

The Impact of Content, Context, and Creator on User Engagement in Social Media Marketing

Roope Jaakonmäki
University of Liechtenstein
roope.jaakonmaeki@uni.li

Oliver Müller
IT University of Copenhagen
olmy@itu.dk

Jan vom Brocke
University of Liechtenstein
jan.vom.brocke@uni.li

Abstract

Social media has become an important tool in establishing relationships between companies and customers. However, creating effective content for social media marketing campaigns is a challenge, as companies have difficulty understanding what drives user engagement. One approach to addressing this challenge is to use analytics on user-generated social media content to understand the relationship between content features and user engagement. In this paper we report on a quantitative study that applies machine learning algorithms to extract textual and visual content features from Instagram posts, along with creator- and context-related variables, and to statistically model their influence on user engagement. Our findings can guide marketing and social media professionals in creating engaging content that communicates more effectively with their audiences.

1. Introduction

Over the past decade, social media has become a popular channel through which to strengthen customers' relationships with products, brands, and companies [20], [22], [27]. In a recent survey of 3,700 marketers, 96 percent of respondents answered that they use social media for marketing [36].

However, as the number of end users and marketers who are active on social media increases [35], it becomes increasingly difficult for companies to stand out from the crowd enough to engage their target audiences. In fact, 91 percent of marketers struggle to answer the question concerning the best ways to engage their target audiences on social media platforms [36]. What's more, measuring the impact of social media marketing campaigns is challenging, as is calculating such campaigns' return on investment [23].

In order to assess the success of social media marketing activities, marketers typically measure the

rate at which users engage with their posts. The engagement rate measures the quantity of responses and interactions that content on social media generates from users [4], [17], [31], [38]. How the engagement rate is calculated varies across social media platforms, but it generally measures the percentage of people who react to a post in some way, such as by "liking" it or commenting on it.

The factors that drive social media engagement can be divided broadly into three groups: those that are related to the post's creator (e.g., the creator's sex, age, number of followers) [24], [21]; the post's context (e.g., time, location) [16], [41]; and certain features of the content, such as, textual content (e.g., words, tags), visual content (e.g., images, videos), and audio content. While researchers have applied various methods to study how users engage with textual content [2], [6], [8], [9], [15], [21], [24], [26], [28], [34], [38], only a few have focused on posts' visual content [4], [5], [24].

Against this background, we follow a holistic approach to study engagement in social media marketing by statistically modeling the influence on user engagement of the textual and visual features of content on user engagement while controlling for features related to creator and context. We use machine-learning algorithms to extract the textual and visual features of content from a dataset of more than 13,000 Instagram posts from professional bloggers and to identify the most important features with regards to user engagement. To the best of our knowledge, our study is among the first to use a data-analytic approach to identify automatically the most significant features that drive social media engagement. Our results can help social media marketers and users understand the most effective approach to engaging social media communities.

The remainder of this paper is structured as follows. Section 2 gives a brief review of social media marketing-related topics and summarizes existing research on drivers of engagement on social media. Section 3 explains the methodology and introduces the dataset. Section 4 presents our empirical results, while

Section 5 discusses our findings, implications, and limitations. Section 6 concludes with suggestions for future research.

2. Background

2.1. Social media and influencer marketing

2.1.1. Social media marketing. In 2016 the number of social network users reached 2.22 billion, a number that is expected to increase to 2.72 billion by 2019 [35]. Because social media provides an inexpensive way to interact and engage with these large numbers of potential customers, social media marketing has become a valuable channel for marketers [42]. The purposes of using social media marketing include branding, promotion, market research, customer service, and customer relationship management activities [11], [23], [42].

Social media marketing has three levels of maturity [42]: *trial*, which includes testing various platforms but not yet considering them an integral part of the company's marketing mix; *transition*, where social media marketing activities are somewhat unplanned but are becoming more systematic; and the *strategic phase*, in which marketers have a formal process to plan and execute social media marketing activities with clearly defined objectives and metrics.

