Predicting Social Trends from Non-photographic Images on Twitter

Mehrdad Yazdani
Qualcomm Institute,
California Institute for Telecommunication and Information Technology,
University of California - San Diego,
La Jolla, California 92037
Email: myazdani@ucsd.edu

Lev Manovich
Ph.D. Program in Computer Science,
The Graduate Center,
City University of New York,
New York, New York
Email: manovich.lev@gmail.com

Abstract*—Humanists use historical images as sources of information about social norms, behavior, fashion, and other details of particular cultures, places and periods. Dutch Golden Era paintings, works by French Impressionists, and 20th century street photography are just three examples of such images. Normally such visuals directly show objects of interests such as social scenes, city streets, or peoples dresses. But what if masses of images shared on social networks contain information about social trends even if these images do not directly represent objects of interest? This is the question we investigate in our study. In the last few years researchers have shown that aggregated characteristics of large volumes of social media are correlated with many socio-economic characteristics and can also predict a range of social trends. The examples include flu trends, success of movies, and measures of social well-being of populations. Nearly all such studies focus on text content, such as posts on Twitter and Facebook. In contrast, we focus on images. We investigate if features extracted from Tweeted images can predict a number of socio-economic characteristics. Our dataset is one million images shared on Twitter during one year in 20 different U.S. cities. We classify the content of these images using the state-of-the-art Convolutional Neural Network GoogLeNet and then select the largest category that we call "image-texts" - non-photographic images that are typically screen shots of websites or text-message conversations. We construct two features describing patterns in image-texts: aggregated sharing rate per year per city, and the sharing rate per hour over a 24-hour period aggregated over one year in each city. We find that these features are correlated with self-reported social well-being responses from Gallup surveys, and also median housing prices, incomes, and education levels. These results suggest that particular types of social media images can be used to predict social characteristics not readily detectable in images.

***Keywords*—Social Media, Twitter, Social Images, Spatio-temporal, Social Indicators, Art History**

I. Introduction

Many fields in the humanities and social sciences (such as history, art history, anthropology, and sociology), rely on visual documents to learn about the past. Visual media created in particular times and places is an important (and sometimes only) source of detailed information about social relations, habits, and details of people's lives in these places and periods. For example, the Dutch Golden Era paintings and prints from the 17th century show us the scenes of everyday life, the appearance of houses inside and outside, the costumes of people from different trades, and other details of that era. Similarly, paintings of French Impressionists created in the 1870s and 1880s inform us about public spaces, leisure activities and gender relations in the growing Paris metropolis. Of course, like the many images that are shared today on social networks (such as Instagram), historical artworks often idealized and aestheticized their subjects. Nevertheless, they remain invaluable sources of information about the past.

The invention of photography and cinema in the 19th century greatly increased the number of available visual documents and the range of their geographic coverage. The prominent examples of photo documentation include "How the Other Half Lives" (Jacob Riis, 1890), works produced by American photographers funded by The Farm Security Administration project in 1930s-1940s and street photography by Garry Winogrand and Lee Friedlander from the 1960s.

The emergence of social media networks in the middle of the 2000s represents a distinct new stage in this history. Today hundreds of millions of people share billions of images daily on Facebook, Instagram, Twitter and other social networks. At the same time, advances in automatic detection of faces, objects and types of scenes in photographs allow us to find particular types of content in massive amounts of user-generated images. Finally, new visualization tools can help us navigate large image collections and do this filtering in real time. All this makes user-generated visual media useful to humanists and social scientists.

When we use historical images as sources of information, we typically choose images that directly show what we are interested in - genre scenes, look of city streets, places of work and worship, and so on. We can certainly use social media images in the same way. Most user-shared photos have date and time stamps, and a significant proportion also has geospatial information. Therefore, we can download large numbers of photos using social networks' APIs, and then use object and scene detection algorithms and photos metadata to select images showing particular objects and scenes in particular locations and times.

Visual social media is characterized by a massive scale and high temporal and geospatial coverage. In many geographic areas, millions of images are shared daily. Therefore, we may also wonder if in addition to the content of shared images, we can use aggregated image statistics as the source of information about lives and feelings of the populations. The examples of

