You Are When You Tweet: Automatic Segmentation of Consumers Based On Social Media Activity

William Rand[1] , Anthony Weishampel[2]
[1] NC State University, Poole College of Management, wmrand@ncsu.edu
[2] NC State University, Department of Statistics, acweisha@ncsu.edu

Social Media is an important marketing channel for companies to directly engage with consumers, and is serving an increasingly important role in customer service. Social media platforms enable the firms to measure brands awareness and gauge customers' opinions. Rather than merely listening, many firms directly engage with the consumers through social media platforms. This has led to an increase in direct interactions between the firms and individual consumers that has revolutionized Customer Relationship Management (CRM), creating a type of Social CRM. Social CRM systems tie traditional CRM data to the identity of individuals on social media [1]. With the widespread presence of social media, Social CRM helps foster beneficial consumer relationships for firms on an individualized scale. The increased interactions present new questions to firms, such as deciding whether or not to engage with a specific individual and how to prioritize responses.

For instance, sometimes the sheer volume of potential interactions is overwhelming. With the vast number of people reaching out to the firms on social media, companies often cannot respond to every individual. Ma et al. showed that it was useful to respond to customers on social media, but that responding to them may encourage future negative word-of-mouth [2]. Thus being able to identify which users have a high potential customer lifetime value (CLV) and which users are most likely to positively impact overall awareness of the brand, is vital to optimize the firm's customer service response and to make the best use out of their Social CRM efforts.

There are cases when the user will provide the relevant information transparently that will enable the firm to classify high CLV individuals. For example the user could include their occupation in their public profile, which may help assess the future value of the customer, or this information can be elicited from the user's previous social media posts. However, there are many cases where the user's profile and history does not include the desired features. Our modeling framework provides a way for assessing how a user of social media should be prioritized regardless of what information is provided in their publicly available social media content, by examining their behavioral activity.

One of the most active platforms where consumers voice their opinions about brands is Twitter. Darmon et al. demonstrated that it was possible to predict a user's Twitter behavior based on recent past behavior and that different groups of users on Twitter exhibited very different types of behavior [3]. The variation in different the groups' Twitter behavior could be sufficient to classify the users whose information cannot be easily extracted from their provided profile information and previous tweets. It may also be the case that an individual's social media behavior alone may not be sufficient to predict these features, but it may provide additional evidence, which can assist in prioritization.

Once a model of a user's behavior has been created, and a similar model of a group of know user's behavior has been created, then by comparing the behavioral models among the different groups and the focal individual's models, we can classify the group status of an unknown individual. Ultimately this research can be used to create a tool that will help marketers identify how important it is to respond to an individual on social media.

To assess our framework, detailed information about 4,776 Twitter users who directly tweeted to a firm using the official Twitter handle of that firm were collected. The firms that were chosen as the seeds for the search were 100 of the most popular brands in English-speaking countries. For each user their timelines (i.e., all of the tweets from the focal user), their profile biography, and their self-described location were collected.

When prioritizing the users, there are many relevant user characteristics that the firm can use. This study investigates three features of the users.

- Are they geographically relevant customers, e.g., are they in a market where the firm's product / service is sold?
- Do they have a potentially high customer lifetime value (CLV)?

- Will they provide an overall increase in positive word-of-mouth about the brand?

The first two characteristics focus on the user's potential to be a high CLV consumer. The third feature is an estimate of the user's effect on the firm's digital word of mouth (WOM). It has been well documented that the digital word of mouth affects the firm's marketing effects and sales [4][5][2] and is an important feature to include when prioritizing users.

The user's Twitter behavior was transformed into a time series, consisting of whether or not the user tweeted during an eight-hour time window. The time series of the users is modeled through a machine learning method known as causal state modeling (CSM) (also called computational mechanics) [6]. By modeling the behavior of each user from individual social data traces, the CSM approach is well equipped to capture heterogeneity among individuals since each model can be as varied as necessary. CSMs assume that the time series data is generated by a conditionally stationary stochastic process. The CSM takes a large state space and identifies the structure of actions within that space. It then collapses this large space down to a Markovian state space that represents the causal states that create that structure, this representation is known as an epsilon-machine. A user's CSM is inferred through implementing the Causal State Splitting Reconstruction (CSSR) algorithm [6].

CSMs were constructed for both individual-level models of user behavior and group-level models. The individual-level models are created by constructing a separate CSM for every user using their own time series data. The group-level models were built by concatenating the time series of the users of that group and applying the CSSR algorithm to the resulting time series. This creates a model that represents the overall behavior of the group, since the CSSR essentially creates an "average" model.

One method of classification to determine the group for a particular user is choosing the group model which best predicts the user's behavior. After modeling the individual hold out user's time series in each group-level CSM, the user is classified to the group corresponding to the most accurate predictions. In other words, the group level model which best predicts the user's behavior is the group to which the user is classified. We call this method *Group-*

Prediction.

Another approach is to create a causal state model for the user in the testing set and then measure the distance between the individual user's model and the group level models. The distance metric between two models is the sum of the absolute differences between the probability of an event occurring between the two models, for all combinations of outcomes $(0000, 0001, \ldots, 1111)$ in the transition matrix. We call this method *Group-Distance*.

