality ($p < .01$) in all three OSNs. It seems that German readers do not value very emotional articles highly. This finding can be caused by cultural differences or the fact that good journalism pursues goals and practices, such as objectivity, neutrality, and fact verification [48]. As the effect of the message's sentiment is controversially discussed in the previous research [2, 39, 7, 17], future research should try to identify the reasons for this dissenting result.

In line with previous research [7], we do find that sadness has a negative impact on sharing in Google+ and Twitter ($p < .05$). We further find that anger ($p < .01$) is a good predictor for the virality of the content in all three OSNs. If anger increases by 1 unit, the log expected number of Tweets, One-ups, and Shares increases by 0.119, 0.26, and 0.283 respectively. Awe has an impact on sharing via Facebook, and surprise on Google+. The users of all three OSNs value interesting articles ($p < .01$ for Twitter and Facebook, $p < .05$ for Google+). Contradictory to our expectations and in line with [42], Facebook users share less articles with practical utility ($p < .1$) while the effects for Twitter and Google+ are insignificant.

Table 4: Estimation results of negative binomial regression, standard errors in parentheses ($*$: $p < .10$, $**$: $p < .05$, $***$: $p < .01$)

Dimension	**Tweets**			**One-ups**			**Shares**		
Positivity	0.031	(0.030)		0.047	(0.088)		0.012	(0.055)	
Emotionality	-0.106	(0.018)	***	-0.191	(0.050)	***	-0.114	(0.024)	***
Anger	0.119	(0.021)	***	0.260	(0.046)	***	0.283	(0.047)	***
Anxiety	-0.004	(0.025)		0.040	(0.050)		-0.033	(0.055)	
Awe	0.013	(0.022)		0.029	(0.049)		0.158	(0.055)	***
Sadness	-0.048	(0.022)	**	-0.124	(0.051)	**	-0.052	(0.049)	
Surprise	0.028	(0.023)		0.088	(0.048)	*	0.076	(0.050)	
Practical utility	0.015	(0.022)		-0.008	(0.044)		-0.106	(0.054)	*
Interest	0.070	(0.023)	***	0.106	(0.047)	**	0.185	(0.051)	***
# of pictures	0.007	(0.002)	***	0.006	(0.005)		0.020	(0.007)	***
# of videos	-0.038	(0.033)		-0.190	(0.088)	**	-0.066	(0.095)	
Female first author	-0.297	(0.066)	***	-0.172	(0.120)		0.027	(0.179)	
Writing complexity	-0.007	(0.001)	***	-0.013	(0.003)	***	-0.014	(0.003)	***
News agency	-0.653	(0.048)	***	-0.828	(0.102)	***	-0.972	(0.117)	***
First author fame	1.81e-5	(8.91e-6)	**	4.03e-5	(2.40e-5)	*	6.6e-5	(2.27e-5)	***
Position	-0.120	(0.044)	***	0.357	(0.107)	***	0.440	(0.120)	***
Section: cars	-0.622	(0.138)	***	-0.244	(0.312)		-0.934	(0.414)	**
Section: career	0.627	(0.165)	***	0.742	(0.292)	**	0.652	(0.357)	*
Section: society	-0.072	(0.099)		0.242	(0.196)		-0.256	(0.302)	
Section: lifestyle	0.113	(0.168)		0.876	(0.284)	***	0.142	(0.406)	
Section: politics	0.302	(0.095)	***	0.495	(0.175)	***	-0.072	(0.276)	
Section: local	-0.031	(0.300)		0.725	(0.662)		1.020	(0.673)	
Section: travel	0.066	(0.130)		0.576	(0.348)	*	0.027	(0.379)	
Section: humor	-0.689	(0.260)	***	-0.338	(0.408)		-0.727	(0.541)	
Section: sports	-0.876	(0.0928)	***	-1.067	(0.196)	***	-1.801	(0.273)	***
Section: technology	0.865	(0.122)	***	1.915	(0.225)	***	0.229	(0.377)	
Section: business	0.360	(0.090)	***	0.409	(0.175)	**	-0.467	(0.270)	*
Section: science	0.281	(0.144)	*	1.302	(0.252)	***	0.372	(0.421)	
Publication time	-5.48e-4	(3.30e-4)	*	0.002	(7.07e-4)	***	1.71e-4	(7.68e-4)	
Time of day: 03:00	-0.151	(0.137)		-0.427	(0.333)		-0.186	(0.420)	
Time of day: 06:00	-0.208	(0.169)		-0.261	(0.353)		-0.362	(0.440)	
Time of day: 09:00	0.107	(0.103)		0.170	(0.218)		0.781	(0.221)	***
Time of day: 12:00	-0.235	(0.0912)	**	0.139	(0.186)		0.515	(0.189)	***
Time of day: 15:00	-0.304	(0.090)	***	-0.068	(0.190)		0.371	(0.197)	*
Time of day: 18:00	-0.340	(0.089)	***	-0.181	(0.183)		0.017	(0.187)	
Time of day: 21:00	-0.318	(0.096)	***	-0.002	(0.203)		0.343	(0.212)	
Day of week: Mon	-0.071	(0.067)		0.095	(0.152)		-0.0991	(0.176)	
Day of week: Tue	0.101	(0.072)		0.398	(0.144)	***	0.263	(0.169)	
Day of week: Thu	-0.051	(0.064)		0.073	(0.142)		0.215	(0.155)	
Day of week: Fri	0.003	(0.067)		0.310	(0.147)	**	0.435	(0.161)	***
Day of week: Sat	-0.116	(0.126)		0.181	(0.290)		-0.292	(0.314)	
Day of week: Sun	-0.161	(0.119)		-0.090	(0.276)		-0.718	(0.301)	**
Constant	3.078	(0.303)	***	-0.564	(0.692)		3.892	(0.808)	***
# of articles	-3.8e-4	(1.25e-4)	***	-4.39e-4	(3.01e-4)		-0.001	(3.39e-4)	***
Monthly reach	1.2e-4	(7.49e-6)	***	1.1e-4	(1.64e-5)	***	2.0e-4	(1.88e-5)	***
$Ln\ \alpha$	0.090	(0.031)	***	1.203	(0.058)	***	1.293	(0.029)	***
N		4278			4278			4278	
Pseudo R^2		0.071			0.102			0.055	

