Chic or Social: Visual Popularity Analysis in Online Fashion Networks

Kota Yamaguchi  
Tohoku University  
Sendai, Miyagi, Japan  
kyamagu@vision.is.tohoku.ac.jp

Tamara L. Berg  
University of North Carolina  
Chapel Hill, NC, USA  
tlberg@cs.unc.edu

Luis E. Ortiz  
Stony Brook University  
Stony Brook, NY, USA  
leortiz@cs.stonybrook.edu

ABSTRACT

From Flickr to Facebook to Pinterest, pictures are increasingly becoming a core content type in social networks. But, how important is this visual content and how does it influence behavior in the network? In this paper we study the effects of visual, textual, and social factors on popularity in a large real-world network focused on fashion. We make use of state of the art computer vision techniques for clothing representation, as well as network and text information to predict post popularity in both in-network and out-of-network scenarios. Our experiments find significant statistical evidence that social factors dominate the in-network scenario, but that combinations of content and social factors can be helpful for predicting popularity outside of the network. This in-depth study of image popularity in social networks suggests that social factors should be carefully considered for research involving social network photos.

Categories and Subject Descriptors
I.2.10 [Vision and Scene Understanding]: Perceptual reasoning

General Terms
Experimentation; Human Factors; Measurement

Keywords
Social multimedia; Online fashion networks

1. INTRODUCTION

Nearly every blog or social network utilizes a combination of images, text, and other modalities (e.g. location) to convey information and promote interaction. In many online communities the amount of visual data is quite vast, sometimes representing the main source of content. Despite the active research in social multimedia and social popularity hypothesis [1], it is not well-studied how much the content quality and the network is influencing to the resulting visual content popularity. In this paper, we take a big-data approach to quantitatively study social influence in a real-world online social network specialized in fashion.

We purposely choose a fashion-focused online community, Chictopia, for the following reasons: 1) The network is large and real-world, with over 175k users and 600k pictures, 2) Content in this network is mainly visual, consisting of “outfit of the day” pictures uploaded by users, 3) The community is focused on a single topic (fashion), which yields relatively consistent user based popularity in contrast to general photos with diverse categories [7], 4) The relevant data is publicly and readily available online, 5) We can take advantage of state of the art computer vision techniques for recognizing clothing in fashion images [10] to help extract the visual content most relevant to popularity.

In a network focused around fashion and style, one might assume that visual content would be the most influential factor for popularity. However, we find that social factors dominate both visual and textual factors in prediction models, even in a community where outfits are purportedly rated based on their fashion style. Furthermore, studying the effects of content outside of the social network, we find that this social bias does not appear, but rather the social and content information together provide a good predictive factor for popularity. These insights are useful for multimedia researchers and engineers seeking to exploit human behavior in social network applications.

The following list summarizes our contributions:

- Models to predict popularity of outfit pictures incorporating visual, social, and textual factors.
- A new computer vision feature for representing outfit style based on clothing parsing (semantic segmentation of clothing items).
- A large-scale empirical study of social vs. content influence on popularity in a real-world, uncontrolled fashion network.
- A large crowdsourcing effort to simulate the socially isolated condition.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Figure 1: An example of a Chictopia post.
Figure 2: Distribution of voting, commenting, and bookmarked based popularity in Chictopia. The number of votes shows a slight kink perhaps due to front-page highlighting or special promotion by the website.

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Modality</th>
<th>Vector</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>User identity</td>
<td>Network</td>
<td>Sparse</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Node degrees</td>
<td>Network</td>
<td>Dense</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Previous posts</td>
<td>Metadata</td>
<td>Dense</td>
<td>1</td>
</tr>
<tr>
<td>Content</td>
<td>Tag TF-IDF</td>
<td>Textual</td>
<td>Sparse</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Style descriptor</td>
<td>Visual</td>
<td>Dense</td>
<td>441</td>
</tr>
<tr>
<td></td>
<td>Parse descriptor</td>
<td>Visual</td>
<td>Dense</td>
<td>1060</td>
</tr>
<tr>
<td></td>
<td>Color entropy</td>
<td>Visual</td>
<td>Dense</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Image composition</td>
<td>Visual</td>
<td>Dense</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>Date bias</td>
<td>Metadata</td>
<td>Sparse</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 1: Summary of the content models.

