The Impact of Visual Attributes on Online Image Diffusion

Luam Totti
Federal University of Minas Gerais (UFMG)
Belo Horizonte, MG, Brazil
luamct@dcc.ufmg.br

Felipe Costa
Federal University of Minas Gerais (UFMG)
Belo Horizonte, MG, Brazil
felipealco@dcc.ufmg.br

Sandra Avila
RECOD Lab., DCA / FEEC / UNICAMP
Campinas, SP, Brazil
sandra@dca.fee.unicamp.br

Eduardo Valle
RECOD Lab., DCA / FEEC / UNICAMP
Campinas, SP, Brazil
dovalle@dca.fee.unicamp.br

Wagner Meira Jr.
Federal University of Minas Gerais (UFMG)
Belo Horizonte, MG, Brazil
meira@dcc.ufmg.br

Virgílio Almeida
Federal University of Minas Gerais (UFMG)
Belo Horizonte, MG, Brazil
virgilio@dcc.ufmg.br

ABSTRACT

Little is known on how visual content affects the popularity on social networks, despite images being now ubiquitous on the Web, and currently accounting for a considerable fraction of all content shared. Existing art on image sharing focuses mainly on non-visual attributes. In this work we take a complementary approach, and investigate resharing from a mainly visual perspective. Two sets of visual features are proposed, encoding both aesthetical properties (brightness, contrast, sharpness, etc.), and semantical content (concepts represented by the images). We collected data from a large image-sharing service (Pinterest) and evaluated the predictive power of different features on popularity (number of reshares). We found that visual properties have low predictive power compared that of social cues. However, after factoring-out social influence, visual features show considerable predictive power, especially for images with higher exposure, with over 3:1 accuracy odds when classifying highly exposed images between very popular and unpopular.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*Data mining*

General Terms

Content diffusion, popularity prediction, image popularity.

1. INTRODUCTION

Online social networks have evolved from textual blogging tools to complex real-time systems of creation, consumption and diffusion of different media. More recently, the ubiquity of digital cameras has contributed to a rapid growth of image-sharing services such as Instagram, Tumblr and Pinterest. Image sharing is not restricted to dedicated services: Facebook, for example, reports visual information corresponding to the majority of reshared content [15], with more than 300 million images processed every day [41].

Therefore, image-sharing services have recently drawn the attention of researchers from many disciplines. The ability to predict image popularity (amount of views and reshares) has impact on advertising, viral marketing, and infrastructure capacity planning. Most works, however, approach prediction exclusively from a social perspective, focusing on the network structure, influence propagation, and temporal analysis [20, 1, 4, 42].

In this work, we take a complementary approach and evaluate the impact of visual attributes on image popularity. We define two sets of visual features, aesthetic and semantic, and analyze their impact on popularity (measured by the number of reshares) of images on Pinterest, a social network of large and increasing audience.

Image aesthetics is the perception of beauty by viewers [13]. It is challenging to extract features representing beauty, due to its abstract and subjective nature, but existing works show some consensus on what makes images more visually appealing [23]. Guided by that prior art, we have carefully choosen image features that encode important aesthetics properties. Our methodology comprises the design, implementation and evaluation of several of those features.

The semantic of images, understood as the identification of concepts represented in the image, stands on the other side of the spectrum of image analysis. Semantic analysis is a challenging open problem of Computer Vision, since visually similar images may portrait completely distinct concepts, and, conversely, similar concepts have much visual variability. The concepts to identify may be the concrete presence of certain classes of objects (e.g., people, cars), the nature of the image (e.g., landscape, interior scene), and even abstract notions (e.g., entertainment, violence). Perhaps due to the challenges of automatically identifying those concepts, semantic analysis is rarely employed on the study of image-sharing social networks. In this work, we have employed semantic features extracted with a state-of-the-art technique [38]. Each image receives a semantic feature vector of 85 dimensions, each quantifying the confidence on the presence of a concept the system was trained to recognize.

WebSci'14, June 23–26, 2014, Bloomington, IN, USA.

ACM 978-1-4503-2622-3/14/06$15.00. http://dx.doi.org/10.1145/2615569.2615700.

We also take into account social-network aspects, like number of followers, category tags, etc. The social attributes are both used as predictors, allowing to compare their predictive power to those of the image features, and also as a nuisance factor for the latter.

The original contributions of this work are:

- One of the first efforts to employ visual analysis in popularity and diffusion on online social networks from a mainly visual perspective;
- A compilation of aesthetics and semantics features, selected and implemented for the task. Those features are useful *per se*, and may be applied to other visual analysis tasks, such as recommendation, or retrieval. Social-network features were also selected and implemented;
- Collection of Pinterest resharing data, that we made publicly available[1]. Pinterest has a large and fast increasing popularity, presenting an interesting case for research.

2. RELATED WORK

Information diffusion on online services is a vastly researched topic [3, 4, 8, 9, 27]. Most of those works focus on designing metrics and models to quantify observable patterns of information diffusion.

Fewer researchers have looked into the users' motivations behind content endorsement actions such as 'retwetting', 'repinning' or 'liking'. Macskassy and Michelson proposed several models for explaining resharing behavior on Twitter, showing that users tend to retweet content on topics different to those of their own tweets, a behavior the authors called anti-homophily [33]. Suh et al. presented a large-scale analysis on how context features are associated to retweetability [43], concluding, among other findings, that the presence of URLs and hashtags correlate positively with retweeting. Stieglitz and Dang-Xuan extended that work by investigating how sentiment alignment affects retweetability on politically engaged content [42]. They found that neutral tweets are less likely to be reshared than polarized tweets (positive or negative), although no significant distinction could be found between those two alignments. Zarella [48], in a series of blog posts, presents practical advice for content creators. Analysing his data, he suggests the inclusion of hashtags, images, and URLs on tweets to increase resharing.

Existing art on image-sharing services tends to focus on social-network aspects, such as user influence, and social ties. Anagnostopoulos et al. developed a statistical test to distinguish causal social influence from simple correlation by examining the spread of picture tags in Flickr [1]. Lerman and Jones investigated photo propagation on Flickr, concluding that the social environment of users plays a significant role on the diffusion of the images [26]. Cha et al. support and extend those results by considering multiple hops on the social network around each user [10]. This paper, by focusing on image visual content, is complementary to all those works that focus on social aspects and user interaction.