The effectiveness of social media marketing is typically measured using proxies rather than monetarily, as linking social media marketing activities to key financial indicators is difficult [23]. Depending on the goal, these proxy measures can include web traffic generated, clicks, repeat visits, number of new followers, search volume, mentions in other social media channels, and peer-to-peer recommendations [10]. Paine [31] suggests using engagement as a key metric, dividing engagement into different phases, starting with clicking and liking, continuing with commenting, following, re-tweeting, and hash-tagging, and finally evolving into advocacy.

2.1.2. Influencer marketing. Companies and marketers use social media platforms not only to push information about products to customers but also as a medium for customer-to-customer communication about product-related information, opinions, attitudes, and purchase and post-purchase experiences [29]. In fact, user-generated social media content has evolved into a major factor in influencing consumer behavior over the last years [23]. Therefore, it is not surprising that marketing concepts like word-of-mouth (WOM) and influencer marketing are gaining popularity among social media marketers. WOM can be defined as “the

act of consumers talking among themselves about a product or service” [39, p. 280], while influencer marketing can be seen as “the practice of identifying key decision makers in a target audience and encouraging them to use their influence to spread WOM” [39, p. 277]. Thus, an influencer is a third party who significantly shapes the opinions and purchasing decisions of other customers [7]. For example, influencers may post photos of themselves with products or brands on a social media platform, accompanied by brand-related hashtags, and be paid or receive a free product from the brand in return as a compensation. Influencers are often popular and well-connected on social media. Although high popularity and connectedness do not guarantee that a person has significant influence and vice-versa [33], these qualities are essential for influencers.

One of the main platforms for WOM and influencer marketing is Instagram, which reached 400 million users in 2015 [25]. According to a social media marketing industry report, Instagram increased its position most significantly among the top platforms used by experienced social media marketers, increasing from 28 percent in 2014 to 36 percent in 2015 [36]. Moreover, 52 percent of marketers are planning to increase their Instagram marketing activities in the near future [36]. Instagram data is also suitable for analyzing the influence of content, since the posts consist of both picture and text.

2.2. What drives social media engagement

In a survey of more than 1,500 marketers [32], 72 percent stated that their top social media priority is to create more engaging content, and their second highest priority (65%) is to improve their understanding of what content is effective. These priorities are aligned with our research aim to identify the factors that drive engagement in social media marketing. Several researchers have addressed particular aspects of this question from a variety of perspectives, but holistic research about what characterizes influential post is still scarce. We divide the features that may influence engagement into three categories—creator, context, and content—and elaborate in the following sections on the current state of research in these areas.

2.2.1. Creator-related features. Many researchers have studied creator-related features (e.g., the creator's number of followers, age, sex) for specific social media communities. For example, Suh et al. [38] found (not surprisingly) that the number of both followers and followees affects the number of times a tweet is retweeted on Twitter. Experience and age also

influence engagement. Arguello et al. [2] found, for example, that posts on online communities were less likely to get a reply if newcomers wrote them. The same seems to hold for Twitter, as the age of a Twitter account increases the number of retweets [38].

The gender of the account holder is another factor that influences engagement on social media platforms. Gilbert et al. [21], for example, discovered in their study of Pinterest users that females get more repins than men, although male Pinterest account holders attract more followers than women do [21].

2.2.2. Contextual features. Most of the research done on contextual features (e.g., time, location) has been conducted and published by practitioners. For example, TrackMaven analyzed the Instagram posts of 123 companies that are on the US Fortune 500 list [41] and found that Sunday is the most effective day of the week for posting and that the time of the day does not have a significant effect on the number of interactions. Similarly, Ellering gathered and analyzed sixteen social media studies and found no best time to post [16].

2.2.3. Content features. Content features can be divided into the categories of text, visual, and audio content. Several researchers have studied the textual content's effect on popularity. Berger and Milkman [6] analyzed *New York Times* articles and found that messages that include high-arousal positive emotions (awe) and negative emotions (anger or anxiety) are more likely to go viral than is content with other types of emotions. Similarly, Lee et al. [28] observed that a message that includes persuasive content (e.g., emotional and philanthropic) increases engagement, while informative content (e.g., product prices, availability, or features) reduces engagement when used separately but increases engagement when combined with some persuasive content.