- Studies of within-network and outside of network scenarios, including empirical findings of asymmetry in content popularity prediction for these scenarios

Related Work: Our work is related to a growing interest in popularity prediction of online content. Work in this direction has mainly looked into early social reaction to content and the prediction of popularity growth in videos [3] [10] [14], news [10] [18], and discussion forums [9]. Some recent work has looked into visual influence on popularity or behavior [2] [6] [17], or in the reverse direction, to categorize visual contents using social information [8]. Social influence on popularity seen in these studies is also consistent with social browsing behavior on Flickr [11] [17]. We distinguish ourselves from the previous work in that: we focus on a community devoted to fashion, and in particular to rating outfits, our main goal is to evaluate how much social influence affects popularity for in and out of network scenarios, we take advantage of recent state-of-the-art computer vision approaches to recognize clothing (rather than general visual features or content), and we take a big-data approach using real-world data from a social networking site focused on a single topic (fashion).

2. DATASET

We collect data from chictopia.com, a social fashion network where users post pictures of their daily outfits along with a title, description, and several labels. Figure 1 shows an example of picture and metadata we can observe. We initially collect 617,708 posts from Chictopia. To compare visual features consistently across images, we run a state of the art pose detector to automatically select pictures with a standing person detected. This leaves us with 328,604 pictures, dating from March 2008 to Dec 2012, with 34,327 unique users.

Figure 3: Style descriptor (left) and parse descriptor (right). Style descriptor extracts visual information from patches while parse descriptor extracts information from the predicted clothing parse (semantic assignment of pixels to garment labels).

The observable popularity measures in this data are the number of votes, comments, and bookmarks associated with each post. As is the case with any web content, Chictopia popularity reveals a long-tailed distribution. Figure 2 plots the popularity histogram from all 617K posts. To deal with the difficulty in such distribution, we consider log-votes as the popularity of posts in our experiments.

3. CONTENT REPRESENTATION

We represent a post as a vector of the quantized information sources available for the post. Two major components are modeled for each post: social factors and content factors. Social factors capture information related to the user and their social status within the network, whereas content factors are to capture the quality of the content. In addition to social and content factors, we extract timestamp information to account for the popularity change due to seasons and site growth over time. We use a sparse indicator to represent the month of the post (a 58 dimensional vector in our experiments). Table 1 summarizes all of the terms of our model. All factors are appropriately normalized.

3.1 Social factors

User identity: We represent the identity of users using a sparse indicator vector. To constrain feature dimensionality we restrict this indicator vector to the top-1000 most frequent users. Posts from the same user will have the same feature vector.

Node degree: We use counts and log-counts of friends, followers, and followees of the user as a six-element feature vector. Posts from the same user will all have the same feature. Note that we empirically tested other network features, but none produced a better prediction than node degree.

User Expertise: We use the number of previous posts from the same user as a scalar feature, related to user expertise.

3.2 Content factors

Tag TF-IDF: In Chictopia, a user can label each individual post with various structured tags indicating general style, occasion, colors, brands, clothing types, in addition to free-form words. We first extract unigrams and bigrams from all the tags and treat them as a document, and compute TF-IDF weights. As in the case of user identity, to constrain the dimensionality of this feature, we only consider the 1000 most frequent n-grams found in the training samples.

Style descriptor: The style descriptor is a state-of-the-art visual clothing representation proposed in 19 for fashion image retrieval, and considered a comprehensive representation of the overall fashion style. The descriptor is based
4. IN-NETWORK POPULARITY

Experimental protocol: We first apply a linear regression analysis on the log-votes of the posts using social, content, and a combination of social and content factors. For this analysis, we adopt $R^2$ and Spearman coefficients as measures of fitness of prediction to the truth. These measures are evaluated on a statistical bootstrapping protocol with our 328K posts: We randomly resample posts, subsample this dataset to 10,000 posts (for computational tractability), and evaluate the above measures with a 90%-10% train-test split. This process is repeated 100 times to derive statistical significance.

We also apply a classification based analysis, in which we predict a binary indicator of being Top $K$% in popularity. In this experiment, we vary our threshold for 25% and 75% percentiles of the votes in the training samples to see the difference between the most popular and least popular posts. The performance is measured in terms of accuracy.

