

# “Nobody comes here anymore, it’s too crowded”; Predicting Image Popularity on Flickr

Philip J. McParlane<sup>\*</sup>  
School of Computing Science  
University of Glasgow  
Glasgow, UK  
p.mcparlane.1@research.gla.ac.uk

Yashar Moshfeghi  
School of Computing Science  
University of Glasgow  
Glasgow, UK  
Yashar.Moshfeghi@glasgow.ac.uk

Joemon M. Jose  
School of Computing Science  
University of Glasgow  
Glasgow, UK  
Joemon.Jose@glasgow.ac.uk

## ABSTRACT

Predicting popular content is a challenging problem for social media websites in order to encourage user interactions and activity. Existing works in this area, including the recommendation approach used by Flickr (called “interestingness<sup>1</sup>”), consider only click through data, tags, comments and explicit user feedback in this computation. On image sharing websites, however, many images are annotated with no tags and initially, an image has no interaction data. In this case, these existing approaches fail due to lack of evidence. In this paper, we therefore focus on image popularity prediction in a cold start scenario (i.e. where there exist no, or limited, textual/interaction data), by considering an image’s context, visual appearance and user context. Specifically, we predict the number of comments and views an image has based on a number of new features for this propose. Experimenting on the MIR-Flickr 1M collection, we are able to overcome the problems associated with popularity prediction in a cold start, achieving accuracy of up to 76%.

**Categories and Subject Descriptors:** H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing

**Keywords:** image, popularity prediction, flickr

## 1. INTRODUCTION

Popularity prediction is becoming an ever more important aspect of web 2.0 applications in an information overload age. Given that a user cannot view all of the data uploaded to a social media website, recommending the most interest-

<sup>\*</sup>This research was supported by the the European Community’s FP7 Programme under grant agreements nr 288024 (LiMoSiNe)

<sup>1</sup><http://www.flickr.com/explore/interesting>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM ICMR 2014

Copyright 2014 ACM 978-1-4503-2782-4/14/04. ...\$10.00.

ing content is an important task. Additionally, predicting the popularity of a web object (i.e. document, image *etc*) also allows a company or website to make decisions more strategically, e.g. manage their resources (e.g. servers) better and target their advertisements more effectively. Overall, both the user benefits from a more enjoyable experience and the company benefits from a monetary saving or gain.

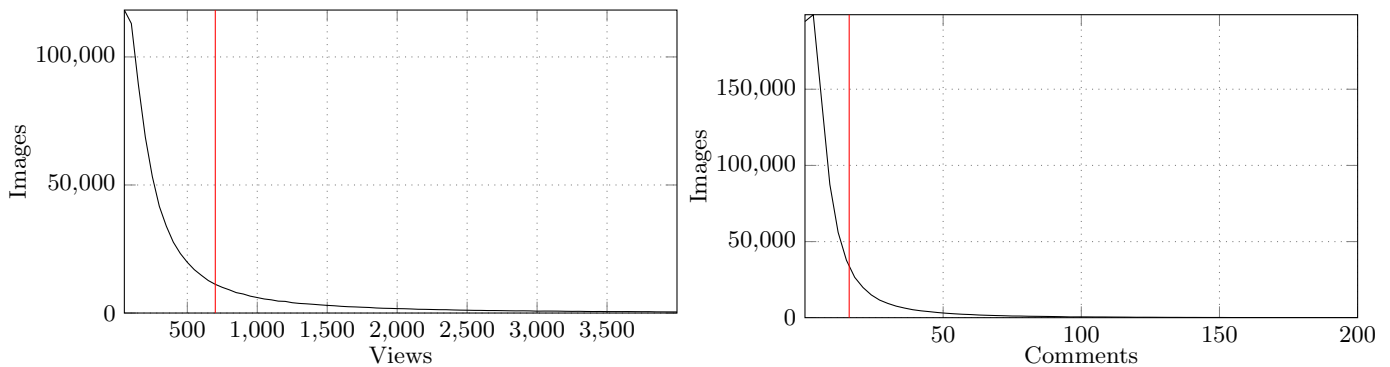
Due to the importance of popularity prediction, many works have been proposed in bookmarking [12, 9], video [20] and social [8, 1] domains. More recently, work has attempted to predict the popularity of images on Flickr<sup>2</sup> [16]. However, this approach, as well as the others mentioned, rely on the interactions of users (e.g. clicks, ratings. *etc.*) to predict the popularity of a given item/document. The majority of images are annotated with less than four tags [19] and initially contain no interaction. Therefore, an approach which requires no or little interaction data is desirable. This paper is an attempt towards alleviating this problem. Specifically, we focus on predicting the future popularity of images based on their context, and contents as well as the contextual information of the user who uploaded them rather than interaction data. For this purpose, we introduce 16 features to classify image popularity.