Table 1: Dimensions, sources, and values for data coding

OSN Statistics		
# of Tweets	Web crawler	$\mathbb{N}_0^+$
# of One-ups	Web crawler	$\mathbb{N}_0^+$
# of Shares	Web crawler	$\mathbb{N}_0^+$
Content characteristics		
Positivity	SentiStrength	$\mathbb{R}$
Emotionality	SentiStrength	$\mathbb{R}^+$
Anger	Human classifier	$\{1 \dots 5\}$
Anxiety	Human classifier	$\{1 \dots 5\}$
Awe	Human classifier	$\{1 \dots 5\}$
Sadness	Human classifier	$\{1 \dots 5\}$
Surprise	Human classifier	$\{1 \dots 5\}$
Practical utility	Human classifier	$\{1 \dots 5\}$
Interest	Human classifier	$\{1 \dots 5\}$
Article features		
# of pictures	Human classifier	$\mathbb{N}_0^+$
# of videos	Human classifier	$\mathbb{N}_0^+$
Author characteristics		
Female first author	Human classifier	$\{0, 1\}$
Writing complexity	Flesch Reading Test	$\{0 \dots 100\}$
News agency	Human classifier	$\{0, 1\}$
First author fame	Bing search	$\mathbb{N}_0^+$
Attention competition		
Position	Web crawler	$\{0, 1\}$
Section (13)	Web crawler	$\{0, 1\}$
Publication time	Web crawler	Datetime
Time of day (8)	Web crawler	$\{0, 1\}$
Day of week (7)	Web crawler	$\{0, 1\}$
Monthly reach	Media agency	$\mathbb{N}_0^+$
# of articles	Web crawler	$\mathbb{N}_0^+$

Table 2: Dataset statistics for selected dimensions

Dimension	Mean	Std. dev.
# of Tweets	11.88	19.90
# of One-ups	0.99	3.86
# of Shares	36.19	175.42
Positivity	-0.15	0.89
Emotionality	1.40	2.16
Anger	2.28	1.16
Anxiety	1.98	1.07
Awe	1.72	0.90
Sadness	2.04	1.12
Surprise	2.47	1.02
Practical utility	1.77	0.96
Interest	3.02	0.92
# of pictures	3.37	9.14
# of videos	0.12	0.45
Writing complexity	67.64	15.37
First author fame	571.34	1967.92
# of articles	1859.38	336.09
Monthly reach	6205.76	2915.20

We find support that article complexity has a positive impact on virality, which is in line with the study by Milkman and Berger [7]. We also find support for the notion that the first author's fame and the article location on the website is a strong, positive predictor for virality in all three OSNs. In line with the study by [44], we find that the number of pictures used in the article has a positive impact on the sharing likelihood in Facebook and in Twitter ($p < .01$). We also controlled for the source of the article's content and used a dummy variable for news agencies as source. If news agencies are the source of the information, the article is less frequently shared in all three OSNs ($p < .01$) as we already hypothesized.

