Noname manuscript No.
(will be inserted by the editor)

Can Computers Learn from the Aesthetic Wisdom of the Crowd?

Christian Bauckhage · Kristian Kersting

Received: date / Accepted: date

Abstract The social media revolution has led to an abundance of image or video data on the Internet. Since this data is typically annotated, rated, or commented upon by large communities, it provides new opportunities and challenges for computer vision. Social networking and content sharing sites seem to hold the key to the integration of context and semantics into image analysis. In this paper, we explore the use of social media in this regard. We present empirical results obtained on a set of 127,593 images with 3,741,176 tag assignments that were harvested from Flickr, a photo sharing site. We report on how users tag and rate photos and present an approach towards automatically recognizing the aesthetic appeal of images using confidence-based classifiers to alleviate effects due to ambiguously labeled data. Our results indicate that user generated content allows for learning about aesthetic appeal. In particular, established low-level image features seem to enable the recognition of beauty. A reliable recognition of unseemliness, on the other hand, appears to require more elaborate high-level analysis.

1 Introduction

The summer of 2010 marked humankind's entry into the Zettabyte age. Since then, the amount of information on the Internet reportedly exceeds a Zettabyte (10^{21} Bytes) and its proliferation continues at a rapid pace [16]. Eric Schmidt, former CEO of Google, famously estimated that the volume of data recorded on the Internet is currently increasing at a rate of about 2.5 Exabytes ($2.5 \cdot 10^{18}$ Bytes) per day [26]. While this trend is remarkable in itself, it becomes even more noteworthy when we consider that one of its major drivers did not exists a mere decade ago: *social media*.

C. Bauckhage[1,3] · K. Kersting[2,3]
[1]B-IT, University of Bonn, Germany
[2]IGG, University of Bonn, Germany
[3]Fraunhofer IAIS, Sankt Augustin, Germany
E-mail: firstname.lastname@iais.fraunhofer.de

1.1 The Social Media Revolution

The term *social media* commonly refers to interactive web platforms that enable communities of people to share, modify, co-create, or discuss content. Expressing their interests and opinions in form of videos, photos, podcasts, or (micro)blogs, users of social networks and interactive services actively create a multitude of the data on the Internet today. Moreover, this kind of user-generated content is typically enriched by meta data such as annotations, geo information, links to related resources, or ratings and comments by other users. Finally, since access to and production of this content increasingly happens via mobile devices, it appears that social media create a pervasive information overlay on top of the real world, that is revolutionizing the way we work, shop, consume news, or maintain friendships or hobbies.

The scale of this revolution and its impact on business and society can hardly be understated. To illustrate this, Tab. 1 lists a few recent statistics about four of today's most prominent social media services. Noting that neither service did exists prior to 2004, it is striking that each has amassed millions of users within only a few years after going public. The social networking site Facebook is arguably the epitome of a successful social media service. As of this writing, it has attracted more than 800,000,000 members worldwide and thus currently ranks as the third largest human community in the world (after the People's Republic of China and the Republic of India). Yet, Facebook is merely one among many social networking sites. While communities that are centered around specific interests or activities or cater to local peculiarities may be less well known than Facebook, they typically attract millions of users, too. For

facebook	twitter	YouTube	flickr
800,000,000 users (active)	150,000,000 users (mostly inactive)	490,000,000 unique visitors (monthly)	51,000,000 users (registered)
250,000,000 photos/day	200,000,000 tweets/day	48 hours of video/minute	7,500,000 photos/day
350,000,000 mobile users	26,000,000 mobile users	400,000,000 mobile views/month	mostly via iPhone

Table 1 Statistics about some of the major drivers behind the social media revolution. The figures in this table (number of users, daily growth rates, mobile usage) reflect the situation as of December 2011 and were distilled from company web sites and business- or technology blogs.

example, Qzone, the largest social network in a China, currently boasts more than 480,000,000 active members and LinkedIn, a platform for professional networking, has more than 160,000,000 users worldwide.

Twitter is a microblogging platform where users openly share texts or *tweets* of up to 140 characters. The service is used from all over the world and its users collectively produce about 200 million new tweets every day. The Twitter community is known to react quickly and thoroughly to breaking news. In fact, events such as the 2008 Bombay attacks or the 2009 emergency landing of a plane in New York's Hudson river were covered on Twitter before they reached the mainstream media.

YouTube, a video sharing platform, has become the world's largest repository of moving pictures. Currently, its users upload about 48 hours of video every single minute. In other words, YouTube users post twice as many moving pictures per minute than a traditional TV station could broadcast in 24 hours.

Compared to these giants, Flickr, a photo sharing community, is of moderate size in terms of number of users and growth rates. While Facebook has also become the largest repository of digital images on the planet because its users upload about 250,000,000 new photos every single day, the members of Flickr post only a mere 7,500,000 new pictures per day. Yet, they constitute a very active and ambitious community who frequently comment, rate, and approve of each others photographic work. Sites like Flickr therefore offer great potential with respect to analyzing *how* large groups of people perceive and feel about content.

1.2 Social Media Analysis

Figures like the above are truly amazing and the rate of adoption of social media appears unparalleled [18]. It is therefore hardly surprising that, despite being a rather recent phenomenon, social media and their use have stirred considerable scientific interest and already attracted a large body of research. Much of this work is centered around the ideas of social media monitoring and *social media analysis*.

Social media analysis deals with *empirical* data that is retrieved from social media services. The typical approach is to consider huge amounts of data and to apply statistics and pattern recognition or data mining methods in order to gain insights from online activities and conversations of large communities of users. Social media analysis therefore resorts to the "unreasonable effectiveness of data" [17] and makes use of the "4th paradigm: data-intensive discovery" [19] to develop an understanding of human behavior, communication, preferences, sentiments, and relations on the individual as well as on the population level.

Surveying the literature on social media analysis, it appears that current research is exploratory and data driven rather than hypothesis driven. Although the field could thus be criticized for lack of a proper sociological, psychological, or philosophical foundation, it has nevertheless led to results that are of practical interest in various disciplines. Due to the timeliness with which social media generally react to news and since interactions among populations of users induce explicit or implicit social ties, social media analysis offers new insights into social dynamics in a networked world. It allows for reconstructing the developments of the Arab spring [12] or US electoral campaigns [33,45] as well as for studying political activism and discourse [8,24]. It reveals social incentives or contagion and uncovers hidden relations [1,29,35,44,52,59]. Empirical models of how word of mouth diffuses through latent networks may inform business or marketing strategies [23,32] and marketing experts are interested in user sentiments [5,28] and their perception of events [30,53]. Finally, studying the dynamics of topics discussed in social media reveals seasonal patterns [2,37,58] which, for example, can be used to track the spread of diseases and plan public health management [40,46].