A final approach is to use the full heterogeneity of the CSM approach and model the group as a set of separate individuals. In this apporach, a CSM was created for every user in the training set. Then a CSM is built for the testing user and the same distance metric from the group-level analysis is applied to identify the five closest training set CSMs to the testing user's CSM. A k-nearest neighbor (KNN with k=5) method is then used to classify the testing user. We call this method *Individual-KNN*.

The TransCSSR algorithm is an extension of the CSSR methods; however an input time series is required in addition to the observed series. The input series was created for each of the levels. These input series are time series of whether the group had low, medium, or high twitter activity, during a particular time window. Once the input time series were formatted, the transSCCR algorithm was run for each user in the testing set at every different input time series level. The users are classified by which input series best predicts the users behavior. We call this method *TransCSSR-Prediction*.

Using the CSSR algorithm, CSMs were created for every level of each of the three features (See Figure 1). From Figure 1 it appears that the CSMs do not differ much by the different levels of the geography, though there were some clear differences in the other two features. However, the real test was the classification rate. The Group-Prediction method never did very well, but the two distances methods were able to achieve decent results, as seen in Table 1.

The performance of the methods, *Individual-Level*, *Group-Level*, and *TransCSSR* are

Variable (Predicted Levels)	Group-Distance		Individual-KNN		TransCSSR-Prediction	
	Precision	Recall	Precision	Recall	Precision	Recall
Geography (American or Unknown)	0.7956	0.9857	0.6133	0.9274	0.8026	0.4929
CLV (High Value)	0.1071	0.1563	0.1481	0.0430	0.1143	0.1649
Internet WOM (Brand Influencer)	0.3185	0.2829	0.5436	0.2918	0.4789	0.4058

Table 1: Classification results for the group-level distance metric and the K-NN individual level methods

reported in Table 1. This table illustrates how well the model did on a hold-out sample of users. Precision measures out of how many users the model labelled as positive actually were positive, and recall measures out of how many positives were in the hold-out sample the model labelled accurately. The performance of the methods varied between which variable it was predicting and according to which model was being used. All of the models performed best when predicting the user's geography, achieving both high levels of precision and recall. The models were the next best at predicting reoccurring WOM with the *Individual-Level* model acheiving the highest percision. Finally none of the models did very well at predicting CLV. It seems to be that a user's activity is more closely related to geography than to either of the other two measures, and that wealth seems to have little impact on activity. WOM being in the middle range is probably related since user's who are more likely to be active in the future are probably more likely to just be active on Twitter overall.

In our results (Table 1) we show that Causal State Models can be used to model a user's activity on Twitter. Using CSMs of the user's Twitter behavior, there is some evidence to suggest that latent characteristics about the user can be extracted. We examined three different approaches for modeling hidden characteristics: (1) a model which aggregates all individuals who have the same characteristic (e.g., all Americans) and then computes a distance from an unknown user to this aggregate model, (2) an approach where every individual is modeled separately and then an unknown user is classified by a k-NN (k-Nearest Neighbors) approach and (3) the TransCSSR algorithm to build a CSM with an input time series. Our results show that in general the second and third methods are superior to the first, and that some features, e.g., likelihood that the individual will speak positively about the focal subject, is easier to classify than others. We suggest that there are many ways in which our modeling efforts could be improved. For instance, one possibility for improvement would be to model the topics of conversations of the users rather than just the timing of social media activity. We will explore how these improvements can fit within the general framework of social media activity prioritization that we have developed in future work.

References

[1] Neil Woodcock, Nick Broomfield, Geoff Downer, and Michael Starkey. *The evolving data architecture of social customer relationship management.* Journal of Direct, Data and Digital Marketing Practice, 12(3): 249-266, 2011.

[2] Liye Ma, Baohong Sun, Sunder Kekre. *The Squeaky Wheel Gets the Grease-An Empirical Analysis of Customer Voice and Firm Intervention on Twitter*. Marketing Science, 34(5):627-645, 2015.

[3] David Darmon, Jared Sylvester, Michelle Girvan, and William Rand. *Predictability of user behavior in social media: Bottom-up v. top-down modeling.* In Social Computing (SocialCom), 2013 International Conference on, pages 102107. IEEE, 2013.

[4] Judith Chevalier and Dina Mayzlin. *The Effect of Word of Mouth on Sales: Online Book Reviews*. Journal of Marking Research 43(3): 345-354, 2006

[5] Wenjing Duan, Bin Gu, Andrew Whinston. *Do online reviews matter? An empirical investigation of panel data*. Decision Support Systems 45: 1007-1016, 2008

[6] Cosma Shalizi and Kristina Shalizi. *Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences.* UAI 2004 Proceedings of the 20th conference on Unvertainty in artificial intelligence: 504-511, 2004