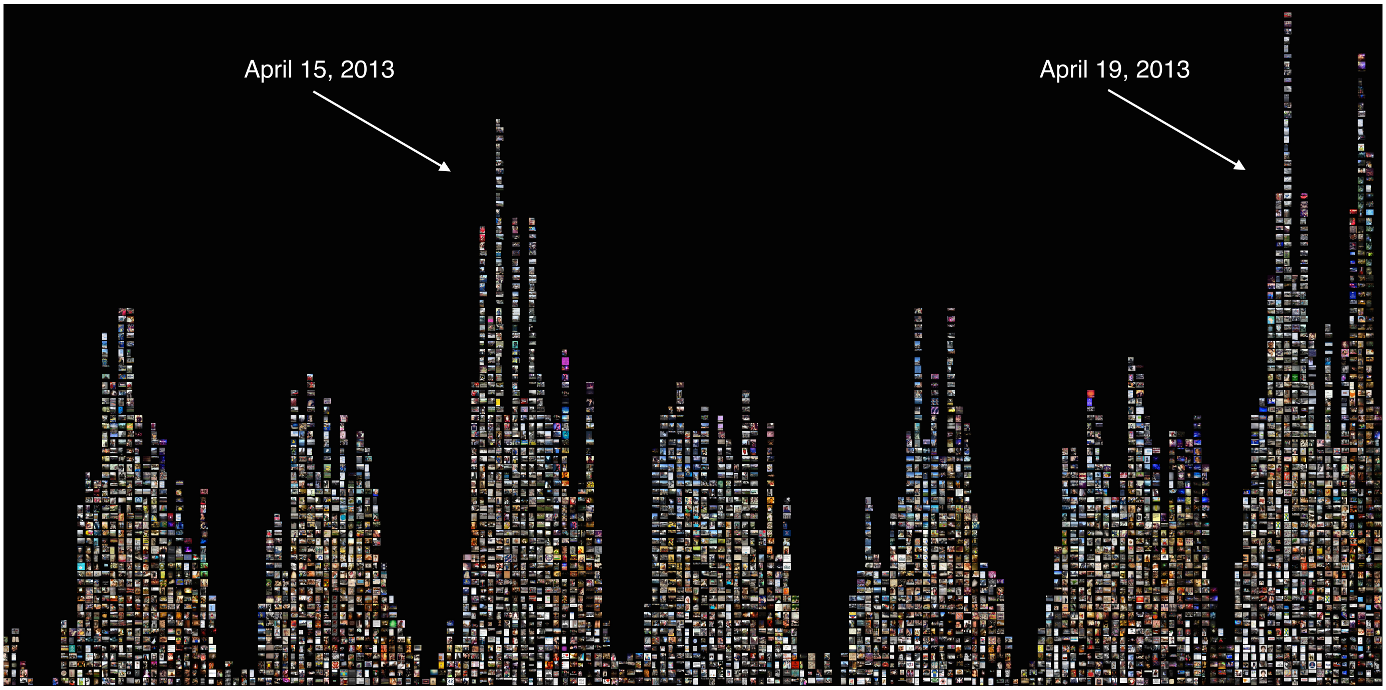

Fig. 1. Hourly volume of images shared on Twitter in Boston during the week of the Boston Marathon bombing. The bombing took place on April 15, 2013 and the perpetrators were captured on April 19, 2013. Each bin shows the volume of images shared on Twitter during a one hour period. Bins are filled in with random samples of images shared during the corresponding hours sorted by hue mode.

such statistics may include the volume of shared images over a 24-hour period, a week, or other temporal periods, the average number of photos shared per user, or the ratios of different types of content. If we use such statistics, we may also ask if the images which do not show concrete scenes, objects, faces or groups of people may be as informative of social characteristics as the images that do show them. These are the questions investigated in this paper.

II. Learning about socio-economic indicators from Twitter images

The popularity of online social media platforms such as Facebook and Twitter have created new unprecedented research opportunities for quantitatively studying social and cultural behaviors and networks. Progress has already led to important results in health care [1], economics [2], [3], digital humanities [4] and many other fields. One interesting and important area of this research is predicting social well-being in urban areas using social media. Social well-being measures are important for determining societies' overall well-being, as evidenced in behavioral economics [5] and psychology [6].

Normally to measure the well-being of a city a sample of people are asked to answer questions on a survey and aggregate statistics are then computed from their answers [7]. While such methods have been shown to provide a useful measure of a city's overall well-being [8], they are costly and have limitations in spatial and temporal resolution. The same applies to other socio-economic indicators which are often updated only yearly. However, the volume and velocity of social media provides the opportunity to indirectly measure social characteristics, people's opinions, and reactions to events. While the majority of researchers focused on using text media such as tweets, we have started to explore the use of social media images. For example, in Figure 1 we show examples of images shared on Twitter during the week of the Boston Marathon Bombing. Each "bin" corresponds to one hour and the peaks correspond to days of the week. The patterns in volumes of shared images reflect the reaction of people in Boston to the events. The day of the event and the day when the perpetrators were captured have highest volume of images. However, not all images are related to the events. Instead, as we have shown in our earlier project where we analyzed Instagram images shared in the center of Kiev during a week of the 2014 Ukranian Revolution, the images directly related to the revolution were only a portion of all shared images [9]. But the remaining always present portion also contains useful information. For example, researchers have shown that the volume of Instagram images has distinct temporal patterns in different global cities [10].