Two very recent works have taken into account visual information. Khosla et al. [25] analyzed the predictive power of both visual and social features in the resharing of images on Flickr. Their prediction models successfully predicted, to some extent, image popularity on different settings, such as *one-image-per-user* or *user-specific*. Important differences from our work are the service studied (Flickr vs. Pinterest), and our analysis of visual features accross the spectrum of strong predictive social-features, like number of followers, an approach in which we use social properties as a nuisance factor. Cheng et al. [11] aimed at predicting cascades of reshares for images on Facebook, modeling resharing as a temporal process, and using the past to predict the future, in a scheme that reveals interesting insights on resharing process. They were able to predict well, for images that were reshared k times in the past, whether or not they would be reshared $2k$ times in the future. Our work is different both in the metric of popularity we are trying to predict, and in the visual features evaluated, since their work does not explore the aesthetic features. Remark that those works were unpublished at the time we completed our experiments, we came in contact with preprints as we were finishing the writing of the paper. As we will show, our work supports the conclusion of both works, that visual features are less predictive of popularity than user and network features.

The aesthetic features we evaluate were proposed on works focused on aesthetics assessment. Early works employed low-level features explicitly designed to quantify perceptual quality. Such features vary from simple channel statistics to complex blur estimation and region segmentation techniques. The works of Datta et al. [13] and Ke et al. [24] stand as the first efforts to infer aesthetic quality by applying Machine Learning techniques on those features, showing that aesthetics can be successfully inferred to some extent. Later works extended and improved the features [14, 22, 31], offered insights for the handling of images in specific corpora (e.g., paintings [28], images with faces [29]), and integrated image-enhancing systems [5]. Although those works yield good and interpretable results, custom-designed features cannot be exhaustive due to the diversity of both perceptual attributes and image corpora. Therefore, recent works have introduced more general visual features as an alternative to the hand-crafted ones. Marchesotti et al. [35] employ GIST and SIFT low-level descriptors, with a bag-of-visual-words and Fisher Vector mid-level descriptors, to infer aesthetic quality more accurately at the cost of less interpretable results. Attempting to achieve both accuracy and interpretability, Marchesotti and Perronnin [34] employed Machine Learning on images and associated textual comments to automatically discover and learn visual attributes.

3. DATA COLLECTION

The data used in this work was entirely collected from Pinterest[2], a recent image-sharing web service, brought to prominence as the fastest-growing large commercial social network [40]. In 2011 alone, the service grew 4000% in number of visits. At the end of 2013, Pinterest had the highest growth rate among all sharing channels, including Facebook, and, as of March 2014, it stands as the fourth most popular social network in number of unique accesses per month [16].

Despite Pinterest drawing much attention from mass media, few academic works aimed at understanding its dynamics [37, 17]. We believe to be the first work to study image

[1]https://github.com/luamct/WebSci14

[2]http://www.pinterest.com

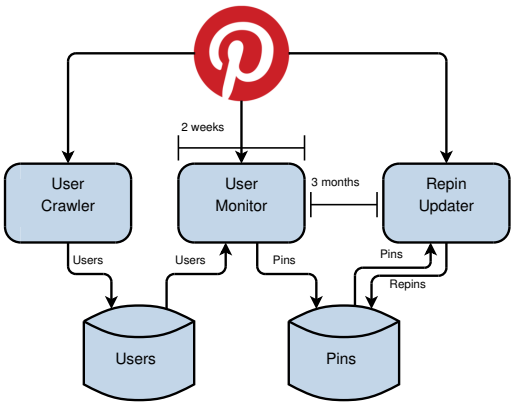

Figure 1: Flowchart of the data collection process. Because Pinterests has no data collection API, a scheme was created to obtain a measure of repins over a certain time span.

popularity on Pinterest. By making the collected data available, we hope to incite researchers to further investigate that network and its dynamics.

3.1 Pinterest Platform

Pinterest uses the metaphor of a pin-board as a collection of images (pins) within some topic of interest. Users can post on their boards by 'pinning' images from the Internet, uploading local content or 'repinning' existing pins (much like retweeting in Twitter). Users follow specific boards and may comment, like, or repin any posted pin. Although users have a followers and followees count on their profile, those values are a simplification, given that users do not follow other users directly, but individual boards. Both counts are calculated using an "at least one" logic, i.e., if a user A follows at least one board from user B, then A is counted as a follower of B , and B is counted as a followeee of A.

Pinterest adopted unusual strategies to estimulate the sharing of higher-quality content. From its conception, the service was promoted for "people with good taste", with sign-up available at invitation only. Designers intentionally avoided providing ranks of popular users, or even recent trends, to discourage competition among users, or the usage of Pinterest as a news media. Those strategies, along with a clean and elegant interface design, successfully enforced the importance of visually-appealing content above personal or informational content. That also makes Pinterest fairly agnostic to external events or trends when compared to other services [36]. Those characteristics make Pinterest particularly suited for this work, since our main goal is to investigate how visual properties affect image popularity.

In this work we consider the number of repins an image has received as an assessment of its popularity. Although the reasons driving resharing actions are numerous [7], it is vastly accepted in literature that resharing can be seen as an endorsement action, and, therefore, is a reasonable candidate for quantifying popularity [43, 7, 20].

3.2 Data Acquisition

Having no official available public API, the data was collected with HTTP requests over the publicly available information, emulating a regular user browsing the service. Each request was able to retrieve at most 50 pins, due to the layout of the pages returned by Pinterest. Such restrictions imposed some limitations on the collection process, both in terms of volume and completeness of the final dataset. Since our goal is to investigate image popularity, we required repin information about each pin. However it would be unfair to claim that a pin is more popular (has more repins) than other if each has been been exposed for a different duration. At the same time, Pinterest web interface does not provide the precise date the pin was posted, making impossible for us to select pins posted in a given time span. For that reason we performed our collection as a multiple step process:

1. Collected Pinterest user handlers through a breadth-first search starting with a few manually selected users.
2. Monitored the collected users over a span of time for collecting timestamped content.
3. Collected the number of repins of the pins collected on step 2, by later revisiting their Pinterest pages.

