# Analyzing Content and Customer Engagement in Social Media with Deep Learning

(*The bulk of this work was done by a student.*)


## Abstract

In the present study, we investigate the effect of social media content on subsequent customer engagement (likes and reblogs) using a large-scale dataset from Tumblr. Our study focuses on company-generated posts, which consist of two main information sources: visual (images) and textual (text and tags). We employ state-of-the-art machine learning approaches including deep learning to extract data-driven features from both sources that effectively capture their semantics in a systematic and scalable manner. With such semantic representations, we develop novel complexity, similarity, and consistency measures of social media content. Our empirical results show that proper visual stimuli (e.g., beautiful images, adult-content, celebrities, etc.), complementary textual content, and consistent themes have positive effects on the engagement, and that content demanding significant concentration levels (e.g., video, images with complex semantics, text with diverse topics, complex sentences, etc.) have the opposite effects. This work contributes to the literature by exemplifying how unstructured multimedia data can be translated into insights. Our framework for semantic content analysis, particularly for visual content, illustrates how to leverage deep learning methods to better model and analyze multimedia data for effective marketing and social media strategies.


---


*Key words*: semantic content analysis, social media, online advertisement, customer engagement, deep learning, topic modeling, Big Data, representation learning, machine learning, Tumblr

## 1. Introduction

Online social networks have attracted hundreds of millions of users and emerged as one of the most important media platforms for companies to deliver their content to existing and potential customers. As a result, companies have substantially increased their investments and activities on social media engagement. Thus, content design or engineering has become increasingly important, with the goal of developing content that better engages target users and achieves desired goals of the marketer. In general, social media content consists of two main types of information: *textual* (text body and tags) and *visual* (images and videos). In the present paper, we systematically analyze both textual and visual content of company-generated posts utilizing advanced machine learning algorithms, and investigate their effect on customer engagement.

Many studies on social media have focused on analyzing the textual content and its effect on customer engagement [33, 30, 18]. While these studies provided interesting findings, we argue that much existing research on social media has substantial limitation in that *visual* content is largely ignored. The old saying "a picture is worth a thousand words" has never been more true in the world of social media. Customers are faced with massive content overload in social media, forcing shorter attention spans and increasing the cost of identifying the information they want.[1] Images increase the odds of a post getting noticed, especially on mobile devices with small screen sizes. Moreover, studies have shown that people are more likely to remember the content of a post when it's accompanied by a striking image, which often conveys concepts better than any text does [28, 8].

The demonstrated impact of visual content begs on the research question of this paper: what kind of image attracts more customer engagement? There is a large body of literature in advertisement and marketing addressing this important question in the context of

[1] `http://www.eremedia.com/ere/social-media-and-short-attention-spans`

print or website ads [7, 21, 31, 25, 35, 26, 39]. This research has led to the development of visual complexity theory [6] and its effect on customer behavior. However, such settings are contextually quite different from digital images in social media posts that are often accompanied with detailed textual content. Moreover, most of this work has investigated theories about visual content primarily in laboratory settings, whereas systematically analyzing the empirical consequences of visual content in real-world social media settings has been lacking. Lastly, visual features were generated manually in previous papers, which can be subjective and susceptible to human errors, thus rendering such analysis infeasible on massive image datasets from social media.

To address the research question, we develop novel content features using state-of-the-art machine learning methods including deep learning [17], Latent Dirichlet Allocation (LDA) topic modeling [2], and word2vec [20]. For visual content, most existing research has focused on *feature complexity* based on basic visual features such as colors, brightness, and edges, which are constructed from raw *pixel* values of images, but not the actual *semantics* of the images. Moreover, the work by [36] has shown that visual preferences are influenced by the semantic content–that is the the linguistic meaning–in the image. To this end, we propose a novel *semantic complexity* for images and distinguish it from feature complexity. We posit that semantic complexity plays an important role in how people actually perceive an image, which in turn affects customer engagement. By utilizing deep learning, we obtain semantic representations of images (instead of pixels) and compute their semantic complexity. Furthermore, we measure *content similarity* between visual and textual content and *content consistency* of a post with respect to the blog's average content, which have been made possible thanks to deep learning. For textual content, we measure textual complexity at a *global* and *local* level for a given post following cognitive and psychology theories of text comprehension [9, 10, 12]. That is, textual complexity is measured from the overall topics of the text (global) using topic modeling as well as from the sequence of individual words (local) with word2vec.

In the present paper, we investigate how the three main variables–content complexity, similarity, and consistency–as well as other relevant features will have impacts on the popularity of social media contents measured as user engagements. We apply linear models with fixed effects to control for company-level, time-invariant fixed effects. For the empirical analysis, we use a novel large-scale dataset from Tumblr, which is one of the leading social media platforms [4]. Many top brands maintain their official Tumblr blogs from which they create posts.[2] Users can engage with company-generated content on Tumblr by *liking* or *reblogging* (i.e., sharing). One attractive feature for Tumblr is that the majority of posts contain images, which makes it a good fit for visual content analysis.

Our paper has significant contributions to both academic literature and industry. First, we develop new features and concepts of content complexity, similarity, and consistency by exploiting advanced machine learning methods. Second, our research contributes to the information systems and marketing literature with more emphasis on advertisement content design, where there has been limited work in the current literature. Third, our large-scale empirical analysis on Tumblr conveys the effects of both visual and textual content on customer engagement. Last but not least, not only our findings can be readily applied to social media strategies by marketers, but also the proposed framework for semantic content analysis demonstrates how to adopt advanced machine learning methods to better model and analyze content data, especially multimedia data.

[2] `http://brands.tumblr.com`

## 2. Hypothesis Development

Successful advertisements should be effective in communication, further influencing consumer behavior. Existing literature in advertising communication suggests hierarchy models in consumer responses [16, 32, 13]. In this paper, we evaluate the effectiveness of social media advertisements on customer engagement in the form of likes and reblogs. Thus, we divide the impact of company-generated posts on social media engagement into three parts: attention, comprehension, and preference.