Can we use characteristics of shared images to predict socio-economic indicators? Previous studies have used sentiment analysis techniques on aggregated texts in Twitter posts. They found that the aggregated Twitter sentiment is correlated with specific social and behavior patterns in urban areas [11]. Other studies found that the temporal and spatial changes in Twitter sentiments also correspond to several socio-economic indicators [7], [12]. Still others have shown that the sentiments of tweets correspond to heart disease rates in urban areas [13].

While estimating the sentiment of a single post remains a challenging open problem, when millions of posts are aggregated, sentiment measures can be correlated with socio-economic indicators. This is especially remarkable considering the vast diversity of topics in social media. However, this research has several limitations. For example, most methods extract bag-of-words features from text in posts and use specific language databases to measure sentiment [14]. Such databases are available for some languages but not for others. Another issue is that social media is not limited to text. Increasingly, images and videos are becoming the dominant medium for users to create and share content [15]. Therefore, if we can complement existing natural language processing (NPL) techniques used on texts of tweets with computer vision analysis of sentiment or other relevant characteristics of images, this would provide an additional signal to improve prediction of socio-economic characteristics of a populated area. Unlike text, the analysis of images does not require specific linguistic databases, and therefore the same techniques can be used on images shared in any area.

A number of publications have reported on developing machine learning methods for visual sentiment analysis [16], [17], [18]. Most of these methods involve collecting a large number of images to train a classifier that uses the accompanying texts as a "target" for sentiment measure. That is, the image captions or tags are used to calculate the sentiment which in turn is used as a label. These methods use unsupervised machine learning techniques for building the classifiers, and the image features used are often uninterpretable (this is often referred to as the "black-box" machine learning).

In this study we adapt a different approach. We classify content of images and test if the distribution of one of the contents type is correlated with socio-economic indicators. The indicator values are from the same cities for which we collect images. This approach takes advantage of recent improvement in the precision of content analysis of images due to Deep Neural Networks. For example, the winner of the 2014 ImageNet challenge presented by a team of researchers from Google has error rates (6.8%) that are almost as low as human performance (5.1%). Google has released this trained deep convolutional neural network called GoogLeNet as open source using the Caffe framework [19].

The dataset used in this study is one million Tweeted images publicly shared along with geolocations in 20 different U.S. cities during 2013. We use GoogLeNet to classify the content of these images. While this classifier has 1,000 categories, here we only focus on one type of images that have been classified as "web site, website, internet site, site." We refer to these images as "image-texts." A sample of images classified in this way is shown in Figure 2. This category includes all images that are screen shots, text memes, texts on neutral backgrounds, or other non-photographic images. They do not directly show anything in the physical world. We compute the volume of image-texts aggregated for the whole year per city and also for each hour in a 24 hour cycle. We then compare these statistics with several important quantitative socio-economic indicators, such as median housing prices, incomes, education levels and social well-being. If we find any significant correlations, this means that the rate of sharing of these images is not random but related in some ways to the socio-economic conditions of the corresponding populated areas or (in the case of well-being) self-perceptions of the inhabitants of these areas.

Particular types of photographic content can directly reveal physical differences between places. For example, photographs can show buildings typical of poor and rich areas in a city. The numbers of such images may directly correlate to economic measures such as income. However, since such areas may look different from city to city, if we learn to correctly detect such images in one place this does not mean that we can use the same algorithm to detect these images in other places. Since people often share photos they have taken much earlier, or photos of locations they dream of visiting, this presents another problem. It would be ideal if instead we can identify a broad and easily detected category of images that can be correlated with some socio-economic characteristics of many areas. We find image-texts to be this category.

