Using images shared on Twitter to predict selected socio-economic characteristics of urban populations

Abstract

Traditional methods of recording socio-economic information about populations, such as censuses and surveys, have poor temporal resolution and are costly to conduct. More recently, computational models that use text features from online social network posts can predict several key socio-economic variables at high accuracies and bypass the aforementioned limitations. However, even these models so far only use text (such as Tweets), ignoring another key type of social media: images. In this paper we explore features from visual social media to develop computational models that estimate several socio-economic characteristics. We extract simple features, such as color histograms and number of faces, from over 7 million images posted on Twitter in 2013 across 60 U.S. cities. We find that aggregated characteristics of these images can be used to accurately predict income, housing prices, education levels, and financial well-being indicators. Our results suggest that images shared on online social networks reflect socio-economic characteristics and that this data can compliment existing computational models that use only text.

Collecting information about populations in particular geographic areas is of vital importance for governments, public agencies, and companies, to assist in budget allocations, defining policies, determining markets, and numerous other decisions. For example, "objective" measures such as income distributions and education levels, play a crucial role in determining the social state of a population. Increasingly institutions are beginning to compliment such objective measures with "subjective" measures such as well-being and "happiness" to more completely capture the nuances of societal needs. The UN General Assembly in particular passed a resolution in 2011 stating that its member countries should measure the happiness of their populations to aid in public policy creation. Typically, measurements of subjective well-being and happiness are either obtained directly by way of surveys (Kahneman et al. 2006; Lett et al. 2004), or inferred indirectly from a composite of socio-economic measures (that is, a predefined utility function) such as weighted combination of housing prices, health distributions, and crime. In surveys, responders answer questions such as "On a scale from 1 to 4, how happy are you?" (Oswald and Wu 2010). The authors in (Mitchell et al. 2013) combine census data, FBI crime data, the BRFSS and Gallup surveys to create an overall composite measure of well-being. All these measurements and utility estimations can aid policymakers in determining the needs of a population in various parts of a country.

Several computational studies have shown that large-scale social media content can be used to predict a variety of socioeconomic and cultural characteristics of populations (Schmidt 2012; Einav and Levin 2014; Bollen, Mao, and Zeng 2011; Manovich 2011). For example, (Bollen, Mao, and Zeng 2011) found that measurements of aggregate moods using Tweets can predict sufficiently well the value of the Dow Jones Industrial Average over time. Other topics for prediction from social media have included flu trends, product sales, and results of elections (Gayo-Avello 2013). Typically, these studies extract features from the text content of a large number of Twitter posts, usually in the millions. One popular type of feature is the so-called "bag-of-words" that counts frequency statistics for all words appearing in the posts (Rajaraman et al. 2012). This feature can be used to calculate sentiments of tweets. The sentiment scores are then used to predict population characteristics. For example, (Golder and Macy 2011) carried out sentiment analysis using aggregate text features to measure diurnal mood patterns in Twitter posts worldwide. (Mitchell et al. 2013; Frank et al. 2013; Eichstaedt et al. 2015) developed computational models that correlate sentiment with income, crime, education levels, financial well-being, and health statistics. Collectively, these studies suggest that there is valuable signal in the text of social media posts.

However, using only the bag-of-words feature set has several limitations. One limitation is that estimating sentiment based on bag-of-words has dependence on specific languages (for example, LIWC, a popular linguistics database, is only available in English (Tausczik and Pennebaker 2010)). Since social media is a global phenomenon, a purely language-based analysis will require as many lexical databases as there are languages. Therefore this approach at present is not applicable to all countries. Another limitation with only using bag-of-words features is that non-textual data that is shared, such as images and video, are ignored. Increasingly, images and videos are becoming the chosen medium for users to create and share content (Cen-