In this paper we consider two measures of image popularity based on the number of *views* and *comments* an image has. We consider these two measures as popularity metrics as they reflect user interest in an image and cover both implicit (i.e. views) and explicit (i.e. comments) aspects of user feedback. As shown in Figure 1, for 865,833 images, both these metrics follow a power law distribution where the majority of the images get little or no attention and the minority of them (tail of the distribution) receive a high level of attention. This makes the prediction of these images a difficult task. Understanding the underlying factors that resulted in such a dynamic can potentially be very beneficial to increase the status of the images in social networking websites. In order to do so, we qualitatively analyse each feature for predicting an image’s popularity.

The contributions of this paper are as follows:

1. We consider the exploitation of new context, content and user features, which can be inferred in a cold start scenario, for the task of image popularity prediction.
2. We introduce measures of image “popularity” based on the number of views and comments an image has had.

<sup>2</sup><http://www.flickr.com/>



**Figure 1: View (left) and comments (right) distribution. The red line is the top 20% popularity threshold.**

3. We exhaustively combine and compare these features through extensive experimentation, testing on the MIR-Flickr 1M image collection.

The rest of this paper is as follows: In Section 2 we summarise the related works in popularity prediction. We define image popularity in Section 3 before introducing the collection and proposed features for popularity prediction in Section 4. Our experimental procedure and results are then detailed in Sections 5 and 6 before concluding in Section 7.

## 2. BACKGROUND WORK

In the following we detail related works in the field of text based popularity prediction before considering those in the multimedia domain.

### *Popularity Prediction on the Web.*

Document popularity prediction first started for Web documents where the idea was to predict the future popularity of a given article. Lerman *et al.* [12] explore the popularity prediction of such content using a stochastic model of user behaviour. Testing on a Digg<sup>3</sup> collection, the authors show that an article’s popularity can be predicted from its initial user votes. Jamali *et al.* [9] exploit the implicit network behaviour on Digg to predict the popularity of submitted content. Specifically, the authors derive features from comments and social network data in a classification and regression framework. Chen *et al.* [5] predict the perceived popularity of web content with respect to a user in order to reduce loading time. Szabo *et al.* [20] study the popularity of YouTube videos and Digg posts showing that the early view patterns reflect long-term user interest.

These works, however, focus on the prediction of web documents and articles which are able to draw evidences from extensive textual data. In this work, we instead consider document prediction in an image domain, where there exists little textual evidence.

### *Popularity Prediction on Social Networks.*

An important domain where interest has increased in recent years, is that of user generated content (Web 2.0). Due to the overwhelming amount of content that is published every second on such websites, popularity plays an important content filtering role. For example, the growth of Twitter

has opened up a new area for forecasting document popularity. Hong *et al.* [8] study the importance of retweets for the prediction of popular messages on Twitter. Specifically, the authors attempt to classify tweets based on a number of content, temporal and context features. Bandari *et al.* [1] exploit properties of tweets, such as the source of the article, the category, subjectivity in the language and the entities mentioned, to predict their popularity.

Similarly, some works have focused on the popularity of images on image sharing websites, such as Flickr. Niu *et al.* [16] introduce a weighted bipartite graph model, called Incomplete Network-based Inference (INI), to predict image popularity based on network relationships. Cha *et al.* [4] study how the popularity of pictures evolves over time, showing that even popular photos propagate slowly and that they do not spread widely. Valafar *et al.* [21] focus on indirect fan-owner interactions in photos on Flickr, showing that there exists no strong relationship between an image’s age and popularity and that photos gain the majority of their fans in the first week since upload.

These works however focus only on network and interaction data thus failing in a cold start scenario i.e. where there exist no or little tag/interaction data. In this work we instead exploit an image’s visual appearance, context and user context for the task of popularity prediction.