Attention attraction is essential for social media advertisement. Since customers have shorter attention spans in social media, posts that can attract more attention will increase the chance of social engagement. Posts with more images or animated GIF images (moving stimuli) are likely to attract attention more easily than posts with only text [24]. Furthermore, feature complexity, measured as the compressed file size of an image, plays a central role in advertisement attention based on visual complexity theory [6]. An image with higher feature complexity implies more variation at the level of individual pixel values (e.g., more complex colors and luminance), resulting in a prominent stimuli. Thus we expect to see:

Hypothesis 1. *In a social media post, the number of images, number of GIF images and feature complexity of images have positive effects on customer engagement.*

The second step of social media engagement is comprehension–that is, understanding and interpreting the content. People in online social media have limited time and may not expend a lot of effort to absorb the advertisement [26]. Hence, consumers are likely to find social media content that is easier to interpret to be more attractive. For visual content, we emphasize the importance of semantic complexity, which is different from feature complexity. For example, an image with a single simple object, which corresponds to low semantic complexity, can have large variations in pixel values leading to high feature complexity. Combined with consumers' limited cognitive capacity, advertisements with semantically concise visual content would improve the information process.

For textual content, we follow text comprehension theory in psychology, which is described as both local micro-level (processing individual words) and global macro-level (organizing full meaning of text) processes [12]. The use of out-of-context words, which is measured by order complexity, impedes the job of comprehending the text from the local perspective [15]. For the global level, text consisting of fewer sentences and focusing on a smaller number of topics is preferred in social media [22].

Lastly, previous research has shown that differences between visual and textual content have a negative impact on cognition of the advertisement [7, 21]. Thus we propose:

Hypothesis 2. *In a social media post, semantic complexity and the number of salient objects of an image have negative effects on customer engagement.*

Hypothesis 3. *In a social media post, the number of sentences, topic and order complexity of text have negative effects on customer engagement.*

Hypothesis 4. *In a social media post, semantic similarity between an image and text & tags has a positive effect on customer engagement.*

The next step is preference. There is abundant theoretical and empirical research in marketing and behavioral economics showing that consumers tend to have stable preference. In fact, this is also one of the underlying bases for content filtering recommendation systems [1]. In addition, studies on other social media have found that posting about a consistent theme attracts more audience [3, 38]. Thus, we posit that:

Hypothesis 5. *In a social media post, semantic consistency of visual and textual content has a positive effect on customer engagement.*

## 3. Tumblr Data and Post Content Characteristics

Our Tumblr dataset consists of 35,651 posts created by a panel of 183 official company blogs from various industry sectors over a six-month period between May 2014 and October 2014. The list of companies is given in the Table 1. Among the collected posts, 88.4% are photo posts with text, 7.4% are pure text posts, and the remaining 4.2% are video posts. A total of 53,417 images was collected from all photo posts. For each post, our data also contains time-series information on the two kinds of engagement measures, likes and reblogs, which are collected through April 2015.

### 3.1. Visual Features

**3.1.1. Feature Complexity** Visual complexity theory [6] is based in the idea that visual stimuli like images are a composite of different elements such as colors, luminance, and edges. For example, an image with more variations in colors or brightness would be more complex than one with fewer colors or uniform brightness. Compressed file size is one of the most widely used and effective measures of such visual complexity [26, 19]. Images are stored on a computer using *compression algorithms* (e.g., JPEG and GIF), that essentially remove redundancies in an image to reduce the amount of memory required, as compared with the original image. That is, an increase in the number or variation of these pixel-based features increases the size of the image in terms of computer memory required and its visual complexity. We use the compressed file size of images as the *feature complexity.*

**3.1.2. Semantic Complexity with Deep Learning** While feature complexity effectively captures how visually complex an image is in its appearance, it does not consider the semantic content. To analyze the semantic content of an image, one would need to detect and classify objects in the image. Conventional techniques for such a task have been limited in their ability to process raw image data. Constructing an image classifier requires careful feature engineering efforts based on considerable domain knowledge, which is one of the main reasons studies have relied on basic or manual image features. However, the recent breakthrough of deep learning approaches has made it possible to accurately analyze the semantic aspect of images in a systematic and scalable manner.

In order to extract meaningful semantic features from images, we employ one of the most successful and popular deep learning approaches called deep Convolutional Neural Network (CNN) [14, 11]. The deep CNN model typically consists of multiple layers, where each layer transforms the representation from the previous layer into a more abstract representation. Generally, the objective is to accurately classify objects that appear in the image via the composition of such multiple transformations. The key aspect of deep learning is that it automatically discovers robust representations needed for accurate classification–that is, the layers are not designed by humans, but are learned from the data. The top-five accuracy on a benchmark dataset is boosted to 92% with deep learning, whereas conventional methods achieved 72% accuracy even with the best handcrafted features.[3]

Our deep CNN model is trained on a proprietary Flickr[4] dataset of more than 1.2 million images with 1,700 object categories–that is, the prediction is a 1,700-dimensional vector of confidence scores between 0 and 1 corresponding to each object category. The object categories are general enough to cover various types of objects and concepts (e.g., animals, people, electronics, food, furnishing, nature, vehicles, etc.). Images collected from company-generated posts are given as input to the trained model to obtain their predictions.

In order to measure the semantic content complexity of images, we construct complexity variables based on the CNN predictions. Specifically, we employ the Shannon Diversity

[3] `http://image-net.org/challenges/LSVRC`

[4] Online photo service owned by Yahoo!