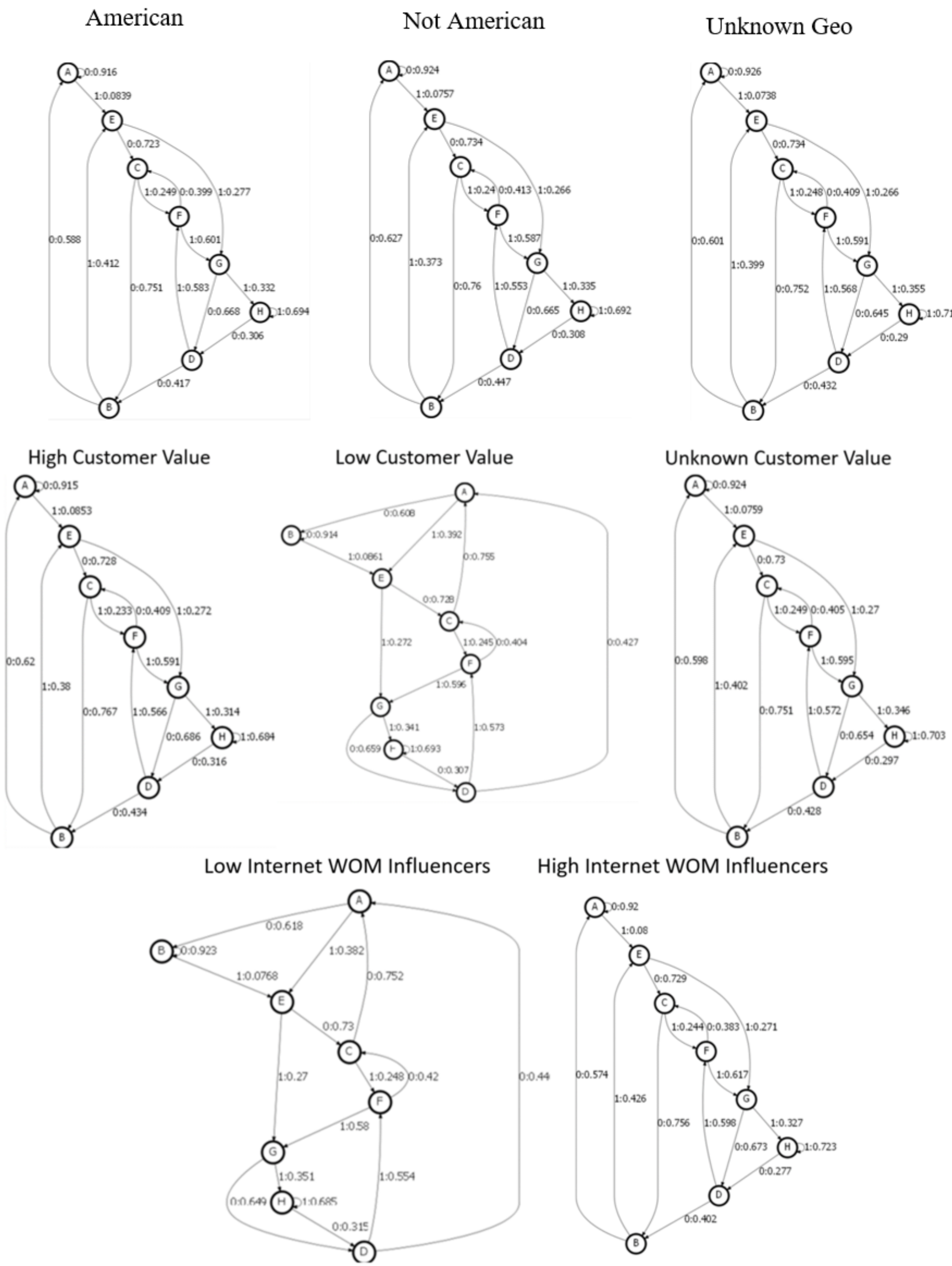

Figure 1: CSMs for the Groups-Levels

Simplicity is not Key: Automatically Identifying Concepts in Firm-Generated Social Media Images that Engage Consumers

Gijs Overgoor and William Rand
University of Amsterdam and NC State University

Introduction
Social media platforms are becoming one of the main channels for achieving a variety of key marketing objectives, from creating awareness to sales conversion (Batra and Keller 2016; Kumar et al. 2016; Kumar et al. 2013). Investing in and developing a social media community strengthens customer-firm relationships and positively impacts firm's revenues and profits (Kumar et al. 2016). However, more and more firms are actively improving their social media marketing strategies by creating fan pages on various social media platforms and generating content. Therefore, managers face the daunting challenge of producing appealing brand content that is less likely to be crowded out and more likely to be engaging to consumers (The Content Marketing Institute 2015). According to a study by media agency Havas (2017) 60% of the content generated by brands is declared as "poor, irrelevant or fails to deliver". As the amount of online firm-generated content (e.g., Instagram posts, brand tweets) continues to increase, it becomes more and more challenging to attain and hold the consumers' attention. To create content that is appealing to consumers requires insight in the popularity or consumer engagement. Understanding the drivers behind consumer engagement towards firm generated content will improve understanding of consumer interests and behavior, which in turn might enable firms to increase return on investment on social media marketing investment.

Although recent studies shed some light on the determinants of consumer engagement in social media, such as arousal (Berger and Milkman 2012), interactivity (de Vries, Gensler and Leeflang 2012), positivity (Hewett et al. 2016), and media persuasiveness (Stephen, Sciandra and Inman 2015), there is very little research on the engagement of consumers towards predominantly visual content. Visual content poses a new dimension of challenges for the content manager. Studies in scene perception (Potter 1976; Oliva 2005) show that observers understand and comprehend the visual information of a scene within 100 milliseconds, so it is crucial that content managers create visual content that is appealing to the observer at the first look. Therefore, there is a need for empirical evidence that shows the drivers of engagement such that it will help firms be more effective with their visual content on social media.