### *Aesthetic Prediction .*

Related to image popularity, recent works have also considered an image’s aesthetic value: Dhar *et al.* [6] attempt to select images with the highest aesthetic value in a collection. In particular the authors study the effect compositional, content and sky-illumination attributes have on perceived aesthetics, evaluated on 16,000 images in a crowd-sourced experiment. Katti *et al.* [10] consider aesthetics in images from a cognitive science perspective. Categories existing in image interestingness, such as colour and structure, are defined with their cognitive load computed through experimentation on 30,000 Flickr images.

These works, however, consider an image’s aesthetic value and how an image’s visual appearance effect this. In this work, we instead focus on what makes an image *popular*, which differs from measuring aesthetic value.

### *Image Context and Contents.*

Drawing evidences and conclusions from an image’s appearance has been researched in a wide number of areas

<sup>3</sup><http://www.digg.com/>

over the past decades, from face recognition [22], to image annotation [14, 11], to image retrieval [13]. Additionally, an image’s context (e.g. time and location) has been exploited for tag recommendation and visualisation purposes: Zhang *et al.* [23] cluster tags based on geolocation and temporal trends allowing for the construction of tag cluster visualisations. McParlane *et al.* [2] exploit the daily, monthly and yearly trends of tags for tag recommendation purposes.

Exploiting these image context, content and user features, as well as their combination, however, has not yet been studied for the purposes of image popularity prediction. In this work we introduce 16 features for this purpose in a classification framework.

### 3. MEASURING POPULARITY

In order to determine an image’s popularity, we consider two aspects of an image’s interaction log for this purpose:

1. **Comments:** the number of user comments a given image has received contributes to its popularity. This feature is considered *explicit*, requiring effort from the user. We classify an image to have a *high* or *low* number of comments. The computation of these classifications is detailed in the following section.
2. **Views:** the number of views a given image has received also contributes to its popularity. This feature is considered *implicit*, requiring little effort from the user. We classify an image to have a *high* or *low* number of views. The computation of these classifications is also detailed in the following section.

In order to determine what constitutes “high” or “low” views/comments, we split our collection in two, using the Pareto Principle (or 80-20 rule) to compute thresholds, as used by existing work [3]. The Pareto Principle is often used to describe the skewness in a distribution. In our work, we use this principle to select the threshold to split between images with *high* (20%) and *low* (80%) comments and views. An image is classified as having *high* views or comments if it exists in the top 20% of the given population. The thresholds for views and comments are shown in the long tail distributions in Figure 1 (i.e.  $\geq 700$  views and  $\geq 16$  comments).

In image popularity prediction, we aim to predict whether a new image will receive a *high* or *low* number of views and comments in the future. Therefore, the problem can be formalized as that of binary classification based on a number of features. In the following section we introduce a number of features representing an image, based on its contents, context, tags and user context.

### 4. COLLECTION AND FEATURES

In our work, we classify images based on four broad feature types in order to predict whether it is popular or not. These feature categories are as follows:

1. **Image Context:** computed from the time an image is taken, its method of capture, size and orientation.
2. **Image Content:** computed based on the content (i.e. pixels) of an image. Specifically, these features are computed using scene classification, face detection and dominant colour techniques.

3. **User Context:** computed based on the user account used to upload an image i.e. their gender, account type, contact and upload information.

4. **Tags:** Finally, we compute image features based on the tags annotated by the user.

#### 4.1 Collection

This work uses the MIR-Flickr 1M collection [15] which contains 1 million Flickr images under the creative commons license. For this work, we extend the meta-data of this collection to include additional details of the user (e.g. their name) and the image context (e.g. the camera used). Therefore, as some images have been deleted from Flickr since the time this collection was released, we are left with 865,833 images taken by 38,314 users, where 96% of the images are taken between 2007 and 2010. For each image in our collection, we compute the following features:

#### 4.2 Image Context Features

For image context, we classify based on the time taken timestamp, the specified camera make, its GPS co-ordinates (if these exist), and details regarding its web context. We therefore introduce the following contextual features to represent an image:

1. **Time:** images are either taken in the *morning* (06:00 to 11:59), *afternoon* (12:00 to 17:59), *evening* (18:00 to 23:59) or *night* (00:00 to 05:59). This feature is a four dimensional binary vector based on this classification.
2. **Day:** images are either taken at the *weekend* (Fri-Sun) or on a *weekday* (Mon-Thu). This feature is a two dimensional binary vector based on this classification.
3. **Season:** images are either taken in *winter*, *spring*, *summer* or *autumn*. This feature is a four dimensional binary vector based on this classification.
4. **Device:** images are classified as taken on either a *mobile phone* or *camera*. This is achieved by manually classifying the 100 most popular camera makes (from an image’s meta data) by an assessor into each category. This feature is a two dimensional binary vector based on this classification.
5. **Size:** images are classified as *large*, *average* or *small* based on the size of the original image in pixels. We select thresholds so each set contains a similar number of users. This feature is a three dimensional binary vector based on this classification.
6. **Flash:** images are classified as flash *on*, *off* or *unknown*, based on whether the flash fired, inferred from the Exchangeable image file format (EXIF) meta-data. This feature is a three dimensional binary vector based on this classification.
7. **Orientation:** images are classified based on their orientation, or specifically the relationship between their height and width. An image is either: *landscape* (i.e. width > height), *portrait* (i.e. width < height) or *square* (i.e. width = height). This feature is a three dimensional binary vector based on this classification.

### 4.3 Image Content Features

For image content, we classify based on the visual appearance of an image, i.e. its pixels. We classify using state-of-the-art methods as described below:

1. **Scene #1:** images are classified to be one of the following scenes: *city*, *party*, *home*, *food* or *sports*. We classify images by training a multi-class SVM on the popular image feature GIST, which has been used in the past to classify an image’s scene with state-of-the-art performance [18]. To build the relevant training collections, we use those 25k images which were manually classified, as one of the given scenes, by Mechanical Turk users for the ImageCLEF 2009 task. A full description of how these images were classified can be found in [17]. From this, we trained a multi-class SVM using 5-fold cross validation to classify all the images in our collection. Best performance was achieved using the Radial basis function (RBF) kernel with parameters  $C = 2$  and  $\gamma = 2^{-3}$ , where 52.6% accuracy for classifying for the 5 scenes is achieved. This feature is a five dimensional binary vector based on this classification.
2. **Scene #2:** images are also classified to be one of the following scenes: *indoor*, *outdoor*, *macro* or *portrait*. To classify images, we use the same process as before, training a multi-class SVM on the GIST feature extracted from those ImageCLEF 2009 images manually classified as one of the given scenes. Best performance was achieved using a Radial basis function (RBF) kernel with parameters  $C = 3$  and  $\gamma = 2^{-7}$ . where 47.4% accuracy for classifying for the 4 scenes is achieved. This feature is a four dimensional binary vector based on this classification.
3. **# Faces:** using the popular HAAR Cascade methods of face detection [22], we count the number of faces in each image, classifying as: *0*, *1*, *2* or *3+*. This feature is a four dimensional binary vector based on this classification.
4. **Colour:** images are classified as being *white*, *black*, *red*, *green* and *blue* based on the dominant colour in an image. This is implemented by averaging the RGB values of each pixel in an image, and selecting the colour with minimal Euclidean distance from each of the pre-defined colour’s RGB values i.e. white (255,255,255), black (0,0,0), red (200,0,0), green (0,200,0) and blue (0,0,200). This feature is a five dimensional binary vector based on this classification.

### 4.4 User Context Features

For user context, we classify images based on the user for which an image is taken by. We classify each image for the following categories:

1. **Gender:** as Flickr does not disclose a user’s gender, we infer this by classifying based on their first name, if this exists. To do so, we collect publicly available 1990 US census data detailing the most popular male<sup>4</sup> and female<sup>5</sup> names. A user is classified as *male* or *female* if their first

<sup>4</sup><http://census.gov/genealogy/www/data/1990surnames/dist.male.first>

<sup>5</sup><http://census.gov/genealogy/www/data/1990surnames/dist.female.first>

name exists in each of these lists, or otherwise, *unknown*. In the case where a name is unisex (e.g. Stacey), we use the gender which is most popular for the name in question. This feature is a three dimensional binary vector based on this classification.

2. **Account:** we classify users based on whether they have a *pro* or *free* Flickr account. A subscription to a pro Flickr account offers more storage space and no advertisements. We use this feature to weakly infer whether a photographer is a professional or a hobbyist. This feature is a two dimensional binary vector based on this classification.
3. **# Images:** we classify a user as having uploaded a *high*, *average* or *low* number of photographs to Flickr. We select thresholds so each set contains a similar number of users. This feature is a three dimensional binary vector based on this classification.
4. **# Contacts:** we classify a user as having a *high*, *average* or *low* number of Flickr contacts. We select thresholds so each set contains a similar number of users. This feature is a three dimensional binary vector based on this classification.