Index to measure variety in content. Let $\mathbf{p} \in [0,1]^d$ be the predicted confidence scores for a given image (i.e., $d = 1,700$). Then, the *semantic complexity* is defined as

$$complexity = -\sum_{i=1}^{d} \mathbf{p}_i \log(\mathbf{p}_i). \tag{1}$$

Note that $\sum_{i=1}^{d} \mathbf{p}_i = 1$ and $complexity = 0$ when $\mathbf{p}_i = 1$ for some $i$.

**3.1.3. Other Relevant Image Features** Deep learning is not restricted to image classification but has been widely adopted to other computer vision tasks. We employ deep learning to extract other image content features including aesthetic [5] and adult-content [27] scores, celebrity information [23], and the number of salient objects [40]. Both aesthetic and adult-content scores range from 0 to 1, where higher scores imply better image quality and more adult-content, respectively. The celebrity detection model can detect more than 450 celebrities with precision, and we count the number of prominent objects (0, 1, 2, 3, 4+) in an image with the salient object detection model. In order to illustrate the constructed visual features, we list four example images in Table 2.

### 3.2. Textual Features

**3.2.1. Topic Complexity via Topic Modeling** To capture topics of the text and tags at a global level, we employ LDA topic modeling [2], following its successful applications in the business literature [30, 34, 29]. The basic assumption of LDA is that a given text document consists of a small number of latent topics and that the words appearing in the document is the realization of the underlying topics. There are two outputs from the LDA: (i) keyword sets for each topic and (ii) topic distribution for each post. We use the later output of the LDA model and compute its complexity by Eq (1) as the *topic complexity* for each post (similar to image semantic complexity). Specifically, $\mathbf{p}$ in Eq (1) is set to be the topic distribution of text/tags for a given post, which gives larger complexity values with more diverse topics. We systematically set the number of topics based on the *perplexity* criteria [37], and further confirmed that resulting keywords form intuitively coherent topics.[5]

**3.2.2. Order Complexity via Word2vec** Recently, a neural-network-inspired model called *word2vec* has been proposed that embeds words in a vector space in a manner that captures a large number of syntactic and semantic word relationships [20]. Specifically, word2vec yields $d$-dimensional vector representations of words so that words used in many similar contexts are close to each other in the vector space. This representation is learned by maximizing the predicted probability of words co-occurring within a small window of consecutive words (e.g., five words before/after the focal word).

In contrast with LDA, which captures document-level associations, word2vec focuses on local context information. That is, word2vec predicts a nearby word given a particular word (focal word → nearby words), whereas LDA globally predicts words at the document level (document → topics → words). Another important difference is that the order of words has a significant impact in word2vec, whereas LDA uses a document/word-frequency matrix representation (bag-of-words) that ignores such ordering.

We train our word2vec model using the text corpus from company-generated posts with $d = 100$, which was chosen by cross-validation with respect to the accuracy of the model.[6] From the trained word2vec model, we can compute the probability $p_s$ of sentence $s$ in a given post as the pairwise composite log probability:

$$\log p_s = \sum_{i=1}^{T} \sum_{j \neq i, j=i-b}^{i+b} \log p(s_j | s_i),$$

where $T$ is the number of words in sentence $s$; $b$ is the window size; $s_i$ is the vector representation of the $i$-th word in sentence $s$; and $p(s_j|s_i)$ is a neural network model.[7] High

[5] Topics and keywords constructed from tags and text are available at: `http://diamond.mccombs.utexas.edu/tumblr-lda-keywords-tags.txt` and `http://diamond.mccombs.utexas.edu/tumblr-lda-keywords-text.txt`

[6] We note that other reasonable values of $d$ give almost identical empirical results.

[7] We refer the reader to [20] for details of the model.

$p_s$ implies that the sentence $s$ is quite likely to appear based on the neighboring words, and sentences with low $p_s$ would be less expected for the reader in the current context. Thus, we define the order complexity of a post as: $\textit{order-complexity} = 1 - \frac{1}{N}\sum_{s=1}^{N} p_s$, where $N$ is the number of sentences in a given post.

The proposed order complexity can be considered a measure of readability considering the likelihood of the sentence. Thus, we compute it only for the text of each post, but not for tags, since the ordering of tags can be ambiguous. In terms of readability measures, researchers have traditionally used scores based on the total number of syllables, words, and sentences such as the Flesch score [30]. However, these scores do not consider the actual order of words, in contrast with our order complexity measure.

### 3.3. Content Consistency

Tumblr is initially proposed for the purpose of blogging, where often a steady and continuous readership is formed based on the post contents. We examine the average content of a company blog and measure a post's similarity to it. That is, we evaluate whether an individual post is similar or distinctive to the company blog's usual semantic content.

More specifically, for post $i$ of a given blog, we compute the average content $\mathbf{c}_i^{avg}$ as: $\mathbf{c}_i^{avg} = \frac{1}{|\Omega_i|}\sum_{j\in\Omega_i}\mathbf{c}_j$, where $\Omega_i$ is the set of posts created by that blog excluding post $i$. For text and tags, we set $\mathbf{c}_i$ as the corresponding topic distributions computed via LDA, as discussed in Section 3.2.1. For images, $\mathbf{c}_i$ is set to be the predicted labels obtained from the deep CNN model as discussed in Section 3.1.2. We emphasize that the average image content would be difficult to compute without the representation obtained by deep learning, since images come with various resolutions and formats. Finally, we measure the content consistency of post $i$ as the cosine similarity between $\mathbf{c}_i$ and $\mathbf{c}_i^{avg}$ as:

$$\textit{ContentConsistency} = \frac{\mathbf{c}_i^{\top} \cdot \mathbf{c}_i^{avg}}{\|\mathbf{c}_i\| \cdot \|\mathbf{c}_i^{avg}\|}. \tag{2}$$

### 3.4. Visual and Textual Content Similarity

The interaction or relationship between the two distinct types of content is an important factor of customer engagement. However, quantifying the relationship between pixel-based images and character-based text is not a straightforward task. Here, we propose a novel content similarity measure between visual and textual content of a given post, with the aid of machine learning methods discussed in previous sections. To measure similarity between the two different content types, we need a common representation, which has been made possible due to deep learning approaches. Specifically, we represent each image as a collection of the predicted labels obtained from our deep CNN models and gather a separate "image corpus" using such representation. From the existing text corpus and the image corpus constructed with deep learning, we build a LDA topic model and obtain topic distributions of both images and text.[8]. Finally, we measure the content similarity between the image and text of a given post as the cosine similarity (given in Eq (2)) between the two corresponding topic distributions $\mathbf{p}_{image}$ and $\mathbf{p}_{text}$.