Burke et al. [8] studied textual discourse in the online community context and discovered that a short group or topic introduction in messages increases community response. In a later study, Burke and Kraut [9] found that politeness increases the number of replies in technical groups, but rudeness is more effective in generating replies in political groups. Arguello et al. [2] reported that posting on topic, introducing oneself, asking questions, and using simple language and shorter text increased replies in an online community.

Another prevailing trend in social media is the use of hashtags and URLs. For example, hashtags and URLs on Twitter have strong relationships with retweets [38]. However, a study about Facebook content found that the number of links in a post

decreases the number of comments [34]. According to TrackMaven, posts by Fortune 500 companies that use more than eleven hashtags provide the most interactions [41].

Comparing social media posts with and without visual content, Adobe found that posts with images perform the best in engaging the audience in social media [1]. Bakhshi, Shamma, and Gilbert [4] performed probably the first study to look at which visual features of a social media posts drive engagement. Their research indicated that pictures that include a human face are significantly more likely to receive likes and comments than are photos without a face. They also found that the number of faces in the photo and the persons' age and gender do not influence engagement. In another study, Bakhshi et al. [5] found that filtered photos attract more views and comments than those without alterations. More specifically, another study found that different Instagram filters have different effects on the engagement rate [41].

Practitioners have generated other findings on visual features. For instance, the social media marketing analytics company Curalate [14] analyzed eight million Instagram photos and discovered that using light instead of dark images, blue as the dominant color instead of red, duck-face selfies instead of realistic selfies, low saturation instead of vibrant colors, and a single dominant color instead of multiple dominant colors generates more likes. Moreover, Nielsen Norman Group found that users pay more attention to photos with real people, big photos, and images that carry information and tend to ignore images that are too stimulating [30].

Finally, one widely acknowledged finding in the marketing industry is that the gender and physical attractiveness of a model in an ad seem to influence people's perception of the ad and the marketed product [3]. One might assume that the same findings also apply in the social media marketing context.

To the best of our knowledge, the audio features of social media content have not yet been studied systemically.

3. Methods and data

We follow a quantitative approach to investigating the relationship between creator-related, contextual, and content features of Instagram posts and engagement. While engagement can be quantified in many ways, depending on the social media platform, we measure engagement as the sum of likes and comments. The number of likes indicates the extent of interest and approval, and the number of comments signals the level of verbal interaction, which also

signals user interest. In this section, we describe how we collected and analyzed the data, and present a statistical overview of our dataset.

3.1. Research process and data sampling

This study exploits an Instagram dataset from an anonymous German marketing and advertising company to determine which factors are most influential on user engagement. Our research process (Figure 1) started with collecting, processing, and cleaning up the dataset, which consisted of a random sample of Instagram posts. Then we gathered available creator-related, contextual, and content variables from the sample. Next, we filtered out the least common variables to reduce the number of variables and created a data frame for the regression. After filtering the variables, we performed a least absolute shrinkage and selection operator (LASSO) regression analysis on the data frame to identify the most influential features. Finally, we interpreted the results and compared them with the findings from existing research.

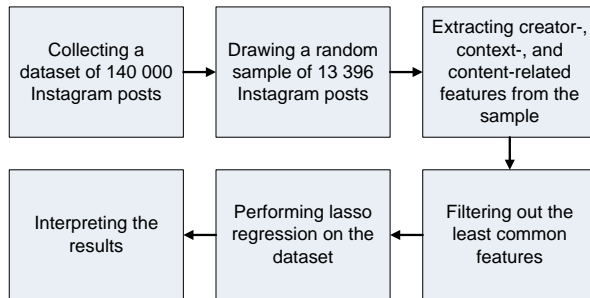


Figure 1. Research process

3.2. Extraction of features

We extracted creator- and context-related features from the metadata provided by the Instagram API. The creator-related data included variables like gender, age, country of residence, number of followers, and number of previous posts. Context-related features included the time and date of the post.

To extract the content-related features—that is, the text and visual features—we wrote several Python scripts using the Natural Language Tool Kit (NLTK) and the Clarifai Image Recognition API.