Our results highlight that the predictors for virality are more complex than previous research suggests and that there are subtle differences between different communication channels. Hence, more research is needed in this area to better understand the phenomenon of virality.

5. SUMMARY AND CONCLUSION

Although there is a huge bulk of literature on word of mouth, social contagion, and viral marketing, research on the impact of content characteristics on sharing behavior is scarce. We therefore investigate how content characteristics impact the sharing likelihood of news articles in three different OSNs: Twitter, Facebook, and Google+. We examined a random sample of 4,278 articles from the most popular news and magazine websites in Germany published in 2012. We used human classifiers as well as text mining tools to categorize the content and enriched the data with control variables.

Our analysis reveals commonalities and subtle differences between the three OSNs examined which highlights that different sharing patterns should be expected from different audiences. We find that emotionality has a negative effect on virality which means that articles without strong emotions are more frequently shared in OSNs. Articles with interesting and anger evoking content go viral in all three OSNs. In line with previous research, we find that sadness is negatively related to content virality in Twitter and Google+ while awe positively influences the likelihood of articles to be shared on Facebook. Articles with high practical utility are less shared on Facebook. Twitter and Google+ users are more likely to share articles related to politics, business, technology, and science. Google+ users are more likely to share articles related to lifestyle and career. In contrast, articles about sports are less likely to be shared in all three OSNs. Finally, we find that articles based on the reports of news agencies are less likely to go viral in all three OSNs, indicating that users prefer original articles.

Our results are valuable for journalists as they can use the results to write more viral articles. If the journalists do not know how to write viral articles purposefully, publishers could at least adapt their social media strategies. As we mention in the introduction, content providers face the problem of selection of appropriate communication channels due to emergence of new sharing, communication and networking services. As our results indicate, Google+ and Twitter users are more interested in technology, science, business and politics related news and content. Content providers might then adapt their website design and prominently feature such content with social plugins from Google+ and Twitter. Otherwise, the spreading of the content that is less appealing to Google+ and Twitter users might be fostered in other, more suitable, OSNs. Most content providers start now to apply systems which track users' journeys (from where a user comes and where she goes) allowing to automate their marketing decisions [18]. Based on the insights provided by our study, publishers could also develop prediction tools that automatically optimize the placement of advertisements which ultimately could increase profits [31, 18].

Our study does not come without limitations. First and most importantly, sharing behavior does not necessarily relate to reading behavior. For example, some readers may shy away from sharing articles that evoke sad feelings although it might be interesting to read them. Therefore future research could study the interrelation between reading and sharing content. Secondly, we can only give advice from a macro perspective. There might be segments within the user population that show completely different patterns of sharing. The high variances and standard errors in our results suggest a more sophisticated analysis. Researchers and practitioners might especially be interested in understanding the role of opinion leaders as they ignite diffusion processes. Thirdly it seems fruitful to understand the relation between individual characteristics and content sharing behavior. It is likely that for example narcissistic people share other content than introverted people. To examine this research question, it would however be necessary to have information on sharing behavior and on the psychographics of the sharing individuals.

In conclusion, there are a number of potentially interesting avenues for future research. The intersection between psychology, sociology, and computer science provides promising new questions that also have value for business practice.

6. REFERENCES

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *Proceedings of the 2008 international conference on web search and data mining*, pages 207–218. ACM, 2008.

[2] M. D. Angelis, A. Bonezzi, A. M. Peluso, D. D. Rucker, and M. Costabile. On braggarts and gossips: a self-enhancement account of word-of-mouth generation and transmission. *Journal of Marketing Research*, 49(4):551–563, 2012.

[3] S. Aral and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.

[4] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace. The effects of the social structure of digital networks on viral marketing performance. *Information Systems Research*, 19(3):273–290, 2008.

[5] J. Berger. *Contagious: Why things catch on.* Simon and Schuster, 2013.

[6] J. Berger and R. Iyengar. How interest shapes word-of-mouth over different channels. *Available at SSRN 2013141*, 2012.

[7] J. Berger and K. L. Milkman. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205, 2012.

[8] J. Berger and E. M. Schwartz. What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, 48(5):869–880, 2011.

[9] J. A. Chevalier and D. Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.

[10] C. M. Chung and P. R. Darke. The consumer as advocate: self-relevance, culture, and word-of-mouth. *Marketing Letters*, 17(4):269–279, 2006.

[11] J. Coleman, E. Katz, and H. Menzel. The diffusion of an innovation among physicians. *Sociometry*, pages 253–270, 1957.