### 4.5 Text Features

We represent each image in our collection by the tags it contains, using a *tf-idf* approach adopted in [7]. Due to the sparsity of tags in images [19], we represent each tag by its tag co-occurrence vector. Firstly we formulate tag co-occurrence:

If we assume that in total  $k$  unique tags represent the images in a collection of size  $n$ , the tag co-occurrence matrix would be a square matrix  $C_k$  where the value of the element  $c_{t_i, t_j}$  represents the number of images that contain both the  $i$ th and  $j$ th tags in the vocabulary. Following the approach defined in [7], we normalise each row by scaling by its maximum value. To reduce computational complexity we consider only those tags used by  $> 10$  users (i.e. 32,865 tags).

We define the representation of each tag  $t_i$  as a vector  $t'_j = (c_{t_{i1}}, c_{t_{i2}}, \dots, c_{t_{ik}})$  where each dimension corresponds to  $t_i$ ’s normalised co-occurrence value with another tag. This co-occurrence vector represents the *tf* part of *tf-idf*, where *idf* is the vector of *inverse document frequencies*. For the *idf* vector, each element computes  $\log(n/n^{(t_j)})$ , for each tag in the collection, where  $n^{(t_j)}$  is the number of images containing tag  $t_j$ . Therefore, given one or more tags, an image can be represented by  $(C \times q) \cdot \text{idf}$  where  $q$  is a number of tags in an image’s tag list. For multiple tags, the corresponding contributions are *added*. From this, we introduce the following textual feature to represent an image:

1. **Tags(m):** given  $m$  random tags from an image, we compute the *tf-idf* vector described above. In this work, we consider the cold start scenario testing with  $m = \{1, 2, 3\}$ . This feature is a 32,865 (i.e. the number of tags) dimensional vector of real values based on the output *tf-idf* vector.

## 5. EXPERIMENTS

In the following, we detail our evaluation procedure as well as define the various systems.

## 5.1 Evaluation Procedure

We investigate whether the number of views and number of comments can be predicted, given the image’s context, content and user context. For this purpose, we classify images based on the described *tf-idf* textual feature, the *seven* image context features, *four* content features, and the *four* user context features defined in Section 4. For the number of views and number of comments, we transform the values for each metric into a binary classification (+1/-1 or high/low), by using the method described in Section 3.

We learn a model to discriminate between the two classes using SVMs trained with a radial-basis function (RBF) kernel, which, based on our analysis, in the majority of cases, outperformed polynomial kernel SVMs. We also tried other models such as Bayesian logistic regression and decision trees but they underperformed with respect to the SVMs. We test on 1000 randomly selected images from the MIR Flickr 1M collection using 10-fold cross validation. These images are taken by 784 users, viewed on average 519 times and commented on average 9.1 times.

## 5.2 Systems

In our experiments, we compare the effectiveness of popularity prediction for the following systems:

1. **Baseline:** Due to the lack of work in image performance prediction, in our experiments we use a naive baseline which predicts image popularity with 50% accuracy, based on our test collection containing a 50/50 split of popular vs unpopular images.
2. **Text( $m$ ):** In this system, we classify images based on the *tf-idf* representation of  $m$  random tags. We consider this the “*oracle*” approach.
3. **Ind( $f$ ):** This system classifies images based on an individual feature,  $f$ , allowing for their effectiveness comparison.
4. **Comb( $l$ ):** Finally, we also consider the combination of image and user features, where the vectors are concatenated. With  $l = \{all, context, content, user\}$ , representing the different feature types e.g. Comb(context) is the system with classifies based on the combination of all context features.

## 6. RESULTS

In the following section we detail the results of our experiment for image popularity prediction. Firstly, we consider the “predictability” of our introduced measures before studying the prediction results of our experiment.

### Popularity Measure Effectiveness.

Figure 2 shows the classification accuracy averaged over all images in the test set. Firstly, it can be observed that we are able to successfully classify an image’s popularity, achieving accuracy of up to 76% when classifying comment count and 59% when classifying an image’s view count. Therefore, we highlight that the number of comments is more highly correlated with the discussed features than image views. The explicit nature, and higher effort required, for commenting in comparison to viewing an image, makes comment count a more effective measure of image popularity.