A content manager's main focus is to create content that stops the consumer when browsing or scrolling through the social media content. Similar to advertisements, firm-generated content needs to catch the consumers attention and be engaging enough for a consumer to press the like button or share a comment. In advertising literature, we see an emphasis on simplicity to increase engagement on the one hand (Aitchinson 1999; Book and Schik 1997) and emphasis on complexity to increase engagement on the other hand (Nelson 1985, Putrevu, Tan and Lord 2004). More in-depth research in the effects of visual complexity and its effect on attitude towards advertisements) has shown positive as well as negative impacts for different visual complexity measures (Pieters, Wedel, and Batra 2010). Inspired by this research, the divide in advertising literature and recent advances in computer science we aim to empirically explain the effect of visual complexity on the engagement with firm generated content on Instagram answering to the following research question: How do different types of visual complexity impact consumer engagement on social media?

This study attempts to fill some of the central gaps in extant work by developing a framework for extracting information from visual content and using this to study the user engagement with this content. Specifically, we extract visual complexity measures from the image content and explore their individual and combined impact on the user engagement. By expanding the visual complexity framework as proposed by Pieters, Wedel, and Batra (2010) we give brand managers control over the different types of complexity, which will then create a decision-making tool for designing and/or modifying that maximizes social media effectiveness. Theoretically we contribute to the advertisement literature as we show in particular the image features that benefit from simplicity versus complexity and vice versa. By investigating the visual complexity framework on a large-scale social media data-set we contribute to the need for exploration of visual content on social media. Our methods for extracting information from visual content on social media opens up possibilities for marketers to obtain rich sources of information and use this into predictive models. It provides a conceptual framework for automatically extracting visual complexity measures from images on a large scale, providing content managers with information on the content they generate as well as advise on what is most likely to be engaging to consumers.

Visual Complexity

People's perceptions, preferences and behavior with regards to visual objects, scenes and display are influenced by visual complexity (Machado et al. 2015). Several perceptual features that contribute to visual complexity invoke cognitive and affective responses and they impact preference and liking (Palumbo et al. 2014; Pecchinenda et al. 2014; Chatterjee 2004; Leder et al. 2004). Based on these features an image can be perceived more or less complex. Derived from past research (Pieters, Wedel, and Batra 2010), we distinguish three different levels of visual complexity. These different levels of complexity are based on the machine learning difficulty, the ease of human perception and the visual processes they require in the human brain. First, the low-level measures are based on low-level visual processes in the primary visual cortex (Palmer 1999) evoked by the complexity of the image. It represents pixel-level variation and unprocessed or unstructured image information without regards to semantics. Second, the mid-level complexity measures consist of information about variation in the design of images - edges and color variation. The high-level measures are based on the design and semantics of an image (i.e. objects and similarity). Both the mid- and high-level visual complexity evoke mid-level visual processes based on objects and pattern recognition (Palmer 1999). We automatically extract and construct these measures from the image using image processing and machine learning techniques. After extraction, we measure the impact of the different visual complexity measures on the consumer engagement - number of likes and number of comments. In figure 1, we conceptualize the process. Figure 1 shows how we extract the visual complexity measures from the image and subsequently use them to measure their impact on the engagement. The level of complexity is derived from the type of visual process in the human brain and the level of machine learning the extraction process requires. Increasing machine learning difficulty leads to a decreasing ease of human perception of the model, in other words understanding of what exactly happens in the process decreases when the machine learning difficulty increases.

Low-level complexity: The detail and variation there is in the three basic visual features (color, luminance and edges) across an image constitute the complexity of an image. This complexity can be constructed by examining the image on a pixel-level. An image that is low in complexity shows little variation per pixel, whereas a high complexity image shows big variation per pixel. Based on algorithmic information theory, research has shown that the minimum length of the code required to describe a visual image constitutes a good measure of its complexity (Leeuwenberg 1969; Simon 1972). Image compression techniques (Shapiro and Stockman 2001) reduce the amount of memory needed relative to the original image by stripping an im- age of its redundancies, this is standard for image compression (Wallace 1992). Simple images contain many redundancies and can be compressed compactly whereas compression of complex images barely decrease size. The Instagram images are compressed and their size ranges from a little over 1 kilobyte to 400 kilo-byte. We will use the JPEG file size as a measure for low-level complexity.

Mid-level complexity: We construct low-level complexity in more detail using image processing techniques. An edge in an image is a boundary or contour at which a significant change occurs in some physical aspect of an image, such as the surface reflectance, illumination or the distances of the visible surfaces from the viewer. Edges in an image usually indicate changes in depth, orientation, illumination, material, object boundaries etc. We perceive that the quantity of edges relates to the complexity of the image. We use canny edge detector (Canny 1986) to detect edges within an image. Every pixel in the image will be classified as either 0 (not on an edge) or 1 (on an edge). Thus, we propose an edge-complexity measure which is the total number of pixels on an edge divided by the total number of pixels.