[12] A. Dobele, A. Lindgreen, M. Beverland, J. Vanhamme, and R. Van Wijk. Why pass on viral messages? because they connect emotionally. *Business Horizons*, 50(4):291–304, 2007.

[13] R. Garg, M. D. Smith, and R. Telang. Measuring information diffusion in an online community. *Journal of Management Information Systems*, 28(2):11–38, 2011.

[14] M. Gladwell. *The tipping point: How little things can make a big difference*. Hachette Digital, Inc., 2006.

[15] D. Godes and D. Mayzlin. Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4):545–560, 2004.

[16] M. Guerini, C. Strapparava, and G. Özbal. Exploring text virality in social networks. In *ICWSM*, 2011.

[17] L. K. Hansen, A. Arvidsson, F. Å. Nielsen, E. Colleoni, and M. Etter. Good friends, bad news-affect and virality in twitter. In *Future information technology*, pages 34–43. Springer, 2011.

[18] I. Heimbach, D. S. Kostyra, and O. Hinz. Marketing automation. *Business & Information Systems Engineering*, 57(2):129–133, 2015.

[19] T. Hennig-Thurau, K. P. Gwinner, G. Walsh, and D. D. Gremler. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52, 2004.

[20] S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, pages 256–276, 2006.

[21] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker. Seeding strategies for viral marketing: an empirical comparison. *Journal of Marketing*, 75(6):55–71, 2011.

[22] O. Hinz and M. Spann. The impact of information diffusion on bidding behavior in secret reserve price auctions. *Information Systems Research*, 19(3):351–368, 2008.

[23] J. Y. Ho and M. Dempsey. Viral marketing: Motivations to forward online content. *Journal of Business Research*, 63(9):1000–1006, 2010.

[24] O. R. Holsti. *Content analysis for the social sciences and humanities*. Addison-Wesley Reading, MA, 1969.

[25] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[26] Z. Katona, P. P. Zubcsek, and M. Sarvary. Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, 48(3):425–443, 2011.

[27] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[28] P. Kennedy. *A guide to econometrics*. MIT press, 2003.

[29] J. G. Lee, S. Moon, and K. Salamatian. An approach to model and predict the popularity of online contents with explanatory factors. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 623–630. IEEE, 2010.

[30] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *ICWSM*, 10:90–97, 2010.

[31] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proceedings of the 19th international conference on World wide web*, pages 621–630. ACM, 2010.

[32] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506. ACM, 2009.

[33] J. Leskovec, M. McGlohon, C. Faloutsos, N. S. Glance, and M. Hurst. Patterns of cascading behavior in large blog graphs. In *SDM*, volume 7, pages 551–556. SIAM, 2007.

[34] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[35] T. Mazzarol, J. C. Sweeney, and G. N. Soutar. Conceptualizing word-of-mouth activity, triggers and conditions: an exploratory study. *European Journal of Marketing*, 41(11/12):1475–1494, 2007.

[36] J. E. Phelps, R. Lewis, L. Mobilio, D. Perry, and N. Raman. Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *Journal of advertising research*, 44(4):333–348, 2004.

[37] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 61–70. ACM, 2002.

[38] E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[39] S. Rosen and A. Tesser. Fear of negative evaluation and the reluctance to transmit bad news. *Journal of Communication*, 22(2):124–141, 1972.

[40] B. Ryan and N. C. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology*, 8(1):15–24, 1943.

[41] B. Schiller, I. Heimbach, T. Strufe, and O. Hinz. Development of social network usage in germany since 2012. Technical report, Technische UniversitÃd't Darmstadt, 2014.

[42] C. Schulze, L. Schöler, and B. Skiera. Not all fun and games: Viral marketing for utilitarian products. *Journal of Marketing*, 78(1):1–19, 2014.

[43] A. T. Stephen and D. R. Lehmann. Why do people transmit word-of-mouth? the effects of recipient and relationship characteristics on transmission behaviors. Technical report, working paper, Columbia University, 2009.

[44] T. Strufe. Profile popularity in a business-oriented online social network. In *Proceedings of the 3rd Workshop on Social Network Systems*, page 2. ACM, 2010.

[45] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, 2010.

[46] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[47] M. Trusov, A. V. Bodapati, and R. E. Bucklin. Determining influential users in internet social networks. *Journal of Marketing Research*, 47(4):643–658, 2010.

[48] Y. Tsfati, O. Meyers, and Y. Peri. What is good journalism? comparing israeli public and journalists' perspectives. *Journalism*, 7(2):152–173, 2006.

[49] F. Wu and B. A. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601, 2007.

[50] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 177–186. ACM, 2011.