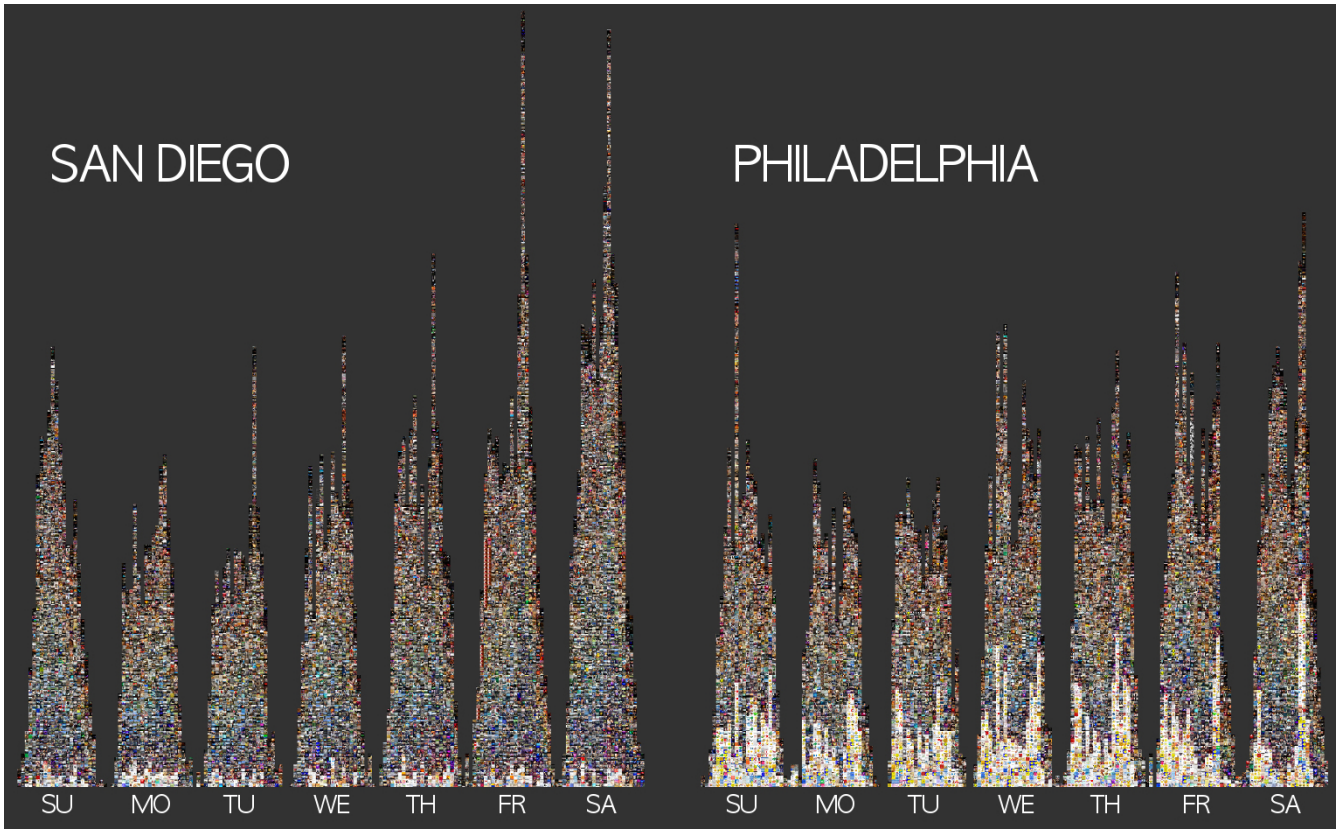

Figure 1: Comparison between the temporal distributions of Tweeted images for one-week using 10,000 random samples from San Diego and Philadelphia. Each "bin" corresponds to one hour in the week. Note that volume of tweets follow distinct temporal patterns for these two cities. Also, as we "zoom-in" to the images, the content of photographs also are unique for cities.

ter 2015). While there are many other limitations with the bag-of-words feature set, here we develop new features that instead use images shared on social media.

Using such images features, in this paper we show that visual social media can also be used to estimate selected socioeconomic characteristics. Specifically, by collecting and analyzing images shared in particular geographic areas we can gain insights into the lives and emotions of people living in these areas. Unlike the studies of shared text messages that rely on the meanings and frequencies of words (and are therefore language-dependent), we will only use color characteristics and the number of faces detected in images. Such features do not require country or city specific semantic databases.

To further motivate that Tweeted images may reflect geospatial patterns, in Figure 1 we show a random 10,000 sample of Tweeted images organized by day and hour from two US cities. Figure 1 can be thought of as an "image histogram" where each bin corresponds to an hour of the week and we place the actual tweeted image in the respective hour-bin that it was tweeted. Just as a traditional histogram, the height of the image histogram suggests that more data (images) is available in that hour bin. Since we controlled for the number of samples (that is, San Diego and Philadelphia each have 10,000 randomly sampled images), we can make fair comparisons. Figure 1 suggests that people in these two cities use Twitter very differently. Furthermore, since we are actually plotting the image itself in the image histogram, we are able to see that the content of images also varies between these two cities. In this paper we try to quantify these features to make more rigorous comparisons between different cities.

Previous studies of large volumes of Twitter messages found distinct diurnal patterns. They showed that the moods of the text in aggregated Tweets (Golder and Macy 2011) and the levels of happiness (Dodds et al. 2011) vary depending on hour of the day, day of the week and season. These patterns also vary between countries (Golder and Macy 2011) and cities (Mitchell et al. 2013). Together, these studies show that both geographic and diurnal characteristics of Tweeted messages are useful in predicting socio-economic characteristics of populations. We use similar methodologies but apply them to images. Several other studies, such as (Abdullah et al. 2015; **?**) have also used social media images for predicting socio-economic characterstics and our work contributes further evidence for this research.

Methods

In our paper, we extract a number of features from over 7 million images shared during 2013 in the 60 largest cities in the lower 48 US states. We then use these features in a linear regression model to predict the values of four socio-economic characteristics of the populations living in these cities. In the following, we first discuss how we constructed our features, followed by a discussion of the linear model we used in the paper.

Data

As recipients of a Twitter Data Grant (2014), we were given unique access to all of Twitter's historical data. For this study, we used all geo-tagged, publicly shared images that were tweeted within the lower 48 US states in 2013. Starting with this initial set of 22 million images, we filtered them to leave only images shared in the 60 largest US cities. To perform this filtering, we used the Yahoo! API to define municipal bounding boxes as was done in (Zhang et al. 2013). The final dataset contains over 7 million images. The metadata includes GPS coordinates, date and time stamp, the optional text (i.e., a tweet) content that accompanies the image, and user name. (We removed user information from the data before starting the analysis).

Color Histograms

To represent the colors and brightness of an image, we extract Hue, Saturation and Value (HSV) information. We use 180 bins for Hue, 256 bins for Saturation and 256 bins for Value (Bradski 2000). For each image $k = 1, \ldots, K$ (where K is the total number of images in our data set) at geographic location g with time stamp t, we have three sets of histograms:

$$\begin{aligned} H_k^{g,t} &= (H_{k,1}^{g,t}, \ldots, H_{k,180}^{g,t})^T \\ S_k^{g,t} &= (S_{k,1}^{g,t}, \ldots, S_{k,256}^{g,t})^T \\ V_k^{g,t} &= (V_{k,1}^{g,t}, \ldots, V_{k,256}^{g,t})^T \end{aligned} \tag{1}$$

where each $H_{k,l}^{g,t}, S_{k,l}^{g,t}, V_{k,l}^{g,t}$ is the lth bin count in the hue, saturation, and value histograms for the kth image. Note that since images can have different dimensions, the bin counts can vary widely. Therefore, we normalize the color histograms by dividing each bin with the total number of bins in the image. This can be computed with $\Sigma_l(\cdot)$ for each histogram channel as follows:

$$\begin{aligned} \hat{H}_k^{g,t} &= (H_{k,1}^{g,t}, \ldots, H_{k,180}^{g,t})^T / \Sigma_l H_{k,l}^{g,t} \\ \hat{S}_k^{g,t} &= (S_{k,1}^{g,t}, \ldots, S_{k,256}^{g,t})^T / \Sigma_l S_{k,l}^{g,t} \\ \hat{V}_k^{g,t} &= (V_{k,1}^{g,t}, \ldots, V_{k,256}^{g,t})^T / \Sigma_l V_{k,l}^{g,t} \end{aligned} \tag{2}$$

We refer to each lth bin in Equation 2 as $\hat{H}_{k,l}^{g,t}, \hat{S}_{k,l}^{g,t}, \hat{V}_{k,l}^{g,t}$. In an attempt to simplify notation, we henceforth drop the hat $\hat{(\cdot)}$ notation and refer to the normalized histograms in Equation 2 as $H_k^{g,t}, S_k^{g,t}, V_k^{g,t}$.