1.3 Aims and Scope

With the work reported here, we extend "traditional" social media analysis towards examining the interplay between user generated content (images), user generated annotations (tags), and user generated ratings (favs). Following the best practices of current research in social media analysis, our work is of exploratory nature and we apply the paradigm of data intensive discovery to a large data set of annotated and rated photographs retrieved from Flickr. First of all, we mine our data and present and discuss empirical observations as to the *tagging and rating behavior of users* that may inform

Fig. 1 Examples of images that were tagged "awful" by the users who uploaded them to Flickr.

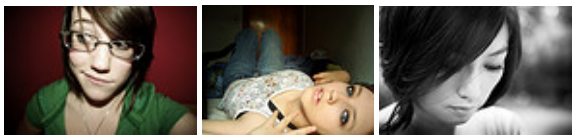


Fig. 2 Examples of images that were tagged "ugly" by the users who uploaded them to Flickr.

future work in sociology, psychology, or media studies. Second of all, we address an emerging problem in image understanding, namely to what extend *crowd generated content* can be used to train statistical classifiers to *automatically distinguish aesthetically pleasing pictures from less appealing ones*.

For research in computer vision, the abundance of annotated pictures on the Internet is an unforeseen but welcome consequence of the rise of social media. Meta data available from interactive web platforms seemingly hold the key to the integration of context and semantics into image analysis. Vast collections of such data provide new perspectives for statistical classification and corresponding efforts towards object recognition are increasing [3,11,15,51].

While object recognition and content categorization are well established topics, the classification of the aesthetic or emotional appeal of a photograph is a novel challenge in the emerging discipline of computational aesthetics. Although the topic was first discussed as a research direction nearly two decades ago [25], corresponding efforts intensified just recently. In times of exploding amounts of image data, aesthetics classification promises to improve targeted advertising, content-based search, or media production but can also identify genres and epochs of paintings [48], match pictures to music [14], or distinguish professional photos from those of amateurs [38,50].

Given the wealth of labeled and rated images produced by web communities of (semi)professional photographers, a growing number of contributors proposes empirical approaches to aesthetics analysis [9,10,49,57]. However, these efforts focus on ranking and retrieval rather than on categorization, and, to the best of our knowledge, largely ignore lessons learned in traditional social media analysis. In particular, pitfalls due to the *tagging behavior* of social media users appear not to have been taken into account. That is, while earlier studies in traditional social media analysis have revealed that people tend to resort to personal preferences and vocabularies when tagging multimedia content [42,54, 56], researchers in computer vision and image retrieval seem unaware of this.

In fact, aesthetics is known to be a subjective experience that still eludes quantification [31,34]. Figures 1 and 2 illustrate how this may hamper automatic classification. All six images were retrieved from the *most interesting* category on Flickr. Even if beauty is in the eye of the beholder, we may consent that the two examples on the left of Fig. 1 are less awful than the one on the right and the pictures in Fig. 2 are hardly ugly. Nevertheless, all six images did indeed result from searching for "awful" or "ugly" pictures, respectively.

Observations like these raise epistemological questions as to the *pragmatics of tags* and the *wisdom of the crowd*: Can *folksonomies*, i.e. collaboratively created tag collections on the Internet, be trusted to be objective? Is the crowd consistent in their behavior? Can and do social media reflect commonly accepted views or are they susceptible to manipulation or herding behavior? Answers to these questions are in dire need but beyond the scope of this paper. While the phenomenon of social media poses challenges to the research community at large and will require interdisciplinary efforts, we embrace a rather pragmatic, statistical point of view and assume that data driven discovery from large enough, unbiased data sets will reveal objective trends. Nevertheless, any approach to supervised learning for aesthetics classification based on social media data must take into account that aesthetic experiences are highly subjective and that a commonly agreed upon theory of their psychological constituents is missing.

1.4 Overview

Our presentation proceeds as follows: Next, we briefly introduce the data set that provides the empirical basis for our investigation. Then, in section 3, we explore the tagging behavior of Flickr users; in particular, we point out observed phenomena with respect to the pragmatics of tag assignments that appear to be specific for communities centered around pictorial content. In section 4, we explore and evaluate an algorithmic approach to the classification of aesthetic content that addresses the problem of subjectively labeled data. Finally, in section 5, we summarize our findings and discuss open questions and promising directions for future research.

2 Data Set and Characteristics

In October 2009, we collected a data set of 127,593 images belonging to 100 different categories. These images were retrieved from Flickr by means of tag-based searches over *most interesting* images. Consequently, our data is not

abominable	546	gorgeous	1799
aggressive	1797	grimy	1609
alarming	216	grubby	1241
alluring	1740	happy	1800
amazing	1799	haunting	1680
amusing	1800	heavenly	1800
appealing	1273	horrible	1800
arousing	392	icky	1800
astonishing	1203	immense	647
astounding	1127	impressive	1800
awful	1620	incredible	1800
beautiful	1800	intense	1800
blissful	1800	interesting	1800
breath-taking	1800	intriguing	1800
brilliant	1800	lovely	1799
brisk	1800	marvelous	1800
calm	960	meek	1678
cavernous	254	mesmerizing	1046
charming	1800	mighty	1657
chubby	1800	mind-expanding	37
chunky	1800	mundane	1800
contaminated	1246	nice	1800
creepy	1800	ominous	1800
crummy	280	optimistic	960
cute	1799	outasight	131
delightful	1800	overpowering	76
desolate	1800	pessimistic	105
disgusting	1800	placid	1800
dismal	1800	pleasant	1800
disturbing	1800	pretty	1800
divine	1800	risky	1800
dreary	1797	romantic	1560
dull	1800	sad	1800
eerie	1679	scary	1799
enchanting	1800	sensuous	1800
energetic	1800	shabby	1800
engrossing	824	splendid	1800
enjoyable	1740	strange	1500
enthralling	58	stunning	1797
enticing	500	super-duper	374
exalted	239	surprising	1261
exciting	1800	suspicious	1784
eye-catching	1800	terrible	1800
fantastic	1680	terrific	1800
fascinating	1800	thunderous	1800
filthy	1800	ugly	1799
flawed	405	uncanny	1800
gargantuan	678	wearisome	177
ghastly	908	wholesome	1101
gigantic	1800	wonderful	1799

Table 2 List of 100 emotional or aesthetic categories originally due to Black et al. [4] and numbers of images retrieved per category.

a truly characteristic of the kind of pictures generally found at Flickr. Rather, it represents a sample of high-quality content uploaded by professional photographers or ambitious hobbyists whose photos feature prominently among Flickr's *most interesting* pictures.