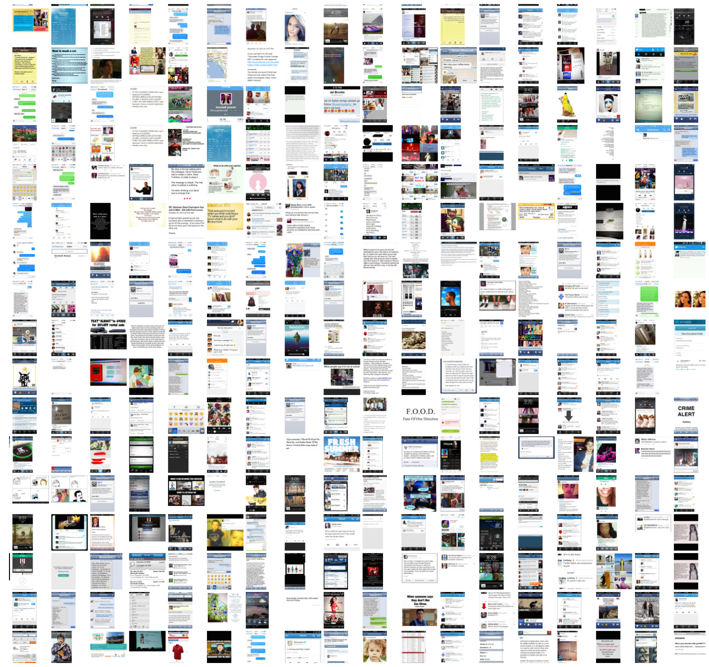

Fig. 2. A random sample of images that are classified as "web site, website, internet site, site" by the GoogLeNet convolutional neural network. We refer to such images as "image-texts." The category includes screen shots of text chats, other types of texts and other kinds of non-photographic images.

III. Data and Methods

A. Data

We were provided with access to publicly available historical tweets with images via the Twitter Data Grant we received in 2014. The grant made available to us all geotagged images shared worldwide during 2011-2014. In this study, we use the subset of publicly tweeted images with GPS coordinates that were shared during 2013 in the lower 48 U.S. states. In addition to the images and GPS coordinates, the dataset also contained other metadata including time stamps (dates and times when images were shared), and the optional texts (also known as "Tweet") that accompanies the images. All metadata made available to us was the same as what can be downloaded using the Twitter API, but we were not limited API download limits or historical time windows.

The total number of publicly available images with geotags from the lower 48 states shared in 2013 is 28 million. To obtain the subset of images Tweeted in each city, we use the Yahoo API to define bounding boxes following the technique in [20]. Table I shows the volume of images for the 60 cities that had the largest numbers of images shared in 2013. The total number of geotagged public images shared in these cities is 7.5 million. For our study, we use 20 top cities, listed in **bold**. We randomly sample 50,000 images from each of these cites. The resulting dataset contains 1 million images.

The next step is to detect content in this dataset and check for correlations between rates of particular content type and external target variables (i.e., socio-economic indicators). As target variables, we use median housing prices [21], education levels (specifically, the proportion of people with Bachelor degrees as reported by the US Census), median incomes (also from the US Census), and social well-being measures from Gallup. Median housing prices come from Zillow.com, the leading U.S. real estate site. Census data comes from 2013 American Community Survey (ACS) that surveys a sample of U.S. population. Gallop obtained well-being rates by also using population samples from each city. Each person was asked a number of questions including a direct question about their well-being. This is considered a "subjective" measure, but it has been shown to correlate with certain objective socio-economic measures of the corresponding areas [8]. We selected these target variables since they are often used in the social media research literature to test for correlations with characteristics of tweets and other types of shared messages [7]. Our goal is to test if the rates of particular types of images maybe be also correlated with these target variables.

City	Volume	City	Volume
New York	1034643	Jacksonville	79850
Los Angeles	810046	Seattle	78139
Houston	405051	Milwaukee	75941
Chicago	334422	Mesa	73567
Dallas	290407	Detroit	71079
Fort Worth	271916	Cleveland	71055
Washington	238254	New Orleans	69473
Philadelphia	229252	Tucson	58937
San Antonio	228038	Baltimore	56520
San Diego	227794	Sacramento	53649
San Francisco	192470	Raleigh	53624
Boston	186484	Wichita	52635
Phoenix	177377	Minneapolis	51944
Austin	167255	Tulsa	50996
Arlington	132146	Omaha	50814
Long Beach	122521	Oakland	50283
Las Vegas	119437	Louisville	50236
Columbus	111506	Memphis	49207
San Jose	109444	Fresno	44687
Tampa	109387	Riverside	44557
Nashville	102341	Virginia Beach	43278
Atlanta	98322	St. Louis	41098
Anaheim	96452	Albuquerque	40291
Denver	96151	Bakersfield	39582
Oklahoma City	94246	Lexington	39100
Charlotte	94024	Corpus Christi	34199
Kansas City	93991	El Paso	32547
Portland	93729	Colorado Springs	30502
Indianapolis	84863	Santa Ana	25750
Miami	83999	Aurora	22048

TABLE I. 60 U.S. CITIES SORTED BY NUMBER OF GEOLOCATED IMAGES PUBLICALY SHARED ON TWITTER IN 2013. THE TOP 20 CITIES USED IN OUR CITY ARE HIGHLIGHTED IN **BOLD**.