Results: In-network column of Table 2 shows the results of regression analysis. The regression models fit significantly better when social factors are present, suggesting a user’s social connections largely dominates the popularity of their posts over the content itself. However, we should note that social factors may also be highly correlated with content quality – users with many followers may tend to wear highly fashionable outfits. Nevertheless, our results indicate that the influence of content quality is considerably smaller than network influence.

Classification reveals an asymmetry between top 25% and top 75% prediction, indicating the prediction of the most popular posts is easier than predicting the least popular posts. We suspect that this is partly due to the consistently better quality of top-rated pictures and partly due to social bias more strongly affecting popular posts. Also, the slightly smaller difference between the social-only and the content-only model in top 75% prediction suggests that popularity is less affected by social influence in least popular posts.

5. OUT-OF-NETWORK POPULARITY

To examine the effects of social and content factors more deeply, we utilize crowdsourcing to emulate a content network without social relationships. We run the popularity voting process in Amazon Mechanical Turk, where no social network exist.

Our task shows the crowd worker 50 random large-resolution images in sequence, and asks them to vote on the picture if they find it chic. For quality control, we measure how long it takes each worker to complete this first step and reject a worker’s votes if they completed too quickly or didn’t display enough variation in their voting procedure. We perform our experiments by randomly selecting 3,000 posts (from our dataset of 328k) and instantiating the above tasks 60 times. We assign 25 workers to each of the 60 resulting tasks. Therefore, any post on each task can obtain up to 25 chic votes.

Using the voting data from the crowd, we apply the same analysis. Our main interest here is, however, the influence of the social factors observed in Chic-topia on crowd popularity.

We use the same bootstrap method from the 3,000 posts to compare the social-only, content-only, and combined models in this experiment. Here, social factors are taken from Chic-topia dataset which has no relationship to crowd popularity.

Results: Results are shown in The out-of-network column of Table 2 shows the result. Given our previous results, we initially expected the social factors to lead to much weaker

Table 2: Regression and top-K% classification on the observed popularity with accompanying 95% bootstrapped confidence intervals on error. For cleaner presentation, the tiny asymmetric difference in confidence intervals are rounded.
predictors. However, the results suggest that social factors (from Chictopia) still lead to comparable predictors in regression. We conjecture that the solid regression result of the social factors is due to the user and content quality correlation.

Also, the combined model (social + content) is significantly better than any single-factorized model in regression. One possible explanation to this result is that the social factors are actually providing complementary information to the content factors in predicting the unbiased popularity from the crowd, as opposed to the biased popularity in the network where social influence is by far the stronger predictor of popularity. We observed the asymmetry of prediction also in the out-of-network condition, but to weaker extent perhaps due to no social bias in voting.

Apart from factorizing content and social influence, we can also use our learned models to predict the popularity of photos. Figure 4 shows an example of the most and least popular pictures predicted by one of our models. There is clearly a distinction in visual quality between the most and the least popular pictures. Perhaps it is also possible to build a system that can predict unbiased content popularity. Such prediction could be useful for many e-commerce applications, such as automatic outfit quality feedback [12], socially-aware fashion recommendation [13]. Stable popularity prediction can benefit in online ad optimization and traffic balancing. It is our future work to use this insight to build a socially-aware multimedia system.

In summary, the out-of-network popularity analysis yield insights that suggest 1) social factors contain not only network information but also some aspect of content evaluation, 2) content factors capture different aspects of popularity than social, and 3) their combination yields better predictions for content popularity.

6. CONCLUSIONS

We presented a vision-based approach to quantitatively evaluate the influence of social and content factors on fashion pictures. Our content representation takes advantage of various sources of information from computer vision, natural language processing, and the network itself. Through experiments, we showed statistical evidence of dominant social influence in networked media and the strength of social factors in unbiased content evaluation. One lesson from our experiments is that any attempt to learn subjective measures such as aesthetics from social content should explicitly consider social influence.

7. ACKNOWLEDGEMENTS

Research supported by Google Faculty Award, “Seeing Social: Exploiting Computer Vision in Online Communities,” NSF Career Award #1054133 and #1054541.

8. REFERENCES