The 100 categories we considered correspond to a list of emotional adjectives first introduced in a pioneering paper on mood-based image retrieval [4]. The lists consist of terms such as "amazing", "beautiful", or "happy" which are typically used to characterize feelings, emotional responses, or aesthetic content. Table 2 lists all 100 adjectives together with the number of images we retrieved per category. Where possible, we collected data from the first 30 pages returned by Flickr where the number of results per page was 60; for some categories, our searches produced less than 30 pages of results so that the number of images per category varies between 37 and 1,800.

No. of images retrieved from Flickr:	127,593
No. of categories :	100
No. of tags:	3,741,176
No. of tags w/o stopwords:	3,581,853
No. of unique tags:	177,415

Table 3 Statistics obtained from our Flickr data set.

Table 3 further summarizes our data. The images we downloaded came along with a total of 3,741,176 tags consisting of more than 3 characters. Removing stopwords from this collection left us with 3,581,853 tags. Collecting them into a dictionary revealed 177,415 distinct tags in the data.

3 Observations on the Usage of Tags and Favs

From an ontological point of view, *folksonomies*, i.e. crowd generated tag collections on the Internet, are notorious for being unreliable or contradictory [42,54,56]. An explanation as to these tendencies might be found in the fact that social media users are primarily interested in the social aspects of participating in an online community. To them, sharing and tagging their photos and commenting on other people's pictures is an act of communication. As such, it will not be restricted to descriptions and semantics, but will also involve *aspects of pragmatics*. Next, we present empirical findings obtained from our data set that illustrate these points.

3.1 Word Count Distributions

Figure 3 shows that tag frequencies or ranked word counts follow a quickly decaying distribution. The most frequent tag in our data is "beautiful" which was found to be assigned to 19,585 images. That is, in addition to the 1,800 images we retrieved when searching for "beautiful", our other 99 searches produced 17,785 additional images tagged as "beautiful". At the same time, more than half of the tags in the dictionary, namely 95,687, occurred just once.

In the literature on social media, long tailed distributions like this are reported frequently. Often, empirical distribution of ranked word frequencies are rashly modeled using discrete power law distributions where

$$p(x) = \frac{a-1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-a} \tag{1}$$

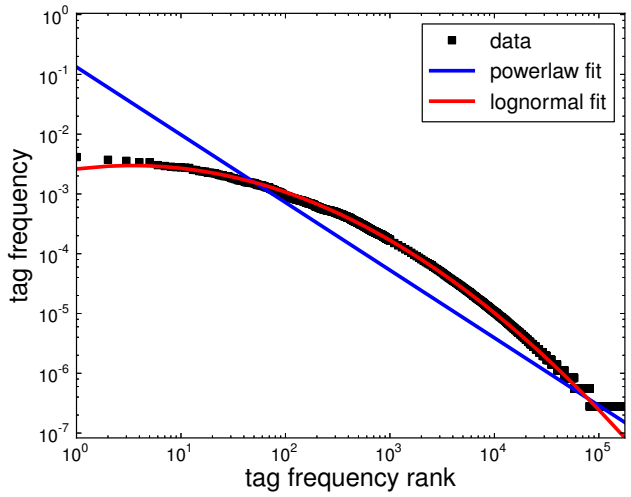


Fig. 3 Tag frequency distribution observed in our data set. It appears that ranked word counts do not accord with a power law but are more accurately modeled by a log-normal distribution ($\mu = 7.06, \sigma = 2.64$). Accordingly, on the population level, the way users choose tags cannot be explained in terms of preferential attachment.

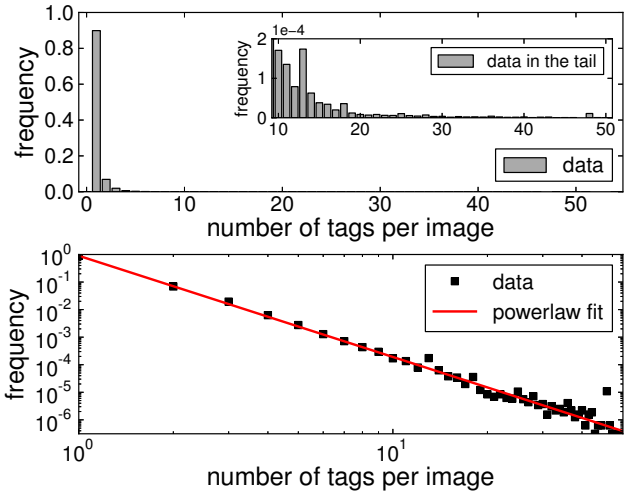


Fig. 4 Tag frequencies per image obtained from our *Flickr* data set. The power law behavior of this distribution ($a = 3.67$) indicates that the data we consider was indeed user generated and not produced by spammers or automated scripts.

with an exponent or scaling parameter $a > 1$. Yet, power laws are not as omnipresent as they are believed to be. On the contrary, the preference for power law models is seldom justified and authors go at great lengths to argue about cut-offs and truncations to justify their use of power law distributions even if they hardly fit the data.

Most likely, the impulse to attribute word count distributions to power laws is due to misunderstandings as to the generality of Zipf's law and its applicability [36]. Zipf's law is an empirical statement about word counts in natural corpora which says that given a collection of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency list. Plotted on a doubly logarithmic scale, distributions like these appear as a straight line. In fact, a variety of web statistics such as file size distributions or in-links of web sites show exactly this behavior [13,36]. Yet, fitting a power law distribution using the rigorous methods in [6] clearly does not well account for the tag frequency distribution we observe in our data (see Fig. 3).