B. Classifying Images using GoogLeNet

In recent years, Deep Learning architectures including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) Recurrent Neural Networks have led to significant advances in computer vision and especially in the area of object recognition [22], [23]. In 2014, Google's research team used the GoogLeNet network [24] to win the ImageNet Large Scale Visual Recognition Challenge [25]. Their error rate was only 6.8% on the test dataset used in the ImageNet competition. As a point of comparison, human error rate is estimated to be 5.1% [26]. More recently, teams from Microsoft and Baidu have made claim of obtaining results even better than GoogLeNet using their CNN and LSTM Deep Learning architectures.

Google has released GoogLeNet under the Caffe Deep Learning framework developed by UC Berkley [19]. Caffe provides an open-source framework that enables easy implementations of Deep Learning architectures using GPUs with interfaces for Python and MATLAB. Caffe also hosts pre-trained models. We use the GoogLeNet model hosted by Caffe to classify the images in our dataset. Since GoogLeNet was trained on the ImageNet data set, GoogLeNet can classify images with 1,000 categories that were defined for the ImageNet competition. To prepare one million images for analysis and then classify them, we use a single Intel Xeon E5-2670 with 2X CPU's (10 cores per CPU), 128GB of RAM, and NVIDIA Quadro K6000 GPU. Most of the time (2-3 days) was used to copy the images. The actual analysis of one million images using GoogLeNet using our computer's GPU took approximately 2 hours, or 100 ms per image.

In our data set, the most popular label is "web site, website, internet site, site." These are non-photographic images that we will refer to as "image-texts." Figure 2 shows a random sample of images from this category. We can see that the images that GoogLeNet puts in this category are usually screen shots, text memes, and other images that are not of the "natural world." Figure 3 shows the rates for this category for the 20 cities in our data. As we can see, the rates range from less than 6% for San Francisco to over 10% for Columbus. Although other image categories can certainly also contain useful signals, for the rest of this study we decided to only use this one most frequent category of image-texts.

C. Measuring Diurnal Patterns of Tweeted Images

Previous research has shown diurnal patterns in tweets can predict some socio-economic indicators. Accordingly, we also check if our image-texts have these diurnal patterns (i.e., different rates throughout a 24 hour period) and if they correlate with any of our target variables.

Using the methodology of [27], we measure the diurnal rates of image-texts for each of the 20 cities in our dataset indicated in bold in Table I. We have $N = 1,000,000$ images containing 50,000 random samples from these 20 cities. For each image $k = 1, \ldots, N$ we have the location g and a time stamp t, and have classified each image with label l using GoogLeNet. To measure diurnal patterns, we compute the following series (similar to [27]) for each image k for location g and for $h = 0, \ldots, 23$:

$$\mathrm{X24}_g(h) = \frac{1}{K_h^g} \sum_{k=1}^{K_h^g} I(l_k^{g,t_h} = l^*) \tag{1}$$

K_h^g are the total number of images for hour h in location g. $I(\cdot)$ is the indicator function that the label of the k-th image is $l^* =$ "web site, website, internet site, site", and t_h are all time stamps corresponding to hour h. In other words, Equation

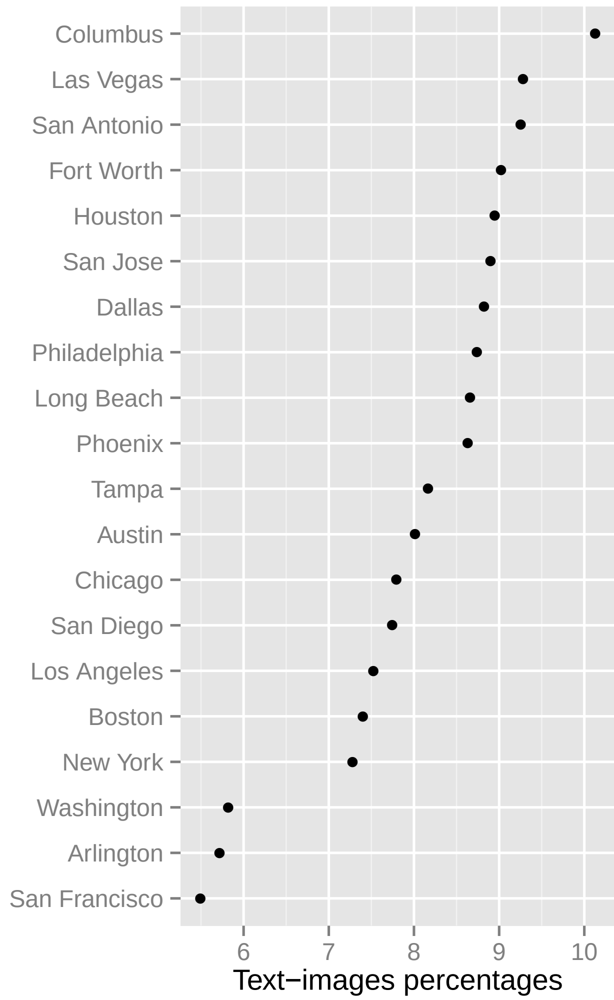

Fig. 3. The proportion of images in each city classified as "web site, website, internet site, site."