For the textual features, we extracted the type and number of words (including hashtags and URLs) and emojis (a symbol expressing an emotion or an idea in electronic messages) used in the posts’ caption fields. In order to reduce the dimensionality of the resulting feature vectors, we considered only those words and

emojis that appeared in at least 1 percent of the posts, which resulted in 312 words and 114 emojis.

To capture the visual features, we used the Clarifai image recognition API, which uses convolutional neural networks to learn complex representations of patterns in images [12], [43]. Clarifai’s API currently consists of more than 11,000 classifiers, including objects (e.g., car, house, river, man, woman), ideas (e.g. education, love, leisure), and feelings (e.g. beautiful, fun) [13]. The API is known for its accuracy, which has been reported to be around 89.3 percent [12]. Figure 2 shows a picture that illustrates the output from classifying a picture with the Clarifai API. Using the Clarifai API for the 13,396 Instagram photos generated 2,061 unique classes, out of which we kept only the 250 most frequent for the regression analysis.



Figure 2. Picture¹ and suggested classifications according to Clarifai API: *water, woman, summer, travel, leisure, sea, relaxation, vacation, young, enjoyment, recreation, fun, ocean, one, beach, girl, tropical, outdoors*

3.3. Regression analysis

After extracting the textual and visual content features, along with the metadata for the creator and context features, we created a data frame (13,396 rows by 768 columns) to serve as input for the subsequent regression analysis. Because of our dataset’s high level of dimensionality, we chose to use LASSO regression, which was first introduced by Tibshirani [40]. LASSO is a linear regression method that performs variable selection by shrinking the coefficients of uninfluential independent variables to exactly zero, which produces a model that includes only the most important independent variables in explaining the dependent variable [18], [40]. Model fitting was performed using the “glmnet” package for R [19].

¹ <https://www.instagram.com/p/BJ00zDTgeBk/?hl=en>

Although the LASSO technique is an advanced regression method that works well with high numbers of features [40], it has some limitations. In particular, when a group of independent variables shows high correlation, LASSO tends to pick one and set the others to zero [18], [44], which may hinder the model's interpretability. We tried other variable-selection and regularization approaches (e.g., elastic net and ridge regression), but LASSO produced the best fit and contained the lowest number of predictors.

3.4. Summary statistics of the dataset

Table 1 provides a statistical summary of some of the independent variables of our regression model. In particular, it presents the most often used words and emojis and the most common image classes.

Table 1. Most common words, emojis, and image classes

Word	Freq.	Emoji	Freq.	Image class	Freq.
"E"	1 875	😍	2 570	Woman	5 539
"Love"	1 469	💕	2 028	People	4 988
"Heute" (today)	1 075	❤️	2 018	Adult	4 497
"Wa"	953	👩	1 268	Fashion	3 424
"ootd"	881	😊	1 218	Portrait	3 399
"Mal" (time)	844	✨	1 178	No person	2 583
"Fashion"	842	💕	1 141	One	2 394
"Schon" (already)	841	💋	1 076	Wear	2 242
"Outfit"	832	😘	939	Indoors	2 154
"Happy"	771	😂	934	Girl	2 124

Table 2 provides additional information about the Instagram posts in our dataset. An average Instagram post received 1,500 likes and thirty-five comments (i.e., 50 times more likes than comments) and was posted by a blogger with almost 60,000 followers. (Our dataset stems from professional bloggers.) The caption of the post averaged twenty-seven words and three emojis, and the Clarifai API detected an average of eight classes in the picture.

Table 2. Summary statistics of Instagram posts

Variable	Mean
Likes	1 546.04
Comments	35.26
Followers	58 185.40
Word count	27.31
Emoji count	3.23
Image classes	7.85

More than 80 percent of the posts were created by female bloggers. The most common day to post was Sunday, and the most common time to post was between 8:00 p.m. and 9:00 p.m.

4. Results

The core functionality of LASSO regression is that it can automatically perform variable selection and explicate the tradeoff between highly accurate models with many predictors and less accurate models with fewer predictors. The plot in Figure 3 visualizes this tradeoff. Each curve in the plot corresponds to one predictor and the value of its coefficient (y-axis), whereas the x-axis represents the amount of deviance of the dependent variable that can be explained by the predictors [19]. The plot shows, for example, that 40 percent of the deviance in engagement can be explained by only 10 predictors, whereas increasing the explanation to 50 percent% of the deviance requires 381 predictors. The full model, with all 768 predictors, explains 51.39 percent of the deviance in engagement.