Figure 2: 12 bin representation of the 180-bin hue histogram. These 12 hues define the edges for the Hue histogram used for computing the hourly time series of 4.

Computing Dominant Colors and Measuring the Diurnal Patterns of Hues

We want to measure changes in the HSV values of images shared over 24 hours in a given city. To be able to analyze these changes reliably, we need to have a significant number of images for every hour. To do this, we aggregate all images shared per hour over the duration of one year (2013). While we can look at these variations in full (i.e. 180- or 256-bin histograms), to make results more interpretable, it would very useful to simplify these histograms. After extensive experimentation, we came up with a new feature which we refer to as the dominant color of an image. We first reduce the full histogram from either 256 or 180 bins to 12 bins. The resulting hues (at fixed Saturation and Value) are shown in Figure 2. Now each bin has a distinct and perceptually meaningful color. We define the dominant color of an image as the mode of the image's 12-bin Hue histogram:

$$\begin{aligned} H_{k,\text{mode}}^{g,t} &= \underset{i}{\text{argmax}}\ H_{k,i}^{g,t} \\ S_{k,\text{mode}}^{g,t} &= \underset{i}{\text{argmax}}\ S_{k,i}^{g,t} \\ V_{k,\text{mode}}^{g,t} &= \underset{i}{\text{argmax}}\ V_{k,i}^{g,t} \end{aligned} \tag{3}$$

Thus the colors of each image k are summarized by the triplet $(H_{k,\text{mode}}^{g,t}, S_{k,\text{mode}}^{g,t}, V_{k,\text{mode}}^{g,t})$ for each geographic location g and time stamp t. In other words, the dominant color is one of the 12 colors that occurs most often in an image. Strictly speaking, since we reduced the histograms to 12 bins, each reduced bin corresponds to a range of values of the original histogram, and so the "dominant color" is not a single color but a range of colors.

In this paper, we focus on measuring only the diurnal patterns, if any exist, using a number of different features. Here what we mean by diurnal patterns are the temporal patterns that emerge for an image feature over the course of 24 hours. (Naaman et al. 2012) compute time series for keywords in tweets that reveal diurnal patterns (for example, they show that the word "lunch" has diurnal patterns corresponding to peaking at noon). Inspired by their approach, we compute the following series for each dominant color Hue_i:

$$\text{X24}_g(h, \text{Hue}_i) = \frac{1}{K_h^g} \sum_{k=1}^{K_h^g} I(H_{k,\text{mode}}^{g,t_h} = \text{Hue}_i) \tag{4}$$

for $h = 0, \ldots, 23$, where K_h^g are the total number of images from hour h in location g, $I(\cdot)$ is the indicator function, and t_h are all time stamps corresponding to hour h.

To put in more simple terms, Equation 4 computes for each hour the percentage of images that have a a specific dominant hue Hue_i. This time series is analogous to the time series that (Naaman et al. 2012) computes for keywords in the text of tweets. Similar to the interpretation that (Naaman et al. 2012) give for the series in Equation 4, $\text{X24}_g(h, \text{Hue}_i)$ represents the expected hourly values for each dominant color and geographic location. To measure variations of $\text{X24}_g(h, \text{Hue}_i)$ across geographies (cities), we compute the entropy for this series for each location g as was done in (Naaman et al. 2012) for keywords. This is done by normalizing $\text{X24}_g(h, \text{Hue}_i)$ for each location g and each color Hue_i to $p(h) = \text{X24}_g(h, \text{Hue}_i)/\Sigma_h \text{X24}_g(h, \text{Hue}_i)$ and calculating $p(h) = -\Sigma_{h=0}^{23} p(h) log(p(h))$. Entropy measures the amount of "spread" or "dispersion" in the series $\text{X24}_g(h, \text{Hue}_i)$. When the time series over the course of 24 hours has peaks high peaks, the entropy has small values.

Aggregated Color Histograms and Pairwise Differences

In addition to dominant colors, we also construct another set of features by aggregating the normalized HSV histograms from Equation 2 for each geographic location and hour. This is done by computing the average for each bin based on location and time over all days in our data as follows: $H_k^{g,t}, S_k^{g,t}, V_k^{g,t}$

$$\begin{aligned}
\bar{H}^g(h,l) &= \frac{1}{K_h}\sum_{k=1}^{K_h} H_{k,l}^{g,t_h} \text{ for } l = 1,\ldots,180 \\
\bar{S}^g(h,l) &= \frac{1}{K_h}\sum_{k=1}^{K_h} S_{k,l}^{g,t_h} \text{ for } h = 1,\ldots,256 \\
\bar{V}^g(h,l) &= \frac{1}{K_h}\sum_{k=1}^{K_h} V_{k,l}^{g,t_h} \text{ for } h = 1,\ldots,256
\end{aligned} \tag{5}$$

for $h = 0 \ldots 23$ and where $H_{k,l}^{g,t}, S_{k,l}^{g,t}, V_{k,l}^{g,t}$ are the normalized bins (as computed in Equation 2). Note that, again as before, the hourly time series for each histogram bin in Equation 5 is analogous to the keyword time series that (Naaman et al. 2012) computes and that we also compute in Equation 4.The variations in values of bins over 24 hours gives us another way to characterize each city's images. Importantly, our final results are not sensitive to the choice of scaling and normalization. After this step we have three sets of HSV histogram features for each hour from 2013 for each geographic location:

$$\begin{aligned}
\bar{H}^g(h) &= (\bar{H}^g(h,1),\ldots,\bar{H}^g(h,180))^T \\
\bar{S}^g(h) &= (\bar{S}^g(h,1),\ldots,\bar{S}^g(h,256))^T \\
\bar{V}^g(h) &= (\bar{V}^g(h,1),\ldots,\bar{V}^g(h,256))^T
\end{aligned} \tag{6}$$

for $h = 0 \ldots 23$. Therefore, for each hour, we have three sets of feature vectors: $\bar{H}^g(h) \in \mathbb{R}^{180}$, $\bar{S}^g(h) \in \mathbb{R}^{256}$, and $\bar{V}^g(h) \in \mathbb{R}^{256}$. To visualize how these features change over each hour, we can use a number of different dimensionality reduction techniques, such as Multi-dimensional Scaling (MDS) and Principle Component Analysis (PCA). In Figure 3 we show the scatter plot of the top two principle components of the Hue aggregated histograms (as computed in Equation 6) over each hour for 17 cities. We observe that the trajectory of the aggregate histogram follows a cyclic pattern and is not randomly distributed. Furthermore, the cyclic trajectory for each city seems to have a unique set of dynamics. The Saturation and Value aggregated histogram features in Equation 6 also reveal cyclic trajectories for each city.