Rather, we found the distribution to be much better explained by a log-normal model where a random variable x is said to be log-normally distributed, if

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\log(x) - \mu\right)^2\right). \qquad (2)$$

The distribution is only defined for positive values, skewed to the left, and often long-tailed. The mean μ and standard deviation σ of $\log(x)$ define the exact form of the curve.

Mitzenmacher [36] provides an excellent review of physical causes and statistical phenomena that give rise to log-normal distributions. While power law distribution can be explained as a consequence of preferential attachment processes, log-normal distributions characterize growth processes or random processes of random duration.

Our findings in Fig. 3 therefore argue against a global form of preferential attachment. In other words, just because many users have assigned a certain tag to their images this does not imply that even more people will use this tag as well. Therefore, concerning the choice of tags there does not seem to be any herding behavior among Flickr users.

However, if we do consider the distribution of the number of tags per image in Fig. 4, we find a power law: the vast majority images in our data set were assigned only few tags and rather few images are tagged more than 10 times. This observation indicates that our data indeed represents user generated content. This is worth stressing, since social media data are known to frequently suffer from phenomena such as *tag spam* [55]. That is, not all content found in social media is necessarily produced by people. Rather, spammers or astroturfers often create fake accounts and personalities and then use scripted access in order to ruthlessly spread their agenda in web communities.

3.2 The Pragmatics of Tags

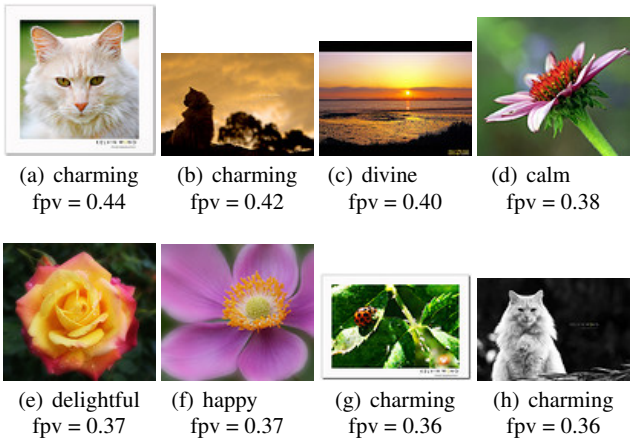
A rather surprising result comes from looking at Tab. 4 which lists the 100 most frequent tags in our collection. Given our choice of categories, a bias towards terms such as "beautiful" or "beauty" is to be expected and well reflected in the table. Interestingly, however, the table also contains terms such as "abigfave" (rank 35.) that do not refer to any natural entity or emotional state. Rather, "abigfave" is the name of one of Flickr's groups.

1.	beautiful	19585	26.	interesting	7308	51.	nice	5425	76.	sea	4271
2.	art	14830	27.	great	7103	52.	travel	5369	77.	tree	4257
3.	photography	13356	28.	landscape	7100	53.	macro	5252	78.	sun	4218
4.	light	12821	29.	clouds	7092	54.	day	5204	79.	flowers	4096
5.	nature	12099	30.	pretty	6959	55.	digital	5134	80.	female	4092
6.	canon	12088	31.	lovely	6705	56.	cool	5054	81.	anawesomeshot	3972
7.	white	10866	32.	fun	6555	57.	happy	5043	82.	summer	3951
8.	nikon	10496	33.	people	6505	58.	urban	5016	83.	perfect	3924
9.	blue	10202	34.	photographer	6347	59.	park	5012	84.	hdr	3901
10.	black	10026	35.	abigfave	6316	60.	yellow	4995	85.	sweet	3896
11.	portrait	9953	36.	beach	6116	61.	hot	4914	86.	2008	3896
12.	girl	9896	37.	stunning	6056	62.	old	4863	87.	good	3884
13.	sky	9746	38.	dark	6038	63.	pink	4846	88.	funny	3826
14.	cute	9069	39.	explore	5991	64.	fashion	4832	89.	london	3797
15.	photo	8995	40.	best	5921	65.	face	4708	90.	colorful	3774
16.	water	8632	41.	sunset	5865	66.	fantastic	4654	91.	winter	3768
17.	love	8513	42.	photos	5736	67.	big	4644	92.	photoshop	3692
18.	color	8497	43.	awesome	5625	68.	wedding	4644	93.	cat	3626
19.	amazing	8133	44.	flower	5612	69.	night	4633	94.	orange	3618
20.	beauty	8132	45.	wonderful	5610	70.	colors	4614	95.	diamondclassphotographer	3592
21.	red	8013	46.	sexy	5561	71.	street	4592	96.	magic	3543
22.	new	7896	47.	flickr	5555	72.	breathtaking	4387	97.	wow	3400
23.	green	7731	48.	eyes	5487	73.	trees	4356	98.	abandoned	3345
24.	woman	7560	49.	life	5478	74.	aplusphoto	4326	99.	model	3343
25.	gorgeous	7334	50.	city	5460	75.	lake	4288	100.	creepy	3342

Table 4 The 100 most frequent tags in our Flickr data set together with their occurrence counts. In addition to tags describing image content or emotional appeal, we find tags referring to photographic techniques (highlighted in blue) as well as names of special interest groups at Flickr (highlighted in red). It thus appears that users generally use tags to tell more about a picture than just what it shows.

Groups are a common feature of social media services and allow users to share specific interest with a circle of like-minded people. All in all, there are 4 group related words among the 100 highest ranking tags in our collection. Each of these groups is an *invitation only* group, i.e. a group where users may posts their photos only after having been invited to do so because the group considered them outstanding. It seems therefore plausible to conclude that users tag their photos thusly not to describe content but to impress others.

Moreover, several tags in Tab. 4 refer to camera brands, photographic techniques, or image processing software. As with the names of exclusive groups, it appears that people use these tags because they want to tell more about a picture than just what it shows. In particular, they frequently share how and under what conditions a picture was taken.