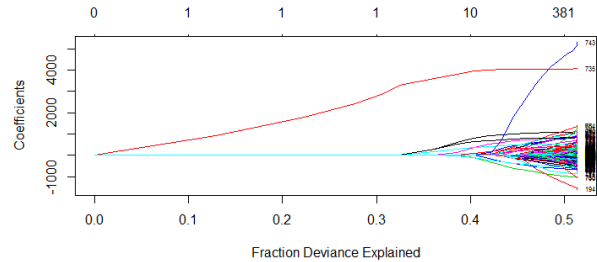


Figure 3. Amount of deviance explained by predictors

Using the cross-validation functionality of the glmnet package in R, we determined the optimal tradeoff between the number of predictors and model accuracy. The results (Figure 4) indicate that the optimal number of variables lies around 383 variables, so about half of the 768 features we extracted from the Instagram posts have no significant influence on engagement.

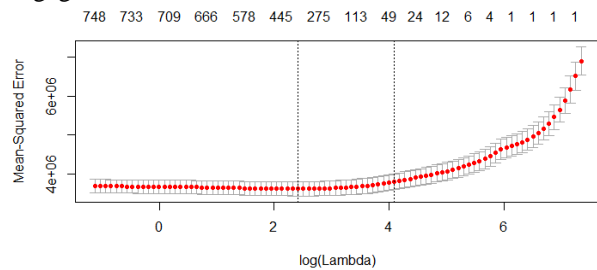


Figure 4. Cross-validation curve with suggested λ values

Table 3 presents the most influential creator- and context-related predictors. The table shows the frequency with which these variables occur in the dataset, as well as their estimated regression coefficients. The coefficients can be interpreted as the additional number of likes and comments the post would be predicted to receive if these features were present. For example, a twenty-year-old woman posting on Friday night at 8:00 p.m. is predicted to receive 1,332 more likes and comments than the average post.

Table 3. Most influential creator- and context-related predictors

Variable	Value	Freq.	Coef.
Posting time a.m.	6:00-7:00 a.m.	109	955
Posting time p.m.	8:00-9:00 p.m.	1 097	188
Posting day	Friday	1 925	23
Gender	Woman	10 997	626
Filter	Clarendon	389	79
Followers	50,000-100,000	2 192	832
Age	Born before 1995	4 162	495

Similarly, Table 4 shows the most influential content-related predictors. For example, a post that includes the caption “Wonderful Switzerland,” a “speak-no-evil” monkey emoji, and a picture of a woman on a mountain is predicted to receive 2,096 more likes and comments than average.

Table 4. Most influential content-related predictors (i.e., words, emojis, and image classes)











Word	Freq.	Coef.
Instagram	136	1 011
Switzerland	128	829
Wonderful	101	448
Video	151	393
Delicious	211	333
Sunday	309	248
Make	164	215
Blonde	121	215
Outfitoftheday	243	213
Christmas	106	194
Emoji	Freq.	Coef.
	205	807
	301	230
	129	204
	190	199
	797	185
	189	173
	311	153
	470	144
	141	130
	183	126

Image class	Freq.	Coef.
Text	401	1 175
Education	210	787
Togetherness	180	520
Mountain	152	355
Woman	5 539	320
Sign	167	300
Ice	131	239
River	90	224
Sand	234	224
Town	106	223

5. Discussion, implications, and limitations

5.1. Contribution to practice

Our results indicate that choosing the right influencer affects user engagement, as the creator-related factors—especially the number of followers and the creator’s age and gender—play the most significant role among all predictors. Similarly, there are certain days and hours (i.e., contexts) during which the audience is more likely to be engaged than at others. Influencer marketing professionals can use this information to choose bloggers and define the launch time of social media marketing campaigns. Our findings regarding content features can also guide content-creation strategies for social media marketing, thereby responding to marketers’ need to improve their understanding of what types of content are the most engaging [1], [36]. For example, our results suggest that pictures with people and scenery and emojis that express positive emotions (e.g., relief, love, joy) increase engagement.