### 3.5. Variable Construction

Table 3 summarizes the variables used in the analysis and their descriptions and descriptive statistics. We observe that the distributions of reblogs and likes are skewed. The additional binary variables, *HasCaption*, *HasIllustration*, and *Symmetry*, are obtained as part of the predicted labels of the deep learning model described in Section 3.1.2. We also include color complexity, which is computed by Eq (1) using the color distribution of an image.[9] Explicit solicitation for likes and reblogs in the text such as "Like/Reblog if ..." are controlled using corresponding binary variables. Figure 2 gives the correlation matrix between the independent variables. The Variance Inflation Factor is 1.32 ensuring that there is no multicollinearity issues with the dataset.

[8] We set the number of topics to 50 as described in Section 3.2.1.

[9] Colors are mapped to their closest color in a standard 16-color palette.

## 4. Model and Empirical Results

In this section, we describe the empirical models and results.

### 4.1. Model

We use linear models with fixed effects for the empirical analysis. In order to control for firm-level, time-invariant unobserved characteristics and temporal trend, we employ both company and time fixed effects.[10] We measure the effects of our constructed independent variables on the two types of customer engagement: likes and reblogs. For the $j$-th post of company blog $i$, our model can be written as:

$$\ln(Engagements_{ij} + 1) = \beta_0 + \boldsymbol{\beta} X_{ij} + \sum_{k=1}^{6} \gamma_k Weekday_k + \sum_{k=6}^{10} \delta_k Month_k + \alpha_i + \epsilon_{ij},$$

where $Engagements_{ij}$ is either $Likes_{ij}$ or $Reblogs_{ij}$; $X_{ij}$ is the independent variables; $\alpha_i$ is the unobserved time-invariant individual company fixed effect; $\epsilon_{ij}$ is a zero mean error term; and $Weekday_k$ and $Month_k$ are dummy variables for the corresponding day of the week (Mon to Sat) and month (Jun to Oct 2014), respectively.

### 4.2. Main Results

Table 4 summarizes the main empirical results. Regarding the visual features, we first observe that *NumImages* has significantly positive effects on customer engagement, which supports **H1**.[11] We also find that *HasGIF* has positive effects on the engagement (**H1**) and that *HasVideo* has the opposite impact on both likes and reblogs. As discussed in Section 2, an animated GIF image tends to attract users' attention quickly and easily. Videos, on the other hand, incur more costs in terms of time and data consumption than passively seeing images do, which may be the reason for the negative coefficient.

[10] Random effects models have also been applied to find mostly consistent results with the fixed effects models.

[11] An additional image increases the number of likes by 12.1% ($e^{0.114} = 1.121$) and reblogs by 10.1% ($e^{0.0965} = 1.101$).

Next, we consider the results on various visual complexity measures. *FeatureComplexity* has a significantly positive effect on the number of reblogs, indicating that "flashy" images motivate user engagements (**H1**). On the other hand, we find that *SemanticComplexity* has a significant and negative effect, indicating that social media users prefer images with semantically simple and straightforward messages than those with complicated meanings (**H2**). We stress that the effects of the two visual complexity concepts (which have opposing effects) are separately estimated in our work, which is only possible thanks to the novel deep learning approach. We do not observe statistically significant effects of *ColorComplexity* and of the number of salient objects in images (*NumObjectsInImage*).

Considering the textual features, *NumSentences* has a significant and negative effect (**H3**). Conversely, *NumTags* has a positive effect. Given that social media platforms leverage tags in the search engine, proper tags will increase the visibility of the focal post. This can explain the positive effect of tags in our analysis. In terms of textual complexity, the results show that people prefer focused topics–negative signs of *TextTopicComplexity* (**H3**). *TextOrderComplexity* has negative effects on reblogs (**H3**), implying that low probability sentences with out-of-context words hinder text comprehension and decrease user engagement. In contrast, we do not observe a statistically significant effect of the Flesch score (*Readability*). Interestingly, all complexity measures have stronger effects on reblog behaviors than on likes in terms of statistical significance and coefficient magnitude.

A well-known social media strategy is to provide incentives (discounts, promotions) to users in exchange for user engagements. In our dataset, however, such explicit solicitations mostly appeal to emotion or support rather than incentives with any economic value. The results show that *AskLike* has the intended effect of increasing the number of likes. To our surprise, *AskReblog* has the opposite effect. One explanation is that, since reblogged posts are integrated into the user's blog, explicitly asking for reblogs without any real incentives can be unfavorable. Another tactic to induce more user reactions is to use questions. However, our results indicate that *HasQuestion* reduces the number of likes and reblogs.

In terms of content consistency, it is found that posts with visually and textually consistent information with respect to the blogs' average content have more user engagement (**H5**). Specifically, the estimated results show that image and text consistency (*ImageConsistency* and *TextConsistency*) have significant and positive effects in both likes and reblogs. The similarity between image and text (*ImageTextSimilarity*) within a post also has significant and positive effect (**H4**). This indicates that social media users prefer content that is internally consistent (i.e., images and text content match up) and externally consistent (i.e., the focal post follows the overall theme of the blog).