(a) charming fpv = 0.44 (b) charming fpv = 0.42 (c) divine fpv = 0.40 (d) calm fpv = 0.38

(e) delightful fpv = 0.37 (f) happy fpv = 0.37 (g) charming fpv = 0.36 (h) charming fpv = 0.36

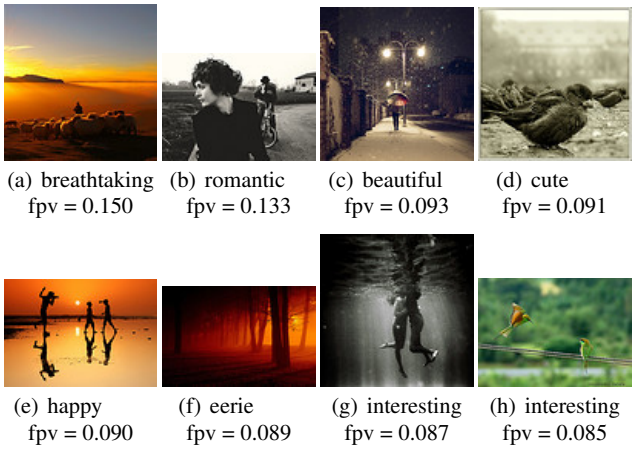
Fig. 5 The 8 most popular images with more than 100 views. The adjective below each image indicates the category it was found in, the fpv values correspond to the number of *favorites* per view. It is noticeable that there are three images of cats and three images of flowers.

3.3 Favs and Image Quality

Flickr is an interactive community whose members comment on each others photographs and discuss their aesthetic or technical merits. In addition, they may also mark other people's photos as a favorite of theirs. Next, we present observation about *favs* assigned to images.

Favs provide contextual information on how a photo is perceived by others. In particular, the number of favs per view (fpv), i.e. the ratio between how often a photo has been called a favorite and how often it has been viewed, provides deeper insights into the Flickr community. Computing this ratio for every image in our data allows us to rank them with respect to popularity. In order to avoid artifacts due low view counts (e.g. a picture may have an fpv of 0.5 because it has been viewed twice and one of the viewers called it a favorite), it is advisable to consider only those images which have attracted a certain minimum number of views.

Figure 5 shows the 8 highest ranking images with more than 100 views. These top ranking pictures certainly are



(a) breathtaking fpv = 0.150 (b) romantic fpv = 0.133 (c) beautiful fpv = 0.093 (d) cute fpv = 0.091

(e) happy fpv = 0.090 (f) eerie fpv = 0.089 (g) interesting fpv = 0.087 (h) interesting fpv = 0.085

Fig. 6 The 8 most popular images with more than 10,000 views. The adjective below each image indicates the category it was found in, the fpv values correspond to the number of *favorites* per view. These pictures are of high photographic and aesthetic quality.

of aesthetic appeal but, on a lighter note, we observe that there are 3 pictures of cats among these top 8. This nicely agrees with common lore which claims that the most popular videos at YouTube are videos of cats and we dare say that social media's obsession with cats may merit further study.

Figure 6 shows another set of 8 top ranking photos, this time, the most popular among all images in our data that attracted more than 10,000 views. Naturally, their fpv values are smaller but what is striking is that these pictures are of distinguished quality. That is to say, their visual rhetoric (composition, contrast, coloring) shows many characteristics of what is commonly consider aesthetically pleasing and artistic [7,20,34,41]. This suggests an effect of the *wisdom of the crowd*. As a community, Flickr users reliably indicate photographic quality. Images that are viewed a lot and declared to be a favorite by many users are indeed appealing.

However, we must verify wether we merely observe herding behavior. Figure 7 therefore plots the ranked fav count distribution and we find that it does not follow a power law but a gamma distribution

$$p(x) = \frac{1}{\theta^{\kappa}\Gamma(\kappa)} x^{\kappa-1} \exp\left(-\frac{x}{\theta}\right) \tag{3}$$

where k and θ are the shape and scale parameter, and $\Gamma(\cdot)$ is the gamma function. Fav assignments therefore seem to happen according to a mixed Poisson process rather than to preferential attachment. That is, just because an image has attracted many favs it will not receive disproportionably many new ones; in other words, Flickr users seem not to follow the herd when they assign favs.

It therefore appears auspicious to consider Flickr data in order to develop systems that are capable of recognizing aesthetic categories. In the next section, we present an approach towards this goal.

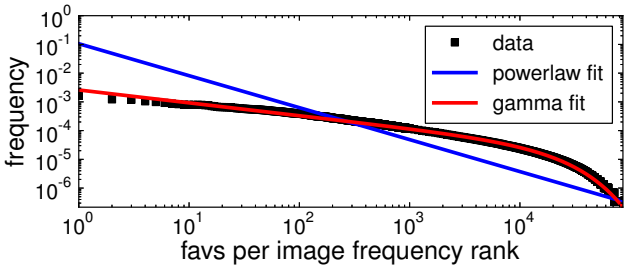


Fig. 7 Fav frequencies per image obtained from our *Flickr* data set. This data is follows a gamma distribution rather than a power law.

4 Learning to Recognize Aesthetic Images

In this section, we explore the use of off-the-shelf techniques for aesthetics recognition. Contrary to previous work, our data basis for classifier training and testing is much larger. Also, we specifically address issues that may arise from the pragmatic and personalized uses of tags which we discussed in the previous section. Similar to previous work, we restrict our investigation to low-level image features for classification. Also, we restrict our investigation to Support Vector Machines (SVMs) which are known for their discriminative power, although Neural Networks, Decision Trees, or Gaussian- and, in particular, Dirichlet mixture models may provide auspicious alternatives. We leave this to future work, because our focus in this paper is on social media analysis rather than on pattern recognition methods.

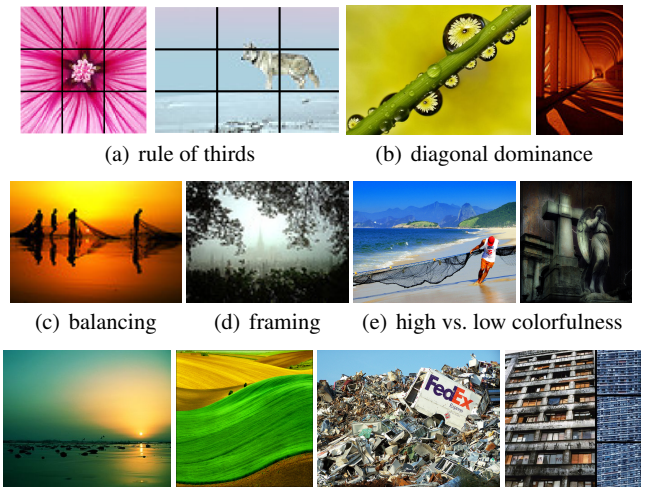
4.1 Features for Aesthetics Classification

Extracting characteristic features from images is usually the first steps in automatic image analysis. The features consider here are a subset of those introduced by Datta et al. [10] and Dunker et al. [14]. Other authors, too, made use of similar features [9,57] and found them to provide good characterizations on low levels of abstraction. They are devised to account for image properties in terms of color or geometry that are commonly agreed to indicate artistic composition (see Fig. 8). Note that we do not compute any features that would describe an image on the object level.