High-level complexity: Images with a higher variation in terms of concepts present are also more complex. Recent advances in computer science have provided us the ability to automatically extract conceptual information from a large number of images. This information has shown to be particularly useful for social media popularity prediction (Gelli et al. 2015; Khosla, Das Sarma, and Hamid 2014; Mazloom et al. 2016). Large scale concept detection plays a central role in constructing high-level complexity measures. Derived from previous research (Pieters, Wedel, and Batra 2010), we develop two main measures for the high-level complexity that can be extracted from an image directly: The quantity of concepts and the dissimilarity of concepts. The concept detection process and construction of the measures will be described more thoroughly in the method section.

Instagram

In this section, we explain why we use Instagram to answer our research question and we will give a quick introduction of some of its main features, as they will explain our dataset and some of the variables we use for modelling the data.

Similar to other social media platforms Instagram allows users to generate content and share this with other users across the platform. Instagram is considered a visual social media platform meaning that its main focus is visual content – images in particular. A user shares (posts) an image with a short description (caption) on their Instagram page that is seen by the user's friends or fans called followers. Typicallym users follow several other users or brands that are (actively) generating content. The followers can show appreciation of the content posted by liking it and they can also comment. After taking a photo a user has several ways to quickly edit it before sharing it on Instagram. One of Instagram's most popular features is the possibility of adding a filter to a photo. These filters add a certain visual effect to the photo, for example black and white or intensifying shadows and brightening highlights. Filters are described by Instagram CEO and founder Kevin Systrom as follows: *"Our filters are a combination of effects – curve profiles, blending modes, color hues, etc. In fact, I usually create them in Photoshop before creating the algorithms to do them on the phone".* Instagram allows users to take, edit and share a photo within seconds.

Additionally, users can make use of hashtags (similar to Twitter) in their description which allows the specific posts to be found by other users and brands can use it to target a specific audience. Additionally, users can tag other users in the image or in the description, which means that they add a particular user to their post. For example, an image that has multiple people in it, when these people are also Instagram users they can be tagged by the user that posted the image. Now the post is now only visible on the page of the user that generated the posts, but it is also visible on the page of the tagged users.

Instagram is one of today's most popular social media platforms with 500 million active monthly users (source: https://instagram-press.com). Its users have shared over 50 billion photos to date and share an average of 95 million photos and videos per day. They "like" about 4.2 billion posts each day. It has also shown to be a particularly interesting platform for brands. In 2016, almost 50% of US brands was using Instagram for social media marketing and this has risen to almost 70% to date. A social media study conducted by Forrester in 2016 reviewed how the top 50 global brands market on social networks. Forrester evaluated 11.8 million user interactions on 2,489 posts made by 249 branded profiles, and collected data on how many top brands use each social network, how many followers they've collected, how often they post, and how often users interact with their posts. They found that the average number of Instagram followers for a top brand in 2016 was already over 1 million. The next section describes the Instagram dataset we have created.

Data

A grand variety of brands have adopted Instagram for their social media marketing. We have selected the brands based on their L2 Digital IQ index (https://www.l2inc.com). The L2 Digital IQ Index is based on the following questions:
1. How does your digital performance benchmark relative to competitors?
2. What near and long-term investments should you make to improve your digital performance? 3. How is your performance progressing over time so that teams and vendors are held accountable?

We have selected the top 1000 highest ranked brands based on this index. Subsequently, we have collected all posts of these brands over a 1-year period, starting on 05/01/2015 and ending 04/30/2014. To ensure an equal comparison between brands we have decided that out of the 1000 brands we only include brands that post at least once a week over the focal period, resulting in 148,226 corresponding to 633 brands across 27 different industries. Some industries can be considered more "visual" (e.g. fashion or travel industries) than others (e.g. financial industry), therefore we intend to analyze overall impact of the complexity measures across all industries as well as industry specific impacts.

The posts considered for this study are firm-generated content only, which means they are generated and posted on the brands fan pages. The visual complexity measures in this study and their impact on consumer engagement lead to managerial implications, particularly relevant to content managers. Therefore, we have not included user-generated content or electronic word-of-mouth because these lead to separate implications.

Method

The engagement of consumers towards firm-generated content on Instagram can be measured in two ways: number of likes and number of comments. The number of likes posts receive is a non-negative integer with a high variance. We perceive it follows a power-law distribution, something we have seen in many cases of social media prediction research (Gelli et al. 2015; Khosla, Das Sarma, and Hamid 2014; Mazloom et al. 2016). The majority of posts receive very few likes whereas a few posts receive up to a million likes. We observe a similar pattern for the number of comments, except the maximum number of comments is lower and we notice a significant increase in the number of zeros (i.e. 25.55 % that have received zero comments vs. 0.47 % that received zero likes).

Model Formulation

The number of likes and comments are both positive integers. Therefore, we model the engagement toward firm-generated content by applying and comparing two models suitable for count data: Poisson and Negative -Binomial. We compare the regular model to the zero-inflated version. A zero-inflated model might be most relevant to the number of comments considering the excess number of zeros.

Poisson: A Poisson distribution is parameterized by λ, which is both the mean and the variance (equidispersion). The equidispersion assumption is often violated, because a distribution of counts usually has a variance that's not equal to its mean (social media popularity in particular). Performing Poisson regression on count data that exhibits this behavior results in a model that doesn't fit well.