## 5. Conclusions and Future Directions

Visual content has grown to be an integral part of social media. However, manipulating and extracting meaningful features from such unstructured data has been mostly limited to simple or manually constructed features. Thus, most previous work on social media has focused only on text data. In the present work, we take a step forward by leveraging deep learning approaches to overcome and bridge this limitation. Specifically, we adopt deep learning techniques to transform unstructured image data into a semantic representation of objects appearing in a given image. This enabled us to construct novel features such as semantic complexity, image and text similarity, and image consistency that previously would have been difficult to compute without human manual intervention.

In this paper, we empirically analyzed $35,651$ Tumblr posts including $53,417$ images from 183 official company blogs to estimate the effects of visual and textual semantic content on customer engagements in terms of likes and reblogs. We leveraged various machine learning techniques such as deep learning and topic modeling to construct relevant features from unstructured visual and textual data sources. We found that proper visual stimuli (such as animated GIF images, beautiful images, adult-content, celebrities, etc.) significantly increases the customer engagement, whereas content that demands significant user concentration levels (such as videos, images with complex semantics, text with diverse topics, complex sentences, etc.) can lead to poor customer engagements given the information-overloading environment customers are facing. We also found that posts with coherent visual and textual content and posts aligning with the theme (or average content of the blog) facilitate customer engagement. This paper is one of the first works to extensively analyze visual content in social media in the context of user engagements.

For possible future extensions of our work, we can analyze the temporal diffusion of the content by looking at the likes and reblogs over time. Even with the same number of likes, some content may be a best seller in a short period of time, and the other may be a steady seller. We can also consider the breadth and depth of the diffusion process. That is, some content may trigger many reblogs that lead to longer propagation paths. We plan to explain various diffusion patterns with content features and social network structures.

We hope that our findings serve as a useful and practical guide for social media content design, where competition for customers is intense. Moreover, the framework of our study illustrates how firms and researchers can analyze rich multimedia data abundant in social media platforms in a scalable and systematic manner using advanced machine learning approaches. We believe that deep learning will increasingly be more effective and prominent in supporting various data-driven decision-making processes, and that our study is one of the first steps toward this direction.

## References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, March 2003.

[3] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 10–17, 2010.

[4] Yi Chang, Lei Tang, Yoshiyuki Inagaki, and Yan Liu. What is Tumblr: A statistical overview and comparison. *ACM SIGKDD Explorations Newsletter*, 16(1):21–29, September 2014.

[5] Sagnik Dhar, Vicente Ordonez, and Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1657–1664, 2011.

[6] Don C Donderi. Visual complexity: A review. *Psychological Bulletin*, 132(1):73–97, 2006.

[7] Julie A Edell and Richard Staelin. The information processing of pictures in print advertisements. *Journal of Consumer Research*, 10(1):45–61, 1983.

[8] Stephanie Geise and Christian Baden. Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, 25(1):46–69, 2015.

[9] Kenneth S. Goodman. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(4):126–135, 1967.

[10] Philip B Gough. One second of reading. *Visible Language*, 6(4):291–320, 1972.

[11] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia (MM)*, pages 675–678. ACM, November 2014.

[12] Walter Kintsch and Teun A van Dijk. Toward a model of text comprehension and production. *Psychological Review*, 85(5):363–394, September 1978.

[13] Philip Kotler. *Marketing Management*. Prentice Hall, 2015.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS)*, pages 1106–1114, 2012.

[15] David LaBerge and S Jay Samuels. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2):293–323, April 1974.

[16] Robert J Lavidge and Gary Albert Steiner. A model for predictive measurements of advertising effectiveness. *Journal of Marketing*, 25(6):59–62, 1961.

[17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, May 2015.

[18] Dokyun Lee, Kartik Hosanagar, and Harikesh Nair. Advertising content and consumer engagement on social media: Evidence from Facebook. *SSRN*, 9 2015.

[19] Penousal Machado, Juan Romero, Marcos Nadal, Antonino Santos, João Correia, and Adrián Carballal. Computerized measures of visual complexity. *Acta Psychologica*, 160:43–57, September 2015.

[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Pro-*

*ceedings of the Advances in Neural Information Processing Systems 26 (NIPS)*, pages 3111–3119, 2013.

[21] Andrew A Mitchell. The effect of verbal and visual components of advertisements on brand attitudes and attitude toward the advertisement. *Journal of Consumer Research*, 13(1):12–24, 1986.

[22] John Morkes and Jakob Nielsen. Applying writing guidelines to web pages. In *Proceedings of the Conference Summary on Human Factors in Computing Systems (CHI)*, pages 321–322. ACM, 1998. ISBN 1-58113-028-7.

[23] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. *Proceedings of the British Machine Vision*, 1(3):6, 2015.

[24] Rik Pieters and Michel Wedel. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, 68(2):36–50, 2004.

[25] Rik Pieters, Michel Wedel, and Jie Zhang. Optimal feature advertising design under competitive clutter. *Management Science*, 53(11):1815–1828, November 2007.

[26] Rik Pieters, Michel Wedel, and Rajeev Batra. The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing*, 74(5):48–60, September 2010.

[27] Srinivasan H Sengamedu, Subhajit Sanyal, and Sriram Satish. Detection of pornographic content in internet images. In *Proceedings of the 19th ACM International Conference on Multimedia (MM)*, pages 1141–1144. ACM, November 2011.

[28] Roger N Shepard. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6(1):156–163, February 1967.

[29] Zhan Shi, Gene Moo Lee, and Andrew B Whinston. Towards a better measure of business proximity: Topic modeling for industry intelligence. *MIS Quarterly (Forthcoming)*, 2015.

[30] Param Vir Singh, Nachiketa Sahoo, and Tridas Mukhopadhyay. How to attract and retain readers in enterprise blogging? *Information Systems Research*, 25(1):35–52, March 2014.