Following [10], we transform RGB images into the HSV color space. The HSV representation of a color pixel describes its hue, saturation and, value (intensity) and is supposed to accommodate human cognition. Accordingly, we consider the following HSV features to characterize color characteristics of an image:

Global hue, saturation, and value are computed as the corresponding averages over all pixels in an image and characterize chromatic purity, dominant color, and intensity.

Central hue, saturation, and value are computed as averages over the pixels in the central rectangle of an image and



(a) rule of thirds (b) diagonal dominance

(c) balancing (d) framing (e) high vs. low colorfulness

(f) low vs. high frequency content

Fig. 8 Principals of photographic composition: (a) applying the "rule of thirds" means to superimpose four imaginary lines over the frame to produce nine rectangular sub-images, the subject of the image is then positioned with respect to this virtual grid; (b) "diagonal dominance" expresses the idea that diagonal lines guide the eye through a picture; (c) "balancing" attempts to evenly balance image subjects; (d) "framing" uses natural frames for emphasis; (e) colorful or less colorful compositions may amplify moods; (f) low-frequent image content is perceived as natural and soothing.

thus meant to account for the photographic *rule of thirds* which states that interesting image content is close to one of the intersections of four imaginary lines superimposed over the image (see Fig. 8).

Colorfulness is a global property of a picture and colorful pictures are usually perceived to be pleasing [21,47]. Again following [10], we compute a color histogram of a picture and determine its distance to a set of histograms obtained from prototypic colorful and less colorful images.

Following [14], we compute *Gabor filter responses for the central region* of an image to characterize its geometry. Our filter bank contains 21 filters of different orientation and phase and can thus detect balanced compositions or diagonal dominance. In addition, it also characterizes the frequency content of an image, i.e. smoothness or roughness, and therefore mimics more elaborate approaches based on Fourier transforms [27,43]. In addition, Gabor filters yield different responses for photos of natural scenes than for pictures of human-made artifacts [22,39] and therefore roughly characterize scene content, too.

In total, for each image I_i, we thus derive a 39 dimensional feature vector $\mathbf{v}(I_i) = \mathbf{v}_i$ to represent an image and to train and test the classifiers described in the next section.

4.2 Classification

We considered different settings of two class classification where we tried to automatically classify an image using three opposing adjectives. In each experiment, the training data was thus split into two classes, where the images in the one class carried labels of *positive* connotation while those in the other class were labeled with *negative* adjectives.

Experimenting with SVMs with radial basis kernels for binary classification (class labels +1 or -1) led to satisfactory predictions of the aesthetic label of an image for the majority of tests cases. However, for several subsets of test images, the classification accuracy was admittedly disappointing (below 70%). Closer inspection of these cases revealed that a substantial number of images in our data appeared to be labeled rather misleadingly.

We frequently observed that for different images of visually similar content different people assigned incoherent labels. For example, we found many portraits of women labeled to be "ugly" or "awful" though objectively this assertion appeared incomprehensible (see Figs. 1 and 2). This phenomenon was in fact peculiar and could not be ascribed to statistical oddities. For instance, among the 1,800 images in the "ugly" category, we found 198 pictures of women as in Fig. 2. Coquetting like this thus distorts a high percentage of our data and is another example of subjective behavior that has to be taken into account in social media analysis.

From the point of view of pattern recognition, ambiguous labels cause considerable class overlap in the feature space. Since manual clean-up is infeasible given the extreme amounts of data in today's applications, we consider an informed postprocessing step to remedy this situation. The basic idea is to soften the binary decision function that is computed by an SVM. If the SVM had been trained to regress examples of the *positive* class to +1 and examples of the *negative* class to -1, we thus compute sigmoid functions

$$y_p(x) = \frac{1}{1+e^{-(x-\mu_p)\sigma_p}} \quad \text{and} \quad y_n(x) = \frac{1}{1+e^{(x-\mu_n)\sigma_n}} \qquad (4)$$

where x is the output of the SVM. The parameters μ_p, μ_n, σ_p, and σ_n govern location and shape of the sigmoid functions and are determined on a verification set of images that is independent both from the training and test data.

Using sigmoidal softening maps the predictions of an SVM to the interval $[0,1]$ so that we may interpret the results y_p and y_n as degrees of belief in whether a pattern belongs to the *positive* or *negative* class, respectively, where the two cases need not be mutually exclusive, i.e. in general $y_p + y_n \neq 1$. This allows us to determine classification accuracy with respect to classification confidence and thus alleviates the effect of overlapping regions in feature space. While pictures in these regions will have low probabilities of belonging to either class, less ambiguous pictures will have higher probabilities of belonging to either the one or the other class.

	experiment 1		experiment 2	
class	SVM	SVM+P	SVM	SVM+P
beautiful	77%	88%	73%	86%
wonderful	91%	97%	89%	96%
divine	60%	80%	84%	93%
ugly	56%	80%	52%	70%
awful	40%	70%	59%	83%
terrible	63%	84%	44%	77%

Table 5 Class specific recognition accuracies. Experiment 1: training with two superclasses of positive and negative pictures; experiment 2: training with images from three pairs of classes of opposing adjectives. Results are shown for classification with regular SVMs only and for SVMs with postprocessing, i.e. with sigmoidal smoothing.

4.3 Experimental Results and Discussion

The results presented here were obtained from experimenting with 10,618 unique pictures from our data set. We focused on 3 adjectives of positive connotation ("beautiful", "wonderful", and "divine") as well as on 3 adjectives of negative connotation ("ugly", "awful", and "terrible").

In a first series of experiments, we explored if it is possible to distinguish pictures that evoke positive emotions from pictures evoking negative feelings. The sets of positive and negative images were unified into two corresponding classes and we randomly subdivide the data into three independent subsets for training, verification and testing. After training, the verification phase was used to determine the parameters μ and σ of both classes such that accuracy exceeds 99% for a classification confidence of 60%.