Negative-Binomial: In the case of overdispersion, where the poisson model fails. The negative binomial distribution, like the Poisson distribution, describes the probabilities of the occurrence of whole numbers greater than or equal to 0. Unlike the Poisson distribution, the variance and the mean are not equivalent. This suggests it might serve as a useful approximation for modeling counts with variability different from its mean. The variance of a negative binomial distribution is a function of its mean and has an additional parameter, θ, called the dispersion parameter. The variance can be described by

$\mathrm{Var}(Y) = \mu + \mu 2/\theta$. As the dispersion parameter gets larger and larger, the variance converges to the same value as the mean, and the negative binomial turns into a Poisson distribution. To test for the most appropriate model we perform a Likelihood Ratio (LR) test between both models. In the presence of Poisson overdispersion the LR test will reject the null hypothesis of theta being equal to infinity. Previous research (Lovett, Peres, and Shachar 2013; Rooderkerk and Pauwels 2016) has utilized the negative binomial to model post popularity as well.

Zero-Inflated and Hurdle models: We will not consider the hurdle type models, because a hurdle models the zeros as if they come from a separate data. generating process. In our case both zero or an arbitrary postive number of likes come from the same process. In case of excess number of zeros, a zero-inflated model, which is a mixture of Bernouilli probabilities and a count model, would be more appropriate. To test whether or not an excess of zeros is the case we perform a Vuong test after modelling both a regular negative-binomial regression and a zero-inflated negative binomial regression.

Variable Operationalization

The construction of variables corresponding to the measures as shown in the conceptual framework in figure 1 are described in the following section.

Dependent Variables

We extract the total number of likes and number of comments from each Instagram post in the dataset. These are the two main measures of engagement toward firm-generated content and we will investigate the impact of the complexity measures on these dependent variables separately.

Independent variables

Low-level complexity: The size of the image after compression, i.e. the number of kilobytes necessary to store the image after performing the jpeg algorithm to remove redundancies.

Mid-level complexity: We use canny edge detector (Canny 1986) to detect edges within an image. Canny edge detector is a pixel-level binary classification where a pixel is classified as 1 if an edge was found. The canny edge detection works by scanning for discontinuities in brightness and assigning ones to pixels where a discontinuity is detected. After performing edge detection on an image, we can easily construct a complexity measure by taking the total number of pixels on an edge divided by the total number of pixels within the image.

High-level complexity: we extract deep neural network features, initially proposed in (Szegedy et al. 2015) and trained to identify 15,293 ImageNet (Deng et al. 2009) concept categories. The deep neural network structure is pre-trained on millions of images to detect the presence of concepts within images. For each of the 15,293 concepts the model was presented at least 200 positive examples and it uses these examples to learn recognition. After we apply the pre- trained deep neural network to the images in our dataset it returns an output that is a distributional representation of the 15,293 concepts and their presence within each image. This representation then allows us to automatically determine how many objects there are within the image. We define a threshold value for correct classification and simply count all the concepts with a higher value than the threshold to construct the quantity of concepts measure. For the concept dissimilarity, we use WordNet Concept Similarity (Pedersen, Patwardhan, and Michelizzi 2004) to calculate similarities between detected concepts. The dissimilarity of concepts is one minus the average similarity between all the concepts detected in the image.

Control variables

Brand followers: The size of the audience to the Instagram images is reflected by the number of followers of the brand posting the images. Upon inspection, we observe that the number of followers is highly correlated with the number of likes. The number of followers of the brand will be included as a brand-level fixed effects. We perceive the number of followers of a brand shows a brands social media ability and it captures part of a brands overall popularity.

Brand activity: The number of posts represents brand activity. We perceive that when brands are actively working on engaging with their consumers it can increase their overall image popularity. Brand activity will also be included into the model as brand-level fixed effects.

McParlane et al. (McParlane, Moshfeghi, and Jose 2014) show how the time of posting affects image popular on social media platform Flickr. We follow their approach by including three time-dependent dummy variables to control for the consumer engagement.

Time of day: Images are either posted in the morning (06:00 am to 11:59 am), afternoon (12:00 pm to 06:00 pm), evening (06:00 pm to 11:59 pm) or night (12:00 am to 05:59 am).

Day of the week: Images are either posted during the weekend (Fri-Sun) or on a weekday (Mon-Thu).

Season: Images are posted during one of the four seasons in winter, spring, summer or autumn.

Additionally, we have information about the post itself that is time-independent as well as brand-independent.

Type of filter: Upon inspection of the data we perceive that the used filter may impact the user engagement.

Number of image tags: An image tag is a reference to some other user (person or brand) within the caption or image itself. We do not have information on the user that is tagged, we do know the number of users that are tagged in the image. The tag itself might lead to an increase in the number of views, because it now includes the audience of the tagged users on top of the brand followers that were already going to be exposed to the content.

Textual Sentiment: Visual and textual information are complementary for popularity prediction (Overgoor et al. 2017). Therefore, we include the positive and negative sentiment scores extracted from the image caption. We use Sentistrength (Thelwall et al. 2010) to calculate positive and negative scores ranging from 1 (no or very low valence) to 5 (very high valence).