Figure 1 summarizes those steps. That process allowed us to collect data with proper timestamps and repin information after the same exposure time on the network. The process started with the collection of approximately 210K user identifiers through a breadth-first search, starting from a small group of manually selected popular users (since no user rank is provided). We understand the limitations of BFS sampling over large networks [18], however due to the lack of a public API, or even numerical identifiers for users and pins, we were left with no better alternative. We then monitored the collected users' activities during the course of two weeks (19th April to 2nd May of 2013) and collected all posted content (around 2 million pins). From that set we randomly selected $10,000$ users and their corresponding pins, consisting of $473,665$ pins, to make processing manageable. To collect repin information we revisied the Pinterest pages for the selected $473,665$ pins after approximately 3 months (July 27), ensuring that the images had roughly the same exposure time in the network.

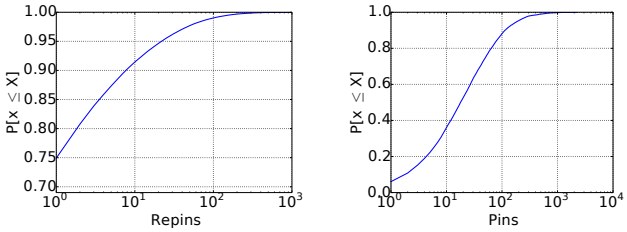

Figure 2: Cumulative distributions for (a) repins per pin and (b) pins per user in the dataset.

To give an overview of the characteristics of the collected data, we show the cumulative distribution of the repins each pin received in Figure 2(a) and of the number of pins posted by each user in Figure 2(b). Figure 2(a) shows a heavy-tailed distribution with 75% of the pins having 1 or less repins and less than 2% of the pins having more than 100 repins. Figure 2(b) shows a less skewed distribution of pins across users, with 50% of the users posting at least 20 pins during the monitored period, but with less than 10% posting

more than 100 pins. Given the highly skewed repins-per-pin distribution (nearly 60% of the pins have 0 repins), and the main objectives of this work, we performed all analyses only with the pins repinned at least once, which reduced the count of pins to 187,796.

4. IMAGE FEATURES

We divided our features into three major groups according to what they encode: visual aesthetics properties, semantic information, and social-network properties. The features employed are summarized on Table 1 and detailed below.

4.1 Aesthetic Features

The development of informative and interpretable visual features for assessing aesthetics properties remains a challenging problem. Even where there is a consensus on what makes images beautiful, efficient features must be designed to properly encode the intended visual notion. The feature selection and design in this work was based upon photography techniques, viewers' intuition, and results from previous works. Features used in different applications, like image retrieval and visual memorability tasks, were also considered [21, 47].

The images were first scaled down to approximately 200,000 pixels while keeping their original aspect ratio. They were then converted to a cylindrical color space (IHSL), which represents color in a more human-friendly way [19].

Channel Statistics: The Hue channel encodes color tonality (i.e., where in the spectrum the color is). Saturation encodes chromatic purity (pure full colors vs. diluted or "pastel" colors). Luminance encodes brightness, the amount of light energy in the color. We compute the mean and standard deviation on pixel values for those three channels. Circular statistics were employed for Hue, since it is an angular measure.

Basic Colors: Colors are one of major components on images. Colors evoke different sentiments and feels on viewers and are deliberately exploited by artists, designers, and photographers. We count the basic colors of each image using the method of Weiber et al. [45].

Dominant Colors: We consider dominant colors as the smallest set of basic colors that occupy 60% of all pixels. The threshold 60% was empirically found to maximize the distance between images with few repins from those with many repins.

Colorfulness: We implement Datta et al. [13] colorfulness metric as an additional quantification of the diversity of colors in the images. This metric divides the RGB color space into 64 equal cubes and computes an histogram over the pixels with those cubes as bins. A hypothetical perfectly colorful image is encoded as a histogram in the same manner and the colorfulness of the target image is taken as the Earth Mover's distance between the two histograms.

Contrast: Proper use of contrast is another important property. Images presenting wider ranges of luminance values are usually perceived as having better contrast. We quantify that by computing a normalized luminance histogram of the image, and taking as metric the size of the minimum contiguous interval of luminance values that concentrates at least 98% of total image luminance (i.e., we count the smallest number of contiguous bins that sum to 0.98).

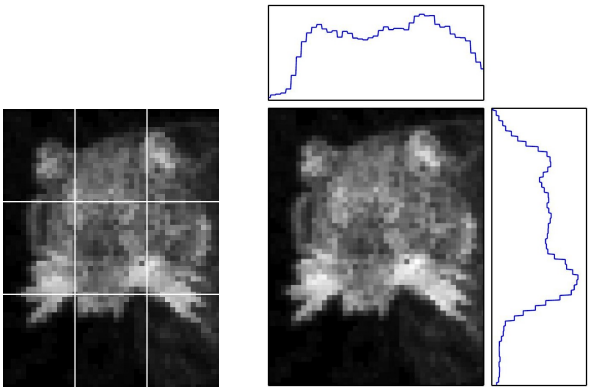

Figure 3: Sharpness-based features. (Left:) Sharpness map superimposed with the 'thirds' grid : region focus is extracted as the mean sharpness of each of those 9 regions. (Right:) The sum of sharpness over rows and columns is employed to measure focus centrality, focus density, and agreement with the 'rule of thirds'.

Aspect Ratio and Resolution: Aspect ratio is given by the ratio between the width w and heigh h of the image, while resolution is given by $w \times h$.

Complexity: Image complexity gives cues about aesthetic value, because simple compositions tend to have more appeal. We quantify that effect by using the number of regions obtained by a segmentation algorithm [12]. Cluttered images tend to segment into many small regions, simpler images tend to generate few large regions.

Texture: Texture, which encodes the perceptual qualities of graininess, smoothness, and directionality, is an important aesthetic cue. Following a similar procedure from previous works [13, 32], we apply a three-level wavelet transform on all three color channels, and summarize information into three features.

Art theory and professional photography rely on rules of spatial composition. Studies show that different compositions trigger different stimuli on observers, also affecting the perceived image quality. The features below explore that.