In a second series of experiments, we evaluated, if our approach distinguishes pictures from classes of opposing adjectives ("beautiful" vs. "ugly", "wonderful" vs. "awful", "divine" vs. "terrible"). Training and verification were done as in the first series of experiments.

Table 5 compares the recognition accuracies of baseline SVM classification to those obtained from postprocessing SVM results using sigmoidal smoothing. For the latter, results with a confidence less than 55% were rejected. Looking at the table, we can summarize our results as follows: i) predicting aesthetic labels using statistical classifiers is possible to a large extent; ii) simple postprocessing consistently improves recognition accuracy and copes well with the problem of subjective tag assignments; iii) for most of the 6 classes in our test, it appears beneficial to train with larger superclasses of positive and negative images rather than with smaller sets of opposingly labeled images only; iv) pictures that evoke negative emotions are classified less reliably than pictures of positive content.

Figures 9 and 10 show examples of pictures that were classified correctly using our approach. Figure 11, on the other hand, shows examples of pictures our system deemed "beautiful" but which were actually tagged as "awful", "terrible", and "ugly". Given these prototypic examples, it appears that a generally robust recognition of unseemliness requires semantic image understanding (e.g. in the case of a "terrible" flower like poison ivy). Beauty, on the other hand, appears to be highly correlated with colorfulness, low frequencies, and local symmetries so that its recognition is possible from rather simple low-level image features only.

Fig. 9 Pictures that were correctly classified as "beautiful", "divine", and "wonderful" in experiment 1.

Fig. 10 Pictures that were correctly classified as "awful", "terrible", and "ugly" in experiment 1.

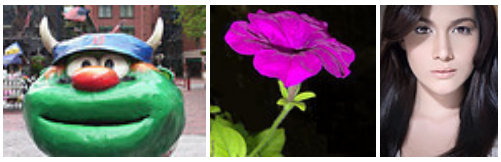


Fig. 11 Examples of pictures that were recognized to be pleasant but had been labeled "awful", "terrible", and "ugly" by their creators.

5 Conclusion and Future Work

Analyzing user behavior in social media yields new insights in the social, political, and economic sciences but for computer science, too, social media provides new opportunities. Techniques for data intensive discovery and learning from large data can now be applied to demanding problems in areas such as computer vision and meta data resulting from interactions among social media users seem to hold the key to the integration of context or semantics into statistical learning. Yet, given the example of large set of annotated Flickr photographs, we showed that care is required when using social media as a data source.

First of all, we found that the tagging behavior of users is highly subjective. From our analysis, it appears that user generated image annotations are meant to covey more than just semantic content descriptions. Rather, there are considerable pragmatic aspects, since tags were found to also express the circumstance under which a picture was taken or to brag about a picture's quality. Our analysis also provided indicators as to a gender specific usage of tags. We found a large percentage of portraits of female users who had used deprecative adjectives as labels. This observation in particu-

lar merits further investigation and comparisons and serves as a prime example as to the type of sociological insights one may gain from analyzing social media.

Second of all, we found the wisdom of the crowd to be at work in our data collection. Analyzing how frequently users declared an image to be a favorite of theirs, we found that quality prevails. Often viewed images that had attracted a lot of favor were found to be of high photographic quality and the good ratings could not be attributed to preferential attachment or *the rich get richer* processes.

Motivated by the photographic quality of the images in our data, we finally explored to what extend social media data may be used to learn to classify the aesthetic or emotional value of an image. We extended SVM based classifiers by a simple postprocessing step to cope with subjective labels and found that computational approaches to photographic aesthetics are indeed feasible. Favorable appeal seems to be highly correlated with image properties such as colorfulness, low frequencies, and local symmetries and can thus be recognized from low-level image features only. Unseemliness, on the other hand, was less reliably classified and it appears that its recognition requires more elaborate image analysis and semantic understanding.

The work reported here focused on possible uses of social media data in computer vision and possibly raises more questions than it answered. Although there is a large body of literature on social media analysis, the interplay between multimedia content and user behavior has found little attention yet. This leads to epistemological questions as to the validity of the paradigm of data intensive discovery and the assumption that crowds of users generate consistent data that allows for training artificial intelligence systems. Answers are in dire need and likely require interdisciplinary efforts.

With respect to recognizing the aesthetical or emotional appeal of images, our results are encouraging but only a first step into this area. Obvious questions to address in future work pertain to different classification mechanisms and their merits. Also, it appears interesting to consider social media data in order to investigate what kind of image properties are responsible for how a picture is perceived. Our current strategy is to compute hundreds of different features from images and to apply Decision Trees and Boosting methods in order to determine the correspondingly most influential features. In this sense, we again resort to data intensive discovery in order to gain an empirical understanding of what kind of beauty is in the eye of the crowd.