[31] Ruth Ann Smith. The effects of visual and verbal advertising information on consumers' inferences. *Journal of Advertising*, 20(4):13–24, 1991.

[32] N V Sreedharan. *Impact of literature for advertising effectiveness in the visual media a study*. PhD thesis, Mahatma Gandhi University, 2014.

[33] Stefan Stieglitz and Linh Dang-Xuan. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4):217–248, June 2013.

[34] Seshadri Tirunillai and Gerard J Tellis. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, August 2014.

[35] Alexandre N Tuch, Javier A Bargas-Avila, Klaus Opwis, and Frank H Wilhelm. Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, 67(9):703–715, September 2009.

[36] Edward A Vessel and Nava Rubin. Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, 10(2):18.1–14, 2010.

[37] Hanna M. Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno. Evaluation methods for topic models. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 1105–1112. ACM, 2009.

[38] Yi-Chia Wang and Robert Kraut. Twitter and the development of an audience: Those who stay on topic thrive! In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1515–1518. ACM, 2012.

[39] Kewen Wu, Julita Vassileva, Yuxiang Zhao, Zeinab Noorian, Wesley Waldner, and Ifeoma Adaji. Complexity or simplicity? Designing product pictures for advertising in online marketplaces. *Journal of Retailing and Consumer Services*, 28:17–27, 2016.

[40] Jianming Zhang, Shugao Ma, Mehrnoosh Sameki, Stan Sclaroff, Margrit Betke, Zhe Lin, Xiaohui Shen, Brian Price, and Radomir Mech. Salient object subitizing. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4045–4054. IEEE, 2015.

# Appendix

## A. Tables and Figures

**Table 1** **List of company blogs by industry category.**

| | |
|---|---|
| Automotive | acura, audicity, bmwusa, chopardclassicracing, departurelane, hondaloves, jeep, kia, landroverusa, lincolnmotorco, mercedesbenz, moversandmakers, sendthemasignal, smartownersbelike |
| Entertainment | aetv, beatsbydre, blackdiamondpa, conversemusic, disney, disneypixar, drmwrks, fox-adhd, gamestop, gettyimages, hashtaglionsgate, hbo, hinl, huffingtonpost, hulu, ifc, latimes, listenforyourself, nbcnews, nbcnightlynews, newmuseum, npr, pbsdigitalstu-dios, pbstv, penguinteen, runningpress, sesamestreet, spotify, theatlantic, thedai-lyshow, theeconomist, ultimateears, vimeo, wmagazine, xbox, youtube |
| Fashion | 10022-shoe, americanapparel, anthropologie, barbour, bergdorfgoodman, calvinklein, capitolcouture, cartier, clubmonaco, dior, dolcegabbana, donnasjournal, fancyfeast, glamour, goodarthlywd, gq, gucci, harpersbazaar, jcrew, katespadeny, lorealparisusa, maccosmetics, makeupforeverusa, maybelline, modcloth, olay, pfflyersstyle, ralphlau-ren, ray-ban, rickysnyc, sephora, stussy, suitsupply, teamtaylorswiftfragrances, tim-berland, topshop, urbanoutfitters, vanssnow, vogue, warbyparker |
| Finance | americanexpress, amexopenforum, bankrate, mastercard, yahoofinance |
| Food | americashamburgerhelper, amstellight, bemoretea, benandjerrys, coca-cola, cut-tysark, dennys, digiorno, dqfanfood, earthsfinestguide, fruttarefruitbars, hellocere-allovers, ihop, jr-watkins, kitkat, kraftrecipes, krispykreme, naturevalley, nowyoure-cooking, officialsubway, oreo, redbull, simplywonderful, skittles, smirnoffice, sprite, tacobell, tgifridays, usmacallan, wonkaicecream, wonkarandoms, zagat |
| Leisure | acehotel, adidasfootball, adidasoriginals, bandh, becausefutbol, enroutemagazine, hol-idayinn, lifeismagnifique, livelymorgue, lomographicsociety, lufthansa, montanamo-ment, nba, qatarairways, reebokclassics, starwoodhotels, takingoff, thescore, trans-formtomorrow, underarmour, visit-florida, whotels, yahoosports |
| Retail | archiemcphee, barbie, ebay, keds, macys, neimanmarcus, patagonia, sanbornca-noecompany, thecorcorangroup10amspecial, theinsidesource, tiffanyandco, tjmaxx, vikingrange, yahooshopping |
| Tech | att, dell, generalelectric, gereports, ibmblr, ibmsocialbiz, madewithcode, marketr, mashablehq, norton, positivelytogether, smartercities, smarterplanet, sonos, sony, txchnologist, volition, yahoo, yahoolabs |

**Figure 1** The number of topics vs. perplexity (lower the better) for text, tags (Section 3.2.1) and images with text & tags (Section 3.4).

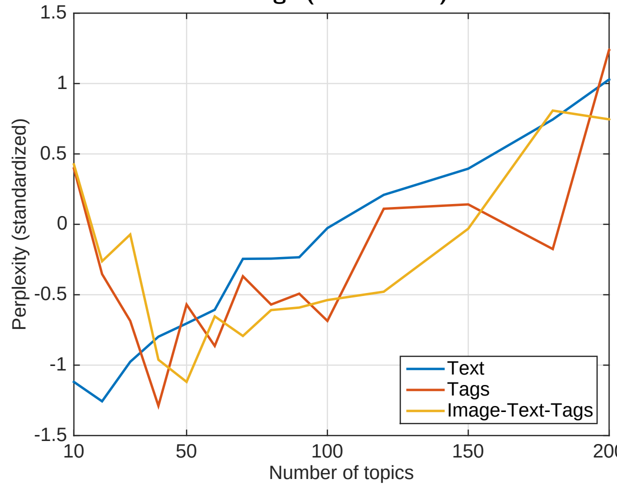



**Figure 2** Correlation of independent variables. Colors in the upper triangular part represent correlations and the lower triangular part shows the same correlations in percentages (between -100 and +100).