1 computes for each hour the percentage of images that are classified as l^*. This time series is analogous to the time series that [27] computed for keywords in the text of tweets. Similar to the interpretation that [27] give for the series in Equation 1, $\mathrm{X24}_g(h)$ represents the expected hourly values for each geographic location for the images that are classified as l^*.

To measure the variations of $\mathrm{X24}_g(h)$ across geographies, we compute the entropy for this series for each location g as was done in [27] for keywords. This is done by normalizing $\mathrm{X24}_g(h)$ for each location g to be a proper probability $p_g(h) = \mathrm{X24}_g(h)/\Sigma_h \mathrm{X24}_g(h)$ and evaluating the following equisation:

$$e_g = -\Sigma_{h=0}^{23} p_g(h) log(p_g(h)). \tag{2}$$

Entropy measures the amount of "spread" or "dispersion" in the series $\mathrm{X24}_g(h)$. Its value is maximum when the time series is perfectly flat, i.e., the rate of image-texts for all hours in a 24-hour cycle is the same. In contrast, Entropy value is close to 0 when the rate varies very strongly between the hours.

IV. Results and Discussion

We have classified one million images from 20 U.S. cities using the GoogLeNet Convolutional Neural Network. The most frequent category is "web site, website, internet site, site." We call these images "image-texts." Figure 2 shows a sample of these images. Figure 3 shows the proportion of this category for all 20 cities among all other images. This proportion varies from about 5% to over 10%. As Figure 2 shows, image-texts are memes, screenshots, and other images that are not directly representative of the real world. However, note that many of them are screenshots of text message conversations on smart phones. So while they do not show real life social interactions or natural environments, they are records of new forms of sociality enabled by networks and mobile phones.

We can see from Figure 3 that the rates of image-texts are different for each city. Furthermore, as Figures 4 and 5 demonstrate, each city also has a unique diurnal pattern of such images. Therefore, both characteristics can be used as features. The first feature is the overall rate of image-texts per city. The second feature is the entropy of the diurnal distribution of the image-text rates for 24-hour cycle per city.

To see if these two features have some connection to the socio-economic indicators, we calculate Pearson correlations between the values of the features and the indicators. Tables II and III show the correlation values. The absolute values of correlations range from 0.47 to 0.64. The values are significant with $p < 0.01$, except for income which has $p < 0.05$. The correlation with "objective" measures (i.e., housing prices, education levels and incomes) are negative, whereas the correlation with "subjective" measure of "social well-being" as reported by Gallup are positive.

These negative correlations suggest that people in cities that are more affluent as measured by objective measures such as housing prices share text-images less frequently. In contrast, people in less affluent cities share text-images more often.

Note that while median housing prices, education levels, and incomes also all have positive correlations between them, there is no significant correlation between these variables and social well-being variable. In other words, the features that we have extracted from image-texts shared on Twitter are far more predictive of social well-being than housing prices, incomes, or education levels reported by the Census. The advantage of using social media is that we can collect new images and update the image features as often as we like. Consequently, we can also update our predictions at any time. In contrast, large scale censuses have significantly lower temporal resolutions since they are conducted much less often. And in the case of surveys that are conducted often, the use very small samples so limits our statistical power.

Housing prices, incomes, and education levels are all examples of "objective" measures because they use single agreed upon scales. For example, having a Bachelor's degree has the same meaning everywhere in the United States. Surveys, on the other hand, are "subjective" measures since they use individual self-reported feelings. However, a survey is still a good method for understanding how people feel about a particular issue since the question is directly posed to individuals and not inferred from some other data.