We also consider measuring the relationship between between the color histograms and time using the method of pairwise differences to have a statistical summary of the diurnal pattern (a feature we later use in our linear models). Here we measure the correlation between the pairwise distances of the aggregated hue histograms (as computed in Equation 6) with the pairwise differences of the hour (this method is related to the distance correlation (Szkely, Rizzo, and Bakirov 2007)). We measure these correlations on the Hue histograms using the Pearson coefficient for each city and show the relationship in Table 1. As seen in this table, all cities have a positive correlation with changes in the hour. However, not all cities have the same correlation and some cities follow a stronger diurnal pattern than others (a perfect correlation would be 1.0).

Face Detection

Another set of features that we use is the presence and number of faces in images. We use the open source (Bradski 2000) face detection algorithm for finding the number of faces $f_k^{g,t}$ in each image (see Table 2 for performance of the face detector that we used). For each image $k = 1,\ldots,K$ (where K is the total number of images in our data set) at geographic location g with time stamp t, we apply OpenCV's face detector and extract the number of faces $f_k^{g,t}$ for each image. $f_k^{g,t} = 0$ indicates no faces in image k, $f_k^{g,t} = 1$ indicates one face, and so on. Given the number of faces in the image, we derive three additional features: face presence FP, single face SF, and multiple faces MF. We define face presence $FP_k^{g,t} \in 0, 1$ as a binary feature that is 1 when there is one or more faces present, and 0 otherwise. We also define single face $SF_k^{g,t} \in 0, 1$ as a binary feature that is 1 if there is exactly a one face present in the image, and 0 otherwise. The single face feature indicates that the picture might be a selfie or of a single person who is posing. Finally, we also define multiple faces $SF_k^{g,t} \in 0, 1$ as a binary feature that is 1 if there are 2 or more faces in the image and 0 otherwise. This feature indicates whether there are multiple people in the image, revealing that perhaps the the people in the image are engaged in a social activity.

As we did in Equation 4, we compute the hourly percentage of images for each of the feature sets that we extract. For example, for the face presence feature set, we compute:

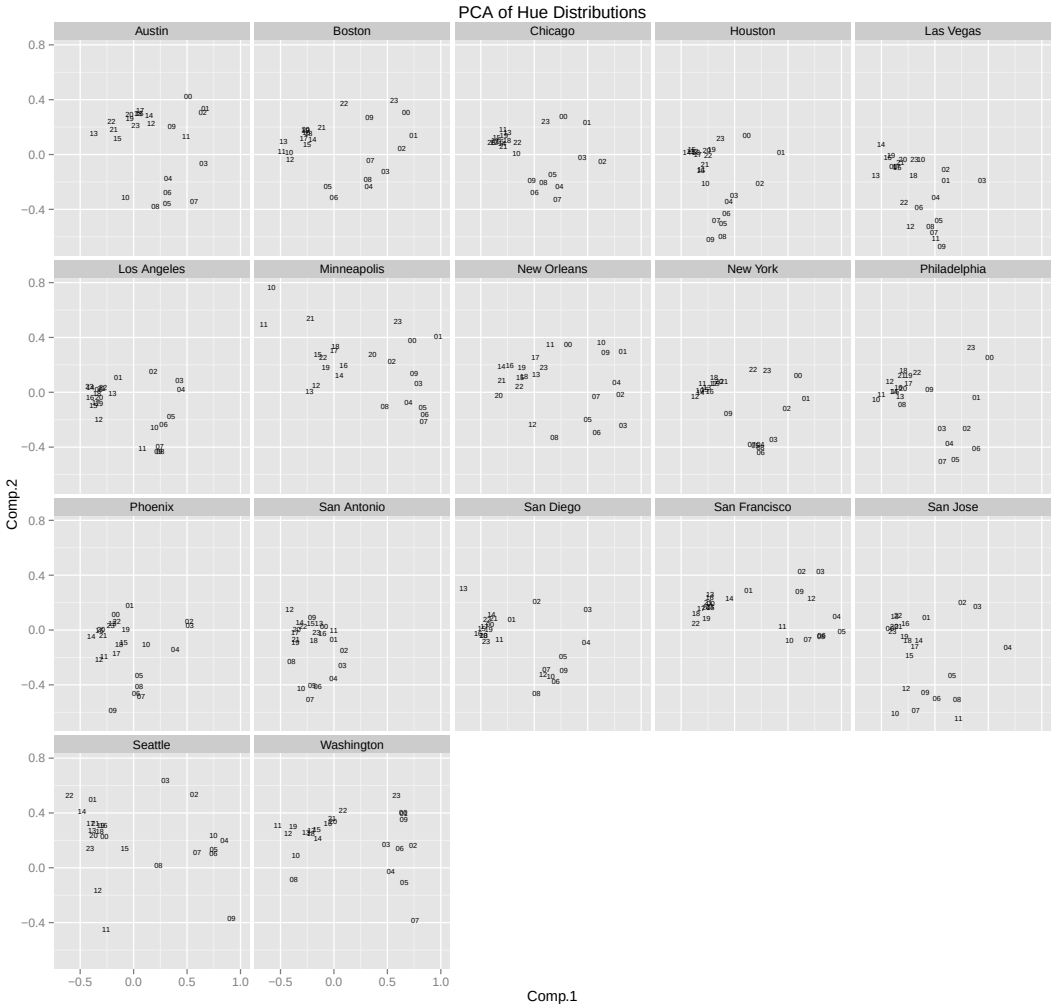

Figure 3: Scatter plot of the PCA of the Hue features computed as in Equation 6. Each point represents the top two PC's of the color histogram for a single hour (indicated by 0 - 23). The temporal variation of the Hue features is distinctly ordered, suggesting temporal changes in the aggregated hue histograms. A perfect circular trajectory suggests a cyclic change in color distributions throughout the day over the course of 24 hours. Deviations of this suggests that a city is less influenced by diurnal patterns.