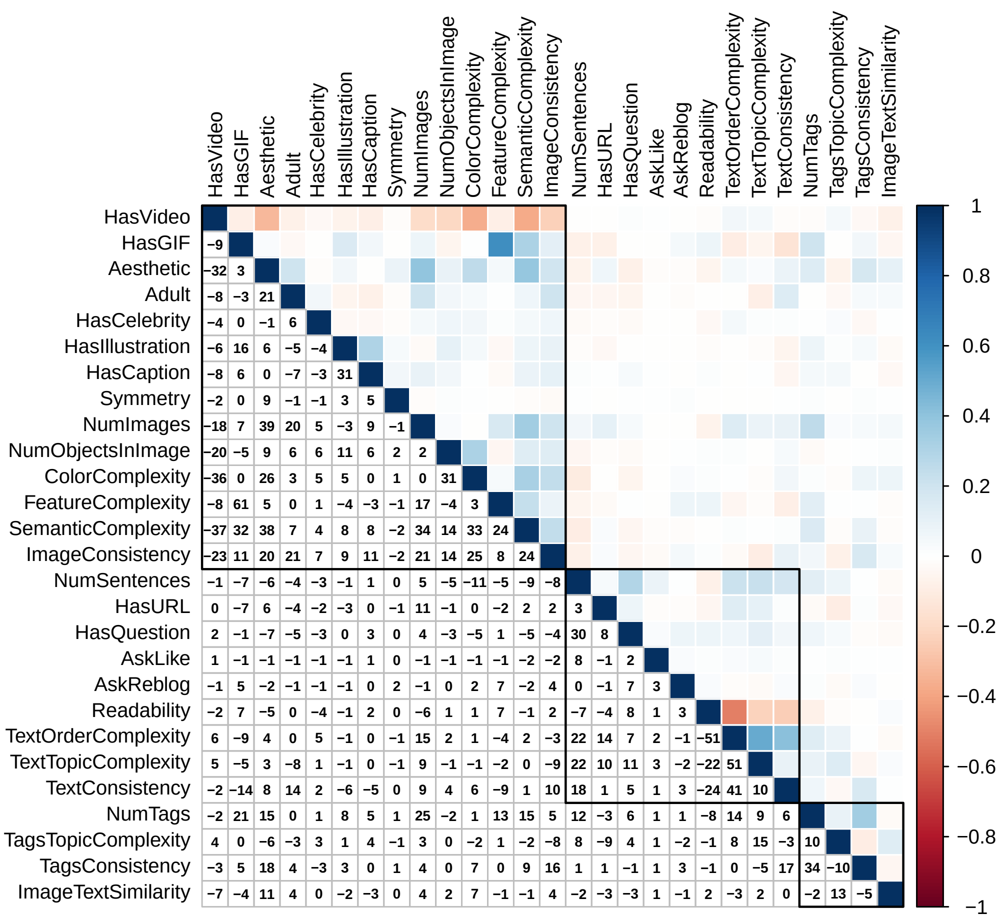


**Table 2** Example images and their feature values. The first image has high aesthetic value, which follows our visual intuition, and a high feature complexity, indicating there are complex pixel structures (e.g., colors, textures). It also has a high semantic complexity, reflecting the fact that it has different types of objects, but the fourth example has the highest semantic complexity since it has more diverse objects than the first example. The second example has a significantly high adult-content score with low feature complexity because it is a grayscale image. The first and fourth examples have many salient objects inside, which the algorithm correctly detected.

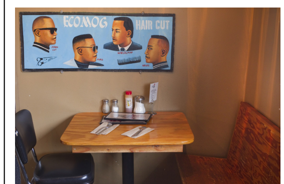
| Example image |   |  |  |   |
|---|---|---|---|---|
| aesthetics | 0.916 | 0.646 | 0.360 | 0.126 |
| adult-content | 0.015 | 0.900 | 0.001 | 0.001 |
| salient objects | 4+ | 2 | 1 | 3 |
| semantic-complexity | 2.458 | 2.749 | 1.583 | 3.240 |
| feature-complexity | 2.113 | 0.595 | 0.201 | 1.613 |