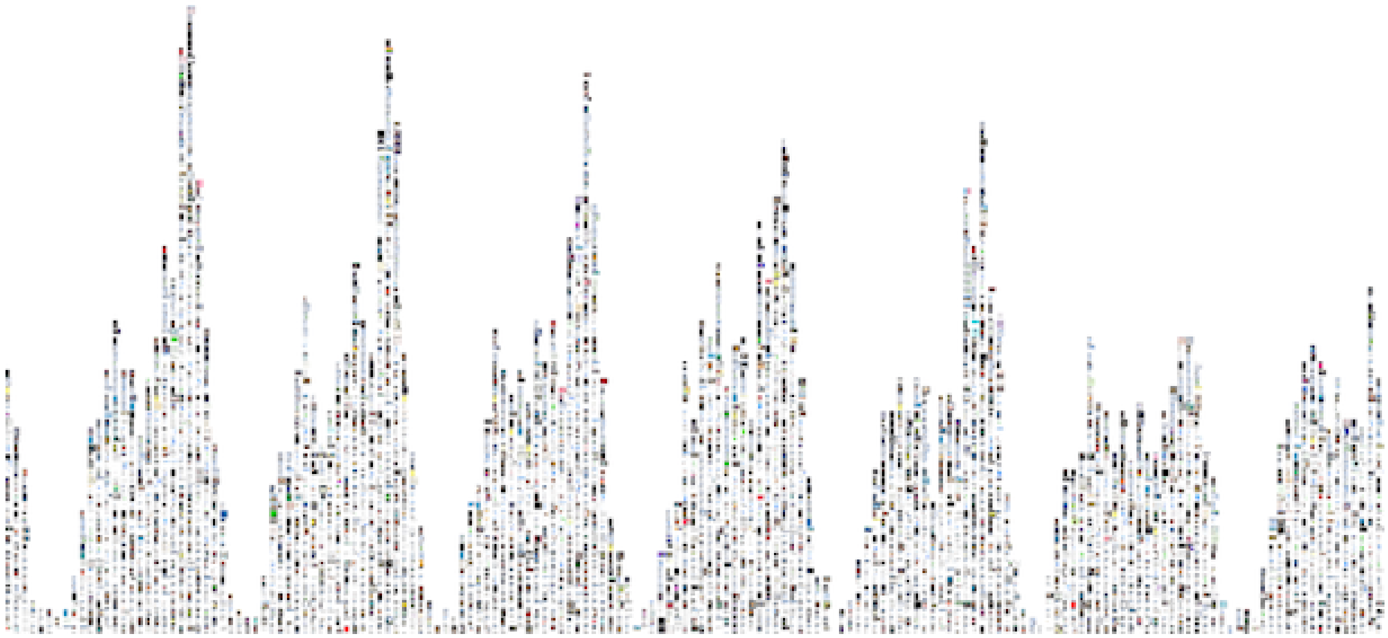

Fig. 4. Weekly and hourly rates of Twitter image-texts from 2013 for Fort Worth, Texas. Each column of images corresponds to one hour for a particular day of the week, starting with Sunday midnight local time. The rates of image-texts has strong diurnal patterns. Notice that the pattern for Sunday to Thursday nights is different from the pattern for Friday and Saturday nights.

Fig. 5. Weekly and hourly rates of Twitter image-texts from 2013 for New York City. Compare with Figure 4 that shows the pattern for Fort Worth, Texas. While the diurnal patterns in New York city are similar to Texas, the nightly peaks are smaller than the ones in Fort Worth.

In the case of social well-being, our data suggest that cities that are more happy with their social lives tend to also share text-images more frequently. We think that the reason for this is due to the fact that the greatest number of text-images are screenshots of text-message conversations. These text-message conversations capture a social activity that users are engaged in. In other words, cities that are more actively engaged in text-message conversations (and also share such conversations on social media), tend to report having better social lives.

Indicator	Correlation	P-value
Median Housing Price	-0.5638	0.007735
Rate of Bachelor's Degree	-0.6413	0.001623
Average Income	-0.4772	0.01805
Social well-being	0.56100	0.001623

TABLE II. PEARSON CORRELATIONS BETWEEN THE PROPORTION OF IMAGES CLASSIFIED AS IMAGE-TEXTS AND FOUR SOCIO-ECONOMIC VARIABLES (FIGURE 2).

V. CONCLUSION

Historical and contemporary images are an important source of information about the details of social life: some examples include built environments, tastes, everyday behavior, communication styles, foods, and fashions. Historians and social scientists in many fields use these images in their research. While social media images can be also used in the same way, their scale and spatiotemporal resolution also provides us with additional signal: the statistics of aggregated images shared in different areas. Such statistics include volume of all shared images and volumes of particular kinds of images shared in particular time periods. Our paper investigates the usefulness of these signals for prediction of key socio-economic indicators.

Using an open source convolutional neural network, we

Indicator	Correlation	P-value
Median Housing Price	-0.5332	0.007735
Rate of Bachelor's Degree	-0.62451	0.001623
Average Income	-0.4709	0.01805
Social well-being	0.5381	0.001623

TABLE III. PEARSON CORRELATIONS BETWEEN THE ENTROPY MEASURES COMPUTED FROM THE SERIES IN EQUATIONS 1 AND 2 AND FOUR SOCIO-ECONOMIC INDICATORS.

classified the content of one million images shared on Twitter in 20 cities in the United States during 2013. Among the 1,000 content categories used for classification, the content most frequently classified in our data is "image-texts" - screen shots, text of chats, and other non-photographic images. We constructed two features of these images: overall share rate per city and the entropy of sharing rates in a 24-hour cycle per city. We found that these two features have significant correlations with four socio-economic indicators we considered: average housing prices, income, education level, and self-reported measures of social well-being. The correlations with first three indicators are negative, whereas the correlation with social well-being measure is positive.