An open issue to be explored in future work concerns the consequences of designing content according to this knowledge. For example, it is possible that, as professional bloggers and marketers increase their use of content that is predicted to be highly engaging, the content will lose effectiveness: If everybody posts pictures with women in front of nature scenes on Friday evenings at 8:00 p.m., a kind of fatigue effect may set in. However, the approach we presented here can easily be repeated with minimal costs in order to monitor such developments in near-real time.

5.2. Contribution to research

To the best of our knowledge, this study is the first to include content variables like words, emojis, and images as independent variables to explain engagement on social media platforms. Even though none of the content features in the social media posts alone

explains more than 2 percent of the deviance in engagement, these features are easy to influence and combine in order to increase a post's impact.

The findings of our research also extend the existing body of literature by confirming the importance of contextual features (e.g., date, time) that practitioners have identified [41] [16] and the high impact of creator-related features (e.g., age, sex, followers) [2], [21]. Of all the features we examined, we found that the number of followers has the most impact on user engagement [4]. Although this finding is not so surprising, unlike context and content features, this variable is difficult for users to influence. As for the content factors that drive engagement, our findings confirm that people in pictures increase the engagement rate and that pictures that include text and scenery have a high impact on the number of likes and comments received [4], [30]. However, we speculate that these content-related features are highly dependent on the industry that uses them and on how they are used.

5.3. Limitations

One limitation of this study is the limited generalizability of the results, as the dataset contains only information from Instagram bloggers from German-speaking countries. A similar approach to the one presented here could easily be used with data from other social media platforms to increase the results' generalizability.

A methodological limitation associated with the study is that we used LASSO regression, which selects only one feature and sets the others to zero in a case when features are highly correlated. Hence, our analysis may have missed features that are highly correlated with those presented in our Results section. In addition, LASSO regression does not provide information regarding the statistical significance of predictors; we can trust it only to discard the insignificant variables and select the significant ones. It is also difficult to argue the predictive accuracy of our model (51.39% of deviance explained), as there are no comparable models reported in the literature to use as benchmarks.

Moreover, using an automated approach to classify pictures sometimes results in misclassification, even though manual checks of the Clarifai results indicated a high level of accuracy.

In future studies, we intend to provide a more comprehensive and precise view of what drives social media engagement. Using the sum of likes and comments might not be the best proxy for engagement because the ratio and weight between comment and

likes is not balanced, so we will consider using other measures for engagement.

Finally, as we used only a static snapshot of data in our analysis, we were not able to capture fully the dynamic nature of engagement on social media platforms. Some of the posts we analyzed might have received more likes and comments after we downloaded the data, which might have caused biases in our analysis.

6. Conclusion

Our approach to identifying and quantifying the factors that influence engagement in social media marketing demonstrates how data analytics can create business value for marketing organizations. Besides directly applying the insights we generated from our analysis, the approach we used can be used in business contexts to maximize the impact of social media activities and increase interaction with potential customers. For instance, our results and approach can guide companies and influence marketers to create more appealing advertisements and successful WOM marketing campaigns by designing engaging content and choosing influential creators and contexts.

This research also creates a foundation for future research on social media engagement. For example, future research may seek to identify additional features that increase the ability to explain and predict engagement and may study whether and how predictors of engagement differ based on the use contexts (e.g., different products, brands, or industries). Finally, our approach might also be used to predict and improve the impact of social media posts in applications outside of marketing (e.g., politics).

7. References

- [1] Adobe Digital Index, The Social Intelligence Report - Q1, 2014, available at: http://www.cmo.com/content/dam/CMO_Other/ADI/ADI_Mobilegeddon/Q2-2015-Social-Intelligence-Report.pdf (retrieved June 8, 2016).
- [2] Arguello, J., B. S. Butler, E. Joyce, R. Kraut, K. S. Ling, C. Rosé, and X. Wang, "Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities", in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Montreal, Québec, 2006, pp. 959-968.
- [3] Baker, M. J., and G. A. Churchill Jr, "The Impact of Physically Attractive Models on Advertising Evaluations", *Journal of Marketing Research*, 14(4), 1977, pp. 538-555.