Results

As expected we see significant impacts for all measures of visual complexity on both types of consumer engagement. The results in table 1 show the general regression results of 148,221 posts from 633 different brands across industries. Comparing these results with the results of the different complexity measures separately (not tabulated) we observe that there is no difference in the coefficients and that the different measures of complexity all separately impact the consumer engagement.

For the low-level complexity measure file size, we see a quadratic relationship with consumer engagement. An increase in file size leads to an increase in the engagement, however the file size squared shows a significant very little negative impact on the engagement. This suggests that at some point the file size negatively impacts engagement.

The mid-level complexity measure, has a positive impact on the number of likes. An increase in percentage of edges within an image leads to an increase in the number of likes. However, we observe the opposite with the number of comments.

The high-level complexity measures show mixed results. On the hand, we observe that a high quantity of concepts positively impacts both measures of consumer engagement whereas the dissimilarity of concepts has a negative impact. This reveals that a higher number of similar concepts detected on average leads to a higher increase in number of likes.

Robustness

We use negative binomial regression to model the number of likes and comments on Instagram posts. The results of the regression are presented in Table 1. We use the Chi-squared Test to find the

significance of the regression model, by computing the reduction in deviance from a null model. For our likes model, the test rejected the null hypothesis of a null model; hence, the model is well-suited to characterize the effects of the described variables. After inclusion of all control variables we reach to an adjusted R^2 value of 0.470 and 0.341 for likes and comments respectively
For our comments model, the test rejected the null hypothesis of a null model) The model for comments is also well-suited to characterize the effects of the independent variables.
We test coefficients of all independent variables for the null hypothesis of a zero-valued coefficient (two-sided) and find that the test rejects the null hypothesis in all cases.

In addition, we performed a likelihood ratio (LR) test to compare the negative binomial model to a Poisson model. Beforehand the structure of the data strongly suggested the use of a negative binomial model, because of expect over-dispersion. The LR test confirms this hypothesis and we reject the null hypothesis of equidispersion. We have also explored the possibility of a zero-inflated negative binomial model, however we found that there was no case of excess zeros in neither the likes or the comments

Table 1. Results

	Dependent variable:	
	Likes	Comments
File Size	0.004***	0.008***
	(0.0004)	(0.0005)
File Size Squared	-0.00002***	-0.00002***
	(0.00000)	(0.00000)
Edges	0.022***	-0.024***
	(0.002)	(0.002)
Quantity	0.082***	0.104***
	(0.008)	(0.010)
Dissimilarity	-0.058***	-0.035***
	(0.008)	(0.010)
Followers	0.001***	0.001***
	(0.00000)	(0.00000)
Posts	0.001***	0.0003***
	(0.00001)	(0.00001)
Caption Positive	0.011**	-0.010*
	(0.005)	(0.006)
Caption Negative	0.0004	0.080***
	(0.007)	(0.009)
Tags	-0.040***	-0.035***
	(0.001)	(0.001)
Filter	-0.583***	-0.785***
	(0.013)	(0.015)
Afternoon	0.508***	0.405***
	(0.014)	(0.017)
Evening	0.775***	0.752***
	(0.014)	(0.016)
Night	0.824***	0.776***
	(0.015)	(0.018)
Weekend	0.076***	-0.057***
	(0.010)	(0.013)
Spring	-0.162***	-0.201***
	(0.016)	(0.019)
Summer	-0.252***	-0.011
	(0.011)	(0.013)
Fall	-0.205***	-0.111***
	(0.011)	(0.013)
Constant	5.252***	1.924***
	(0.027)	(0.032)
Observations	148,221	148,221
Adjusted R^2	0.470	0.314
θ	0.3966*** (0.0012)	0.2799*** (0.0010)

Conclusion

This research shows the impact that different levels of visual complexity have on engagement of consumers with firm-generated imagery on social media. We have identified different types of complexity that can be easily identified and controlled by brand managers. Low-level complexity informs about overall variation and richness of the images. Pieters, Wedel and Bartra (2010) show how this type of complexity hurts the attention and attitude towards an advertisement. We show that this is indeed the case when the low-level complexity gets too high. However, we observe a quadratic relationship with the engagement which shows that a certain level of complexity is needed to stop the consumer and hold their attention and it is beneficial to keep this type of complexity in the mid ranges. The mid-level complexity shows mixed results, the edges impress the consumer and draw their attention, but it does not stimulate the user to go beyond liking by commenting. The high-level complexity type aids the engagement: A higher quantity of similar objects increases engagement. This type of visual complexity has engaging aesthetic qualities (Berlyne 1958) and is “glueing” the consumer to the image. A high quantity of objects will do this, but it is important to make sure that these objects are similar, so the image is not a collection of unrelated objects that will confuse the user.

We have taken existing theory in advertisement and we empirically investigated them on a large scale using social media data. The automatic extraction of the different types of complexity and their impact on engagement give brand managers the tools to improve the images that they post to increase the engagement with their target audience of consumers.

Future Direction

There are some alterations in the research that due to time constraint haven’t been processed and described in this paper. These will be touched upon in the presentation however. The following next steps are implement at this moment.