$$\begin{aligned}
\text{FP24}_g(h) &= \frac{1}{K_h^g}\sum_{k=1}^{K_h^g} FP_k^{g,t_h} \\
\text{SF24}_g(h) &= \frac{1}{K_h^g}\sum_{k=1}^{K_h^g} SF_k^{g,t_h} \\
\text{MF24}_g(h) &= \frac{1}{K_h^g}\sum_{k=1}^{K_h^g} MF_k^{g,t_h}
\end{aligned} \tag{7}$$

for $h = 0, \ldots, 23$, where K_h^g are the total number of images from hour h in location g, and t_h are all time stamps corresponding to hour h. As was the case for the Hue bins discussed above, Equation 7 computes for each hour the percentage of images that either have at least one face (face present), exactly one face (single face), or more than one face (multiple faces). And again, as was the case with hue bins, $\text{FP24}_g(h)$, $\text{SF24}_g(h)$, and $\text{MF24}_g(h)$ represent the expected hourly rates for each geographic location for the face present, single face, or multiple faces features respectively. As was the case for hue bins, $\text{FP24}_g(h)$, $\text{SF24}_g(h)$, and $\text{MF24}_g(h)$ all follow diurnal patterns. We again use entropy to measure the variations of $\text{FP24}_g(h)$, $\text{SF24}_g(h)$, and $\text{MF24}_g(h)$ across different geographies.

Linear Regression on Features

After computing HSV and face features as explained above, we used them as predictors in a multivariable linear regression model for several socio-economic characteristics for each of the 60 cities in our dataset. These characteristics are median housing prices, median incomes, education levels

	city	hue.corr	city	hue.corr
1	albuquerque	0.28	memphis	0.50
2	anaheim	0.59	mesa	0.46
3	arlington	0.54	miami	0.53
4	atlanta	0.43	milwaukee	0.35
5	aurora	0.26	minneapolis	0.65
6	austin	0.47	nashville	0.49
7	bakersfield	0.28	new orleans	0.49
8	baltimore	0.50	new york	0.64
9	boston	0.60	oakland	0.60
10	charlotte	0.51	oklahoma city	0.47
11	chicago	0.58	omaha	0.51
12	cleveland	0.48	philadelphia	0.54
13	colorado springs	0.34	phoenix	0.56
14	columbus	0.45	portland	0.43
15	corpus christi	0.34	raleigh	0.44
16	dallas	0.55	riverside	0.38
17	denver	0.39	sacramento	0.45
18	detroit	0.45	san antonio	0.52
19	el paso	0.36	san diego	0.65
20	fort worth	0.60	san francisco	0.58
21	fresno	0.30	san jose	0.49
22	houston	0.66	santa ana	0.19
23	indianapolis	0.52	seattle	0.55
24	jacksonville	0.54	st. louis	0.35
25	kansas city	0.43	tampa	0.48
26	las vegas	0.44	tucson	0.32
27	lexington	0.30	tulsa	0.47
28	long beach	0.51	virginia beach	0.44
29	los angeles	0.68	washington	0.58
30	louisville	0.42	wichita	0.38

Table 1: The Pearson correlation between the changes of hue histograms with the changes in the hour. Most cities have a strong positive correlation.

(as measured by rate of people with Bachelors degrees), and financial well-being (a rank determined by Gallup). In addition to the features above, the model also incorporates several other features based on rates. For example, we count the specific instance of a feature (such as the number of black and white images in a city) and normalize this by the total number of images in the city. We also use the total number of users in each city, normalized by the total number of images from the respective city. To understand the usefulness of every individual feature as a predictor, we compute the Pearson correlation between each feature and each of the non-rank socio-economic characteristics (median housing prices, median incomes, and education levels). For the financial well-being measure we use the Spearman correlation since we only have relative ranks. Finally, using all these features in combination, we develop a model to predict each of the socio-economic target variables we considered. For all 4 models, we use the same features.

Results

The dataset used in our study is over seven million Twitter images shared in 60 U.S. cities during 2013. First, we extract particular features from these images (hue, saturation, and value histograms, and presence of faces). Next, we show that these features vary in systematic and distinct ways for hours of the days and for each city. Finally, we build a statistical model that uses the image features to predict selected socio-economic characteristics of the areas where the images were shared. The characteristics we used are average income, housing prices, educational levels, and well-being (from Gallup).

	No Faces	Single Face	Two or more faces
Sensitivity	96.26%	45.57%	30.48%
Specificity	52.83%	90.12%	99.49%
Balanced Accuracy	74.55%	67.84%	64.98%

Table 2: Statistical performance of the OpenCV "alt" classifier used in our study.

We start with an exploratory analysis of using over 7 million publicly available geotagged images posted to Twitter in 2013 from 60 cities in the United States. Along with photographs, there are also screenshots from user devices, and even images that originate elsewhere on the Internet, including memes, advertisements, and other graphics. Very few of these have any obvious connection with the time and place of posting. This lack of temporal and geographic characteristics in a significant proportion of Tweeted images makes it challenging to find a strong signal with any local significance.

Our search for a useful signal begins with basic visual features. We calculate HSV (hue, saturation, and value) histograms for all pixels of each image (see Methods). Note that these histograms do not capture the content of images (such as people, places, objects, or texts). Using HSV histograms as features, however, bypasses the limitations of bag-of-words features used for the analysis of text in Tweets. As we have already noted above, such features require language-specific lexical databases, while HSV histograms do not. HSV histograms also have the benefit that they are fast to compute and readily interpretable by people without technical training. Finally, we note that research suggest that humans have emotional responses to colors and that human behavior can be influenced by the colors in the immediate environment. This further suggests then that color patterns of social media images may contain information about people sharing these images, such as their mood, well-being and happiness levels.