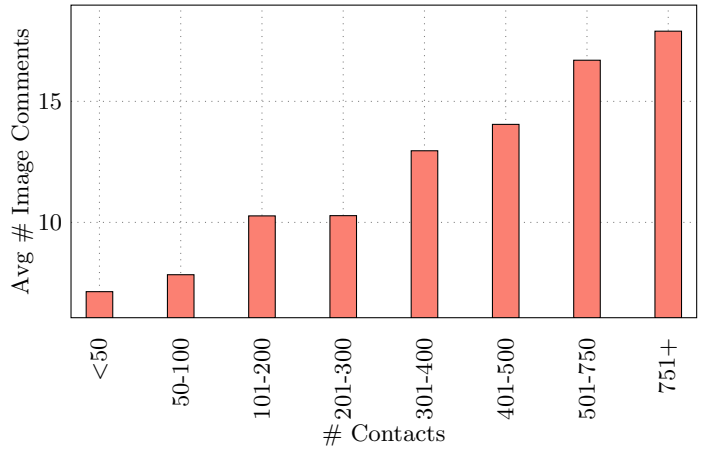


Figure 3: # Comments & contacts relationship

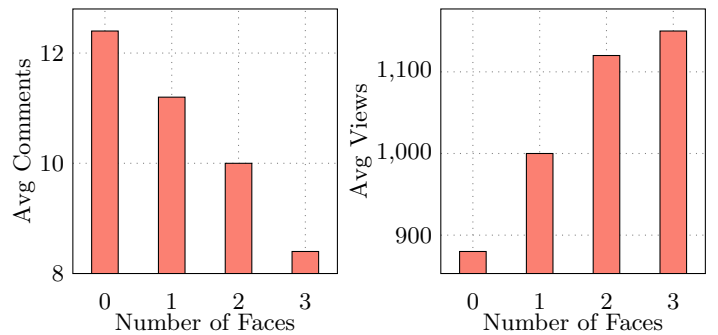


Figure 4: View (left) and comments (right) distribution with popularity threshold.

Viewing an image only requires a click, whereas commenting requires viewing an image, constructing an opinion and inputting a textual message. Due to this effort difference and based on our results, we hypothesise that images with more comments are likely to be of higher interest to a users than those with simply high views. Any future system attempting to predict an image’s popularity should therefore train on comment count over image count.

### Image Popularity Topic Analysis.

In Table 1, we compute the most significant tags, sorted by descending order, for images with high/low comments and views. This is computed as the fraction of images tagged with tag  $t$  in images classified as  $x$ , minus the fraction of images tagged with  $t$  in all images, for the MIR-Flickr 1M collection. By doing so, we identify the tags which occur significantly more in a subset of images sharing a common feature, than in the global set. It can therefore be observed from Table 1 that there exists a relationship between the topical content of an image and its popularity.

Images which the highest views tend to be images of people, especially women, (i.e. *girl*, *portait*, *woman*) in comparison to images with low comments which are often of nature. Despite nature images being the lowest viewed, they are commented the highest. This highlights that there exist many low quality nature photos, yet few high quality