Region Focus: A widely used technique on artistic photography is to limit the depth of field, i.e., to deliberately blur some regions so as to bring focus and attention to the objects of interest. On the other hand, unintentional blur is perceived as poor technique, degrading aesthetic value. We implemented Vu et al. [46] S_3 algorithm for mapping sharpness levels. Although the concept of sharpness can be subjective, the S_3 algorithm achieves good results by combining both spectral analysis and local contrasts to create a sharpness map, quantifying the sharpness of each pixel. We take $Z(x, y)$ as the normalized sharpness of pixel x, y, i.e., Z as the ℓ_1 normalization of the sharpness map. We define nine spatial features as the mean pixel sharpness for each region of a 3×3 grid over the image, as shown in Figure 3. High quality images are expected to concentrate sharpness on inner regions, while poor images are expected to scatter sharpness across more regions.

Focus Centrality: We measure the centrality metric of the sharpness map (explained above). We compute the sum of normalized sharpness for each row ($Z_{row}(y)$) and each column ($Z_{col}(x)$) of the image, as

shown in Figure 3. Rows (columns) centrality is obtained by summing over the sharpness sum of each row (column), attenuated by the squared normalized distance of each row (column) to the center of the image, i.e. $c_{row} = \sum_y Z_{row} \times \left(1 - |y - (h-1)/2| \times (h/2)^{-1}\right)^2$ (analogously for c_{col}). Image sharpness centrality value is the product of the two centralities $c_{row} \times c_{col}$.

Focus Density: From the sharpness map we extract the sharpness density of the image. We measure row (column) spread as the minimum contiguous number of rows (columns) whose total normalized sharpness corresponds to 80% of total image sharpness. For rows, the spread ρ_{row} is $\min_{y_e, y_s} [y_e - y_s]$ subjected to $\sum_{y=y_s}^{y_e} Z_{row}(y) \geq 0.8$ (ρ_{col} is defined analogously). The density measure is given by $1 - \rho_{row} \times \rho_{col}$.

Background Area: Noting that the foreground and background regions of images are mainly defined by boundaries of color and sharpness, we derived a simple but effective background detection algorithm from the sharpness map and segmented image. For each region Q of the segmented image we calculate a vector of four averages $(\bar{Q_Z}, \bar{Q_L}, \bar{Q_a}, \bar{Q_b})$, corresponding, respectively, to the mean pixel value for the sharpness Z, and for each one of the channels of a La*b* color space. We then employ a 2-means clustering on the regions over the vectors $(\bar{Q_Z}, \bar{Q_L}, \bar{Q_a}, \bar{Q_b})$ in order to find the two major regions. Finally, we take the region with lower mean sharpness as the background. Figure 4 illustrates the steps of the algorithm. The final metric is the fraction of the image occupied by the background.

Rule of Thirds: is a commonly guideline for good composition, stating that objects of interest should be placed near to one of the four intersections of the 'thirds' of the image (Figure 5). Agreement to the rule of thirds is measured by the density of sharpness around the 'thirds', i.e., as the sum of normalized sharpness of pixels ponderated by a Gaussian window centralized on the'thirds'. More formally, the agreement for each horizontal axis ($y_a = h/3$, $y_a = 2h/3$) is given by $\sum_y Z_{row}(y)\, \mathrm{N}_{y_a, \beta^{-1}}(y)$, where $\mathrm{N}_{\mu,\sigma^2}(y)$ is the value at y of a Gaussian distribution with mean μ and variance σ^2 (the contribution of horizontal axes $x_a = w/3$ and $x_a = 2w/3$ is defined analogously). The final metric is the sum of the contributions of the four axes. The concentration parameter $\beta = (\sigma^2)^{-1}$ controls the spread of the Gaussian, the bigger it is, the more strict the metric is in terms of proximity to the axes. We set $\beta = 160$ in our experiments. Figure 5 illustrates the process.

4.2 Semantic Features

For our semantic features, we employ one supervised image classification for each concept, and use the confidence scores given by the concept classifiers as a feature vector. Image classification is a challenging, and highly active research topic, with an extended range of applications. A typical classification scheme consists mainly of three steps: (i) low-level local features extraction, (ii) mid-level global feature extraction, and (iii) supervised classification. Those steps are explained below.

4.2.1 Low-level Features

The low-level local features are extracted directly from the image pixels, sampling different regions of the image. Although purely perceptual, the local descriptors provide invariance properties that make them good building blocks for more complex representations. The regions may be densely sampled on a grid of overlapping windows of different scales, or they may be sparsely sampled by a detector of regions of interest.

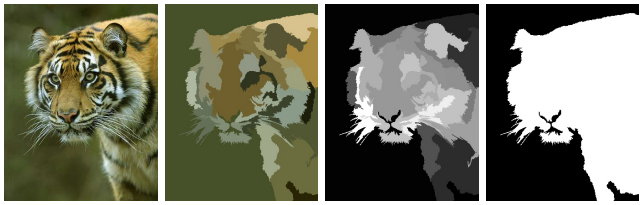

Figure 4: Background detection algorithm. From left to right: (1) Original image; (2) After image segmentation, the mean color $(\bar{Q_L}, \bar{Q_a}, \bar{Q_b})$ on each region Q is computed on the La*b* color space; (3) After image segmentation, the mean sharpness $\bar{Q_Z}$ is computed using the sharpness map Z; (4) Regions are clustered with a 2-means over the vectors $(\bar{Q_Z}, \bar{Q_L}, \bar{Q_a}, \bar{Q_b})$, the region with lesser overrall sharpness is chosen as background.

In this work we adopt the widely used SIFT local descriptor [30], which has consistently shown good results on image classification tasks. It is invariant to scale and rotation, besides being invariant to affine illumination changes. We extract the descriptors on a dense spatial grid with a step-size of half the patch-size, over 8 scales separated by a factor of 1.2, with the smallest patch-size set to 16 pixels. As a result, roughly 8000 descriptors are extracted from each image in the dataset. Each SIFT descriptor had its dimension reduced from 128 to 64 by applying Principal Component Analysis (PCA).