We first compute histograms with 180 bins for the hue channel and 256 bins for the saturation and value channels. Next, we simplify the representation of each image by increasing the bin width of the histograms so that each of three HSV histograms have 12 bins. We then select the mode (i.e., the tallest histogram bin out of 12 possible bins) for each of the reduced histograms. These three numbers give us an approximation of the dominant color of each image.

Figure 4 shows the hourly percentage of images that have a particular dominant hue (from our 12-bin hue histogram)

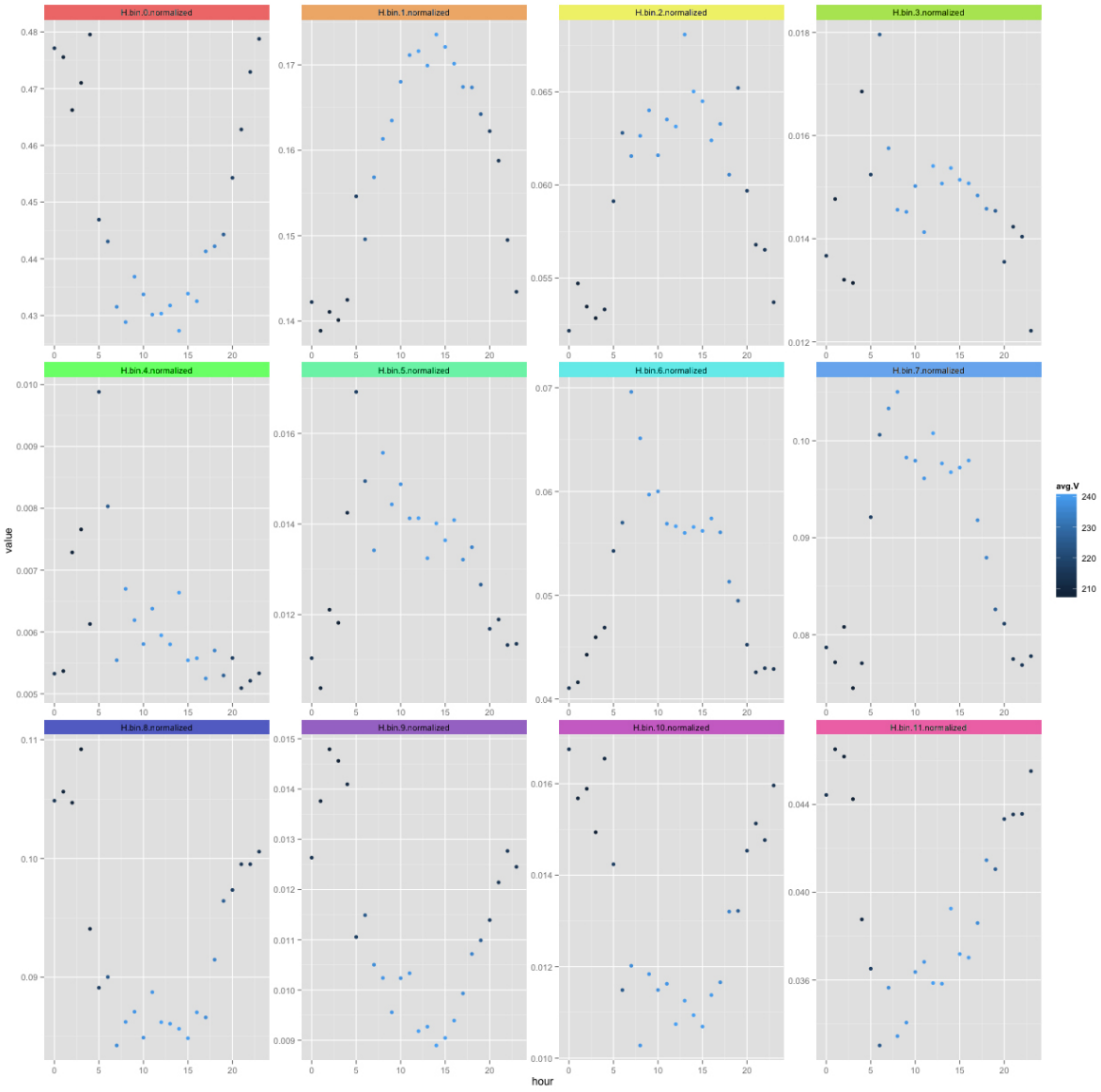

Figure 4: Distribution of images with specific dominant hue for each hour aggregated over 2013. The distribution of each dominant hue follows a diurnal pattern where orange, yellow, and blue colors peak during the afternoon hours whereas violet-based hues (bottom row) follow an opposite distribution. See main text for more details.

for New York City (a similar analysis for text of Tweets is shown in (Naaman et al. 2012)). In addition, the tone of each point (light to dark blue) corresponds to the average brightness (mode of value channel) of the images shared each hour. We see that the distribution of dominant hues and the average value of images follow a distinct diurnal pattern. Images that have hue bin 0 as the dominant color and low value are perceived as "black", and we see that in the middle of the day, the rate of these images decreases drastically. Images with hue bin 1 (corresponding to "yellow-orange") peak during the middle of the day. The remaining hue and value mode bins follow similar temporal patterns. Other cities also have diurnal patterns, with unique dynamics for every city (see Figure 1 for a comparison between San Diego and Philadelphia for example).

Note that this analysis has drastically reduced the visual content of each image to just three numbers (mode bins of 12 bin HSV histograms), and thus a great deal of information has been lost. To remedy this, we also analyze full hue histogram distributions in relation to time of day. To do this, we compare the pairwise differences between the color histograms for each hour with the pairwise differences of the respective hours. Put simply, this analysis answers the question whether, for a given city, a small (or large) change

in hour means a small (or large) change in color. Figure 5 shows the change in color by the change in hour for every pair of hours over the course of a day for a selection of cities in our data. We found that for every city, there is a relationship between the changes in image colors and the time of day. We also found that while particular patterns vary from city to city, the strength of the relationship is bigger for certain cities than for others.