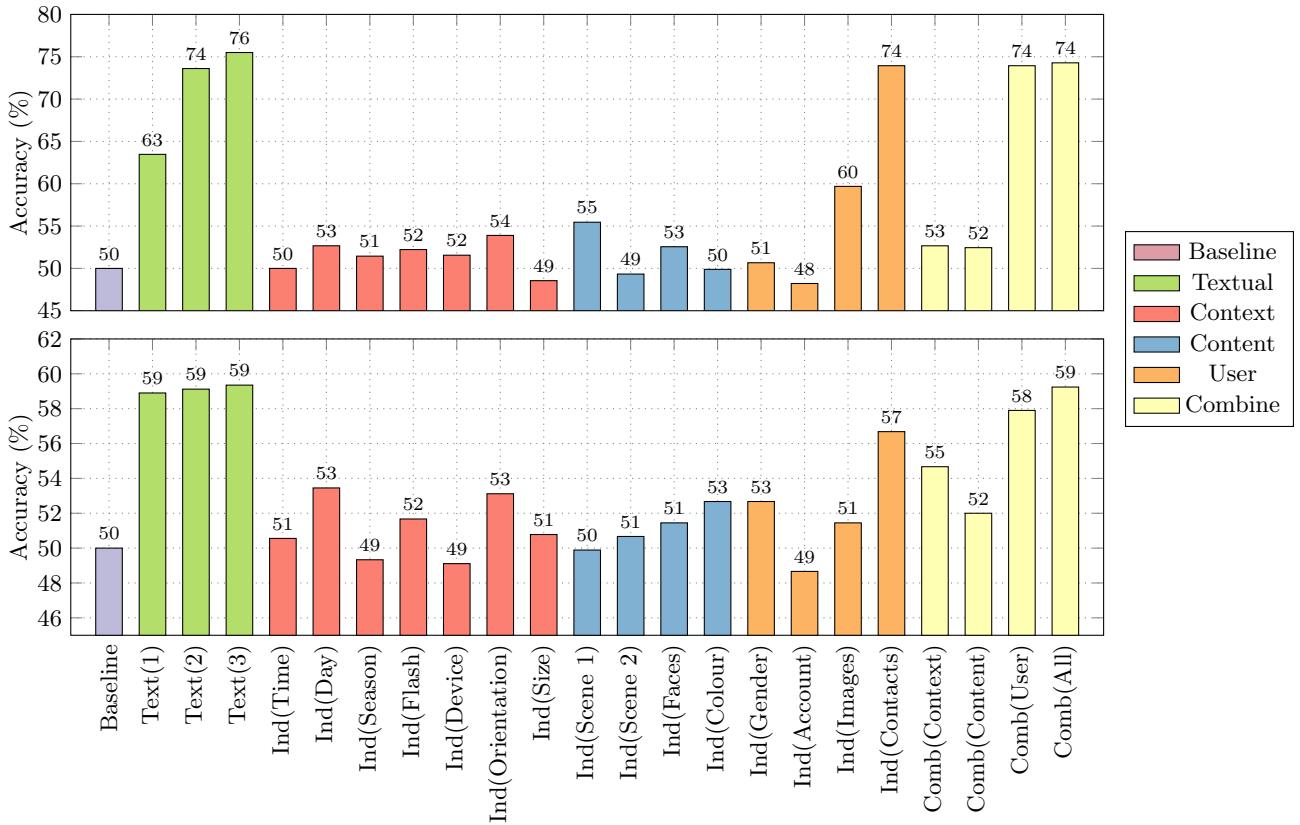


Figure 2: SVM accuracy when classifying comments (top) and views (bottom).

$x$	Top Tags ( $t$ )
Views	<b>High</b> girl, portrait, woman, hdr, explore
	<b>Low</b> flower, macro, cat, flowers, nature
Com	<b>High</b> abigfave, aplusphoto, anawesomeshot, nature
	<b>Low</b> 2008, california, art, graffiti, sanfrancisco

Table 1: Most significant tags for high/low comments and views. Comments is abbreviated to “com”.

nature photos which provoke much discussion and interest. The paradox captured by this paper’s title (“Nobody comes here anymore, it’s too crowded”), highlights this need to predict popular images in order to filter out these images of lower quality, and promote those of high quality to the user. Finally, our hypothesis that user comments are a more reliable measure of popularity is confirmed in Table 1 where the highest commented images contain many Flickr awards (e.g. *abigfave*, *aplusphoto*, *anawesomeshot*).

### Individual Feature Performance.

For *context*, the day type and orientation are the most discriminative features for both comments and view prediction; whereas for *content*, the features do not follow a trend for both metrics. All content features are able to offer some prediction improvement over our baseline, however, the improvement is minimal. We hypothesise that this is because images that grab the attention of the user are different, in

some way, from the majority; therefore the visual appearance of images cannot be effectively used for popularity prediction. However, we do identify a meaningful relationship between the number of faces present in an image and its popularity, as shown in Figure 4. We observe that images with less faces attract less views but have more comments; conversely those images with many faces (e.g. party photos) contain many views but few comments highlighting a high *browsing* and low *discussion* motivation for images with multiple people.

By far, the most important feature for image popularity prediction, however, concerns the user themselves where the number of contacts and images they have correlates highly with view and comment count. We observe that an image’s popularity is linked closely to the user’s popularity and activity, opposed to their contents and context. Further, by relying solely on the number of contacts a user has, we are able to achieve comparable popularity prediction performance in comparison to the case where multiple tags are present, overcoming the cold start scenario and highlighting the correlation between contact and comment count; this relationship is also captured in Figure 3. Finally, the number of images a user has uploaded is also strong correlated with image popularity, where by exploiting this feature we achieve 60% prediction accuracy.