4.2.2 Mid-level Features

Even with a highly robust and comprehensive extraction of low-level features, bridging the semantic gap between pixel values and real concepts and entities requires substantially more complex representations. Mid-level features play that role by aggregating low-level descriptors into a global and richer image representation, in a scheme known as Bags of visual Words (BoW). In the BoW model, unsupervised learning is employed to quantize the low-level feature space, establishing a codebook of representative visual appearances. Then, the feature vector of an image is created by encoding its low-level features in relation to the codebook, and pooling over all codes, in order to create a single feature vector. The BoW model and its extensions are active research areas [6, 2, 38].

In this work, we employ as mid-level representation the state-of-the-art Fisher Vectors [38], an extension to the BoW model that encodes how much the first and second moments of the low-level descriptors present in the image deviate from the global distribution found on the dataset. In Fisher Vectors, the codebook is learned with an Expectation–Maximization algorithm to estimate a Gaussian mixture model (GMM) over one million low-level descriptors sampled from the training set. The mid-level feature vector is the sum of the Fisher scores, over the learned GMM, of each low-level feature. The details of the representation go beyond the scope of this work and can be found in [39].

4.2.3 Supervised Learning

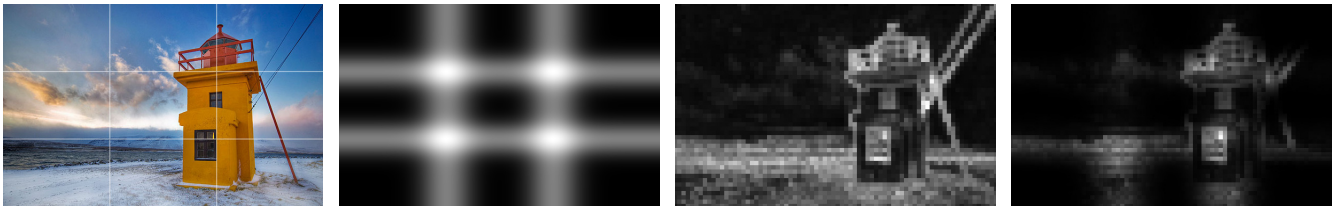

Figure 5: The 'rule of thirds' metric. From left to right: (1) Original image with the axes drawn; (2) Attenuation map corresponding to the contributions of the four axis, each axis contribution being ponderated by a Gaussian window around it; (3) Sharpness map Z generated using the S_3 algorithm; (4) Pixel-wise product of sharpness map and attenuation map. The final accordance metric is the sum of all pixels in the product map.

Finally, supervised learning is applied over the mid-level representation, in order to learn a statistical model for each concept, using a training set of annotated images. In this work we adopt the ImageCLEF 2012 Photo Annotation dataset [44] as our training set. The dataset consists of 25,000 images, of which we employ the training set of 15,000 instances. The dataset contains 94 concepts including natural elements (e.g., day, snow, fire), environment (e.g., coast, plant, bird), people (e.g., baby, female, small group), and human elements (e.g., car, bicycle, air vehicle). We excluded 9 concepts that we considered to be too related to aesthetics properties (e.g., quality_noblur, style_overlay, etc), leaving us with 85 semantic concepts.

When employing the BoW model, Support Vector Machines (SVM) are often the classifier of choice, due to its ability to learn in very high-dimensional spaces. We use it to perform one-versus-all classification using a linear kernel, since previous works show that Fisher Vectors do not benefit from the slower non-linear kernels [38]. A different classification model is learned for each concept.

The final semantic feature vector for an image is the concatenation of the z-score normalized confidences output by the trained model for that image.

4.3 Social Features

To better understand the predictive power of visual features, we also employ features extracted from metadata about users, images and the social network. We call them social features, since they are mainly derived from the users information and interaction with the service. For each pin P posted by user U on the pinboard B we define the features shown on Table 1.

The category of a pin is defined as the category of the board B in which it is pinned. Pinterest offers 33 different categories (e.g., Architecture, Cars, Food and Drink, Women's Fashion, etc.). Previous versions of the service allowed users to leave boards uncategorized, so around 43% of the boards on the dataset still have no category. For the pins in those uncategorized boards we assign an extra *empty* value.

Users may post pins in different ways, such as uploading images, pinning an image from an external domain, or repinning an image already in Pinterest. Binary feature *is_repin* is true only for repins. We also measure, for each user U, the fraction of pins that are repins.

We include two more pin-specific features: the length in characters of the description provided by the creator, and the day of the week the pin was posted. We also include the total number of pins in the board B where the pin was posted.

Given the important role creators play in the diffusion of their messages [1], we employ many user and social features. User profile gives *Gender*, which can be *empty* if is not provided by the user (often the case for institutional and commercial accounts). The binary feature *has_website* is true for users that list an website in their profile (also indicates commercial accounts that use Pinterest as a visual display for products on sale) [37]. Pinterest deals with products by adding a dollar sign ($) in the description of pins that represent products on sale, an information we encode in the binary feature *is_product*.

Feature *#user_followees* is the number of users the pin creator U follows. Since users follow specific boards, that number refers to all users that have at least some board followed by user U. Although the service offers board granularity for following, in practice users tend to follow either all or no boards of the followees [37].

Feature *category_entropy* encodes how general users are regarding the categories of their posted content. Users may specialize in posting on a few categories, or they may post content on many categories. We quantify this by calculating the Shannon entropy of the distribution of categories used on all pins posted by user U. As mentioned, pins that belong to uncategorized boards are also considered uncategorized. The feature *uncategorized* calculates the percentage of uncategorized pins posted by user U.

Finally, the features *#boards* and *#pins* hold the total number of boards user U has created, and the number of pins U has posted.

5. EVALUATION

To evalute the impact of the different features on image diffusion, we employ a classification scheme (using supervised learning) to discriminate between two classes of very popular and very unpopular pins (excluding from the analysis the middle ground of average popular images). The experimental design divides the dataset into a training and testing sets. Accuracy on testing is used as a measure of the features predictive power. A 5-fold cross validation is employed to partition the training and data sets on the experiments, and the average accuracy is reported.