In addition to color features, we also consider the presence of faces in images as another signal. We choose faces from a variety of possible high-level visual features for several reasons. First, face detection algorithms now achieve accurate results, and several well-performing open source algorithms are readily available (we use an OpenCV face detection algorithm). Therefore, if the presence of faces is a valuable signal, such analysis can be carried out by any researcher and not only teams at major software companies that have proprietary object and scene detection algorithms that require large infrastructures to run. Second, the face features we use, such as the number of faces detected in an image, are readily interpretable by non-experts.

Figure 5 shows the distribution of the presence of faces in all 7 million images aggregated over 24 hour period for all 60 U.S. cities in our dataset. For most of these cities, we see that during the working hours (8am to 6pm), there are few images that contain multiple faces. This number increases in the evening, suggesting that at that time people are more likely to get together with friends and family and share their experiences via posted photos. The presence of faces over the course of the day therefore reflects social engagement. The particular pattern of engagement is unique to each urban area. Therefore, we may expect that the presence of faces in images can tell us something about the social behavior of populations and may correlate with popular variables in social science.

Our analysis of HSV color features and faces distributions demonstrate that images shared on Twitter contain systematic patterns. Furthermore, these patterns vary depending on the geographic area. We will now show that the measurements of these patterns can be used to develop computational models for predicting specific socio-demographic characteristics of urban areas. In accordance with our commitment to interpretability of the analysis, we develop simple linear models using small combinations of easily interpretable features. Our target variables are housing prices (Zillow 2015), education levels from the US Census, median incomes also from the US Census, and financial well-being from the Gallup survey for each of our 60 cities.

	Housing	Incomes	Bachelor degree	Financial well-being
Single Face Rate	-0.556	-0.436	-0.684	0.276
Hue Pairwise Corr.	0.377	0.226	0.357	-0.412

Table 3: Single Variable Correlations

We begin by looking at possible correlations between single image features and single socio-economic variables. Table 3 shows correlation measures for two image features and four socio-economic variables for all 60 cities in our dataset (in all cases we report Pearson correlation, except for financial well being, since Gallup reports this measure as a rank). The image features are 1) the proportion of images with a single face and 2) hue pairwise correlation measures (recall that this is a measure of the diurnal change of the hue distribution of colors for images). Since we do not have a priori reasons to expect that simple image features would reflect any larger social patterns, the presence of any correlation is not only surprising but also highly relevant in determining socio-economic indicators. For example, the single face rate is negatively correlated with the first three indicators but has a positive correlation with financial well-being. This suggests that the cities where people share more photos with more than one face have more expensive housing, higher incomes and more people with bachelors degrees.

Looking at the hue pairwise feature, we note that correlation relationship with target variables is exactly the opposite. It is positively correlated with the first three variables and negatively correlated with financial well-being. This result suggests that the more people are satisfied with their finances, the less the hue distributions of images in their city follows a diurnal pattern. On the other hand, images shared in cities with more expensive housing, higher income and more people with bachelors degrees follow diurnal patterns more strongly. Note that the range of correlations we have for these individual image-related features are comparable to correlations measured using text-only features found in earlier studies of Twitter messages and socio-economic characteristics of cities (Mitchell et al. 2013). We also looked at measuring correlations between other socio-economic indicators, but the features that we consider in this paper had weakly correlations and we do not report them here for further consideration.

Having confirmed the presence of correlations between single image features and socio-economic variables, we will now see if we can better predict these characteristics using multiple image features together. The features used are the diurnal measures of HSV and the rate of images with no faces, one face, and more than one face (for details on each of these features, see the Methods). Our intention here is not to create the most accurate model, but rather to investigate if image features are reflective of socio-economic indicators. For this reason, we limit models to simple features and linear regression models (since Gallup reports ranks, we use the ranks of features in the linear regression model that we use). Table shows Adjusted R^2 values. The values range

	Housing	Incomes	Bachelor degree	Financial well-being
Adjusted R^2	0.6305	0.6822	0.5436	0.4584

Table 4: Results of the linear regression model that uses multiple image features to predict socio-economic characteristics