- [4] Bakhshi, S., D. A. Shamma, and E. Gilbert, "Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram", in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Toronto, ON, 2014, pp. 965-974.
- [5] Bakhshi, S., D. A. Shamma, L. Kennedy, and E. Gilbert, "Why We Filter Our Photos and How It Impacts Engagement", in: Proceedings of the Ninth International Conference on Web and Social Media, AAAI, Oxford, England, 2015, pp. 12-21.
- [6] Berger, J., and K. L. Milkman, "What Makes Online Content Viral?", *Journal of Marketing Research*, 49(2), 2012, pp. 192-205.
- [7] Brown, D., and N. Hayes, *Influencer Marketing: Who Really Influences Your Customers?*, Routledge, Amsterdam, Netherlands, 2008.
- [8] Burke, M., E. Joyce, T. Kim, V. Anand, and R. Kraut, "Introductions and Requests: Rhetorical Strategies That Elicit Response in Online Communities", in: Proceedings of the Third Communities and Technologies Conference, Springer London, Michigan State University, MI, 2007, pp. 21-39.
- [9] Burke, M., and R. Kraut, "Mind Your Ps and Qs: The Impact of Politeness and Rudeness in Online Communities", in: Proceedings of the Conference on Computer-Supported Cooperative Work, ACM, San Diego, CA, 2008, pp. 281-284.
- [10] Castronovo, C., and L. Huang, "Social Media in an Alternative Marketing Communication Model", *Journal of Marketing Development and Competitiveness*, 6(1), 2012, pp. 117-131.
- [11] Chen, Y., S. Fay, and Q. Wang, "The Role of Marketing in Social Media: How Online Consumer Reviews Evolve", *Journal of Interactive Marketing*, 25(2), 2011, pp. 85-94.
- [12] Clarifai, Technology, 2016, available at: <https://www.clarifai.com/technology> (retrieved June 13, 2016).
- [13] Clarifai, The ultimate guide to visual recognition APIs, 2016, available at: <http://blog.clarifai.com/the-ultimate-visual-recognition-api-guide/#.V1r9Idl97IV> (retrieved June 13, 2016).
- [14] Curalate, Pixel perfect: Which image qualities drive engagement on Instagram, 2013, available at: <http://www.curalate.com/blog/6-image-qualities-that-drive-more-instagram-likes/> (retrieved June 8, 2016).
- [15] Cvijikj, I. P., and F. Michahelles, "A Case Study of the Effects of Moderator Posts within a Facebook Brand Page", in: Proceedings of the Third International Conference on Social Informatics, Springer Berlin Heidelberg, Singapore, 2011, pp. 161-170.
- [16] Ellering, N., What 16 Studies Say About the Best Times to Post on Social Media, 2016, available at: <http://coschedule.com/blog/best-times-to-post-on-social-media/> (retrieved June 8, 2016).
- [17] Facebook, How is Engagement Rate Defined?, 2016, available at: <https://www.facebook.com/help/178043462360087> (retrieved June 6, 2016).
- [18] Friedman, J., T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent", *Journal of Statistical Software*, 33(1), 2010, pp. 1-22.
- [19] Friedman, J., Hastie, T., and Tibshirani, R., Package 'glmnet', 2016, available at: <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf> (retrieved June 2, 2016).
- [20] Gandomi, A., and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics", *International Journal of Information Management*, 35(2), 2015, pp. 137-144.
- [21] Gilbert, E., S. Bakhshi, S. Chang, and L. Terveen, "I Need to Try This?: A Statistical Overview of Pinterest", in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Paris, France, 2013, pp. 2427-2436.
- [22] He, W., S. Zha, and L. Li, "Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry", *International Journal of Information Management*, 33(3), 2013, pp. 464-472.
- [23] Hoffman, D. L., and M. Fodor, "Can You Measure the ROI of Your Social Media Marketing?", *MIT Sloan Management Review*, 52(1), 2010 pp. 41-49.
- [24] Hu, Y., L. Manikonda, and S. Kambhampati, "What We Instagram: A First Analysis of Instagram Photo Content and User Types", in: Proceedings of the 8th International Conference on Weblogs and Social Media, The AAAI Press, Ann Harbor, US, 2014, pp. 595-598.
- [25] Instagram, Celebrating a Community of 400 Million, 2015, available at: <http://blog.instagram.com/post/129662501137/150922-400million> (retrieved June 6, 2016).
- [26] Jamali, S. and H. Rangwala, "Digging Digg: Comment Mining, Popularity Prediction, and Social Network Analysis", in: Proceedings of the International Conference on Web Information Systems and Mining, IEEE, Shanghai, China, 2009, pp. 32-38.
- [27] Laroche, M, M. R. Habibi, and M. O. Richard, "To Be or Not to Be in Social Media: How Brand Loyalty is Affected by Social Media?", *International Journal of Information Management*, 33(1), 2013, pp. 76-82.