5.1 Popularity Prediction

When dealing with popularity on social networks, precise predictions are extremely difficult to obtain due to a multitude of factors. Fortunately, for most purposes it's suffi-

	Name	#	Brief Description
Aesthetics	*Channel Statistics*	6	Mean and standard deviation of each channel: Hue, Saturation and Brightness.
	Colorfulness	1	Image's distance (EMD) from a hypothetical perfectly colored image.
	Basic Colors	11	Amount of pixels of each basic color: black, blue, brown, green, gray, orange, pink, purple, red, white, yellow.
	Dominant Colors	1	Minimum number of basic colors that cover 60% of the image.
	Aspect Ratio	1	Width divided by height.
	Resolution	1	Width multiplied by height.
	Contrast	1	Measure of the dispersion of histogram of Luminance pixels.
	Texture	3	Roughness and smoothness of the image texture measured by the Wavelet transformation of each channel.
	Complexity	1	Number of regions after *mean-shift* segmentation.
	Region Focus	9	Mean sharpness value in each region on a 3×3 grid over the image.
	Focus Centrality	1	Concentration of sharpness values around the center of the image.
	Focus Density	1	Dispersion of values on the histograms of sharpness on each dimension.
	Background Area	1	Percentage of background area.
	Rule of Thirds	1	Accordance to the rule of thirds.
Semantics	*Concepts*	85	SVM detection confidence for: *view, celestial: (sun, moon, stars), (portrait, closeupmacro, indoor, outdoor), style: (pictureinpicture, circularwarp, graycolor, overlay), combustion: (flames, smoke, fireworks), sentiment: (happy, calm, inactive, melancholic, unpleasant, scary, active, euphoric, funny), gender: (male, female), age: (baby, child, teenager, adult, elderly), flora: (tree, plant, flower, grass), water: (underwater, seaocean, lake, riverstream, other), weather: (clearsky, overcastsky, cloudysky, rainbow, lighting, fogmist, snowice), lighting: (shadow, reflection, silhouette, lenseffect), scape: (mountainhill, desert, forestpark, coast, rural, city, graffiti), relation: (familyfriends, coworkers, strangers), fauna: (cat, dog, horse, fish, bird, insect, spider, amphibianreptile, rodent), timeofday: (day, night, sunrisesunset), quality: (noblur, partialblur, completeblur, motionblur, artifacts), setting: (citylife, partylife, homelife, sportsrecreation, fooddrink), transport: (cycle, car, truckbus, rail, water, air), quantity: (none, one, two, three, smallgroup, biggroup).*
Social	*Category*	(34)	Pin's category defined as the pin's board category.
	Is Repin	(2)	Whether the pin was itself a repin from another pin already in Pinterest.
	Is Product	(2)	Whether the pin is depicts a product for sale.
	Desc. length	1	The size in number of character of the pin's description.
	Day of the Week	(7)	Day of the week the pin P was posted.
	#Board pins	1	Number of pins posted on board B.
	Gender	(3)	The gender of user U as registered in the profile.
	Has Website	(2)	Whether the user U has a website registered in his profile.
	#User Followees	1	Number of followees of user U.
	Category Entropy	1	Entropy of the categories of all of user U's pins.
	Uncategorized	1	Percentage of uncategorized pins of user U.
	Repined	1	Percentage of repins within all pins of user U.
	#Boards	1	Number of boards created by user U.
	#Pins	1	Number of pins posted by user U.

Table 1: Extracted features for a given pin P posted by a user U on a board B. The columns # refers to the dimensionality of the feature vector. The values between parenthesis indicate categorical variables that can assume the number of values shown (*Gender* and *Category* can be unknown, explaining the extra possible value). The *Concepts* employed in semantic analysis are listed hierarchically for readability and contextualization: the detection algorithm actually employs a flat labeling using the concatenation of category and subcategory (e.g.: celestial_sun, celestial_moon, celestial_stars).

cient to foresee whether an image will be *highly popular* or *unpopular*. Therefore, we reduce the problem to a binary classification task into *unpopular* and *popular* pins. More exactly, letting r_i be the number of repins a pin i has received, we define threshold values λ_- and λ_+, and label a pin i as *unpopular* if $r_i < \lambda_-$, and as *popular* if $r > \lambda_+$. The pins between the thresholds are excluded from the analysis. To balance the classes, we set λ_- and λ_+ according to a separation parameter Δ that represents the percentage of the data discarded in the middle section. For example, $\Delta = 0.7$ means that the pins in the top and bottom 15%-rank of repins were used respectively as the popular and unpopular classes, while the remaining 70% of the pins were ignored.

For all classification results we employed a Random-Forest ensemble of 200 tree estimators with strong randomization on both attribute and cut-off choices. Since the task is a

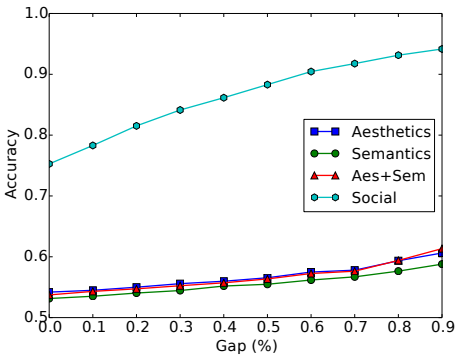

Figure 6: Accuracy of classification into *very popular* and *unpopular* for varying values of the gap Δ separating those two classes, and for different feature sets.

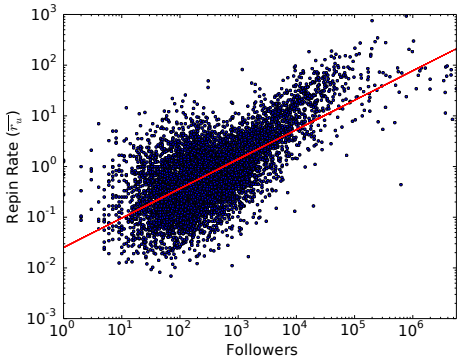

Figure 8: Linear Regression on the log of the mean repin rate given the log of the number of followers of the user.

balanced binary classification, we used accuracy as the evaluation metric, and performed a 5-fold cross-validation over the dataset, in order to obtain the averages and standard deviations of accuracy over the 5 runs.

Figure 6 shows the accuracy for increasing values of the gap Δ. The *Aes+Sem* employs early fusion of both semantical and aesthetical features, concatenating the respective features. Not surprisingly, the social parameters are much more informative in the prediction of popularity than the visual features. This is probably because some social features implicitly encode user popularity, an important factor to predict future posted pin popularity. The aggregated effect of visual features performs a little better than random, but their impact varies widely for different classes of users as we will show later in this section (see Fig. 9).