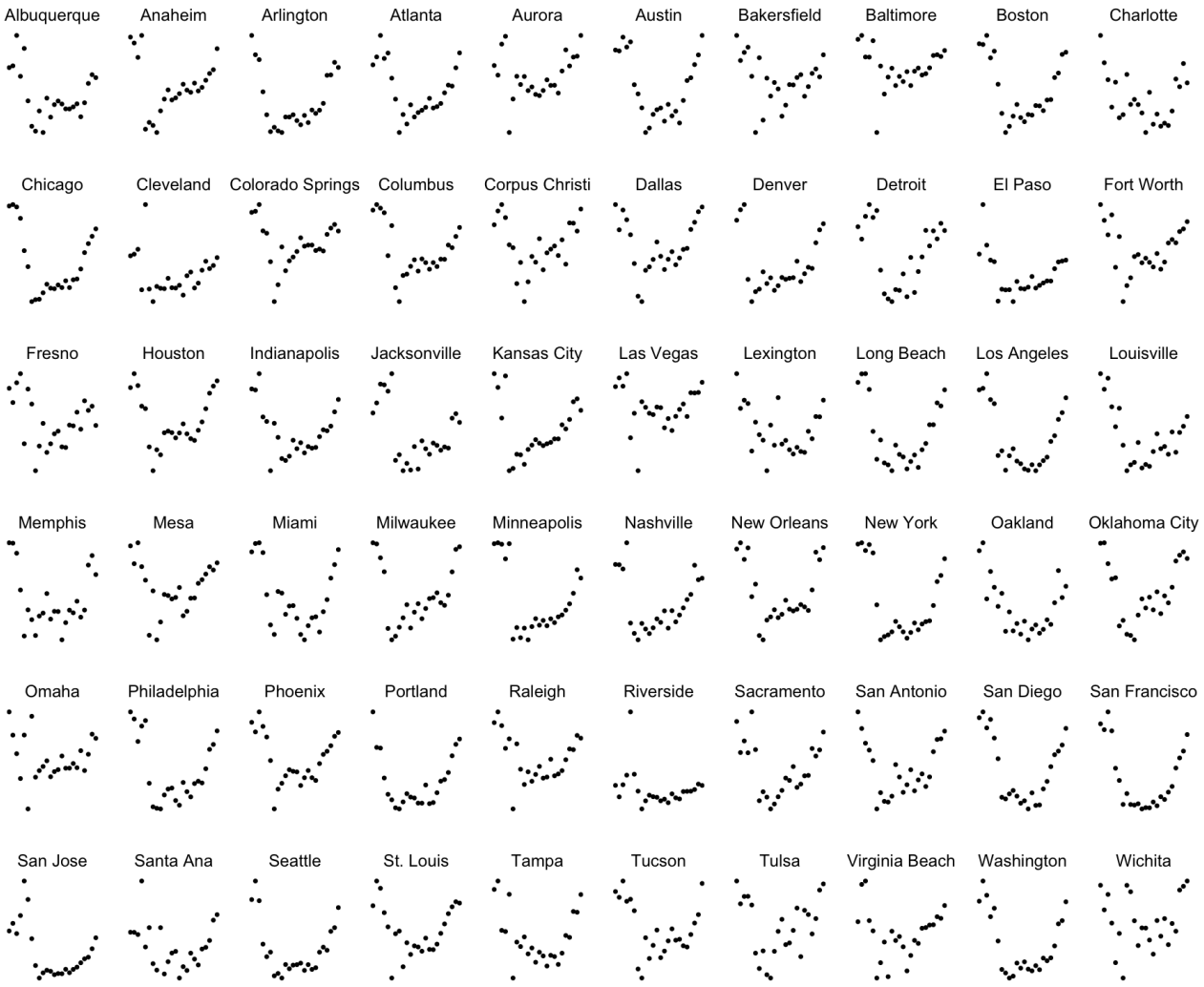

Figure 5: Distribution of the presence of faces in all 7 million images aggregated over 24 hour period for all 60 U.S. cities in our dataset. The presence of faces for each city follows a unique diurnal pattern.

from 0.45 (financial well-being) to 0.68 (median income). These results suggest that in addition to text features, image features from social media can also be used as predictors for various key socio-economic indicators.

Conclusions

The results in our paper puts forth that images shared in social media contain features that reflect socio-economic indicators. A natural next step is to augment existing computational methods that use text only with additional image features like the ones shown here. As discussed above, the large-scale activity of users on social media contains signals that can be used for predicting and understanding socio-econonmic variables. Like the web, social media networks were initially dominated by text. However, over the last five years the role of images and video is increasing in importance. New social networks have emerged that focus only on these media types (Pinterest, Instagram, Vine, Snapchat, and others). We believe that social media images reflect characteristics of the places and their populations lifestyles, feelings, and views. In many cases, they may contain information not available in text messages or tags, and this information can be a powerful predictor of many socio-economic characteristics of a place. Our work has begun to explore this previously ignored signal, and the initial results are encouraging.

Acknowledgments

TBD

References

Abdullah, S.; Murnane, E. L.; Costa, J. M.; and Choudhury, T. 2015. Collective smile: Measuring societal happiness from geolocated images. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 361–374. ACM.

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.

Bradski, G. 2000. Opencv. *Dr. Dobb's Journal of Software Tools*.

Center, P. R. 2015. Social networking fact sheet.

Dodds, P. S.; Harris, K. D.; Kloumann, I. M.; Bliss, C. A.; and Danforth, C. M. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one* 6(12):e26752.

Eichstaedt, J. C.; Schwartz, H. A.; Kern, M. L.; Park, G.; Labarthe, D. R.; Merchant, R. M.; Jha, S.; Agrawal, M.; Dziurzynski, L. A.; Sap, M.; et al. 2015. Psychological language on twitter predicts county-level heart disease mortality. *Psychological science* 26(2):159–169.

Einav, L., and Levin, J. 2014. Economics in the age of big data. *Science* 346(6210):1243089.

Frank, M. R.; Mitchell, L.; Dodds, P. S.; and Danforth, C. M. 2013. Happiness and the patterns of life: A study of geolocated tweets. *Scientific reports* 3.

Gayo-Avello, D. 2013. A meta-analysis of state-of-the-art electoral prediction from twitter data. *Social Science Computer Review* 0894439313493979.

Golder, S. A., and Macy, M. W. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878–1881.

Kahneman, D.; Krueger, A. B.; Schkade, D.; Schwarz, N.; and Stone, A. A. 2006. Would you be happier if you were richer? a focusing illusion. *science* 312(5782):1908–1910.

Lett, H. S.; Blumenthal, J. A.; Babyak, M. A.; Sherwood, A.; Strauman, T.; Robins, C.; and Newman, M. F. 2004. Depression as a risk factor for coronary artery disease: evidence, mechanisms, and treatment. *Psychosomatic medicine* 66(3):305–315.

Manovich, L. 2011. Trending: the promises and the challenges of big social data. *Debates in the digital humanities* 460–475.

Mitchell, L.; Frank, M. R.; Harris, K. D.; Dodds, P. S.; and Danforth, C. M. 2013. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one* 8(5):e64417.

Naaman, M.; Zhang, A. X.; Brody, S.; and Lotan, G. 2012. On the study of diurnal urban routines on twitter. In *ICWSM*.

Oswald, A. J., and Wu, S. 2010. Objective confirmation of subjective measures of human well-being: Evidence from the usa. *Science* 327(5965):576–579.

Rajaraman, A.; Ullman, J. D.; Ullman, J. D.; and Ullman, J. D. 2012. *Mining of massive datasets*, volume 77. Cambridge University Press Cambridge.

Schmidt, C. W. 2012. Trending now: using social media to predict and track disease outbreaks. *Environ Health Perspect* 120(1):30–33.

Szkely, G. J.; Rizzo, M. L.; and Bakirov, N. K. 2007. Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35(6):pp. 2769–2794.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

Zhang, A. X.; Noulas, A.; Scellato, S.; and Mascolo, C. 2013. Hoodsquare: Modeling and recommending neighborhoods in location-based social networks. In *Social Computing (SocialCom), 2013 International Conference on*, 69–74. IEEE.

Zillow. 2015. Zillow real estate research.