- [28] Lee, D., H. Kartik, and N. Harikesh, "The Effect of Social Media Marketing Content on Consumer Engagement: Evidence from Facebook", Working paper, The Wharton School, University of Pennsylvania, 2014, pp. 1-51.
- [29] Mangold, W. G., and D. J. Faulds, "Social Media: The New Hybrid Element of the Promotion Mix", *Business Horizons*, 52(4), 2009, pp. 357-365.
- [30] Nielsen, J., *Photos as Web Content*, 2010, available at: <https://www.nngroup.com/articles/photos-as-web-content/> (retrieved June 8, 2016).
- [31] Paine, K. D., *Measure What Matters: Online Tools for Understanding Customers, Social Media, Engagement, and Key Relationship*, John Wiley & Sons, Westford, MA, US, 2011.
- [32] Pulizzi, J., and Handley, A., *B2B Content Marketing: 2016 Benchmarks, Budgets and Trends — North America*, 2015, available at: <http://contentmarketinginstitute.com/research/content-marketing-research/> (retrieved June 8, 2016).
- [33] Romero, D. M., W. Galuba, S. Asur, and B. A. Huberman, "Influence and Passivity in Social Media", in: *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer Berlin Heidelberg, Athens, Greece, 2011, pp. 18-33.
- [34] Sabate, F., J. Berbegal-Mirabent, A. Cañabate, and P. R. Leberherz, "Factors Influencing Popularity of Branded Content in Facebook Fan Pages", *European Management Journal*, 32(6), 2014, pp. 1001-1011.
- [35] Statista, *Number of Social Network Users Worldwide from 2010 to 2019 (in billions)*, 2016, available at: <http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> (retrieved June 8, 2016).
- [36] Stelzner, M. A., *Social Media Marketing Industry Report: How Marketers Are Using Social Media to Grow Their Businesses*, 2015, available at: <https://www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2015.pdf> (retrieved June 8, 2016).
- [37] Sterne, J., *Social media metrics: How to Measure and Optimize Your Marketing Investment*, John Wiley & Sons, Hoboken, NJ, US, 2010.
- [38] Suh, B., L. Hong, P. Pirolli, and E. H. Chi, "Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network", in: *Proceedings of the Second International Conference on Social Computing*, IEEE, Minneapolis, MN, 2010, pp. 177-184.
- [39] Thorne, L., *Word-of-mouth Advertising, Online and Off: How to Spark Buzz, Excitement, and Free Publicity for Your Business or Organization with Little or No Money*, Atlantic Publishing Company, Ocala, FL, 2008.
- [40] Tibshirani, R., "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, pp. 267-288.
- [41] Trackmaven, *The Fortune 500 Instagram Report*, 2014, available at: http://pages.trackmaven.com/rs/trackmaven/images/TM_Fortune500_Instagram_Report.pdf (retrieved June 6, 2016).
- [42] Tuten, T. L., and M. R. Solomon, *Social Media Marketing*, Sage, London, UK, 2014.
- [43] Zeiler, M. D., and R. Fergus, "Visualizing and Understanding Convolutional Networks", in: *Proceedings of the European Conference on Computer Vision*, Springer International Publishing, Zürich, Switzerland, 2014, pp. 818-833.
- [44] Zou, H., and T. Hastie, "Regularization and Variable Selection via the Elastic Net", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 2005, pp. 301-320.