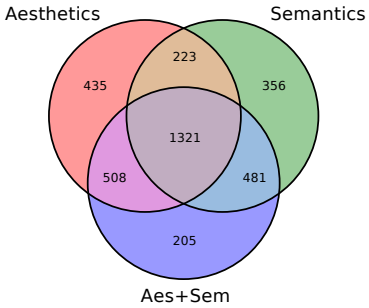

Figure 7: Venn diagram for the correctly classified images using different sets of features.

A particularly intriguing result is the similarity of the curves for aesthetics and for semantics features. Given their very distinct nature and derivation, that result is unexpected. To understand this behavior, Figure 7 shows a Venn diagram of the correctly classified images for all combinations of feature groups (separation $\Delta = 0.7$). We sampled a test set of $4,500$ images for that analysis and trained the classifier with the remaining images. The numeric values labeling each region represent the number of images correctly classified using the corresponding set of features. Although there are 1,321 images correctly classified by all sets of features, 435 images were only correctly classified by the aesthetics features and 356 images were only correctly classified by the semantics features. Furthermore, by merging the two sets of features the classifier is able to identify correctly 205 new images but at the same time misclassify 223 images that were properly identified by the two sets of features separately. That suggests an interesting feature complementarity that could be leveraged in future works. It is still unknown why the classifier was unable to exploit that, given that we employed ensemble techniques with randomized choice of features for the composing tree estimators. Further investigation is required to illuminate that point.

5.2 Factoring-out Social Influence

To better understand the impact of visual features throughout the spectrum of users, we proposed to treat the user popularity (measured as their average number of repins per pin) as a *nuisance factor* and check whether we could improve popularity classification after removing the effects of that variable.

Let f_u be the number of followers user u has and $\bar{r_u}$ be the repin rate of user u, i.e., the average number of repins each pin of user u received. In order to treat f_u as a nuisance parameter we use part of the training set to fit a standard linear least squares model on $log(\bar{r_u}) \sim log(f_u)$ (see Figure 8). By doing this we obtain a regression function $h(f_u)$ that estimates the average number of repins/pin for a user u given their number of followers. Although the function was fitted to user data, we can transfer what we learned to each pin i by providing f_i as argument, which is the number of followers of the board pin i was posted. The predicted value $h(f_i)$ represents the expected number of repins pin i should have received considering only its exposure level.

We then apply a data transformation $\delta_i = r_i - h(f_i)$ for each pin i in order to remove the influence of the number of followers over the number of repins. Basically we are taking the repin residue in log scale of the regression prediction. Finally, we perform the binary classification task as before, but using the residues δ_i instead of r_i. By doing this we are attempting to explain the deviation of the observed number of repins from the expected number of repin given a number of followers.

Figure 8 shows the regressed linear function with each point being a user in the data and the coordinates given by the number of followers and the repin rate (average repins per pin). Figure 9 shows the classification performance of visual features for the transformed variable δ_i. Compared

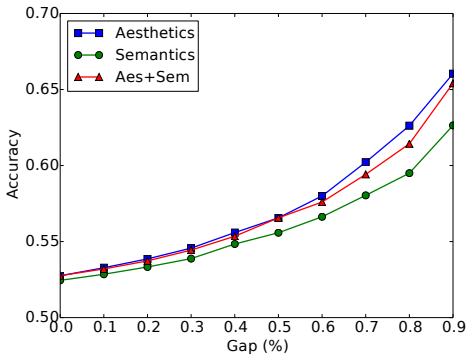

Figure 9: Factoring-out social influences: for different feature sets, accuracy of classification into *very popular* and *unpopular* for varying values of the gap Δ separating those two classes. The classes are defined on the the residue δ_i obtained by subtracting the influence of number of *user* followers, and indicate the deviation from the expected number of repins given the number of followers.

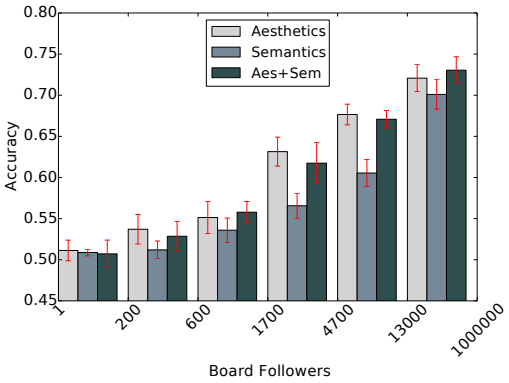

Figure 10: Blocking by number of *board* followers, for different feature sets. The bars plot the accuracy of binary classification for a gap $\Delta = 80$. The classes are defined on the residue δ_i obtained by subtracting the influence of number of board followers, and indicate the deviation from the expected number of repins for that amount of board followers. Error bars are standard deviations.

to Figure 6 the improvement is considerable, with the visual features attaining near 3:1 accuracy odds for the larger gaps.

5.3 Controlling for the Number of Followers

The fact that popularity indeed acts as a nuisance factor for the prediction ability of visual features is further confirmed in Figure 10, where we investigate the impact of the visual features for pins from boards with different number of followers. Each group of bars represent the classification accuracies of only pins from boards with followers within the values in the x axis. The followers intervals were chosen to be as logarithmically separated as possible while maintaining roughly the same number of pins within each interval. The error bars represent the standard deviation within the 5 cross-validation folds. The results show that visual features predict popularity better for pins with higher exposure. It is currently unknown in which direction the causality goes: are visually minded boards more likely to become popular? Or do visually-minded users gravitate towards popular boards, where they are prone to find visually appealing content?

Another interesting question is what explains diffusion of the less-exposed pins, since the visual attributes seem less important in this case. Those are still open questions. Given Pinterest's unusually high regards for visually appealing content, performing those same analyses on a different online service would probably bring interesting insights on those questions.