Instead of the three-level categorization of the complexity measures above we will distinguish two categories of visual complexity: Feature Complexity and Design Complexity. These different categories of complexity can be related to the gist of a scene as described by Oliva (2005). The author distinguishes perceptual and conceptual gist, where the former describes the basic image properties the brain uses to provide a structural representation of the scene and the latter includes the semantic information that is inferred while viewing the scene or shortly after.

The feature complexity measures are based on low-level visual processes in the primary visual cortex (Palmer 1999) evoked by the complexity of the image. It represents pixel-level variation and unprocessed or unstructured image information without regards to semantics. We distinguish three measures consisting of information about variation in the basic features of images – edge density, luminance and color variety. These three measures can be controlled by the content manager in such a way that engagement is optimized. A preview in the results show that some of them follow an inverted u-shape with the engagement, which means an optimal solution can be found.

The design complexity measures are based on the design and semantics of an image (i.e. objects and similarity). The complexity evokes mid-level visual processes based on objects and pattern recognition. The design complexity will be very similar to the high-level feature complexity described in this paper.

Additionally we aim to provide managerial implications focusing on the control and optimization of the measures.

References

Batra, Rajeev and Kevin Lane Keller (2016), “Integrating Marketing Communications: New Findings, New Lessons, and New Ideas,” *Journal of Marketing*, 80(6), 122-145.

Cameron, A Colin and Pravin K Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Canny, John (1986). “A computational approach to edge detection”. In: *IEEE Transactions on pattern analysis and machine intelligence* 6, pp. 679–698.

Chatterjee, Anjan (2004). “Prospects for a cognitive neuroscience of visual aesthetics”. In: *Bulletin of Psychology and the Arts* 4.2, pp. 56–60.

Deng, J. et al. (2009). “ImageNet: A Large-Scale Hierarchical Image Database”. In: *CVPR*.

Gelli, Francesco et al. (2015). “Image popularity prediction in social media using sentiment and context features”. In: *Proceedings of the 23rd ACM interna- tional conference on Multimedia*. ACM, pp. 907–910.

Khosla, Aditya, Atish Das Sarma, and Raffay Hamid (2014). “What makes an image popular?” In: *Proceedings of the 23rd international conference on World wide web*. ACM, pp. 867–876.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” *Journal of Marketing*, 80 (1), 7-25.

Kumar, V., Vikram Bhaskaran, Rohan Mirchandani, and Milap Shah (2013), “Practice Prize Winner—Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey,” *Marketing Science*, 32 (2), 194- 212.

Leder, Helmut et al. (2004). “A model of aesthetic appreciation and aesthetic judgments”. In: *British journal of psychology* 95.4, pp. 489–508.

Leeuwenberg, EL (1969). “Quantitative speci cation of information in sequential patterns.” In: *Psychological review* 76.2, p. 216.

Lovett, Mitchell J, Renana Peres, and Ron Shachar (2013). “On brands and word of mouth”. In: American Marketing Association.

Machado, Penousal et al. (2015). “Computerized measures of visual complexity”. In: *Acta psychologica* 160, pp. 43–57.

Mazloom, Masoud et al. (2016). “Multimodal Popularity Prediction of Brand- related Social Media Posts”. In: *ACM Multimedia Amsterdam, Netherlands*.

McParlane, Philip J, Yashar Moshfeghi, and Joemon M Jose (2014). “Nobody comes here anymore, it’s too crowded; Predicting Image Popularity on Flickr”. In: *Proceedings of International Conference on Multimedia Retrieval*. ACM, p. 385.

Overgoor, Gijs et al. (2017). “A Spatio-Temporal Category Representation for Brand Popularity Prediction”. In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*. ACM, pp. 233–241.

Palmer, Stephen E (1999). *Vision science: Photons to phenomenology*. MIT press.

Palumbo, Letizia et al. (2014). “Examining visual complexity and its influence on perceived duration”. In: *Journal of vision* 14.14, pp. 3–3.

Pecchinenda, Anna et al. (2014). “The pleasantness of visual symmetry: Always, never or sometimes”. In: *PloS one* 9.3, e92685.

Pedersen, Ted, Siddharth Patwardhan, and Jason Michelizzi (2004). “WordNet::Similarity: measuring the relatedness of concepts”. In: *Demonstration papers at HLT-NAACL 2004*. Association for Computational Linguistics, pp. 38–41.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010). “The stopping power of advertising: Measures and e ects of visual complexity”. In: *Journal of Marketing* 74.5, pp. 48–60.

Rooderkerk, Robert P and Koen H Pauwels (2016). “No comment?! The drivers of reactions to online posts in professional groups”. In: *Journal of Interactive Marketing* 35, pp. 1–15.

Shapiro, Linda and George C Stockman (2001). “Computer vision. 2001”. In: *ed: Prentice Hall*.

Simon, Herbert A (1972). “Complexity and the representation of patterned se- quences of symbols.” In: *Psychological review* 79.5, p. 369.

Szegedy, Christian et al. (2015). “Going Deeper with Convolutions”. In: *CVPR*.

Thelwall, Mike et al. (2010). “Sentiment strength detection in short informal text”. In: *J. Am. Soc. Inf. Sci. Technol.* 61.12, pp. 2544–2558. Wallace, Gregory K (1992). “The JPEG still picture compression standard”. In: *IEEE transactions on consumer electronics* 38.1, pp. xviii–xxxiv