References

1. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Discovering value from community activity on focused question answering sites: A case study of stack overflow. In: ACM KDD (2012)
2. Bauckhage, C.: Insights into internet memes. In: AAAI ICWSM (2011)
3. Bauckhage, C., Alpcan, T., Wetzker, R., Umbrath, W.: Image retrieval and web 2.0 – where can we go from here? In: IEEE ICIP (2008)
4. Black, J., Kahol, K., Trapathi, P., Kuchi, P., Panchanathan, S.: Indexing natural image for retrieval based on kansei factors. In: Human Vision and Electronic Imaging IX, *Proc. SPIE*, vol. 5292 (2004)
5. Bollen, J., Mao, H., Pepe, A.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: AAAI ICWSM (2011)
6. Clauset, A., Shalizi, C., Newman, M.: Power-law distributions in empirical data. SIAM Rev. **51**(4), 51–94 (2007)
7. Clemens, B., Rosenfeld, D.: Photographic Composition. Van Nostrand Reinhold Company (1979)
8. Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., Menczer, F.: Political polarization on twitter. In: AAAI ICWSM (2011)
9. Datta, R., Fedorovskaya, E., Luong, Q.T., Wang, J., Li, J., Luo, J.: Aesthetics and emotions in images. IEEE Signal Process. Mag. **28**(5), 94–115 (2001)
10. Datta, R., Joshi, D., Li, J., Wang, J.: Studying aesthetics in photographic images using a computational approach. In: ECCV (2006)
11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE CVPR (2009)
12. Dewey, T., Kaden, J., Marks, M., Matsushima, S., Zhu, B.: The impact of social media on social unrest in the arab spring. Tech. rep., Stanford University (2012)
13. Downey, A.: Lognormal and pareto distributions in the internet. Comput. Commun. **28**(7), 790–801 (2005)
14. Dunker, P., Nowak, S., Begau, A., Lanz, C.: Content-based mood classification for photos and music: a generic multi-modal classification framework and evaluation approach. In: ACM MIR (2008)
15. Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vision **88**(2), 303–338 (2010)
16. Gantz, J., Reinsel, D.: IDC iView: Extracting value from chaos. Tech. rep., EMC Corporation (2011)
17. Halevy, A., Norvig, P., Pereira, F.: The unreasonable effectiveness of data. IEEE Intell. Syst. **24**(2), 8–12 (2009)
18. Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., Weitzner, D.: Web science: An interdisciplinary approach to understanding the web. Commun. ACM **51**(7), 60–69 (2008)
19. Hey, T., Tanslev, S., Tolle, K. (eds.): The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research (2009)
20. Hill, C., Helmers, M. (eds.): Defining Visual Rhetorics. Lawrence Erlbaum Associates (2004)
21. Hogg, J. (ed.): Psychology and the Visual Arts. Penguin Books (1969)
22. Hyvärinen, A., Hurri, J., Hoyer, P.: Natural Image Statistics. Springer (2009)
23. Jansen, B.: Classifying ecommerce information sharing behaviour by youths on social networking sites. J. of Information Science **37**(2), 120–136 (2011)
24. Jisun, A., Cha, M., Gummadi, K., Crowcroft, J.: Media landscape in twitter: A world of new conventions and political diversity. In: AAAI ICWSM (2011)
25. Kato, T.: Database architecture for content-based image retrieval. In: Image Storage and Retrieval Systems, *Proc. SPIE*, vol. 1662 (1992)
26. Kirkpatrick, M.: Google CEO Schmidt: "People Aren't Ready for the Technology Revolution". http://readwriteweb.com (2010)
27. Koch, M., Denzler, J., Redies, C.: $1/f^2$ characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. PLoS One **5**(8), e12,268 (2010)

28. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! In: AAAI ICWSM (2011)
29. Kunegis, J., Lommatzsch, A., Bauckhage, C.: The slashdot zoo: Mining a social network with negative edges. In: ACM WWW (2009)
30. Lanagan, J., Smeaton, A.: Using twitter to detect and tag important events in live sports. In: AAAI ICWSM (2011)
31. Leder, H., Belke, B., Oberst, A., Augustin, D.: A model of aesthetic appreciation and aesthetic judgements. Brit. J. Psychol. **95**(4) (2004)
32. Leskovec, J., Adamic, L., Huberman, B.: The Dynamics of Viral Marketing. ACM Tans. Web **1**(1), 5 (2007)
33. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the Dynamics of the News Cycle. In: ACM KDD (2009)
34. Maquet, A.: The Aesthetic Experience: An Anthropologist Looks at the Visual Arts. Yale University Press (1988)
35. Meeder, B., Karrer, B., Sayedi, A., Ravi, R., Borgs, C., Chayes, J.: We know who you followed last summer: Inferring social link creation times in twitter. In: ACM WWW (2011)
36. Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. Internet Mathematics **1**(2), 226–251 (2004)
37. Naveed, N., Sizov, S., Staab, S.: Att: Analyzing temporal dynamics of topics and authors in social media. In: ACM WebSci (2011)
38. Obrador, P., Moroney, N.: Low level features for image appeal measurement. In: S. Farnand, F. Gaykema (eds.) Image Quality and System Performance, *Proc. SPIE*, vol. 7242 (2009)
39. Oliva, A., Torralba, A.: Building the gist of a scene: the role of global image features in recognition. In: Progress in Brain Research. Elsevier (2006)
40. Paul, M., Dredze, M.: You are what you tweet: Analyzing twitter for public health. In: AAAI ICWSM (2011)
41. Peters, G.: Aesthetic primitives of images for visualization. In: IEEE IV (2007)
42. Peterson, E.: Beneath the metadata: Some philosophical problems with folksonomy. D-Lib Magazine **12**(11) (2006)
43. Redies, C., Hänisch, J., Blickhan, M., Denzler, J.: Artists portray human faces with the fourier statistics of complex natural scenes. Network: Computation in Neural Systems **18**(3), 235–248 (2007)
44. Romero, D., Galuba, W., Asur, S., Huberman, B.: Influence and passivity in social media. In: ACM WWW (2011)
45. Romero, D., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In: ACM WWW (2011)
46. Signorini, A., Segre, A., Polgreen, P.: The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic. PLoS ONE **6**(5), e19,467 (2011)
47. Solso, R.: Cognition and the Visual Arts. MIT Press (1996)
48. Spehr, M., Wallraven, C., Fleming, R.: Image statistics for clustering paintings according to their visual appearance. In: Int. Symp. Comp. Aesthetics in Graphics, Visualization, and Imaging (2009)
49. Thurau, C., Bauckhage, C.: Archetypal images in large photo collections. In: IEEE ICSC (2009)
50. Tong, H., Li, M., Zhang, H.J., He, J., Zhang, C.: Classification of digital photos taken by photographers or home users. In: Pacific Rim Conf. Multimedia (2004)
51. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **30**(11), 1958–1970 (2008)
52. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. PNAS **109**(16), 5962–5966 (2012)
53. Weng, J., Lee, F.: Event detection in twitter. In: AAAI ICWSM (2011)
54. Wetzker, R., Alpcan, T., Bauckhage, C., Umbrath, W., Albayrak, S.: An unsupervised hierarchical method for automated document categorization. In: IEEE/WIC/ACM WI (2007)
55. Wetzker, R., Zimmermann, C., Bauckhage, C.: Detecting trends in social bookmarking systems: A del.icio.us endeavor. Int. J. on Data Warehousing and Mining **6**(1), 38–57 (2010)
56. Wetzker, R., Zimmermann, C., Bauckhage, C., Albayrak, S.: I tag, you tag: Translating tags for advanced user models. In: ACM WSDM (2010)
57. Wong, L.K., Low, K.L.: Saliency-enhanced image aesthetic classification. In: IEEE ICIP (2009)
58. Wu, F., Huberman, B.: Novelty and collective attention. PNAS **104**(45), 17,599–17,601 (2007)
59. Wu, S., Hofman, J., Mason, W., Watts, D.: Who says what to whom on twitter. In: ACM WWW (2011)