**Table 3 Descriptive statistics of dependent and independent variables.**

| | Description | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|---|
| Dependent Variables | | | | | |
| *Likes* | Number of likes | 531.3 | 6,748 | 0 | 430,745 |
| *Reblogs* | Number of reblogs | 445.0 | 6,967 | 0 | 521,709 |
| Independent Variables | | | | | |
| Visual | | | | | |
| *HasVideo* | Post has video | 0.0416 | 0.200 | 0 | 1 |
| *HasGIF* | Post has GIF image | 0.153 | 0.360 | 0 | 1 |
| *Aesthetic* | Image aesthetic score | 0.339 | 0.218 | 0 | 1 |
| *AdultContent* | Image adult-content score | 0.0566 | 0.153 | 0 | 0.995 |
| *HasCelebrity* | Image has celebrity | 0.0339 | 0.181 | 0 | 1 |
| *HasIllustration* | Image is an illustration[a] | 0.080 | 0.272 | 0 | 1 |
| *HasCaption* | Image has caption | 0.138 | 0.345 | 0 | 1 |
| *Symmetry* | Image is symmetric | 0.006 | 0.080 | 0 | 1 |
| *NumImages* | Number of images | 1.481 | 1.676 | 0 | 10 |
| *NumObjectsInImage* | Number of objects in image | 0.971 | 1.011 | 0 | 4 |
| *ColorComplexity* | Image color complexity | 1.068 | 0.616 | 0 | 2.289 |
| *FeatureComplexity* | Image feature complexity | 1.412 | 3.617 | 0 | 93.36 |
| *SemanticComplexity* | Image semantic complexity | 2.958 | 1.660 | 0 | 6.356 |
| *ImageConsistency* | Image consistency | 0.268 | 0.239 | 0 | 0.993 |
| Textual | | | | | |
| *NumSentences* | Number of sentences | 2.536 | 4.426 | 0 | 206 |
| *HasURL* | Text has URL | 0.116 | 0.320 | 0 | 1 |
| *HasQuestion* | Text has question | 0.103 | 0.304 | 0 | 1 |
| *AskLike* | Explicit solicitation for likes | 0.004 | 0.063 | 0 | 1 |
| *AskReblog* | Explicit solicitation for reblogs | 0.001 | 0.031 | 0 | 1 |
| *Readability* | Flesch readability score[b] | 69.239 | 23.841 | -92.31 | 147.39 |
| *TextOrderComplexity* | Text order complexity | 0.482 | 0.212 | 0 | 1 |
| *TextTopicComplexity* | Text topic complexity | 0.569 | 0.467 | 0 | 2.958 |
| *TextConsistency* | Text consistency | 0.544 | 0.313 | 0 | 1 |
| *NumTags* | Number of tags[c] | 6.884 | 4.370 | 0 | 30 |
| *TagsTopicComplexity* | Tags topic complexity | 0.659 | 0.621 | 0 | 3.689 |
| *TagsConsistency* | Tags consistency | 0.643 | 0.334 | 0 | 1 |
| Visual & Textual | | | | | |
| *ImageTextSimilarity* | Image and text similarity | 0.039 | 0.113 | 0 | 1 |

[a] Image is an illustration such as a diagram, drawing, sketch, or cartoon.

[b] Lower values imply a post is harder to read.

[c] The maximum tag count is 30 by Tumblr's design decision.

**Table 4 Linear fixed effects results for Likes and Reblogs.**

| Variables | (1) Likes | | (2) Reblogs | |
|---|---|---|---|---|
| *HasVideo* | -0.201* | (0.119) | -0.313** | (0.133) |
| *HasGIF* | 0.308*** | (0.0911) | 0.424*** | (0.104) |
| *Aesthetic* | 0.326*** | (0.0860) | 0.575*** | (0.101) |
| *AdultContent* | 0.326*** | (0.0908) | 0.229** | (0.0884) |
| *HasCelebrity* | 0.246*** | (0.0473) | 0.205*** | (0.0657) |
| *HasIllustration* | -0.0132 | (0.0413) | -0.0477 | (0.0513) |
| *HasCaption* | -0.0146 | (0.0374) | -0.0159 | (0.0520) |
| *Symmetry* | 0.0662 | (0.0645) | 0.0752 | (0.0714) |
| *NumImages* | 0.114*** | (0.0184) | 0.0965*** | (0.0239) |
| *NumObjectsInImage* | -0.00429 | (0.0103) | -0.0203 | (0.0124) |
| *ColorComplexity* | 0.0239 | (0.0282) | -0.0120 | (0.0329) |
| *FeatureComplexity* | 0.0103 | (0.00871) | 0.0165* | (0.00847) |
| *SemanticComplexity* | -0.0278*** | (0.0102) | -0.0396*** | (0.0120) |
| *ImageConsistency* | 0.256*** | (0.0636) | 0.267*** | (0.0760) |
| *NumSentences* | -0.00628** | (0.00313) | -0.0114*** | (0.00365) |
| *HasURL* | -0.0421 | (0.0532) | -0.0972* | (0.0559) |
| *HasQuestion* | -0.0602** | (0.0294) | -0.135*** | (0.0331) |
| *AskLike* | 0.199* | (0.103) | 0.153 | (0.121) |
| *AskReblog* | -1.457** | (0.679) | -1.490** | (0.740) |
| *Readability* | 0.000319 | (0.000376) | -0.000805 | (0.000793) |
| *TextOrderComplexity* | -0.150 | (0.125) | -0.504*** | (0.146) |
| *TextTopicComplexity* | -0.00129 | (0.0199) | -0.0535** | (0.0268) |
| *TextConsistency* | 0.159*** | (0.0459) | 0.191*** | (0.0575) |
| *NumTags* | 0.0318*** | (0.00555) | 0.0261*** | (0.00610) |
| *TagsTopicComplexity* | -0.0345* | (0.0193) | -0.0536** | (0.0256) |
| *TagsConsistency* | 0.0938 | (0.0618) | 0.113 | (0.0793) |
| *ImageTextSimilarity* | 0.174** | (0.0786) | 0.304*** | (0.104) |
| weekday = 1 | -0.0963*** | (0.0361) | -0.0443 | (0.0437) |
| weekday = 2 | -0.0348 | (0.0313) | 0.0231 | (0.0368) |
| weekday = 3 | -0.0396 | (0.0313) | 0.0453 | (0.0378) |
| weekday = 4 | -0.0616* | (0.0320) | -0.0213 | (0.0430) |
| weekday = 5 | 0.000600 | (0.0315) | 0.0476 | (0.0439) |
| weekday = 6 | 0.0103 | (0.0329) | 0.0385 | (0.0381) |
| month = 6 | 0.0951** | (0.0387) | 0.904*** | (0.0727) |
| month = 7 | 0.0724 | (0.0510) | 0.863*** | (0.0849) |
| month = 8 | 0.0504 | (0.0634) | 0.828*** | (0.0880) |
| month = 9 | -0.00857 | (0.0534) | 0.818*** | (0.0840) |
| month = 10 | -0.00166 | (0.0507) | 0.829*** | (0.0717) |
| Constant | 3.594*** | (0.143) | 2.343*** | (0.150) |
| Observations | 35,651 | | 35,651 | |
| R-squared | 0.077 | | 0.113 | |
| Number of blogs | 183 | | 183 | |

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1