6. CONCLUSION

In this work we investigated content popularity on Pinterest, a relatively recent online image-sharing service that has a large and growing audience. As expected, social parameters, containing important hints about the popularity of *users*, have the most predictive power over the popularity of *pins*. At first, the aggregated effect of visual features seemed a little better than random, but a finer investigation revealed that the predictive power of visual features is considerable over the pins that have greater exposition (those pinned on boards with more followers) reaching over 3:1 accuracy odds for the pins with larger exposition. Although that does not seem much, when compared to the 4:1 to 20:1 accuracy odds of social features, one has to keep in mind that visual features operate at a much lower level and are intrinsically very imprecise, due to the fact they are the result of automated algorithms. Therefore the predictive power we obtained hints at a lower bound on what could be obtained with future advanced visual features either designed or learned for the task.

As an aditional contribution, we proposed and implemented several features that we made available for the scientific community. Visual recomendation and other image tasks may take advantage of the visual properties extracted in this work.

As future works we would like to uncover the user behavior that explains the correlation between image exposure and visual features predictive power. Exploring our results with different social networks, like Instagram and Vine, would also be very valuable in order to understand different content-sharing behaviors across those services.

7. ACKNOWLEDGEMENTS

This work is partially supported by CAPES, CNPq, FAPEMIG, FAPESP and InWeb - the National Institute of Science and Technology for the Web. Sandra Avila is supported by a grant from Brazilian Samsung Research Institute.

8. REFERENCES

[1] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. KDD '08.

[2] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo. Pooling in image representation: the visual codeword point of view. *CVIU*, 2013.

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: Quantifying influence on twitter. WSDM '11, 2011.

[4] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. WWW '12. ACM, 2012.

[5] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proc. of the International Conference on Multimedia*, MM '10.

[6] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. CVPR'10.

[7] D. Boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proc. of HICSS*, 2010.

[8] M. Cha, F. Benevenuto, H. Haddadi, and P. K. Gummadi. The world of connections and information flow in twitter. *Trans. on Systems, Man, and Cybernetics, Part A*, 42(4), 2012.

[9] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM*, 2010.

[10] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. WWW '09.

[11] J. Cheng, L. Adamic, A. Dow, J. Kleinberg, and J. Leskovec. Can cascades be predicted? In *WWW'14*.

[12] D. Comaniciu, P. Meer, and S. Member. Mean shift: A robust approach toward feature space analysis. *Trans. on Pattern Analysis and Machine Intelligence*, 2002.

[13] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. ECCV, 2006.

[14] S. Dhar, V. Ordonez, and T. Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR*, 2011.

[15] A. Dow, L. Adamic, and A. Friggeri. The Anatomy of Large Facebook Cascades. In *ICWSM*, 2013.

[16] eBiz MBA. http://www.ebizmba.com/articles/social-networking-websites, Mar. 2014.

[17] E. Gilbert, S. Bakhshi, S. Chang, and L. Terveen. "i need to try this"?: A statistical overview of pinterest. In *Proceedings of SIGCHI*, 2013.

[18] M. Gjoka, M. Kurant, and C. T. Butts. A Walk in Facebook: Uniform Sampling of Users in Online Social Networks. *CoRR*, 2009.

[19] A. Hanbury. Constructing cylindrical coordinate colour spaces. *Pattern Recogn. Lett.*, Mar. 2008.

[20] L. Hong, O. Dan, and B. D. Davison. Predicting popular messages in twitter. In *Proc. WWW*, 2011.

[21] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *CVPR*, 2011.

[22] W. Jiang, A. Loui, and C. Cerosaletti. Automatic aesthetic value assessment in photographic images. In *ICME*, 2010.

[23] D. Joshi, R. Datta, Q.-T. Luong, E. Fedorovskaya, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images: A computational perspective. In *IEEE Signal Processing Magazine*, 2011.

[24] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, 2006.

[25] A. Khosla, A. D. Sarma, and R. Hamid. What makes an image popular? In *WWW*, April 2014.

[26] K. Lerman and L. Jones. Social browsing on flickr. *CoRR*, abs/cs/0612047, 2006.

[27] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. KDD '09.

[28] C. Li and T. Chen. Aesthetic visual quality assessment of paintings. *J. Sel. Topics Signal Processing*, 2009.

[29] C. Li, A. C. Gallagher, A. C. Loui, and T. Chen. Aesthetic quality assessment of consumer photos with faces. In *ICIP*, 2010.

[30] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of ICCV*, 1999.

[31] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. Proc. ECCV'08.

[32] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM Multimedia*, 2010.

[33] S. A. Macskassy and M. Michelson. Why do people retweet? anti-homophily wins the day! In *ICWSM'11*.

[34] L. Marchesotti and F. Perronnin. Learning beautiful (and ugly) attributes. In *BMVC*. IEEE, 2013.

[35] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. ICCV'11.

[36] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proc. of SIGKDD*, 2012.

[37] R. Ottoni, J. P. Pesce, D. B. L. Casas, G. F. Jr., W. M. Jr., P. Kumaraguru, and V. Almeida. Ladies first: Analyzing gender roles and behaviors in pinterest. In *ICWSM*, 2013.

[38] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *Proceedings of ECCV*, 2010.

[39] J. Sánchez, F. Perronnin, T. Mensink, and J. J. Verbeek. Image classification with the fisher vector: Theory and practice. *IJCV*, 105, 2013.

[40] P. Sloan. www.cnet.com/news/pinterest-crazy-growth-lands-it-as-top-10-social-site, Jan. 2012.

[41] C. Smith. http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9, Sept. 2013.

[42] S. Stieglitz and L. Dang-Xuan. Political communication and influence through microblogging: An empirical analysis of sentiment in twitter messages and retweet behavior. In *Proc. of HICSS*, 2012.

[43] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proc. SOCIALCOM'10*.

[44] B. Thomee and A. Popescu. Overview of the imageclef 2012 flickr photo annotation and retrieval task. In *CLEF*, 2012.

[45] J. van de Weijer, C. Schmid, and J. Verbeek. Learning color names from real-world images. In *CVPR*, 2007.

[46] C. Vu and D. M. Chandler. S3: A spectral and spatial sharpness measure. In *2009 First International Conference on Advances in Multimedia*, 2009.

[47] W.-N. Wang, Y.-L. Yu, and S.-M. Jiang. Image retrieval by emotional semantics: A study of emotional space and feature extraction. In *SMC*, 2006.

[48] D. Zarella. The social media scientist. http://danzarrella.com/, June 2013.