In our data, it appears that the most frequent type of image-text category are screen shots of text conversations. This suggests that populations that report having higher rates of social well-being also, in the aggregate, share their digital social experiences (that is, text message conversations) over Twitter at much higher rates.

For our future work, we plan to extract content-based features from images shared in more cities and also other countries. We also will look at other content categories. Additionally, we also want to interview some Twitter users who share image-texts to better understand their reasons for sharing these images.

Our analysis relies on statistical aggregation: using single aggregated socio-economic measures for a whole city (i.e., average income, average housing price, and average education level), and single aggregated statistics computed from hundreds of thousands of images shared in that city. One obvious limitation of this approach is that we don't consider the range of socio-economic groups in a city, and their different feelings and behaviors. Most large American cities have high level of economic inequality. So while the statistical approach can predict the "average" social well-being of a city, it does not tell us about the range of feelings across many groups and classes living in this city. This is an important limitation of the present study. While it can be overcome if we have well-being measures on the neighborhood level, we also need to take into account the fact that people in many less affluent areas may not share enough images for us to compute reliable statistics.

If we compare the methods and results of our study with the traditional use of historical images to learn about social life and its details, we find that the approach we investigated has both strengths and weaknesses. The strength is the ability to learn about social well-being (and possibly other social characteristics) by using images that do not directly show particular physical objects, scenes, people or situations. However, if we do want to learn about details of people's lives as opposed to overall feelings of well-being, we would still need to look at images that show these physical details. So rather than replacing the traditional use of concrete images, our new method should be seen as complementary.

ACKNOWLEDGMENTS

We are grateful to Twitter for awarding us their Data Grant and preparing for us the dataset of all geo-tagged images shared publicly on Twitter worldwide between 2011 and 2014. We also thank Cherie Huang, Damon Crockett, Agustin Indaco, John Graham, and Joseph Keefe for their assistance. Finally, we also thank the Qualcomm Institute at Calit2 for support of this research.

REFERENCES

[1] C. W. Schmidt, "Trending now: using social media to predict and track disease outbreaks," *Environ Health Perspect*, vol. 120, no. 1, pp. 30–33, 2012.

[2] L. Einav and J. Levin, "Economics in the age of big data," *Science*, vol. 346, no. 6210, p. 1243089, 2014.

[3] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[4] L. Manovich, "Trending: the promises and the challenges of big social data," *Debates in the Digital Humanities*, pp. 460–475, 2011.

[5] D. Kahneman, A. B. Krueger, D. Schkade, N. Schwarz, and A. A. Stone, "Would you be happier if you were richer? a focusing illusion," *science*, vol. 312, no. 5782, pp. 1908–1910, 2006.

[6] H. S. Lett, J. A. Blumenthal, M. A. Babyak, A. Sherwood, T. Strauman, C. Robins, and M. F. Newman, "Depression as a risk factor for coronary artery disease: evidence, mechanisms, and treatment," *Psychosomatic medicine*, vol. 66, no. 3, pp. 305–315, 2004.

[7] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth, "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place," 2013.

[8] A. J. Oswald and S. Wu, "Objective confirmation of subjective measures of human well-being: Evidence from the usa," *Science*, vol. 327, no. 5965, pp. 576–579, 2010.

[9] L. Manovich, A. Tifentale, M. Yazdani, and J. Chow, "The exceptional and the everyday: 144 hours in kiev," in *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 2014, pp. 72–79.

[10] N. Hochman and L. Manovich, "Zooming into an instagram city: Reading the local through social media," *First Monday*, vol. 18, no. 7, 2013.

[11] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, no. 6051, pp. 1878–1881, 2011.

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

[13] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap *et al.*, "Psychological language on twitter predicts county-level heart disease mortality," *Psychological science*, vol. 26, no. 2, pp. 159–169, 2015.

[14] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[15] P. R. Center. (2014) Social networking fact sheet. [Online]. Available: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

[16] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[17] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 223–232.

[18] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li, "Visual sentiment prediction with deep convolutional neural networks," *arXiv preprint arXiv:1411.5731*, 2014.

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

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

[21] Zillow. (2015) Zillow real estate research. [Online]. Available: http://www.zillow.com/research/data/

[22] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 1918–1921.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv preprint arXiv:1409.4842*, 2014.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, pp. 1–42, 2014.

[27] M. Naaman, A. X. Zhang, S. Brody, and G. Lotan, "On the study of diurnal urban routines on twitter." in *ICWSM*, 2012.