### Combination Performance.

Combining evidences from an image’s context, content and user context gives the best results in most cases. Specif-

ically, combining all three gives the highest popularity prediction when comparing against all the features individually, and the other combination approaches. Further, we observe that by combining all three evidences, we are able to match or outperform (except Text(3) for comments) the case where tags are present for popularity prediction, thus highlighting the merit of our approach where there lacks textual evidence.

## 7. CONCLUSION AND FUTURE WORK

Predicting the popularity of a web object has become an important task in recent years for social media websites in order to filter an ever expanding data set and maximise company profits. Instead of relying on interaction and textual data, however, as adopted by existing work, this paper considered the challenging task of *image popularity prediction* in a cold start scenario i.e. where there exists no or little textual/interaction data. Evidences were instead computed from an image's context, contents and user context.

The findings of our experiments showed that we were able to predict, with up to 76% accuracy, whether an image with receive high or low user comments in the future. A user's context was seen to be the most effective feature set, where we showed that the popularity of an image is closely related to the user's popularity and activity level on Flickr, as well as its topical content (i.e. tags). Finally, by combining evidences from an image's context, content and user context we achieved highest popularity prediction accuracy highlighting that the features are complementary for this purpose and reliable in a cold start. This work leaves a number of interesting research questions, such as how we effectively weight these features, and how we combine them with textual evidences for popularity prediction, which is left as future work.

## 8. REFERENCES

- [1] R. Bandari, S. Asur, and B. A. Huberman. The pulse of news in social media: Forecasting popularity. In *ICWSM*, 2012.
- [2] P. J. McParlane, Y. Moshfeghi and J. M. Jose. On Contextual Photo Tag Recommendation. In *ACM SIGIR*, 2013.
- [3] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *ACM SIGCOMM IMC '07*.
- [4] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *ACM WWW '09*.
- [5] X. Chen and X. Zhang. A popularity-based prediction model for web prefetching. *Computer*, 36(3):63–70, 2003.
- [6] S. Dhar, V. Ordonez, and T. Berg. High level describable attributes for predicting aesthetics and interestingness. In *IEEE CVPR '11*.
- [7] N. Garg and I. Weber. Personalized, interactive tag recommendation for flickr. In *ACM RecSys '08*.
- [8] L. Hong, O. Dan, and B. D. Davison. Predicting popular messages in twitter. In *ACM WWW*, 2011.
- [9] S. Jamali and H. Rangwala. Digging digg: Comment mining, popularity prediction, and social network analysis. In *Web Information Systems and Mining*, 2009. *WISM 2009. International Conference on*, pages 32–38, 2009.
- [10] H. Katti, K. Y. Bin, T. S. Chua, and M. Kankanhalli. Pre-attentive discrimination of interestingness in images. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 1433–1436, 2008.
- [11] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1106–1114, 2012.
- [12] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 621–630, New York, NY, USA, 2010. ACM.
- [13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [14] A. Makadia, V. Pavlovic, and S. Kumar. Baselines for image annotation. *International Journal of Computer Vision*, 90(1):88–105, Oct. 2010.
- [15] B. T. Mark J. Huiskes and M. S. Lew. New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative. In *MIR '10*, pages 527–536, New York, NY, USA, 2010. ACM.
- [16] X. Niu, L. Li, T. Mei, J. Shen, and K. Xu. Predicting image popularity in an incomplete social media community by a weighted bi-partite graph. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 735–740, 2012.
- [17] S. Nowak and P. Dunker. Overview of the clef 2009 large-scale visual concept detection and annotation task. In C. Peters, B. Caputo, J. Gonzalo, G. Jones, J. Kalpathy-Cramer, H. Müller, and T. Tsirikla, editors, *Multilingual Information Access Evaluation II. Multimedia Experiment, volume 6242*, 2010.
- [18] A. Oliva and A. Torralba. Chapter 2 building the gist of a scene: the role of global image features in recognition. In L. M. J.-M. A. S. Martinez-Conde, S.L. Macknik and P. Tse, editors, *Visual Perception Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception*, 2006.
- [19] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *ACM WWW*, 2008.
- [20] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, Aug. 2010.
- [21] M. Valafar and R. Rejaie. Beyond friendship graphs: A study of user interactions in flickr. In *In Proc. ACM SIGCOMM WOSN*, 2009.
- [22] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR volume 1*, 2001.
- [23] H. Zhang, M. Korayem, E. You, and D. J. Crandall. Beyond co-occurrence: Discovering and visualizing tag relationships from geo-spatial and temporal similarities. In *ACM WSDM 2012*.